J. Tiago de Oliveira

STATISTICAL ANALYSIS OF EXTREMES

PENDOR

STATISTICAL ANALYSIS OF EXTREMES

Collection /
Colecção: MATHESIS

Scientific Coordination /
Coordenação Científica: Prof. Doutor Carlos Braumann

From J. Tiago de Oliveira in Pendor /
De J. Tiago de Oliveira in Pendor

1. Collected Works – Obras, Vol. I
   Books on Mathematics / Livros de Matemática

2. Collected Works – Obras, Vol. II
   Historical and Philosopical, Didactic and Expository and Miscellaneous Papers /
   Textos Históricos e Filosóficos, Didácticos e Expositórios e Miscelânea.

J. Tiago de Oliveira

STATISTICAL ANALYSIS OF EXTREMES

PENDOR

Título: *Statistical Analysis of Extremes*

Autor: J. Tiago de Oliveira

© Pendor 1997

Editor: Ed. Pendor Ap. 497
7005 Évora Codex

Colecção: Mathesis, 1

Execução Gráfica: Colibri, Artes Gráficas

Depósito legal: 111 740/97

ISBN 972-8163-13-4

Local e Data: Évora, 23-6-1997

Acknowledgements / Patrocínios:
Universidade de Évora
Departamento do Ensino Superior do Ministério da Educação
Fundação Calouste Gulbenkian
Fundação Mário Soares
JNICT
Sociedade Portuguesa de Estatística
Sociedade Portuguesa de Matemática

TABLE OF CONTENTS

Pág.

Preface i

Overview – General Overview 1
References 10

PART I – Probabilistic patterns of univariate statistical extremes
Chapter I – Introduction 11
Chapter II – Asymptotics of univariate extremes 20
Chapter III – Attraction structure and speed of convergence 52
Annex I – On the "Duality" between extremes and sums 69
Annex II – On the Khintchine's convergence of types theorem 71
Annex III – On the δ-method 73
Exercises – Part I 76
References – Part I 83

PART II – Statistics for univariate extremes
Chapter IV – A quick exploration of extreme data 87
Chapter V – Statistical analysis of Gumbel model 98
Chapter VI – Statistics for Fréchet distribution 124
Chapter VII – Statistical decision for Weibull distribution (for minima) 133
Chapter VIII– Analytic statistical choice for univariate extremes..... 161
Chapter IX – Statistical analysis with partial information 179
Annex IV – On Extreme values and tail estimation 209
Exercises – Part II 225
References – Part II 234

PART III – Multivariate extremes
Chapter X – Bivariate extremes; probabilistic results 245
Chapter XI – Bivariate extremes; models and statistical decision 288
Chapter XII – Multivariate extremes 323

Annex V – On the quadrants method 333
Exercises – Part III 335
References – Part III 341

PART IV – Stochastic sequences and processes of extremes
Chapter XIII– Extremes in random sequences and stochastic processes 345
Chapter XIV– Extremal sequences and processes; basics and statistics 357
Chapter XV – Other extremes processes; basics and statistics 380
Annex VI – On the choice of models for sequences of extremes 406
Exercises – Part IV 408
References – Part IV 411

PART V – Complements
Chapter XVI – GEA, *G*eneral *E*xtremes *A*nalysis, a guide in natural language 417
Chapter XVII–Some case studies 421
Exercises – Part V..... 429
References – Part V..... 430



José Tiago da Fonseca Oliveira, 22/12/1928 – 23/6/1992

PREFACE

The Statistical Extremes Theory, after a slow development for half a century, has been developing very quickly in the last ten years; it may be said to be an exploding area.

Although the probabilistic aspects as well as the stochastic processes of extremes have recently been dealt with, statistical decision methods have not actually been described and integrated in a book, even though the area has a large and steadily increasing number of papers. This was done thirty years ago, according to the current state of knowledge at that time, by Gumbel. This is an analogous project: to integrate, in one text, the current state of knowledge of essential statistical methodology for univariate and multivariate extremes and some stochastic processes, based on the fundamental probabilistic results.

We had to choose between including, up to the time of going to press, everything that can be considered important from the statistical decision point of view — the book thus being somewhat heterogeneous and unbalanced, with differential developments — and fixing a basic standard, with homogeneity, but not giving all the fundamental (but differential) knowledge at the time of going to press. We have chosen the former course — with its inevitable heterogeneity — to put the essential tools into the hands of professional people, and also to raise, for pure and applied statisticians involved in research, unsolved questions with, perhaps, hints for the analysis of some unsolved questions.

The comparison with Gumbel's (1958) book shows how far we have advanced and how many unsolved questions still remain.

Some leading ideas helped to write this text. We assumed an empiricist-classical point of view in Statistics, and so Bayesian results for extremes have not been dealt with — they are scarce!, apart from Reliability Theory — although we must stress that they need to be developed. Also p-values for decision, very controversial, have not been used but the elements for their application are clear and the adaptation is easy.

We considered as important questions: the choice of models, the evaluation of quantiles and the probabilities of over- and underpassing thresholds (or quantiles) for which the tools are evidently the already classical estimation and testing theory for parameters, moments, etc ... We have stressed, when possible, the efficiency behaviour of quantile estimators, the trilema choice of univariate asymptotic models in the univariate case and practical approximations, the tolerance intervals and prediction results, the choice of bivariate models, the difficulties of statistical decision for stochastic processes in continuous time when discretely sampled, etc... These theoretical questions seem to be the most important ones regarding applications.

The book contains exercises for Parts I, II, III, IV and V as well as references to all parts. To a certain extent, we may conjecture that the joint set of references covers 2/3 to 3/4 of the essential papers on Statistical Extremes Theory; but this is a personal opinion. Exercises are also an integral part of the book either by containing more information (even unpublished results) that can be used afterwards or by an attempt to clarify some results in the text.

In the short Case Studies Chapter, we have, in some cases, used a partial sample of about 2/3 of the sample size and compared its results with those of the total sample; when possible, this is a confirmatory procedure that is to be recommended.

Parts of this book have been piloted and written in graduate courses, short courses, seminars series and research stays in the Technion (Haifa), Universidade de Lisboa, Universidad de Madrid, University of North Carolina at Chapel Hill, Universidade de São Paulo at Ribeirão Preto, University of Surrey, Imperial College of Science and Technology (London), Université de Paris, Université Pierre et Marie Curie (Paris — VI), Iowa State University of Science and Technology (Ames), Bern Universität, University of California at Santa Barbara and Université Libre de Bruxelles. I would like to thank the colleagues and students that attended and "suffered" the lectures for their comments, remarks, as well as the kind friendship.

I would also like to thank the many colleagues with whom I have talked and had clarifying discussions and who pressed me to write this book.

I would also like to thank Mr. Mario Silva and Ms. Fátima Correia for typing the first draft and the final text, Mr. Vitor Teodoro for the drawings, Prof. Manuela Neves and a group of students who helped in the numerical calculations and Mr. Paul Covill who made the final revision of the text improving my English writing.

The last, but certainly not the least thanks, are due to the Gulbenkian Foundation — through Professors Ferrer Correia and Pina Martins — for the support for the final revision and typing of the book.

Many thanks to all !

OVERVIEW

GENERAL OVERVIEW

Preliminaries

The purpose of this general overview is to introduce informally and to recall basic notions as well as to fix notations to be used throughout the book. It will be assumed that readers know the "Introduction to the Theory of Statistics" by A. M. Mood, F. Graybill and D. Boas, McGraw-Hill (1950), or "Statistical Inference" by Silvey, Chapman and Hall (1975); knowledge of the already classic "Bibles of Statistics" such as "Mathematical Methods of Statistics" by H. Cramér, Princeton University Press (1946), "Probability Theory and Mathematical Statistics" by Marek Fisz, Wiley (3rd ed., 1963), or, even more, "The Advanced Theory of Statistics" (3 vol.) by M.G. Kendall and A. Stuart, Griffin (1958, 1961, 1966), would be much better. But we will try to stick to the assumed level of knowledge; the general results that go beyond the basic books will be given a reference in the text; results of the *Statistical Extremes Theory* will be proved or referred to in the text and, sometimes, given as exercises. We will not separate as problems the unsolved questions that are scattered throughout the book: '*a* not *the* solution', 'an open question', 'empirical evidence', etc., are some ways of describing them.

A review of basic notions and notations

Let X be a random variable; we will denote by $F_X(x) = \text{Prob}\{X \le x\}$, or simply by $F(x)$ when clear, its *distribution function*; we will also define the *survival function* $S_X(x) = \text{Prob}\{X > x\}$, denoted also by $S(x)$ when possible. Recall that $F(x) + S(x) = 1$.

In almost all cases we will suppose F (or S) to be absolutely continuous, i.e., such that there exists a *probability density* (or *density* for short) $f(t)$ such that $f(t) \ge 0$ and $\int_{-\infty}^{+\infty} f(t)\, dt = 1$ with

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \qquad \text{and} \qquad S(x) = \int_{x}^{+\infty} f(t)\, dt .$$

Recall that, always, $F(x)$ is non-decreasing and right-continuous, $F(-\infty) = 0$, $F(+\infty) = 1$, and consequently $S(x)$ is non-increasing and also right-continuous, $S(-\infty) = 1$ and $S(+\infty) = 0$. Note that we are excluding improper random variables, for which $P\{X = -\infty\} > 0$ or $P\{X = +\infty\} > 0$. The degenerated random variable X at a, i.e., taking the value a with probability 1, has $F(x) = H(x - a)$ and $S(x) = 1 - H(x - a)$, where $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$ is the Heaviside (unit jump) function at 0.

In some cases the random variable X may have bounds. If there exists a number $\bar{w}$, the least upper bound or *right-end point*, such that $F(\bar{w}) = 1$, and $F(x) < 1$ for $x < \bar{w}$, we know that $X \leq \bar{w}$ with probability one; analogously there may be a greatest lower bound or *left-end point* $\underline{w}$, such that $X \geq \underline{w}$ with probability one. $\bar{w}$ and $\underline{w}$ could be defined analogously in terms of $S(x)$.

A number x such that $F(x) = p$ $(0 < p < 1)$, when it exists, is the *quantile* χ_p of probability p or *p-quantile*. Usually, the p-quantiles are uniquely defined, although they can form an interval if the distribution and survival functions are constant over this interval or equivalently if the density is zero in the interval or also do not exist with this definition if $F(x)$ or $S(x)$ have jumps; in fact, we will be interested, almost always, in quantiles for $F(x)$ or $S(x)$ continuous. More formally let us introduce the notion of *quantile function* $Q(v) = F^{-1}(v) = \inf \{x \mid F(x) \geq v\}$. $Q(v)$ is continuous *iff* $F(x)$ is. Also we have $(F(Q(v)^-) \leq v \leq F(Q(v))$ with $F(Q(v)) = v$ if $F(x)$ is continuous at $Q(v)$. Then we define uniquely $\chi_p = Q(p)$. Sometimes, the quantiles for $p \leq 0.01$ or $p \geq 0.99$, the small or large probability quantiles, are incorrectly called "extreme" quantiles. In practical action we will be primarily interested in those so called "extreme" quantiles. The meaning of the quantiles is straightforward; for example : $\chi_{1/2} = \tilde{\mu}$, the *median*, is such that Prob $\{X \leq \tilde{\mu}\}$ = Prob $\{X > \tilde{\mu}\} = 1/2$. The empirical meaning of 'median' is also clear : in a large independent sample of same population we may expect that about half of the values of the sample will be below $\tilde{\mu}$ and about half above. For the *first* and *third quartiles* $\chi_{1/4}$ and $\chi_{3/4}$, the *deciles* (p = 1/10, ... , 9/10), or the *percentiles* (p = 1/100, ..., 99/100), the probabilistic interpretation and the empirical meanings are analogous.

Let us introduce a useful notion. The function

$$\mu(x) = f(x) / [1 - F(x)] = -\frac{d}{dx} \log (1 - F(x)) = -\frac{d \log S(x)}{dx}$$

(where log denotes, as usual, the natural logarithm) is called the *intensity function*, *mortality function*, *hazard rate* or *failure rate*. Its empirical meaning is obvious: it is the conditional density of X at x, knowing that $X \geq x$.

Let us now begin the consideration of relevant statistics. Let $(X_1, \ldots, X_n)$ be an i.i.d. sample from a population, that is, n independent and identically distributed observations or random variables with distribution function F(x), supposed to be absolutely continuous for the sake of simplicity.

The statistics frequently considered may be location statistics, such as the average, the empirical or sample quantiles, the maximum or the minimum of a sample, etc.; dispersion statistics, such as the standard deviation, the range, the absolute deviation, etc.; the asymmetry and the kurtosis statistics, all of which are well known and will not be defined in the text; but statistics as the absolute maximum or minimum, the mid-range = mid-point = centre of a sample, the extremal quotient, and shape statistics will be defined, when used for the first time.

Let us now introduce an idea to which engineers are accustomed when they think of events that, on average, repeat themselves with a "random cycle" of some time units (years, months, etc.,): the idea of the *return period.*

Let A be an event of (permanent) probability p (= 1 – q) and consider a sequence of i.i.d. trials in which the event can happen or not happen. The random variable T denoting the first time when A happens has a distribution given by the elementary probabilities

$$\text{Prob}\{T = j\} = \text{Prob}\{\text{A does not happen in the first } j-1 \text{ trials}\} \times \text{Prob}\{\text{A happens at the } j\text{ – th trial}\} = q^{j-1}\,p.$$

It is immediate that $\sum_{j=1}^{\infty} q^{j-1}\,p = 1.$

We have the mean value

$$M(T) = \sum_{1}^{\infty} j\,q^{j-1}\,p = p\,\frac{d}{dq}\left(\sum_{1}^{\infty} q^j\right) = p\,\frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{1}{p}$$

and the variance

$$V(T) = \sum_{1}^{\infty} j^2\,q^{j-1}\,p - \left(\frac{1}{p}\right)^2 =$$

$$= \sum_{1}^{\infty} j(j-1)\,q^{j-1}\,p + \sum_{1}^{\infty} j\,q^{j-1}\,p - \left(\frac{1}{p}\right)^2 =$$

$$= p\,q\,\frac{d^2}{dq^2}\left(\sum_1^{\infty} q^j\right) + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}\,.$$

The median time $\tilde{\mu}(T)$ to the return of A, approximately given by $\sum_1^{\tilde{\mu}} p\ q^{j-1}$ $= 1 - q^{\tilde{\mu}(T)} = 1/2$, is $\tilde{\mu}(T) \approx \frac{-\log 2}{\log q}$, asymptotic to $\log 2 \times \frac{1}{1-q} = \log 2 \times M(T)$ as $q \to 1$.

The value $M(T) = 1/p$ is called the *return period* of the event A in a sequence of trials or experiments; the probability p, of A, is estimated by the average duration between the realizations of the event A. More concretely, if the event did occur k times in the trials of orders $n_1 < n_2 < \ldots < n_k$, the estimate of T is

$$\bar{T}_k = \frac{n_1 + (n_2 - n_1) + \ldots + (n_k - n_{k-1})}{k} = \frac{n_k}{k}\,.$$

As $\bar{T}_k$ is an average of k i.i.d. random variables $T_j = n_j - n_{j-1}$ $(n_0 = 0)$, we know that $\sqrt{k}\,\frac{\bar{T}_k - M(\bar{T}_k)}{\sqrt{V(T)}} = \sqrt{k}\,\frac{\bar{T}_k - 1/p}{\sqrt{q/p^2}} = \sqrt{k}\,\frac{\bar{T}_k / T - 1}{\sqrt{1 - 1/T}}$, by the Central Limit Theorem, is asymptotically normal and we can easily obtain, for large k, asymptotic confidence intervals for T. Also $\sqrt{k}\,\frac{\bar{T}_k / T - 1}{\sqrt{1 - 1/\bar{T}_k}}$ is asymptotically normal.

In practice we are, in general, interested in the underpassing (as droughts) or overpassing (as floods) of a fixed threshold a that can be considered a catastrophe level; for instance, then we have for underpassing $p = F(a)$ and for overpassing $p = 1 - F(a)$. To design a dam for a duration D, say 50 years, we need to know the point a such that $T(a) = \frac{1}{1 - F(a)} \geq 50$ or $p(A) = 1 - F(a) < 1/50$. If we know $F(x)$ or $T(x)$, we have to solve for a, the equation, $\frac{1}{1 - F(a)} = D$ (for overpassings) or $\frac{1}{F(a)} = D$ (for underpassings), i.e., to obtain the $(1 - 1/D)$ -quantile or $(1/D)$ -quantile respectively. Suppose $F(x)$ is not known and we need to estimate $T(a)$ with some confidence coefficient $1 - \alpha$. Let λ_α be such that $N(\lambda_\alpha) = 1 - \alpha$, where $N(.)$ denotes the standard normal distribution function. We have, then, approximately

$$\text{Prob}\left\{\sqrt{k}\,\frac{\bar{T}_k / T - 1}{\sqrt{1 - 1/\bar{T}_k}} < \lambda_\alpha\right\} \approx N(\lambda_\alpha) = 1 - \alpha \qquad \text{or, equivalently,}$$

Prob $\{\bar{T}_k / (1 + \frac{\lambda_\alpha \sqrt{1 - 1/\bar{T}_k}}{\sqrt{k}} < T\} \approx 1 - \alpha$; giving thus the (approximate but safe) confidence region $T > \bar{T}(1 + \frac{\lambda_\alpha \sqrt{1 - 1/\bar{T}_k}}{\sqrt{k}})$.

But in general, we do not have sufficient large samples to obtain *empirically* one value a such that D belongs to the region for small α ($\alpha = .05$ leads to $\lambda_{.05} = 1.64$).

Then this is done usually by substituting the exact *unknown* distribution function by one of the approximations that are the asymptotic distributions of extremes or using another direct approach based on the asymptotic distributions of extremes. The question will be discussed in Part II, Annex IV.

Some time ago, to avoid long or difficult computations, it was usual for overpassings, as $\frac{-\log u}{1-u} \to 1$ when $u \to 1$, and because we were dealing with large values of $F(a)$, to substitute $\frac{1}{1 - F(a)} = T(a)$ by its approximation $\tilde{T}(a) = (-\log F(a))^{-1}$, the Fuller approximation. As $1 - u \le e^{-u} \le 1 - u + u^2/2$ ($u > 0$), putting $u = -\log F(a)$ we get $\tilde{T}(a) \le T(a) \le \tilde{T}(a)(1 - 1/2\,\tilde{T}(a))^{-1}$, which shows that $\tilde{T}(a)$ is a conservative underevaluation of $T(a)$ with a relative error smaller or equal to $1 / 2\,\tilde{T}(a)$. Analogously, for minima, as the return period is $T(a) = 1/F(a)$, the (conservative) approximation is $\tilde{T}(a) = (-\log(1 - F(a))^{-1}$ with the same inequalities and the same bound to the relative error.

Note that, today, this approximation is not necessary because the $(1 - 1/D)$ -quantile or $1/D$ -quantile are easily obtained with the aid of a pocket calculator.

The notion of the return period of some event, for its first realization, can be easily extended. If we are interested in the random time T of the s-th occurrence of the event A with (permanent) probability p, we have the basic equation :

Prob $\{T = j\}$ = Prob {A occurs $s - 1$ times in the 1st $(j - 1)$ trials)

x Prob {A occurs at the j-th trial} $= \binom{j-1}{s-1} p^{s-1} q^{j-s} p = \binom{j-1}{s-1} p^s q^{j-s}$.

Evidently $\sum_{j=s}^{+\infty} \binom{j-1}{s-1} p^s q^{j-s} = 1$.

Its mean value is

$$M(T) = \sum_{j=s}^{+\infty} j \binom{j-1}{s-1} p^s q^{j-s} = \frac{s}{p}$$

and the variance is

$$V(T) = M(T^2) - M^2(T) = M(T(T-1)) + M(T) - M^2(T) = \frac{sq}{p^2},$$

as is easily seen using the relation $\sum_{j=s}^{+\infty} \binom{j-1}{s-1} p^s q^{j-s} = 1$, valid for any s (positive integer), and with some minor algebra. The general formulae, for $s = 1$, gives the initial results, as should happen.

This result for the random time T of the s-th exceedance could be expected: in fact, T is the sum of s *independent* times of first exceedance, which may be helpful for the consideration of higher moments.

Some History of Statistical Extremes Theory

Statistical Extremes Theory — a growing field in the area of Statistical Sciences — deals essentially with the largest or smallest values of samples, the overpassing or underpassing of some bounds which can give rise to catastrophes such as floods, droughts, breaking of materials, large waves, etc.

And to catastrophes we can make some references in classical books: the great age of Methuselah, the Great Flood, the safety of Noah's ark (i.e., building against a catastrophic risk) etc. in *Genesis* (the *Bible*, vol. I); the Deluge, Noah's ark, the storm of stones etc. in the *Koran*; and even the catastrophic powers of Zeus (the Thunderer), of Poseidon (the Lord of the Earthquake), and other Gods and Heroes as seen in Homer's *Iliad.*

These classical references give a hint on the possible applications of the theory: to Engineering (structural safety, dam and port building, fatigue of materials, ...), to Geophysical Sciences (earthquakes, gusts of wind, snowfalls, floods, droughts, large waves, ...), etc.

The first fundamental paper on extremes is the one by Dodd (1923) : in this paper, the author presents the distribution of the maximum and the minimum of a sample, and also gives important approximations allowing the approximate computation of quantiles.

Shortly afterwards, we have a paper by Fisher and Tippett (1928), where through a vague proof of the stability equation, the authors derive the three asymptotic distributions and introduce the theme of the rate (or speed) of convergence. At the same time Fréchet (1927) obtains, through the stability relation, the Fréchet distribution. De Finneti (1932) deals with the first aspects of a law of large numbers for extremes.

Later von Mises (1936) presented a different approach through the use of the concept of force of mortality or hazard rate, very important in Demography and Reliability. Gumbel (1935), with approximate proofs, discussed the subject and extended it to second, third, ... extremes. Gumbel (1937) used these ideas in the study of greatest ages.

Weibull's (1939) paper presents a physical approach, different from the usual ones, but connected with the stability relation, leading to the Weibull distribution.

Finally Gnedenko (1943) solves the most important problems of the asymptotic behaviour of statistical extremes, giving the necessary and sufficient conditions for the validity of the law of large numbers, proves correctly that the classes of asymptotic distributions and stable distributions for extremes coincide, obtains the necessary and sufficient conditions for attraction to each of the three asymptotic distributions, and gives examples of no-attraction.

These are the first steps of the probabilistic theory of extremes, obviously in the independent and identically distributed case; since then the flow of the papers has been increasing each year and it is very difficult to refer to one paper without referring to a dozen or more at the same level and in the same area. Statistical decision theory, weakening of the i.i.d., conditions, multivariate extremes, extremes in random sequences processes and fields are being studied intensively, but there is still a gap between the techniques at our disposal and the existing needs. Suffice it to call attention, as very important books, to the introductory Gumbel (1954), Gumbel (1958), Galambos (1978), Leadbetter, Lindgren and Rootzén (1983) and Tiago de Oliveira (1984).

Additional Notations

X or X_i (latin capital letters, indexed or not) — random variables or vectors;

$(X_1, \ldots, X_n)$ — the sample (to be observed)
$(x_1, \ldots, x_n)$ — the sample (observed)
} — when a distinction is needed;

$\{X_i\}$ — a sequence of random variables or vectors ($i = 1, 2, \ldots$);

$X(t)$, $t \in T$ — a stochastic process (or field), indexed by T : this may be $\{1,2, ...,n\}$ representing a sample, $\{1,2, \dots , n, \dots\}$ representing a sequence, $[a, b]$ representing a realization of a continuous time (vector) process, $[a,b] \times [a',b']$ representing a bidimensional (vector) field, $\{t_1 < \dots < t_n\}$ representing a finite sample of a stochastic process (sometimes a point process), etc.;

$M(X)$ — the mean (or expected) value;

$V(X)$ — the variance;

$C(X, Y)$ — the covariance;

$\max_1^n \{X_i\} = \max (X_1, \dots , X_n)$ — the maximum of a sample;

$\min_1^n \{X_i\} = \min (X_1, \dots , X_n)$ — the minimum of a sample;

$$F_n(x) = \text{Prob } \{\max_1^n \{X_i\} \leq x \} ;$$

$$S_n(x) = \text{Prob } \{\min_1^n \{X_i\} > x \} ;$$

$L(x) = \lim_{n \to \infty} F_n(a_n + b_n x)$ for convenient $\{a_n, b_n > 0\}$, when $L(x)$ exists;

$S(x) = \lim_{n \to \infty} S_n(a_n + b_n x)$ for convenient $\{a_n, b_n > 0\}$, when $S(x)$ exists;

when it will clear $\lim_{n \to \infty}$ or $\lim_{t \to \infty}$ or $\lim_{t \to 0}$ will be denoted simply by lim;

$L(x)$ and $S(x)$ are always supposed proper and in general non-degenerated in an almost sure random variable.

$A^c = \{x \mid x \notin A\}$ — the complement of the set A ;

i.i.d. — independent and identically distributed random variables (vectors);

$N(x)$ — the standard normal distribution function $N(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt$;

$E(x)$ — the standard exponential distribution function :

$$E(x) = 0 \quad \text{if} \quad x \le 0, \; E(x) = 1 - e^{-x} \quad \text{if} \quad x \ge 0;$$

$U(x)$ — the standard uniform distribution function in $[0,1]$:

$$U(x) = 0 \quad \text{if} \quad x \le 0, \; U(x) = x \quad \text{if} \quad 0 \le x \le 1, \quad U(x) = 1 \quad \text{if} \quad x \ge 1;$$

$H(x) = 0$ if $x < 0$, $H(x) = 1$ if $x \ge 1$, the Heaviside (unit jump) function;

$S_n(x \mid x_i)$ or $S_n(x)$, for short, — the sample distribution function for the sample $(x_1 , \dots , x_n)$,

$$S_n(x \mid x_i) = S_n(x) = \#\,(x_i \le x) / n = \frac{1}{n} \sum_{i=1}^{n} H\,(x - x_i) \; ;$$

$f_+ = \max\,(0,f)$, $f_- = \max\,(0,-f) = -\min\,(0,f)$ with $f_+ - f_- = f$ and $f^+ + f^- = |f|$;

$[x]$ – the integer part of x ;

$r(x) = x - [x]$ — the fractional part of x , $0 \le r\,(x) < 1$, $r\,(x+1) = r\,(x)$;

$a_n \sim b_n$ or $a(t) \sim b(t)$ means "asymptotic to" , i.e., that $a_n / b_n \to 1$ as $n \to \infty$ or $a(t)/b(t) \to 1$ as $t \to t_0$ (eventually $+\infty$, 0^+, etc) ;

$f = o(g)$ means that $f(t)/\,g(t) \to 0$ as $t \to t_0$; $o(1)$ means a vanishing a function or sequence ;
$f = o(g)$ is equivalent to $f(t)/\,g(t) = o(1)$;

$f = O(g)$ means that $f(t)/\,g(t) \to c$ $(0 < c < +\infty)$ when $t \to t_0$.

Notations for order statistics are introduced in chapter I.

B. DE FINETTI (1932). Sulle legge di probabilitá degli estreme, *Metron*, 9, 127-138

E. L. DODD (1923). The greatest and least variate under general laws of error, *Trans. Am. Math. Soc.*, 25, 525-539

R. A. FISHER and L.H.C. TIPPETT (1928). Limiting forms of the frequency distribution of the largest and smallest values of a sample, *Proc. Camb. Phil. Soc.*, 24, 180-190

M. FRÉCHET (1927). Sur la loi de probabilité de l´écart maximum, *Ann. Soc. Polon. Math.* (Cracovie), 6, 93-116

J. GALAMBOS (1978). *The Asymptotic Theory of Extreme Order Statistics,* Wiley, New-York; (1987), 2nd ed., Krieger Publ.

B. V. GNEDENKO (1943). Sur la distribution limite du terme maximum d´une série aléatoire, *Ann. Math.*, 44, 423-453

E. J. GUMBEL (1935). Les valeurs extrêmes des distributions statistiques, *Ann. Inst. H. Poincaré*, 4, 115-168

E. J. GUMBEL (1937). *La Durée Extrême de la Vie Humaine*, Hermann et Cie, Paris

E. J. GUMBEL (1954). *Statistical Theory of Extreme Values and Practical Applications*, Nat. Bur. Standards, Appl. Math. Series, #33, Washington

E. J. GUMBEL (1958). *Statistics of Extremes*, Columbia University Press, New York

M. R. LEADBETTER, GEORG LINDGREN and HOLGER ROOTZÉN (1983). *Extremes and Related Properties in Random Sequences and Processes*, Springer-Verlag, Heidelberg

J. TIAGO DE OLIVEIRA (ed.) (1984). *Statistical Extremes and Applications*, D. Reidel, Dordrecht

R. VON MISES (1936). La distribution de la plus grande de n valeurs, *Rev. Math. Union Interbalkanique*, 1, 141-160 and also in *Selected Papers*, vol.II, 1954, 271-294, Am. Math. Soc., Providence, Rhode Island

W. WEIBULL (1939). A statistical theory of the strength of materials, *Proc. Royal Swed. Inst. Eng. Res.* # 151, Stockholm

PART I

Probabilistic patterns of univariate statistical extremes

Chapter I

INTRODUCTION

The general framework

The purpose of STATISTICAL ANALYSIS of EXTREMES is to deal with, analyse and predict (or forecast) aspects of natural phenomena that correspond to the largest or smallest values of sampled data, or to over - or underpassing some level that sometimes lead to natural disasters. Floods, large fire claims, heavy rains, gusts of wind, and large waves, are examples of maxima or largest values of samples, as well as droughts, breaking strength of materials, failure of equipment or apparatus, low temperatures, etc. are examples of minima or smallest values of samples, of extremes of samples or extreme order statistics; many times not only the first but the second, third, etc. extremes are relevant.

Those problems can be analysed using suitable or available samples — according to the manageability of technique, error allowance for models, convenience and simplicity, etc — as univariate samples, multivariate samples or excursion(s) of random sequences or of stochastic processes (if time is continuous). Evidently, from the philosophical point of view, as a rule, natural phenomena should be considered as multivariate random sequences or stochastic processes (according to the discreteness or the continuity of the time set). This procedure, however sound it may be, by its complication would not allow for a (practical) description of the observations (the past and/or present) but also to predict about the next observations (the future) to legitimate sound inference and/or decision design.

In fact, in data analysis, we can sometimes proceed through a sequence of models, adapted to the case. Case studies, as in the last chapter and in the statistical chapters of the book, may be of help to practitioners.

We have not tried to prove all results, to keep the book to a manageable size (though not necessarily completely balanced) and to allow for easier use for data analysis and design; essential references are given for the proved (sometimes in a new or different way) or unproved results.

Basically, the applications will be based on the use of asymptotic distributions as approximations (better or worse) to the description of extremes, useful for forecasting: so obtaining asymptotic distributions, as well some reference to the speed of convergence, plays an essential role in the book.

If the approximation is good we will, with good efficiency, estimate or predict "extreme" quantiles (for large or small probabilities) for the design of structures, evaluation of extreme conditions, forecasting catastrophes, etc..

As a rule, the behaviour of sums and of extremes (maxima and minima) of samples are different and not strongly related, although there are some connexion as regards convergence of averages and of extremes. But, as a general rule, in the i.i.d. case, characteristic functions for sums vs. distribution functions for maxima (or survival functions for minima), the linear combination $aX + bY$ vs. the maximal combination $\max (X + a, Y + b)$, seem to play a similar or dual role; but the duality almost stops here. The Central Limit Theorem leads to the normal distribution (at least when the variance exists) or more generally to the infinitely divisible distributions with a central role for the normal, but the Extremal Limit Theorem leads to three asymptotic distributions (which can be integrated in one form, with a shape parameter) where the Gumbel distribution plays a central role; not only are both distributions indefinitely differentiable but, also, sums of Gumbel random variables are asymptotically normal as well as maxima or minima of normal random variables have asymptotically the Gumbel distribution. It does not seem that we can extend the analogies and this "experimental" digression will play only a vague role — behind the scenes — in the book. But see the Annex I, of Part I, for more details that extend to bivariate distributions.

Let us consider one example to clarify the use of the theory of statistical extremes. Suppose that we are considering the levels (or discharges) of a river at some fixed point and that we are interested in the yearly floods or droughts at that point. As the flood (drought) is the maximum (minimum) of the discharges we would be led to consider the exact statistical behaviour of the maxima (minima). The difficulties of dealing with them and the fact the yearly maximum considered is obtained from a large number of (daily) observations suggest using the asymptotic theory of maxima, whose distribution is known except for two or three parameters. The validity of this approximate substitution will be considered and referred to. A set of 60 or more years of observations in general gives sufficient data to estimate those unknown parameters with adequate accuracy and allows subsequent use of the asymptotic distribution function of the maxima or minima.

For bivariate extremes the situation is similar, although more complicated. Suppose that we are measuring the discharges of a river at two different points every day, say, at noon. It is evident that the measurements are positively correlated, the correlation being stronger and stronger as the points are closer and closer. From this series of daily measurements (pairs) we can obtain a sample sequence of pairs of maxima (the floods). We will see later how to deal with those samples of (yearly) pairs of maxima, considered as random pairs.

As one year is a relatively large sample, we will use the large sample (asymptotic) distribution of pairs of maxima as a substitute for its very complex and unknown actual distribution, but the representation of the large sample distribution of pairs of maxima has some unknown parameters and a dependence function.

The sample of pairs of maxima obtained in 50, 60, ..., 100 years of observations can then be used to solve the statistical decision problems concerning the unknown elements of the representation, specifically those that are most important in engineering applications: estimation of the parameters and testing hypotheses about them, testing for the independence of the maxima, and estimation of the dependence function.

We will not refer specially to the behaviour of minima (droughts, for example) because it is similar to that of maxima; we will recall the technique of conversion of results concerning maxima into results concerning minima.

The stochastic model

Let $\{X_1,..., X_n\}$ be a set of multidimensional random vectors with the joint distribution function $F_n(x_1,..., x_n) = \text{Prob}\{X_1 \le x_1,..., X_n \le x_n\}$ and with survival function $S_n(x_1, ..., x_n) = \text{Prob}\{X_1 > x_1, ..., X_n > x_n\}$; the relations between $F_n(x_1, ..., x_n)$ and $S_n(x_1, ..., x_n)$ are known in the continuity points of the margins being extended afterwards by right continuity for F_n and S_n.

We have $\text{Prob}\{\max_1^n \{X_i\} \le x\} = F_n(x, ..., x) = F_n(x)$ and

$\text{Prob}\{\min_1^n \{X_i\} > x\} = S_n(x, ..., x) = S_n(x)$; in any point x we see that

$$\text{Prob}\{\max_1^n X_i > x\} = 1 - F_n(x, \dots, x) = 1 - F_n(x) \quad \text{and}$$

$$\text{Prob}\{\min_1^n X_i > x\} = 1 - S_n(x, \dots, x) = 1 - S_n(x).$$

Evidently as $\max_1^{n+1} X_i \geq \max_1^n X_i$ we have $F_n(x) \geq F_{n+1}(x)$, and as

$\min_1^{n+1} X_i \leq \min_1^n X_i$ we obtain $S_{n+1}(x) \leq S_n(x)$.

Note that if $h(.)$ is an increasing continuous function then $\max_1^n \{h(X_i)\} = h(\max_1^n X_i)$ and $\min_1^n \{h(X_i)\} = h(\min_1^n X_i)$; if $h(.)$ is a decreasing continuous function we obtain $\max_1^n \{h(X_i)\} = h(\min_1^n X_i)$ and $\min_1^n \{h(X_i)\} = h(\max_1^n X_i)$.

When we take $h(u) = -u$ we obtain the essential equalities

$$\max_1^n \{X_i\} = -\min_1^n \{-X_i\} \quad \text{and} \quad \min_1^n \{X_i\} = -\max_1^n \{-X_i\}.$$

These relations translate maxima results into minima results and vice-versa. In this book we will almost always deal with maxima results, the conversion to minima results following immediate; the exceptions will be, essentially, the study of the Weibull distribution for minima and the study of the dependence function with (standard) exponential margins in bivariate extremes.

Notice that these simple results in $h(.)$ — a very general transformation — may after be helpful in dealing with data, by a convenient transformation.

Let us apply the results above to the i.i.d. case; we have $F_n(x) = F_1^n(x) = (1 - S_1(x))^n$ and $S_n(x) = S_1^n(x) = (1 - F_1(x))^n$.

Order statistics and exceedances

Let $(X_1, \dots, X_n)$ be an i.i.d. univariate sample where $F(x)$ is the distribution function of the X_i. Let us order the sample as $X'_{1,n} \leq X'_{2,n} \leq \dots \leq X'_{n,n}$ (*ascending order statistics*) or $X''_{1,n}$

$\geq X''_{2,n} \geq \ldots \geq X''_{n,n}$ (*descending order statistics*) ; note that $X'_{k,n} = X''_{n+1-k,n}$ and that $X'_{1,n} = X''_{n,n}$ is the minimum of the sample, $X'_{n,n} = X''_{1,n}$ is the maximum of the sample, $X'_{2,n} = X''_{n-1,n}$ and $X'_{n-1,n} = X''_{2,n}$ are second minimum and maximum (second extremes), $X'_{k,n} = X''_{n+1-k,n}$ and $X'_{n+1-k,n} = X''_{k,n}$ are the k-th minimum and maximum. When possible and clear we will use the simple notations X'_k and X''_k.

The distribution function of $X'_{k,n} = X''_{n+1-k,n}$ is

$$F_{k,n}(x) = \text{Prob}\,\{X'_{k,n} \leq x\} = \text{Prob}\,\{k \text{ or more } X_i \leq x\} =$$

$$= \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j} ,$$

by the binomial distribution.

If F(x) has a probability density f(x) = F'(x), the probability density of $X'_{k,n} = X''_{n+1-k'}$ is

$$F'_{k,n}(x) = F(x)^{k-1} (1 - F(x))^{n-k} f(x) = k\binom{n}{k} F(x)^{k-1} (1 - F(x))^{n-k} .$$

Analogously the distribution function of $(X'_{k,n}, X'_{m,n}) = (X''_{n+1-k,n}, X''_{n+1-m,n})$, can be written with $1 \leq k < m \leq n$; its probability density is immediately seen to be, for $X'_{k,n} = x$, $X'_{m,n} = y$,

$= 0$ if $y \leq x$

$$= \frac{n!}{(k-1)!(m-k-1)!(n-m)!} F^{k-1}(x) (F(y) - F(x))^{m-k-1} (1 - F(y))^{n-m} f(x) f(y) \text{ if } y \geq x.$$

From these distribution functions and their multivariate extensions — where concomitants can intervene, but which will not be dealt with in this book — we can obtain the asymptotic

behaviour of the sample quantiles which are *central* order statistics, i.e., in a sequence of sample the k_n-th ascending order statistic $X'_{k_n,n}$ is such that $k_n/n \to p$ with $0 < p < 1$.

Let us now suppose that we are thinking of a second sample of m random variables with the same distribution function F(x). The probability that *exactly* J = j of the random variables of the second sample are smaller or equal to $X'_{k,n} = X''_{n+1-k,n} = x$ is evidently $\binom{m}{j} F(x)^j (1 - F(x))^{m-j}$

and, thus, the probability — *before* the first sample — that exactly j out of m observations are smaller or equal to the k-th ascending order statistic $X'_{k,n}$ of the first sample is

$$P(j,m|k,n) = \int_{-\infty}^{+\infty} \binom{m}{j} F^j(x) (1 - F(x))^{m-j} \; k\binom{n}{k} F^{k-1}(x) (1 - F(x))^{n-k} \; f(x)\, dx$$

$$= k \binom{m}{j} \binom{n}{k} \int_{-\infty}^{+\infty} F^{j+k-1}(x) (1 - F(x))^{m-j+n-k} \; f(x)\, dx$$

$$= \frac{k \binom{n}{k} \binom{m}{j}}{(k + j) \binom{n+m}{k+j}}$$

with $$\sum_{j=0}^{m} \frac{k \binom{n}{k} \binom{m}{j}}{(k + j) \binom{n+m}{k+j}} = 1 .$$

Note the symmetry $P(j,m|k,n) = P(m - j, m \mid n + 1 - k , n)$ and that this last probability is that exactly j observations of the second sample will be larger than or equal to $X''_k = X'_{n+1-k'}$.

These situations, for the second sample, are called *exceedances* : the underpassing of $X'_{k,n}$ or the overpassing of $X''_{k,n}$.

Let us now compute the mean value and variance of the random number of exceedances J (below $X'_{k,n}$ or above $X''_{k,n}$). We have

$$\mu\,(m|k,n) = M(J) = \sum_{j=0}^{m} j\, P(j,m|k,n) = \frac{m\, k}{n+1}$$

and

$$\sigma^2 (m|k,n) = M(J^2) - M^2(J) = M(J(J-1)) + M(J) - M^2(J)$$

$$= \frac{m(m+n+1)}{(n+1)^2 (n+2)} \; k\,(n+1-k)\,.$$

The variance is smaller for $k = 1$ or $k = n$ and the ratio of the variances $\sigma^2 (m|1,n) = \sigma^2 (m|n,n)$ to the variance at $k = \frac{n+1}{2}$ (the median if n is odd) is $\frac{4n}{(n+1)^2} \sim 4 / n$, which shows that, for exceedances, extremes are more reliable than central values, contrary to the usual belief.

Let us now consider, more generally, the random time T of the *s-th* exceedance (the underpassing of $X'_{k,n}$ or the overpassing of $X''_{k,n}$) in a (possibly) infinite second sample (m not fixed as before) . Thus the random order T in which the exceedance happens does not relate to a fixed event — as in the (classical) return period situation — but to a random event, defined previously to the observation of the first sample of n random variables. Consequently we will not need the knowledge the distribution of $X'_{k,n}$ or of $X''_{k,n}$, as was the case in the return period situation, because the set is defined previously to the observations : briefly speaking, we have an inverse binomial with an underlying random probability. The approach will, naturally, be non-parametric as for exceedances.

For the *s-th* underpassing of $X'_{k,n}$ at the $T = j$ observation we have

$$P(j|s,k,n) = \text{Prob}\,\{T = j|s,k,n) = \int_{-\infty}^{+\infty} [\,\binom{j-1}{s-1} F^s(x)\;(1-F(x))^{j-s}\,]\;x$$

$$x\,[\,n\binom{n}{k}\;F(x)^{k-1}\,(1-F(x))^{n-k}\,f(x)\,]\,dx,$$

where the meaning of the two brackets is clear from the previous study.

It is immediate that

$$P(j|s,k,n) = \frac{k\;\binom{n}{k}\binom{j-1}{s-1}}{(k+s)\binom{n+j}{k+s}}$$

with $\sum_{j=s}^{+\infty} P(j|s,k,n) = 1$ obviously. This probability is, clearly, also the probability that $X''_{k,n}$ is overpassed at the j-th observation of the second sample the s-th time.

The mean value and variance of T are

$$\mu\,(s|k,n) = M(T) = \sum_{j=s}^{+\infty} j\;\; P(j|s,k,n) = \frac{n\;\; s}{k\;\; -\;\; 1}$$

and

$$\sigma^2\,(s|k,n) = M(T^2) - M^2(T) = M(T(T+1)) - M(T) - M^2(T) =$$

$$= \frac{ns(n\;\; -\;\; k\;\; +\;\; 1)\;\; (k\;\; +\;\; s\;\; -\;\; 1)}{(k-1)^2\;\; (k-2)}\;\; .$$

For $k = 1$ we have $\mu(s|1,n) = +\infty$ which is easily understandable and for $k = n$ we get $\mu(s|1,n) = \frac{n\;\; s}{n\;\; -\;\; 1}$, results that could be obtained directly. Also for $k = 1$ and $k = 2$ we have $\sigma^2\,(s|1,n) = +\infty$ (as expected) and $\sigma^2\,(s|2,n) = +\infty$.

It is convenient to compare the two different approaches of this section: exceedances and the random return period, both for the second sample, previously to the knowledge of the first sample, that, allowing for the use of non-parametric methods, overcomes the necessity of knowledge of the underlying continuous F(x). In the first case (exceedances) we study the number of underpassings or overpassings of some order statistics of the first sample in a finite number of observations as the second sample has an infinite size. For the random return period we accept, admit, or consider an infinite second sample (thus allowing for infinite i.i.d. observations) and we are dealing with a random stopping time. The analogy with the direct and inverse binomial is obvious.

A note on the asymptotic behaviour of the sample quantiles

As it is well known the usual estimator of the unique theoretical quantile χ_p $(0 < p < 1)$ is the order statistic $X'_{[np]+1}$ called the *empirical* or *sample quantile* χ^*_p which we will denote also by $Q_n(p)$; note that $X'_1 \le Q_n(p) \le X'_n$ if $(0 < p < 1)$. In the case of existence of a

density $f(x) = F'(x)$ we know that $\chi_p^* \xrightarrow{P} \chi_p$ and $\sqrt{n}\ f(\chi_p)\ \frac{\chi_p^* - \chi_p}{\sqrt{p(1-p)}}$ is asymptotically standard normal if $0 < f(\chi_p) < \infty$; the joint distribution of the set of quantiles $(\chi_{p_1}^*, \ldots, \chi_{p_n}^*) =$ $(Q_n(p_1) \ldots, Q_n(p_k))$, is such that $\{\sqrt{n}\ f(\chi_{p_i})\ \frac{\chi_{p_i}^* - \chi_{p_i}}{\sqrt{p_i(1-p_i)}}\}$ is asymptotically multinormal with standard margins and correlation coefficients, for $p_i < p_j$, $\rho_{ij} = \sqrt{\frac{p_i}{p_j}\ \frac{1-p_j}{1-p_i}}$ (>0); for details see Cramér (1946).

Chapter II

ASYMPTOTICS OF UNIVARIATE EXTREMES

Introduction

This chapter can be divided into two different parts: the first (laws of large numbers) which expresses, in a formal way, the intuitive ideas relative to the behaviour of sample maxima and minima and the way they converge to some values (finite or infinite) and the second (asymptotic distributions of extremes) dealing with the behaviour of extremes in large samples, which is, in fact, the basis for the development of the whole book. Although the first part is of almost entirely theoretical importance only, as its practical application is essentially nil, it helps to pave the way for the following developments. Practically speaking hardly anyone will use the laws of large numbers, but many of the practical applications in which extremes have to be taken into account, such as the design of dams, the study of fatigue fractures, etc, are based on the asymptotic distributions of extremes.

General introduction to the laws of large numbers

We say that a sequence of random variables $\{X_k, k = 1,2,...\}$ verifies *a* law of large numbers if there exists a sequence of functions $\{\varphi_k\}$ (possibly one function φ), a sequence of constants $\{c_k\}$, and a finite constant c such that $\varphi_k(X_k, c_k) \xrightarrow{P} c$ as $k \to \infty$; c can evidently be taken as zero.

The interpretation of this definition is easy : it means that for large k, $\varphi_k(X_k, c_k)$ is close in probability to c, i.e. , given $\in (>0)$ and $\delta(0 < \delta < 1)$ we know that for $k > N(\in, \delta)$ we have

$$\text{Prob}\{c + \in \leq \varphi_k(X_k, c_k) \leq c + \in\} \geq 1 - \delta,$$

thus obtaining an approximate evaluation of the values of X_k for large k. How good the approximation is depends on how large k must be, for small ϵ and δ; sometimes the approximation is bad even for very large *k*, in real–life conditions. Be that as it may, it gives a *deterministic* approximation that may be useful for rough calculations.

Such a large framework, depending also on the choice of $\{\varphi_k\}$, has not aroused interest in itself and, until now, studies have specialized in two forms of φ_k: the additive and the multiplicative laws of large numbers, denoted ALLN and MLLN respectively.

The sequence $\{X_k\}$ satisfies the ALLN if there exist $\{c_k\}$ such that $X_k - c_k \xrightarrow{P} 0$. Evidently the c_k are not uniquely defined because if $\{\epsilon_k\} \to 0$ *also* $X_k - (c_k + \epsilon_k) \xrightarrow{P} 0$.

If $h(x) = \alpha + \beta x$ $(\beta \neq 0)$ is a linear transformation then $h(X_k) - (\alpha + \beta a_k) \xrightarrow{P} 0$, i.e., $\tilde{X}_k = h(X_k)$ also satisfies the ALLN if with coefficients $\tilde{a}_k = \alpha + \beta a_k$.

The sequence $\{X_k\}$ satisfies the MLLN if there exist $\{c_k \neq 0\}$ such that $X_k / c_k \xrightarrow{P} 1$. It is evident, also, that c_k is not uniquely defined, as $X_k / (c_k / \eta_k) \xrightarrow{P} 1$ if $\eta_k (> 0) \to 1$. Evidently $\{\beta X_k\}$ $(\beta \neq 0)$ satisfies the MLLN with $\{\beta c_k\}$.

Putting aside the case where $\{c_k\}$ oscillates permanently in sign, we can suppose that we have c_k ultimately (i.e., for all $k > k_0$) positive or negative. If c_k is ultimately positive we can define $\tilde{c}_k = 1$ if $c_k \leq 0$ and $\tilde{c}_k = c_k$ if $c_k > 0$ which is such that $X_k / \tilde{c}_k \xrightarrow{P} > 1$; also if c_k is ultimately negative we can define $\tilde{c}_k = -1$ if $c_k > 0$, $\tilde{c}_k = c_k$ if $c_k < 0$ which is such that $X_k / \tilde{c}_k \xrightarrow{P} 1$; we can take it that we always have $c_k > 0$ or $c_k < 0$. But if $X_k / c_k \xrightarrow{P} 1$ and $c_k < 0$ then $(-X_k)/(-c_k) \xrightarrow{P} 1$, and so we can — with a change of sign for X_k if needed — always take $c_k > 0$; an alternative was always to take $c_k > 0$ and have $X_k / c_k \xrightarrow{P} -1$ or $X_k / c_k \xrightarrow{P} 1$. Note that the choice between $c_k > 0$ and $c_k < 0$ is easy. If MLLN is valid with $c_k > 0$ we have, $F_k(x)$ denoting Prob $\{X_k \leq x\}$, $F_k(0) = \text{Prob}\{X_k \leq 0\}$ $\leq \text{Prob}\{X_k \leq c_k (1 - \epsilon)\} = \text{Prob}\{X_k / c_k - 1 < -\epsilon\} \leq \text{Prob}\{|X_k / c_k - 1| > \epsilon\} \to 0$ and $F_k(0) \to 0$; in reverse, if the MLLN is valid with $c_k < 0$ we have $F_k(0) = \text{Prob}\{X_k \leq 0\} \geq$ $\text{Prob}\{X_k \leq c_k (1 - \epsilon)\} \geq \text{Prob}\{|X_k / c_k - 1| < \epsilon\} \to 1$. $F_k(0) \to 0$ and $F_k(0) \to 1$ are the indicators for the choice of the sign of c_k. $F_k(0) \to 0$ means that $\{X_k > 0\}$ has a probability close to 1 when k is large and so X_k can, then, be considered practically positive; dually, if

$F_k(0) \to 1$, X_k, for large k, is practically negative. This frame is sufficient for the study of maxima, minima being dealt with by decreasing transformations.

Let us obtain some simple consequences :

1) if $X_k \geq 0$ (i.e., $F_k(0) = 0$) verifies the MLLN, with coefficients $c_k > 0$, then $\log X_k$ verifies the ALLN with coefficients $\log c_k$; more generally if $\{X_k\}$ verifies the MLLN with $c_k > 0$ then $\log \max (c_k , X_k)$ verifies the ALLN with $\log c_k$ as coefficients; conversely, if X_k verifies the ALLN with coefficients c_k then $\exp X_k$ (>0) verifies the MLLN with coefficients $\exp c_k (>0)$;

2) if $\{X_k\}$ verifies the ALLN, with $c_k > 0$, then $\{X_k\}$ verifies the MLLN with the same coefficients c_k : it is a simple consequence of $X_k - c_k = c_k (X_k / c_k - 1) \xrightarrow{P} 0$

if $\lim c_k > 0$; also if $c_k < 0$ and $\lim c_k < 0$ the same happens, $(- X_k)$ and $(- c_k)$ are in the previous conditions and so $(-X_k / - c_k) \xrightarrow{P} 1$.

Let us now obtain the necessary and sufficient conditions to have ALLN for $\{X_k\}$ with coefficients $\{c_k\}$ and the MLLN for $\{X_k\}$ with coefficients $\{c_k > 0\}$.

The condition $X_k - c_k \xrightarrow{P} 0$ is equivalent to $\text{Prob}\{|X_k - c_k| \leq \epsilon\} \to 1$, $\forall \epsilon > 0$, i.e., $F_k (c_k + \epsilon) - F_k ((c_k - \epsilon)^-) \to 1$ and so $F_k (c_k + \epsilon) \to 1$ and $F_k ((c_k - \epsilon)^-) \to 0$; as $F_k ((c_k + \epsilon)^-) \leq F_k (c_k - \epsilon) \leq F_k ((c_k - \epsilon')^-) \to 0$ for $\epsilon' < \epsilon$ we get that $X_k - c_k \xrightarrow{P} 0$ implies $F_k (c_k + \epsilon) \to 1$ and $F_k (c_k - \epsilon) \to 0$; the converse is immediate and so ALLN is valid iff $F_k (c_k + \epsilon) \to 1$ and $F_k (c_k - \epsilon) \to 0$, $\forall \epsilon > 0$. By the same technique we see that the MLLN is valid (with $c_k > 0$) iff $F_k ((1 + \epsilon) c_k) \to 1$ and $F_k ((1 - \epsilon) c_k) \to 0$, $\forall \epsilon > 0$.

The laws of large numbers (LLN) for extremes

It is intuitive, at least in the i.i.d. case, that the maxima and minima of a sequence of samples should converge to the right–end point $\bar{w}$ (i.e., $F(x) < 1$ if $x < \bar{w}$, $F(x) = 1$ if $x \geq \bar{w}$) and to the left-end point $\underline{w}$ (i.e., $F(x) = 0$ if $x < \underline{w}$, $F(x) > 0$ if $x > \underline{w}$).

Let us suppose $\bar{w} < +\infty$. Then Prob $\{|\max_1^k \{X_i\} - \bar{w}| \leq \epsilon\} = 1 - F^k((w-\epsilon)^-)$ by the definition of $\bar{w}$, which also implies $F^k((\bar{w}-\epsilon)^-) \to 0$ and so $\max\{X_i\} \xrightarrow{P} \bar{w}$. Suppose, now, that $\bar{w} = +\infty$; then Prob $\{\max_1^k \{X_i\} > M\} = 1 - F^k(M) \to 1$. In both cases we have shown that $\max_1^k \{X_i\} \xrightarrow{P} \bar{w}$. In the same way we can show that $\min_1^k \{X_i\} \xrightarrow{P} \underline{w}$ (finite or infinite). These results are independent of the knowledge of the distribution function $F(x)$, whether it is continuous, discrete or a mixture of both. They were shown in the proper (i.e., with zero probability at the points $\pm\infty$ if we are dealing with the complete real line) and non–degenerate (i.e., a random variable not taking some finite value with probability 1) cases; they can be extended to these situations but there seems little point in doing so.

But, with some knowledge of $F(x)$, we can seek some *deterministic* sequences, essentially for $\underline{w} = -\infty$ and $\bar{w} = +\infty$, that in some way 'measure' the type of increase (for maxima) and of decrease (for minima) of i.i.d. samples.

Although we could try to use a convenient sequence of increasing or decreasing functions, we will only consider the situations leading to an ALLN or MLLN. As stated, we stress that we are dealing only with an i.i.d. situation with distribution function $F(x)$, the increasing sample being denoted by $\{X_k, k = 1,2,...\}$.

We say that $\{X_k\}$ verifies the ALLN for maxima, or that $\max_1^k \{X_i\}$ verifies the ALLN, if there exist constants c_k (not uniquely defined, as said before) such that $\max_1^k \{X_i\} - c_k \xrightarrow{P} 0$. From the above we see that if $\bar{w} < +\infty$ we can take $c_k = \bar{w}$. Let us consider the case $\bar{w} = +\infty$. We have ALLN iff $F^k(c_k + \epsilon) \to 1$ and $F^k(c_k - \epsilon) \to 0$: this will be shown to be equivalent, when $\bar{w} = +\infty$, to $\frac{1 - F(y+x)}{1 - F(y)} \to 0$ if $x > 0$. From $F^k(c_k + \epsilon) \to 1$ or $k \log F(c_k + \epsilon) \to 0$ as $\frac{\log u}{u-1} \to 1$ when $u \to 1$ we get $k(1 - F(c_k + \epsilon)) \to 0$ and $F^k(c_k - \epsilon) \to 0$ is equivalent to $k(1 - F(c_k - \epsilon)) \to +\infty$. Note that we can always take c_k as non–decreasing, substituting c_k by $\max(c_1, ..., c_k)$: it is evident that if $F^k(c_k + \epsilon) \to 1$ also $F^k(\max(c_1, ..., c_k) + \epsilon) \to 1$. We need to prove, now, that $F^k(\max(c_1, ..., c_k) - \epsilon) \to 0$ also : as $F^k(c_k - \epsilon) \to 0$, for $k > N(\delta)$ ($\delta > 0$, fixed) we have $F^k(c_k - \epsilon) < \delta$. Take then $k > N(\delta)$ and we have

$$F^k(\max(c_1, \dots, c_k) - \epsilon) = \max_1^k (F^k(c_i - \epsilon)) = \max\left(\max_{i=1}^{N(\delta)} F^k(c_i - \epsilon), \max_{i=N(\delta)+1}^{k} F^k(c_i - \epsilon)\right).$$

But $F^k(c_i - \epsilon) \le F^i(c_i - \epsilon) < \delta$ if $i > N(\delta)$ and as $N(\delta)$ if fixed $\max_{i=1}^{N(\delta)} F^k(c_i - \epsilon) \to 0$

and so for $k > N'(\delta)$ we have $\max_{i=1}^{N(\delta)} F^k(c_i - \epsilon) < \delta$.

Consequently for $k \ge \max(N(\delta), N'(\delta))$ we have $F^k(\max(c_1, \dots, c_k) - \epsilon) < \delta$ and $F^k(\max(c_1, \dots, c_k) - \epsilon) \to 0$. We can thus suppose $c_k \le c_{k+1}$ and $c_k \to +\infty$. Situations like guaranteeing that for $k > N(\delta)$ some event or property happens will be described "as for large k" – its formalization is made as previously.

Then for y large we can obtain k such that $c_k \le y \le c_{k+1}$ and so we get $1 - F(c_k + \eta) \ge 1 - F(y + \eta)$ and $1 - F(c_{k+1} - \eta) \le 1 - F(y - \eta)$ so that

$$\frac{1 - F(y+\eta)}{1 - F(y - \eta)} \le \frac{1 - F(c_k + \eta)}{1 - F(c_{k+1} - \eta)}$$ and as $k(1 - F(c_k + \epsilon)) \to 0$ and $k(1 - F(c_k - \epsilon)) \to \infty$

we see that $\frac{1 - F(y + \eta)}{1 - F(y - \eta)} \to 0$ as $y \to \infty$.

Let us now prove the converse, also introducing one way of calculating the $\{c_k\}$. Let us define c_k such that $F(c_k^-) \le 1 - 1/k \le F(c_k)$ and so $c_k \to \infty$; $\{c_k\}$ is non–decreasing. From the condition $\frac{1 - F(y + x)}{1 - F(y)} \to 0$ as $y \to \infty$ or under the form $\frac{1 - F(y + \eta)}{1 - F(y - \eta)} \to 0$ as $y \to \infty$ (used above) we see that for $\epsilon > \eta > 0$ we have $k(1 - F(c_k + \epsilon)) \le \frac{1 - F(c_k + \epsilon)}{1 - F(c_k)} \to 0$ and so $k(1 - F(c_k + \epsilon)) \to 0$; also $k(1 - F(c_k - \epsilon)) \ge \frac{1 - F(c_k - \epsilon)}{1 - F(c_k^-)} \to +\infty$. Thus we have shown that when $\bar{w} = +\infty$ the ALLN is valid iff $\frac{1 - F(y + x)}{1 - F(y)} \to 0$ as $y \to \infty$ and that *one* system of coefficients $\{c_k\}$ is given by $F(c_k^-) \le 1 - 1/k \le F(c_k)$.

As a summary we have seen that :

The ALLN for i.i.d. maxima is valid either when $\bar{w} < +\infty$ *or* $\bar{w} = +\infty$ *iff*

$$\frac{1 - F(y + x)}{1 - F(y)} \to 0 \text{ if } y \to + \infty \text{ for any } x > 0.$$

It is now easy to obtain the MLLN. If $0 < \bar{w} < + \infty$ it is obvious that $\max_1^k X_i / \bar{w} \xrightarrow{P} 1$.

Consider then the case $\bar{w} = + \infty$. Then $\max_1^k X_i / c_k \xrightarrow{P} 1$, $(c_k > 0)$, as $\bar{w} = + \infty$, is equivalent to $\max_1^k X_i^+ / c_k \to + \infty$ where $X^+ = \max(0,X)$ and so equivalent to the ALLN for $\max_1^k \{\log X_i^+\} - \log c_k \xrightarrow{P} 0$.

Denoting by $F^+(x) = \text{Prob}\{\log X_i^+ \le x\}$, i.e., $F^+(x) = 0$ if $x < 0$, $F^+(x) = F(e^x)$ if $x \ge 0$ the conditions for the ALLN and for obtaining the $\{c_k\}$ for $\max_1^k \{\log X_i^+\}$ are $\frac{1 - F^+(y + x)}{1 - F^+(y)}$ $\to 0$ as $y \to \infty$ and $F^+(\log c_k^-) \le 1 - 1/k \le F^+(\log c_k)$ or, returning to $F(.)$, and substituting e^y and e^x by y and $x(>1)$, $\frac{1 - F(y \,.\, x)}{1 - F(y)} \to 0$ as $y \to \infty$ and $F(c_k^-) \le 1 - 1/k \le F(c_k)$.

As a summary we have seen that :

The MLLN for i.i.d. maxima is valid either when $0 < \bar{w} < + \infty$ *or when* $\bar{w} = + \infty$ *iff* $\frac{1 - F(y \,.\, x)}{1 - F(y)} \to 0$ *if* $y \to + \infty$ *for any* $x > 1$.

The conditions on $F(x)$ can be written as conditions on $S(x)$.

For minima we can obtain the ALLN and MLLN from the previous results using the fact that $\min_1^k \{X_i\} = - \max_1^k \{- X_i\}$, or directly.

Then if $\underline{w} > - \infty$ the ALLN is valid and also if $\underline{w} > 0$ the MLLN is valid. Consider now the case for $\underline{w} = - \infty$; the ALLN is valid iff $\frac{F(y)}{F(y+x)} \to 0$ as $y \to - \infty$ with $x > 0$; *one* system of coefficients is $\{c_k\}$ (non–increasing) given by $F(c_k^-) \le 1/k \le F(c_k)$; the MLLN has no meaning

if we continue to use $c_k > 0$; if we accept it as $c_k < 0$ then it is valid iff $\frac{F(y)}{F(y+x)} \to 0$ as $y \to -\infty$, $x > 1$; *one* system of $\{c_k\}$ $(c_k < 0)$ is given by $F(c_k^-) \le 1/k \le F(c_k)$.

The statements for the ALLN and MLLN for i.i.d. minima are the conversion of the ones for i.i.d. maxima with the substitution of $\bar{w}$ *by* $\underline{w}$ *and* $1 - F(x)$ *by* $F(-x)$.

The initial study connected with LLN is the one by Dodd (1923), using different terminology; the modern aspect appears, with unnecessary restrictions, in de Finneti (1932) and Gnedenko (1943); the essential aspects are contained in Galambos (1978) and Tiago de Oliveira ed. (1984).

Some examples concerning the LLN for extremes

Let us now give some examples that can clarify for maxima the border between the non–validity of the LLN, the validity of ALLN (and consequently of the MLLN if $\bar{w} > 0$) and the validity of MLLN.

1. Consider, as the first example, the uniform distribution in [a,b]. As $\bar{w} = b$ the ALLN for maxima is valid but the MLLN is valid only if $b > 0$; for minima as $\bar{w} = a$ the ALLN is valid and the MLLN is valid with $c_k = a$ if $a \neq 0$, if we accept the possibility of $c_k < 0$.

We will now consider cases where $\bar{w} = +\infty$ (and $\underline{w} = -\infty$).

2. Take as the second example the standard exponential distribution $E(x) = 0$ if $x \le 0$, $E(x) = 1 - e^{-x}$ if $x \ge 0$ ($\underline{w} = 0$, $\bar{w} = +\infty$). As $\frac{1 - E(y+x)}{1 - E(y)} = e^{-x} \neq 0$ the ALLN is not valid but the MLLN is valid because $\frac{1 - E(y \,.\, x)}{1 - E(y)} = e^{-(x-1)y} \to 0$ when $y \to +\infty$ with $x > 1$ (with $c_k = \log k$); for minima the ALLN is trivially valid as $\underline{w} = 0$ but the MLLN is not.

3. Consider now the standard normal distribution with $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$ ($\underline{w} = -\infty$, $\bar{w} = +\infty$), which is symmetrical and so we need to deal only with maxima.

We have

$$\frac{1-N(y+x)}{1-N(y)} = \frac{\int_{y+x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt}{\int_{y}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt} \sim \frac{e^{-(y+x)^2/2}}{e^{-y^2/2}} \to 0$$

and the ALLN is valid; also MLLN is valid because

$$\frac{1-N(y\,.\,x)}{1-N(y)} = \frac{\int_{y.x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt}{\int_{y}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt} \sim x\, \frac{e^{-y^2x^2/2}}{e^{-y^2/2}} \to 0$$

because $x > 1$; for $c_k = N^{-1}(1 - 1/k)$ we can use the easily obtained approximation $\tilde{c}_k = \sqrt{2 \log k}$ using $k(1 - N(\tilde{c}_k + \epsilon)) \to 0$, $k(1 - N(\tilde{c}_k - \epsilon)) \to +\infty$ and l'Hôpital's rule.

4. The geometric or Pascal distribution $F(x) = 0$ if $x < 0$, $F(x) = 1 - e^{-\theta[x]}$ for $x \geq 0$ (with $[x]$ denoting the integer part of x) is a discrete distribution with jumps at the integers $x = 0,1,2,...$, $\underline{w} = 0$, $\bar{w} = +\infty$ (sometimes written as Prob $\{X = m\} = (1 - p)\, p^{m-1}$, with $\theta = -\log p$).

Then $\frac{1 - F(y + x)}{1 - F(y)} = \frac{e^{-\theta[y+x]}}{e^{-\theta[y]}}$ which for $x = 1$ takes the constant value $e^{-\theta}$ and so the ALLN is not valid but $\frac{1 - F(y\,.\,x)}{1 - F(y)} = \frac{e^{-\theta[y\,.\,x]}}{e^{-\theta[y]}} \leq e^{-\theta\{(x-1)\,y-1\}} \to 0$ as $y \to \infty$ and so the MLLN is valid.

Let us consider three examples connected with the asymptotic distributions of extremes, to be dealt with in the next sections: the Gumbel distribution and the Fréchet distribution for maxima, and the Weibull distribution for minima. They seem to be, in applications, the most important ones.

5. For the reduced Gumbel distribution we have $\Lambda(x) = \exp(-e^{-x})$, $\underline{w} = -\infty$, $\bar{w} = +\infty$.

Then for maxima as $\frac{1 - \Lambda(y + x)}{1 - \Lambda(y)} \sim e^{-x}\, e^{-e^{-y}(e^{-x}-1)} \to e^{-x}$ and the ALLN is not valid; but the MLLN is valid because $\frac{1 - \Lambda(y\,x)}{1 - \Lambda(y)} \to 0$ when $y \to +\infty$ and $x > 1$. The

coefficients are $c_k = -\log\log\frac{k}{k-1}$ or $\tilde{c}_k = \log k$; this result could be expected because, for large x, $\Lambda(x)$ and $E(x)$ behave similarly. For minima we have $\frac{\Lambda(y)}{\Lambda(y+x)} \to 0$ as $y \to -\infty$, $x > 0$ and so ALLN is valid, a set of coefficients being $c_k = -\log\log k$; accepting $c_k < 0$ the MLLN is also true as $\frac{\Lambda(y)}{\Lambda(y \cdot x)} \to 0$ as $y \to -\infty$ if $x > 1$: the coefficients c_k (≤ 0) may be the same.

6. Consider now the reduced Fréchet distribution for maxima $\Phi_\alpha(x) = 0$ if $x \leq 0$, $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $\alpha > 0$, if $x \geq 0$ The ALLN is not valid as $\frac{1-\Phi_\alpha(y+x)}{1-\Phi_\alpha(y)} \to 1$ as $y \to +\infty$ ($x > 1$), nor is the MLLN as $\frac{1-\Phi_\alpha(y \cdot x)}{1-\Phi_\alpha(y)} \to x^{-\alpha}$ as $y \to +\infty$ ($x > 1$). For minima we have $\frac{\Phi_\alpha(y)}{\Phi_\alpha(y+x)} \to 0$ for $y \to 0$ and so the ALLN is valid, a system of coefficients being $c_k = (\log k)^{-1/\alpha}$; here the MLLN has direct meaning and is valid because we have $\frac{\Phi_\alpha(y)}{\Phi_\alpha(y \cdot x)} \to 0$ as $y \to 0^+$, $x > 1$, the coefficients being the same.

7. For the reduced Weibull distribution for minima $W_\alpha(x) = 0$ if $x \leq 0$, $W_\alpha(x) = 1 - \exp(-x^\alpha)$ if $x \geq 0$, $\alpha > 0$, we have $\frac{1-W_\alpha(y+x)}{1-W_\alpha(y)} = \exp\{y^\alpha - (y+x)^\alpha\} \to 0$ iff $\alpha > 1$, and so only if $\alpha > 1$ the ALLN is valid with coefficients $c_k = (\log k)^{1/\alpha}$; the MLLN is valid as $\frac{1-W_\alpha(y)}{1-W_\alpha(y \cdot x)} \to 0$ as $y \to \infty$ with $x > 1$, the coefficients being the same. For minima we get $\frac{W_\alpha(y)}{W_\alpha(y+x)} \to 0$ ($x > 0$), and so the ALLN is valid with coefficients $c_k = (-\log(1-1/k))^{1/\alpha}$ or more simply $c_k = k^{-1/\alpha}$; the MLLN is not valid as $\frac{W_\alpha(y)}{W_\alpha(y \cdot x)} \to x^{-\alpha}$ as $y \to 0^+$ ($x > 1$). Notice that $W_1(x) = E(x)$, the standard exponential distribution, and so the results could be expected.

Note that the non–validity of the MLLN in the last two cases is associated with the attraction conditions for the Fréchet and Weibull distribution of extremes, as will be seen in one of the next sections.

Although we have analysed — as an exercise — the ALLN and MLLN in each case, part of the conclusions could be reached if we used the relations between the LLN given in a previous section.

8. Consider now, finally, as a counter-example, the distribution function $F(x) = 0$ if $x \leq e$, $F(x) = 1 - 1/\log x$ if $x \geq e$; it has $\underline{w} = e$ and $\bar{w} = +\infty$.

For maxima as $\frac{1 - F(y + x)}{1 - F(y)} \to 1$ when $y \to +\infty$ $(x > 0)$ the ALLN is not valid as well as the MLLN because $\frac{1 - F(y \,.\, x)}{1 - F(y)} \to 1$ when $y \to +\infty$ $(x > 1)$. For minima as $\underline{w} = e$ the ALLN is trivially valid as well as the MLLN.

The asymptotic distributions of extremes – some examples

The results that are contained in this and the following sections constitute the initial core of *Statistical Extremes Theory* and are the basis for many applications. So we will proceed at a slower pace, with considerable independence from what was said about the LLN.

As has been said many times, the distribution of maxima and minima in a i.i.d. univariate sample $(X_1, \ldots, X_k)$ is given by

$$\text{Prob}\{\max_1^k \{X_i\} \leq x\} = F^k(x) \text{ and } \text{Prob}\{\min_1^k \{X_i\} \leq x\} = 1 - (1 - F(x))^k ;$$

for survival functions (which satisfy $F(x) + S(x) = 1$) we have

$$\text{Prob}\{\max_1^k \{X_i\} > x\} = 1 - (1 - S(x))^k \text{ and } \text{Prob}\{\min_1^k \{X_i\} > x\} = S^k(x).$$

From these formulae we see that $F(x)$ and $S(x)$ play symmetrical roles for maxima and minima.

In the case where $F(x)$ or $S(x)$ are known, we could proceed to a classical probabilistic and statistical analysis or, if $F(x)$ or $S(x)$ are known to be continuous, symmetrical, etc., in some cases we could have recourse to non–parametric methods. But, in many cases we do not even know if the observations come from some known parametric family of distributions and, in

general, we are not under the i.i.d. hypothesis. So we have to resort, in practical applications, to asymptotic results of the type: if for some $\{F_k(x)\}$ *attraction coefficients* $\{\lambda_k , \delta_k > 0\}$ exist such that $F_k(\lambda_k + \delta_k x) \xrightarrow{w} L(x)$, *with* $L(x)$ *continuous (and so uniform convergence)*, as $F_k(\lambda_k + \delta_k x)$ is close to $L(x)$ we will take $L\left(\frac{y - \lambda'}{\delta'}\right)$ as an approximation to $F_k(y)$, (λ', δ') not being necessarily (λ_k , δ_k) to allow for a better fit in statistical analysis. It must be said, at this preliminary stage, that the i.i.d conditions are not essential and can be weakened — as a rule the margins of the sequence $\{X_1,...,X_k,...\}$ should not be very different and the correlation/ association between X_i and X_j must wane out as the distance between i and j increases; two examples will shed some light on this question.

As the relation $\min_1^k\{X_i\} = -\max_1^k\{-X_i\}$ is true we can deal only with maxima or with minima and translate the results if necessary; we will deal, almost always, with maxima.

Thus, our purpose now is to obtain limiting (proper and non–degenerate) distributions of $F^k(\lambda_k + \delta_k x)$. When the asymptotic or limiting distributions were degenerate — Laws of Large Numbers — the three previous examples (5., 6. and 7.) correspond to the three possible limiting distributions and the last one to a case where there does not exist a limiting distribution, as we shall see later. Let us say, as will be shown in the next section, that if $\tilde{L}(x)$ *is a possible limiting* distribution for maxima then $\underset{\sim}{L}(x) = 1 - \tilde{L}(-x)$ is a possible limiting distribution for minima; but that the same distribution can't have the corresponding distributions $\underset{\sim}{L}$ and $\tilde{L}$, as seen in some examples below; also it can happen that one of $\underset{\sim}{L}$ or $\tilde{L}$ does not exist but the other one exists or, even, that both do not exist. Let us finally recall that (λ_k , δ_k) are not uniquely defined: by Khintchine's convergence of types theorem, (λ_k , δ_k) and (λ_k', δ_k'), such that $(\lambda_k' - \lambda_k)/\delta_k \to A$ and $\delta_k'/\delta_k \to B$ $(-\infty < A < +\infty, 0 < B < +\infty)$ as $k \to \infty$, lead to limiting distributions of the same type, respectively $L(x)$ and $L(A + Bx)$ (and conversely), the sets $\{(\lambda_k' , \delta_k')\}$ being thus equivalent for limiting purposes; in many cases we have $A = 0$ and $B = 1$, the *total* equivalence. For Khintchine's convergence of types theorem see the Annex II to Part I.

A calculation facility is convenient: as for maxima $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$ is equivalent to $k \log F(\lambda_k + \delta_k x) \to \log \tilde{L}(x)$, as $F(\lambda_k + \delta_k x) \to 1$ if $0 < \tilde{L}(x) < 1$, and $\frac{\log u}{u - 1} \to 1$ as $u \to 1$, the convergence relation $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$ is equivalent to $k(1 - F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$; also $1 - (1 - F(\lambda_k + \delta_k x))^k \xrightarrow{w} \underset{\sim}{L}(x)$ is equivalent, as $F(\lambda_k + \delta_k x)) \to 0$, to $k F(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ in the region $0 < \underset{\sim}{L}(x) < 1$. Those equivalent convergence relations (which are not uniform) $k(1 - F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$ and $k F(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ are very convenient for calculations; corresponding relations can be written for survival functions.

The three reduced limiting distributions for maxima $(L(x))$ are $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$ if $x \le 0$ with $\alpha > 0$, $\Psi_\alpha(x) = 1$ if $x \ge 0$ (Weibull distribution), $\Lambda(x) = \exp(-e^{-x})$ (Gumbel distribution) and $\Phi_\alpha(x) = 0$ if $x < 0$, $\Psi_\alpha(x) = \exp\{-x^{-\alpha}\}$ if $x \ge 0$ with $\alpha > 0$) (Fréchet distribution); the corresponding distributions for minima $\underset{\sim}{L}(x))$ are $W_\alpha(x) = 1 - \Psi_\alpha(-x)$, $1 - \Lambda(-x) = 1 - \exp\{-e^x\}$ and $1 - \Phi_\alpha(-x)$, with the same denominations. Notice that the following *stability equations* will be shown: $\tilde{L}^k(x) = L(\alpha_k + \beta_k x)$ and $\underset{\sim}{S}^k(x) = \underset{\sim}{S}(\alpha_k + \beta_k x)$ for convenient (α_k, β_k). Note that $(\lambda_{k+1} + \delta_{k+1})$ are, in both cases, also attraction coefficients by Khintchine's convergence of types theorem; this is left as exercise.

We will now give some examples that lead to some of the distributions above : note that each of $\tilde{L}(x)$ or of $\underset{\sim}{L}(x)$ is a limiting distribution for itself, as will be seen in some examples. All these examples will have a full justification in the next sections.

1. Take $F(x) = \Lambda(x)$. Then $F^k(\lambda_k + \delta_k x) = \Lambda^k(\lambda_k + \delta_k x) = \Lambda(\lambda_k + \delta_k x - \log k) = \Lambda(x)$ when $\lambda_k = \log k$ and $\delta_k = 1$ and so $\tilde{L}(x) = \Lambda(x)$; $\Lambda(x)$ is thus stable for maxima. Recall, once more, that equivalent coefficients could be used; in general, we will try the simpler and more manageable ones.

Suppose we are now dealing with minima of Gumbel distribution, we should have $k \Lambda(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ and it is easy to see that if $\lambda_k = -\log\log k$, $\delta_k = 1/\log k$ then $1 - (1 - F(\lambda_k + \delta_k x))^k \to 1 - \Lambda(-x)$. What is important is the behaviour of the right tail for maxima and the left tail for minima, and they can be very different which is not the case: note that

for $x > 0$ the ratio of the right tail to the left one is $(1 - \Lambda(x))/\Lambda(-x) \to 1$ as $x \to +\infty$; only when this ratio is 1 the distribution would be the pair $\tilde{L}(x)$ and $\underset{\sim}{L}(x) = 1 - \tilde{L}(x)$; if the limit of the ratio is $c(0 < c < +\infty)$ then $\tilde{L}$ and $\underset{\sim}{L}$ would be connected in the same way apart from a power transformation.

2. Consider the exponential distribution: $F(x) = E(x) = W_1(x)$ where $E(x) = 0$ if $x < 0$, $E(x) = 1 - e^{-x}$ if $x \geq 0$. As $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$ is equivalent to $k(1 - F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$, we must have $k e^{-\lambda_k - \delta_k x} \to -\log \tilde{L}(x)$ when $\lambda_k + \delta_k x > 0$; a solution is $\lambda_k = \log k$, $\delta_k = 1$ and $\tilde{L}(x) = \Lambda(x)$.

As far as minima are concerned, we should have $k(1 - e^{-\lambda_k - \delta_k x}) \to -\log(1 - \tilde{L}(x))$ for $\lambda_k + \delta_k x > 0$; taking $\lambda_k = 0$, $\delta_k = 1/k$ we get $\tilde{L}(x) = E(x) = W_1(x)$ and the exponential distribution is stable for minima.

3. Take now the Pareto distribution: $F(x) = 0$ if $x \leq 1$, $F(x) = 1 - x^{-\alpha}$ if $x \geq 1$, with $\alpha > 0$. To have $F^k(\lambda_k + \delta_k x) \to \tilde{L}(x)$, we must have $k(1 - F(\lambda_k + \delta_k x)) = k(\lambda_k + \delta_k x)^{-\alpha} \to -\log \tilde{L}(x)$ when $\lambda_k + \delta_k x \geq 1$; a solution is $\lambda_k = 0$, $\delta_k = k^{1/\alpha}$ and the limiting distribution is $\tilde{L}(x) = \Phi_\alpha(x)$.

For minima we must have $k\left(1 - \frac{1}{(\lambda_k + \delta_k x)^\alpha}\right) \to -\log(1 - \underset{\sim}{L}(x))$ when $\lambda_k + \delta_k x \geq 1$; if we take $\lambda_k = 1$, $\delta_k = 1/\alpha k$ we get $\underset{\sim}{L}(x) = E(x) = W_1(x)$, the exponential distribution, which is a Weibull form for minima and not a Fréchet one.

4. When we consider the uniform distribution in $[0, 1]$, $F(x) = 0$ if $x \leq 0$, $F(x) = x$ if $0 \leq x \leq 1$ and $F(x) = 1$ if $x \geq 1$, the condition is $k(1 - \lambda_k - \delta_k x) \to -\log(1 - \tilde{L}(x))$ when $0 \leq \lambda_k + \delta_k x \leq 1$; taking $\lambda_k = 1$ and $\delta_k = 1/k$ we get $\tilde{L}(x) = \Psi_1(x) = e^x$ if $x \leq 0$, $\Psi_1(x) = 1$ if $x \geq 0$, i.e., the Weibull distribution for maxima.

For minima we should have $k(\lambda_k + \delta_k x) \to -\log(1 - \underset{\sim}{L}(x))$ when $0 \leq \lambda_k + \delta_k x \leq 1$ which is obtained with $\lambda_k = 0$, $\delta_k = 1/k$ and $\underset{\sim}{L}(x) = W_1(x)$, the Weibull distribution for minima. Here situations are in correspondence because the uniform distribution is symmetrical about the mid–point 1/2.

We could study more cases of classical distributions but it does not seem necessary except for the normal, the logistic being a simple exercise. As seen, (λ_k , δ_k) were given with a hint about the way we could have arrived at them; the standard techniques will be explained in the next section.

5. Consider now the standard normal distribution function $N(x)$. It can be shown very easily, using the attraction conditions to be given later, that the possible limits of $N^k(\lambda_k + \delta_k x)$ cannot be either $\Psi_\alpha(x)$ or $\Phi_\alpha(x)$, Weibull and Fréchet distributions. We should thus have, if the limit exists, $N^k(\lambda_k + \delta_k x) \to \Lambda(x)$ or $k(1 - N(\lambda_k + \delta_k x)) \to e^{-x}$, where as known $N(\lambda_k + \delta_k x) \to 1$ and thus $\lambda_k + \delta_k x \to +\infty$.

A very well known result is that $1 - N(x) \sim \frac{N'(x)}{x}$ when $x \to +\infty$ and thus the previous condition is equivalent to

$$\log k + \log N'(\lambda_k + \delta_k x) - \log(\lambda_k + \delta_k x) + x \to 0 \quad \text{as} \quad k \to \infty$$

or

$$\log k - \log\sqrt{2\pi} - \frac{1}{2}(\lambda_k + \delta_k x)^2 - \log(\lambda_k + \delta_k x) + x = o(1);$$

dividing by $\log k$, we also see that $\frac{(\lambda_k + \delta_k x)^2}{2 \log k} \to 1$ as $\frac{\log(\lambda_k + \delta_k x)}{(\lambda_k + \delta_k x)^2} \to 0$ and thus $2\log(\lambda_k + \delta_k x) - \log 2 - \log\log k \to 0$ and so $\log(\lambda_k + \delta_k x) = \frac{1}{2}(\log 2 + \log\log k) + o(1)$. Substituting above $\log(\lambda_k + \delta_k x)$ we get

$$(\lambda_k + \delta_k x)^2 = 2\{\log k + x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log k\} + o(1) =$$

$$= 2\log k \times \{1 + \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log k}{\log k} + o(\frac{1}{\log k})\},$$

$$\lambda_k + \delta_k x = \sqrt{2\log k}\,(1 + \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log k}{2\log k} + o(\frac{1}{\log k}))$$

and thus we can take

$$\lambda_k = \sqrt{2\log k}\,(1 - \frac{\log 4\pi + \log\log k}{4\log k}) = \sqrt{2\log k} - \frac{\log 4\pi + \log\log k}{2\sqrt{2\log k}}$$

and $\delta_k = \dfrac{1}{\sqrt{2 \log k}}$.

The computation is easy, although tedious. But the verification is simple using $\sqrt{2 \log k} = t \to +\infty$.

6. Let us study the last distribution function considered in the previous section : $F(x) = 0$ if $x \le e$, $F(x) = 1 - \dfrac{1}{\log k}$ if $x \ge e$.

By the attraction conditions to be given later, it is immediately excluded that, for some (λ_k , δ_k) , $F^k (\lambda_k + \delta_k x)$ would converge to $\Psi_\alpha(x)$ or $\Phi_\alpha(x)$; the exclusion of $\Psi_\alpha(x)$ as a possible limit comes from the fact $\bar{w} = +\infty$; the other exclusion is almost as easy.

Let us show, directly, that $F^k (\lambda_k + \delta_k x)$ cannot converge (weakly) to $\Lambda(x)$ or, equivalently, that $k(1 - F (\lambda_k + \delta_k x))$ does not converge to $- \log \Lambda(x) = e^{-x}$. If it did converge we would have $F (\lambda_k + \delta_k x) \to 1$ and so it is sufficient to study the function in the right tail: we should thus have $\dfrac{k}{\log (\lambda_k + \delta_k x)} \to e^{-x}$ $(\lambda_k + \delta_k x > 1)$ or $\dfrac{\log (\lambda_k + \delta_k x)}{k} \to e^x$.

For $x = 0$ we have $\dfrac{\log \lambda_k}{k} \to 1$ and so $\lambda_k \to +\infty$ and $\lambda_k > 0$ ultimately.

Then we would have $\dfrac{\log (\lambda_k + \delta_k x)}{k} = \dfrac{\log \lambda_k + \log (1 + \delta_k / \lambda_k . x)}{k} \to e^x$ or $\dfrac{1}{k} \log (1 + \delta_k / \lambda_k . x) \to e^{-x} - 1$. Let a (≥ 0) be one of the possible limits of δ_k / λ_k : using the subsequence of $\{k\}$ such that $\delta_k / \lambda_k \to a$ if $a < +\infty$ we get $\dfrac{1}{k} \log (1 + \delta_k/\lambda_k . x) \to 0 \neq e^x - 1$; suppose now $a = +\infty$ and $x > 0$: then $\dfrac{\log (1 + \delta_k /\lambda_k . x)}{\log (\delta_k /\lambda_k . x)} \dfrac{\log (\delta_k /\lambda_k . x)}{k} \to e^x - 1$ is impossible. Thus if $F(x) = 0$ if $x \le e$, $F(x) = 1 - 1/\log x$ if $x \ge e$, it does not exist (λ_k , δ_k) such that $F^k(\lambda_k + \delta_k x)$ has a proper and nondegenerate limiting distribution. For minima we should have $k F (\lambda_k + \delta_k x) \to - \log (1 - \underset{\sim}{L} (x))$ or equivalently $k(1 - \dfrac{1}{\log (\lambda_k + \delta_k x)}) \to - \log (1 - \underset{\sim}{L}(x))$ and so $\dfrac{1}{\log (\lambda_k + \delta_k x)} \to 1$ or $\lambda_k + \delta_k x \to e$. Let us take $\lambda_k = e$ and $\delta_k = e/k$: we get $k(1 - \dfrac{1}{(\log e + ex/k)}) = \dfrac{k \log (1 + x/k)}{1 + \log (1 + x/k)} \to x = - \log (1 - \underset{\sim}{L} (x))$.

This shows that we have $\underset{\sim}{L}(x) = 1 - W_1(x) = E(x)$, which is the exponential distribution, an Weibull distribution for minima with $\alpha = 1$.

7. Let us now consider, finally, an example that shows how misleading some intuitive approximations can be. Consider the geometric or Pascal distribution $F(x) = 0$ if $x \leq 0$, $F(x) = 1 - e^{\theta[x]}$ if $x \geq 0$. We could expect that $F(x)$ would behave practically like the exponential distribution to which it is very similar, say a discretized version. As $\bar{w} = +\infty$ we can exclude the Weibull distribution as a limit; by other conditions below we can exclude the Fréchet distribution as a limit. We still have the Gumbel distribution as a possible limit which we could expect to be the one as happens with exponential distribution.

The condition $k(1 - F(\lambda_k + \delta_k x)) \to e^{-x}$ is equivalent, to $k\, e^{-\theta[\lambda_k + \delta_k x]} \to e^{-x}$ or $\theta[\lambda_k + \delta_k x] - \log k \to x$. If this happens we will show that $\theta\, \delta_k = 1$, defining uniquely δ_k, which is impossible by Khintchine's theorem. In fact, for $x = 0$ we also get $[\lambda_k] - \log k \to 0$ and so $\lambda_k \to +\infty$, as could be expected. By subtraction we get $\theta\{[\lambda_k + \delta_k x] - [\lambda_k]\} \to x$; denoting by $r(\alpha) = \alpha - [\alpha]$ the fractional part of α, the previous relation can be written $\theta\{\delta_k x - (r(\lambda_k + \delta_k x) - r(\lambda_k))\} \to x$ and with $x > 0$ (to simplify) we get $\theta\, \delta_k = \dfrac{r(\lambda_k + \delta_k x) - r(\lambda_k)}{x} \to 1$. As $|r(\alpha) - r(\beta)| < 1$, for k fixed, $\in (> 0)$ fixed and $k > N(\in)$ we have $1 - \in - \dfrac{1}{x} \leq \sigma\, \delta_k \leq 1 + \in + \dfrac{1}{x}$ and letting $x \to \infty$, we obtain finally $\delta_k = 1/\theta$, a *unique* result that is impossible so rejecting the Gumbel limit. But let us go further in the analysis. Accepting $\delta_k = 1/\theta$ we should have then $r(\lambda_k + x/\theta) - r(\lambda_k) \to 0$ as $k \to \infty$, $\forall\, x$. Suppose we fix x such that $0 < \varphi = x/\theta < 1$: we have $r(\lambda_k) = \tau_k$ $(0 \leq \tau < 1)$ and $r(\lambda_k + \varphi) = \tau_k + \varphi$ when $\tau_k + \varphi < 1$ and $\tau_k + \varphi - 1$ when $1 \leq \tau_k + \varphi\ (< 2)$ and so $r(\lambda_k + \varphi) - r(\lambda_k)$ has as possible limits φ and $\varphi - 1$, showing, once more, that a Pascal distribution cannot have, for maxima, a Gumbel distribution as a limit.

8. Also a Poisson distribution does not have a limiting distribution for maxima.

9. In principle all lattice distributions, i.e., discrete distributions with successive jumps at equal steps, can be dealt with by the use of a theorem of Gnedenko (1943) and in a simpler way. We have used those elementary proofs to clarify the issue from the start.

Let us now give two simple examples showing that i.i.d. conditions can be weakened.

Take a random i.i.d. sequence $\{X_i, i = 1,2, ...\}$ where each random variable has the distribution function F(x) and that $\max_1^k \{X_i\}$ has a limiting distribution $\underset{\sim}{L}(x)$ (as said before $\underset{\sim}{L}$ can only be Ψ_α , Λ, Φ_α apart from location-dispersion parameters), i.e., there exist attraction coefficients $(\lambda_k , \delta_k > 0)$ such that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$. Now define the new dependent sequence $\{X_i^* = \max (X_i , X_{i+1}), i = 1,2, ...\}$. We have

$$F_k^*(x) = \text{Prob}\{\max_1^k \{X_i^*\} \le x\} = \text{Prob}\{\max_1^{k+1} \{X_i\} \le x\} = F^{k+1}(x) \text{ and thus } \lambda_k^* = \lambda_{k+1}$$

$\delta_k^* = \delta_{k+1}$ is a system of attraction coefficients for the new (dependent) sequence leading to the same limiting distribution $\tilde{L}(x)$. We could also use the coefficients (λ_k , δ_k), by Khintchine's theorem as said before.

As concrete examples for $\{\max_1^k X_i^*\}$, consider the cases where F(x) is one of the three limiting distributions:

1. if $F(x) = \Psi_\alpha$ we have the stability relation $\Psi_\alpha^k(x) = \Psi_\alpha(k^{1/\alpha} x)$, $\Psi_\alpha^k(k^{-1/\alpha} x) = \Psi_\alpha(x)$; so we can take $\lambda_k = 0$ and $\delta_k = k^{-1/\alpha}$ and thus we can use $\lambda_k^* = \lambda_k = 0$ and $\delta_k^* = \delta_k = k^{-1/\alpha}$;

2. if $F(x) = \Lambda(x)$ we have the stability equation $\Lambda^k = \Lambda(x - \log k)$ or $\Lambda^k(\log k + x) = \Lambda(x)$; so we can take $\lambda_k = \log k$, $\delta_k = 1$ and thus we can use $\lambda_k^* = \lambda_k = \log k$ and $\delta_k^* = \delta_k = 1$;

3. if $F(x) = \Phi_\alpha(x)$ we have the stability equation $\Phi_\alpha^k(x) = \Phi_\alpha^k(k^{-1/\alpha} x)$ and so $\Phi_\alpha^k(k^{1/\alpha} x) = \Phi_\alpha^k(x)$; so we can take $\lambda_k = 0$, $\delta_k = k^{1/\alpha}$; and thus we can also use $\lambda_k^* = \lambda_k = 0$, $\delta_k^* = \delta_k$.

Let us note that this is a special case of 1–dependence as $X_i^* = \max(X_i, X_{i+1})$ and $X_{i+1}^* = \max(X_{i+1}, X_{i+1})$ are dependent but $X_i^* = \max(X_i, X_{i+1})$ and $X_j^* = \max(X_j^*, X_{j+1})$ are independent if $|j - i| > 1$; all the cases of m–dependence (X_i and X_j independent if $|i - j| > m$) lead to the same situation.

Let us now consider the weakening of the identical distributions condition. Consider a sequence $\{X_i\}$ of i.i.d. random variables with distribution function F(x) and such that

$$\text{Prob}\{(\max_1^k\{X_i\} - \lambda_k)/\delta_k \le x\} = F^k(\lambda_k + \delta_k x) \to \tilde{L}(x) \text{ when } k \to \infty .$$

Define the new independent, but not identically distributed, sequence $\{X_i^*\}$ by the equation

$$X_i^* = X_i + (-1)^i.$$

Let $F_k^*(x) = \text{Prob}\{(\max_1^k\{X_i^*\} \le x)$. We have for the distribution of $\max_1^k\{X_i^*\}$:

for $k = 2p$, $F_k^*(x) = \text{Prob}\{X_1 - 1 \le x, X_3 - 1 \le x, \ldots, X_{2p-1} - 1 \le x\}$ x

$$\text{x Prob}\{X_2 + 1 \le x, \ldots, X_{2p} + \le x\} = F^p(x + 1)\, F^p(x - 1);$$

for $k = 2p+1$ $\qquad F_{2p+1}^*(x) = F^{p+1}(x + 1)\, F^p(x - 1)$.

Do we have coefficients $(\lambda_k^*, \delta_k^*)$ such that $F_k^*(\lambda_k^* + \delta_k^* x)$ converges to a limiting distribution function, possibly to $\tilde{L}(x)$? This is a special case of a problem (periodic disturbances, dealt with in the last section) where we alternately add or subtract 1, and can be studied directly. We leave it to the reader.

For simplicity of the example — but where all the ingredients are present — we will study only the cases where $F(x) = \tilde{L}(x)$, i.e., where F(x) is $\Psi_\alpha(x)$, $\Lambda(x)$ or $\Phi_\alpha(x)$ and we intend to obtain $\tilde{L}(x)$ as a limit.

Consider the case where $F(x) = \Psi_\alpha(x)$. As is known, $\lambda_k = 0$ and $\delta_k = k^{1/\alpha}$, i.e., the stability equation takes the form $\Psi_\alpha^k(x) = \Psi_\alpha(k^{1/\alpha} x)$. We will compute, separately, $(\lambda^*_{2p}, \delta^*_{2p})$ and $(\lambda^*_{2p+1}, \delta^*_{2p+1})$ such that $F_k^*(\lambda_k^* + \delta_k^* x) \to \Psi_\alpha(x) = \exp\{(-\min(x,0))^\alpha\}$. Let us note that $\bar{w}^*_{2p} = 1$, $\bar{w}^*_{2p+1} = -1$ and so the right-end point of $\max_1^k X_i^*$ is 1 for $k > 1$. For $k = 2p$ if we take $\lambda_k^* = 1$ and $\delta_k^* = (2p)^{-1/\alpha}$ and if for $k = 2p+1$ we take the same expressions, we get $F_k^*(\lambda_k^* + \delta_k^* x) \to \Psi_\alpha(x)$. In general we have $\lambda_k^* = 1$, $\delta_k^* = [k/2]^{-1/\alpha}$. Another proof of the result can be given by showing that $F^k_{2p}(x)$ and $F^k_{2p+1}(x)$ are both equivalent for maxima to $W^p_\alpha(x)$; see the next chapter. This is left as an exercise as well as for the immediate examples.

When $F(x) = \Lambda(x)$ we have $\lambda_k = \log k$ and $\delta_k = 1$ as seen. The relation $F_k^*(\lambda_k^* + \delta_k^* x) \to \Lambda(x)$ gives a solution $\lambda_k^* = \log \frac{e + e^{-1}}{2} + \log k$, $\delta_k^* = 1$.

For Fréchet distribution , $F(x) = \Phi_\alpha(x)$ we have $\lambda_k^* = 0$ and $\delta_k^* = k^{1/\alpha}$ and also $\lambda_k^* = 0$ and $\delta_k^* = k^{1/\alpha}$.

In all cases, the coefficients can be the same owing to the special (limiting) form of $F(x)$ $(= \tilde{L}(x))$.

The asymptotic distributions of extremes

After these examples — where, for instance, logistic distribution, Cauchy distribution and other 'textbook' distribution have not been dealt with here for reasons to be seen shortly — we will proceed to obtain the limiting distributions of maxima from i.i.d. samples $\tilde{L}(x)$ which, as was said before, are $\Psi_\alpha(x)$, $\Lambda(x)$ and $\Phi_\alpha(x)$.

We say that a random sequence $\{X_i\}$ of i.i.d. random variables with distribution function *$F(x)$ is attracted for maxima to $\tilde{L}(x)$* if there exist coefficients $(\lambda_k, \delta_k > 0)$ — already called

attraction coefficients (for maxima) and not uniquely defined — such that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$.

As the limiting distributions are continuous the convergence is uniform, as is well known. We will say also that *F(x) is in the domain of attraction of* $\tilde{L}(x)$ which will be denoted by $F \in \mathcal{D}(\tilde{L})$. Correspondingly we will say that *F(x) is stable for maxima* (or *max–stable*) if $F^k(x) = F(\alpha_k + \beta_k x)$ for convenient α_k and $\beta_k > 0$. Evidently $\Psi_\alpha(x)$, $\Lambda(x)$ and $\Phi_\alpha(x)$ are max-stable with $\alpha_k = 0$, $\beta_k = k^{1/\alpha}$, $\alpha_k = -\log k$, $\beta_k = 1$ and $\alpha_k = 0$, $\beta_k = k^{-1/\alpha}$. It is immediate that if $F(x)$ is max–stable, $F(x)$ is a limiting distribution or is attracted to $F(x)$: in fact if we take $\lambda_k = -\alpha_k/\beta_k$ and $\delta_k = 1/\beta_k$ we obtain $F^k(\lambda_k + \delta_k x) = F(x)$ and the convergence is verified. The crux of the proof is to show, in reverse, that the limiting distributions are max–stable, to solve the corresponding functional equation, and to obtain one (or more) systems of attraction coefficients.

The conversion of these definitions for minima is immediate and can be done in terms of either $F(x)$ or of the survival function $S(x)$.

We will continue, out of habit, to use $F(x)$, the conversion to $S(x)$ being the reproduction of the previous definitions, but with the introduction of some additional notation which would not be used in the sequel; but this is an interesting exercise because it stresses the duality.

We say that *F(x) is attracted for minima to* $\underset{\sim}{L}(x)$ or *is in the domain of a attraction of* $\underset{\sim}{L}(x)$, denoted by $F \in \mathcal{D}(\underset{\sim}{L})$ (the symbol $\mathcal{D}$ does not introduce ambiguity because of the context and of the pair of dual symbols $\tilde{L}$ and $\underset{\sim}{L}$) if there exist *attraction coefficients* (for minima) $\{(\lambda_k, \delta_k), \delta_k > 0\}$ not uniquely defined, such that $1 - (1 - F(\lambda_k + \delta_k x))^k \xrightarrow{w} \underset{\sim}{L}(x)$; as said, and to be shown, the limiting $\underset{\sim}{L}(x)$ are $1 - \Psi_\alpha(-x)$, $1 - \Lambda(-x)$ and $1 - \Phi_\alpha(-x)$. The stability relation is either $(1 - F(x))^k = 1 - F(\alpha_k + \beta_k x)$ or, using survival functions, $S^k(x) = S(\alpha_k + \beta_k x)$. We say that F (or S) is *stable for minima* (or *min–stable*) if the last relation hold. There is no difference in the definitions, arising from $F(x) + S(x) = 1$. An exercise is to compute the expressions of (α_k, β_k) for min–stability. It is also evident that min–stable distribution (or survival functions) are also limiting ones. The result comes from the solution obtained for maxima.

Let us, then, study maxima. We will show that:

The limiting distributions of maxima are max–stable.

In fact, suppose that there exist $\{(\lambda_k, \delta_k), \delta_k > 0\}$ such that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$. Then we have $F^{mk}(\lambda_{mk} + \delta_{mk} x) \xrightarrow{w} \tilde{L}(x)$ also, and thus $F^k(\lambda_{mk} + \delta_{mk} x) \to \tilde{L}^{1/m}(x)$. Let

α'_m, $\beta'_m > 0$ be a limit of a subsequence of $(\lambda_{mk} - \lambda_k)/\delta_k \to \alpha'_m$, $\delta_{mk}/\delta_k \to \beta'_m$, as $k \to \infty$ through the subsequence; then by Khintchine's theorem we have $F^k(\lambda_{mk} + \delta_{mk} x) \to \tilde{L}(\alpha'_m + \beta'_m x) = \tilde{L}^{1/m}(x)$ or putting $x' = \alpha'_m + \beta'_m x$ we get the stability equation

$$\tilde{L}^m(x') = \tilde{L}\left(-\frac{\alpha'_m}{\beta'_m} + \frac{1}{\beta'_m} x\right) = \tilde{L}(\alpha_m + \beta_m x)$$

and the result is independent of the subsequence. Notice that (α_m, β_m) can be defined by $\alpha_m = -\lim(\lambda_{mk} - \lambda_k)/\delta_{mk}$ and $\beta_m = \lim \delta_k/\delta_{mk}$ when $k \to +\infty$.

Thus we now have to solve the stability functional equation. The first step is to extend for any positive real t the stability equation, i.e., that there exist $\{(\alpha(t), \beta(t)), \beta(t) > 0\}$ such that $\tilde{L}^t(x) = \tilde{L}(\alpha(t) + \beta(t) x)$.

When $k \to +\infty$ we have $F^{[kt]}(\lambda_{[kt]} + \delta_{[kt]} x) \xrightarrow{w} \tilde{L}(x)$ and as $[kt]/k \to t$ also, we have

$$[F^k(\lambda_{[kt]} + \delta_{[kt]} x)]^t \to \tilde{L}(x);$$

consequently $(\lambda_{[kt]} - \lambda_k)/\delta_k$ and $\delta_{[kt]}/\delta_k$ converge as $k \to +\infty$. Let us define $\alpha(t)$ and $\beta(t)$ as

$$\alpha(t) = -\lim(\lambda_{[kt]} - \lambda_k)/\delta_{[kt]} \quad \text{and} \quad \beta(t) = \lim \delta_k/\delta_{[kt]}:$$

we get, as desired, $\tilde{L}^t(x) = \tilde{L}(\alpha(t) + \beta(t) x)$, which shows the continuity of $\tilde{L}(x)$.

To solve this functional equation we must know that the functional equation $u(ts) = u(t)\,u(s)$ $(t, s > 0)$ has the solution $u(t) = 0$ or $u(t) = t^{\rho}$ (*).

Then the relation

$$\tilde{L}^{st}(x) = \tilde{L}(\alpha(st) + \beta(st)\,x) = (\tilde{L}^{s}(x))^{t} = \tilde{L}^{t}(\alpha(s) + \beta(s)\,x))$$

$$= \tilde{L}(\alpha(t) + \beta(t)\,(\alpha(s) + \beta(s)\,x))$$

which exchanging s and t is equal to $\tilde{L}(\alpha(s) + \beta(s)\,(\alpha(t) + \beta(t)\,x))$. This leads to the equation

$$\alpha(st) = \alpha(t) + \alpha(s)\,\beta(t) = \alpha(s) + \alpha(t)\,\beta(s)$$

$$\beta(st) = \beta(s)\,\beta(t)\,.$$

The second equation can have the solution $\beta(t) = 0$ or $\beta(t) = t^{\rho}$. But $\beta(t) = 0$ leads to $\tilde{L}^{t}(x) = \tilde{L}\,(\alpha(t))$ and should be disregarded.

We have, then, as $\beta(t) = t^{\rho}$ the relations $\alpha(t)\,(1 - s^{\rho}) = \alpha(s)\,(1 - t^{\rho})$.

(*) — This is true for $u(t)$ continuous, monotonic or measurable; we will prove it for the continuous case. If $u(t) = 0$ the theorem is proved. Suppose that there exists $t_o\,(> 0)$ such that $u(t_o) \neq 0$ and, thus, $u(1) = 1$. Then for $u(t_o^m) = u(t_o)\,u(t_o^{m-1})$ we get $u(t_o^m) = u(t_o)^m$ and as $u(t_o^{m/n}) = u(t_o^{1/n})^m = u(t_o)^{m/n}$ we see that for every rational $r > 0$ we have $u(t_o^r) = u(t_o)^r$. But the positive rationals are dense in the positive reals and as $u(t)$ is continuous, we get $u(t_o^s) = u(t_o)^s$. Putting $t_o^s = t$ we get $u(t) = u(t_o)^{\log t/\log t_o}$ and putting $u(t_o)^{1/\log t_o} = e^{\rho}$ we get $u(t) = t^{\rho}$ as could be expected.

And here, we can have two cases : $\rho = 0$ or $\rho \neq 0$.

If $\rho = 0$ we get $\alpha(st) = \alpha(s) + \alpha(t)$ and so $e^{\alpha(t)}$ satisfies the equation for u(t); then $\alpha(t) = -\theta \log t$ and the functional equation reads as $\tilde{L}^t(x) = \tilde{L}(-\theta \log t + x)$ whose solution is $\tilde{L}(y) = \exp\{-\exp-\{y/\theta - \log(-\log \tilde{L}(0))\}\}$ which is a Gumbel distribution when $\theta > 0$. For $\theta \leq 0$ the solution is not a distribution function.

Suppose now that $\rho \neq 0$. Then we have (as $t^\rho \neq 1$ for $t \neq 1$) we have (for $t, s \neq 1$) $\frac{\alpha(s)}{1-s^\rho} = \frac{\alpha(t)}{1-s^t} = c$ and so we have the equation $\tilde{L}^t(x) = \tilde{L}(c(1-t^\rho) + t^\rho x)$.

Consider that $\rho > 0$. For $t \to 0$ we get $1 = \tilde{L}(c)$ but for no $x_0 < c$ is $\tilde{L}(x_0) = 1$ because otherwise we would have $\tilde{L}(x_0) = 1 = \tilde{L}^t(x_0) = \tilde{L}(c + t^\rho(x_0 - c))$ which for $t^\rho = \frac{y-c}{x_0-c} > 0$ (implying $y < 0$) would lead to $\tilde{L}(x_0) = \tilde{L}(y) = 1$ for any y; this shows that c is the right–end point. Let us put $x - c = \varphi (\leq 0)$. We get $\tilde{L}^t(c+\varphi) = \tilde{L}(c+t^\rho \varphi)$. Put $\tilde{L}(c+\varphi) = e^{-h(\varphi)}$ where $h(\varphi)$ is continuous and non-increasing from $+\infty$ to 0 in $[-\infty, 0]$. We get, for $\varphi \leq 0$,

$$t\,h(\varphi) = h(t^\rho \varphi) \text{ or } h(\varphi) = h(-1)(-\varphi)^{1/\rho} \quad \text{and thus}$$

$$\tilde{L}(x) = 1 \quad \text{for} \quad x \geq c$$

$$\tilde{L}(x) = \tilde{L}(c + (x-c)) = e^{-h(x-c)} = e^{-h(-1)(c-x)^{1/\rho}}$$

which is a Weibull distribution with location parameter $c(X \leq c$ with probability 1), dispersion parameter $h(-1)^{-\rho}$, and shape parameter $1/\rho$.

If we are dealing with the case $\rho < 0$, when we take $t \to +\infty$ we get $0 = \tilde{L}(c)$ and so c is the left–end point of $\tilde{L}(x)$.

Put also $x - c = \varphi (\geq 0)$: the equation reads as $\tilde{L}(c+\varphi) = \tilde{L}(c+t^\rho \varphi)$. Now put $\tilde{L}(c+\varphi) = e^{-h(\varphi)}$. The equation gives $t\,h(\varphi) = h(t^\rho \varphi)$ and analogously $h(\varphi) = h(1)\,\varphi^{1/\rho}$ and thus

$$\tilde{L}(x) = 0 \quad \text{if} \quad x < c$$

$$\tilde{L}(x) = \tilde{L}(c + (x-c)) = e^{-h(x-c)} = e^{-h(1)(x-c)^{1/\rho}}$$

which is a Fréchet distribution with left–end point c, dispersion parameter $h(1)^{-\rho}$, and shape coefficient $1/\rho$. We have thus proved that the classes of max–stable and limiting distributions for

maxima coincide; the changes of the location–dispersion parameters are integrated in the type (or equivalence) changes.

Note that, by the proof given above connected with the stability equation, we have shown that $F^k(\lambda_k + \delta_k x) \xrightarrow{w} \tilde{L}(x)$, k (>0) integer, and $F^t(\lambda_t + \delta_t x) \xrightarrow{w} \tilde{L}(x)$, t (>0) real, are equivalent and this will be used sometimes later.

Notice that the $\tilde{L}$ are continuous and, consequently, *the convergence is uniform.*

For minima, let $F^*(x) = \text{Prob}\{-X \le x\} = \text{Prob}\{X \ge -x\} = S(-x)$ in the continuity set of S.

We have $\text{Prob}\{(\min_1^k \{X_i\} - \lambda_k)/\delta_k < x\} = \text{Prob}\{\min_1^k \{X_i\} < \lambda_k + \delta_k x\} =$

$$= \text{Prob}\{\max_1^k \{-X_i\} > -(\lambda_k + \delta_k x)\} = 1 - \text{Prob}\{\max_1^k \{-X_i\} \le -(\lambda_k + \delta_k x)\} =$$

$$= 1 - F^{*k}(-\lambda_k - \delta_k x) \text{ and if } F^{*k}(-\lambda_k + \delta_k y) \to \tilde{L}(y) \text{ we get}$$

$$\text{Prob}\{(\min_1^k \{X_i\} - \lambda_k)/\delta_k < x\} \to 1 - \tilde{L}(-x) \text{ as } \tilde{L}(y) \text{ is continuous.}$$

$\tilde{L}$ is thus associated with the limiting distribution, if it exists, of $F^*(x)$ and not of $F(x)$. The set of limiting distribution functions for minima is the one already described.

Another proof could be obtained using the survival functions, which is left as an exercise.

We have then proved the *Extremal Limit Theorem*:

The reduced asymptotic distributions of maxima $\tilde{L}(x)$, when they exist, are $\Psi_\alpha(x)$, $\Lambda(x)$ and $\Psi_\alpha(x)$ and the reduced asymptotic distribution of minima $\underset{\sim}{L}(x)$, when they exist are $1 - \Psi_\alpha(-x)$, $1 - \Lambda(-x)$ and $1 - \Psi_\alpha(-x)$.

Notice that the right–tail behaviour of $F^*(x)$ defining the maxima of $\{-X_i\}$ corresponds to the left–tail behaviour of $F(x)$ corresponding to the minima.

Consider a distribution function (proper and non–degenerate) and a sequence of real numbers $\{u_k\}$ such that $k(1 - F(u_k)) \to \tau$ or $k\,(1 - F(u_k)) = \tau + o\,(1) = \tau\,(1 + o(1))$; the u_k are, evidently, a function of τ and a functional of F.

It is immediate that, if $\{X_i\}$ are i.i.d. with distribution function F, then $= \text{Prob}\{\max_1^k \{X_i\} \leq u_k\} = F^k(u_k) = (1 - \frac{\tau(1+o(1))}{k})^k \to e^{-\tau}$ as $k \to \infty$; the inverse is also valid as seen by taking logarithms; the case $\tau = +\infty$ is dealt with analogously. The connection between the changing (non–decreasing in general) levels for overpassing and the previous results corresponds for instance to taking $u_k\,(\tau\,,\,x) = \lambda_k + \delta_k\,x$ and $\tau = e^{-x}$, etc.. Note that u_k does not necessarily exist; for $F(x) = 0$ if $x \leq 0$, $F(x) = x/2$ if $0 \leq x \leq 1$ and $F(x) = 1$ for $x \geq 1$ (a jump of 1/2 at $x = 1$) we see that for $\tau = 1$ we can take neither $u_k \geq 1$, because we would get $k(1 - F(u_k)) = 0$, nor $0 \leq u_k < 1$, because for $k(1 - u_k/2) \to 1$ we should have taken $u_k = 2 - \frac{2 + o(1)}{k}$. Also if $F(x) = 1 - e^{-[x]}$ for $x \geq 0$, $F(x) = 0$ for $x < 0$, it is easy to see that for no $\{u_k\}$ we can obtain $k(1 - F(u_k)) \to \tau$, as $\log k - [u_k]$ is oscillating. But for $F(x) = 0$ if $x \leq e$, $1 - 1/\log x$ if $x \geq e$, we get $u_k(\tau) = e^{k/\tau}$.

The classical theory already developed used $u_k = \lambda_k + \delta_k\,x$ with τ depending on x, giving the desired results when $\tau = (-x)^{\alpha}$ $(x < 0)$, $\tau = e^{-x}$ or $\tau = x^{-\alpha}$ $(x > 0)$. In some cases we can transform $u_k(\tau)$ into an equivalent linear function of x (x and τ related as above), but other cases exist, like the last example in the previous paragraph, where such a transformation for a linear function cannot be made.

Finally let us speak of a general form that integrates in one expression the three limiting distributions of maxima, called von Mises-Jenkinson formula

$$G(z|\theta) = \exp\{-(1 + \theta\,z)_+^{-1/\theta}\}\,,\ \theta \text{ real, i.e.,}$$

$$G(z\,|\,\theta) = \exp\{-(1 + \theta\,z)_-^{-1/\theta}\}\,,\ \theta \text{ real,}$$

where $1 + \theta\,z \geq 0$ with the natural truncation when $1 + \theta\,z < 0$: if $\theta > 0$, for $z < -1/\theta$ we have $G(z|\theta) = 0$ and if $\theta < 0$ for $z > -1/\theta$ we have $G(z|\theta) = 1$.

If $\theta < 0$ we have immediately $G(-(1 + x)/\theta|\theta) = \Psi_{-1/\theta}(x)$ where $\alpha = -1/\theta > 0$; if $\theta > 0$ we have $G(-(1 - x)/\theta|\theta) = \Phi_{-1/\theta}(x)$ where $\alpha = 1/\theta > 0$ and finally we have $G(z|0^+) = G(z|0^-) = 1(z)$.

The graphics of the densities of G'(z|θ), for maxima, are given in Fig.II.1. They will be given for each form in the chapters V, VI and VII (for minima)

Fig. II.1 — Graphs of the densities G'(z|θ)

The graphs of the densities for minima G'(–z|θ) are the miror images of the previous ones.

It should be noted that when $\theta \to 0$ the graphs of G'(z|θ) converge to G'(z|0) = Λ'(z) ; this corresponds to the fact that for large $\alpha = 1/\theta$ Fréchet and Weibull distributions are very close to the Gumbel distribution.

This integrated formula will be useful for the statistical choice of models in chapter VIII.

The basic texts are Fisher and Tippet (1928), Fréchet (1927), de Finetti (1932), Gumbel (1935), von Mises (1936) and Gnedenko (1943); exposés can be found in Galambos (1978) and de Hann (1976).

It is of interest to consider the joint behaviour of $(\min_1^k\{X_i\}, \max_1^k\{X_i\})$. Its joint distribution function is

$$\text{Prob}\{\min_1^k \{X_i\} \le x, \max_1^k \{X_i\} \le y) = \text{Prob}\{\min_1^k \{X_i\} \le \min(x, y)\},$$

$$\max_1^k \{X_i\} \le y\} = F^k(y) - (F(y) - F(\min(x, y))^k .$$

Supposing that

$$\text{Prob}\{\min_1^k \{X_i\} \le \lambda_k' + \delta_k' x\} = 1 - (1 - F(\lambda_k' + \delta_k' x))^k \to \tilde{L}(x)$$

and

$$\text{Prob}\{\max_{1}^{k}\{X_i\} \le \lambda_k + \delta_k\, y\} = F^k(\lambda_k + \delta_k\, y)) \to \tilde{L}(y)$$

we get

$$\text{Prob}\{\min_{1}^{k}\{X_i\} \le \lambda_k^{'} + \delta_k^{'}\, x,\ \max_{1}^{k}\{X_i\} \le \lambda_k + \delta_k\, y\} =$$

$$= F^k(\lambda_k + \delta_k\, y) - \{F(\lambda_k + \delta_k\, y) - F\,(\min(\lambda_k^{'} + \delta_k\, x, \lambda_k + \delta_k\, y))\}^k .$$

The first summand converges to $\tilde{L}$ (y). Let us thus study the second one which can be written as

$$F^k\,(\lambda_k + \delta_k\, y)\,\{1 - \frac{F\,(\min\,(\lambda_k^{'} + \delta_k^{'}\, x,\ \lambda_k + \delta_k\, y))}{F\,(\lambda_k + \delta_k\, y)}\}^k$$

As the first factor converges to $\tilde{L}(y)$, denote by $\varphi\,(x\,,\,y)$ the limit of the second factor. We have, if $\varphi\,(x\,,\,y)$ exists,

$$\text{Prob}\{\min_{1}^{k}\{X_i\} \le \lambda_k^{'} + \delta_k^{'}\, x,\ \min_{1}^{k}\{X_i\} \le \lambda_k + \delta_k\, y\} \to \tilde{L}(y)$$

$$\text{But}\quad \varphi\,(x\,,\,y) = \lim\,(1 - \frac{F\,(\min\,(\lambda_k^{'} + \delta_k^{'}\, x,\ \lambda_k + \delta_k\, y))}{F\,(\lambda_k + \delta_k\, y)})^k$$

$$= \exp\,\{-\lim\,\frac{k\,F(\min\,(\lambda_k^{'} + \delta_k^{'}\, x,\ \lambda_k + \delta_k\, y))}{F\,(\lambda_k + \delta_k\, y)}\}$$

As said before $F(\lambda_k + \delta_k\, y) \to 1$ for y in the support of $\tilde{L}(y)$; on the other hand we know that for x and y in the supports of $\underset{\sim}{L}(x)$ and $\tilde{L}\,(y)$, $\lambda_k^{'} + \delta_k^{'}\, x \to \underline{w}$ and $\lambda_k + \delta_k\, y \to \bar{w}$

and for large k we have $\lambda_k^{'} + \delta_k^{'}\, x < \lambda_k + \delta_k\, y$ and, thus, we get

$$\varphi(x, y) = \exp\{-\lim k\, F(\lambda_k' + \delta_k' x)\} = \exp\{-[-\log(1 - \underset{\sim}{L}(x))] = 1 - \underset{\sim}{L}(x)$$

and so

$$\text{Prob}\{\min_1^k \{X_i\} \le \lambda_k' + \delta_k' x,\ \max_1^k \{X_i\} \le \lambda_k + \delta_k y\} \to \underset{\sim}{L}(x)\ \tilde{L}(y).$$

Consequently the reduced extremes are asymptotically independent (in the sense defined in the proof). For more details on asymptotic independence of sample extremes see Geffroy (1958/59).

From asymptotic independence, it is evident that the distributions of the reduced range, centre, and absolute maximum depend only on the prevailing reduced extremes. If they are of the corresponding asymptotic form their distributions are, for large samples, those of $Y - X$, $\frac{1}{2}(Y + X)$ and $\max(-X, Y)$.

Similarly it was proved by Rosengard (1962, 1966), Tiago de Oliveira (1961), and Rossberg (1963), (1965), that the sample extremes and mean, under very general conditions, are asymptotically independent.

The asymptotic distributions of the m–th extremes

The asymptotic theory of m–th extremes was, essentially, initiated in a paper by Gumbel (1935).

Consider a sequence of i.i.d. random variables $\{X_i\}$ with distribution function $F(x)$ and suppose that

$$\text{Prob}\{\max_1^k (X_i - \lambda_k)/\delta_k \le x\} = F^k(\lambda_k + \delta_k x) \to \tilde{L}(x).$$

Let us see what happens to the m–th order maximum $X'_{k-m+1} = X''_m$ in the notation of chapter I.

We have $F_{m,k}(x) = \text{Prob}\{X''_m \le x\} = \text{Prob}\{\text{at least } k–m+1 \text{ of the } X_i \le x\} =$

$$\sum_{k-m+1}^{k} \binom{k}{j} F(x)^j (1 - F(x))^{k-j} = \sum_{i=o}^{m-1} \binom{k}{k-i} F^{k-i}(x) (1 - F(x))^i$$

and so

$$F_{m,k}(\lambda_k + \delta_k x) \to \tilde{L}_m(x) = \lim \sum_{i=o}^{m-1} \binom{k}{k-i} F^{k-i}(\lambda_k + \delta_k x) (1 - F(\lambda_k + \delta_k x))^i$$

$$= \tilde{L}(x) \sum_{i=o}^{m-1} \frac{(-\log \tilde{L}(x))^i}{i!} \quad \text{as } F^{k-i}(\lambda_k + \delta_k x) \to \tilde{L}(x) \qquad \text{and}$$

$$\binom{k}{k-i} (1 - F(\lambda_k + \delta_k x))^i = \binom{k}{i} (1 - F(\lambda_k + \delta_k x))^i$$

$$= \frac{k(k-1)\ldots(k - i + 1) (1 - F(\lambda_k + \delta_k x))^i}{i!} \sim [k(1 - F(\lambda_k + \delta_k x))]^i / i!$$

and as seen before $k(1 - F(\lambda_k + \delta_k x)) \to -\log \tilde{L}(x)$ as $k \to +\infty$. Evidently $\tilde{L}_1(x) = \tilde{L}(x)$.

Analogously we can obtain the distribution of the m–th minimum: suppose that $1 - (1 - F(\lambda_k + \delta_k x))^k \to -\log \underset{\sim}{L}(x)$ or equivalently that $k F(\lambda_k + \delta_k x)) \to -\log (1 - \underset{\sim}{L}(x))$.

Then

$$F^*_{m,k}(x) = \text{Prob}\{X'_m \le x\} = 1 - \text{Prob}\{X'_m > x\} = 1 - \sum_{j=o}^{m-1} \binom{k}{j} F^j(x) (1 - F(x))^{k-j}$$

and consequently

$$F^*_{m,k}(\lambda_k + \delta_k x) = 1 - \sum_{j=o}^{m-1} \binom{k}{j} F^j (\lambda_k + \delta_k x) (1 - F(\lambda_k + \delta_k x))^{k-j} .$$

But $(1 - F(\lambda_k + \delta_k x))^{k-j} \to 1 - \underset{\sim}{L}(x)$ and

$$\binom{k}{j} F^j (\lambda_k + \delta_k x) = \frac{k(k-1) \ldots (k - i + 1)}{j!} F^j (\lambda_k + \delta_k x) \sim \frac{(k F(\lambda_k + \delta_k x))^j}{j!} \to$$

$$\rightarrow \left(\frac{-\log\,(1-\underset{\sim}{L}(x))}{j\,!}\right)^{j} \qquad \text{and we get}$$

$$\underset{\sim}{L}_m(x) = 1 - (1-\underset{\sim}{L}_m(x)) \sum_{j=0}^{m-1} \left(\frac{-\log\,(1-\underset{\sim}{L}(x))}{j\,!}\right)^{j} .$$

Using the survival function $\underset{\sim}{S}(x) = 1 - \underset{\sim}{L}(x)$ we get

$$\underset{\sim}{S}_m(x) = \underset{\sim}{S}(x) \sum_{j=0}^{m-1} \left(\frac{-\log \underset{\sim}{S}(x)}{j\,!}\right)^{j},$$

dual of the expression for $\underset{\sim}{L}_m(x)$; evidently $\underset{\sim}{S}_1(x) = \underset{\sim}{S}(x)$.

To compute $\underset{\sim}{L}_m(x)$ and $\underset{\sim}{S}_m(x)$ from the three possible expression of $\underset{\sim}{L}_1(x)$ and $\underset{\sim}{S}_1(x)$ is a simple exercise.

In what has been said we have always supposed that we were dealing with the m–th maximum or m–th minimum, with m fixed. We can even allow m to be a function m(k) such that $m(k)/k \rightarrow 0$ as $k \rightarrow +\infty$ and obtain similar results: see Smirnov (1952) as the first step, and Mejzler (1984) for the general overview.

We will not develop this points further because it does not seem to be particularly interesting.

Let us now recall that maxima in between themselves and minima in between themselves are associated; but, as shown the maximum and the minimum of a sample are asymptotically independent; a m_1–th maximum and a m_2–minimum are also asymptotically independent by a simple extension of the proof.

An overview of some extensions

Here we will sketch, briefly, without proof, results extending the three limiting distributions under conditions which are not i.i.d., for which we have given already some examples.

If the identically distributed condition is relaxed, but independence is maintained, the initial papers of Juncosa (1948) and Mejzler (1949), detailed in Mejzler (1984), have shown that under general reasonable conditions, the three limiting distributions are still valid. But this is not completely general: as a counter-example, take any distribution function F(x) and consider the

sequence of independent random variables $\{X_i\}$ with distribution functions $F_i(x) = (F(x))^{2^{-i}}$. It is immediate that

$$\text{Prob}\{\max(X_1, \ldots, X_k) \le x\} = \prod_1^k F_i(x) = (F(x))^{1-2^{-k}} \to F(x)$$

as $k \to \infty$, with a system of attraction coefficients $\lambda_k = 0$, $\delta_k = 1$. Thus any distribution function can be the limiting distribution function of $\max(X_1, \ldots, X_k)$ and so some moderate condition about the set of $\{F_i(.)\}$ must be introduced to maintain the classical limiting distribution of maxima (and minima, obviously). See the above–mentioned papers for more details, and in the case where $F_i(x) = F(x + \theta_i)$ with $\theta_i = \theta_{i+P}$ (periodic disturbances of period P), see Tiago de Oliveira (1976).

Also if the independence condition is lifted, but i.d. margins are maintained, the asymptotic distribution of maxima may be different from the three given forms. As a counter example take the sequence $Y_0, Y_1, \ldots, Y_i, \ldots$ i.i.d. with a standard normal distribution function and define

$$X_i = \sqrt{\rho}\, Y_0 + \sqrt{1-\rho}\; Y_i \quad (i = 1,2, \ldots)\,, \rho > 0.$$

The X_i are multinormal with standard margins and (constant) correlation coefficient ρ (> 0). Then

$$\max_1^k X_i = \sqrt{\rho}\, Y_0 + \sqrt{1-\rho}\; \max_1^k Y_i$$

and, as by the ALLN we have $\max_1^k X_i - \sqrt{2 \log k} \xrightarrow{P} 0$, we see that

$$\frac{\max_1^k X_i - \sqrt{1-\rho}\;\sqrt{2 \log k}}{\sqrt{\rho}} - Y_0 \xrightarrow{P} 0$$

and so

$$\text{Prob}\{\max_1^k X_i \le \sqrt{1-\rho}\;\sqrt{2 \log k} + \sqrt{\rho}\; x\} \to N(x).$$

This example suggests that the correlation in dependent sequences must be waning out at a reasonable rate to obtain the classical limiting forms for maxima.

The case where we deal with a random (dependent) sequence with the same marginal distributions (F(.)) — i.e., stationarity of order 1 — and where the sequence of maxima is attracted by the same distribution as the i.i.d. sequence with the same distributions F(.), has been initially considered by Watson (1954) and Newell (1964) for the case of m–dependence, i.e., (X_i, X_j) independent if $|i - j| > m$, and later by others — this point will be developed in part IV; see references therein.

The application to Meteorology and Oceanography of m–dependence is obvious as after a few days (2 or 3) there is practical independence for the observations.

For a relaxing of both i. and i.d. conditions, but having yet one of the limiting distributions, see Tiago de Oliveira (1978).

Chapter III

ATTRACTION STRUCTURE AND SPEED OF CONVERGENCE

Introduction

From the utilitarian standpoint this chapter has very little importance: it deals with the i.i.d. case, where the underlying distribution F(x) is known, and we want to know if the sequence of random variables $\{X_k\}$ is attracted for maxima, to one of the three limiting distributions, how to obtain *some* set of attraction coefficients and, finally, how quick is the convergence.

In applications, observations are often not i.i.d., but under some conditions, for large samples, the limiting distributions of maxima of a sequence of non-i.i.d. observations are still the same as in the i.i.d. case. But even if they were independent , F(x) would not be known (in general) in order to analyse if the maxima are attracted to Ψ_α , Λ or Φ_α and thus to obtain a set (not necessarily the best in any sense) of attraction coefficients (λ_k, δ_k) . The results here, from a utilitarian standpoint – immediate utility – are of little importance. The analysis of the speed of convergence follows the same lines.

Moreover, the practical importance is not nil; on the contrary. The last section (extensions) of the previous chapter says that the limiting distributions $\tilde{L}(x)$ for maxima (or $\underset{\sim}{L}(x)$ for minima) are valid for outside the i.i.d. case and in many cases (say for instance the discharges of a river) the total sample can be divided into "chunks" (say the hydrological year) where the maximum (the flood) has a distribution function, at least approximately stable, close to one of the $\tilde{L}((x-\lambda)/\delta)$ which will then be used to study maxima, thus avoiding the "computable" values (λ_k, δ_k) and in some cases α, substituted by its statistical estimation. As in oceanographical and hydrological situations data a few days apart (sometimes less than two days) behave in an independent – like way, the knowledge of the speed of convergence gives an idea how the approximation $\tilde{L}((x-\lambda)/\delta)$ is valid.

In consequence, so far *through analogy*, this chapter, for the i.i.d. case plus the previous extensions section, illustrates the use of asymptotic distributions in dealing with a dependent, finite set of data. And this is its practical, though not utilitarian, justification.

The attraction conditions and attraction coefficients

Let there be a sequence of i.i.d. random variables $\{X_i\}$ with distribution function F(x).

We will prove the following statements for $F \in \mathcal{D}(\tilde{L})$; the dual statements of $F \in \mathcal{D}(\underset{\sim}{L})$ can be obtained by a simple conversion of the previous results. The quantile function $Q(v) = \inf\{x|F(x) \geq v\}$; Q(v) is very important in this study and, by the results given here, also in the problems of the right tail estimation using the largest values of a sample (Annex V to Part II).

The statements for $F \in \mathcal{D}(\tilde{L})$, taking $t > 0$, are:

– F is attracted, for maxima, to the Weibull distribution $\Psi_\alpha(x)$ *iff* $\bar{w} < +\infty$ *and* $\frac{1 - F(\bar{w} - tx)}{1 - F(\bar{w} - t)} \to x^\alpha$ *as* $t \to 0^+$; *a system* (λ_k, δ_k) *of attraction coefficients is* $(\bar{w} - Q(1-1/k))$.

– F is attracted, for maxima, to the Gumbel distribution $\Lambda(x)$ *iff*

$$\frac{Q(1 - tx) - Q(1 - t)}{Q(1 - te) - Q(1 - t)} \to \log x \text{ as } t \to 0^+, \text{ or}$$

$$\frac{Q(1 - tx) - Q(1 - t)}{Q(1 - ty) - Q(1 - t)} \to \frac{\log x}{\log y} \text{ as } t \to 0^+ \text{ if } y \neq 1, \text{ or}$$

$$\frac{Q(1 - (1 - v)x) - Q(v)}{Q(1 - (1 - v)e) - Q(v)} \to \log x \text{ as } v \to 1^-, \text{ or, also}$$

$$\frac{Q(v^x) - Q(v)}{Q(v^e) - Q(v)} \to \log x \text{ as } v \to 1^-;$$

a system of attraction coefficients is $\lambda_k = Q(1 - 1/k)$, $\delta_k = Q(1 - 1/ek) - Q(1 - 1/k)$.

– F is attracted, for maxima, to the Fréchet distribution $\Phi_\alpha(x)$ *iff* $\bar{w} = +\infty$ *and* $\frac{1 - F(t)}{1 - F(tx)} \to x^\alpha$ *as* $t \to +\infty$; *a system of attraction coefficients is* $(\lambda_k, \delta_k) = (0, Q(1-1/k))$ *(ultimately positive as* $\bar{w} = +\infty$*).*

Analogously to the attraction condition, for maxima, to Gumbel distribution given in terms of the quantile function we have also

– F is attracted, for maxima, to Fréchet distribution $\Phi_\alpha(x)$ *iff*

$$\frac{Q\,(1-tx)-Q\,(1-t)}{Q(\,1-t\,y)-Q\,(1-t)} \to \frac{x^{-1/\alpha}-1}{y^{-1/\alpha}-1} \quad as\ t \to 0^+ \quad if \quad y \neq 1$$

and to Weibull distribution $\Psi_\alpha(x)$ *iff*

$$\frac{Q\,(1-tx)-Q\,(1-t)}{Q(\,1-t\,y)-Q\,(1-t)} \to \frac{x^{-1/\alpha}-1}{y^{-1/\alpha}-1} \quad as\ t \to 0^+ \quad if\ y \neq 1\,.$$

If we do use the integrated von Mises-Jenkinson form we can say that $F(.)$ *is attracted, for maxima, to* $G(.|\theta)$ *iff* $\frac{Q\,(1-tx)-Q\,(1-t)}{Q(1-t\,y)-Q\,(1-t)} \to \frac{x^{-\theta}-1}{y^{-\theta}-1}$ *(to* $\frac{\log x}{\log y}$ *if* $\theta = 0$*) as* $t \to 0^+$ *if* $y \neq 1$.

Notes :

1 . There are other conditions for the attraction to the Gumbel distribution, but these seem to be the most operational;

2 . A function SV(x) is said to be of *slow variation* (or *slowly varying*) at $+\infty$ if it is defined at least in a right half-line (i.e., for all $x > x_0$) and $SV(tx)/SV(x) \to 1$ as $x \to +\infty$; the latter nomenclature can be used to say that F(x) is attracted to the Weibull distribution if $\bar{w} < +\infty$ and $SV(x) = x^\alpha(1 - F(\bar{w} - 1/x))$ is slowly varying at $+\infty$ and that F(x) is attracted to the Fréchet distribution if $\bar{w} = +\infty$ and $SV(x) = x^\alpha(1 - F(x))$ is slowly varying at $+\infty$;

3 . To obtain the Fréchet distribution as a limit we must have $\bar{w} = +\infty$ and to obtain the Weibull distribution as a limit we must have $\bar{w} < +\infty$; the Gumbel distribution can be attained either with $\bar{w} < +\infty$ or $\bar{w} = +\infty$;

4 . The conditions for Weibull and Fréchet distributions to be limiting distributions for maxima are easy to work out; this is not the case for the Gumbel distribution;

5 . At the end of the section we will give conditions using F'(x), when it exists, which are in general very handy to use.

6 . Dually we obtain the statements for $F \in \mathcal{D}(\underset{\sim}{L})$, supposing, also, $t > 0$. They are :

$-F$ *is attracted, for minima, to the Weibull distribution* $W_\alpha(x) = 1 - \Psi_\alpha(-x)$ *iff* $\underline{w} = -\infty$ *and* $\frac{F(\underline{w} + t\,x)}{1 - F(\underline{w} + x)} \to t^\alpha$ *as* $x \to 0^+$; *a system of attraction coefficients is* $(\underline{w}\,,\ Q(1/k) - \underline{w})$.

– *F is attracted, for minima, to the Gumbel distribution* $1 - \Lambda(-x)$ *iff*

$$\frac{Q(t\,x) - Q(t)}{Q(e\,t) - Q(t)} \to \log x \quad as \quad x \to 0^+ \;, \; or \; \frac{Q(1 - v^x) - Q(1 - v)}{Q(1 - v^e) - Q(1 - v)} \to \log x \; as \; v \to 1^- \;;$$

a system of attraction coefficients is $(Q(1/k), Q(1/k) - Q(1/ek))$.

– *F is attracted, for minima, to the Fréchet distribution* $1 - \Phi_\alpha(-x)$ *iff* $\underline{w} = -\infty$ *and* $\frac{F(t)}{F(t\,x)} \to x^\alpha$ *as* $t \to -\infty$ *; a system of attraction coefficients is* $(0\,, -Q(1/k))$ $(-Q(1/k)$ *is ultimately positive as* $\underline{w} = -\infty)$.

7 . Before going further, let us recall that each of the limiting distributions is attracted to itself; this is shown immediately using the above criteria. Notice also that the conditions on $\bar{w}$ and $\underline{w}$ can help to eliminate one case for each $F(x)$; we used this in the previous examples.

Proofs :

We will follow a different order in the proofs for convenience. Let us consider the necessary and sufficient condition for attraction of maxima to $\Phi_\alpha(x)$.

Suppose $\bar{w} = +\infty$ and $\frac{1 - F(t)}{1 - F(t\,x)} \to x^\alpha$ as $t \to +\infty$. To prove the result we have to show that for some $\{\lambda_k\,, \delta_k\}$, $F^k(\lambda_k + \delta_k\,x) \to \Phi_\alpha(x)$ or, equivalently as said, that $k(1 - F(\lambda_k + \delta_k\,x)) \to x^{-\alpha}$ if $x > 0$, and $k(1 - F(\lambda_k + \delta_k\,x)) \to +\infty$ if $x < 0$.

Let us take $\lambda_k = 0$ and $\delta_k = Q(1 - 1/k)$. As $\bar{w} = +\infty$, we have $\delta_k \to +\infty$ and for $x < 0$, $\delta_k\,x \to -\infty$, implying $k(1 - F(\delta_k\,x)) \to +\infty$ for $x < 0$.

Consider now $k(1 - F(\delta_k\,x))$ for $x > 0$. As $\lim k(1 - F(\delta_k\,x)) = \lim\{k(1 - F(\delta_k))$ $\frac{1 - F(\delta_k\,x)}{1 - F(\delta_k)}\}$ and by hypothesis when $k \to +\infty$, $\delta_k \to +\infty$ the second factor $\frac{1 - F(\delta_k\,x)}{1 - F(\delta_k)}$ $\to x^{-\alpha}$; it remains to show that $k(1 - F(\delta_k)) \to 1$. By the definition of δ_k one gets $F(\delta_k^-) \leq$ $1 - 1/k = F(\delta_k)$ and so $k(1 - F(\delta_k)) \leq 1$; but $F(\delta_k\,x) \leq F(\delta_k^-) \leq 1 - 1/k$ for $0 < x < 1$ and so $\frac{1 - F(\delta_k)}{1 - F(\delta_k\,x)} \leq k(1 - F(\delta_k))$ and $\frac{1 - F(\delta_k)}{1 - F(\delta_k\,x)} \to x^\alpha$ and thus $k(1 - F(\delta_k)) \to 1$ with $k \to \infty$, as desired. We have shown up to now that if the attraction conditions for $\Phi_\alpha(x)$ are valid we can take as attraction coefficients $\lambda_k = 0$, $\delta_k = Q(1 - 1/k)$ obtaining as limit $\Phi_\alpha(x)$.

Let us now prove the converse. As $F^k(\lambda_k + \delta_k x) \to \Phi_\alpha(x)$ or, equivalently, $k(1 - F(\lambda_k + \delta_k x)) \to x^{-\alpha}$ we have also

$$[k\beta]\ (1 - F(\lambda_k + \delta_k x)) \to \beta\ x^{-\alpha},$$

and thus, with $x = \beta^{1/\alpha}\ z$,

$$[k\beta]\ (1 - F(\lambda_k + \delta_k\ \beta^{1/\alpha}\ z) \to\ z^{-\alpha},$$

from which, with

$$[k\beta]\ (1 - F(\lambda_{[k\beta]} + \delta_{[k\beta]}\ z) \to\ z^{-\alpha},$$

by Khintchine's convergence of types theorem we get $(\lambda_{[k\beta]} - \lambda_k)/\delta_k \to 0$ and $\delta_{[k\beta]}/\delta_k \to \beta^{1/\alpha}$.

Take now, for fixed $\beta > 1$, $\lambda_{[k\beta]} = \lambda_k$, $\delta_{[k\beta]} = \delta_k \beta^{1/\alpha}$. Define, now, a integer sequence $\{k(s)\}$ by $k(1) = [k\beta]$, $k(s+1) = [k(s) \cdot \beta]$, and so we get $\lambda_{k(1)} = \lambda_k$, $\delta_{k(s)} = \delta_k \beta^{(s-1)/\alpha}$ $(\to +\infty$ as $\beta > 1)$. Then $\lambda_{k(1)}/\delta_{k(s)} \to 0$ and we can write, as $\lambda_{k(s)}$ can be taken to be zero,

$$F^{k(s)}\ (\delta_{k(s)}\ x) \to \Phi_\alpha(x).$$

Let us fix, now, x, choose y — to increase indefinitely — and obtain s such that $\delta_{k(s)}\ x \le y \le \delta_{k(s+1)}\ x$. Then we have

$$1 - F(\delta_{k(s+1)}\ x) \le\ 1 - F(y) \le 1 - F(\delta_{k(s)}\ x)$$

and so

$$\frac{1 - F(\delta_{k(s+1)}\ x)}{1 - F(\delta_{k(s)}\ t\ x)} \le \frac{1 - F\ (y)}{1 - F\ (t\ y)} \le \frac{1 - F(\delta_{k(s)}\ x)}{1 - F(\delta_{k(s+1)}\ t\ x)}$$

and as

$$\frac{k\ (s+1)}{k(s)} = \frac{\beta\ k(s) - r}{k(s)} \to \beta$$

$(0 \le r < 1$ is the fractional part) and as $k(s)\ (1 - F(\delta_{k(s)}\ x)) \to x^{-\alpha}$ as $k \to +\infty$ we get, finally letting $y \to +\infty$ and so $k(s) \to +\infty$,

$$\frac{1}{\beta}\ t^\alpha \le \lim_{y \to \infty} \frac{1 - F\ (y)}{1 - F\ (y\ t)} \le \beta\ t^\alpha$$

and thus, as $\beta > 1$, as close to 1 as wished,

$$\frac{1 - F(y)}{1 - F(y\,t)} \to t^{\alpha} \text{ as } y \to \infty .$$

We have shown, also, that as attraction coefficients we can take $(0, Q(1-1/k))$ in choosing δ_k such that $k(1 - F(\delta_k)) \to 1$.

The conversion of this result concerning $\Phi_\alpha(x)$ to $\Psi_\alpha(x)$ is very easy. As seen, $\bar{F}^k(\lambda_k + \delta_k x) \to \Phi_\alpha(x)$ is equivalent to $\bar{F}^k(\delta_k x) \to \Phi_\alpha(x)$; in addition, if we define $\bar{F}(x) = F(\bar{w} - 1/x)$ we see that with $\bar{\delta}_k = Q(1 - 1/k)$, $\bar{F}^k(\bar{\delta}_k x) \to \Phi_\alpha(x)$ iff $F^k(\bar{w} - 1/\bar{\delta}_k x) \to \Phi_\alpha(x)$ or $F^k(\bar{w} - 1/\bar{\delta}_k x) \to \Phi_\alpha(-1/x) = \Psi_\alpha(x)$.

Consequently for F to be attracted, for maxima, to $\Psi_\alpha(x)$ we must have $\bar{w} < +\infty$, satisfy the attraction condition $\frac{1 - F(\bar{w} - t\,x)}{1 - F(\bar{w} - t)} \to x^{\alpha}$ as $t \to 0^+$, and take as attraction coefficients $\lambda_k = \bar{w}$, $\delta_k = \bar{\delta}_k^{-1}$, or, more simply, $\delta_k = \bar{w} - Q(1 - 1/k)$.

Let us now go to the proof of the attraction condition for maxima, to the Gumbel distribution.

We have from $F^k(\lambda_k + \delta_k x) \to \Lambda(x)$ also $F^{[k\,t]}(\lambda_{[k\,t]} + \delta_{[k\,t]} x) \to \Lambda(x)$ and thus $(F^k(\lambda_{[k\,t]} + \delta_{[k\,t]} x))^{[k\,t]/k} \to \Lambda(x)$ or $F^k(\lambda_{[k\,t]} + \delta_{[k\,t]} x) \to \Lambda^{1/t}(x) = \Lambda(x + \log t)$ and so $F^k(\lambda_{[k\,t]} + \delta_{[k\,t]}(x - \log t)) \to \Lambda(x)$. By the Khintchine's convergence of types theorem we get

$$\frac{\lambda_{[k\,t]} - \delta_{[k\,t]} \log t - \lambda_k}{\delta_k} \to 0$$

and

$$\delta_{[k\,t]} / \delta_k \to 1,$$

or equivalently $(\lambda_{[k\,t]} - \lambda_k)/\delta_k \sim \log t$ and $\delta_{[k\,t]} \sim \delta_k$.

Taking $t = e$ we get $\delta_k \sim \lambda_{[k\,e]} - \lambda_k$ and so the conditions are

$$\frac{\lambda_{[k\,t]} - \lambda_k}{\lambda_{[k\,e]} - \lambda_k} \to \log t \text{ and } \delta_k \sim \lambda_{[k\,e]} - \lambda_k ,$$

the last evidently defining δ_k, asymptotically.

Thus if $F \in \mathcal{D}(\Lambda)$, with attraction coefficients (λ_k , δ_k), we know that

$$\frac{\lambda_{[k\,t]} - \lambda_k}{\delta_k} \sim \frac{\lambda_{[k\,t]} - \lambda_k}{\lambda_{[k\,e]} - \lambda_k} \to \log t \quad \text{with} \quad \delta_k \sim \lambda_{[k\,e]} - \lambda_k \ ,$$

The relation $k(1 - F(\delta_k + \lambda_k x)) \to e^{-x}$ suggests the use of $\lambda_k^* = Q(1 - 1/k)$, $\delta_k^* = \lambda_{[ke]}^* - \lambda_k^*$, with Q the quantile function. Let us show that if $(\lambda_{[k\,t]}^* - \lambda_k^*)/\delta_k^* \to \log t$ then $F \in \mathcal{D}(\Lambda)$ and by the Khintchine's convergence of types theorem any (λ_k, δ_k) such that $F^k(\lambda_k + \delta_k x) \to \Lambda(x)$ and $(\lambda_k^*, \delta_k^*)$ are equivalent for the limit. For fixed t, from

$$\frac{\lambda_{[k\,t]}^* - \lambda_k^*}{\delta_k^*} \to \log t \quad \text{we get, for large } k,$$

$$\lambda_k^* + \delta_k^* (\log t - \epsilon) < \lambda_k^* = Q\left(1 - \frac{1}{[k\,t]}\right) < \lambda_k^* + \delta_k^* (\log t + \epsilon) \ ,$$

and so $F(\lambda_k^* + \delta_k^* (\log t - \epsilon)) \le F(Q(1 - 1/[k\,t])) \le F(\lambda_k^* + \delta_k^* (\log t + \epsilon))$.

From the RHS inequality, as $v \le F(Q(v))$, we get

$$1 - 1/[k\,t] \le F(\lambda_k^* + \delta_k^* (\log t + \epsilon)) \ ;$$

raising to power k and letting $k \to \infty$ we get

$$e^{-1/t} \le \underline{\lim}\ F^k(\lambda_k^* + \delta_k^* (\log t + \epsilon))$$

and so $$\exp\{-e^{-(z-\epsilon)}\} \le \underline{\lim}\ F^k(\lambda_k^* + \delta_k^* z) \ .$$

For the LHS inequality, as $\lambda_k^* + \delta_k^* (\log t + \epsilon) < \lambda_k^*$ we get, from $F(Q(v)^-) \le v$,

$$F(\lambda_k^* + \delta_k^* (\log t - \epsilon)) \le 1 - 1/[k\,t] \ ;$$

raising to power k and letting $k \to +\infty$ we get

$$\overline{\lim}\ F^k(\lambda_k^* + \delta_k^* (\log t - \epsilon)) \le e^{-1/t}$$

and in the same way

$$\overline{\lim}\ F^k(\lambda_k^* + \delta^*\ z) \le \exp\{-e^{-(z+\epsilon)}\}\ .$$

Consequently

$$\Lambda(z-\epsilon) \le \underline{\lim}\ F^k(\lambda_k^* + \delta_k^*\ z) \le \overline{\lim}\ F^k(\lambda_k^* + \delta_k^*\ z) \le \Lambda(z+\epsilon)$$

and thus

$$F^k(\lambda_k^* + \delta_k^*\ z) \to \Lambda(z)\ .,$$

and any system of attraction coefficients is equivalent to $(\lambda_k^*\ ,\ \delta_k^*)$.

As a consequence we see that $F \in \mathcal{D}(\Lambda)$ iff

$$\frac{Q(1 - 1/[k\ x]) - Q(1 - 1/k)}{Q(1 - 1/[k\ e]) - Q(1 - 1/k)} \to \log x \text{ as } k \to \infty$$

and we can use as attraction coefficients $\lambda_k^* = Q(1 - 1/k)$ and $\delta_k^* = \lambda_{[k\,e]}^* - \lambda_k^*$

Let us now give a continuous form to this discrete result. We will show that the condition above is equivalent to

$$\frac{Q(1 - t\ x) - Q(1 - t)}{Q(1 - t\ e) - Q(1 - t)} \to \log x \text{ as } t \to 0^+.$$

Let $k = [1/k] + 1$; we will show that

$$\frac{Q(1 - t\ x) - Q(1 - t)}{\delta_k^*} \to -\log x \text{ as } t \to 0^+\ ;$$

by division we obtain

$$\frac{Q(1 - t\ x) - Q(1 - t)}{Q(1 - t\ e) - Q(1 - t)} \to \log x \text{ as } t \to 0^+.$$

From $1/t < k \le 1/t + 1$ we get $1/k < t \le 1/(k-1)$, $x/k < t\ x \le x/(k-1)$ and so $[(k/x) + 1]^{-1} < t\ x < [(k-1)/x]^{-1}$ and as $Q(1-\xi)$ is non-increasing in ξ we have

$$\frac{\lambda^*_{[(k-1)/x]} - \lambda^*_{k+1}}{\delta^*_k} \leq \frac{Q(1 - t\,x) - Q(1 - t)}{\delta^*_k} \leq \frac{\lambda^*_{[(k/x)+1]} - \lambda^*_{k-1}}{\delta^*_k} .$$

But

$$\frac{\lambda^*_{[(k/x)+1]} - \lambda^*_{k-1}}{\delta^*_k} = \frac{\lambda^*_{[k/x]+1} - \lambda^*_{[k/x]}}{\delta^*_k} + \frac{\lambda^*_{[k/x]} - \lambda^*_k}{\delta^*_k} + \frac{\lambda^*_k - \lambda^*_{k-1}}{\delta^*_k} .$$

The last summand converges to zero and the second to $\log(1/x) = -\log x$. The first summand can be written as $\dfrac{\lambda^*_{[k/x]+1} - \lambda^*_{[k/x]}}{\delta^*_{[k/x]}} \cdot \dfrac{\delta^*_{[k/x]}}{\delta^*_k}$. The first factor converges to zero and the second to 1 as seen in the beginning of the proof because $\delta_{[kx]}/\delta_k \to 1$ and so $\delta^*_{[k/x]}/\delta^*_k \to 1$.

Thus the RHS of the inequality converges to $-\log x$ and the same can be proved for the LHS. Thus $\dfrac{Q(1 - t\,x) - Q(1 - t)}{\delta^*_k} \to -\log x$ and so the continuous condition for attraction is

$$\frac{Q(1 - t\,x) - Q(1 - t)}{Q(1 - t\,e) - Q(1 - t)} \to \log x \quad \text{as} \quad t \to 0^+ .$$

This condition was given by Mejzler (1949) and was also stated by Marcus and Pinsky (1969) independently in another form; other important texts are de Haan (1970), (1971) and Balkema and de Haan (1972).

The proof of the results concerning atraction to $\Phi_\alpha(x)$, $\Psi_\alpha(x)$ and $G(z \mid \theta)$ in terms of the quantile function runs in the same lines.

Sufficiency conditions and examples

Before going on to some examples, some old and revisited and some new, let us state only sufficiency conditions — which are almost necessary — for attraction, but supposing the existence of the derivative $F'(x)$, at least in a right half-line (i.e., for $x > x_0$). The proofs are simple and can be found in Gnedenko (1943) and de Haan (1976).

— *If* $\bar{w} < +\infty$, $F(x)$ *has positive density* $F'(x)$ *in some interval* $]x_0, \bar{w}[$, *and* $\lim_{x \uparrow \bar{w}} \frac{(\bar{w} - x) F'(x)}{1 - F(x)} = \alpha$, *then* $F(x)$ *is attracted to* $\Psi_\alpha(x)$ *and we can use the coefficients* $\lambda_k = \bar{w}$, $\delta_k = \bar{w} - Q(1 - 1/k)$, *as* $F(x)$ *defines uniquely* $Q(1 - 1/k)$ $(\to \bar{w})$ *for large* k;

— *If* $F''(x) < 0$ *for some interval* $]x_0, \bar{w}[$, $(\bar{w} < +\infty$ *or* $\bar{w} = +\infty)$ *and* $\frac{d}{dx}(\mu(x)^{-1}) \to 0$, *where* $\mu(x)$ *is the force of mortality or hazard rate, or* $\frac{F''(x)(1 - F(x))}{(F'(x))^2} \to -1$ *as* $x \uparrow \bar{w}$ *then* $F(x)$ *is attracted to* $\Lambda(x)$ *and we can use the coefficients* $\lambda_k = Q(1 - 1/k)$, *and* $\delta_k = 1/k\, F'(\lambda_k)$ *(von Mises criterion)*;

— *If* $\bar{w} = +\infty$, $F(x)$ *has a positive density for* $x > x_0$, *and* $\frac{x F'(x)}{1 - F(x)} \to \alpha$, *as* $x \to +\infty$, *then* $F(x)$ *is attracted to* $\Phi_\alpha(x)$ *and we can use the attraction coefficients* $\lambda_k = 0$, $\delta_k = Q(1 - 1/k)$.

These maxima results, in the same way as those following, can be converted to minima results.

But also:

— *If* $F(x)$ *is attracted for maxima to* $\Lambda(x)$ *then* $\delta_k / \lambda_k \to 0$;

— *If* $\bar{w} = +\infty$, *and* $F(x)$ *is attracted for maxima to* $\Lambda(x)$ *then* $\frac{1 - F(tx)}{1 - F(t)} \to 0$ *as* $t \to +\infty$, *with* $t > 0$; *this result implies the MLLN;*

— *If* $\ldots x_1 < x_2 < x_k < \ldots \bar{w}$ $(\leq +\infty)$ *are jump points of* $F(x)$ *and* $\frac{1 - F(x_k^-)}{1 - F(x_k)} \geq 1 + \beta$, $\beta > 0$, *then* $F(x)$ *is not attracted for maxima to* $\Lambda(x)$;

— *If* $w^- = +\infty$, $F^k(\lambda_k + \delta x) \to \Lambda(x)$ $(\delta_k = \delta$ *constant) is equivalent to*

$\frac{1 - F(\log t)}{1 - F(\log t\, x)} \to x^\alpha (\alpha = 1/\delta)$ *as* $t \to \infty$ *or* $\frac{1 - F(y)}{1 - F(y + k)} \to e^{k/\delta}$ *as* $y \to +\infty$;

the proof of the last statement is easily reduced to the Fréchet limiting situation by the use of the transform $Y = \log X$ and some adjustments.

Let us now go to the miscellaneous examples. We will not obtain the attraction coefficients, but will only decide on attraction.

1. Let us consider the classical normal distribution. As $\bar{w} = +\infty$ it cannot be attracted to $\Psi_\alpha(x)$. But $\frac{1 - N(t)}{1 - N(t\,x)} \sim x^{-1}\, e^{t^2(x^2-1)/2}$ which does not converge to x^α and so $\Phi_\alpha(x)$ is excluded. The Mejzler criterion for the convergence to $\Lambda(x)$ in this example is not very operational. Let us consider the von Mises criterion $\frac{N''(x)\,(1 - N(x))}{(N'(x))^2} \to -1$ as $x \to +\infty$. But $1 - N(x) \sim N'(x)/x$; as used before, and $N''(x) = -x\,N'(x)$ and so $\frac{N''(x)\,(1 - N(x))}{(N'(x))^2} \to -1$, and $N(x)$ is attracted, for maxima, to $\Lambda(x)$.

2. Consider the geometric distribution $F(x) = 1 - e^{-\lambda[x]}$ if $x \geq 0$, $F(x) = 0$ if $x < 0$. The jump points are the non-negative integers and we have $\frac{1 - F(k^-)}{1 - F(k)} = \frac{1 - F(k-1)}{1 - F(k)} = e^\lambda > 1$ and so $F(x)$ is not attracted to $\Lambda(x)$, as opposed to the exponential. As $\bar{w} = +\infty$ the limit could be $\Phi_\alpha(x)$ if $\frac{1 - F(t)}{1 - F(t\,x)} \to x^\alpha$ as $t \to +\infty$. But $\frac{1 - F(t)}{1 - F(t\,x)} = e^{\lambda\{[tx]-[t]\}} = e^{\lambda\{[tx-r]-[t-t']\}}$, with $0 \leq r, r' < 1$, which does not converge to x^α, and so the distribution is not attracted for maxima to $\Phi_\alpha(x)$.

3. Let us now analyse the Poisson distribution. We have $F(x) = \sum_{j=o}^{[x]} e^{-\lambda} \frac{\lambda^j}{j!} = \sum_{o}^{[x]} p_j$ with jumps at the integers $j = 0,1, \ldots,$. Then $\frac{1 - F(j^-)}{1 - F(j)} = \frac{1 - F(j-1)}{1 - F(j)} = 1 + \frac{p_j}{1 - F(j)} \geq$

$$1 + \frac{\lambda^j/j\,!}{\sum_{k=j}^{[\infty]} \lambda^k/k\,!} = 1 + \frac{1}{\sum_{o}^{[\infty]} \frac{j\,!}{(r+j)\,!}\lambda^r} = 1 + \frac{1}{\sum_{o}^{[\infty]} \binom{x+j-1}{r} \frac{\lambda^r}{r!}} > 1 + \frac{1}{\sum_{o}^{[\infty]} \lambda^r/r!} = 1 + e^{-\lambda}$$

and Poisson distribution is not attracted for maxima to Gumbel distribution; the non-attraction to Fréchet distribution is left as an exercise.

4. If we have $F(x)$ such that there exist α and $\beta > 0$ for which $e^{\alpha+\beta x}(1 - F(x)) \to 1$ as $x \to +\infty$ $(\bar{w} = +\infty)$, then $F(x)$ is attracted for maxima to $\Lambda(x)$ as seen from either that $\frac{1 - F(y)}{1 - F(y+t)} \to e^{\beta t}$ and so $\delta_k = \beta^{-1}$ with λ_k to be determined. By observing that $e^{\alpha+\beta(\lambda_k+\delta_k x)}(1 - F(\lambda_k + \delta_k x)) \to 1$ and also that we should have $k(1 - F(\lambda_k + \delta_k x)) \to e^{-x}$ which implies $e^{\alpha+\beta(\lambda_k+\delta_k x)}/k \to e^x$ or $e^{\alpha+\beta\lambda_k}/k \to 1$ and $e^{\beta\delta_k x} \to e^x$ giving thus $\lambda_k = \frac{\log k - \alpha}{\beta}$, $\delta_k = \frac{1}{\beta}$ as a system of attractions coefficients.

5. Consider, finally, the lognormal distribution $LN(x) = N(\log x)$ for $x \geq 0$, $LN(x) = 0$ for $x \leq 0$. As $\bar{w} = +\infty$ we can try the attraction for maxima to $\Lambda(x)$ or $\Phi_\alpha(x)$. Let us use the von Mises criterion:

$$\frac{LN''(x)\,(1 - LN(x))}{(LN'(x))^2} = \rightarrow -1$$ as happened before for the normal tail.

Tail equivalent distributions

The subject of equivalence of distributions is relevant, like the use of attraction conditions, to evaluate what can be the asymptotic behaviour of maxima (or minima) of some distribution by substituting it with another one which is easier to manipulate.

We say that $F(x)$ and $F_1(x)$ are *tail equivalent for maxima* if $\bar{w}(F) = \bar{w}(F_1) = \bar{w}$ and $\frac{1 - F(x)}{1 - F_1(x)} \rightarrow 1$ as $x \uparrow \bar{w}$, and *tail equivalent for minima* if $\underline{w}(F) = \underline{w}(F_1) = \underline{w}$ and $\frac{F(x)}{F_1(x)} \rightarrow 1$ as $x \downarrow \underline{w}$; we could, instead of the limit 1, put a constant c ($0 < c < +\infty$), thus substituting the equality of limits for a power relation using c. This is left as an exercise.

Let us deal with maxima.

If $F(x)$ and $F_1(x)$ are tail equivalent for maxima, then if one of the distributions is attracted to $\tilde{L}(x)$ the other distribution is also attracted to the same limit $\tilde{L}(x)$ and with the same coefficients.

Suppose that $F^k(\lambda_k + \delta_k x) \rightarrow \tilde{L}(x)$ or $k(1 - F(\lambda_k + \delta_k x)) \rightarrow -\log \tilde{L}(x)$; as $\lambda_k + \delta_k x \rightarrow \bar{w}$ we have $k(1 - F_1(\lambda_k + \delta_k x)) \frac{1 - F(\lambda_k + \delta_k x)}{1 - F_1(\lambda_k + \delta_k x)} \rightarrow -\log \tilde{L}(x)$ and by the condition of equivalence $k(1 - F_1(\lambda_k + \delta_k x)) \rightarrow -\log \tilde{L}(x)$ or $F_1^k(\lambda_k + \delta_k x) \rightarrow \tilde{L}(x)$ as said. Let us prove the converse: if $F^k(\lambda_k + \delta_k x) \rightarrow \tilde{L}(x)$ and $F_1^k(\lambda_k + \delta_k x) \rightarrow \tilde{L}(x)$ (the same coefficients) we have $\frac{1 - F(\lambda_k + \delta_k x)}{1 - F_1(\lambda_k + \delta_k x)} \rightarrow 1$ and as $\lambda_k + \delta_k x \rightarrow \bar{w}$ we get $\frac{1 - F(y)}{1 - F_1(y)} \rightarrow 1$ as $x \uparrow \bar{w}$.

The statement for minima is obvious:

If $F(x)$ and $F_1(x)$ are tail equivalent for minima and one of then is attracted to $\underset{\sim}{L}(x)$, then the other is also attracted to $\underset{\sim}{L}(x)$ with the same coefficients and conversely.

It is immediate that the logistic, the exponential, and Gumbel distributions are tail equivalent for maxima, which shows that in many cases attraction can simply be obtained by seeking equivalence to one of the $\tilde{L}(x)$; the Cauchy distribution is tail equivalent to $\Phi_1(\pi x)$. But the normal distribution, although attracted to the Gumbel distribution, is not tail equivalent to it (not the logistic or the exponential); for more details see Resnick (1971) and Tiago de Oliveira and Epstein (1972).

The convergence of quantiles

Consider a sequence of distribution functions $\{F_k(x)\}$ such that $F_k(x) \xrightarrow{w} L(x)$, $L(x)$ being continuous and, thus, the convergence being uniform. For large k, given ϵ , we know that

$$L(x) - \epsilon \leq F_k(x) \leq L(x) + \epsilon$$

and so

$$F_k^{-1}(L(x) - \epsilon) \leq x \leq F_k^{-1}(L(x) + \epsilon) \quad \text{and with} \quad L(x) = p$$

we get

$$F_k^{-1}(p - \epsilon) \leq L^{-1}(p) \leq F_k^{-1}(p + \epsilon) .$$

Taking in the RHS inequality $p - \epsilon$ instead of p and in the LHS inequality $p + \epsilon$ instead of p, we get

$$L^{-1}(p - \epsilon) \leq F_k^{-1}(p) \leq L^{-1}(p + \epsilon) ;$$

and so, by the continuity of $L(x)$, we get

$$F_k^{-1}(p) \to L^{-1}(p) .$$

Applying the previous results to maxima, where the $\tilde{L}(x)$ are continuous, we see that, as $F_k(x) = F^k(\lambda_k + \delta_k x)$, we get for

$$\chi_p = \tilde{L}^{-1}(p), \quad \frac{F^{-1}(p^{1/k}) - \lambda_k}{\delta_k} \to \chi_p$$

which shows the practical result $F^{-1}(p^{1/k}) \approx \lambda_k + \delta_k \chi_p$.

For instance if $F(x) = 1/(1 + e^{-x})$ is the standard logistic, attracted for maxima to Gumbel distribution, as we can take $\lambda_k = \log k$ and $\delta_k = 1$ we get $F^{-1}(p^{1/k}) \approx \log k + \chi_p$, $\chi_p = \Lambda^{-1}(p) = -\log(-\log p)$; the probability error is $|F^k(\log k + \chi_p) - p| = |(1 - \frac{\log p}{k})^{-k} - p| \to 0$ as $k \to +\infty$ and the linear error is $(\log + \chi_p) - F^{-1}(p^{1/k})| = |\log \frac{k(p^{-1/k} - 1)}{-\log p}| \to 0$ as $k \to +\infty$.

For the exponential we have, in the same conditions $\lambda_k = \log k$, $\delta_k = 1$, the probability error is $|F^k(\log k + \chi_p) - p| = |(1 + \frac{\log p}{k})^k - p| \to 0$ as $k \to \infty$, and also the linear error $|F^{-1}(\log k + \chi_p) - (p^{1/k})| = |\log \frac{k(1 - p^{1/k})}{-\log p}| \to 0$ as $k \to +\infty$.

Speed of convergence

The way a sequence of distributions of maxima converge to its limit is a very important question : either it converges quickly to the limit and this limit can be used as an approximation to the real distribution, or the approach is slow and the limit, from the statistical standpoint, has little relevance : if an error of $\epsilon = 10^{-2}$ is obtained in one case for $k = 50$ the approximation can be used for moderate samples — the approximation for small samples being practically speaking an illusion — but if the error $\epsilon = 10^{-2}$ is a attained only for $k \geq 10^6$ the result has no practical use.

Suppose that $F^k(\lambda_k + \delta_k x) \to \tilde{L}(x)$. We can think of two different approaches, briefly touched upon in the examples of the preceding section.

We may be interested in $p_k = \sup_x |F^k(\lambda_k + \delta_k x) - \tilde{L}(x)|$: this maximum probability error gives an evaluation of the computation of the probability of overpassing $\lambda_k + \delta_k x$ — i.e. of $1 - F^k(\lambda_k + \delta_k x)$ — by evaluating it by $1 - \tilde{L}(x)$. If $p_k \leq \eta$ and χ_p is p-quantile of $\tilde{L}(x)$ we see that $p - \eta < F^k(\lambda_k + \delta_k \chi_p) < p + \eta$, and if η is very small in comparison with p (in general close to 1) we have good approximations to design, etc... It is evident that $p_k = p_k(\lambda_k, \delta_k)$ and so an open question is to determine the best (λ_k, δ_k), i.e., the values that minimize $p_k(\lambda, \delta)$.

The other error — the linear one — needs some care in its definition. The idea is to study the difference of the quantiles of $F^k(\lambda_k + \delta_k x)$ and of $\tilde{L}(x)$, i.e., to compute $d_k = \sup_x |\lambda_k + \delta_k Q(p^{1/k}) - \chi_p|$. But a simple example shows that this definition can lead to results of no

practical use. Suppose $F(x)$ is the exponential distribution which (with $\lambda_k = \log k$, $\delta_k = 1$) is attracted for maxima to $\Lambda(x)$. As $F^k(\lambda_k + \delta_k x) \to \Lambda(x)$ we have, approximately $F^k(y) \approx \Lambda((y - \lambda_k)/\delta_k)$ and so the exact p-quantile is $Q(p^{1/k}) = -\log(1 - p^{1/k})$ and the approximate one is $\lambda_k + \delta_k \Lambda^{-1}(p) = \log k - \log(-\log p)$. The maximum resulting linear error is then $d_k = \sup_p |\log k - \log(-\log p) + \log(1 - p^{1/k})| = \sup_p \left| \log \frac{k(1 - p^{1/k})}{-\log p} \right| = +\infty$, the value attained when $p \to 0$. A careful study of the linear error (dependent also on (λ_k, δ_k)) was not made but it leads to the computations being made in a shorter interval $\epsilon \le p \le 1 - \epsilon$' chiefly because, for maxima, we are essentially interested in values of p close to 1 and not to 0.

The relative probability error $\sup_x \left| \frac{F^k(\lambda_k + \delta_k x) - \tilde{L}(x)}{1 - \tilde{L}(x)} \right| = \sup_x \left| \frac{1 - F^k(\lambda_k + \delta_k x)}{1 - \tilde{L}(x)} - 1 \right|$

for the interval $0 < \tilde{L}(x) < 1$ is also an open problem.

Although the more general results are due to Davis (1982) and Tiago de Oliveira (1991) we will only describe the statements of Galambos (1978); some important results are the ones of Balkema, de Hann and Resnick (1984), Galambos (1984) and Beirlant and Willekens (1999).

If $F^k(\lambda_k + \delta_k x) \to \tilde{L}(x)$ *denote by* $z_k(x) = k(1 - F(\lambda_k + \delta_k x))$ *and for* $x > \underline{w}$, $\rho_k(x) = z_k(x) + \log \tilde{L}(x)$ *: then for* $x > \underline{w}$ *and* $z_k(x) \le k/2$ *we have*

$$|F^k(\lambda_k + \delta_k x) - \tilde{L}(x)| \le \tilde{L}(x) [r_{1,k}(x) + r_{2,k}(x) + r_{1,k}(x)\ r_{2,k}(x)]$$

where

$$r_{1,k}(x) = \frac{2 z_k^2(x)}{k} + \frac{2 z_k^4(x)}{k^2} \cdot \frac{1}{1 - q}$$

$$r_{2,k}(x) = |\rho_k(x)| + \frac{\rho_k^2(x)}{2} \cdot \frac{1}{1 - s}$$

with $q < 1$, $s < 1$ *such that* $z_k^2(x) \le 3 k q/2$ *and* $|\rho_k(x)| < 3 s$.

As can be seen, this statement does not apply for all x but only for a part of the domain of $F(x)$ although is valid for all admissible sets $\{(\lambda_k, \delta_k)\}$.

The dual statement for minima is :

If $1-(1-F(\lambda_k+\delta_k x))^k \to \underset{\sim}{L}(x)$ *denote by* $z_k(x) = k\,F(\lambda_k+\delta_k x)$ *and for* $x < \bar{w}$, $z_k(x) = z_k(x) + \log(1-L(x))$ *: then for* $x < \bar{w}$, $z_k(x) \le k/2$ *we have*

$$|1-(1-F(\lambda_k+\delta_k x))^k - \underset{\sim}{L}(x)| \le (1-\underset{\sim}{L}(x))\,[r_{1,k}(x) + r_{2,k}(x) + r_{1,k}(x)\,r_{2,k}(x)]$$

$r_{1,k}$ *and* $r_{2,k}$ *having the same definition as before.*

Davis (1982) gave a different approach in probability error evaluation using, essentially, the approximation — $k \log F^k(\lambda_k+\delta_k x) \sim k(1-F(\lambda_k+\delta_k x))$ in the interval $0 < \tilde{L}(x) < 1$. In Tiago de Oliveira (1991) we sketched a similar result, but with a more direct approach, that we will explain.

As $a^k - b^k = (a-b)\sum_{j=o}^{k-1} a^j\, b^{k-1-j}$ we have $|a^k - b^k| \le k\,|a-b|\,\max(a^{k-1}, b^{k-1})$ for $0 \le a, b \le 1$. If $F^k(\lambda_k+\delta_k x) \to \tilde{L}(x)$, as we see that $|F^k(\lambda_k+\delta_k x) - \tilde{L}(x)| = |F^k(\lambda_k+\delta_k x) - (\tilde{L}^{1/k}(x))^k| \le k\,|F(\lambda_k+\delta_k x) - \tilde{L}^{1/k}(x)| \times \max(F^{k-1}(\lambda_k+\delta_k x), \tilde{L}^{1-1/k}(x)) \le k\,|(F(\lambda_k+\delta_k x) - \tilde{L}^{1/k}(x)|$. As the third factor (max) in the before last expression converges to $\tilde{L}(x)$ we could substitute it by $\tilde{L}(x)$ but this is practically irrelevant because we are interested in the large quantiles ($\tilde{L}(x) \approx 1$).

This the basic result is $|F^k(\lambda_k+\delta_k x) - \tilde{L}(x)| \le k\,|F(\lambda_k+\delta_k x) - \tilde{L}^{1/k}(x)|$, *the RHS being the principal part of the error, giving thus the order of convergence of* $|F^k(\lambda_k+\delta_k x) - \tilde{L}(x)|$ *to zero.*

In the interval $0 < \tilde{L}(x) < 1$, introducing $\tau_k(x) = F(\lambda_k+\delta_k x)/\tilde{L}^{1/k}(x) - 1$, where, as it is immediate, $k\,\tau_k(x) \to 0$ we can give a formulation analogous to the one of Davis (1982) $|F^k(\lambda_k+\delta_k x) - \tilde{L}(x) \le k\,|\tau_k(x)|\,\tilde{L}(x)\max((1+\tau_k(x))^{k-1}, 1)$ and we see, once more, that the order of convergence of $F^k(\lambda_k+\delta_k x)$ to $\tilde{L}(x)$ to is the one of $k\,\tau_k(x) \to 0$

For the relative error of the tail evaluation we have

$$\left|\frac{1-F^k(\lambda_k+\delta_k x)}{1-\tilde{L}(x)} - 1\right| \le k\left|\frac{1-F(\lambda_k+\delta_k x)}{1-\tilde{L}^{1/k}(x)} - 1\right|\frac{1-\tilde{L}^{1/k}(x)}{1-\tilde{L}(x)}$$

$$\le \left|\frac{1-F(\lambda_k+\delta_k x)}{1-\tilde{L}^{1/k}(x)} - 1\right|\left(1+\frac{1-1/k}{2}\,(1-\tilde{L}(x))\right)$$

using the development of $1 - \tilde{L}^{1/k}(x)$ in the (alternating) Taylor series on $1 - \tilde{L}(x)$.

Davis (1982) essential result can be obtained from $| F^k(\lambda_k + \delta_k x) - \tilde{L}(x)| \leq k | F(\lambda_k + \delta_k x) - \tilde{L}^{1/k}(x) |$, as $\tilde{L}^{1/k}(x) = \exp (\frac{1}{k} \log \tilde{L}(x)) = 1 + \frac{1}{k} \log \tilde{L}(x) + O (\frac{1}{k^2})$, under the form $|F^k(\lambda_k + \delta_k x) - \tilde{L} (x)| \leq k |F(\lambda_k + \delta_k x) + \log \tilde{L}(x) + O(1/k) |$ which shows that the order of convergence is, at most, $O(1/k)$ and is of that order only if $k(1 - F(\lambda_k + \delta_k x) + \log \tilde{L}(x) = O(1/k)|$. The convergence is thus slow in general.

Finally it should be noted that in some cases, as for the normal distribution, a sequence of von Mises-Jenkinson forms $G(z| \theta_m)$, with $\theta_m \rightarrow 0$ conveniently chosen, can give a better approximation to $F^k(\lambda_k' + \delta_k' x)$ ((λ_k' , δ_k') also convenient) than $G(z|0) = \Lambda(z)$, the Gumbel distribution. Although theoretically very interesting this point it has a small statistical interest because we can make the statistical choice of a distribution that fits better the data (see Chapter IV and VIII).

In Part I we used the traditional notation where $k > 0$ (integer) is the index of a sequence; in the next parts, except for the probabilistic chapter of Part III, we will use n (>0) integer, which will be the sample size.

ANNEX I

On The "Duality" between Extremes and Sums

For simplicity we will deal with some "duality" between sums (or averages) and maxima, the translation to minima being obvious from the relation $\min_{1}^{n} \{ X_i \} = - \max_{1}^{n} \{ - X_i \}$.

The "duality" is expressed by the two columns in correspondence, where there are various gaps. F(.) , F(. , .), ... and $\varphi(.)$, $\varphi(. , .)$, ... will denote the distribution functions and the characteristic functions.

Sums	Maxima
$S_k = \sum_{1}^{k} X_i$	$M_k = \max_{1}^{k} \{X_i\}$
$\varphi_X(t) = M_X (e^{itX})$: ch. f. of X	$F_X(t) = \text{Prob}\{X \le x\}$: d. f. of X
X_i indep. : $\varphi_{S_k}(t) = \prod_{1}^{k} \varphi_i(t)$	X_i indep. : $F_{M_k}(x) = \prod_{1}^{n} F_i(x)$
X_i i. i.d. : $\varphi_{S_k}(t) = \varphi^k(t)$	X_i i. i.d. : $F_{M_k}(x) = F^k(x)$
(X, Y) indep: $\varphi_{aX+bY}(t) = \varphi_X(a\,t)\,\varphi_Y(b\,t)$	(X, Y) indep: $F_{\max(X+a,\ Y+b)} = F_X(x - a)\,F_Y(x - b)$
(+ , .)	(max , +)
$M(aX + b) = a\,M(X) + b$	...
$V(aX + b) = a^2\,V(X)$	...

If $\{X_i\}$ i.i.d. have μ, σ^2 then

$\varphi_{(s_k - k\mu)/\sqrt{k}\,\sigma}(t) \to (e^{-t^2/2})$

(Central Limit Theorem); in the general case "sometimes" the ch. f. of the normal law may be substituted by that of an indefinitely divisible law; $\mu = \varphi'_X(0)/i$, $\sigma^2 = \varphi'_X(0)^2 - \varphi''_X(0)$, in the usual case.

If (X,Y) has a binormal distribution with standard normal margins and correlation coefficient then $Z = aX + bY$ has a normal distribution $N(x/\sigma(a,b))$ with $\sigma(a,b) = a^2 + b^2 + 2\rho\, a\, b)$.

If $\rho = 0$ (independence), $\sigma(a,b) = 1$ iff $a^2 + b^2 = 1$; in that case $C(X, Z) = a$.

If $\{X_i\}$ are i.i.d. there "sometimes" exist $(\lambda_k, \delta_k > 0)$ such that $\text{Prob}\{(M_k - \lambda_n)/\delta_n \le x\} = F^k(\lambda_k + \delta_k x) \to \tilde{L}(x)$, $\tilde{L}(x)$ then being $\Psi_\alpha(x)$, $\Lambda(x)$ or $\Phi_\alpha(x)$;
for $\Lambda(x)$ we have $n(1 - F(\lambda_k)) \to 1$, $k(1 - F(\lambda_k + \delta_k)) \to e^{-1}$ or $\delta_n \sim 1/n\, F'(\lambda_n)$; there are corresponding results for Φ_α and Φ_α .

If (X, Y) has a bivariate distribution with reduced Gumbel margins then $Z = \max(X - a, Y - b)$ has a Gumbel distribution $\Lambda(z - \lambda(a, b))$ with $\lambda(a, b) = \log\{(e^{-a} + e^{-b})\, k(b - a)\}$

If $k(w) = 1$ (independence) , $\lambda(a , b) = 0$ if $e^{-a} + e^{-b} = 1$; in that case $\text{Prob}\{Z \le X - a\} = \text{Prob}\{X - b \le X - a\} = e^{-a}$

ANNEX II

ON THE KHINTCHINE'S CONVERGENCE OF TYPES THEOREM

As this theorem has been referred to and will be important to the proof of asymptotic univariate resuls — and to subsequent generalizations — we will give a proof of it. Let $\{V_k(x)\}$ be any sequence of distribution functions and suppose that for two sets $\{(\lambda_k, \delta_k), \delta_k > 0\}$ and $\{(\lambda_k^*, \delta_k^*), \delta_k^* > 0\}$ we have $V_k(\lambda_k + \delta_k x) \xrightarrow{w} L(x)$ and $V_k(\lambda_k^* + \delta_k^* x) \xrightarrow{w} L^*(x)$. Then:

$L^*(x) = L(A + B x)$ *with* $(\lambda_k^* - \lambda_k)/\delta_k \to A$, $\delta_k^*/\delta_k \to B (-\infty < A < +\infty, 0 < B < +\infty)$ *and conversely.*

Essentially the theorem says that no sequence of distribution functions is attracted or converges after convenient positive linear transformations to distribution functions not belonging to the same type, if the limits are proper and non–degenerate.

We will closely follow Feller (1966) for the proof. Assume that $(\lambda_k^* - \lambda_k)/\delta_k \to A$, $\delta_k^*/\delta_k \to B$ and we will show that $L^*(x) = L(A + B x)$. In fact, given $\varepsilon > 0$, for k sufficiently large, we have

$$V_k(\lambda_k + \delta_k(A + B x - \varepsilon)) \leq V_k(\lambda_k^* + \delta_k^* x) \leq V_k(\lambda_k + \delta_k(A + B x + \varepsilon)).$$

Letting $k \to +\infty$ and at the points x that are continuity points of $L^*(x)$ and $A + B x \pm \varepsilon$ that are continuity points of L(x) — which excludes at most a denumerable set — we get

$$L(A + B x - \varepsilon) \leq L^*(x) \leq L(A + B x + \varepsilon).$$

Letting now $\varepsilon \to 0$ we get $L^*(x) = L(A + B x)$ except possibly at a denumerable set and thus everywhere owing to the right–continuity of distribution functions.

Let us now assume that $L^*(x) = L(A + B\,x)$. As L and L^* are proper and non–degenerate, choose two continuity points $x' < x''$ of $L^*(x)$ $(0 < L^*(x') < L^*(x'') < 1)$; then $A + B\,x'$ and $A + B\,x''$ are continuity points of $L(x)$. For sufficiently large k we have from $V_k(\lambda_k + \delta_k\,x) \to L(x)$ and $V_k(\lambda_k^* + \delta_k^*\,x) \to L^*(x)$ that there exist also continuity points $y' < y''$ of $L(x)$ such that

$$V_k(\lambda_k + \delta_k\,y') < V_k(\lambda_k^* + \delta_k^*\,x') \le V_k(\lambda_k^* + \delta_k^*\,x'') < V_k(\lambda_k + \delta_k\,y'')$$

and so $$\lambda_k + \delta_k\,y' < \lambda_k^* + \delta_k^*\,x' < \lambda_k^* + \delta_k^*\,x'' < \lambda_k + \delta_k\,y'' \ .$$

Thus $(0 <) \dfrac{\delta_k^*}{\delta_k} < \dfrac{y'' - y'}{x'' - x'}$ and consequently for some subsequence of k we have $\delta_k^*/\delta_k \to B'$ $(> 0$ because of the non–degeneracy). Also $y' - \delta_k^*/\delta_k\ x' < (\lambda^* - \lambda_k)/\delta_k < y'' - \delta_k^*/\delta_k\ x''$ and so, as $(\lambda_k^* - \lambda_k)/\delta_k$ remains bounded and in a subsequence of the previous one, it has a limit A'.

Evidently A' and B' are uniquely determined and we have $A' = A$, $B' = B$.

For the situation described where $L^*(x) = L(A + B\,x)$ we say that $L(x)$ and $L^*(x)$ belong to the same type and the coefficients (λ_k, δ_k) and $(\lambda_k^*, \delta_k^*)$ are *equivalent*; if $L^*(x) = L(x)$ or $A = 0, B = 1$ will say that they are *totally equivalent.*

ANNEX III

On the δ-Method

Essentially the *δ-method* is based on the fact that if $\{X_k\}$ and $\{Y_k\}$ are sequences of random variables such that either $X_k - Y_k \xrightarrow{P} 0$ or $X_k / Y_k \xrightarrow{P} 1$ and $\{X_k\}$ has a limiting proper and non-degenerate distribution, the same happens to $\{Y_k\}$; see Cramér (1946), p. 254-255. Also if $\{(X_k, Z_k)\}$ and $\{(Y_k, Z_k)\}$ are sequences of random pairs and $X_k - Y_k \xrightarrow{P} 0$ or $X_k / Y_k \xrightarrow{P} 1$ and $\{(X_k, Z_k)\}$ has a limiting (bivariate) proper and non-degenerate distribution, the same happens to $\{(Y_k, Z_k)\}$. The multivariate generalization is obvious. In many applications the δ-method leads to a linearization (use of the terms up to the first order of Taylor development). For details see Tiago de Oliveira (1982).

The δ-method was underlying some reasoning in this Part and will be used more extensively in the following Parts.

Let us give two examples, used in the text, that help to clarify the use of the δ-method.

Example 1:

We wish to study the asymptotic behaviour of $(\bar{x}, s)$ and to show that $C(\bar{x}, s) \sim \beta_1 \sigma^2/2k$, as obtained by Gumbel and Carlson (1956).

By the Central Limit Theorem , if $\{X_i\}$ have moments up to the fourth order, we know that

$$A_k = \sqrt{k}\ \frac{\frac{1}{k}\sum_1^k x_i - \mu}{\sigma} \quad \text{and} \quad B_k = \sqrt{k}\ \frac{\frac{1}{k}\sum_1^k (x_i - \mu)^2 - \sigma^2}{\sqrt{\beta_2 - 1}\ \sigma^2}$$

are asymptotically standard normal and the random pair is asymptotically binormal with standard margins and correlation coefficient $\rho = \dfrac{\beta_1}{\sqrt{\beta_2 - 1}}$. But B_k, as $x_i - \mu = (x_i - \bar{x}) + (\bar{x} - \mu)$, takes the form

$$B_k = \sqrt{k}\ \frac{s^2 + (\bar{x} - \mu)^2 - \sigma^2}{\sqrt{\beta_2 - 1}\ \sigma^2}\ .$$

Then as $\sqrt{k}\ (\bar{x} - \mu)^2 = \sigma^2 A_k^2 / \sqrt{k} \xrightarrow{P} 0$ and so

$$\left(\sqrt{k}\ \frac{\bar{x} - \mu}{\sigma}\ ,\ \sqrt{k}\ \frac{s^2 - \sigma^2}{\sqrt{\beta_2 - 1}\ \sigma^2}\right)$$

is also an asymptotically binormal pair with standard margins and correlation coefficient $\rho = \beta_1 / \sqrt{\beta_2 - 1}$. Also $s^2 \xrightarrow{P} \sigma^2$ and, thus, $\dfrac{s^2 - \sigma^2}{2\,\sigma(s - \sigma)} \xrightarrow{P} 1$.

Consequently $\left(\sqrt{k}\ \dfrac{\bar{x} - \mu}{\sigma}\ ,\ 2\sqrt{k}\ \dfrac{s - \sigma}{\sqrt{\beta_2 - 1}\ \sigma}\right)$ is also an asymptotically binormal pair with standard margins, correlation coefficient ρ, and

$$C(\bar{x}, s) \sim \rho\ \frac{\sigma}{\sqrt{k}}\ .\ \frac{\sqrt{\beta_2 - 1}\ \sigma}{2\sqrt{k}} = \frac{\beta_1 \sigma^2}{2k}\ .$$

Example 2:

Let $f(x, y)$ be a twice-differentiable function and $\{(X_k, Y_k)\}$ a sequence of random pairs such that $\left(\sqrt{k}\ \dfrac{X_k - \mu_x}{\sigma_x}\ ,\ \sqrt{k}\ \dfrac{Y_k - \mu_Y}{\sigma_y}\right)$ is asymptotically binormal with standard margins and correlation coefficient ρ. Then:

$\sqrt{k}\ \{f(X_k, Y_k) - f(\mu_x, \mu_y)\}$ is asymptotically normal with mean value zero and variance V below.

In fact, as

$$\sqrt{k}\ \{f(X_k, Y_k) - f(\mu_x, \mu_y)\} = \sqrt{k}\ \{(X_k - \mu_x)\ \frac{\partial f}{\partial x}\Big|_{\mu_x, \mu_y} + (Y_k - \mu_y)\ \frac{\partial f}{\partial y}\Big|_{\mu_x, \mu_y}$$

$$+ \text{ 2nd order terms } \},$$

as the second order terms are $o_p(n^{-1/2})$ we see that $\sqrt{k}\ \{f(X_k, Y_k) - f(\mu_x, \mu_y)\}$ has the same asymptotic distribution of the linearized form

$$\sqrt{k}\ \{(X_k - \mu_x)\ \frac{\partial f}{\partial x}\Big|_{\mu_x, \mu_y} + (Y_k - \mu_y)\ \frac{\partial f}{\partial y}\Big|_{\mu_x, \mu_y}\}$$

and is thus normal with mean value zero and variance

$$V = \sigma_x^2 \left(\frac{\partial f}{\partial x}\Big|_{\mu_x, \mu_y}\right)^2 + 2\,\rho\,\sigma_x\ \sigma_y\ \frac{\partial f}{\partial z}\Big|_{\mu_x, \mu_y}\ \ \frac{\partial f}{\partial y}\Big|_{\mu_x, \mu_y} + \sigma_y^2 \left(\frac{\partial f}{\partial y}\Big|_{\mu_x, \mu_y}\right)^2 .$$

The multivariate generalization is obvious.

This result presupposes that not all the first derivatives are null; if this is not the case (i.e., all first derivatives are null) then we have different asymptotic distributions, such as the χ^2 — see Tiago de Oliveira (1982).

Exercises — Part I

1.1 — Show that $\{\min_{1}^{k} X_i\} \xrightarrow{P} \underline{w}$.

1.2 — Study the Cauchy distribution for the LLN.

1.3 — Prove the conditions for the validity of MLLN.

1.4 — Consider a sample of n observations $\{X_i\}$ and a possible future sample of m observations $\{Y_j\}$, all of them i.i.d. Show that

$$\text{Prob}\{\min_{1}^{n} X_i \leq \min_{1}^{m} Y_j \leq \max_{1}^{m} Y_j \leq \max_{1}^{n} X_i\} = \frac{n(n-1)}{(n+m)(n+m-1)} .$$

1.5 — Show that $\max_{1}^{k}\{X_i\} \leq a + \max_{1}^{k}\{(X_i - a)_+\}$. If the $\{X_i\}$ have the margins $F_i(x)$ then

$$M(\max_{1}^{k} X_i) \leq a + \sum_{1}^{k} \int_{a}^{+\infty} (1 - F_i(x))\, dx \; ; \text{ no assumption of independence is made.}$$

1.6 — Obtain the constants $(\alpha_m , \beta_m > 0)$ that satisfy the stability equation for maxima and for minima.

1.7 — Obtain the transformations interchanging $\tilde{L}(x)$ between themselves, $\underset{\sim}{L}(x)$ between themselves and between $\underset{\sim}{L}(x)$ and $\underset{\sim}{L}(x)$; compare the different graphs of the densities to see the analogies.

1.8 — Translate maxima results to minima results and vice-versa; use, in particular, the relation between the distribution function and the survival function and with the reduced Gumbel distribution.

1.9 — $F(x) = 0$ if $x \leq 0$, $F(x) = 1 - e^{-\sqrt{x}}$ is attracted for maxima to the Gumbel distribution; obtain one system of attraction coefficients.

I.10 — Show that if $\bar{w} = +\infty$, and for some u and $v > 0$, as $x \to +\infty$, $x^u \exp(x^v)(1 - F(x)) \to c$, $0 < c < +\infty$, then $\{(\max_1^k X_i - \lambda_k)/\delta_k\}$ is attracted to $\Lambda(x)$; a system of coefficients can be $\lambda_k = (\log(k\,c))^{1/v} - \dfrac{u \log\log(k\,c)}{v^2 (\log(k\,c))^{(v-1)/v}}$ and $\delta_k = \dfrac{(\log(k\,c))^{(1-v)/v}}{v}$.

J. Villaseñor (1976), Ph.D. Thesis

I.11 — Show that the conditions given for $F \in \mathcal{D}(\Lambda)$ are equivalent.

I.12 — Obtain the continuous condition on $Q(v)$ for F to be attracted to the Gumbel distribution of minima.

I.13— The maximum likelihood estimators of (λ, δ) for the Gumbel distribution $\Lambda((x - \lambda)/\delta)$ are (Chapter V, Part II)

$$\hat{\delta} = \bar{x} - \frac{\sum_1^n x_i\, e^{-x_i/\hat{\delta}}}{\sum_1^n e^{-x_i/\hat{\delta}}}$$

$$\hat{\lambda} = -\hat{\delta} \log\left(\frac{1}{n} \sum_1^n e^{-x_i/\hat{\delta}}\right).$$

Show that if $\Phi(\theta) = \int_{-\infty}^{+\infty} e^{-\theta x}\, d\,\Lambda((x - \lambda)/\delta)$ then

$$\delta = -\Phi'(0) + \frac{\Phi'(1/\delta)}{\Phi(1/\delta)}$$

$$\lambda = -\delta \log \Phi(1/\delta),$$

corresponding to the limiting structure of the maximum likelihood estimators.

I.14 — If $F \in \mathcal{D}(\Lambda)$ then $F^k(x) \approx \Lambda\left(\dfrac{x - \lambda_k}{\delta_k}\right)$. Define then $\Phi_k(\theta)$. Can this be used to obtain attraction coefficients? Can this be extended to the Weibull and Fréchet distributions ?

I.15 — Show that Fuller approximations to the return period of overpassing the level a for the Gumbel and Fréchet distributions for maxima are

$$\tilde{T}(a) = [-\log \Lambda(\frac{a - \lambda}{\delta})]^{-1} = e^{(a-\lambda)/\delta} \quad \text{and} \quad \tilde{T}(a) = [-\log \Phi_\alpha(\frac{a - \lambda}{\delta})]^{-1} = (\frac{a - \lambda}{\delta})^\alpha$$

$(a > \lambda)$ for large a and that the return period of underpassing the level a for the Weibull distribution is $\tilde{T}(a) = [-\log(1 - W_\alpha(\frac{a-\lambda}{\delta}))]^{-1} = (\frac{a-\lambda}{\delta})^{-\alpha}$ $(a > \lambda)$ for small $a - \lambda$.

I.16— Show, for the Gumbel distribution, that $\tilde{T}((1+\eta)\,a)$ and $\tilde{T}(a)$ verify the relation $\tilde{T}((1+\eta)\,a) = e^{\eta\,\lambda/\delta}\ \tilde{T}^{1+\eta}(a)$. This approximation, with a very small relative error, shows the nonsense of the rule "to tame your river and build the dam strong enough to withstand anything, double the largest flood so far observed". Then $\eta = 1$, $\tilde{T}(2\,a) = e^{\lambda/\delta}\,\tilde{T}^2(a) > \tilde{T}^2(a)$, as $\lambda > 0$, and so if the number of years of observation is 50 years, we have the return period $\tilde{T}(2\,a) > 2500$ years which is not a realistic planning time.

Note : See in Part II, for each limiting distribution, the status of statistical analysis of extreme data for risk evaluation and control and design planning.

I.17— Using the fact that, for the standard normal distribution, $\frac{1 - N(x)}{N'(x)/x} \to 1$ as $x \to \infty$, and defining c_n by $c_n = n\,N'(c_n)$, show that $N^n(c_n + x/c_n) \to \Lambda(x)$.

I.18— Using Mill's inequalities for the standard normal distribution

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) N'(x) < 1 - N(x) < \frac{1}{x}\ N'(x) \quad \text{for } x > 0,$$

show that $N^k(\lambda_k + \delta_k\,x) \to \Lambda(x)$ as $k \to \infty$ if $\lambda_k\,\delta_k \to 1$ and $\frac{k}{\sqrt{2\pi}\,\lambda_k}\ e^{-\lambda_k^2/2} \to 1$ as $k \to \infty$, thus directly obtaining the results in the text.

I.19— Let $\{X_k\}$ be a sequence of i.i.d. reduced Gumbel random variables. Fix an integer m and define $Y_k = \max(Y_i, X_{k+1}, \ldots, X_{k+m-1})$. Show that Prob $\{Y_k - \log m \le x\} = \Lambda(x)$ and, also, that

$$\text{Prob}\ \{\max_{1}^{k}\ (Y_i - \log(k + m - 1) \le x\} = \Lambda(x).$$

Note that the sequence $\{Y_k\}$ is $(m-1)$-dependent.

I.20— Let $\{X_i\}$ be a sequence of independent random variables such that $a_i\,X_i + b_i$ $(a_i > 0)$ have the distribution function $F \in \mathcal{D}(\Lambda)$; let a system of attraction coefficients for F be (λ_k, δ_k). Show then, as

$$F(y) = 1 - \frac{\exp(-(y - \lambda_k)/\delta_k) + o(1)}{k}, \quad \text{that}$$

$$\text{Prob}\{(\max(X_1, ..., X_k) - \lambda_k')/\delta_k' \le x\} \to \Lambda(x)$$

$$\text{if} \quad \frac{1}{k} \sum_{1}^{k} e^{-(\eta_i + \eta_i x)} \to e^{-x} = 1$$

where $\xi_i\ (= \xi_i(k)) = a_i\ \delta_k'/\delta_k$ and $\eta_i\ (= \eta_i(k)) = (a_i\ \lambda_k' + b_i - \lambda_k)/\delta_k$.

We can have the same attraction coefficients (i.e., $(\lambda_k'\ ,\ \delta_k')) = (\lambda_k\ ,\ \delta_k)$) if $\eta_i = 0$ and $\xi_i = 1$ or $a_i = 1$ and $b_i = 0$, which is the already classical case: but if $a_i = 1$ and $b_i = 0$ is false there may be a (non-equivalent) change in the attraction coefficients.

I.21— Compute the mean value and variance of Ψ_α and Φ_α (when they exist) up to $0(\alpha^{-1})$ and $0(\alpha^{-2})$ and obtain thus the linear transformation such that the transformed variables converge to the corresponding ones of the reduced Gumbel distribution (whose mean value and variance are γ and $\frac{\pi^2}{6}$); see, for more details, in Part II, the Gumbel distribution.

I.22— Show that if Z_α has the reduced Weibull distribution for maxima, with shape parameter $\alpha > 0$, then $\alpha(1 + Z_\alpha)$ has the Gumbel distribution for maxima as a limit when $\alpha \to \infty$. Show also that if Z_α has the reduced Fréchet distribution, with shape parameter $\alpha > 0$, then $\alpha(-1 + Z_\alpha)$ has the Gumbel distribution for maxima as a limit when $\alpha \to +\infty$.

I.23— The graphs suggest that $\Phi_\alpha(z)$ and $W_\alpha(z)$ are very close. Show, although, that

$$\sup_{z \ge 0} |W_\alpha(z) - \Phi_\alpha(z)| = |W_\alpha(1) - \Phi_\alpha(1)| = 1 - 2e^{-1} = .2642411,$$

with $W_\alpha'(1) = \Phi_\alpha'(1)$. Note that $\sup_{z \ge 0} = |W_\alpha'(z) - \Phi_\alpha'(z)| = +\infty$, for $\alpha < 1$ owing to the fact $W_\alpha'(0) = +\infty$, but for $\alpha \ge 1$ we have $\sup_{z \ge 0} = |W_\alpha'(z) - \Phi_\alpha'(z)| < +\infty$. Compute this bound.

I.24— As known, for the Fréchet distribution (for maxima) it can be shown that $\bar{w}_F = +\infty$ and that we can take $\lambda_k = 0$, i.e., $F^k(\delta_k z) \to \Phi_\alpha(z)$ as $k \to \infty$. Show that if Prob $\{\max(X_1, ..., X_k) / \delta_k \le z\} = F^n(\delta_n z) \to \Phi_\alpha(z)$, with $k \to \infty$, then with $Y_i = \log(\max(X_i, 0)) = \log(X_i)_+$

we have also Prob $\{\alpha\,(\max_{1}^{k} Y_i - \log \delta_k) \le z\} \to \Lambda(z)$. Clearly the result stresses the importance of homogeneous estimators.

I.25— As known, for the Weibull distribution (for minima), it can be shown that $\underline{w}_F < +\infty$ and that we can take $\lambda_k = \underline{w}_F$, i.e., $1 - (1 - F(\underline{w}_F + \delta_k z))^k \to W_\alpha(z)$, as $k \to \infty$; Show that if $Y_i = -\log (X_i - \underline{w}_F)$ then Prob $\{\alpha\,(\max_{1}^{k} Y_i + \log \delta_k) \le z\} \to \Lambda(z)$ also as $k \to \infty$. Consequently, in estimation, the left-end point is very important.

I.26— Using the relation between the Gumbel and Fréchet distributions, show that $F^n(\lambda_n + a\,z) \to \Lambda(z)$ if and only if $\frac{1 - F(x)}{1 - F(x + t)} \to e^{t/a}$ and so $1 - F(\log x)$ is a slowly varying function with index $1/a$.

Gnedenko (1943)

I.27—Show, in sequence to the previous exercise, that if X is normal (μ, σ) then $Y = X^2$ is also attracted to Gumbel distribution with $\delta_k = a = 2\sigma^2$.

I.28— For the examples 1 to 5 and for the attraction coefficients given in the section "The asymptotic distributions of extremes – some examples" (Chapter II) compute numerically the probability and linear errors for probability levels $p = .95, .99, .999$ and values of $k = 10, 100, 1000$. Study the behaviour of example 6.

I.29— Use attraction conditions to show that both $F(x) = 0$ if $x < e$, $F(x) = 1 - \frac{1}{\log x}$ if $x \ge e$, and the Pascal distribution are not attracted to the Fréchet distribution.

I.30— Obtain the constants $(\alpha_t , \beta_t > 0)$ such that the stability equation for $G(z|\theta)$, i.e., $G^t(\alpha_t + \beta_t z|\theta) = G(z|\theta)$ is valid.

I.31— Show using the criteria that $G(z|\theta)$ is attracted to $G(z|\theta)$ as could be expected; i.e., Ψ_α to Ψ_α , Λ to Λ and Φ_α to Φ_α.

I.32— Show that the quantiles of $G(z|\theta)$ converge to those of the Gumbel distribution $\Lambda(z)$ when $\theta \to 0^+$ or $k \to 0^-$.

I.33— Study the case where the $\{X_i\}$ are i.i.d. and $\{X_i^*\}$, with $X_i^* = X_i + (-1)^i$, are both attracted for maxima (or for minima) to the same $\tilde{L}(x)$ $(\underset{\sim}{L}(x))$.

I.34— Show that if $\{X_i\}$ are i.i.d. such that $\bar{w} < +\infty$ and Prob $\{(\max_1^k X_i - \lambda_k)/\delta_k \le x\} \to \tilde{L}(x)$, as $k \to \infty$, for $X_i^* = \frac{1}{\bar{w} - X_i}$ there exist λ_k^* and δ_k^* such that Prob $\{(\max_1^k X_i^* - \lambda_k^*)/\delta_k^* \le x\}$ also converges to some $\tilde{L}(x)$.

I.35— Obtain conditions such that if $\{X_i\}$ are independent and such that if $\{(X_i - \lambda_i)/\delta_i\}$ have the same distribution $\tilde{L}(x)$, then $\max_1^k \{X_i\}$ also has the same asymptotic distribution $\tilde{L}(x)$; compute one system of attraction coefficients, asymptotically.

I.36— Using the fact thats, if $F \in \mathcal{D}(\tilde{L})$ then $F(\lambda_k + \delta_k x) \to 1$ and if $F \in \mathcal{D}(\underset{\sim}{L})$ then $F(\lambda_k + \delta_k x) \to 0$, - show, by the use of Khintchine's convergence of types theorem, that not only $(\lambda_{k+1}, \delta_{k+1})$ are also attraction coefficients for the same $\tilde{L}$ or $\underset{\sim}{L}$, i.e., $(\lambda_{k+1} - \lambda_k)/\delta_k \to 0$ and $\delta_{k+1}/\delta_k \to 1$, but, also, that $F^{n+k_n}(\lambda_n + \delta_n x) \to L(x)$ iff k_n/n; interpret the result.

I.37— Let $\{X_i\}$ be i.i.d. random variables with distribution function $F(x)$ attracted to $\tilde{L}(x)$ for maxima with attraction coefficients $\lambda_k, \delta_k > 0$ and to $\underset{\sim}{L}(x)$ for minima (with attraction coefficients $\lambda_k', \delta_k' > 0$); recall that $\underset{\sim}{L}(x)$ may be different $1 - \tilde{L}(x)$, i.e., that the limiting distributions can not be in correspondence. It was shown that as $k \to \infty$

$$\text{Prob}\{\min_1^k X_i \le \lambda_k' + \delta_k' x, \max_1^k X_i \le \lambda_k + \delta_k y\} \to \underset{\sim}{L}(x)\ \tilde{L}(y).$$

Consider the range $R_k = X_k' - X_1'$ and show that, if $\delta_k'/\delta_k \to b$,

$$\text{Prob}\{(R_k - (\lambda_k - \lambda_k'))/\max(\delta_k, \delta_k') \le z\} \to \int\int_D d\,\underset{\sim}{L}(x)\, d\tilde{L}(y),$$

with $D = \{\frac{y}{\max(1,b)} - \frac{b\,x}{\max(1,b)} \le z\}$. Conclude the dominance of X_k' if $b = 0$, of X_1' if $b = +\infty$, and a joint influence (a convolution) if $0 < b < +\infty$; compare with the

results of Chapter VIII (Analytical Statistical Choice for Univariate Extremes) of Part II for the Gumbel statistic Q.

Obtain specific results for $\Lambda\left(\frac{x-\lambda}{\delta}\right)$, $\Phi_\alpha\left(\frac{x-\lambda}{\delta}\right)$ and $W_\alpha\left(\frac{x-\lambda}{\delta}\right)$ and other common distributions. Note that the δ-method can also be used.

Gumbel (1958), *Galambos* (1987)

I.38— Obtain corresponding results for the mid-point $(X_1^{'} + X_k^{'})/2$.

Gumbel (1958), *Galambos* (1987)

I.39— Obtain the corresponding results for the extremal quotient $EQ_k = X_k^{'} / X_1^{'}$ when the X_i are supposed to be non-negative ($\text{Prob}\{X_i > 0\} = 1$). Note that $\log EQ_k$ is the range of $\{\log X_i\}$.

Gumbel (1958)

I.40— Show that if $F \in \mathcal{D}(\tilde{L})$, with attraction coefficients (λ_k, δ_k) and $G(x)$ is such that $\frac{1-F(x)}{1-G(x)} \to c$ $(0 < c < +\infty)$ when $x \to w_F$ then $G \in \mathcal{D}(\tilde{L})$ and a system of attraction coefficients is $(\lambda_{[ck]}, \delta_{[ck]})$.

I.41— Use the previous result to show that if X has a distribution function $F \in \mathcal{D}(\tilde{L})$ then the truncated random variable $X_a = \max(X, a)$, with $a < w_F$, has a distribution function $F_a \in \mathcal{D}(\tilde{L})$ also. Compute c.

A. A. BALKEMA and L. DE HAAN (1972). On the von Mises condition for the domain of attraction of exp $(-e^{-x})$, *Ann. Math. Statist.* , 43, 1352-1354

A. A. BALKEMA, L. DE HAAN and S. RESNICK (1984). Uniform rates of convergence to extreme value distributions, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 341-346, D. Reidel, Dordrecht

J. BEIRLANT and E. WILLEKENS (1990). Rapid variation with remainder and rates of convergence, *Adv Appl. Probab.*, 22, 787-801

RICHARD A. DAVIS (1982) . The rate of convergence in distribution of the maxima, *Statist. Neerl.*, 36, #1, 31-35

E. L. DODD (1923). The greatest and least variate under general laws of error, *Trans. Am. Math. Soc.*, 25, 525-539

W. FELLER (1966). *An Introduction to Probability Theory and its Applications*, vol. II, Wiley, New York

B. DE FINETTI (1932). Sulle legge di probabilitá degli estreme, *Metron*, 9, 127-138

R. A. FISHER and L. M. C. TIPPET (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Phil. Soc.* , 24, 180-190

M. FRÉCHET (1927). Sur la loi de probabilité de l´écart maximum, *Ann. Soc. Polon. Math.* (Cracovie), 6, 93-116

J. GALAMBOS (1978). *The Asymptotic Theory of Extreme Order Statistics,* Wiley, New-York; (1987), 2nd ed., Krieger Publ.

J. GALAMBOS (1984). Introduction, order statistics, exceedances, laws of large numbers, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 13-17, D. Reidel, Dordrecht

J. GALAMBOS (1984). Rates of convergence in extreme value theory, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 347-352, D. Reidel, Dordrecht

J. GEFFROY (1958/59). Contribution a l´étude de la théorie des valeurs extrêmes, *Publ. Inst. Statist. Univ. Paris* , 7/8, 37-185, Thèse de Doctorat d´Etat

B. V. GNEDENKO (1943). Sur la distribution limite du terme maximum d´une série aléatoire, *Ann. Math.*, 44, 423-453

E. J. GUMBEL (1935). Les valeurs extrêmes des distributions statistiques, *Ann. Inst. H. Poincaré*, 4, 115-168

E. J. GUMBEL (1955). The calculated risk in flood control, *Appl. Sc. Research*, (Holland), 5, A, 273-280

E. J. GUMBEL and P.G. CARLSON (1956). On the asymptotic covariance between the sample mean and standard deviation, *Metron,* XVIII, # 1-2, 3-9

L. DE HAAN (1970). On regular variation and its applications to weak convergence of sample extremes, *Math. Centre Tracts*, # 32, Amsterdam

L. DE HAAN (1971). A form of regular variation and its applications to the domain of attraction of the double exponential distribution, *Z. Wahrsch. verw. Gebiete*, 17, 241-258

L. DE HAAN (1976). Sample extremes; an elementary introduction, *Statist. Neerl.*, 30, 161-172

M. L. JUNCOSA (1949). On the distribution of the minimum of a sequence of mutually independent random variables, *Duke Math. J.*, 16, 609-618

M. R. LEADBETTER, GEORG LINDGREN and HOLGER ROOTZÉN (1983). *Extremes and Related Properties in Random Sequences and Processes*, Springer-Verlag, Heidelberg

DAVID MEJZLER (1949). On a theorem of B. V. Gnedenko, *Sb. Trudov Inst. Math. Akad. Nauk. SSR*, 12, 31-35 (in russian)

DAVID MEJZLER (1984). Asymptotic behaviour of extreme order statistics in the non-identically distributed case, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 535-547, D. Reidel, Dordrecht

MICHAEL MARCUS and MARK PINSKY (1969). On the domain of attraction of $\exp(-e^{-e^{-x}})$, *J. Math. Analysis Appl.*, 28, 440-449

G. F. NEWELL (1964). Asymptotic extremes for m-dependent random variables, *Ann. Math. Statist.*, 35, 1322-1325

S. I. RESNICK (1971). Tail equivalence and its applications, *J. Appl. Prob.*, 8, 136-156

A. ROSENGARD (1962). Étude des lois limites jointes et marginales de la moyenne et des valeurs extrêmes d´un échantillon, *Publ. Inst. Statist. Univ. Paris*, 11, 3-53, Thèse de Doctorat de 3e cycle

A. ROSENGARD (1966). Contribution a l´étude des liaisons limites entre les différents statistiques d´un échantillon, *Université de Paris* , Thèse de Doctorat d´Etat

H. J. W. RÖSSBERG (1963). Uber des asymptotische verhalten der rand-und zentralglieder einer variationsreihe, *Publ. Inst. Math. Acad. Soc. Hungary*, 8, 463-468

H. J. W. RÖSSBERG (1965). Die asymptotiche unabahängikeit der kleinen und grossten werte einer stichprobe von stichprobenmittel, *Math. Nachr.*, 28, 305-318

N. V. SMIRNOV (1952). Limit distributions for the terms of a variational series, *Amer. Math. Soc. Transl.*, 67, 1-64

J. TIAGO DE OLIVEIRA (1961). The asymptotic independence of the sample mean and extremes, *Rev. Fac. Ciências de Lisboa*, A, VIII, 215-222

J. TIAGO DE OLIVEIRA (1976). Asymptotic behaviour of maxima with periodic disturbances, *Ann. Inst. Statist. Math*, 28, 19-23

J. TIAGO DE OLIVEIRA (1978). Approximate distribution for sequences of maxima, *Metron*, XXXVI, 3-21

J. TIAGO DE OLIVEIRA (1982). The δ-method for the obtention of asymptotic distributions — applications, *Publ. Inst. Statist. Univ. Paris* , XXVII, 49-70

J. TIAGO DE OLIVEIRA (ed.) (1984). *Statistical Extremes and Applications*, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1991). Perspectivas sobre a estatística de extremos e problemas em aberto, to be publ. in *Rev. Real Acad. Ciencias Exactas, Fisicas y Naturales*, Madrid

J. TIAGO DE OLIVEIRA and B. EPSTEIN (1972). Tail comparison for maxima; applications to mixtures, unpubl. paper, *Technion*

R. VON MISES (1936). La distribution de la plus grande de n valeurs, *Rev. Math. Union Interbalkanique*, 1, 141-160 and also in *Selected Papers*, vol.II, 1954, 271-294, Am. Math. Soc., Providence, Rhode Island

G. S. WATSON (1954). Extreme values in samples for m-dependent stationary stochastic processes, *Ann. Math. Statist.*, 25, 798-800

PART II

Statistics for univariate extremes

Chapter IV

A QUICK EXPLORATION OF EXTREME DATA

Introduction

Statistical decision for univariate extremes is currently at a moderate stage of development. Some problems have been effectively solved, but not all of the essential ones. On the way, many problems to be solved will appear, constituting a happy hunting ground for researchers. Curiously, even, Bayesian methods have not been strongly developed for statistical extremes, except essentially for some Reliability questions.

The analysis that follows has the following steps: once it is *assumed* that the useful distribution is of one of the three forms — Fréchet, Gumbel or Weibull — we will discuss statistical decision using the *assumed* distribution. This chapter will be devoted to quick exploration techniques, or even as a rough guide on how to assume one of the three models. The last chapter will deal with optimal analytical techniques, comparing them, as far as possible, with the quick ones.

Note that as a matter of practical convenience we will deal with statistical decision for Gumbel and Fréchet distributions for maxima and for Weibull distribution for minima, as the latter is very important in Reliability.

The statistical techniques presented, in the all but the two last chapters, deal with complete samples and an assumed distribution; the penultimate chapter will deal with statistical choice of models and the last one with some cases of incomplete information.

Clearly the Gumbel distribution for maxima with δ (> 0) known, the Fréchet distributions for maxima with λ and α (> 0) known, and the Weibull distributions for minima with λ and α (> 0) known are easily reduced to the exponential distribution (only with a scale parameter) by evident transformations, and the whole technique of the exponential distribution can be transformed to these special distributions of extremes.

Assumption of a model from previous knowledge

In fact, this section could also be called "assumption by tradition". This essentially involves a variable mixture: a convenient usage of accumulated experimentation, of limiting results to be considered as being finitely capable of fitting data, as well as some intuition. The situation is completely analogous to the assumption of normal (multinormal) distribution, where the Central Limit Theorem plays an important role as does long-term accumulated experience with measurements (*), to the assumption of binomial or Bernoulli distribution for use in repeatable trials with an approximately stable probability of the event under consideration, to the assumption of Poisson either as behaviour of small (random) numbers or as a limit of binomial distribution(**), to the assumption of the exponential distribution in the (first) analysis of inter-arrival times (thus implying the Poisson distribution for the associated counting process), to the assumption of geometric, inverse binomial or Pascal distributions in stopping time experiments with the event probabilities approximately stable, etc ... The situation here is very similar.

Let us suppose that we are dealing with maxima (or dually with minima) of i.i.d. samples, with the same size n, and underlying distribution F(.).

Suppose that $F(.) \in \mathcal{D}(\tilde{L}(.))$, that is F(.) belongs to the domain of attraction of $\tilde{L}(.)$ (Weibull, Gumbel or Fréchet distributions), i.e., there exist sequences $\{\lambda_n, \delta_n; \delta_n > 0\}$ such that $F^n(\lambda_n + \delta_n x) \to \tilde{L}(x)$, the convergence being at every point of $\mathbb{R}$ by the continuity of $\tilde{L}(.)$, and the sequences not being unique, as said before.

Then as convergence is uniform, i.e.

$$\sup_x | F^n(\lambda_n + \delta_n x) - L(x)| = \sup_y | F^n(y) - \tilde{L}\left(\frac{y - \lambda_n}{\delta_n}\right) | \to 0 ,$$

for large n, we can assume that there exists $(\lambda, \delta; \delta > 0)$ such that $F^n(y) \approx \tilde{L}\left(\frac{y - \lambda}{\delta}\right)$.

Also $\tilde{L}(.)$ satisfies the stability equation $\tilde{L}^t(x) = \tilde{L}(\alpha_t + \beta_t x)$, $\beta_t > 0$.

(*) Recall that Poincaré (a mathematician) said Lippman (a physicist) told him that physicists used normal distribution supposing it to be a theorem, while mathematicians studied it supposing it to be experimentally verified.

(**) Recall the justification of the Poisson distribution used by "Student" for the counts of particles in a haemacytomaters.

The heuristic reasoning used by Fisher and Tippett (1927/28, p. 180), shows *approximate* stability of maxima.

Thus, with the uniformity of convergence and the (approximate) stability of maxima, we can use the approximation $F^n(y) \approx \tilde{L}((y - \lambda)/\delta)$.

But what happens when F(.) is one of those "awkward" distributions that are not attracted to one of the $\tilde{L}$'s ? This, in general, is a consequence of high tails (the right-tail for maxima), and the use of "slowing down" transformations as the log(.) or log log (.) can be useful; recall the example given in Part I.

Note that the approximation to $F^n(y)$ depends on the size of the sample (n); accumulated experience and simulation can give some idea when the fit is good or even reasonable.

It should be noted that in the heuristics given before we assumed i.i.d. samples of the same size. If the samples are i.i.d. but not of the same size, a not uncommon situation, a correction can be made and the data fit analysed.

Essentially if two samples are of sizes n and n' we have under general conditions :

$\lambda_{n'} = \lambda_n + \delta_n \log \frac{n'}{n}$, $\delta_{n'} = \delta_n$ *for the Gumbel distribution,*

$\lambda_{n'} = \lambda_n = 0$, $\delta_{n'} = \delta_n (n'/n)^{1/\alpha}$ *for the Fréchet distribution and*

$\lambda_{n'} = \lambda_n = \bar{w}$, $\delta_{n'} = \delta_n (n/n')^{1/\alpha}$ *for the Weibull distribution*

for maxima; see Tiago de Oliveira (1977) and Ramachandran (1974). If we have more than two samples we will shift their parameters to those of a sample of average size.

If the i. or i.d. on both conditions fail we have seen (Part I) that usually the asymptotic results are still valid and the same approximating technique can be used; but, in general, the sample sizes should be larger chiefly when correlation is positive.

Some rules of thumb, integrating accumulated experience, can be suggested.

For maxima :

floods — Gumbel and Fréchet distribution; see Gumbel (1954), (1958) and Thom (1960)

extreme winds — Gumbel and Fréchet distribution; see Thom (1960)

largest waves — Fréchet distribution; see Thom (1970)

longest lives — Gumbel distribution; see Gumbel (1958).

For minima :

droughts — Weibull distribution; see Gumbel (1958) and (1962)

breaking strength of materials — Weibull distribution; see Freudhental and Gumbel (1953) and Weibull (1939).

As a consequence, and with these guide-lines coming from accumulated experience, we can give the first steps in the approximation of $F^n(y)$ by $\tilde{L}((y-\lambda)/\delta)$ and of $1-(1-F(y))^n$ by $1-\tilde{L}(-(y-\lambda)/\delta)$ where $\tilde{L}(.)$ can be $\Psi_\alpha(.)$, $\Lambda(.)$ or $\Phi_\alpha(.)$.

But, also, other quick and more experimental techniques are availables, as follows.

The probability paper for Gumbel distribution

Consider that it is assumed that the i.i.d. sample has the underlying Gumbel distribution

$$\Lambda(x|\lambda, \delta) = \exp\left(-e^{-(x-\lambda)/\delta}\right)$$

which can be written by taking logarithms twice :

$$z = \frac{x-\lambda}{\delta} = -\log(-\log \Lambda)$$

or

$$z = \lambda + \delta\,[-\log(-\log \Lambda)]\,;$$

z is the reduced variate and x the observed one.

Recall that the return period or return time for the level x is

$$T(x|\lambda, \delta) = (1-\Lambda((x-\lambda)/\delta))^{-1}\,.$$

For the probability paper, the iterated logarithms can be built on one of the scales of the paper so that we need only to plot (x,Λ). We do not have only scales for x and Λ but also for T as been in the usual probability paper that follows.

Fig. IV.1 — Gumbel maxima probability paper

The next figure illustrates the reverse probability paper where the scales of (x,Λ) are exchanged, as engineers often do.

Fig. IV.2 — Gumbel maxima probability paper (reverse)

So far we have ignored the problem of determining an estimate of $\Lambda((x_i' - \lambda)/\delta)$ to plot. One could use, i/n as plotting position to go with x_i', the i-th order statistic in a sample of n, which is the usual plotting position for the sample distribution function, used for example in the Kolmogorov-Smirnov bounds. This is equivalent to estimating $\Lambda((x_i' - \lambda)/\delta)$ as the proportion of x_j' s less than or equal to x_i'. Since 1 (and also 0) do not appear on the Λ-scale, we cannot plot $\Lambda((x_n' - \lambda)/\delta)$. If we estimate $\Lambda((x_i' - \lambda)/\delta)$ by $(i - 1)/n$, the proportion of x_j' s less than x_i', we cannot plot $\Lambda((x_1' - \lambda)/\delta)$. As a compromise it was suggested to "split the difference" and use $(i - 1/2)/n$ as the estimate of $\Lambda((x_i' - \lambda)/\delta)$. Another procedure is to use the mean value of $\Lambda((x_i' - \lambda)/\delta)$ as a plotting position, which leads to the value $i/(n + 1)$. Blom (1958) devoted almost an entire book to the problem of obtaining "optimum" plotting positions, based on the idea of Chernoff and Lieberman (1954) that the plotting position should depend on the quantity to be estimated. For example Blom (1958, p. 145) shows that for the *normal distribution*, plotting position $(i - 3/8)/(n+1/4) = (i - .375)/(n+.25)$ leads to a practically unbiased estimate of σ with a mean square error about the same as that of the unbiased best linear estimate, while the plotting position $(i - .5)/n$ leads to a biased estimate of σ with nearly minimum mean square error.

Note that some symmetry of plotting positions must be verified : if $p_{i,n}$ is the plotting position for the x_i' then we should have $p_{i,n} + p_{n+1-i,n} = 1$ which does not happen with the plotting positions i/n and $(i - 1)/n$ but happens with the two other ones. Another perspective, to be used further in the next section, leads to the plotting position $p_i = (i - 0.3)/(n + 0.4)$, also symmetrical; for details see Tiago de Oliveira (1983). Taking $p_{i,n} = (i + A)/(n + B)$ the symmetry condition leads to $B = 2A + 1$ and $p_{i,n} = (i + A)/(n + 1 + 2A)$ and as $0 < p_{i,n} < 1$ for each i

and n we must have $-1 < A$; for $A = -1/2$, $A = 0$, $A = 3/8$ and $A = -.3$ we obtain the given plotting positions. We have $\max_i \left| \frac{i + A}{n + 1 + 2A} - \frac{i + A'}{x + 1 + 2A'} \right| \leq \frac{|A - A'|}{n}$. Note that for very large A all the plotting positions are close to 1/2 which means that it is acceptable to use $|A| < 1$ to have a reasonable spread of the plotting positions in $]0,1[$; in that case the upper bound $|A - A'|/n$ of the maximum error is bounded above by $2/n$ and is practically irrelevant for sample sizes $n \geq 50$, and all $p_{i,n}$ are practically equivalent (*) .

In the graph that follows we will use the classical Gumbel plotting position i/(n+1). Fig. IV.3 that follows was taken from Gumbel (1956) and illustrates the procedure for estimating $\lambda^* = 778.5$ and $\delta^* = 2.839$ for maximum atmospheric pressure (in mm) at Bergen, Norway for the period between 1857 and 1926. If the fitted straight line is extrapolated beyond the fitted values, one would "predict", for example, that a pressure of 793 mm corresponds to a probability of 0.994. That is, pressures of this magnitude have less than 1 chance in 100 of being exceeded in any particular year.

Fig. IV.3 - Annual maxima of atmospheric pressures, Bergen, Norway, 1857-1926 (reverse probability paper)

The use of Gumbel probability paper for quick statistical choice between Weibull, Gumbel and Fréchet distributions for maxima

Here we will describe a graphical technique taken from Tiago de Oliveira (1983). As was shown in the previous section, Z being a reduced random variable, the probability paper for Gumbel distribution was made as follows : from $p = \Lambda(z)$, which can be written as $\Lambda^{-1}(p) = z$, if we note by $y = \Lambda^{-1}(p)$ and graduate the y-axis in the functional scale p, the relation $p = \Lambda(z)$ can be written as the first diagonal $y = z$. In addition, if the true distribution is $G(z|\theta) = \exp\{-(1 + \theta z)_+^{-1/\theta}\}$, the von Mises-Jenkinson formula, the relation $p = \Lambda(y) = G(z|\theta)$ may be written as $y = \Lambda^{-1}(G(z|\theta))$ and the relation between (z,y) (or (z,p)) is the curve in the plane $y = \log(1 + \theta z)/\theta$, for $1 + \theta z > 0$.

(*) Also if we choose A such that for $F(x_n') = p_{n,n}$, $p_{n,n}^n \to 1/2$ (i.e., x_n' is approximately the median of $F^n(x)$) we get $A = \log 2 - 1$ and $p_{i,n} = (i - .307)/(n + .386)$; we get the same result for minima with $1 - (1-p_{i,n})^n \to 1/2$.

Let us now return to the plotting positions. If $p_{i,n}$ is the plotting position for the order statistic z_i ($= z'_{i,n}$ with greater rigour) the distance between the plotted point $(z'_i , p'_{i,n})$ and the first diagonal is $|\Lambda^{-1}(p_{i,n}) - z'_i|$ horizontally or vertically or $|\Lambda^{-1}(p_{i,n}) - z'_i| / \sqrt{2}$ orthogonally. The distance $|\Lambda^{-1}(p_{i,n}) - z'_i|$ is minimized in mean if $\Lambda^{-1}(p_{i,n})$ is the median of z'_i (with Gumbel distribution) or $p_{i,n}$ is the median of Beta $(i, n-1-i)$ distribution; a good approximation is to take the one given before, $p_{i,n} = \frac{i - 0.3}{n+0.4}$; note that we are not using the least squares approach but we are minimizing $\max_i |\Lambda^{-1}(p_{i,n}) - z'_i|$.

Having chosen one of these plotting positions, let us consider the general case, i.e. observations with unknown location and dispersion parameters λ and δ, $z_i = (x_i - \lambda)/\delta$ being the reduces values. Then if $\chi_\theta(p)$ is the p-quantile of $G(z|\theta)$, i.e. $G(\chi_\theta(p)|\theta) = p$, the p-quantile of $G((x - \lambda)/\delta|\theta)$ is, evidently, $\lambda + \delta\, \chi_\theta(p)$. Consider then the random variable

$$v = \frac{x - (\lambda + \delta\, \chi_\theta(r))}{(\lambda + \delta\, \chi_\theta(s)) - (\lambda + \delta\, \chi_\theta(r))} = \frac{z - \chi_\theta(r)}{\chi_\theta(s) - \chi_\theta(r)}$$

with $0 < r < s < 1$; v is obviously independent of the location and dispersion parameters.

Clearly we are going to plot *estimated* v'_i where the r and s quantiles are estimated by the sample quantiles by $x'_{[nr]+1}$ and $x'_{[ns]+1}$.

The relation $p = \Lambda(y) = G(z|\theta)$ can then be written as $y = \Lambda^{-1}(G(\chi_\theta(r) + (\chi_\theta(s) - \chi_\theta(r))v|\theta)) = y_\theta(v)$, for r and s fixed. Taking $v = 0$ and $v = 1$ we see that all curves $y_\theta(v)$ pass through the points $(0, \chi_o(r))$ and $(1, \chi_o(s))$, or $(0,r)$ and $(0, s)$ if we use the functional scale for y. Thus between $v = 0$ and $v = 1$ the curves $y_\theta(v)$ will with difficulty be separated from the first diagonal, chiefly taking into account that the plotted points, even using real and not estimated reduced observations, do not fall exactly on the first diagonal for Gumbel distribution. In summary, the points between $v = 0$ and $v = 1$ are *lost* for separation of models; it is, thus, natural to use about 1/3 of the observations for the zone $v \in [0,1]$ and 2/3 outside; a smaller percentage in this zone would introduce instability in the implicit estimation of λ and δ by the quantiles, as the denominator would be small.

With this rule of thumb, we have taken $\chi_o(r) = 0$ and $\chi_o(s) = 1$ or $r = \Lambda^{-1}(0) = e^{-1} = 0.3678794$ and $s = \Lambda^{-1}(1) = \exp(-e^{-1}) = 0.6922006$ (note that, as said before, $r \approx 1/3$ and $s \approx 2/3$); practically we can take $r = 0.368$ or even $r = 0.37$ and $s = 0.692$ or even $s = 0.69$.

Let us now compute the curves $y_\theta(v)$. As $\chi_\theta(p) = \dfrac{(-\log p)^{-\theta} - 1}{\theta}$ and so $\chi_\theta(e^{-1}) = 0$ and $\chi_\theta(\exp(-e^{-1})) = (e^\theta - 1)/\theta$ we get

$$y_\theta(v) = \frac{\log\,(1 + (e^\theta - 1)v)}{\theta}$$

which is evidently defined when $1 + (e^\theta - 1)\, v > 0$; note that $y_\theta(v)$ is convex for $\theta < 0$ (Weibull model) and concave for $\theta < 0$ (Fréchet model), $y_0(v)$ (Gumbel model) being a straight line.

Thus in a Gumbel probability paper we plot a bundle of curves $(v, y_\theta(v))$ for different values of θ and the quick statistical choice is made according to the location of the ploted points (v_i', p_i) close to the diagonal (Gumbel distribution) or in the Fréchet or Weibull zones as shown in the Fig. IV.4. Estimation of the parameters is to be made by the analytical methods of the next three chapters.

As an example consider the Table I of maximum wind velocities in km/h between 1941-1970 in Lisbon.

TABLE I

1941	129.0	1951	96.0	1961	86.0
1942	117.0	1952	72.0	1962	91.0
1943	100.0	1953	98.0	1963	96.0
1944	100.0	1954	85.0	1964	89.0
1945	132.0	1955	124.0	1965	90.0
1946	94.0	1956	108.0	1966	89.0
1947	108.0	1957	102.0	1967	89.0
1948	113.0	1958	102.0	1968	84.0
1949	96.0	1959	112.0	1969	107.0
1950	113.0	1960	107.0	1970	111.0

The values of the maximum likelihood estimators are $\hat{\lambda} = 94.71$ and $\hat{\delta} = 12.49$; the 1/e - quantile and exp (– 1/e) -quantile needed for computation of the v_i' are Q(.367) = 96.0 and Q(.692) = 108.0; we took, as said before, $p_i = 1/(n+1)$.

Fig. IV.4 - Plotting of (v_i', p_i) for the maximum wind velocities in km/h in Lisbon, 1941-1970; Fréchet and Weibull zones shown.

As regards m-th extreme, the simplest way is to try the statistical choice for maxima (or minima) because then the limiting distribution is well known and was given in Part I.

Simple statistical choice between the extreme models

We will use a simple statistic proposed in Gumbel (1965) for the study of the Fréchet distribution. In Tiago de Oliveira and Gomes (1984) we developed the asymptotic theory of this statistic for all distributions. Let us denote for a sample $(x_1, \ldots, x_n)$, from the distribution $G(z|\theta)$ by $x_1' = \min(x_i)$, $x'_{[n/2]+1} = \text{med}\{x_i\}$ and $x_n' = \max\{x_i\}$ the maximum, the medium and the minimum of the sample. We consider the statistic

$$Q_n = Q_{n,\theta} = \frac{x_n' - x'_{[n/2]+1}}{x'_{[n/2]+1} - x_1'} = \frac{\max\{x_i\} - \text{med}\{x_i\}}{\text{med}\{x_i\} - \min\{x_i\}}$$

which is, evidently, location and dispersion parameter-free.

We have shown in Tiago de Oliveira and Gomes (1984) that if $\theta = 0$, $G(z|0^+) = G(z|0^-) = \Lambda(z)$, there exist $\beta_{n,0}$ and $\alpha_{n,0} > 0$ such that $\frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} \xrightarrow{w} \Lambda(z)$; a choice is $\beta_{n,0} = \frac{\log n + \log\log 2}{\log\log n - \log\log 2}$, $\alpha_{n,0} = 1/\log\log n$; if $\theta > 0$, $G(z|\theta) = \Phi_{1/\theta}(\theta z + 1)$, there exist $\beta_{n,\theta}$ and $\alpha_{n,\theta} > 0$ such that $\frac{Q_n - \beta_{n,\theta}}{\alpha_{n,\theta}} \xrightarrow{w} G(z|\theta) = \Phi_{1/\theta}(\theta z + 1)$; a choice is $\beta_{n,\theta} = \frac{n^\theta - (\log 2)^{-\theta}}{(\log 2)^{-\theta} - (\log n)^{-\theta}}$; $\alpha_{n,\theta} = \theta(n \log 2)^\theta$ or more simply $\beta'_{n,\theta} = (n \log 2)^\theta$, $\alpha'_{n,\theta} = \theta \beta'_{n,\theta}$; if $\theta < 0$, $G(z|\theta) = \Psi_{-1/\theta}(-\theta z - 1)$, there exist $\beta_{n,\theta}$ and $\alpha_{n,\theta} > 0$ such that $\frac{Q_n - \beta_{n,\theta}}{\alpha_{n,\theta}} \xrightarrow{w} 1 - \Lambda(-z)$; a choice is $\beta_{n,\theta} = \frac{n^\theta - (\log 2)^{-\theta}}{(\log 2)^{-\theta} - (\log n)^{-\theta}}$; $\alpha_{n,\theta} = -\theta(\log 2)^{-\theta}$

$(\log n)^{\theta-1}$ or even more simply for $\theta < 1$, $\beta'_{n,\theta} = (\frac{\log n}{\log 2})^{\theta}$ and similarly $\alpha'_{n,\theta} = \alpha_{n,\theta} = -\frac{\theta}{\log n}\beta'_{n,\theta}$.

Two intuitive conclusions may be drawn from these results:

1 — There is a "discontinuity" between $\theta \geq 0$ and $\theta < 0$; for $\theta \geq 0$ the influent term is the maximum but for $\theta < 0$ the influent term is the minimum; this explains, to a certain extent, why we cannot have a simple common expression for $\beta_{n,\theta}$ and $\alpha_{n,\theta}$.

2 — As for $\theta \geq 0$, $\alpha_{n,\theta} \to \infty$ and for $\theta < 0$, $\beta_{n,\theta} \to \infty$ 0 it seems natural to decide for the Fréchet distribution ($\theta > 0$) if Q_n is large and for the Weibull distribution ($\theta < 0$) if Q_n is small; then the decision rule will be:

Choose $0 < b < a < +\infty$ *and decide for the Gumbel distribution when* $b \leq \frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} \leq a$, *for the Fréchet distribution when* $\frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} > a$, *and for the Weibull distribution when* $\frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} < b$, *where* $\beta_{n,\theta} = \frac{\log n + \log\log 2}{\log\log n - \log\log 2}$ *and* $\alpha_{n,\theta} = (\log\log n)^{-1}$.

Let us compute, for $b < a$, the probabilities of correct decision; using the Khintchine convergence of types theorem :

1) $\theta = 0$: $\quad \text{Prob}\{b \leq \frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} < a\} \to \Lambda(a) - \Lambda(b)$;

2) $\theta > 0$: $\quad \text{Prob}\{\frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} > a\} \to 1 - G(-1/\theta|\theta) = 1$:

3) $\theta < 0$: $\quad \text{Prob}\{\frac{Q_n - \beta_{n,0}}{\alpha_{n,0}} < b\} \to 1 - G(-1/\theta|\theta)$.

The technique is thus consistent in the usual sense, whatever $0 < b < a < +\infty$ may be; in fact if the decision technique is such that for $\theta = 0$ the limit probability of correct decision is $\Lambda(a) - \Lambda(b)$, but if $\theta \neq 0$ the limit probability of correct decision is 1. We could improve the situation by putting the significance level $\alpha_n = 1 - (\Lambda(a_n) - \Lambda(b_n)) \to 0$ with convenient restrictions in b_n and a_n. But this does not seem to be needed because it is well known that for large α (i.e, for small θ in von Mises-Jenkinson form) both the Fréchet and the Weibull distributions are very close to the Gumbel distribution. Following on from the previous remark, it seems natural to solve only the statistical trilemma ($\theta < 0$, $\theta = 0$ or $\theta > 0$) and not a statistical

dilemma or one-sided test of hypothesis of $\theta = 0$ vs. $\theta > 0$ or $\theta = 0$ vs. $\theta < 0$. The latter would correspond to the partial acceptance of $\theta \geq 0$ or $\theta \leq 0$ which lead to a one-sided test as a consequence of assuming stronger previous knowledge or a stronger plausibility. Thus we will not deal with one-sided tests.

It remains now to obtain the one equation for (b , a) and the best seems, from what has been said, that the length of the Gumbel decision interval $a - b$ should be the smallest possible with the condition $\Lambda(a) - \Lambda(b) = 1 - \alpha$. The use of Lagrange multipliers gives the equations

$$\Lambda(a) - \Lambda(b) = 1 - \alpha$$

$$\Lambda'(a) = \Lambda'(b) .$$

The values corresponding to the usual significance levels are :

α	b	a
.050	– 1.561334	3.161461
.025	– 1.719620	3.841321
.010	– 1.893530	4.740459
.001	– 2.222951	7.010001

Examples of quick exploration of data, as well as the analysis of univariate models and statistical choice, can be found in the "Some Case Studies" chapter.

Chapter V

STATISTICAL ANALYSIS OF THE GUMBEL MODEL

Introduction

In what follows we will assume that the samples of maxima or minima we are dealing with have a Gumbel distribution for maxima $\Lambda((x-\lambda)/\delta)$, sometimes denoted $\Lambda(x|\lambda,\delta)$, or for minima $1-\Lambda(-(x-\lambda)/\delta)$. We will deal only with maxima, minima being dealt with, as said, by the relation $\min(x_i) = -\max(-x_i)$. Recall that the Gumbel distribution plays a central role in extreme value theory, being sometimes called *the* extreme value distribution; it is the limit of both Fréchet and Weibull distributions according to $\theta \downarrow 0$ or $\theta \uparrow 0$ in the von Mises-Jenkinson formula and those distributions can be transformed to a Gumbel distribution by simple (but not always convenient) logarithmic transformations.

Characteristics of the Gumbel distribution

As the distribution $\Lambda((x-\lambda)/\delta)$ is not symmetrical, the natural location parameter need not be the mean; also the convenient scale parameter will not be the standard deviation. Define the *standardized* or *reduced* random variable as

$$Z = (X-\lambda)/\delta \quad \text{or} \quad X = \lambda + \delta Z .$$

The graph of the probability density $\frac{1}{\delta} e^{-(x-\lambda)/\delta} \exp(-e^{(x-\lambda)/\delta})$ for $\lambda = 0$ and $\delta = 1$ is given in Fig.V.1.

Fig. V.1 – Reduced Gumbel density (for maxima)

The moment generating function of Z is $m_Z(\theta) = \Gamma(1-\theta)$ (for $\theta < 1$) and that of X is $m_X(\theta) = e^{\lambda\theta}\ m_Z(\delta\theta) = e^{\lambda\theta}\ \Gamma(1-\delta\theta)$ $(\delta\theta < 1)$; the corresponding characteristic functions are $\Phi_Z(t) = \Gamma(1-it)$ and $\Phi_X(t) = e^{i\lambda t}\ \Gamma(1-i\delta t)$.

All the moments exist as follows from the infinite differentiability of $\Gamma(1-it)$. Let us recall that when the mean value μ exists, the central moments (i.e., the moments about μ) have the expressions, if they exist,

$$\mu_i = M((X-\mu)^i)$$

with the variance $\sigma^2 = \mu_2$ and the frequently used linearly invariant measures of skewness and kurtosis,

$$\beta_1 = \mu_3/\sigma^3 \quad , \quad \beta_2 = \mu_4/\sigma^4 \quad .$$

In our case we should consider the random variables Z and $X = \lambda + \delta Z$ with distribution functions $\Lambda(z)$ and $\Lambda((x-\lambda)/\delta)$, and the moments and coefficients should be noted $\mu_{Z,i}$ and $\mu_{X,i}$, the values of β_1 and β_2 being independent of λ and δ. We shall, for simplicity, drop the index Z always and we have

$$\mu_i = \mu_{Z,i} = M(Z-\mu)^i$$

$$\mu_{X,i} = M(Z-\mu_X)^i = \delta^i\ \mu_i$$

with
$$\mu_X = M(X) = \lambda+\delta M(Z) = \lambda+\delta\mu = \lambda+\delta\mu_Z \quad .$$

Note also that an analogous relation is valid for quantiles :

$$\text{if } \Lambda(\chi_p) = \Lambda(\chi_{Z,p}) = p \quad \text{we have} \quad \chi_{X,p} = \lambda + \delta\,\chi_{Z,p} = \lambda + \delta\,\chi_p \quad .$$

For the mode μ', as for the median $\tilde{\mu}$ which is the .5–quantile, we have also

$$\mu'_X = \lambda + \delta\,\mu'_Z = \lambda + \delta\,\mu'$$

and
$$\tilde{\mu}_X = \lambda + \delta\,\tilde{\mu}_Z = \lambda + \delta\,\tilde{\mu}$$

We now have the following values for the coefficients associated with Z, see Gumbel (1958):

$\mu_1 = \gamma = .57722\ldots$ = Euler's constant;

$\mu' = 0 = \chi_{1/e}$;

$\tilde{\mu} = -\log\log 2 = .36651\ldots$;

$\chi_p = -\log\log(p^{-1})$;

$\mu_2 = \sigma^2 = \pi^2/6 = 1.64493\ldots$;

$\sigma = \sqrt{\pi^2/6} = 1.28255\ldots$;

$\mu_3 = 2.40411\ldots$;

$\mu_4 = 3\pi^4/20 = 14.61136\ldots$;

$\beta_1 = 1.13955$;

$\beta_2 = 5.40000$.

The return period, for the overpassing the level (or threshold) a, is, as said before,

$$T(a) = (1 - \Lambda((a - \lambda)/\delta))^{-1} ;$$

on average $T(a)$ trials will be made until the maximum overpasses a, after the last overpassing. Note that the probability of overpassing a before or at the mth trial (year, for instance, if we are thinking of dams) is $1 - (1 - p)^m$ where $p = 1 - \Lambda((a - \lambda)/\delta)$. Thus is can be written as $1 - (1 - \frac{1}{T(a)})^m$.

The probability that the overpassing will occur before or at the return period is

$$1 - (1 - \frac{1}{T(a)})^{[T(a)]} = 1 - e^{-1} = .63212 \quad \text{for large } T(a).$$

The maximum likelihood estimator of $T(a)$ is $T(a) = N_k/k$, where N_k is the time to the k–th occurrence of the overpassing. This is not very convenient and it is better to estimate $T((a - \lambda)/\delta)$ by $T((a - \hat{\lambda})/\hat{\delta})$, $(\hat{\lambda}, \hat{\delta})$ being the maximum likelihood estimators of (λ,δ) (to be dealt with in the next section), or with other estimators of (λ, δ).

A simple and useful conservative approximation for $T(a)$, given by Fuller (see Tiago de Oliveira (1972) is

$$T(a) \approx \tilde{T}(a) = \exp\{\frac{a - \lambda}{\delta}\}$$

with the bounds $\tilde{T}(a) \leq T(a) \leq \tilde{T}(a)\ (1-1/2\ \tilde{T}(a))^{-1}$.

Usually, when $T(a) = 100$, engineers call a "the largest flood in 100 years" or "the 100–years flood"; the expressions in quotes are misnomers because the largest flood in 100 years is a random variable. In a sense, *on average*, we expect to have to wait 100 years to observe a maximum yearly flood greater than or equal to a.

From $\Lambda((a-\lambda)/\delta)) = 1 - \dfrac{1}{T((a-\lambda)/\delta)}$ we see that "the 100–year flood" is the .99 quantile.

If a random variable has a Gumbel distribution with parameters $\lambda, \delta > 0$ the interval $[\lambda + b\,\delta, \lambda + a\,\delta]$ has the probability $\Lambda(a) - (b)$ and its length is $\delta(a - b)$. The shortest interval with a fixed probability β is given by the equations (obtained by the Lagrange multipliers method)

$$\Lambda(a) - \Lambda(b) = \beta = 1 - \alpha$$

$$\Lambda'(a) - \Lambda'(b) = 0$$

or

$$\Lambda(a) - \Lambda(b) = \beta = 1 - \alpha$$

$$a + e^{-a} = b + e^{-b}.$$

A table for some relevant values of is given below.

TABLE

α	β	b	a	$\Lambda(b)$	$\Lambda(a)$	a – b
1.	0.	0.0	0.0	1/e	1/e	0.0
.50	.50	– .651256	.831143	.146908	.646908	1.482399
.20	.80	–1.125023	1.787968	.045946	.845946	2.912991
.10	.90	–1.369180	2.479152	.019602	.919602	3.848332
.05	.95	–1.561344	3.161461	.008521	.958521	4.722805
.01	.99	–1.893530	4.740459	.001305	.991305	6.633989

As could by the expected from the skewness of the distribution the shorted interval is not symmetrical; for instance the shorted intervals containing 90%, 95% and 99% of the population are roughly $[\lambda - 1.4\,\delta, \lambda + 2.5\,\delta]$, $[\lambda - 1.6\,\delta, \lambda + 3.2\,\delta]$ and $[\lambda - 1.9\,\delta , \lambda + 4.7\delta]$. The interval with $a = 2$, $b = 5$, i.e., $[\lambda - 2\,\delta, \lambda + 5\,\delta]$, contains more than 99% of the population.

Point Estimation

The earliest methods of point estimation are the method of moments and the method of maximum likelihood. They were evidently used for Gumbel distribution. Thus these procedures will be given first, followed by some other treatments involving best linear unbiased estimators and best linear invariant estimators.

1 – Maximum Likelihood Estimators (MLE)

The logarithm of the likelihood of a i.i.d. sample $(x_1, x_2 , \ldots , x_n)$ of n observations with the distribution function $\Lambda((x-\lambda)/\delta))$ is

$$\log L(\lambda , \delta|x_i) = - n \log \delta - \sum_1^n e^{-(x_i-\lambda)/\delta} - \sum_1^n (x_i-\lambda)/\delta \quad .$$

Setting the partial derivatives of log L with respect to λ and δ equal to zero, we get, with easy algebra, the maximum likelihood equations:

$$\hat{\delta} = \bar{x} - \sum_1^n x_i\, e^{-x_i/\hat{\delta}} \,/\, \sum_1^n e^{-x_i/\hat{\delta}}$$

$$\hat{\lambda} = - \hat{\delta} \log \left(\sum_1^n e^{-x_i/\hat{\delta}} / n \right) .$$

The second equation is very easy to deal with once the first one is solved. But this one, involving only $\hat{\delta}$, is intractable without computers. One could use an iterative procedure, starting with the moment estimators for λ and δ to follow. However, Kimball developed a simple procedure to get approximate solutions. This is described by him in the sections of Gumbel (1958) which he wrote, summarizing some of his papers. Panchang and Aggarwal (1962) in the "Poona Report" were among the first to use the iterative procedure indicated above, with the use of computers.

Maximum likelihood estimators are asymptotically efficient and unbiased with the variance–covariance matrix given by the inverse of

$$-n \begin{bmatrix} M\left(\frac{\partial^2 \log g}{\partial \lambda^2}\right) & M\left(\frac{\partial^2 \log g}{\partial \lambda \, \partial \delta}\right) \\ M\left(\frac{\partial^2 \log g}{\partial \lambda \, \partial \delta}\right) & M\left(\frac{\partial^2 \log g}{\partial \delta^2}\right) \end{bmatrix}$$

with $g = g(x \mid \lambda, \delta) = \frac{1}{\delta} \Lambda'\left(\frac{x-\lambda}{\delta}\right)$, the density of Gumbel distribution with parameters (λ, δ).

It is known that:

The estimators $(\hat{\lambda}, \hat{\delta})$ are asymptotically binormal with mean values (λ,δ), the asymptotic variance–covariance matrix is

$$\frac{1}{n} \begin{bmatrix} \left(1 + \frac{6(1-\gamma)^2}{\pi^2}\right) \delta^2 & \frac{6(1-\gamma)}{\pi^2} \delta^2 \\ \frac{6(1-\gamma)}{\pi^2} \delta^2 & \frac{6}{\pi^2} \delta^2 \end{bmatrix}$$

and the asymptotic correlation coefficient is

$$\rho(\hat{\lambda}, \hat{\delta}) = \left(1 + \frac{\pi^2}{6(1-\gamma)^2}\right)^{-1/2} = .31307 .$$

2 – Moments Estimators (ME)

From the relations

$$M(X) = \lambda + \delta\gamma$$

$$V(X) = \frac{\pi^2}{6} \delta^2$$

we get as estimators by the method of moments

$$\delta^* = \frac{\sqrt{6}\, s}{\pi} = .77970\, s$$

$$\lambda^* = \bar{x} - \frac{\sqrt{6}\,\gamma}{\pi}\, s = \bar{x} - .45501\, s$$

where $\bar{x}$ and s^2 are the average ($\sum_1^n x_i/n$) and the empirical variance ($\sum_1^n (x_i - \bar{x})^2/n$).

(λ^*, δ^*) is asymptotically binormal but its efficiency is very low (48%); see Tiago de Oliveira (1963). The usefulness of this method, now that computers are used, is to give a simple first estimate as a seed for iteration to solve the maximum likelihood estimator equations.

It should be noted that Gumbel (1958) and Posner (1965) developed methods close to the method of moments (called "simplified" and "modified"), using the plotting positions and a fitted linear regression.

Mann (1968), by simulation and also using Lieblein's (1954) simulation, has shown that, for estimating quantiles, the Posner (1965) method gives poor estimators for either small or large samples, although the Gumbel (1958) method is relatively good for small samples. The BLUE methods – see below – are preferable.

3 – Best Linear Unbiased Estimators (BLUE)

Lieblein (1954) was the first to recognize that as Gumbel plotting used order statistics, information was wasted if estimators did not use order statistics. He then considered estimating the linear combination

$$\theta = \theta\,(a\,,\,b) = a\,\lambda + b\,\delta,$$

with, as special cases, $\lambda = \theta\,(1,0)$, $\delta = \theta\,(0,1)$ and the percentile

$$\theta_p = \theta\,(1\,,\,\Lambda^{-1}(p)) = \lambda + \delta\,\Lambda^{-1}(p)$$

where $\Lambda^{-1}(p) = -\log(-\log p)$. He proposed to estimate θ by θ^*, a linear combination of order statistics in which the weights are determined to yield minimum variance unbiased estimators. More specifically, θ^* based on double censored data is given by

$$\theta^* = \sum_{i=r+1}^{N-s} w_i\, x_i\,,$$

where, obviously, $w_i = a\,a_i + b\,b_i$. Note that for $s = r = 0$ we have the complete sample.

The values of a_i and b_i for which $M(\theta^*) = \theta$ and $V(\theta^*)$ is minimized, are functions of n, r, s and may be determined by the generalized least squares procedure of Lloyd (1952), based on the generalized Gauss–Markov theorem. As the distribution of order statistics Z_i' is parameter–free, the mean values, variances, and covariances depend only on the standardized distribution function and the sample size. For example

$$M(Z_i'^{\,k}) = \frac{n\ !}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} z^k \, \Lambda^{i-1}(z) \, [1-\Lambda(z)]^{n-i} \, \Lambda'(z) \, dz$$

Those moments were given numerically by Lieblein (1953).

We start with the fact, observed earlier, because $X_i' = \lambda + \delta Z_i'$, that

$$M(Z_i'^{\,k}) = \lambda + \delta \mu_i, \quad C(X_i', X_j') = \delta^2 \sigma'_{ij}$$

where $\mu'_i = M(Z_i')$ and σ'_{ij} is the (i,j) the element of the variance–covariance matrix of $Z = (Z_1', Z_2', \ldots, Z_n')^T$. The generalized least squares theorem then states that if X is the vector of observations,

$$Z = (Z_1', Z_2', \ldots, Z_n')^T$$

having $M(X) = A\,\theta$ with $\theta = [\lambda, \delta]^T$, and $C(X) = C(\lambda + \delta z) = [\sigma'_{ij}]\,\delta^2 = V\,\delta^2$ where $A = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \mu'_1 & \mu'_2 & & \mu'_n \end{bmatrix}^T$, both A and V being known and $\det(V) \neq 0$, the least squares estimator of the vector $\theta = [\lambda, \delta]^T$ is the value that minimizes

$$(X - A\,\theta)^T \, V^{-1} \, (X - A\,\theta).$$

The minimum solution is the one that satisfies the so–called normal equations,

$$A^T V^{-1} A\,\theta^* = A^T V^{-1} X \qquad \text{and so}$$

$$\theta^* = (A^T V^{-1} A)^{-1} A^T V^{-1} X$$

and the variance–covariance matrix of θ^* being

$$C(\theta^*) = (A^T V^{-1} A)^{-1} \delta^2$$

The estimator θ^* is unbiased and the θ_i^* have minimum variance in the class of unbiased linear statistics.

Calculation of the mean values variances and covariances of the order statistics Z_i' and the weights a_i, b_i are tedious and require computers even for small values of n. They are now tabulated by Mann (1967) (Appendix C) to N = 2(1)25 , r = 0 , s = 0 (1) N–2, after some previous work by Lieblein and Zellen (1956).

4 – Best Linear Invariant Estimators (BLIE)

The unbiased condition of the previous section can be relaxed by defining loss as the mean squared error divided by δ^2 and by considering weighted sums of Z_i', i = 1,2, ..., r ≤ n, as estimators of (λ , δ) and of the quantiles with mean squared error, invariant under translations and multiplications. The estimators among those with smallest mean squared error were called the best linear invariant estimators. The BLUE of Lieblein and Zelen (1956) are also invariant. Mann (1969) has shown that those estimators of (λ , δ), and of the quantiles, are linear functions of the BLUE of λ and δ. The mean squared errors of the BLIE are uniformly less than those of the corresponding BLUE, Mann (1968). She says that when r << n, the ratios of the mean losses of the BLUE relative to those of the corresponding ones of the BLIE are large; for r = 2, n = 2(1)25 the mean losses of the BLUE of λ and δ increase with n from .659 to 8.24 and .712 to .980, respectively. The corresponding ranges of losses when the BLIE are used are .657 to 4.49 and .416 to .495.

Furthermore, for m = 2, the losses of the BLIE of the quantiles and δ are evidently less than or equal to those of the corresponding MLE, which are also invariant. Tables of weights for obtaining the BLIE and the losses of the estimators are available in Mann (1967, Appendix C).

Asymptotically efficient linear estimators of location and dispersion parameters were obtained and shown to be asymptotically normal by Chernoff *et al.* (1967). They are invariant and quite efficient with respect to the BLUE, even for fairly small n , Mann (1968). Johns and Lieberman (1966) used this procedure to determine the weights in the asymptotically efficient linear estimators of λ and δ, but there exist (or existed) computer difficulties in comparing the losses of those methods as moment calculations (mean values, variances and covariances) of order

statistics involve large rounding errors, requiring multiple–precision computing techniques. Some comparison, based on limited computation, appears in Mann (1968).

5– Further methods of linear point estimation

Other approximating methods exist to generate the weights for linear estimators of (λ , δ).

We will enumerate some of them, giving sufficient references; in general they do not need, as does Johns and Lieberman's (1966) procedure, the tedious and error – prone computation of the first and second moments of order statistics.

A method of splitting the sample at random in blocks of 4,5 or 6, computing the averages of the blocks, estimating the parameters (λ , δ) from those averages with convenient coefficients, and averaging the estimates, was developed by Lieblein and Zellen (1956) and also studied in Tiago de Oliveira (1972). It was shown there that the asymptotic efficiency for the estimator $\chi^*_{X,p} = \lambda^* + \chi_p \delta^*$ of the quantile χ_p has the expression

$$\text{eff}(\chi^*_{X,p}) = \frac{1 + .60793\ (.42278 + \chi_p)^2}{6\ (.19117 + .06274\ \chi_p + .13196\ \chi_p^2)}$$

having as $\chi_p \to \pm\infty$ the limit value 0.77, with $\text{eff}(\chi^*_{X,1/e}) = .97$ for the mode and $\text{eff}(\chi^*_{X,p}) > .77$ for $\chi_p > -.99802$ whose probability is $\Lambda(-.98) = .066$. See in Fig. V.2 the graph of the efficiency.

Fig. V.2 – Asymptotic efficiency of Lieblein — Zellen method

Let us also refer to a method due to Downton (1966), developed for Gumbel minima distribution, now translated to the maxima data.

Denoting by $\bar{x} = \sum_1^n x_i/n$ and $\bar{w} = \frac{2}{n(n-1)} \sum_1^n i\ x'_i$, where x'_i are the usual order statistics, we know that $\bar{x}$ and $\bar{w}$ have mean values $\lambda + \gamma\delta$ and $\lambda + (\gamma + \frac{n-1}{n+1} \log 2)\ \delta$. The unbiased estimators

$$\lambda^* = \bar{x} - \gamma\,\delta^*$$

$$\delta^* = \frac{n+1}{(n-1)\log 2}\,(\bar{w} - \bar{x})$$

(note that $\bar{w} > \bar{x}$) have the asymptotic efficiency for the estimation of the p–quantile

$$\text{eff}(\chi_{X,p}) = \frac{1 + .60793\,(.42278 + \chi_p)^2}{1.112825 + .459414\,\chi_p^2 + .804621\,\chi_p^2}$$

with a graph analogous to the previous one.

See also Tiago de Oliveira (1972) for details. In this paper it was also searched the asymptotically best sample quantiles π and π' $(0 < \pi < \pi' < 1)$ such that the variance of the estimators $\lambda^* = a\,x'_{\pi} + (1-a)\,x'_{\pi'}$, , $\delta^* = b\,(x'_{\pi'} - x'_{\pi})$, was minimum. The correct results (the ones in the paper had a computation error) are $\pi = .07$, $\pi' = .76$, $a = -.43065$, $b = .44032$ and the asymptotic efficiency is $\text{eff}\,(\lambda^*, \delta^*) = 40.8\%$. For more details see in the Chapter IX the section "Estimation using two or three quantiles" .

Blom (1958) and in Saharan and Greenberg (1962) developed the nearly best unbiased linear estimators (NBULE) making use of the quantiles $\Lambda^{-1}(i/(n+1))$ with relation to the order statistics X'_i, which Hassanein (1964) applied to the Gumbel distribution. The method can be extended to any distribution having only location and dispersion parameters, directly or by transformation.

Lieblein (1954) proposed a technique using a finite number (in general 3) of empirical quantiles. Hassanein (1964) increased the number of quantiles to be used for the estimation of λ and δ. With the sub–sample $(x'_{a1}, x'_{a2}, \ldots, x'_{an})$ where $1 \le a_1 < a_2 < \ldots < a_k \le n$, for $k = 2(1)10$, and with the results of Ogawa (1951), he determined the quantiles to be used which maximize the joint efficiency of the estimators.

Recently Kubat and Epstein (1983) sought the sets of 2 (or 3) order statistics x'_r and $x'_{r'}$ with $(1 \le r = [n\,\pi] < r' = [\,n\,\pi\,] \le n)$ and coefficients a and b, such that (λ^*, δ^*) with $\lambda^* = a\,x'_r + (1-a)\,x'_{r'}$, , $\delta^* = b\,(x'_{r'} - x'_r)$ were asymptotically unbiased, asymptotically binormal, and with the best efficiency that could be attained to estimate a quantile χ_p with $\pi < p < \pi'$.

Let us refer in greater detail to the method of block partitioning, which seems simple to execute and does not depend on knowledge of the coefficients of linear combination of the observations, although it needs, once and for all, the computation of efficiency and the best choice of splitting points; see Kubat (1982) for other modifications. The method can be extended to Fréchet and Weibull distributions, as will be seen later.

Take $r < s$ such that $1 < r < s < n$. The averages are

$$\bar{x}_1 = \frac{1}{r}\sum_{1}^{r} x'_i < \bar{x}_2 = \frac{1}{s-r}\sum_{r+1}^{s} x'_i < \bar{x}_3 = \frac{1}{n-s}\sum_{s+1}^{n} x'_i \text{ , with probability 1 .}$$

It can be shown that $(\bar{x}_1 , \bar{x}_2 , \bar{x}_3)$ is asymptotically trinormal and we can seek coefficients a_j and b_j $(j = 1,2,3)$ such that the random pair

$$\lambda^* = \lambda^*(x_i) = \sum_{1}^{3} a_i \bar{x}_i$$

$$\delta^* = \delta^*(x_i) = \sum_{1}^{3} b_i \bar{x}_i$$

which is always asymptotically binormal has minimum variance; the condition of invariance (i.e., such that $\lambda^*(\alpha + \beta x_i) = \alpha + \beta \lambda^*(x_i)$, $\delta^*(\alpha + \beta x_i) = \beta \delta^*(x_i)$, $(\beta > 0)$ implies $\sum_{1}^{3} a_i = 1$) and $\sum_{1}^{3} b_i = 0$. Then we can seek the best r and s in terms of efficiency and thus compute the a_i and b_i.

Further comments

As is well known, the method of maximum likelihood is asymptotically the best and its use is to be recommended.

The other methods (from the moments and the various linear estimators) are not asymptotically efficient and can be considered either as quick methods, if we use only a calculator (and tables if needed), or as a seed for the iterative solution of the $\hat{\delta}$ equation. If we do not use tables of coefficients, the most practical methods are those of moments — as said with efficiency

48% – or the linear methods with 2 quantiles of Tiago de Oliveira (1972) and of Kubat and Epstein (1983) or of Downton (1966).

Confidence Interval Estimation

In principle we should always use the maximum likelihood estimator, either for the p–quantile also with a confidence interval, or for the parameter point.

The maximum likelihood estimator $\hat{\lambda} + \chi_p \hat{\delta}$ is asymptotically normal with mean value $\lambda + \chi_p \delta$ and variance $(1 + \frac{6}{\pi^2}(1 - \gamma + \chi_p)^2)^{1/2} \frac{\delta^2}{n}$.

As

$$\sqrt{n} \frac{(\hat{\lambda} + \chi_p \hat{\delta}) - (\lambda + \chi_p \delta)}{(1 + \frac{6}{\pi^2}(1 - \gamma + \chi_p)^2)^{1/2} \delta}$$

is asymptotically standard normal and, as $\hat{\delta} \xrightarrow{P} \delta$, by the δ-method we know that:

$\sqrt{n} \frac{(\hat{\lambda} + \chi_p \hat{\delta}) - (\lambda + \chi_p \delta)}{(1 + \frac{6}{\pi^2}(1 - \gamma + \chi_p)^2)^{1/2} \hat{\delta}}$ *is asymptotically standard normal, which gives a simple confidence interval for* $\lambda + \chi_p \delta$ *independent of the parameter* δ as did not happen with the first asymptotic relation.

We could also compute the bias of estimator of the p–quantile $M(\hat{\lambda} - \lambda) + M(\hat{\delta} - \delta) \chi_p$ as well as its variance $V(\hat{\lambda} + \chi_p \hat{\delta}) = V(\hat{\lambda}) + 2\, C(\hat{\lambda}, \hat{\delta})\, \chi_p + V(\hat{\delta})\, \chi_p^2$. It seems that the bias converges to zero as $O(n^{-1})$ when the variance converges to zero as $O(n^{-1})$.

As a result of the asymptotic binormality of $(\hat{\lambda}, \hat{\delta})$, the asymptotic confidence interval with significance level α for (λ, δ) is given by the curve in the (λ, δ) plane

$$(\hat{\lambda} - \lambda)^2 - 2(1 - \gamma)(\hat{\lambda} - \lambda)(\hat{\delta} - \delta) + [\frac{\pi^2}{6} + (1 - \gamma)^2](\hat{\delta} - \delta)^2 \leq - \frac{2}{n} \log \alpha \,.\, \delta^2,$$

that is,

$$(\hat{\lambda}-\lambda)^2 - 0.84556\,(\hat{\lambda}-\lambda)\,(\hat{\delta}-\delta) + 1.82367\,(\delta-\delta)^2 \le -\frac{2}{n}\log\alpha\,.\,\delta^2;$$

recall that the quadratic form of a binormal distribution has the exponential distribution; see Cramér (1946).

Also from $\hat{\delta} \xrightarrow{P} \delta$ and by the δ-method we can have asymptotically a part of the ellipsis (as $\delta > 0$)

$$(\hat{\lambda}-\lambda)^2 - 2(1-\gamma)\,(\hat{\lambda}-\lambda)\,(\hat{\delta}-\delta) + [\frac{\pi^2}{6} + (1-\gamma)^2]\,(\hat{\delta}-\delta)^2 \le -\frac{2}{n}\log\alpha\,.\,\hat{\delta}^2.$$

The other estimators given in "Point Estimation" can be used to obtain an interval estimation or to test hypotheses about the parameters. *Most of them are asymptotically normal with known asymptotic variances and covariances, so that one can compute asymptotic confidence intervals*; in some cases we may resort to Monte Carlo simulation. For more details see Johns and Lieberman (1966) and Herbach (1970). The reliability function $R(x) = \exp\{-\exp\{-\nu/\delta\}\}$, where $\nu = \lambda - x$, is an increasing function of ν/δ and a lower confidence limit on ν/δ can be used to determine a lower confidence limit on $R(x)$, as $\text{Prob}\{L < \nu/\delta\} = \text{Prob}\{\exp\{-\exp(-L)\} < \exp\{-\exp(-\nu/\delta)\} = R(x) = \beta$. They proposed to use their asymptotically efficient linear estimators of ν and δ based on the first r order statistics in a sample of n. Let $Y_i' = X_i'/a$. Then for estimates of ν, δ we use

$$Y_a = \sum_{i=1}^{r} a_i\, Y_i' \quad , \quad Y_b = \sum_{i=1}^{r} b_i\, Y_i' \quad .$$

Let $V_a = (Y_a - \nu)/\delta$, $V_b = Y_a/\delta$. The joint distribution of (V_a, V_b) is parameter–free. Next define a function $L(t)$ with the property that for fixed β, $0 < \beta < 1$, assuming that $\text{Prob}\{V_b > 0\} = 1$, we have

$$\text{Prob}\,(L(t) < t\,V_b - V_a) = \beta \quad \text{for all} \quad t.$$

Then

$$\text{Prob}\,\{L(Y_a / Y_b) < \nu/\delta\} = \text{Prob}\{Y_a / Y_b) < L^{-1}\,(\nu/\delta)\} =$$

$$= \text{Prob}\,\{[\frac{\nu}{\delta} < L^{-1}\,(\frac{\nu}{\delta})\,V_b - V_a]\} = \text{Prob}\{L\,[\,L^{-1}\,(\frac{\nu}{\delta})] < L^{-1}(\frac{\nu}{\delta})V_b - V_a\,\} = \beta\ .$$

Thus $L(Y_a / Y_b)$ being a lower confidence limit for ν/δ with confidence coefficient β, correspondingly $\exp\{-\exp\{-L(Y_a/Y_b)\}$ is a lower confidence limit for $R(x)$, also with confidence coefficient β.

Finally they show that, asymptotically, L and $\exp\{-\exp\{-L\}\}$ are efficient bounds for ν/δ and $R(x)$. For $n = 10, 15, 20, 30, 50, 100$ and with four values of r for each n, they computed Monte Carlo simulations of the distribution of $tV_b - V_a$ for *fixed* values of $t = Y_a/Y_b$. Thus they generated tables from which one can obtain the bounds for $R(x)$ for specific x for $\beta = .5, .75, .90, .95$ and $.99$. To get a lower bound on $R^{-1}(w)$ we find in the table a value of $\exp\{-\exp\{-L(Y_a/Y_b)\}\}$ corresponding to w which is $Y_a - t\,Y_b$ where t is the table value of Y_a/Y_b associated with $\exp\{-\exp\{-L(Y_a/Y_b)\}\}$ for the appropriate confidence level.

Tests of Hypotheses

Hypothesis testing on parameters can be handled, as usual, by relating the acceptance region to the confidence region.

Consequently, the asymptotic test of $(\lambda , \delta) = (\lambda_0 , \delta_0)$ vs. $(\lambda, \delta) \neq (\lambda_0 , \delta_0)$ with significance level α, is given by the acceptance region : accept (λ_0 , δ_0) if this point falls in the interior of the ellipsis given for the confidence region, and reject otherwise.

Concretely, we will use as the acceptance region, for the significance level α,

$$(\hat{\lambda} - \lambda_0)^2 - 2(1-\gamma)\,(\hat{\lambda} - \lambda_0)\,(\hat{\delta} - \delta_0) + \left(\frac{\pi^2}{6} + (1-\gamma)^2\right)\,(\hat{\delta} - \delta_0)^2 \leq -\frac{2}{n}\,\log\alpha\,.\,\delta_0^2\,.$$

Here we do not need to substitute δ_0^2 by $\hat{\delta}^2$ in the RHS because in the test δ is supposed known.

Evidently we could obtain, by the classical method, the locally most powerful and locally most powerful unbiased tests if one of the parameters, or a relation between them, is fixed; but we would obtain analogous difficulties and, finally, resort to tests based on maximum likelihood estimation with optimal asymptotic behaviour. They can be let as exercises.

Evidently we could also use other estimators such as the ones described previously, but their efficiencies would in general be low.

Point and Interval Prediction

Let us begin by considering the point prediction of the maximum of (the next) m observations from a sample of n. The likelihood of the sample $(x_1, ..., x_n)$ and the (future) maximum $W = w$ is evidently

$$L(\lambda , \delta \mid x_i , w) = \frac{1}{\delta^n} e^{-n(\bar{x}- \lambda)/\delta} \exp \left(\sum_1^n e^{-(x_i-\lambda)/\delta} \right) . \frac{m}{\delta} e^{-(w-\lambda)/\delta} e^{-m(w-\lambda)/\delta}.$$

We could seek, following Tiago de Oliveira (1966) and (1968), the quasi–linearly invariant predictor $p(x_1, ..., x_n)$ (e.g., the predictor such that $p(\alpha + \beta x_i) = \alpha + \beta p(x_i)$, $\beta > 0$) such that $M(w-p(x_i))^2$ is minimum. The result is not manageable, even using computers, as can be seen from the expression

$$p(x_i) = \frac{\int_0^{+\infty} \xi^{n-1} \frac{e^{-n\bar{x}\xi}}{(\sum_1^n e^{-\xi x_i})^n} \left(\gamma + \log m - \log \sum_1^n e^{-\xi x_i} + \frac{\Gamma'(n)}{\Gamma(n)} \right) d\xi}{\int_0^{+\infty} \xi^{n} \frac{e^{-n\bar{x}\xi}}{(\sum_1^n e^{-\xi x_i})^n} d\xi}$$

As the $M(W) = \lambda + (\gamma + \log m)\, \delta$ and $(\hat{\lambda} , \hat{\delta})$ is asymptotically optimal it is then natural to use, as point predictor, the function $p(x_i) = \hat{\lambda} + k \hat{\delta}$ which is quasi–linearly invariant because $\hat{\lambda} (\alpha + \beta x_i) = \alpha + \beta \hat{\lambda} (x_i)$ and $\hat{\delta} (\alpha + \beta x_i) = \beta \delta(x_i)$.

Using the mean-square error as a measure of distance between the prediction and the (to be) observed value of $W = w$, we have

$$MSE(p) = M(W - (\hat{\lambda} + k \hat{\delta})^2) = M(W-M(W))^2 + M(\hat{\lambda} + k \hat{\delta} - M(W))^2 =$$

$$= \frac{\pi^2}{6} \delta^2 + M((\hat{\lambda}+ k \hat{\delta}) - (\lambda + (\gamma + \log m)\, \delta)^2 .$$

The mean–square error of the point predictor $p(x_i) = \hat{\lambda} + k \hat{\delta}$ *is asymptotically a minimum for* $k = \gamma + \log m$ *and its variance is that of the quantile* $\chi = \gamma + \log m$, *i.e., asymptotically.*

$$\frac{\pi^2}{6} \delta^2 + \left[1 + \frac{6}{\pi^2} (\gamma + \log m)^2 \right] \frac{\delta^2}{n} .$$

The determination of prediction regions, also dealt with in the same papers, leads to the same difficulties in computing integrals. Thus we will once more use the maximum likelihood estimators.

Let us now seek, then, one–sided and two–sided prediction intervals based on $(\hat{\lambda}, \hat{\delta})$, for a prediction level ω, in general close to 1.

The most important one–sided prediction interval for the prediction level ω for W is $W \leq \hat{\lambda} + c\,\hat{\delta}$ with c such that

$$\text{Prob}\{W \leq \hat{\lambda} + c\,\hat{\delta}\} = \omega$$

which as

$$\text{Prob}\{W \leq y\} = \Lambda^m((y-\lambda)/\delta) = \Lambda\,((y-\lambda)/\delta) - \log m]$$

is reduced to

$$M(\Lambda\,(\frac{\hat{\lambda} + c\,\hat{\delta} - \lambda}{\delta} - \log m)) = \omega\ .$$

Evidently the first approximation to c is from the relation $\Lambda(c - \log m) = \omega$ or $c = \chi_\omega + \log m$. Consider now the approximation $c = \chi_\omega + \log m + \frac{\varphi}{\sqrt{n}} + o\,(n^{-1/2})$. Let us put $\sqrt{n}\,\frac{\hat{\lambda} - \lambda}{\delta} = A_n$, $\sqrt{n}\,(\frac{\hat{\delta}}{\delta} - 1) = B_n$. Then we get

$$M(\Lambda\,(\frac{\hat{\lambda} + c\,\hat{\delta} - \lambda}{\delta} - \log m)) = M(\Lambda\,(c - \log m + \frac{A_n + c\,B_n}{\sqrt{n}})) = \omega.$$

Developing c through $\chi_\omega + \log m + \frac{\varphi}{\sqrt{n}} + o\,(n^{-1/2})$ we get the equation, by easy algebra,

$$\varphi + M(A_n) + (\chi_\omega + \log \omega)\,M(B_n) + o(n^{-1/2}) = 0,$$

and as $M(A_n) \to 0$ and $M(B_n) \to 0$ we get $\varphi = 0$. Thus $c = \chi_\omega + \log m + o\,(n^{-1/2})$, or $c = \chi_\omega + \log m$ with a linear error of the order of $o(n^{-1/2})$. The prediction level differs from ω in terms of $(o\,(n^{-1/2}))^2 = o\,(n^{-1})$ which shows that $c = \chi_\omega + \log m$ gives a good approximation to the quantile with an error in the prediction level of the order $o(n^{-1})$. Consequently:

The one-sided prediction interval with prediction level ω, *to the order of* $o(n^{-1})$, *is* $\hat{\lambda} + (\chi_\omega + \log m)\,\hat{\delta}$ *with* $\Lambda(\chi_\omega) = \omega$.

A subsequent summand ψ/n of order n^{-1} could be obtained in the same way supposing now that $\sqrt{n}\, M(A_n) \to 0$ and $\sqrt{n}\, M(B_n) \to 0$; we have then $\psi = \frac{1 + \log \omega}{2}\,(1 + \frac{6}{\pi^2}\,(1 - \gamma + \chi_\omega + \log m)^2)$ and so the second order approximation to c, with a linear error of the order of $o(n^{-1})$, is

$$c = \chi_\omega + \log m + \frac{1 + \log \omega}{2n}\,\left(1 + \frac{6}{\pi^2}\,(1 - \lambda + \chi_\omega + \log m)^2\right)$$

and the prediction level differs from ω by a quantity of $o(n^{-3/2})$.

The other one–sided interval can be dealt with in the same way.

Let us sketch two–sided prediction intervals with prediction level ω and asymptotically shortest average length. If $a' \le b'$ are quantities such that

$$\text{Prob}(\hat{\lambda} + a'\,\hat{\delta} \le X \le \hat{\lambda} + b'\,\hat{\delta}) = w$$

and $b' - a'$ is minimum, they are given by the equations

$$a' + e^{-a'} = b' + e^{-b'}.$$

It is then easy to show that if $a \le b$ and are such that

$$\text{Prob}(\hat{\lambda} + a\,\hat{\delta} \le W \le \hat{\lambda} + b\,\hat{\delta}) = \omega$$

we shall have $a = a' + \log m$, $b = b' + \log m$, apart from terms of order n^{-1} with the length $(b - a)\,\delta = (b' - a')\,\delta$ which is asymptotically shortest on average.

Other estimators of (λ, δ) can be considered as well to form predictors.

Miscellaneous Results

Discrimination

Consider that we have two populations whose distribution functions are $\Lambda((x - \lambda)/\delta)$ and $\Lambda((x' - \lambda')/\delta)$, where the scale parameter is the same.

Consider that we want to decide if some observation y belongs to the first or the second population (i.e., has the location parameter λ or λ'). Supposing λ, λ' and δ to be known, we are in a situation analogous to usual Neyman–Pearson theory for testing hypothesis. Denoting by A and $A' = A^c$ the acceptance regions for λ and λ', the misclassification errors of the test are

$$\text{Prob}\{A \mid \lambda'\} = \int_A d\Lambda\left(\frac{y-\lambda'}{\delta}\right)$$

and $$\text{Prob}\{A' \mid \lambda\} = \int_{A'} d\Lambda\left(\frac{y-\lambda}{\delta}\right) = 1 - \int_A d\Lambda\left(\frac{y-\lambda}{\delta}\right).$$

To be fair we stipulate these misclassification errors to be equal, i.e.,

$$\int_A \left[d\Lambda\left(\frac{y-\lambda}{\delta}\right) + d\Lambda\left(\frac{y-\lambda'}{\delta}\right)\right] = 1$$

and minimize the (equal) misclassification errors $\int_A d\Lambda\left(\frac{y-\lambda'}{\delta}\right)$.

This test, by the Neyman–Pearson theorem, has the acceptance region for λ,

$$\frac{\frac{1}{\delta}\Lambda'\left(\frac{y-\lambda'}{\delta}\right)}{\frac{1}{\delta}\Lambda'\left(\frac{y-\lambda}{\delta}\right) + \frac{1}{\delta}\Lambda'\left(\frac{y-\lambda'}{\delta}\right)} \leq k$$

or

$$e^{-y/\delta}\left(e^{\lambda/\delta} - e^{\lambda'/\delta}\right) \leq c\left(\lambda, \lambda', \delta\right).$$

Suppose that $\lambda > \lambda'$. Then the acceptance region A is given by $y > c(\lambda, \lambda', \delta)$, where c is defined by

$$\Lambda\left(\frac{c-\lambda}{\delta}\right) + \Lambda\left(\frac{c-\lambda'}{\delta}\right) = 1$$

and the misclassification error is given by

$$\int_A d\Lambda\left(\frac{y-\lambda'}{\delta}\right) = 1 - \Lambda\left(\frac{c-\lambda'}{\delta}\right) ;$$

if $\lambda < \lambda'$ the changes are obvious.

As λ, λ' and δ are not known it is natural to estimate them and substitute in the previous relations.

For a sample of $(x_1, ..., x_n)$ of the first population and of $(x'_1, ..., x'_{n'})$ of the second one, the maximum likelihood estimators are

$$\hat{\lambda} = -\hat{\delta} \log \frac{\sum_1^n e^{-x_i/\hat{\delta}}}{n} \quad , \quad \lambda' = -\hat{\delta} \log \frac{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}{n'}$$

and

$$\hat{\delta} = \frac{n}{n+n'}\left(\bar{x} - \frac{\sum_1^n x_i e^{-x_i/\delta}}{\sum_1^n e^{-x_i/\delta}}\right) + \frac{n'}{n+n'}\left(\bar{x}' - \frac{\sum_1^{n'} x'_i e^{-x'_i/\hat{\delta}}}{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}\right).$$

Thus we substitute λ, λ' and δ by their estimators in the fair test above, taking for A the discrimination region corresponding to the largest estimated location parameter.

For more details and the analysis of the asymptotic behaviour (consistency) see Tiago de Oliveira (1973).

Tolerance intervals

The concept of tolerance intervals was introduced by Wilks (1941) as a procedure for the prediction of the values of a population. In Themido (1985), parameter–free tolerance intervals for the Gumbel distribution are obtained and a comparison, for large samples, is made with the distribution–free tolerance intervals proposed by Wilks, a comparison based on the respective expected lengths.

Let $(x_1, ..., x_n)$ be an i.i.d. sample of one–dimensional continuous random variables, and $L_i = L_i(x_1, ..., x_n)$, $i = 1, 2$, be statistics such that $L_1(x_i) < L_2(x_i)$. The random interval $[L_1, L_2]$ is a (β, ω) – tolerance interval if

$$\text{Prob}\{\text{Prob}[L_1 \leq X \leq L_2] \geq \beta\} = \omega,$$

ω being the tolerance level of the interval; if, instead of this relation, we have

$$M(\text{Prob}[L_1 \leq X \leq L_2]) = \alpha,$$

the interval $[L_1, L_2]$ is an α–average tolerance interval.

In the first case for given β and $\omega(0 < \beta, \omega < 1)$, we seek tolerance limits L_1 and L_2, based on the sample, such that $[L_1, L_2]$ contains, with probability ω, at least $100\,\beta\,\%$ of the populations or such that we have a confidence of $100\,\omega\,\%$ that the probability that a future observation will lie between L_1 and L_2 is at least β; for the second case we must find L_1 and L_2 such that the average or mean coverage of the interval is $\alpha(0 < \alpha < 1)$.

If we have $L_1 = -\infty$ or $L_2 = +\infty$ we obtain one–sided tolerance intervals.

Regarding the distribution function of the X_i we should consider two situations :

— the non–parametric case where the functional form of the distribution of the X_i is unknown, apart from the fact that it is (absolutely) continuous;

— the parametric case where the functional form of the distribution of the X_i is known and only the values of a finite number of parameters of the distribution are unknown.

For the determination of tolerance intervals for one–dimensional distributions in the non–parametric case, an exact solution, using order statistics, was given by Wilks (1941).

In the parametric case it is natural to expect that tolerance limits better than those of Wilks can be obtained.

For the Gumbel distribution $\Lambda(x \mid \lambda, \delta) = \Lambda((x - \lambda)/\delta), -\infty < \lambda < +\infty, 0 < \delta < +\infty$, let us use as tolerance limits the maximum likelihood estimators of the quantiles of probabilities $p_i = \Lambda(k_i)$, $i = 1,2$, $L_i = L_i(x_1, ..., x_n) = \hat{\lambda} + k_i \hat{\delta}$, with k_1 and k_2 $(k_1 < k_2)$ real coefficients independent of n, a choice made because $\hat{\lambda}(x_i)$ and $\hat{\delta}(x_i)$ are optimum estimators of λ and δ in terms of asymptotic efficiency. Recall that $\hat{\lambda} + k\,\hat{\delta}$ are the unique quasi–linearly invariant functions of $\hat{\lambda}$ and $\hat{\delta}$.

The (random) proportion of the Gumbel distribution covered by the tolerance interval $[L_1, L_2]$ is

$$P_{1,2} = \Lambda\left(\frac{\hat{\lambda}(x_i)}{\delta} + k_2 \frac{\hat{\delta}(x_i)}{\delta}\right) - \Lambda\left(\frac{\hat{\lambda}(x_i)}{\delta} + k_1 \frac{\hat{\delta}(x_i)}{\delta}\right) =$$

$$= \Lambda(\hat{\lambda}(Z_i) + k_2\,\hat{\delta}(Z_i)) - \Lambda(\hat{\lambda}(Z_i) + k_1\,\hat{\delta}(Z_i)),$$

where $(z_1, \ldots, z_n)$ is the corresponding random sample from the reduced Gumbel distribution. This follows from the invariance of maximum likelihood estimators and from the form of the quasi–linearly invariant tolerance limits considered. Thus, the probability $\hat{P}_{1,2}$ of $[L_1, L_2]$ is parameter–free and the (β, ω) — tolerance interval can be written as $\text{Prob}\{\hat{P}_{1,2} \geq \beta\} = \omega$. Using the δ–method (Tiago de Oliveira (1982)) we can show that the asymptotic distribution of $\hat{P}_{1,2}$ is normal, as usual, with mean value $\Lambda(k_2) - \Lambda(k_1)$ and variance $O(n^{-1})$. Even if we take $k_1(n) = k_i + a_i \sqrt{n} + o(n^{-1/2})$, the a_i can be computed to give a better approximation; details of an analogous approach can be found in Chapter VIII.

For each pair (β, ω), $0 < \beta, \omega < 1$, there is an infinity of pairs (k_1, k_2) such that the tolerance limits L_1 and L_2 define a (β, ω) – tolerance interval for the Gumbel distribution. We can determine the minimum length parameter–free (β, ω) tolerance interval.

Also, in terms of the α–average tolerance interval, $0 < \alpha < 1$, we can seek the one with minimum length.

The parameter–free (β, ω) — tolerance limits and the parameter–free α–average tolerance limits, being estimators of quantiles for the Gumbel distribution converge in probability, as $n \to +\infty$, to the quantiles which are the end-points of the shortest quantile intervals with probability β and α respectively. Tolerance intervals thus correspond to the intervals above but with random end–points.

We can also deal with one–sided parameter–free (β, ω) tolerance interval; in this case we have only one coefficient, k_1 or k_2 for lower or upper intervals respectively.

In the comparison for large samples, between distribution-free and parameter–free two–sided tolerance interval for the Gumbel distribution, with respect to the mean length criterion, the parameter–free tolerance interval using the information about the distribution are, as expected, better than distribution–free tolerance interval.

For the proofs, a tabulation of (k_1, k_2), and other details, see Themido (1985).

Multisample analysis

Consider two samples $(x_1, ..., x_n)$ and $(x'_1, ..., x'_n)$, for which we assume that they have a Gumbel distribution with parameters λ and λ' and the same parameter δ. We want to test if $\lambda = \lambda'$. This is the first step of Multisample Analysis for Extremes, analogous to Analysis of Variance, but practically unsolved for extremes, and the only case that will be dealt with here.

Analogously to what happened to discrimination we have

$$\hat{\lambda} = -\hat{\delta} \log \frac{\sum_1^n e^{-x_i/\hat{\delta}}}{n} \quad , \quad \hat{\lambda}' = -\hat{\delta} \log \frac{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}{n'}$$

and

$$\hat{\delta} = \frac{n}{n+n'} \left(\bar{x} - \frac{\sum_1^n x_i e^{-x_i/\hat{\delta}}}{\sum_1^n e^{-x_i/\hat{\delta}}} \right) + \frac{n'}{n+n'} \left(\bar{x}' - \frac{\sum_1^{n'} x'_i e^{-x'_i/\hat{\delta}}}{\sum_1^{n'} e^{-x'_i/\hat{\delta}}} \right)$$

if λ is supposed different from λ' and

$$\hat{\hat{\lambda}} = -\hat{\hat{\delta}} \log \frac{\sum_1^n e^{-x_i/\hat{\hat{\delta}}} + \sum_1^{n'} e^{-x'_i/\hat{\hat{\delta}}}}{n+n'}$$

$$\hat{\hat{\delta}} = \frac{n\bar{x} + n'\bar{x}'}{n+n'} - \frac{\sum_1^n x_i e^{-x_i/\hat{\hat{\delta}}} + \sum_1^n x'_i e^{-x'_i/\hat{\hat{\delta}}}}{\sum_1^n e^{-x_i/\hat{\hat{\delta}}} + \sum_1^{n'} e^{-x'_i/\hat{\hat{\delta}}}}$$

if we suppose $\lambda' = \lambda$.

The likelihood ratio criterion leads to the test statistic

$$-2\log\frac{\max\limits_{\lambda,\delta}[L_n(x\,|\,\lambda,\delta)\,L_n,(x'\,|\,\lambda,\delta)]}{\max\limits_{\lambda,\lambda',\delta}[L_n(x\,|\,\lambda,\delta)\,L_n,(x'\,|\,\lambda,\delta)]}=$$

$$=2\{(n+n')\,\log\frac{\hat{\hat{\delta}}}{\hat{\delta}}+\frac{n(\bar{x}-\hat{\hat{\lambda}})+n'(\bar{x}\,'-\hat{\hat{\lambda}})}{\hat{\hat{\delta}}}-\frac{n(\bar{x}-\hat{\lambda})+n'(\bar{x}\,'-\hat{\lambda}')}{\hat{\delta}}\}$$

which is asymptotically a χ^2 with one degree of freedom in the hypothesis tested.

Overpassing probability of a level

Let a be a level and suppose that X has a Gumbel distribution $\Lambda((x-\lambda)/\delta)$. Then $P(a) = P(a\,|\,\lambda,\delta) =$ Prob $\{X > a\} = 1 - \Lambda((a-\lambda)/\delta)$. We will take as estimator of $P(a\,|\,\lambda,\delta)$, using the estimators $(\hat{\lambda}\,,\hat{\delta}\,)$ of $(\lambda\,,\delta)$,

$$\hat{P}(a) = P(a\,|\,\hat{\lambda},\hat{\delta})$$

and by the use δ–method we have that $\sqrt{n}\,(\hat{P}\,(a) - P(a\,|\,\lambda\,,\delta))$ has the asymptotically normal distribution of the linearized expression.

$$\sqrt{n}\,[(\hat{\lambda}-\lambda)\,\frac{\partial P(a)}{\partial\lambda}+(\hat{\delta}-\delta)\frac{\partial P(a)}{\partial\,\delta}]=$$

$$=\sqrt{n}\,(1-P(a))\,(-\log(1-P(a))\,[\frac{\hat{\lambda}-\lambda}{\delta}+\frac{a-\lambda}{\delta}\,\frac{\hat{\delta}-\delta}{\delta}]\,.$$

Its asymptotic mean value is zero and the variance is

$$V\,(a\,|\,\lambda,\delta)=(1-P(a))^2\;(\log(1-P(a)))^2\;\{1+\frac{6}{\pi^2}\,(1-\gamma+\frac{a-\lambda}{\delta})^2\}\,.$$

Also $\sqrt{n}\,\dfrac{\hat{P}\,(a)-P(a)}{\sqrt{V\,(a\,|\,\lambda,\delta)}}$ is asymptotically standard normal even with $V(a\,|\,\lambda,\delta)$ substituted by $\hat{V}\,(a) = V(a\,|\,\hat{\lambda}\,,\hat{\delta}\,)$.

Then, λ_α denoting as usual the solution of $N(x) = 1-\alpha$, we have

$$\text{Prob}\left\{\sqrt{n}\,\frac{\hat{P}(a) - P(a)}{\sqrt{\hat{V}(a)}} \geq -\lambda_\alpha\right\} \to 1-\alpha \quad \text{as} \quad n \to \infty$$ and thus :

The one–sided asymptotic confidence interval for P(a) with significance level α is

$$0 \leq P(a) \leq \hat{P}(a) + \frac{\lambda_\alpha}{\sqrt{n}}\sqrt{\hat{V}(a)}\ .$$

A worked example

Sneyers (1977) gives the following n = 35 maximal yearly precipitations from 1938 to 1972 at Uccle, in 1 mm units, for the durations of 24 h, 60 min, 10 min and 1 min.

Table

Year	24 h	1 min	10 min	60 min	Year	24 h	1 min	10 min	60 min
1938	33.8	2.5	6.5	14.0	1956	35.6	3.0	8.5	18.8
1939	27.7	1.0	8.5	12.8	1957	34.2	1.6	9.8	12.0
1940	60.0	.5	5.0	12.9	1958	24.3	2.0	5.5	12.0
1941	24.0	.9	8.4	11.9	1959	20.3	1.2	9.8	11.6
1942	72.3	1.5	13.2	20.6	1960	48.0	2.0	9.5	15.3
1943	50.7	4.4	11.9	29.1	1961	32.4	1.5	11.5	19.2
1944	18.7	1.0	3.8	6.2	1962	59.6	2.9	12.7	42.8
1945	41.2	3.0	13.0	21.1	1963	60.4	3.7	9.0	13.0
1946	26.6	3.3	11.1	11.2	1964	27.0	2.7	13.0	15.7
1947	27.2	2.0	13.0	18.0	1965	45.8	2.0	12.2	15.4
1948	23.8	1.8	6.5	15.6	1966	39.8	2.9	9.5	14.3
1949	19.8	1.0	5.7	8.7	1967	21.6	3.0	11.9	13.1
1950	34.3	2.0	13.3	23.8	1968	19.7	2.1	8.3	14.9
1951	28.2	4.0	12.2	12.2	1969	54.4	2.3	15.3	25.8
1952	51.1	2.0	8.4	29.0	1970	29.1	2.2	13.8	17.1
1953	37.5	1.0	5.0	9.9	1971	41.6	1.6	7.0	21.2
1954	34.3	2.0	6.9	12.5	1972	26.0	2.8	8.7	16.3
1955	22.2	1.6	6.2	9.6					

Experience, says Sneyers – and the plotting confirms it – shows that for small durations the maximal precipitation follows a Gumbel distribution but for large durations it follows a Fréchet distribution.

The ML method applied to 1 min duration data gives $\hat{\lambda} = 1.709286$ and $\hat{\delta} = .778273$, close to 1.694 and .7777 or 1.691 and .7484 obtained by Sneyers by other methods.

From the values of $(\hat{\lambda}, \hat{\delta})$ we can solve other decision problems such as the ones discussed before.

To test the goodness of fit of a *known* (or *fixed*) distribution F(x) we can use the Kolmogoroff -Smirnov test statistic

$$K S_n = \sup_x | S_n(x) - F(x) | = \max_1^n \{ \max (| F(x_i') - \frac{i}{n} | , | F(x_{i+1}') - \frac{i}{n} |) \} ,$$

where $S_n(x) = S_n(x | x_i)$ denotes the sample distribution function $(x_1' \leq x_2' \leq \ldots \leq x_n')$, the order statistics of the i.i.d. sample and is taken $F(x_{n+1}') = 1$. We accept F(x), at the significance levels .05 and .01, according to $\sqrt{n}\ K S_n \leq 1.36$ or $\sqrt{n}\ K S_n \leq 1.63$, rejecting otherwise. When we assume *model* $F(x|\theta)$, we estimate θ by $\hat{\theta}$ (say) and instead of the unknown (parameterized) $F(x|\theta)$ we compute $\hat{K}S_n = \sup_x | S_n(x) - F(x|\hat{\theta}) |$ and act in the same way as before. Although the procedure is not completely justified, it can be expected to be sufficiently approximate when n is large.

In our case, assuming the Gumbel model, with the values of $(\hat{\lambda}, \hat{\delta})$ above, we have

$$\hat{K}S_n = \sup_x | S_n(x) - \Lambda((x - \hat{\lambda}) / \hat{\delta}) | = .0102432$$

and as n = 35 we have $\sqrt{n}\ \hat{K}S_n = .06059$, accepting thus the Gumbel model.

Chapter VI

STATISTICS FOR FRÉCHET DISTRIBUTIONS

Introduction

The general Fréchet distribution for maxima has the form

$$\Phi_\alpha (x|\lambda, \delta) = \Phi_\alpha \left(\frac{x - \lambda}{\delta} \right) = 0 \qquad \text{if} \qquad x \leq \lambda$$

$$\Phi_\alpha (x) = \exp \left\{ - \left(\frac{x - \lambda}{\delta} \right)^{-\alpha} \right\} \qquad \text{if} \qquad x \geq \lambda .$$

It is immediate that $Y = \left(\frac{X - \lambda}{\delta} \right)^{-\alpha}$ has the standard exponential distribution. Also if $\lambda = \lambda_0$ is known (it can be taken as zero, for convenience, by subtraction) $\tilde{Y} = \log (X - \lambda_0)$ has a Gumbel distribution of maxima with location and dispersion parameters $\log \delta$ and $1/\alpha$. Thus theory for Gumbel distribution can be applied, and the maximum likelihood estimators of (λ, δ) (if $\lambda_0 = 0$) are given by the equations.

$$\frac{n}{\hat{\alpha}} - \sum_1^n \log x_i + n \frac{\sum_1^n x_i^{-\hat{\alpha}} \log x_i}{\sum_1^n x_i^{-\hat{\alpha}}} = 0$$

$$\hat{\delta}\,\hat{\alpha} = \frac{n}{\sum_1^n x_i^{-\hat{\alpha}}} .$$

The solution of the first equation needs an initial seed which can be obtained using the fact that $\{\log x_i\}$ has dispersion parameter $1/\alpha$ of a general Gumbel distribution among other ways.

The random pair $(\hat{\delta}, \hat{\alpha})$ *is asymptotically normal with mean values* δ *and* α, *variances* $\delta^2/n\,\alpha^2 \cdot [1 + 6/\pi^2\,(1-\delta)^2]$, *and* $6\,\alpha^2/n\,\pi^2$, *covariance* $= \dfrac{6\,(1-\gamma)\,\delta}{n\,\pi^2\,\alpha^2}$ *and correlation coefficient* $\rho = -(1 + \dfrac{\pi^2}{6\,(1-\gamma)^2})^{1/2} = -.31307.$

This should be expected from the formulae given before for the Gumbel distribution.

The asymptotic confidence region with a significance level is α' given by

$$\alpha^2(\hat{\delta}-\delta)^2 + 2(1-\gamma)\,\delta(\hat{\delta}-\delta)(\hat{\alpha}-\alpha) + (\frac{\pi^2}{6} + (1-\gamma)^2)\,\delta^2\,(\frac{\hat{\alpha}-\alpha}{\alpha})^2 \le -2\log\alpha'\cdot\delta^2$$

where, to avoid big changes of notation, the α' in the RHS of the inequality is the significance level. The substitution of α and δ by $\hat{\alpha}$ and $\hat{\delta}$ except in the differences $(\hat{\alpha}-\alpha)$ and $(\hat{\delta}-\delta)$ gives a part of an ellipsis in the quadrant $(\alpha > 0, \delta > 0)$. From the same result we can make the rule for asymptotic tests of hypotheses.

The estimator of the quantile $\chi_{X,p} = \delta/(-\log p)^{1/\alpha}$ *is evidently taken as* $\hat{\chi}_{X,p} = \hat{\delta} / (-\log p)^{1/\hat{\alpha}}$; *it is asymptotically normal with mean value* $\chi_{X,p}$ *and variance*

$$(-\log p)^{-2/\alpha}[1 + \frac{6}{\pi^2}(1-\gamma-\log(-\log p))^2]\;\delta^2/n\;\alpha^2$$

or, equivalently from the asymptotic point of view,

$$(-\log p)^{-2/\hat{\alpha}}[1 + \frac{6}{\pi^2}(1-\gamma-\log(-\log p))^2]\;\hat{\delta}^2/n\;\hat{\alpha}^2\ .$$

As the mean value of the maximum of the next m *observations is* $m^{-1/\alpha}\,\Gamma(1-1/\alpha)\,\delta$ *(if* $\alpha > 1$*) we can take as estimator* $m^{-1/\hat{\alpha}}\Gamma(1-1/\hat{\alpha})\,\hat{\delta}$; *this estimator is asymptotically normal with the same mean value referred to and variance*

$$\frac{\Gamma^2(1-1/\alpha)}{m^{2/\alpha}\cdot\;\alpha^2\;n}[1 + \frac{6}{\pi^2}(\gamma + \frac{\Gamma'(1-1/\alpha)}{\Gamma(1/\alpha)})^2]\ .$$

Equivalently, from the asymptotic point of view, we can substitute in the variance δ and α by $\hat{\delta}$ and $\hat{\alpha}$. The asymptotic variance exists only if $\alpha > 2$, as it will be seen shortly.

It is worth noting that if X has a Fréchet distribution of maxima with $\lambda = \lambda_0 = 0$ and dispersion and location parameters δ and α, then $1/X$ has a Weibull distribution of minima (to

be studied in the next chapter) with, also, $\lambda = \lambda_0 = 0$, and dispersion and shape parameters $1/\delta$ and α; so the techniques for the Fréchet distribution of maxima and Weibull distribution of minima can be exchanged.

As far as applications are concerned, as said before, they have been done for largest waves and floods of rivers.

Characteristics of the Fréchet distribution

From the Fréchet distribution

$$\Phi_\alpha\left(\frac{x-\lambda}{\delta}\right) = 0 \qquad \text{if} \qquad x \le \lambda$$

$$\Phi_\alpha(x) = \exp\left\{-\left(\frac{x-\lambda}{\delta}\right)^{-\alpha}\right\} \qquad \text{if} \qquad x \ge \lambda$$

we can get the usual coefficients if they exist. So we will compute them for the reduced random variable $Z = (X - \lambda)/\delta$, the transfer to the general case being obvious as seen in the previous chapter.

The graph of the probability densities $\Phi'_\alpha(z)$ $(z \ge 0)$ for different values of $\alpha > 0$ are given in fig. VI.1.

Fig.1 – Reduced Fréchet densities (for maxima)

For $\Phi'_\alpha(z) = \alpha z^{-1-\alpha}(\exp(-z^{-\alpha}))$, the mode is $\mu' = \left(\frac{\alpha}{1+\alpha}\right)^{1/\alpha}$, the median $\mu = (\log 2)^{-1/\alpha}$, and the quantile $\chi_p = (-\log p)^{-1/\alpha}$ with $\chi_{1/e} = 1$.

As regards moments, they exist only if $r < \alpha$, and have the expression $\mu'_r = \Gamma(1 - r/\alpha)$. Consequently we have as mean value $\mu = \Gamma(1 - 1/\alpha)$ if $1 < \alpha$ and as variance $\sigma^2 = \Gamma(1 - 2/\alpha) - \Gamma^2(1 - 1/\alpha)$ if $2 < \alpha$, the corresponding values for $X = \lambda + \delta z$ being obvious. The computation of β_1 and β_2, which depend on α, is left as an exercise. Notice that the negative order moments always exist.

An important point to recall is that if Z has the distribution function $\Phi_\alpha(z)$ then $\text{Prob}\{\alpha\,(Z-1) \le y\} = \Phi_\alpha(1 + y/\alpha) \to \Lambda(y)$ as $\alpha \to \infty$. This means that for large α we can use the Gumbel distribution as an approximation to the Fréchet distribution; see the exercises related to the approximation of $\Phi_\alpha(1 + y/\alpha)$ by $\Lambda(z)$.

The moment generating function of Z has no explicit expression, but that of $\log Z$ is $M(e^{\theta \log Z}) = M(Z^\theta) = \Gamma(1 - \theta/\alpha)$, defined only for $\theta < \alpha$, as could be expected from the moments, $M(Z^\theta)$ giving the cumulants associated with the moments.

Estimation, testing, and point prediction

Let us now consider the statistical decision for the 3–parameters case. The likelihood being

$$L(\lambda, \delta, \alpha | x_i) = 0 \qquad \text{if} \quad \min\,(x_1, \ldots, x_n) \le \lambda$$

$$L(\lambda, \delta, \alpha | x_i) = \frac{\alpha^n \, \delta^{n\alpha}}{\prod_1^n (x_i - \lambda)^{\alpha+1}} \exp\{-\delta^\alpha \sum_1^n (x_i - \lambda)^{-\alpha}\} \quad \text{if} \quad \min(x_1, \ldots, x_n) > \lambda\,,$$

the maximum likelihood estimators are given by

$$\hat{\delta}\,\hat{\alpha} = \frac{n}{\sum_1^n (x_i - \hat{\lambda})^{-\hat{\alpha}}}$$

$$\frac{(\hat{\alpha}+1)}{\hat{\alpha}} \sum_1^n (x_i - \hat{\lambda})^{-1} = \frac{n \sum_1^n (x_i - \hat{\lambda})^{-\hat{\alpha}-1}}{\sum_1^n (x_i - \hat{\lambda})^{-\hat{\alpha}}}$$

$$\frac{1}{n} \sum_1^n \log(x_i - \hat{\lambda}) - \frac{1}{\hat{\alpha}} = \frac{\sum_1^n (x_i - \hat{\lambda})^{-\hat{\alpha}} \log(x_i - \hat{\lambda})}{\sum_1^n (x_i - \hat{\lambda})^{-\hat{\alpha}}}$$

For this system we have to solve, first, the last two equations and immediately compute $\hat{\delta}$.

Note that the solutions must be admissible, i.e., they must satisfy $\hat{\lambda} \leq \min(x_1, \ldots, x_n)$, $\hat{\alpha} > 0$ and $\hat{\delta} > 0$ being obvious from the relations defining them. To obtain a first approximation to $(\hat{\lambda}, \hat{\alpha})$ we can follow Gumbel (1965). Denoting by $\min\{x_i\}$, $\text{med}\{x_i\}$ and $\max\{x_i\}$ the minimum, the median and the maximum of the sample, we have the first approximations $(\hat{\lambda}_1, \hat{\alpha}_1)$

$$Q_n = \frac{\max\{x_i\} - \text{med}\{x_i\}}{\text{med}\{x_i\} - \min\{x_i\}} = Q\left(\frac{1}{\hat{\alpha}_1}, n\right)$$

$$\hat{\lambda}_1 = \frac{n^{1/\hat{\alpha}_1}\,\text{med}\{x_i\} - \max\{x_i\}}{n^{1/\hat{\alpha}_1} - 1} .$$

Gumbel (1965) gives a table of the function $Q(1/\alpha, n)$ ($1/\alpha$ there being called λ) for values of $\alpha > 1$. The statistic Q_n was considered as a tool for the quick statistical choice of extreme models in Chapter IV.

We will now compute the asymptotic sampling characteristics of the maximum likelihood estimators $(\hat{\lambda}, \hat{\delta}, \hat{\alpha})$.

The variance-covariance matrix is

$$\frac{1}{n} \cdot \begin{bmatrix} \frac{(\alpha+1)^2\,\Gamma(1+2/\alpha)}{\delta^2} & \frac{\alpha(\alpha+1)\,\Gamma(1+1/\alpha)}{\delta^2} & \frac{(\alpha+1)\,[\Gamma(1+1/\alpha) + \Gamma'(1+1/\alpha)]}{\alpha\delta} \\ \frac{\alpha(\alpha+1)\,\Gamma(1+1/\alpha)}{\delta^2} & \frac{\alpha^2}{\delta^2} & \frac{1-\gamma}{\delta} \\ \frac{(\alpha+1)\,[\Gamma(1+1/\alpha) + \Gamma'(1+1/\alpha)]}{\alpha\delta} & \frac{1-\gamma}{\delta} & \frac{1}{\alpha^2}\left(\frac{\pi^2}{6} + (1-\gamma)^2\right) \end{bmatrix}^{-1}$$

as given by Tiago de Oliveira (1973).

The asymptotic confidence region with level of significance α' is given by

$$\frac{(\alpha+1)^2\,\Gamma(1+2/\alpha)}{\delta^2}(\hat{\lambda}-\lambda)^2 + \frac{\alpha^2}{\delta^2}(\hat{\delta}-\delta)^2 + \frac{1}{\alpha^2}\left(\frac{\pi^2}{6} + (1-\gamma)^2\right)(\hat{\alpha}-\alpha)^2 +$$

$$+ \frac{2\,\alpha(\alpha+1)\,\Gamma(1+1/\alpha)}{\delta^2}(\hat{\lambda}-\lambda)\,(\hat{\delta}-\delta) +$$

$$+ \frac{2(\alpha+1)\,[\Gamma(1+1/\alpha) + \Gamma'(1+1/\alpha)]}{\alpha\delta}\,(\hat{\lambda}-\lambda)\,(\hat{\alpha}-\alpha) +$$

$$+ \frac{2(1-\gamma)}{\delta}\,(\hat{\delta}-\delta)\,(\hat{\alpha}-\alpha) \le \frac{\chi^2_3(\alpha')}{n}$$

where $\chi^2_3(\alpha')$ is the point α' for a χ^2 with 3 degrees of freedom.

The usual tests of hypotheses can be made using this asymptotic confidence region.

The estimator of the quantile $\chi_{X,p} = \lambda + (-\log p)^{-1/\alpha}\,\delta$ *is* $\hat{\chi}_{X,p} = \hat{\lambda} + (-\log p)^{-1/\hat{\alpha}}\,\hat{\delta}$. *It is asymptotically normal with mean value* $\chi_{X,p}$ *and asymptotic variance*

$$V(\hat{\lambda}) + (-\log p)^{-2/\alpha}\,V(\hat{\delta}) + \frac{\delta^2}{\alpha^4}\,(-\log p)^{-2/\alpha}\,(\log(-\log p))^2\,V(\hat{\alpha})$$

$$+ 2\,(-\log p)^{-1/\alpha}\,C(\hat{\lambda},\hat{\delta}) + \frac{2\,\delta(-\log p)^{-1-\alpha}\,\log(-\log p)}{\alpha^2}\,C(\hat{\lambda},\hat{\alpha}) +$$

$$+ \frac{2\,\delta}{\alpha^2}\,(-\log p)^{-2/\alpha}\,\log(-\log p)\,C(\hat{\delta},\hat{\alpha}).$$

The least squares predictor of the mean value of the maximum of the next m *observations* $\lambda + m^{-1/\alpha}\,\Gamma(1-1/\alpha)\,\delta$ *is given by* $\hat{\lambda} + m^{-1/\hat{\alpha}}\,\Gamma(1-1/\hat{\alpha})\,\hat{\delta}$ *(if* $1 < \alpha$*) and its asymptotic variance is*

$$V(\hat{\lambda}) + m^{-2/\alpha}\,\Gamma^2(1-1/\alpha)\,V(\hat{\delta}) + \frac{\delta^2\,m^{-2/\alpha}}{\alpha^4}\,[\Gamma'(1-1/\alpha) + (\log m)\,\Gamma(1-1/\alpha)]^2\,V(\hat{\alpha})$$

$$+ 2\,m^{-1/\alpha}\,\Gamma(1-1/\alpha)\,C(\hat{\lambda},\hat{\delta}) + \frac{2\,m^{-1/\alpha}}{\alpha^2}\,[\Gamma'(1-1/\alpha) + (\log m)\,\Gamma(1-1/\alpha)]\,C(\hat{\lambda},\hat{\alpha})$$

$$+ \frac{2\,m^{-2/\alpha}\,\Gamma(1-1/\alpha)\,\delta}{\alpha^2}\,[\Gamma'(1-1/\alpha) + (\log m)\,\Gamma(1-1/\alpha)]\,C(\hat{\delta},\hat{\alpha})$$

(if $2 < \alpha$*)* ; the mean-square error of the predictor is obtained by adding $V(X) = [\Gamma(1-2/\alpha) - [\Gamma^2(1-1/\alpha)]\,\delta^2$.

The computation of the asymptotic variances is easily made from the formulae given, substituting $(\lambda, \delta, \alpha)$ by $(\hat{\lambda}, \hat{\delta}, \hat{\alpha})$.

Overpassing probability for a level

The probability of overpassing a level $a\ (>\lambda)$ is

$$P(a) = P(a\,|\lambda, \delta, \alpha) = \text{Prob}\,\{X > a\} = 1 - \Phi_\alpha\left(\frac{a-\lambda}{\delta}\right)$$

which is estimated by substituting $(\lambda, \delta, \alpha)$ by its maximum likelihood estimators $(\hat{\lambda}, \hat{\delta}, \hat{\alpha})$, already given. Then

$$\sqrt{n}\,(\hat{P}(a) - P(a))$$

is asymptotically equivalent, by the δ-method, to its linearized form

$$\sqrt{n}\,[(\hat{\lambda}-\lambda)\frac{\partial P(a)}{\partial\lambda} + (\hat{\delta}-\delta)\frac{\partial P(a)}{\partial\delta} + (\hat{\alpha}-\alpha)\frac{\partial P(a)}{\partial\alpha}] =$$

$$= \sqrt{n}\,(1 - P(a))\,(-\log(1 - P(a))\,\{\alpha\frac{\hat{\lambda}-\lambda}{a-\lambda} + \alpha\frac{\hat{\delta}-\delta}{\delta} - \log\frac{a-\lambda}{\delta}\,(\hat{\alpha}-\alpha)\},$$

and so $\sqrt{n}\,\{\hat{P}(a) - P(a)\}$ is asymptotically normal with mean value zero and variance

$$V(a\,|\lambda, \delta, \alpha) = V(a) = (1 - P(a))^2\,(\log(1 - P(a))^2\ \times$$

$$\lim_{n\to\infty} n\,\{\alpha^2\frac{V(\hat{\lambda})}{(a-\lambda)^2} + \frac{\alpha^2 V(\hat{\delta})}{\delta^2} + (\log\frac{a-\lambda}{\delta})^2\ V(\hat{\alpha}) +$$

$$+ 2\frac{\alpha^2}{(a-\lambda)\,\delta}\,C(\hat{\lambda}, \hat{\delta}) - 2\,\alpha\frac{\log((a-\lambda)/\delta)}{a-\lambda}\,C(\hat{\lambda}, \hat{\alpha}) - 2\frac{\alpha}{\delta}\,\log\frac{a-\lambda}{\delta}\,C(\hat{\delta}, \hat{\alpha})\}.$$

Substituting $(\lambda, \delta, \alpha)$ by their maximum likelihood estimators $(\hat{\lambda}, \hat{\delta}, \hat{\alpha})$, we also get $\hat{V}(a) = V(a|\hat{\lambda}, \hat{\delta}, \hat{\alpha}) \sim V(a)$ and, thus $\sqrt{n}\,\dfrac{\hat{P}(a) - P(a)}{\sqrt{\hat{V}(a)}}$ is asymptotically standard normal, and so the one-sided asymptotic confidence interval with significance level α' for $P(a)$ is

$$0 \le P(a) \le \hat{P}(a) + \frac{\lambda_{\alpha'}}{\sqrt{n}}\sqrt{\hat{V}(a)}$$

as before.

When λ is known (0 for convenience) we have $P(a) = P(a \mid \delta, \alpha) = 1 - \Phi_\alpha(a/\delta)$, which corresponds to the case of some rivers and wave heights. Then the variance–covariance matrix of the maximum likelihood estimators $(\hat{\delta}, \hat{\alpha})$ is

$$\frac{1}{n}\begin{bmatrix} \frac{\alpha^2}{\delta^2} & \frac{1-\gamma}{\delta} \\ \frac{1-\gamma}{\delta} & \frac{1}{\alpha^2}\left(\frac{\pi^2}{6}+(1-\gamma)^2\right) \end{bmatrix}^{-1} = \frac{1}{n}\begin{bmatrix} \frac{\delta^2}{\alpha^2}\left(1+\frac{6}{\pi^2}(1-\gamma)^2\right) & \frac{6(1-\gamma)}{\pi^2}\delta \\ \frac{6(1-\gamma)}{\pi^2}\delta & \frac{6\alpha^2}{\pi^2} \end{bmatrix}$$

and thus with $\hat{P}(a) = P(a \mid \hat{\delta}, \hat{\alpha})$, $\sqrt{n}\,(\hat{P}(a) - P(a))$ is asymptotically equivalent to

$$\sqrt{n}\,\left[(\hat{\delta} - \delta)\frac{\partial P(a)}{\partial \delta} + (\hat{\alpha} - \alpha)\frac{\partial P(a)}{\partial \alpha}\right] =$$

$$= \sqrt{n}\,(1-P(a))\,(\log(1-P(a)))\left[\alpha\frac{\hat{\delta}-\delta}{\delta} + \log\frac{\delta}{a} - (\hat{\alpha}-\alpha)\right].$$

Thus $\sqrt{n}\,(\hat{P}(a) - P(a))$ is asymptotically normal with zero mean value and variance

$$V(a) \sim (1-P(a))^2\,(\log(1-P(a)))^2\,\{1-\gamma+\alpha\,\log\frac{\delta}{a}\};$$

$\hat{V}(a)$, as before, denotes $V(a)$ when (δ, α) are substituted by $(\hat{\delta}, \hat{\alpha})$, and from the practical point of view, for large n, we can use the fact that

$\sqrt{n}\,\dfrac{\hat{P}(a) - P(a)}{\sqrt{\hat{V}(a)}}$ *is asymptotically standard normal. The one–sided asymptotic confidence interval for the probability of overpassing the level a is*

$$0 \le P(a) \le \hat{P}(a) + \frac{\lambda_{\alpha'}}{\sqrt{n}}\sqrt{\hat{V}(a)}$$

with significance level α' .

A worked example :

Let us use the data given in the worked example for the Gumbel distribution, but for 24h duration, which is said to verify the Fréchet distribution.

We then get $\hat{\lambda} = -10.615183$, $\hat{\delta} = 38.998362$, and $\hat{\alpha} = 4.318996$. The Kolmogoroff -Smirnov statistic is $\hat{KS}_n = \sup_x | S_n(x) - \Phi^{\wedge}_{\alpha}((x - \hat{\lambda}) / \hat{\delta}) | = .0681721$ and as $n = 35$ we have $\sqrt{n}\ \hat{KS}_n = .40331$, accepting thus the Fréchet model.

Chapter VII

STATISTICAL DECISION FOR WEIBULL DISTRIBUTION

(FOR MINIMA)

Introduction

The Weibull distribution for minima is one of the asymptotic distributions for minima when the initial distribution has a left-end point (recall that the other possible asymptotic distribution is the Gumbel distribution for minima). Thus any Weibull random variable (for minima) has a finite location parameter λ (a *threshold*) and it takes the form

$$W_\alpha(x|\lambda,\delta) = W_\alpha((x-\lambda)/\delta) = 1 - \exp\{-(\frac{x-\lambda}{\delta})^\alpha\} \quad \text{for} \quad x \geq \lambda$$

$$W_\alpha(x|\lambda,\delta) = W_\alpha((x-\lambda)/\delta) = 0 \quad \text{for} \quad x \leq \lambda.$$

For $\alpha = 1$ we have the exponential distribution and for $\alpha = 2$ we have the Rayleigh distribution important in delay-times and Oceanography, respectively. Evidently $W_\alpha(z) = 1 - \Psi_\alpha(-z)$.

It is immediate that $Y = ((X-\lambda)/\delta)^\alpha$ has the standard exponential distribution. Also if $\lambda = \lambda_o$ is known (it can be taken as zero, for convenience, by subtraction) $Y = -\log(X - \lambda_o)$ has a Gumbel distribution for maxima with the location and dispersion parameters $-\log\delta$ and $1/\alpha$. Thus the theory for Gumbel distribution can be applied and the maximum likelihood estimators are (if $\lambda_o = 0$)

$$\frac{n}{\hat{\alpha}} - \sum_1^n \log x_i - n \frac{\sum_1^n x_i^{\hat{\alpha}} \cdot \log x_i}{\sum_1^n x_i^{\hat{\alpha}}} = 0$$

$$\hat{\delta} = \frac{1}{n} \sum_{1}^{n} x_i^{\hat{\alpha}}.$$

The solution of the first equation needs an initial seed which can be obtained either using the fact that $\{-\log x_i\}$ has a general Gumbel distribution where $1/\alpha$ is the dispersion parameter among other ways.

The random pair $(\hat{\delta}, \hat{\alpha})$ *has a binormal asymptotic behaviour with mean values* (δ, α), *variances* $(1 + \frac{6}{\pi^2}(1-\gamma)^2)\,\delta^2/n\alpha^2$ *and* $6\,\alpha^2/n\,\pi^2$, *covariance* $\frac{6(1-\gamma)\,\delta}{n\,\pi^2}$, *and correlation coefficient* $\rho = (1 + \frac{\pi^2}{6(1-\gamma)^2})^{-1/2} = .31307$.

The confidence intervals; tests, estimators of quantiles, predictors, etc., can be dealt with by the usual techniques, as done before in previous chapters.

Weibull (1939) gave an empirical derivation of this distribution in an analysis of dynamic breaking strengths by requiring only that is should meet certain practical criteria. That paper, and other related papers by Weibull dealing with the modelling of failure (or survival) data, seem to have found a large audience chiefly among those concerned with reliability analysis. The use of the Weibull distribution as a model for "time-to-failure" has thus been widespread. It arises naturally also in the analysis of droughts, in Oceanography, in Survival Analysis; see, for instance, Gumbel (1954).

The Weibull distribution for maxima has been used to model maximum temperatures, maximum wind speeds and maximum earthquake magnitudes; see Jenkinson (1955) and Yegulalp and Kuo (1974).

Other methods of estimation, as the ones given for the Gumbel distribution, can be used by the reduction to $y_i = -\log x_i$.

Recall that if X has a Fréchet distribution with dispersion and shape parameters δ and α, respectively, then 1/X has a two-parameter Weibull distribution with shape parameter α and dispersion parameter δ^{-1}. Thus, when data have not been censored, estimation procedures that have been derived for the two-parameter Fréchet distribution can be used to estimate parameters of the Weibull distribution by simply taking reciprocals of the observed values or directly when appropriate.

Methods applicable for Weibull censoring on the right will be appropriate for censoring Fréchet data sets on the left, i.e., the smallest data values, etc.

For these censorings, and for obtaining best invariant estimators of $-\log \delta$ and $1/\alpha$ among those which are linear in the Gumbel order statistics (logarithms of the Weibull order statistics in reverse order), weights have been calculated and tabulated by Laue (1974) for $n = 1(1)\ 25$, $r = 2(1)n$. For zero censoring they are equal to the weights calculated and tabulated by Mann (1967) for best linearly invariant estimation from the smallest $r \leq n$ Gumbel order statistics.

From best linearly invariant estimators of $-\log \delta$ and α^{-1} one can obtain estimators of the location and dispersion parameters and best linearly estimators of any quantile of the distribution. Invariant estimators and maximum likelihood equations yield estimators of α^{-1} and $-\log \delta$ that tend to be very nearly equal in value, ant the constants in the variance-covariance matrix of the best linearly unbiased estimators can be used effectively to remove the bias from maximum likelihood estimators if approximate unbiased estimators are desired. All these types of estimators are asymptotically unbiased, asymptotically efficient and asymptotically normal.

Characteristics of the Weibull distribution

From the definition of the distribution function

$$W_\alpha(x|\lambda,\delta) = 1 - \Psi_\alpha(-x|\lambda,\delta)$$

we can get, when they exist, the usual coefficients. We will obtain them for the reduced random variable $Z = (X - \lambda)/\delta$, the transfer of the general case being immediate.

The computation of β_1 and β_2, also depending on α, is left as an exercise.

The graphs of probability densities $W'_\alpha(z) = \alpha\, z^{\alpha-1} \exp(-z^\alpha)$ $(z \geq 0)$ for different values of $\alpha > 0$ are given in Fig. VII.1.

Fig.VII.1 - Reduced Weibull densities (for minima).

Evidently as $W_\alpha(z) = 1 - \Psi_\alpha(-z)$ the density $W'_\alpha(z)$ is the mirror image of $\Psi'_\alpha(z)$.

We have

$$\mu'_r = M(Z^r) = \Gamma(1 + r/\alpha)$$

and so the mean value is $\mu = \Gamma(1 + r/\alpha)$, the variance $\sigma^2 = \Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)$, the corresponding values for X being obvious.

Also we have the mode $\mu' = 0$ if $\alpha \leq 1$, $\mu' = (1 - 1/\alpha)^{1/\alpha}$ if $\alpha \geq 1$, the median $\tilde{\mu} = (\log 2)^{1/\alpha}$ and the p-quantile $\chi_p = (-\log(1-p))^{1/\alpha}$ with $\chi_{1-1/e} = 1$.

Note that $\alpha(1 - Z)$, with Z having the distribution function $W_\alpha(z)$, has mean value $\alpha(1 - \Gamma(1 + 1/\alpha)) = \gamma - (\frac{\pi^2}{6} + \gamma^2) / 2\alpha + O(\alpha^{-2})$ and variance $\alpha^2(\Gamma(1 + 2/\alpha)) - \Gamma^2(1 + 1/\alpha) = \frac{\pi^2}{6} + O(\alpha^{-1})$ which suggests that $\alpha(1 - Z)$ converges to a Gumbel distribution of maxima when $\alpha \to \infty$ and so, for large α, we can use it as an approximation. The exact proof is left as an exercise.

The moment generating function of Z does not have an explicit expression but that of $\log Z$ (which is a Gumbel random variable of minima) is

$$M(e^{\theta \log Z}) = M(Z^\theta) = \Gamma(1 + \theta/\alpha) .$$

The computation of β_1 and β_2, also depending on α, is left as an exercise.

Estimation and testing

Some difficulties can appear for the Weibull distribution depending on the interplay of the location (λ) and shape parameter (α).

If $\lambda = \lambda_o$ is known, as said before, there is no problem with its reduction to the Gumbel distribution for maxima.

Let us consider first the situation when α is known and (λ, δ) unknown — a real location-dispersion case. It is very easy to verify that if $\alpha < 1$, $\hat{\lambda} = x'_1 = \min(x_1, \ldots, x_n)$ and *any positive statistic* $\delta^*(x_1, \ldots, x_n) > 0$ are maximum likelihood estimators of (λ, δ)!

For $\alpha = 1$ (exponential distribution) we have $\hat{\lambda} = x'_1 = \min(x_1, \ldots, x_n)$ and $\hat{\delta} = \bar{x} - \hat{\lambda}$, but owing to the influence of the minimum $\hat{\lambda}$ the asymptotic behaviour of $(\hat{\lambda}, \hat{\delta})$ is not binormal, because $\hat{\lambda}$ also has the same exponential distribution with location parameter λ and dispersion parameter δ/n. Note, also, that $P(\hat{\lambda} > \lambda) = 1$.

It can remarked that the minimum-variance unbiased estimators of λ and δ are

$$\lambda^* = (n\, x'_1 - \bar{x})/(n-1) = \hat{\lambda} - \hat{\delta}/(n-1)$$

and

$$\delta^* = n(\bar{x} - x'_1)/(n-1) = n\,\hat{\delta}/(n-1);$$

see Epstein and Sobel (1954) for details.

For $1 \leq \alpha \leq 2$ the maximum likelihood estimators exist but the Cramér-Rao bound, using the second derivatives, does not. Other bounds, valid throughout the range of α, were found in Tiago de Oliveira and Littauer (1972). Are smaller than the Cramér-Rao bound in the range $\alpha > 2$, as shown in the paper by Tiago de Oliveira (1968).

In prediction results we will develop these bounds.

For $1 \leq \alpha \leq 2$ the maximum likelihood estimator $\hat{\lambda}$ is not asymptotically normal, and whether or not it is efficient is an open question; see Woodroofe (1974). For known $\alpha < 2$, Tiago de Oliveira (1983) has shown that $\delta^{*\alpha} = \sum_{i=1}^{n} (x_i - x'_1)^{\alpha}/n$ has all the asymptotic properties of the maximum likelihood estimator of α. If α is known but λ and δ are unknown, the tables of Harter and Dubey (1967) can be used to test hypotheses concerning the mean value and variance.

The case where $\alpha > 2$ is simple and regular; see Cramér (1946). The maximum likelihood estimators of (λ, δ) for $\alpha = \alpha_o (>2)$ known, are given by the equations

$$\frac{\sum_{1}^{n} (x_i - \hat{\lambda})^{\alpha_o - 1}}{\sum_{1}^{n} (x_i - \hat{\lambda})^{\alpha_o}} = \frac{\alpha_o - 1}{n\,\alpha_o} \sum_{1}^{n} \frac{1}{x_i - \hat{\lambda}}$$

$$\hat{\delta} = \left(\frac{1}{n} \sum_{1}^{n} (x_i - \hat{\lambda})\right)^{1/\alpha_o}.$$

The random pair $(\hat{\lambda}, \hat{\delta})$ is asymptotically normal with mean values (λ, δ), variances $\alpha_o^2 \delta^2/n\Delta$, $(\alpha_o-1)^2 \Gamma(1 - 2/\alpha_o)\delta^2/n\Delta$, covariance $-\alpha_o(\alpha_o-1)\Gamma(1 - 1/\alpha_o)\delta^2/n\Delta$, and correlation coefficient $\rho = -\Gamma(1 - 1/\alpha_o)/\sqrt{\Gamma(1 - 2/\alpha_o)}$, where Δ denotes

$$\alpha_o^2 (\alpha_o-1)^2 [\Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o)].$$

From these results the usual tests and confidence intervals can be made. The estimator $\hat{\lambda} + \chi_p \hat{\delta}$ *of the p-quantile* $\lambda + \chi_p \delta$ *is asymptotically normal with mean value* $\lambda + \chi_p \delta$ *and variance*

$$((\alpha_o - (\alpha_o-1)\Gamma(1 - 1/\alpha_o)\chi_p)^2 + (\alpha_o-1)^2 [\Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o)]\chi_p)\delta^2/n\Delta$$

$$= \left(\frac{(\alpha_o - (\alpha_o-1)\Gamma(1 - 1/\alpha_o)\chi_p)^2}{\Delta} + \frac{\chi_p^2}{\alpha_o^2}\right)\delta^2/n$$

$$= \left(\frac{(1 - \Gamma(2 - 1/\alpha_o)\chi_p)^2}{(\alpha_o-1)^2 (\Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o))} + \frac{\chi_p^2}{\alpha_o^2}\right)\delta^2/n.$$

For prediction, the mean value of the minimum of the next m *observations is* $\lambda + \Gamma(1 + 1/\alpha_o)/m^{1/\alpha_o}\ \delta$. *As before we use as predictor* $\hat{\lambda} + \Gamma(1 + 1/\alpha_o)/m^{1/\alpha_o}\ \hat{\delta}$ *and the mean square error of the predictor is, as we can take* $\chi_p = \Gamma(1 + 1/\alpha_o)$,

$$(\Gamma(1 + 2/\alpha_o) - \Gamma^2(1 + 1(\alpha_o))\delta^2 +$$

$$\left(\frac{1 - \Gamma(2 - 1/\alpha_o)\ \Gamma(1 + 1(\alpha_o))^2}{(\alpha_o-1)^2 (\Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o))} + \frac{\Gamma^2(1 + 1/\alpha_o)}{\alpha_o^2}\right)\delta^2/n.$$

Let us now consider the 3-parameter Weibull distribution. The state of affairs in this case is very awkward, chiefly because we have to estimate α with a global behaviour changing according to its real value; see Dubey (1967) and (1967') and Mann (1968) to get the first feelings of the difficulties of the problem, chiefly connected with the non-existence of maximum likelihood

estimators for all values of α. Cohen and Whitten (1982) advise that one should attempt to obtain maximum likelihood estimators for the three-parameter Weibull distribution unless there is reason to expect $\alpha > 2.2$.

With all these restrictions, let us write the maximum likelihood equations. From the likelihood of the sample $(x_1, \ldots, x_n)$

$$L(\lambda, \delta, \alpha | x_i) = 0 \qquad \text{if} \qquad \min(x_1, \ldots, x_n) \leq \lambda$$

$$L(\lambda, \delta, \alpha | x_i) = \frac{\alpha^n}{\delta^{\alpha n}} \prod_1^n (x_i - \lambda)^{\alpha-1} \exp\{-\sum_1^n (\frac{x_i - \lambda}{\delta})^\alpha\} \quad \text{if } \min(x_1, \ldots, x_n) > \lambda$$

we get

$$\hat{\delta}^{\hat{\alpha}} = \frac{1}{n} \sum_1^n (x_i - \hat{\lambda})^{\hat{\alpha}}$$

$$\frac{\hat{\alpha} - 1}{\hat{\alpha}} \sum_1^n \frac{1}{x_i - \hat{\lambda}} = \frac{n \sum_1^n (x_i - \hat{\lambda})^{\hat{\alpha}-1}}{\sum_1^n (x_i - \hat{\lambda})^{\hat{\alpha}}}$$

$$\frac{1}{\hat{\alpha}} + \frac{1}{n} \sum_1^n \log(x_i - \hat{\lambda}) = \sum_1^n (x_i - \hat{\lambda})^{\hat{\alpha}} \log(x_i - \hat{\lambda}) / \sum_1^n (x_i - \hat{\lambda})^{\hat{\alpha}}.$$

For this system we have to solve for $(\hat{\lambda}, \hat{\alpha})$, as a first step, the last two equations and then compute $\hat{\delta}$. If $\alpha > 2$ the system is asymptotically trinormal and the previous techniques applied to the Fréchet distribution can be used. Those equations can be compared, and are similar, to those for the Fréchet distribution of maxima, as could be expected. Also it can be shown that:

The random triple $(\hat{\lambda}, \hat{\delta}, \hat{\alpha})$ is asymptotically trinormal with mean values $(\hat{\lambda}, \hat{\delta}, \hat{\alpha})$ if $\alpha > 2$ and variance-covariance matrix analogous to the one for Fréchet distribution.

The real important situations are considered below, particularly in the subsection "The maximum likelihood estimator when $\alpha_0 > 2$".

The usual techniques of confidence interval estimation and hypothesis testing can be applied as before. Prediction will be discussed in the next section.

Lemon (1975) modified the likelihood equations so that one need iteratively solve only two equations for estimates of λ and α, which together then specify an estimate of δ. Cohen and Whitten (1982) suggest other modifications to the maximum likelihood methods which involve simplification of one of the equations.

For estimation of the three Weibull parameters, the method of moments seems to work well, in general, as estimators can always be obtained. The tables of Dubey (1967) applying to the two-parameter distribution, and the Monte Carlo results of Cohen and Whitten (1982) applying to the three-parameter distribution, indicate that the moment estimators are very efficient for $\alpha = 4.7$ and become progressively less efficient as α converges towards 2 or towards $+\infty$. Cohen and Whitten (1983) provide a table which facilitates estimation of α from the distribution skewness coefficient β_1. These authors also give several modifications to the moment equations which simplify estimation of all three parameters.

Another iterative procedure for estimating λ is one by Mann and Fertig (1973). If α is not too large, this method also provides a lower confidence bound for λ. Also we can use the proposed statistic to test $\lambda = 0$ vs. $\lambda > 0$, which is very important in fatigue failure research. The basic statistic for this test is a ratio of selected sums of differences of successive ordered logarithms of sample observations, each difference being divided by its expected value. To obtain a point estimate of λ (which is median unbiased) or a lower confidence bound, the initial guess of λ must be subtracted from each observation and iterated until the modified statistic is numerically equal to a percentile of the statistics distribution. The method has two advantages: the test statistic, under $\alpha = 0$, has a Beta distribution, and it is appropriate for use with censored data. The later advantage can also be claimed for a similar statistic by Weissman (1978) and for certain linear estimators of $-\log \delta$ and α^{-1} that can be applied to the ordered logarithms of the observed data after an estimate of λ is subtracted from each of the original values.

Wyckoff, Bain and Engelhardt (1980) suggest the use of a simplified linear estimator of α^{-1}, after having substituted λ by a simple estimator, as an approximation of the best linear unbiased estimator. This seems to yield appropriate and quite efficient estimates for α^{-1}, and ultimately for $\log \delta$, if $\alpha \leq 3.0$ when there is no censoring (or if α is somewhat less than 3.0 when there is censoring). For larger values of α, estimates of α^{-1} obtained by this method have a substantial negative bias unless sample size is extremely large.

Besides the linear estimators of $\log \delta$ and α^{-1}, there are other simple estimators of the Weibull parameters that are appropriate for use under certain conditions. For $\alpha \leq 1$, a special case suggested by Zanakis (1979) of the estimator of λ previously proposed by Dubey (1966),

namely, $\lambda' = (x_1' x_n' - x_2'^2)/(x_1' + x_n' - 2x_2')$, has relatively small bias and mean-square error. For $\alpha > 1$, a statistic investigated by Kappenmann (1981) has smaller bias and mean-square error than the last one for estimating λ, but is also considerably more difficult to compute, involving a weighted average of all the order statistics.

Zanakis and Mann (1982) compared two simple estimators of α based on a few order statistics :

$$\alpha_1^* = \log [(x_k' - x_j')/(x_j' - x_i')] / 3.643$$

and

$$\alpha_2^* = \log [(x_m' - \lambda')/(x_h' - \lambda')] / 2.989$$

with $1 \le i < j < k \le n$, $1 \le h < m \le n$ defined appropriately and λ' given above.

They found both estimators to be good for small values of α, but having more negative bias than estimators based on the procedures prescribed by Wyckoff, Bain and Engelhardt (1986). The simple estimators are useful for first guesses in maximum-likelihood estimation and in other situations in which quick results are necessary.

Point prediction

As before, the mean-square point predictor of the minimum of the next m observations is evidently the estimator of the mean value of $W = \min (Y_1, ..., Y_n)$. The distribution of W is, as seen before, $W_\alpha(\frac{x - \lambda}{\delta m^{-1/\alpha}})$ if the distributions of Y_j, as well as those of x_j, are $W_\alpha ((x - \lambda)/\delta)$. The mean value of W is

$$M(W) = \lambda + m^{-1/\alpha} \Gamma(1 + 1/\alpha) \delta ,$$

and thus a point predictor of W can be obtained substituting $(\lambda, \delta, \alpha)$ by their estimators. But difficulties exist when λ is not known and α may be ≤ 2. If $\alpha = \alpha_0$ is known and $\alpha_0 > 2$ (practically speaking $\alpha_0 > 2.1$), there are no difficulties with the maximum likelihood estimation; likewise if $(\lambda, \delta, \alpha)$ are all unknown but it is known that $\alpha > 2$. The general theory used before for the Fréchet distribution can also be applied here.

If $\lambda = \lambda_o$ is known then $-\log(x - \lambda_o)$ has a Gumbel distribution with location and dispersion parameters $-\log \delta$ and α^{-1} and estimation can be dealt with by the methods of Chapter V.

The theory of best quasi-linear prediction when $\alpha = \alpha_o$ leads to an unmanageable formula for the predictor, following Tiago de Oliveira (1966) as we will recall briefly.

Later we will make considerable use of Tiago de Oliveira and Littauer (1976).

The likelihood of the observed sample is

$$\bar{L}_{\alpha_o}(\lambda, \delta \mid x_i) = 0 \qquad \text{if} \quad \min(x_i) > \lambda$$

$$\bar{L}_{\alpha_o}(\lambda, \delta \mid x_i) = \prod_1^n \{\frac{1}{\delta} W'_{\alpha_o}(\frac{x_i - \lambda}{\delta})\} \qquad \text{if} \quad \min(x_i) > \lambda$$

and the likelihood of the sample and the future minimum W from m observations is

$$L_{\alpha_o}(\lambda, \delta \mid x_i, w) = \bar{L}_{\alpha_o}(\lambda, \delta \mid x_i, \ldots, x_n) \frac{1}{\delta m^{-1/\alpha_o}} W'_{\alpha_o}(\frac{w - \lambda}{\delta m^{-1/\alpha_o}})$$

where $W_\alpha(z)$ was defined above.

According to the paper previously cited, the mean-square predictor has the expression

$$p_m(x_i) = \frac{\int_0^{+\infty} d\xi\, \xi^{n-1} \int_{-\infty}^{+\infty} d\eta\, \bar{L}(\eta + \xi x_1, \ldots, \eta + \xi x_n)[\mu_m(\eta + \xi x_1, \ldots, \eta + \xi x_n) - \eta]}{\int_0^{+\infty} d\xi\, \xi^n \int_{-\infty}^{+\infty} d\eta\, \bar{L}(\eta + \xi x_1, \ldots, \eta + \xi x_n)}$$

where $L(z_1, \ldots, z_n; v)$ denotes the likelihood of the first sample and the minimum of the second sample, $\bar{L}(x_1, \ldots, x_n) = \int_{-\infty}^{+\infty} d\,v\; L(x_1, \ldots, x_n; v)$ is the marginal likelihood of the sample, and

$$\mu_m(x_i) = \int_{-\infty}^{+\infty} d\,v\; v\, L(x_1, \ldots, x_n; v) / \bar{L}(x_1, \ldots, x_n)$$

is the conditional mean. As it is irrelevant, from the expression of $p_m(x_i)$, we can always suppose $\lambda = 0$ and $\delta = 1$. This predictor was shown in that paper to be quasi-linear, i.e., $p_m(\lambda + \delta x_i) = \lambda + \delta p_m(x_i)$, for every λ and $\delta\ (> 0)$.

Note that, in our case, we have

$$\mu_m(x_i) = \mu_{\alpha_o,\, m} = \int_{-\infty}^{+\infty} d\, v\, m^{1/\alpha_o}\ v\ W'_{\alpha}(m^{1/\alpha_o}\ v) = \Gamma(1 + 1/\alpha_o)/m^{1/\alpha_o} = \mu_{\alpha_o,1}/m^{1/\alpha_o}\ .$$

and $\bar{L}(x_i) = L_{\alpha_o}(x_i)$.

As $W'_{\alpha_o}(z) = 0$ if $z \le 0$, the integration takes place in the region $\alpha + \beta x_i \ge 0$, that is, $\alpha \ge -\beta\, l\,(x_i)$ where $l\,(x_i) = \min(x_1, \ldots, x_n)$.

The predictor can be written as

$$p_{\alpha_o,\, m}\ (x_i) = \frac{\displaystyle\int_0^{+\infty} d\xi\ \xi^{n-1} \int_{l\,(x_i)}^{+\infty} d\eta\ (\mu_{\alpha_o,\, m} - \eta) \prod_{i=1}^{n} W'_{\alpha_o}(\eta + \xi\, x_i)}{\displaystyle\int_0^{+\infty} d\xi\ \xi^{n} \int_{l\,(x_i)}^{+\infty} d\eta \prod_{i=1}^{n} W'_{\alpha_o}(\eta + \xi\, x_i)}\ .$$

Putting $\zeta = \xi + \eta\, l\,(x_i)$ we get

$$p_{\alpha_o,m}(x_i) = l\,(x_i) + \frac{\displaystyle\int_0^{+\infty} d\xi\ \xi^{n-1} \int_0^{+\infty} d\zeta (\mu_{\alpha_o,m} - \zeta) \prod_{i=1}^{n} W'_{\alpha_o}(\zeta + \xi\,(x_i - l\,(x_i)))}{\displaystyle\int_0^{+\infty} d\xi\ \xi^{n} \int_0^{+\infty} d\zeta \prod_{i=1}^{n} W'_{\alpha_o}(\zeta + \xi\,(x_i - l\,(x_i)))}\ .$$

It is evident that computation facilities will not be useful at all, as the expression above implies the computation of the predictor $p_{\alpha_o,\, m}(x_i)$ for any n (sample size) and for any (practically) observable sample mean, except for special cases. For example, in the case of the exponential distribution ($\alpha_o = 1$), we have

$$W'_1(z) = 0 \quad \text{if} \quad z \leq 0$$
$$= e^{-z} \quad \text{if} \quad z \geq 0,$$

$\mu_{1,m} = 1/m$, and thus we obtain

$$p_{1,m}(x_i) = l(x_i) + \left(\frac{1}{m} - \frac{1}{n}\right)[\bar{x} - l(x_i)].$$

Lower bounds for the mean-square error

As the maximum likelihood estimators are not regular or do not exist if $\alpha \leq 2$, it is natural to search for a quasi-linear predictor and thus to try to obtain a lower bound for the mean-square error of any quasi-linear statistic $f(x_i)$ for $\lambda + \mu_{\alpha_o,m}\ \delta$.

The mean-square error of a quasi-linear statistic $f(x_i)$ is

$$M(f(x_i) - W)^2 = \delta^2\, M(f(z_i) - v)^2$$

where $x_i = \lambda + \delta z_i$ and $w = \lambda + \delta v$ (i.e., z_i and v are the reduced values of the observations). From now on we will deal, thus, with reduced observations ($\lambda = 0$, $\delta = 1$) because we are going to evaluate efficiencies and the factor δ^2 cancels out.

The values of $M(f(z_i) - v)^2$ can be split simply by putting $f(z_i) - v = (f(z_i) - \mu_{\alpha_o,m}) - (v - \mu_{\alpha_o,m})$ and so, by independence between the first and second sample, we have

$$M(f(z_i) - v)^2 = M(f(z_i) - \mu_{\alpha_o,m})^2 + M(v - \mu_{\alpha_o,m})^2$$

$= D^2_{n,m}(f) + \sigma^2_m(\alpha_o)$ = mean-square error of $f(z_i)$ about $\mu_{\alpha_o,m}$ + variance of v.

Thus we will study $D^2_{n,m}(f)$. It was shown in Tiago de Oliveira (1968), using the Schwarz inequality conveniently, that, in the general case, if we denote by $\varphi(\lambda, \delta)$ $(-\infty < \lambda < \infty,\ 0 < \delta < \infty)$

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \mu_m(\lambda + \delta z_1, \ldots, \lambda + \delta z_n)\, \bar{L}(z_1, \ldots, z_n)\, dz_1 \ldots dz_n$$

and by $w_n(\lambda, \delta)$

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\bar{L}^2(z_1, \ldots, z_n)}{\delta^n \bar{L}(\lambda + \delta z_1, \ldots, \lambda + \delta z_n)} \, dz_1 \ldots dz_n .$$

then

$$B^2 = \sup_{\lambda, \delta} \frac{(\lambda - \varphi(\lambda, \delta) + \varphi(0,1)\,\delta)^2}{w_n(\lambda, \delta) - 2\delta + \delta^2}$$

is a lower bound for the mean-square error $D^2_{n,m}(f) = M(f(z_i) - \mu_m(z_i))^2$ of a quasi-linear statistic $f(z_i)$ (when $\lambda = 0, \delta = 1$). In the expression of $w_n(\lambda, \delta)$ in the integrals the indeterminate ratio $\frac{0}{0}$ is taken to be equal to zero.

B^2 is indeterminate for $\lambda = 0, \delta = 1$. In our case, owing to independence, we have

$$\varphi(\lambda, \delta) = \mu_{\alpha_o, m} \quad \text{and} \quad w_n(\lambda, \delta) = w^n_{\alpha_o}(\lambda, \delta)$$

where

$$w_{\alpha_o}(\lambda, \delta) = \int_{-\infty}^{+\infty} \frac{W'_{\alpha_o}(z)^2}{\delta\, W'_{\alpha_o}(\lambda + \delta z)} \, dz$$

and thus we have

$$B^2 = \sup_{\lambda, \delta} \frac{(\lambda + \delta\, \mu_{\alpha_o, m} - \mu_{\alpha_o, m})^2}{w^n_{\alpha_o}(\lambda, \delta) - 2\delta + \delta^2} .$$

As B^2 is almost impossible to obtain, we will put $\lambda = 0$ above and compute

$$\bar{B}^2 = \sup_{\delta} \frac{\mu^2_{\alpha_o, m}\,(1 - \delta)^2}{w^n_{\alpha_o}(0, \delta) - 2\delta + \delta^2} \leq B^2 .$$

and as

$$w_{\alpha_o}(0, \delta) = 1/[\,\delta^{\alpha_o}\,(2 - \delta^{\alpha_o})] \qquad \text{for} \quad 0 < \delta^{\alpha_o} < 2 ,$$

we get the (smaller) lower bound

$$\bar{B}^2 = \frac{\mu^2_{\alpha_o,m}}{1 + n\,\alpha_o^2} \sim \frac{\mu^2_{\alpha_o,m}}{n\,\alpha_o^2} = \frac{\mu^2_{\alpha_o,1}}{n\,\alpha_o^2\, m^{2/\alpha_o}}$$

and an (under) evaluation of efficiency is

$$\lim \frac{\bar{B}^2}{D^2_{n,m}(f)} = \frac{\mu^2_{\alpha_o,1}}{\alpha_o^2\, m^{2/\alpha_o} \lim n\, D^2_{n,m}(f)} .$$

For $\alpha_o = 1$, $\bar{B}^2 \sim D^2_{n,m}(f)$ when $f(x_i) = l(x_i) + (\frac{1}{m} - \frac{1}{n})\,[\bar{x} - l(x_i)]$ is the best predictor for the exponential distribution, already given.

In case of regularity, as happens for $\alpha_o > 2$, we can obtain bounds analogous to the Cramer-Rao bounds for estimation. In the last paper referred to it was shown, for the case of independence with the density $g(z)$ for the (reduced) z_i and the mean μ_m for v, to have the expression

$$B'^2 \sim \frac{S_o\,\mu_m^2 - 2\,S_1\,\mu_m + S_2 - 1}{n(S_o\,S_2 - S_1^2 - S_o)}$$

with

$$S_o = \int_{-\infty}^{+\infty} x^p\, \frac{g'(x)^2}{g(x)}\, dx .$$

Putting $g(z) = W'_{\alpha_o}(z)$ and $\mu_m = \mu_{\alpha_o,m}$ we get, as

$$S_p = [(\alpha_o - 1)^2 + p\,(\alpha_o + p - 2)]\,\Gamma\,(1 + \frac{p-2}{\alpha_o}) ,$$

$$B'^2 \sim \frac{\alpha_o^2 - 2\,\alpha_o(\alpha_o - 1)\,\Gamma\,(1 - 1/\alpha_o)\,\mu_{\alpha_o,m} + (\alpha_o - 1)^2\,\Gamma(1 - 2/\alpha_o)\,\mu^2_{\alpha_o,m}}{n\,\alpha_o^2\,(\alpha_o - 1)^2\,[\Gamma\,(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o)]} .$$

Asymptotically we have $\bar{B}^2 \le B'^2$ and so the (under-) evaluation of the efficiency is to be taken as

$$B'^2 / D^2_{n,m}(f) .$$

In fact, as the maximum likelihood estimators are the best when regular ($\alpha_o > 2$), we are going to obtain a set of simple predictors, chiefly for $\alpha_o \leq 2$, such that they have reasonably good efficiency for some interval in $]0,2]$.

Some predictors for the shape parameter $\alpha_o \leq 2$

It is desirable to explore different predictors for different ranges of α_o because of computational difficulties. Let us begin by examining the region $\alpha_o \leq 2$, with a suggestion from the result for the exponential ($\alpha_o = 1$).

The simple predictor for the exponential distribution suggests the use of the statistic

$$l(x_i) + a(\alpha_o)(\bar{x} - l(x_i))$$

as predictor, where $a(\alpha_o)$ is to be determined so as to minimize the mean-square error, which for reduced values is

$$D^2_{n,m} = M(l(x_i) + a(\alpha_o)(\bar{x} - l(x_i)) - \mu_{\alpha_o,m})^2 .$$

This predictor is suggested also by the fact that the lower limit λ has $l(x_i)$ as a sufficient estimator (but an over-estimator).

The sample average is a quasi-linear statistic converging to $\mu_{\alpha_o,1} = \lambda + \delta\,\Gamma(1 + 1/\alpha_o)$, with variance $\delta^2\, b(\alpha_o)/n$ with

$$b(\alpha_o) = [\Gamma(1 + 2/\alpha_o) - \Gamma^2(1 + 1/\alpha_o)] ;$$

$l(x_i)$ has the mean value $\mu_{\alpha_o,n} = \lambda + \delta\,\Gamma(1 + 1/\alpha_o)/n^{1/\alpha_o}$, its variance being $\delta^2\, b(\alpha_o)/n^{2/\alpha_o}$; the correlation coefficient $\rho_{\alpha_o,n} = \rho(\bar{x}, l(x_i))$ converges to zero; see Tiago de Oliveira (1962) and Rosengard (1966).

Thus, the mean-square error $D^2_{n,m}$ for reduced observations as a function of a, can be written as $M[l(x_i) + a(\bar{x} - l(x_i)) - \mu_{\alpha_o,m}]^2 = M[l(x_i) - \mu_{\alpha_o,m}]^2 + 2\,M[(l(x_i) - \mu_{\alpha_o,m})(\bar{x} - l(x_i))]\,a +$

$M[\bar{x} - l(x_i)]^2\ a^2$. Using the decomposition $l(x_i) - \mu_{\alpha_o,m} = l(x_i) - \mu_{\alpha_o,n} + \mu_{\alpha_o,n} - \mu_{\alpha_o,m}$ and $\bar{x} - l(x_i) = \bar{x} - \mu_{\alpha_o,1} + \mu_{\alpha_o,1} - \mu_{\alpha_o,n} + \mu_{\alpha_o,n} - l(x_i)$, we have

$$M[(l(x_i) - \mu_{\alpha_o,m})^2] = \frac{b(\alpha_o)}{n^{2/\alpha_o}} + (\mu_{\alpha_o,n} - \mu_{\alpha_o,m})^2 ,$$

$$M[(l(x_i) - \mu_{\alpha_o,m})(\bar{x} - l(x_i))] = \rho_{\alpha_o,n} \frac{b(\alpha_o)}{n^{2/\alpha_o + 1/2}} - \frac{b(\alpha_o)}{n^{2/\alpha_o}} +$$

$$(\mu_{\alpha_o,n} - \mu_{\alpha_o,m})(\mu_{\alpha_o,1} - \mu_{\alpha_o,n}) ,$$

$$M[(\bar{x} - l(x_i))^2] = \frac{b(\alpha_o)}{n} + (\mu_{\alpha_o,1} - \mu_{\alpha_o,n})^2 + \frac{b(\alpha_o)}{n^{2/\alpha_o}} - 2\rho_{\alpha_o,n} \frac{b(\alpha_o)}{n^{1/\alpha_o + 1/2}} .$$

The minimum of the quadratic is attained at some value whose limit is

$$-a(\alpha_o) = \mu_{\alpha_o,m} / \mu_{\alpha_o,1} = m^{-1/\alpha_o} ,$$

thus the predictor is $l(x_i) + m^{-1/\alpha_o}(\bar{x} - l(x_i))$. The asymptotic values of the mean-square error are then :

$$D^2_{n,m} = \frac{\mu^2_{\alpha_o,m}\, b(\alpha_o)}{\mu^2_{\alpha_o,1}\, n} = \frac{b(\alpha_o)}{n\, m^{2/\alpha_o}} \qquad \text{if } \alpha_o < 2 ,$$

$$D^2_{n,m} \sim \frac{m - 2\sqrt{m} + 2 - \pi/4}{m\, n} \qquad \text{if } \alpha_o = 2 ,$$

$$D^2_{n,m} \sim \frac{\Gamma(1 + 2/\alpha_o)(1 - m^{-1/\alpha_o})^2}{n^{2/\alpha_o}} \qquad \text{if } \alpha_o > 2 .$$

Consequently, the asymptotic efficiency is, since $\bar{B}^2 \sim \frac{\mu^2_{\alpha_o,m}}{n\, \alpha^2_o}$,

$$\text{eff}(\alpha_o) = \frac{\Gamma^2(1 + 1/\alpha_o)}{\alpha^2_o\, b(\alpha_o)} \qquad \text{if } \alpha_o < 2 .$$

$$\text{eff}(\alpha_o) = \frac{\pi}{16(m - 2\sqrt{m} + 2 - \pi/4)} \qquad \text{if } \alpha_o = 2 ,$$

$$\text{eff}(\alpha_o) = 0 \qquad \text{if } \alpha_o > 2 .$$

Zero efficiency for $\alpha_o > 2$ results from the slow convergence of $l(x_i)$ to the parameter λ, an effect that in this region overcomes the more rapid convergence of the average $\bar{x}$. The efficiency for $\alpha_o > 2$ is independent of m. A short table of the efficiency for the above predictor is, for $m = 1$.

α_o	1/3	1/2	1	2-
eff (α_o)	0.5	0.8	1	0.91 .

The predictor for $1/2 < \alpha_o < 2$ is thus reasonably efficient. Outside this range for the shape parameter, another predictor must be found.

Consider now the usage of a sample quantile $Q(x_i)$ and the minimum value $l(x_i)$. It is apparent (by the argument given above), that, since the variance of a quantile is of order 1/n, a linear combination

$$l(x_i) + a(\alpha_o)\,(Q(x_i) - l(x_i))$$

may be efficient only in the region $\alpha_o \leq 2$; it will be useful in the region $\alpha_o < 1/2$. For known $\alpha = \alpha_o$, let us denote by υ the quantile of probability $\omega = W_{\alpha_o}(\upsilon)$. It is well known (Cramér (1946)) that the sample quantile $Q(x_i)$ for probability ω, with reduced observations, is asymptotically normal with expected value υ and asymptotic variance of the order of $\omega(1 - \omega)/n\, W'^{2}_{\alpha_o}(\upsilon)$. As

$$W_{\alpha_o}(z) = 1 - e^{-z^{\alpha_o}} , \quad z \geq 0 ,$$

the asymptotic variance is of the order of

$$\frac{e^{(\upsilon^{\alpha_o})} - 1}{n\,\alpha_o^2\,\upsilon^{2(\alpha_o - 1)}} .$$

The mean-square error, for reduced values, is

$$M[(l(x_i) + a(\alpha_o)\,(Q(x_i) - l(x_i)) - \mu_{\alpha_o,m})^2] = M[(l(x_i) - \mu_{\alpha_o,m})^2] +$$

$$2\,M[(l(x_i) - \mu_{\alpha_o,m})\,(Q(x_i) - l(x_i))]\; a(\alpha_o) + M[(Q(x_i) - l(x_i))^2]\; a^2(\alpha_o)$$

with

$$M[(l(x_i) - \mu_{\alpha_o,m})^2] = \frac{b(\alpha_o)}{n^{2/\alpha_o}} + (\mu_{\alpha_o,n} - \mu_{\alpha_o,m})^2 ,$$

$$M[(l(x_i) - \mu_{\alpha_o,m})(Q(x_i) - l(x_i))] = \rho'_{\alpha_o,n}\ \tau_{\alpha_o,n}\ \frac{b(\alpha_o)^{1/2}}{n^{1/\alpha_o}} - \frac{b(\alpha_o)}{n^{2/\alpha_o}}$$

$$+ (\mu_{\alpha_o,n} - \mu_{\alpha_o,m})\ (\upsilon_{\alpha_o,n} - \mu_{\alpha_o,n}) ,$$

$$M[(Q(x_i) - l(x_i))^2] = \tau^2_{\alpha_o,n} + (\upsilon_{\alpha_o,n} - \mu_{\alpha_o,n})^2$$

$$+ \frac{b(\alpha_o)}{n^{2/\alpha_o}} - 2\rho'_{\alpha_o,n}\ \tau_{\alpha_o,n}\ \frac{b(\alpha_o)^{1/2}}{n^{1/\alpha_o}} ,$$

where we have made a decomposition analogous to the previous one and where $\upsilon_{\alpha_o,n}$, $\tau^2_{\alpha_o,n}$ and $\rho'_{\alpha_o,n}$ denote the asymptotic mean value and the variance of the sample quantile and the correlation coefficient between the sample quantile and the minimum. As might be expected, $\rho'_{\alpha_o,n} \to 0$ (Rosengard (1966))* .

Denoting by υ_{α_o} the limiting value of $\upsilon_{\alpha_o,n}$, we see that the limiting value of $a(\alpha_o)$ that minimizes the mean-square error is

$$a(\alpha_o) = \mu_{\alpha_o,m} / \upsilon_{\alpha_o}$$

and the asymptotic mean-square error for $\alpha_o < 2$ is, for reduced values,

$$D^2_{n,m} = \frac{\mu^2_{\alpha_o,m}\ (e^{\upsilon^{\alpha_o}} - 1)}{n\ \alpha_o^2\ \upsilon^{2/\alpha_o}} .$$

For $\alpha_o = 2$ we have

---

* In Cramér (1946), p. 368, it is shown implicity that $\upsilon_{\alpha_o,n} - \upsilon_{\alpha_o} = O(n^{1/2})$.

$$D^2_{n,m} \sim \{ \frac{\pi(e^{\upsilon^2} - 1)}{16\,\upsilon^4} + (1 - \frac{\sqrt{\pi}}{2\sqrt{m}\,\upsilon})^2 \} .$$

We will only consider the problem for $\alpha_o < 2$. The case for $\alpha_o = 2$ must be dealt with numerically. The minimum mean-square error is attained for $\upsilon^{\alpha_o} = 1.593624$, and we have approximately

$$D^2_{n,m} \sim \frac{\mu^2_{\alpha_o,m}}{n\,\alpha^2_o} \left(\frac{e^{1.593624} - 1}{(1.593624)^2}\right) = \frac{1.5441386\,\mu^2_{\alpha_o,m}}{n\,\alpha^2_o}$$

so that the asymptotic efficiency is

$$\frac{1}{1.5441386} = 0.65 .$$

Taking the approximation $\upsilon^{\alpha_o} = 1.60$, the (constant) efficiency is

$\frac{1}{1.5441532} = 0.65$ also. The predictor will then be

$$l(x_i) + \mu_{\alpha_o,1}(1.60\,m)^{-1/\alpha_o}\,(Q(x_i) - l(x_i))$$

where $Q(x_i) = \chi^*_{.796}$ is then the sample quantile for probability $\omega = 1 - e^{-\upsilon^{\alpha_o}} = 1 - e^{-1.60} = 0.796$.

This technique, whose asymptotic efficiency is independent of m and α_o, is to be used for values of $\alpha_o \le 0.40$ where the efficiency of the predictor $l(x_i) + a(\alpha_o)(\bar{x} - l(x_i))$ is smaller than 0.65. For $\alpha_o = 2$ a better predictor will be considered.

The maximum likelihood predictor when $\alpha_o > 2$

The maximum likelihood estimators of λ and δ for known $\alpha_o (> 2)$ are given by the equations

$$\frac{\sum_1^n (x_i - \hat{\lambda})^{\alpha_o - 1}}{\sum_1^n (x_i - \hat{\lambda})^{\alpha_o}} = \frac{\alpha_o - 1}{n\,\alpha_o} \sum_1^n (x_i - \hat{\lambda})^{-1}$$

and

$$\hat{\delta}^{\alpha_o} = \frac{1}{n} \sum_{1}^{n} (x_i - \hat{\lambda})^{\alpha_o} \quad ,$$

where the variances and covariance are asymptotically

$$V(\hat{\lambda}) \sim \frac{\alpha_o^2}{n\,\Delta}\,\delta^2 \quad , \quad V(\hat{\delta}) \sim \frac{(\alpha_o - 1)^2\,\Gamma\,(1 - 2/\alpha_o)}{n\,\Delta}\,\delta^2$$

$$C(\hat{\lambda}, \hat{\delta}) \sim -\frac{\alpha_o(\alpha_o - 1)\,\Gamma\,(1 - 1/\alpha_o)}{n\,\Delta}\,\delta^2$$

$$\Delta = \alpha_o^2\,(\alpha_o - 1)^2\,[\,\Gamma\,(1 - 2/\alpha_o) - \Gamma^2\,(1 - 1/\alpha_o)\,] \ .$$

The predictor of the quantile $\lambda + \mu_{\alpha_o,m}\,\delta$ to be used must be $\hat{\lambda} + \mu_{\alpha_o,m}\,\hat{\delta}$ with the mean-square error

$$D^2_{n,m}(\hat{\lambda} + \mu_{\alpha_o}\,,_m\,\hat{\delta}) \sim \frac{\alpha_o^2 - 2\alpha_o(\alpha_o - 1)\,\Gamma(1-1/\alpha_o)\,\mu_{\alpha_o,m} + (\alpha_o - 1)^2\,\Gamma(1-2/\alpha_o)\,\mu^2_{\alpha_o,m}}{n\alpha_o^2\,(\alpha_o - 1)^2\,[\Gamma(1-2/\alpha_o) - \Gamma^2(1-1/\alpha_o)]} \ .$$

This is exactly equal to the asymptotic value of B'^2 and so the efficiency is 1. This technique requires the use of computers for the determination of $\hat{\lambda}$ and $\hat{\delta}$. Since, however, it is desirable to have simpler and more practical techniques when computer facilities are not available or, also, to initiate iteration, we consider in the next section some all-range predictors in this category. Such predictors can be quite useful since the range $\alpha_o \leq 2$ may be important for practical applications.

Some all-range predictors

The difficulties found above in respect to computability and efficiency of the proposed predictors can suggest the consideration of the traditional approach, using quantiles and moments, since their variances are of order n^{-1}.

For example, consider the quantile $q(x_i)$ for the probability $\omega = W_{\alpha_o}(\mu_{\alpha_o,m})$, whose asymptotic efficiency for the bound $\bar{B}^2$ is given as

$$\frac{\Gamma^2 (1 + 1/\alpha_o)}{(e^{\Gamma^{\alpha_o}(1+1/\alpha_o)/m} - 1)\ m^2}$$

which is approximated by $\Gamma^{2-\alpha_o}(1 + 1/\alpha_o)/m$ for large m. For $m = 1$ (one-step predictor) we have the following table of efficiencies :

α_o	0.25	0.5	1	2
eff (α_o)	0.60	0.64	0.58	0.52 .

This short table shows that the sample quantile for probabilities $W_{\alpha_o}(\mu_{\alpha_o,m})$, whose efficiency fades out with m, is not useful in the range $\alpha_o \leq 2$.

For $\alpha_o > 2$ the efficiency is given, using the bound B'^2, by

$$\frac{\mu_{\alpha_o,m}^{2(\alpha_o - 1)} [\alpha_o^2 - 2\alpha_o(\alpha_o - 1)\ \Gamma(1-1/\alpha_o)\mu_{\alpha_o,m} + (\alpha_o - 1)^2 [\ \Gamma(1-2/\alpha_o)\mu_{\alpha_o,m}^2]}{(e^{\mu_{\alpha_o,m}^{\alpha_o}} - 1)\ (\alpha_o - 1)^2\ [\ \Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o)]} .$$

The short table of efficiencies for $m = 1$ is :

α_o	2.5	3.0	3.5	4.0	4.5	5.0
eff (α_o)	.53	.55	.57	.58	.59	.60 ,

which shows that the quantile predictor has an almost constant efficiency of about 60%.

Consider now the use of a predictor based on the method of moments $\bar{x} + a(\alpha_o)\ s$, where s denotes the standard deviation of the first sample. The mean-square error for reduced values is

$$D^2_{n,m} = M[(\bar{x} + a(\alpha_o)\ s - \mu_{\alpha_o,\,m})^2] = M[(\bar{x} - \mu_{\alpha_o,\,m})^2] +$$

$$2M[(\bar{x} - \mu_{\alpha_o}, m)\, s]\, a(\alpha_o) + M(s^2)\, a^2(\alpha_o) .$$

The covariance between $\bar{x}$ and s is asymptotic to $[\beta_1(\alpha_o)\, b(\alpha_o)]/2n$ by the δ-method, where $\beta_1(\alpha_o)$ is the skewness coefficient.

Let $\sigma_n(\alpha_o)$ be the mean value of s. We have

$$\sigma_n(\alpha_o) - \sqrt{b(\alpha_o)} \sim -\frac{(\beta_2(\alpha_o) + 3)\sqrt{b(\alpha_o)}}{8n}$$

where $\beta_2(\alpha_o)$ is the kurtosis coefficient (*) and

$$D^2_{n,m} = \frac{b(\alpha_o)}{n} + (\mu_{\alpha_o,1} - \mu_{\alpha_o,m})^2 + 2[C(\bar{x}, s) + (\mu_{\alpha_o,1} - \mu_{\alpha_o,m})\sigma_n(\alpha_o)]\, a(\alpha_o) + b(\alpha_o)\, a^2(\alpha_o) .$$

The value at which the quadratic attains the minimum, as

$$C(\bar{x}, s) \sim \frac{\beta_1(\alpha_o)\, b(\alpha_o)}{2n} ,$$

converges to

$$a(\alpha_o) = -\frac{\mu_{\alpha_o,1} - \mu_{\alpha_o,m}}{\sqrt{b(\alpha_o)}}$$

so that the asymptotic value of the mean-square error is

$$D^2_{n,m} = \frac{b(\alpha_o)}{n}\left\{1 - \frac{\beta_1(\alpha_o)(\mu_{\alpha_o,1} - \mu_{\alpha_o,m})}{\sqrt{b(\alpha_o)}} + \frac{(\beta_2(\alpha_o) + 3)(\mu_{\alpha_o,1} - \mu_{\alpha_o,m})^2}{4\, b(\alpha_o)}\right\}$$

(*) A short proof of the asymptotic value of σ_n is given. From Cramér (1946), p. 353, we have the following relation,

$$M(s - \sigma_n(\alpha_o))^2 = \frac{(\beta_2(\alpha_o) - 1)\, b(\alpha_o)}{4n} + O(n^{-2}) .$$

Developing the square and taking mean values, we have

$$\frac{n-1}{n}\, b(\alpha_o) - \sigma^2_n(\alpha_o) = \frac{(\beta_2(\alpha_o) - 1)\, b(\alpha_o)}{4n} + O(n^{-2}) .$$

and so

$$\sigma^2\, n(\alpha_o) - b(\alpha_o) = -\frac{\beta_2(\alpha_o) + 3}{4n}\, b(\alpha_o) + O(n^{-2})$$

from which the desired result follows.

where

$$\beta_1(\alpha_o) = [\Gamma(1+3/\alpha_o) - 3\,\Gamma(1+2/\alpha_o)\,\Gamma(1+1/\alpha_o) + 2\,\Gamma^3(1+1/\alpha_o)] \,/\, b^{3/2}(\alpha_o)$$

and

$$\beta_2(\alpha_o) = (\Gamma(1+4/\alpha_o) - 4\,\Gamma(1+3/\alpha_o)\,\Gamma(1+1/\alpha_o) + 6\,\Gamma(1+2/\alpha_o)\,\Gamma^2(1+1/\alpha_o) - 3\Gamma^4(1+1/\alpha_o)]/b^2(\alpha_o)$$

Then the predictor is

$$\bar{x} - \left(\frac{\mu_{\alpha_o,1} - \mu_{\alpha_o,m}}{b(\alpha_o)^{1/2}}\right) s \ .$$

The asymptotic efficiency with respect to B^2 is

$$\frac{\mu^2_{\alpha_o,m}}{\alpha_o^2\, b(\alpha_o)\left\{1 - \dfrac{\beta_1(\alpha_o)\,(\mu_{\alpha_o,1} - \mu_{\alpha_o,m})}{\sqrt{b(\alpha_o)}} + \dfrac{(\beta_2(\alpha_o)+3)\,(\mu_{\alpha_o,1} - \mu_{\alpha_o,n})^2}{4\,b(\alpha_o)}\right\}}$$

and the asymptotic efficiency for $\alpha_o > 2$, with respect to B'^2, is given by

$$\frac{\alpha_o^2 - 2\alpha_o(\alpha_o - 1)\,\Gamma(1 - 1/\alpha_o)\,\mu_{\alpha_o,m} + (\alpha_o - 1)^2\,\Gamma(1 - 2/\alpha_o)\,\mu^2_{\alpha_o,m}}{\alpha_o^2\,(\alpha_o - 1)^2\, b(\alpha_o)\left[\Gamma\left(1 - \frac{2}{\alpha_o}\right) - \Gamma^2\left(1 - \frac{1}{\alpha_o}\right)\right]\left\{1 - \dfrac{\beta_1(\alpha_o)\,(\mu_{\alpha_o,1} - \mu_{\alpha_o,m})}{\sqrt{b(\alpha_o)}} + \dfrac{(\beta_2(\alpha_o)+3)\,(\mu_{\alpha_o,1} - \mu_{\alpha_o,m})^2}{4b(\alpha_o)}\right\}} \ .$$

A short table for the first efficiency for $m = 1$ is :

α_o	0.5	1.0	2
eff (α_o)	0.80	1.00	0.91 .

An example

Consider the natural and simple predictor $l(x_i) + m^{-1/\alpha_o}(\bar{x} - l(x_i))$.

We will experiment with the predictor generating a sample of n Weibull observations, computing the predictor, generating a new sample of m under the same conditions and comparing the minimum of the second sample with the value of the predictor.

The experiments are given in Table I for

$\alpha_o = 1.2$:	n = 50,	m = 20, 40
$\alpha_o = 1.2$:	n = 100,	m = 50
$\alpha_o = 1.2$:	n = 500,	m = 50, 100, 200, 300, 400
$\alpha_o = 1.5$:	n = 50,	m = 20, 40
$\alpha_o = 1.5$:	n = 100,	m = 50
$\alpha_o = 1.5$:	n = 500,	m = 50, 100, 200, 300, 400
$\alpha_o = 2.0$:	n = 50,	m = 20, 40
$\alpha_o = 2.0$:	n = 100,	m = 50
$\alpha_o = 2.0$:	n = 500,	m = 100, 200, 300, 400 .

The values l_1', l_2', ... , denote the observed smallest values in the second sample of m (generated observations).

TABLE I

α_o	n	m	x_n	x_m	l	l'_1	l'_2	l'_3	l'_4	l'_5	l'_6	l'_7
1.2	50	20	1.106482	0.618368	0.087202	0.171168	0.013614	0.067171	0.090015	0.170916		
	50	40	0.999691	0.864131	0.037770	0.082242	0.012288	0.026883	0.027323	0.057124		
	100	50	0.993607	0.870766	0.058743	0.094630	0.009625	0.075735	0.089076			
	500	50	0.955362	1.053420	0.017990	0.053974	$\geq 1'$					
	500	100	0.986448	0.811142	0.005847	0.026973	0.009693	0.021219				
	500	200	0.948326	0.995920	0.009625	0.020975	$\geq 1'$					
	500	300	0.995920	0.995920	0.009625	0.016349	$\geq 1'$					
	500	400	0.938055	0.927183	0.000687	0.007048	0.004685					
1.5	50	20	0.802585	1.157086	0.041248	0.097344	$\geq 1'$					
	50	40	0.864939	0.919777	0.091047	0.157814	0.020942					
	100	50	0.952770	0.885156	0.080250	0.144538	0.041757					
	500	50	0.936661	0.987915	0.032320	0.098952	0.083899					
	500	100	0.902707	0.901279	0.024205	0.065282	0.044599	0.059561				
	500	200	0.931957	0.893814	0.009254	0.036226	0.010214					
	500	300	0.906875	0.894213	0.003972	0.024120	$\geq 1'$					
	500	400	0.948098	0.958196	0.004700	0.022078	0.001939	0.018320				
2.0	50	20	0.936665	0.870119	0.254734	0.407218	0.138128	0.288591	0.386233			
	50	40	0.818688	0.914252	0.089966	0.205187	0.018976	0.090137	0.102450			
	100	50	0.822177	0.981068	0.100348	0.202430	$\geq 1'$					
	500	100	0.895189	0.837045	0.056219	0.140116	0.021859	0.023695	0.056533	0.063959	0.099370	0.118113
	500	200	0.918345	0.862408	0.076011	0.135573	0.010172	0.045231	0.054819	0.064132	0.093357	0.097550
	500	300	0.901596	0.894982	0.066431	0.114653	0.065947	0.076226	0.085539	0.095738	0.099727	
	500	400	0.897788	0.884023	0.047098	0.089632	0.031186	0.046646	0.055452	0.076625	0.086074	

The table as well a great part of this section was taken from Tiago de Oliveira and Littauer (1976).

Downpassing probability of level

The technique that follows is the one used for Fréchet distribution. In what follows we will suppose that $\alpha = \alpha_o$ (> 2) is known and, as said before, we have

$$V(\hat{\lambda}) \sim \frac{\alpha_o^2}{n\Delta}\,\delta^2 \quad , \quad V(\hat{\delta}) \sim \frac{(\alpha_o - 1)^2\,\Gamma\,(1 - 2/\alpha_o)}{n\,\Delta}\,\delta^2$$

and

$$C(\hat{\lambda}, \hat{\delta}) \sim - \frac{\alpha_o\,(\alpha_o - 1)\;\Gamma\,(1 - 1/\alpha_o)}{n\,\Delta}\,\delta^2$$

where $\Delta = \alpha_o^2 (\alpha_o - 1)^2 \ (\Gamma(1 - 2(\alpha_o) - \Gamma^2 (1 - 1/\alpha_o))$; $\alpha_o > 2$ guarantees the existence and regularity of the maximum likelihood estimators.

Now, we will be interested in the downpassing a of level a . $P(a) = P(a \,|\lambda,\delta) = W_{\alpha_o}(\frac{a-\lambda}{\delta})$ is estimated by $\hat{P}(a) = P(a \,|\hat{\lambda}, \hat{\delta})$.

As before, $\sqrt{n}\ (\hat{P}(a) - P(a))$ is asymptotically equivalent to

$$\sqrt{n}\ \{(\hat{\lambda} - \lambda)\frac{\partial P(a)}{\partial \lambda} + (\hat{\delta} - \delta)\frac{\partial P(a)}{\partial \delta}\}$$

$$= \sqrt{n}\ (1 - P(a)) \log (1 - P(a))\ \alpha_o \{\frac{\hat{\lambda} - \lambda}{a - \lambda} + \frac{\hat{\delta} - \delta}{\delta}\}$$

whose asymptotic variance is

$$V(a) = (1 - P(a))^2 (\log (1 - P(a)))^2 \{1 + \frac{\alpha_o^2}{\Delta}\ (\frac{\alpha_o - \delta}{a - \lambda} - (\alpha_o - 1)\ \Gamma(1 - 1/\alpha_o)^2)$$

estimated as $\hat{V}(a)$, substituting (λ, δ) by $(\hat{\lambda}, \hat{\delta})$.

As $\sqrt{n}\ (\frac{\hat{P}(a) - P(a)}{\sqrt{\hat{V}(a)}})$ is asymptotically standard normal, the one-sided asymptotic confidence interval with significance level α' is

$$0 \le P(a) \le \hat{P}(a) + \frac{\lambda_{\alpha'} \sqrt{\hat{V}(a)}}{\sqrt{n}}$$

as before.

In the case where the location parameter is known, taken as zero for convenience, we have

$$P(a) = P(a \,|\delta,\alpha) = W_\alpha(a/\delta)$$

and the variance-covariance matrix of $(\hat{\delta}, \hat{\alpha})$ is

$$\frac{1}{n}\begin{bmatrix} \frac{\alpha^2}{\delta^2} & -\frac{1-\gamma}{\delta} \\ -\frac{1-\gamma}{\delta} & \frac{1}{\alpha^2}(\frac{\pi^2}{6} + (1-\gamma)^2) \end{bmatrix}^{-1} = \frac{1}{n}\begin{bmatrix} \frac{\delta^2}{\alpha^2}\ (1 + \frac{6(1-\gamma)^2}{\pi^2}) & -\frac{6(1-\gamma)}{\pi^2}\delta \\ -\frac{6(1-\gamma)}{\pi^2}\delta & \frac{6}{\pi^2}\alpha^2 \end{bmatrix}$$

The results is analogous to the one for Fréchet distribution (for maxima) with the exchange of $\frac{1-\gamma}{\delta}$ for $-\frac{1-\gamma}{\delta}$ in the both matrices.

Then, evidently $\sqrt{n}\,(\hat{P}(a) - P(a))$ is asymptotically equivalent to

$$\sqrt{n}((\hat{\delta}-\delta)\frac{\partial P(a)}{\partial \delta} + (\hat{\alpha}-\alpha)\frac{\partial P(a)}{\partial \alpha}) = \sqrt{n}(1-P(a))\log(1-P(a))\,(\alpha\frac{\hat{\delta}-\delta}{\delta} - (\log\frac{a}{\delta})(\hat{\alpha}-\alpha))$$

with asymptotic zero mean value and variance

$$V(a) = (1 - P(a))^2\,(\log(1-P(a)))^2\,(1 + \frac{6}{\pi^2}\,(1-\gamma+\alpha\log\frac{a}{\delta})^2)$$

also analogous to the result for the Fréchet distribution. In the same way we see that:

The one-sided asymptotic confidence interval for P(a) is

$$0 \le P(a) \le \hat{P}(a) + \frac{\lambda_{\alpha'}}{\sqrt{n}}\sqrt{\hat{V}(a)}\,,$$

with significance level α' *, where* $\hat{V}(a)$ *, as before, is the expression of V(a) with* (δ, α) *substituted by* $(\hat{\delta}, \hat{\alpha})$.

For an application to earthquakes see Tiago de Oliveira (1984''') and references therein.

A worked example

Consider the data given in van Montfort and Otten (1978) of failure times corresponding to a complete sample of n = 24 observations, obtained in that order :

i	t_i
1	6.0
2	8.6
3	17.8
4	18.0
5	27.5
6	33.5
7	50.5
8	51.5

i	t_i
9	69.0
10	74.0
11	74.0
12	89.0
13	109.0
14	118.0
15	119.0
16	138.0

i	t_i
17	141.0
18	144.0
19	146.0
20	150.0
21	151.0
22	153.0
23	153.1
24	153.2

Supposing $\lambda = 0$, we obtain $\hat{\delta} = 101.023905$ and $\hat{\alpha} = 4.31899592$ (with 4 iterations!) from which the usual decision problems can be dealt with.

The Kolmogoroff -Smirnov statistic is $\hat{KS}_n = \sup_x | S_n(x) - \Phi^{\wedge}_{\alpha}(x / \hat{\delta}) | = .14182353$ and as $n = 24$ we have $\sqrt{n}\ \hat{KS}_n = .69479$, accepting thus the Weibull model for minima.

When this is the case — and it seems reasonable in the context of failure times — as $X = -\log T$ has a Gumbel maxima distribution with parameters $\lambda' = -\log \delta$ and $\delta' = 1/\alpha$ and, also, $X = 1/T$ has a Fréchet maxima distribution with parameters $\lambda'' = 0$, $\delta'' = \delta^{-1}$ and $\alpha'' = \alpha$, as seen before, we can use these transformations.

Chapter VIII

ANALITIC STATISTICAL CHOICE FOR UNIVARIATE EXTREMES

Introduction

As said before, the three limiting distribution of univariate maxima, under reduced form $\Psi_\alpha(z)$ (Weibull distribution), $\Lambda(z)$ (Gumbel distribution), and $\Phi_\alpha(z)$ (Fréchet distribution), can be imbedded in the general von Mises-Jenkinson formula

$$G(z|\theta) = \exp\left(-(1+\theta z)_+^{-1/\theta}\right), \quad -\infty < \theta < +\infty .$$

It is immediate that for $\theta < 0$ we have $\Psi_{-1/\theta}(z) = G\left(-\frac{1+z}{\theta} \mid \theta\right)$, a Weibull distribution with $\alpha = -1/\theta$, and for $\theta > 0$ we have $\Phi_{1/\theta}(z) = G\left(\frac{z-1}{\theta} \mid \theta\right)$, a Fréchet distribution with $\alpha = 1/\theta$; evidently, for $\theta = 0$, we obtain a Gumbel distribution $\Lambda(z)$ as $G(z|0^+) = G(z|0^-) = \Lambda(z)$.

We will describe first a method of statistical choice between the three models ($\theta < 0$, $\theta = 0$ or $\theta > 0$) which is asymptotically optimal and leads to a statistic that gives asymptotically a locally most powerful unbiased test of $\theta = 0$ vs. $\theta \neq 0$.

We will begin by giving a locally optimal approach, as in Tiago de Oliveira (1981) and (1982), enlarged to a more general set-up in Tiago de Oliveira (1983) and then a variational approach, the two ways being coincident in the statistical technique used, which illuminates the problem from two points of view; application to data comes from Tiago de Oliveira (1981) and Fransén and Tiago de Oliveira (1984).

We will begin with the analysis for samples of reduced values and then extend it to the general case.

For other details see Tiago de Oliveira (1981), (1982), (1984) and (1986) and Fransén and Tiago de Oliveira (1984).

Afterwards, we will sketch other analyses of the problem by van Montfort (1973), van Montfort and Otten (1978), Galambos (1980), Gomes (1982) and Pickands (1975); the use of probability paper and of Q_n was detailed in the Chapter IV, "Quick Exploration of Extreme Data", in this book.

The locally optimal approach

Consider a sample of (independent) reduced values $(z_1, \ldots, z_n)$ with distribution function $G(z|\theta)$, from which we want to decide for the Weibull model $(\theta < 0)$, the Gumbel model $(\theta = 0)$, or the Fréchet model $(\theta > 0)$.

Let $g(z|\theta) = G'(z|\theta)$ denote the probability density,

$$L_n(z|\theta) = L(z_1, \ldots, z_n|\theta) = \prod_1^n g(z_i|\theta)$$

the likelihood of the sample, and $v(z) = \dfrac{\partial \log g(z|\theta)}{\partial \theta}\Big|_{\theta=0} = \dfrac{z^2}{2} - z - \dfrac{z^2}{2} e^{-z}$.

The locally optimal test of Gumbel vs. Fréchet models $(\theta = 0$ vs. $\theta > 0)$ is given by the acceptance region $V_n(z) = \sum_1^n v(z_i) < a_n'$, the locally optimal test of Gumbel vs. Weibull models $(\theta = 0$ vs. $\theta < 0)$ is given by the acceptance region $V_n(z) = \sum_1^n v(z_i) > b_n'$, and the asymptotically unbiased locally optimal test of Gumbel vs. Fréchet or Weibull models $(\theta = 0$ vs. $\theta \neq 0)$ is given by the acceptance region $b_n < V_n(z) < a_n$; see Tiago de Oliveira (1981) for the proof of the asymptotic unbiasedness using only $V_n(z)$, where it is proven that the part containing the second derivatives, coming from the Neyman-Pearson theorem, converges to zero: the two-sided test is thus an intersection of the two one-sided tests. Thus:

The procedure is to decide for the Weibull model if $V_n(z) < b_n$, *for the Gumbel model if* $b_n \leq V_n(z) \leq a_n$ *and for the Fréchet model if* $a_n < V_n(z)$ *and we should have Prob {accept* $\theta = 0 | \theta = 0\} \rightarrow 1 - \alpha$ *and*

$$\frac{d\ Prob\ \{accept\ \theta = 0\ |\theta\ \}}{d\ \theta}\Big/_{\theta=0} \rightarrow 0\ .$$

The first attempt — see Tiago de Oliveira (1981) and (1984) — was to seek quantities $b_n < a_n$ that asymptotically split, in a balanced way, the significance level α, i.e., such that

$$\text{Prob}\{V_n(z) \leq b_n\} \to \alpha/2,$$

$$\text{Prob}\{V_n(z) \geq a_n\} \to \alpha/2,$$

and

$$\text{Prob}\{b_n < V_n(z) < a_n\} \to 1 - \alpha .$$

As, for $\theta = 0$, $v(z)$ has mean value zero and variance $\sigma^2 = \Gamma^{(4)}(1)/4 + \Gamma^{(3)}(1) + \Gamma''(1) = 2.42361$, we know that $V_n(z)/\sqrt{n}\,\sigma$ is asymptotically standard normal.

Putting $c_n(\alpha) = \sqrt{n}\,\sigma\,\lambda(\alpha/2)$, where $\lambda(p)$ is the solution of the equation $N(x) = 1 - p$, $N(.)$ being the standard normal distribution function, we see immediately that $a_n \sim c_n(\alpha)$ and $b_n \sim -c_n(\alpha)$. The approximation $-b_n = a_n = c_n(\alpha)$ was used in the first papers on the subject and applications, such as those referred to before, and Fransén and Tiago de Oliveira (1984) was also written under this pattern. We will show that a considerable improvement can be made.

Let us denote by $\mu(\theta)$ and $\sigma^2(\theta)$ the mean value and the variance of $v(z)$ with respect to the distribution function $G(z|\theta)$; $\mu(0) = 0$ and $\sigma^2(0) = \sigma^2$. It can be shown that $\mu(\theta)$ exists only when $-1 < \theta < 1/2$ and the variance exists only when $-1 < \theta < 1/4$. As $\mu'(0) = \sigma^2(>0)$ we see that $\mu(\theta)$ increases in the neighbourhood of $\theta = 0$. We have $\mu(\theta) = 2.42361\,\theta + 4.15362\,\theta^2 + O(\theta^3)$ and $\sigma^2(\theta) = 2.42361 + 25.26470\,\theta + 275.03597\,\theta^2 + O(\theta^3)$.

For $-1 < \theta < 1/4$ the Central Limit Theorem is valid and, thus, $(V_n(z) - n\mu(\theta))/\sqrt{n}\,\sigma(\theta)$ is asymptotically standard normal.

Under this formulation, and denoting by $\tilde{P}_n(\theta)$ the normal approximation based on the Central Limit Theorem, we obtain that the probabilities of correct decision are:

—for the Weibull model :

$$P_n(\theta) = \text{Prob}\{V_n(z) \leq -c_n(\alpha)\,|\theta\} =$$

$$\text{Prob}\left\{\frac{V_n(z) - n\mu(\theta)}{\sqrt{n}\,\sigma(\theta)} \leq \frac{-c_n(\alpha) - n\mu(\theta)}{\sqrt{n}\,\sigma(\theta)} \,/\, \theta\right\}$$

if $-1 < \theta < 0$, is approached asymptotically by

$$\tilde{P}_n(\theta) = N\left(\frac{-c_n(\alpha) - n\mu(\theta)}{\sqrt{n}\,\sigma(\theta)}\right),$$

which, as $\mu(\theta) < 0$ for small values of $\theta < 0$, leads to $\tilde{P}_n(\theta) \to 1$;

—for the Gumbel model :

$$P_n(0) = Prob\{-c_n(\alpha) < V_n(z) < c_n(\alpha) \mid \theta = 0\} =$$

$$= Prob\{\frac{c_n(\alpha)}{\sqrt{n}\,\sigma} < \frac{V_n(z)}{\sqrt{n}\,\sigma} < \frac{c_n(\alpha)}{\sqrt{n}\,\sigma} \mid \theta = 0\},$$

is approached asymptotically by

$$\tilde{P}_n(0) = N(\lambda(\alpha/2)) - N(-\lambda(\alpha/2)) = 1 - \alpha;$$

—for the Fréchet model :

$$P_n(\theta) = Prob\{V_n(z) \geq c_n(\alpha) \mid \theta\} =$$

$$Prob\{\frac{V_n(z) - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)} \geq \frac{c_n(\alpha) - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)} / \theta\}$$

if $0 < \theta < 1/4$ *, is approached asymptotically by*

$$\tilde{P}_n(\theta) = 1 - N(\frac{c_n(\alpha) - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)} \mid \theta),$$

which, as $\mu(\theta) > 0$ *for small values of* $\theta > 0$ *, leads to* $\tilde{P}_n(\theta) \to 1$ *;*

But there is a remark to be made. As

$$\text{Prob}\{\text{accept } \theta = 0 \mid \theta\} = \text{Prob}\{-c_n(\alpha) < V_n(z) < c_n(\alpha) \mid \theta\} \approx \tilde{P}_n(\theta)$$

$$= N(\frac{c_n(\alpha) - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)}) - N(\frac{-c_n(\alpha) - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)}),$$

we have

$$\tilde{P}_n(0) = N(\lambda(\alpha/2)) - N(-\lambda(\alpha/2)) = 1 - \alpha;$$

but

$$\frac{d\,\tilde{P}_n(\theta)}{d\,\theta}\Big|_{\theta=0} = -2\,\lambda(\alpha/2)\,N'(\lambda(\alpha/2))\,\frac{\sigma'(0)}{\sigma(0)} < 0$$

for the usual values of α and thus the symmetric test chosen is not asymptotically unbiased.

This statistical decision procedure is thus *consistent* in the same way as tests are said to be consistent, the Gumbel model $(\theta = 0)$ having a central position here. Recall that the procedure is unbiased and locally optimal at $\theta = 0$.

It can be changed to be *strongly consistent* , i.e., to be such that the probability of correct decision converges to 1, *whatever* θ may be, for $-1 < \theta < 1/4$. But in this approach we do not have asymptotic unbiasedness. We will summarize it.

Consider then as sequence $\{d_n\}$ $(d_n > 0)$ *and use the statistical decision procedure:*

decide for the Weibull model $(\theta < 0)$ *if* $V_n(z) \leq -\sqrt{n}\ \sigma d_n$,

decide for the Gumbel model $(\theta = 0)$ *if* $|V_n(z)| < \sqrt{n}\ \sigma d_n$,

and *decide for the Fréchet model* $(\theta > 0)$ *if* $V_n(z) \geq \sqrt{n}\ \sigma d_n$.

A necessary and sufficient condition for strong consistency is that $d_n \to +\infty$ *but* $d_n / \sqrt{n} \to 0$.

Evidently the conditions $d_n \to \infty$ and $d_n/\sqrt{n}$ avoid the "over-rejection" of the Gumbel model of the previous procedure (with $c_n(\alpha)$ as shown by the strong consistency.

Optimization of $\{d_n\}$, whatever it may mean, has not been solved.

A way to obtain $\{d_n\}$ is, after choosing a (moving) level of significance $2\alpha_n (0 < \alpha_n < 1/2$, $\alpha_n \to 0)$, to define d_n by

$$\text{Prob}\ \{|V_n(z)| < \sqrt{n}\ \sigma d_n \mid \theta = 0\} \approx N(d_n) - N(-d_n) = 1 - 2\alpha_n$$

or, asymptotically, as is known,

$$d_n = \sqrt{-2 \log \alpha_n} - \frac{\log(-\log \alpha_n) + \log(4\pi)}{2\sqrt{-2 \log \alpha_n}} .$$

As $\alpha_n \to 0$ we see that $d_n \to +\infty$; the condition $d_n/\sqrt{n} \to 0$ leads to $(\log \alpha_n)/n \to 0$. A simple solution is to take $\alpha_n = 1/n$, thus leading to a probability of incorrect rejection of $\theta = 0$ of order $2/n$; for details see Tiago de Oliveira (1981) and (1984).

But let us return to the initial question seeking asymptotic unbiasedness.

Let us study the decision rule : choose $b_n < a_n$ and

decide for the Weibull distribution if $V_n \leq b_n$,

decide for the Gumbel distribution if $b_n < V_n < a_n$,

and decide for the Fréchet distribution if $a_n \leq V_n$.

We will seek the best (b_n, a_n) for unbiasedness and some useful asymptotic approximation, as done in the previous step.

We will study the rule centering it on the Gumbel distribution, which is the pivot of the system. We have

$$P_n(k) = \text{Prob}\{\text{accept Gumbel df} \mid \theta\} = \text{Prob}\{b_n < V_n < a_n \mid \theta\}$$

$$= \text{Prob}\left\{\frac{b_n - n\,\mu(\theta)}{\sqrt{n}\ \sigma(\theta)} < \frac{V_n - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)} < \frac{a_n - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)}\right\}$$

if $\mu(\theta)$ and $\sigma(\theta)$ exist, and by the Central Limit Theorem we have

$$P_n(\theta) \approx \tilde{P}_n(\theta) = N\left(\frac{a_n - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)}\right) - N\left(\frac{b_n - n\,\mu(\theta)}{\sqrt{n}\,\sigma(\theta)}\right).$$

For simplicity of notation we will denote by $\mu_o = 0$, $\mu'_o = \sigma_o^2$, μ''_o, ... and σ_o, σ'_o σ''_o, ... the values of $\mu(0)$, $\mu'(0)$, $\mu''(0)$, ... and $\sigma(0)$, $\sigma'(0)$, $\sigma''(0)$, ...

Let us introduce the convenient notations $p = \sigma'_o / \sigma_o^2 = 3.34804$, $\bar{b}_n = b_n/\sqrt{n}\ \sigma_o$ and $\bar{a}_n = a_n/\sqrt{n}\ \sigma_o$.

By successive derivation we get

$$\tilde{P}_n(0) = N(\bar{a}_n) - N(\bar{b}_n),$$

$$\tilde{P}'_n(0) = -\sqrt{n}\,\sigma_o\,\{N'(\bar{a}_n)(1 + p\,\bar{a}_n/\sqrt{n}) - N'(\bar{b}_n)(1 + p\,\bar{b}_n/\sqrt{n})\},$$

and

$$\tilde{P}''_n(0) = n\,\sigma_o^2\,\{N''(\bar{a}_n)(1 + p\,\bar{a}_n/\sqrt{n})^2 - N''(\bar{b}_n)(1 + p\,\bar{b}_n/\sqrt{n})^2\}$$

$$-(\sqrt{n}/\sigma_o)\{N'(\bar{a}_n)(\mu''_o + \sigma''_o\,p\,\bar{a}_n/\sqrt{n} - 2\,\sigma_o^3(1 + p\,\bar{a}_n/\sqrt{n}))$$

$$- N'(\bar{b}_n)(\mu''_o + \sigma''_o\,p\,\bar{b}_n/\sqrt{n} - 2\,\sigma_o^3(1 + p\,\bar{b}_n/\sqrt{n}))\}$$

which will be needed to obtain the asymptotically optimal solution and a sequence of approximations.

For the limit solution, with significance level α, we have

$\tilde{P}_n(0) = N(\bar{a}_n) - N(\bar{b}_n) = 1 - \alpha,$

$\tilde{P}'_n(0) = -\sqrt{n}\,\sigma_o\,[N'(\bar{a}_n)(1 + p\,\bar{a}_n/\sqrt{n}) - N'(\bar{b}_n)(1 + p\,\bar{b}_n/\sqrt{n})] = 0$ which gives *the basic equations for the problem*

$$N(\bar{a}_n) - N(\bar{b}_n) = 1 - \alpha$$

$$N'(\bar{a}_n)(1 + p\,\bar{a}_n/\sqrt{n}) - N'(\bar{b}_n)(1 + p\,\bar{b}_n/\sqrt{n})] = 0 \ .$$

They must be solved numerically and do not lead to a symmetrical solution. We will get, first, a limiting solution and then a sequence of approximations that give a good simple solution.

Letting $n \to \infty$ in the basic equations, with $\bar{b}_n \to \bar{b}$ and $\bar{a}_n \to \bar{a}$, we get

$$N(\bar{a}) - N(\bar{b}) = 1 - \alpha$$

$$N'(\bar{a}) = N'(\bar{b})$$

which lead to $\bar{b} = -\lambda$, $\bar{a} = \lambda$, with $\lambda = \lambda(\alpha/2)$ being defined by the equation $N(\lambda) = 1 - \alpha/2$.

A remark can now be made : as $\bar{a}_n \to \lambda$, $\bar{b}_n \to -\lambda$ we see that

$$\tilde{P}''_n(0)/n\,\sigma_o^2 \to 2\,N''(\lambda) \qquad \text{or} \qquad \tilde{P}''_n(0) \sim -2\,\sigma_o^2\,\lambda\,N'(\lambda) \quad n < 0,$$

showing that we have a maximum of $\tilde{P}_n(\theta)$ at $\theta = 0$. But the decision procedure with the limiting solution $\bar{b} = -\lambda$, $\bar{a} = \lambda$, although verifying the condition $\tilde{P}_n(0) = 1 - \alpha$, is slightly biased because $\tilde{P}'_n(0) = -2\,\sigma_o\,p\,\lambda\,N'(\lambda) < 0$.

We will define approximations to $\bar{b}_n$ and $\bar{a}_n$ developing them in powers of $1/\sqrt{n}$ as

$$\bar{b}_{nj} = -\lambda + \bar{b}'_1 / n^{-1/2} + \bar{b}'_2 /n + \ldots + \bar{b}'_j\, n^{-j/2}$$

$$\bar{a}_{nj} = \lambda + \bar{a}'_1 / n^{-1/2} + \bar{a}'_2 /n + \ldots + \bar{a}'_j\, n^{-j/2} ;$$

obviously we have

$$\bar{b}_n = \bar{b}_{nj} + O\,(n^{-(j+1)/2}) ,$$

$$\bar{a}_n = \bar{a}_{nj} + O\,(n^{-(j+1)/2}) ,$$

$$\bar{b}_{n,j+1} = \bar{b}_{n,j} + \bar{b}'_{j+1} / n^{(j+1)/2} ,$$

and

$$\bar{a}_{n,j+1} = \bar{a}_{n,j} + \bar{a}'_{j+1} / n^{(j+1)/2} .$$

For convenience we will introduce the functions :

$$\bar{b}_j(\tau) = -\lambda + \sum_1^j \bar{b}'_r\, \tau^r$$

$$\bar{a}_j(\tau) = \lambda + \sum_1^j \bar{a}'_r\, \tau^r ;$$

we evidently have $\bar{b}_{nj} = \bar{b}_j\,(1/\sqrt{n})$ and $\bar{a}_{nj} = \bar{a}_j\,(1/\sqrt{n})$.

$\bar{b}_{nj}$ and $\bar{a}_{nj}$ will be the j-th approximations to $\bar{b}_n$ and $\bar{a}_n$; the 0-th approximation is evidently $\bar{b} = -\lambda$ and $\bar{a} = \lambda$. We will denote by $\tilde{P}_{nj}(0)$ and $\tilde{P}'_{nj}(0)$ the values of $\tilde{P}_n(0)$ and $\tilde{P}'_n(0)$ when we use the j-th approximation (in the initial notation, when we take $\bar{b}_n = \bar{b}_{nj}$ and $\bar{a}_n = \bar{a}_{nj}$) .

The deviation and the slope, at origin ($\theta = 0$), of the decision procedure are

$$\tilde{\Delta}_{nj} = |\tilde{P}_{nj}(0) - (1-\alpha)| \text{ and } \tilde{P}'_{nj}(0) .$$

As we have seen for $j = 0$ we have $\tilde{\Delta}_{no} = 0$ and $\tilde{P}'_{n}(0) = -2\,\sigma_o\, p\,\lambda\, N'(\lambda)$.

For each j , accepting the previous $\bar{b}_{ni}$ and $\bar{a}_{ni}$ $(i < j)$ we could develop in series of $1/\sqrt{n}$ the terms of the two basic equations, put equal to zero the first non-null term of the series in each equation, thus obtaining two linear equations in $\bar{b}'_j$ and $\bar{a}'_j$, and evaluate $\tilde{\Delta}_{nj}$ and $\tilde{P}_{nj}(0)$ by the asymptotics of the next non-zero term. We will use a systematic method. Denote by

$$f_j(\tau) = N(\bar{a}_j(\tau)) - N(\bar{b}_j(\tau)) - (1-\alpha)$$

$$g_j(\tau) = N'(\bar{a}_j(\tau))\,(1 + p\,\tau\,\bar{a}_j(\tau)) - N'(\bar{b}_j(\tau))\,(1 + p\,\tau\,\bar{b}_j(\tau)) ;$$

we have $\tilde{\Delta}_{nj} = |f_j(1/\sqrt{n})|$ and $P'_{nj}(0) = -\sqrt{n}\;\sigma_o\; g_j(1/\sqrt{n})$.

For $j = 1$ we have

$$f_1(0) = 0,\ f'_1(0) = (\bar{a}'_1 - \bar{b}'_1)\,N'(\lambda),\ \ldots$$

$$g_1(0) = 0,\ g'_1(0) = (\bar{a}'_1 + \bar{b}'_1)\,N''(\lambda) + 2\,p\,\lambda\,N'(\lambda) .$$

The equations defining $\bar{b}'_1$ and $\bar{a}'_1$ are then $f'_1(0) = 0$ and $g'_1(0) = 0$ and we get $\bar{a}'_1 = \bar{b}'_1 = p$ and so $\bar{b}_{n1} = -\lambda + p/\sqrt{n}$ and $\bar{a}_{n1} = \lambda + p/\sqrt{n}$. Then we get, using $\bar{a}'_1 = \bar{b}'_1 = p$,

$$f''_1(0) = 2\,p^2\,N''(\lambda) = -2\,p^2\,\lambda\,N'(\lambda) ,\quad g''_1(0) = 0 ,$$

$$g'''_1(0) = 4\,p^3\,(\lambda^3 - 2\lambda)\,N'(\lambda) \text{ and so}$$

$$\tilde{\Delta}_{n1} \sim \frac{|f''_1(0)|}{2}\,\left(\frac{1}{\sqrt{n}}\right)^2 = \lambda\,N'(\lambda)\,p^2/n$$

and $\tilde{P}'_{n1}(0) \sim -\sqrt{n}\ \sigma_o \frac{g_1'''(0)}{6}(\frac{1}{\sqrt{n}})^3 = -2(\lambda^3 - 2\lambda)\ N'(\lambda)\ p^3\ \sigma_o/3n$.

The development of $[N(\bar{a}_{n1}) - N(\bar{b}_{n1}) - (1-\alpha)]$ up to terms of second order in $1/\sqrt{n}$ shows that

$$\tilde{\Delta}_{n1} \le \frac{p^2}{n} \max_{\lambda} |N''(\lambda)| = \frac{p^2}{n} |N''(\pm 1)| = \frac{1}{\sqrt{2\ \pi\ e}} \cdot \frac{p^2}{n}$$

$$= 0.2419707\ \frac{p^2}{n} = 2.7123396/n.$$

Note that $\tilde{\Delta}_{n1}$ and $\tilde{P}_{n1}(0)$ are both $O(n^{-1})$. In the same way we can evaluate the next order approximations. For $j = 2$ we get through the analysis of $f_2(\tau)$ and $g_2(\tau)$, $\bar{b}'_2 = -p^2\ \lambda/2$ and $\bar{a}'_2 = p^2\ \lambda/2$ and then, as $f''_2(0) = 4\ p^2\ \lambda^2\ N'(\lambda)^2$ we get $\tilde{\Delta}_{n2} \sim 2\ p^2\ \lambda^2\ N'(\lambda)^2/n$ with a small reduction of the deviation measured by

$$\tilde{\Delta}_{n2}\ /\tilde{\Delta}_{n1} \sim 2\ \lambda\ N'(\lambda) \le 2\ \lambda\ N'(\lambda)|_{\lambda=1} = .48394$$

but with a deviation of the same order $O(n^{-1})$; and, as $g'''_2(0) = 2\ p^3\ (-\lambda^3 - 2\ \lambda^2 + \lambda + 4)\ N'(\lambda)$ and so

$$\tilde{P}''_{n2}(0) \sim -\sqrt{n}\ \sigma_o \frac{g'''(0)}{6}(\frac{1}{\sqrt{n}})^3 = \sigma_o \frac{p^3\ (\lambda^3 + 2\ \lambda^2 - \lambda - 4)\ N'(\lambda)}{3n}$$

and

$$\tilde{P}''_{n2}(0) / \tilde{P}'_{n1}(0) \sim \frac{-\lambda^3 - 2\ \lambda^2 + \lambda + 4}{2\ (\lambda^3 - 2\ \lambda)} \quad (\to -1/2 \text{ as } \lambda \to +\infty)$$

and so $\tilde{P}''_{n2}(0) = O(n^{-1})$ as $\tilde{P}'_{n1}(0)$; there is not a great improvement in the use of the second order approximation.

Consequently the procedure, asymptotically unbiased to order $0(n^{-1})$, is to take as bounds for $V_n(z)/n$ the values

$$b_n = \sigma_o \, \bar{b}_n / \sqrt{n} = \sigma_o \, (-\lambda + p/\sqrt{n} - p^2 \, \lambda/2n) / \sqrt{n}$$

$$a_n = \sigma_o \, \bar{a}_n / \sqrt{n} = \sigma_o \, (\lambda + p/\sqrt{n} + p^2 \, \lambda/2n) / \sqrt{n}$$

or to decide for the Gumbel distribution if

$$-\lambda + p/\sqrt{n} - p^2 \, \lambda/2n \leq \frac{V_n(z)}{\sqrt{n} \, \sigma_o} \leq \lambda + p/\sqrt{n} + p^2 \, \lambda/2n$$

and correspondingly for the Weibull or Fréchet distributions.

The acceptance interval for the Gumbel distribution (in terms of $V_n(z)/n$) thus increases from $a_0 - b_0 = 2\lambda$ to $a_2 - b_2 = 2\lambda\,(1 + \frac{p^2}{2n})$ from $O(1)$ to $O(n^{-1})$.

Let us now generalize this statistical choice procedure, when introducing location λ and dispersion $\delta(>0)$ parameters.

If the underlying distribution is the general Gumbel distribution $\Lambda((x - \lambda)/\delta)$, the maximum likelihood estimators $\hat{\lambda}$ and $\hat{\delta}$, for the sample $(x_1, \ldots, x_n)$, are given by the equations

$$\hat{\delta} = \bar{x} - \frac{\sum_1^n x_i \exp(-x_i / \hat{\delta})}{\sum_1^n \exp(-x_i / \hat{\delta})}$$

and

$$\hat{\lambda} = -\hat{\delta} \log\left(\frac{\sum_1^n \exp(-x_i / \hat{\delta})}{n}\right).$$

Now we will use "estimated" reduced values $\hat{z}_i = (x_i - \hat{\lambda})/\hat{\delta}$ and compute the "estimated $V_n(z)$" as $V_n(\hat{z})$, given by

$$\hat{V}_n(x) = \sum_1^n v((x_i - \hat{\lambda})/\hat{\delta}) = \frac{1}{2\,\hat{\delta}^2}\left\{\sum_1^n x_i^2 - 2\hat{\delta}\sum_1^n x_i - \frac{n\sum_1^n x_i^2 \exp(-x_i/\hat{\delta})}{\sum_1^n \exp(-x_i/\hat{\delta})}\right\}$$

$$= \frac{1}{2}\left\{\sum_1^n ((x_i - \bar{x})/\hat{\delta})^2 - n\,\frac{\sum_1^n ((x_i - \bar{x})/\hat{\delta})^2 \exp(-(x_i - \bar{x})/\hat{\delta})}{\sum_1^n \exp(-(x_i - \bar{x})/\hat{\delta})}\right\}.$$

Either by the δ-method, Tiago de Oliveira, (1982'), but with a lengthy algebra, or following Tiago de Oliveira (1981), it can be shown that, for the Gumbel model, $\hat{V}_n(x)/n$ is asymptotically normal with mean value zero and variance $\hat{\sigma}^2/n$ with $\hat{\sigma}^2 = 2.09797$; note that $\hat{\sigma}^2 < \sigma^2$, a reduction of 13.4 % . More generally $\hat{V}_n(x)/n$ is asymptotically normal with mean value $\mu(\theta)$ and variance $\hat{\sigma}^2(\theta)/n$, if the underlying distribution is $G(z|\theta)$; $\hat{\mu}(0) = 0$. Also $\hat{\mu}'(0) = \hat{\sigma}^2(0) = \hat{\sigma}^2$ and also $\hat{\mu}(\theta)$ is increasing in the neighbourhood of $\theta = 0$. The expansions of $\hat{\mu}(\theta)$ and $\hat{\sigma}(\theta)$ are $\hat{\mu}(\theta) = 2.09797\,\theta + 4.58653\,\theta^2 + O(\theta^3)$ and $\hat{\sigma}^2(\theta) = 2.09797 + 19.91015\,\theta + 248.60225\,\theta^2 + O(\theta^3)$.

Thus we can apply the reasoning used previously with the reduced sample and formulate a consistent or a strongly consistent decision procedure.

The strongly consistent procedure is to choose $\{d_n\}$ *such that* $d_n \to \infty$ *and* $d_n/\sqrt{n} \to 0$ *and decide for the Weibull model* $(\theta < 0)$ *if* $\hat{V}_n(x) \le -\sqrt{n}\,\hat{\sigma}\,d_n$ *, decide for the Gumbel model* $(\theta = 0)$ *if* $|\hat{V}_n(x)| < \sqrt{n}\,\hat{\sigma}\,d_n$ *and decide for the Fréchet model* $(\theta > 0)$ *if* $\hat{V}_n(x) \ge \sqrt{n}\,\hat{\sigma}\,d_n$.

Evidently for a (classical) consistent decision procedure we should use the asymptotically locally optimal approach given for the reduced case with the substitution of σ by $\hat{\sigma}$ and of p by $\hat{p} = \hat{\sigma}'/\sigma_0^2 = 3.27601$.

The method of analysis described above can be related to the likelihood ratio test of $\theta = 0$ vs. $\theta \neq 0$.

For more details and other information see Tiago de Oliveira (1981), although connected with the initial symmetrical approach.

The example associated with Table I and Fig. IV.4 in Chapter IV (with $\hat{\lambda} = 94.71$ and $\hat{\delta} = 12.49$) gives $\hat{V}_n / \sqrt{2.09797\, n} = -1.21$ leading, thus, to the acceptance of Gumbel model, as suggested by Fig. IV.4.

The variational approach

We begin by considering the reduced case. From a sample $(z_1, \ldots, z_n)$ we want to find a statistic $t_n(z) = t(z_1, \ldots, z_n)$ which approaches zero (0) in mean-square if $\theta = 0$ and deviates at the maximum positive speed if $\theta \neq 0$ in the neighbourhood of $\theta = 0$. The basic idea is to use $t_n(z)$ to decide for the Weibull model if $t_n(z) \leq B_n$, for the Gumbel model if $B_n < t_n(z) < A_n$ and for the Fréchet model if $A_n \leq t_n(z)$; naturally we can expect $B_n < 0 < A_n$.

Once more $L_n(z|\theta) = L_n(z_1, \ldots, z_n|\theta) = \prod_1^n g(z_i|\theta)$ denotes the likelihood of the sample; the mean value and variance of $t_n(z)$ for the value θ of the shape parameter are denoted by

$$\nu_n(\theta) = \int_{-\infty}^{+\infty} t_n(z)\, L_n(z|\theta)\, dz$$

and

$$\tau_n^2(\theta) = \int_{-\infty}^{+\infty} t_n^2(z)\, L_n(z|\theta)\, dz - \nu_n^2(\theta)$$

if they exist.

The conditions described above, with a scaling condition on the variance at $\theta = 0$, for convenience, are

$$\nu_n(0) = \int_{-\infty}^{+\infty} t_n(z)\, L_n(z|0)\, dz = 0$$

$$\tau_n^2(0) = \int_{-\infty}^{+\infty} t_n^2(z)\, L_n(z|0)\, dz = C \quad (\text{a positive constant})$$

and we want

$$\nu_n'(0) = \frac{d\,\nu_n(\theta)}{d\,\theta}\Big|_{\theta=0} = \int_{-\infty}^{+\infty} t_n(z) \sum_1^n v(z)\, L_n(z \mid 0)\; d\,z$$

to be a positive maximum ; as before we have

$$v(z) = \frac{\partial \log g(z \mid \theta)}{\partial\, \theta}\Big|_{\theta=0} = \frac{z^2}{2} - z - \frac{z^2}{2}\, e^{-z}.$$

The Lagrangean for the Calculus of Variations, with Lagrange multipliers A and B for the side conditions $\nu_n(0) = 0$ and $\tau_n^2(0) = C$, is

$$\bar{L}(t_n) = \{ t_n(z) + A\, t_n(z) + B\, t_n^2(z) \}\, L_n(z \mid 0).$$

The Euler-Lagrange equation for t_n is

$$\frac{\partial\, \bar{L}(t_n)}{\partial\, t_n} = 0$$

giving thus, as $L_n(z \mid 0) > 0$ for $-\infty < z < +\infty$,

$$\sum_1^n v(z_i) + A + 2\, B\, t_n(z) = 0.$$

Multiplying by $L_n(z \mid 0)$ and integrating the equation on z we get $A = 0$ and so $t_n(z)$ is proportional to $V_n(z) = \sum_1^n v(z_i)$.

Denoting, as before, by $\mu(\theta)$ and $\sigma^2(\theta)$ the mean value and the variance of $v(z)$ with respect to $g(z \mid \theta)$, we see that taking $t_n(z) = V_n(z)$ we have

$$\nu_n(\theta) = n\, \mu(\theta) \qquad (\nu_n(0) = 0)$$

and

$$\tau_n^2(\theta) = n\, \sigma^2(\theta) \qquad (\tau_n^2(0) = n\, \sigma^2)\ .$$

Then as $\nu_n'(0) = \tau_n^2(0) = n\sigma^2 > 0$ the variance increases with θ.

The variational approach is thus coincident with the locally optimal one and the analysis can proceed on the same lines as before.

The variational approach could be continued by seeking a statistic $t_n(x) = t(x_1, \ldots, x_n)$, independent of the location and dispersion parameters (i.e., $t(\alpha + \beta x_1, \ldots, \alpha + \beta x_n) = t(x_1, \ldots, x_n)$) with the property of quickest deviation from zero, as before, but the result does not seem workable, as happened before in the search of the best quasi-linear estimators and predictors in Chapters V, VI and VII.

Other analytic approaches

Let us describe, in a sketchy way, some approaches tried by Jenkinson (1955), van Montfort (1973), Pickands (1975), van Montfort and Otten (1978), Galambos (1980), Gomes (1982) and Tiago de Oliveira and Gomes (1984).

Jenkinson (1955) suggested the following decision statistic : for a sample of $2n$, we split it in n pairs whose maxima are taken and use the ratio of the variances of the $2n$ observations and the (produced) sample of n maxima of pairs. This statistic is, clearly, independent of the location and dispersion parameters and converges to a function of the shape parameter. A modification of this statistic was studied in Tiago de Oliveira and Gomes (1984).

van Monfort (1973) developed the following statistic W to test Gumbel vs. Fréchet distribution : define

$$W_n = \frac{1}{2} \log \frac{1 + r_n}{1 - r_n}$$

where r_n is the correlation coefficient between $\Psi_n(i + 1/2) = -\log(-\log((i + 1/2)/n))$ and $L_i = (x'_{i+1} - x'_i)/(\eta_{i+1} - \eta_i)$ $(i = 1,2, \ldots, n-1)$ where the x'_i are the order statistics of a sample $(x_1, \ldots, x_n)$ and η_i are the mean values of x'_i for the Gumbel distribution under reduced form $(\alpha = 0, \delta = 1)$. Evidently r, and thus W, are location and dispersion free. Simulation has shown that W is approximately normal with the mean value and variance to be fitted to the simulated critical points. The rejection region is $W > a$.

In the statistical choice proposed by van Montfort and Otten (1978) there is defined the statistic

$$A_n = \sqrt{n} \sum_{2}^{n} \frac{(\Delta_i - \bar{\Delta})}{\sigma_\Delta} W_i$$

where

$$W_i = \frac{L_i}{L_2 + \ldots + L_n}, \quad L_i = (x_i' - x_{i-1}') / (\eta_i - \eta_{i-1}), \quad x_i',$$

and η_i with the same meaning as in van Montfort (1973),

$$\Delta_i = -\gamma - (S_i^* - S_{i-1}^*) / (S_i - S_{i-1}) \quad (i = 2, \ldots, n), \quad \gamma = 0.57722 \quad \text{(the Euler constant)},$$

$$S_i = i\binom{n}{i} \sum_{j=0}^{n-i} (-1)^j \binom{n-j}{j} \frac{\log (i + j)}{i + j}, \quad S_i^* = i\binom{n}{i} \sum_{j=0}^{n-i} (-i)^j \binom{n-j}{j} \frac{\log^2(i+j)}{2(i + j)}, \quad \bar{\Delta} = \frac{1}{n - 1} \sum_{2}^{n} \Delta_i$$

and $\sigma_\Delta^2 = \frac{1}{n - 1} \sum_{2}^{n} (\Delta_i - \bar{\Delta})^2$.

The rule is to decide for $\theta > 0$ (Fréchet) if A_n is large, for $\theta < 0$ (Weibull) if it is small and for $\theta = 0$ (Gumbel) if it is intermediate. To this statistic A_n, a normal distribution is fitted giving conservative values, as shown by simulation.

Pickands (1975) gives a procedure to estimate the upper tail of a distribution, with an approach close to the one of Galambos (1980) testing procedure, as follows: for large a he shows that the conditional distribution $\frac{1 - F(x + a)}{1 - F(a)}$, if F is continuous, is close to $1 - L(x/a \mid c) = \exp\{-\int_0^{x/a} ((1 + ct)_+)^{-1} dt\}$ if F(.) is attracted to a limiting extreme value distribution, where $c < 0$, $c = 0$, $c > 0$ correspond to $\theta > 0$ (Fréchet), $\theta = 0$ (Gumbel) and $\theta < 0$ (Weibull) distributions. Then, letting $x_1'' \geq x_2'' \geq \ldots \geq x_k'' \geq \ldots \geq x_n''$ be the (descending) order statistics of a sample of n i.i.d. observations with distribution function F(x), and taking the upper 4M observations $(x_1'' \geq x_2'' \geq \ldots x_{4M}'')$, we use

$$\hat{c} = \log ((x_M'' - x_{2M}'') / (x_{2M}'' - x_{4M}'')) / \log 2$$

and

$$\hat{a} = (x''_{2M} - x''_{4M}) \Big/ \int_0^{\log 2} e^{\hat{c}\,u}\, du$$

as estimators of c and a if $M \xrightarrow{P} \infty$ but $M/n \xrightarrow{P} 0$. A choice of M is to take $d_M = \min_{1 \le j \le n/4} d_j$, with $d_j = \sup_{0 \le x < +\infty} | \hat{F}_j(x) - P_j(x \mid \hat{a}, \hat{c}) |$, $\hat{F}_j(x)$ being the usual empirical distribution function of $x''_1 - X''_{4i}$ ($i = 1,2, \ldots, 4j-1$), and $P_j(x \mid \hat{a}, \hat{c})$ is computed with $\hat{a}, \hat{c}$ corresponding to $M = j$. $1 - F(x)$ is then approximated by $1 - \bar{G}_M(x \mid \hat{a}, \hat{c})$.

The distribution functions P(.) are called *the generalized Pareto distributions*.

For more details connected to tail estimation see the Annex IV.

In Galambos (1980) another approach was presented, for the same type of decision for Gumbel vs. Fréchet models. It can be shown that the maximum of an i.i.d. sequence of random variables $\{X_j\}$, unbounded to the right and with mean value, is attracted to $\Lambda(.)$ if

$$\text{Prob}\{X > t + \mu(t)\, x\} \,/\, \text{Prob}\{(X > t)\} \to e^{-x} \text{ as } t \to +\infty,$$

where $\mu(t) = \int_t^{+\infty} (1 - F(y))\, dy \,/\, (1 - F(t))$ is the conditional mean of $X-t$ if $X > t$. Thus for large t, the $Y_j = (X_j - t)/\mu(t)$ are asymptotically standard exponential, and $\mu(t)$ can be estimated by $\mu^*(t)$, the average of the values $x_j - t > 0$. Then it is suggested to use for t the m-th maximum ($t = x'_{n+1-m} - x''_m$) and to proceed to the common tests for exponentiality; in the case of rejection of $\Lambda(.)$ we will accept the Fréchet model.

The Gumbel statistic Q_n for estimation of θ for the Fréchet distribution has already been used for quick exploration of extremes (Chapter IV). Gomes (1982) studied, by simulation, the behaviour of the statistic $Q_n = \dfrac{x'_n - x'_{[n/2]+1}}{x'_{[n/2]+1} - x'_1}$ where x'_n and x'_1 are obviously the maximum and the minimum of the sample and $x'_{[n/2]+1}$ the median of the sample. The asymptotic

distribution of this statistic was obtained in Tiago de Oliveira and Gomes (1984) and has different behaviours according to the exact value of θ : a Gumbel distribution of minima if $\theta < 0$, a Gumbel distribution of maxima if $\theta = 0$ and Fréchet distribution of maxima if $\theta > 0$, as said in chapter IV. But there does not exist an optimal procedure for statistical choice based on Q_n .

Chapter IX

STATISTICAL ANALYSIS WITH PARTIAL INFORMATION

Introduction

So far we have dealt with complete samples which were supposed to have Weibull, Gumbel or Fréchet distributions, or, otherwise, that one of them could be accepted as a good approximation to the distribution of the underlying population.

Now we will deal with some sparse results in which the samples are not complete, such as the following :

1 – We know only the upper (lower) part of a sample, i.e., the upper (lower) m observations, and we want to estimate the upper (lower) tail of the distribution; see Weismann (1978) and (1984) and references therein;

2 – We know that the sample of maxima can be split into m subsamples and we use only the k maxima of each subsample; see Hüsler and Tiago de Oliveira (1988) for $k = 1$;

3 – We use two or three quantiles to estimate the parameters; see Tiago de Oliveira (1972) and Kubat and Epstein (1980);

4 – We split the sample into disjoint parts, compute averages, and estimate the parameters; see Kubat (1982), Kubat and Epstein (1980), Hüsler and Schupbach (1986), and Neves (1988);

5 – We use excesses over threshold in the sequence of observations; see Smith (1984), Rösbjerg and Knudsen (1984).

The use of the extreme m observations

Let $x_1'' \geq x_2'' \geq \ldots \geq x_n''$ be the decreasing order statistics in a sample of size n from a distribution function F and suppose that $F \in \mathcal{D}(\tilde{L})$, i.e., that

$$F^n(\lambda_n + \delta_n x) \rightarrow \tilde{L}(x) \; (n \rightarrow \infty)$$

for some constants λ_n and $\delta_n > 0$. We are concerned here with drawing inferences about quantities associated with the right tail of the real F (.) when n is large, using asymptotic extreme value theory.

We are thus interested in

$\bar{w} = \sup \{x|F(x) < 1\}$, the right - end point;

$F(x)$ for large $x \; (< \bar{w})$;

$\chi_p = Q(p) = F^{-1}(p)$ with $p = 1 - c/n$ (c constant)

and, hence, in λ_n and $\delta_n > 0$.

Similarly we can be interested in the corresponding quantiles for the lower tail which can be dealt with by the usual transformation from maxima to minima.

The estimation method used up to now, and developed at length in Gumbel (1958), is to divide the original (unordered) data into subsamples. The maxima (or minima) of the subsamples are then used to estimate the parameters of $\tilde{L}$, one of the three extreme value limit distributions. This method is natural when some period exists in the data, as in environmental series, and the natural subsample is the data period. This method is sometimes called the *yearly data* or *(natural) subsamples method.*

But another approach is possible. Instead of splitting all observations into n/m samples of m observation each and taking the maximum of each sample, we can take the m largest observations of the sample of n and decide with them. Note that the underlying hypotheses are different : in the (natural) subsamples method we *only* suppose that maxima have the same extreme value distribution, while in the largest maxima method we suppose that *all* observations are i.i.d. with the distribution F(.), attracted by one of the extreme value distributions; this will be discussed in the Annex IV.

It will be seen in the next section that if the sample is i.i.d. with a Gumbel distribution then the subsample method is, in general, more efficient than the method of the largest m maxima using the same number of observations (m maxima of n/m subsamples or m largest maxima both of a sample of size n).

Let us then describe the Weissman method.

Define $x''_{ni} = (x''_i - \lambda_n)/\delta_n$. Then, as $F^n(\lambda_n + \delta_n x) \to \tilde{L}(x)$ as $n \to \infty$, the vector $(x''_{ni} : i = 1, ..., n)^T$ converges in distribution to $(M_i = \tilde{\lambda}^{-1}(Z_1 + ... + Z_i) : i = 1, ..., n)^T$, where $\tilde{\lambda}(x) = -\log \tilde{L}(x)$ and the $\{Z_i\}$ is a sequence of i.i.d. exponential random variables all with mean value 1 (David (1981), p. 266). Hence for large n, the m largest order statistics, up to a linear transformation, are distributed approximately as $(M_1, ..., M_m)$. There are only three possible $\tilde{\lambda}(x)$ to consider :

Gumbel : $\tilde{\lambda}(x) = e^{-x}$,

Fréchet : $\tilde{\lambda}(x) = x^{-\alpha}$ $(x > 0)$, $(\alpha > 0)$,

Weibull : $\tilde{\lambda}(x) = (-x)^{\alpha}$ $(x < 0)$, $(\alpha > 0)$.

Then for large values of y , if $F^n(\lambda_n + \delta_n y) \to \tilde{L}(y)$ as $n \to \infty$, we have

$$F(y) \approx \tilde{L}^{1/n}((y - \lambda_n)/\delta_n) = e^{-\tilde{\lambda}((y - \lambda_n)/\delta_n)/n} ;$$

estimation of $F(y)$ is thus obtained by substitution of $(\hat{\lambda}_n, \hat{\delta}_n)$ for (λ_n , δ_n).

The problem is now to estimate (λ_n , δ_n) from the m largest observations.

In the first case(Gumbel distribution), $\tilde{\lambda}^{-1}(y) = -\log y$ and thus we have that

$$D_i = M_i - M_{i+1} \ , -\log \frac{Z_1 + ... + Z_i}{Z_1 + ... + Z_i + Z_{i+1}} \ , \text{ and } \frac{Z_i}{i}$$

have the same distribution.

The spacings D_i are thus independent exponential random variables with mean values $1/i$, independent of M_m $(i = 1, ... , m-1)$. In the other cases the ratios M_i / M_{i+1} are independent . Independent and exponential spacings among order statistics are a well known characterization of the exponential distribution. The latter is an important member of the domain of attraction of the Gumbel distribution.

One can plot x''_i vs. $M\{M_i\}$ $(i = 1, ... , m)$ for each of the three models and decide which (if any !) fits the data, i.e., that they are approximately in a straight line. If m is at our disposal it

can be determined as the largest value for which $x_1'' > \ldots > x_m''$ fit the model. Once the model is chosen the parameters can be estimated. Maximum likelihood estimators and minimum variance estimators are discussed in Weissman (1978). For the Gumbel model the maximum likelihood estimators are

$$\hat{\delta}_n = \frac{1}{m}\sum_{i=1}^{m-1}(x_i'' - x_m'') = \frac{1}{m}\sum_{i=1}^{m} x_i'' - x_m'' = \bar{x}'' - x_m'' = \frac{1}{m}\sum_{i=1}^{m-1} i\,(x_i'' - x_{i+1}''))$$

$$\hat{\lambda}_n = x_m'' + \hat{\delta}_n \log m\,,$$

and the minimum variance estimators are

$$\delta_n^* = \frac{1}{m-1}\sum_{i=1}^{m-1}(x_i'' - x_m'') = \frac{m}{m-1}\,\hat{\delta}_n$$

$$\lambda_n^* = x_m'' + \delta_n^*(S_n - \gamma)$$

where $S_1 = 0$, $S_m = 1 + \frac{1}{2} + \ldots + \frac{1}{m-1}$ $(m > 1)$ and $\gamma = .57722$ is Euler's constant. Then $2(m-1)\,\delta_n^*/\delta_n$ is (asymptotically) a $\chi^2(2m-2)$ variate and thus confidence intervals for δ_n can be obtained. The distribution of $U_m = (x_m'' - \lambda_n)/\delta_n^*$ is parameter-free and thus confidence intervals for λ_n can be obtained provided the percentage points are available. The latter are tabulated in Weissman (1978).

For the Gumbel distribution we have, as seen in the attraction conditions Chapter,

$$(Q(1 - c/n) - \lambda_n)/\delta_n \to -\log c \qquad \text{as } n \to \infty\,.$$

and so

$$\hat{Q}(1 - c/n) = \hat{\lambda}_n + \hat{\delta}_n\,(-\log c)$$

is the approximate maximum likelihood estimator for $Q(1 - c/n)$.

Similar estimates are obtained in Weissman (1980) for threshold (type I) censoring. Particularly, only values larger than some a (fixed) are observable and then m is random. It turns out that a plays the role of x''_m in the formulae for $\hat{\lambda}_n$, $\hat{\delta}_n$ and $\hat{Q}$.

In principle, these asymptotic results hold for every distribution function $F(.) \in \mathcal{D}(\Lambda(.))$ (exponential, gamma, normal, lognormal, logistic, Weibull, to name some). But for finite n, better approximations can be obtained for distributions with exponential tail. Recently Boos (1983) used $\hat{Q}(1 - c/n)$ to estimate large quantiles for various distributions. He found that for some of them, as regards the mean-square error, this method is better for some distributions but worse for others; for more details see Weissman (1984).

For the Fréchet distribution the transformation $Y''_i = \log (X''_i - \lambda)$ transforms the model to the Gumbel one. For the Weibull distribution the transformation $Y''_i = - \log (\lambda - X''_i)$ has the same effect. But in both cases the location parameter, unknown in general, appears and has to be estimated. An example follows.

The three-parameter case for the Weibull distribution for minima is found in life testing, strength distributions, fatigue failures, etc., the available data being

$$x''_1 \le x''_2 \le \ldots \le x''_m$$

with $m << n$.

Thus we assume that $F(x) = 1 - \exp\{-((x - \lambda)/\delta)^\alpha)$ for $x > \lambda$, $F(x) = 0$ if $x \le \lambda$; for this distribution we have $\lambda_n = \lambda$ and $\delta_n = \delta\ n^{-1/\alpha}$. The maximum likelihood estimators for the three-parameters $(\lambda, \delta, \alpha)$ are

$$\hat{\delta} = \left(\frac{n}{m}\right)^{1/\hat{\alpha}} (x'_m - \hat{\lambda}) ,$$

$$\hat{\alpha} = m \Big/ \sum_{i=1}^{m} \log \frac{x'_m - \hat{\lambda}}{x'_i - \hat{\lambda}} \quad (> 0) , \quad \text{and}$$

$$\frac{1}{m}\sum_{i=1}^{m}\log\frac{x'_m-\hat{\lambda}}{x'_i-\hat{\lambda}}+\frac{m}{\sum_{i=1}^{m}(x'_m-\hat{\lambda})/(x'_i-\hat{\lambda})}=1.$$

The last equation shows that $1/\hat{\alpha}<1$ or $\hat{\alpha}>1$ and so the method does not work for $\alpha \leq 1$. We cannot use x'_1 as an estimator of λ because $x'_1 \xrightarrow{a.s.} \lambda$ and so the two other equations would give $\hat{\alpha}=0$ and $\hat{\delta}=+\infty$.

Maximum likelihood estimation is discussed by Hall (1982) and Smith and Weissman (1983).

Confidence intervals for λ are suggested by Weissman (1981, 1982). Using the convergence in distribution of $(M_1, M_2, \ldots, M_m)^T$ conveniently adapted for minima we see that

$$\frac{x'_1-\lambda}{x'_m-x'_1}$$

converge in distribution to a random variable whose distribution function is

$$1-(1-(\frac{x}{1+x})^{\alpha})^{m-1}$$

for $x \geq 0$ and 0 for $x \leq 0$.

If α is known then the quantile $\chi_{m,\alpha}(p)$ of this distribution can be obtained in a closed form. For large values of n, $(x'_1-\lambda)/(x'_m-\lambda)$ can be used as a pivotal function to form confidence intervals for λ. If α is unknown we can use the pivotal functions

$$\log\frac{x'_m-\lambda}{x'_1-\lambda}\Big/\sum_{i=1}^{m-1}\log\frac{x'_m-\lambda}{x'_1-\lambda}\quad(m\geq 3)$$

or

$$\log\frac{x'_i-\lambda}{x'_1-\lambda}\Big/\log\frac{x'_m-\lambda}{x'_1-\lambda}\quad(1<i<m\leq n)$$

to obtain confidence intervals for λ. Under Weibull distributions for minima both have limiting distributions which do not depend on any parameter. The quantiles of these distributions are

tabulated in Weissman (1981) and (1982). Simulation has shown good performance when $\alpha \leq 1$, which is useful since in this case ($\alpha < 1$) the maximum likelihood method fails.

Gomes (1981) investigated the maximum likelihood estimators of (λ, δ) using the m largest observations in each of n samples from $\Lambda(x)$ (Weissmann (1978) treated the case $n = 1$). That is, for $m = 2$, she considered the pairs $x''_1 \geq x''_2$. The results agree with those of Tiago de Oliveira (1972) and others when $m = 1$. She then investigated the properties of these estimators for $m = 1, 2, 3$ and $m = 5(5)30$, by a Monte Carlo simulation based on between 4500 and 10 000 replicates for each case, and found that $\hat{\lambda}$ is positively biased for $m = 1$ and negatively biased for $m > 1$ and $\hat{\delta}$ is always negatively biased. In addition, she gave a similar treatment for estimating (λ, δ) using moment estimators, simple best linear unbiased estimators and simple best linear invariant estimators. For $m = 3$, a simple linear estimator is one of the form

$$\sum_{i=1}^{m} \sum_{k=1}^{3} w_{kj} \, x''_{kj}$$

where x''_{kj} is the kth largest order statistics in the jth sample. The moment estimator λ^* is negatively biased for $m = 1$ and positively biased for $m > 1$ and δ^* is always positively biased. The bias properties of maximum likelihood and of moment estimators are the reverse of one another. The method of moments gave a slightly better estimator λ^* for small values of n, but yields very poor estimators δ^*.

Other references are Smith (1986) and Gomes (1984).

Finally some reference must be made to Hill's (1975) estimator for Fréchet distributions, $\Phi_\alpha(x/\delta)$ (i.e., $\lambda = 0$) and correspondingly to an analogous estimator for the Weibull distribution $W_\alpha(x/\delta)$ (also $\lambda = 0$).

For the Fréchet distribution $\Phi_\alpha(x/\delta)$ the transformed random variable $Y = \log X$ follows the Gumbel distribution $\Lambda(\frac{y - \log \delta}{1/\alpha})$. The m largest values x''_i (i = 1, 2, ... , m) correspond to the m largest values $y''_i = \log x''_i$. Using the ML estimators based on the largest m observations, of Weissman (1978), we have

$$1/\hat{\alpha} = \frac{1}{m} \sum_{1}^{m} y''_i - y''_m$$

$$\log \hat{\delta} = y''_m + (1/\hat{\alpha}) \log m$$

or equivalently

$$\hat{\alpha} = (\log \frac{(\prod_{1}^{m} x''_i)^{1/m}}{x''_m})^{-1} .$$

$$\hat{\delta} = m^{1/\hat{\alpha}} . x''_m .$$

$\hat{\alpha}$ is the well-known Hill estimator of the index of Fréchet distribution. See also Diaz (1985).

For the Weibull distribution $W_\alpha(x/\delta)$ the transformed random variable $Y = -\log X$ (a decreasing transformation) has the Gumbel distribution $\Lambda(\frac{y + \log \delta}{1/\alpha})$. Then the smallest m observations $x'_1 \leq x'_2 \leq \ldots \leq x'_m$ give rise to the m largest values $y''_i = -\log x'_i$. In the same way we have

$$(1/\hat{\alpha}) = \frac{1}{m} \sum_{1}^{m} y''_i - y''_m$$

$$(-\log \hat{\delta}) = y''_m + (1/\hat{\alpha}) \log m$$

or equivalently

$$\hat{\alpha} = [\log (x'_m / (\prod_{1}^{m} x''_i)^{1/m})]^{-1}$$

$$\hat{\delta} = m^{-1/\hat{\alpha}} x''_m .$$

This result can be useful for right-censored life tests (where naturally $\lambda = 0$).

Evidently quantile estimators and predictors can be made as usual.

For more detail connected to tail estimation see the Annex IV.

Largest observations vs. m subsamples maxima methods; a comparison

In this section, corresponding to item 2 of the Introduction, we will closely follow Hüsler and Tiago de Oliveira (1988). Suppose that we have a sample of $n = km$ observations and we can use only the largest m of all observations or the m maxima of natural blocks of data — in fact we are comparing what engineers call, respectively, the *largest observations method* with the *block (yearly) maxima method*. Let us denote the $n = km$ random variables by x_{ij}, $i = 1, \dots, k$ and $j = 1, \dots, m$, where m denotes the number of possible blocks (e.g. years) and k the number of observed random variables per block (year). Let $y_j = \max\{x_{ij}, 1 \le i \le k, j = 1, \dots, m\}$ and assume that all x_{ij} have the Gumbel distribution $\Lambda(\frac{x - \lambda}{\delta})$ and so the block (yearly) distribution of the y_j is $\Lambda(\frac{x - \lambda}{\delta} - \log k) = \Lambda(\frac{x - (\lambda + \delta \log k)}{\delta})$; $\lambda + \delta \log k$ is thus the block (year) location parameter (*).

We may use the m values y_j for estimating the values λ and δ, which is the classical method, described e.g. in Gumbel (1958). As said previously, Weissman (1978) proposed an estimation based on the m-largest observations of all $n = km$ values x_{ij}, which are

$$x_1'' > x_2'' > \dots > x_m''$$

with probability one.

At this point let us analyze the efficiency of Weissman's method with respect to the classical one (**).

(*) From the practical point of view, as in the previous section, we are assuming that the x_{ij} are i.i.d. and attracted to a Gumbel distribution and so their distribution can be approximated also by a Gumbel distribution. What is essential in the subsamples (or block) maxima method is that the maxima have a Gumbel distribution, even if the observations of the subsample are dependent.

(**) Notice that there are situations where only the maxima y_j per block (year) are recorded; in this case Weissman's method is not applicable, only the classical one.

We are now assuming, for the sake of comparison, that both methods can be used in a given application, i.e. that the x''_i , $i \leq m$, as well as y_j, $j = 1, \ldots, m$, can be observed and n is sufficiently large to the use of the Gumbel approximation. The comparison uses asymptotic results for both methods.

Both methods lose much information with respect to the (underlying) full sample of $n = km$; the efficiencies of both methods with respect to the full sample tend to zero as $k \to \infty$ since the procedures are based on very special subsamples.

Then the parameters $(\lambda + \delta \log k, \delta)$ of the y_j, $(j = 1, \ldots, m)$ have the usual maximum likelihood estimators $(\hat{\lambda} + \hat{\delta} \log k, \hat{\delta})$.

From the asymptotic ($m \to \infty$) variance-covariance matrix of $(\hat{\lambda} + \hat{\delta} \log k, \hat{\delta})$, given in Chapter V, we get for the asymptotic variance-covariance matrix $\hat{\Sigma}$ of $(\hat{\lambda}, \hat{\delta})$,

$$\hat{\Sigma} = \frac{\delta^2}{m} \begin{bmatrix} 1 + \frac{6}{\pi^2}(1 - \gamma - \log k)^2 & -6(1 - \gamma - \log k)/\pi^2 \\ -6(1 - \gamma - \log k)/\pi^2 & 6/\pi^2 \end{bmatrix}$$

and, for $k > 1$,

$$\rho(\hat{\lambda}, \hat{\delta}) = -\left(1 + \frac{\pi^2}{6(1 - \gamma - \log k)^2}\right)^{-1/2} \to -1, \text{ as } k \to +\infty ,$$

but slowly.

Consider now estimation based on the m largest values $x''_1 > x''_2 > \ldots > x''_m$ of the i.i.d. sample of $\{x_{ij}\}$ $(i = 1, \ldots, k;\ j = 1, \ldots, m)$.

For fixed k and m we get the usual maximum likelihood estimators (λ, δ) (with a slight change of notations) from the censored sample of m largest observations of a (possible) sample of $n = km$. We have, as $k \to +\infty$,

$$\lambda^* = x''_m - \delta^* \log k + O_p(k^{-1})$$

$$\delta^* = \bar{x}''_m - x''_m + O_p\left(\frac{\log k}{k}\right)$$

with $\bar{x}''_m = \frac{1}{m}\sum_{1}^{m} x''_i$.

The asymptotic variance-covariance matrix of (λ^*, δ^*) is

$$\sum{}^* \sim \delta^2 \begin{bmatrix} \sigma_m^2 + \frac{m-1}{m^2}(\log m)^2 & \frac{m-1}{m^2}\log m \\ \frac{m-1}{m^2}\log m & \frac{m-1}{m^2} \end{bmatrix}$$

with

$$\rho(\lambda^*, \delta^*) = (1 + m^2\,\sigma_m^2/(m-1)(\log m)^2)^{-1/2} \qquad \text{for } m \geq 2$$

where $\sigma_m^2 = \frac{\pi^2}{6} - S'_m$, with $S'_m = \sum_{1}^{m-1} j^{-2}$, $S'_m = 0$

Clearly $\rho(\lambda^*, \delta^*) \to 1$, as $k \to \infty$, also slowly. To compare these two procedures we will use the well-known Cramér efficiency. It is defined as

$$\text{eff}((\lambda^*, \delta^*)/(\hat{\lambda}, \hat{\delta})) = \det\left(\hat{\sum}\right) / \det\left(\sum{}^*\right)$$

$$= \frac{6}{\pi^2}\,\frac{1}{(m-1)\,\sigma_m^2} \downarrow \frac{6}{\pi^2} = .60793$$

as $m \to \infty$.

We could also study the efficiency as defined in Tiago de Oliveira (1982), defined as the worst possible for the estimation of all quantiles. It was shown that this efficiency is the smallest root of $\det\left(\hat{\sum} - \mu \sum{}^*\right) = 0$. But as $\mu_{max} \cdot \mu_{min} = \text{eff}((\lambda^*, \delta^*)/(\hat{\lambda}, \hat{\delta})) \approx \frac{\pi^2}{6}$, we see that μ_{min} is (approximately) smaller than $\sqrt{\pi^2/6} = .77970$ for large m . It was shown, in the

basic paper referred to, that $(\hat{\lambda}, \hat{\delta})$ — the yearly maxima method — is better than (λ^*, δ^*) — the largest maxima method — if the quantile corresponds to a probability larger than $p = .9$, for $m \geq 7$ using the asymptotic approximation, and in practice for every p if $m \geq 15$ which is the general case.

This shows a definite superiority of the yearly maxima method over some wide-spread thinking, when both methods can be applied. Evidently a symbiosis of both methods can be useful in some cases.

These results are based on the exact assumption that the $\{x_{ij}\}$ have the Gumbel distribution Λ. But these results can hold even with the assumption that the distribution $\{x_{ij}\}$ of belongs to the domain of attraction of Λ with a sufficiently fast rate of convergence of F^n to Λ. This will certainly be true if F is sufficiently close to Λ in an intuitive sense.

One might think, if the Gumbel distribution is valid, that the largest maxima method can always be applied for small $m's$. But the assumption of the independence of the x_{ij}'s is rarely satisfied in applications. The crucial point for this method is mainly the local dependence of the x_{ij}'s. Often the largest values occur in clusters, which prevents the application of the largest maxima method; but the yearly maxima method k can still be applied.

We can mention that these results may useful for the design of experiments: when only a small amount of information is available, i.e. if only m instead of $n = km$ observations can be recorded, a choice of design must be made, and if we are interested in large quantiles the block method should be used, chiefly because it guarantees the practical independence of the blocks.

An analysis similar to the one given here was made by Reiss (1987) for the Fréchet distribution, with special reference to the Hill estimator.

Estimation using two or three quantiles:

It is important, at this point, to recall that $\chi_p^* \xrightarrow{P} \chi_p$, $\sqrt{n}\ f(\chi_p) \dfrac{\chi_p' \chi_p}{\sqrt{p(1-p)}}$ is asymptotically standard normal if $0 < f(\chi_p) < \infty$, that (χ_p^*, χ_q^*), where $0 < p < q < 1$ conveniently reduced is an asymptotically binormal pair with standard margins and correlation coefficient $\rho = \sqrt{\dfrac{p}{q}\ \dfrac{1-q}{1-p}} > 0$. Its extension to more sample quantiles is immediate. See in Chapter I "A note on the asymptotic behaviour of sample quantiles"

Consider now the Gumbel distribution $\Lambda((x - \lambda)/\delta)$. Then as $\chi_p = \lambda + \delta(-\log(-\log p))$, the estimation equations are

$$Q_p(n) = \chi_p^* = \lambda^* - \delta^* \log(-\log p)$$

$$Q_q(n) = \chi_q^* = \lambda^* - \delta^* \log(-\log q),$$

and so the estimators are

$$\lambda^* = [\log(-\log p)\, Q_q - \log(-\log q)\, Q_p] / \log\left(\frac{\log p}{\log q}\right)$$

$$\delta^* = (Q_q - Q_p) / \log\left(\frac{\log p}{\log q}\right).$$

As the variance-covariance matrix of (Q_q, Q_p) is

$$V \sim \frac{1}{n}\begin{bmatrix} \dfrac{p(1-p)}{f^2(\chi_p)} & \dfrac{p(1-q)}{f(\chi_p)\, f(\chi_q)} \\ \dfrac{p(1-q)}{f(\chi_p)\, f(\chi_q)} & \dfrac{q(1-q)}{f^2(\chi_q)} \end{bmatrix} = \frac{\delta^2}{n}\begin{bmatrix} \dfrac{1-p}{p(\log p)^2} & \dfrac{1-q}{q \log p \log q} \\ \dfrac{1-q}{q \log p \log q} & \dfrac{1-p}{q(\log q)^2} \end{bmatrix}$$

with $\det(V) \sim \dfrac{\delta^4}{n^2} \dfrac{(1-q)(q-p)}{pq^2(\log p)^2 (\log q)^2}$, the variance-covariance matrix of (λ^*, δ^*) as

$$[Q_q, Q_p]^T = \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix} (\lambda^*, \delta^*)^T,$$

is

$$V^* = \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix}^{-1} V \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log p) \end{bmatrix}^{-1 T}$$

with $\det(V^*) = \det(V) / \det^2 \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix}$.

As the Cramér-Rao bound for the estimators of (λ, δ) — corresponding to the ML estimators $(\hat{\lambda}, \hat{\delta})$ — is

$$\hat{V} = \frac{\delta^2}{n} \begin{bmatrix} 1 + \frac{6(1-\gamma)^2}{\pi^2} & \frac{6(1-\gamma)}{\pi^2} \\ \frac{6(1-\gamma)}{\pi^2} & \frac{6}{\pi^2} \end{bmatrix}$$

the Cramér efficiency is

$$\det(\hat{V}) / \det(V^*) = \frac{6}{\pi^2} \frac{(\log(\frac{\log p}{\log q}))^2 \; p\, q^2 \, (\log p)^2 \, (\log q)^2}{(1-q)^2 \; (q-p)^2} .$$

The maximum efficiency is obtained for $p = .07$ and $q = .76$ and its value is 40.8%.

One could also compute the efficiency as defined in Tiago de Oliveira (1982), connected with the study of quantile estimators that follows.

The quantile estimator of the ξ-quantile χ_ξ is

$$\chi^*_\xi (p,q) = c_1 Q_p + c_2 Q_q$$

with $\quad c_1 + c_2 = 1$

and $\quad c_1 \chi_p + c_2 \chi_q = \chi_\xi$ or $(-\log p)^{c_1} (-\log q)^{1-c_1} = -\log \xi$

to be quasi-linearly invariant.

It is asymptotically normal with mean value χ_ξ and asymptotic variance

$$V(\chi^*_\xi(p, q)) \sim \frac{1}{n} \{c_1^2 \frac{p(1-p)}{f^2(\chi_p)} + 2 c_1 (1-c_1) \frac{p(1-q)}{f(\chi_p) f(\chi_q)} + (1-c_1)^2 \frac{q(1-q)}{f^2(\chi_q)} \} =$$

$$\frac{\delta^2}{n} \{c_1^2 \frac{1-p}{p(\log p)^2} + 2 c_1 (1-c_1) \frac{1-q}{q \log p \; \log q} + (1-c_1)^2 \frac{1-q}{q(\log q)^2} \} .$$

We can compute its asymptotic efficiency with respect to the Cramér-Rao bound, i.e., with respect to the ML estimator $\hat{\chi}_\xi = \hat{\lambda} + \xi \hat{\delta}$.

As $V(\hat{\chi}_\xi) \sim \frac{\delta^2}{n} \{1 + (1 - \gamma + \chi_\xi)^2\} = \frac{\delta^2}{n} \{1 + (1 - \gamma - \log(-\log \xi))^2\}$, the asymptotic efficiency is

$$\lim_{n \to \infty} \frac{V(\hat{\chi}_\xi)}{V(\chi^*_\xi (p, q))}$$

and it can be maximized in (p, q) ($0 < p < q < 1$).

Neves (1986) has made the calculations and the optimal results are as follows with p, q, and c_1 (recall that $c_2 = 1 - c_1$) as functions of ξ ; see table I.

Table I

ξ	p	q	c_1	eff(%)
0.01	0.066	0.933	1.143720	65
0.02	0.072	0.943	1.104300	65
0.03	0.008	0.051	0.339046	67
0.05	0.014	0.086	0.360599	75
0.10	0.030	0.180	0.412024	82
0.20	0.070	0.350	0.459707	82
0.30	0.130	0.510	0.524218	81
0.40	0.210	0.640	0.574626	82
0.50	0.310	0.740	0.613838	83
0.60	0.430	0.820	0.653152	84
0.70	0.560	0.880	0.678641	83
0.80	0.710	0.930	0.723894	79
0.85	0.777	0.948	0.716725	73
0.89	0.010	0.755	- 0.314795	68
0.90	0.010	0.759	- 0.341731	68
0.95	0.013	0.776	- 0.562649	68
0.99	0.021	0.793	- 1.115900	67

Notice that we do not always have $p < \xi < q$, as might be expected from a naive analysis. This is connected to the fact that the graph of efficiency has two relative maxima and the absolute maximum changes with ξ . For more details and the graphs see Neves (1986). The efficiency is thus reasonable for the usual quantiles.

Having thus sketched the way of analysing the use of quantiles, we will in the next cases/ distributions concentrate only on the efficiency for quantiles estimators.

The Fréchet distribution has the form $\Phi_\alpha((x-\lambda)/\delta) = 0$ if $x \le \lambda$, $\Phi_\alpha((x-\lambda)/\delta) = \exp\{-((x-\lambda)/\delta)^\alpha\}$ if $x \ge \lambda$. Supposing λ known ($\lambda = \lambda_o = 0$) for convenience we know that the increasing transformation $Y = \log X$ leads to a Gumbel distribution $\Lambda(\frac{y - \log \delta}{1/\alpha})$. The empirical quantiles of the $\{y_i\}$ are the corresponding ones of the $\{x_i\}$, and so the estimator of the ξ-quantile χ_p is given by $\log \chi^*_\xi = c_1 \log Q_p + (1 - c_1) \log Q_q$, i. e., $\chi^*_\xi = Q_p^{c_1} Q_q^{1-c_1}$; recall that this is in correspondence with the second condition which, here, takes the form $c_1 \log \chi_p + (1 - c_1) \log \chi_q = \log \chi_\xi$ or $\chi_\xi = \chi_p^{c_1} \chi_q^{1-c_1}$, defining $c_1(p, q, \xi)$.

χ^*_ξ is homogeneous, i.e., if the $\{x_i\}$ are multiplied by $c(>0)$ the same happens to χ^*_ξ; as the asymptotic variance of the χ^*_ξ is the same as for χ^*_ξ for the Gumbel distribution above, the δ-method shows that $\chi^*_\xi(p,q)$ is also asymptotically normal with mean value and asymptotic variance $\frac{\delta^2}{(-\log \xi)^{2/\alpha}\, \alpha^2 n} \{1 + (1 - \gamma - \log(-\log \xi))^2\}$; the efficiency is the same as before. Notice that the variance depends on δ and α and not only on α as could be expected ($1/\alpha$ is the dispersion parameter of the transformed variable y). This is, evidently, connected with the homogeneity but the not quasi-linearity of the estimator.

If for the Fréchet distribution we have (λ,δ) unknown, but $\alpha = \alpha_o$ is known, the situation is completely analogous to the location-dispersion situation above. The ξ-quantile is $\chi_\xi = \lambda + (-\log \xi)^{-1/\alpha_o} \cdot \delta$ and $c_1 = c_1(p, q, \xi)$ is defined by $c_1 \chi_p + (1 - c_1) \chi_q = \chi_\xi$ or $c_1(-\log p)^{-1/\alpha_o} + (1 - c_1)(-\log q)^{-1/\alpha_o} = (-\log \xi)^{-1/\alpha_o}$. The estimator χ^*_ξ of the ξ-quantile is

$$\chi^*_\xi = c_1 Q_p + (1 - c_1) Q_q$$

which is asymptotically normal with mean-value χ_ξ and asymptotic variance

$$V(\chi^*_\xi) \sim \frac{1}{n} \{c_1^2 \frac{p(1-p)}{f^2(\chi_p)} + 2 c_1 (1 - c_1) \frac{p(1-q)}{f(\chi_p)\, f(\chi_q)} + (1 - c_1)^2 \frac{q(1-q)}{f(\chi_q)^2}\}$$

$$= \frac{\delta^2}{\alpha_o^2\, n} \{c_1^2 \frac{1-p}{p(-\log p)^{2+2/\alpha_o}} + 2c_1(1-c_1) \frac{1-q}{q(\log p \,.\, \log q)^{1+1/\alpha_o}}$$

$$+ (1-c_1)^2 \frac{1-q}{q(-\log q)^{2+2/\alpha_o}} \},$$

and the ML estimator $\hat{\chi}_\xi = \hat{\lambda} + (-\log \xi)$, associated with the Cramér-Rao bound, has mean value χ_ξ and asymptotic variance

$$V(\hat{\chi}_\xi) \sim \frac{\delta^2}{n} \{ \frac{(-\log \xi)^{-2/\alpha_o}}{\alpha_o^2} + \frac{(1 - \Gamma(2 + 1/\alpha_o)\,(-\log \xi)^{-1/\alpha_o})^2}{(\alpha_o + 1)^2\,(\Gamma(1 + 2/\alpha_o) - \Gamma^2(1 + 1/\alpha_o))} \}.$$

The asymptotic efficiency can then be computed as before. For $\alpha_o = 3$ we have

Table II ($\alpha_o = 3$)

ξ	p	q	c_1	eff (%)
0.010	0.0399	0.8632	1.062470	64
0.015	0.0500	0.8500	1.064880	63
0.018	0.0045	0.0320	0.379776	69
0.020	0.0050	0.0356	0.362109	71
0.050	0.0132	0.0924	0.407834	80
0.100	0.0300	0.1890	0.465436	82
0.200	0.0770	0.3720	0.550790	82
0.300	0.1500	0.5180	0.613496	84
0.400	0.2450	0.6280	0.655708	86
0.500	0.3570	0.7150	0.688727	87
0.600	0.4780	0.7870	0.713217	84
0.700	0.6050	0.8490	0.732972	78
0.740	0.6500	0.9000	0.788330	73
0.750	0.0043	0.5952	- 0.399820	73
0.800	0.0050	0.6090	- 0.558671	73
0.900	0.0070	0.6320	- 1.155680	72
0.950	0.0082	0.6448	- 1.901690	72
0.990	0.0095	0.6591	- 4.454210	71

We can make comments similar to those for the Gumbel distribution.

Let us recall that, as seen before, $\Phi_\alpha(1 + \frac{x-\lambda}{\alpha\,\delta}) = \Phi_\alpha(\frac{x-(\lambda-\alpha\,\delta)}{\alpha\,\delta}) \to \Lambda(\frac{x-\lambda}{\delta})$ as $\alpha \to +\infty$. Then $\lambda - \alpha\,\delta$ takes the place of λ and $\alpha\,\delta$ that of δ in the previous formulation, and the ξ-quantile is $\tilde{\chi}_\xi = \lambda - \alpha(1 - (-\log\xi)^{-1/\alpha})\,\delta$ and for fixed α_o the ML estimator of $\tilde{\chi}_\xi$ is $\tilde{\tilde{\chi}}_\xi = \tilde{\lambda} - \alpha_o(1 - (-\log\xi)^{-1/\alpha_o})\,\tilde{\delta}$ where $\tilde{\lambda}$ and $\tilde{\delta}$ are the ML estimators of the *new* parameters (λ,δ). $\tilde{\tilde{\chi}}_\xi$ is also asymptotically normal with mean value $\tilde{\chi}_\xi$ and variance asymptotic to

$$\frac{\alpha_o^2\;\delta^2}{n}\left\{\frac{(-\log\xi)^{-2/\alpha_o}}{\alpha_o^2} + \frac{(1-\Gamma(2+1/\alpha_o)\,(-\log\xi)^{-1/\alpha_o})^2}{(\alpha_o+1)^2\;(\Gamma(1+2/\alpha_o) - \Gamma^2(1+1/\alpha_o))}\right\};$$

when $\alpha_o \to +\infty$ we see that $\tilde{\chi}_\xi \to \lambda - \log(-\log\xi)\,\delta$ and the variance asymptotic is to $\frac{\delta^2}{n}\{1 + \frac{6}{\pi^2}(1 - \gamma - \log(-\log\xi))^2\}$ which are the corresponding values for the Gumbel distribution $\Lambda((x-\lambda)/\delta)$, as could be expected from

$$\Phi_{\alpha_o}(1 + \frac{x-\lambda}{\alpha_o}) \to \Lambda(\frac{x-\lambda}{\delta}) \quad \text{as} \quad \alpha_o \to \infty.$$

Consider now the three-parameter case for the Fréchet distribution.

It is not possible to obtain linear combinations of three quantiles (Q_p, Q_q, Q_r), with constant coefficients depending on (p,q,r), to estimate parameters; they should always depend also on α being estimated, and this could be presumed because it appears in a weak form in the estimation of the ξ-quantile when $\lambda = \lambda_o$ is known. The parameter α can be estimated by the location-dispersion-free ratio.

$$\frac{Q_r - Q_q}{Q_q - Q_p} \xrightarrow{P} g(\alpha) = \frac{(-\log r)^{-1/\alpha} - (-\log q)^{-1/\alpha}}{(-\log q)^{-1/\alpha} - (-\log p)^{-1/\alpha}}.$$

Thus the estimator α^* can be given by $g(\alpha^*) = \frac{Q_r - Q_q}{Q_q - Q_p}$ which, by the δ-method, can be shown to be asymptotically normal with mean value α and variance $O(n^{-1})$; the triple $(\lambda^*, \delta^*, \alpha^*)$ is also asymptotically trinormal with mean value $(\lambda, \delta, \alpha)$ and variance-covariance matrix $O(n^{-1})$.

A simple solution is to take $\frac{\log p}{\log q} = c > 1$ and $\frac{\log r}{\log q} = c^{-1} < 1$; we get

$$\frac{Q_r - Q_q}{Q_q - Q_p} = c^{1/\alpha} \qquad \text{and} \qquad p = q^c \, , \, q = r^c \; ;$$

any choice of c gives a system for the estimation of α.

The best choice of (p, q, r) for each ξ, or even the choice of (p, q, r) that maximizes the asymptotic efficiency, has not yet been studied and should not be expected to be very efficient.

The method only seems useful, at present, to obtain the first estimates of $(\lambda, \delta, \alpha)$ from the equations

$$Q_p = \chi_p^* = \lambda^* + (-\log p)^{-1/\alpha^*} \, \delta^*$$

$$Q_q = \chi_q^* = \lambda^* + (-\log q)^{-1/\alpha^*} \, \delta^*$$

$$Q_r = \chi_q^* = \lambda^* + (-\log r)^{-1/\alpha^*} \, \delta^*$$

to be used to seed the solution of the ML equations and then to estimate the quantiles; using these estimators to estimate the quantiles their asymptotically normal behaviour can be obtained, through simple but lengthy computations, by the use of the δ-method.

Recall that for the Fréchet distribution we must always have $\lambda^* \leq X_1^{'}$, which imposes a new condition on the $(X_1^{'} , Q_p, Q_q)$ or $(X_1^{'} , Q_p, Q_q, Q_r)$ to allow the use of the method; note that $\text{Prob}\{\lambda^* \leq X_1^{'} \} \to 1$.

Consider finally the Weibull distribution (for minima) $W_\alpha((x - \lambda)/\delta)$ and suppose that $\lambda = \lambda_o$ is known ($\lambda_o = 0$ for convenience) . Then by the transformation $Y = -\log X$ we get a Gumbel random variable with parameters $(-\log \delta, 1/\alpha)$. As the transformation is decreasing we have $\chi_{1-\xi}(\text{Gumbel}) = -\log \chi_\xi$ (Weibull), and so the estimator of the ξ-quantile is

$$\chi_\xi^* = Q_p^{d_1} \; Q_q^{1-d_1}$$

where $d_1(p,q,\xi) = c_1(1-q, 1-p, 1-\xi)$; χ^*_ξ is also homogeneous; as the asymptotic variance of $-\log \chi^*_\xi$ is the same as that for $\chi^*_{1-\xi}$ in the Gumbel distribution, the δ-method shows that $\chi^*_\xi(p,q)$ is also asymptotically normal with mean value χ_ξ and asymptotic variance $V(\chi^*_\xi(p,q))$

$$\sim \frac{\delta^2}{(-\log(1-\xi))^{2/\alpha}\,\alpha^2\,n}\{1 + (1 - \gamma - \log(-\log(1-\xi)))^2\}$$, the efficiency being the same as before with the exchange of (ξ, p, q) by $(1-\xi, 1-p, 1-q)$.

The same could be obtained by noting that if X has the distribution $W_\alpha(x/\delta)$ then $1/X$ has the distribution $\Phi_\alpha(x/(1/\delta))$.

Suppose now that we have the Weibull distribution (for minima) with $\alpha = \alpha_o > 2$ known, i.e., the distribution is $W_{\alpha_o}((x-\lambda)/\delta)$ whose quantiles are $\chi_\xi = \lambda + (-\log(1-\xi))^{1/\alpha_o}\,\delta$. Consider the estimator $\chi^*_\xi = \chi^*_\xi(p, q) = c_1 Q_p + c_2 Q_q$ with $c_1 + c_2 = 1$ and $c_1(-\log(1-p))^{1/\alpha_o} + c_2(-\log(1-q))^{1/\alpha_o} = (-\log(1-\xi))^{1/\alpha_o}$. Evidently χ^*_ξ is asymptotically normal with mean value χ_ξ and asymptotic variance

$$V(\chi^*_\xi) \sim \frac{\delta^2}{\alpha_o^2\,n}\{c_1^2\,\frac{p}{(1-p)(-\log(1-p))^{2-2/\alpha_o}} + 2\,c_1(1-c_1)$$

$$\frac{p}{(1-p)\,(\log p\ \log q)^{1-1/\alpha_o}} + (1-c_1)^2\,\frac{q}{(1-q)(-\log(1-q))^{2-2/\alpha_o}}\}.$$

The ML estimator, for $\alpha_o > 2$, associated with the Cramér-Rao bound $\hat{\chi}_\xi = \hat{\lambda} - (-\log(1-\xi))^{1/\alpha_o}\,\hat{\delta}$ has the mean value χ_ξ and asymptotic variance

$$V(\hat{\chi}_\xi) \sim \frac{\delta^2}{n}\{\frac{(-\log(1-\xi))^{2/\alpha_o}}{\alpha_o^2} + \frac{(1 - \Gamma(2 - 1/\alpha_o)\,(-\log(1-\xi))^{-1/\alpha_o})^2}{(\alpha_o - 1)^2\ \{\Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o)\}}\}.$$

Note the similarity of this result (for $\alpha_o > 2$) with the corresponding one for the Fréchet distribution already given.

Let us obtain an asymptotic result analogous to the one connected with the convergence $W_\alpha(\lambda + \frac{x-\lambda}{\alpha\ \delta}) \to \Lambda\ (\frac{x-\lambda}{\delta})$.

In our case we have $1 - W_\alpha\ (1 - \frac{x-\lambda}{\delta}) = 1 - W_\alpha\ (-\frac{x-(\lambda+\alpha\,\delta)}{\alpha\ \delta}) \to \Lambda\ (\frac{x-\lambda}{\delta})$, $x \to \infty$, and the ξ-quantile is $\tilde{\chi}_\xi = \lambda + \alpha(1-(-\log \xi)^{1/\alpha})\ \delta$ corresponding to the substitution of λ by $\lambda + \alpha\,\delta$, of δ by $\alpha\,\delta$ and ξ by $1-\xi$, as we are, for W_α, dealing with minima and for Λ dealing with maxima.

$\tilde{\tilde{\chi}}_\xi = \tilde{\lambda} + \alpha_o(1 - (-\log \xi)^{1/\alpha_o})\ \tilde{\delta}$ is the ML estimator of $\tilde{\chi}_\xi$ for $\alpha_o > 2$, $(\tilde{\lambda}, \tilde{\delta})$ being the ML estimators of the new $(\lambda\ ,\ \delta)$.

Thus, as the transformations are linear, we know that $\tilde{\chi}_\xi$ is asymptotically normal with mean value $\tilde{\chi}_\xi$ and variance

$$V(\tilde{\tilde{\chi}}_\xi) \sim \frac{\alpha_o^2\ \delta^2}{n}\ \{\frac{(-\log \xi)^{2/\alpha_o}}{\alpha_o^{\,2}} + \frac{(1 - \Gamma(2 - 1/\alpha_o)\ (-\log \xi)^{1/\alpha_o})^2}{(\alpha_o - 1)^2\ \ (\Gamma(1 - 2/\alpha_o) - \Gamma^2(1 - 1/\alpha_o))}\ \}.$$

Letting $\alpha_o \to +\infty$, and corresponding to the fact that $1 - W_{\alpha_o}(1 - \frac{x-\lambda}{\alpha_o}) \to \Lambda\ (\frac{x-\lambda}{\delta})$, we see that $\tilde{\chi}_\xi \to \lambda - \log\ (-\log \xi))\ \delta$ and the variance is asymptotic to $\frac{\delta^2}{n}\ \{1 + \frac{6}{\pi^2}\ (1 - \gamma - \log\ (-\log \xi))^2\}$ which are the corresponding values for the Gumbel distribution, as happened for the Fréchet distribution before.

Consider now the three-parameter case for the Weibull distribution (of minima). The situation is completely analogous to the previous one for the Fréchet distribution (for minima) with the added difficulty of not having regular ML estimators if $\alpha \le 2$. The equations are analogous to the ones for the Fréchet distribution and quantile estimation is also analogous.

We must recall that any estimator λ^* should be such that $\lambda^* \le X_1^{'}$, which imposes conditions on $(X_1^{'}, Q_p, Q_q)$ or $(X_1^{'}, Q_p, Q_q, Q_r)$ to allow the use of the method; although we know that $\text{Prob}\{\lambda^* \le X_1^{'}\} \to 1$.

See also the exercises concerning the use of X_1' and of one or two empirical quantiles to estimate the parameters of the Fréchet distribution or the Weibull distribution (for minima) and their ξ-quantiles.

Estimation of the parameters using block partitions of the sample :

As a matter of convenience, suppose we have the i.i.d. sample $(x_1, \dots, x_n)$, the ordered sample $(x_1' \le x_2' \le \dots \le x_n')$, we choose probability levels (p, q) with $0 < p < q < 1$, and we split the sample into three blocks:

$$x_1' \le \dots \le x_{[np]}' \qquad \text{with average } \bar{x}_1 = \sum_{1}^{[np]} x_i' / [np] ,$$

$$x_{[np]+1}' \le \dots \le x_{[nq]}' \qquad \text{with average } \bar{x}_2 = \sum_{[np]+1}^{[nq]} x_i' / ([nq] - [np]) ,$$

$$x_{[nq]+1}' \le \dots \le x_n' \qquad \text{with average } \bar{x}_3 = \sum_{[np]+1}^{n} x_i' / (n - [nq]) ;$$

we are supposing n sufficiently large such that $1 < [np] < [nq] < n$ or $p > 1/n$ and $q - p > 1/n$.

As can be expected, the triple $(\bar{x}_1, \bar{x}_2, \bar{x}_n)$ has an asymptotic trinormal distribution. Their mean values are $(\lambda + \delta \mu_1 , \lambda + \delta \mu_2 , \lambda + \delta \mu_3)$ and the variance-covariance matrix is

$$\frac{\delta^2}{n} \Omega^{-1} \sim \frac{\delta^2}{n} \begin{bmatrix} \sigma_1^2/p & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2/(q-p) & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2/(1-q) \end{bmatrix}$$

where $F(\frac{x - \lambda}{\delta})$ is the form of the distribution function of the x_i . We have, for the mean values,

$$\mu_1 = \frac{1}{p} \int_0^p F^{-1}(w) \, d\, w \ ,$$

$$\mu_2 = \frac{1}{q-p} \int_p^q F^{-1}(w)\, d\, w, \qquad \text{and}$$

$$\mu_3 = \frac{1}{1-q} \int_q^1 F^{-1}(w)\ \ d\, w,$$

and for the terms of the variance-covariance matrix :

$$\sigma_1^2 = \frac{1}{p} \int_o^p (F^{-1}(w))^2\, d\, w\ - \mu_1^2 + (\chi_p - \mu_1)^2\, (1-p)\,,$$

$$\sigma_2^2 = \frac{1}{q-p} \int_p^q (F^{-1}(w))^2\, d\, w - \mu_2^2 + \frac{1}{q-p} \{p(1-p)\, (\mu_2 - F^{-1}(p))^2 +$$

$$+ q(1-q)\, (F^{-1}\,(q) - \mu_2)^2 + 2\, p(1-q)\, (\mu_2 - F^{-1}(p))\, (F^{-1}(q) - \mu_2)\}\,,$$

$$\sigma_3^2 = \frac{1}{1-q} \int_q^1 (F^{-1}(w))^2\, d\, w - \mu_3^2 + q\, (\mu_3 - F^{-1}(q))^2\,,$$

$$\sigma_{12} = (F^{-1}\,(p) - \mu_1)\, [\mu_2 + \frac{1-q}{1-p}\ F^{-1}(q) - \frac{1-p}{q-p}\ F^{-1}(p)]\,,$$

$$\sigma_{13} = (\mu_3 - F^{-1}(q)\, (F^{-1}(p) - \mu_1)\,, \text{ and}$$

$$\sigma_{23} = (\mu_3 - F^{-1}(q))\, (-\mu_2 + \frac{q}{q-p}\ F^{-1}(q) - \frac{p}{q-p}\ F^{-1}(p))\,.$$

Notice that these values are expressed in the reduced distribution function.

Our purpose is to determine, first, coefficients (l_1, l_2, l_3) and (d_1, d_2, d_3) such that

$$\lambda^* = \sum_1^3 l_i\ \bar{x}_i \qquad \text{and} \qquad \delta^* = \sum_1^3 d_i\ \bar{x}_i$$

are the least-squares estimators of (λ, δ). Denoting by $[\mu]^T = (\mu_1, \mu_2, \mu_3)$, $[1]^T = (1, 1, 1)$, $[\bar{x}]^T = (\bar{x}_1\ ,\ \bar{x}_2\ ,\ \bar{x}_3)$, and by Ω , as written before, the inverse of the variance-covariance

matrix, with $\Gamma = \Omega\,([1]\,[\mu]^T - [\mu]\,[1]^T)\,\Omega/\Delta$, where $\Delta = [1]^T\,\Omega\,([1]\,.\,[\mu]^T\,\Omega\,[\mu] - [1]^T\,\Omega\,[\mu]^2$, we have

$$\lambda^* = -[\mu]^T\ \Gamma\,[\bar{x}]$$

$$\delta^* = [1]^T\ \Gamma\,[\bar{x}]\ ,$$

the coefficients of the linear combinations being, in obvious notation,

$$[l\,]^T = -[\mu]^T\,\Gamma \qquad \text{and} \qquad [d]^T = [1]^T\,\Gamma\,.$$

The pair (λ^*, δ^*) is asymptotically binormal with mean values (λ, δ) and variance-covariance matrix

$$\frac{\delta^2}{n\Delta}\begin{bmatrix} [\mu]^T\,\Omega\,[\mu] & [1]^T\,\Omega\,[\mu] \\ [1]^T\,\Omega\,[\mu] & [1]^T\,\Omega\,[1] \end{bmatrix}$$

and the asymptotic efficiency relative to the Cramér-Rao bounds (corresponding to the ML estimators) is

$$\text{eff} = \Delta/\delta^4 \times \{M(\frac{\partial \log f}{\partial \lambda})^2\,.\,M(\frac{\partial \log f}{\partial \delta})^2 - M(\frac{\partial \log f}{\partial \lambda}\ \frac{\partial \log f}{\partial \delta})^2\}$$

where $f\,(\frac{x-\lambda}{\delta}) = \frac{d\,F\,((x-\lambda)/\delta\,)}{d\,x} = \frac{1}{\delta}\ F'\,(\frac{x-\lambda}{\delta})\,.$

Notice that the coefficients, when a shape parameter exists, should be written more correctly as $l_i(\alpha)$ and $d_i(\alpha)$.

The study of the maxima efficiency for the Fréchet distribution can be summarized in the following short table (the μ_i exist only for $\alpha_o > 2$) :

	max. eff (%)	p	q
$\alpha_o = 3$	84.3	.10	.82
$\alpha_o = 6$	88.7	.11	.72
$\alpha_o = 8$	90.9	.07	.45
$\alpha_o = 9$	91.9	.07	.44
$\alpha_o = 10$	92.6	.06	.41

Recall that the result for $\alpha_0 = 10$ is very similar to the one corresponding to Gumbel distribution, as known.

We can also, in the same way, study right, left and doubly censored samples. It corresponds, in the first and second case, to choosing three probabilities $0 < p < q < r < 1$ and discarding the right-block average $\bar{x}_4$ or the left-block average $\bar{x}_1$ (obvious notations), which is equivalent to taking $l_4 = d_4 = 0$ or $l_1 = d_1 = 0$. Clearly the triples $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ and $(\bar{x}_2, \bar{x}_3, \bar{x}_4)$ are asymptotically trinormal with a variance-covariance matrix which is the obvious sub-matrix of that for a 4-block partition, analogous to the one given. By the technique above we can obtain, when possible, the asymptotic relative efficiency of estimators (λ^*, δ^*) which are linear combinations of the averages, quasi-linearity invariant, i.e., for $\alpha_0 > 2$ in both the Fréchet and Weibull (for minima) cases and in Gumbel case, and optimize for $0 < p < q < r < 1$, and then compute the best coefficients that, as before, depend on α_0. Note that the best efficiency using one of the three averages of the simply censored sample can be better than the use of three averages in a non-censored sample; this seems a paradox but is easily explained. When using a simply censored sample we are partitioning it into four blocks and discarding the right or the left one : so comparison should be made with the efficiency of a 4-block partition (not a 3-block one !) and then the efficiency of the uncensored sample is better.

A doubly censored sample corresponds to choosing four probabilities $0 < p < q < r < s < 1$, and to discharging the extreme averages $\bar{x}_1$ and $\bar{x}_5$. The remaining triple $(\bar{x}_2, \bar{x}_3, \bar{x}_4)$ is asymptotically trinormal with a variance-covariance matrix which is the obvious sub-matrix of the one corresponding to the 5-block partition. The best choice of the coefficients of the linear combination can be made, as before, by seeking the probabilities that maximize efficiency and then computing the best coefficients, depending as always on α_0.

Let us, finally, consider the case when we have the three-parameters $(\lambda, \delta, \alpha > 2)$ unknown the Gumbel case is contained in the model for $\alpha = +\infty$ or, practically speaking, α large) and a non-censored sample with a 3-block partition.

We will not make all the computations but only sketch the method, along the lines of what was done previously.

Supposing α known, we obtained the best estimators of (λ, δ)

$$\lambda^* = \sum_1^3 l_i(\alpha)\, \bar{x}_i$$

$$\delta^* = \sum_1^3 d_i\ (\alpha)\ \bar{x}_i$$

in the beginning of the section.

The ratio statistic $\dfrac{\bar{x}_3 - \bar{x}_2}{\bar{x}_2 - \bar{x}_1} \rightarrow g(\alpha) = \dfrac{\mu_3 - \mu_2}{\mu_2 - \mu_1}$ is location-dispersion- free.

Then, by the δ-method, the ratio statistic is asymptotically normal with mean value $g(\alpha)$ and thus, by the same method, α^*, given by $g(\alpha^*) = \dfrac{\bar{x}_3 - \bar{x}_2}{\bar{x}_2 - \bar{x}_1}$, is asymptotically normal with variance $O(n^{-1})$. Also the triple $(\lambda^*, \delta^*, \alpha^*)$ is asymptotically normal with mean value $(\lambda, \delta, \alpha)$ and variance-covariance matrix $O(n^{-1})$. But, in fact, $\lambda^* = \sum_1^3 l_i\ (\alpha)\ \bar{x}_i$ and $\delta^* = \sum_1^3 d_i\ (\alpha)\ \bar{x}_i$, written above, depend on α and we should use $\lambda^{**} = \sum_1^3 l_i\ (\alpha^*)\ \bar{x}_i$, $\delta^{**} = \sum_1^3 d_i\ (\alpha^*)\ \bar{x}_i$, the triple $(\lambda^{**}, \delta^{**}, \alpha^*)$ being, also by the δ-method, asymptotically trinormal with mean value $(\lambda, \delta, \alpha)$ and variance-covariance matrix of $O(n^{-1})$. The new coefficients $l_i\ (\alpha^*)$, $d_i\ (\alpha^*)$ are not the best ($\alpha > 2$ supposed known), although they can be expected to be close to the best ones as α^* should be close to $(\alpha > 2)$. But the asymptotic efficiency of the method drops down to values around 0.1 and 0.2 and so the method is poor, as could be expected. We could seek functions $\bar{l}_i\ (\alpha)$, $\bar{d}_i\ (\alpha)$ such that $\sum_1^3 \bar{l}_i\ (\alpha^*)\ \bar{x}_i$, $\sum_1^3 \bar{d}_i\ (\alpha^*)\ \bar{x}_i$, $\alpha^*)$ have maximum asymptotic efficiency. This is an open and important problem which is very practical and allowing a quick decision; recall that as we want quasi-linear invariance we must have $\sum_1^3 \bar{l}_i\ (\alpha) = 1$ and $\sum_1^3 \bar{d}_i\ (\alpha) = 0$. The extension of the previous reasonings to censored samples is immediate as well as to the unsolved problem.

Estimation using exceedances over thresholds

This section contains two different approaches, the peaks-over-threshold (P.O.T.) method and the all-excesses-over-threshold (A.E.O.T.) method.

The P.O.T. method can be connected with the notion of the return period, already given, the papers by Gumbel (1955) on calculated risk, Tiago de Oliveira (1984''') on large earthquakes, and Rosbjerg and Knudsen (1984) on extreme sea states.

Let us fix a threshold u and consider in the sequence of random variables $\{X_i\}$ the *exceedances* over u, i.e., the random variables $X_j > u$. If the X_i has the distribution function $F(x)$, the exceedance $X_j\ (>u)$ has the distribution function $F(x|u) = \frac{F(x) - F(u)}{1 - F(u)}$ $(u \leq x \leq \bar{\omega}_F)$; we will suppose $F(x)$ continuous.

In Chapter II we considered a sequence of thresholds (there called levels) such that $n(1 - F(u_n)) \to \tau$ as $n \to +\infty$, and sketched a brief way of obtaining the asymptotic distribution of maxima. In the same way, see Leadbetter, Lindgren and Rootzén (1983), p.32, we can prove that if $n(1 - F(u_n)) \to \tau$ as $n \to +\infty$ then

$$\text{Prob}\ \{\#\ \{X_i > u_n\} \leq k\} \to e^{-\tau} \sum_{S=0}^{k} \frac{\tau^S}{S!},$$

showing asymptotically the Poissonian character of excesses of large thresholds in i.i.d. case. The exceedance may be considered a large storm, large earthquake, large flood, etc.

We will *assume* that the large exceedances over a level x are i.i.d. with a Poisson distribution with mean value $\beta(1 - F(x|u))$ per cycle (e.g., 1 year), and so the mean value of exceedances of u per cycle is $\nu = \beta$ as $F(u|u) = 0$ and in t cycles βt. Thus the level x_T such that the mean value of exceedances in T cycles is 1 is given by $1 = \beta T(1 - F(x_T|u))$ or $F(x_T|u) = 1 - 1/\beta T$; x_T is evidently the *design value* for a return period of T (for the excess variables).

Also

$$\text{Prob}\ \{\max\ X_i \text{ in } t \text{ cycles} \leq x\}$$

$$= \text{Prob}\ \{\#\ \{X_i > x\} \text{ in } t \text{ cycles equal to } 0\} = e^{-\beta t(1 - F(x|u))}.$$

Let $x(L,R)$ be the level to be exceeded with probability R in L cycles. In the same way we have

$$1 - R = e^{-\beta L(1 - F(x(L,R)|u))}$$

or $F(x(L, R)|u) = 1 + \frac{\log(1 - R)}{\beta\ L}$ which with $F(x_T|u) = 1 - \frac{1}{\beta\ T}$ leads to the non-parametric relation

$$1 - R = e^{-L/T}$$

which connects the design value x_T for one exceedance on average in T cycles to the level $x(L,R)$ that can be exceeded in cycles with probability R (the risk). Thus the two levels are connected and x_T can be used for design, with a new interpretation.

Notice that the hypothesis that we were dealing with exceedances over large thresholds was necessary to justify the use of the Poisson approximation.

In fact exceedances cluster in general, and this Poisson distribution applies to sequenced approximately independent clusters (with convenient β).

Let us now consider the peaks of each cluster and *assume*, as an example that has been useful, that $F(x)$ is the exponential distribution and so $F(x|u) = 1 - e^{-(x-u)/\delta}$, $x \geq u$, $\delta > 0$, independent of cluster distribution. Then we have $x_T = u + \delta \log (\beta T)$.

Thus if we have, in t_1 cycles, n_1 clusters above u we get

$$\hat{\beta} = n_1/t_1$$

and if peaks of the clusters are x_i we get

$$\hat{\delta} = \frac{1}{n_1} \sum_{1}^{n_1} (x_i - u) = \frac{1}{n_1} \sum_{1}^{n_1} x_i - u = \bar{x} - u .$$

Once more by the δ-method, as $\hat{\beta}$ is asymptotically normal with mean value β and variance β/t_1 , $\hat{\delta}$ is asymptotically normal with mean value δ and variance δ^2/n_1 and zero covariance (independence), we get $\hat{x}_T = u + \hat{\delta} \log (\hat{\beta} T)$ asymptotically normal with mean value x_T and asymptotic variance $V(\hat{x}_T) \sim \frac{\delta^2}{\beta\ t_1} (1 + (\log (\beta\ T)))^2$.

Evidently the corresponding results can be obtained when using Gumbel and Fréchet distributions or distributions attracted to them.

We will finally deal with the A.E.O.T. method. We will use the papers by Smith (1984) and Davison (1984).

Let us assume an i.i.d. sample $\{X_i\}$ with distribution function $F(x)$, fix a threshold u, and select only the exceedances $X_i > u$. To a certain extent we are choosing, on average, the $n(1 - F(u))$ largest values, and so the results are analogous to those concerning $m = n(1 - F(u))$ largest values, dealt with in second section of this chapter. What is important, now, is that we are assuming that the largest values come from an i.i.d. sample which, even forgetting the seasonality, is not necessarily valid because large values tend to cluster and independence for large values is necessarily false. Note that in the P.O.T. method, what was assumed was the (practical) independence of clusters plus a stable distribution of the peaks of the exceedances, which is much more reasonable. Because of this, the treatment will necessarily be sketchy.

Here we will use some preasymptotic form, the generalized Pareto distribution; note that the θ here corresponds to the θ of the von Mises-Jenkinson from $G(z|\theta)$ and is symmetric to the θ of the usual generalized Pareto distribution. The generalized Pareto distribution has the form

$$G(y/\delta|\theta) = 1 - (1 + \theta\, y/\delta)_+^{-1/\theta} \qquad \text{if} \qquad \theta \neq 0$$

$$= 1 - e^{-y+/\delta} \qquad \text{if} \qquad \theta = 0$$

where $\delta = \delta(u) > 0$ and $-\infty < \theta < +\infty$; also $0 \leq y < +\infty$ for $\theta \geq 0$ and $0 \leq y \leq -\delta/\theta$ for $\theta < 0$.

Let $\bar{F}(\Delta|u) = F(u + \Delta|u)$ for $0 \leq \Delta \leq \bar{w}_F - u$ be the distribution of the *excess* $X - u$ over the threshold u. Pickands (1975) proved that $F(.)$ is attracted to $G(.|\theta)$ iff

$$\lim_{u \to \bar{w}_F} \inf_{0<\delta<+\infty} \sup_{\Delta} |\bar{F}(\Delta|u) - G(\Delta/\delta(u)|\theta)| = 0,$$

and so the generalized Pareto distribution $G(\Delta/\delta(u)|\theta)$ is an approximation to $\bar{F}(\Delta|u)$, the distribution of the excess.

Then *assuming* all the excesses $X - u$ to be independent and that $\theta > -1/2$, we can obtain the ML estimators $(\hat{\delta}, \hat{\theta})$ of (δ, θ) which are asymptotically normal with mean values (δ, θ) and the variance-covariance matrix asymptotically

$$\frac{1}{m}\begin{bmatrix} 2\,\delta^2(1+\theta) & \delta(1+\theta) \\ \delta(1+\theta) & (1+\theta)^2 \end{bmatrix}$$

where m is the number of exceedances over u. We skip the results for $\theta \leq -1/2$.

Thus, once a high threshold u is fixed, *assuming* independence, the m exceedances have a binomial distribution with exceedance (or survival) probability $S(u) = 1 - F(u)$. Then with the observed excesses Δ assumed independent and with a generalized Pareto distribution we can, conditionally on m, obtain the ML estimators $(\hat{\delta}, \hat{\theta})$. An approximation to the design value x_T is then given by $F(x_T)$ as $F(x_T) = F(u + (x_T - u)) - F(u) + F(u) \approx S(u)\, G(\frac{x_T - u}{\delta(u)} | \theta) + F(u) = 1 - 1/T$, as in the first part of the section, or equivalently with $x_T = u + \Delta_T$, with Δ_T is given by

$$G(\Delta_T / \delta(u) | \theta) = 1 - \frac{1}{S(u)\, T}$$

or $\Delta_T = \frac{(S(u)\, T)^{\theta} - 1}{\theta}\, \delta(u)$ and thus $x_T = u + \frac{(S(u)\, T)^{\theta} - 1}{\theta}\, \delta(u)$ if $\theta \neq 0$ and $x_T = u + \log (S(u)\, T)\, \delta(u)$ if $\theta = 0$; clearly $\hat{x}_T$ is asymptotically standard normal with mean value x_T and variance of $O(n^{-1})$. The last formula can be compared with $x_T = u + \log (\beta T)$, given previously for the exponential distribution, which corresponds to the Gumbel limiting distribution $(\theta = 0)$; for large u, β will be in general smaller than 1 and thus comparable with $S(u)$.

This second part of the section needs a thorough revision owing to the hypothesis made.

For more details connected to tail estimation see Annex IV.

ANNEX IV

On Extreme Values and Tail Estimation

Introduction

The estimation of the tail of a distribution, using extreme order statistics, sometimes referred as a possibility, has been approached in different ways. We will describe the essential features of these, with some comments and a comparison between them as well as some remarks about some open questions; we have drawn heavily on the review paper by Tiago de Oliveira (1991).

As there is complete duality between the maxima and the minima, and so between the right and left tail, we will only analyse the right tail estimation using the largest order statistics.

Under this umbrella two different problems can be dealt with :

a) estimation of the large quantiles (for probabilities close to 1)

and

b) estimation of the probability of overpassing some given level.

As was shown by Einmahl (1990), the problem of right tail estimation is only interesting when we are beyond (or near) the largest value of the sample, thus constituting a *statistical extrapolation* and, from this perspective, some hypothesis needs to be made. For statistical extrapolation we must suppose that the distribution function of the i.i.d. sample lies in the domain of attraction of some asymptotic distribution of maxima. On the other hand, if we are interested only in the behaviour of the distribution in the interior of the support of the sample, i.e., the closed interval [minimum, maximum] , the utilization of the usual sample distribution function is completely sufficient and for this *statistical interpolation* we do not need to add any hypothesis except the usual (and convenient one) of continuity; in fact, it is a classical non-parametric approach.

Let us recall that a distribution function F(z) is attracted, for maxima, to G (z | θ), noted F(.) $\in \mathcal{D}$ (G (. | θ)), if there exist attraction coefficients $\{\lambda_n, \delta_n (>0)\}$, not uniquely defined, such that for an i.i.d. sample $(x_1, \ldots x_n)$, with the distribution function F(x), we have:

$$P\{(\max(x_1, ..., x_n) - \lambda_n)/\delta_n \le z\} = P\{\max_1^n (x_i) \le \lambda_n + \delta_n z\} = F^n(\lambda_n + \delta_n z) \to G(z \mid \theta)$$

as $n \to \infty$; as the $G(z \mid \theta)$ are continuous the convergence is uniform. It should be noticed that if $F(.) \in \mathcal{D}(\Phi_\alpha)$ we have $\bar{w} = +\infty$ and we can take $\lambda_n = 0$ and if $F \in \mathcal{D}(\Psi_\alpha)$ we have $\bar{w} < +\infty$ and we can take $\lambda_n = \bar{w}$.

The necessary and sufficient conditions for attraction to one of the asymptotic distributions of maxima have been given before, the most practical being the one to be recalled in the description of the third approach although some others will appear in the other two approaches.

For some time the limiting relation $F^n(\lambda_n + \delta_n z) \to G(z \mid \theta)$ as $n \to \infty$ led to the approximation $F^n(\lambda_n + \delta_n z) \approx G(z \mid \theta)$ or $F^n(y) \approx G((x - \lambda)/\delta \mid \theta)$ where the index n was omitted because θ, λ and δ were to be estimated from a sample of maxima. This way gave an approximation to $F^n(y)$, and sometimes $G^{1/n}((y - \lambda)/\delta \mid \theta)$, equal to $G(y - \lambda')/\delta' \mid \theta)$ by stability, was used as a (very) rough approximation to $F(y)$. Thus the idea of using the largest observations of a sample to estimate $F(y)$ came naturally. As a side remark, it should be noted that, for instance, if $F^n(\lambda_n + \delta_n z) \to \Lambda(z)$ as $n \to \infty$, for n not sufficiently large, a better approximation to $F^n(y)$ can be obtained using, instead of $G((y - \lambda)/\delta \mid 0) = \Lambda((y - \lambda)/\delta)$, a penultimate distribution function $G((y - \lambda)/\delta \mid \theta)$ [with λ, δ and θ to be estimated from the sample of i.i.d. maxima]. Such is the case for the normal distribution when the criterion used for comparison of the asymptotic approximation is the maximum absolute error in the probability evaluation, i.e., when we try to minimize $\sup_y |N^n(\lambda_n + \delta_n y) - G(y \mid \theta)|$ where we have a sequence of $\theta = \theta_n \uparrow 0$. The global approach used here integrates this penultimate approximation naturally by *estimating*, from the sample of i.i.d. maxima, the θ, λ and δ that fits the data better.

Before continuing an important remark must be made. When we are dealing with maxima *only*, in many cases, we can split the sequence of observations (the concrete time-series) into natural blocks for which the maxima are surely independent, thus having a sample of independent maxima; this was the case for various examples in this Part II. These maxima, as was said for stationary sequences, for markovian sequences, for m-dependent sequences, for strongly mixing sequences and even for some evolutionary sequences, as will be shown in Part IV, have, under some more strong or mild conditions on association (many times, on the correlation), the same asymptotic distribution for maxima; the same happens for analogous continuous time stochastic processes for which we sometimes have records, many times discretized for the statistical analysis. The natural block used for hydrological data is the natural hydrological year, but for oceanographical and meteorological data the block used, in general, is the civil (!) year. This is the

well-known "yearly method". But if we use the largest values of the time-series it is obvious that the observations are not independent, not even stationary, because — even for an additive model — they have underlying the effects of trend and of cyclic variation, which have to be removed giving many times the largest values at *other* time points. Even when assumed to be stationary sequences, large (and also small) values occur by clusters and so in general the 2nd, 3rd, ... maximum occur at time points close to that of the (1st) maximum, as the association (correlation) is very strong; this drawback can be avoided by "cleaning" the time series from the cluster effect, which means that once the 1st maximum at instant t is obtained, data should be "cleaned" from observations obtained in the instants (time steps) $t-m$ to $t+m$, and the 2nd maximum sought in the remaining data, etc.; *this is the natural procedure if we are dealing with a stationary m-dependent sequence*. For other situations something similar must be found. Another difficulty appears when we are using the largest values in a time-series split in the "wrong" blocks in which some number of largest values are to be used; if the block is, for instance, a civil year (and not a natural one like the hydrologic year), it may happen that a storm occurs by the end/beginning of the year and the maxima registered for the two sucessive years correspond to observations made at 9 p.m. on 31st December and to 3 a.m. on the following 1st January, which clearly, belong to the same cluster against the "cleaning" procedure sketched above, thus giving rise to errors in planning and design.

The "cleaning" procedure is thus essential to avoid errors; a good example of the "independentization" of observations in a time series was done by Dijk and de Haan (1990); the behaviour of the auto-correlation coefficient is sometimes a good tool to steer the procedure of choosing m.

The three approaches to tail estimation

We will now consider a large sample of i.i.d. observations, which is supposed to have a distribution function $F(.)$ such that $F(.) \in \mathcal{D}(G(.\mid \theta))$. This means, *for practical applications*, that the removal of the trend and cyclic variation and the cluster "cleaning" have been made. So far, it seems that only two situations exist where independence can be assumed directly: largest ages of death for men, for women or, for both, in some country if, in the years considered, no war, epidemic, national disaster or famine struck the region under study; and large fires data (with compensation for inflation) in some region if we discard fires with "epidemics" like forest fires, fires from bombing, and large fires in built-up areas. In these cases the "pooling" of data is natural, giving rise to a (very) large sample from which a relatively large set of largest values can be used to estimate tail properties. Possibly analogous cases exist but the conditioning to *assume*

independence is very clear. Once the estimator of F (.) is obtained, we can estimate either the desired *exceedance probability* or the *quantile.*

The first approach is due to Pickands (1975) and has been used with variations in Smith (1987), Smith and Weismann (1987), Smith (1989) and Pickands (1989). Basically it consists of the approximation of the asymptotics of the distribution to the *excesses over a threshold* by a generalized Pareto distribution which was obtained in the important Pickands (1975) paper. The method, then, corresponds to the P.O.T. (peaks over threshold) method sometimes used by engineers, against the yearly method.

The second approach is based on Weismann (1978) where the joint asymptotic distribution of a fixed number of the largest order statistics is used.

Both approaches have been used *as if* the *asymptotic or preasymptotic* distribution was the *actual* one, evidently raising the already classical criticism to the usage of $G((y-\lambda)/\delta \mid \theta)$, with λ, δ and θ, estimated, for the actual distribution $F^n(y)$ of the maximum of n observations when, in fact, both methods are approximations; note that the use of penultimate distributions referred to, is one way to compensate by using an approximation to the distribution of independent maxima. A guide to this approach was given by Tiago de Oliveira (1981) and (1984") where the theory of statistical choice for $\theta < 0$ (Weibull), $\theta = 0$ (Gumbel) or $\theta > 0$ (Fréchet) was developed; this statistical choice can give a better way of approaching the description of the statistical behaviour of independent maxima, even when generated by a dependence process, as happens in the yearly method.

A third and new approach has been developed essentially by Dekkers and de Haan (1989), Dekkers, Einmahl and de Haan (1989) and Dijk and de Haan (1990); it is connected with some previous results for the estimation of the shape parameter for Fréchet distribution, such as the Hill estimator.

This approach, although steered by the asymptotic distribution of maxima, *does not* use it as the *actual* distribution, in contrast to the previous ones that, at some point, substitute the actual (unknown) distribution by its asymptotic approximation.

Notice that the initial assumption that all observations are i.i.d. and their distribution function is attracted to some asymptotic distribution of extremes $G(z \mid \theta)$ leads to the use of upper order statistics — sometimes called extremal, sometimes intermediate, see Reiss (1989) — which are evidently dependent; independence thus is at the beginning but *not* in the observations choosen for the statistical analysis.

The asymptotic normality of Pickands' (1975) estimator of θ has been proved in various papers, such as that of Dekkers and de Haan (1989).

Complementary results can be found in Tiago de Oliveira ed., (1984), especially the papers by Weismann (1984), Smith (1984) and Davison (1984), and also in Davison and Smith (1990).

This Annex applies results of Chapters III and IX for the estimation of large quantiles and small exceedance probabilities; but it must be remarked that optimal procedures for the choice of the number and the orders of large quantiles, the efficiency evaluation, the speed of convergence, etc, are yet open problems.

The first approach

Let $(x_1, \ldots, x_n)$ be a sample of i.i.d. random variables (observations) with distribution function $F(x)$, attracted to some $G(z \mid \theta)$) and let $\bar{w} = \inf \{x \mid F(x) = 1\}$ be the right-end point of $F(.)$. As we are going to use the upper order statistics, let us recall that x''_k denotes the k-th maximum, i.e., the ascending order statistics $x'_{n+1-k} = x_{n+1-k:n}$; the maximum is obviously x''_1 and the minimum is x''_n.

Recall that Pickands (1975) gave the name of standard *generalized Pareto distribution* to the distribution function $P(x \mid \theta) = 1 - (1 + \theta x)_+^{-1/\theta}$ where $P(x \mid 0)$, defined as $P(x \mid 0^+) = P(x \mid 0^-) = 1 - e^{-x_+}$, is the standard exponential distribution.

If $\theta \geq 0$ we have for $P(x \mid \theta)$ the right-end point $\bar{w} = +\infty$ but for $\theta < 0$ the right-end point is $\bar{w} = -1/\theta$; in general $\bar{w} = 1/(-\theta)_+$ for the standard generalized Pareto distribution.

Evidently if X has the distribution function $F(x)$, the conditional survival function of the *excess* $X - u$, when $X \geq u$, is given by

$$S(x \mid u) = P(X > u + x \mid X > u) = \frac{1 - F(u + x)}{1 - F(u)} \qquad \text{for } x \geq 0$$

and the conditional distribution function of the excess is evidently

$$F(x \mid u) = 1 - S(x \mid u) = \frac{F(x + u) - F(u)}{1 - F(u)} \qquad \text{for } x \geq 0.$$

Pickands (1975) proved the following result:

$$\lim_{u \uparrow \bar{w}} | F(x/u) - P(x/\delta(u)/\theta) | = \lim_{u \uparrow \bar{w}} | S(x/u) - (1 + \theta\, x/\delta(u))_{+}^{-1/\theta} | = 0$$

iff $F(.) \in \mathcal{D}(G(./\theta))$.

Notice that the scale parameter $\delta(u)$ depends on u.

Also in the same paper it was proved that :

If m_n *is an integer such that* $1 \leq 4\ m_n \leq n$ *and* $m_n \to \infty$ *but* $m_n / n \to 0$

then $\theta^* = \log \left(\dfrac{x''_{m_n} - x''_{2m_n}}{x''_{2m_n} - x''_{4m_n}} \right) / \log 2$

and $\delta^* = \theta^* (x''_{2m_n} - x''_{4m_n}) / (2^{\theta^*} - 1)$ *are such that*

$$\sup_x | F(x/x''_{4m_n}) - P(x/\delta^*/\theta^*) | \xrightarrow{a.s.} 0 \quad \textit{and also that} \quad \theta^* \xrightarrow{P} \theta .$$

This theorem shows that for the (large) random threshold x''_{4m_n} the conditional distribution of the random excess $x - x''_{4m_n}$ is asymptotically approximated by a Pareto generalized distribution.

A rule for the choice of m_n is also given.

Later Dekkers and de Haan (1989) proved the asymptotically normality of θ^* (see the third approach, first result, of this Annex).

Smith (1987) (with a change of sign in θ) used Pickands (1975) result with either a moving or a random threshold u_n. Evidently, conditional on a given u, the excesses $x_i - u$ are independent random variables and its number (the number of exceedances of u) is random (a binomial variate if u is not random). Thus u should not be neither too large to have few exceedances or too small to have almost all the sample giving rise to exceedances. Then the excesses are taken as having, as *actual* distribution, a generalized Pareto one, and estimation is made by the method of maximum likelihood which is only regular if $\theta > -1/2$, with an extension if $\theta \leq -1/2$. Let us denote by $(\hat{\theta}, \hat{\delta})$ the maximum likelihood estimators. Then Smith (1987), assuming that the excesses have an *actual* generalized Pareto distribution, shows, with N denoting the random number of excesses, that:

$\sqrt{N}\ (\hat{\delta}/\delta - 1\ ,\ \hat{\theta} - \theta)$ is asymptotically binormal with zero mean values and variance-covariance matrix $V = (1+\theta)\begin{bmatrix} 2 & 1 \\ 1 & 1+\theta \end{bmatrix}$ *which is the Cramér-Rao bound .*

It is then shown that if the excesses do not have the actual generalized Pareto distribution but

$$F(x \mid u) - P(x / \delta(u) \mid \theta) = O(1/\sqrt{N}) \qquad \text{then}$$

$\sqrt{N}\ (\hat{\delta}/\delta - 1\ ,\hat{\theta} - \theta)$ is asymptotically binormal with mean values not necessarily zero but with the same variance-covariance matrix V .

The non-zero mean values allow for the existence of a bias in the approximation; but knowledge of this bias depends on the structure and the form of the tail $1 - F(x)$. The study is made for each form of the asymptotic distribution (Weibull, Gumbel or Fréchet) in particular, with additional conditions.

The distribution function for values $y = u + x$, $x \geq 0$, naturally using $1 - \hat{F}(u) = N/n$, is then estimated by $1 - \hat{F}(u + x) = \frac{N}{n}(1 + \hat{\theta}\, x / \hat{\delta})_+^{-1/\hat{\theta}}$ or $\hat{F}(y) = 1 - \frac{N}{n}(1 + ((y-u)/\hat{\delta})_+^{-1/\hat{\theta}})$,

for $y \geq u$ which is asymptotically normal by the use of the δ-method; see Tiago de Oliveira (1982').

The quantile for probability $p = 1 - \varepsilon$ is estimated by solving back the equation $\hat{F}(y) = 1 - \varepsilon$ and so is given by $\hat{y}(1-\varepsilon) = u + \hat{\delta}\ [(\frac{n\,\varepsilon}{N})^{-\hat{\theta}} - 1]$ which is also asymptotically normal by the δ-method.

The choice of the threshold u_n (random or moving) and the way that $u_n \uparrow \bar{w}$ is partially discussed, distribution by distribution, using some additional assumptions on the type of variation of the tail $1 - F(y)$.

Here are some examples :

If $F \in \mathcal{D}(\Phi_\alpha)$, that is *iff* $\bar{w} = +\infty$ and $L(x) = x^\alpha (1 - F(x))$ is of *slow variation* (i.e., $\frac{L(x\,t)}{L(x)} \to 1$ when $x \to \infty$) such that $L(t\,x) / L(x) = 1 + k(t)\ \phi(x) + 0(\phi(x))$, where $k(t) = \frac{C}{\rho}(t^\rho - 1)$ $(\rho < 0)$, then the previous asymptotic result is valid with the mean values $(\frac{\mu(1+\theta)(1-2\theta\rho)}{1+\theta-\theta\rho}\ ,\ \frac{-\mu(1+\theta)\theta(1+\rho)}{1+\theta-\theta\rho})$ and variance-covariance matrix V given; here we denote by $\mu = \text{plim}\ \frac{\sqrt{N}\ C\ \phi(u_N)}{\alpha - \rho}$. Recall that $\theta = 1/\alpha$ and $\delta(u) = u/\alpha$.

The Hill estimator for α is given by

$$\alpha_n^* = \frac{N}{\sum_1^N \log x_i'' - N \log u}$$

and the asymptotic ratio of $\frac{MSE\,(\hat{\alpha}_n)}{MSE\,(\alpha_n^*)}$ is

$$(\alpha + 1)^2 \left(\frac{1 + \rho}{\alpha + 1 - \rho}\right)^{2\alpha/(\alpha - 2\rho)},$$

which shows, that according to the value of ρ, one or other of the estimators is better.

The choice of a deterministic sequence $\{u_n\}$ must be such that $n(1 - F(u_n)) \to \infty$, $\sqrt{n(1 - F(u_n))}\ C\,\phi(u_n)) \to \mu\,(\alpha - \rho)$, a result of which is of doubtful usefulness because $F(x)$ is generally unknown, as well as μ and α; in the case where u is a random threshold we choose a sequence of N_n such that $N_n \to \infty$, $N_n/n \xrightarrow{P} 0$ and use $u_n = x''_{N_n+1}$.

The case for Ψ_α, when $\alpha > 2$, has some similarities with the previous one. The necessary and sufficient condition to be attracted to $\Psi_\alpha(x)$ is that $\bar{w} < +\infty$ and $L(x) = x^\alpha(1 - F(\bar{w} - 1/x))$ is also of slow variation.

Assuming the same condition as before, with $\rho < 0$, and defining μ by $\mu = \lim \sqrt{N}\ C(\phi((\bar{w} - u_n)^{-1})/(\alpha - \rho)$, we see that it is also valid with the mean values

$$\left(\frac{\mu(1 + \theta)\,(1 + 2\,\theta\,\rho)}{1 + \theta + \theta\,\rho}\ ,\ \frac{-\,\mu\,\theta\,(1 + \theta)\,(1 - \rho)}{1 + \theta + \theta\,\rho}\right)$$

and also the same variance-covariance matrix V. We have, then, $\theta = -1/\alpha$ and $\delta = (\bar{w} - u)/\alpha$. This result is essentially due to Smith and Weismann (1985).

The estimator of the right-end point $\bar{w}$ is given by $\hat{\bar{w}} = u_n + \hat{\alpha}\,\hat{\delta} = u_n - \hat{\delta}/\hat{\theta}$ which is also asymptotically normal. The random threshold $u_n = x''_{N_n+1}$ must be such that $\bar{w} - u_n = O_p((N/n)^{1/\alpha}) = O_p((n/N)^\theta)$, but θ is unknown except for the bounds $-1/2 < \theta < 0$.

The cases for $\alpha \in\ ]0, 2]$ are analogous but more complex.

Now consider the asymptotic distribution Λ. Although three approaches are presented in Smith (1987), we will describe only the second one because it corresponds to the P.O.T. methods used by engineers, which directly assume $\theta = 0$. The estimator of δ is then $\hat{\delta} = \frac{1}{N} \sum_{1}^{N} \log x_i'' - u_n$ where the x_i'' are the order statistics larger than u_n. Then for the estimator of the distribution function

$$\hat{F}(y) = 1 - \frac{N}{n} \exp\left(\frac{y - u_n}{\hat{\delta}}\right)$$

it can be shown that

$$\sqrt{N}\left(\frac{1 - \hat{F}(u_n + z\,\phi(u_n))}{1 - F(u_n + z\,\phi(u_n))} - 1\right)$$

is asymptotically normal with mean value $\mu z - \frac{\mu z^2}{2}$ and variance $1 + z^2$, where we define μ by $\mu = \text{plim } \sqrt{N}\,\phi(u_n)/n$ and $\phi(t)$ is such that

$$1 - F(x) = c(x) \exp\left(-\int_{-\infty}^{x} \frac{d\,t}{\phi(t)}\right), \quad x < \bar{w}$$

with $c(x) \to 1$ as $x \to \bar{w}$, ϕ is a positive differentiable function, and $\phi'(x) \to 0$ as $x \to \bar{w}$.

Notice that, in this result, we have an asymptotic evaluation of the relative error of the approximation by the exponential distribution. The fact that the result, once more, depends on knowledge (at least global) of the initial F leads to difficulties.

The second approach

This approach stems essentially from Weissmann's (1978) paper and assumes that the number of upper order statistics used is fixed, although the sample size increases indefinitely. The sample $(x_1, \ldots, x_n)$ is assumed, once more, to be i.i.d. with a distribution function which is in the domain of attraction of one of the three asymptotic distributions supposed known, analogously to what happened before.

Let us then suppose that $F \in \mathcal{D}(\Lambda)$, i.e., that $F^n(\lambda_n + \delta_n x) \to \Lambda(x)$. Denoting by $y_{i,n} = (x_i'' - \lambda_n)/\delta_n = y_i$ for short, it can be shown that:

The vector $(y_1, \dots, y_m)$, *if* $F \in \mathcal{D}(\Lambda)$ *with attraction coefficients* λ_n *and* δ_n (>0), *has the limiting density* $\exp(-e^{-y_m} - \sum_1^m y_i)$ *with* $y_1 \geq y_2 \geq \dots \geq y_m$.

Thus we can approximate the actual distribution of the x_i'' (or y_i) (i = 1,2, ... , m) by the asymptotic one, then having the likelihood

$$\frac{1}{\delta^m} \exp\left(e^{-\frac{x_m'' - \lambda}{\delta}} - \sum_1^m \frac{x_i'' - \lambda}{\delta}\right).$$

The random pair $(\sum_1^m x_i'', x_m'')$ is a sufficient statistic for (λ, δ) and the maximum likelihood estimators $(\hat{\lambda}, \hat{\delta})$ and minimum variance unbiased estimators (λ^*, δ^*) of (λ, δ) are

$$\hat{\delta} = \frac{1}{m} \sum_1^m x_i'' - x_m'' \,, \quad \hat{\lambda} = x_m'' - \log m \,.\, \hat{\delta}$$

and

$$\delta^* = \frac{1}{m-1} \sum_1^{m-1} x_i'' - x_m'' = \frac{m}{m-1}\,\delta^* \,, \quad \lambda^* = x_m'' + (S_m - \gamma)\,\delta^*$$

with $S_m = \sum_1^{m-1} \frac{1}{i}$. We have $V(\delta^*) = \frac{\delta^2}{m-1}$ and

$$V(\delta^*) = \left[\frac{(S_m - \gamma)^2}{m-1} + S_m'\right] \delta^2 \quad \text{where} \quad S_m' = \frac{\pi^2}{6} - \sum_1^{m-1} \frac{1}{i^2}.$$

By a result from Mejzler (1949), also given by de Haan (1971), we have, for the quantile $F^{-1}(1 - c/n)$, the basic relation: $(F^{-1}(1 - c/n) - \lambda_n)/\delta_n \to \log c$ which shows that for λ_n we can take $F^{-1}(1 - 1/n)$ and for $\delta_n = F^{-1}(1 - e/n) - F^{-1}(1 - 1/n)$, as known since Gnedenko's (1943) paper. Thus

The quantile $F^{-1}(1 - c/n)$, *if* $F \in \mathcal{D}(\Lambda)$, *can be estimated by* $\hat{F}^{-1}(1 - c/n) = \hat{\lambda}_n + \hat{\delta}_n \log c$ *or by* $F^{*-1}(1 - c/n) = \lambda_n^* + \delta_n^* \log c$.

In the case where $F \in \mathcal{D}(\Phi_\alpha)$, as previously stated $\bar{w} = +\infty$ and we can take $\lambda_n = 1$, and as $\Phi_\alpha(z) = \Lambda(\alpha \log z)$ $(z > 0)$, $\log X$ $(X > 0)$ is attracted to Λ. Reducing to the previous case and afterwards returning to the observed values we have

If $F \in \mathcal{D}(\Phi_\alpha)$ *(α unknown) we have*

$$\hat{\alpha} = \left(\frac{1}{m} \sum_{1}^{m} \log x''_i - \log x''_m\right)^{-1} \quad \text{and} \quad \hat{F}^{-1}(1 - c/n) = (m/c)^{1/\hat{\alpha}} x''_m \ .$$

For the case $F \in \mathcal{D}(\Psi_\alpha)$ we know that $\bar{w} < +\infty$ and $\Psi_\alpha(z) = \Lambda(-\alpha \log(-z))$ for $z < 0$ and then

If $\bar{w}$ *is known and* $F \in \mathcal{D}(\Psi_\alpha)$, *then we have*

$$\hat{\alpha} = \left(\log(\bar{w} - x''_m) - \frac{1}{m} \sum_{1}^{m} (\log \bar{w} - x''_i)\right)^{-1} \quad \text{and}$$

$$\hat{F}^{-1}(1 - c/n) = \bar{w} - \left(\frac{c}{m}\right)^{1/\hat{\alpha}} (\bar{w} - x''_m) \ .$$

The case where $\bar{w}$ is unknown was not considered.

The third approach

As stated previously, the third approach is different from the previous ones,in that it does not use the approximation given by an asymptotic or preasymptotic distribution as the *actual* one, as was done in the first and second approaches in one way or another.

The essential result for large quantile estimation, is based in the attraction conditions using the quantile function $Q(v) = F^{-1}(v)$, the inverse function $F^{-1}(v) = \inf\{x \mid F(x) \geq v\}$, given by Mejzler (1949) and de Hann (1970). We have the basic result:

$F(.) \in \mathcal{D}(G(. \mid \theta))$ *iff* $\dfrac{Q(1 - t\,x) - Q(t)}{Q(1 - t\,y) - Q(t)} \to \dfrac{x^{-\theta} - 1}{y^{-\theta} - 1}$

$\left(\dfrac{\log x}{\log y} \quad \text{if} \quad \theta = 0\right)$ *locally uniformly when* $t \to 0$.

Let us denote by $F_n(x)$ the sample distribution function and by $Q_n(v) = F_n^{-1}(v)$ the sample quantile function, which are both in correspondence with $F(x)$ and $Q(v)$. It is immediate that $Q_n(1 - m_n / n) = x''_{m_n}$. Thus if we are seeking the p-quantile of $F(x)$, i.e., the solution of $F(x) = p = 1 - \varepsilon$, this is equivalent to seeking $Q(1 - \varepsilon)$. Evidently if we want to design a dam with a probability ε' of being overpassed for n' units of time under the i.i.d. hypothesis, we have to solve the equation $F^{n'}(x) = 1 - \varepsilon'$ or $F(x) = (1 - \varepsilon')^{1/n'}$, that is, use $Q(1 - \varepsilon)$ where $\varepsilon = 1 - (1 - \varepsilon')^{1/n'}$ is much closer to zero as ε' is small and n' large.

Our purpose is thus to estimate $Q(1 - \varepsilon)$ from a sample of n i.i.d. observations whose distribution function is attracted to $G(. \mid \theta)$ using the some largest order statistics.

As we can write

$$Q(1-\varepsilon) = \frac{Q(1-\varepsilon) - Q(1-\frac{m_n}{n})}{Q(1-\frac{m_n}{n}) - Q(1-\frac{2\,m_n}{n})} \left[Q(1-\frac{m_n}{n}) - Q(1-2\frac{m_n}{n})\right] + Q(1-\frac{m_n}{n})$$

replacing Q by Q_n where convenient, as $\frac{m_n}{n} \to 0$ by hypothesis, and, substituting the first quotient by its limit given by the attraction conditions, as $\varepsilon = \frac{n}{m_n}\ \frac{m_n\,\varepsilon}{n}$,

$$Q_n^*(1-\varepsilon) = \frac{(m_n/\varepsilon\, n)^{\theta_n^*} - 1}{1 - 2^{-\theta_n^*}} \left[x''_{m_n} - x''_{2m_n}\right] + x''_{m_n}$$

where θ_n^* is the estimate of θ given by Pickands (1975)

$$\theta_n^* = \log \frac{x''_{m_n} - x''_{2m_n}}{x''_{2m_n} - x''_{m_n}} \Big/ \log 2 .$$

Only three largest order statistics (intermediate as $m_n \to \infty$ also, but at a slower rate) are sufficient to deal directly with the estimation of the quantile of probability $1 - \varepsilon$.

Then Dekkers and de Haan (1989) show the following results :

If $Q(t)$ has a positive derivative and a positive function $b(t)$ exists such that

$$t^{1+\theta}\frac{x^{1+\theta}Q(1-tx)-Q(1-t)}{b(t)} \to \pm \log x \quad as \quad t \to 0$$

then $F(.) \in \mathcal{D}(G(.|\theta))$ *and* $\sqrt{m_n}(\theta_n^* - \theta)$ *is asymptotically normal with mean value zero and variance* $\theta^2(2^{2\theta+1}+1)/(2(2^\theta-1)\log 2)^2$ *for sequences* $m_n \to \infty$ *such that* $m_n = o(n/g^{-1}(1/n))$ *with* $g(t) = t^{1+2\theta}Q'(1-t)^2/b^2(t)$ *(when* $\theta = 0$ *the asymptotic variance is* $\frac{3}{4(\log 2)^4}$ *as follows from l'Hôpital's rule).*

This allows us to conclude the consistency of θ_n^*. As we are dealing with intermediate order statistics that are asymptotically normal — see Reiss (1989) — $Q_n^*(1-\varepsilon)$ is expressed in them and also in θ_n^*, all asymptotically normal, by the δ-method we can expect that $Q_n^*(1-\varepsilon)$ is asymptotically normal. In fact we have

If $Q'(t)$ *exists and* $t^{1+\theta}Q'(1-t)$ *is of slow variation when* $t \downarrow 0$ *then*

$$\sqrt{2m_n}\ \frac{x''_{m_n} - Q(1-\varepsilon_n)}{x''_{m_n} - x''_{2m_n}}$$

is asymptotically normal with mean value zero and variance $\theta^2\, 2^{2\theta+1}/(2^\theta-1)^2$ *if* $\varepsilon_n \to 0$ *and* $m_n = [n\varepsilon_n]$ *(when* $\theta = 0$ *we have for the variance* $2/(\log 2)^2$*).*

This results allows us as before to obtain confidence intervals for the quantile of probability $1-\varepsilon_n$.

In Dekkers, Einmahl and de Haan (1989) there was introduced a general estimator of θ also asymptotically normal that can be used for all values of θ, an estimator which is a modification of Hill's estimator only applicable for $\theta > 0$. Denoting by

$$M_n = \frac{1}{m_n}\sum_{1}^{m_n}(\log x''_i - \log x''_{m_n})^2 ,$$

it is shown that

If $Q'(t)$ *exists and* $t^{1+\theta}Q'(1-t)$ *is of slow variation when* $t \downarrow 0$ *then*

$$\sqrt{m_n}\ \frac{x''_{m_n} - Q(1-\varepsilon_n)}{x''_{m_n} \cdot M_n}$$

is asymptotically normal, with mean value zero and variance $(1 - \min(0, \theta))^2$ *if* $\varepsilon_n \to 0$, $n\,\varepsilon_n \to \infty$ *and* $m_n = [n\,\varepsilon_n]$.

Let us finally study the estimator of the exceedance probabilities.

In Dijk and de Haan (1990) the problem is approached in a somewhat similar way to the generalized Pareto distribution. Consider a level y_n such that, with an i.i.d. sample from a population with distribution function $F(.) \in \mathcal{D}(G(.\mid\theta))$, we have $n(1 - F(y_n)) \to 0$. As is well known, the attraction of $F(.)$ to $G(.\mid\theta)$ is equivalent to $n(1 - F(\lambda_n + \delta_n z)) \to (1 + \theta z)_+^{-1/\theta}$. With $m_n \to \infty$, $\frac{m_n}{n} \to 0$, we have also $\frac{n}{m_n}(1 - F(\lambda_{n/m_n} + \delta_{n/m_n} z)) \to (1 + \theta z)_+^{-1/\theta}$, and so

$$1 - F(y_n) \approx \frac{m_n}{n}\left(1 + \theta\,\frac{y_n - \lambda_{n/m_n}}{\delta_{n/m_n}}\right)_+^{-1/\theta}$$

which suggests the estimator

$$F_n^*(y_n) = 1 - \frac{m_n}{n}\left(1 + \theta_n^*\,\frac{y_n - \lambda_{n/m_n}}{\delta_{n/m_n}}\right)_+^{-1/\theta_n^*}$$

when we assume $\quad y_n = \lambda_{n/m_n} + \delta_{n/m_n}\,\frac{c_n^\theta - 1}{\theta}$

(note that $\quad c_n \sim \frac{m_n}{n(1 - F(y_n))}$).

Then the asymptotic normality of $F_n^*(y_n)$ is proved.

Let us put $q_\theta(x) = \int_1^x s^{\theta-1} \log s\ ds$ [$\sim (x^\theta \log x)/\theta$ if $\theta > 0$, $\sim (\log x)^2/2$ if $\theta = 0$ and $\sim 1/\theta^2$ if $\theta < 0$], $c_n^* = \frac{m_n}{n(1 - F_n^*(y_n))}$ and denote by θ_n^* the estimator of θ given in Dekkers, Einmahl and de Haan (1989). Then we can prove that:

If $m_n \to \infty$, $m_n/n \to \infty$, $\dfrac{q_\theta(c_n)}{c_n^\theta \sqrt{m_n}} \to 0$,

$$\frac{\sqrt{m_n}}{q_\theta(c_n)} \left(\frac{c_n^\theta - 1}{\theta}\right)^2 \sup_{u \geq \lambda_{n/m_n}} \left| 1 + \theta + \frac{(1 - F(u))\, F''(u)}{F'(u)^2} \right| \to 0$$

as $n \to \infty$, *and the conditions that allow the estimation of* θ *by* θ_n^* *are valid then*

$$\frac{\sqrt{m_n}\; c^{*\,\theta_n^*}}{q_{\theta_n^*}(c_n^*)} \left[\frac{1 - F(y_n)}{1 - F_n^*(y_n)} - 1\right]$$

is asymptotically normal with mean value zero and variance $1 + \theta^2$ *if* $\theta \geq 0$,

$$\frac{(2 - 3\theta)^2}{1 - 2\theta} + \frac{(1 - 2\theta)^2 (5 - 11\theta)}{(1 - 3\theta)(1 - 4\theta)} - \frac{4(2 - 3\theta)(1 - 2\theta)}{1 - 3\theta} \quad \text{if} \quad \theta < 0.$$

The choice of m_n is strongly dependent on the initial distribution function F; for instance for a Cauchy distribution we must have $m_n = O(n^{4/5})$ but for the normal distribution we must have $m_n = O((\log n)^2)$.

Some final comments

Although the results contained in the third approach may be difficult to utilize with regard to the estimation of the exceedance probability, this one seems to be the most appropriate procedure as it only assumes the *existence* of one asymptotic distribution of maxima to the *actual distribution*, *all observations supposed i.i.d.*, instead of *using* the asymptotics directly and afterwards, when possible, correcting for bias. Although for this procedure there is a study of optimal m_n by Dekkers and de Haan (1991), the results are so different that there does not seem to exist a general rule to be used universally. To compare the first and the third method, one way could be the comparison of the asymptotic MSE's. In fact as the MSE for the first method is of the order of $\frac{A}{N}$ or, as $N \sim n(1 - F(u_n))$, of the order of $\frac{A}{n(1 - F(u_n))}$, and for the third method the MSE (for the first θ^* the MSE equals the variance) is of the order of $\frac{B}{m_n}$ (the exact values of the constants A and B are given in the text), we have $\frac{MSE_1}{MSE_3} = O\left(\frac{m_n}{n(1 - F(u_n))}\right)$. But, in fact, comparison is difficult as the optimal (or approximately optimal) u_n in the first

method depends on the *assumed* or *supposed* behaviour of $1 - F(x)$, as well as A depending on the constants of slow variation but m_n in the third method depends on the behaviour of Q(v) through g(t).

If we consider $F(x) = \Phi_\alpha(x)$ we see $m_n \sim B\, n^{2/3}$ and $n(1 - \phi_\alpha(v_n)) \sim A\, n^{2/3}$, m_n being optimal for the second estimator in the third approach, and the methods are of the same order.

We can also devise a statistical choice procedure to choose between $\theta < 0$, $\theta = 0$ and $\theta > 0$, analogous to the one developed by Tiago de Oliveira (1981); see Tiago de Oliveira (1991') for details.

Exercises — Part II

II.1—If $Z = \max(X - a, Y - b)$ with (X,Y) independent with the reduced Gumbel distribution as well as z, show that Prob $\{X - a \leq Z \leq z\} = e^{-b}\ \Lambda(z)$.

II.2— Consider one of the sets of data given in the book (or elsewhere); verify using the probability paper if they can be fitted to the Gumbel, Weibull or Fréchet distributions; compare with the use of the choice statistic; estimate the parameters; take logarithms and exponentials of the same set of data and analyse the fit. Is there any apparent contradiction ? Explain it.

II.3— Consider the Gumbel distribution $\Lambda((x - \lambda)/\delta)$. Obtain the estimators, confidence interval for λ, and tests corresponding to $\lambda_o = \lambda_o$ vs. $\lambda > \lambda_o$, $\lambda < \lambda_o$ and $\lambda \neq \lambda_o$. Study the statistics used.

Passos Coelho and *Gil* (1963)

II.4— Consider the prediction problem for $\Lambda(x - \lambda)$ from a sample $(x_1, \ldots, x_n)$ to a second sample (to be observed) $(Y_1, \ldots, Y_n)$. Show that the least squares predictor of $W = \max(Y_1, \ldots, Y_m)$ is $p(\hat{\lambda}) = \hat{\lambda} + k$ with $k = \gamma + \log \frac{m}{n} + \Gamma'(n)/\Gamma(n)$ with a mean square error of prediction of $E^2 = \frac{\pi^2}{6} + \Gamma''(n)/\Gamma(n) - (\Gamma'(n)/\Gamma(n))^2$.

The shortest prediction interval for W, with level ω, is given by

$$c_\omega \leq W - \hat{\lambda} \leq d_\omega$$

where c_ω and d_ω are given by the prediction level and the relation

$$\frac{c_\omega}{n+1} + \log\left(1 + \frac{m}{n} e^{-c\omega}\right) = \frac{d_\omega}{n+1} + \log\left(1 + \frac{m}{n} e^{-d\omega}\right)$$

or, approximately, $c_\omega + m\, e^{-c\omega} = d_\omega + m\, e^{-d\omega}$.

Tiago de Oliveira (1963)

II.5— Consider two populations P and P' whose distribution functions are $\Lambda(x - \lambda)$ and $\Lambda(x - \lambda')$, with the parameters λ and λ' estimable from samples $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$.

By the likelihood ratio criterion, show that if $V_n = \frac{1}{n} \sum_1^n e^{-x_i}$ and $V'_{n'} = \frac{1}{n'} \sum_1^{n'} e^{-x'_i}$ we should discriminate for P or P' a new observation z by the rule

$$V_n^n \, (V'_{n'} - \frac{V'_{n'} - e^{-z}}{n' + 1})^{n'+1} \lesseqgtr c(\lambda, \lambda') \, V'^{n'}_{n'} \, (V_n - \frac{V_n - e^{-z}}{n + 1})^{n+1} ,$$

c being computed to lead to equal chances of misclassification; c depends on $\lambda' - \lambda$. For large values of n and n' we have using the notation $\hat{\lambda} = -\log V_n$ and $\hat{\lambda}' = -\log V'_n$ the approximation,

$$\hat{\lambda} - e^{\hat{\lambda} - z} \lesseqgtr \log c + \hat{\lambda}' - e^{\hat{\lambda}' - z} .$$

II.6— Develop sequential probability ratio tests for $\Lambda(x - \lambda)$. Show that to test $\lambda = 0$ vs. $\lambda = \lambda_1$, as the logarithm of the sequential probability ratio is $Z_n = n [\lambda_1 + (1 - e^{\lambda_1}) V_n]$ ($V_n = \frac{1}{n} \sum_1^n e^{-x_i}$), then α and β being the approximate values of the errors of first and second kinds, the test rule is to accept $\lambda = 0$ if $V_n \leq b_n(\lambda_1)$, to continue sampling if $b_n(\lambda_1) < V_n < a_n(\lambda_1)$, and to reject $\lambda = 0$ if $a_n(\lambda_1) \leq V_n$, if $\lambda_1 < 0$, and with the inequalities reversed if $\lambda_1 > 0$; we have denoted by $a_n(\lambda) = (\log A - n\lambda) / (1 - e^{\lambda})$, $b_n(\lambda) = (\log B - n\lambda) / (1 - e^{\lambda})$ where $A = \frac{1 - \beta}{\alpha}$ and $B = \frac{1 - \beta}{\alpha}$. Also obtain the operating characteristic and the average sample number of the sequential test.

Mexia (1967/68)

II.7— Form a two-sided sequential test of $\lambda = 0$ vs. $\lambda = \bar{\lambda} < 0$ and $\lambda = \bar{\bar{\lambda}} > 0$ by "intersection" of two one-sided tests as before with $\lambda_1 = \bar{\lambda}$ and $\lambda_1 = \bar{\bar{\lambda}}$, "intersection" meaning accept $\lambda = 0$ if both one-sided tests aceept it, reject it if any of the one-sided tests rejected it, and continue sampling otherwise. Obtain the special feature of the graph of the sequential probability ratio test and evaluate approximately the operating characteristic and the average sample number.

Mexia (1967/68)

II.8— Develop a sequential estimation procedure for the parameter λ for $\Lambda(x - \lambda)$.

Mexia (1967/68)

III.9— Consider the last six exercises and solve the analogous questions for the exponential distribution, which is equivalent to taking $Y = e^{-X}$; as X is $\Lambda(x - \lambda)$ then Y has the exponential distribution with zero location and $\delta = e^{-\lambda}$. The reverse method could be followed: to deduce from the results for the exponential distribution the results for $\Lambda(x - \lambda)$.

III.10— Obtain the locally most powerful one-sided tests for $\lambda = \lambda_0$ vs. $\lambda = \lambda_0$ (or $\lambda < \lambda_0$) with fixed δ ($\delta = 1$, for simplicity), for $\delta = \delta_0$ vs. $\delta = \delta_0$ (or $\delta < \delta_0$) with fixed λ ($\lambda = 0$, for simplicity); obtain the corresponding locally most powerful unbiased two-sided tests in both cases and study asymptotic behaviour.

III.11— For the Gumbel distribution consider a pair of estimators (λ^*, δ^*) of (λ, δ) and assume that $V(\lambda^*) \sim A\ \delta^2/n$, $V(\delta^*) \sim B\ \delta^2/n$, $C(\lambda^*, \delta^*) \sim C\ \delta^2/n$.

a) Compute the asymptotic efficiency for the estimation of any quantile and obtain its minimum (worst asymptotic efficiencies);

b) Choose a pair of central quantiles (of probabilities $0 < p < q < 1$); evaluate the worst asymptotic efficiency; seek the best pair (p, q), i.e., the one giving the largest value to the smallest efficiency.

III.12— For the Gumbel distribution consider an estimator (λ^*, δ^*) of (λ, δ), asymptotically standard binormal, with mean values λ and δ, and $V(\lambda^*) \sim A\ \delta^2/n$, $V(\delta^*) \sim B\ \delta^2/n$ and $C(\lambda^*, \delta^*) \sim C\ \delta^2/n$. Compute the asymptotic variance of the estimator (or predictor) $\lambda^* + \chi\ \delta^*$ of the quantile $\lambda + \chi\ \delta$ and obtain its asymptotic efficiency with respect to the estimator (or predictor) $\hat{\lambda} + \chi\ \hat{\delta}$ based on the maximum likelihood. In particular, compute this efficiency when (λ^*, δ^*) are based on two quantiles, and analyse the choice of the quantiles with different criteria (optimization for $\chi = a$, best asymptotic efficiency for $\chi \to \pm\infty$, etc).

III.13— Supposing that the samples $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$ have a Gumbel distribution with parameters (λ, δ) and (λ', δ'), determine the likelihood ratio test for $\lambda = \lambda'$, $\delta = \delta'$ and obtain its asymptotic behaviour. Using the approximate relation $(\bar{x} - \hat{\lambda})/\hat{\delta} \approx \gamma$ simplify the expression and determine its asymptotic behaviour (use the δ-method).

III.14— Suppose that $X_1, \ldots, X_n$ have a Gumbel distribution with parameters $\lambda + \beta i$ ($i = 1, \ldots, n$) and δ. Obtain the least squares estimators of the parameters (λ, β, δ) of this linear regression with additive Gumbel errors ($X_i = \lambda + \beta i + \delta Z_i$, where the Z_i are reduced Gumbel random variables) and its asymptotic behaviour, and formulate a test of $\beta = 0$ (no regression). Generalize, substituting the term βi by βc_i, with the c_i known.

II.15— Determine the equations defining the values (k_1, k_2) that minimize the mean lenght of the (β,ω) and α-average tolerance intervals for Gumbel distribution.

Themido (1985)

II.16— In the search for one-sided and two-sided tolerance intervals for Gumbel distribution obtain the first-order development of the coefficients k_i (i.e., the a such that $k_i(n) = k_i + a(n^{-1/2}) + o(n^{-1/2})$), using the δ-method.

Themido (1985)

II.17— Generate 50 plus 20 random numbers for the Gumbel (Fréchet) distribution for maxima and study, in the way used for the Weibull distribution of minima, possible predictors of maxima of the next 20 observations based on the initial sample of 50 observations.

II.18— Using the method of moments or of the two quantiles or of Downton or of blocks, obtain the test of $(\lambda , \delta) = (\lambda_0 , \delta_0)$ vs. $(\lambda , \delta) \neq (\lambda_0 , \delta_0)$.

II.19— Consider the following situation: the random variables X_{ij} ($i = 1, ..., m$, $j = 1 , ... , p$) have a Gumbel distribution with location parameters λ_{ij} and unit ($\delta_{ij} = 1$) dispersion parameter. For the sample of $n = m\,p$, test the periodicity of the λ_{ij} , i.e., $\lambda_{ij} = \lambda_j$, using the likelihood ratio criterion. Show that the statistic obtained is translation invariant in the hypothesis tested (i.e., for the transformation $x_{ij} \rightarrow x_{ij} + a_j$), and obtain its asymptotic behaviour; obtain the corresponding formulation after transformation to exponential margins.

II.20— Let $\{X_k\}$ be i.i.d. with distribution function $\Lambda((x - \lambda)/\delta)$. Now consider the dependent sequence $M_1 = x_1$, $M_k = \max (M_{k-1} , x_k) = \max (x_1, ... , x_k)$ for $k > 1$.

a) Show that the sequence $\{M_1, ..., M_n\}$ is increasing, i.e., $M_1 < M_n$, with probability $1 - 1/n$.

b) Show that

$$M(M_k) = \lambda + \delta((\gamma + \log k) , V(M_k) = \frac{\pi^2}{6}\,\delta^2 \quad \text{and for} \quad p \geq 1$$

$$C(M_k, M_{k+p}) = [\frac{\pi^2}{6} + \int_{x \leq y + \log p/k} \int (x - \gamma)\,(y + \log p/k - x)\, d\,\Lambda(x)\, d\,\Lambda(y)]\delta^2$$

$$= [\int_{x \leq y + \log p/k} \int (x - \gamma)\,(y + \log p/k) + \frac{k}{p + k}\,\log\frac{p + k}{k}]\,\delta^2 \;;$$

II.21— Obtain the least squares estimators of (λ , δ) in the previous exercise, if $M_1 < M_n$.

II.22— Obtain maximum likelihood estimators of δ for the Fréchet and Weibull distributions, supposing $\lambda = 0$ and α known. Compare these results with the estimators of λ for the Gumbel distribution with $\delta = 1$. Reduce all the results to the study of the dispersion parameter for exponential distribution.

II.23— Compute β_1 and β_2 for the Fréchet distribution.

II.24—Draw the graphs of the densities of $\alpha^{-1} \Phi'_\alpha(1 + y / \alpha)$ and verify that for $\alpha > 6$ is very close to $\Lambda'(y)$.

II.25— Compute β_1 and β_2 for the Weibull distribution of minima.

II.26—Show that if Z has the distribution function $W_\alpha(z)$ then $\text{Prob}\{\alpha(1 - Z) = y\}$ converges to $\Lambda(y)$. Draw the corresponding graphs.

II.27—Compute tables, similar to the one of Chapter V, for the shortest interval for the same probabilities and different values of the shape parameter (α) for the Fréchet and Weibull distributions.

II.28— Obtain prediction of maxima of the next m observations for a Weibull distributions of maxima.

II.29— Obtain tests and estimators using two and three central quantiles for the Gumbel and Fréchet distributions of maxima and the Weibull distribution of minima.

Tiago de Oliveira and Littauer (1976), Epstein and Kubat (1980), Neves (1986)

II.30— Determine the expression of the efficiency for the estimation of quantiles when using the method of moments.

II.31— For a Fréchet distribution (for maxima) with location parameter known (zero for convenience) and shape parameter known, obtain maximum likelihood and least squares (and approximations), homogeneous estimators (i.e., such that $\delta^*(\alpha x_i) = \alpha \delta^*(x_i)$, $\alpha > 0$) of the quantiles, homogeneous predictors, and their asymptotic distributions, as well as the behaviour of overpassing probability.

Neves (1987)

II.32— For a Weibull distribution (for minima) with location parameter known (zero for convenience) and shape parameter known, obtain maximum likelihood and least squares (and approximations), homogeneous estimators (i.e., such that $\delta^*(\alpha x_i) = \alpha \delta^*(x_i)$, $\alpha > 0$) of the

quantiles, homogeneous predictors, and their asymptotic distributions, as well as the behaviour of underpassing probability.

II.33— Compute the variance-covariance matrix of maximum likelihood estimators for the Fréchet distribution (for maxima) with $\alpha > 0$.

II.34— Compute the variance-covariance matrix of maximum likelihood estimators for the Weibull distribution (for minima) with $\alpha > 2$.

II.35— Seek the probability $p(0 < p < 1)$ such that $\lambda^* = X_1^{'} - a(Q_n(p) - X_1^{'})$, $\delta^* = b(Q_n(p) - X_1^{'})$, where $Q_n(p)$ is the sample quantile (where $[np] > 0$) and $X_1^{'}$ is the sample minimum and $a, b > 0$, are the best estimators of (λ,δ) for the Fréchet distribution with $\alpha = \alpha_0$ known, 'best' being taken either in the meaning of minimizing the generalized asymptotic variance or of minimizing the asymptotic variance of some quantile); note that the estimators are quasi-linearly invariant. Compute the efficiency of (λ^*,δ^*).

II.36— Do the same exercise for the Weibull distribution (for minima).

II.37— Seek the probabilities $p, q\ (0 < p < q < 1)$ such that for Fréchet distribution $\lambda^* = X_1^{'} - a_1(Q_n(p) - X_1^{'}) - a_2(Q_n(q) - X_1^{'})$, $\delta^* = b_1(Q_n(p) - X_1^{'}) + b_2(Q_n(q) - X_1^{'})$, $\alpha^* = g^{-1}\left(\frac{Q_n(q) - X_1^{'}}{Q_n(p) - X_1^{'}}\right)$, where $X_1^{'}$, $Q_n(p)$, $Q_n(q)$ have the usual meaning, and we suppose $0 < [np] < [nq]$, and $a_i, b_i > 0$, are such that $(\lambda^*,\delta^*, \alpha^*)$ minimizes either the generalized asymptotic variance or the asymptotic variance of some quantile; note that $(\lambda^*,\delta^*, \alpha^*)$ are quasi-linearly invariant. Compute the efficiency of $(\lambda^*, \delta^*, \alpha^*)$. $g(\alpha)$ can be defined by $\frac{Q_n(q) - X_1^{'}}{Q_n(p) - X_1^{'}} \xrightarrow{P} g(\alpha) = \frac{\chi_q}{\chi_p}$.

II.38— Do the same exercise for the Weibull distribution (for minima).

II.39— Consider the Weibull distribution for minima with $\alpha = 1/2$, i.e.,

$$W_{1/2}(x|\lambda, \delta) = 0 \qquad \text{if } x \le \lambda$$

$$W_{1/2}(x|\lambda, \delta) = 1 - \exp\left(-\left(\frac{x - \lambda}{\delta}\right)^{1/2}\right) \qquad \text{if } x \ge \lambda.$$

Show that if $\lambda = \lambda_0$ is known then $\hat{\delta} = (\frac{1}{n}\sum_1^n \sqrt{x_i - \lambda_0})^2$ is asymptotically normal. If λ and δ are unknown parameters, consider $\hat{\lambda} = \min(x_1, ..., x_n)$ and $\hat{\delta} = (\frac{1}{n}\sum_1^n \sqrt{x_i - \hat{\lambda}})^2$. $\hat{\delta}$ is a modified estimator of δ where in the previous expression we substituted λ_0 by $\hat{\lambda}$ which converges quickly to λ. Show that $(\hat{\lambda}, \hat{\delta})$ is an estimator of (λ, δ) with independent Weibull and normal margins. Use the δ-method.

Tiago de Oliveira (1982')

II.40— Consider $0 < p < q < r < 1$ such that $\beta = \frac{\log p}{\log q} = \frac{\log q}{\log r} > 1$, and define the statistic $Q^* = \frac{X'_{[nr]} - X'_{[nq]}}{X'_{[nq]} - X'_{[np]}} \xrightarrow{P} \beta^\theta$, where θ is the shape parameter in the von Mises-Jenkinson formula; note that Q^* is positive iff $[np] < [nq] < [nr]$. Then $k^* = \frac{\log Q^*}{\log \beta}$ is an estimator of θ which is asymptotically normal. Use θ^* for statistical choice between the Weibull, Gumbel and Fréchet distributions, as done in the test, and compare it with $\hat{V}_n(x)$.

II.41— Consider the choice statistic

$$\hat{V}_n(x) = \frac{1}{2}\{\sum_1^n (x_i - \bar{x})^2/\hat{\delta}^2 - n \frac{\sum_1^n ((x_i - \bar{x})/\hat{\delta})^2 e^{-(x_i-\bar{x})/\delta}}{\sum_1^n e^{-(x_i-\bar{x})/\delta}}\},$$

with

$$\hat{\delta} = \bar{x} - \frac{\sum_1^n x_i e^{-x_i/\hat{\delta}}}{\sum_1^n e^{-x_i/\hat{\delta}}}.$$

It was been shown that if $\theta = 0$ (Gumbel distribution) $\hat{V}_n(x)/\sqrt{n}\,\hat{\delta}$ is asymptotically standard normal. Consider then the following modified approximate statistics :

a) as $\frac{1}{n}\sum_1^n e^{-(x_i-\bar{x})/\delta} \approx M(e^{-(x-(\lambda+\gamma\delta))/\delta}) = e^\gamma$, study

$$\hat{V}_{n1}(x) = \frac{1}{2}\left\{ \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{\hat{\delta}^2} - e^{-\gamma} \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{\hat{\delta}^2} e^{-(x_i-\bar{x})/\delta} \right\}$$

and show its asymptotically normality;

b) as $\frac{1}{n}\sum_{1}^{n} \frac{(x_i - \bar{x})^2}{\hat{\delta}^2} e^{-(x_i - \bar{x})/\hat{\delta}} \approx M\left\{\left(\frac{x-(\lambda + \gamma\delta)}{\delta}\right)^2 e^{-(x-(\lambda+\gamma\delta))/\delta}\right\} = \frac{\pi^2}{6}$ study the statistic $\hat{V}_{n,2} = \frac{n}{2}(s^2/\hat{\delta}^2 - \pi^2/6)$ and show its asymptotic normality; note that this is a comparison between the estimates $\hat{\delta}^2$ and $\delta^{*2} = \frac{6}{\pi^2} s^2$, the first approximation to $\hat{\delta}$.

c) Study the statistic $\hat{V}_{n,3}$, obtained by substituting $\hat{\delta}$ by δ^* in $\hat{V}_{n,1}$, and show its asymptotic normality;

d) Study the statistic $\hat{V}_{n,4}$ when (λ, δ) have been estimated by the method of moments, i.e., $\lambda^* = \bar{x} - \gamma\delta^*$, $s^2 = \frac{6}{\pi^2}\delta^{*2}$, and show its asymptotic normality;

e) Compare the efficiencies of these statistics with respect to $\hat{V}_n$.

II.42— Compare the Prob{accept k = 0|k} for the statistic $\hat{V}_n(x)$ and for the statistic Q_n for the same significance level (α = .05, .02, .01, .) and different values of k

II.43— Analyse also the same question for the statistics given in exercise II.41.

II.44— Compute Δ_n and $\tilde{P}_n'(0)$, or its orders, for the approximation of order n^{-2} for the solution of the statistical trilemma of the locally optimal approach.

II.45—Show that efficiency of largest observations vs. subsamples methods changes with the value of the shape parameter k of the von Mises-Jenkinson formula, which shows some lack of robustness.

II.46— For the techniques with partial information, seek the optimal quantiles (for quantile estimation and for block estimation) that minimize the asymptotic variance of the estimators (λ^*, δ^* or α^* if it exists) and compare with those with maxima (global) asymptotic variance.

II.47— Owing to the importance of λ in $\Phi_\alpha(\frac{x-\lambda}{\delta})$ and $W_\alpha(\frac{x-\lambda}{\delta})$, which can be estimated by $X_1' = \min_1^n X_i$, consider the (new) random variables $Y_i = X_i - X_1'$ ($\min_1^n Y_i = 0$).

Suppose that they have distributions $\Phi_\alpha(y/\delta)$ and $W_\alpha(y/\delta)$ (which they have approximately), and study the estimators $(\alpha^{**}, \delta^{**})$ of (α, δ) based on the positive Y_i (in number of $n-1$ with probability one).

Obtain sufficient conditions for the statistics (α^*, δ^*) to be estimators of (α, δ). Recall that $\Phi_\alpha(y/\delta)$ and $W_\alpha(y/\delta)$ can be easily transformed to have a Gumbel distribution.

II.48— Develop the theory of "excesses over thresholds" for the limiting distributions of maxima both for the P.O.T. and A.E.O.T. methods.

L. J. BAIN (1973). Results for one or more independent censored samples for Weibull or extreme-value distribution, *Technometrics*, 15, 279-288

G. BLOM (1958). *Statistical Estimation and Transformed Beta-Variables*, Wiley, New York

D. D. BOOS (1983). Using extreme value theory to estimate large percentiles, Techn. Rep., Dept. Statist., *North Carolina State University*

H. CHERNOFF and G. J. LIEBERMAN (1954). Use of normal probability paper, *J. Amer. Statist. Soc.*, 49, 778-785

H. CHERNOFF, L. GASTWIRTH and M. V. JOHNS (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation, *Ann. Math. Statist.*, 38, 52-71

A. C. COHEN Jr and B. J. WITTEN (1982). Modified maximum likelihood and modified moments estimators for the three-parameter Weibull distribution, *Commun. Statist. - Theor. Meth.*, 11, 2631-2656

H. CRAMÉR (1946). *Mathematical Methods of Statistics*, Princeton University Press, New Jersey

H. A. DAVID (1981). *Order Statistics* (2nd ed), Wiley, New York

ANTHONY C. DAVISON (1984). Modelling excesses over high thresholds, with an application, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 461-482, D. Reidel, Dordrecht

ANTHONY C. DAVISON and R.L. SMITH (1990). Models for exceedance over high thresholds, (with discussion), *J. Roy. Statist. Soc.*, 8, 52, 393-442

A.L.M. DEKKERS, J.H. J. EINMAHL and L. de HAAN (1989). A moment estimator for the index of an extreme value distribution, *Ann. Statist*, 17, 1883-1855

A.L.M. DEKKERS and L. de HAAN (1989). On the estimation of the extreme value index and large quantile estimation, *Ann. Statist*, 17, 1795-1832

A.L.M. DEKKERS and L. DE HAAN (1991). Optimal choice of the sample fraction in extreme – value estimation, *Centre for Mathematics and Computer Science* , Amsterdam

JEAN-PIERRE DIAZ (1985). Estimation et prédiction de valeurs extrêmes dans le domaine d´atraction de la loi de Fréchet, *C. R. Acad. Sc. Paris*, t. 301, ser. I, 541-544

V. DIJK and L. DE HAAN (1990). On the estimation of the exceedance probability of a high level, Report 9051/A , Econometric Institute, *Erasmus University* , Rotterdam

F. DOWNTON (1966). Linear estimates of parameters in the extreme value distribution, *Technometrics*, 8, 3-17

S. D. DUBEY (1966). Hiper-efficient estimators of the location parameter of Weibull law, *Naval Res. Logist. Quat.*, 13, 253-264

S. D. DUBEY (1967). Revised tables for asymptotic efficiencies of the moment estimators for the parameters of the Weibull laws, *Naval Res. Logist. Quat.*, 14, 261-267

S. D. DUBEY (1967). Some percentile estimators for Weibull parameters, *Technometrics*, 9, 112-129

S. D. DUBEY (1967'). On some permissible estimators of the location parameter of the Weibull and certain other distributions, *Technometrics*, 9, 293-307

J.M.J. EINMAHL (1990). The empirical distribution function as a tail estimator, *Statist Neerland.*, 44, 79-82

B. EPSTEIN and M. SOBEL (1954). Some theorems relevant to life testing from an exponential distribution, *Ann. Math. Statist.*, 25, 373-381

R. A. FISHER and L.H.C. TIPPET (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Proc. Camb. Phil. Soc.*, 24, 180-190

A. FRANSÉN and J. TIAGO DE OLIVEIRA (1984). Statistical choice of univariate extreme models, part II, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 373-394, D. Reidel, Dordrecht

A.M. FREUDENTHAL and E.J. GUMBEL (1953). Statistical interpretation of fatigue tests, *Proc. Roy. Soc. London, A*, 216, 309-332

J. GALAMBOS (1980). A statistical test for extreme value distributions, *Coll. Math. Soc. Janos Bolyai*, # 32, 221-230, Budapest

B. V. GNEDENKO (1943). Sur la distribution limite du terme maximum d´une série aléatoire, *Ann. Math.*, 44, 423-453

M.I. GOMES (1978). Some probabilistic and statistical problems in extreme value theory, *University of Sheffield*, Ph. D. Thesis

M.I. GOMES (1981). An i-dimensional limiting distribution of largest values and its relevance to the statistical theory of extremes, in *Statistical Distributions in Scientific Work*, vol.6, C. Taillie *et al* eds., 389-410, D. Reidel, Dordrecht

M.I. GOMES (1982). A note on statistical choice of extremal models, *Act. IX Jorn. Matem. Hispano-Lusas*, Salamanca, II, 653-656

M.I. GOMES (1984). Concomitants in a multidimensional extreme model, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 353-354, D. Reidel, Dordrecht

E. J. GUMBEL (1954). *Statistical Theory of Extreme Values and Some Practical Applications*, Nat. Bur. Standards, Appl. Math. Series, # 33, Washington

E. J. GUMBEL (1955). The calculated risk in flood control, *Appl. Sc. Research* (Holland), 5, A, 273-280.

E. J. GUMBEL (1958). *Statistics of Extremes*, Columbia University Press, New York.

E. J. GUMBEL (1962). Statistical theory of extreme values (main results), in *Contributions to Order Statistics,* Ahmed E. Sarhan and Bernard G. Greenberg eds., 56-93, Wiley, New York

E. J. GUMBEL (1965). A quick estimation of the parameters in Fréchet distribution, *Rev. Int. Statist. Inst.*, 33, 349-363

E. J. GUMBEL and G. J. LIEBERMAN (1954). Some applications of extreme value methods, *Amer. Statist.*, 8, 14-17

L. DE HAAN (1971). *On regular variation and its applications to weak convergence of sample extremes, Mathmatische Centrum,* Amsterdam

P. HALL (1982). On estimating the end-point of a distribution, *Ann. Statist.*, 10, 556-568

H. L. HARTER and S. D. DUBEY (1967). Theory and tables for tests of hypothesis concerning the mean and the variance of a Weibull population, *Aerospace Res. Labs. Rep.*, 67-0059, Wright-Paterson AFB, Ohio

K. M. HASSANEIN (1964). Estimation of the parameters of the extreme value distribution by order statistics, *University of North Carolina*, Chapell Hill

K. M. HASSANEIN (1968). Analysis of extreme value data by sample quantiles for very large samples, *J. Amer. Statist. Assoc.*, 63, 877-888

L. HERBACH (1970). An exact asymptotically efficient confidence bound for reliability in the case of Gumbel population, *Technometrics*, 12, 700-701

L. HERBACH (1970). Introduction, Gumbel model, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 49-80, D. Reidel, Dordrecht

B. M. HILL (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, 3, 1163-1174.

J. HÜSLER and M. SCHÜPBACH (1986). On simple block estimators for the parameters of the extreme-value distribution, *Commun. Statist.-Simul. Comput.*, 15, 61-76

J. HÜSLER and J. TIAGO DE OLIVEIRA (1988). The usage of largest observations for parameter and quantile estimation for Gumbel distribution; an efficiency analysis, *Publ. Inst. Statist. Univ. Paris*, XXXIII, # 1, 41-56

A. F. JENKINSON (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quat. J. Roy. Meteor. Soc.*, 87, 158-171

M. V. JOHNS Jr and G. J. LIEBERMAN (1966). An exact asymptotically efficient confidence bound for reliability in the case of Weibull distribution, *Technometrics*, 8, 135-175

R.F. KAPPENAM (1981). Estimation of Weibull location, *J. Statist. Comp. Simul.*, 13, 245-254

B. F. KIMBALL (1949). An approximation to the sampling variance of an estimated maximum value of given of given frequency based on the fit of doubly exponential distribution of maximum values, *Ann. Math. Statist.*, 27, 758-767

P. KUBAT (1982). Simple large sample estimators of scale and location parameters based on blocks of order statistics, *Trab. Estad. Inv. Oper.*, 33, 86-118

P. KUBAT and B. EPSTEIN (1980). Estimation of quantiles of location scale distribution based on 2 or 3 order statistics, *Trab. Estad. Inv. Oper.*, 22, 575-581

R. V. LAUE (1974). Unpublished paper, *Bell Labs.*, New Jersey

M. R. LEADBETTER, GEORG LINDGREN and HOLGER ROOTZÉN (1982). *Extremes and Related Properties in Random Sequences and Processes*, Springer-Verlag, Heidelberg

G. H. LEMON (1975). Maximum likelihood estimation of the three parameters Weibull distribution based on censored samples, *Technometrics*, 17, 247-254

J. LIEBLEIN (1953). On the exact evaluation of the variance and covariance of order statistics in samples from the extreme value distribution, *Ann. Math. Statist.*, 24, 282-287

J. LIEBLEIN (1954). A new method of analysing extreme value data, *Nat. Advisory Comm. for Aviation*, Techn. Note # 3053, Washington

J. LIEBLEIN and M. ZELEN (1956). Statistical investigation of the fatigue life of deep groove ball bearings, *J. Res. Nat. Bur. Stand.*, 57, 273-316

E. H. LOYD (1962). Generalized least squares theorem, in *Contributions to Order Statistics*, Ahmed E. Sarhan and Bernard G. Greenberg eds., 20-27, Wiley

NANCY R. MANN (1967). Results on the location and scale parameter estimation with applications to extreme-value distribution, *Aerospace Res. Labs.*, Rep 67-0023, Wright-Patterson AFB, Ohio

NANCY R. MANN (1967). Tables for obtaining the best linear invariant estimates of the parameters of the Weibull distributions, *Technometrics*, 9, 629-645

NANCY R. MANN (1968). Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions, *Technometrics*, 10, 231-256

NANCY R. MANN and KENNETH W. FERTIG (1973). Tables for obtaining Weibull confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme-value distribution, *Technometrics*, 15, 87-101

NANCY R. MANN and KENNETH W. FERTIG (1975). A goodness of fit test for the two parameters vs three parameters Weibull; confidence bounds for threshold, *Technometrics*, 17, 237-245

DAVID MEJZLER (1949). On a theorem of B. V. Gnedenko, *Sb. Trudov Inst. Math. Akad. Nauk. Ukrain . RSR*, 12, 31-35 (in russian)

J. T. MEXIA (1967). Studies on the extreme double exponential distribution-II, Sequential estimation and testing for the location parameter of Gumbel distribution, *Rev. Fac. Ciências de Lisboa*, A, XII, 5-14

M. A. J. VAN MONTFORT (1973). An asymmetric test on the type of the distribution of extremes, *Medelelugen Landbbouwhoge-School Wageningen*, # 73-18, 1-15

M. A. J. VAN MONTFORT and ALBERT OTTEN (1978). On testing a shape parameter in the presence of a scale parameter, *Math. Operations Forsch. Statist., Ser. Statistics*, 9, 91-104

MANUELA NEVES (1986). Estimação de quantis das distribuições de Gumbel e Fréchet com parâmetros de escala e de localização desconhecidos, baseados em duas estatísticas ordinais, *Centro de Estatística e Aplicações*, # 25/86, Lisboa

MANUELA NEVES (1987). Predição homogénea para a distribuição de Fréchet, in publ. *Act. XII Jorn. Matem. Luso-Espanholas* , Braga

MANUELA NEVES (1988). Estimação por blocos dos parâmetros da distribuição de Fréchet, *Universidade Nova de Lisboa*, Tese de Doutoramento

J. OGAWA (1951). Contributions to the theory of systematic statistics, *Osaka Math. J.*, 3, 175-213

G.M. PANCHANG and V. P. AGGARWAL (1962). Peak flow estimation by the method of maximum likelihood, Techn. Mem. HLO 2, *Central Water and Power Research Station*, Poona, India

D. PASSOS COELHO and T. PINHO GIL (1963). Studies on extreme double exponential distribution, I - The location parameter, *Rev. Fac. Ciências de Lisboa*, A, 9, 37-46

J. PICKANDS III (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, 3, 119-131

J. PICKANDS III (1989). Consistent estimation of extreme quantiles, Preprint, *University of Pennsylvania*

E.C. POSNER (1965). The applications of extreme value theory to error free communication, *Technometrics*, 7, 517-519.

G. RAMACHANDRAN (1974). Extreme value theory and large fire losses, *The Astin Bull.*, VII, 293-310.

R. D. REISS (1987). Estimating the tail index of a claim size distribution, *Bläter Deutschen Gesllsch. Versicher. Math.*, 18, 21-25

R.- D. REISS (1989). *Approximate Distribution of Order Statistics*, Springer-Verlag

D. ROSBJERG and J. KNUDSEN (1984). POT-estimation of extreme sea states and the benefit of using wind data, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 611-620, D. Reidel, Dordrecht

AHMED E. SARHAN and BERNARD G. GREENBERG (eds.) (1962). *Contributions to Order Statistics*, Wiley, New York

M. SCHÜPBACH and J. HÜSLER (1983). Simple estimators for the parameters of the extreme-value distribution based in censored data, *Technometrics*, 25, 189-192.

R. L. SMITH (1984). Threshold methods for sample extremes, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 621-638, D. Reidel, Dordrecht

R. L. SMITH (1986). Extreme value theory based on the r largest annual events, *J. Hydrology*, 86, 27-43

R. L. SMITH (1987). Estimating tails of probability distribution, *Ann. Statist,* 15, 1174-1207

R. L. SMITH (1989). Approximations in extreme value theory, Preprint, Centre for Stochastic Processes, *University of North Carolina, at Chappel Hill*

R. L. SMITH and I. WEISMANN (1985). Maximum likelihood estimation of the lower tail of a probability distribution, *J. Roy. Statist. Soc.*, B, 47, 285-298

R. L. SMITH and I. WEISMANN (1987). Large deviations of tail estimators based on Pareto distribution, *J. Appl. Prob.*, 24, 619-630.

R. SNEYERS (1977). L'intensité maximale des précipitations en Belgique, *Inst. Royal Météor. Belgique*, ser. B, nº 86

H. C. S. THOM (1960). Distribution of extreme winds in the United States, *J. Structural Div., Proc. ASCE*, 86

H. C. S. THOM (1971). Asymptotic extreme-value distributions of wave heights in the open ocean, *J. Marine Res.*, 29, 19-27

M. TERESA THEMIDO (1985). Intervalos de tolerância equivariantes para a distribuição de Gumbel, *Universidade de Lisboa*, Tese de Mestrado

J. TIAGO DE OLIVEIRA (1963). Decision results for the parameters of the extreme value distribution (Gumbel) based on the mean and the standard deviation, *Trab. Estad. Inv. Oper.*, XIV, 61-81, Madrid

J. TIAGO DE OLIVEIRA (1966). Quasi-linearly invariant prediction, *Ann. Math. Statist.*, 37, 1684-1687

J. TIAGO DE OLIVEIRA ed., (1966'). Statistical extremes; general introduction; preprint of the *NATO ASI on Engineering Applications of Statistics Extremes*, Faro

J. TIAGO DE OLIVEIRA (1968). Efficiency evaluation for quasi-linear invariant predictors, *Rev. Belge Statist. Rech. Oper.*, 9, 1-9

J. TIAGO DE OLIVEIRA (1972). Statistics for Gumbel and Fréchet distributions, in *Structural Safety and Reliability*, A. M. Freudenthal ed., 91-105

J. TIAGO DE OLIVEIRA (1973). Quasi-linear discrimination, in *Discriminant Analysis and Applications*, T. Cacoullos ed., 291-309, Academic Press, New York

J. TIAGO DE OLIVEIRA (1975). Statistical decision for extremes, *Trab. Estad. Inv. Oper.*, XXVI, 453-471, Madrid

J. TIAGO DE OLIVEIRA (1977). Asymptotic distributions of univariate and bivariate m-th extremes, in *Recent Developments in Statistics*, J. R. Barra *et al.* eds., 613-619, North-Holland, Amsterdam

J. TIAGO DE OLIVEIRA (1981). Statistical choice of univariate extreme models, in *Statistical Distributions in Scientific work*, C. Taillie *et al.* eds., vol.6, 367-387, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1982). Efficient estimation for quantiles of Weibull distributions, *Rev. Belge Statist. Rech. Oper.*, 22, 3-10

J. TIAGO DE OLIVEIRA (1982). A definition of estimator efficiency in k-parameter case, *Ann. Inst. Statist. Math.*, 34 A, 411-421

J. TIAGO DE OLIVEIRA (1982'). The δ-method for obtention of asymptotic distributions, applications, *Publ. Inst. Statist. Univ. Paris*, XXVII, 49-70

J. TIAGO DE OLIVEIRA (1983). Gumbel distribution, in *Encyclopedia of Statistical Sciences*, III, N.L. Johnson and S. Katz eds., 552- 558, Wiley, New York

J. TIAGO DE OLIVEIRA (1984). Initiation of statistical decision for Weibull distribution, *Cuad. Bioestad. Aplic. Informat.*, 2, 495-499

J. TIAGO DE OLIVEIRA (1984'). Probability paper and plotting positions - a new approach, *XIV Reunion Nacional SEIOEI*, I, 175-186, Granada

J. TIAGO DE OLIVEIRA (1984"). Univariate extremes: statistical choice, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 91-107, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1984"'). Weibull distributions and large earthquake modelling, in *Probabilistic Methods in the Mechanics of Solids and Structures* (IUTAM Symposium in honour of Dr. W. Weibull), 81-89, S. Eggwertz and N.C. Lind eds., Springer-Verlag, Heidelberg

J. TIAGO DE OLIVEIRA (ed.) (1984""). *Statistical Extremes and Applications*, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1986). Statistical choice of univariate extreme models, (summary), *Mem. Acad. Ciências Lisboa, cl. Ciências*, XXVII, 107-110

J. TIAGO DE OLIVEIRA (1991). Largest values and tail estimation, statistical choice, to be publ in *Mem. Acad. Ciências Lisboa*

J. TIAGO DE OLIVEIRA (1991'). Tail Estimation and Extremes, to be publ in *Homenage to Ahmed A. Sarhan, Alexandria, Egipt*

J. TIAGO DE OLIVEIRA and M. I. GOMES (1984). Two test statistics for choice of univariate extremes models, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA and S. B. LITTAUER (1976). Mean square invariant forecasters for the Weibull distributions, *Naval Res. Logist. Quat.*, 23, 487-511

W. WEIBULL (1939). A statistical theory of the strength of materials, *Proc. Roy. Swed. Inst. Eng. Res.*, # 151, Stockholm

I. WEISSMAN (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, 73, 812-815

I. WEISSMAN (1981). Confidence intervals for the threshold parameter, *Commun. Statist.-Theor. Meth.*, A10, 549-557

I. WEISSMAN (1981'). Confidence intervals for the threshold parameter-II: unknown shape parameters, *Commun.Statist.-Theor. Meth.*, 11, 2461-2467

I. WEISSMAN (1984). Statistical estimation in extreme value theory, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 109-115, D. Reidel, Dordrecht

S. S. WILKS (1941). Determination of a sample sizes for setting tolerance limits, *Ann. Math. Statist.*, 12, 91-96

M. WOODROFE (1974). Maximum likelihood estimation of a translation parameter of a truncated distribution - II. *Ann. Statist.*, 3, 474-448

T.M. YEGULALP and J. T. KUO (1974). Statistical prediction of the occurrence of maximum magnitude earthquakes, *Bull. Seismol. Soc. America*, 64, 393-414

J. WYCKOFF, L. J. BAIN and M. ENGELHARDT (1979) and (1981). Some complete and censored sampling results for the three-parameter Weibull Distribution. *J. Statist. Comp. Simul.*, 11, 139-151 and 13, 149

S. M. ZANAKIS (1979). A simulation study of some simple estimators for the three-parameter Weibull distribution, *J. Statist. Comp. Simul.*, 9, 101-116

S. M. ZANAKIS and NANCY R. MANN (1982). A good simple percentile estimator of the Weibull shape parameter for use when all three parameters are unknown, *Naval Logist. Res. Quat.*, 29, 419-428

S. M. ZANAKIS and J. KYPARISIS (1986). A review of maximum likelihood estimation methods for the three-parameter Weibull distribution, *J. Statist. Comp. Simul.*, 25, 53-73

PART III

Multivariate extremes

Chapter X

BIVARIATE EXTREMES; PROBABILISTIC RESULTS

Introduction

Bivariate asymptotic distributions of maxima are useful for the analysis of many concrete problems such as the greatest ages of death for men and women, each year, whose distribution, naturally, splits in the product of the margins, by independence; floods at two different places on the same river, each year; bivariate extreme meteorological data (pressures, temperature, wind velocity, etc.), each week; largest waves, each week, etc.. The same can be said for the study of minima. Extensions to multivariate distributions, more complex, will be made later.

Evidently, the aim of study of asymptotic distributions of bivariate extremes is to obtain the asymptotic probabilistic behaviour, and also to provide bivariate models of (asymptotic) extremes that fit observed data. But, as will seen in the next chapter, only some problems have been solved and the methods founds so far cover a much smaller area than the theory for univariate extremes.

The distribution of bivariate extremes

The theory of limiting bivariate (and multivariate) extremes, to a certain extent, follows the same lines of limiting univariate extremes. For simplicity of the exposé we now consider the bivariate case; the multivariate case will appear as an extension in a following chapter.

Let $(X_1 , Y_1), \ldots , (X_k , Y_k), \ldots$ be a sequence of i.i.d. random pairs with distribution function $F(x , y) = \text{Prob}\{X \leq x , Y \leq y\}$ and survival function $S(x , y) = \text{Prob}\{X > x , Y > y\} = 1 + F(x , y) - F(x , + \infty) - F(+ \infty, y)$; the inversion of the last relation expresses $F(x , y)$ in $S(x , y)$ by $F(x , y) = 1 + S(x , y) - S (x , - \infty) - S(- \infty, y)$ because, as known, we have $F(x , + \infty) + S(x , - \infty) = 1$ and $F(+ \infty, y) + S(- \infty, y) = 1$. The survival function plays the same rôle for minima as the distribution function played for maxima.

The random pair $(\max_1^k X_i , \max_1^k Y_i)$ has the distribution function

$$F_k(x , y) = \text{Prob}\{\max_1^k X_i \le x, \max_1^k Y_i \le y \} = F^n(x , y) .$$

The probability of $(\max_1^k X_i , \max_1^k Y_i)$ being an observed point is, evidently,

$$\pi_k = k \, \text{Prob}\{\max_1^k X_i = X_1 , \max_1^k Y_i = Y_1\} = k \int_{-\infty}^{+\infty} \int F^{k-1}(x , y) \, d F (x , y) .$$

If $F(x , y)$ splits into its margins $F(x , y) = F(x , + \infty) . F (+ \infty, y))$, we have $\pi_k = 1/k$. It is well known (see Fréchet (1951)) that $\max (0 , F(x , + \infty) + F (+ \infty, y) - 1) \le F(x , y) \le \min (F(x , + \infty) , F (+ \infty, y))$ with the bounds attained; if $F(x , y)$ is equal to the LHS we have $\pi_k = 0$, and for the RHS we get $\pi_k = 1$ because in the first case we have $F(x , + \infty) + F (+ \infty, y)) = 1$ with probability 1, and in the second we have $F(x , + \infty) = F (+ \infty, y)$ with probability 1.

For the random pair $(\min_1^k X_i , \min_1^k Y_i)$ we have $S_k(x , y) = S^k(x , y)$ in an analogous way,

and the distribution function of the pair of minima is

$$\text{Prob}\{\min_1^k X_i \le x, \min_1^k Y_i \le y\} = 1 + S^k(x , y) - S^k(x , - \infty) - S^k(- \infty , y) =$$

$$1 + (1 + F(x , y) - F(x , + \infty) - F(+ \infty, y))^k - (1 - F(x , + \infty))^k - (1 + F(+ \infty, y))^k .$$

The probability π'_k of $(\min_1^k X_i , \min_1^k Y_i)$ being an observed point is also $1/k$ for the

independence case and 0 and 1 if $F(x , y)$ is equal to the LHS or RHS of the corresponding Fréchet inequality.

As before we will stick essentially to the study of maxima, the conversion to minima being immediate by the relations $\min_1^k X_i = -\max_1^k (-X_i)$ and $\min_1^k Y_i = -\max_1^k (-Y_i)$; note that the distribution function of $(-X, -Y)$ is $\text{Prob}\{-X \le x, -Y \le y\} = \text{Prob}\{X \ge -x, Y \ge -y\} = S(-x, -y)$ in the continuity points (with right-continuity continuation in the discontinuity set) and thus, here, all over the plane because the limiting distributions are continuous.

Note, as an hint for applications, if (X_i, Y_i) are floods of the same river at two different points they must be dealt with as maxima; but if (X_i, Y_i) are droughts at the same two points they must be dealt with as minima , i.e. $(-X_i, -Y_i)$ will be dealt with as maxima.

The limiting distribution of maxima

An important question, both theoretical and applied, is to find out, analogously, as was done in the univariate case, if there exist positive linear transforms of $(\max_1^k X_i, \max_1^k Y_i)$ such that the reduced random pair

$$\left(\left(\max_1^k X_i - \lambda_k\right)/\delta_k , \left(\max_1^k Y_i - \lambda'_k\right)/\delta'_k\right),$$

whose distribution function is

$$\text{Prob}\left\{\left(\max_1^k X_i - \lambda_k\right)/\delta_k \le x , \left(\max_1^k Y_i - \lambda'_k\right)/\delta'_k \le y\right\} = F^k(\lambda_k + \delta_k x , \lambda'_k + \delta'_k y) \xrightarrow{w} L(x, y),$$

i.e., has a limiting, non-degenerate and proper distribution.

As a consequence we know that $F^k(\lambda_k + \delta_k x , +\infty) \xrightarrow{w} L(x , +\infty)$ and $F^k(+\infty, \lambda'_k + \delta'_k y) \xrightarrow{w} L(+\infty, y)$ — see Deheuvels (1984).

Evidently, as we dealing with maxima, the marginal distribution functions $L(x , +\infty)$ and $L(+\infty, y)$ must be of any of the three forms — Weibull, Gumbel or Fréchet. As they are easily translated from one to another, as seen before, we will suppose that the margins have a reduced Gumbel distribution function, i.e., $L(x , +\infty) = \Lambda(x)$, $L(+\infty , y) = \Lambda(y)$. As the margins are

continuous, so is $L(x , y)$ and, thus, the convergence $F^k(\lambda_k + \delta_k\ x\ ,\ \lambda'_k + \delta'_k\ y) \xrightarrow{w} L\ (x, y)$ is also uniform. For notational convenience, we will denote by $\Lambda(x , y)$ the limiting distribution functions $L(x , y)$ when the margins have a reduced Gumbel distribution function.

The special case where we have minima with an reduced exponential distribution will also be considered, sometimes by duality, but the kernel of the Chapter, as well as the next ones, is the case of maxima with reduced Gumbel margins.

Let us show that $\Lambda(x , y) = e^{-(e^{-x}+e^{-y})k(y-x)}$ where $k(.)$, called the *dependence function*, must satisfy some conditions to be seen later.

The proof of this result is simple.

As $F^k(\lambda_k + \delta_k\ x\ ,\ \lambda'_k + \delta'_k\ y) \to \Lambda\ (x, y)$ also

$F^{km}(\lambda_{km} + \delta_{km}\ x\ ,\ \lambda'_{km} + \delta'_{km}\ y)) \to \Lambda\ (x, y)$ or

$F^k(\lambda_{km} + \delta_{km}\ x\ ,\ \lambda'_{km} + \delta'_{km}\ y)) \to \Lambda^{1/m}\ (x, y)$.

By Khintchine's convergence of types theorem we know that there exist coefficients α_m, $\beta_m > 0$, $\alpha'_m , \beta'_m > 0$ such that

$$\Lambda^{1/m}\ (x, y) = \Lambda\ (\alpha_m + \beta_m\ x\ ,\ \alpha'_m + \beta'_m\ y) .$$

Using this relation for the margins (i.e., putting $y = +\infty$ or $x = +\infty$) we get $\Lambda\ (\alpha_m + \beta_m\ x) = \Lambda^{1/m}\ (x)$, $\Lambda\ (\alpha'_m + \beta'_m\ y) = \Lambda^{1/m}\ (y)$ and thus $\alpha_m = \alpha'_m = \log m$, $\beta_m = \beta'_m = 1$, and so

$$\Lambda(x, y) = \Lambda^m\ (\log m + x\ ,\ \log m + y) .$$

Let us take $m = [k\ t]$, $t > 0$, the integer part of $(k\ t)$. We have

$$\Lambda(x, y) = \Lambda^{[kt]}\ (\log\ [k\ t] + x\ ,\ \log\ [k\ t] + y) .$$

Substituting $x - \log k$ for x and $y - \log k$ for y , we get

$$\Lambda(x - \log k\ ,\ y - \log k) = \Lambda^k\ (x\ ,\ y)$$

and also $\Lambda(x - \log k\ ,\ y - \log k) = \Lambda^{[kt]}\ (\log\ [k\ t] - \log k + x\ ,\ \log\ [k\ t] - \log k + y)$

so we have $\Lambda(x\ ,\ y)^{k/[kt]} = \Lambda(\log\ [k\ t]/k + x\ ,\ \log[k\ t]/k + y)$ and

letting $k \to \infty$, by continuity, we get the final stability relation $(t > 0)$:

$$\Lambda(x, y) = \Lambda^t(\log t + x, \log t + y).$$

For $x + \log t = 0$ it gives

$$\Lambda(x, y) = [\Lambda(0, y - x)]^{e^{-x}},$$

which can be put under the form

$$\Lambda(x, y) = \exp\{-(e^{-x} + e^{-y})\, k(y - x)\}$$

as said before. The dependence function expresses the probabilistic inter-relation, or association, between the margins.

An actual characterization of the distribution function $\Lambda(x, y)$ can be made in the following way. It is immediate that a random pair (X, Y) with distribution function $\Lambda(x, y)$ is such that $\max(X + a, Y + b)$ has a Gumbel distribution function with a location parameter. But, also, the converse is true. In fact putting $z - a = x$, $z - b = y$ we have $L(x, y) = \Lambda(z - \varphi(z - x, z - y))$ independent of z, from which we get $\varphi(\xi, \eta) = \xi + \varphi(0, \eta - \xi)$, and then $(1 + e^{-w})\, k(w) = e^{\varphi(0,-w)}$. Thus:

$\Lambda(x, y)$ is the only distribution function $L(x, y)$ such that $Prob\{\max(X + a, Y + b) \le z\} = L(z - a, z - b) = \Lambda(z - \varphi(a, b))$.

Another approach may be the following: let us call the *structure function* of a bivariate distribution function $F(x, y)$ with continuous margins $A(x) = F(x, +\infty)$ and $B(y) = F(+\infty, y)$ the function $\bar{S}(\xi, \eta)$ such that

$$\bar{S}(A(x), B(y)) = F(x, y);$$

$\bar{S}(\xi, \eta)$ is a bivariate distribution function with uniform margins, defined in the unit square $[0, 1] \times [0, 1]$ and is continuous.

Then any structure function of the limiting distribution functions of bivariate extremes verifies the functional equation $(0 < \omega < +\infty)$

$$\bar{S}^{\omega}(\xi, \eta) = \bar{S}(\xi^{\omega}, \eta^{\omega});$$

Let S_0 denote the (initial) structure function of the distribution function with margins $A(x)$ and $B(y)$ and suppose, as before, that the marginal limiting distributions are $\Lambda(x)$ and $\Lambda(y)$, with $\lambda_k, \delta_k > 0$, $\lambda'_k, \delta'_k > 0$, as a system of attraction coefficients. We then have

$$\xi_k = F^k(\lambda_k + \delta_k x, +\infty) \to \Lambda(x)$$

$$\eta_k = F^k(+\infty, \lambda'_k + \delta'_k y) \to \Lambda(y)$$

and $F^k(\lambda_k + \delta_k x, \lambda'_k, \delta'_k y) = \bar{S}_0^k(\xi_k^{1/k}, \eta_k^{1/k})$. Let us now show that $\bar{S}_0^k(\xi_k^{1/k}, \eta_k^{1/k})$ has a limit $\bar{S}(\xi, \eta)$ when $\xi_k \to \xi$, $\eta_k \to \eta$ if and only if $\bar{S}_0^k(\xi^{1/k}, \eta^{1/k}) \to \bar{S}(\xi, \eta)$. The 'if' part is obvious; let us show the converse.

As $\xi_k \to \xi$, $\eta_k \to \eta$ for $\eta > N(\varepsilon)$, we have

$$|\xi_k - \xi| \le \varepsilon, \ |\eta_k - \eta| \le \varepsilon$$

and, thus,

$$(\xi - \varepsilon)^{1/k} \le \xi_k^{1/k} (\xi + \varepsilon)^{1/k}$$

$$(\eta - \varepsilon)^{1/k} \le \eta_k^{1/k} (\eta + \varepsilon)^{1/k}$$

so that

$\bar{S}_0^k((\xi - \varepsilon)^{1/k}, (\eta - \varepsilon)^{1/k}) \le \bar{S}_0^k(\xi_k^{1/k}, \eta_k^{1/k}) \le \bar{S}_0^k((\xi + \varepsilon)^{1/k}, (\eta + \varepsilon)^{1/k})$, and every possible limit of $\bar{S}_0^k(\xi_k^{1/k}, \eta_k^{1/k})$ is between $\bar{S}(\xi - \varepsilon, \eta - \varepsilon)$ and $\bar{S}(\xi + \varepsilon, \eta + \varepsilon)$; owing to the continuity of $\bar{S}(\xi, \eta)$ we have the desired result.

We must now characterize $\bar{S}(\xi, \eta)$. As $\bar{S}_0^k(\xi^{1/k}, \eta^{1/k}) \to \bar{S}(\xi, \eta)$ and $\bar{S}_0^{k'k}(\xi^{1/k'k}, \eta^{1/k'k}) \to \bar{S}(\xi, \eta)$, putting $\xi' = \xi^{1/k'}$, $\eta' = \eta^{1/k'}$ we have

$$\bar{S}^{k'k}(\xi^{1/k'k}, \eta^{1/k'k}) = [\bar{S}^k(\xi'^{1/k}, \eta'^{1/k})]^{k'},$$

and letting $k \to \infty$ we have

$$\bar{S}(\xi, \eta) = \bar{S}^{k'}(\xi', \eta')$$

that is $\bar{S}^{k'}(\xi', \eta') = \bar{S}(\xi'^{,k'}, \eta'^{,k'})$, and subsequently $\bar{S}^{m/k}(\xi', \eta') = \bar{S}(\xi'^{,m/k}, \eta'^{,m/k})$. By the continuity of $\bar{S}(\xi, \eta)$, for every $\omega > 0$, we have

$$\bar{S}^{\omega}(\xi, \eta) = \bar{S}(\xi^{\omega}, \eta^{\omega}).$$

The solution of this functional equation can be written under the form

$$\bar{S}(\xi, \eta) = (\xi . \eta)^{k(\log \frac{\log \xi}{\log \eta})} .$$

Taking, as stated, the reduced margins to be $\Lambda(x) = \exp(-e^{-x})$ and $\Lambda(y) = \exp(-e^{-y})$, we have the following fundamental result:

Any limiting distribution function of bivariate pairs with reduced Gumbel margins is of the form

$$\Lambda(x, y) = \exp\{-(e^{-x} + e^{-y})\, k\,(y - x)\} = \{\Lambda(x)\, \Lambda(y)\}^{k(y-x)}, \text{ and conversely}$$

When the asymptotic extremal pair $(\tilde{X}, \tilde{Y})$ does not have reduced margins, that is $\tilde{X}$ and $\tilde{Y}$ have location parameters λ_x, λ_y and dispersion parameters $\delta_x > 0, \delta_y > 0$, the distribution function of $(\tilde{X}, \tilde{Y})$, as $((\tilde{X} - \lambda_x)/\delta_x, (\tilde{Y} - \lambda_y)/\delta_x)$ has reduced margins, is $\Lambda((\tilde{x} - \lambda_x)/\delta_x, (\tilde{y} - \lambda_y)/\delta_y)$.

We have shown in two ways that the limiting and the stable distribution functions of extremal pairs form the same class.

Let us now deal briefly with minima. As $\underset{\sim}{X}_k = \min(X_1, \ldots, X_k) = -\max(-X_1, \ldots, -X_k)$ and $\underset{\sim}{Y}_k = \min(Y_1, \ldots, Y_k) = -\max(-Y_1, \ldots, -Y_k)$, the asymptotic distribution function of minima with reduced Gumbel margins $1 - \exp(-e^{x})$ and $1 - \exp(-e^{y})$, if they exist, and correspondence with $\{\Lambda(x) . \Lambda(y)\}^{k(y-x)}$, is

$$\Psi(x, y) = 1 - \exp(-e^{x}) - \exp(-e^{y}) + \exp(-(e^{x} + e^{y})\, k(x - y)).$$

Using the fact that $\min_1^k(Y_i) = -\max_1^k(-Y_i)$ we can also prove easily that the limiting distribution of $(\min_1^k X_i , \max_1^k Y_i)$, if it exists, is

$$\Theta(x , y) = \Lambda(y) - \Lambda(-x , y) = \exp(-e^{-y}) - \exp[-(e^{x} + e^{-y})\, k(y + x)] .$$

We have obtained the dependence function k(w) obviously continuous and non-negative for $\Lambda(x , y)$ to be a distribution function as well as the corresponding functions in other equivalent representations. These results are well known. They can be found, with different forms for the margins, in Finkelshteyn (1953), Tiago de Oliveira (1958), Geffroy (1958/59) and Sibuya (1960), with a synthesis of the results in Tiago de Oliveira (1962/63). Subsequent results are in Tiago de Oliveira (1975), (1980) and (1984). Berman (1962) gives also some limiting results and Galambos (1978) contains also a recent account.

Let us now obtain some inequalities for $\Lambda(x , y)$ or k(w) which will simplify some of the derivations below. Although $\Lambda(x , y)$ is a continuous function, it does not necessarily have derivatives, and consequently we cannot expect all bivariate maxima random pairs with distribution function $\Lambda(x , y) = [\Lambda(x)\, \Lambda(y)]^{k(y-x)}$ to have a planar density. From the Boole-Fréchet inequality — Fréchet (1951) —

$$\max(0 , \Lambda(x) + \Lambda(y) - 1) \leq \Lambda(x , y) \leq \min(\Lambda(x) , \Lambda(y))$$

we have, replacing x and y by $x + \log k$ and $y + \log k$, raising to the power k and letting $k \to \infty$, the limit inequality

$$\Lambda(x)\, \Lambda(y) \leq \Lambda(x , y) \leq \min(\Lambda(x) , \Lambda(y))$$

or

$$\exp\{-(e^{-x} + e^{-y}) \leq \Lambda(x , y) \leq \exp(-e^{-\min(x,y)}) .$$

The LHS inequality shows that we have positive association, that is $\text{Prob}\{X \leq x , Y \leq y\} + \text{Prob}\{X > x , Y > y\}$ is larger or equal to the corresponding values the case of for independence, for all x and y, which means that if we have a large (small) value of X, the value of Y, tends also to be large (small). $\Lambda(x) . \Lambda(y)$ is, evidently, the independence situation; $\min(\Lambda(x) , \Lambda(y))$ is called the *diagonal or complete dependence case* where we have, for reduce values, $\text{Prob}\{Y = X\} = 1$. As regards k(w), we have

$$(1/2 \leq) \frac{\max(1\ ,\ e^{w})}{1 + e^{w}} \leq k(w) \leq 1;$$

$k_D(w) = \dfrac{\max(1\ ,\ e^{w})}{1 + e^{w}} = \dfrac{1}{1 + e^{-|w|}}$ is the dependence function corresponding to the diagonal case and $k_I(w) = 1$ is the dependence function for the independence situation. Positive association, shown by Sibuya (1960), is a result that could be expected.

Remark that for $\bar{S}(\xi, \eta)$ defined above we have $\xi\,\eta \leq \bar{S}(\xi, \eta) \leq \min(\xi, \eta)$.

If there exists a planar density almost everywhere, i.e., if $k''(w)$ exists almost everywhere, $k(w)$ must satisfy the relations:

$$k(-\infty) = k(+\infty) = 1\ ,$$

$$[(1 + e^{w})\ k(w)]' \geq 0\ ,$$

$$[(1 + e^{-w})\ k(w)]' \leq 0\ ,$$

and

$$(1 + e^{-w})\ k''(w) + (1 - e^{-w})\ k'(w) \geq 0\ ,$$

easily obtained by derivation; the corresponding conditions for the general case are, as $\Delta^2_{x,y}\ \Lambda(x\ ,\ y) \geq 0$:

$k(-\infty) = k(+\infty) = 1$,

$(1 + e^{w})\ k(w)$ a non-decreasing function ,

$(1 + e^{-w})\ k(w)$ a non increasing function ,

and

$$\Delta^2_{x,y}\ [(e^{-x} + e^{-y})\ k\ (y - x)] \leq 0\ .$$

Note that as $(1 + e^{w})\ k(w)$ and $(1 + e^{-w})\ k(w)$ are monotonic (non-decreasing and non-increasing), we know that $k'(w)$ exists everywhere except for a denumerable set of points of $\mathbb{R}$.

We will follow the proof of Tiago de Oliveira (1962/63).

As is well known, $\Lambda(x , y)$ is such that

1) $\Lambda(x , +\infty) = \Lambda(x)$, $\Lambda(+\infty, y) = \Lambda(y)$ (margin conditions)

2) $\Delta^2 \Lambda (x , y) \geq 0$. (non-negativity condition)

As $\Lambda (x , y)$ is a continuous function, $k(w)$ is also a continuous function. Condition 1) is easily seen to be equivalent to

I) $k(-\infty) = k(+\infty) = 1$.

Consider condition 2) . It can be written as

$$\exp \{-(e^{-x} + e^{-y}) k(y - x)\} + \exp \{-(e^{-\xi} + e^{-\eta}) k(\eta - \xi)\}$$

$$\geq \exp \{-(e^{-x} + e^{-\eta}) k(\eta - x)\} + \exp \{-(e^{-\xi} + e^{-y}) k(y - \xi)\}$$

for $x \leq \xi , y \leq \eta$.

We will show that 2) is equivalent to

II) $(1 + e^{w})\ k(w)$ a non-decreasing function ,

III) $(1 + e^{-w})\ k(w)$ a non increasing function ,

IV) $(e^{-x} + e^{-y})\, k(y - x) + (e^{-\xi} + e^{-\eta})\ k(\eta + \xi)\}$

$$\geq (e^{-x} + e^{-\eta})\, k(\eta - x)\} + (e^{-\xi} + e^{-y})\, k(y - \xi)\}$$

for $x \leq \xi , y \leq \eta$.

II) and III) are obtained taking, respectively, $y \to -\infty$ and $x \to +\infty$. To show IV), let us put

$$Z = \exp(- e^{-\xi}) , \ A = (e^{-(x-\xi)} + e^{-(y-\xi)})\ k(y - x) ,$$

$$B = (1 + e^{-(\eta-\xi)})\, k(\eta - \xi) , \ C = (e^{-(\eta-\xi)} + e^{-(\eta-\xi)})\ k(\eta - x) ,$$

$$D = (1 + e^{-(y-\xi)}\ k(y - \xi) .$$

Conditions 2) and IV) are written as $(A, B, C, D \geq 0)$

2) $Z^{-A} + Z^{-B} \geq Z^{-C} + Z^{-D}$

IV) $A + B \geq C + D$.

Letting ξ vary with x, y, η but such that $x - \xi, y - \xi, \eta - \xi$ are fixed, A, B, C, D are also fixed. We then have to prove the full equivalence of 2) and II) under the new form for fixed A, B, C, D and $0 \leq Z \leq 1$.

Let us then consider the function $f(Z) = Z^{-A} + Z^{-B} - Z^{-C} - Z^{-D}$. As $f(Z) \geq 0$ by 2) and $f(1) = 0$ we have $f'(1) \leq 0$ so that IV) is true.

Let us now prove the converse, i.e., that II), III) and IV) imply 2). Conditions II) and III) give $A \geq D, C \geq B$.

Let us prove now 2). Supposing IV) to be true, we have

$$Z^{-A} . Z^{-B} \geq Z^{-C} . Z^{-D} \qquad \text{and thus}$$

$$\max_{\substack{A+B \leq C+D \\ B \leq C, D \leq A}} (Z^{-C} + Z^{-D}) \leq \max_{B \leq C \leq A} (Z^{-C} + \frac{Z^{-A} Z^{-B}}{Z^{-C}}) .$$

The RHS function of $C \in [B, A]$ has a minimum at $C = \frac{A + B}{2}$ and the maximum value in the interval is attained at $C = A$ and $C = B$ with the value $Z^{-B} + Z^{-A}$. We have then proved $Z^{-A} + Z^{-B} \geq Z^{-C} + Z^{-D}$ as desired.

If we are dealing with minima with exponential margins, corresponding to the transformation $\xi = e^{-x}, \eta = e^{-y}$, the asymptotic distribution for $\xi, \eta \geq 0$ takes the form $\Psi(x, y) = 1 - e^{-\xi} - e^{-\eta} + e^{-(\xi+\eta)A(\eta/(\xi+\eta))}$ or the survival function $S(\xi, \eta) = e^{-(\xi+\eta)A(\eta/(\xi+\eta))}$.
It can be shown that the conditions for k(.) are equivalent to

$$A(0) = A(1) = 1,$$

$$0 \leq -A'(0), A'(1) \leq 1$$

and $A(u)$ *convex in* $[0, 1]$.

The functions $k(w)$ and $A(u)$ are related by $k(w) = A(\frac{1}{1 + e^{w}})$ or $A(u) = k(\log \frac{1 - u}{u})$.
We have $A(u) = 1$ for independence and $A(u) = \max(u, 1 - u)$ in the diagonal case, with, also, $\max(u, 1 - u) \leq A(u) \leq 1$.

Another formulation is

$$A(0) = A(1) = 1,$$

$$max(u,1-u) \leq A(u) \leq 1$$

and $A(u)$ *convex in* $[0,1]$.

Complementary basic results

Some properties can be ascribed to the set of $k(w)$. *The first one is a symmetry property,* i.e., if $k(w)$ is a dependence function, then $k(-w)$ is also a dependence function. The proof is immediate if we consider the conditions in the differentiable case (where a planar density does exist) and slightly longer in the general case. If $k(w) = k(-w)$ then (X, Y) is an *exchangeable pair* and $\Lambda(x, y) = \Lambda(y, x)$.

Also it is immediate that if $k_1(w)$ and $k_2(w)$ are dependence functions, any mixture $\theta\, k_1(w) + (1-\theta)\, k_1(w)$, $0 \leq \theta \leq 1$ is also a dependence function. *The set of dependence functions is, then, convex.* And this convexity property

$$\Lambda_1^{\theta}(x, y) = \Lambda(x, y) \cdot \Lambda_2^{1-\theta}(x, y)$$

is very useful for obtaining models: the mixed model, as well as the Gumbel model, are such examples. We will call this way of modelling the *mix-technique.*

As a generalization, it can be observed that if $G(u)$ is a distribution function in $[0, 1]$, then $k(w)$ given by $(1 + e^{-w})\, k(w) = \int_0^1 (1 + e^{-w/u})^u \, d\, G(u)$ is also a dependence function.

Another method of generating models is the following *max-technique.* Let (X, Y) be an extreme random pair, with dependence function $k(w)$ and reduced Gumbel margins, and consider the new random pair $(\tilde{X}, \tilde{Y})$ with $\tilde{X} = \max(X + a, Y + b)$, $\tilde{Y} = \max(X + c, Y + d)$. To have reduced Gumbel margins we must have $(e^a + e^b)\, k(a-b) = 1$ and $(e^c + e^d)\, k(c-d) = 1$. Then we have

$$\tilde{k}(w) = \frac{[e^{\max(a+w,c)} + e^{\max(b+w,d)}]\; k[\max(a+w, c) - \max(b+w, d)]}{1 + e^w}$$

with (a , b) and (c , d) satisfying the conditions given before. This *max-technique* will be used towards the end of the paper to generate the biextremal and natural models.

Let us stress that independence has a very important position as a limiting situation. If we denote by $P(u , v)$ the function defined by $\text{Prob}\{X > x , Y > y\} = P(F(x , +\infty) , F(+\infty, y))$, Sibuya (1960) has shown that the necessary and sufficient condition to have limiting independence is that $P(1 - s , 1 - s)/s \to 0$ as $s \to 0$. He also showed that the necessary and sufficient condition for having the diagonal case as a limiting situation is that $P(1 - s , 1 - s)/s \to 1$ as $s \to 0$. With the first result we can easily show that the maxima of the binormal distribution has independence as a limiting distribution if $|\rho| < 1$.

Also Geffroy (1958/59) showed that a sufficient condition for limiting independence is that

$$\frac{1 + F(x , y) - F(x , w_y) - F(w_x , y)}{1 - F(x , y)} \to 0 \quad \text{as} \quad x \to \bar{w}_x \quad \text{and} \quad y \to \bar{w}_y ,$$

$\bar{w}_x$ and $\bar{w}_y$ being the right end-points of the supports of X and Y.

Sibuya conditions (and Geffroy sufficient condition) are easy to interpret: we have limiting independence if $\text{Prob}\{X > x , Y > y\}$ is a vanishing summand of $\text{Prob}\{X > x \text{ or } Y > y\}$ and the diagonal case as limit if $\text{Prob}\{X > x , Y > y\}$ is the leading summand of $\text{Prob}\{X > x \text{ or } Y > y\}$.

We have asymptotic independence when $\bar{S}(\xi , \eta) = \xi \eta$, that is, when $\bar{S}_o^k (\xi^{1/k}, \eta^{1/k}) \to \xi \eta$. The sufficient condition for independence given by Geffroy can be proved in a simpler way, as follows. It is equivalent to :

If $\dfrac{1 - \bar{S}_o (\xi , \eta)}{2 - \xi - \eta} \to 1$ *when* $\xi , \eta \to 1$ *then* $\bar{S}(\xi , \eta) = \xi \eta$.

As $\bar{S}_o (\xi , \eta) = 1 - (2 - \xi - \eta)\, h (\xi , \eta)$, with $h (\xi , \eta) \to 1$ when $\xi , \eta \to 1$, we have $\bar{S}_o^k (\xi^{1/k}, \eta^{1/k}) = \{1 - (2 - \xi^{1/k} - \eta^{1/k})\, h (\xi^{1/k}, \eta^{1/k})\}^k \to \xi \eta$ because

$$h (\xi^{1/k} , \eta^{1/k})\}^k \to 1 \quad \text{when} \quad \eta \to \infty .$$

Its meaning is straightforward, and, in general, we can expect extremal pairs to be asymptotically independent. Mardia (1964) proved that the sample extremes of a bivariate pair are independent under very general conditions.

A simple example now shows that we can have non-independence cases. Let (X , Y , Z) be three independent random variables with the same distribution function F. The distribution function of the pair $(\max(X , Z), \max (Y , Z))$ is $P(x , y) = \text{Prob}\{\max(X , Z) \leq x , \max (Y , Z) \leq y\} = F(x) F(y) F(\min(x , y))$ and the one for the maxima in a sample of k pairs it is $P_k(x , y) = P^k(x , y)$. Its margins are

$$P_k(x , + \infty) = F^{2k}(x) , \quad P_k(+ \infty, y) = F^{2k}(y) .$$

Supposing now that for λ_k and $\delta_k > 0$, we have

$$F^k(\lambda_k + \delta_k \ z) \rightarrow \Lambda(z)$$

we obtain

$$P_k (\lambda_{2k} + \delta_{2k} \ x , \lambda_{2k} + \delta_{2k} \ y)$$

$$= F^k(\lambda_{2k} + \delta_{2k} \ x) F^k(\lambda_{2k} + \delta_{2k} \ y) F^k(\lambda_{2k} + \delta_{2k} \ \min(x , y))$$

and we have

$$P_k (\lambda_{2k} + \delta_{2k} \ x , \lambda_{2k} + \delta_{2k} \ y) \rightarrow \Lambda^{1/2} (x) \Lambda^{1/2} (y) \Lambda^{1/2} (\min(x , y)) ,$$

a case with does not correspond to independence.

It can be remarked that $P(x , y)$, as well as the limit, does not have a planar density.

Sometimes it is usual to study the equidistribution and equidensity curves (if they exist). The latter are very difficult to deal with so that we will restrict ourselves to the former.

The equations of the equidistribution curves

$$\Lambda(x , y) = \alpha \qquad (0 < \alpha < 1)$$

have the form

$$(e^{-x} + e^{-y}) \ k(y - x) = - \log \ \alpha \qquad (0 < - \log \alpha < + \infty) .$$

From these equations we see that the translations parallel to the first diagonal

$$x \rightarrow x + a = x'$$

$$y \rightarrow y + a = y'$$

transform the first curve in the other equidistribution curve

$$(e^{-x'} + e^{-y'})\, k(y' - x') = (-\log\ \alpha)\, e^{-a}\ .$$

Consequently we can centre the study of those curves in the median curve $\Lambda(x\ ,\ y) = 1/2$, that is

$$(e^{-x} + e^{-y})\, k(y - x) = \log 2\ .$$

As $k(-\infty) = k(+\infty) = 1$, the asymptotes to the median curves are

$$x = -\log\log 2 \qquad \text{and} \qquad y = -\log\log 2\ .$$

As these asymptotes are independent of the dependence function, the equidistribution curves behave similarly for large values of x or y, as could be expected, and, then, like the median curve for independence. The relations deduced from Fréchet inequalities yield the double inequality

$$e^{-\min(x,y)} \leq \log 2 \leq e^{-x} + e^{-y}\ ,$$

showing that any median curve is, necessarily, between the median curves corresponding to the independence

$$e^{-x} + e^{-y} = \log 2$$

and to the diagonal case

$$\min(x\ ,\ y) = -\log\log 2$$

which degenerates in part of the asymptotes.

Consider now the random variable $W = Y - X$, the difference of reduced extremes. We have, evidently, $M(W) = 0$ and $V(W) = \frac{\pi^2}{3}\ (1 - \rho)$. Recall that as $V(X) = V(Y) = \pi^2/6$ the covariance $C(X\ ,\ Y) = \frac{\pi^2}{6}\ \rho$ exists as well as the correlation coefficients ρ.

When $\rho = 1$ we have $V(W) = 0$; W is a random variable almost surely equal to zero and we have $Y = X$ almost surely. We then have the diagonal case.

Let us denote by $D(w) = \text{Prob}\{Y - X \leq w\}$; we have

$$D(w) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{x+w} dy \, \frac{\partial^2 \Lambda}{\partial x \, \partial y} = \int_{-\infty}^{+\infty} dx \, \Lambda_1(x, x+w)$$

where

$$\Lambda_1(x, y) = \frac{\partial \Lambda(x, y)}{\partial x} = \Lambda(x, y)\{e^{-x} k(y-x) + (e^{-x} + e^{-y}) k'(y-x)\}$$

so that

$$D(w) = \int_{-\infty}^{+\infty} dx \, e^{-e^{-x}(1+e^{-w})k(w)} e^{-x} \{k(w) + (1+e^{-w}) k'(w)\} = \frac{1}{1+e^{-w}} + \frac{k'(w)}{k(w)}$$

as is obvious.

The relations between $D(w)$ and $k(w)$ are immediate. Integration between 0 and w of the differential relation gives

$$(1+e^{w}) \; k(w) = 2\, k(0) \exp \int_{0}^{w} D(w) \, dw,$$

and letting $w \to -\infty$ $(k(-\infty) = 1)$ we obtain

$$k(w) = \frac{\exp \int_{-\infty}^{w} D(w) \, dw}{1+e^{w}} .$$

Note that $\int_{-\infty}^{0} D(w) \, dw$ and $\int_{0}^{+\infty} (1 - D(w)) \, dw$ exist and are equal because W has mean value zero.

The conditions for $k(w)$ are immediately translated into conditions for $D(w)$ so that

$$D(w) = \text{Prob}\{Y - X \leq w\}$$

verifies the following relations (apart from being a distribution function):

$$\int_{-\infty}^{0} D(w)\, d\, w = \int_{0}^{+\infty} [1 - D(w)]\, d\, w \qquad \text{or} \qquad \int_{-\infty}^{+\infty} w\, d\, D(w) = 0$$

and

$$D'(w) \geq D(w)\,[1 - D(w)] \ .$$

We can also write

$$D(w) = k'(w)/k(w) + L(w) \ ,$$

where $L(w) = (1 + e^{-w})^{-1}$ is the standard logistic distribution function.

We have $D(w) = L(w)$ in the independence case and $D(w) = H(w)$ where $H(w)$ is the Heaviside jump function at 0, $H(w) = 0$ if $w < 0$ and $H(w) = 1$ if $w \geq 0$ — for the diagonal case. In the case of exchangeability $(k(w) = k(-w))$ we have $D(w) + D(-w) = 1$, and W has a symmetric distribution.

As said before, $k'(w)$ exists everywhere except for a denumerable set of points and thus $D(w)$ is defined everywhere: directly almost everywhere and at the discontinuity set by the right-continuity of the distribution function $D(.)$.

Let us denote by $ID(w) = \int_{-\infty}^{w} D(t)\, d\, t$; we have seen that

$$k(w) = \frac{e^{ID(w)}}{1 + e^{w}} \ .$$

The conditions on the distribution function $D(w)$, in the case of the existence of planar density, are then

$$\int_{-\infty}^{+\infty} w\, d\, D(w) = 0$$

or equivalently

$$\int_{-\infty}^{+\infty} (H(w) - D(w))\, d\, w = 0$$

and

$$D'(w) \geq D(w)\,(1 - D(w)) \ .$$

More generally, the last condition is substituted by

$$\Delta^2_{x,y}\ [\exp\,(ID(y - x) - y] \leq 0$$

which is equivalent to

$$e^{ID(w)} - e^{ID(w-\alpha)} \leq e^{-\beta}\,[e^{ID(w+\beta)} - e^{ID(w+\beta-\alpha)}\,]$$

with $\alpha\,,\,\beta \geq 0$, equivalent

$$e^{ID(w)}\,(1 - D(w)) \leq e^{ID(w-\alpha)}\,(1 - D(w - \alpha))$$

or that

$$e^{ID(w)}\,(1 - D(w))$$

is a non-increasing function.

Note that $ID(-\infty) = 0$, $ID(w) - w \to +\infty$, $ID(w + \beta) - ID(w) \leq \beta$ $(\beta > 0)$, $e^{ID(w)} - e^{ID(w-\alpha)} \leq e^w - e^{w-\alpha}$ $(\alpha \geq 0)$, $\max\,(0\,,\,w) \leq ID(w) \leq \log\,(1 + e^w)$ (from Boole-Fréchet inequalities) , $\int_{-\infty}^{0} D(t)\,d\,t = \int_{0}^{+\infty} (1 - D(t))\,d\,t$ (from the null mean value) , $D(w)\ e^{ID(w)} \leq e^w$ and $(1 - D(w))\ e^{ID(w)} \leq 1$ (by convenient integration of $D' \geq D(1 - D)$) .

Let $\underline{w}$ and $\bar{w}$ ($\underline{w} \leq 0 \leq \bar{w}$, $\bar{w} = \underline{w} = 0$ in the diagonal case and $\underline{w} < 0 < \bar{w}$ in the other cases) be the left and right end-points of $D(w)$.

For $w > \bar{w}$, if $\bar{w} < +\infty$, we have

$$ID(w) = \int_{-\infty}^{w} D(t)\,d\,t = \int_{+\infty}^{\bar{w}} D(t)\,d\,t + w - \bar{w}\ ;$$

as

$$ID(w) - w = \int_{-\infty}^{\bar{w}} D(t)\, d\, t - \bar{w} \to 0$$

we have

$$\bar{w} = \int_{-\infty}^{\bar{w}} D(t)\, d\, t\,,$$

a result that is also true $\bar{w} = +\infty$.

Consider now, for $\underline{w} > -\infty$,

$$\int_{\underline{w}}^{+\infty} (1 - D(t))\, d\, t = \int_{\underline{w}}^{0} (1 - D(t))\, d\, t + \int_{0}^{+\infty} (1 - D(t))\, d\, t =$$

$$-\underline{w} - \int_{\underline{w}}^{0} D(t)\, d\, t + \int_{0}^{+\infty} (1 - D(t))\, d\, t = -\underline{w}$$

which is also true $\underline{w} = -\infty$.

If both $\underline{w}$ and $\bar{w}$ are finite we have

$$\bar{w} = \int_{\underline{w}}^{w} D(t)\, d\, t \quad \text{and} \quad \underline{w} = -\int_{\underline{w}}^{\bar{w}} (1 - D(t))\, d\, t\,.$$

The symmetry condition (exchangeability) $k(w) = k(-w)$ is equivalent to $D(w) + D(-w) = 1$ or $ID(w) = w + ID(-w)$.

From the conditions on $D(w)$ we can give new methods of generating bivariate models of maxima, as follows, essentially for absolutely continuous distribution functions. The most natural is the one that follows.

In $[\underline{w}, \bar{w}]$, the support of $D(.)$, we can define the function $\Psi(w) = \dfrac{D'(w)}{D(w)\,(1 - D(w))} \geq 1$. Notice that as a consequence every point of $[\underline{w}, \bar{w}]$ is a point of increase of $D(.)$, and thus

the quantiles are uniquely defined. From the absolute continuity we have $D(\underline{w}) = 0$ and $D(\bar{w}) = 1$. Note that $\underline{w} < 0 < \bar{w}$, as the mean value is zero.

Let us fix ν as the median. Integrating the differential equation $D' = D(1 - D)\,\Psi$ we get, with $I\Psi(w) = \int_{-\infty}^{w} \Psi(t)\,dt$,

$$D(w|\nu) = \frac{1}{1 + e^{-(I\Psi(w) - I\Psi(\nu))}},$$

and thus the relation between $D(w|\nu)$ and $D(w|\nu')$ is

$$D(w|\nu') = \frac{D(w|\nu)}{D(w|\nu) + e^{-K}\,(1 - D(w|\nu))}$$

with $K = \int_{\nu'}^{\nu} \Psi(t)\,dt$ and the same support.

As $\Psi(w) \geq 1$ it is immediate that $\int_{\nu}^{w} \Psi(t)\,dt \geq (w - \nu)$ for $w > \nu$. We get, thus

$$D(w|\nu) \geq \frac{1}{1 + e^{-(w-\nu)}}$$

and for $w < \nu$, analogously, we obtain

$$D(w|\nu) \leq \frac{1}{1 + e^{-(w-\nu)}},$$

results not known previously.

Having fixed, temporarily, the median at ν ($\underline{w} < \nu < \bar{w}$), the condition $\int_{-\infty}^{+\infty} w\,d\,D(w|\nu) = \int_{-\infty}^{+\infty} [H(w) - D(w|\nu)]\,dw = 0$ reads as

$$I(K) = 0 \text{ where}$$

$$I(K) = \int_{\underline{w}}^{\bar{w}} [\, H(w) - \frac{D(w|v)}{D(w|v) + e^{-K}\ (1 - D(w|v))}\,]\, d\, w\ .$$

As $I(K) \to \underline{w}$ when $K \to +\infty$, $I(K) \to \bar{w}$ when $K \to -\infty$ and $I(K)$ is increasing, we see that the solution of $I(K) = 0$ is unique. The unique solution K_o of $I(K) = 0$ leads to

$$\int_{v'}^{v} \Psi(t)\, d\, t = K_o$$

and, as $\Psi(w) \geq 1$, the only solution for $D(.)$ is immediately obtained. It is thus sufficient to give an integrable function $\Psi(w) \geq 1$ in every finite interval $[\underline{w}, \bar{w}]$ to obtain the unique $D(.)$ related to it. For simplicity we can take $v = 0$.

Correlation results

The results concerning correlation that follow and the regression results to be given later, to a certain extent, illuminate some other features of the situation.

Let us now compute some widely used correlation coefficients.

The classical correlation coefficient always exists as shown before and is $\rho = \frac{6}{\pi^2}\ C\,(X\ ,\ Y)$.

A known expression for $C\,(X\ ,\ Y)$ is

$$C\,(X\ ,\ Y) = \int_{-\infty}^{+\infty}\int [F(x\ ,\ y) - F(x\ ,\ +\infty)\ F(+\infty,\ y)]\, d\, x\ \ d\, y$$

which, in our case, takes the form

$$C\,(X\ ,\ Y) = \int_{-\infty}^{+\infty}\int [\Lambda(x\ ,\ y) - \Lambda(x)\ \ \Lambda(y)]\, d\, x\ \ d\, y\ ,$$

and because of the positive association we have $C(X\ ,\ Y) \geq 0$, Substituting the expressions of $\Lambda(x\ ,\ y)$, $\Lambda(x)$ and $\Lambda(y)$ and using the change of variables

$$e^{-v} = e^{-x} + e^{-y}$$

$$w = y - x ,$$

we obtain

$$C(X , Y) = \int_{-\infty}^{+\infty} d w \int_{-\infty}^{+\infty} d v \, [e^{-e^{-v} k(w)} - e^{-e^{-v}}] .$$

Let us now compute the inner integral. It can be written as

$$\int_{0}^{+\infty} (1 - e^{-e^{-v}}) \, d v - \int_{-\infty}^{0} e^{-e^{-v}} \, d v + \int_{-\infty}^{0} e^{-e^{-v} k(w)} \, d v - \int_{0}^{+\infty} (1 - e^{-e^{-v} k(w)}) \, d v .$$

The first two integrals, being an expression for $M(X)$, add to γ. Changing v to $v = z + \log k(w)$ in the two last integrals, we have after simple computations for the inner integral

$$\gamma + \int_{-\infty}^{0} e^{-e^{-z}} \, d z - \int_{0}^{+\infty} (1 - e^{-e^{-z}}) \, d z - \int_{-\log k(w)}^{0} d z = - \log k(w)$$

because the two first integrals add to $-\gamma$.

We then have the expression:

$$C(X , Y) = - \int_{-\infty}^{+\infty} \log k(w) \, d w \geq 0$$

and thus the classical correlation coefficient for $\Lambda(x , y)$ is

$$\rho = - \frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log k(w) \, d w \; ;$$

when needed we will write $\rho(k)$ to indicate the dependence function.

From the expression of $D(w)$ we can obtain, once more in the differentiable case, also the expression $\rho = - \frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log k(w) \, d w$. Let us give the supplementary proof.

Let us recall first that the existence of the variance V(W) implies that $\lim_{w\to\infty} w^2 D(w) = 0$ and $\lim_{w\to\infty} w^2(1 - D(w)) = 0$, and, as the same happens for the logistic $1/(1 + e^{-w})$, those conditions are equivalent to $\lim_{w\to+\infty} w^2 \frac{k'(w)}{k(w)} = 0$. As the variance of the logistic ($\rho = 0$) is $\pi^2/3$, we have the general relation (M(W) being zero)

$$V(W) = \frac{\pi^2}{3}(1 - \rho) = \int_{-\infty}^{+\infty} w^2 \, d\,D(w) = \int_{-\infty}^{+\infty} w^2 \, d\left(\frac{1}{1 + e^{-w}}\right) + \int_{-\infty}^{+\infty} w^2 \, d\left(\frac{k'(w)}{k(w)}\right)$$

so that

$$-\frac{\pi^2}{3}\rho = \int_{-\infty}^{+\infty} w^2 \, d\left(\frac{k'(w)}{k(w)}\right).$$

Integration by parts gives

$$\frac{\pi^2}{3}\rho = 2\int_{-\infty}^{+\infty} w \, \frac{k'(w)}{k(w)} \, d\,w,$$

the integrated parts being zero as a consequence of

$$\lim_{w \to \pm\infty} w^2 \frac{k'(w)}{k(w)} = 0.$$

The last integral is equal, by integration by parts, to

$$\int_{-\infty}^{+\infty} \log k(w) \, d\,w,$$

with the integrated part $w \log k(w) \to 0$ when $w \to \pm\infty$, as follows from the Fréchet derived inequalities.

We can summarize the results by giving the equivalent expressions for ρ:

$$\rho = 1 - \frac{3}{\pi^2}\int_{-\infty}^{+\infty} w^2 \, d\,D(w) = \frac{6}{\pi^2}\int_{-\infty}^{+\infty} w\,\{D(w) - L(w)\}\,d\,w$$

$$= \frac{6}{\pi^2} \cdot \int_{-\infty}^{+\infty} (\mathbb{L}(t) - \mathbb{D}(t))\, d\, t = - \frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log k(w)\, d\, w \ .$$

As $k(w) \leq 1$ $(-\log k(w) \geq 0)$ we have $0 \leq \rho$, as could be expected from the positive association. It is very easy to show that for the diagonal case

$$k_D(w) = \frac{\max(1, e^w)}{1 + e^w}$$

we have $\rho = 1$. The value of ρ does not identify the dependence function (or the distribution): ρ is the same for $k(w)$ and $k(-w)$. But $\rho = 0$, as $k(w) \leq 1$, implies $k(w) = 1$, or independence. Now writing ρ under the form

$$\rho = 1 - \frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log \frac{k(w)}{k_D(w)}\, d\, w \ ,$$

we see, analogously, that $\rho = 1$ as $k(w) \geq k_D(w)$ implies $k(w) = k_D(w)$, or the diagonal case.

For the non-parametric correlation coefficients used, see Fraser (1957) and Konijn (1956) for details.

With the expression

$$\chi = 12 \int_{-\infty}^{+\infty} \int \Lambda(x, y)\, d\, \Lambda(x)\, d\Lambda(y) - 3$$

for the grade correlation coefficient we immediately find that the grade correlation coefficient has the expression

$$\chi = 12 \int_{-\infty}^{+\infty} \frac{e^w}{(1 + e^w)^2\, (1 + k(w))^2}\, d\, w - 3 \ .$$

It is immediate also that as $k(w) \leq 1$ the integral in the final expression for χ is greater than

$$\int_{-\infty}^{+\infty} \frac{e^w}{(1 + e^w)^2} \cdot \frac{d\, w}{4} = \frac{1}{4} \qquad \text{so that} \qquad \chi \geq 0 \ .$$

Other expressions for χ are

$$\chi = 12 \int_{-\infty}^{+\infty} \frac{e^w}{(1 + e^w)^2 (1 + k(w))^2} \, d\, w - 3 = 12 \int_{-\infty}^{+\infty} \frac{e^w}{(1 + e^w + e^{ID(w)})^2} \, d\, w - 3 .$$

For the difference-sign correlation coefficient

$$\tau = 4 \int_{-\infty}^{+\infty} \int \Lambda(x , y) \, d \, \Lambda(x , y) - 1 \ ,$$

in the case of differentiability, we show that the expression for the difference-sign correlation is

$$\tau = \int_{-\infty}^{+\infty} \frac{k'(w)2}{k(w)} \, d\, w - 2 \int_{-\infty}^{+\infty} \frac{e^w}{(1 + e^w)^2} \log k(w) \, d\, w \ ,$$

which can also take the form

$$\tau = 1 - \int_{-\infty}^{+\infty} D(w) \, (1 - D(w)) \, d\, w = \int_{-\infty}^{+\infty} (D^2 \, (w) - L^2 \, (w)) \, d\, w \ .$$

By integration by parts of $\int_{-\infty}^{+\infty} D(w) \, (1 - D(w)) \, d\, w$ as $\int_{-\infty}^{+\infty} w \, d \, D(w) = 0$, we get the simple expression

$$\tau = 1 - \int_{-\infty}^{+\infty} w \ d \ D^2 \, (w) \ .$$

Finally for the medial correlation coefficient

$$\nu = 4 \, [\Lambda(\tilde{\mu} \, , \tilde{\mu} \,) - \Lambda(\tilde{\mu} \,) \, \Lambda(\tilde{\mu} \,)] = 4 \, \Lambda(\tilde{\mu} \, , \tilde{\mu} \,) - 1,$$

$\tilde{\mu} = -\log \log 2$ being the median, we see that the medial correlation coefficient has the expression $\nu = 4^{1-k(0)} - 1 \ (\geq 0)$.

The probability of concordance, that is, the probability that the two pairs $(X_1 \, , Y_1)$ and $(X_2 \, , Y_2)$ are such that $X_1 - X_2$ and $Y_1 - Y_2$ have the same sign, being given by $(1 + \tau \,)/2$, is always greater than 1/2.

As an index of dependence we can, also, use

$$\Delta(k) = \sup_{x,y} |\Lambda(x , y) - \Lambda(x)\ \Lambda(y)| = \sup_{x,y} [\Lambda(x , y) - \Lambda(x)\ \Lambda(y)]$$

as follows from positive association.

Computations show easily that if $k(w)$ is differentiable and w_o is such that $k'(w_o) = o$ the value of $\Delta(k)$ is, with $k_o = k(w_o)$,

$$\Delta(k) = k_o^{k_o/(1-k_o)} - k_o^{1/(1-k_o)} ;$$

for $k(w) = k(-w)$ we have $w_o = 0$. It is very easy to see from the expression of $k'(w)$ has only an interval of solutions so that $k(w)$ has the same value for all w_o in the interval (a minimum). The same result an be obtained directly without recourse to the hypothesis of differentiability, as follows.

Denoting by $Z = \Lambda(x)\ \Lambda(y)$, the expression of $\Delta(k)$ can be written (as $k(w) \leq 1$)

$$\sup_{\substack{-\infty \leq w \leq +\infty \\ 0 \leq Z \leq 1}} (Z^{k(w)} - Z)$$

which for fixed $Z\ (\leq 1)$ has the maximum for the minimum $k_o = k(w_o)$ of $k(w)$. The maximum of $Z^{k(w_o)} - Z$ is given by the previous value. Then the index of dependence

$$\Delta(k) = \sup_{x,y} [\Lambda(x , y) - \Lambda(x)\ \Lambda(y)]$$

has the value $\Lambda(k) = k_o^{k_o/(1-k_o)} - k_o^{1/(1-k_o)}$, k_o being the minimum of k.

It can be recalled that the classical correlation coefficient is linearly invariant (independent of the margin parameters) and that all the others are transformation invariant (non-parametric).

Consider, finally, the case for bivariate minima with standard exponential margins. Then the mean values and variances are equal to 1; the covariance (evidently equal to the correlation coefficient) is, using the same formula as the one at the beginning, transformed to survival functions, $(\xi , \eta \geq 0)$

$$\rho = C(\xi , \eta) = \int_0^{+\infty} \int (S(\xi , \eta) - S_o(\xi , \eta))\ d\xi\ d\eta$$

where $S_o(\xi, \eta) = e^{-\xi} e^{-\eta}$ is the survival function for independence. By the change of variables $s = \xi + \eta$, $u = \eta/(\xi + \eta)$ we get

$$\rho = \int_0^1 d\,u \int_0^{+\infty} (e^{-sA(u)} - e^{-s})\, s\, d\, s = \int_0^1 (1/A^2(u) - 1)\, d\, u$$

$$= \int_0^1 A^{-2}(u)\, d\, u - 1 \geq 0 \qquad \text{as} \quad (1/2 \leq) \max(u, 1-u) \leq A(u) \leq 1$$

with the value $\rho = 0$ for independence $(A(u) = 1)$ and $\rho = 1$ for the diagonal case $(A(u) = \max(u, 1-u))$ as should be expected. Recall that ρ is linearly invariant although in this case only scale invariance matters.

The other dependence indicators, being non-parametric, have the same expression, after transformation from $k(w)$ to $A(u)$.

We have then:

for the grade correlation $\chi = 12 \int_0^1 \frac{d\,u}{(1 + A(u))^2} - 3$

for the difference-sign correlation

$$\tau = \int_0^1 \left(\frac{A'(u)}{A(u)}\right)^2 u(1-u)\, d\, u - 2 \int_0^1 \log A(u)\, d\, u\,;$$

for the medial correlation $\nu = 4^{1 - A(1/2)} - 1\,;$

for the index of dependence $\Lambda(A) = A_o^{A_o/(1-A_o)} - A_o^{1/(1-A_o)}$ where A_o $(\geq 1/2)$ denotes the minimum of $A(u)$.

As $\xi = e^{-X}, \eta = e^{-Y}$, $u = (1 + e^{w})^{-1}$, we see that $B(u) = \text{Prob}\{\frac{\eta}{\xi + \eta} \leq u\} =$ $\text{Prob}\{Y - X \geq \log\frac{1-u}{u}\} = 1 - D(\log\frac{1-u}{u}) = u + \frac{u(1-u)\,A'(u)}{A(u)}$ and

$$\tau = 1 - \int_0^1 \frac{B(u)\,(1 - B(u))}{u\,(1-u)}\, d\,u = 1 - \int_0^1 \log\frac{1-u}{u}\, d\,B^2(u) \text{ because } \int_0^1 \log\frac{1-u}{u}\, d\,B(u) = 0$$

as follows from $\int_{-\infty}^{+\infty} w\, d\, D(w) = 0.$

Regression results

Let us now discuss some results on regression in the case of reduced Gumbel margins. For convenience, we will only consider the regression of Y on X , the regression of X on Y being dealt with in the same way with the substitution of k(w) by k(– w) as said before. The linear regression line for reduced margins is evidently $L_y(x) = \gamma + \rho(k)\,(x - \gamma)$, $\rho(k)$ being given before.

As $\text{Prob}\{Y \le y \mid X = x\} = \Lambda(y \mid x)$ is given, in the case of existence of a density, by

$$\Lambda(y \mid x) = \frac{1}{\Lambda'(x)} \frac{\partial\, \Lambda(x\ ,\ y)}{\partial\ x} = \exp\{e^{-x} - (e^{-x} + e^{-y})\, k(y-x)\} \times \{k(y-x)$$

$$+ (1 + e^{-(y-x)})k'\,(y - x)\}$$

the regression line is

$$\bar{y}\,(x \mid k) = \int_{-\infty}^{+\infty} y\, d\, \Lambda(y \mid x)$$

and as $\gamma = \int_{-\infty}^{+\infty} y\, d\, \Lambda(y)$ we get, by integration by parts,

$$\bar{y}\,(x \mid k) = \gamma + \int_{-\infty}^{+\infty} [\Lambda(y) - \Lambda(y \mid x)]\, d\, y$$

$$= \gamma + \int_{-\infty}^{+\infty} \{\exp(-e^{-x}\, e^{-w}) - [k(w) + (1 + e^{-w})k'(w)]\,.$$

$$\exp\{e^{-x}\,[(1 + e^{-w})\, k\,(w) - 1]\}\}\, d\, w$$

$$= \gamma + e^x + \int_{-\infty}^{+\infty} [e^{-e^{-x} e^{-w}} - (1 + e^{-w})\, k(w)\, e^{-e^{-x}[(1+e^{-w})\, k(w) - 1]}\, d\, w\,.$$

The correlation ratio is given by

$$R^2 (y \mid x\,;\, k) = \frac{6}{\pi^2} \int_{-\infty}^{+\infty} (\bar{y}\,(x \mid k) - \gamma)^2\, d\, \Lambda(x)$$

$$= \frac{6}{\pi^2} \int_{-\infty}^{+\infty}\int d\, v\, d\, w \Big[\frac{e^{-v}\, e^{-w}}{1 + e^{-v} + e^{-w}} - 2\, \frac{e^{-v}\, ((1 + e^{-w})\, k(w) - 1)}{e^{-v} + (1 + e^{-w})\, k\,(w)} +$$

$$+ \frac{((1 + e^{-v})\, k(v) - 1)\, ((1 + e^{-w})\, k(w) - 1)}{(1 + e^{-w})\, k\,(v) + (1 + e^{-w})\, k(w) - 1} \Big]\,.$$

Median regression can be defined as the solution of the equation $\Lambda(\tilde{y} \mid x) = 1/2$; it takes the form $\tilde{y}\,(x \mid k) = x + \varphi(x)$ where $\varphi(x)$ is given by $e^{-x}\,(1 + e^{-\varphi})\, k(\varphi) - \log\,[k(\varphi) + (1 + e^{-\varphi})\, k'(\varphi)] = (\log 2)\, e^{-x}$. In two of the models (logistic and mixed) described below the curves are approximately linear — see Gumbel and Mustafi (1968).

Evidently where we have margin parameters we must substitute x and y by $(x - \lambda_x)/\delta_x$ and $(y - \lambda_y)/\delta_y$ in all the three cases above. For instance, for linear regression we have $\frac{L_y(x) - \lambda_y}{\delta_y}$ $= \gamma + \rho\, (\frac{x - \lambda_x}{\delta_x} - \gamma)$ or $L_y(x) = \lambda_y + \gamma\, \delta_y + \rho\, \delta_y \frac{x - \lambda_x - \gamma\, \delta_x}{\delta_x}$, which can take the usual form $L_y(x) = \mu_y + \rho\, \frac{\sigma_y}{\sigma_x}\, (x - \mu_x)$.

Let us return, from now on, to the study of regression for reduced values of (X , Y).

As is well known, see Cramér (1946), the mean-square error of the linear regression for reduced Gumbel margins is

$$MSE(L_y(X)) = M_{X,Y}(Y - \gamma - \rho(X - \gamma))^2 = \frac{\pi^2}{6}\, (1 - \rho^2)$$

when the mean-square error of the general regression is

$$MSE(\bar{y}(X)) = M_{X,Y}(Y - \bar{y}\,(X))^2 = \frac{\pi^2}{6} - M_X(\bar{y}\,(X) - \gamma)^2 = \frac{\pi^2}{6}\, (1 - R^2)$$

which is, as known, equal to the variance $V(Y - \bar{Y})$.

The mean-square error reduction in percentage is

$$NL = \frac{MSE(L_y) - MSE(\bar{y})}{MSE(L_y)} = \frac{R^2 - \rho^2}{1 - \rho^2} ;$$

it can be considered the index of non-linearity, evaluating the improvement in using non-linear regression — see Tiago de Oliveira (1974) .

It can observed that in all the cases studied so far, NL is very small and so the improvement is negligible. This is easily understandable because $L_y(x)$ and $\bar{y}(x)$ are, in fact, very different for $x < -2$ or $x > 5$ whose total probability is very low (about .007 or less) .

The bounds for $k(w)$ can give bounds for $\bar{y}(x)$, using the expression of $\bar{y}(x \mid k) - \gamma$ given before, but the bounds are so large that they are not useful. In fact we have

$$-\gamma - 1 + x - e^{e^{-x}} \int_{e^{-x}}^{+\infty} \frac{e^{-t}}{t}\, dt \le \bar{y}(x \mid k) - \gamma \le e^x (1 - e^{e^{-x}}) + \int_{e^{-x}}^{+\infty} \frac{e^{-t}}{t}\, dt .$$

Let us see how we can, if needed, improve the regression lines, through the use of orthogonal polynomials with respect to $\Lambda(x)$, that we are going to construct.

Let $\{\Psi_p(x)\}$ denote the complete set of orthogonal polynomials with respect to $\Lambda(x)$, i.e.

$$\int_{-\infty}^{+\infty} \Psi_p(x)\ \Psi_q(x)\ d\Lambda(x) = \delta_{pq}$$

where δ_{pq} is the Kroneker symbol ($\delta_{pp} = 1$, $\delta_{pq} = 0$ if $p \neq q$), $\Psi_p(x)$ having the degree p. Recall that $\{(x - \gamma)^p\}$ is a complete set of polynomials (not orthogonal) in any bounded interval.

Let us take $\Psi_0(x) = 1$ and $\Psi_{k+1}(x) = \sum_0^k a_{kj}\, \Psi_j(x) + b_{k+1}(x - \gamma)^{k+1}$. The orthonormality conditions give for $t = 0,1, \ldots , k + 1$

$$\int_{-\infty}^{+\infty} \Psi_{k+1}(x)\ \Psi_t(x)\ d\Lambda(x) = \delta_{k+1,t}$$

which for $t = 0, \quad 1, \ldots, k$ give

$$0 = \sum_{0}^{k} a_{kj} \ \delta_{jt} + b_{k+1} \int_{-\infty}^{+\infty} (x - \gamma)^{k+1} \ \Psi_t(x) \ d\Lambda(x) = a_{kt} + b_{k+1} \ \varphi_{k+1,t}$$

and thus,

$$\Psi_{k+1}(x) = b_{k+1} \left(- \sum_{0}^{k} \varphi_{k+1,t} \ \Psi_t(x) + (x - \gamma)^{k+1}\right) .$$

The normalization equation, for $k + 1$, gives

$$\int_{-\infty}^{+\infty} \Psi^2_{k+1}(x) \, d \, \Lambda(x) = 1$$

or

$$b^2_{k+1} \int_{-\infty}^{+\infty} ((x - \gamma)^{k+1} - \sum_{o}^{k} \varphi_{k+1,t} \ \Psi_t(x))^2 \, d\Lambda(x) = 1$$

or

$$b^2_k = 1/(\mu_{2k+2} - \sum_{o}^{k} \varphi^2_{k+1,t}) ;$$

for simplicity we will take $b_k = (\mu_{2k+2} - \sum_{0}^{k} \varphi^{\,2}_{k+1,t})^{-1/2}$ (the positive root for the radical).

We have, immediately,

$$\Psi_o(x) = 1, \ \Psi_1(x) = \sqrt{\frac{6}{\pi^2}} \ (x - \gamma) ,$$

$$\Psi_2(x) = \frac{6\,(x-\gamma)^2/\pi^2 - \beta_1\sqrt{6/\pi^2}\,(x-\gamma) - 1}{\sqrt{\beta_2 - \beta_1^2 - 1}}\ ,\ \text{etc.}$$

where $\beta_1 = \mu_3/\mu_2^{3/2} = 1.13958$ and $\beta_2 = \mu_4/\mu_2^2 = 5.4$, as known.

Suppose that we want to develop $\bar{y}(x) - \gamma = \sum_0^\infty \rho_p\ \Psi_p(x)$.

We have

$$\frac{6}{\pi^2}\,R^2 = M(\bar{y}(x)-\gamma)^2 = \int_{-\infty}^{+\infty} \Big(\sum_0^\infty \rho_p\ \Psi_p(x)\Big)^2\ d\Lambda(x) =$$

$$\sum_{p,q} \rho_p\ \rho_q \int_{-\infty}^{+\infty} \Psi_p(x)\ \Psi_q(x)\ d\Lambda(x) = \sum_0^\infty \rho_p^2\ .$$

Let us now compute the ρ_j . We have

$$\int_{-\infty}^{+\infty} (\bar{y}(x)-\gamma)\,\Psi_j(x)\,d\Lambda(x) = \sum_0^\infty \rho_p \int_{-\infty}^{+\infty} \Psi_p(x)\,\Psi_j(x)\ d\Lambda(x)$$

$$= \sum_0^\infty \rho_p\ \delta_{pj} = \rho_j\ ,\ \text{i.e.,}$$

$$\rho_j = \int_{-\infty}^{+\infty} (\bar{y}(x)-\gamma))\,\Psi_j(x)\,d\Lambda(x)$$

Thus we have:

$$\rho_o = \int_{-\infty}^{+\infty} 1\,.\,(\bar{y}(x)-\gamma)\,d\Lambda(x) = 0$$

$$\rho_1 = \int_{-\infty}^{+\infty} \Psi_1(x)\,(\bar{y}(x)-\gamma)\,d\Lambda(x) = a_{11} \int_{-\infty}^{+\infty} (x-\gamma)\,(\bar{y}(x)-\gamma)\,d\Lambda(x) =$$

$$= a_{11} \frac{\pi^2}{6} \rho = \sqrt{\frac{\pi^2}{6}} \cdot \rho$$

$$\frac{\pi^2}{6} R^2 = \frac{\pi^2}{6} \rho^2 + \sum_2^\infty \rho_j^2 \qquad \text{or} \qquad R^2 = \rho^2 + \frac{6}{\pi^2} \sum_2^\infty \rho_j^2 .$$

For the independence case we get $\bar{y}(x) = \gamma$, and so as $\rho = 0$,

$$\frac{\pi^2}{6} R^2 = \sum_2^\infty \rho_j^2 \quad , \quad \rho_j = \int_{-\infty}^{+\infty} (\gamma - \gamma) \Psi_j(x) \, d\Lambda(x) \Rightarrow R^2 = \rho^2 = 0 ,$$

and for the diagonal case we obtain $\bar{y}(x) = x$, so

$$\rho_j = \int_{-\infty}^{+\infty} (x - \gamma) \Psi_j(x) \, d\Lambda(x) = \frac{1}{a_{11}} \int_{-\infty}^{+\infty} \Psi_1(x) \, \Psi_j(x) \, dx =$$

$$= \frac{1}{a_{11}} \delta_{pj} \Rightarrow \rho_j = 0 , \qquad \text{except} \qquad \rho_1 = \frac{1}{a_{11}} = \sqrt{\frac{\pi^2}{6}} ,$$

and so

$$\frac{\pi^2}{6} R^2 = \frac{\pi^2}{6} \Rightarrow R^2 = 1 .$$

Miscellaneous results

The results contained in this section were, in part, given in Tiago de Oliveira (1962/63) and (1964). They refer to bivariate pairs with reduced Gumbel margins.

Let us prove that

If (X, Y) is an extremal pair, the distribution of (Y \|X) is extremal if and only if (X , Y) is an independent pair.

It is evident that if (X, Y) is an independent pair (Y | X) = Y is an extremal variate. Let us prove the converse.

Denoting by (as in the previous consideration of regression)

$$\Lambda(y \mid x) = e^{-(e^{-x} + e^{-y})\, k(y-x) + e^{-x}} \quad \{k(y-x) + (1 + e^{-y+x})\, k'(y-x)\}$$

the distribution function of $(Y \mid X)$, we must have

$$\Lambda(y \mid x) = \Lambda\left(\frac{y - \nu(x)}{\tau(x)}\right) .$$

This relation takes the form

$$(e^{-x} + e^{-y})\, k(y-x) - e^{-x} - \log\{k(y-x) + (1 + e^{-(y-x)})\, k'(y-x)\} = e^{-(y-\nu(x)/\tau(x))} .$$

Taking now $x_1 < x_2$ such that $\tau(x_1) > 0$, $\tau(x_2) > 0$, putting $y = w + x_1$ and $y = w + x_2$ and subtracting, we obtain

$$(1 + e^{-w})\, k(w) = 1 + \alpha\, e^{-\beta w} - \alpha'\, e^{-\beta' w} ,$$

$\beta = \tau(x_1)^{-1} > 0$, $\beta' = \tau(x_2)^{-1} > 0$ and $\alpha \geq 0$ and $\alpha' \geq 0$ are functions of x_1 and x_2.

Take $\beta \geq \beta'$; $k(w)$ can be written as

$$k(w) = \frac{1 + e^{-\beta w}\,(\alpha - \alpha'\, e^{(\beta - \beta')w})}{1 + e^{-w}}$$

and as $k(-\infty) = 1$ we obtain

$$k(-\infty) = \lim_{w \to -\infty} \alpha\, e^{(1-\beta)w} = 1$$

so that $\beta = 1$, $\alpha = 1$. Consequently we have

$$k(w) = 1 - \alpha' \frac{e^{-\beta' w}}{1 + e^{-w}} \qquad (\beta' \leq 1) ,$$

Condition II) gives $\alpha' \leq 0$ so that $\alpha' = 0$ and $k(w) = 1$. The case $\beta' \leq \beta$ is dealt with in the same way.

From the expression $\rho = -\frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log\, k(w)\, d\, w$, we see that $\rho = 0$ if and only if $k(w) = 1$, that is, if the extremal pair is independent.

An extremal pair (X , Y) is independent if and only if any pair (X_θ , Y_θ), $X_\theta = \cos\theta X + \operatorname{sen}\theta Y + a$, $Y_\theta = -\operatorname{sen}\theta X + \cos\theta Y + b$ *for any* $\theta \neq 0, \pi/2, \pi, 3\pi/2$ *have equal variances.*

The result follows immediately from the fact that

$$V(X_\theta) = \frac{\pi^2}{6}\,(1 + \rho \operatorname{sen} 2\theta)$$

$$V(Y_\theta) = \frac{\pi^2}{6}\,(1 - \rho \operatorname{sen} 2\theta)\,.$$

A side result is that

$Z = a + a' X + a'' Y$ *has an extremal distribution when (X , Y) is an extremal independent pair if and only if* $a' = 0$ *or* $a'' = 0$.

One half of the proof is straightforward. Let us prove the converse. We can suppose that X, Y, and Z have reduced extremal distributions, only having to transform linearly Z if this is not the case (the coefficients of X and Y being proportional to a' and a''). As $M(X) = M(Y) = M(Z) = \gamma$ we obtain

$$Z - \gamma = a'\,(X - \gamma) + a''\,(Y - \gamma)\,.$$

The computation of the second and fourth central moments immediately gives

$$a'^2 + a''^2 = 1$$

$$\beta_2\,(a'^4 + a''^4) + 6\,a'^2\,a''^2 = \beta_2\,,$$

and so

$$(6 - 2\,\beta_2)\,a'^2\,a''^2 = 0 \;;\; \text{as } \beta_2 = 5.4 \text{ we have } a' = 0 \text{ or } a'' = 0.$$

As an immediate consequence the pair (X,Y) ($X = a + a' X_0 + a'' Y_0$, $Y = b + b' X_0 + b'' Y_0$) is an extremal pair, (X_0 , Y_0) being an independent extremal pair, if and only if $a'' = b' = 0$ or $a' = b'' = 0$.

Another result, corresponding to the characterization of $\Lambda(x , y)$, *is that* $Z = \max(\alpha X + a , \beta Y + b)$ $(\alpha , \beta > 0)$ *is extremal, (X , Y) being a reduced extremal pair, if and only if* $\alpha = \beta$.

As $\Lambda(z) = \text{Prob}\{\max(\alpha X + a , \beta Y + b) \leq z\} = \Lambda(\frac{z-a}{\alpha} , \frac{z-b}{\beta})$ we must seek conditions such that

$$\Lambda(\frac{z-a}{\alpha} , \frac{z-b}{\beta}) = \Lambda(\frac{z-\lambda}{\delta})$$

or such that $(z = \lambda + \delta v)$

$$\Lambda(v) = \Lambda(\frac{\lambda-a}{\alpha} + \frac{\delta}{\alpha} v, \frac{\lambda-b}{\beta} + \frac{\delta}{\beta} v) = \Lambda(a' + \alpha' v , b' + \beta' v) ,$$

that is

$$e^{-v} = (e^{-a'-\alpha' v} + e^{-b'-\beta' v}) k (b' + \beta' v - a' - \alpha' v) .$$

Multiplying this equation by e^{v} and letting $v \to \pm\infty$ we have a limit if and only if $\alpha' = \beta' = 1$ or $\alpha = \beta = \delta$ and then

$$e^{\lambda/\delta} = (e^{a/\delta} + e^{b/\delta}) k (\frac{a-b}{\delta}) ,$$

which relates λ and δ.

$\Lambda(x , y|\theta) = exp(-(e^{-x} + e^{-y}) k(y - x|\theta))$ cannot have sufficient statistics (of rank 1) if for some $\theta = \theta_0$ we have independence.

As is well known, if $\Lambda(x , y|\theta)$ does have a sufficient statistic for θ (of rank 1), its density

$$\frac{\partial^2 \Lambda(x , y|\theta)}{\partial x \partial y} = \exp\{-(e^{-x} + e^{-y}) k(y - x|\theta)\} . \{e^{-x} P(y - x|\theta) + e^{-2x} Q(y - x|\theta)\}$$

with

$$P(w|\theta) = (1 + e^{-w}) k'' (w|\theta) + (1 - e^{-w}) k'(w|\theta) \geq 0$$

and

$$Q(w|\theta) = \{k (w|\theta) + (1 + e^{-w}) k'(w|\theta)\} . \{e^{-w} k(w|\theta) - (1 + e^{-w}) k'(w|\theta)\} \geq 0$$

would be of the form

$$\frac{\partial^2 \Lambda(x\ ,\ y|\theta)}{\partial\ x\ \ \partial\ y} = e^{-a(\theta)b(x,y) - c(x,y) - d(\theta)}\ .$$

For $\theta = \theta_o = 0$ we will suppose independence and thus $P(w|0) = 0$ and $Q(w|0) = e^{-w}$; we can suppose $a(0) = d(0) = 0$ by a simple transformation and, as $a'(0)$ is a factor of the variance, we can suppose $a'(0) = 1$.

Then, putting $\theta = 0$, we obtain $c(x\ ,\ y)\ = x + y + e^{-x} + e^{-y}$, so that the condition for sufficiency can be written

$$(e^{-x} + e^{-y})\ k(y - x|\theta) - \log\ \{P(y - x|\theta) + e^{-x}\ Q(y - x|\theta)\}$$

$$= a(\theta)\ b(x\ ,\ y) + y + e^{-x} + e^{-y} + d(\theta)\ .$$

Denoting by $h(w) = \dfrac{\partial\ k\ (w|\theta)}{\partial\ \theta}\Big|_{\theta = 0}$ and deriving the relation above in order to θ, we obtain

$$(e^{-x} + e^{-y})\ h(y - x) = e^{y-x}\ \{\frac{\partial\ P\ (y - x|\theta)}{\partial\ \theta}\Big|_{\theta = 0} + \ e^{-x}\ \frac{\partial\ Q\ (y - x|\theta)}{\partial\ \theta}\Big|_{\theta = 0}\} + b(x\ ,\ y) + d'(0)$$

so that

$$b(x\ ,\ y) = e^{-x}\ \alpha(y - x) + \beta(y - x)\ ,$$

and the sufficiency condition has the form

$$(e^{-x} + e^{-y})\ \ k(y - x|\theta) - \log\ \{P(y - x|\theta) + e^{-x}\ Q(y - x|\theta)\} =$$

$$y + e^{-x} + e^{-y} + a(\theta)\ (e^{-x}\ \alpha\ (y - x) + \beta\ (y - x)) + d(\theta)\ .$$

Putting $y = w + x$ and letting $x \to +\infty$ we see that we must have $P(w|\theta) = 0$ for any θ , which implies independence.

As regards correlation, we can also give some results.

The first one is that symmetrization reduces correlation.

If $\Lambda_1(x\ ,\ y)$ and $\Lambda_2(x\ ,\ y)$ are two distribution functions of bivariate reduced extremes with dependence functions $k_1(w)$ and $k_2(w)$ and correlation coefficients $\rho_i = -\dfrac{6}{\pi^2}\ \displaystyle\int_{-\infty}^{+\infty} \log\ k_i(w)\ d\ w\ ,$

consider now the dependence function $\bar{k}(w) = (k_1(w) + k_2(w))/2$ — also a dependence function by the mix-technique — and as $\frac{k_1 + k_2}{2} \geq \sqrt{k_1 k_2}$ we obtain

$$\bar{\rho} = -\frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log \bar{k}(w)\, dw \leq -\frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log \sqrt{k_1(w)\ k_2(w)}\, dw = (\rho_1 + \rho_2)/2.$$

Reference should be made to Tiago de Oliveira (1991').

Also if $k_o \leq k(w)$, we know that $\max(k_o, \frac{\max(1, e^w)}{1 + e^w}) \leq k(w)$

and thus

$$\int_{-\infty}^{+\infty} \log\ k(w)\, dw \geq \int_{-\infty}^{+\infty} \log \max(k_o, \frac{\max(1, e^w)}{1 + e^w})\, dw =$$

$$= 2 \int_{o}^{+\infty} \log \max(k_o, \frac{e^w}{1 + e^w})\, dw = 2 \Big[\int_{o}^{\log(k_o/(1-k_o))} \log\ k_o\, dw$$

$$+ \int_{\log(k_o/(1-k_o))}^{+\infty} \log \frac{e^w}{1 + e^w}\, dw \Big] = 2 \int_{1}^{1/k_o} \frac{\log(\xi - 1)}{\xi}\, d\xi,$$

and, thus, denoting by

$$\varphi(k_o) = -\frac{12}{\pi^2} \int_{1}^{1/k_o} \frac{\log(\xi - 1)}{\xi}\, d\xi,$$

we see that

$$\varphi(1) = 0 \leq \rho \leq \varphi(k_o) \leq \varphi(1/2) \leq = 1.$$

$\varphi(k_o)$ is a decreasing function of k_o, and the relation for k_o can be rewritten as $1/2 \leq k_o \leq \varphi^{-1}(\rho) \leq 1$; and as the index of dependence $\Delta(k_o)$ is also a decreasing function of k_o, we get

$$\Delta(1) = 0 \leq \Delta(\varphi^{-1}(\rho)) \leq \Delta(k_o) \leq \Delta(1/2) = 1/4$$

and so the knowledge of ρ gives a lower bound for $\Delta(k_o)$.

An inequality with the reverse direction was not obtained although such a result seems possible as we know that for $\rho = 0$ we have $k_0 = 1$ ($k(w) = 1$, independence) and for $\rho = 1$ we have $k_0 = 1/2$ ($k(w) = \frac{\max(1, e^w)}{1 + e^w}$, diagonal case) against the fact that by the previous inequality for $\rho = 1$ we have $k_0 = 1/2$ and $\Delta(1/2) = 1$ but for $\rho = 0$ we have $k_0 = 1$ and $0 \le \Delta(k_0) \le 1$.

An important and unexpected relation between ρ and τ is

$$(1 - \tau)^2 \le \frac{4\pi^2}{9} (1 - \rho) .$$

In fact, as we have $\rho = 1 - \frac{3}{\pi^2} \int_{-\infty}^{+\infty} w^2 \, d\, D(w)$ and $\tau = 1 - \int_{-\infty}^{+\infty} w \, d\, D^2(w)$, the Schwarz inequality gives the desired result.

The relation can be written as $1 - \tau \le \frac{2\pi}{3} \sqrt{1 - \rho}$ or $1 - \frac{2\pi}{3} \sqrt{1 - \rho} \le \tau$. Then we get

$$1 - \min \left(\frac{2\pi}{3} \sqrt{1 - \rho} , 1 \right) \le \tau \le 1$$

which improves the usual $0 \le \tau \le 1$ if $\frac{2\pi}{3} \sqrt{1 - \rho} \le 1$ or $\rho > 1 - (\frac{2\pi}{3})^2 = .77203$.

Also, given τ, we get $0 \le \rho \le 1 - \frac{9}{4\pi^2} (1 - \tau)^2$.

Let $k_1(w)$ and $k_2(w)$ be the dependence functions and $D_1(w)$ and $D_2(w)$ the distribution functions of the reduced extremes of the distribution functions $\Lambda_1(x , y)$ and $\Lambda_2(x , y)$.

The difference between the correlation coefficients is

$$\rho_2 - \rho_1 = - \frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log \frac{k_2(w)}{k_1(w)} \, d\, w .$$

Let us denote by $Q_1 = \inf \frac{k_2(w)}{k_1(w)}$ and $Q_2 = \sup \frac{k_2(w)}{k_1(w)}$; as for $w \to \pm \infty$ we know that $k_2(w)/k_1(w) \to 1$ and by the relation $k_D(w) = 1/1 + e^{-|w|} \le k_1(w)$, $k_2(w) \le 1 = k_I(w)$, we get the inequalities $1/2 \le Q_1 \le 1 \le Q_2 \le 2$.

As we have $Q_1\, k_1(w) \le k_2(w) \le Q_2\, k_I(w)$ and by the previous relations we get $\max(k_D(w)\,,\, Q_1\, k_1(w)) \le k_2(w) \le \min(1,\, Q_2\, k_1(w))$, then

$$-\frac{6}{\pi^2}\int_{-\infty}^{+\infty}(\log\ \min(1, Q_2\, k_I(w)) - \log k_1(w))\ d\,w \le \rho_2 - \rho_1 \le$$

$$-\frac{6}{\pi^2}\int_{-\infty}^{+\infty}(\log\ \max(k_D(w)\,,\, Q_1\, k_1(w)) - \log k_1(w))\ d\,w\,;$$

$$-\frac{6}{\pi^2}\int_{-\infty}^{+\infty}(\log\ \min(Q_2\,,\, 1/k_1(w))\ d\,w \le \rho_2 - \rho_1 \le -\frac{6}{\pi^2}\int_{-\infty}^{+\infty}(\log \max(\frac{k_D(w)}{k_1(w)}\,,\, Q_1)\ d\,w$$

with $k_D(w) \le k_1(w)\,,\, k_2(w) \le 1$.

Then $1/k_1(w) \le 1/k_D(w)$ and, thus, $\min(Q_2\,,\, 1/k_1(w)) \le \min(Q_2\,,\, 1/k_D(w))$ and so

$$-\frac{6}{\pi^2}\int_{-\infty}^{+\infty}\log\min(Q_2\,,\, 1/k_D(w))\ d\,w \le -\frac{6}{\pi^2}\int_{-\infty}^{+\infty}\log\min(Q_2\,,\, 1/k_1(w))\ d\,w\,.$$

But

$$\frac{6}{\pi^2}\int_{-\infty}^{+\infty}\log\min(Q_2\,,\, 1+e^{-|w|})\ d\,w = \frac{12}{\pi^2}\int_{-\infty}^{+\infty}\log\min(Q_2\,,\, 1+e^{-w})\ d\,w =$$

$$= \frac{12}{\pi^2}\Big[\int_{0}^{-\log(Q_2-1)}\log Q_2\ d\,w + \int_{-\log(Q_2-1)}^{+\infty}\log(1+e^{-w})\ d\,w =$$

$$= \frac{12}{\pi^2}\{-\log(Q_2-1)\,.\,\log Q_2 + \int_0^{Q_2-1}\frac{\log(1+t)}{t}\ d\,t\Big]$$

$$= \frac{12}{\pi^2}\{-\log(Q_2-1)\,.\,\log Q_2 + \log(1+t)\ \log t\,\Big|_0^{Q_2-1} - \int_0^{Q_2-1}\frac{\log t}{1+t}\ d\,t\}$$

$$= -\frac{12}{\pi^2}\int_0^{Q_2-1}\frac{\log t}{1+t}\ d\,t = -\frac{12}{\pi^2}\int_1^{Q_2}\frac{\log(\xi-1)}{\xi}\ d\,\xi = \varphi(Q_2^{-1}) \quad \text{and so}$$

$$-\varphi(Q_2^{-1}) \le \rho_2 - \rho_1 \ .$$

But also

$$1 \le 1/k_1(w)) \quad \text{so} \quad \max\left(\frac{k_D(w)}{k_1(w)},\ Q_1\right) \ge \max(k_D(w)\ ,\ Q_1)$$

and then

$$-\frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log\max\left(\frac{k_D}{k_1},\ Q_1\right) dw \le -\frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log\max(k_D\ ,\ Q_1)\ d\,w\ ;$$

$$= \frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log\min\left(\frac{1}{k_D},\ \frac{1}{Q_1}\right) d\,w = \varphi(Q_1)$$

and, thus, we get

$$-\varphi(Q_2^{-1}) \le \rho_2 - \rho_1 \le \varphi(Q_1) \ .$$

Also, the exchange of k_1 and k_2 (exchanging Q_1 and Q_2 for Q_2^{-1} and Q_1^{-1}) gives the same result and, so, $|\rho_2 - \rho_1| \le \max(\varphi(Q_2^{-1})\ ,\ \varphi(Q_1))$ and, as $\varphi(Q)$ is an increasing function for $Q^{-1} > 1$, we get

$$|\rho_2 - \rho_1| \le \varphi(\min(Q^{-1}\ ,\ Q_2)) \ .$$

The distance between Λ_1 and Λ_2, analogous to the index of dependence, is

$$\sup_{x,y} |\Lambda_1(x\ ,\ y) - \Lambda_2(x\ ,\ y)| = \max\left\{(Q_2 - 1)\ Q_2^{-\frac{Q_2}{Q_2 - 1}}\ ,\ (1 - Q_1)\ Q_1^{\frac{Q_1}{1 - Q_1}}\right\}\ ,$$

where Q_1 and Q_2^{-1} are analogous to k_o in the beginning, and the technique of computation is the same. As $(Q_2 - 1)\ e^{-Q_2/(Q_2 - 1)} \le Q_2 - 1$ and $(1 - Q_1)\ Q_1^{\,Q_1/(Q_1 - 1)} \le 1 - Q_1$ are rough approximations, a very rough approximation is $\sup|\Lambda_1 - \Lambda_2| \le \max\ (Q_2 - 1\ ,\ 1 - Q_1)$. If we consider the two non-differentiable models, biextremal with

$$k_1(w\ |\theta) = 1 - \frac{\min(\theta\ ,\ e^w)}{1 + e^w}$$

and its dual with

$$k_2(w\,|\theta) = 1 - \frac{\min(1\ ,\ \theta\ e^w)}{1 + e^w}\quad ,$$

where $0 \le \theta \le 1$, both with correlation coefficients

$$\rho_1(\theta) = \rho_2(\theta) = -\frac{6}{\pi^2}\int_0^\theta \frac{\log\ t}{1-t}\,d\,t\,,$$

we have

$$Q_2 = \sup_w \frac{k_2(w\,|\theta)}{k_1(w\,|\theta)} = 1 + \theta\,(1-\theta)\quad ,\quad Q_1 = \inf_w \frac{k_2(w\,|\theta)}{k_1(w\,|\theta)} = 1/Q_2$$

so that, in this case,

$$\sup_{x,y} |\,\Lambda_1(x\ ,\ y) - \Lambda_2(x\ ,\ y)\,| = (Q_2 - 1)\,Q_2^{-\frac{Q_2}{Q_2 - 1}}\ ;$$

this distance is an increasing function of $Q_2 = 1 + \theta(1-\theta)$ whose maximum value is $Q_2 = 1 + 1/4$. So, with equal correlation coefficients for the biextremal model and its dual, we have the maximum distance equal to

$$\frac{1}{4}\ x\ (\frac{5}{4})^{-\frac{5/4}{1/4}}\ ;\ = 4^4/5^5 = .08192\ .$$

Another expression for the difference of the correlation coefficients is

$$\rho_2 - \rho_1 = \frac{3}{\pi^2}\int_{-\infty}^{+\infty} w^2\ d(D_1(w) - D_2(w)) = \frac{6}{\pi^2}\int_{-\infty}^{+\infty} w\ (D_2(w) - D_1(w))\,d\,w\ .$$

Also for the difference-sign correlation we have

$$\tau_2 - \tau_1 = \frac{3}{\pi^2}\int_{-\infty}^{+\infty} w\ d(D_1^2(w) - D_2^2(w)) = \frac{3}{\pi^2}\int_{-\infty}^{+\infty} w\ (D_2^2(w) - D_1^2(w))\,d\,w\ .$$

From the relations written between ρ and τ we get

$$|\rho_1 - \rho_2| = \frac{9}{4\pi^2} \max\{(1-\tau_1)^2, (1-\tau_2)^2 = \frac{9}{4\pi^2}(1 - \min(\tau_1, \tau_2))^2$$

and, conversely, we obtain

$$(\tau_1 - \tau_2) \leq \max(\min(\frac{2\pi}{3}\sqrt{1-\rho_1}, 1), \min(\frac{2\pi}{3}\sqrt{1-\rho_2}, 1))$$

$$= \min(1, \frac{2\pi}{3}\max(\sqrt{1-\rho_1}, \sqrt{1-\rho_2})) = \min(1, \frac{2\pi}{3}\sqrt{1 - \min(\rho_1, \rho_2)}.$$

Chapter XI

BIVARIATE EXTREMES; MODELS AND STATISTICAL DECISION

Introduction

As regards statistical decision for bivariate extremes, we will use an approach similar to the one used for univariate extremes: we will replace, for large samples, $F^n(x , y)$ by an asymptotic approximation

$$L\left(\frac{x - \lambda}{\delta} , \frac{x - \lambda'}{\delta'}\right) ,$$

where $L(x , y)$ will have the expressions given before, thus reducing the difficulty of not knowing $F^n(x , y)$ to the simpler (!) difficulties of not knowing the dependence structure ($k(.)$ or $A(.)$ for instance) and four independent margin parameters. Recall that we have the results similar to the ones used to justify the substitution of $F^n(x)$ by $L\left(\frac{x - \lambda}{\delta}\right)$ in the univariate case :

The stability and the uniformity of convergence are valid as well as results similar to those of Watson and others, showing that even in the case of some dependence between the successive observations the limiting distributions are still the same, are also true.

In the asymptotic approximation of $L\left(\frac{x - \lambda}{\delta} , \frac{y - \lambda'}{\delta'}\right)$ to the actual and unknown $F^n(x , y)$, we will use, in principle, a dependence function, although Geffroy's sufficient condition could suggest the independence model. The existence of a strong correlation between the maxima (or minima), a correlation not completely destroyed in some applications by the relatively large number of observations from which the maxima (or minima) are taken, suggests the use of an asymptotic bivariate extremes distribution function as a better approximation to the actual (unknown) distribution function of bivariate extremes. This stand is also justified by the fact that the discrepancy between the independence and the dependence cases may be very large in a probabilistic sense but not expressed by the usual indicators. As an example we can cite the fact that the uniform distance between the biextremal model with standard Gumbel margins, where the distribution function is $\Lambda(x , y \mid \theta) = \exp\{- e^{-x} + e^{-y} - \min(e^{-x} , \theta\, e^{-y})\}$ (note that

$\Lambda(x , y \mid 0) = \Lambda(x) . \Lambda(y)$ — independence — and that $\Lambda(x , y \mid 1) = \Lambda(\min(x , y))$, — the diagonal case—) , and the independence case is

$$\sup_{x,y} | \Lambda(x , y|\theta) - \Lambda(x , y|0) | = \theta (1 + \theta)^{-1-1/\theta} ;$$

consequently if we take the independence model as an approximation to a biextremal model with θ close to 1, i.e., close to the diagonal case, we could obtain probability evaluations for quadrants with an error around 1/4, the value of the distance for $\theta = 1$, but that can differ by almost 1 if the set used is a strip containing the first diagonal !

Experience will, finally, decide the fitness and accuracy of those approximations. But, for the moment, as the (m,N) natural model shows, we can have any finite number of parameters, forgetting the four of the margins. And it also can be shown that for no one-parameter models (with standard margins), do there exist sufficient statistics. An alternative is the intrinsic estimation of the dependence function (k or A), as will be seen.

For some models we do not have the best test, the best estimation procedure, etc. (we may have one, but not necessarily the best) although in other situations nothing at all is known. An example: the statistical choice of bivariate extreme models — so important for applications — is now beginning to be considered! In general, the few papers up to now have chosen one model from the beginning or compare two of them by the use of the Kolmogoroff-Smirnov statistic or another test; see, for instance, Gumbel and Golstein (1964).

As a general rule in statistical decision for bivariate maxima with Gumbel margins, the papers Tiago de Oliveira (1970), (1975), (1975'), (1980), (1984) and (1989) contain results that are improved here.

As regards modelling, we have to deal with some models which have a planar density and with others, which appear naturally, that have a singular component.

Finally, let us recall that as the margins in both cases (Gumbel and exponential) have all moments, we also have crossed moments, etc., a fact that will be used.

Note that for all models to be considered, we only give the correlation coefficients and index of dependence when they have a compact form. The correlation coefficients can then easily be used for statistical decision. The non-parametric correlation coefficients and the index of dependence are the same for the corresponding $\Lambda(x , y \mid \theta)$ and $S(x' , y' \mid \theta)$; the classical correlation coefficient is different.

It is important to note that separation between models is difficult for moderate samples. A graphical explanation is the essence of exercise III.23.

The differentiable models; statistical decision

Until now only two models of bivariate extremes with a planar density (with a point exception in one case) have been considered and studied in some depth in the literature: the logistic model and the mixed model, both appearing in Gumbel (1962) and (1966). Both are symmetrical but nothing precludes the existence of asymmetric models, which is the case in the majority of non-differentiable models to be considered in the next section. Here, as well as for non-differentiable models, we will always study them under the form of bivariate maxima with standard Gumbel margins, i.e., with the distribution function $\Lambda(x , y) = (\Lambda(x)\,\Lambda(y))^{k(y-x)}$ with the dependence function $k(w)$; at the end of each case we will consider the corresponding forms for bivariate minima with standard exponential margins, i.e., with the survival function $S(x' , y') = \exp\{ -(x' + y')\, A\,(\frac{y'}{x' + y'})\}$ $(x' , y' \geq 0)$ obtained by the transformation $X' = e^{-X}$, $Y' = e^{-Y}$ and with $A(u) = k(\log(\frac{1-u}{u}))$, as known.

The *logistic model* has the distribution function $\Lambda(x , y|\theta) = \exp\{- e^{-x/(1-\theta)} + e^{-y/(1-\theta)})^{1-\theta}\}$, $0 \leq \theta < 1$ (for $\theta = 0$ we have independence and for $\theta = 1^-$ we get the diagonal case, which is the only situation without density). It is so called because

$$\text{Prob}\,\{Y - X \leq w|\theta\} = D(w|\theta) = (1 + e^{-w/(1-\theta)})^{-1}$$

corresponding to the dependence function

$$k(w|\theta) = \frac{(1 + e^{-w/(1-\theta)})^{1-\theta}}{1 + e^{-w}} ;$$

note that we must have $0 \leq \theta \leq 1$, as comes from the inequality $\frac{\max(1 + e^{w})}{1 + e^{w}} \leq k(w) \leq 1.$

The margins are exchangeable as shown by the expression of $\Lambda(x , y|\theta)$, by the fact that $k(- w|\theta) = k(w|\theta)$ or even because $D(w|\theta) + D(- w|\theta) = 1.$

The equidistribution curves are very easy to draw. The median curve has the equation

$$e^{-x/(1-\theta)} + e^{-y/(1-\theta)} = (\log 2)^{1/(1-\theta)}$$

which is similar, with similarity ratio $1/(1-\theta)$, to the equidistribution curve for the independence case.

The classical correlation coefficient is $\rho(\theta) = \theta(2-\theta)$, increasing from $\rho(0) = 0$ to $\rho(1) = 1$, the difference-sign correlation coefficient is $\tau(\theta) = \theta$, increasing from $\tau(0) = 0$ to $\tau(1) = 1$, the medial correlation coefficient is $\nu(0) = 4^{(1-2^{-\theta})} - 1$, increasing from $\nu(0) = 0$ to $\nu(1) = 1$. Linear regression of Y in X is, evidently, $L_y(x|\theta) = \gamma + \theta(2-\theta)(x-\gamma)$. It can be shown that the index of dependence

$$\sup_{x,y} |\Lambda(x, y|\theta) - \Lambda(x)\, y|0)| = (1-2^{-\theta})\, 2^{-\theta/(2^{\theta}-1)} ;$$

increases from 0 (at $\theta = 0$) to $1/4 = .25$ (at $\theta = 1$). It is, thus intuitive that the distance between the independence ($\theta = 0$) and the assumed model for $\theta > 0$ is small in general, and for small samples it will naturally be difficult to separate models.

Owing to the importance of independence, as a reference because in general we can expect association, we can formulate as confirmation of dependence an L.M.P. test of $\theta = 0$ vs $\theta > 0$. As for the sample $\{(x_1, y_1), ..., (x_n, y_n)\}$ the likelihood is

$$L(x_i, y_i|\theta) = \prod_1^n \frac{\partial^2 \Lambda(x_i, y_i|\theta)}{\partial x_i \, \partial y_i} ,$$

the rejection region has the form

$$\frac{\partial}{\partial \theta} \left(\sum_1^n \frac{\partial^2 \Lambda(x_i, y_i|\theta)}{\partial x_i \, \partial y_i} \right) \Big|_{\theta = 0} > c_n$$

or

$$\sum_1^n \nu(x_i, y_i) \geq c_n$$

where

$$\nu(x, y) = x + y + x\, e^{-x} + y\, e^{-y} + (e^{-x} + e^{-y} - 2)\, \log(e^{-x} + e^{-y}) + \frac{1}{e^{-x} + e^{-y}} .$$

In the case of independence, $v(X, Y)$ has mean value zero but infinite variance; evidently $1/n \cdot \sum_{1}^{n} v(x_i, y_i) \xrightarrow{a.s.} 0$, in the case of independence, by Khintchine's theorem. Although then the usual Central Limit Theorem is not applicable, it was shown by Tawn (1988) that $\sum_{1}^{n} v(x_i, y_i)/\sqrt{(n \log n)/2}$ is asymptotically standard normal. Estimation by the maximum likelihood is possible and Cramér — regular for $\theta \neq 0$, having for $\theta = 0$ the difficulties arising from the infinite variance of $v(X, Y)$. Thus we have $c_n \sim \sqrt{(n \log n)/2}\ N^{-1}(1 - \alpha)$, where $N(.)$ denotes the standard normal distribution function and α the significance level of the one-sided tests.

As θ is a dependence parameter, it is natural to use the correlation coefficients whose estimators are, as is well known, independent of the location and dispersion parameters of the margins, to test $\theta = 0$ vs. $\theta > 0$. It was shown, in Tiago de Oliveira (1964) and (1975), that the most efficient is the classical correlation coefficient. Recall that, in the case of independence, $\sqrt{n}\, r$ is asymptotically standard normal. As $\rho(\theta) = \theta(2 - \theta)$, a point estimator of θ is given by $r = \theta^*(2 - \theta^*)$ or $\theta^* = 1 - \sqrt{1 - r}$, which leads to no truncation of the estimator if $r \geq 0$; otherwise we will take $\theta^* = 0$ (independence).

A quick approach may be the quadrants method (see Annex V); as the medians of the margins, in reduced form, are $\tilde{\mu} = -\log\log 2 = 0.3665129$ we have $p(\theta) = \text{Prob}\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\} = \exp(-2^{1-\theta} \log 2)$, so the probability that both components are at the same side of the medians, i.e., $\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\}$ or $\{X > \tilde{\mu}, Y > \tilde{\mu}\}$ is $2p(\theta)$. If N denotes the number of sample pairs with components both smaller or both larger than the margin medians $\tilde{\mu}$, we know that $\sqrt{n}\, \dfrac{N/n - 2p(\theta)}{\sqrt{2p(\theta)\,(1 - 2p(\theta))}}$ is asymptotically standard normal. The point estimator of θ is then given by

$$\theta^{**} = 1 - \frac{\log(\log(2n/N)/\log 2)}{\log 2}$$

and the condition $0 \leq \theta^* \leq 1$ is equivalent to $1/2 \leq N/n \leq 1$; as the RHS inequality is trivial, if the LHS one is violated (i.e., $n > 2N$) we get $\theta^{**} < 0$ and we must assume independence (if the model is valid).

Notice that $p(\theta) = \exp\{(-2^{1-\theta} \log 2)\}$ increases from $p(0) = 1/4$ to $p(1) = 1/2$.

It is easy, by the use of the δ-method, to show that $\log 2 \cdot \sqrt{\frac{n\,N}{n-N}}\ \log\frac{2\,n}{N}\,(\theta^{**}-\theta)$ is asymptotically standard normal.

The logistic model for bivariate minima with standard exponential margins is immediately obtained by the transformation $X' = e^{-X}$, $Y' = e^{-Y}$ $(A(u) = k(\log(\frac{1-u}{u}))$ has the following survival function:

$$S(x', y'|\theta) = P(X' \geq x', Y' \geq y') = \exp\{-\{x'^{1/(1-\theta)} + y'^{1/(1-\theta)}\}^{1-\theta}\}$$

for $x', y' \geq 0$; also we get $A(u|\theta) = (u^{1/(1-\theta)} + (1-u)^{1/(1-\theta)})^{1-\theta}$ and the equisurvival median curve is $x'^{1/(1-\theta)} + y'^{1/(1-\theta)} = (\log 2)^{1/(1-\theta)}$; for $\theta = 1$ we obtain a straight lines. We also have $B(u|\theta) = \text{Prob}\{\frac{Y'}{X'+Y'} \leq u\} = 1 - D(\log\frac{1-u}{u}|\theta) = \frac{u^{1/(1-\theta)}}{u^{1/(1-\theta)} + (1-u)^{1/(1-\theta)}}$.

The function $v(x', y')$ is $v(x', y') = -(\log x' + x' \log x' + \log y' + y' \log y') + (x'+y'-2)\log(x'+y') + 1/(x'+y')$.

The classical correlation coefficient $\rho = \int_0^1 A^{-2}(u|\theta)\,du - 1$, by the transformation $u = (1 + (\frac{1-t}{t})^{1-\theta})^{-1}$, takes the form $\rho(\theta) = (1-\theta)\int_0^1 t^{-\theta}(1-t)^{-\theta}\,dt - 1 = (1-\theta)\,B(1-\theta, 1-\theta) - 1 = (1-\theta)\frac{\Gamma^2(1-\theta)}{\Gamma(2(1-\theta))} - 1$ (as $B(.,.)$ is the beta function), which increases from $\rho(0) = 0$ to $\rho(1^-) = 1$ as $(1-\theta)\,\Gamma(1-\theta) \to 1$ for $\theta \to 1^-$.

Let us now consider the *mixed model*. Its distribution function is $\Lambda(x, y|\theta) = \exp(-e^{-x} - e^{y} + \theta/(e^{x}+e^{y}))$, $0 \leq \theta \leq 1$. When $\theta = 0$ we have independence but for $\theta = 1$ it does not degenerate, in the diagonal case, as happens with the logistic model.

The model has the dependence function $k(w|\theta) = 1 - \theta\frac{w}{(1+e^{w})^2} = k(-w|\theta)$, and so the model is exchangeable, as also follows from

$$D(w|\theta) = \frac{e^{w}}{1+e^{w}} \cdot \frac{(1+e^{w})^2 - \theta}{(1+e^{w})^2 - \theta\ e^{w}} = 1 - D(-w|\theta)\ .$$

The dependence function $k(w|\theta)$ is the $(1-\theta, \theta)$ mixture of the dependence functions 1 and $1 - \frac{e^w}{(1+e^w)^2}$; the inequality $\frac{\max(1, e^w)}{1+e^w} \le k(w) \le 1$ shows that the domain $[0,1]$, of cannot be enlarged.

The equidistribution curves are drawn as usual. The median curve takes the form $e^{-x} + e^{-y} - \theta/(e^x + e^y) = \log 2$.

The classical correlation coefficient has the expression $\rho(\theta) = \frac{6}{\pi^2} \cdot (\arccos(1-\theta/2))^2$ which increases from $\rho(0) = 0$ to $\rho(1) = 2/3$, the medial correlation coefficient is $\nu(\theta) = 2^{\theta/2} - 1$ increasing from $\nu(0) = 0$ to $\nu(1) = 1$; the model is, thus, not fit to express a strong dependence, as confirmed below by the index of dependence. Linear regression of Y is X is evidently $L_y(x|\theta) = \gamma + \frac{6}{\pi^2} (\arccos(1-\theta/2))^2 (x - \gamma)$. We have also

$$\sup_{x,y} |\Lambda(x, y|\theta) - \Lambda(x)\, y|0)| = \frac{\theta}{4-\theta}(1-\theta/4)^{4/\theta}$$

increasing from 0 (at $\theta = 0$) to $3^3/4^4 = .105$ at = 1. The smaller variation of the correlation coefficient and of the index of dependence shows that the deviation from independence is smaller and more difficult to detect.

The LMP test of $\theta = 0$ vs. $\theta > 0$ is given by the rejection region

$$\sum_1^n v(x_i, y_i) \ge c_n$$

with

$$v(x, y) = 2\frac{e^{2x} \cdot e^{2y}}{(e^x + e^y)^3} - \frac{e^{2x} + e^{2y}}{(e^x + e^y)^2} + \frac{1}{e^x + e^y} .$$

As before, the situation is marred by the fact that for independence, although $v(X, Y)$ has mean value zero, the variance is infinite. The usual Central Limit Theorem cannot be applied but, as before, $1/n. \sum_1^n v(x_i, y_i) \xrightarrow{a.s.} 0$. But as shown also by Tawn (1987), $\sum_1^n v(x_i, y_i) / \sqrt{(n \log n)/15}$ is also asymptotically standard normal; then $c_n \sim \sqrt{(n \log n)/15}\ N^{-1}(1-\alpha)$. Estimation by the maximum likelihood is also possible and Cramér-regular for $\theta \neq 0$ with analogous difficulties for $\theta = 0$, as before.

Correlation coefficients, independent of the margin parameters, are then to be used to test $\theta = 0$ vs. $\theta > 0$. Once more the usual correlation coefficient is the best. The point estimator of θ, if $0 \leq r \leq 2/3$, is given by $\theta^* = 2(1 - \cos\sqrt{\frac{\pi^2}{6} r})$ with the natural truncation otherwise.

A quick approach may also use the quadrants method. As the medians of the margins are $\tilde{\mu} = -\log\log 2$ we have $p(\theta) = \text{Prob}\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\} = 2^{\theta/2}/4$ which varies from $p(\theta) = 1/4$ to $p(1) = \sqrt{2}/4$. If N/n denotes the fraction of sample pairs both smaller or both larger than the margin medians, we know that $\sqrt{n}\,\frac{N/n - 2p(\theta)}{\sqrt{2p(\theta)\,(1 - 2p(\theta))}}$ is asymptotically standard normal. If $1/2 \leq N/n \leq \sqrt{2}/4$ the estimator of θ is then given by $\theta^{**} = 2(1 + \frac{\log N/n}{\log 2})$ being truncated as before when the double inequality does not apply and, once more by the δ-method, we see that $\frac{\log 2}{2}\sqrt{\frac{n\,N}{n - N}}\,(\theta^{**} - \theta)$ is asymptotically standard normal.

The mixed model for bivariate minima with standard exponential margins immediately has the survival function

$$S(x', y'|\theta) = \exp\{-(x' + y') + \frac{\theta\, x' + y'}{x' + y'}\}, \quad x' + y' \geq 0;$$

we have $A(u|\theta) = 1 - \theta\ u(1 - u)$ and the equisurvival median curve is $x' + y' - \theta\frac{x'\ y'}{x' + y'} = \log 2$. We have $B(u|\theta) = 1 - D(\log\frac{1 - u}{u}|\theta) = \frac{u - \theta\ u(1 - u)^2}{1 - \theta\ u(1 - u)}$. The function $v(x', y')$ is $v(x', y') = 2\ \frac{x'.\ y'}{(x' + y')^3} - \frac{x'^2 + y'^2}{(x' + y')^2} + \frac{x'.\ y'}{x' + y'}$. The classical correlation coefficient is simply

$$\rho(\theta) = \frac{8}{(4 - \theta)\sqrt{4\,\theta - \theta^2}}\ \text{tg}^{-1}\sqrt{\frac{\theta}{4 - \theta}} - \frac{2 - \theta}{4 - \theta}$$

increasing from $\rho(0) = 0$ to $\rho(1) = \frac{1}{3}(\frac{4\pi}{3\sqrt{3}} - 1) = .4727997$.

Finally, we discuss the *statistical choice* between the two non-separated models (logistic and mixed), non-separated because they have in common the independence case ($\theta = 0$ in both cases). As developed in Tiago de Oliveira (1984) and generalized in Tiago de Oliveira (1985), the decision rule (with v_L and v_M denoting the v functions for the logistic and mixed models, given before) is:

decide for independence if $\sum_1^n v_L(x_i, y_i) \le a_n$, $\sum_1^n v_M(x_i, y_i) \le b_n$;

decide for the logistic model if $\sum_1^n v_L(x_i, y_i) > a_n$

and $\sum_1^n \frac{v_L(x_i, y_i)}{a_n} > \sum_1^n \frac{v_M(x_i, y_i)}{b_n}$;

decide for the mixed model if $\sum_1^n v_M(x_i, y_i) > b_n$

and $\sum_1^n \frac{v_M(x_i, y_i)}{b_n} > \sum_1^n \frac{v_L(x_i, y_i)}{a_n}$.

It was shown, in Tawn (1987), that the asymptotic distribution of the pair $(\sum_1^n v_L(x_i, y_i)/\sqrt{(n \log n)/2}$, $\sum_1^n v_M(x_i, y_i)/\sqrt{(n \log n)/15})$ is asymptotically binormal independent with standard margins, and so we should take, with the usual meanings, $a_n \sim \sqrt{(n \log n)/2}\, N^{-1}(1 - \alpha/2)$ and $b_n \sim \sqrt{(n \log n)/15}\, N^{-1}(1 - \alpha/2)$.

Let us now make some comments.

It must be strongly emphasized that the tests and estimators based on the correlation coefficients are independent of the values of the margin parameters. However, in the other cases we have to "estimate" reduced values by using $\hat{x}_i = (x_i - \hat{\lambda}_x)/\hat{\delta}_x$ and $\hat{y}_i = (y_i - \hat{\lambda}_y)/\hat{\delta}_y$ in the Gumbel margins case, the effect of which is not known.

It is natural to consider, as a first step in the statistical analysis of random pairs of maxima or minima, a test for independence, although in many cases we may expect its rejection, maxima or minima being almost always dependent, as observed above.

We can now summarize the results for a test of independence given in Tiago de Oliveira (1965').

The comparison of the usual estimators of the classical, difference-sign, grade, and medial correlation coefficients for one - sided tests, owing to the non-negativity of the values of those

coefficients for the bivariate extremal distributions, shows, if we are in the model with bivariate Gumbel margins, that:

a) — the difference-sign and grade correlation are asymptotically equivalent as tests for independence with respect to any (reasonable) system of alternatives considered, and as the Spearman correlation is always asymptotically equivalent to the grade correlation, the three are equivalent in this case; note that as the grade correlation can be optimal for testing for independence — see Tiago de Oliveira (1985') — with $F(x , y|\theta)$ such that $F(x , +\infty|\theta) = F_1(x)$, $F(+\infty, y|\theta) = F_2(y)$, $F(x , y|0) = F_1(x)\, F_2(x)$ only if $\frac{\partial F(x , y|\theta)}{\partial \theta}|_{\theta=0} = F_1(x)\,(1 - F_1(x))\, F_2(y)\,(1 - F_2(y))$ which is not the case; no one of the correlation coefficients tests above can be optimal;

b) — for the two differentiable models it was shown that the classical correlation coefficient

$$r = \sum (x_i - \bar{x})\,(y_i - \bar{y}) / \{ \sum (x_i - \bar{x})^2\,(y_i - \bar{y})^2 \}^{1/2}$$

gives the best test of the ones above based on correlation coefficients; as $\sqrt{n}\, r$ is asymptotically standard normal we will accept the hypothesis of independence if $\sqrt{n}\, r \leq \lambda_\alpha$ where $N(\lambda_\alpha) = 1 - \alpha$;

c) — the Gumbel and Mustafi (1967) test is equivalent to use of medial correlation;

d) — the strip method devised by Posner, Rodemich, Ashlock and Lurie (1969) also has an efficiency smaller than that of the use of the classical product-moment method.

Evidently, some of those results — the non-parametric ones — are immediately extended.

There exist asymmetric extensions of the logistic model; see Tawn (1988). Also there exists another asymmetric model under study, presented in Tiago de Oliveira (1989"), with high correlations ($\rho > .77$); it has a non-differentiable extension such that $0 \leq \rho \leq 1$. It gave a good fit to Fox River.

The non-differentiable models; statistical decision

We know essentially three forms of the non-differentiable models — i.e., with a singular component, thus not having a planar density: the Gumbel model, the biextremal model, and the natural model, the last one being recently generalized by the author (see the next chapter for some

details); but no statistical results have appeared so far except when it degenerates into the natural model.

The *Gumbel model* has the distribution function is $\Lambda(x, y|\theta) = \exp\{-(e^{-x} - e^{y} - \theta \min(e^{-x}, e^{-y})\}$, $0 \le \theta \le 1$; its dependence function is $k(w|\theta) = 1 - \theta \dfrac{\min(1, e^{-w})}{1 + e^{-w}}$ which is the $(1-\theta, \theta)$ mixture of $k_I(w) = 1$ (independence) and $k_D(w) = \dfrac{\max(1, e^{-w})}{1 + e^{-w}}$ (diagonal case). The model is exchangeable as $k(w|\theta) = k(-w|\theta)$ and we have $D(w|\theta) = \dfrac{1-\theta}{1-\theta+e^{-w}}$ if $w < 0$, $D(w|\theta) = \dfrac{e^{w}}{1-\theta+e^{w}}$ if $w > 0$, with a jump of $\dfrac{\theta}{2-\theta}$ at $w = 0$, which is probability $P(Y = X)$; if $w \neq 0$ we have $D(w|\theta) + D(-w|\theta) = 1$. This model was introduced in Tiago de Oliveira (1971) and is a conversion, to Gumbel margins, of a bivariate model with exponential margins, due to Matshall and Olkin (1967).

The median equidistribution curve is, evidently, $e^{-x} + e^{-y} - \theta \min(e^{-x}, e^{-y}) = \log 2$. The classical correlation coefficient can take the form $\rho(\theta) = \dfrac{12}{\pi^2} \int_0^{\theta} \dfrac{\log(2-t)}{1-t}\, dt$, the difference-sign correlation coefficient is $\tau(\theta) = \theta/(2-\theta)$, the grade correlation coefficient is $\chi(\theta) = \dfrac{3\theta}{4-\theta}$, and the medial correlation coefficient is $\nu(\theta) = 2^{\theta} - 1$, all the coefficients increasing from 0 to 1 as θ varies from 0 to 1. The index of dependence is

$$\sup_{x,y} |\Lambda(x, y|\theta) - \Lambda(x)\, y|0)| = \frac{\theta}{2-\theta}(1-\theta/2)^{2/\theta}$$

increasing also from 0 to 1/4 as θ increases from 0 to 1.

The use of formulae given for curvilinear regression leads to

$$\bar{y}(x|\theta) = \gamma + \log(1-\theta) + \int_{(1-\theta)e^{-x}}^{+\infty} e^{-t}/t \,.\, dt - (1-\theta)\exp(\theta\, e^{-x}) \int_{e^{-x}}^{+\infty} e^{-t}/t \,.\, dt$$

$$= x + \int_0^{(1-\theta)e^{-x}} \frac{1-e^{-t}}{t}\, dt - (1-\theta)\exp(\theta\, e^{-x}) \int_{e^{-x}}^{+\infty} \frac{e^{-t}}{t}\, dt\,.$$

The form of $\bar{y}(x|\theta)$ is very different from the linear regression $\gamma + \rho(\theta)(x - \gamma)$, but the non-linearity index NL oscillates between 0 (from $\theta = 0$ to $\theta = 1$) and .006; the improvement is very small in the passage from the (simple) linear regression to the (more complex) curvilinear regression: in fact, a large difference between the two regressions only happens for very small or very large values of X, whose probability is very small: see Tiago de Oliveira (1974).

Let us now consider statistical decision for this model.

The estimation of the margin parameters is avoided using the correlation coefficients. The test for independence can be done, as before, based on the fact that if $\theta = 0$ (independence) $\sqrt{n}\, r$ is asymptotically standard normal.

The method of quadrants can be used, once more with $p(\theta) = \text{Prob}\{X \leq \tilde{\mu}\ ,\ Y \leq \tilde{\mu}\} = 2^{\theta}/4 = 2^{\theta-2}$, increasing from $p(0) = 1/4$ to $p(1) = 1/2$. Then $\sqrt{n}\ \dfrac{N/n - 2p(\theta)}{\sqrt{2p(\theta)\,(1 - 2p(\theta))}}$ is asymptotically standard normal and $\theta^{**} = \dfrac{\log N/n}{\log 2} + 1$ (supposing that $N/n \geq 1/2$) is an estimator of θ, with truncation if necessary. Also it is immediate to show that $\sqrt{n}\ \dfrac{\log 2}{\sqrt{2^{1-\theta} - 1}}\ (\theta^{**} - \theta)$ and $(\log 2)\sqrt{\dfrac{n\ N}{n - N}}\ (\theta^{**} - \theta)$ are asymptotically standard normal.

The Gumbel model for bivariate minima with standard exponential margins was presented, in a different approach, as said before, by Marshall and Olkin (1967). We have $S(x', y'|\theta) = e^{-x' - y' + \theta \min(x', y')}$, $0 \leq \theta \leq 1$; we have $A(u|\theta) = 1 - \theta \min(u, 1 - u)$ a mixture $(\theta, 1 - \theta)$ of independence $(A(u) = 1)$ and diagonal case $(A(u) = \max(u, 1 - u))$ as before, naturally; and the equisurvival median curve is $x' + y' - \theta \min(x', y') = \log 2$.

We have $B(u|\theta) = 1 - D\left(\log \dfrac{1-u}{u}\,|\theta\right) = \dfrac{(1 - \theta)\,u}{1 - \theta\ u}$ if $u < \dfrac{1}{2}$, $B(u|\theta) = \dfrac{u}{1 - \theta(1 - u)}$ if $u \geq \dfrac{1}{2}$ with a jump of $\theta/(2 - \theta)$ at $u = 1/2$, as could be expected from the result for $D(w|\theta)$.

The classical correlation coefficient is $\rho(\theta) = \theta/(2 - \theta)$ increasing from $\rho(0) = 0$ to $\rho(1) = 1$.

Let us now consider the *biextremal model*. It appears naturally in the study of extremal processes; see Tiago de Oliveira (1968).

It can defined directly as follows: consider (X, Z) independent reduced Gumbel random variables and form the new pair (X, Y) with $Y = \max(X - a, Z - b)$, a and b such that

$\text{Prob}\{Y \le y\} = \text{Prob}\{\max(X - a\,,\, Z - b) \le y\} = \Lambda(y)$, which implies $e^{-a} + e^{-b} = 1$. Putting $e^{-a} = \theta, e^{-b} = 1 - \theta\ (0 \le \theta \le 1)$, we have $Y = \max(X + \log\theta\,,\, Z + \log(1 - \theta))$. This was the initial use of the max-technique to generate new models and is a particular case of a natural model, to be defined below, with $a = 0, b = +\infty, c = -\log\theta$ and $d = -\log(1 - \theta)$. The distribution function is $\Lambda(x,y|\theta) = \exp\{-\max(e^{-x} + (1 - \theta)\,e^{-y}\,,\, e^{-y})\}$, $(0 \le \theta \le 1)$ and the dependence function is immediately $k(w|\theta) = 1 - \min(\theta\,,\, e^{w})/(1 + e^{w}) = (1 - \theta + \max(\theta\,,\, e^{w}))/(1 + e^{w})$ and we have $D(w|\theta) = 0$ if $w < \log\theta$ and $(1 + (1 - \theta)\,e^{-w})^{-1}$ if $w \ge \log\theta$ with a jump of θ at $w = \log\theta$, which is equal to the $\text{Prob}\{Y = X + \log\theta\}$, and so the singular part is concentrated at the line $y = x + \log\theta$, with probability θ. Evidently $Y \ge X + \log\theta$ with probability one. For $\theta = 0$ and $\theta = 1$ we obtain independence and the diagonal case. Notice that $k(w|\theta) \ne k(-w|\theta)$ and so the model is not exchangeable.

The classical correlation coefficient of the biextremal model is

$$\rho(\theta) = \frac{6}{\pi^2} \int_0^{\theta} \frac{\log t}{1 - t}\, d\,t = 1 + \frac{6}{\pi^2}\, L(\theta),$$

$L(\theta) = \int_1^{\theta} \frac{\log w}{w - 1}\, d\,w$ being the Spence integral; $\rho(\theta)$ increases from $\rho(0) = 0$ to $\rho(1) = 1$. Linear regression is thus $L_y(x) = \gamma + \rho(\theta)\,(x - \gamma)$.

The grade correlation, difference-sign correlation, and medial correlation coefficients are easily seen to be $\chi(\theta) = 3\theta/(2 + \theta)$, $\tau(\theta) = \theta$ and $\nu(\theta) = 2^{\theta} - 1$, all increasing from 0 to 1 as θ varies from 0 to 1.

The index of dependence is

$$\sup_{x,y} |\,\Lambda(x\,,\, y|\theta) - \Lambda(x,\, y|0)| = \theta/(1 + \theta)^{(1+\theta)/\theta}\,,$$

increasing from 0 to 1/4 as θ increases from 0 to 1.

The curvilinear regressions — here we do not have exchangeability — are

$$\bar{y}(x|\theta) = x + \log\theta + \int_0^{(1-\theta)/\theta.e^{-x}} \frac{1 - e^{-t}}{t}\, d\,t =$$

$$= x + \log(1-\theta) + \int_{\frac{1-\theta}{\theta} e^{-x}}^{+\infty} \frac{e^{-t}}{t}\, dt$$

and

$$\bar{x}(y|\theta) = y - \log\theta - (1-\theta)\exp(\theta\, e^{-\bar{x}}) \int_{\theta\, e^{-x}}^{+\infty} \frac{e^{-t}}{t}\, dt\,.$$

The comparison with the linear regressions $L_y(x|\theta) = \gamma + \rho(\theta)(x-y)$ and $L_x(y|\theta) = \gamma + \rho(\theta)(x-\gamma)$ shows that the non-linearity index varies from 0 to 0.068 in the first curvilinear regression and from 0 to 0.007 in the second one; see Tiago de Oliveira (1974). We do not, therefore, have a substantial improvement in using curvilinear regression, although the comparison between $\bar{y}(x|\theta)$ and $L_y(x|\theta)$ and between $\bar{x}(y|\theta)$ and $L_x(y|\theta)$ shows a great difference. The explanation is as before: the small probability mass attached to the regions where the regressions (linear and curvilinear) differ strongly.

The test of independence ($\theta = 0$ vs. $\theta > 0$) can be made by using the classical correlation coefficient as usual.

The quadrants method can be applied as $p(\theta) = \text{Prob}\{X \le \tilde{\mu}\ ,\ Y \le \tilde{\mu}\} = 2^{\theta}/4$, equal to the expression for the Gumbel model; we can use the techniques described before.

As regards the estimation of θ, as $y_i - x_i \ge \log\theta$ and $0 \le \theta \le 1$, the best estimator of the left-end point $\log\theta$ is $\log\tilde{\theta} = \min(y_i - x_i\ ,\ 0)$ and so $\tilde{\theta} = e^{\min(y_i - x_i,\ 0)} = \min(e^{y_i - x_i}\ ,\ 1)$. We have, for $z < 1$, $\text{Prob}\{\tilde{\theta} \le z|\theta\} = \text{Prob}\{\min_1^n (e^{y_i - x_i}) \le z|\theta\} = 1 - \text{Prob}\{\min_1^n (y_i - x_i) > \log z|\theta\}$

$= 1 - (1 - D(\log z|\theta))^n$, and so have $\text{Prob}\{\tilde{\theta} \le z|\theta\} = 0$ if $z < \theta$, $\text{Prob}\{\tilde{\theta} \le z|\theta\} = 1 - \left(\frac{z}{1-\theta+z}\right)^n$ if $\theta \le z < 1$ and $\text{Prob}\{\tilde{\theta} \le z|\theta\} = 1$ if $z \ge 1$; the jumps have the values $1 - (1-\theta)^n$ at $z = 0$ and $\left(\frac{1-\theta}{2-\theta}\right)^n$ at $z = 1$; if $\theta \ne 1$ we see that $\tilde{\theta} \xrightarrow{P} \theta$; see Tiago de Oliveira (1970) and (1975') for other details.

The biextremal model for minima has the survival function $S(x', y'|\theta) = e^{-\max(x'+(1-\theta)y',\, y')}$, $0 \le \theta \le 1$, $x', y' \ge 0$; we obtain $A(u|\theta) = \max(u\ ,\ 1 - \theta\, u))$ and $B(u|\theta) = 1 - D\left(\log\frac{1-u}{u}|\theta\right) =$

$\frac{(1-\theta)\,u}{1-\theta\,u}$ if $u < \frac{1}{1+\theta}$, $B(u|\theta) = 1$ if $u \geq \frac{1}{1+\theta}$ with a jump of θ at $u = 1/(1+\theta)$. The equisurvival median curve is $\max(x' + (1-\theta)\,y', y') = y' + \max(x' - \theta\,y', 0) = \log 2$. The classical correlation coefficient takes the value $\rho(\theta) = \theta$.

Let us now consider the *natural model*. Using the max-technique, from a pair $(\bar{X}, \bar{Y})$ of independent reduced Gumbel random variables we get a new (dependent) random pair $X = \max(\bar{X} - a, \bar{Y} - b)$, $Y = \max(\bar{X} - c, \bar{Y} - d)$ with Gumbel margins if and only if $e^{-a} + e^{-b} = 1$ and $e^{-c} + e^{-d} = 1$. The dependence function is $k(w|a, b, c, d) = (e^{-\min(a,c+w)} + e^{-\min(b,d+w)})\,(1 + e^{w})$, whose tails coincide with those of the diagonal case $k_D(w) = \frac{\max(1, e^{w})}{1 + e^{w}}$. It is easy to see that $W = Y - X = \min(a + \bar{W}, b) - \min(c + \bar{W}, d)$ is such that $a - c \leq W \leq b - d$ if $a + d < b + c$ and $b - d \leq W \leq a - c$ if $b + c < a + d$. The case $a + b = b + c$ is irrelevant, because W is then an almost sure random variable with $W = a - c = b - d$, and as W must be zero (its mean value is zero) we have $a = c$, $b = d$, $X = Y$.

From now on we will suppose $a + d < b + c$, the case $b + c < a + d$ being dealt with by exchange of X and Y. Let us put $\alpha = a - c$, $b - d = \beta$; by the fact that $\bar{W}$ has a mean value of zero we have $\alpha < 0 < \beta$. Using the relations $e^{-a} + e^{-b} = 1$ and $e^{-c} + e^{-d} = 1$, with $a - c = \alpha$ and $b - d = \beta$ we get the final expression

$$k(w|\alpha,\beta) = \frac{(e^{\beta} - 1)\max(1, e^{\alpha - w}) + (1 - e^{\alpha})\max(1, e^{\beta - w})}{(e^{\beta} - e^{\alpha})\,(1 + e^{-w})}$$

or, in more detail,

$$k(w|\alpha,\beta) = \frac{1}{1 + e^{w}} \quad \text{if} \quad w < \alpha,$$

$$= \frac{e^{\beta} - 1 + (1 - e^{\alpha})\,e^{\beta - w}}{(e^{\beta} - e^{\alpha})\,(1 + e^{-w})} \quad \text{if} \quad \alpha \leq w \leq \beta,$$

$$= \frac{1}{1 + e^{-w}} \quad \text{if} \quad \beta < w;$$

as said before for $w < \alpha$ and for $w > \beta$ we have $k(w|\alpha,\beta) = k_D(w)$.

Note that $k(w|-\beta, -\alpha) = k(-w|\alpha, \beta)$; we get an exchangeable model if $\alpha = -\beta$. As said before note also that for $\alpha = \log \theta$, $\beta = +\infty$, we get the biextremal model and $\alpha = -\infty$, $\beta = -\log \theta$ its dual.

Using now the parameters (α, β) $(\alpha < 0 < \beta\,,\ \alpha \le W \le \beta$ with probability one), we see that $\Lambda(x, y|\alpha, \beta) = \exp\{-[(e^{\beta} - 1)\max(e^{-x}\,, e^{\alpha-y}) + (1 - e^{\alpha})\max(e^{-x}\,, e^{\beta-y})]/(e^{\beta} - e^{\alpha})\}$, $-\infty < \alpha \le 0 \le \beta < +\infty$. $D(w|\,\alpha, \beta)$ is given by :

$$D(w|\,\alpha, \beta) = 0 \qquad \text{if} \quad w < \alpha$$

$$= \frac{e^{\beta} - 1}{e^{\beta} - 1 + (1 - e^{\alpha})\, e^{\beta - w}} \qquad \text{if} \quad \alpha \le w \le \beta\,,$$

$$= 1 \qquad \text{if} \quad \beta \le w\ ;$$

with jumps of $\dfrac{(e^{\beta} - 1)\, e^{\alpha}}{e^{\beta} - e^{\alpha}}$ at $w = \alpha$ and of $\dfrac{(1 - e^{\alpha})}{e^{\beta} - e^{\alpha}}$ at $w = \beta$; $[\alpha, \beta]$ is, thus, the support of $W = Y - X$, $R = \beta - \alpha$ the range and so when we have $-\infty < \alpha \le 0 \le \beta < +\infty$ there exists, as $\alpha \le W = Y - X \le \beta$, a strong stochastic connection between X and Y; this model thus has flexibility to express various situations.

The classical correlation coefficient can be written is

$$\rho(\alpha, \beta) = 1 - \frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log \frac{k(w|\alpha,\beta)}{k_D(w)}\, d\,w =$$

$$= 1 - \frac{6}{\pi^2} \int_{\alpha}^{\beta} \log \frac{e^{\beta} - 1 + (1 - e^{\alpha})\ e^{\beta - w}}{(e^{\beta} - e^{\alpha})\max(1 + e^{-w})}\ d\,w\ ;$$

it is immediate that $\rho(\alpha, \beta) \to 1$ as $\alpha, \beta \to 0$ and that $\rho(\alpha, \beta) \to 0$ as $-\alpha, \beta \to +\infty$.

The difference-sign correlation coefficient is

$$\tau(\alpha, \beta) = 1 - \int_{\alpha}^{\beta} D(w|\alpha,\beta)\,(1 - D(w|\alpha,\beta)\, d\,w =$$

$$= 1 - \frac{(e^{\beta} - 1)\,(1 - e^{\alpha})}{(e^{\beta} - e^{\alpha})} = \frac{e^{\beta+\alpha} - 2\ e^{\alpha} + 1}{e^{\beta} - e^{\alpha}}\ ;$$

the medial correlation coefficient takes the form

$$\nu(\alpha, \beta) = 2^{(e^{\beta+\alpha} - 2\, e^{\alpha} + 2)/(e^{\beta} - e^{\alpha})} - 1\,.$$

The equidistribution median curve is

$$(e^{\beta} - 1)\max(e^{-x}, e^{\alpha-y}) + (1 - e^{\alpha})\max(e^{-x}, e^{\beta-y}) = \log 2 .(e^{\beta} - e^{\alpha}) .$$

The index of dependence is

$$(1 - k_o)\ \exp\{\frac{k_o}{1 - k_o}\log k_o\} \quad \text{where} \quad k_o = \min(\frac{1}{e^{\alpha} + 1}, \frac{e^{\beta}}{e^{\beta} + 1}) .$$

The linear regressions with reduced margins are given by the straight lines $L_y(x) = \gamma + \rho(\alpha, \beta)\ (x - \gamma)$ and $L_x(y) = \gamma + \rho(\alpha, \beta)\ (y - \gamma)$. For the curvilinear regression, as the exchange of X and Y is equivalent to the exchange of $-\alpha$ and β, it is sufficient to *compute the curvilinear regression*

$$\bar{y}(x|\alpha, \beta) = x + \beta - \frac{(e^{\beta} - 1)}{(e^{\beta} - e^{\alpha})} \exp(\frac{1 - e^{\alpha}}{e^{\beta} - e^{\alpha}} e^{-x}) . \int_{\frac{1-e^{\alpha}}{e^{\beta}-e^{\alpha}}e^{-x}}^{\frac{e^{\beta}}{e^{\alpha}}\ \frac{1-e^{\alpha}}{e^{\beta}-e^{\alpha}}e^{-x}} e^{\frac{-t}{t}} dt .$$

Analogously to the situation for the biextremal model, we may expect that regression curves are a good approximation, also in the same sense, to the curvilinear regression.

As $\alpha \le w_i \le \beta$, with $\alpha < 0 < \beta$, we can take as estimators of α and β the statistics $\alpha^* = \min\ (0, y_i - x_i)$ and $\beta^* = \min\ (0, y_i - x_i)$. We have exchangeability $(k(w|\alpha,\beta) = k(-w|\alpha,\beta))$ iff

$$k(w|\alpha,\beta) = \frac{1}{1 + e^{w}} \qquad \text{if} \quad w < -\beta$$

$$= \frac{e^{\beta}}{1 + e^{\beta}} \qquad \text{if} \quad -\beta \le w \le \beta$$

$$= \frac{1}{1 + e^{-w}} \qquad \text{if} \quad \beta \le w$$

and, consequently

$$D(w|\beta) = 0 \qquad \text{if} \quad w < \beta$$

$$= \frac{1}{1 + e^{-w}} \qquad \text{if } -\beta \le w \le \beta$$

$$= 1 \qquad \text{if } \beta \le \alpha ,$$

with jumps of $(1 + e^{\beta})^{-1}$ at $-\beta$ and β .

Also we have

$$\rho(\beta) = 1 - \frac{12}{\pi^2} \int_0^{\beta} \frac{w}{1 + e^{-w}} \, dw .$$

$$\tau(\beta) = \frac{2}{1 + e^{\beta}} .$$

$$\nu(\beta) = 2^{2/(1 + e^{\beta})} - 1 ,$$

$$\bar{y}(x|\beta) = x + \beta - \frac{e^{\beta}}{1 + e^{\beta}} \exp \frac{e^{-x}}{1 + e^{\beta}} \int_{\frac{e^{-x}}{1+e^{\beta}}}^{\frac{e^{\beta-x}}{1+e^{\beta}}} e^{-\frac{t}{t}} dt .$$

and $\beta^* = \max|x_i|$ as estimator of β .

Passing to standard exponential margins and introducing the new parameters $\alpha' = e^{-\alpha} \ge 1 \ge \beta' = e^{-\beta}$, we get

$$S(x', y'|\alpha', \beta') = \exp\{-\frac{(1 - \beta') \max(\alpha' x', y') + (\alpha' - 1) \max(\beta' x', y')}{\alpha' - \beta'}\} ;$$

the dependence function

$$A(u|\alpha', \beta') = \frac{(1 - \beta') \max(\alpha'(1 - u), u) + (\alpha' - 1) \max(\beta'(1 - u), u)}{\alpha' - \beta'} , \text{ i.e ,}$$

$$A(u|\alpha', \beta') = 1 - u \text{ if } 0 \le u \le \frac{\beta'}{\beta' + 1}$$

$$= \frac{\alpha'(1 - \beta') + (\alpha'\beta' - 1) u}{\alpha' - \beta'} \qquad \text{if } \frac{\beta'}{\beta' + 1} \le u \le \frac{\alpha'}{\alpha' + 1}$$

$$= u \quad \text{for} \quad \frac{\alpha'}{\alpha' + 1} \leq u \leq 1.$$

$B(u|\alpha', \beta')$ is equal to 0 if $u < \beta'/(1 + \beta')$, to $\dfrac{(\alpha' - 1)\, u}{(1 - \beta')\, \alpha' - (1 - \alpha'\, \beta')\, u}$ if $\beta'/(1 + \beta') \leq u < \alpha'/(1 + \alpha')$ and to 1 when $\alpha'/(1 + \alpha') \leq u \leq 1$; the jumps at $u = \beta'/(1 + \beta')$ and $u = \alpha'/(1 + \alpha')$ are respectively $(\alpha' - 1)\beta'/(\alpha' - \beta')$ and $(1 - \beta')/(\alpha' - \beta')$. The support for $\dfrac{Y'}{X' + Y'}$ is thus $\left[\dfrac{\beta'}{\beta' + 1} , \dfrac{\alpha'}{\alpha' + 1} \right]$ or equivalently the angle $\beta' X' \leq Y' \leq \alpha' X'$ and not all the first quadrant as in the other examples.

Let us make a quick reference to a generalization of the natural model, the *(2,N) natural model*, a particular case of the *(m,N) natural model* (for m=2) described briefly in the next Chapter, still under study.

Let $Z_1, ..., Z_n$ be N independent reduced Gumbel random variables and define $X = \max_1^n (Z_j - p_j)$, $Y = \max_1^n (Z_j - p_j)$.

The random pair (X, Y) has reduced Gumbel margins iff $\sum_1^N e^{-p_j} = 1$ and $\sum_1^N e^{-q_j} = 1$; then we have the distribution function

$$\Lambda(x, y | p, q) = \text{Prob}\{X \leq x, Y \leq y\} = \exp\{\sum_{j=1}^{N} \max(e^{-x-p_j}, e^{-y-q_j})\},$$

with the dependence function

$$k(w|p, q) = \frac{\sum_1^N \max (e^{-p_j}, e^{-q_j - w})}{1 + e^{-w}}.$$

The model thus has $2(N - 1)$ parameters. The corresponding model for bivariate minima with standard exponential margins has the survival function

$S(x', y'|p', q') = \exp\{-\sum_{1}^{N} \max(x' p_j', y' q_j')\}$, where $p_j' = e^{-p_j}$, $q_j' = e^{-q_j}$, $\sum_{1}^{N} p_j' = 1$, $\sum_{1}^{N} q_j' = 1$ with the dependence function

$$A(u|p, q) = \sum_{1}^{N} \max(p_j'(1-u), q_j' u).$$

Intrinsic estimation of the dependence functions

We have given the necessary and sufficient conditions on $k(w)$ and $A(u)$ for $\Lambda(x, y) = \exp\{-(e^{-x} + e^{-y}) k(y - x)\}$ to be a distribution function of bivariate maxima with reduced Gumbel margins and for $S(x', y') = \exp\{-(x' + y') A(\frac{y'}{x' + y'})\}$, $x', y' \geq 0$ to be the survival function of bivariate minima with standard exponential margins. Let us recall that, for $A(u)$, $u \in [0, 1]$, we have $A(0) = A(1) = 1$, $0 \leq -A'(0), A'(1) \leq 1$ and $A(u)$ convex.

We will call an *intrinsic estimator,* sometimes also called *non-parametric*, of $A(u)$ ($k(w)$) an estimator (i.e., a statistical function) converging in probability to the function $A(u)$ ($k(w)$), for every $u \in [0, 1]$ ($-\infty \leq w \leq +\infty$) that satisfies the necessary and sufficient conditions to be an $A(u)$ ($k(w)$). In Tiago de Oliveira (1989') and in Deheuvels and Tiago de Oliveira (1989) a first, and heavy, approach was made. Later, in Tiago de Oliveira (1991), a new and simpler approach was obtained. The procedure presented in Tiago de Oliveira (1989'), whose study was continued in Deheuvels and Tiago de Oliveira (1989), is more complicated than this one; also the estimator given there does not have all moments for any size sample, as opposed to what happens with the new one.

Let us consider, for simplicity, the intrinsic estimation of $A(u)$ from an i.i.d. sample of n pairs $\{(x_i, y_i)\}$ with survival function $S(x, y)$, satisfying *all* necessary and sufficient conditions given above; the conversion to the estimation of $k(w)$ is immediate.

For convenience of the exposé we will use $\Omega(u) = 1 - A(u)$; we have the necessary and sufficient conditions: $\Omega(0) = \Omega(1) = 0$, $0 \leq \Omega'(0), -\Omega'(1) \leq 1$ and $\Omega(u)$ concave and also from $\max(u, 1-u) \leq A(u) \leq 1$ we obtain $0 \leq \Omega(u) \leq \min(u, 1-u) = l(u)$.

Let us consider to estimate $\Omega(u)$, the statistical function

$$\bar{\Omega}_n(u) = \bar{\Omega}_n(u | x_i', y_i', \alpha) = \frac{c(\alpha)}{n} \sum_1^n \min\left(\frac{1-u}{\alpha + x_i'}, \frac{u}{\alpha + y_i'} \right),$$

where we will take $\alpha(=\alpha_n) \to 0$.

As $\min\left(\frac{1-u}{\alpha + x_i'}, \frac{u}{\alpha + y_i'} \right) \le \frac{l(u)}{\alpha}$, we see that each summand has all the moments.

We will seek function(s) $c = c(\alpha)$ and $\alpha = \alpha_n$ sufficient to guarantee that $\bar{\Omega}_n(u) \xrightarrow{m.s.} \Omega(u)$.

From the properties of each summand we can prove that $\bar{\Omega}_n(0) = \bar{\Omega}_n(1) = 0$ and $\bar{\Omega}_n(u)$ is concave, but the two other conditions on $\Omega'(u)$ are not necessarily verified (the conditions on the derivatives are valid for *each* summand only if $0 < c(\alpha) \le \alpha$, which will not be the case). As a consequence we will use the truncation $\Omega_n^*(u) = \min(\bar{\Omega}_n(u), l(u))$, which is a simple intrinsic estimator of $\Omega(u)$.

Let us begin by considering $\bar{\Omega}_n(u)$ in the complete dependence case:

In the diagonal case, where $X' = Y'$ with probability one, $X'(=Y')$ has the standard exponential distribution.

Then we have

$$M(\bar{\Omega}_n(u)) = M(c(\alpha) \min\left(\frac{1-u}{\alpha + x'}, \frac{u}{\alpha + y'} \right)) = (c(\alpha) \int_0^{+\infty} \frac{e^{-x'}}{\alpha + x'} dx)\, l(u) =$$

$$= (c(\alpha)\, e^{\alpha} \int_{\alpha}^{+\infty} \frac{e^{-t}}{t} dt)\, .\, l(u) = c(\alpha)\, J_1(\alpha)\, l(u) =$$

and

$$V(\bar{\Omega}_n(u)) = \frac{1}{n} V(c(\alpha) \min\left(\frac{1-u}{\alpha + x'}, \frac{u}{\alpha + y'} \right)) = \frac{c^2(\alpha)}{n} [J_2(\alpha) - J_1^2(\alpha)]\, l^2(u)$$

where $J_1(\alpha) = e^{\alpha} \int_{\alpha}^{+\infty} \frac{e^{-t}}{t} dt$ and $J_2(\alpha) = e^{\alpha} \int_{\alpha}^{+\infty} \frac{e^{-t}}{t^2} dt$; note that $J_1'(\alpha) = J_1(\alpha) - 1/\alpha$ and $J_2'(\alpha) = J_2(\alpha) - 1/\alpha^2$. It is immediate that $J_1(\alpha) \sim -\log \alpha$ and $J_2(\alpha) \sim 1/\alpha$ as $\alpha \to 0$.

Clearly the unbiased estimator of $l(u)$ leads to the choice $c(\alpha) J_1(\alpha) = 1$ and the condition that $V(\bar{\Omega}_n(u)) \to 0$ takes the form $\frac{1}{n} \left(\frac{J_2(\alpha)}{J_1^2(\alpha)} - 1 \right) \to 0$ or, as $\frac{J_1^2(\alpha)}{J_2(\alpha)} \to 0$ if $\alpha \to 0$, $\frac{1}{n} \frac{J_2(\alpha)}{J_1^2(\alpha)} \sim \frac{1}{n \alpha (\log \alpha)^2} \to 0$ as $\alpha = \alpha_n \to 0$.

Let us analyse other choices of $c(\alpha)$ which may be simpler and more convenient. Evidently we must have $c(\alpha) J_1(\alpha) \to 1$ and $\frac{c^2(\alpha)}{n} (J_2(\alpha) - J_1^2(\alpha)) \to 0$ or $c(\alpha) J_1(\alpha) \to 1$ and, also, $\frac{1}{n \alpha (\log \alpha)^2} \to 0$. The mean-square error of the estimator is thus

$$MSE = (c(\alpha) J_1(\alpha) - 1))^2 l^2(u) + \frac{c^2(\alpha)}{n} (J_2(\alpha) - J_1^2(\alpha)) l^2(u) ,$$

which converges to zero under the conditions stated above.

We could seek the function $c(\alpha)$ minimizing the MSE but this has no practical interest.

As $\int_0^{\alpha} \frac{1 - e^{-t}}{t} dt - \int_{\alpha}^{+\infty} \frac{e^{-t}}{t} dt = \gamma + \log \alpha$ we see immediately that $J_1(\alpha) + \log \alpha \to -\gamma$ as $\alpha \to 0$. But the choice $c(\alpha) = 1/(-\gamma - \log \alpha)$ $(c(\alpha) J_1(\alpha) \to 1)$ has $c(\alpha) \leq 0$ for values of $\alpha \approx 1$. Let us take then $c(\alpha) = \frac{1}{a - \log \alpha}$. The bias is then $(c(\alpha) J_1(\alpha) - 1) l(u) = \left(\frac{J_1(\alpha)}{a - \log \alpha} - 1 \right)$ $l(u) \sim \frac{\gamma + a}{\log \alpha} l(u)$; a must be chosen such that $a - \log \alpha > 0$ and as we are going to deal with small values of α, we can suppose $\alpha \leq 1$ and then take $a > 0$; note that this is equivalent to using $\alpha' = e^{-a} \alpha$ instead of α.

We can then compute asymptotically the mean-square error of the estimator. We have

$$\text{MSE} = \{[(\frac{a+\gamma}{a-\log\alpha})^2 + o(\frac{1}{(a-\log\alpha)^2})] + [\frac{1}{n\,\alpha(a-\log\alpha)^2} + o(\frac{1}{n\,\alpha(a-\log\alpha)^2})]\}l^2(u)$$

where the first term is the square of the bias and the second corresponds to the asymptotic evaluation of the variance $\frac{J_2(\alpha) - J_1^2(\alpha)}{n\,(a-\log\alpha)^2}$ because $J_2(\alpha) \sim 1/\alpha$ and $J_1(\alpha) \sim -\log\alpha$, as stated previously.

As $\alpha \to 0$, and so $-\log\alpha \to +\infty$, the condition on $\alpha = \alpha_n$ is to be such that $n\,\alpha(a-\log\alpha)^2$ or $n\,\alpha(\log\alpha)^2 \to +\infty$.

The discussion of the dominant terms in the MSE shows immediately that the squared bias is dominant if $n\,\alpha_n \to \infty$, is of the same order the variance if $n\,\alpha_n \to H\ (0 < H < +\infty)$, and the variance term is dominant if $n\,\alpha_n \to 0$.

The simplest form for $\alpha = \alpha_n$ is to take $\alpha_n = Rn^{-p}\ (p > 0)$. The condition $n\,\alpha_n(\log\alpha_n)^2 \to +\infty$ immediately imposes $0 < p \leq 1$ and the square of the bias and variance are asymptotic to, respectively, $\frac{(a+\gamma)^2}{p^2\,(\log n)^2}$ and $\frac{1}{R\,p^2\,n^{1-p}\,(\log n)^2}$. Their ratio is asymptotic to $(a+\gamma)^2\,R\,n^{1-p}$ having the squared bias (asymptotic) dominance if $p < 1$ and (asymptotic) equivalence if $p = 1$. In both cases the MSE is of the order of $(\log n)^{-2}$, and it is simpler to take $p = 1$ and so $\alpha_n = R/n$. Then $c(\alpha_n) = \frac{1}{a - \log R - \log n}$ and the condition $c(\alpha_n) > 0$ implies $a - \log R > 0$. We can take $R = 1(\alpha_1 = 1)$ and dispose of a. Taking for simplicity $a = 1$ we get the simple expression $\alpha_n = 1/n$ and $c(\alpha_n) = \frac{1}{1 - \log\alpha_n} = \frac{1}{1 + \log n}$, and then the squared bias is $\frac{(1+\gamma)^2}{(1+\log n)^2} + o(\frac{1}{(1+\log n)^2})$, the variance is $\frac{1}{(1+\log n)^2} + o(\frac{1}{(1+\log n)^2})$ and the MSE is $\frac{1+(1+\gamma)^2}{(1+\log n)^2} + o(\frac{1}{(1+\log n)^2})$. The convergence, being of the order of $1/(\log n)^2$, is thus very slow. Notice that the condition $c(\alpha) < \alpha$ which was needed for the properties of $\Omega'(0)$ and $\Omega'(1)$ is not verified.

Clearly as $\bar{\Omega}_n(u) \xrightarrow{\text{m.s.}} \Omega(u) = l(u)$ then $\bar{\Omega}_n(u) \xrightarrow{P} l(u)$ and so $\Omega_n^*(u) = \min(\bar{\Omega}_n(u), l(u)) \xrightarrow{P} l(u)$.

As a further note, consider two choices: $\alpha_n = R\, n^{-p}$ and $\alpha_n' = R'\, n^{-p'}$ $(0 < p\,,\, p' \leq 1)$.

The ratio of the variances corresponding to α_n and α_n' is

$$\sim \frac{1/R\; p^2\; n^{1-p}\; (\log n)^2}{1/R'\; p'^2\; n^{1-p'}\; (\log n)^2} = \frac{R'\; p'^2}{R\; p^2}\; n^{p-p'}\;,$$

which converges to zero (variance of smaller asymptotic order) if $n^{p-p'} \to 0$ or $p < p'$, showing that although the variance would be of smaller order with smaller p this effect is outweighed by the bias effect and so the choice of $\alpha_n = \frac{1}{1 + \log n}$ is reasonable.

Consider now the general case

$$\bar{\Omega}_n(u) = \frac{c(\alpha)}{n} \sum_1^n \min\left(\frac{1-u}{\alpha + x_i'}\,,\, \frac{u}{\alpha + y_i'} \right)$$

where, as seen before, we will take $\alpha_n = 1/n$ and $c(\alpha_n) = \frac{1}{1 - \log \alpha_n} = \frac{1}{1 + \log n}$.

We have

$$Q_\alpha(z) = \text{Prob}\{c(\alpha)\, \min\left(\frac{1-u}{\alpha + X'}\,,\, \frac{u}{\alpha + Y'} \right) \geq z\} =$$

$$= \text{Prob}\{X' \leq c(\alpha)\frac{(1-u)}{z} - \alpha\,,\, Y' \leq c(\alpha)\,\frac{u}{z} - \alpha\}$$

$$= 1 - S\left(\frac{c(1-u)}{z} - \alpha,\, 0\right) - S\left(0\,,\, \frac{c\,,\, u}{z} - \alpha\right) + S\left(\frac{c(1-u)}{z} - \alpha, \frac{c\,.\,u}{z} - \alpha\right).$$

Then the mean value of $c(\alpha)\, \min\left(\frac{1-u}{\alpha + X'}\,,\, \frac{u}{\alpha + Y'} \right)$ is given by $\int_0^{+\infty} z\, d(1 - Q_\alpha(z))$

$$= \int_0^{+\infty} z\, d\, S\left(\frac{c(1-u)}{z} - \alpha\,,\, 0\right) + \int_0^{+\infty} z\, d\, S\left(\frac{c\,u}{z} - \alpha,\, 0\right) - \int_0^{+\infty} z\, d\, S\left(\frac{c(1-u)}{z} - \alpha\,,\, \frac{c\,u}{z} - \alpha\right)$$

$$= I_1(\alpha) + I_2(\alpha) - I_3(\alpha)\,.$$

Let us then compute the summands. We have

$$I_1(\alpha) = \int_0^{+\infty} z\, d\, S(\frac{c(1-u)}{z} - \alpha\,,\, 0\,) = \int_0^{+\infty} z\, d\, (e^{-((c(1-u)/z) - \alpha)_+}) =$$

$$\int_{(c(1-u)/z)-\alpha>0}^{+\infty} z\, d\, (e^{-((c(1-u)/z) - \alpha)}) = e^{\alpha}\, c(1-u) \int_{\alpha}^{+\infty} \frac{e^{-\xi}}{\xi}\, d\, \xi$$

and

$$I_2(\alpha) = \int_0^{+\infty} z\, d\, S(0\,,\, \frac{c\,u}{z} - \alpha\,) = \int_0^{+\infty} z\, d\, (e^{-(((c\,u)/z) - \alpha)_+}) =$$

$$\int_{((c\,u)/z)-\alpha>0}^{+\infty} z\, d\, (e^{-(((c\,u)/z) - \alpha)}) = e^{\alpha}\, c\,u \int_{\alpha}^{+\infty} \frac{e^{-\xi}}{\xi}\, e\, \xi$$

and so $I_1(\alpha) + I_2(\alpha) = e^{\alpha}\, c(\alpha)\, J_1(\alpha) \to 1$.

Now compute $I_3(\alpha)$. We will show that $I_3(\alpha) \to A(u)$ and so $M(\bar{\Omega}_n(u)) = I_1(\alpha) + I_2(\alpha) - I_3(\alpha) \to 1 - A(u) = \Omega(u)$.

We have

$$I_3(\alpha) = \int_0^{+\infty} z\, d[\exp\{-[(\frac{c(1-u)}{z} - \alpha)_+ + (\frac{c\,u}{z} - \alpha)_+]\, A\, (\frac{(\frac{c\,u}{z} - \alpha)_+}{(\frac{c(1-u)}{z} - \alpha)_+ + (\frac{c\,u}{z} - \alpha)_+})\}].$$

Let us suppose, for convenience, that $0 \le u \le 1 - u \le 1$. Then $(\frac{c(1-u)}{z} - \alpha)_+$ is different from zero when $c(1-u) > \alpha\, z$ and $(\frac{c\,u}{z} - \alpha)_+$ is different from zero when $c\,u > \alpha\, z$. Then $I_3(\alpha)$ can be written as

$$I_3(\alpha) = \int_0^{cu/\alpha} z\, d[e^{-(c/z - 2\,\alpha)\, A\, ((cu/z) - \alpha)/((c/z) - 2\,\alpha))}] + \int_{cu/\alpha}^{c(1-u)/\alpha} z\, d\, (e^{-((c(1-u)/z) - \alpha)}) .$$

The second integral is

$$c(\alpha)\, e^{\alpha}\,(1-u)\int_{\alpha}^{((1-u)/u)\alpha}\frac{e^{-\xi}}{\xi}\,d\xi=\frac{e^{\alpha}}{1-\log\alpha}\,(1-u)\int_{\alpha}^{((1-u)/u)\alpha}\frac{e^{-\xi}}{\xi}\,d\xi\to 0 \text{ as } \alpha\to 0\,.$$

Let us consider the first integral. It can be written as

$$-\int_{0}^{cu/\alpha} z\, d\left(1-e^{-(c/z-2\alpha)\,A\,((cu/z-\alpha)/(c/z-2\alpha))}\right),$$

which, by integration by parts, takes the form

$$-\left[z\left(1-e^{-(c/z-2\alpha)\,A\,((cu/z-\alpha)/(c/z-2\alpha))}\right)\Big|_{0}^{cu/\alpha}\right]+$$

$$+\int_{0}^{cu/\alpha}\left(1-e^{-(c/z-2\alpha)\,A\,((cu/z-\alpha)/(c/z-2\alpha))}\right)dz$$

$$=-\frac{c(\alpha)\,u}{\alpha}\left(1-e^{-(1/u-2)\alpha}\right)+\int_{0}^{cu/\alpha}\left(1-e^{-(c/z-2\alpha)\,A\,((cu/z-\alpha)/(c/z-2\alpha))}\right)dz\,.$$

The first summand, as $c(\alpha)=\dfrac{1}{1-\log\alpha}$, converges to zero. Let us, then, study the integral. Now put $\xi=c/z$. The integral takes the form

$$=\int_{\alpha/u}^{+\infty} c(\alpha)\left(1-e^{-(\xi-2\alpha)\,A\,((\xi u-\alpha)/(\xi-2\alpha))}\right)\frac{d\xi}{\xi^{2}}\,.$$

But as $u\le 1-u$ $(u\le 1/2)$ we have $\dfrac{\xi u-\alpha}{\xi-2\alpha}<u$ and as $A(u)/u$ is non-increasing we obtain

$$\frac{A\left(\dfrac{\xi u-\alpha}{\xi-2\alpha}\right)}{\dfrac{\xi u-\alpha}{\xi-2\alpha}}\ge\frac{A(u)}{u}\quad\text{or}\quad(\xi-\alpha/u)\,A(u)\le(\xi-2\alpha)\,A\left(\frac{\xi u-\alpha}{\xi-2\alpha}\right);$$

and from $\dfrac{A(u)}{1-u}$ non-decreasing we get $(\xi-2\alpha)\,A\left(\dfrac{\xi u-\alpha}{\xi-2\alpha}\right)\le\left(\xi-\dfrac{\alpha}{1-u}\right)A(u)$ and, thus, the integral is bounded above and below by

$$\int_{\alpha/u}^{+\infty} c(\alpha)\,(1 - e^{-(\xi - \alpha/(1-u))A(u)})\,\frac{d\xi}{\xi^2} \quad \text{and} \quad \int_{\alpha/u}^{+\infty} c(\alpha)\,(1 - e^{-(\xi - (\alpha/u))A(u)})\,\frac{d\xi}{\xi^2}\ .$$

Putting $\xi A(u) = t$ we get, respectively, the following bounds above and below :

$$A(u)\,c(\alpha) \int_{\alpha A/u}^{+\infty} (1 - e^{\alpha A/(1-u) - t})\,\frac{dt}{t^2} \quad \text{and} \quad A(u)\,c(\alpha) \int_{\alpha A/u}^{+\infty} (1 - e^{(\alpha A/u) - t})\,\frac{dt}{t^2}\ .$$

But $\int_{p\alpha}^{+\infty} (1 - e^{q\alpha - t})\,\frac{dt}{t^2}$, with $p, q > 0$, by integration by parts, has the expression $\frac{1 - e^{(q-p)\alpha}}{p^{\alpha}} + e^{q\alpha} \int_{p\alpha}^{+\infty} \frac{e^{-t}}{t}\,dt$ and with $c(\alpha) = \frac{1}{1 - \log \alpha}$ we see immediately that $c(\alpha) \int_{p\alpha}^{+\infty} (1 - e^{q\alpha - t})\,\frac{dt}{t^2} \to 1$, and so both bounds converge to $A(u)$. Thus, as said initially, $I_3(\alpha) \to A(u)$, and so the mean value $I_1(\alpha) + I_2(\alpha) - I_3(\alpha) \to 1 - A(u) = \Omega(u)$. The proof when $1 - u \le u$ leads to the same result.

By the same technique we can show that

$$V(c(\alpha) \min\,(\frac{1-u}{\alpha + X'}, \frac{u}{\alpha + Y'})) \sim \frac{c^2(\alpha)}{\alpha}\,C(u)\,,$$

when $C(u)$ is a bounded function of u.

The condition of convergence is, once more, $V(\bar{\Omega}_n(u)) \sim \frac{C(u)}{n\alpha(\log \alpha)^2} \to 0$ or $n\alpha(\log \alpha)^2 \to 0$; as before for the diagonal case the previous analysis applies.

The final estimator is, as before, $\Omega_n^*(u) = \min\,(\bar{\Omega}_n(u)\,,\, l\,(u)) \xrightarrow{P} \Omega(u) = 1 - A(u)$.

We have, then, for bivariate minima with standard exponential margins, as $\alpha_n = 1/n$,

$$\bar{A}_n(u) = 1 - \frac{1}{1 + \log n} \sum_{1}^{n} \min(\frac{1-u}{1 + n x_i}, \frac{u}{1 + n y_i})\,,$$

and, more generally,

$$A_n^*(u) = 1 - \min\left(l(u),\ \frac{1}{1+\log n}\sum_1^n \min\left(\frac{1-u}{1+n\,x_i'},\ \frac{u}{1+n\,y_i'}\right)\right),$$

Recall that $\bar{A}_n(u)$ can be used if $0 \le -\bar{A}_n'(0)$, $\bar{A}_n(1) \le 1$ that is iff

$$0 \le \sum_1^n \frac{1}{1+n\,x_i'},\ \sum_1^n \frac{1}{1+n\,y_i'} \le 1 + \log n,$$

and, in that case, we get

$$\bar{S}_n(x, y) = e^{-(x+y)\,\bar{A}_n\,(y/(x+y))} = e^{-(x+y) + 1/(1+\log n)\,.\,\sum_1^n \min(x/(1+n\,x_i'),\ y/(1+n\,y_i'))}.$$

As $\bar{A}_n(u) \xrightarrow{P} A(u)$ and $A_n^*(u) = \max(\bar{A}_n(u),\ 1 - l(u)) \xrightarrow{P} A(u)$, we see that $\bar{S}_n(x', y') = \exp\{-(x'+y')\ \bar{A}_n(\frac{y'}{x'+y'})\} \xrightarrow{P} S(x', y')$ and also $S_n^*(x', y') = \exp\{-(x'+y')\ A_n^*(\frac{y'}{x'+y'})\} \xrightarrow{P} S(x', y')$.

For bivariate maxima with reduced Gumbel margins we have $\bar{k}_n(w) = \bar{A}_n\left(\frac{1}{1+e^w}\right)$ and $k_n^*(w) = A_n^*\left(\frac{1}{1+e^w}\right)$. Denoting by $\{(x_i, y_i)\}$ the sample we have (with the transformation (x, y) to $(x', y') = (e^{-x}, e^{-y})$ as stated,

$$\bar{k}_n(w) = 1 - \frac{1}{(1+\log n)(1+e^w)}\sum_1^n \min\left(\frac{e^w}{1+e^{-n\,x_i}},\ \frac{1}{1+e^{-n\,y_i}}\right) \quad \text{and}$$

$$\bar{A}_n(x, y) = \exp\{-(e^{-x}+e^{-y}) + \frac{1}{1+\log n}\sum_1^n \min\left(\frac{e^{-x}}{1+n\,e^{-x_i}},\ \frac{e^{-y}}{1+n\,e^{-y_i}}\right)\}.$$

Pickands (1981) gave a much more complex estimator, converging with probability one but with a very tedious calculation.

Miscellaneous statistical results:

A ready-to-wear model:

In Ramberg, Dudewicz, Tadikamalle and Mytyka (1979) a family of quantile functions $Q(p) = F^{-1}(p)$ is proposed that the authors say fit the data well; this family, of the form (where $0^0 = 1$)

$$Q(p) = \lambda + \delta\,(p^{\alpha} - (1-p)^{\beta}) ,$$

extends Tukey's λ-functions and so contains the logistic distribution. As the latter plays a central role for D(w) functions — corresponding to independence — it seems natural to study the quantile functions Q(p) that verify the conditions to be a D(w) function. First, as D(w) has all moments $\mu'_k = \int_0^1 Q^k(p)\,d\,p$, $k \geq 0$, we see immediately that we must have $\alpha \geq 0$ and $\beta \geq 0$; in the contrary case for k integer, such that $k\alpha + 1 < 0 \leq (k-1)\alpha + 1$ or $k\beta + 1 < 0 \leq (k-1)\beta + 1$, the integrals $\int_0^1 p^{k\alpha}\,d\,p$ or $\int_0^1 (1-p)^{k\beta}\,d\,p$ would diverge. Then for Q(p) to be non-decreasing we must have $\delta > 0$ (the case $\delta = 0$ leads to the almost sure random variable at λ, a case that can also be obtained with $\alpha = 0$ and $\beta = 0$) . It is immediate that the conditions $\int_{-\infty}^{+\infty} w\,d\,D(w) = 0$ and $D'(w) \geq D(w)\,(1 - D(w))$ are equivalent to $\int_0^1 Q(p)\,d\,p = 0$ and $(0 \leq)$ $p(1-p)\,Q'(p) \leq 1$. The condition $\int_0^1 Q(p)\,d\,p = 0$ gives

$$\lambda + \delta\,(\frac{1}{\alpha + 1} - \frac{1}{\beta + 1}) = 0 ,$$

and so

$$Q(p) = \delta\,\{(p^{\alpha} - \frac{1}{\alpha + 1}) + ((1-p)^{\beta} - \frac{1}{\beta + 1})\} ,$$

$$Q(0) = \delta\,\{(0^{\alpha} - \frac{1}{\beta + 1}) - (1 - \frac{1}{\beta + 1})\} ,$$

$$Q(1) = \delta\,\{(1 - \frac{1}{\alpha + 1}) - (0^{\beta} - \frac{1}{\beta + 1})\} .$$

The support is $[Q(0), Q(1)]$ and the range $R = \delta(2 - 0^{\alpha} - 0^{\beta})$, i.e., $R = 2\delta$ if $\alpha > 0$ and $\beta > 0$, $R = \delta$ if $\alpha > 0$ and $\beta = 0$ or $\alpha = 0$ and $\beta = 1$ and $R = 0$ in the degenerated case where $W(=Y - X) = 0$ with probability 1.

Let us now consider the condition $p(1 - p)\, Q'(p) \leq 1$ for each $p \in [0, 1]$ which is equivalent to $\delta \sup_{0 \leq p \leq 1} [\alpha (p^{\alpha}(1 - p) + \beta (1 - p)^{\beta}] \leq 1$ relating the three parameters $\alpha \geq 0, \beta \geq 0$ and $\delta > 0$.

We will consider only a limiting case when $\delta \to +\infty$ that already includes the logistic distribution; note that, as we have shown with the natural model, we can have cases where the range of $W = Y - X$ is a finite one. As $\dfrac{p^{\alpha} - 1/(\alpha + 1)}{\alpha}$ has a limit $(1 + \log p)$ if $\alpha \to 0$, then $\delta(p^{\alpha} - \dfrac{1}{\alpha + 1})$ has a limit $A(1 + \log p)$ iff $\delta\alpha \to A$ $(0 \leq A < +\infty)$; analogously we shall have $\delta\beta \to B$ $(0 \leq A < +\infty)$ and the limiting quantile function is $A(1 + \log p) - B(1 + \log (1 - p))$, $A \geq 0$, $B \geq 0$.

Note that $A = 0$ gives the exponential $p = 1 - e^{-(1 + w/B)}$ if $w \geq -B$ and $p = 0$ for $w \leq -B$; for $B = 0$ we obtain $p = e^{(w/A) - 1}$ if $w \leq A$ and $p = 1$ if $w \geq A$; for $A = B$ we get the logistic distribution $p = (1 + e^{-w/A})^{-1}$.

The condition $0 \leq p(1 - p)\, Q'(p) \leq 1$ takes the form $0 \leq A(1 - p) + Bp \leq 1$ and so we get the conditions on A and B, $0 \leq A$, $B \leq 1$.

The estimation of A and B can easily be made by least squares, i.e., associating the ordered statistics, $w'_{i,n}$ to the plotting positions $p_{i,n}$, and thus minimizing $\sum_{1}^{n} [w'_{i,n} - (A(1 + \log p_{i,n}) - B(1 + \log(1 - p_{i,n}))]^2$ and obtaining systematic (i.e., linear on the ordered statistics) estimators. There is no special reason to use this method, which also would introduce the question of plotting positions.

Let us compute some correlation coefficients. We have

$$\rho = 1 - \frac{3}{\pi^2} \int_{-\infty}^{+\infty} w^2 \, d\, D(w) = 1 - \frac{3}{\pi^2} \int_{0}^{1} Q^2(p) \, d\, p$$

$$= 1 - \frac{3}{\pi^2} \{A^2 + B^2 - 2(1 - \frac{\pi^2}{6}) A B\}$$

as

$$\int_0^1 \log p \; d\,p = \int_0^1 \log (1-p) \; d\,p = -1, \int_0^1 (1+\log p)^2 \, d\,p = \int_0^1 (1+\log (1-p))^2 \, d\,p = 1$$

and

$$\int_0^1 \log p \; \log (1-p) \; d\,p = \frac{\partial^2 B(a\,,\,b)}{\partial\, a\, \partial\, b} \Big|_{a=b=1} = 2 - \pi^2/6.$$

We have then $\rho(A\,,\,B) = 1 - AB - \frac{3}{\pi^2}\,(A-B)^2 = 1 + (\frac{12}{\pi^2} - 1)\,AB - \frac{3}{\pi^2}\,(A+B)^2$.

The logistic model $(A = B)$ has then $\rho = 1 - A^2$ ($\theta = 1 - A$ was the parameter previously used) oscillating between $0(A = 1)$ and $1(A = 0)$.

Also $\tau = 1 - \int_{-\infty}^{+\infty} w \, d\, D^2(w) = 1 - \int_0^1 Q(p)\, d\, p^2 = 1 - \frac{A+B}{2}$. Then taking $\tilde{\rho}^* = \max(0,$ $\rho^*)$ and $\tilde{\tau}^* = \max(0, \tau^*)$ where ρ^* and τ^* are the usual estimators of ρ and τ, and the truncation by zero results from the fact that $\rho \geq 0$ and $\tau \geq 0$, the equations $\tilde{\rho}^* = \rho(A^*\,,\,B^*)$ and $\tilde{\tau}^* = \tau\,(A^*\,,\,B^*)$ do not define uniquely A^* and B^* because of the symmetry of $\rho(A\,,\,B)$ and $\tau(A\,,\,B)$ in A and B, but give two alternatives for $(A^*\,,\,B^*)$ corresponding to the choice for A^* of the smallest or the largest of a second degree equation and dually for B^*.

A test of the logistic model as $\rho = 1 - A^2$ and $\tau = 1 - A$ is a test of $\rho = \tau(2 - \tau)$.

The use of quantiles can estimate A and B in an unique way.

As A and B play symmetrical roles it is natural to use quantiles of probabilities p and $q = 1 - p\ (> p \Leftrightarrow p \leq 1/2)$ and solve the equations

$$Q^*(p) = A^*(1 + \log p) - B^*(1 + \log q)$$

$$Q^*(q) = A^*(1 + \log q) - B^*(1 + \log p)\ .$$

We get

$$A^* = \frac{(1+\log q)\, Q^*(q) - (1+\log p)\, Q^*(p)}{(1+\log q)^2 - (1+\log p)^2}$$

$$B^* = \frac{(1+\log q)\, Q^*(p) - (1+\log p)\, Q^*(q)}{(1+\log q)^2 - (1+\log p)^2}$$

(A^*, B^*) as linear combinations of the quantiles $Q^*(p)$ and $Q^*(q)$ have an asymptotically binormal distribution which can be obtained by the δ-method. It does not seem worthwhile to advance this line of thought until it is felt that this "ready-to-wear" approach is a *good one.*

The estimation of the association between the margins

The fact that we may use an intrinsic estimator of $A(u)$ or $k(w)$, even supposing that we are using $\bar{A}_n(w)$ or $\bar{k}_n(w)$, and not $A_n^*(u)$ or $k_n^*(w)$, does not facilitate the evaluation of the association. It is sufficient to recall that $\rho(k) = -\frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log k(w)\, dw$ or $\rho(A) = \int_0^1 A^{-2}(w)\, du - 1$ to see that they are awkward to obtain except for numerical computation.

As $\max(u, 1-u) \le A(u) \le 1$ and $\frac{\max(1, e^w)}{1+e^w} \le k(w) \le 1$, we can use as association indicators the ratios of areas between $A(u)$ or $k(u)$ and independence suitably normalized:

$$\xi(A) = \frac{\int_0^1 (1 - A(u))\, d\,u}{\int_0^1 (1 - \max(u, 1-u))\, d\,u} = 4\int_0^1 (1 - A(u))\, d\,u$$

and

$$\xi(k) = \frac{\int_{-\infty}^{+\infty} (1 - k(w))\, d\,w}{\int_{-\infty}^{+\infty} \left(1 - \frac{\max(1, e^w)}{1+e^w}\right) d\,w} = \frac{1}{2\log 2}\int_{-\infty}^{+\infty} (1 - k(w))\, d\,w\ ;$$

note that the last coefficient could be suggested by the use of the approximation $1 - k(w)$ to $-\log k(w)$ which, as $1/2 \le k(w) \le 1$, has the following bounds

$$0 \le -\log k(w) - (1 - k(w)) \le -\log(1/2) - (1/2) = .1931471\ ,$$

and so the renormalization substitutes the coefficient $\frac{6}{\pi^2} = .6079271$ by the coefficient $\frac{1}{2 \log 2} = .7213475$; it is immediate, by integrating the relation above, that $0 \leq 2 \log 2 \,.\, \xi(k) \leq \frac{6}{\pi^2} \rho(k)$.

We get the estimators

$$\bar{\xi}_n(A) = 4 \int_{-\infty}^{+\infty} (1 - \bar{A}_n(w)(u))\, d\, u$$

$$= \frac{4}{1 + \log n} \sum_1^n \int_o^1 \min\left(\frac{1 - u}{1 + n\, x_i'}, \frac{u}{1 + n\, y_i'}\right) d\, u =$$

$$= \frac{2}{1 + \log n} \sum_1^n \frac{1}{2 + n\,(x_i' + y_i')}$$

and

$$\bar{\xi}_n(k) = \frac{1}{2 \log 2} \int_{-\infty}^{+\infty} (1 - \bar{k}_n(w))\, d\, w$$

$$= \frac{1}{2 \log 2.\,(1 + \log n)} \sum_1^n \int_{-\infty}^{+\infty} \frac{1}{1 + e^w} \min\left(\frac{e^w}{1 + e^{-x_i}}, \frac{1}{1 + e^{-y_i}}\right) d\, w$$

$$= \frac{1}{2 \log 2.\,(1 + \log n)} \sum_1^n \left[\frac{1}{1 + n\, e^{-x_i}} \log\left(1 + \frac{1 + n\, e^{-x_i}}{1 + n\, e^{-y_i}}\right) + \right.$$

$$\left. \frac{1}{1 + n\, e^{-y_i}} \log\left(1 + \frac{1 + e^{-y_i}}{1 + n\, e^{-x_i}}\right)\right]$$

A worked example :

Consider the data for the flood discharges of the Ocmulgee River in Georgia, USA at Hawkinsville (upstream) and Macon (dowstream), for the 40 years 1910/1949, contained in Table I taken from Gumbel and Goldstein (1964).

Table I

Year	Hawkinsville (1000 ft^3/s)	Macon (1000 ft^3/s)	Year	Hawkinsville (1000 ft^3/s)	Macon (1000 ft^3/s)
1910	18.8	28.8	1930	50.0	64.4
1911	5.9	8.5	1931	12.2	10.7
1912	44.4	44.8	1932	16.2	19.6
1913	52.0	51.0	1933	19.9	19.0
1914	5.9	4.8	1934	17.4	16.9
1915	20.1	19.1	1935	13.5	22.7
1916	40.4	47.8	1936	61.0	65.3
1917	27.0	25.4	1937	25.8	33.3
1918	14.3	14.3	1938	33.0	31.0
1919	40.0	31.0	1939	37.9	33.9
1920	45.2	66.2	1940	13.3	14.2
1921	30.0	37.0	1941	6.9	7.3
1922	44.0	48.6	1942	57.0	73.4
1923	30.3	28.3	1943	41.6	44.8
1924	15.2	21.0	1944	46.8	50.2
1925	79.0	72.5	1945	26.2	40.4
1926	19.3	28.3	1946	35.4	57.6
1927	7.6	7.9	1947	34.8	32.6
1928	42.4	47.1	1948	28.2	24.0
1929	70.5	73.4	1949	68.0	84.0

The margins data can be accepted to have a Gumbel distribution and the ML estimators of the parameters are $\hat{\lambda}_H = 23.709$, $\hat{\delta}_H = 15.057$, $\hat{\lambda}_M = 26.378$ and $\hat{\delta}_M = 17.042$. The empirical Kolmogoroff-Smirnov statistics for the margins are $KS_H = .09442$ and $KS_M = .06228$ with $\sqrt{n}\, KS_H = .597$ and $\sqrt{n}\, KS_M = .394$ accepting thus the Gumbel model for the margins. The better analysis is to use the $\hat{V}_n$ statistics for each margin. The medians are $\tilde{\mu}_H = 30.15$ and $\tilde{\mu}_M = 31.80$; note that the years ranked 20 and 21 in both stations are not the same. The number of concordant pairs (for the quadrants method) is $N = 34$; the value given in Gumbel and Goldstein (1964) is $N = 33$ because they estimate the medians through an inefficient estimation of the margin parameters. The correlation coefficient is $\rho^* = .942$.

The estimation equation $2p^{**} = \frac{N}{n} = \frac{N_1 + N_2}{n}$ imposes, as $\frac{1}{4} \le p^{**} \le \frac{1}{2}$, $\frac{1}{2} \le \frac{N}{n} \le 1$ which is acceptable as $N/n = 34/40 = .85$ and $p^{**} = N/2n = .425$. Clearly this value of p^* rejects independence as can be seen.

As for the mixed model $p_M(\theta) = 2^{-2+\theta/2}$, $\max p_M(\theta) = p_M(1) = 2^{-3/2} = .354 < .425$, the mixed model should be rejected or, accepting it, take $\theta = 1$; the distance between .354 and .425 suggests avoiding the truncation. Although Posner, Rodemich, Ashlock and Lurie (1969) did find an application of the mixed model — see the data and the strip method in their paper — we may expect the model to be useful only for weak correlations.

Consider now the logistic model. As $p_L(\theta) = \exp\{-2^{1-\theta} \log 2\}$ has the maximum $p_L(1) = 1/2$, the model can be accepted. In that case we have $\theta^{**} = .696$. As $\rho_L(\theta) = \theta(2-\theta)$ we get $\theta^* = 1 - \sqrt{1-\rho^*} = .759$ close to θ^{**} as $\theta^* - \theta^{**} = .063$. But better estimators of θ must be found.

From the practical point of view it seems that, for the differentiable models, the logistic one seems to give a good fit in general. But it must be recalled that Tiago de Oliveira (1987) has shown that the logistic and the natural models can be very close and difficult to separate for moderate samples. The latter is easily generalizable for asymmetric situations and even for more dimensions, which is not the case with the logistic model.

Chapter XII

MULTIVARIATE EXTREMES

Introduction

The theory of multivariate extremes is currently at a much less advanced stage than that of bivariate extremes with regard to both properties and modelling.

When thinking of floods at different points of the same river (or hydrographic basin), high tides in different ports of the same geographical area, or gusts of wind in different stations of a meteorological system, it is very easy to see how modelling is definitely needed but delayed — it must be recalled that there are still many open questions for bivariate extremes. The importance of models stems not only from their use in description but also from the important need of ways to predict future, possibly dangerous, events. As a simple example, think of a flood that can submerge river-bank villages, with the usual loss of lives and property, if warning is not given in time. It seems that we are still very far from this stage of knowledge.

Even so we will obtain the limiting forms (with Gumbel margins), obtain some inequalities, give independence conditions, and advance some models.

The probabilistic theory of multivariate extremes

The approach to multivariate extreme distributions is similar to that outlined previously for bivariate extreme distributions.

Let $F(x_1, \ldots, x_m)$ be a multivariate distribution function of a random m-dimensional vector $(X_1, \ldots, X_m)$. The distribution function of the maximum of each coordinate, in a sample of n independent random vectors with the distribution function $F(x_1, \ldots, x_m)$, is

$$F^n(x_1, \ldots, x_m) \ .$$

Evidently the point whose m coordinates are the maxima of each coordinates is a virtual point, not observed in general.

Suppose now that for the one-dimensional margins $F_i(x_i) = F(+\infty, \ldots, x_i, \ldots, +\infty)$ there exist attraction coefficients $\lambda_i^{(n)}$ and $\delta_i^{(n)}$ (>0) such that

$$F_i^n(\lambda_i^{(n)} + \delta_i^{(n)} x_i) \underset{n \to \infty}{\to} \Lambda(x_i)$$

and that

$$\Lambda(x_1, \ldots, x_m) = \lim_{n \to \infty} F^n(\lambda_1^{(n)} + \delta_1^{(n)} x_1, \ldots, \lambda_m^{(n)} + \delta_m^{(n)} x_m)$$

exists.

It was proved in Tiago de Oliveira (1958) and Geffroy (1958/59), in a way similar the one used above, that :

$\Lambda(x_1, \ldots, x_m)$ is a stable distribution function, i.e., a distribution function of the form

$$\Lambda(x_1, \ldots, x_m) = \exp\{-(e^{-x_1} + \ldots + e^{-x_m})\, k(x_2 - x_1, \ldots, x_m - x_1)\}$$

$$= \{\Lambda(x_1) \ldots \Lambda(x_m)\}^{k(x_2 - x_1, \ldots, x_m - x_1)} .$$

For other details see Gumbel (1961), Tiago de Oliveira (1962/63), and Deheuvels (1984).

Evidently, if the margins are Weibull or Fréchet for maxima or Weibull, Gumbel or Fréchet for minima, the usual transformations can reduce them to the case above. Special mention can be made of the situation for minima with standard exponential margins. Putting $y_i = e^{-x_i}$, as before, we have for the survival function $(y_i \geq 0)$,

$$S(y_1, \ldots, y_m) = \exp\{-(y_1 + \ldots + y_m)\, k(\log \frac{y_1}{y_2}, \ldots, \log \frac{y_1}{y_n})\}$$

which, by analogy with the A function of the bivariate extremes, can be written as

$$S(y_1, \ldots, y_m) = \exp\{-(y_1 + \ldots + y_m)\, A(\frac{y_2}{y_1 + y_2}, \ldots, \frac{y_m}{y_1 + y_m})\} .$$

We can now obtain inequalities connecting $\Lambda(x_1, \ldots, x_m)$ with the margins. As the one-dimensional margins are $\Lambda(x_i)$, from the Fréchet (1940) inequality

$$\max\,(0, \sum_i \Lambda(x_i) - m + 1) \;\le\; \Lambda(x_1, \ldots, x_m) \;\le\; \min\,(\Lambda(x_1), \ldots, \Lambda(x_m))\,,$$

using the stability condition

$$\Lambda^k(x_1 + \log k\,, \ldots, x_m + \log k) = \Lambda(x_1, \ldots, x_m)\,,$$

we get, with $k \to \infty$, that

$\Lambda(x_1, \ldots, x_m)$ verifies the double inequality

$$\Lambda(x_1) \ldots \Lambda(x_m) \;\le\; \Lambda(x_1, \ldots, x_m) \;\le \min\,[\Lambda(x_1)\,, \ldots, \Lambda(x_m)]\,.$$

To simplify notations we are noting $\sum_i^m$, $\prod_i^m$, $\sum_{\substack{i,j=n \\ i\neq j \\ i,j=1}}$, ... by $\sum_i$, $\prod_i$, $\sum_{i\neq j}$, ...

This double inequality cannot be improved because the upper and lower bounds are respectively the asymptotic distributions of multivariate extremes, the diagonal and the independence case.

Suppose now that the bivariate margins

$$\Lambda_{ij}\,(x_i\,, x_j) = [\,\Lambda(x_i)\,\Lambda(x_j)\,]^{\,k_{ij}(x_j - x_i)} \qquad (i \neq j)$$

are known and *compatible.*

A lower bound is obtained from a Gumbel inequality $(1 - S_m)\,\binom{m-1}{2} \le \binom{m}{2} - S_2$ — see Fréchet (1940) — giving

$$\Lambda(x_1, \ldots, x_m) \;\ge 1 - \sum_{i\neq j} (1 - \Lambda_{ij}\,(x_i\,, x_j))\,/\,2\,(m-1)$$

which, with stability, leads to

$$\Lambda(x_1, \ldots, x_m) \;\ge \{\prod_{i\neq j} \Lambda_{ij}\,(x_i\,, x_j)\}^{\frac{1}{2\,(m-1)}}\,.$$

As $\Lambda(x_1, \ldots, x_m) \le \Lambda_{ij}\,(x_i\,, x_j)$ we obtain also

$$\Lambda(x_1, \ldots, x_m) \leq \{ \prod_{i \neq j} \Lambda_{ij}(x_i, x_j) \}^{\frac{1}{m(m-1)}}$$

but a sharper inequality is

$$\Lambda(x_1, \ldots, x_m) \leq \min_{i \neq j} \{\Lambda_{ij}(x_i, x_j)\} .$$

A Bonferroni inequality gives

$$\Lambda(x_1, \ldots, x_m) \leq \frac{1}{2} \sum_{i \neq j} \Lambda_{ij}(x_i, x_j) - (m-2) \sum_{i} \Lambda(x_i) + \frac{(m-1)(m-2)}{2}$$

and stability leads to

$$\Lambda(x_1, \ldots, x_m) \leq \frac{\{ \prod_{i \neq j} \Lambda_{ij}(x_i, x_j) \}^{1/2}}{(\prod_{i} \Lambda(x_i))^{m-2}} .$$

As a consequence we can state that

$\Lambda(x_1, \ldots, x_m)$ verifies the following inequalities, when the bivariate margins $\Lambda_{ij}(x_i, x_j)$ are known and compatible :

$$\{ \prod_{i \neq j} \Lambda_{ij}(x_i, x_j) \}^{\frac{1}{2(m-1)}} \leq \Lambda(x_1, \ldots, x_m) ;$$

$$\Lambda(x_1, \ldots, x_m) \leq \min_{i \neq j} \{\Lambda_{ij}(x_i, x_j)\} \leq \{ \prod_{i \neq j} \Lambda_{ij}(x_i, x_j) \}^{\frac{1}{m(m-1)}} ;$$

$$\Lambda(x_1, \ldots, x_m) \leq \frac{\{ \prod_{i \neq j} \Lambda_{ij}(x_i, x_j) \}^{1/2}}{(\prod_{i} \Lambda(x_i))^{m-2}} .$$

An immediate consequence is that :

If the bivariate margins split in the product of the univariate margins then $\Lambda(x_1, \ldots, x_m) = \Lambda(x_1) \ldots \Lambda(x_m)$. Thus the necessary and sufficient condition for global independence is the independence of the bivariate margins.

This follows immediately substituting $\Lambda_{ij}(x_i, x_j) = \Lambda(x_i)\ \Lambda(x_j)$ in the first and third inequalities.

As a corollary we may verify Berman's (1961) statement that Geffroy's sufficient condition of independence of the two margins implies asymptotic independence. Thus we can state that

$\frac{1 - F_{ij}(x_i, x_j)}{2 - F_i(x_i) - F_i(x_j)} \to 1$ *when* $x_i \to \bar{w}_i$, $x_j \to \bar{w}_j$ ($\bar{w}_p$ *being the right-end points of the margins) is a sufficient condition for the asymptotic independence* $\Lambda(x_1, \ldots, x_m) = \Lambda(x_1) \ldots \Lambda(x_m)$.

Finally let us suppose that the $(m-1)$ - dimensional margins $\Lambda^{(i)}(x) = \Lambda(x_1, \ldots, x_{i-1}, +\infty, x_{i+1}, \ldots, x_n)$ are known and *compatible*.

From the relation

$$\max\left(\sum_i \Lambda^{(i)}(x) - (m-1), 0\right) \le \Lambda(x_1, \ldots, x_m) \le \min_i \{\Lambda^{(i)}(x)\}$$

the use of stability leads to

$$\prod_i \Lambda^{(i)}(x) \le \Lambda(x_1, \ldots, x_m) \le \min_i \{\Lambda^{(i)}(x)\} .$$

Another Gumbel inequality $(1 - S_m)(m-1) \le m - S_{m-1}$ — *see also Fréchet (1940)* — leads to $\sum_i \Lambda^{(i)}(x) - 1 \le (m-1)\,\Lambda(x_1, \ldots, x_m)$ and stability implies

$$\prod_i \Lambda^{(i)}(x) \le \{\prod_i \Lambda^{(i)}(x)\}^{\frac{1}{m-1}} \le \Lambda(x_1, \ldots, x_m) .$$

Another relation derived similarly from

$$\Lambda(x_1, \ldots, x_m) \le \frac{\sum_i \Lambda^{(i)}(x)}{m} \quad \text{is} \quad \Lambda(x_1, \ldots, x_m) \le \{\prod_i \Lambda^{(i)}(x)\}^{1/m} ,$$

this upper bound being greater than $\min_i \{\Lambda^{(i)}(x)\}$ obtained previously .

Consequently we can state :

If the margins $\Lambda^{(i)}(x) = \Lambda(x_1, \ldots, x_{i-1}, +\infty, x_{i+1}, \ldots, x_m)$ *are known and compatible,* $\Lambda(x)$ *verifies the following sequence of inequalities :*

$$\prod_i \Lambda^{(i)}(x) \leq \{\prod_i \Lambda^{(i)}(x)\}^{\frac{1}{m-1}} \leq \Lambda(x_1, \ldots, x_m) \leq \{\min_i \Lambda^{(i)}(x)\} \leq \prod_i \Lambda^{(i)}(x)\}^{1/m}.$$

Generalizations of bivariate models

The previous inequalities introduce some limitations on the formulation of multivariate extreme models.

Evidently, if the bivariate margins $\Lambda_{ij}(x)$ $(i \neq j)$ are known, as the upper bound $\min_{i \neq j} \{\Lambda_{ij}(x_i, x_j)\}$ implies the non-existence of a planar density (and such a situation will be considered in the next section), we can try, as a generalization, the lower bound

$$\{\prod_{i \neq j} \Lambda_{ij}(x_i, x_j)\}^{1/2(m-1)},$$

but the conditions for this function to be a distribution function should be verified in each case.

Here we will give only a few scattered models.

A triextremal model can be defined as follows :

Consider three independent reduced Gumbel random variables (Z_1, Z_2, Z_3) and consider the triple (X_1, X_2, X_3) defined as $X_1 = Z_1$, $X_2 = \max(Z_1 - a, Z_2 - b)$, $X_3 = \max(X_2 - c, Z_3 - d)$. To force (X_1, X_2) to have a biextremal distribution with reduced Gumbel margins $\Lambda(x_1, x_2 \mid \alpha)$, we must have $e^{-a} + e^{-b} = 1$ and $e^{-a} = \alpha (0 \leq \alpha \leq 1)$; to impose (X_2, X_3) also to have a biextremal distribution with reduced Gumbel margins $\Lambda(x_2, x_3 \mid \beta)$, we must have $e^{-c} + e^{-d} = 1$ and $e^{-c} = \beta (0 \leq \beta \leq 1)$; .

Thus the joint distribution of (X_1, X_2, X_3) is $\Lambda(x_1, x_2, x_3 \mid \alpha, \beta) = \text{Prob}\{X_1 \leq x_1, X_2 \leq x_2, X_3 \leq x_3\} = \text{Prob}\{Z_1 \leq x_1, Z_1 \leq a + x_2, Z_1 \leq a + c + x_3, Z_2 \leq b + x_2, Z_2 \leq b + c + x_3, Z_3 \leq d + x_3\} = \exp\{-\max(e^{-x_1}, e^{-a} e^{-x_2}, e^{-a} e^{-c} e^{-x_3}) - \max(e^{-b} e^{-x_2}, e^{-b} e^{-c} e^{-x_3}) - e^{-d} e^{-x_3}\} = \exp\{-\max(e^{-x_1}, \alpha e^{-x_2}, \alpha \beta e^{-x_3}) - (1 - \alpha) \max(e^{-x_2}, \beta e^{-x_3}) - (1 - \beta) e^{-x_3}\}$ and we obviously get $\Lambda(+\infty, x_2, x_3) = \Lambda(x_2, x_3 \mid \beta)$, $\Lambda(x_1, x_2, +\infty) = \Lambda(x_1, x_2 \mid \alpha)$ and $\Lambda(x_1, +\infty, x_3) = \Lambda(x_1, x_3 \mid \alpha\, \beta)$ so that naturally $0 \leq \alpha\ \beta \leq 1$.

For trivariate minima with standard exponential margins the survival function is (for $y_i \geq 0$)
$S(y_1, y_2, y_3) = \exp\{-\max(y_1, \alpha y_2, \alpha\beta y_3) - (1-\alpha)\max(y_2, \beta y_3) - (1-\beta) y_3\}$.

The multiextremal distribution – defined analogously to the biextremal one – which is important in the extremal processes that will follow is given by

$$\Lambda(x_1, \ldots, x_m) = \exp\left\{-\sum_1^m a_i \max\left(\frac{e^{-x_i}}{a_1 + \ldots + a_i}, \ldots, \frac{e^{-x_m}}{a_1 + \ldots + a_m}\right)\right\}$$

with $a_i \geq 0$. If we consider the biextremal pair (X_p, X_q) $(1 \leq p < q \leq m)$ it has the dependence parameter $\theta_{p,q} = \frac{a_1 + \ldots + a_p}{a_1 + \ldots + a_q}$; asymptotic independence exists if all $\theta_{p,q} \to 0$.

The generalization of the Gumbel model is

$$\Lambda(x_1, \ldots, x_n) = \exp\{-a_1 \sum_i e^{-x_i} - a_2 \sum_{i<j} \max(e^{-x_i}, e^{-x_j}) - \ldots - \ldots - a_m \max(e^{-x_1}, \ldots, e^{-x_m})$$

These distributions appeared in Tiago de Oliveira (1980).

A generalization of the logistic model, as this one can be defined by the rule

$$(-\log \Lambda(x_1, x_2))^{1/(1-\theta)} = (-\log \Lambda(x_1))^{1/(1-\theta)} + (-\log \Lambda(x_2))^{1/(1-\theta)},$$

could be

$$(-\log \Lambda(x_1, \ldots, x_m))^{1/(1-\theta)} = \sum_i (-\log \Lambda(x_i))^{1/(1-\theta)} = \sum_i e^{-x_i/(1-\theta)};$$

as the correlation coefficients between all margins would always be equal to $\theta(2-\theta)$, it does not seem useful to fit data. Another formal characterization of the bivariate logistic model could be extended to multivariate extremes and does not impose, as does the previous one, both symmetry and the quality of correlations between the margins, a fact which do not fit real (oceanographic, meteorological, hydrological, etc.) data.

A full generalization of the (bivariate) *natural model*, presented in Tiago de Oliveira (1987), is as follows: consider N independent reduced Gumbel random variables $Z_1, \ldots, Z_N$ and a $(m \times N)$ matrix $A = [a_{ij}]$ $(i = 1, \ldots, m\ ;\ j = 1, \ldots, N)$; now define the random vector $(X_1, \ldots, X_m)$ by

$$X_i = \max_{j=1}^{N} (Z_j - a_{ij}).$$

The condition $\text{Prob}\{X_i \le z\} = \Lambda(z)$ gives

$$\text{Prob}\{Z_j \le a_{ij} + z\} = \prod_{j=1}^{N} \Lambda(a_{ij} + z) = \Lambda(z) \quad \text{or} \quad \sum_{j=1}^{N} e^{-a_{ij}} = 1 ,$$

which imposes m conditions on the m N parameters a_{ij} (apart from being ≥ 0); we thus have m(N-1) non-negative parameters.

The joint distribution of $(X_1, \ldots, X_m)$ is

$$\text{Prob}\{X_i \le z_1, \ldots, X_m \le z_m\} = \text{Prob}\{Z_j \le a_{ij} + z_i\} = \prod_{j=1}^{N} \Lambda(\min_{i=1}^{m} (a_{ij} + z_i))$$

$$= \exp\{-\sum_{j=1}^{N} e^{-\min_{i=1}^{m} (a_{ij}+x_i)}\} = \exp\{-\sum_{j=1}^{N} \max (e^{-a_{ij}} . e^{-x_i})\} .$$

The choice of the matrix $A = [a_{ij}]$ is equivalent, putting $b_{ij} = e^{-a_{ij}}$, to the choice of a stochastic matrix $B = [b_{ij}]$ because $\sum_{j=1}^{N} b_{ij} = 1$ $(b_{ij} \ge 0)$.

It is evident that a permutation of the lines of A (or B) leads to a vector $(X'_1, \ldots, X'_m)$ with the same distribution of $(X_1, \ldots, X_m)$ except for the permutation of the indices.

The bivariate distribution of the pair (X_p, X_q) is

$$\Lambda_{p,q}(x_p, x_q) = \exp\{-\sum_{j=1}^{N} e^{-\min(a_{pj} + x_p , a_{qj} + x_q)}\}$$

$$= \exp\{-\sum_{j=1}^{N} \max (b_{pj} e^{-x_p} , b_{qj} e^{-x_q})\} .$$

It can be shown that $\min_{j=1}^{N} (a_{qj}) \le X_p - X_q \le \max_{j=1}^{N} (a_{pj})$. .

The statistical problems for multivariate extremes have not yet been analysed with some depth except concerning pairwise behaviour (for instance, using correlation coefficients of different

types) but certainly giving different estimates of the a_{ij} (or b_{ij}) *if* the identifiability problem is solved positively. In fact we can expect that the sample size will be much larger than the number $m(N-1)$ of parameters.

Some notes on linear regression

Consider once more the m-dimensional extreme value reduced margins distribution and let us discuss, briefly, the best least squares linear prediction of one of the components of $(X_1, X_2, ..., X_m)$ based on some other $p\ (< m)$ components. For notational simplicity we will suppose that the prediction variables are the first p ones $(X_1, ..., X_p)$ and the variable to be predicted is the last one (X_m).

We thus have to seek the coefficients $\alpha_o, \alpha_1, ..., \alpha_p$ such that the predictor $\hat{X}_m = \alpha_o + \sum_1^p \alpha_i X_i$ approaches best X_m in the mean-square sense, i.e., where

$$MSE(\hat{X}_m) = M(\hat{X}_m - X_m)^2 = M(X_m - \alpha_o - \sum_1^p \alpha_i X_i)^2 = \min .$$

We get

$$\alpha_o = \gamma (1 - \sum_1^p \alpha_i)$$

$$\rho_{mj} = \sum_{i=1}^p \alpha_i \rho_{ij}$$

where, in the last p equations, the ρ_{ij} denote the correlation coefficients between X_i and X_j.

Thus we have $\hat{X}_m = \gamma + \sum_{i=1}^p \alpha_i (X_i - \gamma)$ with $(\alpha_1, ..., \alpha_p)$ given by the equations $\rho_{mj} = \sum_{i=1}^p \alpha_i \rho_{ij}$; $\hat{X}_m$ is obviously unbiased.

Clearly we are using reduced values; if we have margin parameters (λ_i, δ_i) we will substitute X_i by $(X_i - \hat{\lambda}_i)/\hat{\delta}_i$.

The MSE, with reduced margins assumed, is

$$MSE = \frac{\pi^2}{6} \left(1 - \sum_{i=1}^{p} \alpha_i \rho_{mi} \right),$$

and so the relative reduction of variance is $\sum_{i=1}^{p} \alpha_i \rho_{mi}$.

Using matrix notation, where $[\rho_m] = (\rho_{m1}, \ldots, \rho_{mp})^T$ is the vector of the correlations between the prediction variables and X_m, $[\alpha] = (\alpha_1, \ldots, \alpha_m)^T$ is the vector of coefficients, and $R = [\rho_{ij}]$ $(i, j = 1, \ldots p)$ is the correlation matrix between the prediction variables, we have, from the equations $R\,[\alpha] = [\rho_m]$, $[\alpha] = R^{-1}[\rho_m]$, $MSE = \frac{\pi^2}{6} (1 - [\rho_m]^T R^{-1}[\rho_m])$ and relative reduction of variance $[\rho_m]\, R^{-1}[\rho_m]$.

ANNEX V

On the Quadrants Method

Consider a sample of n i.i.d. random pairs $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ whose distribution function is $F(x, y|\theta)$, where θ is a (unique) dependence parameter and the margins $F(x, +\infty|\theta)$ and $F(+\infty, y|\theta)$ are parameter-free.

Let us denote by ξ and η the medians the margins $(F(\xi, +\infty|\theta) = (F(+\infty, \eta|\theta) = 1/2)$, by N_1, N_2, N_3, N_4 $(N_1 + N_2 + N_3 + N_4 = n)$ the sample cardinals $N_1 = \#(x_i > \xi, y_i > \eta)$, $N_2 = \#(x_i \leq \xi, y_i > \eta)$, $N_3 = \#(x_i \leq \xi, y_i \leq \eta)$ and $N_4 = \#(x_i > \xi, y_i \leq \eta)$, and by $p_1(\theta)$, $p_2(\theta)$, $p_3(\theta)$, $p_4(\theta)$ the probabilities $p_1(\theta) = \text{Prob}(X > \xi, Y > \eta) = p(\theta)$, $p_2(\theta) = \text{Prob}(X \leq \xi, Y > \eta) = 1/2 - p(\theta)$, $p_3(\theta) = F(\xi, \eta|\theta) = p(\theta)$ and $p_4(\theta) = (X > \xi, Y \leq \eta) = 1/2 - p(\theta)$.

The maximum likelihood estimator of θ based on (N_1, N_2, N_3, N_4) is given by the equation $p(\theta^{**}) = \dfrac{(N_1 + N_3)}{2n}$; denoting by $N = N_1 + N_3$ we have $p(\theta^{**}) = N/2n$.

As $\sqrt{n}\ \dfrac{N/n - 2\,p(\theta)}{\sqrt{2\,p(\theta)\,(1 - 2\,p(\theta))}}$ is asymptotically standard normal we know, by the δ-method — see Tiago de Oliveira (1982) — that $\sqrt{2n}\ \dfrac{p'(\theta)}{\sqrt{p(\theta)\,(1 - 2\,p(\theta))}}\ (\theta^{**} - \theta)$ is also asymptotically standard normal, as well as $\dfrac{2n^{3/2}\,p'(\theta^{**})\,(\theta^{**} - \theta)}{\sqrt{N(n - N)}}$.

If we consider $F(x, y|\theta) = \Lambda(x, y|\theta)$ we have $\xi = \eta = -\log\log 2$ and we get $p(\theta) = (\xi, \eta|\theta) = \exp\{(-e^{-\xi} + e^{-\eta})\,k(\eta - \xi|\theta)\} = \exp(-2\log 2 \,.\, k(0|\theta))$. Thus the estimator θ^{**} is given by the equation $k(0|\theta^{**}) = \dfrac{\log(2n/N)}{\log 4}$. But, for any (one-parameter) model, we have $\bar{k} = \sup_{\theta} k(0|\theta)$, $\underline{k} = \inf_{\theta} k(0|\theta)$; suppose that $\bar{k} = k(0|\bar{\theta})$ and $\underline{k} = k(0|\underline{\theta})$ have unique solutions: then we shall truncate the estimator and take $\theta^{**} = \bar{\theta}$ if $N/2n > \bar{k}$ and $\theta^{**} = \underline{\theta}$ if $N/2n < \underline{k}$.

A test of independence vs. dependence (i.e. if $p(0) = 1/4$ vs. $p(0) > 1/4$) is given by the rejection region $\sqrt{n}\,(2N/n) - 1) \geq \lambda_\alpha$ where $\lambda_\alpha = N^{-1}(1 - \alpha)$, N(.) being the standard normal distribution function. This acts as a test independent of the model *but* only based on N_1, N_2, N_3, N_4: essentially it can confirm dependence.

We can, evidently, extend the method to use other quantiles in the margins or with estimated parameters in the margins.

This method is due to Gumbel and Mustafi (1967), but presented here with modifications.

Exercises — Part III

III.1— Obtain the inequality $\frac{\max(1,e^w)}{1+e^w} \leq k(w) \leq 1$ or $\max(u, 1-u) \leq A(u) \leq 1$ from the Boole-Fréchet inequality and a passage to the limit.

III.2— Show that the sets $\{k(w)\}$ or $\{A(u)\}$ are convex, closed and symmetrical (corresponding to exchangeability $k(w) = k(-w)$ or $A(u) = A(1-u)$).

III.3— Obtain the conditions that $k(w)$ or $A(u)$ must satisfy in the absolutely continuous case.

III.4— The equidistribution median curve for the case with reduced Gumbel margins is between the curves $e^{-x} + e^{-y} = \log 2$ and $\max(e^{-x}, e^{-y}) = \log 2$; are these curves convex ? : study the corresponding situation for the equisurvival median curve (case of $A(u)$).

III.5— For bivariate distributions attracted for maxima to $\Lambda(x,y)$, show that if there is negative dependence (i.e., $F(x, y) \leq F(x, +\infty)\ F(+\infty, y)$) then $\Lambda(x,y) = \Lambda(x)\ \Lambda(y)$ (*asymptotic attraction to independence*).

III.6—We can expect that for bivariate samples $(\max_1^n X_i, \max_1^n Y_i)$ are asymptotically independent.

Obtain conditions for this, but show that in the singular case where Prob $\{X + Y = a\} = 1$ this is not the case. Recall that $\min_1^n Y_i = -\max_1^n (-Y_i)$, and use Sibuya's necessary and sufficient condition or Geffroy's sufficient condition for independence in maxima.

III.7—In the characterization of $A(u)$, show that if $A(0) = A(1) = 1$ and $A(u)$ convex in $[0,1]$ the conditions

a) $0 \leq -A'(0), A'(1) \leq 1$,

b) $\max(u, 1-u) \leq A(u) \leq 1$ and

c) $A(u)/(1-u)$ non-decreasing and $A(u)/u$ non-increasing are equivalent.
Show also that $A'(0) \leq A'(u) \leq A'(1)$, $A(u) \geq 1 + A'(0)\,u$, and
$A(u) \geq 1 - A'(1)(1-u)$.

III.8—Obtain the relation between the function $P(u, v)$ defined by Sibuya and the structure function $\bar{S}(\xi, \eta)$. Convert the results of Sibuya on $P(u, v)$ concerning asymptotic independence and asymptotic complete dependence (diagonal case) to $\bar{S}(\xi, \eta)$.

III.9— The condition of $k''(w)$, for the bivariate distribution function $\Lambda(x,y)$, implies that $\varphi(w) = (1 + e^{w})\, k(w)$ and $\psi(w) = (1 + e^{-w})\, k(w)$ are such that $\varphi''(w) \geq \varphi'(w)$ and so $\varphi'(w) \geq 0$ and $\psi''(w) + \psi'(w) \geq 0$ and so $\psi'(w) \leq 0$, which eliminates the two monotonicity conditions.

III.10— Show that if (X, Y) has the distribution function $\Lambda(x,y)$ then the pair (V,W) with $e^{-V} = e^{-X} + e^{-Y}$, $W = Y - X$ has the distribution function

$$\text{Prob}\{V \leq v, W \leq w\} = \Lambda(v) + \left(\frac{k'(w)}{k(w)} + \frac{1 - e^{-w}}{1 + e^{-w}}\right) \Lambda(v - \log k(w))$$

$$+ \int_{-\infty}^{w} \frac{e^{-\rho}}{(1 + e^{-\rho})^2} (e^{-v} k(\rho) - 1)\, \Lambda(v - \log k(\rho))\, d\rho \ .$$

III.11— For the distribution function $D(w)$, show that $D^{a}(w + b)$, if $a \geq 1$, is also a $D(w)$-function for convenient b.

III.12—Show that if (X, Y) is a bivariate pair with reduced Gumbel margins, and so distribution function $\Lambda(x, y)$, then

$$Prob\{X \leq x,\ Y - X \leq w\} = D(w)\, \Lambda(x,\ x + w) \ .$$

III.13— Translate maxima results to minima results and vice-versa; use, in particular, the relation between the distribution function and the survival function.

III.14— Show that $\Lambda(w + \gamma)$ and $\dfrac{w + \theta}{2\theta}$ for $0 \leq \theta \leq 2$, $|w| < \theta$ are $D(w)$ - functions.

III.15— Suppose that (X, Y) is an independent random pair with Gumbel margins whose parameters are (λ_X, δ) and (λ_Y, δ) (the same dispersion parameter). Compute Prob $\{Y > X\}$. Solve the corresponding question for minima with exponential margins and compare the results.

III.16— Compute $k(.)$ for the use of max-technique ; what is the corresponding technique if $S(x,y)$ is used ? Interpret the statistical implications for small samples of the fact that the index of dependence is $\leq 1/4$; what does this imply for the discrimination between models ?

III.17— Show that for general bivariate distribution of maxima with reduced Gumbel margins Prob $\{X \le Y\} = \frac{1}{2} - \frac{k'(0)}{k(0)}$ so $\left|\frac{k'(0)}{k(0)}\right| \le 1/2$ and that for minima with standard exponential margins we have, correspondingly, Prob$\{X' \le Y'\} = \frac{1}{2} - \frac{A'(1/2)}{4\,A(1/2)}$ and so $\left|\frac{A'(1/2)}{A(1/2)}\right| \le 2$.

III.18— Let $G_1(x)$ and $G_2(y)$ be distribution function such that $G_1^k(\lambda_k + \delta_k x) \to \Lambda(x)$ and $G_2^k(\lambda_k' + \delta_k' y) \to \Lambda(y)$. As stated, we know by Fréchet inequalities, that any bivariate distribution function $F(x, y)$ such that $F(x, +\infty) = G_1(x)$, $F(+\infty, y) = G_2(y)$ satisfies the inequalities $\underline{D}(x, y) = \max(0, G_1(x) + G_2(y) - 1) \le F(x, y) \le \min(G_1(x), G_2(y)) = \overline{D}(x, y)$. Show that the family of distribution functions $\{\theta \underline{D}(x, y) + (1 - \theta)\overline{D}(x, y)\}$ is attracted, with the same attraction coefficients for the margins, to $\exp\{-[\theta(e^{-x} + e^{-y}) + (1 - \theta)\max(e^{-x}, e^{-y})\}$, which is the Gumbel model.

III.19— Show that the bivariate model of maxima with Weibull margins and shape parameter $\alpha = 1$ ($\xi = -e^{-x}$, $\eta = -e^{-y}$) is $\Psi_1(\xi, \eta) = \exp\{(\xi + \eta)\, k(\exp\frac{\xi}{\eta})\}$ and the corresponding logistic is $\Psi_{1L}(\xi, \eta) = \exp\{((-\xi)^{1/(1-\theta)} + (-\eta)^{1/(1-\theta)})^{1-\theta}\}$.

III.20— Show that if we have a reduced sample $\{(x_i, y_i)\ ,\ i = 1, 2, \ldots, n\}$, and we expect to observe x, the MSE linear predictor of y ($y^* = \alpha + \beta x + \sum_1^n \varphi_i\ x_i + \sum_1^n \psi_i\ y_i$) is given by the usual linear regression $L_y(x) = \gamma + \rho(x - \gamma)$, the sample having an intervention only through the computation of $\rho^* = r$, as usual .

III.21— For bivariate extremes with reduced Gumbel margins, the linear regression line of y in x is $L_y(x) = \gamma + \rho(x - \gamma)$. Show that all regression lines are in the area : if $x \le \gamma$ then $x \le L_y(x) \le \gamma$ and if $x \ge \gamma$ then $\gamma \le L_y(x) \le x$.

III.22— Show that, as all regression lines are in the area described above, statistical choice of bivariate model through linear regression is difficult even for moderate sample sizes.

III.23— Draw the graphs of $A(u|\theta)$ for different models (and values of $\theta = 0.$, .5 and 1.0). They are all in in the triangle bounded by $\max(u\ ,\ 1 - u) \le A(u|\theta) \le 1$. If they are not close to independence or to the diagonal case it is difficult to distinguish them.

III.24— Using the explicit expression of the correlation coefficients, show that :

for the logistic model : $\rho(\theta) = \theta(2-\theta)$, $\nu(\theta) = 2^{-2^{1-\theta}} - 1$;

for the mixed model : $\rho(\theta) = \frac{6}{\pi^2}(\arccos(1-\theta/2))^2 = \frac{24}{\pi^2}(\operatorname{arctg}\sqrt{\frac{\theta}{4-\theta}})^2$, $\nu(\theta) = 2^{\theta/2}-1$;

for the Gumbel model : $\rho(\theta) = \frac{12}{\pi^2}\int_0^\theta \frac{\log(2-t)}{1-t}\,dt$, $\nu(\theta) = 2^\theta - 1$;

for the biextremal model : $\rho(\theta) = -\frac{6}{\pi^2}\int_0^\theta \frac{\log t}{1-t}\,dt$, $\nu(\theta) = 2^\theta - 1$.

III.25— Analyse the data of Table I, contained in Gumbel and Goldstein (1964):

TABLE I

Oldest ages at death, Sweden

Year	Men	Women	Year	Men	Women
1905	100.88	102.54	1932	102.55	104.87
6	101.17	106.13	33	103.17	105.98
7	104.65	103.46	34	103.98	103.38
8	105.12	102.12	35	106.09	105.32
9	102.57	101.69	36	103.43	103.77
10	101.70	102.92	37	105.72	105.86
11	100.49	102.78	38	103.24	104.27
12	100.90	106.15	39	103.25	105.45
13	103.06	105.02	40	103.40	105.71
14	102.63	103.56	41	101.66	106.15
15	102.69	106.52	42	106.48	104.71
16	100.82	101.50	43	101.26	103.83
17	102.52	104.01	44	105.12	105.19
18	100.08	105.01	45	104.88	105.03
19	101.67	104.52	46	102.41	105.88
20	101.41	103.94	47	104.22	107.49
21	101.76	103.14	48	102.88	105.83
22	102.57	104.33	49	103.57	103.41
23	101.63	102.32	50	105.12	105.64
24	103.47	103.56	51	103.80	103.53
25	105.48	103.86	52	102.94	107.90
26	104.01	105.87	53	103.00	104.42
27	105.83	103.31	54	106.50	104.85
28	105.00	104.37	55	103.36	103.97
29	102.78	102.72	56	103.15	107.89
30	102.61	105.01	57	102.54	104.46
31	105.55	104.40	58	104.92	104.12

Test the fit of the Gumbel distribution to each set of oldest ages of death (men, women) and, after that, test for (the expected) independence.

Add to Table I the data of Table II, given in Fransén and Tiago de Oliveira (1984).

TABLE II

Oldest ages at death, Sweden

Year	Men	Women	Year	Men	Women
1959	104.23	104.77	1965	105.28	104.31
60	103.59	106.13	66	104.93	104.98
61	103.74	107.10	67	105.27	105.83
62	103.00	104.56	68	105.92	106.35
63	104.25	110.07	69	101.81	107.58
64	104.12	106.15	70	104.02	105.42

Do the same tests as before and compare the estimators. Also compare the estimators of the margins for Table I and Tables I and II. Use the previous result to estimate the probability that women have greater ages of death than men (take δ^* as the average of the δ's for Tables I and II) and test if women "live longer" than men.

III.26— Consider any river at two points with a sample of $k \geq 50$ yearly floods. Check if the margins are Weibull, Gumbel or Fréchet distributed. If they are not Gumbel distributed, make the necessary transformations to be so, and then :

– estimate (λ, δ) for both points;

– choose the bivariate model for the "estimated" reduced margins (exactly or approximately).

III.27— Obtain subroutines for statistical choice of bivariate maxima models with Gumbel margins or for bivariate minima with exponential margins.

III.28— Show that the "estimated" best linear predictor $\frac{y^* - \hat{\lambda}_y}{\hat{\delta}_y} = \gamma + r \,.\, (\frac{x - \hat{\lambda}_x}{\hat{\delta}_x} - \gamma)$ is to be expected to be very close to $\frac{y^* - \bar{y}}{s_y} = r \,.\, (\frac{x - \bar{x}}{s_x})$ as

$$\bar{x} - \hat{\lambda}_x - \gamma \hat{\delta}_x \approx 0 \,,\, \bar{y} - \hat{\lambda}_y - \gamma \hat{\delta}_y \approx 0 \quad \text{and} \quad \hat{\delta}_y / \hat{\delta}_x \approx s_y / s_x \,.$$

III.29— Study the estimator $\tilde{\theta}$ of θ in the biextremal case with Gumbel margins (moments, convergence, asymptotic distribution, etc.) .

III.30— Show that from the sample distribution function of the w_i, $D^*(w) = \frac{1}{n} \sum_{1}^{n} H(w - w_i)$ $\xrightarrow{P}$ $D(w)$ we can obtain estimators $k^*(w)$ of $k(w)$ and $D^*(w)$ of $D(w)$, both being non-parametric but not intrinsic, which is a more restrictive condition. It is easy to prove this by showing that $A(u)$ is non-convex (using exponential margins).

III.31— Consider the trivariate model with Gumbel margins

$$\Lambda(x_1, x_2, x_3) = \exp \{ - (e^{-x_1} + e^{-x_2} + e^{-x_3}) + \theta_3 \min(e^{-x_1}, e^{-x_2})$$

$$+ \theta_2 \min (e^{-x_1}, e^{-x_3}) + \theta_1 \min (e^{-x_2}, e^{-x_3}) - \tau \min (e^{-x_1}, e^{-x_2}, e^{-x_3}) \} .$$

Using the inequalities for the margins, show that

$$0 \leq \theta_1, \theta_2, \theta_3 \leq 1 \quad \text{and} \quad \max (0, \theta_1 + \theta_2 + \theta_3 - 1) \leq \tau \leq (\theta_1 + \theta_2 + \theta_3)/2$$

and so $\theta_1 + \theta_2 + \theta_3 \leq 2$.

S. M. BERMAN (1961) . Convergence to bivariate limiting extreme value distributions, *Ann. Inst. Statist. Math.*, XIII, 217-233, Tokyo

H. CRAMÉR (1946). *Mathematical Methods of Statistics*, Princeton University Press, New Jersey

P. DEHEUVELS (1984). Probabilistic aspects of multivariate extremes, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 117-130, D. Reidel, Dordrecht

P. DEHEUVELS and J. TIAGO DE OLIVEIRA (1989). On the non-parametric estimation of the bivariate extreme-value distributions, *Statist. and Prob. Letters*, 8, 315-323

B. V. FINKELSHTEYN (1953) . Limiting distributions of extremes terms of the variational series of a two-dimensional random variable, *Dokl. Ak. Nauk. SSSR*, 91 (in Russian), 209-211

D. A. S. FRASER (1957). *Non-Parametric Methods in Statistics*, Wiley, New York

M. FRÉCHET (1940). *Les Probabilités Associées a un Système d'Événements Compatibles et Dépendents*, Act. Scient. Ind., # 859, Herman et Cie, Paris

M. FRÉCHET (1951). Sur les tableaux de corrélation dont les marges sont données, *Ann. Univ. Lyon*, ser III, XIV-A, Math. et Astr., 53-77

J. GALAMBOS (1978). *The Asymptotic Theory of Extreme Order Statistics,* Wiley, New-York; (1987), 2nd. ed. Krieger Publ.

J. GEFFROY (1958/59). Contribution a l'étude de la théorie des valeurs extrêmes, *Publ. Inst. Statist. Univ. Paris*, 7/8, 37-185, Thèse de Doctorat d'Etat

E. J. GUMBEL (1961). Multivariate extremal distributions, *Bull. Int. Statist. Inst.*, 33rd sess., Paris

E. J. GUMBEL (1962). Statistical theory of extreme values (main results), in *Contributions to Order Statistics,* Ahmed R. Sarhan and Bernard G. Greenberg eds., 56-93, Wiley, New York

E. J. GUMBEL (1966). Two systems of bivariate extremal distributions, *Bull. Int. Statist. Inst.*, 41th. sess., Paris

E. J. GUMBEL and NEIL GOLDSTEIN (1964) . Analysis of empirical bivariate extremal distributions, *J. Amer. Statist. Assoc.*, 59, # 307, 794-816

E. J. GUMBEL and C. K. MUSTAFI (1967). Some analytical properties of bivariate extremal distributions, *J. Amer. Statist. Assoc.*, 62, 569-588

E. J. GUMBEL and C. K. MUSTAFI (1968). Conditional medians for bivariate extreme distributions, Tech. Report., *Columbia University* (New York), Dep. of Industrial Engineering

H. S. KONIJN (1956). On the power of certain tests for independence for bivariate populations, *Ann. Math. Statist.*, 27, 300-323

K. V. MARDIA (1964). Asymptotic independence of bivariate extremes, *Calcutta Statist. Assoc. Bull.*, 13, 172-178

ALBERT W. MARSHALL and INGRAM OLKIN (1967). A multivariate exponential distribution, *J. Amer. Statist. Assoc.*, 62, # 317, 30-44

J. PICKANDS III (1981). Multivariate extreme value theory, *Bull. Int. Statist. Inst.*, 43rd sess., II, 859-878, Buenos Aires

EDWARD C. POSNER, EUGENE R. RODEMICH, JOHN C. ASHLOCK and SANDRA LURIE (1969). Application of an estimator of high efficiency in bivariate extreme value theory, *J. Amer. Statist. Assoc.*, 64, # 328, 1403-1414

JOHN S. RAMBERG, EDWARD J. DUDEWICZ, PANDU R. TADIKAMALLA and EDWARD F. MYTYTKA (1979). A probability distribution and its uses in fitting data, *Technometrics*, 21, #2, 201-214

M. SIBUYA (1960). Bivariate extremal statistics, I, *Ann. Inst. Statist. Math.*, XI, 195-210, Tokyo

J. TAWN (1988). Bivariate extreme value theory-models and estimation, *Biometrika*, 75, 397-415

J. TIAGO DE OLIVEIRA (1958). Extremal distributions, *Rev. Fac. Ciências de Lisboa*, 2 ser., A, Mat, VII, 215-227

J. TIAGO DE OLIVEIRA (1962/63). Structure theory of bivariate extremes; extensions, *Estudos Mat., Estatist. e Econometria*, Lisboa, VII, 165-194

J. TIAGO DE OLIVEIRA (1964). L'indépendance des distributions extrêmales bivariées, *Publ. Inst. Statist. Univ. Paris*, XIII, 137-141

J. TIAGO DE OLIVEIRA (1965). Optimality of grade correlation, *An. Fac. Ciências Porto*, XLVIII, 39-48

J. TIAGO DE OLIVEIRA (1965'). Statistical decision for bivariate extremes, *Portugaliae Math.*, 24, 145-154

J. TIAGO DE OLIVEIRA (1968). Extremal processes; definition and properties, *Publ. Inst. Statist. Univ. Paris*, XII, 25-36

J. TIAGO DE OLIVEIRA (1970). Biextremal distributions: statistical decision, *Trab. Estad. Inv. Oper.*, vol. XXI, 107-117, Madrid

J. TIAGO DE OLIVEIRA (1971). A new model of bivariate extremes; statistical decision, in *Studi di Probabilitá, Statistica e Ricerca Operativa in onore di Giuseppe Pompilj*, Tip. Oderisi, Gubbio, 1-13

J. TIAGO DE OLIVEIRA (1974). Regression in the nondifferentiable bivariate extreme models, *J. Amer. Statist. Assoc.*, 69, 816-818

J. TIAGO DE OLIVEIRA (1975). Bivariate and multivariate extreme distributions, in *Statistical Distributions in Scientific Work*, G. P. Patil, S. Kotz and J. K. Ord eds., 1, 355-361, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1975'). Statistical decision for extremes, *Trab. Estad. Inv. Oper.*, XXVI, 453-471, Madrid

J. TIAGO DE OLIVEIRA (1980). Bivariate extremes: foundations and statistics, *Multivariate Analysis-V,* P. R. Krishnaiah ed., 349-368, North-Holland, Amsterdam

J. TIAGO DE OLIVEIRA (1984). Bivariate models for extremes: statistical decision, in *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 131-153, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1985). Statistical choice of non-separated models, *Trab. Estad. Inv. Operat.* , Madrid, XXXVI, 136-152

J. TIAGO DE OLIVEIRA (1987). Comparaison entre les modèles bivariés logistique et naturel pour les maxima et extensions, *C. R. Acad. Sc. Paris,* 305, ser I, 481-484

J. TIAGO DE OLIVEIRA (1989). Statistical decision for bivariate extremes, *Extreme Value Theory*, Proceedings, Oberwolfach 1987, 246-261, Springer

J. TIAGO DE OLIVEIRA (1989'). Intrinsic estimation of dependence structure for bivariate extremes, *Statist. and Prob. Letters*, 8, 213-218

J. TIAGO DE OLIVEIRA (1989''). Bivariate models for floods of rivers, presented at the *47th session of the I.S.I.*, Paris

J. TIAGO DE OLIVEIRA (1991). Intrinsic estimation of the dependence functions in bivariate extremes; a new and simpler approach, to be publ. in *Commun. Statist.-Theor. Meth.*

J. TIAGO DE OLIVEIRA (1991'). Some notes on correlation for bivariate extremes, presented at the *48th session of the I.S.I.*, Cairo

PART IV

Stochastic sequences and processes of extremes

Chapter XIII

EXTREMES IN RANDOM SEQUENCES AND STOCHASTIC PROCESSES

Introduction

There are various physical applications modelled by a family of random variables X(t) depending on a discrete or continuous parameter t , usually interpreted as time. This situation arises when we study the development in time of some physical system subject to random influences, by means of a discrete or continuously registering instrument. Our observations can take the form of a set of discrete ordinates or of a curve in a diagram with the time t, as abscissa, and the observed value X(t) as ordinates. This cloud of points, or curve, can be regarded as a *realization* , a *version*, a *trajectory* or a *sample function* of the basic random sequence or stochastic process X(t). The random variables X(t) and X(u), for different time moments t and u will, in general, be correlated. We will not deal with random fields

We will, then, successively, consider the behaviour of maxima in random sequences and stochastic processes having as a target to obtain, under convenient conditions, the same limiting distributions. It is intuitive that some waning out of the dependence must be imposed. Let us clarify this by a simple example. Consider a sequence of i.i.d. standard normal random variables $Y_0, Y_1, Y_2, \ldots, Y_n$ and consider the (composed) sequence $X_i = \sqrt{\rho}\, Y_0 + \sqrt{1-\rho}\; Y_i$, $i \geq 1$, $0 \leq \rho \leq 1$. It is evident that the sequence $\{X_i\}$ is a multinormal sequence with standard margins and (constant) correlation ρ (the inverse is also true). It is evident that $\max_1^n X_i = \sqrt{\rho}\; Y_0 + \sqrt{1-\rho}\; \max_1^n Y_i$. As by the Weak Law of Large Numbers $\max_1^n Y_i - \sqrt{2 \log n} \xrightarrow{P} 0$ we see that $(\max_1^n X_i - \sqrt{1-\rho}\; \sqrt{2 \log n}) / \sqrt{\rho} - Y_0 \xrightarrow{P} 0$, so that the asymptotic distribution of $(\max_1^n X_i - \sqrt{1-\rho}\; \sqrt{2 \log n}) / \sqrt{\rho}$ is standard normal. Thus, *if the correlation is constant (and > 0) and the process is gaussian, the asymptotic distribution is normal and not extremal.* As said, some conditions must be imposed on the dependence to obtain the

Extremal Limit Theorem: essentially that, in some sense, dependence must converge adequately to zero.

Some applications

Consider first the following situation which may occur in some Reliability studies. Suppose that the output of some electronic device should, ideally, be a given function $f(t)$ of time, but the observed output will be $X(t) = f(t) + \varepsilon(t)$, where the deviation $\varepsilon(t)$ may be regarded as the noise, produced by various internal and external disturbances. In some cases, when the probability structure of the noise $\varepsilon(t)$ is known, at least in general terms, it may be possible to find the probability that, during a given observation period $t = 1, 2, \ldots, n$ or $0 \leq t \leq T$, $X(t)$ will always stay within some limit(s), possibly functions of time. We may even be prepared to accept exceedances of some level, provided that they do not occur too often, or perhaps that the excursion of $X(t)$ outside the limit(s) will be a small fraction of the total observation time $\{1, 2, \ldots, n\}$ or $[0, T]$. We are, then, interested in the occurrence and frequency of certain extreme values of $X(t)$, particularly the ones falling outside the limit(s). Also it may be important to find the probability that the maximum or minimum of $X(t)$ during the observation period will over- or underpass some value.

An example from the statistical analysis of random loads: an airplane travelling through gusty winds, or a car driving along a rough road, or a ship in a rough sea, will be exposed to loads fluctuating in a random way. To estimate the service life of such structures, it may be important to study the extreme values of the load curve $X(t)$ to avoid ruptures, in particular by studying the frequency and the size of the peaks of the load curve. In some of these applications, it will be important to investigate the random length $t_2 - t_1$ and the random size $X(t_2) - X(t_1)$ of the fluctuations, t_1 corresponding to a maximum of $X(t)$ while t_2 denotes the following minimum.

Analogous questions of the sizes of fluctuations in a random the construction of ships, antennas, etc., where the height and length of storm-generated ocean waves, gusts of wind, etc., are important, and may be investigated by probabilistic methods.

Analogous problems of extreme values are encountered in Meteorology, Hydrology and Hydrography, where $X(t)$ may signify the wind velocity or the river discharges or the sea level at some given point. Once more it is important, for various practical problems, to have information about the height and the frequency of the peaks of the $X(t)$ curve, and even the shape of the curve in the neighbourhood of the peaks.

Finally we can mention some economic applications. If X(t) denotes, for instance, a price index, or some measure of business activity at time t, analogous extreme value problems may often be studied. A particularly interesting case occurs, as will be seen below, in insurance practice, in particular the ruin problem.

The pioneer writer in this field, from the point of view of Engineering applications, was Rice (1944/45). Although the area is still under intensive study we can cite, as basic, Cramér and Leadbetter (1967). Leadbetter, Lindgren and Rootzén (1983), and some of the papers in Tiago de Oliveira (ed.) (1984) and references therein.

We will not give proofs, only sketch some of them, because we are going to refer to limiting results of theoretical type which, in fact, reduce to univariate distributions for applications. The next chapters will deal with actual sequences and processes of extremes.

Extremes in stochastic sequences

Associated with the asymptotic behaviour of maxima (minima) in random sequences is the problem of "high" ("low") exceedances of increasing (decreasing) levels $\{u_n\}$. We will deal, once more, only with maxima, minima being dealt with by symmetry.

A basic result is the following:

For the random sequence $\{X_n\}$, *if* $F_{1,\dots,n}(x_1, \dots, x_n) = Prob\{X_1 \le x_1, \dots, X_n \le x_n\}$ *is such that* $F_{1,\dots,n}(\lambda_n + \delta_n x, \dots, \lambda_n + \delta_n x) \to L(x)$ *and also* $F_{1,\dots,n}(\lambda_n + \delta_n x, \dots, \lambda_n + \delta_n x) - F^k_{1,\dots,[n/k]}(\lambda_n + \delta_n x, \dots, \lambda_n + \delta_n x) \to 0$, *then* $L(x)$ *is max-stable, and thus must be one of the* $\Psi_\alpha(x)$, $\Lambda(x)$ *or* $\Phi_\alpha(x)$.

This condition, as said before, imposes for $\{F_{1,\dots,n}(x_1, \dots, x_n)\}$ *a sort of independence-like limiting behaviour.*

For stationary sequences the condition $D(u_n)$ for the (real) sequence u_n which reads as

$$|F_{i_1,\dots,i_p,j_1,\dots,j_q}(u_n, \dots, u_n) - F_{i_1,\dots,i_p}(u_n, \dots, u_n) F_{j_1,\dots,j_q}(u_n, \dots, u_n)| \le \alpha_{n,l}$$

for any integers $1 \le i_1 < \dots < i_p < j_1 < \dots < j_q \le n$ with $j_1 - i_p \ge l$ where $\alpha_{n,l} \to 0$ as $n \to \infty$ and $l = l_n = o(n)$, is very important. This condition is, evidently, also a kind of asymptotic-like independence weaker than the one before, owing to the stationarity added. It is equivalent to impose $\alpha_{n,[n\lambda]} \to 0$ for each $\lambda > 0$.

It can be shown that:

If $\{X_n\}$ *is a stationary sequence,* $\{\max_1^n X_i\}$ *has a limiting distribution, and for all sequences* $\lambda_n + \delta_n x$ $(-\infty < x < +\infty)$ *the condition* $D(\lambda_n + \delta_n x)$ *is valid for each* x*, then the limiting distribution function is of one of the forms* $\Psi_\alpha(x)$, $\Lambda(x)$ *or* $\Phi_\alpha(x)$.

For stationary normal sequences $D(u_n)$ is valid if $\rho_n \log n \to 0$ (Berman condition).

Concerning stationary sequences, we can go further as regards domains of attraction by comparison with i.i.d. sequences with the same margins $F = F_1$. Denoting by $u_n(\tau)$ a sequence such that $n(1 - F(u_n(\tau))) \to \tau$ and supposing that for some τ_o condition $D(u_n(\tau_o))$ is valid and Prob $\{\max_1^n X_i \le u_n(\tau_o)\}$ converges, then for any τ we have Prob $\{\max_1^n X_i \le u_n(\tau)\} \to e^{-\theta\tau}$ for a convenient θ $(0 \le \theta \le 1)$.

If for any τ, Prob $\{\max_1^n X_i \le u_n(\tau)\} \to e^{-\theta\tau}$, θ is called the *extremal index* and we can compare $\max_1^n X_i$ with $\max_1^n \hat{X}_i$ when $\{\hat{X}_i\}$ is an i.i.d. sequence with margins F_1. It can be shown that:

If the stationary sequence $\{X_n\}$ *has extremal index* θ *and* Prob $\{\max_1^n \hat{X}_i \le \lambda_n + \delta_n x\} \to L(x)$ *then* Prob $\{\max_1^n X_i \le \lambda_n + \delta_n x\} \to L^\theta(x)$ *which is of the same type of* $L(x)$. The attraction coefficients used are the same.

Note that in many cases we have $\theta = 1$; but if we take $Y_1, \ldots, Y_n$ i.i.d. with $F(x)$ and if we define $X_i = \max(Y_i, Y_{i+1})$ we have Prob $\{X_i \le x\} = F^2(x)$ and $F^{2n}(\lambda_n + \delta_n x) \to L(x)$ (i.e., if $\max_1^n \hat{X}_i$ has limiting distribution function $L(x)$) so that Prob $\{\max_1^n X_i \le \lambda_n + \delta_n x\} = F^{n+1}(\lambda_n + \delta_n x) \to L^{1/2}(x)$ and $\theta = 1/2$.

For details see Leadbetter, Lindgren and Rootzén (1983).

Global extremes of Gaussian stationary processes

In many applications, as in those mentioned above, the random mechanism generating the stochastic process under observation can be assumed to have the dependence structure waning out, at least as long as we consider observation periods of reasonable length. A precise probabilistic characterization of this asymptotic independence property could be the following. Consider the random variables $X(t_1 + h)$, $X(t_2 + h)$, ..., $X(t_n + h)$, where the t_j are any time points, h is a real number and $\{t_j\}$ and $\{t_j + h\}$ belong to the observation time interval. If the joint distribution of any such group of random variables associated with the $X(t)$ process is independent of h, we say that the process is (strictly) stationary. If, moreover, all these distributions are multinormal, we are dealing with a *normal* or *Gaussian stationary process*. For this class of processes, some of the extreme value problems indicated above can be solved at least in part, and we shall give an account of some results so for obtained in this direction.

From the stationarity it follows immediately that the mean value $\mu(t)$, the variance $\sigma^2(t)$, and all other moments of $X(t)$, when they exist, are independent of t. It also follows that the covariance $C(X(t), X(u))$ is a function of the time difference $t - u$ and thus an even function, so that we may write

$$C(X(t), X(u)) = C(|t - u|), C(0) = \sigma^2$$

From now on we will suppose, for simplicity, that we are dealing with the standardized Gaussian process $X(t) = (X(t) - \mu)/\sigma$, *denoted simply now by* $X(t)$, $\rho(t - u) = C(t - u)/\sigma^2$ *being the correlation function.*

The *correlation function* $\rho(t)$ has the *spectral representation*

$$\rho(t) = \int_0^\infty \cos \lambda t \; d F(\lambda) \; ,$$

where $F(\lambda)$ is a distribution function. In the most important applications, $F(\lambda)$ is absolutely continuous ($f(\lambda) = F'(\lambda) \geq 0$), so that we have

$$\rho(t) = \int_0^\infty \cos \lambda t \; f(\lambda) \, d \lambda \; .$$

The function $f(\lambda)$ is called the *spectral density of the process*. The spectral moments of $\rho(t)$

$$\lambda_k = \int_0^\infty \lambda^k \, f(\lambda) \, d\lambda \qquad (\lambda_0 = 1)$$

are not necessarily finite. We shall here assume that λ_k is finite for $k \leq 4$. This implies that $\rho(t)$ and its four first derivatives tend to zero as t tends to infinity ; we shall further assume that $\rho(t)$ itself tends to zero at least as rapidly as some negative power of t. Then for small values of t we have the expansion

$$\rho(t) = 1 - \frac{1}{2!} \lambda_2 t^2 + \frac{1}{4!} \lambda_4 t^4 + o(t^4) .$$

Recall that $\rho(t)$ (or $C(t)$) being an even function, the odd moments are null. We can now discuss some extreme value problems for $X(t)$ processes under the conditions given above. Suppose that the observation period is $[0, T]$. As $X(t)$ has mean value 0 it can take positive and negative values (supposing non-degeneracy $X(t) \neq 0$).

Let then $u > 0$ be some level. We can seek the probability that $X(t)$ exceeds u at least once in $[0, T]$. Or, denoting by $U(T)$ the (random) number of upcrossings of u (i.e., crossings of u upwards), we can seek its properties, and even its distribution. We will give the expression of the mean value of $U(T)$ and the asymptotic behaviour of $U(T)$. Other results can be seen in Cramér and Leadbetter (1967) and in Leadbetter, Lindgren and Rootzén (1983).

The mean value of $U(T)$, for the level u, is given by

$$M(U(T)) = (\sqrt{\lambda_2} / 2\pi) \, . \, T \, e^{-u^2/2} \, .$$

From this expression we see that the mean number of upcrossings of the level u decreases quickly to zero, as u becomes large. For a negative level $-u$, the same expression gives the mean number of downcrossings of the level $-u$.

If we denote by $D(T)$ and $C(T)$ the number of downcrossing and of crossings ($C(T) \geq U(T) + D(T)$, because of tangencies) we have from the stationarity $M(D(T)) = M(U(T)) = \frac{1}{2} M(C(T))$. Every upcrossing of the level u corresponds to a peak higher than u for $X(t)$. The formula above thus given the mean number of upcrossings or peaks, showing that they become scarce as u becomes large. Writing

$$\mu = M(U(1)) = (\sqrt{\lambda_2} / 2\pi) \, . \, e^{-u^2/2}$$

we have

$$M(U(1/\mu)) = 1,$$

and thus the average number of peaks higher than u in a time interval of length $1/\mu$ is precisely one: the *return period* of these high peaks is $1/\mu$. The return period of downcrossings is also $1/\mu$ and that of the crossings is $2/\mu$.

We can even give a much more specific result concerning the asymptotic time distribution of peaks higher than a level u. As u tends to infinity, it can be shown that the time distance between two consecutive peaks higher than u has, asymptotically, an exponential distribution with probability density $\mu\, e^{-\mu t}$ (the return period $1/\mu$ being the scale parameter), and that the successive time intervals between such peaks are asymptotically independent. The occurrence of such high peaks will, asymptotically, constitute a Poisson process and if $c > 0$ is a number independent of u, the probability of exactly m peaks higher than u in a time interval of length c/μ converges, as $u \to \infty$, to the Poisson limit $c^m\, e^{-c}/m!$.

From this result, the limiting probability distribution for the largest ordinate of $X(t)$ in a large observation period $[0, T]$ is easily deduced. If we denote $\tilde{X}(T) = \max_{0 \le t \le T} \{X(t)\}$, the largest value assumed by $X(t)$ for $0 \le t \le T$, we have the following relation:

$$\text{Prob}\{\, \tilde{X}(T) \le \sqrt{2 \log T} + \frac{z + \log\left(\sqrt{\lambda_2}\,/2\pi\right)}{\sqrt{2 \log T}} \,\} \to \Lambda(z) = \exp(-e^{-z})$$

as $T \to \infty$. Note that this result is very similar to that given for *independent* normal sequences, the already classical case.

For the smallest value of $X(t)$ in $[0, T]$ there is, of course, a corresponding relation.

When u is large, any peak higher than u of $X(t)$ will have the following appearance: the upcrossing of the level u will be followed, after an average time asymptotic to $\left(\sqrt{2\pi/\lambda_2}\right) / u$, by a downcrossing of the same level. Half-way between an upcrossing and the next downcrossing there will be a maximum of the curve, and the excess of this maximum over the u level is small when u is large.

Let us deal, briefly, with the length and size of the fluctuations of $X(t)$. Suppose first that it is known that $X(t)$ has a local maximum at some point t_1. The probability distribution of the ordinate $X(t_1)$, relative to this hypothesis, has an explicit analytic expression with a mean value equal to $\lambda_2 / 2\sqrt{2\pi\, \lambda_4}$.

Let us now suppose that $X(t)$ has two consecutive minima at $t = t_1$ and $t = t_3$, and a maximum at $t = t_2$, where $t_1 < t_2 < t_3$. The wave of $X(t)$ in $[t_1, t_3]$ has the length $t_3 - t_1$, its height (or amplitude) being defined as $X(t_2) - [X(t_1) + X(t_3)]/2$. The probability distributions of the length and height of the wave are known and and their mean values are:

$$\text{Mean length} = 2\pi\sqrt{\lambda_2 \lambda_4}$$

$$\text{Mean height} = \lambda_2\sqrt{2\pi\ \lambda_4}\ .$$

As $\lambda_0 = 1$ we have $\lambda_4 \geq \lambda_2^2$. If the spectral density $f(\lambda)$ differs appreciably from zero for large λ, the spectrum of the process contains high frequencies, and λ_4 will be large. The mean length and height of the waves will then be small, which means that $X(t)$ will show a large number of fluctuations of small duration and size.

In the opposite extreme case, when λ_4 only exceeds λ_2^2 by a small amount, the spectrum will be concentrated in a small interval about some central frequency λ^*, and the length of the $X(t)$ waves will then all be approximately equal to $2\pi/\lambda^*$, while the size will be variable, with a mean of approximately $\sqrt{2\pi}$.

Global extremes in a Wiener-Lévy processes

Let $X(t)$ be the Wiener-Lévy stochastic process, the first approach given to the theory of Brownian motion. We will seek the stochastic behaviour of the global maximum $\tilde{X}(t) = \max_{0\leq t\leq T} X(t)$. The result to be given is due to Lévy. We will briefly follow Papoulis (1965).

As is well known, the Brownian motion $X(t)$ has $X(0) = 0$ and $\text{Prob}\{X(t) \leq x\} = N(x/\sigma\sqrt{t})$ where $N(x)$, as before, denotes the standard normal distribution function and σ is the dispersion parameter of the process (of $X(1)$).

It can be shown that the distribution function of $\tilde{X}(t)$ is the same as that of $|X(t)|$, that is,

$$\text{Prob}\{\tilde{X}(t) \leq x\} = 0 \quad \text{if } x < 0$$

$$= 2\,N(x/\sigma\sqrt{T}) - 1 \quad \text{if } x > 0$$

owing to the symmetry $[N(x) + N(-x) = 1]$ of $N(.)$.

We can roughly sketch the proof of this result. The basic fact is the reflection symmetry principle (of D. André); if a trajectory $X(t)$ in any instant $(<T)$ attains the level $d(>0)$ the probabilities of $X(t)$ being later, at some instant, in positions $d + a$ or $d - a$ are equal, or equivalently, that for each trajectory crossing the line $x = d\ (>0)$ there is a symmetrical (reflected) trajectory on this line. Then

$$\text{Prob}\{\tilde{X}(T) \geq d\} = \text{Prob}\{X(T) \geq d\} + \text{Prob}\{X(T) < d\ ,\ \tilde{X}(T) \geq d\}$$

$$= 2\ \text{Prob}\ \{X(T) \geq d\}$$

because, by André's principle. we have

$$\text{Prob}\{X(T) < d,\ \tilde{X}(T) \geq d\} = \text{Prob}\{X(T) \geq d\ ,\ \tilde{X}(T) \geq d\} = \text{Prob}\ \{X(T) \geq d\}\ ;$$

it is evident that for minima the result is symmetrical to the line $x = 0$.

Local extremes in Gaussian stationary processes

Cartwright and Longuet-Higgins (1956) have shown that if $X(t)$ is a twice-differentiable Gaussian stationary process and $p(\xi, \eta, \zeta)$ is the trinormal probability density of the random triple (X, X', X''), the density of X, *known* to be a maximum, is given by

$$f(\xi) = \frac{\int_{-\infty}^{0} d\zeta\ p(\xi, 0, \zeta)\ \zeta}{\int_{-\infty}^{+\infty} d\zeta \int_{-\infty}^{0} d\zeta\ p(\xi, 0, \zeta)\ \zeta}\ .$$

The reasoning, based on Rice (1944/45), is roughly as follows: let $[t, t + dt]$ be an (infinitesimal) time interval. If there is a maximum in $[t, t + dt]$ we must have $X'(t') = 0$ in that interval and $X''(t') < 0$. As the trajectories $X'(t)$ are almost linear, we have approximately $|X'(t)| \leq |X''(t)|\, dt$ in the interval and, consequently, the straight line close to $X'(t)$, passing by η and with the slope ζ, has a zero at $t - \eta/\zeta$ so that $0 \leq -\eta/\zeta \leq dt$. The density of ζ, relative to this strip, is then $(\zeta < 0)$

$$\frac{\int_{-\infty}^{0} d\zeta \int_{0}^{-\zeta\, dt} d\eta\; p(\xi, \eta, \zeta)}{\int_{-\infty}^{+\infty} d\xi \int_{-\infty}^{0} d\zeta \int_{0}^{-\zeta\, dt} d\eta\; p(\xi, \eta, \zeta)}$$

whose limit, as $dt \downarrow 0$, is given by

$$\frac{\int_{-\infty}^{0} d\zeta\; p(\xi, 0, \zeta)\, \zeta}{\int_{-\infty}^{+\infty} d\xi \int_{-\infty}^{0} d\zeta\; p(\xi, 0, \zeta)\, \zeta} .$$

This heuristic derivation — evidently not rigourous — gives a good idea of the behaviour of $X'(t)$ on a local scale. A correct — and longer! — proof can be given.

For a Gaussian stationary process, with zero means, the covariance matrix of (X, X', X'') is, as is well known,

$$\begin{bmatrix} \sigma^2 & 0 & -\sigma^2 \lambda_2 \\ 0 & \sigma^2 \lambda_2 & 0 \\ -\sigma^2 \lambda_2 & 0 & \sigma^2 \lambda_4 \end{bmatrix}$$

where the λ_j, as before, denote the jth moments of the spectral density of the standardized process $X(t)/\sigma$. For the standard process $X(t)/\sigma$, as the density $p(\xi, \eta, \zeta)$ is the trinormal one with zero means and the covariance matrix given, the density of $X(t)/\sigma$ is

$$p(x) = \frac{1}{\sqrt{2\pi}} \{\varepsilon\, e^{-x^2/2}\, \varepsilon^2 + (1 - \varepsilon^2)^{1/2}\, x\, e^{-x^2/2} \int_{-\infty}^{x(1-\varepsilon^2)^{1/2}/\varepsilon} e^{-x^2/2}\, dx\}$$

where $\varepsilon^2 = 1 - \lambda_2^2/\lambda_4$ (≥ 0 by Schwartz inequality) and $X(t)$ evidently has the density

$$f(\xi) = p(\xi/\sigma)/\sigma .$$

The distribution of the observed maximum then depends on only two parameters: a dispersion parameter σ and a quantity ε ($0 \leq \varepsilon \leq 1$) which can be interpreted as the relative spectrum width. For $\varepsilon = 0$ we have the Rayleigh (minima) density

$$p(x) = 0 \qquad \text{if} \qquad x < 0$$

$$= x\, e^{-x^2/2} \qquad \text{if} \qquad x < 0$$

and for $\varepsilon = 1$ we obtain the normal density

$$p(x) = \frac{1}{\sqrt{2\pi}}\ e^{-x^2/2}\ ;$$

ε is thus a measure of non-normality.

For the reduced minimum $\underset{\sim}{X}$ — which is easily obtained by reflection on $x = 0$ — we have

$$q(x) = \frac{1}{\sqrt{2\pi}}\ \{\varepsilon\, e^{-x^2/2}\ \varepsilon^2 - (1-\varepsilon^2)^{1/2}\ x\ e^{-x^2/2} \int_{-\infty}^{-x(1-\varepsilon^2)^{1/2}/\varepsilon} e^{-x^2/2}\ dx\}\ .$$

Some further results

As our main interest has been directed towards Gaussian processes, we will only give some scattered examples of extreme value properties for other kinds of stochastic processes.

Consider the class of processes $X(t)$ *with independent increments* such that $X(0) = 0$; for $t \geq 0$ the characteristic function of $X(t)$ is

$$M(e^{iy\,X(t)}) = \exp\left[-\frac{1}{2}\ \sigma^2\, t\, y^2 + t \int_{-\infty}^{+\infty} (e^{iyv} - 1 - i\, y\, v)\ d\, G(v)\right],$$

where σ is a constant and $G(v)$ is bounded and non-decreasing. If $G(v)$ is identically zero, we have the Wiener-Lévy, or Brownian motion, process dealt with previously. On the other hand, if $\sigma = 0$ and G is constant except for a jump of magnitude 1 at $v = 1$, we have a centered Poisson process. For the general class of processes with independent increments, with the characteristic function given above, the same formula for $\tilde{X}(T)$ holds asymptotically for large T, with σ^2 replaced by $\sigma^2 + \int_{-\infty}^{+\infty} v^2\, d\, G\,(v)$. In particular this result applies to the Poisson process.

Consider now a process with independent increments and a slightly modified form of the characteristic function:

$$M(e^{iy\,X(t)}) = \exp\,[\,-t \int_0^\infty (e^{iyv} - 1 - (1+\lambda)\,i\,y\,v)\;d\,G(v)]$$

where $\lambda > 0$ is a constant, while $G(v)$ is a distribution function with a mean value 1. This process occurs in Insurance Risk Theory, where G is the distribution function of the amount of a claim, while λ represents the "loading" of the risk premium. $X(t)$ is here the total amount of loaded risk premiums paid during the time interval $[0, t]$, minus the total amount of claims that have occurred during the same period. If we assume that $X(0) = u$ is the initial capital of the risk business, and consider this business to be ruined if $X(t)$ ever takes a value ≤ 0, the ruin probability $\Psi(u)$ is the probability that the minimum value of $X(t)$ for some positive t will be ≤ 0. A fundamental result of Risk Theory asserts that

$$\Psi(u) \leq e^{-Ru}\,,$$

where R is the smallest positive root of the equation

$$\int_0^\infty e^{Rv}\,d\,G(v) = 1 + (1+\lambda)\,R\,.$$

As a final example consider the *integral of a stationary process*. We write for $t > 0$

$$I\,X(t) = \int_0^t X(u)\,d\,u\,,$$

where $X(u)$ is a gaussian stationary process satisfying the conditions above. Moreover, we assume here that the spectral density $f(\lambda)$ is continuous and positive at the point $\lambda = 0$. We then have $M(IX(t)) = 0$ and $V(I\,X(t)) = M(IX^2(t)) \sim \pi\,t\,f(0)$ as $t \to \infty$.

For the normalized process $\overline{I\,X}\,(t) = I\,X(t) / \sqrt{V(I\,X(t))}$ it can then be shown that we have, with $I\,\tilde{X}\,(t) = \max_{0 \leq t \leq T} \overline{I\,X}\,(t)$,

$$\text{Prob}\{\,\sqrt{(1-\varepsilon)\log\log T} < I\,\tilde{X}(T) < \sqrt{(1+\varepsilon)\log\log T}\,\} \to 1$$

for any $\varepsilon > 0$, as $T \to \infty$.

For other details see the papers by Berman (1962), (1964), (1964'), (1974) and (1982) and Pickands (1967), (1967'), (1968), (1969) and (1969') as well as the references given in the first section.

Chapter XIV

EXTREMAL SEQUENCES AND PROCESSES; BASICS AND STATISTICS

Introduction

We present a definition of extremal (maxima) random sequences and stochastic processes [the latter being distinct from the limiting constructions given by Dwass (1964) and Lamperti (1964)] following Tiago de Oliveira (1968); the insistence will be more on an *actual* characterization than on a limiting one. Extremal sequences are the usual sequences of maxima, and extremal processes will be obtained as stochastic processes (in continuous time) of independent extremes (maxima) with extremal (maxima) margins. The limiting approaches will be described in a special section.

The definition being the weak one (in law), as seen essentially in the identifiability problem, the properties described are, chiefly, the second order properties for the case studied, with Gumbel margins.

It will be clear that extremal processes are surely continuous to the right and almost surely with discontinuous trajectories.

We will always assume the marginal distributions of extremal sequences and processes to be the Gumbel distribution ($\Lambda(z) = \exp(-e^{-z})$), for reduced random variables). The other extremal sequences and processes are easily converted to Gumbel margins by logarithmic transformations if they are maxima and also using symmetry considerations when minima; its form will, naturally, be given in the limiting characterization section.

The definition of processes by a non-limiting technique is analogous to that of Mann (1953) for Brownian motion, using the know "duality", for independence, between sums and maxima (characteristic functions versus distribution functions), now extended to stochastic processes with continuous time. We also obtain stochastic processes, connected with maxima, whose one-

dimensional finite behaviour was the one described asymptotically by Newel (1962), for diffusion processes, like the Orstein-Uhlenbeck Brownian motion. The structure of the jump process is obtained as well as a natural representation.

Let us recall some results.

A general Gumbel random variable has the distribution function $\Lambda((x-\lambda)/\delta)$ where $\Lambda(z) = (\exp(-e^{-z}))$ is the distribution function of a reduced random variable ($\lambda = 0, \delta = 1$); we also have $\Lambda'(z) = e^{-z}\,\Lambda(z)$. The first moments are $\mu = \lambda + \gamma\,\delta$ ($\gamma = 0.57722\ldots$, Euler's constant), $\sigma^2 = \pi^2/6\,.\,\delta^2 = 1.64493\ldots\,\delta^2$, $\sigma = \pi/6\,.\,\delta = 1.28825\ldots\,\delta$, $\beta_1 = \mu_3/\sigma^3 = 1.13955$, and $\beta_2 = \mu_4/\sigma^4 = 5.4$; see Gumbel (1958).

An extreme random pair (Z_1, Z_2) with reduced Gumbel margins has the distribution function $\Lambda(x,y) = \text{Prob}\{Z_1 \le x\,,\, Z_2 \le y\} = \exp\{-(e^{-x} + e^{-y})\,k\,(y-x)\} = \{\Lambda(x)\,\Lambda(y)\}^{k(y-x)}$ where the dependence function $k(.)$ satisfies some conditions for $\Lambda(.\,,.)$ to be a distribution function with reduced Gumbel margins; see Tiago de Oliveira (1962/63). In particular we have $\frac{\max(1,e^w)}{1+e^w} \le k(w) \le 1$, the lower bound corresponding to the diagonal case where we have $\text{Prob}\{Y = X\} = 1$ (for reduced margins) and $k(w) = 1$ corresponding to independence.

The correlation coefficient has the expression

$$\rho = -\frac{6}{\pi^2} \int_{-\infty}^{+\infty} \log k(w)\; d\,w\; (\ge 0)\,,$$

$\rho = 0$ being equivalent to independence and $\rho = 1$ to the diagonal case.

The probability $D(w) = \text{Prob}\{Z_2 - Z_1 \le w\}$, (Z_1, Z_2) being reduced Gumbel random variables, has the expression $D(w) = \frac{k'(w)}{k(w)} + \frac{e^w}{1+e^w}$ a.e., as $k'(.)$ exists a.e.,; in the independence case ($k(w) = 1$) we have the logistic distribution and in the diagonal case we have $D(w) = H(w)$, where $H(w)$ is the Heaviside jump function ($H(w) = 0$ if $w < 0$ and $H(w) = 1$ if $w \ge 0$). Note that $\partial^2 \Lambda/\partial\,x\,\partial\,y$ does not exist in all cases — as $k''(.)$ does not always exist; this is the case in all stochastic processes (sequences) connected with extremes, as happens in extremal sequences and processes and in EMS and EME sequences.

In particular, for $k_\theta(w) = 1 - \frac{\min(\theta, e^w)}{1+e^w}$ $(0 \le \theta \le 1)$, we have $D_\theta(w) = 0$ if $w < \log\theta$ (< 0) (i.e., $\text{Prob}\{Z_2 \ge Z_1 + \log\theta\} = 1$) and $D_\theta(w) = \frac{1}{1+(1-\theta)\,e^w}$ if $w \ge \log\theta$ with a

jump of θ at $w = \log \theta$, the jump being the probability of $Y = X + \log \theta$; the correlation coefficient is

$$\rho(\theta) = \frac{6}{\pi^2} R(\theta)$$

where $R(\theta) = \int_0^\theta \frac{-\log t}{1-t} \, dt$, $R'(\theta) \geq 0$ with $R(0) = 0$, $R(1) = \frac{6}{\pi^2}$ so that for $\theta = 0$ and $\theta = 1$ we have, obviously, the independence and diagonal cases. The expression of $k_\theta(w)$ corresponds to the *biextremal model.*

The other coefficients are the difference-sign correlation $\tau(\theta) = \theta$, the grade correlation $\chi(\theta) = 3\theta/(2 + \theta)$, and the medial correlation $\nu(\theta) = 2^\theta - 1$, all increasing from 0 to 1 as θ increases from $\theta = 0$ to $\theta = 1$.

Extremal sequences

Let $E_1, E_2, \ldots, E_n, \ldots$ be a sequence of i.i.d. reduced Gumbel random variables and let $Z_1 = E_1$, $Z_2 = \max(Z_1, E_2), \ldots, Z_n = \max(Z_{n-1}, E_1) = \max(E_1, \ldots, E_n)$.

Evidently $\text{Prob}\{Z_1 \leq z_1, \ldots, Z_n \leq z_n\} = \text{Prob}\{Z_i \leq z_i, i = 1, \ldots, n\} = \text{Prob}\{E_1 \leq \min_1^n z_i\}$, $E_2 \leq \min_2^n z_i, \ldots\} = \prod_{i=1}^n \Lambda(\min(z_i, \ldots, z_n))$. This corresponds to putting $t_1 = 1, \ldots, t_n = n$ in the extremal processes, so we will only sketch some results. We can now consider the parameterized sequence $X_n = \lambda + \delta Z_n$ with real λ and $\delta(> 0)$. We have $\text{Prob}\{X_i \leq x_1, \ldots, X_n \leq x_n\} = \text{Prob}\{Z_i \leq (x_i - \lambda)/\delta, \ldots, Z_n \leq (x_n - \lambda)/\delta\} = \prod_{i=1}^n \Lambda((\min(x_i, \ldots, x_n) - \lambda)/\delta)$.

Let us now consider a periodic subsequence $X'_j = X_{jp}$, $p(> 0)$ integer. We can show, in two ways, that $\{X'_j\}$ and $\{X_j\}$ have the same distribution apart from the parameters, which are (λ, δ) for the sequence $\{X_n\}$ and $(\lambda', \delta') = (\lambda + \delta \log p, \delta)$ for the sequence $\{X'_n\}$. A simple one is to consider the sequence $E'_1 = \max(E_1, \ldots, E_p) - \log p$, $E'_2 = \max(E_{p+1}, \ldots, E_{2p}) - \log p$, etc. It

is immediate that the E_j' are independent random variables with the reduced Gumbel distribution. If we denote by $\{Z_i' = E_1', Z_2' = \max(E_1', E_2'), \ldots, Z_n' = \max(E_1', \ldots, E_n')$, we see that

$$Z_j' = Z_{jp} - \log p \qquad \text{and so} \qquad X_j' = \lambda + \delta(Z_j' + \log p) = \lambda' + \delta Z_j',$$

and then $\text{Prob}\{X_j' \le x_1, \ldots, X_n' \le x_n\} = \prod_{i=1}^{n} \Lambda((\min_i (x_i, \ldots, x_n) - \lambda')/\delta)$, which is the same as that of $\{X_j\}$ with the new parameters.

Thus a change of timescale — or a periodic selection at instants p, 2p, ... — leads to extremal sequences with the same properties. In particular we have $M(X_j) = \lambda + \delta(\gamma + \log j)$, $C(X_j, X_l) = \delta^2 R(\frac{\min(j, l)}{\max(j, l)})$, $M(X_j') = M(X_{jp}) = \lambda + \delta(\gamma + \log(p_j)) = \lambda' + \delta \log j$, and $C(X_j', X_l') = \delta^2 R(\frac{\min(j, l)}{\max(j, l)})$. Periodic sub-sampling does not alter the structure of the $\{X_j\}$ apart from the location effect of the period.

Another proof is analogous to the one given for extremal processes where this effect is integrated in a general change of the time scale.

An ergodic (type) theorem, analogous to the one for extremal processes, can be written. It is $b/a \to \infty$,

$$\frac{1}{\log(b/a)} \sum_{a+1}^{b} \frac{X_k}{k} - \delta \log \sqrt{a\,b} \xrightarrow{ms} \lambda + \gamma\,\delta.$$

The proof is not important being analogous to the one to follow and statistically irrelevant.

Extremal processes; distribution and properties

A stochastic process $Z(t)$ $(t \ge 0)$ is called an *extremal process* (of maxima) if :

1) $Z(t + \theta) = \max(Z(t), E(t,\theta))$, where $X(t)$ and $E(t,\theta)$ are independent random variables $(\theta \ge 0)$;

2) $E(t,\theta)$ and $E(t',\theta')$ are independent random variables if the time intervals $]t, t + \theta]$ and $]t', t' + \theta']$ are disjoint;

3) the processes is time-homogeneous, that is, the distribution of $E(t,\theta)$ depends only on θ and is independent of t;

4) $X(0) = -\infty$.

If G(x) is the distribution function of X(1) we have

$$Prob\{Z(t) \leq x\} = G^t(x) \text{ and } Prob\{E(t,\theta) \leq x\} = G^\theta(x).$$

Let $F(x,t)$ be the distribution function of $X(t)$ and denote by $F_1(x,\theta)$ the distribution function of $E(t,\theta)$, independent of t by assumption 3).

Thus we have, by 1) and 2),

$$F(x, t+\theta+\theta') = F(x, t)\, F_1(x, \theta+\theta') = F(x, t+\theta)\, F_1(x, \theta')$$

$$= F(x, t)\, F_1(x, \theta)\, F_1(x, \theta')$$

so that

$$F_1(x+\theta+\theta') = F_1(x, \theta)\, F_1(x, \theta')$$

As the solution of this functional equation, for measurable functions, is the exponential function, we have

$$F_1(x, \theta) = F_1^{\theta}(x, 1)$$

As $F(x, t) = F(x, 0)\, F_1(x, t) = F(x, 0)\, F_1^t(x, 1)$ and, by 4). $F(x, 0) = 1$ we obtain

$$F(x, t) = F_1^t(x, 1) = G^t(x)$$

as

$$F(x, 1) = F_1(x, 1) = G(x).$$

This result was obtained asymptotically by Newell (1962) under other conditions.

Let us finally introduce a new assumption, that one-dimensional margins have Gumbel distributions, by:

5) $G(x) = \Lambda(x)$ (standardization of the time unit).

We have then

The one-dimensional distribution of $Z(t)$ *is* $\Lambda^t(x) = \Lambda(x - \log t)$;

the two-dimensional distribution of $(Z(t_1), Z(t_2))$ $(t_1 < t_2)$ *is*

$$\Lambda(\min(x_1, x_2) - \log t_1) \cdot \Lambda(x_2 - \log(t_2 - t_1)) ;$$

and, in general, the n-dimensional distribution of $(Z(t_1), Z(t_2), \ldots, Z(t_n))$ *is*

$$\prod_{i=1}^{n} \Lambda(\min(x_1, \ldots, x_n) - \log(t_i - t_{i-1})) \quad (t_0 = 0 < t_1 < \ldots < t_n) .$$

It is also immediate that, with H(w) denoting the Heaviside distribution function of the almost sure random variable equal to zero, then

X(t) is a Markoff process with transition probability function

$$P(x_1, t_1 ; x_2, t_2) = H(x_2 - x_1) \ \Lambda^{t_2 - t_1}(x_2) .$$

From this transition probability function we get, denoting by $q(x_1, t_1)$ and $Q(t_1, x_1, x_2)$ the *process intensity* and the (relative) *transition probability function* (using formula (8.8.5) of M. Fisz (1962)),

$$P(x_1, t_1 ; x_2, t_2) = [1 - q(x_1, t_1)(t_2 - t_1)] H(x_2 - x_1)$$

$$+ q(x_1, t_1) Q(t_1, x_1, x_2)(t_2 - t_1) + o(t_2 - t_1) ,$$

the following result:

An extremal process is a purely discontinuous (step) Markoff process where

$$P(x_1, t_1, x_2, t_2) = [1 - e^{-x_1}(t_2 - t_1)] H(x_2 - x_1) + o(t_2 - t_1) ,$$

so that

$$q(x_1, t_1) = e^{-x_1} \qquad \text{and} \quad Q(t_1, x_1, x_2) = H(x_2 - x_1)(1 - e^{x_1 - x_2}) .$$

For this purely discontinuous Markoff process we can immediately write the classical Kolmogoroff-Chapman integral equations.

Let $T(a)$ denote the *upcrossing time* of the level a. As Prob $\{T(a) > t\}$ = Prob $\{Z(t') < a, t' \leq t\} = \Lambda^t(a) = \Lambda(a - \log t)$, we can say that:

The distribution function of the upcrossing time is

$$Prob\,\{T(a) \le t\} = 1 - \Lambda(a - \log t)\,.$$

Some more results can also be given.

A first one, relating the chance experiments between points of time, is that for $s \le u \le t$ we have $E(s\,,\,t) = \max\,(E(s\,,\,u)\,,\,E\,(u\,,\,t))$, as follows from the two expressions of $Z(t)$ in $Z(s)$, directly and by way of $Z(u)$.

As $Z(t) - \log t = E(0\,,\,t) - \log t$ *is a reduced Gumbel random variable we see that* $\frac{Z(t)}{\log t} \overset{ms}{\to} 1$.

Finally the least squares predictor of $Z(t)$, knowing $Z(t_1) = z_1, \ldots, Z(t_n) = z_n (0 < t_1 < t_2 < \ldots < t_n < t)$, as $Z(t)$ is markovian, it depends only on z_n is given by

$$p(t\,;\,z_n\,,\,t_n) = z_n + \int_{z_n}^{+\infty} (1 - \Lambda(z - \log\,(t - t_n)))\,d\,z\ (\ge z_n)$$

whose mean-square error is

$$\int_{-\infty}^{+\infty} \left[\int_{z_n}^{+\infty} (1 - \Lambda(y - \log\,(t - t_n)))\,d\,y^2 - [p^2(t\,;\,z_n\,,\,t_n) - z_n^2\,]\,d\,\Lambda^{t-\,t_n}\,(z_n)\right.\,.$$

From the practical point of this is not very rewarding and it is simpler to use the linear predictor $Z(t) = z_n + \log t/t_n$.

Notice that, as $Z(t)$ is not decreasing, we know that $\tilde{Z}\,(t) = \max\limits_{0 \le s \le t} Z(s) = Z(t)$.

The limiting characterizations

We will sketch another characterization of extremal processes; for details and enlargements see Dwass (1964, 1966, 1973), Lamperti (1964), Resnick (1973, 1975) and Resnick and Rubinovitch (1973); for other approaches see references in Galambos (1978) and Leadbetter, Lindgren and Rootzén (1983).

Let $\{X_j\}$, $j = 1,2, \ldots$ be a sequence of i.i.d. random variables with distribution function $F(x)$ and such that there exist λ_n and $\delta_n (> 0)$ such that $\text{Prob}\{\max_1^n (X_i - \lambda_n)/\delta_n \le x\} = F^n(\lambda_n + \delta_n x) \xrightarrow{w} L(x)$, where the proper and non-degenerate limiting distribution is necessarily a Gumbel distribution function $\Lambda(x) = \exp(-e^{-x})$, or a Fréchet distribution function $\Phi_\alpha(x) = 0$ if $x \le 0$, $\Phi_\alpha(x) = \exp(-e^{-\alpha})$ if $x \ge 0$ with $\alpha > 0$, or even a Weibull distribution function $\Psi_\alpha(x) = \exp(-(-(x)^\alpha)$ if $x \le 0$, $\Psi_\alpha(x) = 1$ if $x \ge 0$ with $\alpha > 0$.

Let us define, for each $n = 1,2, \ldots$ and $t \ge 0$, stochastic processes

$$M_n(t) = (\max(X_1, \ldots, X_{[nt]}) - \lambda_n)/\gamma_n \quad \text{for } t > 1/n$$

$$= (X_1 - \lambda_n)/\delta_n \quad \text{for } 0 \le t \le 1/n ,$$

$[u]$ denoting as usual the largest integer $\le u$.

As $\text{Prob}\{M_n(t) \le x\} = F^{[nt]}(\lambda_n + \delta_n x) = [F^{[n]}(\lambda_n + \delta_n x)]^{[nt]/n} \to L^t(x)$

and more generally, for $0 < t_1 < t_2 < \ldots t_k$,

$$\text{Prob}\{M_n(t_1) \le x_1 , M_n(t_2) \le x_2, \ldots , M_n(t_k) \le x_n\} \to$$

$$L^{t_1}(\min_1^k (x_i))\, L^{t_2 - t_1}(\min_2^k (x_i))\, L^{t_3 - t_2}(\min_3^k (x_i)) \ldots L^{t_k - t_{k-1}}(x_k) ,$$

we define an extremal reduced process $M(t)$ as a limit of $M_n(t)$ which extends to Fréchet and Weibull margins the extremal process defined with Gumbel margins.

It is evident that all "margin-free" properties such as the jump process described below. (to be Poisson-distributed) and all non-parametric correlation coefficients have the same behaviour as in extremal process with Gumbel margins, which will continue to be considered in what follows. Note, in particular, that second order properties are not always valid for extremal processes with Fréchet or Weibull margins.

Second order properties of extremal processes

The second order properties of extremal processes (with Gumbel margins, to be dealt with in the rest of the chapter) depend on the form of the covariance (or correlation) function, as the margins have mean value and variance.

From the expression of the two-dimensional distribution function of $(Z(t_1), Z(t_2))$ passing to the reduced margins $\xi_1 = x_1 - \log t_1$ and $\xi_2 = x_2 - \log t_2$, we see that the mean value function of $Z(t)$ is $\mu(t) = \gamma + \log t$ and the dependence function ($t_1 \le t_2$) is the biextremal one

$$k(w;\, t_1/t_2) = \frac{(1 - t_1/t_2) + \max(e^w,\, t_1/t_2)}{1 + e^w} = 1 - \frac{\min(t_1/t_2,\, e^w)}{1 + e^w},$$

with parameter $\theta = t_1/t_2$ $(0 \le \theta \le 1)$. Consequently :

An extremal (Gumbel) location-dispersion free process has the mean value $\gamma + \log t$, *variance* $\pi^2/6$ *and the correlation function*

$$\rho(t_1,t_2) = \rho(t_1/t_2) = 1 + \frac{6}{\pi^2} \int_{t_1/t_2}^{1} \frac{\log \alpha}{1 - \alpha}\, d\alpha = -\frac{6}{\pi^2} \int_{0}^{t_1/t_2} \frac{\log \alpha}{1 - \alpha}\, d\alpha = \frac{6}{\pi^2} R(t_1/t_2)$$

the processes is integrable, continuous and indefinitely differentiable in mean square except at origin (0,0) where ρ *is not continuous.*

The mean-square continuity is, evidently, consistent with the almost sure discontinuity, the discontinuities being moving (random) discontinuities; the graph of the process is a non-decreasing random step function.

The jumps of an extremal process

As we showed, the transition probability function is

$$P(x_1, t_1 ; x_2, t_2) = H(x_2 - x_1)\, \Lambda^{t_2 - t_1}(x_2) \quad (t_1 < t_2).$$

Consequently the probability of zero jumps in the time interval $]t_1, t_2]$ is given by

$$p_0(t_1, t_2) = \int_{-\infty}^{+\infty} \Lambda^{t_2 - t_1}(x)\, d\, \Lambda^{t_1}(x) = t_1 / t_2,$$

because the probability of zero jumps in $]t_1, t_2]$ if $Z(t_1) = x_1$ (which implies $Z(t_2) = x_1$) is $\Lambda^{t_2 - t_1}(x_1)$; the probability that we have at least one jump is then

$$p_{>0}(t_1, t_2) = 1 - t_1 / t_2.$$

Notice that for every $t > 0$ we have $\text{Prob}\{Z(t) > -\infty\} = p_{>0}(0, t) = 1$ so that:

For $t > 0$, $Z(t)$ is finite with probability one.

If $p_N(t_1, t_2)$ denotes the probability of N jumps in the interval $]t_1, t_2]$, the probability of zero jumps in the interval $]t_2, t_2 + \theta]$ is given by

$$p_0(t_2, t_2 + \theta) = \frac{t_2}{t_2 + \theta} = 1 - \frac{\theta}{t_2} + o(\theta),$$

and as

$$p_{>0}(t_2, t_2 + \theta) = \frac{\theta}{t_2} + o(\theta),$$

as is natural and will be proved later, then

$$p_1(t_2, t_2 + \theta) = p_{>0}(t_2, t_2 + \theta) + o(\theta).$$

The basic equation is

$$p_{N+1}(t_1, t_2 + \theta) = p_{N+1}(t_1, t_2) \cdot p_0(t_2, t_2 + \theta)$$

$$+ p_N(t_1, t_2) \cdot p_1(t_2, t_2 + \theta) + o(\theta).$$

which, as derivatives exist, can be written as

$$p_{N+1}(t_1, t_2) + \frac{\partial p_{N+1}(t_1, t_2)}{\partial t_2} \theta = p_{N+1}(t_1, t_2)\left(1 - \frac{\theta}{t_2}\right)$$

$$+ p_N(t_1, t_2)\frac{\theta}{t_2} + o(\theta).$$

and gives, for $\theta \to 0$,

$$\frac{\partial p_{N+1}(t_1, t_2)}{\partial t_2} = \frac{1}{t_2}\left(p_N(t_1, t_2) - p_{N+1}(t_1, t_2)\right).$$

If we denote by $\Delta(z; t_1, t_2) = \sum_0^\infty p_N(t_1, t_2)\, z^N$ the generating function of the probabilities, then $\Delta(1; t_1, t_2) = 1$ and $\Delta(z; t_1, t_1) = 1$.

The generating function Δ, from the difference-differential equation for the probabilities, satisfies the partial differential equation

$$\frac{1}{z}\left(\frac{\partial \Delta}{\partial t_2} - \frac{\partial\, p_0(t_1, t_2)}{\partial t_2}\right) = \frac{1}{t_2}\left(\Delta - \frac{\Delta - p_0(t_1, t_2)}{z}\right).$$

From $p_0(t_1, t_2) = t_1 / t_2$ we obtain

$$\frac{1}{\Delta}\frac{\partial \Delta}{\partial t_2} = \frac{z-1}{t_2},$$

whose solution is evidently

$$\Delta(z; t_1, t_2) = \left(\frac{t_2}{t_1}\right)^{z-1} = e^{(z-1)\log(t_2/t_1)}$$

owing to the known values of $\Delta(1; t_1, t_2) = 1$ and $\Delta(z; t_1, t_1) = 1$. But Δ is the generating function of the Poisson distribution with parameter $\nu = \log(t_2 / t_1)$, and so

The jumps of $Z(t)$ in the interval $]t_1, t_2]$ have a Poisson distribution with parameter $\nu = \log(t_2 / t_1)$ (>0); the mean number and variance of jumps increase logarithmically with time.

Thus in the finite interval $]0, t]$ we must expect an infinity of jumps; the natural science interpretation is easy: in the process beginning at $-\infty$ $(Z(0) = -\infty)$ we should except an infinite number of jumps to obtain any finite value.

Let us show, finally, that

$p_1(t, t+\theta) = \frac{\theta}{t} + o(\theta)$; as $p_{>0}(t, t+\theta) = \frac{\theta}{t} + o(\theta)$, it is sufficient to show that $p_{>1}(t, t+\theta) = o(\theta)$. As we have two or more jumps on $]t, t+\theta]$, let $\tilde{\theta}$ denote the mid-point of the instants of the first and last jump and $\pi(\tilde{\theta})$ its distribution function. We can clearly write

$$p_{>1}(t, t+\theta) = \int_0^{\theta} p_{>0}(t, t+\tilde{\theta})\; p_{>0}(t+\tilde{\theta}, t+\theta)\, d\pi(\tilde{\theta})$$

$$= \int_0^{\theta} \frac{\tilde{\theta}}{t+\tilde{\theta}} \cdot \frac{\theta - \tilde{\theta}}{t+\theta}\, d\pi(\tilde{\theta}) \le \frac{\theta^2}{t^2} \int_0^{\theta} \frac{\tilde{\theta}}{\theta} \cdot \frac{\theta - \tilde{\theta}}{\theta}\, d\pi(\tilde{\theta}) \le \frac{\theta^2}{t^2}$$

with $p_{>1}(t, t+\theta) = o(\theta)$ as we wished to show.

The following representation (i.e., a stochastic process with the same finite-dimensional distributions) can be given:

For $t \geq t_0$ *the extremal process* $Z(t)$ *is equivalent in law to*

$$Z(t) = Z(t_0) \text{ for } t \in [t_0, t_1[,$$

$$Z(t) = Z(t_1) \text{ for } t \in [t_1, t_2[$$

$$\ldots$$

where $t_1 < t_2 < \ldots$ *are the jump times after* t_0, *having the densities* t_0/t_1^2 *for* $t_1 \geq t_0$, t_1/t_2^2 *for* $t_2 \geq t_1$, ... *and* $Z(t_1) = x_1$, $X(t_2) = x_2$, ... *are the values at the jump times* $t_1 < t_2 < \ldots$ *with the distribution functions*

$$\frac{\Lambda^{t_1 - t_0}(x_1) - \Lambda^{t_1 - t_0}(x_0)}{1 - \Lambda^{t_1 - t_0}(x_0)} \quad \text{for } x_1 > x_0,$$

$$\frac{\Lambda^{t_2 - t_1}(x_2) - \Lambda^{t_2 - t_1}(x_1)}{1 - \Lambda^{t_2 - t_1}(x_1)} \quad \text{for } x_2 > x_1,$$

$$\ldots$$

The proof is immediate because the probability of no jumps in the time interval $[t_0, t_1[$ is t_0/t_1 and the probability that, as there is a jump, its value is smaller than x_1 is given by the transition probability function truncated at x_0.

The study of the times of the jumps, as will be seen, is not helpful for statistical decision making.

The logarithmization of time and the ergodic theorem for extremal processes

As $\log t$ appears frequently in the expressions connected with extremal processes, we can consider the use of a new time variable $\tilde{t} = \log t$ (logarithmic time) defining a new process $\tilde{Z}(\tilde{t}) = Z(e^{\tilde{t}})$, whose characterization could be transposed from that of $Z(t)$. We will give a different one, equivalent but more natural for the case. Let us just remark that $\text{Prob}\{ \tilde{Z}(\tilde{t}) \leq x \} = \Lambda(x - \tilde{t})$.

Consider then the (transposed) extremal process:

1) $\tilde{Z}(\tilde{t})$ is defined over all the real (time) line;

2) $\tilde{Z}(\tilde{t}_2) = \max(\tilde{Z}(\tilde{t}_1), \tilde{E}(\tilde{t}_1, \tilde{t}_2))$ where $\tilde{Z}(\tilde{t}_1)$ and $\tilde{E}(\tilde{t}_1, \tilde{t}_2))$ are independent random variables;

3) $\tilde{Z}(\tilde{t})$ has the Gumbel distribution $\Lambda(x - \tilde{t})$ for convenient standardization of the time (or change of origin) and change of scale (i.e., a linear change of the clock).

Note immediately, that from 3) we have $\tilde{Z}(-\infty) = -\infty$ with probability one.

Let us denote by $\Psi(x_1, \tilde{t}_1 ; x_2, \tilde{t}_2) = \text{Prob}\{\tilde{Z}(\tilde{t}_1) \le x_1), \tilde{Z}(\tilde{t}_2) \le x_2\}$; we have $\Psi(x_1, \tilde{t}_1 ; x_2, \tilde{t}_2) = \text{Prob}\{\tilde{Z}(\tilde{t}_1) \le x_1, \tilde{Z}(\tilde{t}_2) \le x_2, \tilde{E}(\tilde{t}_1, \tilde{t}_2) \le x_2\} = \Lambda(\min(x_1, x_2) - \tilde{t}_1) \cdot G(x_2 ; \tilde{t}_1, \tilde{t}_2)$ using hypothesis 3) for the distribution of $\tilde{Z}(\tilde{t}_1)$.

With hypothesis 3) we have also, putting $x_1 = +\infty$,

$$\Lambda(x_2 - \tilde{t}_2) = \Lambda(x_2 - \tilde{t}_1) \cdot G(x_2 ; \tilde{t}_1, \tilde{t}_2)$$

and so

$\Psi(x_1, \tilde{t}_1 ; x_2, \tilde{t}_2) = \Lambda(\min(x_1, x_2) - \tilde{E}_1)\, \Lambda(x_2 - \log(e^{\tilde{t}_2} - e^{\tilde{t}_1}))$, as could be expected from the relation $t = e^{\tilde{t}}$.

Thus the mean value function is $\gamma + \tilde{t}$, the variance $\frac{\pi^2}{6}$ and the correlation function

$$\tilde{\rho}(\tilde{t}_2 - \tilde{t}_1) = \rho(e^{\tilde{t}_2} / e^{\tilde{t}_1}) = \frac{6}{\pi^2} R(e^{-|\tilde{t}_2 - \tilde{t}_1|}),$$

so $\tilde{Y}(\tilde{t}) = \tilde{Z}(\tilde{t}) - \gamma - \tilde{t}$ is a weakly stationary process.

$\tilde{Y}(\tilde{t})$ is thus an ergodic process (of second order) and we know that $\frac{1}{\tilde{b} - \tilde{a}} \int_{\tilde{a}}^{\tilde{b}} \tilde{Y}(\tilde{t})\, d\tilde{t} \xrightarrow{ms} 0$ *as* $\tilde{b} - \tilde{a} \to +\infty$.

Passing to the original clock $t = e^{\tilde{t}}$, $b = e^{\tilde{b}}$, $a = e^{\tilde{a}}$, we get

$$\frac{1}{\log(b/a)} \int_a^b \frac{Z(t) - \gamma - \log t}{t} dt \xrightarrow{ms} 0 \quad \text{as} \quad b/a \to +\infty$$

or, in a simpler form,

$$\frac{1}{\log (b/a)} \int_a^b \frac{Z(t)}{t}\, d\,t - \log \sqrt{a\,b} \xrightarrow{ms} \gamma$$

when $b/a \to +\infty$. We have for the variance of $\frac{1}{\log (b/a)} \int_a^b \frac{Z(t)}{t}\, d\,t$ the expression

$$R(a/b) + \frac{2}{(\log (b/a))^2} \int_{a/b}^{1} \left(\log \frac{b}{a} + \frac{\log u}{2} \right) \frac{(\log u)^2}{1-u}\, d\,u \ .$$

Notice that if the margins are Fréchet or Weibull, the correlation coefficient between $Z(t_1)$ and $Z(t_2)$ $(0 < t_1 < t_2)$ also depends only on t_1/t_2 (and α), when the second moments exist.

The parameterization of extremal processes

Let $Z(t)$, $t \geq 0$ be a (reduced) extremal process and consider the two (general) extremal processes

$$X_1(t) = \lambda + \delta\, Z(t) \qquad (\lambda \text{ real}, \delta > 0) \qquad \text{and}$$

$$X_2(t) = \delta\, Z(\beta\, t) \qquad (\delta\,, \beta > 0)\,.$$

Their mean values $\mu_1(t) = \lambda + \delta\,(\gamma + \log t)$ and $\mu_2(t) = \delta\,(\gamma + \log(\beta\, t))$ are equal if $\lambda = \delta \log \beta$; also the covariance function in both processes is $R\left(\frac{\min(s\,,\,t)}{\max(s,\,t)}\right)\delta^2$. This suggests that $X_1(t)$ and $X_2(t)$ can have identical distribution. In fact we have (with $t_0 = 0$)

$$\text{Prob}\,\{X_1(t_i) \leq x_i\} = \text{Prob}\{Z(t_i) \leq (x_i - \lambda)/\delta\}$$

$$= \prod_{i=1}^{n} \Lambda \left(\min_{j=i}^{n} \frac{x_j - \lambda}{\delta} \right) - \log (t_i - t_{i-1}))$$

$$= \prod_{i=1}^{n} \Lambda \{ (\min_{j=i}^{n} x_j - \lambda)/\delta - \log (t_1 - t_{i-1}) \}$$

and also

$$\text{Prob}\{X_2(t_i) \le x_i\} = \text{Prob}\{Z(\beta\, t_i) \le x_i/\delta\}$$

$$= \prod_{i=1}^{n} \Lambda\left(\min_{j=i}^{n} \frac{x_j}{\delta} - \log\left(\beta(t_i - t_{i-1})\right)\right)$$

$$= \prod_{i=1}^{n} \Lambda\left(\left(\min_{j=i}^{n} x_j - \delta \log \beta\right)/\delta - \log\left(t_i - t_{i-1}\right)\right)$$

which are equal if $\lambda = \delta \log \beta$ as said before.

Evidently any splitting (λ_1, β_1) with $\lambda = \delta \log \beta = \lambda_1 + \delta \log \beta_1$ is admissible.

Owing to linearity we will prefer the parameterization in $(\lambda\ ,\ \delta)$; any change of time unit (scale) is given by $\lambda = \delta \log \beta$.

Statistics for extremal sequences

Let $\{Z_n\ ,\ n \ge 1\}$ be a extremal sequence, as described before, with mean values $\mu_i = M(Z_i) = \gamma + \log i$ and covariance function $\sigma_{ij} = (Z_i\ ,\ Z_j) = R\left(\frac{\min(i\ ,\ j)}{\max(i\ ,\ j)}\right.$ with $R(u) = -\int_0^u \frac{\log t}{1 - t}\ d\,t$ $(R(0) = 0\ ,\ R'(u) > 0, R(1) = \pi^2/6$, and consider the sequence $\{X_n\ ,\ n \ge 1\}$, with $X_n = \lambda + \delta\ Z_n$, λ real and $\delta > 0$; as said before we can have a subsequence $\{X_{np}\}$ of $\{X_n\}$ with p (>0) an integer, obtained from the original by periodic sampling, with the convenient change of parameters (the location one, in this case). We know then that $M(X_i) = \lambda + \delta\, \mu_i$ and $C(X_i\ ,\ X_j) = \delta^2\, \sigma_{ij} = \delta^2\, R\left(\frac{\min(i\ ,\ j)}{\max(i\ ,\ j)}\right)$.

As $X_i = \lambda + \delta\, Z_i = \lambda + \delta\, \mu_i + \varepsilon_i$ with $\varepsilon_i = \delta(Z_i - \mu_i)$ where $M(\varepsilon_i) = 0$ and $C(\varepsilon_i, \varepsilon_j) = \delta^2\, \sigma_{ij} = \delta^2\, R\left(\frac{\min(i\ ,\ j)}{\max(i\ ,\ j)}\right)$, we know that the least-squares (unbiased) estimators of $(\lambda\ ,\ \delta)$ are given by

$$\begin{bmatrix} \lambda^* \\ \delta^* \end{bmatrix} = (A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1} [X]$$

where $A^T = \begin{bmatrix} 1 & \dots & 1 \\ \mu_1 & \dots & \mu_n \end{bmatrix}$, $\sum = \delta^2 [\sigma_{ij}]$ (variance-covariance matrix), $[X]^T = [X_1, \dots, X_n]$ (the transposed sample vector) and $\det (A^T \sum^{-1} A)$ assumed to be $\neq 0$; see Silvey (1975). The least squares estimator of the quantile $\varphi = \lambda + \chi \delta$ is, as known, $\varphi^* = \lambda^* + \chi \delta^*$, which is unbiased ($M(\varphi^*) = \varphi$) and has the variance $V(\varphi^*) = \delta^2 [1 \; \chi] (A^T \sum^{-1} A)^{-1} \begin{bmatrix} 1 \\ \chi \end{bmatrix} = V(\lambda^*) + 2\chi$ $C(\lambda^*, \delta^*) + \chi^2 \; V(\delta^*)$ with $(A^T \sum^{-1} A)^{-1})^{-1} = \begin{bmatrix} V(\lambda^*) & C(\lambda^*, \delta^*) \\ C(\lambda^*, \delta^*) & V(\delta^*) \end{bmatrix}$.

From the form of $\sum = [R(\min(i,j)/\max(i,j))]$, the least squares estimators (λ^*, δ^*) do not seem very easy to manage.

Another pair of (unbiased) estimators is given by minimizing

$$\sum_1^n \epsilon_i^2 = [X - \lambda - \delta \mu]^T [X - \lambda - \delta \mu] \qquad \text{as}$$

$$\begin{bmatrix} \lambda^{**} \\ \delta^{**} \end{bmatrix} = (A^T A)^{-1} A^T [X] = \begin{bmatrix} n & \sum_i \mu_i \\ \sum_i \mu_i & \sum_i \mu_i^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & \dots & 1 \\ \mu_1 & \dots & \mu_n \end{bmatrix} \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix}$$

$$= \begin{bmatrix} \mu & \sum_i \mu_i \\ \sum_i \mu_i & \sum_i \mu_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i X_i \\ \sum_i \mu_i X_i \end{bmatrix}$$

$$= \frac{1}{n \sum_i \mu_i^2 - (\sum_i \mu_i)^2} \begin{bmatrix} \sum_i \mu_i^2 \cdot \sum_i X_i - \sum_i \mu_i \cdot \sum_i \mu_i X_i \\ - \sum_i \mu_i \cdot \sum_i X_i + n \sum_i \mu_i X_i \end{bmatrix}.$$

The estimator of φ is, then, $\varphi^{**} = \lambda^{**} + \chi\,\delta^{**}$ and

$$V(\varphi^{**}) = [1\ \chi]\,(A^T A)^{-1}\,(A^T \Sigma\, A)\,(A^T\ A)^{-1} \begin{bmatrix} 1 \\ \chi \end{bmatrix} .$$

The efficiency of φ^{**} with respect to φ^{*} is given by

$$\frac{V(\lambda^*) + 2\,\chi\ C(\lambda^*, \delta^*) + \chi^2\ V(\delta^*)}{V(\lambda^{**}) + 2\,\chi\ C(\lambda^*, \delta^*) + \chi^2\ C(\lambda^*, \delta^*)}$$

where

$$\begin{bmatrix} V(\lambda^{**}) & C(\lambda^{**}, \delta^{**}) \\ C(\lambda^{**}, \delta^{**}) & V(\delta^{**}) \end{bmatrix} = (A^T A)^{-1}\,(A^T \Sigma\, A)\,(A^T\ A)^{-1} .$$

The general efficiency of the system $(\lambda^{**}, \delta^{**})$ with respect to (λ^*, δ^*) can be defined in two ways:

a) as the value of

$$\frac{\det(A^T \Sigma^{-1} A)^{-1}}{\det\{(A^T A)^{-1}\,(A^T \Sigma\, A)\,(A^T\ A)^{-1}\}} = \frac{\det^2(A^T\ A)}{\det\,(A^T\ \Sigma\, A)\,\det(A^T\ \Sigma^{-1} A)} ;$$

see Cramér (1946);

b) as the smallest root of

$$\det\left(\begin{bmatrix} V(\lambda^*) & C(\lambda^*, \delta^*) \\ C(\lambda^*, \delta^*) & V(\delta^*) \end{bmatrix} - \xi \begin{bmatrix} V(\lambda^{**}) & C(\lambda^{**}, \delta^{**}) \\ C(\lambda^{**}, \delta^{**}) & V(\delta^{**}) \end{bmatrix} \right) = 0 ;$$

see Tiago de Oliveira (1972) and (1982).

Both definitions impose a lot of numerical calculation, beginning with the inversion of $\sum = [R(\min(i,j)/\max(i,j)]$, and do not seem easy to manipulate. If $n = 2$ we get for both estimators.

$$\begin{bmatrix} \lambda^* \\ \delta^* \end{bmatrix} = \begin{bmatrix} \lambda^{**} \\ \delta^{**} \end{bmatrix} = \frac{1}{\log 2} \begin{bmatrix} (\gamma + \log 2)\, X_1 - \gamma X_2 \\ X_2 - X_1 \end{bmatrix}$$

and both efficiencies are equal to 1.

In general we use the estimators:

$$\lambda^* = \frac{(\gamma \sum_i \log i + \sum_i (\log i)^2) \cdot \sum_i X_i - (n\gamma + \sum_i \log i) \cdot \sum_i (\log i)\, X_i}{n \sum_i (\log i)^2 - (\sum_i \log i)^2}$$

$$\delta^* = \frac{n \sum_i (\log i)\ X_i - \sum_i \log i \cdot \sum_i X_i}{n \sum_i (\log i)^2 - \sum_i (\log i)^2} .$$

whose variance-covariance matrix was given above.

Statistics for extremal processes

Consider now the extremal process $Z(t)$ $(t \geq 0)$ (Tiago de Oliveira, 1968) whose mean value function is $\mu(t) = \gamma + \log t$, covariance function $\sigma(s,t) = C(Z(s), Z(t)) = R(\min(s,t)/\max(s,t))$ where, as before, $R(u) = -\int_0^u \frac{\log t}{1-t}\, dt$, $R(0) = 0$, $R'(0) > 0$, $R(1) = \pi^2/6$; the correlation function is $\rho(s,t) = \frac{6}{\pi^2} R(s,t)$.

Let us consider the extremal process $X(t) = \lambda + \delta Z(t)$, $t \geq 0$. It is well known (see above) that there is an essential *un*identifiability problem between the use of $X(t)$, $X_1(t) = \lambda_1 + \delta Z(\beta_1 t)$ and $X_2(t) = \delta Z(\beta_2 t)$ with the equalities $\lambda = \lambda_1 + \delta \log \beta_1 = \delta \log \beta_2$. We have chosen the

present formulation because then the parameters are linear. We will suppose that $X(t)$ is known (observed) for the interval $[a , b]$ or $]a , b[$, which is irrelevant, with $0 < a < b$.

We could approach the problem by an analogous procedure to the least squares, substituting the sums by integrals, i.e., as $X(t) = \lambda + \delta Z(t) = \lambda + \delta (\gamma + \log t) + \varepsilon (t)$ ($\varepsilon (t) = \delta (Z(t) - \gamma - \log t)$, with mean value zero), by minimizing $\int_a^b \int (X(t) - \lambda - \delta(\gamma + \log s))\ \bar{\sigma}(s , t)$ $(X(t) - \lambda - \delta(\gamma + \log t))\ d\, s\, d\, t$ where $\bar{\sigma} (s , t)$ is the "inverse" of the covariance function $\sigma(s , t)$ analogous to the relation between $\sum^{-1}$ and $\sum$. The solution is

$$\begin{bmatrix} \lambda^* \\ \delta^* \end{bmatrix} = \frac{1}{\bar{\sigma}_0\ \bar{\sigma}_2 - \bar{\sigma}_1^2} \begin{bmatrix} \int_a^b \int (\bar{\sigma}_2 - \bar{\sigma}_1 \mu(t))\ \bar{\sigma}(s , t)\ X(s)\ d\, s\ d\, t \\ \int_a^b \int (\bar{\sigma}_0 \mu(t) - \bar{\sigma}_1)\ \bar{\sigma}(s , t)\ X(s)\ d\, s\ d\, t \end{bmatrix}$$

where $\bar{\sigma}_0 = \int_a^b \int \bar{\sigma} (s , t)\, d\, s\ d\, t$, $\bar{\sigma}_1 = \int_a^b \int \bar{\sigma} (s , t)\, \mu(s)\, d\, s\ d\, t$ and $\bar{\sigma}_2 = \int_a^b \int \bar{\sigma} (s , t)\, \mu(s)\, \mu(t)\, d\, s\ d\, t$; (λ^*, δ^*) are unbiased estimators of (λ, δ). The same result can be obtained be seeking the function $f(t)$ for which $\int_a^b f(t)\ X(t)\ d\, t$ is the least-squares (unbiased) estimator of $\varphi = \lambda + \chi\, \delta$, which can be shown, by a variational method, to be $\varphi^* = \lambda^* + \chi\, \delta^*$. But the result is unhelpful because the determination of $\bar{\sigma} (s , t)$ is practically impossible.

As before, for extremal sequences, it is natural to seek the estimators $(\lambda^{**}, \delta^{**})$ such that the (empirical) mean-square distance

$$\int_a^b (X(t) - \lambda - \delta\, \mu (t))^2\ (d\, t) = \int_a^b (X(t) - \lambda - \delta(\gamma + \log t))^2\ d\, t$$

is a minimum. We get the also unbiased estimators

$$\begin{bmatrix} \lambda^{**} \\ \delta^{**} \end{bmatrix} = \frac{1}{\int_a^b dt \cdot \int_a^b \mu^2(t)\,dt - \left(\int_a^b \mu(t)\,dt\right)^2}$$

$$\begin{bmatrix} \int_a^b \mu^2(t)\,dt \cdot \int_a^b X(t)\,dt - \int_a^b \mu(t)\,dt \cdot \int_a^b \mu(t)\,X(t)\,dt \\ \int_a^b dt \cdot \int_a^b \mu(t)\,X(t)\,dt - \int_a^b \mu(t)\,dt \cdot \int_a^b X(t)\,dt \end{bmatrix}$$

Evidently we could choose two functions $l(t)$ and $d(t)$ and take $\left(\int_a^b l(t)\,X(t)\,dt, \int_a^b d(t)\,X(t)\,dt\right)$ as estimators of (λ, δ) with the condition of quasi-linearity or of unbiasedness; $(\lambda^{**}, \delta^{**})$, is one of these choices, optimization being practically impossible.

The complex structure of these estimators suggests the use of the ergodic theorem, when possible. As shown before,

$$\frac{1}{\log(b/a)} \int_a^b \frac{Z(t)}{t}\,dt - \log\sqrt{ab} \xrightarrow{ms} \gamma \text{ as } b/a \to +\infty$$

or, under a more general form,

$$\frac{1}{\log(b/a)} \int_a^b \frac{X(t)}{t}\,dt - \delta \log\sqrt{ab} \xrightarrow{ms} \lambda + \gamma\delta .$$

If we take $\log\sqrt{ab} = \chi - \gamma$ with $b/a \to +\infty$ or $a = e^{2(\chi-\gamma)}/b$, $b \to +\infty$ we see that

$$\frac{1}{2(\log b - (\chi - \gamma))} \int_{e^{2(\chi-\gamma)}/b}^b \frac{X(t)}{t}\,dt \xrightarrow{ms} \lambda + \chi\delta \quad \text{when} \quad b \to +\infty.$$

As, then, $b/a = b^2/e^{2(\chi-\gamma)}$ increases with b, if we know $X(t)$ in an interval $[A, B]$ we can estimate $\varphi = \lambda + \chi\delta$ taking $b = B$ and $a = e^{2(\chi-\gamma)}/B$ if $a > A$. Thus the estimation is obtained by taking $\varphi^* = \frac{1}{2(\log B - (\chi - \gamma))} \int_{e^{2(\chi-\gamma)}/b}^b \frac{X(t)}{t}\,dt$ when $AB \le e^{2(\chi-\gamma)}$ or $\chi \ge \gamma + \log\sqrt{AB}$, thus allowing, the estimators of large quantiles, which is important in the applications.

The variance of this estimator of $\lambda + \chi\,\delta$ is given by

$$\frac{V(\varphi^*)}{\delta^2} = \frac{1}{4\,(\log B - (\chi - \gamma))^2} \int_{e^{2(\chi-\gamma)/B}}^{B}\int \frac{\sigma(s,t)}{s\,t}\,ds\,dt$$

$$= \frac{1}{2\,(\log B - \chi + \gamma)^2} \int_{e^{2(\chi-\gamma)/B \le s \le t \le B}}\int R(s/t)\,\frac{ds\,dt}{s\,t}$$

$$= \frac{1}{2\,(\log B - \chi + \gamma)^2} \int_{e^{2(\chi-\gamma)/B^2}}^{1} \frac{R(u)}{u}\,du \int_{e^{2(\chi-\gamma)/Bu}}^{B} \frac{dt}{t}$$

$$= \frac{1}{2\,(\log B - \chi + \gamma)^2} \int_{e^{2(\chi-\gamma)/B^2}}^{1} \frac{R(u)}{u}\,du\,\log\frac{B^2\,u}{e^{2(\chi-\gamma)}}$$

$$= \frac{1}{2\,(\log B - \chi + \gamma)^2} \int_{e^{2(\chi-\gamma)/B^2}}^{1} \frac{R(u)}{u}\,(2(\log B - \chi + \gamma) + \log u)\,du$$

$$= \frac{1}{2\,(\log B - \chi + \gamma)^2} \{R(u)\,[2(\log B - \chi + \gamma)\log u) + \frac{\log^2 u}{2}\,]\,|_{e^{2(\chi-\gamma)/B^2}}^{1} -$$

$$- \int_{e^{2(\chi-\gamma)/B^2}}^{1} (2(\log B - \chi + \gamma)\,\log u + \frac{\log^2 u}{2})\,R'(u)\,du\}$$

$$= \frac{1}{2\,(\log B - \chi + \gamma)^2} \{(2(\log B - \chi + \gamma)^2\,R\,(\frac{e^{2(\chi-\gamma)}}{B^2})$$

$$- \int_{e^{2(\chi-\gamma)/B^2}}^{1} (2(\log B - \chi + \gamma)\log u) + \frac{\log^2 u}{2})\,R'(u)\,du\} = R(e^{2(\chi-\gamma)/B^2})$$

$$- \frac{1}{2\,(\log B - \chi + \gamma)^2} \int_{e^{2(\chi-\gamma)/B^2}}^{1} (2(\log B - \chi + \gamma)\log u + \frac{\log^2 u}{2})\,\frac{\log u}{1-u}\,du$$

$$= R(e^{2(X-Y)}/B^2) - \frac{1}{2(\log B - X - Y)^2} \int_{e^{2(X-Y)}/B^2}^{1} (2(\log B - X + Y)\log u + \frac{\log^2 u}{2}) \frac{\log u}{1-u}$$

$$= R(e^{2(X-Y)}/B^2 - \int_{e^{2(X-Y)}/B^2}^{1} \frac{\log^2 u}{1-u}\, du - \frac{1}{4(\log B - X + Y)^2} \int_{e^{2(X-Y)}/B^2}^{1} \frac{\log^3 u}{1-u}\, du$$

Analogously we could chose a point C $(0 < A < C < B)$ and study the stochastic integrals

$$\frac{1}{\log C/A} \int_A \frac{X(t)}{t}\, dt \quad \text{and} \quad \frac{1}{\log B/C} \int_C^B \frac{X(t)}{t}\, dt$$

To have $\min(\log \frac{C}{A}, \log \frac{B}{C})$ the largest possible we must take $C = \sqrt{AB}$ with $\frac{C}{A} = \sqrt{B/A} = \frac{B}{C}$. The convex combination of those two integrals with coefficients $\frac{\log \sqrt{CB} - (X-Y)}{\log \sqrt{B/A}}$ and $\frac{(X-Y) - \log \sqrt{AC}}{\log \sqrt{B/A}}$ estimates, unbiasedly, and mean square, the quantile $\lambda + X\delta$ and consequently, for $X = Y + \log t$, $\lambda^* + (Y + \log t)\delta^*$ is the natural linear predictor of $X(t)$.

An important question in the integrals above is the assumed existence of jumps in the interval of observation $[A,B]$. Suppose that $X(t)$ has no jumps in $[A,B]$ a fact that has probability A/B. Noting by X the value of $X(t)$ in the interval the estimator

$$(\int_A^B \ell(t)\, X(t)dt, \int_A^B d(t)\, X(t)dt) = (X \int_A^B \ell(t)dt, X \int_A^B d(t)dt) = (X,0) \quad (!)$$

as by the quasilinearity we must have $\int_A^B \ell(t)dt = 1$ and $\int_A^B d(t)dt = 0$.

The probability of being capable to estimate is, then, $1 - A/B$.

In that case is easy to show that $\frac{X(B) - X(a)}{\log B/A} \xrightarrow{ms} \delta$ if $B/A \to +\infty$ but an analogous estimator of λ based in the extreme values $X(A)$ and $X(B)$ does not converge in m.s. to λ.

The study of the jump process does not seem to iluminate the structure of the location parameter (λ) or of the change of time unit (β). In fact, the extremal process $Z(t)$ - and the processes $\lambda+\delta Z(t)$ or $\delta Z(\beta t)$ - have a random number of jumps $N(t_1,t_2)$ in the interval $[t_1,t_2]$ $(0 < t_1)$ with a Poisson distributions with parameter (= mean value = variance) $\log \frac{t_2}{t_1}$; in $[0,t]$ there is an infinity of jumps with probability one. If t_1 is a time instant (be it a jump instant or not) the probability that the next jump instant is $\leq t_2$ is given by $1 - t_1/t_2 (t_1 < t_2)$. In any case a change of time unit does not alter the structure.

The mean value of inter-jump times is

$$M(t_{i+1}-t_2) = \int_{t_i}^{+\infty} (t_{i+1}-t_1)d(1 - t_i(t_{i+1}) = +\infty$$

which shows that the averages of inter-jump times is not useful. Also the median of t_{i+1} is $2\,t_i$, independent of the time unit.

As $M(t_{i+1}-t_i)^\alpha = t_i^\alpha \int_0^{+\infty} \xi^\alpha \, d(\frac{\xi}{1+\xi}) = t_i^\alpha \frac{\alpha\,\pi}{\sin \pi}$ if $\alpha < 1$ and nothing is gained because the effect of a change of time unit is cancelled; variance exists if $\alpha < 1/2$. Also we have

$$\text{Prob}\{\frac{t_{i+1}-t_i}{t_i} \leq \xi\} = \frac{\xi}{\xi+1} .$$

For other details see Deheuvels (1981, 1983) and references therein.

Chapter XV

OTHER PROCESSES OF EXTREMES; BASICS AND STATISTICS

Extreme-Markovian-stationary (EMS) sequences

Let us introduce the extreme-Markovian-stationary (EMS) sequences analogous to normal-Markovian-stationary ones and study some of their properties. In this section and in the following one will follow Tiago de Oliveira (1972').

Let $\{E_k , k = 1,2,...\}$ be a sequence of independent reduced Gumbel random variables and consider the sequence

$$Z_1 = E_1$$

$$Z_k = \max (Z_{k-1} + a , E_k + b) , \quad k \geq 2 .$$

We see immediately that *all* Z_k *have distribution function* $\Lambda(x)$ *iff we have* $e^a + e^b = 1$. Putting $e^a = \theta$ $(0 \leq \theta \leq 1)$ we get the EMS sequence defined as

$$Z_1 = E_1$$

$$Z_k = \max (Z_{k-1} + \log \theta , E_k + \log (1 - \theta)) , k \geq 2 \ (0 \leq \theta \leq 1).$$

The distribution function of the pair (Z_k , Z_{k+1}) $(k \geq 1)$ is Prob $\{Z_k \leq x , Z_{k+1} \leq y\}$ = Prob $\{Z_k \leq x , Z_k \leq y - \log \theta , E_{k+1} \leq y - \log(1 - \theta)\} = \Lambda(\min (x , y - \log \theta)) \Lambda(y - \log(1 - \theta))$ $= \Lambda(x , y \mid \theta) = \exp \{- (e^{-x} + e^{-y}) + \min (e^{-x} , \theta\ e^{-y}))$ where the dependence function is $k(w \mid \theta) = 1 - \min (\theta , e^w) / (1 + e^w)$.

The transition distribution function Prob $\{Z_k \leq y \mid Z_{k-1} = x\}$ is given as

$P(y \mid x; \theta)$ = Prob $\{\max(X + \log \theta , E_k + \log(1 - \theta)\} \leq y) = 0$ if $y < x + \log \theta$ and $P(y \mid x; \theta) =$ $\exp \{-(1 - \theta) e^{-y}\}$ if $y \geq x + \log \theta$, that is $P (y \mid x ; \theta) = H(y - x - \log \theta)$. $\exp - \{(1 - \theta) e^{-y}\}$, H being the Heaviside jump function. The transition density, using the Dirac δ pseudo-function, is

$$[(1 - \theta) e^{-y}\ \ H(y - x - \log \theta) + \delta(y - x - \log \theta)] \exp\{-1 - \theta) e^{-y}\} .$$

The distribution function of $(Z_1, \dots, Z_n)$ is given by the iterative formula

$\Lambda_n(x_1, \dots, x_n \mid \theta) = \text{Prob}\{Z_1 \le x_1, \dots, Z_n \le x_n\} = \text{Prob}\{Z_1 \le x_1, \dots, Z_{n-1} \le x_{n-1}, Z_{n-1} \le x_n - \log\theta, E_n \le x_n - \log(1-\theta)\} = \Lambda_{n-1}(x_1, \dots, x_{n-2}, \min(x_{n-1}, x_n - \log\theta) \mid \theta)\, \Lambda(x_n - \log(1-\theta))$.

The distribution function of (Z_p, Z_q) $(p < q)$ is given by

$$\Lambda_{p,q}(x, y \mid \theta) = \text{Prob}\{Z_p \le x, Z_q \le y\}$$

$$= \Lambda_{p,q-1}(x, y - \log\theta \mid \theta) \cdot \Lambda(y - \log(1-\theta)) \quad \text{if} \quad p < q-1,$$

$$= \Lambda(\min(x, y - \log\theta))\, \Lambda(y - \log(1-\theta)) = \Lambda(x, y \mid \theta) \quad \text{if} \quad p = q-1.$$

The iteration shows that the distribution function of (Z_p, Z_q) is $\Lambda(x, y \mid \theta^{q-p})$, so that the correlation coefficient between Z_p and Z_q is $\frac{6}{\pi^2} R(\theta^{q-p})$.

The expression of Z_q in Z_p $(q > p)$ is

$Z_q = \max\left[Z_p + (q-p)\log\theta, \max_{p+1}^{q}(E_j + (q-j)\log\theta) + \log(1-\theta)\right]$. Thus:

The EMS sequence Z_n is stationary with the distribution function given by the iteration

$$\Lambda(x \mid \theta) = \Lambda(x), \quad \Lambda_n(x_1, \dots, x_n \mid \theta) =$$

$$\Lambda_{n-1}(x_1, \dots, x_{n-2}, \min(x_{n-1}, x_n - \log\theta) \mid \theta) \cdot \Lambda(x_n - \log(1-\theta)).$$

Its mean value is γ, the variance $\pi^2/6$ and the correlation coefficient $6/\pi^2 \cdot R(\theta)$; the correlation coefficient between Z_p and Z_q $(q > p)$ is $\frac{6}{\pi^2} R(\theta^{q-p})$.

The distribution function of $\max(Z_1, \dots, Z_n)$ being

$$\text{Prob}\{\max(Z_1, \dots, Z_n) \le x\} = \exp\{-(\theta + n(1-\theta))\, e^{-x}\},$$

we see that

$$\frac{\max(Z_1, \dots, Z_n)}{\log(\theta + n(1-\theta))} \xrightarrow{P} 1 \quad \text{as} \quad n \to \infty.$$

Also $\frac{\max(Z_1, \dots, Z_n)}{\log n} \xrightarrow{P} 1$ and the mean value and variance of $\max(Z_1, \dots, Z_n)$ are $\gamma + \log(\theta + n(1-\theta))$ and $\pi^2/6$.

Let us now obtain some results which may be useful for the description and analysis of an EMS sequence.

Let us first show that:

The ergodic theorem in mean-square for the EMS sequence is valid.

As $M(\frac{1}{n}\sum_{1}^{n} Z_k) = \gamma$ we have only to show that $V(\frac{1}{n}\sum_{1}^{n} Z_k) \to 0$. But

$$V(\frac{1}{n}\sum_{1}^{n} Z_k) = \frac{\pi^2/6}{n^2}\sum_{i,j}^{n} \rho_{ij} = \frac{\pi^2/6}{n^2}(n + 2\sum_{1}^{n}(n-k)\rho_k)$$

where the correlation coefficient ρ_k is $\rho_{i,\,i+k} = 6/\pi^2 \cdot C(Z_i, Z_{i+k})$ by the stationarity. As $\rho_k = \frac{6}{\pi^2} R(\theta^k)$ we have

$$\frac{1}{n}\sum_{1}^{n}(1 - \frac{k}{n})\, R(\theta^k) \to 0,$$

which shows the desired result.

The results that follow suggests the trend of the next section concerning statistical decision.

As the EMS sequence is stationary the correlogram technique can be applied; from $\theta^k \downarrow 0$ we can expect the empirical correlogram to tend quickly to zero.

By contrast, *the periodogram technique is not useful as could be expected.* Let T be the trial period. The quantities $A_n(T) = \frac{1}{n}\sum_{1}^{n} Z_k \sin\frac{2\pi k}{T}$ and $B_n(T) = \frac{1}{n}\sum_{1}^{n} Z_k \cos\frac{2\pi k}{T}$ converge in mean square to zero , as their mean values converge to zero and $V(A_n(T))$ and $V(B_n(T))$ converge to zero. A simple way is to show that

$$V(A_n(T)) + V(B_n(T)) = \frac{\pi^2/6}{n^2}\sum_{i,j} \rho_{ij} \cos\frac{2\pi(i-j)}{T} =$$

$$\frac{\pi^2/6}{n^2}(n + 2\sum_{1}^{n}\rho_k (n-k) \cos\frac{2\pi k}{T} \to 0,$$

which is evident because $\rho_k = 6/\pi^2 R(\theta^k) \to 0$ (as follows from the ergodic theorem).

We can also consider the associated sequence of ups and downs. Denoting by K_n the number of times that $Z_{k-1} \leq Z_k$ for $(Z_1, Z_2, \ldots, Z_n)$ we have

$$K_n = \sum_2^n H(Z_k - Z_{k-1}) .$$

As $Z_n = \max(Z_{k-1} + \log\theta , E_k + \log(1-\theta))$, we have $H(Z_k - Z_{k-1}) = H(E_k + \log(1-\theta) - Z_{k-1})$. Thus $M(\frac{K_n}{n-1}) = \frac{1-\theta}{2-\theta}$ and $V(\frac{K_n}{n-1}) = \frac{1}{(n-1)^2}\ [(n-1)\,\sigma^2 + 2\sum_1^n (1 - \frac{k}{n-1})\ \rho_k^{'}]$ with $\sigma^2 = V(H(Z_k - Z_{k-1})) = \frac{1-\theta}{(2-\theta)^2}$ and $\rho_k^{'} = -\frac{(1-\theta)^2\,\theta^{k-1}}{1 + (1-\theta)\,(1+\theta^{k-1})}$ as

$$C(H(Z_2 - Z_1),\ H(Z_{k+2} - Z_{k-1})) = \text{Prob}\,(Z_1 \leq Z_2 ,\ Z_{k+1} \leq Z_{k+2}) - (\frac{1-\theta}{2-\theta})^2$$

$$= -\ \frac{(1-\theta)^3\,\theta^{k-1}}{(2-\theta)^2\,(1 + (1-\theta)(1+\theta^{k-1}))} .$$

As $\rho_k^{'} \to 0$ when $k \to \infty$ *we see that* $\frac{K_n}{n-1} \overset{ms}{\to} \frac{1-\theta}{2-\theta}$.

We can generalize, by defining another type of stochastic sequences of extremes analogous to linear processes in gaussian processes theory.

Let $\{E_n , -\infty < n < +\infty\}$ be a sequence of independent reduced Gumbel random variables and define

$$Z_k = \max_{0 \leq j < +\infty} (E_{k-j} + a_j) .$$

Then $\text{Prob}\,\{Z_k \leq x\} = \text{Prob}\,\{E_{k-j} \leq x - a_j\} = \pi_0^{\infty} \Lambda(x - a_j) = \exp\{-e^{-x} \sum_0^{\infty} e^{a_j}\}$ has Gumbel reduced margins iff $\sum_0^{\infty} e^{a_j} = 1$.

Taking then any set of probabilities $p_j \geq 0\ (\sum_0^{\infty} p_j = 1)$, we must have

$$Z_k = \max_{0 \leq j < \infty} (E_{k-j} + \log p_j) ,$$

the EMS sequence being obtained by taking $p_j = (1-\theta)\theta^j$.

For a sequence $\{E_n, 0 \le n < +\infty\}$ of independent reduced Gumbel random variables we can also define an extreme sequence by choosing $\beta_n(j) \ge 0$ such that $\sum_1^n \beta_n(j) = 1$ and putting

$$Z_n = \max_{1 \le k < n} (E_k + \log \beta_n(k)).$$

Statistical decision for EMS sequences

Let us return to EMS sequences with reduced margins. From the relation $Z_k - Z_{k-1} \ge \log \theta$ $(k \ge 2)$ we get the maximum likelihood estimator

$$\hat{\theta}_n = \exp\left(\min_{2 \le k \le n} (Z_k - Z_{k-1})\right).$$

We can easily obtain the distribution of $\hat{\theta}_n$ and show that $\text{Prob}\{\hat{\theta}_n = \theta\} \to 1$. Note that $\hat{\theta}_n \ge \theta$, but we can have $\hat{\theta}_n > 1$, thus suggesting truncation if this is the case.

Then $\text{Prob}\{\hat{\theta}_n \le a\} = 0$ for $a < \theta$. Let us then obtain $\text{Prob}\{\hat{\theta}_n \le a\}$ for $a \ge \theta$, if $\theta < 1$ by computing $Q_n(a) = \text{Prob}\{\hat{\theta}_n > a\}$ for $a \ge \theta$.

As $Z_k - Z_{k-1} > \log a \ (\ge \log \theta)$ we see that

$$Q_n(a) = \text{Prob}\{Z_k - Z_{k-1} > \log a\} =$$

$$\text{Prob}\left\{E_2 > E_1 + \log \frac{a}{1-\theta}, \ E_3 > E_2 + \log a, \ldots, E_n > E_{n-1} + \log a\right\} =$$

$$= \text{Prob}\{t_1 > a\, q_1(\theta)\, t_2, \ t_2 > a\, t_3, \ldots, t_{n-1} > a\, t_n\},$$

with $t_i = e^{-E_i}$ standard exponential and $q_1(\theta) = (1-\theta)^{-1}$, which, iteratively, gives

$$Q_n(a) = \left(\prod_2^n q_k(a)\right)^{-1} \quad \text{where} \quad q_k(a) = 1 + a\, q_{k-1}(a),$$

so that, as $q_1(a) = (1-\theta)^{-1}$, we have $q_k(a) = \dfrac{a^k}{\theta(1-\theta)} + \dfrac{1-a^{k-1}}{1-a}$, $q_k(\theta) = (1-\theta)^{-1}$,

$$Q_n(\theta) = (1-\theta)^{n-1}$$

and thus $\text{Prob}(\hat{\theta}_n = \theta) = 1 - (1-\theta)^{n-1} \to 1$ if $\theta > 0$ and $\text{Prob}(\hat{\theta}_n = 1) = 1$ if $\theta = 1$. This estimator has for $a = \theta$

$$\frac{d\ \text{Prob}(\hat{\theta}_n \le a)}{d\ a} \sim n\ (1-\theta)^{n-2} \quad \text{if} \quad 0 \le \theta < 1$$

$$= 0 \quad \text{if} \quad \theta = 1;$$

asymptotically $\hat{\theta}_n$ is better than the common maximum likelihood estimators whose order is, usually, $n^{1/2}$.

In the more general case, we have an EMS sequence $X_n = \lambda + \delta Z_n$ with general margins, where λ and $\delta(>0)$ are the location and dispersion parameters. We will show, using the results given in the paper, how to obtain quick estimators for λ, δ and θ.

We have $\bar{X}_n = \lambda + \delta \bar{Z}_n \xrightarrow{P} \lambda + \gamma\delta$ by the ergodic results,

$$F_n = \frac{K_n}{n-1} \xrightarrow{P} \frac{1-\theta}{2-\theta}$$

by the ups and downs sequence,

$$\Delta_n = \min(X_k - X_{k-1}) = \delta\ \min(Z_k - Z_{k-1}) \xrightarrow{P} \delta\ \log\theta,$$

and also

$$\frac{\max(X_1, \ldots, X_n)}{\log n} = \frac{\lambda + \delta\ \max(Z_1, \ldots, Z_n)}{\log n} \xrightarrow{P} \delta.$$

As, for $0 \le \theta \le 1$, we have $0 \le \dfrac{1-\theta}{2-\theta} \le \dfrac{1}{2}$, decreasing with θ, we will take

$$F_n^* = F_n \quad \text{if} \quad 0 \le F_n \le 1/2$$

$$F_n^* = 1/2 \quad \text{if} \quad 1/2 \le F_n\ (\le 1),$$

and use

$$F_n^* = \frac{1 - \theta_n^*}{2 - \theta_n^*} \quad \text{or} \quad \theta_n^* = 1 - \frac{F_n^*}{1 - F_n^*} .$$

Note that $F_n = 0$ ($F_n^* = 0$) gives $\theta_n^* = 1$ (diagonal case) and $F_n \geq 1/2$ ($F_n^* = 1/2$) gives $\theta_n^* = 0$ (independence case).

Once θ_n^* is known we must estimate λ and δ. A natural choice is the ergodic theorem thus giving one equation

$$\bar{X}_n = \lambda_n^* + \gamma \, \delta_n^* .$$

For a second equation, we could use either

$$\min (X_k - X_{k-1}) = \delta_n^* \log \theta_n^*$$

(as $\min (X_k - X_{k-1}) = \delta \min (Z_k - Z_{k-1}) \xrightarrow{P} \delta \log \theta$)

or

$$\max (X_1 , \ldots , X_n) = \lambda_n^* + (\gamma + \log (\theta_n^* + n(1 - \theta_n^*)) \delta_n^* .$$

As the use of the first relation, for δ_n^*, imposes one more condition (i.e., $\min(X_k - X_{k-1}) < 0$), we will use the second relation, with the ergodic theorem, to estimate λ_n^* and δ_n^*.

Thus we have the system

$$\theta_n^* = 1 - \frac{F_n^*}{1 - F_n^*}$$

$$\delta_n^* = \frac{\max_1^n (X_i) - \bar{X}_n}{\log (\theta_n^* + n(1 - \theta_n^*))} .$$

and $\lambda^* = \bar{X}_n - \gamma\, \delta_n^*$

to estimate the parameters. Note that the denominator of δ_n^* is always ≥ 0 for $n > 1$ and that $[\log(\theta_n^* + n(1 - \theta_n^*))] / \log n \xrightarrow{P} 1$.

If we substitute the denominator of δ_n^* by $\log n$ we thus have the estimator $\delta_n^{**} = (\max_1^n(X_i) - \bar{X}_n)/\log n$ (independent of θ_n^*) and also the new estimator $\lambda_n^{**} = \bar{X}_n - \gamma\, \delta_n^{**}$. We will use these estimators.

As an example we will apply these simple estimators to two random sequences $\{E_j\}$ and $\{Z_j\}$ of 25 terms where $\{E_j\}$ is a sequence of independent reduced Gumbel random variables obtained from Goldstein (1963), and $\{Z_j\}$ is an EMS sequence with $\theta = 1/2$, i.e.,

$$Z_k = \max(Z_{k-1}, E_k) - \log 2 :$$

j	1	2	3	4	5	6	7
E_j	1.412	–0.296	–0.031	1.388	1.657	–0.382	–0.175
Z_j	1.412	0.719	0.026	0.695	0.964	0.271	–0.422

j	8	9	10	11	12	13	14
E_j	–0.380	–0.692	2.135	0.130	0.040	1.927	1.456
Z_j	–1.073	–1.385	1.442	0.749	0.056	1.234	0.763

j	15	16	17	18	19	20	21
E_j	0.472	1.140	–0.930	–0.793	–0.913	–0.610	0.616
Z_j	0.070	0.447	–0.246	–0.939	–1.606	–1.303	–0.077

j	22	23	24	25
E_j	–0.920	1.434	1.586	3.688
Z_j	–0.770	0.741	0.893	2.995

Note that in the two cases we have $\lambda = 0$, $\delta = 1$. Assuming this, for the $\{E_j\}$ sequence we have $K_{25} = 13$ so $F_{25} = 13/24$ ($\geq 1/2$) and so $\theta^*_{25} = 0$ (independence); for the $\{Z_j\}$ sequence we have $K_{25} = 10$ and so $F_{25} = 10/24$, $F^*_{25} = F_{25}$ and $\theta^*_{25} = 2/7 = .286$ which is a long way from $\theta = 1/2$!

In the general case we can add the estimators of λ and δ.

For the $\{E_j\}$ sequence we have $\max_1^{25} (E_j) = 3.688$, $\bar{E}_{25} = .51716$ and so

$$\delta^{**}_{25} = .985 \qquad \lambda^{**}_{25} = -.051 ,$$

which are not very far from the exact values. Clearly, assuming $\theta = 0$, we should have used habitual ML estimators.

Consider, now, the $\{Z_j\}$ sequence. We have estimated, before, θ by $\theta^*_{25} = 2/7 = .286$.

As we have $\max_1^{25} (X_j) = 2.995$ and $\bar{X}_{25} = .226$ we get $\delta^{**}_{25} = .860$ and $\lambda^{**}_{25} = -.271$.

Once more the estimates are not close to the exact values!

The estimation problem has to be reconsidered.

Extreme-Markovian stationary (EMS) processes

After the definition of extremal processes and EMS sequences, we will define the EMS processes and relate them to the extremal processes, consider the associated maximum process, and show that its asymptotic behaviour is similar to that of an extremal process. This is analogous to the relation between the Wiener-Levy process and the integrated Orstein-Uhlenbeck process, as could be expected, to a certain extent, from the "duality" between maxima and sums.

In this section we will follow Tiago de Oliveira (1973).

An extreme-Markovian-stationary (EMS) process Z(t) can be characterized by means of the following axioms:

- Z(t) is a stationary process defined for $t\ (-\infty < t < +\infty)$;
- for $s \le t$, $Z(t) = \max\ (Z(s) + \varphi(t-s),\ E(s,t) + \psi(t-s))$, Z(s) and E(s , t) being independent (*);
- the random variables E(s , t) have a reduced Gumbel distribution;
- E(s , t) and E(s' , t') are independent if $]s,t[\cap]s',t'[= \varnothing$;
- Z(0) is a reduced Gumbel random variable.

We will now deduce some basic results from the axioms.

From the stationarity we get

$$\text{Prob}\ (Z(t) \le x) = \text{Prob}\ (Z(0) \le x) = \Lambda\ (x)\ .$$

Now using the second axiom we get

$$\Lambda(x) = \text{Prob}\ (Z(t) \le x) = \Lambda(x - \varphi(t-s))\ \ \Lambda(x - \psi(t-s)) \quad \text{so that}$$

$$e^{\varphi(t-s)} + e^{\psi(t-s)} = 1\ .$$

Now taking $s \le u \le t$ we get

$$Z(t) = \max\ (Z(s) + \varphi(t-s)\ ,\ E(s\ ,\ t) + \psi(t-s))$$

$$= \max\ (Z(s) + \varphi(u-s) + \varphi(t-u)\ ,\ E(s\ ,\ u) + \psi(u-s) + \varphi(t-u))\ ,\ E(u\ ,\ t) + \psi(t-u))$$

so that

$$\varphi(t-s) = \varphi(t-u) + \varphi(u-s)$$

$$E(s\ ,\ t) + \psi(t-s) = \max(E(s\ ,\ u)\ + \psi(u-s) + \varphi(t-u)\ ,\ E(u\ ,\ t) + \psi(t-u))\ .$$

As $e^{\varphi(t-s)} \le 1$ the first relation gives, as $\varphi(w) \le 0$,

$$\varphi(w) = -\beta\ w\ (\beta \ge 0)$$

(*) – We could instead of this formulation introduce the random variables E'(s , t) – E(s , t) + ψ(t – s) with a Gumbel distribution (not reduced), adopting conveniently the axious and the proofs. This is left as an exercise.

so that

$$\psi(w) = \log (1 - e^{-\beta w})$$

and, consequently, we get for $s \le t$, $Z(t) = \max (Z(s) - \beta(t - s), E(s, t) + \log (1 - e^{-\beta (t-s)}))$.

Note that the random variables $E(s, t)$ satisfy the relation

$$E(s, t) = \max \left(E(s, u) + \log \frac{e^{-\beta s} - e^{-\beta u}}{e^{-\beta s} - e^{-\beta t}}, E(u, t) + \log \frac{e^{\beta t} - e^{\beta u}}{e^{\beta t} - e^{\beta s}}\right)$$

for $0 < s \le u \le t$. It is immediate that

$$Z(t) = Z_0 (e^{\beta t}) - \beta t,$$

where $Z_0(t')$ is a (reduced) extremal process.

The joint distribution function of $(Z(t_1), \ldots, Z(t_n))$ is easily shown to be

$$\Lambda_n(x_1, t_1; \ldots; x_n, t_n) = \text{Prob}\{Z(t) \le x_1, \ldots, Z(t_n) \le x_n\}$$

$$= \Lambda_{n-1}(x_1, t_1; \ldots; x_{n-2}, t_{n-2}; \min (x_{n-1}, x_n + \beta (t_n - t_{n-1})), t_{n-1}) \cdot$$

$$\cdot \Lambda(x_n - \log (1 - e^{-\beta (t_n - t_{n-1})}).$$

From the previous result we get

$$\Lambda_2(x_1, t_1; x_2, t_2) = \exp (-(e^{-x_1} + e^{-x_2}) - \min (e^{-x_1}, e^{-\beta (t_2 - t_1)} e^{-x_2})$$

so that the dependence function is

$$k(w) = 1 - \max (e^{-\beta (t_2 - t_1)}, e^w) / (1 + e^w),$$

a biextremal one with the parameter $\theta = e^{-\beta (t_2 - t_1)}$.

The correlation coefficient is then

$$\frac{6}{\pi^2} R (e^{-\beta (t_2 - t_1)}),$$

now being continuous in the diagonal.

Also the conditional distribution function of the EMS process is

$$\text{Prob}\,(Z(t) \le y) \mid Z(s) = x) \;= 0 \qquad \text{if} \quad y < x - \beta(t-s)$$

$$= \Lambda(y - \log\,(1 - e^{-\beta\,(t-s)})) \qquad \text{if} \quad y \ge x - \beta(t-s)\ ,$$

with a jump of

$$\Lambda(x - \log\,(^{\beta\,(t-s)} - 1)) \qquad \text{at} \qquad y = x - \beta(t-s)\ ,$$

for $s \le t$.

Because $Z(t)$ is a Markovian process, the least squares predictor of $Z(t)$ when $Z(t_1) = x_1\ ,\ \ldots\ ,\ Z(t_n) = x_n$ $(0 < t_1 < t_2 \ldots < t_n < t)$ is the conditional mean value of $Z(t)$ when $Z(t_n) = x_n$,

$$p(t\ ;\ x_n\ ,\ t_n) = x_n - \beta(\,t - t_n) + \int_{x_n}^{+\infty} [\,1 - \Lambda(\,x - \log\,(e^{\,\beta(t\,-\,t_n)} - 1))]\ d\,x$$

and the conditional mean-square error is, after a simple algebra,

$$\int_{x_n}^{+\infty} [1 - \Lambda(x - \log\,(e^{\,\beta(t\,-\,t_n)} - 1))]\ d\,x^2 - [(p\,(t\ ;\ x_n\ ,\ t_n) + \beta(t - t_n))^2 - x_n^2\,]\ .$$

Recall that also for the EMS process, from the stationarity, we have the mean value γ and the variance $\pi^2/6$.

Let us now study $\tilde{Z}\,(t) = \max\limits_{0 \le s \le t} Z(s)$, $(0 \le s \le t\,)$. The definition has effective meaning because of the relationship between $Z(t)$ and $Z_o(t)$. The correlation coefficient being continuous on the diagonal, Theorem C, p. 510 of Loéve (1961), shows that $Z(t)$ has many separability sets, one of them being the set of non-negative rationals.

We can then compute Prob $\{\tilde{Z}\,(t) \le x\}$. Fixing $h > 0$, for

$$F_n(x) = \text{Prob}\,\{\,\bigcap_{o}^{n}\,\{Z(k\,h) \le x\}\}$$

we have

$$F_n(x) = \text{Prob}\ \{ \bigcap_{o}^{n-1} \{Z(k\,h) \le x\} \cap \max\,(Z(n-1)\,h) - \beta\,h,\ E((n-1)\,h,\ n\,h) + \log(1 - e^{\beta h})) \le x\}\}$$

$$= F_{n-1}(x)\ .\ \exp\{-(1 - e^{-\beta\,h})e^{-x}\}$$

$$= F_o(x)\ .\ \exp\,(-\,n(1 - e^{-\beta\,h})\,e^{-x}) = \exp\,\{-\,(1 + n\,(1 - e^{-\beta\,h}))\,e^{-x}\}\ .$$

Taking now $\{h_n\}$ rational and such that $n\,h_n \to t$, we get

$$\text{Prob}\,\{\tilde{Z}(t) \le x\} = \exp(-\,(1 + \beta\,t)\,e^{-x}),$$

so that $\tilde{Z}(t) - \log\,(1 + \beta\,t)$ is a reduced Gumbel random variable.

Analogously we can show, for $s \le t$, that we have

$$\text{Prob}\,\{\tilde{Z}\,(s) \le x\ ,\ \tilde{Z}\,(t) \le y\} = \exp\{-(1 + \beta\,s)\ \max\,(e^{-x}, e^{-y}) - \beta\,(t - s)\,e^{-y}\}.$$

The $\tilde{Z}(t)$ process is evolutionary. Using the reduced margins $\xi = x - \log(1 + \beta\,s)$, $\eta = y - \log(1 + \beta\,t)$ we get the dependence function

$$k(w) = 1 - \min(\theta\ ,\ e^{w})\ /\ (1 + e^{w}) \qquad \text{with} \qquad \theta\ =\ \frac{1 + \beta\ s}{1 + \beta\ t}\ .$$

For large values of s and t we have $\theta \simeq s\,/\,t$ which suggests that $Z(t)$ is asymptotically similar to the extremal process.

Note that the mean value function is $\tilde{\mu}\,(t) = \gamma + \log(1 + \beta\ t)$, the variance is constant $(\pi^2/6)$ and the correlation coefficient is $\frac{6}{\pi^2}\ R\ (\frac{1 + \beta\ s}{1 + \beta\ t})$ for $s \le t$.

For the times $\tau_i = \log\,(1 + \beta\,t_i)$ we get

$$\text{Prob}\,\{Z\,(t_i) \le x_i\ ,\ 0 < t_1 < t_2 < \ldots < t_n\ \} = \prod_{i=1}^{n} \Lambda(\min\,(x_i\ ,\ \ldots\ ,\ x_n) - (\tau_i - \tau_{i-1}))$$

so that, for the new timing, $\tilde{Z}\,(\tau) = \tilde{Z}((e^{\tau} - 1)\,/\,\beta)$ is exactly an extremal process, and for large t_i and $\beta = 1$ (change of the time unit) we have $\theta_i \sim t_i$ so that $\tilde{Z}(t)$ is asymptotically an extremal process.

We see immediately that $Z(t)/\log\,(1 + \beta\,t) \xrightarrow{ms}$ as $t \to \infty$.

For small values of $s \le t$ ($\beta t << 1$) we have $\frac{1 + \beta s}{1 + \beta t} \sim 1$, and $\tilde{Z}(t) - \log(1 + \beta t) \simeq \tilde{Z}(t) - \beta t$ is approximately a reduced Gumbel random variable; the process is stationary only to the first order.

The relationship between $Z(t)$ and $\tilde{Z}(t)$ has a complex joint behaviour of $Z(s)$ and $\tilde{Z}(t)$, as the bivariate distribution is, unexpectedly, not biextremal.

By the technique previously used, we can show that

$$\Psi(x, y|t) = \text{Prob}(Z(t) \le x, \tilde{Z}(t) \le y) = \exp\{-\max(e^{-x-\beta t}, e^{-y}) - \int_0^{\beta t} \max(e^{-h-x}, e^{-y})\, dh\},$$

the margins being $\exp(-e^{-x})$ and $\exp(-(1 + \beta t)e^{-y})$. We have $\Psi(x, y|t) = \Psi(\min(x, y), y|t)$ from the definition. Then the dependence function, for the reduced margins, is

$$k(w) = \frac{1}{1 + e^{-w}} \{\max(e^{-\beta t}, \frac{e^{-w}}{1 + \beta t} + \int_0^{\beta t} \max(e^{-h}, \frac{e^{-w}}{1 + \beta t})\, dh\},$$

clearly not a biextremal one as shown in the more detailed form for $k(w)$

$$k(w) = \frac{1}{1 + e^{w}} \qquad \text{if} \quad w < -\log(1 + \beta t)$$

$$= 1 - \frac{1 + w + \log(1 + \beta t)}{(1 + \beta t)(1 + e^{w})} \qquad \text{if} \quad -\log(1 + \beta t) \le w \le \beta t - \log(1 + \beta t)$$

$$= \frac{e^{w}}{1 + e^{w}} \qquad \text{if} \quad w > \beta t - \log(1 + \beta t)\ ;$$

but it should be noted that this $k(w)$ is a generalized form of the biextremal dependence function, the natural one.

It is also easy to compute

$$G(w|t) = \text{Prob}\{\tilde{Z}(t) - Z(t) \le w\}.$$

Evidently we must have $G(w|t) = 0$ for $w < 0$ and for $t = 0$ we have $G(w|0) = H(w)$.

The general expression for $t > 0$ is

$$G(w|t) = \frac{e^{-\beta t} H(w - \beta t) + \int_0^{\beta t} e^{-h} H(w - h)\, d\, h}{\max(e^{-\beta t}, e^{-w}) + \int_0^{\beta t} \max(e^{-h}, e^{-w})\, d\, h} .$$

This result shows, as $G(w|t) = 1$ for $w > \beta t$, that, with probability one,

$$\tilde{Z}(t) \leq Z(t) + \beta t .$$

We can show this result directly, as follows.

From the basic equation we get $Z(t) \geq Z(s) - \beta (t - s)$ with $s \leq t$, and as $\tilde{Z}(t) = Z(s_t)$ we have

$$Z(t) \geq Z(s_t) - \beta (t - s_t)$$

and thus

$$\tilde{Z}(t) \leq Z(t) + \beta (t - s_t) \leq Z(t) + \beta t .$$

The correlation coefficient between $Z(t)$ and $\tilde{Z}(t)$, as follows from the natural model, is

$$\rho(t) = 1 + 6/\pi^2 \left[\frac{\beta t - \log(1 + \beta t)^2}{2} - \int_0^{\beta t} \log(e^w + \beta t - w)\, d\, w\right] = 1 - 6/\pi^2 \,.\, \beta^2 t^2 + \ldots$$

The statistical decision for an EMS process has not yet been considered. Only a few suggestions will be made here.

Two ways can be used to approach the estimation of the parameters of the general stochastics process $X(t) = \lambda + \delta Z(t)$, with $Z(t)$ a reduced EMS process.

One way is to consider the sequence $X_j = X(t_o + (j - 1) h)$ observing $X(t)$ at equal time steps and considering it as an EMS sequence. In that case we have for the parameters θ of the EMS sequence and β of the EMS process the relation $\theta = e^{-\beta h}$, and then we estimate θ (or β), λ and δ as in the previous section.

Another way is to recall that if $Z(t)$ is an EMS process then $Z(\frac{\log t}{\beta}) + \log t$ is an extremal process for $t \geq 0$ and thus, once β is estimated, we can estimate λ and δ for extremal processes, as before.

Extreme-Markovian-evolutionary (EME) sequences

Let $\{E_j\}$ $(j = 0, 1,2, ...)$ be a sequence of i.i.d. random variables with standard Gumbel margins, and $X_0 = \lambda_0 + \delta_0 Z_0$ a Gumbel random variable with parameters (λ_0 , δ_0), i.e., Z_0 is a standard Gumbel random variable; Z_0 is independent of all $\{E_j\}$.

In this section and the next we will follow Tiago de Oliveira (1986) . Let us consider the auto-regressive sequence

$$X'_{j+1} = \max (a + b X_j, a' + b' E_j) , \quad j = 0, 1, 2, ...$$

where the X'_j are assumed to be Gumbel random variables. Let (λ_j , δ_j) be the parameters of X_j, i.e., $X_j = \lambda_j + \delta_j Z_j$ with Z_j reduced Gumbel random variables. The auto-regressive relation can then be written as

$$\lambda_{j+1} + \delta_{j+1} Z_{j+1} = \max (a + b (\lambda_j + \delta_j Z_j) , a' + b' E_j) , \quad j = 0, 1, 2, ...$$

Thus Prob $\{Z_{j+1} \leq x\}$ is $\Lambda(x)$ if

$$\delta_{j+1} = b \delta_j$$

$$\delta_{j+1} = b'$$

$$e^{(a + b (\lambda_j - \lambda_{j+1})/b \delta_j} + e^{(a' - \lambda_{j+1})/b'} = 1$$

so that $b' = \delta_j = \delta_0$, $b = 1$ and $e^{\lambda_{j+1}/ \delta_0} = e^{(a + \lambda_j)/ \delta_0} + e^{a'/ \delta_0}$, and the auto-regressive relation takes the simpler form

$$X_{j+1} = \max (a + X_j , a' + \delta_0 E_j) .$$

If we introduce, for convenience , the "patterned" sequence $Y_j = (X_j - \lambda_0)/\delta_0$ $(Y_0 = Z_0)$ with $Z_j = (X_j - \lambda_j)/\delta_j = Y_j - (\lambda_j - \lambda_0)/ \delta_0 = Y_j - \eta_j$, $(\eta_j = (\lambda_j - \lambda_0)/ \delta_0 , \eta_0 = 0)$. Y_j then satisfies the auto-regressive equation

$$Y_{j+1} = \max (a_0 + Y_j , a'_0 + E_j) , \quad j = 0, 1, 2, ...$$

with $a_o = a/\delta_o$, $a'_o = (a' - \lambda_o)/\delta_o$, and the relation for $\{\lambda_j\}$ takes the form

$$e^{\eta_{j+1}} = e^{a_o}\, e^{\eta_j} + e^{a'_o}$$

In brief, the EME-sequence $\{X_j\}$, with X_o with parameters (λ_o , δ_o), satisfies the relation

$$X_{j+1} = \max (a_o\, \delta_o + X_j , \lambda_o + a'_o\, \delta_o + \delta_o\, E_j)$$

and the "patterned" sequence verifies

$$Y_{j+1} = \max (a_o + Y_j , a'_o + E_j) ;$$

(a_o , a'_o) are then the *essential* parameters and (λ_o , δ_o) are *incidental* parameters; we can reconstitute X_j by the relation $X_j = \lambda_o + \delta_o Y_j$, and as $Y_j = Z_j + \eta_j$ the margin parameters of X_j are $(\lambda_o + \delta_o\, \eta_j , \delta_o)$.

Consider, now, the difference equation

$$e^{\eta_{j+1}} = e^{a_o}\, e^{\eta_j} + e^{a'_o} , \qquad \text{with } \eta_o = 0 :$$

if $e^{a_o} = 1$ $(a_o = 0)$ we have $e^{\eta_j} = 1 + e^{a'_o}\, j$ $(\eta_j = \log (1 + e^{a'_o} j) > 0)$;

if $e^{a_o} \neq 1$ $(a_o \neq 0)$ we obtain

$$e^{\eta_j} = \frac{e^{a'_o}}{1 - e^{a_o}} + (1 - \frac{e^{a'_o}}{1 - e^{a_o}})\, e^{a_o j} = A + (1 - A)\, e^{a_o\, j} \quad (\geq 0)$$

with $A = \dfrac{e^{a'_o}}{1 - e^{a_o}}$; note that $\lim\limits_{a_o \to 1} e^{\eta_j} = 1 + e^{a'_o} j$, the expression of e^{η_j} for $a_o = 0$.

The condition of stationarity imposes $\eta_{j+1} = \eta_j = ... = \eta_o = 0$ so that $e^{a_o} + e^{a'_o} = 1$; this is the condition for stationarity obtained previously in EMS sequences; $\theta = e^{a_o}$ was there a dependence (essential) parameter.

As we have only fragmented results we will study some detailed features of the EME sequence.

Let us consider the monotonicity behaviour of the "patterned" sequence. As $M(Y_j) = M(\eta_j + Z_j) = \gamma + \eta_j$ we see that Y_j (and thus X_j) are increasing, constant or decreasing in mean according to $\eta_{j+1} > \eta_j$, $\eta_{j+1} = \eta_j$ or $\eta_{j+1} < \eta_j$, i.e., according to $e^{a_o} + e^{a'_o} > 1$, $e^{a_o} + e^{a'_o} = 1$ (stationarity) or $e^{a_o} + e^{a'_o} < 1$.

If $a_o = 0$ we do not have the decreasing behaviour and we get constancy only if $a_o = 0$ and $a'_o = -\infty$, i.e., $Y_{j+1} = Y_j$.

Now compute Prob $\{Y_{j+1} > Y_j\}$. If $a_o > 0$ it is immediate that $Y_{j+1} > a_o + Y_j$ so that Prob $\{Y_{j+1} > Y_j\} = 1$: the EME sequence is increasing with probability one, and so the method of "ups and downs" considered before for EMS sequences for the estimation of $\theta = e^{a_o}$, which should be < 1, cannot be used.

For $a_o = 0$ as $Y_{j+1} \geq Y_j$ we have Prob $\{Y_{j+1} > Y_j\} = 1/(1 + e^{a'_o} + j)$ decreasing with j and as Prob $\{Y_{j+1} > Y_j\} = 1 -$ Prob $\{Y_{j+1} > Y_j\} \to 1$ we see that the sequence stabilizes asymptotically.

When $a_o < 0$ we have Prob $\{Y_{j+1} > Y_j\} = 1 -$ Prob $\{Y_{j+1} \leq Y_j\} = (1 - e^{a_o})\, e^{a'_o} / [(2 - e^{a_o})\, e^{a'_o} + (1 - e^{a_o} - e^{a'_o})\, e^{a_o j}]$ with value $e^{a'_o}/(1 + e^{a'_o})$ at $j = 0$, converging to $(1 - e^{a_o})/(2 - e^{a_o})$ as $j \to \infty$ increases if $e^{a_o} + e^{a'_o} < 1$, behaving in a stable way if $e^{a_o} + e^{a'_o} = 1$ (constancy), and decreasing if $e^{a_o} + e^{a'_o} > 1$.

The "patterned" sequence $\{Y_j\}$ and also $\{X_j\}$ increase in median Prob $\{Y_{j+1} \geq Y_j\} > 1/2$ always if $a_o > 0$, when $j \geq 1 - e^{a'_o}$ if $a_o = 0$ and $a'_o > 0$, but never when $a'_o < 0$, and when $e^{-a_o j} < (e^{a_o} + e^{a'_o} - 1)\, e^{-a_o} + e^{-a'_o}$ if $e^{a_o} + e^{a'_o} > 1$ but never if $e^{a_o} + e^{a'_o} \leq 1$.

We will obtain the bivariate structure of an EME sequence, the multivariate structure being an immediate extension.

Taking $i < j$ we have

$$Y_j = \max(a_o + Y_{j-1}\, ,\, a'_o + E_{j-1}) = \max\,(2\, a_o + Y_{j-2}\, ,\, a_o + a'_o + E_{j-2}\, ,\, a'_o + E_{j-1}) = \ldots$$

$$= \max\,((j - i)\, a_o + Y_i\, ,\, a'_o + \max_{p\,=\,1}^{j\,-\,i}\,((p - 1)\, a_o + E_{j-p}))\,.$$

Then Prob $\{Y_i \leq x\, ,\, Y_j \leq y\} = \Lambda(\min(x\, ,\, y - (j - i)\, a_o - \eta_i))$

$$\times \prod_{p=1}^{j-i} \Lambda(y - a'_o - (p - 1)\, a_o) \quad \text{if} \quad a_o \neq 0$$

and

$$\text{Prob}\,\{Y_i \leq x\, ,\, Y_j \leq y\} = \exp\,\{ - \max(e^{-x}\, ,\, e^{-y})\, e^{\eta_i} - e^{a'_o}\,(j - i)\, e^{-y}\,\}$$

if $a_o = 0$.

The dependence function associated with the bivariate structure of $(Z_i , Z_j) = (Y_i - \eta_i , Y_j - \eta_j)$ is

$$\text{Prob}\{Z_i \le x , Z_j \le y\} = \text{Prob}\{Y_i \le x + \eta_j , Y_j \le y + \eta_j \} = (\Lambda(x)\ \Lambda(y))^{k_{i,j}(y-x)} ,$$

where

$$k_{i,\,j}\,(w) = 1 - \frac{\min(e^{(j-i)\,a_o - \eta_j + \eta_i} , e^{w})}{1 + e^{w}} .$$

The correlation coefficient is, then,

$$\rho_{i,\,j} = \frac{6}{\pi^2}\ R\ (e^{(j-i)\,a_o - \eta_j + \eta_i}) .$$

where $0 \le e^{(j-i)\,a_o - \eta_j + \eta_i} \le 1$. For $a_o = 0$ we get $\rho_{ij} = \frac{6}{\pi^2}\ R\ (\frac{1 + e^{a'oi}}{1 + e^{a'oj}})$ and for $a_o \ne 0$ we have

$$\rho_{ij} = \frac{6}{\pi^2}\ R\ \frac{1 - A + A\ e^{-a'oi}}{1 - A + A\ e^{-a'oj}} ,$$

where the argument of R(.) is between 0 and 1. For any EMS sequence, with $e^{a_o} + e^{a'o} = 1$ we get $\rho_{ij} = \frac{6}{\pi^2}\ R\ (e^{a_o(j-i)})$ evidently with $a_o < 0$ as obtained previously, with $a_o = \log \theta$, $a'_o = \log(1 - \theta)$, $0 \le \theta \le 1$: independence $(\theta = 0)$ gives $\rho_{ij} = 0$.

Let us obtain some more propositions that can be useful for the statistical analysis of the EME sequences.

A first result is that $Y_j / j \xrightarrow{ms} \max (a_o , 0)$ and $Y_{j+1} - Y_j \xrightarrow{ms} a_o$, if $a_o > 0$, as $j \to \infty$. From $Y_j = \eta_j + Z_j$ we obtain

$$M(Y_j/j) = (\gamma + \eta_j)/ j \to \max(a_o , 0),\ V(Y_j / j) = (\pi^2/6)/ j^2 \to 0 \qquad \text{and}$$

$$M(Y_{j+1} - Y_j) = \eta_{j+1} - \eta_j \to a_o \text{ if } > 0 \text{ and } V(Y_{j+1} - Y_j) = \pi^2/3\ (1 - \rho_{j,j+1}) \to 0.$$

As $Y_{j+1} - Y_j = a_o$ (or $a'_o + E_j \le a_o + Y_j$) with probability

$$\frac{e^{\,a_o}\ .\ e^{\eta j}}{e^{a'o} + e^{\,a_o}\ e^{\eta j}} \to e^{\min\,(a_o,o)} \text{ as } j \to \infty \text{ and}$$

$Y_{j+1} - Y_j > a_o$ with the complementary probability, it seems natural to study the statistic $A_n = \min_{1}^{n} (Y_j - Y_{j-1})$.

Let $Q_n(\Delta_1, \ldots, \Delta_n)$ denote Prob $\{Y_1 - Y_o > a_o + \Delta_1, \ldots, Y_n - Y_{n-1} > a_o + \Delta_n\}$ for $\Delta_i \geq 0$. It is immediate that the event $D_n = \{Y_1 - Y_o > a_o + \Delta_1, \ldots, Y_n - Y_{n-1} > a_o + \Delta_n\}$, as $Y_j = a'_o + E_j$, is equivalent to $D'_n = \{a'_o + E_1 > a_o + \Delta_1 + Y_o, E_2 > a_o + \Delta_2 + E_1, \ldots, E_n > a_o + \Delta_n + E_{n-1}\}$, and we get

$$Q_n(\Delta_1, \ldots, \Delta_n) = G_n(e^{a_o + \Delta_1 - a'_o}, e^{a_o + \Delta_2}, \ldots, e^{a_o + \Delta_n})$$

where

$$G_n(\varphi_1, \ldots, \varphi_n) = \int_{\substack{t_o > \varphi_1 t_1 \\ \cdots \\ t_{n-1} > \varphi_n t_n}} e^{-(t_o + t_1 + \ldots + t_n)}\, dt_o\, dt_1 \ldots dt_n,$$

which satisfies the relation

$$G_n(\varphi_1, \ldots, \varphi_n) = \frac{1}{1 + \varphi_1} G_{n-1}((1 + \varphi_1)\varphi_2, \varphi_3, \ldots, \varphi_n).$$

Then

$$Q_n(\Delta, \ldots, \Delta) = \frac{1}{1 + e^{a_o + \Delta - a'_o}} Q_{n-1}((1 + e^{a_o + \Delta - a'_o}) e^{a_o + \Delta}, \ldots, e^{a_o} e^{\Delta})$$

tends to 0 as $n \to \infty$. A_n is thus an estimator of a_o.

Some other results give hints for statistical estimation. It is easily shown that, as

$$M(Y_{j+1} - Y_j - a_o) \overset{ms}{\to} -\min(a_o, 0), \sum_{o}^{n-1} M(Y_{j+1} - Y_j - a_o) = Y_n - Y_o - n a_o$$

$\to \log(1 - \frac{e^{a_o}}{1 - e^{a_o}})$ if $a_o > 0$ and $\sum_{o}^{n-1} M(Y_{j+1} - Y_j) \to a'_o - \log(1 - e^{a_o})$ if $a_o > 0$, as $n \to \infty$.

We have also, for $a_o \leq 0$, Prob $\{\max(Y_o, Y_1, \ldots, Y_n) \leq x\}$

$= \text{Prob}\{\max(Y_o, a'_o + E_o, \ldots, a'_o + E_{n-1}) \leq x\} = \Lambda(x)\Lambda^{n-1}(x - a'_o)$

$$= \exp\{-e^{-x}(1 + (n-1)e^{a'_0})\}$$

and so $\quad \text{Prob}\{\max(Y_0, Y_1, \ldots, Y_n) - \log n \leq x\} = \Lambda(x - a'_0)$.

Some remarks on statistical decision for EME sequences

We have not obtained sufficient results even for simple and quick statistical decision but some remarks can be made.

It is natural, in this initial phase, to divide statistical decision for the EME sequences into two steps: statistical decision concerning the (essential) parameters (a_0, a'_0) and then, supposing (a_0, a'_0) is known, to estimate (λ_0, δ_0) by the least squares method. In principle, the estimators of (a_0, a'_0) must be independent of (λ_0, δ_0) and those of the incidental parameters must be quasi-linear, i.e., such that $\lambda^*_0(\alpha + \beta X_j) = \alpha + \beta \lambda^*_0(X_j)$ and $\delta^*(\alpha + \beta X_j) = \beta \delta^*_0(X_j)$ for $-\infty < \alpha < +\infty$, $0 < \beta < +\infty$, as happens with the least squares method; see Cramér (1946) and Silvey (1975).

Let us now suppose we are dealing with the "patterned" sequence $\{Y_j\}$.

A test of constancy $(a_0 = 0, a'_0 = -\infty)$ is not necessary; to devise tests of independence $(a_0 = -\infty, a'_0 = 0)$ and of stationarity is very important. We will consider only the important case where $a_0 > 0$.

As $Y_j/j \xrightarrow{ms} a_0 (> 0)$, a natural region for deciding $a_0 > 0$ is to accept this hypothesis if $\{X_j > A_j\}$, which is also the Neyman-Pearson test of $\lambda > 0$ against $\lambda \leq 0$ for the distribution $\Lambda(x - \lambda)$. Recall that $\eta_j - a_0 \to \log\left(1 + \frac{e^{a'_0}}{e^{a_0} - 1}\right)$ if $a_0 > 0$, $\eta_j - \log j \to a'_0$ if $a_0 = 0$, and $\eta_j \to a'_0 - \log(1 - e^{a_0})$ if $a_0 < 0$ as $j \to \infty$; η_j increases linearly with j if $a_0 > 0$, logarithmically if $a_0 = 0$, and converges to a constant if $a_0 < 0$.

A_j can be defined by imposing $\text{Prob}\{Y_j \leq A_j | a_0 = 0\} = \alpha$ or $\Lambda(A_j - \log(1 + e^{a'_0} j)) = \alpha$ and so $A_j = \log(1 + e^{a'_0} j) - \log(-\log \alpha)$ still depending on the fixation of the value a'_0.

A last remark: when $a_0 > 0$, as $Y_{j+1} \geq a_0 + Y_j$ we have $Y_j \geq Y_0 + a_0 j$, so that after some steps (depending on the random Y_0 and on (a_0, a'_0)) we will practically always have $Y_{t+1} = a_0 + Y_t$ because $\text{Prob}(a'_0 + E_j > Y_0 + a_0 j) = \frac{e^{a'_0}}{e^{a'_0} + e^{a_0 j}} \to 0$ very quickly. In

practice a'_0, can be estimated only by the first values of $\{Y_t\}$, if possible. The difficulty is analogous to the no-jump situation (non-increasingness) in extremal sequences/processes.

Statistical decision for EME-sequences is still open; the results given here may be helpful in some cases for a first step.

Evidently if we have, or suppose, or assume, that the above processes have margins that are not Gumbel but Frechét or Weibull, by the usual transformations, estimating the convenient parameters, we can reduce them to Gumbel margins.

Sliding extreme (SE) sequences

Consider a (doubly) infinite sequence of *independent* random variables $\ldots X_{-1}, X_0, X_1, \ldots, X_n, \ldots$ which are assumed to have Gumbel distribution.

We will suppose that if the sequence is *stable* the X_i have the parameters (λ, δ), i.e.,

$$\text{Prob}\,\{X_j \leq x\} = \Lambda((x - \lambda)/\delta)\ ,$$

but if the sequence is unstable the parameters are $(\lambda + (\nu + n\,\theta)\delta\,,\,\delta)$ (ν and $\theta > 0$ unknown), i.e., the distribution function of the X_j is

$$\Lambda(((x - \lambda) - (\nu + j\,\theta)\,\delta)/\delta)\,.$$

We will obtain a test of $\theta = 0$ vs. $\theta > 0$, i.e. *stability* vs. (*positive*) *instability*.

Here and in the next section will follow Tiago de Oliveira (1987).

Evidently, if the distributions of the observations are Weibull or Fréchet, the usual log-transformations will reduce them to the present case, as is well known and will be done later.

Thus we can write

$$X_j = \lambda + (\nu + j\,\theta)\,\delta + \delta\,Z_j$$

where $\{Z_j\}$ are independent reduced Gumbel random variables with $\nu = 0$ and $\theta = 0$ in the stable case and $\theta > 0$ in the (positive) unstable case (linear increase of the location parameter).

The underlying idea is that, although we assume independence either in the stable or the unstable cases, independence will act as a reference pattern for significance tests.

The nuisance parameter $\nu > 0$ can be interpreted as meaning that instability began somewhere in the past ($j_0 < 0$), before sampling, or even can absorb a wrong choice of λ.

It is easy to show that if r_n is the correlation coefficient between $(1, 2, \ldots, n)$ and $(X_1, \ldots, X_n)$, $r_n \xrightarrow{P} 0$ if $\theta = 0$ and $r_n \xrightarrow{P} 1$ if $\theta > 0$.

Note that if we take $Y = \bar{w} - e^{-X}$, where X has the Gumbel distribution with parameters (λ, δ), then Y has the Weibull distribution

$$\text{Prob}\{Y \leq y\} = \exp\{-(\frac{\bar{w} - y}{e^{-\lambda}})^{1/\delta}\} \qquad \text{for} \quad y \leq \bar{w}$$

$$= 1 \qquad \text{for} \quad y \geq \bar{w}$$

with the location parameter (right-end point) $\bar{w}$, the dispersion parameter $\tau = e^{-\lambda}$ – decreasing to zero if $\lambda \to +\infty$ and thus increasing the probability $\text{Prob}\{Y > y\} = 1 - \text{Prob}\{Y \leq y\}$, which is relevant to earthquakes – and shape parameter $\alpha = 1/\delta$; if we take $Y = e^X + \underline{w}$, with X also a Gumbel random variable with parameters (λ, δ), then

$$\text{Prob}\{Y \leq y\} = 0 \qquad \text{if} \quad y \leq \underline{w}$$

$$= \exp\{-(\frac{y - \underline{w}}{e^{-\lambda}})^{-1/\delta}\} \qquad \text{if} \quad y \geq \underline{w}$$

a Fréchet distribution with location parameter (left-end point) $\underline{w}$, dispersion parameter $\tau = e^{\lambda}$ – increasing with $\lambda \to \infty$ and thus seeming irrelevant for applications – and shape parameter $\alpha = 1/\delta$.

For earthquake applications Yegulalp and Kuo (1974) have shown that the Weibull distribution gives a good fit, but accepting the Gumbel distribution in some seismic areas; for the area considered in the case study Ramachandran (1980) says that the Gumbel distributions gives a good fit; for some remarks on global modelling for seismic areas see Tiago de Oliveira (1984).

Statistical decision for SE sequences

The likelihood of the sample $(x_1, \ldots, x_n)$ is

$$L = L(x_1, \ldots, x_n) = \frac{1}{\delta^n} \exp\{ - \sum_1^n \frac{x_j - \lambda}{\delta} - \nu - j\theta) \}$$

$$\exp\{ - \sum_1^n e^{-((x_j-\lambda)/\delta - \nu - j\theta)} \},$$

and so the LMP test of $\theta = 0$ vs. $\theta > 0$ (ν assumed to be zero) is given by the rejection region

$$\frac{\partial \log L}{\partial \theta} |_{\theta = 0} \geq c'_n$$

or

$$T_n = \sum_1^n j \exp(- \frac{x_j - \lambda}{\delta}) \leq c_n .$$

In the tested stability ($\nu = 0$, $\theta = 0$), the $\exp(- \frac{x_j - \lambda}{\delta}) = E_j$ being standard exponential random variables, the distribution of T_n is that of $\sum_1^n j E_j$, with the $\{E_j\}$ independent.

It is obvious that if

$$F_n(c) = \text{Prob} \{ \sum_1^n j E_j \leq c \}$$

we have

$$F_n(c) \leq F_{n-1}(c) \leq \ldots \leq F_1(c) = 1 - e^{-c}$$

with

$$F_n(c) \leq F_{n-1}(c) - e^{-c/n} \int_0^c e^{y/n} \, d F_{n-1}(y),$$

and so, for instance,

$$F_2(c) = 1 + e^{-c} - 2 e^{-c/2} .$$

Denoting by $c_n(\alpha)$ the solution of $F_n(x) = 1 - \alpha$, we see that

$$c_n(\alpha) > c_{n-1}(\alpha) .$$

We have

$$c_1(\alpha) = -\log \alpha\,,\quad c_2(\alpha) = -2\log\left(1 - \sqrt{1-\alpha}\,\right)$$

which for $\alpha = .25$, as

$$c_1(.05) = 2.9957323 < c_2(.05) = 7.3522767\,,$$

gives an idea of the initial rate of increase of $c_n(.05)$.

If $\nu \neq 0$ and $\theta = 0$ (instability *before* the sample) the correct statistic would be $T_n\, e^{-\nu}$ (with T_n as before), the correct region would be $T_n\, e^{-\nu} \leq c_n$ smaller than $T_n \leq c_n$, thus giving over-rejection of stability.

T_n has the mean value $n(n+1)/2$, variance $n(n+1)\,(2n+1)/6$, and we can show easily that

$$T'_n = \frac{T_n - n(n+1)/2}{\sqrt{n(n+1)\,(2n+1)/6}}$$

is asymptotically standard normal; but this result is not very useful because we will deal with small values of n in applications.

The test assumes (λ, δ) known (stable case); we can presume it in some cases, such as earthquakes, because for each seismic region the long history gives sufficiently good estimates of the parameters to be used.

If we were dealing with the Weibull distribution, with location dispersion $\bar{w}$, dispersion parameter τ, and shape parameter α, the statistic T_n takes the form

$$T_n = \sum_{1}^{n} j \left(\frac{\bar{w} - y_j}{\tau}\right)^{\alpha}\,;$$

when dealing with the Fréchet distribution with location parameter $\underline{w}$, dispersion parameter τ, and shape parameter α, the statistic T_n is

$$T_n = \sum_{1}^{n} j \left(\frac{y_j - \underline{w}}{\tau}\right)^{-\alpha}\,.$$

The distribution of T_n in all three cases is the same, in the stable case.

Let us now consider a case study, the waning down of the Santa Barbara earthquake of 13 August 1978.

The daily maximum magnitudes for 13th August and the following days until 27th August, with no observation at the 27th, are

Day	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Mag.	5.1	3.1	2.4	2.7	2.6	2.3	2.3	3.1	2.6	2.1	2.3	2.4	2.1	2.7	—

as given in Corbett and Johnson (1982).

Taking the last six daily maxima of the paper, which goes to 30th September (2.0, 3.5, 1.9, 2.0, 2.0, 1.8), the Leiblein-Zellen estimators – see Chapter V for details – are $\lambda^* = 1.980$ and $\delta^* = .314$, and the estimates given by Ramachandran (1980) for the area (case c) are $\lambda^{**} = 1.825$ and $\delta^{**} = 1/3.425 = .292$. For simplicity we will take $\tilde{\lambda} = 2.0$ and $\tilde{\delta} = .3$. Using the first six observed daily maxima *after* the earthquake, in *reverse order* to take account of the expected downward trend, we get

$$T_n = \sum_{1}^{6} j \exp\left(-\frac{x_j - 2}{.3}\right) = 3.3688867 < c_2(0.05);$$

we must conclude that a downward trend exists. The closeness of the values of (λ^*, δ^*) and $(\lambda^{**}, \delta^{**})$ can be interpreted as meaning that after, approximately, two weeks the usual stability was practically attained.

ANNEX VI

On The Choice of Stochastic Processes Models

Up to now we have discussed some quick estimation, not necessarily optimal, if and when some model was assumed. It may be useful to take a look at the choice of the models, even if we do not use sound procedures, but bearing in mind the theory of regression.

Let us recapitulate some second-order features of sequences of extremes, $X_n = \lambda + \delta Z_n$. We have, where $R(u) = -\int_0^u \frac{\log t}{1-t}\, dt$,

for *extremal sequences*: $M(X_n) = \lambda + \delta(\gamma + \log n)$, $V(X_n) = \frac{\pi^2}{6}\delta^2$ and

$$\rho(X_m, X_n) = \frac{6}{\pi^2} R\left(\frac{\min(m,n)}{\max(m,n)}\right) \quad (m, n \geq 1);$$

for *EMS sequences*:

$$M(X_n) = \lambda + \gamma\delta, \quad V(X_n) = \frac{\pi^2}{6}\delta^2$$

and

$$\rho(X_m, X_n) = \frac{6}{\pi^2} R(\theta^{|m-n|}) = \frac{6}{\pi^2} R(e^{-\varphi|m-n|}), \ \varphi = -\log\theta, \ (m, n \geq 1), \ \varphi \geq 0;$$

for *EME sequences*:

$$M(X_n) = \lambda_o + \delta_o(\gamma + \eta_n) \quad (\eta_n = \log(A + (1-A)e^{a_o n}), \qquad V(X_n) = \frac{\pi^2}{6}\delta_o^2$$

and

$$\rho(X_m, X_n) = \frac{6}{\pi^2} R\left(\frac{1 - A + A\, e^{-a_o \min(m,n)}}{1 - A + A\, e^{-a_o \max(m,n)}}\right), \ (m, n \geq 0);$$

and for *sliding sequences*:

$$M(X_n) = \lambda + (\gamma + \nu + \theta_n)\,\delta \quad \text{and} \quad V(X_n) = \frac{\pi^2}{6}\,\delta^2 \quad (-\infty < n < +\infty)\,.$$

We see, thus, that the variances are constant but the mean value varies logarithmically with n for extremal sequences, is constant in n for EMS sequences, has the form A + log(1 + B $e^{a_o n}$) for EME sequences, and varies linearly with n for sliding sequences. Thus a relatively long trajectory can give some hints on the model.

For extremal and EMS processes, the sequences that can be obtained by the use of periodic sampling in the continuous trajectory can give suggestions for the model to be used.

Exercises — Part IV

IV.1—Consider the EMS sequence $\{Z_k\}$. Obtain the expression of

$$\text{Prob}\{\ \min(Z_1, \dots, Z_k) \geq z\} = Q_k\ (z, z|\theta)$$

where

$$Q_k\ (x, y|\theta) = \text{Prob}\{\ \min_{1}^{k-1}\ (Z_1, \dots, Z_{k-1})\ \geq\ x, Z_k \geq y\}$$

$$= Q_{k-1}\ (x, \max\ (x, y - \log\theta))\ \Lambda\ (y - \log\ (1 - \theta))\ .$$

Obtain its expression when $x < y - \log\theta$ and $x > y - \log\theta$.

IV.2— Consider the general EMS sequence $X_k = \lambda + \delta\, Z_k$ and the seek estimators of $(\lambda, \delta, \theta)$.
We have $\min_{2}^{k}\ (X_i - X_{i-1}) \xrightarrow{P} \delta\ \log\theta$ which gives an (over–) estimator of $\Delta = \delta \log\theta$.
Using this result obtain, supposing $\Delta^* = \Delta$, an estimator of (λ, δ) .

IV.3— Translate maxima results to minima results and vice-versa; use, in particular, the relation between the distribution function and the survival function).

IV.4— Study the processes of oldest ages of death for men and women in Sweden (Tables I and Tables I + II), trying to fit any of the random sequences that may have trend and/or oscillations. In particular they may be considered as sliding processes of maxima, with Gumbel margins.

IV.5— Consider a reduced extremal process $Z(t)$, $t > 0$ observed at (non-random) instants $(0 <)$ $t_1 < t_2 < \dots < t_n < \dots$. The best (least squares) predictor of $Z(t_{n+1})$ $(t_n < t_{n+1})$ of the form $Z_1^*(t) = Z(t_n) + a$ has $a = \log(t_{n+1}/t_n)$. The best (i.e., least squares) linear predictor of $Z(t_{n+1})$ based only on $Z(t_n)$ is of the form $Z^{**}(t_{n+1}) = \alpha + \beta\, Z(t_n)$; the minimization of $MSE = M(Z^{**}(t_{n+1}) - Z(t_{n+1}))^2$ leads to

$$Z^{**}(t_{n+1}) = \gamma + \log t_{n+1} + \rho\ (t_n/t_{n+1})\ (Z(t_n) - \gamma - \log t_n)\ .$$

They are both unbiased (i.e., with the same mean value) and

$$\text{MSE}\,(Z^*(t_{n+1})) = \frac{\pi^2}{3}\,(1 - \rho\,(t_n/t_{n+1})) \quad \text{and} \quad \text{MSE}\,(Z^{**}(t_{n+1})) = \frac{\pi^2}{6}\,(1 - \rho^2\,(t_n/t_{n+1}))\ .$$

Note that $Z^{**}(t_{n+1})$ can be compared with the simpler linear predictor $Z^*(t_{n+1}) = Z(t_n) + \log \frac{t_{n+1}}{t_n}$. The efficiency is

$$\frac{\text{MSE}(Z^*(t_{n+1}))}{\text{MSE}(Z^{**}(t_{n+1}))} = \frac{1 + \rho\,(t_n/t_{n+1})}{2} < 1.$$

None of the predictors is invariant for linear transformations, and so they can't be used to predict the general extremal processes $X(t) = \lambda + \delta\,Z(t)$.

IV.6— Consider the extremal processes $X(t) = \lambda + \delta\,Z(t)$, $t \geq 0$, where $Z(t)$ is the reduced extremal process. Suppose observations are made at (non-random) instants ($0 <$) $t_1 < t_2 < \ldots < t_n < \ldots$.

Obtain the expression of the quasi-linear predictor of $X(t_{n+1})$ when $X(t_i)$, $i = 1, \ldots, n$, are known. The quasi-linear predictor of $X(t_{n+1})$ based on the last two observations $X(t_{n-1})$ and $X(t_n)$ $(\geq X(t_{n-1}))$ is, obviously.

$$X^*(t_{n+1}) = X(t_n) + \beta(\,X(t_n) - X(t_{n-1}))\ .$$

The best (least squares) predictor of $X(t_{n+1})$, minimizing

$$\text{MSE}(X^*(t_{n+1})) = \text{M}(X^*(t_{n+1}) - X(t_{n+1})\})^2\ ,$$

is given by

$$\beta = \frac{\text{M}\,((X(t_{n+1}) - X(t_n))\ (X(t_n) - X(t_{n-1})))}{\text{M}((X(t_n) - X(t_{n-1}))^2)}$$

and

$$\text{MSE}\,(X^*(t_{n+1})) = \text{M}\,((X(t_{n+1}) - X(t_n))^2) - \beta^2\ \text{M}((X(t_n) - X(t_{n-1}))^2)\ ,$$

which can be expressed in mean values and covariance of the process.

IV.7— Define an extremal process of Weibull minima, using the conversion between the Weibull distribution of minima and the Gumbel distribution of maxima.

IV.8— Consider a max-compound Poisson stochastic process, $X(t) = \max\,\{Y_i\}$, where $N(t)$ is a Poisson process with intensity ν and $Y_0,\ Y_1, \ldots, Y_k, \ldots$, is a sequence of independent

random variables such that Y_0 has the distribution function $G_0(x)$ and Y_j ($j \geq 1$) have the distribution function $G(x)$. Show that $\text{Prob}\{X(t) \leq x\} = G_0(x) \exp\{-\nu t(1 - G(x))\}$ and that for *no* choice of (G_0, G) can we have $\text{Prob}\{X(t) \leq x\} = \Lambda(\alpha(t) + \beta(t)\, x)$.

IV.9— Analyse the same question for max-filtered Poisson processes and max-renewal point processes.

S. M. BERMAN (1962). Limiting distribution of the maximum term in sequences of dependent random variables, *Ann. Math. Statist.*, 33, 894-908

S. M. BERMAN (1964). Limiting distribution of the maximum of a diffusion process, *Ann. Math. Statist.*, 35, 319-329

S. M. BERMAN (1964'). Limit theorems for the maximum term of a stationary sequence, *Ann. Math. Statist.*, 35, 502-516

S. M. BERMAN (1974). Sojourns and extremes of Gaussian processes, *Ann. Prob.* , 2, 999-1026 (Correction (1980), 8, 999)

S. M. BERMAN (1982). Sojourns and extremes of stationary processes, *Ann. Prob* ., 10, 1-46

D. E. CARTWRIGHT and M. S. LONGUET-HIGGINS (1956) . The statistical distribution of maxima of a random function, *Proc. Roy. Soc. London*, A, 237, 212-232

EDWARD J. CORBETT and CARL E. JOHNSON (1982). The Santa Barbara, Califórnia, earthquake of 13 AUGUST 1978, *Bull. Seismol. Soc. America*, 72, # 6, Dec. 2201-2226

H. CRAMÉR (1946). *Mathematical Methods of Statistics*, Princeton University Press, New Jersey

H. CRAMÉR and M. R. LEADBETTER (1967). *Stationary and Related Stochastic Processes*, Wiley, New York

P. DEHEUVELS (1981). The strong approximation to extremal processes, *Z. Wahrsch. verw. Gebiete*, 58, 1-6

P. DEHEUVELS (1981'). The strong approximation to extremal processes II, *Z. Wahrsch. verw. Gebiete*, 62, 7-15

M. DWASS (1964). Extremal processes, *Ann. Math. Statist.*, vol. 35, 1718-1725

M. DWASS (1966). Extremal processes, II, *Illinois J. Math.*, vol. 10, 381-391

M. DWASS (1973). Extremal processes, III, *Northwerstern Univ. Discussion Paper*

M. FISZ (1962). *Probability Theory and Mathematical Statistics*, Wiley, New York

J. GALAMBOS (1978). *The Asymptotic Theory of Extremes Order Statistics,* Wiley, New-York; (1987), 2nd. ed., Krieger Publ.

N. GOLDSTEIN (1963). Random numbers for extreme values distributions, *Publ. Inst. Statist. Univ. Paris*, XII, 137-158

E. J. GUMBEL (1958). *Statistics of Extremes*, Columbia University Press, New York

J. LAMPERTI (1964). On extreme order statistics, *Ann. Math. Statist.*, 35, 1726-1737

M. R. LEADBETTER, GEORG LINDGREN and HOLGER ROOTZÉN (1983). *Extremes and Related Properties in Random Sequences and Processes*, Springer-Verlag, Heidelberg

M. LOÉVE (1961). *Probability Theory* (2nd ed.), van Nostrand, New York

HENRY B. MANN (1953). *An Introduction to the Theory of Stochastic Process with Continuous Parameter*, Nat. Bur. Standards, Appl. Math. Series, # 24, Washington

G. F. NEWELL (1962). Asymptotic extreme value distribution for one-dimensional diffusion process, *J. Math. and Mech.*, 11, 481-496

A. PAPOULIS (1965). *Probability, Random Variables and Stochastic Processes*, McGraw Hill, New York

J. PICKANDS III (1967). Sample sequences of maxima, *Ann. Math. Statist.*, 39, 1570-1574

J. PICKANDS III (1967'). Maxima of stationary Gaussian processes, *Z. Wahrsch. verw. Gebiete*, 7, 190-233

J. PICKANDS III (1968). Moment convergence of sample extremes, *Ann. Math. Statist.*, 39, 881-889

J. PICKANDS III (1969). Asymptotic properties of the maximum of a stationary Gaussian process, *Trans. Amer. Math. Soc.*, 145, 75-86

J. PICKANDS III (1969'). Upcrossing probabilities for stationary Gaussian process, *Trans. Amer. Math. Soc.*, 145, 51-73

G. RAMACHANDRAN (1980). Extreme value theory and earthquake insurance, *Trans. 21st Int. Congress Actuaries*, 1, 337-353, Switzerland

S. I. RESNICK (1973). Extreme processes and record time values, *J. Appl. Prob.*, 10, 863-868

S. I. RESNICK (1975). Weak convergence to extremal processes, *Ann. Prob.*, 3, 951-1960

S. I. RESNICK and M. RUBINOVITCH (1973). The structure of extremal processes, *Adv. Appl. Prob.*, 5, 287-307

S. O. RICE (1944/45). Mathematical analysis of random noise, *Bell Syst. Techn. J.*, 23, 282-332 and 24, 46-156

S. D. SILVEY (1975). *Statistical Inference*, Chapman and Hall, London

J. TIAGO DE OLIVEIRA (1962/63). Structure theory of bivariate extremes; extensions, *Estudos Mat., Estatist. e Econometria*, VII, 165-194, Lisboa

J. TIAGO DE OLIVEIRA (1968). Extremal processes; definition and properties, *Publ. Inst. Statist. Univ. Paris*, XII, # 2, 25-36

J. TIAGO DE OLIVEIRA (1972). Statistics for Gumbel and Fréchet distributions, in *Structural Safety and Reliability*, A. M. Freudenthal ed., 91-105

J. TIAGO DE OLIVEIRA (1972'). An extreme-markovian-stationary sequence; quick statistical decision, *Metron*, XXX, # 1-4, A, 1-11

J. TIAGO DE OLIVEIRA (1973). An extreme-markovian-stationary process, *Proc. 4th. Conf. Prob. Th.*, Brasov, Editure Academici Republici Socialista Romania, 217-225

J. TIAGO DE OLIVEIRA (1982). A definition of estimator efficiency in k-parameter case, *Ann. Inst. Statist. Math.*, 14, #3, A, 411-421, Tokyo

J. TIAGO DE OLIVEIRA (1984). Weibull distributions and large earthquake modelling, in *Probabilistic Methods in the Mechanics of Solids and Structures*, S. Eggwertz and N. C. Lind eds., 81-89, Springer-Verlag, Heidelberg

J. TIAGO DE OLIVEIRA (ed.) (1984'). *Statistical Extremes and Applications*, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1986). An extreme-markovian-evolutionary sequence; *Trab. Estad. Inv. Oper.*, 36, 291-300

J. TIAGO DE OLIVEIRA (1987). The structure of sliding processes; applications, *Proc. Int. Conf. Structural Failure*, I, Singapure

TUNCEL M. YEGULALP and J. T. KUO (1974). Statistical prediction of the occurrence of maximum magnitude earthquakes, *Bull. Seismol. Soc. America*, 64, #2, Apr, 393-414

PART V

Complements

Chapter XVI

GEA: GENERAL EXTREMES ANALYSIS,

a guide in natural language

Introdution

This chapter is not a program for *GEA*, which would differ according to the language used, nor a set of recipes for the concrete analysis of data. It presupposes knowledge of the corresponding chapters and the writing of a program for the computer used in the appropriate computer language. This guide is, at most, a non-diagram fluxogram of the procedures to be implemented in each computer + language system. In fact, with every system, we can obtain (to some decimal points of accuracy) the results that will be analysed in the Case Studies chapter. In general we will think of maxima (largest values) of samples.

This sequence of procedures depends, evidently, on the *formulae* to be found in Parts II, III and IV. *We assume, throughout, that the exact distribution of extremes is approximated by one of the asymptotic distributions.*

Step 1 — Statistical Choice of Models (SCM)

This initial step, for univariate distributions (or margins in the multivariate case) leads to the (preliminary) model to be fitted to the data and subsequently used. It can be omitted if there exist justifications (such as theoretical results, previous experience, etc.) to justify some assumption and, thus, to act in such a way. Although in this phase we integrate a geometrical plotting of data, its purpose is not to justify the statistical choice made analytically but to *support intuition* and to clarify for experimentalists the use of the assumed model.

Thus, after filing the data — for each margin if the data are multivariate — we can do the following types of *GEA*, which are not mutually exclusive.

For graphical choice, after ordering the univariate (or univariate margins) data, we plot them on a Gumbel probability paper, as said in Chapter IV, and choose one of the 3 models (Weibull, Gumbel and Fréchet) according to the "eye-curvature" of the plotted points.

This graphical choice, although attractive and intuitive, must be confirmed by the use of the analytical choice, through the test statistic $\hat{V}_n$ and the associated decision rule (Chapter VIII). Recall that $\hat{V}_n$ pressuposes that the location-dispersion parameters of a possible Gumbel distribution are estimated as said before.

This trilemma decision, analogous to an LMPU two-sided test, splits the next step in three directions, according to the chosen model or alternative (Weibull, Gumbel or Fréchet). In this case we choose between $\theta < 0$ (Weibull), $\theta = 0$ (Gumbel) and $\theta > 0$ (Fréchet) distributions (in the von Mises-Jenkinson formula), excluding thus the distributions of extremes that are not attracted to any of the limiting distributions.

In the cases where we can assume, with sufficient information, that $\theta \geq \theta_0 \geq 0$ or $\theta \leq \theta_0 \leq 0$, we can have a usual LMP test, also using the test statistic $\hat{V}_n$.

Another (weak) check of the sign of θ ($< 0, = 0, > 0$) can be made by using the statistic Q_n (Chapter VIII), intuitive but not as efficient as $\hat{V}_n$.

A comment : if the graphical choice and the analytical choice (with $\hat{V}_n$ and its ancillary, Q_n) show a clear-cut discrepancy we can conjecture that the asymptotic approximation was not obtained for the sample sixe (n) used, that the i.i.d. condition is not (approximately) verified, or even that there does not exist an attraction to any of the asymptotic distributions. A pragmatical approach must then be tried.

Step 2 — Statistical Analysis of Models (SAM)

After *assuming* (i.e., having made the statistical choice of) one of the asymptotic distributions (of the univariate margins, in the multivariate case) the next essential step is to obtain estimators of the parameters. If the Gumbel distribution is assumed then $(\hat{\lambda}, \hat{\delta})$ have been obtained in the first step, and they can be used for the different problems described in Chapter V and possibly others. If the Gumbel distribution is not assumed (i.e., $\theta \neq 0$) the techniques relative to the assumption of the Fréchet distribution ($\theta > 0$) are described in Chapter VI and

those for the assumption of the Weibull distribution ($\theta < 0$) are dealt with (but for minima) in Chapter VII.

In all these cases, parameters and quantiles are estimated as well as probabilities of over- and underpassing threshold levels, and tests of hypotheses, prediction, discrimination, tolerance intervals and multisample analysis have also been described when possible.

Recall that if the location parameter is known we can easily reduce the Weibull and Fréchet models to the Gumbel one by exp/log transformations.

The tail estimation must be made according the third approach given in Annex IV.

Step 3 — Statistical Decision for Multivariate Models (SDM)

Let us suppose that we have vectors of extremes, whose margins have been studied (Step 1) and assumed to have one of the asymptotic distributions. The case for bivariate extremes is, in part, dealt with theoretically but the multivariate case is very partially covered.

Various bivariate models exist (Chapter X) but it *seems*, for the moment, that essentially the logistic and in second place the biextremal and natural models most often fit to the data; the last two are very easily generalized to the multivariate case.

In the bivariate case, the most important initial step is to test for independence as a reference pattern, although we can in general expect dependence. Once the dependence is confirmed and a *bivariate model assumed* we should estimate the dependence parameters and act accordingly. Chapter XI gives the essential formulae. But it should be recalled, if we are dealing with the multivariate case, that if there exists independence in the bivariate margins than we should always have independence, and we are returning to Step 2.

Sometimes, and at this state of knowledge, the estimation of the dependence function can be made intrinsically and compared with the supposed and assumed bivariate distribution, in an intuitive way.

For the multivariate models, after the bivariate studies are made, we can *try* to assume one the bivariate models for the corresponding margins. The logistic, the biextremal, and the natural model may be the best fit to data.

Step 4 — Stochastic Processes Analysis (SPA)

In principle, although it is not necessarily always the case, we may suppose that every stochastic process (of extremes), by a choice of observation times, is reduced to a convenient random sequence (originated from random or periodic sampling); then, with a discrete time data, we will follow the procedures for random sequences; otherwise we will follow directly the procedures for stochastic processes.

As strong assumptions are made about the margins in general Step 2 may be of interest.

As a first approach, after filing the data and ploting it against a time variable (the data may be a sequence or a process) we try to "see" if it follows globally one of the patterns described in Annex VI.

Then we try to fit of curvilinear regression to the observed sequence: linear in log n for extremal sequences, constant for EMS sequences, asymptotically linear in n or constant according to $a_0 > 0$ or $a_0 < 0$ for EME sequences, and linear in n for sliding sequences; recall that in all cases variance is constant. For stochastic processes fix a net of points, at equal steps, and consider the sequence of observations at each point of the net; it will be convenient to use various nets to *reinforce the assumed model.*

When dealing with random sequences:

a) if the extremal sequence model is assumed, use formulae of Chapter XIV;
b) if the EMS sequence model is assumed, use formulae of Chapter XV;
c) if the EME sequence model is assumed, use formulae of Chapter XV;
d) if the sliding sequence model is assumed, use formulae of Chapter XV.

When dealing with stochastic processes:

a) if the extremal process model is assumed, use formulae of Chapter XIV;
b) if the EMS process is assumed, use formulae of Chapter XV.

Chapter XVII

SOME CASE STUDIES

Introduction

We will deal here with the statistical analysis of data considered in Parts II (univariate extremes data) and III (multivariate extremes data); the situation related to stochastic processes and sequence of extremes (still in its infancy) must, in general, be connected to multivariate extremes data and was, in part, sketched in the Chapters of Part IV and Annex VI and Chapter XVI (*GEA*).

We will begin by analysing some univariate data and then, presupposing the analysis technique, deal with bivariate and multivariate data. We will not of course detail the computation of the estimators of parameters, of quantiles and of exceedance probabilities because they follow directly from the formulae and examples given in Parts II and III.

Maximum wind speed data in Lisbon

The data that follow correspond to maximum wind speeds in Lisbon between 1941 and 1970(*). The data for exploratory and analytical study are contained in Table 1.

TABLE I

Year	km/h	Year	km/h	Year	km/h
1941	129	1951	96	1961	86
1942	117	1952	72	1962	91
1943	100	1953	98	1963	96
1944	100	1954	85	1964	89
1945	132	1955	124	1965	90
1946	94	1956	108	1966	89
1947	108	1957	102	1967	89
1948	113	1958	102	1968	84
1949	96	1959	112	1969	107
1950	113	1960	107	1970	111

(*) The author thanks the Portuguese Institute for Meteorology and Geophysics for having kindly supplied the data.

a) Quick exploration of extremes data:

The calculation of the statistic

$$Q = \frac{x'_n - x'_{[n/2]+1}}{X'_{[n/2]+1} - X'_1} = \frac{x_{max} - x_{med}}{x_{med} - x_{min}},$$

as $x_{max} = 132$, $x_{med} = 100$, $x_{min} = 72$ gives

$$Q_{30} = 1.1428571.$$

As for $\theta = 0$ we have, for $n = 30$,

$$\beta_{30,0} = \frac{\log 30 + \log(\log 2)}{\log\log 30 - \log\log 2} = 1.9078381$$

and

$$\alpha_{30,0} = \frac{1}{\log\log 30} = .8169083,$$

we see that

$$\frac{Q_{30} - \beta_{30,0}}{\alpha_{30,0}} = -.9364343.$$

Recall that $(Q_n - \beta_{n,o})/\alpha_{n,o}$ has asymptotically a reduced Gumbel distribution if the i.i.d. sample has the same distribution we see that $(Q_{30} - \beta_{30,0})/\alpha_{30,0}$ is between the bounds for shortest 95% - probability interval for the Gumbel distribution and therefore we can, initially, assume it as the approximate distribution of maxima.

For the plotting we must know $x'_{[.367n]+1} = x'_{12} = 96$ and $x'_{[.692n]+1} = x'_{21} = 108$ and so the plot in the statistical choice modification of Gumbel probability papers gives the graph that follows, agreeing with the use of the Gumbel model.

Fig. XVII-1 – Plot of maximum wind speed at Lisbon (1941-70)

b) Analytical choice of the model:

Once having seen a general trend for the interpretation of data as (approximately) described by the Gumbel distribution, let us confirm (or reject) it by analytic study.

The ML estimators of (λ, δ) are

$$\hat{\lambda} = 94.71 \quad , \quad \hat{\delta} = 12.49 \, ,$$

and so the statistical choice statistic has the value

$$\hat{V}_{30} = -\, 9.63 \, .$$

Consequently, as $\hat{V}_n / \sqrt{2.09797 \times n}$ has the value,

$$\hat{V}_{30} / \sqrt{2.09797 \times 30} = -\, 1.21 \quad ,$$

which is in the acceptance region for the asymptotically normal distribution of $\hat{V}_n / \sqrt{n \, \hat{\sigma}_o^2}$,

thus confirming the exploratory analysis. For the value of $\hat{V}_n / \sqrt{n \, \hat{\sigma}_o^2} = -\, 1.21$, obviously, we do not need to go to the refined bounds which are, for $\alpha = .05$ and to the order $O(n^{-1})$, $b_{30} = -\, 1.71$ and $a_{30} = 2.91$, confirming the decision.

It would be useful to divide the sample into two parts (about 2/3 and 1/3) and confirm the results.

To a certain extent this was done in the papers by Tiago de Oliveira (1981) and Fransén and Tiago de Oliveira (1984) : in the former, data for greatest ages of death for men and for women in Sweden are studied for the years 1905/1958, and in the latter corresponding data for a larger set of years (1905/1970); and the results are coherent.

Other cases have been considered in the paper by Fransén and Tiago de Oliveira (1984) and can be considered as exercises. We will only follow another case leading to a different decision — the choice of the Fréchet distribution.

It would also be useful to read Whitmore *et al.* (1987) where data for maximum wind speeds are analysed for Canada.

Flood discharges of the North Saskachevan River at Edmonton

In this case we consider not only the raw data but, as a consequence of the rejection of the Gumbel distribution as a possible model, their logarithms. The ordered raw data, for a period of 47 years in 1000 ft^3/s , is (*)

TABLE II

i	x_i	i	x_i	i	x_i	i	x_i
1	19.885	13	30.380	25	40.400	37	61.740
2	20.940	14	31.500	26	42.250	38	65.440
3	21.820	15	32.600	27	44.020	39	65.597
4	23.700	16	32.680	28	44.730	40	66.000
5	24.888	17	34.400	29	44.900	41	74.100
6	25.460	18	35.347	30	46.300	42	75.800
7	25.760	19	35.700	31	50.330	43	84.100
8	26.720	20	38.100	32	51.442	44	106.600
9	27.500	21	39.020	33	57.220	45	109.700
10	28.100	22	39.200	34	58.700	46	121.970
11	28.600	23	40.000	35	58.800	47	185.560
12	30.200	24	40.400	36	61.200		

a) Quick exploration of extremes data:

We will avoid the graphical analysis, which is as before, but will only consider the statistic

$$Q_{47} = \frac{x_{max} - x_{med}}{x_{med} - x_{min}}$$

which as $x_{max} = 185.560$, $x_{med} = 40.400$ and $x_{min} = 19.885$ takes the value $Q_{47} = 7.0757982$; we have $\beta_{47,0} = 2.0317188$ and $\alpha_{47,0} = .7417784$. So

$$\frac{Q_{47} - \beta_{47,0}}{\alpha_{47,0}} = 6.7999815 ,$$

(*) Data taken from van Montfort (1970) .

thus rejecting the acceptance of the Gumbel distribution as an approximate model.

b) Analytical choice of the model:

The values of $(\hat{\lambda}, \hat{\delta})$ are $\hat{\lambda} = 38.151$ and $\hat{\delta} = 17.740$, thus giving $\hat{V}_{47} = 37.77531$ and then $\hat{V}_{47} / \sqrt{2.09797 \times 47} = 3.804$, outside the acceptance region of the asymptotically normal distribution of $\hat{V}_n / \sqrt{n \hat{\sigma}_0^2}$ at a level even smaller than 2‰. But this value leads to the Fréchet distribution in the statistical choice procedure.

c) Analysis of the logarithms of raw data :

The value of $(Q_{47} - \beta_{47,0})/\alpha_{47,0}$ and of $\hat{V}_{47} / \sqrt{2.09797 \times 47}$ suggest the relevance of trying the Fréchet distribution. Denoting by $'x = \log x$ we get $'x_{max} = 5.2234$, $'x_{med} = 3.6988$ and $'x_{min} = 2.9900$ and $'Q_{47} = 2.1509594$ and so $('Q_{47} - \beta_{47,0})/\alpha_{47,0} = .1607495$, which leads us to accept the Gumbel distribution for $\log x$ and thus suggests the use of the Fréchet approximation.

Let us confirm this analytically. We have $'\hat{\lambda} = 3.5489$, $'\hat{\delta} = .3963$, $'\hat{V}_{47} = -1.95849$ and $'\hat{V}_{47} / \sqrt{2.09797 \times 47} = -1.197$, thus leading to the acceptance of Gumbel distribution for the logarithms of raw data and thus the Fréchet distribution for raw data, as seen in b).

Flood discharges of the Fox River at Berlin and Wrightstown

In this analysis we will use flood data the Fox river at two points of its course: Berlin (upstream) and Wrightstown (downstream), both in Wisconsin, USA (*). The set of data of $n = 33$ pairs of observations is given in

(*) Data taken from Gumbel and Mustafi (1967).

TABLE II

Year	Berlin (1000 ft^3/s)	Wrightstown (1000 ft^3/s)	Year	Berlin (1000 ft^3/s)	Wrightstown (1000 ft^3/s)
1918	6.05	16.3	1934	1.91	6.7
19	2.67	13.1	35	4.34	11.1
20	5.15	16.6	36	4.34	6.3
21	2.45	14.2	37	3.26	13.5
22	5.92	20.1	38	6.19	18.0
23	6.05	13.7	39	4.91	18.2
24	4.02	15.5	40	4.72	17.5
25	2.52	8.3	41	3.54	16.6
26	3.44	9.1	42	2.74	19.8
27	3.17	13.3	43	5.08	21.3
28	5.92	15.1	44	2.29	10.8
29	6.62	20.6	45	3.46	15.8
30	3.00	6.6	46	6.90	21.3
31	1.14	3.1	47	3.16	11.0
32	1.91	9.9	48	4.54	10.3
33	2.60	8.9	49	2.00	6.4
			50	4.63	10.9

We can, evidently, expect dependence between the yearly flood discharges at Berlin (B) and Wrightstown (W). The graphs on the probability paper show that the Gumbel distribution fits well to the observed marginal flood discharges. The ML estimators are $\hat{\lambda}_B = 3.21070$, $\hat{\delta}_B = 1.33880$, $\hat{\lambda}_W = 10.89643$ and $\hat{\delta}_W = 4.61371$ and the empirical Kolmogoroff-Smirnov statistics are $KS_B = .07462$ and $KS_W = .09957$ with $\sqrt{n}\ KS_B = .429$ and $\sqrt{n}\ KS_W = .572$ accepting thus the Gumbel model for the margins. Evidently the better analysis is to use the $\hat{V}_n$ statistics for each margin .

The elements for our statistical decision are the values of the estimated correlation coefficient $\rho^* = .693$, the medians $\tilde{\mu}_B = 3.54$ and $\tilde{\mu}_W = 13.5$ and the number $N = 24$ of pairs (x_i , y_i) such that $x_i > \tilde{\mu}_B$ and $y_i > \tilde{\mu}_W$ (first quadrant) or $x_i < \tilde{\mu}_B$ and $y_i < \tilde{\mu}_W$ (third quadrant). This value of N is not coincident with the total of the first and third quadrants (22) given in Gumbel and Mustafi (1967) because the authors, instead of computing the medians directly from the sample, estimate the margin parameters and from them compute estimated medians. The test of independence , as $\sqrt{n}\ \rho^* = \sqrt{33} \times .693 = 3.981$ is larger than the values

$\lambda_{0.05} = 1.64$ and $\lambda_{0.01} = 2.33$, leads to rejection of independence at level 1%. Consequently we will have to estimate θ.

Let us consider first the mixed model. As the larger value of ρ^* in the mixed case is 2/3, we have no point estimator of θ by this method (the best being evidently $\theta^* = 1$, allowing for sampling errors).

The relation $p^{**} = p_M(\theta^{**}) = 2^{-2+\theta^{**}/2} = \frac{N}{2n} = \frac{24}{2x33} = .36364$ gives $\theta^{**} = 1.081$ and thus is not solvable for θ^{**} as we should have $0 \le \theta^{**} \le 1$; also note that the maximum value of $p_M(\theta)$ is $p_M(1) = .35355 < \frac{24}{2 \times 33} = .36364$. The mixed model must therefore be rejected.

Consider now the logistic model. As $0 \le \rho_L(\theta) \le 1$ the equation $\rho_L(\theta^*) = \theta^*(2 - \theta^*) = .693$ can be solved giving $\theta^* = .446$. As $p_L(\theta) = 2^{-2^{1-\theta}}$, the equation $p^{**} = p_L(\theta^{**}) = \frac{N}{2n} = .36364$ gives $\theta^{**} = .4546$, which is close to the previous estimate, justifying to a certain extent the use of the logistic model.

Note that this is a margin (location-dispersion) parameter-free approach. As said, Gumbel and Mustafi (1967) used estimators of the margin parameters and computed the *estimated* reduced difference, comparing it, by the Kolmogorov-Smirnov test, with the logistic and confirming the suggestion to use the logistic model as an approximation.

Recall that estimation of the margin parameters involves as yet unsolved problems with the use of the Kolmogorov-Smirnov test, although it gives a hint for the logistic model, as stated.

The 5% asymptotic confidence interval is given by

$|p_L(\theta) - p^{**}| \le 1.96 \sqrt{p^{**}(.5 - p^{**})} / \sqrt{n}$ which is, in our case,

$|p_L(\theta) - .36364| \le .07598$ or $.28766 \le p_L(\theta) \le .43962$ and so

$.154 \le \theta \le .754$, a very large (asymptotic) confidence interval.

Let us now consider the biextremal model. The equation $\rho_B(\theta^*) = .693$ gives $\theta^* = .55$. The solution from $p_B(\theta^{**}) = 2^{-2+\theta^{**}} = .36364$ is .541; the approximate confidence interval for the significance level 5% is $|p_B(\theta) - .36364| \le .07598$ is $.202 \le \theta \le .814$.

For the Gumbel model the equation $\rho_G(\theta^*) = .693$ gives $\theta^* = .723$ and the confidence interval from $p_G(\theta) = p_B(\theta)$ is also $.202 \leq \theta \leq .814$, which suggests that the Gumbel model is not appropriate to this problem, the solution of $\rho(\theta^*) = .6846$ being close to the upper bound of the confidence interval; in fact the model arises naturally in questions connected with failures. A simple check can be the estimation of θ from the use of Kendall's $\tau_G = \theta/(2 - \theta)$.

The question of discriminating between models, as for the remaining ones (logistic, biextremal) in this example is, as yet, an unsolved problem. Here, though, the solution is simple.

As in the biextremal model we have for the reduced values $Y = X$ with probability θ and $Y > X$ with probability $1 - \theta$, we can expect that, roughly, $n\theta$ observations (x_i, y_i) should be in a straight line $y = ax + b$ and rest should be such that $y > ax + b$. In our case, as $\theta^* = .541$ and $n = 33$ we should have roughly $.541 \times 33 \approx 18$ points in a straight line; the scatter diagram given in Gumbel and Mustafi (1967) shows immediately that this is not the case. Of the four models considered, the only one not yet rejected is, for the moment, the logistic. Also the plotting of *estimated* reduced differences on the logistic paper, given also in Gumbel and Mustafi (1967), speaks in favour of the logistic model.

The application of the strip method — see Posner *et al.* (1969) referred to in Chapter — to the logistic model as a flood model for the Fox River confirms its usefulness.

It should be remarked here that if we *expect* bivariate models to be absolutely continuous here, as in other cases, the logistic model seems to be a good candidate. Also, as shown in Tiago de Oliveira (1987) the discrepancy between the logistic and (2,N) natural model is small and not easily seen for such small samples. The intrinsic estimation will also not give large discrepancies.

Exercises — Part V

V.1—Study the data given in Sneyers (1977) for early maximal precipitations in mm at Uccle for durations of 20 min, 30 min, 2h and 6h (samples of 17 observations):

Year	20 min	30 min	2 h	6 h
1956	14.0	16.2	26.7	32.9
1957	10.2	10.4	19.0	27.0
1958	8.4	9.5	18.5	22.7
1959	10.6	11.2	15.2	18.6
1960	14.4	15.3	15.4	18.3
1961	15.5	17.6	20.1	20.1
1962	19.5	23.5	48.5	48.5
1963	11.8	11.9	16.0	30.5
1964	13.2	13.3	25.3	25.3
1965	15.0	15.0	15.8	25.8
1966	10.2	10.2	15.5	25.3
1967	13.0	13.0	13.4	15.5
1968	9.0	11.4	17.1	17.5
1969	21.0	21.5	31.8	49.0
1970	17.1	17.1	20.0	27.9
1971	10.5	14.0	21.8	41.0
1972	13.0	15.5	20.8	21.7

V.2—Study some of the data contained in the references to Part V (and in particular the lag-correlation) as well as data contained in Parts II, III and IV.

V.3—For Sneyers (1977) data (Chapter V) compare the estimators obtained for the full sample (1938/72) and a partial one (1956/72) like the one given above. Compare the two partial samples (1956/72) and discuss if there is an average relation between the durations and the amount of maximal precipitations.

A. FRANSÉN and J. TIAGO DE OLIVEIRA (1984). Statistical choice of univariate extreme models, Part II, *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 373-394, D. Reidel, Dordrecht

E.J. GUMBEL and C. K. MUSTAFI (1967). Some analytical properties of bivariate extremal distributions, *J. Amer. Statist. Assoc.*, 62, #318, 569-588

EDWARD C. POSNER, EUGENE R. RODEMICH, JONH C. ASHLOCK and SANDRA LURIE (1969). Application of an estimator of high efficiency in bivariate extreme value theory, *J. Amer. Statist. Assoc.* , 64, # 328, 1403-1414

R. SNEYERS (1977). L'intensité maximale des précipitations en Belgique, *Inst. Royal Météor. Belgique*, ser. B, nº 86

J. TIAGO DE OLIVEIRA (1981). Statistical choice of univariate extreme models, *Statistical Distributions in Scientific Work*, vol. 6, G.P. Patil ed., 367-387, D. Reidel, Dordrecht

J. TIAGO DE OLIVEIRA (1987). Comparaison entre les modèles logistique et naturel pour les maxima et extensions, *C. R. Acad. Sc. Paris*, t. 305, ser.I, 481-484

M.A.J. VAN MONTFORT (1970). On testing that the distribution of extremes is of type I when type II is the alternative, *J. Hydrology*, 11, 421-427

G.A. WHITMORE *et al.* (1987). Case studies in data analysis, nº 5, *Canad. J. Statist.*, 15, #4, 311-337

ANNEX

Figures referred in the text

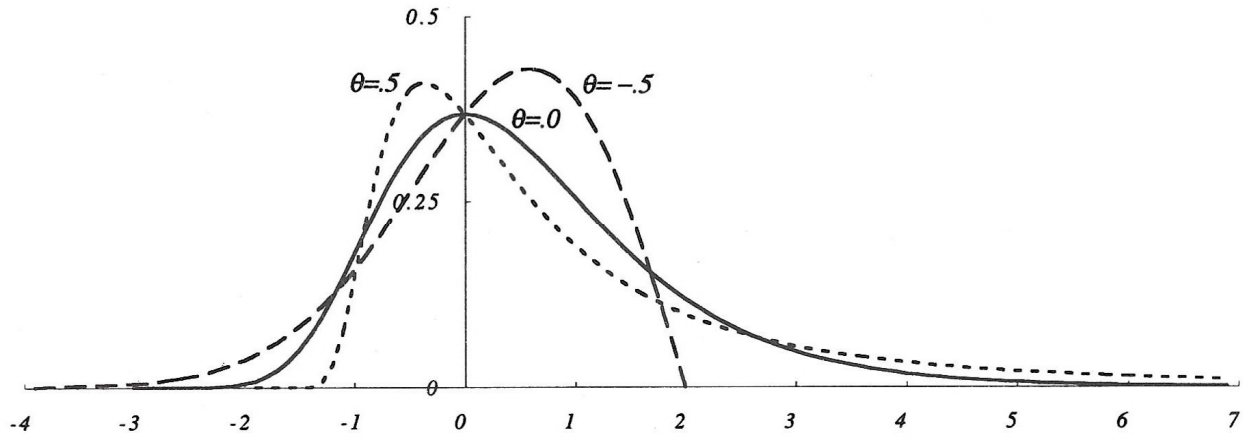

Fig. II.1 — p. 45

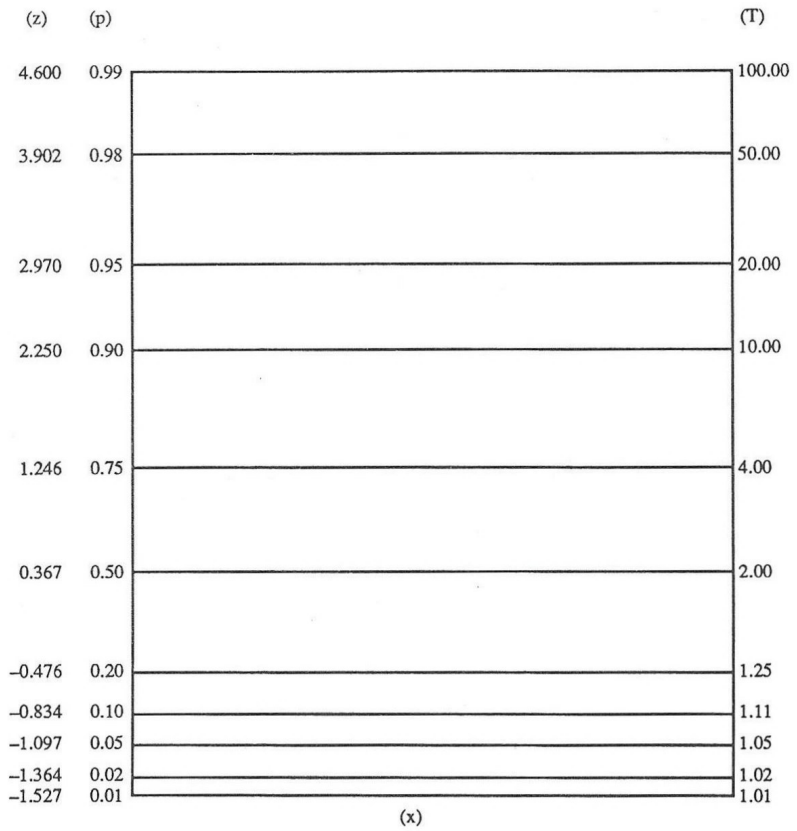

Fig. IV.1 — p. 90

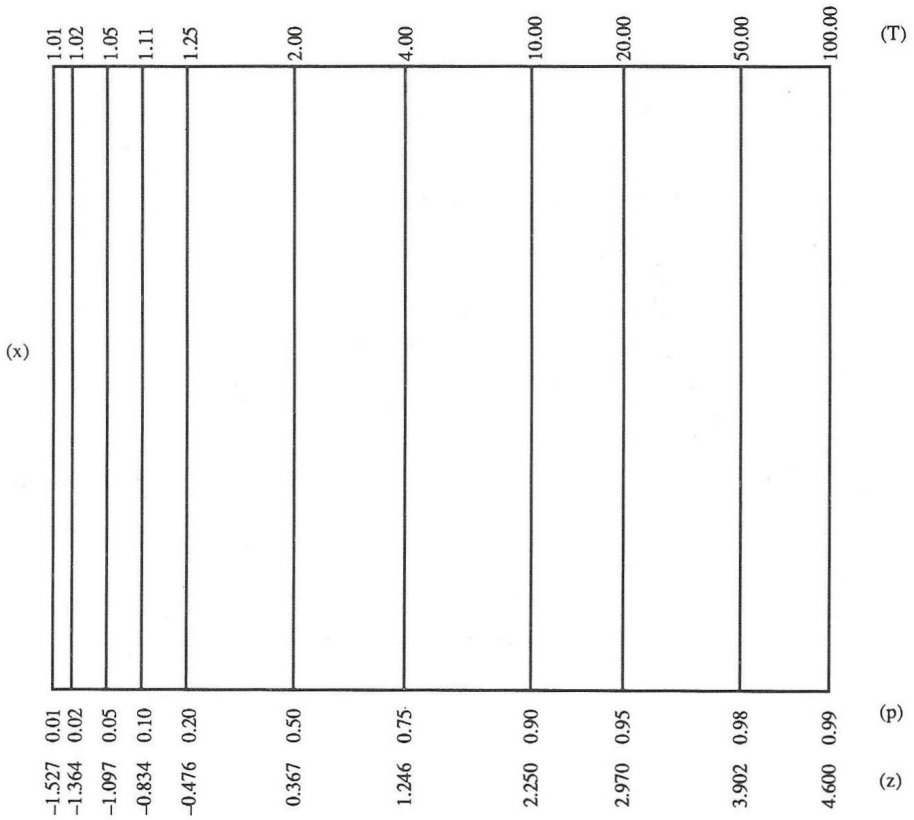

Fig. IV.2 — p. 91

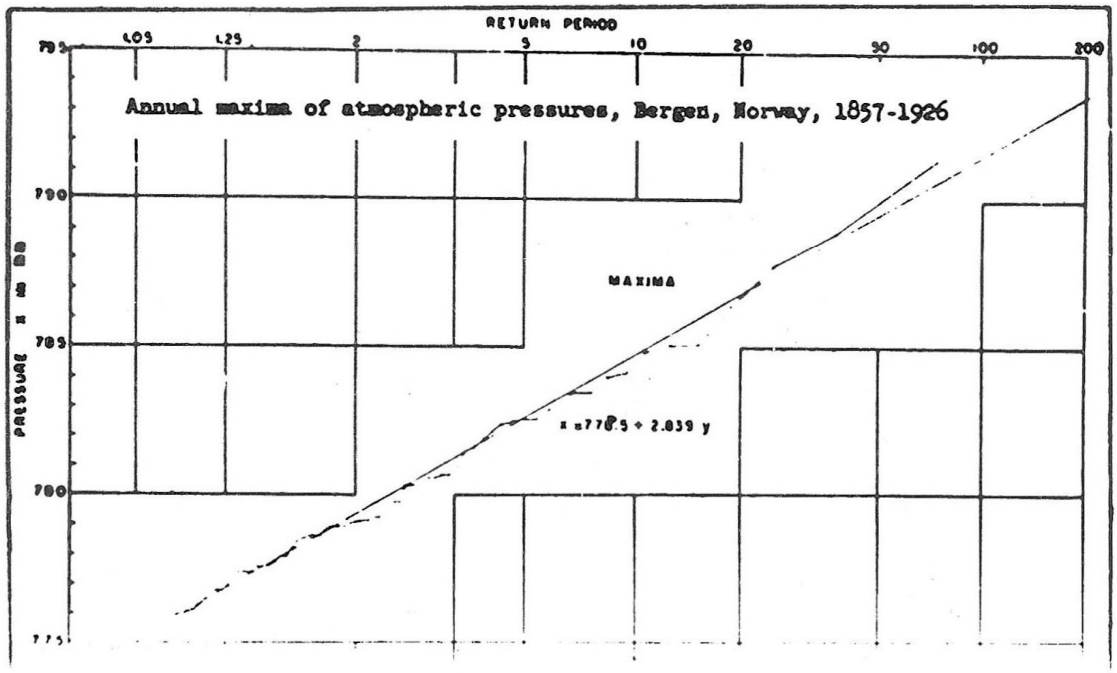

Fig. IV.3 — p. 92

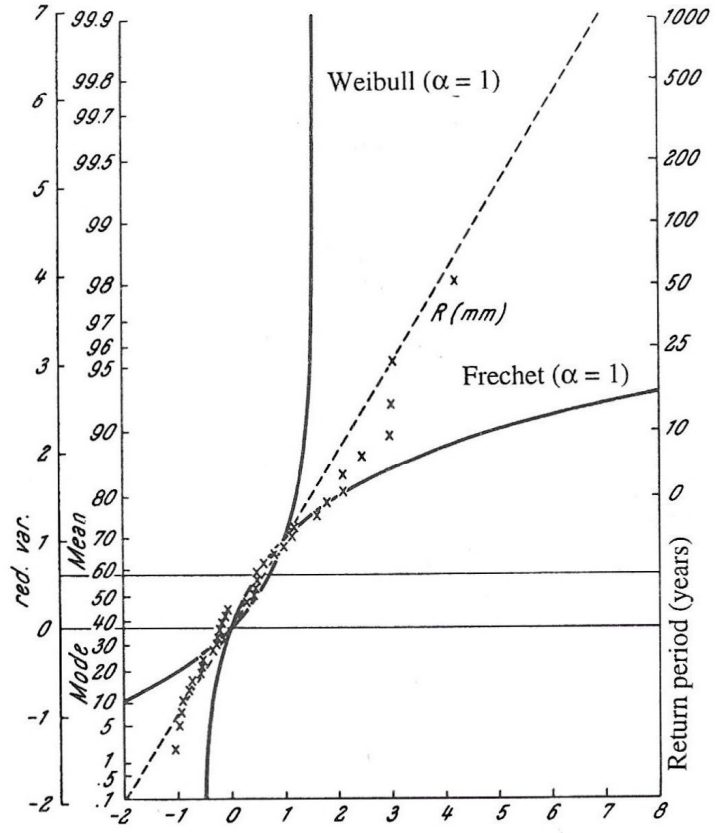

Fig. IV.4 — p. 95

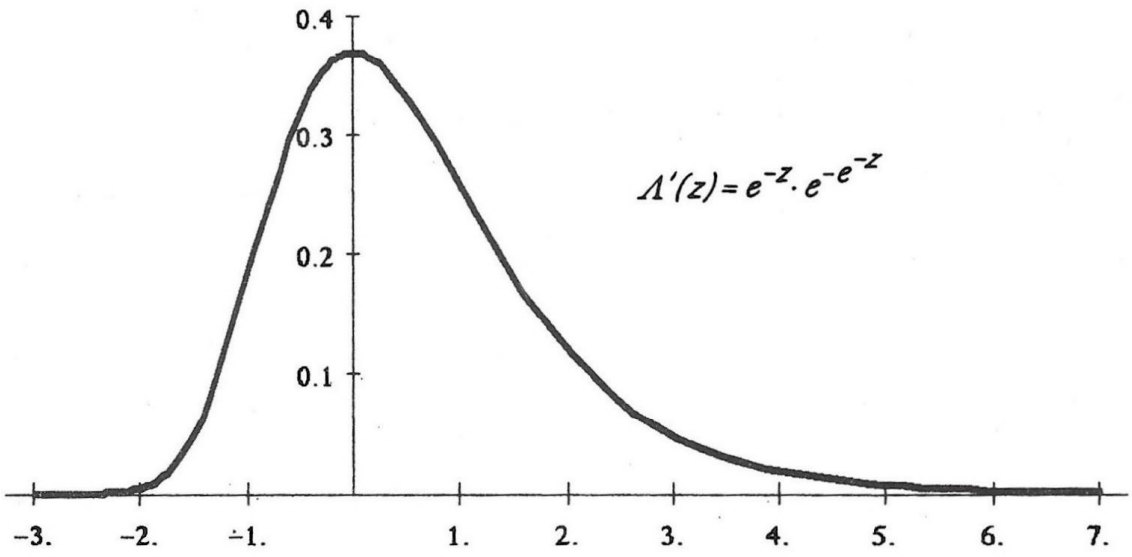

Fig. V.1 — p. 98

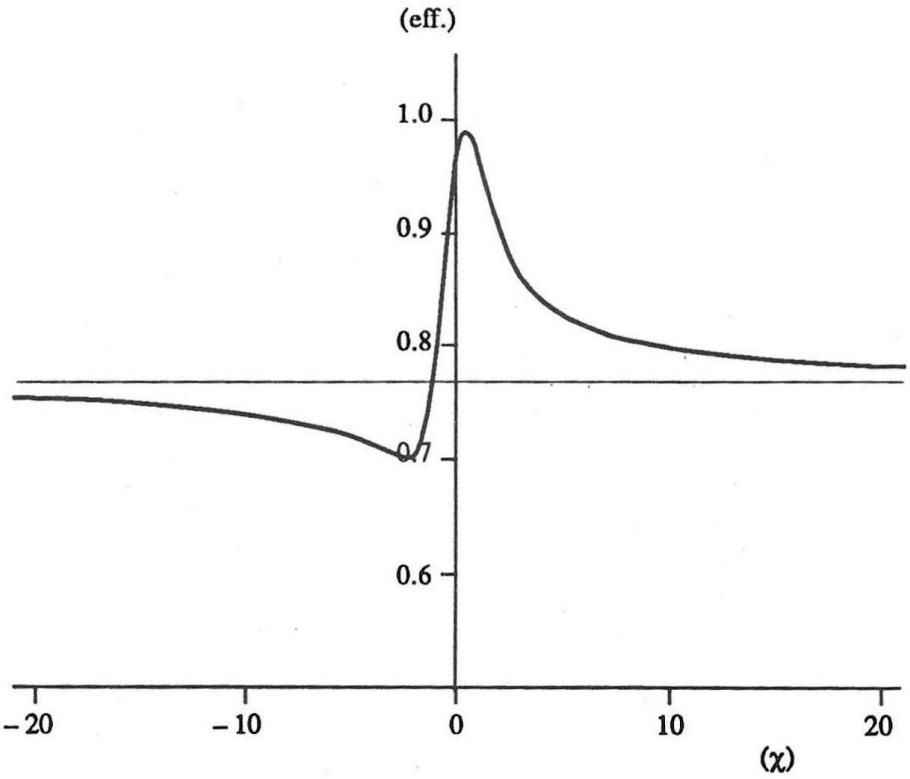

Fig. V.2 — p. 107

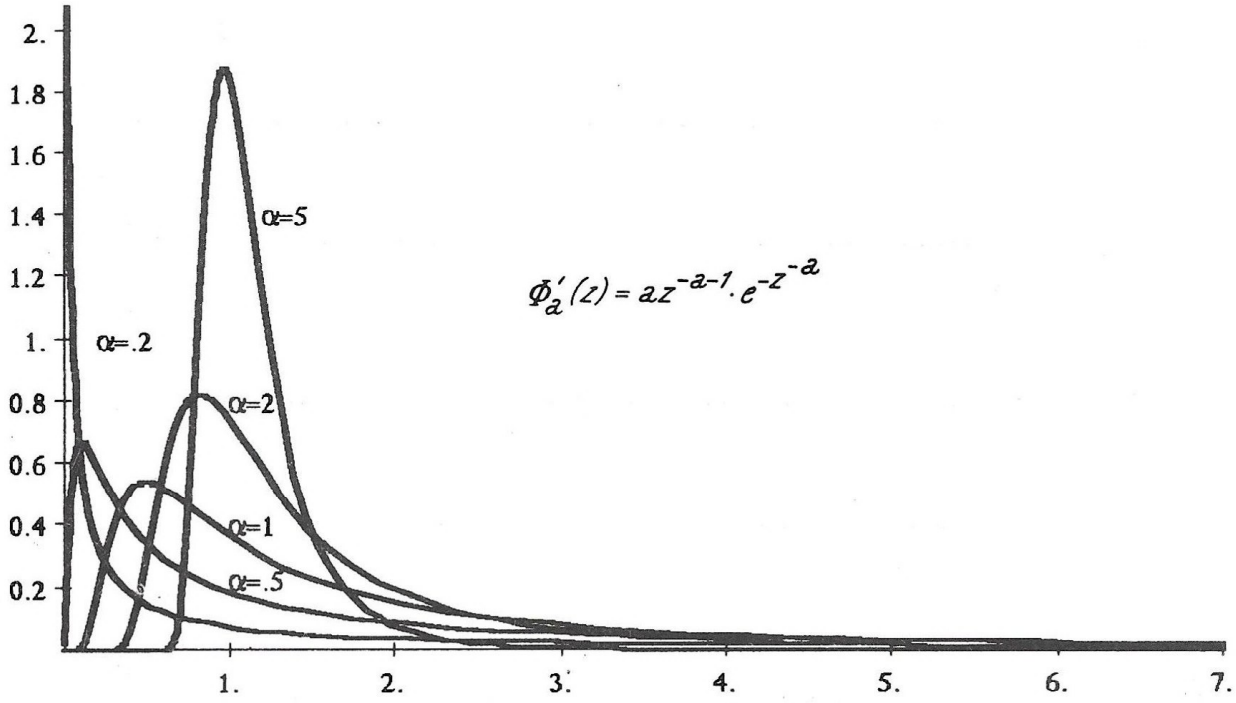

Fig. VI.1 — p. 126

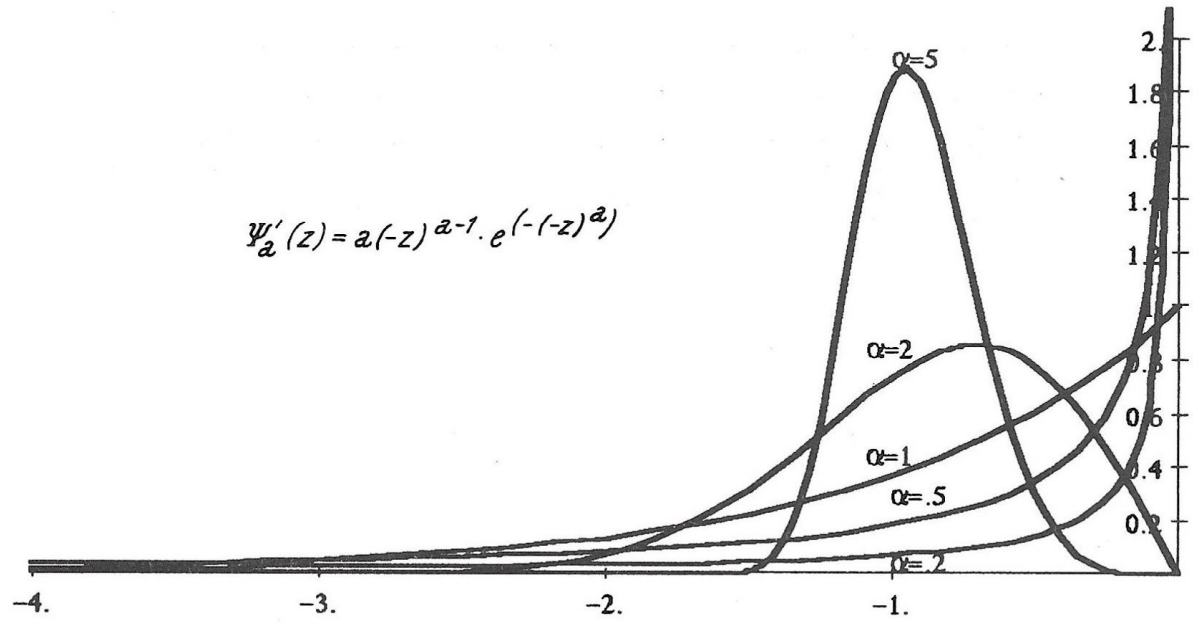

Fig. VII.1 — p. 135

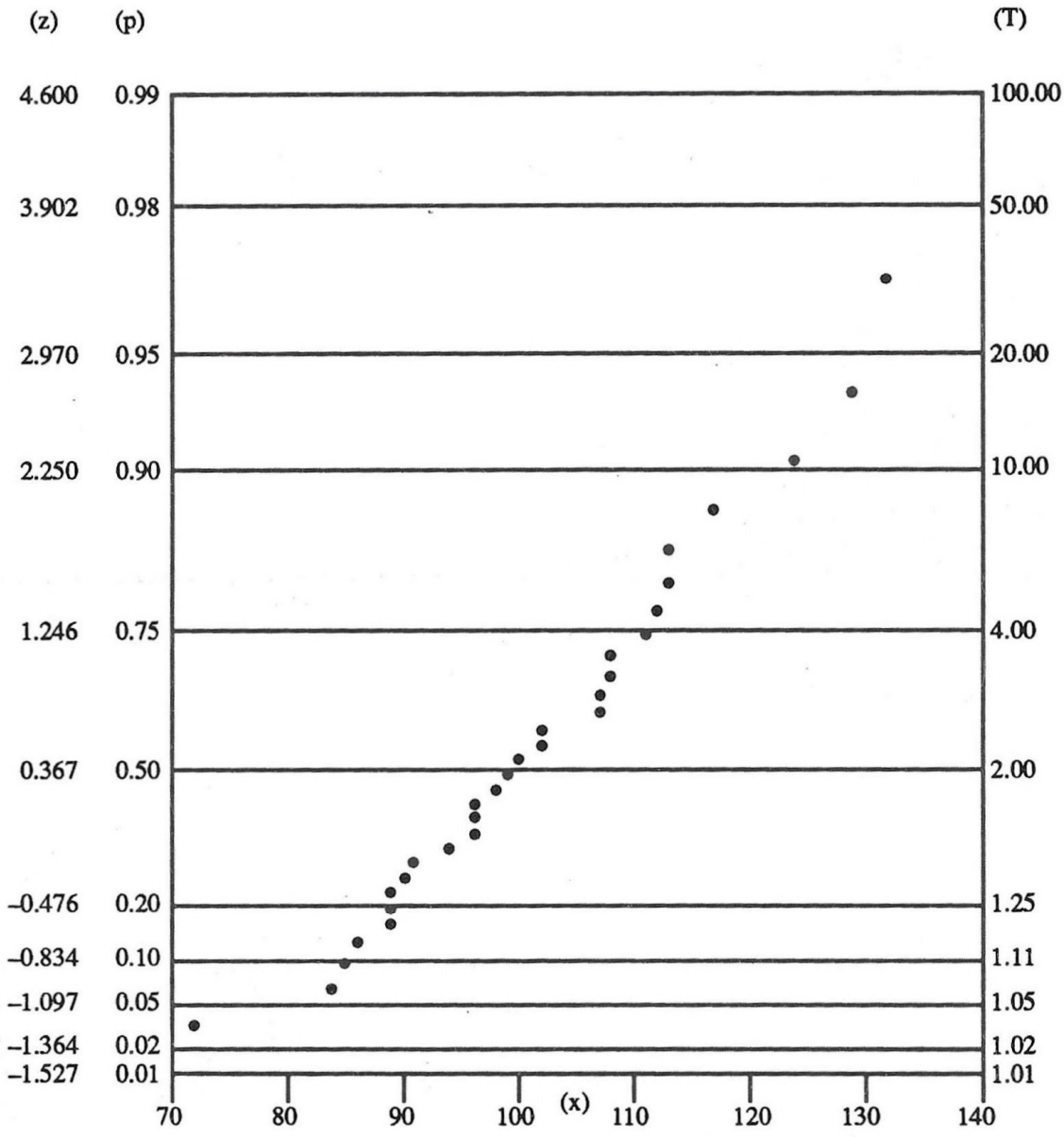

Fig. XVII.1 — p. 422

ERRATA

This book is a fac-simile of the Author's copy of the manuscript. The figures quoted in the text were not inserted in it. Therefore, they are not included in this edition.
The absence of pp. 378-379 in the book is due to a page of the manuscript having been misplaced. That page would have lain between lines 11-12 at the bottom of p. 377.

Lisboa, Junho de 1997

Colibri – Artes Gráficas
Faculdade de Letras
Alameda da Universidade
1699 Lisboa Codex
Telef. / Fax 796 40 38