

**SOCIEDADE PORTUGUESA
DE ESTATÍSTICA**

Publicação semestral

Primavera de 2007



A “Escola de Extremos” em Portugal

Não há caminhos, há que caminhar Andreia Hall	15
Acerca de Testes Estatísticos para Valores Extremos Isabel Fraga Alves	20
Models for Spatial Extremes P. de Zea Bermudez, J. Mendes and K. F. Turkman	27
Semiestabilidade em Teoria de Extremos Maria da Graça Temido	32
Memorial da Escola Maria Ivette Gomes	37

Editorial	2
Mensagem do Presidente	3
Notícias	5
Episódios na História da Estatística	14
Ciência Estatística	
• Livro	52
• Artigos Científicos Publicados	52
• Teses de Mestrado	52
• Teses de Doutoramento	53
• Prémio SPE 2007.....	55
• Prémio Estatístico Júnior 2007	56

Informação Editorial

Endereço: Sociedade Portuguesa de Estatística.
Campo Grande. Bloco C6. Piso 4.
1749-016 Lisboa. Portugal.

Telefone: +351.217500120

e-mail: spe@fc.ul.pt

URL: <http://www.spestatistica.pt>

ISSN: 1646-5903

Depósito Legal: 249102/06

Tiragem: 1000 exemplares

Execução Gráfica e Impressão: Gráfica SobreireNSE

Editor: Fernando Rosado, fernando.rosado@fc.ul.pt

Este Boletim tem o apoio da **FCT** Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Editorial

... partilhemos informação...

O Boletim SPE Outono 2006, da Sociedade Portuguesa de Estatística, foi enviado para todos os sócios – honorários, ordinários e colectivos e também para diversas associações congéneres. Assim, a edição com 1000 exemplares foi amplamente divulgada. Os poucos exemplares que restam podem ainda ser remetidos para alguém ou alguma instituição que os nossos leitores considerem que deva ser eleita e para a qual confirmem que a nossa edição não foi enviada. Enviem as vossas sugestões.

Sobre aquele Boletim, a Opinião dos Leitores foi bastante favorável. É devido um agradecimento aos autores das diversas mensagens. Deseja-se uma participação ainda mais activa dos sócios, com “textos de opinião” para além das (sempre estimulantes!) palavras de congratulação. Como foi referido por alguns sócios, a chegada do Boletim foi também um alerta para encontrar a resposta à pergunta: Quantas quotas tenho em atraso? De facto, ao receber o Boletim de Outono anuncia-se que “o ano já lá vai” e a respectiva quota de sócio poderá ficar em dívida. Por sua vez o Boletim da Primavera desperta o “tempo do pagamento” com vista à participação a preço especial no Congresso anual que se avizinha e ainda, neste ano, com especial importância pela possível acumulação ISI 2007 com XV Congresso SPE.

Mostrando o seu exemplar também cada sócio pode ajudar a trazer mais um sócio... e, se considerar necessário, pode solicitar que lhe seja enviado algum exemplar para oferta (se o stock ainda permitir essa acção...).

Assim, o Boletim alerta cada membro da SPE para o cumprimento dos seus deveres estatutários. Mas, também deve ser usado como seu instrumento para plenamente usufruir dos seus direitos... e uma das maneiras é partilhando ideias e opiniões além de toda a informação estatística fundamental.

Enviem as vossas mensagens para o editor... Partilhemos toda a informação!

O tema central deste Boletim de Primavera é a **Estatística de Extremos** que, com verdadeira propriedade deve intitular-se “Escola de Extremos” em Portugal.

De facto, este Boletim narra os últimos 25 anos da *estatística de extremos* no nosso país.

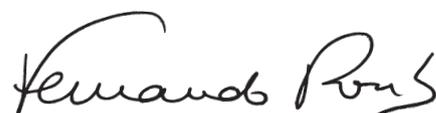
De entre as diversas áreas desse importante tema, alguns dos principais investigadores apresentam-nos o *estado da arte* que, em Portugal, tem Ivette Gomes como a grande obreira do sucesso “extremista neste extremo da Europa”.

Nesta edição é iniciada uma nova secção. Intitulada **Episódios na História da Estatística**, deseja-se que seja simultaneamente testemunho e contributo para a construção de uma história da Estatística em Portugal. Nela podemos relatar breves contribuições que entendamos relevantes para esse objectivo. Como “primeiro episódio” vamos eleger o antecessor deste Boletim SPE – o Boletim Informativo de Estatística e Investigação Operacional. Esse facto, de algum modo, também relata um início, um ponto de partida de um projecto científico envolvendo a comunidade dos estatísticos portugueses que começava a nascer.

Recordamos pioneiros!

Relevamos esse despertar de novos projectos científicos para Portugal! Há 28 anos.

É bem vinda toda a contribuição que os sócios desejem enviar para esta nova secção do Boletim.



Mensagem do Presidente

O Relatório de Actividades de 2006, presente à Assembleia Geral, relata o que de mais importante se passou na SPE durante este primeiro ano de mandato da actual Direcção. Aqui, por limitações de espaço, venho dar-vos conta telegraficamente dos acontecimentos de maior relevo.

Foram constituídas as Comissões Especializadas de Educação e de Nomenclatura Estatística. A primeira supervisionou já a revisão científica de alguns manuais escolares dos ensinos básico e secundário, serviço que a SPE pôs à disposição das editoras interessadas. A segunda está em fase avançada de elaboração de um glossário estatístico em língua portuguesa, para o que se conta também com a colaboração da Associação Brasileira de Estatística.

Realizou-se na Covilhã, com grande sucesso, o XIV Congresso Anual da SPE, a cuja Comissão Organizadora presidiu a Maria Eugénia Ferrão. O livro do minicurso, “Outliers em Dados Estatísticos”, da autoria do Fernando Rosado, foi publicado nessa ocasião. Também ficaram prontas nessa ocasião as Actas do XIII Congresso. O Prémio SPE 2006 foi entregue durante o Congresso à vencedora Sandra Dias.

O XV Congresso (19 a 21 de Agosto de 2007), a realizar em associação com a 56ª Sessão do International Statistical Institute (22 a 29 de Agosto de 2007), continua a ser activamente preparado, sendo a Comissão Organizadora presidida pela Manuela Magalhães Hill. Como já sabem, a Sociedade Portuguesa de Estatística negociou com a Organização do ISI 2007 condições muito vantajosas de participação conjunta nos dois eventos (ver <http://www.spe2007.dmq.eg.iscte.pt>). O tempo passou e é agora altura de tratarem do envio dos resumos das vossas comunicações e de procederem à vossa inscrição. Apelamos à vossa participação massiva e à capacidade de mobilizarem outros participantes na área da Estatística ou das suas variadas aplicações, contribuindo para o sucesso destes eventos e para a afirmação internacional da comunidade estatística portuguesa.

Outra realização importante foi a retoma da publicação do nosso Boletim, sendo este o segundo número sob a supervisão do Fernando Rosado. O seu sucesso depende da vossa colaboração.

Os Seminários com Novos Doutores têm vindo a decorrer na sede da SPE.

Em termos do funcionamento interno da SPE, fizemos as alterações da organização contabilística e documental necessárias à regularização da nossa situação fiscal e estamos a prosseguir o trabalho de reorganização da nossa página web, com algumas dificuldades logísticas que têm sido difíceis de ultrapassar e para as quais pedimos a vossa compreensão.

A SPE tem colaborado com inúmeras outras organizações na divulgação por e-mail de reuniões científicas, seminários (alguns dos quais tiveram mesmo o seu apoio institucional), cursos de verão, bolsas, recrutamento de profissionais estatísticos e outras matérias de interesse para os sócios.

Infelizmente, os e-mails chegam a uma parte reduzida dos sócios por falta de actualização das respectivas fichas. Pedia-vos o favor de procederem a essa actualização, preenchendo e devolvendo o impresso

que está a ser distribuído juntamente com a ficha de pagamento da quota de 2007.

Estão em fase de finalização protocolos de colaboração com a CLAD (Associação Portuguesa de Classificação e Análise de Dados) e a CienciaMetrics. Ainda que sem protocolo, desenrola-se activamente a cooperação com a SPM (Sociedade Portuguesa de Matemática), com o INE (nos nossos Congressos, no ISI 2007 e noutras matérias) e esperamos brevemente vir a colaborar com a Royal Statistical Society. Colaborámos com o Ministério da Educação na discussão do Plano de Acção Nacional para a Matemática. Estabelecemos também acordos com a Porto Editora para a revisão de manuais escolares e para patrocínio dos Prémios Estatístico Júnior, que vão ser relançados.

Com o CIM (Centro Internacional de Matemática) acordámos a realização dos Encontros de Estatística SPE-CIM, os quais têm a colaboração do INE. Em 2007 vão realizar-se dois Encontros:

- Probability and Statistics in Telecommunications (19 de Maio, organizado por António Pacheco, do IST), cujo programa estará brevemente na nossa página web;
- Methodological Issues in Official Statistics (17 de Novembro, organizado por Pedro Corte Real, da UNL e do INE).

Aproveitamos para agradecer às várias instituições que têm apoiado a SPE ou os seus Congressos.

Aos sócios, cujo apoio tem sido fundamental para a actividade da SPE, particularmente os que aceitaram desempenhar várias incumbências a pedido da Direcção, o nosso muito obrigado.

Saudações cordiais,



Notícias

• XV Congresso SPE



Caros Colegas,

No próximo verão, em Agosto,
Não finja estar indisposto
Nem arranje desculpa
(Mais tarde vai pensar “mea culpa”).

Venha ao Congresso da SPE
Em Lisboa, no ISCTE
Onde as técnicas estatísticas
Nem sempre probabilísticas
Serão apresentadas e discutidas
Em sessões descontraídas.

Participe activamente!
Divirta-se e fique bem disposto
E lembre-se que em Agosto
Todas as praias têm demasiada gente...
Os Congressistas e Seus Trabalhos

Manuela Magalhães Hill

Mais informações em: <http://spe2007.dmq.eg.iscte.pt>

• XV Congresso SPE e 56ª Sessão do International Statistical Institute

Neste ano, realizam-se em Lisboa dois eventos de relevo para a Comunidade Estatística portuguesa: a 56ª Sessão do International Statistical Institute – ISI 2007, a grande reunião científica internacional da Estatística (cerca de 3000 participantes), que decorrerá no Centro de Congressos de Lisboa de 22 a 29 de Agosto, e o XV Congresso da Sociedade Portuguesa de Estatística, que decorrerá no ISCTE de 19 a 21 de Agosto.

São duas oportunidades de encontro com especialistas de renome nas diversas áreas da Estatística que por certo permitirão a partilha de experiências e o enriquecimento científico de todos nós, além de contribuir, com a participação massiva dos estatísticos portugueses a que apelamos, para a afirmação internacional da nossa comunidade.

A Sociedade Portuguesa de Estatística negociou com a Organização do ISI 2007, de que também faz parte, condições vantajosas de participação nos dois eventos, quer em termos de valor de inscrição quer em termos de condições de alojamento. Os interessados em alojamento em Lisboa durante o período do Congresso podem beneficiar dos custos reduzidos negociados com os hotéis associados à organização do ISI.

Os sócios e outros participantes que queiram usufruir do combinado ISI+SPE (com custos de inscrição conjunta muito vantajosos) têm de fazer um pré-registo de inscrição, através da página do ISI 2007: www.isi2007.com.pt, mas (e isso é essencial) devem efectuar o pagamento da inscrição conjunta nos dois eventos exclusivamente à SPE, que posteriormente o comunicará à Organização do ISI 2007. Mais informações podem ser obtidas no site do Congresso: www.spe2007.dmq.eg.iscte.pt

Em virtude da antecipação da data de realização do Congresso da SPE, a data limite para envio de resumos foi também antecipada para *20 de Abril de 2007*.

• Advances in Semiparametric Methods and Applications

No âmbito da 56ª Sessão do ISI 2007 vai decorrer a conferência satélite em *Advances In Semiparametric Methods And Applications*, nos dias 20 e 21 de Agosto nas instalações do ISEG – UTL (Instituto Superior de Economia e Gestão). Este evento tem a tutela da sociedade de Bernoulli e o patrocínio do CEMAPRE (Centro de Matemática Aplicada à Previsão e Decisão Económica), centro de investigação do ISEG-UTL.

Os objectivos deste encontro compreendem a exploração dos princípios e das técnicas mais recentes relacionadas com o conceito de smoothing numa abordagem semiparamétrica, focando aplicações nas finanças, micro-econometria, séries temporais, análise espacial, entre outras. Espera-se um programa científico variado e estimulante reflectindo a multidisciplinaridade dos métodos semiparamétricos e das suas aplicações, reunindo investigadores nacionais e internacionais de referência, sendo de destacar a presença de Peter Bickel, Adónis Yatchew, Matt Wand e Wolfgang Härdle.

No programa constam ainda sessões com outros oradores convidados abordando os seguintes temas: *Semiparametric metric views on security and risk; Non- and Semiparametric methods for non-stationary data; Non-and Semiparametric estimation of frontiers and boundaries.*

As submissões de artigos podem ser feitas até 10 de Abril de 2007. Mais informações sobre este evento podem ser encontradas no endereço <http://asma.iseg.utl.pt>.

Isabel Proença

• IASC 07

A **IASC 07** é uma conferência satélite da *56th Session do ISI - International Statistical Institute*, promovida pelo *IASC - International Association for Statistical Computing*, uma das secções do *ISI*.

A **IASC 07** decorrerá na Universidade de Aveiro, de **30 de Agosto a 1 de Setembro de 2007**.

O objectivo do encontro é estimular a interacção de investigadores em tópicos como Estatística Computacional, *Data Mining*, *Knowledge Discovery* e *Statistical Learning*.

Está confirmada a participação de Mário Figueiredo (Portugal), Donato Malerba (Itália), Gilbert Ritschard (Suíça) e Arno Siebes (Holanda), como especialistas convidados. A conferência incluirá também sessões temáticas e comunicações livres, orais ou em poster.

Os interessados podem já efectuar o seu registo on-line <http://www.mat.ua.pt/iasc07/>. Os sócios da SPE beneficiam de redução na taxa de inscrição.

Os autores de comunicações livres que pretendam submeter os seus artigos para publicação no CD-ROM da conferência, devem fazê-lo **até 28 de Fevereiro de 2007**, inserindo os respectivos ficheiros no sistema "*PAPER SUBMISSION*". Alguns artigos incluídos no CD-ROM serão seleccionados para posterior desenvolvimento e publicação numa edição especial da revista *Computational Statistics & Data Analysis*.

Está previsto o alojamento de participantes em residências universitárias ou em hotéis da cidade. Todos os pedidos de reserva deverão ser directamente dirigidos à *Top Atlântico Operated by TopTours*.

Datas Importantes:

Submissão de artigos: 28 de Fevereiro de 2007

Inscrição de taxa reduzida: 31 de Março de 2007

Notificação de aceitação: 2 – 8 de Maio de 2007

Inscrição regular: 31 de Maio de 2007

Conferência **IASC 07**: 30 de Agosto – 1 de Setembro de 2007

Para mais informações: <http://www.mat.ua.pt/iasc07/> ou iasc07@mat.ua.pt

Manuela Souto

• Probability and Statistics in Science and Technology

Realizar-se-á nos dias 30 de Agosto a 1 de Setembro de 2007, no Porto, na FEUP, o congresso “Probability and Statistics in Science and Technology”. Este congresso tem como objectivo promover investigação que contribua para encontrar respostas às necessidades de resolver problemas práticos e de desenvolvimentos nas tecnologias e em diferentes áreas científicas. O encontro centrar-se-á em 3 tópicos principais: Estatísticas espaço-temporais, Simulação e Classificação. No entanto, contribuições de outras áreas da Estatística e das Probabilidades são bem vindas e esperadas. O congresso será presidido pelo Professor Doutor Holger Rootzen, da Universidade de Tecnologia de Chalmers (Suécia), e contará com a palestra de abertura proferida por Sir David Cox e palestras convidadas proferidas por Peter Diggle (U. Lancaster), Gareth Roberts (U. Lancaster) e Trevor Hastie (U. Stanford). Sessões temáticas têm vindo a ser organizadas e estão a cargo de Domenico Marinucci (U. Roma Tor Vergata), Aila Särkkä (U. de Tecnologia de Chalmers), Christophe Andrieu (U. Bristol), Søren Asmussen (U. Aarhus), Sara van de Geer (ETH Zurich) e Gabor Lugosi (U. Pompeu Fabra). Estão previstas sessões de contribuições orais ou em formato de poster, para as quais se deseja, para além da habitual participação estrangeira num encontro de cariz internacional, uma forte participação também da comunidade científica portuguesa que se dedica a esta área.

Data limite de submissão de contribuições: 15/Abril/2007.

Mais informações: <http://paginas.fe.up.pt/~bsconf07/>.

Paula Milheiro

• Conferência IASE - Assessing Student Learning in Statistics

Esta conferência satélite inserida no encontro ISI 56 é sobre o tema Assessing Student Learning in Statistics e é organizada pela International Association for Statistical Education, IASE. Realizar-se-á de 19 a 21 de Agosto de 2007 em Guimarães, ocorrendo imediatamente antes da 56ª Sessão do ISI em Lisboa. Será uma perfeita oportunidade para desfrutarem de apresentações realizadas por diversos indivíduos cujo elo comum entre eles é um interesse especial na avaliação de métodos de aprendizagem em estatística. Desta forma, as apresentações incluirão as mais diversas discussões nesta área de estudos, pelo que, poderemos encontrar nesta conferência artigos relevantes em muitos dos aspectos principais na avaliação dos métodos de aprendizagem em estatística, como por exemplo, como escrever de maneira eficiente questões para exame, estratégias de implementação de exames e métodos alternativos de avaliação, tais como, projectos, trabalhos em ambiente de aulas práticas e trabalhos escritos. O encontro criará assim um ambiente de convívio único onde surgirão oportunidades múltiplas de discussões de como usar os métodos de avaliação para melhorar a aprendizagem em estatística e de como desenvolver e executar estes métodos de avaliação por forma à condução de investigação nesta área de conhecimento. O prazo de submissão de trabalhos terminou, tendo sido seleccionados para apresentação um conjunto de trabalhos extremamente interessantes. As actas deste encontro estarão disponíveis gratuitamente na página de publicações da IASE. Para mais pormenores sobre o programa e como se inscrever nesta conferência satélite consultar a morada <http://www.stat.auckland.ac.nz/iasestat07>

Note que o prazo de inscrição nesta conferência para não apresentadores é de 31 de Maio de 2007.

Organizador Local: Bruno C. de Sousa (Portugal), bruno@mct.uminho.pt

Bruno Sousa

Colóquio e Debate sobre O ENSINO DA ESTATÍSTICA A NÍVEL SUPERIOR

Por iniciativa da Faculdade de Ciências da Universidade de Lisboa (FCUL) e com o apoio do Departamento de Estatística e Investigação Operacional da FCUL, a Sociedade Portuguesa de Estatística, o Centro de Estatística e Aplicações da Universidade de Lisboa, a Sociedade Portuguesa de Matemática e o Instituto de Ciência Aplicada à Tecnologia da FCUL, Feridun Turkman, Antónia Turkman, Eugénia Graça Martins e Luísa Canto e Castro organizaram um Colóquio e Debate sobre “O Ensino da Estatística a Nível Superior” que se realizou no dia 31 de Janeiro de 2007 nas instalações da FCUL.

Sabe-se que...

Nos últimos anos tem-se vindo a observar uma drástica redução, tanto em quantidade como em qualidade, de estudantes que escolhem cursos relacionados com ciências básicas, tais como Matemática, Física e Química. A Estatística, sendo ainda considerada como uma área da Matemática Aplicada, tem sofrido também esse declínio. Por outro lado, as metodologias estatísticas são cada vez mais importantes em todos os domínios, desde as Ciências Experimentais até às Ciências Sociais e Políticas, havendo pois necessidade de formar técnicos capazes de responder a exigências de mercado extremamente vastas. De modo a dar resposta a este desafio é preciso rever o presente estado da Estatística na sociedade e o seu ensino nas universidades, culminando em recomendações para alterações futuras.

Sabe-se também que...

As universidades no Reino Unido têm passado por problemas semelhantes, tendo a “Royal Statistical Society” designado o Professor Fred Smith para presidir a uma comissão com o objectivo de rever o presente estado do ensino da Estatística nas universidades britânicas, não só em departamentos de Estatística e de Matemática como também em departamentos de outras áreas.

É fundamental encontrar-se respostas para questões importantes, a saber:

Qual o tipo de formação que se deve considerar para contemplar vertentes de formação tão variadas como:

- Profissionais de Estatística para empresas, quadros técnicos, indústria, serviços, etc.
- Professores de Matemática dos Ensino Básico e Secundário.
- Investigação

Como e quando é que essas diferentes formações devem ser dadas?

Com estes ingredientes, a Comissão Organizadora estabeleceu o seguinte:

Programa

- Das 14:30 às 16:30

Conferencistas Convidados:

- Fred Smith, Universidade de Southampton, U.K. – State of Statistics in British Universities.
- Dinis Pestana, Universidade de Lisboa – O papel da Estatística na Sociedade, na Ciência e na Tecnologia.
- João Branco, Universidade Técnica de Lisboa – O Ensino da Estatística no Ensino Superior em Portugal – passado e presente.

- Das 17 às 19

Debate sobre o Ensino da Estatística no Ensino Superior moderado por Feridun Turkman.: Como atrair estudantes para prosseguir formação em Estatística?

Este Colóquio e Debate teve cerca de 200 participantes, vindos de todo o País.

Sobre **O Ensino da Estatística nas Universidades do Reino Unido** Fred Smith e Laura Staetsky da Universidade de Southampton apresentaram uma comunicação convidada.

No seu estudo e com a motivação de que a “Estatística vem perdendo importância como disciplina”, partindo da “Análise da base de dados recolhida anualmente desde a década de 70” no Reino Unido, no projecto que integraram tinham como principais objectivos:

- Examinar a condição actual e os modelos de ensino da Estatística nas universidades do Reino Unido.
- Organizar uma base de dados para apoiar futuros debates sobre o ensino da Estatística.

De entre as principais conclusões do seu estudo salientam que:

- O número de pessoas envolvidas com a pesquisa e o ensino da Estatística está em declínio nas universidades do Reino Unido desde 1996.
- O ensino superior de Estatística está intrinsecamente ligado com as Ciências Matemáticas e apenas estes grupos satisfazem os critérios de credenciação da Royal Statistical Society.
- Um resultado importante da análise foi a descoberta da situação difícil em que se encontram os grupo mais fracos.
- A associação entre os grupos fortes e os scores do sistema de avaliação já era esperada dadas as recompensas financeiras e facilidades de financiamento disponíveis para os grupos com maiores escores. As universidades estão agindo tal como agentes económicos.
- Se os grupos mais fracos desaparecem, o Reino Unido ficará sem ensino superior de Estatística em diversas regiões.
- Para 2010, podemos prever um declínio de 7% a 22% no número de professores e pesquisadores da Estatística no grupo das Ciências Matemáticas. Para os demais grupos, a previsão é de estabilidade se for possível recrutar pessoal.
- A Estatística deve-se aliar às Ciências Matemáticas para crescerem em conjunto.
- As Ciências Matemáticas devem apoiar o processo de harmonização dos sistemas de ensino na Europa para poder competir com o padrão das pós-graduações norte-americanas.
- Há necessidade de estabelecer políticas regionais para o ensino superior.

Após a pausa para café, Feridun Turkman moderou um debate orientado para temáticas como:

1. Em Portugal o mercado de trabalho requer profissionais de Estatística? Se sim qual o perfil ideal deste profissional?
2. Como devem as Universidades/Politécnicos responder a esta necessidade - qual o perfil de cursos a oferecer para formar um estatístico profissional?
 - Há necessidade de um 1º ciclo em Estatística? Se sim, deve ser um curso profissionalizante ou mais matematizado?
 - Pode ser a estatística ensinada apenas em 2ºs ciclos para alunos com outras formações de base?
3. Qual deve ser o tipo de ensino da Estatística como base de apoio para outras Ciências?
4. Como atrair mais e melhores alunos para a área de Estatística?
5. Qual deve ser o papel das sociedades científicas na promoção da Estatística como Ciência?

Finalizando esta “notícia síntese” registamos, a todos os níveis, o enorme interesse desta iniciativa que, além de tudo, também teve a vantagem de despertar a vontade de que outras, do mesmo género, possam seguir-se... pois esta... “soube a pouco”!

A seguir, com textos dos próprios, apresentamos uma síntese das intervenções dos outros dois oradores convidados, Dinis Pestana e João Branco, bem como as opiniões de Luís Silva e Giovani Silva, a quem agradecemos.

Fernando Rosado

1) A recusa dos alunos, por exemplo aqueles que escolhem Biologia, em formar-se em Matemática/Estatística. Lecciono a disciplina de Bioestatística há cerca de 10 anos. Num questionário aos alunos de Biologia, Biologia Marinha e Ciências Biológicas e da Saúde, numa pergunta sobre o “gosto pela matemática” (escala ordinal de 1- detesta, até 5 - gosta muito) é muito raro o aluno que afirma gostar de matemática. A maioria distribui-se entre o 2 (não gosta) e o 3 (indiferente). Talvez por isso não escolham licenciaturas na área da Matemática/Estatística.

2) Penso que há necessidade de uma unidade curricular de introdução às Probabilidades e Estatística ao nível da licenciatura. Os alunos que realizam estágios de final de licenciatura ou que avançam para um mestrado, necessitam da aplicação da estatística. O facto de já incluírem uma formação de base nessa área poderá facilitar a realização desses trabalhos e a frequência de uma disciplina de “Estatística Aplicada” ao nível do mestrado. Por exemplo, os alunos de Biologia, têm a Bioestatística como disciplina de Base, e parecem estar mais à vontade numa Estatística Aplicada leccionada num mestrado, do que alunos que nunca tiveram formação nessa área. No entanto, nem sempre é fácil explicar a alunos do primeiro ou segundo ano da licenciatura, qual a utilidade da estatística, pois não sentiram até aí, necessidade de analisar dados numa escala apreciável.

Por outro lado, ao nível do mestrado, são muitos os alunos, das mais variadas áreas (Ensino, Saúde, Psicologia, Geologia, Biologia, Sociologia), que requerem mais alguma formação estatística, no sentido de realizarem as suas teses, pelo que é crucial a existência de formação estatística a esse nível. A realização de cursos livres ou de acções de formação também ajuda esses alunos.

3) Questão para a mesa:

- O que pensam acerca da organização do ensino da estatística na Universidade. Por exemplo no caso da minha universidade, há vários docentes a leccionar disciplinas na área da estatística em vários departamentos, incluindo o de Matemática. Qual seria o modo mais racional de gerir esta situação?

Luís Silva
Universidade dos Açores

“Uma preocupação referida neste debate é o futuro dos cursos de Estatística decorrente da crescente falta de alunos nos mesmos. Apesar de isso estar a ocorrer em Portugal e Inglaterra, existem outros países em que, pelo contrário, o número de cursos tem aumentado nesta área. Por exemplo, no Brasil os cursos de pós-graduação em Estatística e afins têm progredido nos últimos cinco anos, certamente motivados quer pela qualificação crescente do pessoal docente dos Departamentos de Estatística quer pela necessidade crescente de formação complementar de bacharéis formados nesse campo. Quanto a mim, o sucesso desta experiência deve-se ao carácter interveniente dos que promoveram a Estatística na sociedade brasileira, nomeadamente, com a regulamentação da profissão “estatístico” já há largos anos naquele país. Nesse sentido, nós devemos procurar intervir com mais determinação, se quisermos obter a projecção desejável desta área de conhecimento na sociedade civil portuguesa.”

Giovani Silva
Universidade Técnica de Lisboa

O Ensino da Estatística no Ensino Superior em Portugal – passado e presente

A Estatística não é uma daquelas disciplinas que nos deixa apenas um rasto da cultura que adquirimos durante a sua aprendizagem e que usamos só esporadicamente. A cultura estatística é muito abrangente e o recurso a conhecimentos de estatística é muito frequente. A Estatística faz parte integrante da literacia do cidadão, para que este se mova melhor no mundo aleatório que o rodeia, é indispensável ao progresso da actividade científica e alimenta de forma regular e eficaz os processos que sustentam as outras actividades. O ensino e os profissionais do ensino da estatística não podem esquecer esta realidade incontornável.

A crise que hoje se vive e que justifica a nossa presença aqui, não é geral, nem tem as mesmas causas e intensidade em todos os locais onde se tem manifestado. Pelo menos é o que eu deduzo de alguns comentários a esta crise, proferidos por eminentes cientistas:

Both the UK and Australia are in crisis in their training of statisticians. The crisis in Australia is worse, but reflects the same pattern as in UK. Declining funding has forced statistics departments or sections, where they still exist, to concentrate their teaching in low-level courses with a large numbers of students. Since most of these courses are service courses for other departments which do not lead to graduate output of statisticians, the decline in graduate training has been severe. ... The progressive decline in graduate and postgraduate training has occurred against a background of progressively increasing demand for statisticians. This is a disastrous mismatch.

(Murray Aitkin, University of Melbourne)

I appreciate that in US universities the situation is not quite ideal; for example, recruitment of students and junior faculty is becoming increasingly challenging, and the level of fundings is not as high as it might be. But the position of US academic Statistics today is preeminent, relative to that in the rest of the world, to a degree that could hardly have been anticipated twenty years ago.

(Peter Hall, The Australian National University, Canberra)

Se o decréscimo de verbas atribuídas a Departamentos de Matemática e de Estatística é apontado como justificação da crise australiana e inglesa e também de alguma insatisfação relativamente ao desvio da situação ideal que se ambiciona para Estados Unidos da América, tal não parece adaptar-se bem à situação que se vive em Portugal. Aqui o sistema de atribuição de verbas à universidade está em vias de ser alterado de forma mais considerável, mas até agora as alterações que se têm verificado não parecem justificar o que se está a passar. O problema da baixa procura de formação em estatística pelos alunos merece ser cuidadosamente investigado. Será que essa baixa procura acontece em todas as escolas? Eu estou a lembrar-me que, na pesquisa de informação que efectuei para preparar a minha intervenção no colóquio, encontrei um curso, votado à formação de profissionais numa área que inclui a estatística, com uma componente de estatística apreciável (1/5 das disciplinas são inquestionavelmente da área da estatística) e que preencheu todas as várias dezenas de vagas postas à disposição dos alunos.

Este colóquio foi uma iniciativa muito importante no sentido de alertar a comunidade estatística para a natureza e dimensão da crise e, ao mesmo tempo, desafiar essa mesma comunidade a responder às clarividentes questões que foram propostas para o debate, contribuindo assim para encontrar causas e sugerir soluções. Infelizmente eu não estou, nesta breve nota, a dar esse tipo de contribuição concreta, mas não queria deixar de reagir à questão específica que me foi colocada “Há necessidade de um 1º ciclo em Estatística? Se sim, deve ser um curso profissionalizante ou mais matematizado?” Em meu entender formar profissionais é essencial e produzir especialistas é imprescindível. Acho que as duas vertentes deviam coexistir. Se isso não for possível então que a formação matemática, necessária ao verdadeiro especialista, lhe seja oferecida no segundo ciclo. A especialização é vital numa área do saber que está à mercê de todos, para o bem e para o mal.

Para terminar considero que, perante o tremendo desenvolvimento que a Estatística sofreu nestes últimos 25 anos em Portugal, não nos podemos atemorizar com as ameaças que surgem agora no horizonte. Há muito que fazer e muito que reflectir para perceber o que temos vindo a fazer até hoje e o que é necessário fazer a seguir para que a Estatística possa continuar a contribuir, de forma continuada e rigorosa, para o desenvolvimento da sociedade em geral.

João Branco
Universidade Técnica de Lisboa

O Ensino da Estatística

Estatística e Sociedade
Estatística e Ciência

- *Estatística para todos — as preocupações que devemos ter com a formação, reciclagem e especialização de docentes de outros graus de ensino.*

Thomas Carlyle (Chartism, 1839): [...]” a judicious man looks at statistics [...] to save himself from having ignorance foisted on him.” Reconhecia assim, já no século XIX, o papel da Estatística na defesa da cidadania.

Todos devem por isso ter o direito a algum treino em Estatística. Como todos sabemos, a escolha de exemplos apropriados, a apresentação simples das ideias, a capacidade de despertar a curiosidade, não são tarefas simples.

Num sentido positivo, a Estatística é um dos instrumentos com mais sucesso na transformação da informação em conhecimento; numa perspectiva mais redutora, mas a ter em conta, a estatística é uma defesa contra a perversão da informação, e nesse sentido um escudo da Cidadania.

É, por isso, uma peça essencial no “contrato social”, na forma como se ultrapassa a desconfiança entre estados (papel do EUROSTAT, por exemplo, na fiscalização das contas públicas dos estados da UE), na forma como os cidadãos limitam, pela opinião pública informada, o abuso dos poderes do Estado — e uma forma de defesa dos poderes do Estado contra a (perversão da) opinião pública pelos órgãos de comunicação, por exemplo.

Claro que, como tudo, pode ter uma utilização perversa e abusiva, contribuindo pelo contrário em dar uma aparência de solidez e objectividade a mentiras. Basta ver o seu uso em publicidade, e em campanhas eleitorais.

Quer a Sociedade Portuguesa de Estatística quer as instituições de ensino superior — nomeadamente as que intervêm na formação de professores — devem investir na disponibilização de documentação apropriada para a divulgação da Estatística junto da sociedade, e nomeadamente junto das classes etárias mais susceptíveis de incorporarem esses conhecimentos na forma como enfrentam o quotidiano. O grupo Alea tem realizado um trabalho admirável, que devia ser reforçado e diversificado. Seria decerto saudável, também, uma intensificação de contacto (cursos de especialização e de reciclagem) com professores de outros graus de ensino, com insistência na ideia de que o treino elementar da Estatística deve ter, entre outras, a intenção de reforçar a capacidade crítica sobre a informação a que estamos sujeitos.

- *Há empregos para licenciados (1º ciclo) em Estatística?*

Há necessidade de profissionais de Estatística, e um bom profissional consegue decerto colocar-se com facilidade; provavelmente, devido ao mau conhecimento das muitas aplicações da Estatística, terá que ter alguma iniciativa de quais as mais-valias que pode trazer aos potenciais empregadores. Para além dos tradicionais (sector terciário, sobretudo), há um largo campo de actuação na administração central e local, administração hospitalar, agências de informação / comunicação social, etc.

Mas será sempre um mito pretender que só licenciados em Estatística estão devidamente equipados com conhecimentos para exercer profissionalmente (muita da Estatística que se usa é verdadeiramente elementar). Há é que explicar o que podem fazer melhor, e como podem fazer evoluir qualitativamente a utilização da Estatística no seu local de trabalho.

Para os licenciados (1º ciclo) se apresentarem como candidatos credíveis e apelativos no mercado de trabalho, devem ser treinados na recolha e gestão de informação, análise inicial dos dados e apresentação dos resultados de forma rigorosa e apelativa. Muita da formação matemática indispensável para o treino de cientistas pode ser substituída por familiaridade com métodos de Monte Carlo. Deve investir-se fortemente na formação na área de amostragem e estudos de mercado; o controle de qualidade, ao nível necessário para um desempenho útil na área industrial e empresarial, também é uma aposta a fazer neste tipo de formação em Estatística Aplicada.

- *O ensino elementar da Estatística noutras áreas.*

Tudo aquilo que nos rodeia tem uma parte de necessidade e uma parte de incerteza. A evolução da Ci-

ência tem acompanhado por um lado a nossa capacidade de medir melhor — e a Metrologia é uma área demasiado subdesenvolvida, entre nós, nos estudos de ciências experimentais —, e a capacidade de incorporar no nosso conhecimento modelos para o que é aleatório. Felizmente já é reconhecida em todas as áreas a necessidade de uma formação inicial em Probabilidade e Estatística, mas infelizmente o tempo alocado a essa disciplina é em geral insuficiente, porque falha a compreensão de que o que está em causa é o desenvolvimento de novas formas de perspectivar a realidade.

É por isso necessário transmitir — não só aos alunos, como aos nossos colegas responsáveis pela estruturação de cursos de outras áreas — a importância dos conceitos necessários para domesticar o acaso, usar a incerteza em nosso benefício, e em última análise usar bem a informação na criação de conhecimento. É preciso explicar que para usar o acaso como aliado em vez de inimigo é preciso alterar hábitos de pensar, e isso tem que ser feito com a demora adequada. Há que insistir que num curso elementar o que se pode ensinar é limitado, que é mais importante dominar os conceitos fundamentais do que aprender a usar acriticamente packages estatísticos, que dificilmente se poderá incluir numa cadeira semestral mais do que estatística descritiva, os modelos estatísticos univariados mais usuais, uma panorâmica dos aspectos elementares de estimação e de testes de hipóteses, um pouco de regressão, um pouquinho de amostragem, e uma incursão nos aspectos mais simples de análise da variância — e para isto, menos do que 6 horas por semana é condenar os alunos a insucesso e desamor por esta disciplina estruturante do pensamento, mesmo que tenham um aparente sucesso escolar.

Claro que não se pode aspirar em Portugal a situações que se aproximem do que se passa em universidades de escol (em Berkeley, por exemplo, há 4 cadeiras semestrais de Bioestatística). Mas não será facilmente entendível que porventura uma cadeira mais de Estatística, a nível opcional, poderia contribuir para uma formação de qualidade dos alunos?

- *O ensino da Estatística em pós-graduações de outras áreas.*

Por outro lado, defendemos que haja uma formação em metodologia da investigação científica — em que a Estatística ocupa uma posição saliente — mais rigorosa e exigente a nível dos segundos ciclos. Parece-nos de facto essencial que os alunos de segundo ciclo tenham conhecimentos mais profundos de planeamento experimental e de amostragem, e conhecimento das técnicas exploratórias e confirmatórias mais correntes em estatística multivariada. Por outro lado, parece cada vez mais importante, nas áreas científicas, um conhecimento razoável de técnicas robustas e de simulação.

Por outro lado, a estatística industrial, a metrologia, o controle de qualidade, deveriam ser áreas de investimento de segundos ciclos multidisciplinares, em que a Estatística deveria ter um papel.

- *Segundo ciclo, pós-graduações e cursos de especialização em Estatística.*

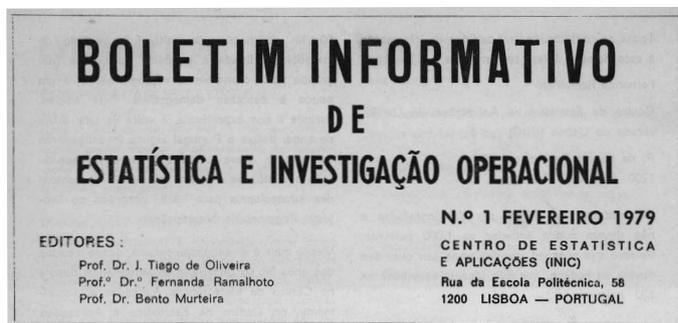
A nível de segundo ciclo pode e deve investir-se quer em Estatística Matemática, essencial para a manutenção do nível científico do País, quer nas novas áreas de “Intelligent Data Analysis”, como complemento das áreas tradicionais de análise de dados, que não devem também ser abandonadas. Faz por outro lado todo o sentido explorar a procura de conhecimentos estatísticos de formados em ciências sociais e da informação e em ciências bio-médicas. Há, por outro lado, um manancial de áreas pluridisciplinares em que a noção de risco é um motor de progresso, que só a nossa incapacidade de deitar a mão a tanta coisa nos impede de explorar. Fica assim o desafio: planear, a longo termo, um conjunto alargado de acções a desenvolver, porventura por consórcios de instituições de ensino superior, que dêem à Estatística a visibilidade que a nossa sociedade merece e de que o nosso País precisa.

Dinis Pestana
Universidade de Lisboa

Episódios na História da Estatística

• O Boletim

O número 1 do Boletim Informativo de Estatística e Investigação Operacional foi publicado em Fevereiro de 1979 pelo Centro de Estatística e Aplicações (INIC).



Tiago de Oliveira, Fernanda Ramalhoto e Bento Murteira foram os editores. Num pequeno formato de 24cm x 17 cm esse primeiro número teve 8 páginas; seis destas contêm informação variada compilada em “Calendário Internacional”, a relativa a congressos, “Acordos Culturais” e “Revistas”. Nesta secção, eram apresentadas cerca de cinco dezenas de “algumas das princi-

pais revistas de Estatística e Investigação Operacional existentes no Instituto Gulbenkian de Ciência em Oeiras” e com a indicação das datas a partir das quais se encontravam disponíveis. A grande maioria das assinaturas foi começada na década de sessenta do século passado, portanto em coincidência com o início da recém criada Fundação Calouste Gulbenkian em 1956.

No editorial, assinado por Fernanda Ramalhoto, é explicada a razão da iniciativa. (Aos editores e em nome de outros) pareceu “ser importante para uma melhor aproximação entre as pessoas de qualquer modo ligadas à Estatística” que se iniciasse essa publicação periódica, “sobre o que se vai passando no país e no estrangeiro” no domínio de interesse, em particular da estatística. Afirmam ainda a sua convicção que “neste momento há já no país um grupo considerável de investigadores e diferentes utilizadores de Estatística e Investigação Operacional que sentem a necessidade de uma melhor utilização dos nossos recursos humanos e técnicos, embora tais recursos a nível mundial ou mesmo europeu sejam ainda surpreendentemente modestos”. Através do Boletim desejavam “fornecer os meios para: Pôr questões, ou responder a questões postas por outros leitores do Boletim [...]; Fazer relatos sobre Conferências nacionais ou estrangeiras [...]; Fazer um apelo para uma possível ajuda num determinado trabalho científico [...]; Iniciar discussão entre os leitores do Boletim sobre o ensino da Estatística e Investigação Operacional no Ensino Secundário; Anunciar empregos [...]; Publicar críticas a livros [...]; Sugerir rubricas que não estejam aqui contempladas”. Citámos um bom projecto de trabalho!

Naquele número 1, como artigo de opinião é apresentado “Alguns Núcleos Recentes de Investigação Estatística em Portugal” da autoria de Tiago de Oliveira. Neste estudo, o autor defende que “os núcleos fundamentais de estudiosos têm estado ligados ao Instituto Superior de Ciências Económicas e Financeiras, à Faculdade de Ciências de Lisboa e, também, de índole bastante mais aplicada, ao Instituto Nacional de Estatística. [...] A actividade do Centro de Estudos de Estatística Económica (do IAC) ligado a gente do ISCEF, e que se estende dos anos 50 a meados dos anos 60, foi essencialmente votada ao estudo de metodologia ligada à pesquisa económica [...]. Virado mais à investigação teórica, desde meados dos anos 50, no Seminário de Matemática, depois no Centro de Matemáticas Aplicadas e, actualmente, no Centro de Estatística e Aplicações (do IAC e, agora do INIC) a gente da Faculdade de Ciências de Lisboa e outros têm prosseguido estudos[...] Constituídos em meados dos anos 40 os Centros de Estudos Económicos e Demográficos do Instituto Nacional de Estatística têm executado alguns estudos ligados à sua definição. Deverão recordar-se, entre outros, os estudos de demografia portuguesa realizados. É bem pouco, quer na área de investigação teórica, quer na das aplicações o que entre nós se tem realizado, muitas vezes fruto do interesse e persistência individual. Países da dimensão do nosso têm, há muito, Institutos de Estatística nas Universidades em ligação com os Institutos de Aplicação. Estamos bem atrasados! Temos, pois, de avançar rapidamente para o futuro, para o que temos gente capaz. Não se pode perder mais tempo!”

Este Boletim número 1 é património partilhado pela Sociedade Portuguesa de Estatística e pela Associação para o Desenvolvimento da Investigação Operacional. De algum modo ligada a um início de projectos científicos... é um episódio na História da Estatística.

Fernando Rosado

A “Escola de Extremos” em Portugal

Não há caminhos, há que caminhar

Andreia Hall, *andrea@mat.ua.pt*
Universidade de Aveiro

1 Enquadramento

O título deste trabalho baseia-se numa inscrição que se encontra na catedral de Toledo dirigida aos peregrinos que ali se dirigem: “Caminantes, não hay caminos, hay que caminar”. Em ciência também não há caminhos, os caminhos vão-se desbravando `a medida que se trabalha.

A Análise de Valores Extremos é uma área das Probabilidades e Estatística que tem merecido uma especial atenção pela comunidade científica Portuguesa. Factores externos `a própria ciência levaram a que um grande número de investigadores optasse por dedicar o seu trabalho a esta área. Até Fevereiro de 2005, e com base na informação recolhida pelas Professoras Antónia Amaral Turkman e Ivette Gomes (Rosado, 2005) doutoraram-se nesta área 32 investigadores portugueses ou residentes em Portugal. Tendo em conta que, no mesmo período, contabilizaram-se no total 139 doutoramentos na área das Probabilidades e Estatística, a proporção de doutorados em Análise de Valores Extremos mostra bem a invulgar adesão a esta área no nosso país. A este facto não é certamente alheio o papel da Professora Ivette Gomes que consegue reunir numa só pessoa capacidades invulgares de investigação e trabalho simultaneamente com o dom de cativar, motivar e apadrinhar pessoas para a investigação.

Dentro da Análise de Valores Extremos têm sido vários os tópicos desenvolvidos pelos investigadores portugueses. Tal como noutras áreas encontramos trabalhos mais ligados `a estimação de parâmetros de interesse e trabalhos mais ligados a questões probabilísticas. Muito embora, em última análise, a Análise de Valores Extremos tenha como objectivo a resolução de problemas reais, a verdade é que a maioria dos trabalhos desenvolvidos em Portugal centra-se em aspectos mais teóricos, quer probabilísticos quer estatísticos. Se eu tivesse que apontar alguma falha na investigação nacional desta área apontaria para a falta de aplicações reais. A minha área de trabalho não foge à regra e todo o trabalho nela desenvolvido tem-se ficado pelas questões teóricas ainda que haja uma constante preocupação de estudar tópicos com aplicabilidade prática. Em concreto a minha área de investigação diz respeito ao estudo dos extremos de sucessões de variáveis aleatórias inteiras, não negativas. A motivação surge do facto de encontrarmos muitos fenómenos que geram dados de contagem que interessa analisar. O número de quartos de hotel ocupados diariamente, o número de indivíduos de determinada espécie num certo habitat, o número de transacções efectuadas diariamente numa praça financeira são apenas alguns exemplos de variáveis que assumem valores inteiros e que têm interesse ser estudadas.

Acontece que, no que respeita a Análise de Valores Extremos, as variáveis discretas levantam problemas inexistentes nas variáveis contínuas. Daí que se torne pertinente considerar variáveis discretas, em particular inteiras não-negativas, só por si.

A natureza discreta dos dados não trás apenas implicações no estudo do seu comportamento extremal. A própria definição e caracterização dos modelos de contagem trás novidades e problemas acrescidos já que a maioria dos modelos para dados contínuos não se aplica a dados discretos. Por exemplo os modelos ARMA apenas admitem distribuições contínuas e foi necessário definir modelos com estruturas semelhantes para dados discretos. Uma forma de obter modelos para dados discretos consiste em partir de modelos contínuos e substituir operadores contínuos por operadores discretos. Um dos procedimentos mais utilizados até agora consiste em substituir a multiplicação por um escalar pela operação thinning em que o escalar assume o papel de uma probabilidade. Assim, para uma variável aleatória discreta X e um escalar $\alpha \in [0, 1]$, tem-se

$$\alpha * X = \sum_{i=1}^X B_i(\alpha)$$

onde $\{B_i\}$ é uma sucessão iid de variáveis aleatórias de Bernoulli com parâmetro α . Muitos dos modelos considerados nos trabalhos acima referidos basearam-se neste procedimento. Para mais informações sobre modelos de dados de contagem ver McKenzie (2003).

É neste contexto que têm surgido na literatura diversos trabalhos dedicados ao estudo dos extremos de sucessões de variáveis discretas, entendendo-se geralmente por variáveis discretas, variáveis inteiras não negativas.

2 Caminhos e Referências

A Teoria clássica de Valores Extremos focou inicialmente a sua atenção no estudo do comportamento do máximo (ou outras estatísticas de ordem extrema) de sucessões de variáveis independentes e identicamente distribuídas (iid). A descoberta do Teorema de Tipos Extremos na primeira metade do século XX permitiu concluir que independentemente da função de distribuição das variáveis, a distribuição limite do máximo das primeiras n , quando n tende para infinito e após normalização linear, tem de pertencer a uma única classe, a classe de distribuições max-estáveis, desde que exista esse limite. A classe de distribuições max-estáveis pode ser definida pela seguinte expressão, dependente de apenas um único parâmetro, a menos de localização e escala:

$$G(x) = \exp(-(1 + \gamma x)^{-1/\gamma}), \quad 1 + \gamma x > 0, \gamma \in \mathbb{R}. \quad (1)$$

Quando $\gamma > 0$ temos a classe das distribuições de Fréchet, quando $\gamma = 0$ obtemos a distribuição de Gumbel (tomando $\lim_{\gamma \rightarrow 0}$ em (1)), e quando $\gamma < 0$ temos a classe das distribuições de Weibull. O parâmetro na expressão anterior designa-se habitualmente por índice de cauda e naturalmente captou a atenção de inúmeros investigadores ao longo das últimas décadas. Uma vez que se trata de um parâmetro associado a um resultado limite (podemos dizer para além de qualquer amostra que se recolha) a sua estimação é necessariamente problemática. Assim, surgiu um enorme número de publicações dedicadas à proposta e caracterização de diferentes estimadores para γ . Neste campo a contribuição dos investigadores Portugueses tem sido notória. Por exemplo, podemos encontrar propostas de estimadores que procuram reduzir o viés em relação aos estimadores mais consagrados na literatura ou então que procuram diminuir a dependência da performance do estimador em relação à escolha do nível, a partir do qual se considera observar uma excedência.

Quando a distribuição dos dados é inteira surgem alguns problemas no estudo das características extremas. Por exemplo, se a distribuição possuir um limite superior de suporte finito (caso das distribuições Binomial, Hipergeométrica, etc.) não faz sequer sentido estudar o comportamento assintótico do máximo. Isto faz com que não exista nenhuma distribuição discreta no domínio de atracção da Weibull ($\gamma < 0$). Felizmente que no domínio de atracção da Fréchet a natureza discreta das observações não trás complicações no estudo do comportamento dos extremos. No entanto, nenhuma das distribuições discretas usuais pertence a este domínio. No domínio de atracção da Gumbel apenas vamos encontrar algumas distribuições discretas com decaimentos bastante lentos, mais lentos que a função exponencial negativa. Nenhuma das distribuições discretas vulgarmente utilizadas se encontra nesta situação. Quando o decaimento da cauda da distribuição é muito rápido (por exemplo como na distribuição de Poisson) não é possível encontrar qualquer tipo de normalização que estabilize o comportamento do máximo. Mas, quando o decaimento da cauda é do tipo exponencial (como acontece com a distribuição Geométrica e Binomial Negativa em geral) é possível encontrar alguma estabilidade na distribuição do máximo: limites superior e inferior max-estáveis, ainda que não exista nenhuma distribuição limite. Clive Anderson foi o primeiro a considerar distribuições discretas com este tipo de comportamento e obteve um resultado bastante útil que se encontra em seguida.

Teorema 1 (Anderson, 1970) *Seja F uma função de distribuição cujo suporte é um conjunto de inteiros suficientemente grande. Então, existe uma sucessão de constantes $\{b_n\}$ tais que*

$$\begin{cases} \limsup_{n \rightarrow \infty} F^n(x + b_n) \leq e^{-e^{-\alpha x}} \\ \liminf_{n \rightarrow \infty} F^n(x + b_n) \geq e^{-e^{-\alpha(x-1)}} \end{cases},$$

para algum $\alpha > 0$ e para todo $x \in \mathbb{R}$, sse

$$\lim_{n \rightarrow \infty} \frac{1 - F(n)}{1 - F(n-1)} = \exp\{-\alpha\}.$$

De facto b_n pode ser escolhido a partir de $b_n = F_c^{-1}(1 - \frac{1}{n})$ onde F_c é qualquer distribuição contínua no domínio de atracção da Gumbel que verifica

$$F_c([x]) = F_x.$$

A modelação de dados reais a partir de sucessões iid é como todos sabemos inadequada em inúmeras situações. O pressuposto de independência é muitas vezes demasiado forte, e o pressuposto de estacionaridade também se pode questionar em muitas situações. Tendo em conta esta realidade, é natural que a investigação na área dos valores extremos tenha procurado abranger situações de não independência e/ou não igualdade de distribuição. A maioria dos resultados que se encontram referem-se à situação de não independência, ou seja ao estudo dos extremos de sucessões estacionárias e é fundamentalmente neste âmbito que o meu trabalho se tem desenvolvido. Uma grande diferença que surge no estudo deste tipo de sucessões decorre do facto de ser natural encontrarmos agrupamentos de excedências de níveis elevados em vez de valores isolados como acontecia no caso iid. No contexto das sucessões estacionárias é possível definir um novo parâmetro de interesse que no caso de existir (e sob condições bastante gerais de regularidade) pode ser interpretado como o inverso do número médio de excedências, quando n tende para infinito. Este parâmetro designa-se por índice extremal.

No estudo dos valores extremos de sucessões estacionárias é usual considerar certo tipo de condições de dependência que, por um lado garantem que a distribuição limite do máximo é ainda do tipo max-estável, e por outro permitem calcular o valor do índice extremal que é em geral difícil de obter. Algumas dessas condições podem-se encontrar em Leadbetter et al. (1983), Leadbetter e Nandagopalan (1989), Chernick et al. (1991), Hsing et al. (1988) e Ferreira (1993).

Com base nalgumas das condições de dependência acima mencionadas foi possível estudar o comportamento do máximo de vários tipos de sucessões estacionárias de dados de contagem. De entre as publicações que daí resultaram destacamos as seguintes:

- McCormick e Park (1992) fizeram o estudo do comportamento extremal dum processo autoregressivo de primeira ordem e dum processo autoregressivo de média móvel, ambos com distribuição marginal Binomial Negativa.
- McCormick e Sun (1993) fizeram o estudo do comportamento assintótico do máximo conjuntamente com a média de sucessões estacionárias com distribuição marginal na classe de Anderson.
Os resultados são aplicados a um processo autoregressivo de primeira ordem e ao comprimento uma fila de espera $M|M|s$.
- Hall (1996) fez o estudo do comportamento extremal de uma sucessão autoregressiva de máximos.
- Hooghiemstra et al. (1998) fornecem limites e aproximações para o comprimento máximo dum fila de espera $M|M|s$.
- Hall (2001, 2003) considerou uma classe geral de médias móveis com distribuição marginal no domínio de atracção da Fréchet e na classe de Anderson, respectivamente. Esta classe define-se por analogia com os processos clássicos MA de ordem infinita, substituindo a multiplicação pela operação *thinning*.
- Hall e Moreira (2006) consideraram uma sucessão particular com estrutura de média móvel cuja distribuição marginal é Geométrica. O modelo em causa foi obtido por substituição do operador multiplicação pelo operador *thinning* num modelo particular contínuo, com distribuição marginal exponencial.

A partir destes trabalhos foi possível desenvolver novos resultados alargando um pouco o âmbito de trabalho. Assim,

- Hall e Scotto (2003) e Hall, Scotto e Ferreira (2004) consideraram o efeito da sub-amostragem nos extremos de médias móveis de contagem no domínio de atracção da Fréchet. O primeiro trabalho considera a sub-amostragem sistemática enquanto o segundo considera o caso mais geral da sub-amostragem periódica.
- Hall e Hüsler (2006) consideraram o efeito da amostragem aleatória nos extremos de sucessões estacionárias com especial destaque à sucessões de dados de contagem.
- Utilizando uma metodologia distinta, Hall e Scotto (2007) consideraram o efeito da amostragem aleatória numa classe geral de médias móveis de dados de contagem no domínio de atracção da Fréchet.
- Hall e Scotto (2006) consideraram uma classe geral de médias móveis periódicas com distribuição marginal na classe de Anderson. Neste trabalho deixamos de ter estacionaridade e passamos a ter sucessões periódicas.
- Hall, Scotto e Cruz (2007) consideraram uma classe geral de médias móveis periódicas com distribuição marginal de Fréchet e também uma classe de médias móveis estacionárias mas com coeficientes aleatórios.

Para finalizar gostaria de fazer uma breve referência ao trabalho de três colegas Portugueses, que está fortemente relacionado com os trabalhos acima mencionados.

Maria da Graça Temido, professora da Universidade de Coimbra, dedicou uma parte relevante do seu estudo à classe de distribuições max-semiestáveis. Trata-se de uma classe de distribuições que inclui a classe de distribuições max-estáveis e que surge como classe de possíveis leis limite do máximo das primeiras k_n observações (em vez das primeiras n observações) para uma escolha adequada da sucessão $\{k_n\}$. De referir que as distribuições da classe de Anderson, que não pertencem ao domínio de atracção de nenhuma lei max-estável, pertencem ao domínio de atracção das max-semiestáveis. De entre os seus trabalhos destaco os seguintes:

- Canto e Castro et al. (2001) e Temido e Canto e Castro (2003) fizeram um estudo inovador da classe de distribuições max-semiestáveis e seus domínios de atracção.
- Com base nos dois trabalhos anteriores, Hall e Temido (2006) fizeram o estudo de alguns modelos de contagem com distribuição marginal na classe de Anderson.

Helena Maria Ferreira, professora da Universidade da Beira Interior possui uma longa e valiosa lista de publicações em diversas áreas da Teoria de Valores Extremos, sendo os seus trabalhos essencialmente de índole probabilística. Irei apenas referir dois dos seus trabalhos por terem sido essenciais ao estudo das sucessões periódicas de dados de contagem. Ferreira e Martins (2003) e Martins e Ferreira (2004) estudaram os extremos de sucessões periódicas e o impacto da sub-amostragem neste tipo de sucessões.

Manuel Gonzalez Scotto, meu colega na Universidade de Aveiro, tem já um número muito significativo de publicações apesar de apenas ter concluído o seu doutoramento em 2000. Os seus trabalhos abordam vários aspectos probabilísticos da Análise de Valores Extremos, com especial atenção ao efeito da sub-amostragem no comportamento extremal de sucessões estacionários. De entre os seus trabalhos destaca-se o estudo do comportamento extremal de séries de Volterra (Scotto e Turkman, 2005) e de equações estocásticas às diferenças (Scotto, 2005).

Referências

- Anderson, C. (1970) Extreme value theory for a class of distributions with applications to some stochastic processes, *Journal of Appl. Probab.* Vol. 7, 99-113.
- Canto e Castro, L., de Haan, L. & Temido, M.G. (2001) Rarely observed sample maxima. *Theory Probab. Appl.*, Vol. 45:4, 652-662.
- Chernick, M., Hsing, T. & McCormick, W. P. (1991) Calculating the extremal index for a class of stationary sequences, *Adv. Appl. Prob.*, Vol. 23, 835-850.

- Ferreira, H. (1993) Joint exceedances under a local dependence condition, *Journal of Appl. Probab.*, Vol. 30, 112-120.
- Ferreira, H. & Martins, A.P. (2003). The extremal index of sub-sampled sequences with strong local condition. *Revstat*. Vol. 1, 16-24.
- Hall, A. (1996) Maximum term of a particular autoregressive sequence with discrete margins, *Comm. Stat. - Theory and Methods* Vol. 25, 721-736.
- Hall, A. (2001) Extremes of integer-valued moving average models with regularly varying tails. *Extremes* Vol. 4:3, 219-239.
- Hall, A. (2003) Extremes of integer-valued moving average models with exponential type tails. *Extremes*, Vol. 6, 361-379.
- Hall, A. & Scotto, M. (2003) Extremes of sub-sampled integer-valued moving average models with heavy-tailed innovations, *Statistics and Probability Letters*, Vol. 63:1, 97-105.
- Hall, A., Scotto, M. & Ferreira, H. (2004) On the extremal behaviour of generalised periodic subsampled moving average models with regularly varying tails, *Extremes*, Vol. 7, 149-160.
- Hall, A. & Hüsler, J. (2006) Extremes of stationary sequences with failures, *Stochastic Models*, Vol. 22, 537-557.
- Hall, A. & Moreira, O. (2006) A note on the extremes of a particular moving average count data model, *Statistics and Probability Letters*, Vol. 76, 135-141.
- Hall, A. & Scotto, M. (2006) Extremes of periodic integer-valued sequences with exponential type tails, *Revstat- Statistical Journal*, 4, 249-273.
- Hall, A. & Scotto, M. (2007) Extremes of randomly sub-sampled time series, (preprint).
- Hall, A., Scotto, M. & Cruz, P. (2007) Extremes of integer-valued moving average sequences, (preprint).
- Hall, A. & Temido, M.G. (2006) On the maximum term of MA and Max-AR models with margins in Anderson's class, *Theory Probab. Appl.* (aceite para publicação).
- Hooghiemstra, G., Meester, L. E. and Hüsler, J. (1998). On the extremal index for the M/M/s queue. *Comm. Statist. Stochastic Models* Vol. 14, 611-621.
- Hsing, T., Hüsler, J. & Leadbetter, M. R. (1988) On the exceedance point process for a stationary sequence, *Probab. Theory Rel. Fields*, Vol. 78, 97-112.
- Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983) *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York.
- Leadbetter, M. R. & Nandagopalan, S. (1989) *On exceedance point processes for stationary sequences under mild oscillation restriction*. In Hüsler, J. & Reiss, R. D., eds., *Extreme Values*, Springer-Verlag, New York.
- Martins, A.P. & Ferreira, H. (2004) Extremes of periodic moving averages of random variables with regularly varying tail probabilities, *SORT*, Vol. 28:2, 161-175.
- McCormick, W. P. and Park, Y. S. (1992) Asymptotic analysis of extremes from autoregressive negative binomial processes. *J. Appl. Probab.* 29, 904-920.
- McCormick, W. P. and Sun, J. (1993) Sums and maxima of discrete stationary processes. *J. Appl. Probab.*, Vol. 30:4, 863-876.
- McKenzie, E. (2003). *Discrete variate time series*. In Shanbhag, D. N. and Rao, C. R., editors, *Handbook of Statistics*, Elsevier Science. 573-606.
- Rosado, F. (Editor) (2005) *Memorial da Sociedade Portuguesa de Estatística*, Edições SPE, Lisboa.
- Scotto, M. G. (2005). Extremes of a class of deterministic sub-sampled processes with applications to stochastic difference equations. *Stoch. Process. Appl.* 115, 417-434
- Scotto, M. G. & Turkman, K. F. (2005). Extremes of Volterra series expansions with heavy-tailed innovations. *Nonlinear Analysis*. Vol. 63, 106-122
- Temido, M.G. & Canto e Castro, L. (2003) Max-semistable laws in extremes of stationary random sequences, *Theory Probab. Appl.*, Vol. 47:2, 365-374.



Acerca de Testes Estatísticos para Valores Extremos

Isabel Fraga Alves, *isabel.alves@fc.ul.pt*
DEIO / CEAUL / FCUL

1. Introdução

Na análise de valores extremos é de importância relevante as suposições que se consideram na cauda da função de distribuição subjacente à amostra de dados em causa. Se o interesse reside em grandes (pequenos) valores, essa ligação é dirigida à cauda superior (inferior), tendo por base o Teorema de Gnedenko que estabelece três domínios de atracção para o máximo linearmente normalizado e correspondentes modelos limite para as maiores observações. Um exemplo típico que evidencia uma conveniente modelação da cauda da distribuição é o da inferência acerca de quantis extremos, usualmente fora do intervalo amostral. Dizemos, por vezes, que se trata de um problema de extrapolação além da amostra. Se nos fixarmos de uma vez por todas em grandes valores, já que o problema dual dos pequenos valores é tratado por simetriação dos dados, uma cauda direita mal “pesada” pode deixar revelar uma sub ou sobre-estimação acerca desse valor, com as correspondentes consequências práticas que se adivinham. Realmente, a inferência estatística acerca de acontecimentos raros pode (e deve!) ser deduzida a partir daquelas observações que são extremas sob determinado tipo de critério.

Existem diferentes formas de definir tais observações e as respectivas abordagens em inferência estatística de valores extremos:

A metodologia clássica de Gumbel que extrai os máximos por blocos, o método POT dos excessos de nível (do inglês Peaks-Over-Threshold) e mais recentemente a metodologia PORT dos excessos de nível aleatório (Peaks-Over-Random-Threshold).

Em todas as abordagens, a inferência estatística é claramente melhorada se fizermos a priori uma escolha estatística acerca decaimento para zero mais apropriado para a cauda da distribuição subjacente: caudas curtas com limite superior do suporte finito, caudas leves de tipo exponencial ou caudas pesadas com limite superior de suporte infinito e que vão polinomialmente para zero.

Tal como referido no início, esta posição é fortemente baseada na Teoria de Valores Extremos (EVT), a qual se apoia no Teorema Fundamental de Gnedenko (1943) para os max-domínios de atracção, e que estabelece a Distribuição generalizada de Valores Extremos (GEV-Generalized Extreme Value, do inglês), como uma versão unificada de todos os comportamentos distribucionais não-degenerados limite, para a sequência do máximo de variáveis aleatórias igualmente distribuídas independentes ou fracamente dependentes. A GEV reduz-se às distribuições Fréchet, Weibull e Gumbel. O Domínio de atracção Fréchet diz respeito a caudas com decaimento polinomial, com limite superior do suporte infinito. O Domínio Weibull contém distribuições de caudas curtas com limite superior do suporte finito. Por sua vez, o caso intermédio do Domínio Gumbel, está dirigido a distribuições de cauda leve com limite superior do suporte finito ou infinito e de tipo exponencial. Assim, separar os procedimentos de inferência estatística de acordo com o domínio de max-atracção mais conveniente para a distribuição amostrada tornou-se uma prática usual ao longo da literatura da especialidade, seguindo quer abordagens paramétricas quer semi-paramétricas. Numa abordagem semi-paramétrica, a única suposição de base é a de que a distribuição está no domínio de atracção da GEV. Neste contexto, qualquer inferência que diga respeito à cauda da distribuição subjacente pode ser baseada naquelas observações situadas acima de um nível aleatório (PORT). Esta última posição contrasta com a possibilidade alternativa de restringir a atenção para um número aleatório de valores amostrais excedendo um dado nível determinístico (POT).

Esta pequena nota tem por objectivo dar uma visão breve acerca dos variados testes publicados no contexto de escolha estatística de condições de cauda para valores extremos, assim como dar conhecimento de algumas abordagens semi-paramétricas baseadas em estatísticas invariantes perante alterações de localização/escala, baseadas nos excessos acima de um patamar aleatório; este último corresponde a uma estatística ordinal intermédia, tomando em consideração a informação crescente acerca da cauda direita, disponibilizada pelo topo da amostra, e permitindo o aumento da dimensão amostral de uma forma

natural. Esta análise para interpretação dos dados de valores extremos representa um importante desafio para o nosso conhecimento em aplicações práticas a conjuntos de dados do mundo real, em áreas como o ambiente, climatologia ou finanças.

2. Preliminares e notação

Quando estamos interessados em modelar grandes observações, confrontamo-nos usualmente com duas classes de modelos para valores extremos:

- A distribuição Generalizada de Valores Extremos (GEVfd)

$$G_\gamma(x) = \begin{cases} \exp\left[-(1+\gamma x)^{-1/\gamma}\right], & \text{para } 1+\gamma x > 0, \text{ se } \gamma \neq 0 \\ \exp(-\exp(-x)), & \text{para } x \in \mathbb{R}, \text{ se } \gamma = 0 \end{cases}$$

- A distribuição Generalizada de Pareto (GPfd)

$$H_\gamma(x) = \begin{cases} 1 - (1+\gamma x)^{-1/\gamma}, & 1+\gamma x > 0, x \in \mathbb{R}^+ \text{ se } \gamma \neq 0 \\ 1 - \exp(-x), & x \in \mathbb{R}^+ \text{ se } \gamma = 0 \end{cases}$$

Note-se que $\lim_{\gamma \rightarrow 0} G_\gamma(x) = G_0(x)$, o *modelo Gumbel*, e $\lim_{\gamma \rightarrow 0} H_\gamma(x) = H_0(x)$, o *modelo exponencial*. A introdução de parâmetros de escala $\delta > 0$ e de localização $\lambda \in \mathbb{R}$, resulta nas famílias *completas* de GEVfd e GPfd, respectivamente $G_\gamma(x; \lambda, \delta) = G_\gamma((x-\lambda)/\delta)$ e $H_\gamma(x; \lambda, \delta) = H_\gamma((x-\lambda)/\delta)$ que desempenham um papel central na inferência estatística de valores extremos.

GEVfd e MAX-Domínio: O artigo fundamental de Gnedenko (1943) estabelece a GEVfd como uma versão unificada de todas as leis limite não-degeneradas do máximo convenientemente normalizado de sequências de v.a.'s i.i.d. X_1, X_2, \dots, X_n com f.d. comum F . Isto é, no caso de existência de sucessões normalizadoras $a_n > 0$ e $b_n \in \mathbb{R}$, tais que, para todo x

$$P\left[\max(X_1, X_2, \dots, X_n) \leq a_n x + b_n\right] = F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G_\gamma(x).$$

Dizemos nestas circunstâncias que F pertence ao domínio de atracção de G_γ e denotamos este facto por $F \in \mathcal{D}(G_\gamma)$. Para $\gamma < 0$, $\gamma = 0$ e $\gamma > 0$ a f.d. G_γ reduz-se às f.d.'s de Weibull, Gumbel e Fréchet, respectivamente.

GPfd e POT-Domínio: O uso da GPfd é sugerido pelo resultado de Balkema e de Haan (1974) e Pickands III (1975), os quais provaram que $F \in \mathcal{D}(G_\gamma)$ se e só se a cauda superior de F está, sob certo sentido, perto da cauda superior de H_γ . Enquanto restringindo a atenção à porção de topo da amostra, a GPfd surge como interveniente, uma vez que aparece como a distribuição limite para os excessos $Y_i := X_i - u > y | X_i > u$, $i = 1, 2, \dots, k$, acima de um patamar suficientemente elevado u (método POT). Para $\gamma < 0$, $\gamma = 0$ e $\gamma > 0$ a f.d. H_γ reduz-se às f.d.'s de Beta, Exponencial e Pareto, respectivamente. Em ambos os casos, o índice de valores extremos γ está intimamente relacionado ao peso da cauda de F . Nesse sentido, o valor $\gamma = 0$ (*cauda exponencial*) pode ser encarado como um ponto de viragem: $\gamma < 0$ refere-se a caudas curtas com limite superior do suporte $x_F = \sup\{x, F(x) < 1\}$ finito, enquanto que para $\gamma > 0$ as f.d.'s são de cauda pesada.

Em muitas ciências aplicadas onde os extremos surgem com papel relevante, é assumido que o índice $\gamma = 0$ e os procedimentos de inferência estatística respeitantes a acontecimentos raros na cauda de F , tais como a estimação de pequenas probabilidades de excedência ou de períodos de retorno, assentam nesta suposição. Além disso, os modelos Gumbel e exponencial são também preferidos devido à maior simplicidade de inferência associada a este tipo de populações.

3. Testes em Valores Extremos

Para analisar valores extremos existem diferentes abordagens, de acordo com as suposições subjacentes para F e as observações específicas retiradas dentro da amostra disponível tendo por propósito uma

posterior inferência. No que se segue, $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ são as e.o.'s associadas a X_1, X_2, \dots, X_n .

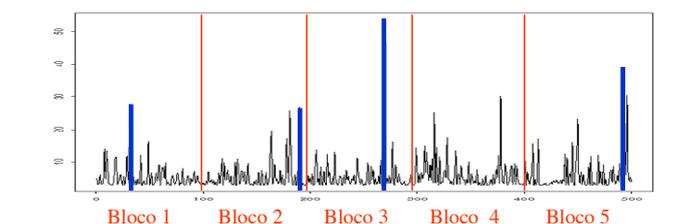
3.1 Abordagens Paramétricas

Neste contexto, a suposição principal é a da existência de uma classe adequada de modelos que descrevam convenientemente a característica aleatória sob estudo. Estas três classes possíveis são motivadas pela TVE, e dependem fundamentalmente do parâmetro γ , e eventualmente de parâmetros de localização λ e escala δ .

• Método de Gumbel ou dos Máximos Anuais (MA):

Suponhamos que é possível obter o máximo para cada um dos blocos de observações igualmente espaçadas. A classe GEVfd, G_γ , é usada para modelar o s máximos dessas k subamostras construídas a partir de um conjunto de dados de dimensão kn , para n suficientemente elevado,

$$Z_i := \max\{X_1^{(i)}, \dots, X_n^{(i)}\}, i = 1, \dots, k.$$



É prática comum neste método clássico de Gumbel tomar os máximos anuais, o que justifica esta designação. Neste contexto MA, o seguinte problema de teste foi tratado extensivamente na literatura, com ênfase principal no teste da hipótese Gumbel para a f.d. dos $\{Z_i\}_{i=1, \dots, k}$

$$H_0 : \{G_\gamma : \gamma = 0\} \quad \text{vs.} \quad H_1 : \{G_\gamma : \gamma \neq 0\}.$$

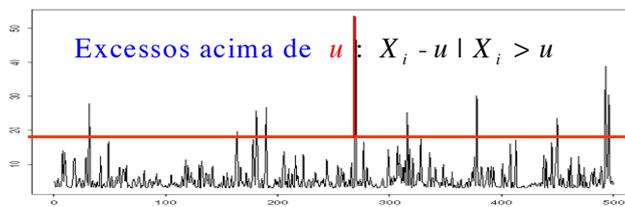
De facto, os hidrologistas têm desde há muito feito uso das distribuições de valores extremos para estimar probabilidades de inundações e a correcta escolha da GEVfd é muito importante, já que os três tipos diferem consideravelmente nas respectivas caudas direitas.

Entre os artigos que se referem a este tema, ou tendo por alternativa hipóteses unilaterais do tipo Weibull $\{G_\gamma : \gamma < 0\}$ ou Fréchet $\{G_\gamma : \gamma > 0\}$, referimos Van Montfort (1970), Bardsley (1977), Otten and Van Montfort (1978), Tiago de Oliveira (1981), Gomes (1982), Tiago de Oliveira (1984), Tiago de Oliveira and Gomes (1984), Hosking (1984), Marohn (1994), Wang et al. (1996) and Marohn (2000). De certa forma relacionado com este problema estão os testes de ajustamento para o modelo Gumbel, o qual recebeu atenção em Stephens (1976), Stephens (1977), and Stephens (1986). Os testes aí considerados são baseados em estatísticas-FDE, como as de Kolmogorov, Cramér-von Mises e Anderson-Darling.

• Método de Excessos de Nível (POT):

Podem acontecer que, ao considerarmos a divisão dos dados de forma anual, alguns anos contenham vários valores que são maiores do que o máximo de outros anos. Este tipo de situação motiva a seguinte abordagem: suponhamos que extraímos as observações da amostra que excedem um nível fixo u . A classe da GPfd, H_γ , é usada quando os excessos são retirados das k excedências de u de um certo conjunto amostral; i.e.,

$$Y_i := X_i - u > y | X_i > u, i = 1, 2, \dots, k \equiv k_u.$$



Uma clara diferença entre os contextos MA e POT é a de que enquanto na segunda as k maiores observações da amostra são tomadas em consideração, na primeira os k máximos anuais não são necessariamente as maiores observações na amostra.

Neste contexto POT o seguinte problema de teste tem vindo a ser considerado, dando prioridade à hipótese Exponencial para distribuição associada aos excessos $\{Y_i\}_{i=1,\dots,k}$:

$$H_0 : \{H_\gamma : \gamma = 0\} \quad \text{vs.} \quad H_1 : \{H_\gamma : \gamma \neq 0\}.$$

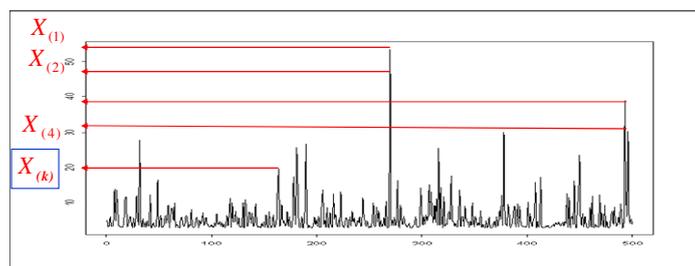
Este teste de hipóteses recebeu igualmente atenção por parte dos hidrologistas; de entre os artigos dedicados a este problema de teste de H_0 vs. H_1 , ou contra alternativas unilaterais Beta $\{H_\gamma : \gamma < 0\}$ ou Pareto $\{H_\gamma : \gamma > 0\}$, mencionamos Van Montfort and Witter (1985), Gomes and Van Montfort (1986) e mais recentemente Brilhante (2004). Em Marohn (2000) ambas abordagens MA e POT são consideradas, no contexto de escolha estatística para as classes GEVfd e GPfd, salientando a propriedade de normalidade assintótica local (LAN) de forma a obter testes uniformemente óptimos para alternativas unilaterais e bilaterais. O ajustamento dos dados à GEVfd foi desenvolvido Castillo and Hadi (1997), tendo sido também estudado em Choulakian and Stephens (2001). Recentemente, e de certa forma relacionado com o anterior, Beirlant et al. (2006) dedicaram-se ao problema de ajustamento para a classe de distribuições de cauda pesada ou tipo-Pareto.

• **Método das Maiores Observações (MO):**

Consideremos agora que as k maiores observações da amostra, depois de convenientemente linearizadas com parâmetros de localização e escala, λ e δ , $\{Z_{(i)} := (X_{(i)} - \lambda) / \delta\}_{i=1,\dots,k}$ são bem modeladas com a f.d.p. conjunta dada por

$$f_\gamma(\mathbf{z}) := f_\gamma(z_{(1)}, \dots, z_{(k)}) = g_\gamma(z_{(k)}) \prod_{i=1}^{k-1} \frac{g_\gamma(z_{(i)})}{G_\gamma(z_{(i)})}, \quad z_{(1)} > \dots > z_{(k)},$$

onde $g_\gamma(z) := \partial G_\gamma(z) / \partial z$ é a f.d.p. da GEVfd.



Em geral, $f_\gamma(\mathbf{z})$ é a forma da f.d.p. conjunta limite não-degenerada das k e.o.'s de topo, de um conjunto de n i.i.d. v.a.'s, quando $n \rightarrow \infty$. Na literatura, aquele modelo recebeu a designação de *GEV-multivariado* ou de *GEV-processo extremal* e foi considerado em Gomes and Alpuim (1986) para a escolha estatística de G_0 contra $\{G_\gamma : \gamma \neq 0\}$, ou contra alternativas unilaterais $\{G_\gamma : \gamma < 0\}$ ou $\{G_\gamma : \gamma > 0\}$. Em Gomes (1989), o problema de testar a hipótese Gumbel, é abordado através de uma combinação de partição da amostra em blocos com a extracção das maiores k observações em cada um dos m blocos, através do que é o denominado modelo GEV-multidimensional, como se segue: um conjunto de vectores aleatórios k -dimensionais i.i.d. $\{X_i\}_{i=1,\dots,m}$ normalizados para λ e δ , onde a f.d.p. conjunta dos vectores $\{Z_{(i)} := (X_{(i)} - \lambda) / \delta\}_{i=1,\dots,k}$ é dada por $f_\gamma(\mathbf{z}_{(i)})$. Note-se que ambas abordagens MA e MO podem ser casos

particulares deste modelo multidimensional, tomando $k=1$ e $m=1$, respectivamente. Em Gomes (1987) é considerada uma amostra censurada dos maiores valores da amostra, cuja dimensão vai para infinito e cuja distribuição limite está na classe GEVfd, tendo por propósito testes de ajustamento. Usando a redução da Gumbel à exponencial, a autora desenvolve testes bilaterais de exponencialidade, para as v.a.'s transformadas, sendo focados os testes de ajustamento Kolmogorov-Smirnov, Cramér-von Mises e Stephens.

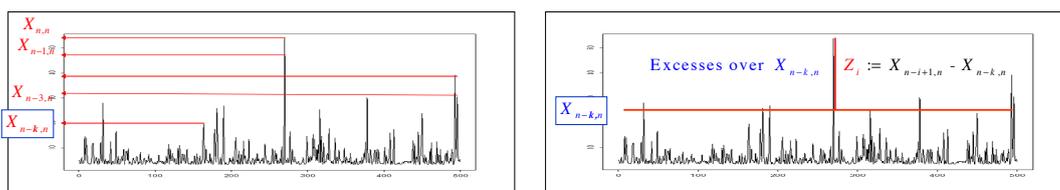
3.2 Abordagens Semi-Paramétricas

Seguindo uma abordagem semi-paramétrica, a única suposição é a de que $F \in \mathcal{D}(G_\gamma)$, para algum γ real. Neste contexto, qualquer inferência respeitante à cauda da distribuição subjacente F pode ser baseada nas k observações acima de um nível aleatório $X_{n-k,n}$. Recordemos que o domínio de atracção Fréchet contém distribuições com cauda polinomial, com limite superior do suporte x_F infinito, enquanto que as do domínio Weibull são de cauda curta e x_F finito. O caso intermédio de f.d.'s no domínio Gumbel abrange simultaneamente x_F finito ou infinito. O problema de escolha estatística neste contexto semi-paramétrico pode ser explicitado através de

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs.} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0}$$

ou contra alternativas unilaterais $F \in \mathcal{D}(G_\gamma)_{\gamma < 0}$ (Domínio Weibull) ou $F \in \mathcal{D}(G_\gamma)_{\gamma > 0}$ (Domínio Fréchet). Testes dirigidos a este problema surgiram na literatura a partir dos artigos de Galambos (1982) e Castillo et al. (1989). Este último consiste num procedimento engenhoso, motivado pelo método visual para escolha da distribuição assintótica pela curvatura da nuvem de pontos associada a um papel de probabilidades. Escolhendo quatro ou mesmo três quantis extremos superiores empíricos, dois declives podem ser determinados e mostra-se que a razão entre estes (que não depende de localização/escala) converge para o ratio de combinações lineares com soma nula, dos quantis limite, o que justifica aquela técnica visual. Outros procedimentos para escolha estatística de max-domínios de atracção podem ser igualmente sugeridas pelos artigos de Hasofer and Wang (1992), Falk (1995), Fraga Alves and Gomes (1996), Fraga Alves (1999), Marohn (1998a,b), Segers and Teugels (2000) e, mais recentemente, Neves *et al.* (2006) e Neves and Fraga Alves (2007).

Nestes últimos dois artigos, os procedimentos de teste são baseados nas observações que excedem um certo nível aleatório $X_{n-k,n}$, através dos excessos $\{Z_i := X_{n-i+1,n} - X_{n-k,n}\}_{i=1,\dots,k}$.



Este contexto representa alguma analogia com a abordagem POT, mas em que $X_{n-k,n}$ joga o papel do nível determinístico u , o que justifica a designação PORT.

3.3 Testes para condições Valores Extremos

Recentemente, testes para $F \in \mathcal{D}(G_\gamma)$, para algum γ real, têm ganhado bastante interesse. Dietrich et al. (2002) e Drees *et al.* (2006) consideraram este caso geral de teste para condições de Valores Extremos, através da adaptação de testes de ajustamento do tipo Kolmogorov-Smirnov e Cramér-von Mises.

BIBLIOGRAFIA

Fraga Alves, M. I. e Neves, C. (2006). Testing extreme value conditions – an overview and recent approaches. In papers (electronic) of *International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo*. June 28-30, 2006. Available at: <http://www.uclm.es/actividades0506/congresos/icmsm2006/articles/AlvesN06.pdf>.

REFERÊNCIAS

- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2:792-804.
- Bardsley, W. E. (1977). A test for distinguishing between extreme value distributions. *Journal of Hydrology*, 34:377-381.
- Beirlant, J., de Wet, T., and Goegebeur, Y. (2006). A goodness-of-fit statistic for Pareto-type behaviour. *Journal of Computational and Applied Mathematics*, 186:99-116.
- Brilhante, M. F. (2004). Exponentiality versus generalized Pareto – a resistant and robust test. *RevStat*, 2(1):1-13.
- Castillo, E., Galambos, J., and Sarabia, J. M. (1989). The selection of the domain of attraction of an extreme value distribution from a set of data. In J. Hausler and R.-D. Reiss, eds., *Extreme value theory (Oberwolfach, 1987) -Lecture Notes in Statistics 51*, pp. 181-190. Springer, Berlin-Heidelberg.
- Castillo, E. and Hadi, A. S. (1997). Fitting the Generalized Pareto distribution to data. *Journal of the American Statistical Association*, 92:1609-1620.
- Choulakian, V. and Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 43(4):478-484.
- Dietrich, D., de Haan, L., and Husler, J. (2002). Testing extreme value conditions. *Extremes*, 5:71-85.
- Drees, H., de Haan, L., and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Journal of Statistical Planning and Inference*, 136, 3498-3538.
- Falk, M. (1995). On testing the extreme value index via the POT-method. *Annals of Statistics*, 23:2013-2035.
- Fraga Alves, M. I. (1999). Asymptotic distribution of Gumbel statistic in a semi-parametric approach. *Portugaliæ Mathematica*, 56(3):282-298.
- Fraga Alves, M. I. and Gomes, M. I. (1996). Statistical choice of extreme value domains of attraction - a comparative analysis. *Communications in Statistics- Theory and Methods*, 25(4):789-811.
- Galambos, J. (1982). A statistical test for extreme value distributions. In e. B.W. Gnedenko et al., ed., *Nonparametric Statistical Inference*, pp. 221-230. North Holland, Amsterdam.
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44:423-453.
- Gomes, M. I. (1982). A note on statistical choice of extremal models. In Proc. IX Jornadas Mat. Hispano-Lusas, Salamanca. 653-655.
- Gomes, M. I. (1987). Extreme value theory – statistical choice. In Colloq. Math. Soc. Janos Bolyai 45, pp. 195-210. Debrecen.
- Gomes, M. I. (1989). Comparison of extremal models through statistical choice in multidimensional backgrounds. In J. Hausler and R.-D.Reiss, eds., *Extreme value theory (Oberwolfach, 1987) - Lecture Notes in Statistics 51*, pp. 191-203. Springer, Berlin-Heidelberg.
- Gomes, M. and Alpuim, M. (1986). Inference in a multivariate generalized extreme value model – asymptotic properties of two test statistics. *Scandinavian Journal of Statistics*, 13:291-300.
- Gomes, M. I. and Van Montfort, M. A. J. (1986). Exponentiality versus generalized Pareto-quick tests. In Proc. III Internat. Conf. Statistical Climatology, pp. 185-195.
- Greenwood, M. (1946). The statistical study of infectious diseases. *Journal of the Royal Statistical Society, Series A*, 109:85-109.
- Hasofer, A. and Wang, J. Z. (1992). A test for extreme value domain of attraction. *Journal of the American Statistical Association*, 87:171-177.
- Hosking, J. (1984). Testing whether the shape parameter is zero in the generalized extreme value distri-

- bution. *Biometrika*, 71:367-374.
- Marohn, F. (1994). On testing the exponential and Gumbel distribution. In J. Galambos, ed., *Extreme Value Theory*, pp. 159-174. Kluwer, Dordrecht.
- Marohn, F. (1998a). An adaptive efficient test for Gumbel domain of attraction. *Scandinavian Journal of Statistics*, 25:311-324.
- Marohn, F. (1998b). Testing the Gumbel hypothesis via the POT-method. *Extremes*, 1:2:191-213.
- Marohn, F. (2000). Testing extreme value models. *Extremes*, 3:363-384.
- Neves, C. and Fraga Alves, M. I. (2007). Semi-parametric approach to Hasofer-Wang and Greenwood statistics in extremes. Test (to appear).
- Neves, C., Picek, J., and Fraga Alves, M. I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction. *Journal of Statistical Planning and Inference*, 136:4:1281-1301.
- Otten, A. and Van Montfort, M. (1978). The power of two tests on the type of distribution of extremes. *Journal of Hydrology*, 37:195-199.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119-131.
- Segers, J. and Teugels, J. (2000). Testing the Gumbel hypothesis by Galton's ratio. *Extremes*, 3:3:291-303.
- Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Annals of Statistics*, 4:357-369.
- Stephens, M. A. (1977). Goodness-of-fit for the extreme value distribution. *Biometrika*, 64:583-588.
- Stephens, M. A. (1986). Tests for the exponential distribution. In R. B. D'Agostinho and M. A. Stephens, eds., *Goodness-of-Fit Techniques*, pp. 421-459. Marcel Dekker, New York.
- Tiago de Oliveira, J. (1981). The selection of the domain of attraction of an extreme value distribution from a set of data. In C. Taillie, ed., *Statistical Distributions in Scientific Work*, Vol. 6, pp. 367-387. D. Reidel, Dordrecht.
- Tiago de Oliveira, J. (1984). Univariate extremes: Statistical choice. In J. Tiago de Oliveira, ed., *Statistical Extremes and Applications*, pp. 91-107. D. Reidel, Dordrecht.
- Tiago de Oliveira, J. and Gomes, M. I. (1984). Two statistics for choice of univariate extreme value models. In J. Tiago de Oliveira, ed., *Statistical Extremes and Applications*, pp. 651-668. D. Reidel, Dordrecht.
- Van Montfort, M. A. J. (1970). On testing that the distribution of extremes is of type i when type ii is the alternative. *Journal of Hydrology*, 11:421-427.
- Van Montfort, M. A. J. and Witter, J. V. (1985). Testing exponentiality against generalized Pareto distribution. *Journal of Hydrology*, 78:305-315.
- Wang, J. Z., Cooke, P., and Li, S. (1996). Determination of domains of attraction based on a sequence of maxima. *Australian Journal of Statistics*, 38:173-181.



Models for Spatial Extremes

P. de Zea Bermudez ¹, J. Mendes ² and K.F. Turkman ³

Let $X(\mathbf{s})$ be a random function with \mathbf{s} taking values in some $D \subset R^d$. What can we say about the extremes of such random functions? In particular, if we have data $\{x(\mathbf{s}_i), i = 1, 2, \dots, n\}$, how can we make inferences on the extremes of such processes?

If $d = 1$, and D is a discrete subset of R so that $X(\mathbf{s})$ is a discrete time series, then the Extreme Value Theory (EVT) answers these questions almost completely. In fact, the EVT provides statistical models for the tail of a probability distribution and also offers specific statistical techniques to make inference about these models. If $\{X_i\}$ are iid random variables with distribution function F then the maxima $M_n = \max(X_1, \dots, X_n)$ suitably normalized converges, as $n \rightarrow \infty$, to one of the three extreme value distributions: Weibull, Gumbel or Fréchet. In this situation it is said that F is in the domain of attraction of one of these extreme value distributions. Because of this asymptotic justification, the Generalized Extreme Value distribution (GEVD) is often used as model for maxima of blocks, such as annual maxima, to estimate tail probabilities. However, this method does not use the data efficiently and many large observations which could also be used in inference are ignored. Another method, which is called Peaks Over Threshold (POT) method is based on modeling the excesses (or exceedances) over a suitably chosen high threshold. The generalized Pareto distribution (GPD) is then used to model the excesses over this high threshold and the theoretical asymptotic justification is given by Pickands(1975). Hence, the tail of the conditional distribution is modeled by the GPD as follows:

$$P(X > x + u | X > u) = \left(1 + k \frac{x}{\sigma}\right)^{-1/k}, \quad (1)$$

where $\sigma > 0$, $k \in (-\infty, \infty)$ and $1 + k \frac{x}{\sigma} > 0$. Here, σ is the scale parameter and depends on the chosen threshold u . The k is called the shape parameter and is invariant to threshold u selected. The choice of the threshold is critical and has to compromise the availability of sufficient amount of data for statistical estimation and the asymptotic theoretical justification. Consequently, it is generally chosen sufficiently high so that the excesses follow approximately a GPD distribution and low enough in order to have an adequate amount of excesses to perform the statistical analysis. For an extensive treatment of the EVT, see Embrecht et al. (1997) and for a review of methods for parameter estimation of the GPD, see de Zea Bermudez and Kotz (2006).

The theory of the POT method is carried out under the assumption that the data is independent and identically distributed and that the limiting GPD distribution is solely characterized by the tail behavior of the marginal distribution F . When dealing with non-stationary data with strong short term serial dependence, the following aspects need to be considered in order to use the POT method:

1. the non-stationarity has to be removed.
2. the threshold exceedances need to be de-clustered in order to identify independent clusters of exceedances.

For addressing the non-stationarity issue, several methods have been suggested in the literature (see, for example, Smith(1989) and Ramesh and Davison(2002)).

¹CEAUL-UL, pcbermudez@fc.ul.pt

²ISEGI-UNL, jmm@isegi.unl.pt

³CEAUL-UL, kfturkman@fc.ul.pt

The usual method of de-clustering is done by estimating the extremal index θ which is an asymptotic parameter measuring the strength of clustering of the extreme values of the series. The GPD model is then fitted to the sample of cluster maxima. For the estimation of the extremal index see Embrechts et al. (1997). In general, the extremal index θ is invariant of the threshold chosen for the GPD model. However, there are processes with short term dependence structures which contradict the invariance property (Ledford and Tawn, 2003), thus indicating that GPD may not be a good model for certain types of serially dependent data (see Eastoe and Tawn (2006 for alternative models).

As the above brief introduction indicates, the tools commonly used for dealing with non-stationary data and de-clustering for local extremes are well understood and studied for serially dependent data. However, for $d \geq 2$, that is when $X(\mathbf{s})$ is a random field, and the data is of the form $\{x(\mathbf{s}_i), \mathbf{s}_i \in D \subset R^d\}$, statistical techniques for dealing with the extremes of such data sets are not very well known. Figure 1 gives the locations of wild fires in Portugal above 5 and 250 hectares respectively in 2004. How can one model the locations, as well as the sizes of these events

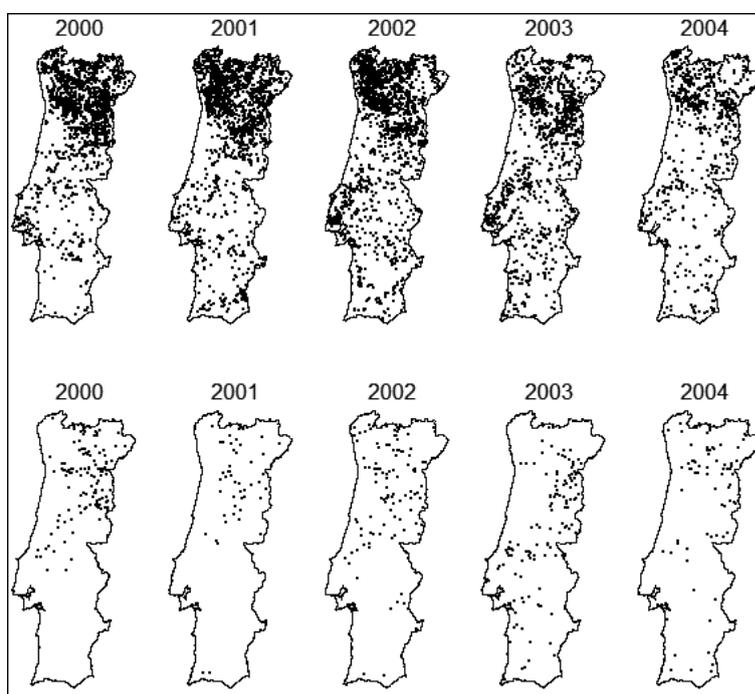


Figure 1: Locations of fires above 5 hectares and locations of fires above 250 hectares

so that one quantifies the risk of large fires at a given time and space? The natural way handle such data set is to de-trend and de-seasonalize the data in time and space to get fairly stationary data, to identify the spatial clusters of large values and finally to fit a GPD model to the maxima of spatial clusters. However, apart from appropriate methods for transforming non-stationarity (which will not be discussed here) one needs to make inferences on the average cluster size or on the spatial extremal index of the field. However, it is not clear what spatial extremal index is. Turkman (2006) gives one possible characterization of spatial extremal index through coordinate wise representation by looking at how three different sequences cluster in different directions. First, at each time point k , the sequence of time plate maxima

$$\zeta_k = \max_{0 \leq i \leq r_1, 0 \leq j \leq r_2} X(i, j, k),$$

are formed. Then, for each fixed k and i , the sequence of column maxima

$$\nu_i = \max_{0 \leq j \leq r_2} X(i, j, 0)$$

are formed. The characterization of clusters for these processes along k -direction and along x -direction respectively, as well as the clustering of $X(0, j, 0)$ along the j -direction can be made in the usual way and Turkman (2006) gives the extremal index of the field in terms of these individual coordinate indexes. However, in order to use these results effectively for spatial-temporal data sets, adequate methods of extremal index estimation should be formulated. Although the adaptation of the runs method to estimate each of the coordinate indexes seems to be straightforward, there are some complicated issues in the choice of a proper threshold for the analysis.

Other possible methods of handling spatial extreme data are suggested by Davison (2007), which we explain now.

In modeling spatial extremes, de Haan and Pereira (2006) suggest starting from the stationary spatial process $X(\mathbf{s}, t)$ traveling in time and look at the marginal convergence

$$\frac{1}{a_T}(\sup_{t \in T} X(\mathbf{s}, t) - b_T) \rightarrow^W Y(\mathbf{s}).$$

The limiting process $Y(\mathbf{s})$ is a max-stable process (see de Haan and Lin(2001)). This max-stable process can then be used as an asymptotic model for the extremal values of the spatial process $X(\mathbf{s})$. If one can derive or make inference on the probability structure of Y , the problem of extrapolation of large values in space to unobserved sites and/or inference on the largest value $\sup_{\mathbf{s} \in S} X(\mathbf{s})$ can be solved. De Haan and Pereira (2006) then look at specific models for the extremal process $Y(\mathbf{s})$ to address the issue of spatial extrapolation. Davis and Mikosch (2006) look at the same problem, but start with a heavy-tailed linear space-time process, continuous in space and discrete in time, and then derive the characteristic behavior of the limiting extremal process. This extremal process is uniquely represented by the model suggested by de Haan and Pereira (2006). Similar models were already suggested by Coles and Tawn (1996) for modeling areal rainfall processes. The problem with this possible method is that only the bivariate distributions of the $Y(\mathbf{s})$ process are known. How can we make inference on the whole probability structure of the $Y(\mathbf{s})$ process based on bivariate distributional models? Davison(2007) suggests using likelihood methods. In principle, for likelihood inference, the full likelihood

$$f(x(\mathbf{s}_1), \dots, x(\mathbf{s}_n)|\Theta)$$

is needed. However, one can make use of the composite likelihood based on the bivariate densities

$$\sum_{i>j} \log f(y(\mathbf{s}_i, \mathbf{s}_j)|\Theta),$$

where $f(y(\mathbf{s}_i, \mathbf{s}_j)|\Theta)$ is a model for the bivariate density of $Y(\mathbf{s})$ at two locations \mathbf{s}_i and \mathbf{s}_j . It is known that if Θ is identifiable from the bivariate densities, then under fairly general conditions, the estimator obtained from () is consistent, asymptotically Gaussian with known covariance structure, giving a good approximation to the estimators obtained from the full likelihood. Davison (2007) also gives suggestions as to how bivariate models can be identified using the data. Even though this method looks very promising, it has one drawback: often spatial variations in the extremes are due to many other factors or covariates which are observable. Therefore, it is highly desirable to incorporate this information in the model. Bayesian Hierarchical Models and the simulation based inference techniques, such as MCMC seem to be the best way of handling such situations. Therefore, we suggest a Bayesian Hierarchical modeling framework as the third alternative for modeling spatial extremes. The basic fundamental assumption taken in Bayesian hierarchical modeling is that, although the extremes are dependent in space (and in time), they are conditionally independent given a hidden spatial temporal random process, as well as other observable covariates. How reasonable this assumption highly depends on each case study and also on the quality of the auxiliary information contained in the covariates.

Thus, we assume that the data

$$\mathbf{x}(\mathbf{s}) = \{x(\mathbf{s}_i, t), i = 1, 2, \dots, n_t, t = 1, 2, \dots, T\}$$

is the realization of a Spatio-temporal process-continuous in space and discrete in time- $X(\mathbf{s}, t)$, $\mathbf{s} = (s_1, s_2) \in D \subset R^2$, $t = 1, 2, \dots$. Such data are called geostatistical (Cressie, 1993), or point referenced (Banarjee, 2004) data.

Let

$$Z(\mathbf{s}_i, t) = X(\mathbf{s}, t) - u,$$

be the excesses over the threshold u .

We consider the following hierarchical model for $Z(\mathbf{s}_i, t)$

1. Level 1: Likelihood

we assume that, conditional on a hidden unobserved spatial process, $\eta(\mathbf{s})$,

$$p(z(\mathbf{s}_i, t), i = 1, 2, \dots, n_t, t = 1, \dots, T | \eta(\mathbf{s})) = \prod_{t=1}^T \prod_{i=1}^{n_t} p(z(\mathbf{s}_i, t) | \eta(\mathbf{s}), \Theta), \quad (2)$$

where

$$\begin{aligned} p(z(\mathbf{s}_i, t) | \eta(\mathbf{s}), \Theta) &= GPD(k(\mathbf{s}_i), \sigma(\mathbf{s}_i)) \\ &= \frac{1}{\sigma(\mathbf{s}_i)} \left(1 + \frac{k(\mathbf{s}_i)}{\sigma(\mathbf{s}_i)} z \right)^{-1-1/k(\mathbf{s}_i)} \end{aligned} \quad (3)$$

Here, n_t is the number of exceedances above threshold u in year t .

2. Level 2: Link functions

$$\log \sigma(\mathbf{s}_i) = \mu_\sigma(\mathbf{s}) = a_0 + a_1 u + a_2 s_1 + a_3 s_2 + \eta(\mathbf{s}_i) + \epsilon_\sigma(\mathbf{s}_i), \quad (4)$$

$$\log \frac{k(\mathbf{s}_i)}{1 - k(\mathbf{s}_i)} = b_0 + b_1 s_1 + b_2 s_2 + \eta(\mathbf{s}_i) + \epsilon_k(\mathbf{s}), \quad (5)$$

In general, the shape parameter k ranges in $(-\infty, \infty)$. However, in the case of extreme fire sizes we have a very precise prior information that the fires sizes are heavy tailed but with finite mean, thus restricting the range of k in $(0, 1)$.

3. Level 3: Parameters $\Theta = (a_0, a_1, a_2, b_0, b_1, b_2, c, \phi, \tau, \tau_1, \tau_2)$ are random hyper parameters, where

- $\eta(\mathbf{s})$ is a isotropic Gaussian spatial process with mean 0 and exponential covariance matrix Σ give by

$$\Sigma_{ij} = c^2 \exp(-\phi d_{ij}) + \tau^2 \mathbf{1}_{\{i=j\}}, \quad (6)$$

where d_{ij} is the distance between \mathbf{s}_i and \mathbf{s}_j , and $\mathbf{1}$ is the indicator function.

- $\epsilon_\sigma(\mathbf{s})$ and $\epsilon_k(\mathbf{s})$ are independent and identically distributed random sequences, independent of each other with means 0 and variances τ_1 and τ_2

4. level 4: Hyper-parameters $\{a_j\}_{j=0}^3, \{b_j\}_{j=0}^2$ are iid normal variables independent of each other, $c, \tau, \phi, \tau_1, \tau_2$ are positive, independent random variables

This model and a slightly simpler version based on areal models (Banarjee et al. (2004)), together with the MCMC inference are fitted to extremes of wild fire data in Portugal during 1984-2004 and the results are very encouraging (de Zea Bermudez et al. (2007)); see also Cooley et al. (2006, 2007) for similar applications. It is therefore our opinion that the most practical way of handling spatial extremes is through Bayesian Hierarchical modeling. Of course, these models are not free of criticism. For example, the spatial dependencies are only introduced through the model parameters, not directly through the likelihood. Also, they are open to questions due to the fact that the posterior marginal distributions of the largest values may not be an extreme value

distribution. However, due to their flexibility, they may prove to be a very powerful and useful tool in extreme value theory for spatial data.

References

Banarjee, S., Carlin, B. P. and Gelfand, A. (2004). Hierarchical modeling and Analysis for spatial data. Chapman and Hall.

Coles, S. and Tawn, J. (1996) Modeling Extremes of the areal rainfall process. *JRSSB* 58:2, 329-347

Cooley, D., Neveau, P., Joneli, V. Rababtel, A. and D. Grancher (2005) A Bayesian hierarchical extreme value model for lichenometry. *Environmetrics*. In press.

Cooley, D., Nychka, D. and P. Naveau (2006) Bayesian spatial model of extreme precipitation return levels. Submitted.

Cressie(1993) statistics for Spatial Data. John Wiley New York.

Davidson (2007) Towards a Geostatistics of Extremes. Workshop on Statistical Extremes and Environmental Risk, Lisbon 15-18, 2007.

Davis, R.A. and Mikoch, T. (2006) Extreme value theory for space-time processes with heavy tailed distributions. Preprint.

de Haan, L. and Lin, T. (2001) On convergence toward an extreme value distribution in $C[0, 1]$. *Ann. probab.* 29, 467-483.

de Haan, L and T. Pereira (2006) Spatial extremes: models for the stationary case. *The Annals of Statistics*, 34, 146-168.

de Zea Bermudez, P. and S. Kotz (2006) Parameter estimation of the Generalized Pareto Distribution Technical Report 6 , CEAUL.

de Zea Bermudez, P., Mendes, J., K.F. Turkman and Pereira, J. (2007) Spatial extremes of wildfire sizes, In preparation.

Eastoe, E.M. and J. Tawn(2006) The distribution for cluster maxima of exceedances of sub-asymptotic thresholds. *JRSSB*. Submitted.

Embrechts, P. Kluppelberg, C. and T. Mikosch(1997) Modelling Extremal Events, Springer

Ledford, A.W. and J. Tawn(2003) Diagnostics for dependence within time series extremes. *JRSSB* 65:2, 521-543

Pickands,J.(1975) Statistical Inference using extreme order statistics, *Ann. Statist.*, 3:119-131

Ramesh, N.I. and Davison, A. (2002) Local models for exploratory analysis of hydrological extremes. *J. of Hydrology*, 256, 106-119.

Smith, R. (1989) Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science* 4: 367-393.

Turkman, K.F. (2006) A note on t he extremal index for space-time processes. *J. Appl. Prob.* 43, 114-126.



Semiestabilidade em Teoria de Extremos

Maria da Graça Temido, *mgtm@mat.uc.pt*
Universidade de Coimbra

Mirei-a e admirei-a pela primeira vez na Primavera de 1996. Era uma criança desconhecida mas com um futuro auspicioso. Nascida na Bulgária, fizera apenas uma viagem à Rússia. Difícil de compreender, reclamava por atenção. A sua única prima ocidental, habituada aos mimos dos apaixonados pelos extremos, mal a conhecia. Ao olhar para ela apercebemo-nos de que a sua solidão estava associada aos traços das suas corcundas e ao humor periódico e limitado. Por ser proveniente de uma família bem mais abastada que a da sua prima famosa, embora com muitos membros convenientemente discretos, foi atraindo atenções. Na Primavera de 1998 eu e a Professora Luisa Canto e Castro já conseguíamos conversar com ela com bastante segurança. Conversar não bastava. Dadas as suas formas pouco convencionais resolvemos trajá-la de forma alternativa, para o que contribuiu a ajuda primordial do Professor Laurens de Haan. Com este novo traje voltou à Rússia.

Ao longo dos anos, as suas formas, as suas corcundas, as sazonalidades do seu carácter e os seus membros discretos, fizeram aumentar o número de interessados em conhecê-la. Dominadora, conciliadora, tem já um domínio de atracção maior que o da prima. Foi aliás o seu carácter conciliador e discreto que lhe valeu nova ida à Rússia em 2005, desta vez, pela minha mão e da Professora Andreia Hall. Depois de ter revelado dar-se bem com um número aleatório de primas, eu e a Professora Helena Ferreira levámo-la também a Paris.

Em Fevereiro de 2007 é uma jovem cheia de potencialidades, a quem eu agradeço a amizade com que me tem presenteado. Principalmente o presente que me ofereceu pouco antes do Natal de 2000.

Estou a falar-vos da classe MSS. A classe das distribuições max-semiestáveis.

$$\longrightarrow F, k_n, r, \alpha, \nu, p, G \longleftarrow$$

A classe das leis max-estáveis sempre foi considerada por alguns estatísticos demasiado restrita para muitas das aplicações da Estatística de Extremos. Com efeito, existe um número considerável de funções de distribuição (f.d.'s) que não pertencem a qualquer domínio de atracção de f.d.'s max-estáveis. Por exemplo, as f.d.'s que, como a Binomial, verificam a condição $\lim_{x \rightarrow w_{\bar{F}}} F(x) < 1$ e outras, como a Binomial Negativa e a Gumbel discretizada, para as quais se tem

$$\lim_{n \rightarrow +\infty} \frac{1 - F(n-1)}{1 - F(n)} = r, \quad (1)$$

com r em $]1, +\infty[$. Assim, a procura de uma distribuição limite não degenerada para o máximo de variáveis aleatórias com função de distribuição (f.d.) discreta que verificasse (1) ou com f.d. contínua multimodal era, sem dúvida, uma questão de interesse. Citamos ainda o exemplo da f.d. multimodal definida por $F(x) = 1 - e^{-x - senx}$, para $x \geq 0$.

Respondendo às questões anteriores, embora porventura não motivada por elas, surgiu na Teoria de Extremos a classe das f.d.'s max-semiestáveis – classe MSS – que engloba a classe MS. As primeiras referências sobre esta classe são Pancheva (1992) e Grinevich (1992, 1993, 1994). Esta nova classe, constituída por todas as funções de distribuição para as quais existem reais $r > 1$, $a > 0$ e b tais que

$$G(x) = G^r(x/a + b), \quad (2)$$

coincide com a classe dos possíveis limites de $F^{k_n}(x/a_n + b_n)$, com $a_n > 0$ e b_n real, isto é, dos possíveis limites em distribuição do máximo, linearmente normalizado, de k_n variáveis aleatórias independentes e identicamente distribuídas, onde $\{k_n\}$ é uma sucessão de inteiros que verifica

$$k_{n+1} \geq k_n \geq 1 \quad \text{e} \quad \lim_{n \rightarrow +\infty} \frac{k_{n+1}}{k_n} = r, \quad (3)$$

com $r \geq 1$. No que se segue diremos que $\{k_n\}$ é uma sucessão de crescimento geométrico. As expressões analíticas das três famílias de f.d.'s max-semiestáveis podem ser unificadas na forma seguinte:

$$G(x) = \begin{cases} \exp\left\{-\left(1 + \gamma x\right)^{-1/\gamma} \nu(\log(1 + \gamma x))\right\} & x \in R, 1 + \gamma x > 0 \text{ e } \gamma \neq 0 \\ \exp\{-e^{-x} \nu(x)\} & x \in R \text{ e } \gamma = 0 \end{cases},$$

onde ν é uma função positiva periódica e limitada.

Notemos que se considerarmos ν constante ou $\{k_n\}$ tal que $r = 1$ obtemos a classe MS.

A classe MSS veio assim perspectivar um aumento significativo das aplicações em Teoria de Extremos, uma vez que contém f.d.'s descontínuas bem como uma vasta gama de f.d.'s multimodais.

O aparecimento do parâmetro r e da função periódica ν e a sua coexistência com os restantes parâmetros de localização e forma, para além de tornar mais complexa a caracterização dos domínios de atracção desta classe, o que foi conseguido em Grinevich (1993), conduz a um processo inferencial com várias adversidades. Na tentativa de encetar a inferência estatística da classe MSS, obtivemos uma caracterização diferente desta classe e dos seus domínios de atracção, bem mais simples do que a de Grinevich. De facto, desta nova representação concluímos que qualquer solução de (2) pode ser obtida a partir da igualdade $-\log(-\log G(a^{-m}x + s_m)) = m \log r + y(x)$, válida para qualquer x em $[0, 1]$ e qualquer m inteiro, onde $s_m = m$ se $a = 1$ e $s_m = (a^{-m} - 1)/(a^{-1} - 1)$ se $a \neq 1$ e $y : [0, 1] \rightarrow [0, \log r]$ é um função contínua à direita, não decrescente e contínua em $x = 1$. Resultados similares são obtidos para a inversa generalizada da função $-\log(-\log G)$. Esta representação dispensa o papel complexo da função periódica que surge na caracterização de Grinevich. Estes resultados enformam Canto e Castro, de Haan e Temido (2002).

O facto da classe MSS conter f.d.'s discretas levou-nos a procurar o comportamento das f.d.'s que pertencem aos seus domínios de atracção. Em Temido (2000) provámos que, para uma f.d. discreta F com sucessão de pontos de descontinuidade $\{x_n\}$, uma condição necessária e suficiente para que exista uma sucessão real $\{u_n\}$ tal que $F^{k_n}(u_n)$ tenha limite não degenerado é $\lim_{n \rightarrow +\infty} \frac{1 - F(x_{n-1})}{1 - F(x_n)} = r$, onde r e $\{k_n\}$ estão relacionados em (3). Este resultado permite-nos provar que se a f.d. F não tiver um salto no extremo superior do suporte e verificar a condição anterior, então só poderá pertencer ao domínio de atracção de uma max-semiestável discreta. Em particular, para qualquer f.d. inteira na classe de Anderson, isto é, com extremo superior do suporte infinito e a verificar (1), ou equivalentemente

$$\exp(-e^{-\alpha(x-1)}) \leq \liminf_{n \rightarrow +\infty} F^n(x + b_n) \leq \limsup_{n \rightarrow +\infty} F^n(x + b_n) \leq \exp(-e^{-\alpha x}),$$

se existirem k_n de crescimento geométrico e b_n tais que $F^{k_n}(x + b_n)$ tenha limite não degenerado, esse limite pertence à família da Gumbel discretizada $G(x) = \exp(-\beta r^{-[x]})$, $x \in R$. Destacamos o exemplo da f.d. Binomial Negativa $BN(m, p)$ que, neste novo contexto, pertence ao domínio de atracção de $G(x) = \exp(-p^{[x]})$, $x \in R$.

Em Temido e Canto e Castro (2002) consideramos sucessões estritamente estacionárias sujeitas a uma condição de independência assintótica que é uma generalização da condição $D(u_n)$ de Leadbetter e uma sucessão de crescimento geométrico. Para este tipo de sucessões, generalizámos o Teorema de Valores Extremos de Leadbetter, isto é, mostramos que a classe dos possíveis limites em distribuição do máximo, linearmente normalizado, das primeiras k_n variáveis aleatórias da sucessão coincide com a classe MSS. De modo a podermos estudar o comportamento extremal de sucessões estritamente estacionárias com f.d. comum no domínio de atracção de alguma max-semiestável, generalizámos o conhecido conceito de índice extremal. Introduzimos também algumas condições de dependência local, que são generalizações das já conhecidas para o caso $k_n = n$. Estas condições permitem-nos estabelecer condições necessárias e suficientes para a convergência em distribuição de M_{k_n} , sob normalização linear. A necessidade de encontrar modelos estacionários de interesse em aplicações reais e que verifiquem as novas condições de independência

assintótica e dependência local, assim como os resultados já referidos sobre a classe de Anderson, levou-nos a considerar um novo enquadramento para alguns dos modelos de margens inteiras estudados em Hall (1998). Com efeito, naquele trabalho são estudados vários modelos estacionários max-autorregressivos e de médias móveis de variáveis aleatórias inteiras onde a multiplicação é substituída por uma operação aleatória designada operação *thinning*. Para alguns destes modelos é provado que se a f.d. marginal verificar (1), então o máximo M_n , linearmente normalizado, tem um comportamento da forma

$$\exp(-\theta e^{-\alpha(x-1)}) \leq \liminf_{n \rightarrow +\infty} P(M_n \leq x + b_n) \leq \limsup_{n \rightarrow +\infty} P(M_n \leq x + b_n) \leq \exp(-\theta e^{-\alpha x}),$$

com $\theta \in]0, 1]$ e $\alpha > 0$, no caso de se verificarem certas condições de independência assintótica e dependência local então existentes na literatura. Surge assim em Hall e Temido (2005) o estudo do comportamento do máximo de modelos com f.d. marginal inteira no domínio de atracção da Gumbel discretizada, concretamente os modelos max-INAR(2) e INMA (*integer valued moving average*) e um modelo muito particular de médias móveis de ordem q . A verificação das novas condições de independência assintótica e de dependência local leva a concluir que o máximo das primeiras k_n variáveis também é atraído em distribuição para a Gumbel discretizada, sob a influência de um índice extremal.

Ao estudar o comportamento limite do máximo, linearmente normalizado, de k_n variáveis aleatórias independentes e, em geral, não identicamente distribuídas, introduzimos uma nova condição de uniformidade de máximos, que generaliza a que já era conhecida na literatura de extremos. Neste caso supomos apenas que k_n tende para infinito. Sob esta hipótese encontramos uma nova classe de distribuições limite para o máximo M_{k_n} . Esta nova classe contém a classe MSS e, obviamente, a classe M das f.d.'s log-côncavas de Meijzler. Atendendo à log-concavidade dos elementos da classe M, averiguámos esta propriedade para os elementos desta classe, razão pela qual lhe chamámos *log-semicôncavas*. Estes resultados encontram-se publicados em Temido (2003).

Em Ferreira e Temido (2005) estudamos a lei limite do máximo de K_n variáveis de uma sucessão estacionária, onde $\{K_n\}$ é uma sucessão de variáveis aleatórias inteiras positivas a verificar certas condições de convergência em probabilidade, de modo a estender os resultados de Ferreira (1994). Como limite de M_{K_n} , linearmente normalizado, obtém-se a esperança matemática da variável $G^{r,N}$, onde G é max-semiestável e N é uma variável inteira relacionada com o limite em probabilidade de K_n .

A iniciação à inferência estatística em modelos max-semiestáveis foi efectuada essencialmente em Temido, Gomes e Canto e Castro (2001). Foi fundamental começar por generalizar o Teorema de Pickands, isto é, provar que uma f.d. F , contínua, pertence a um domínio de atracção max-semiestável se e só se a f.d. dos excessos acima de certos níveis suficientemente elevados, pode ser convenientemente aproximada por uma f.d. pertencente a uma nova classe de Pareto Generalizada. A já referida caracterização dos domínios de atracção da classe MSS, que enforma Canto e Castro, de Hann e Temido (2002), mostrou-se bastante profícua na construção de um estimador convergente para o parâmetro a , na qual se supõe o conhecimento prévio do parâmetro r . A partir do estimador do parâmetro a obtemos um estimador convergente para o período da função periódica.

Com o retomar da actividade docente no Outono de 2000, após a defesa do meu doutoramento, concomitantemente com a preparação de alguns dos artigos referidos acima, a inferência estatística em modelos MSS, inequivocamente uma tarefa difícil e a precisar de meios computacionais diligentes, não foi de imediato reconsiderada. Todavia, os dotes intelectuais e humanos da Professora Luisa Canto e Castro permitiram-lhe encontrar mais uma aluna de doutoramento que logo se interessou pela estimação dos parâmetros da classe MSS. Trabalhadora, metódica e expedita a lidar com o Matlab, a Sandra Dias começou rapidamente a obter resultados. Em Dias e Canto e Castro (2004) as autoras retomaram a sucessão de estatísticas, introduzida em Temido (2000),

$$Z_s(m_N) := \frac{X_{(m_N/s)} - X_{(m_N)}}{X_{(m_N)} - X_{(m_N s)}}.$$

onde $X_{(m_N)} := X_{N-[m_N]+1:N}$ representa uma estatística de ordem de uma amostra de dimensão N e $\{m_N\}$ é uma sucessão de inteiros que verifica $\lim_{N \rightarrow +\infty} m_N = +\infty$ e $\lim_{N \rightarrow +\infty} m_N/N = 0$. Provaram que $\{Z_s(m_N)\}$ converge em probabilidade para a^c , com c natural, se e só se $s = r^c$. Por outras palavras, aquela sucessão de estatística só converge no caso em que s é igual ao verdadeiro valor do parâmetro r ou a uma sua qualquer potência natural. Motivadas por resultados de simulação bem planeados, as autoras propõem também um estimador do parâmetro r . Este estimador, conjuntamente com outras convergências decorrentes da convergência em probabilidade citada acima, permitem-lhes obter estimadores para o período da função ν e do parâmetro γ . Estes resultados, acompanhados pela correspondente ilustração computacional, encontram-se submetidos a revista internacional. Supondo a f.d. limite diferenciável e $\{m_n\}$ suficientemente bem escolhida, aquelas autoras provaram também a normalidade assintótica da sucessão $\{Z_s\}$. Depois de se terem dedicado a outras questões de interesse para o correcto conhecimento da classe MSS, borilando, em alguns casos, resultados de outros autores, estabeleceram uma generalização do conhecido Lema de Khintchine. Concretamente, é estudada a relação entre diferentes f.d.'s limite, e conseqüentemente dos seus parâmetros, quando se consideram diferentes sucessões de crescimento geométrico e diferentes constantes de normalização, dando origem a um conjunto de resultados dos quais também sou co-autora. A estimação da função y e de quantis elevados da f.d. limite max-semiestável G encontra-se publicada em Dias, Canto e Castro e Temido (2006). No trabalho "Max-estável ou Max-semiestável?" são propostos resultados de convergência de uma sucessão de estatísticas de teste, derivados a partir de resultados de Drees (1998) e de Drees, Ferreira e de Haan (2004), que de forma suficientemente robusta identifica uma f.d. max-semiestável. Este trabalho permitiu à Sandra Dias obter o Prémio SPE 2006.

No Verão de 2000 era minha convicção que a estimação da função periódica e dos parâmetros envolventes da classe MSS seria uma tarefa demasiado árdua. Pelos resultados obtidos e pelos obstáculos ultrapassados, os meus parabéns à Sandra Dias e à Professora Luisa Canto e Castro.

É meu entendimento que a classe MSS deve continuar a interessar outros autores da Teoria de Extremos. Com efeito, muitos são os problemas que surgem quando se pretende generalizar ou adaptar a vasta gama de resultados conhecidos nas diferentes áreas desta Teoria ao novo contexto gerado pela classe MSS. Pela complexidade que apresenta e pela falta de similitude com os resultados já conhecidos para a classe MS, a classe MSS abriu um leque surpreendente de campos de trabalho.

Bibliografia

- Canto e Castro, L., de Haan, L. e Temido, M.G. (2002). Rarely observed sample maxima. *Theory Prob. Appl.*, 45, 4, 779-782.
- Dias, S. e Canto e Castro, L. (2004). Contributos para a estimação em modelos max-semiestáveis. Em *Estatística com acaso e necessidade*. Editores, P. Rodrigues, E. Rebelo, F. Rosado, 177-187. Edições SPE.
- Dias, S. e Canto e Castro, L. (2005). Modelação de máximos em domínios de atracção max-semiestáveis. Em *Estatística Jubilar*. Editores Carlos Braumann e outros, 223-234. Edições SPE.
- Dias, S., Canto e Castro, L. e Temido, M. G. (2006). Estimação de quantis elevados de distribuições no domínio de atracção de leis max-semiestáveis. Em *Ciência Estatística*. Editores Luisa Canto e Castro e outros, 289-300. Edições SPE.
- Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.*, 25, 187-210.
- Drees, H., Ferreira, A. e de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability*, 14, 1179-1201.
- Ferreira, H. (1994), *Condições de Dependência Local em Teoria de Valores Extremos*, Tese de Doutoramento, Universidade de Coimbra.
- Ferreira, H. e Temido, M.G. (2005). Maximum of a random number of variables attracted to a max-semistable distribution. *Annales de L'I.S.U.P.*, Vol XLIX, I, 71-80.

Grinevich, I.V. (1992). Max-semistable limit laws under linear and power normalization, *Theory Prob. Appl.*, 37, 720-721.

Grinevich, I.V. (1993). Domains of attraction of the max-semistable laws under linear and power normalizations, *Theory Prob. Appl.*, 38, 640-650.

Hall, A. (1998). *Extremos de Sucessões de Contagem*, Tese de Doutoramento, Universidade de Lisboa.

Hall, A., Temido, M.G. (2005) On the maximum term of MA and Max-AR models with margins in Anderson's class. A publicar em *Theory Prob Appl.*

Temido, M.G. (2000). Leis Limite em Teoria de Valores Extremos - Estabilidade e Semi-estabilidade. Tese de Doutoramento. Universidade de Coimbra.

Temido, M.G. (2003). New limiting distributions of maxima of independent random variables. *Portugaliae Mathematica*, 60, 4, 379-388.

Temido, M.G. e Canto e Castro, L. (2003). Max-semistable laws in extremes of stationary random sequences. *Theory Prob. Appl.*, 47, 2, 365-374.

Temido, M.G., Gomes, M.I. e Canto e Castro, L. (2001). Inferência estatística em modelos max-semiestáveis. Em *Um olhar sobre a Estatística*. Editores Pedro Oliveira e Emília Athayde, 291-305. Edições SPE.



Memorial da Escola

Maria Ivette Gomes, *ivette.gomes@fc.ul.pt*
DEIO e CEAUL, FCUL, Universidade de Lisboa

O desenvolvimento da “*Escola de Extremos*” em Portugal teve sem dúvida como grande responsável a obra científica de Tiago de Oliveira, nesta área. Em *Estatística de Extremos*, essa investigação centrava-se essencialmente em modelos de índole paramétrica, e foi também nesse campo que eu comecei, aquando da minha estadia em Sheffield, onde obtive, em finais de 1978, o grau de Ph.D., sob a orientação de Clive Anderson. Para além de Tiago de Oliveira, que infelizmente faleceu em 1992, e de eu própria, juntou-se a nós o Feridun Turkman, com o grau de Ph.D. obtido em Sheffield em 1980, sob a orientação de Morris Walker, na área de *Extremos em Processos Estocásticos*. Embora nunca se tenham assumido como elementos nesta área, considero que o Dinis Pestana e a Antónia Amaral Turkman têm também tido um papel de relevo na construção deste grupo. Em meados de 1981, enveredámos pela organização do “*NATO Advanced Statistical Institute (ASI) on Statistical Extremes and Applications*”, que decorreu no Vimeiro, no Verão de 1983. Vieram a esse encontro nomes sonantes na área de extremos. No livro, editado por Tiago de Oliveira (Tiago de Oliveira, 1984a), em memória de Émile Gumbel, podem-se encontrar artigos desses investigadores, de que refiro só alguns ainda muito activos na área, Clive Anderson, Paul Deheuvels, Janos Galambos, Laurens de Haan, Ross Leadbetter, Georg Lindgren, James Pickands III, Sid Resnick, Holger Rootzén, Masaaki Sibuya, Jef Teugels, Ishay Weissman. Mas na própria lista de autores de “*contributed papers*” podemos encontrar nomes como Richard Davis, Anthony Davison, Jürg Hüsler e Rolf Reiss.... No Prefácio deste livro pode ler-se: “... *the narrow and shallow stream (of extremes) gained momentum and is now a huge river, enlarging at every moment and flooding the margins*”. E Tiago de Oliveira termina o Prefácio com agradecimentos aos elementos do recém formado Departamento de Estatística Investigação Operacional e Computação (DEIOC) da Faculdade de Ciências da Universidade de Lisboa (FCUL), dizendo: “... *it is a very good group that crossed the desert during the organization time and continues to work on ...*”. Atrevo-me hoje a dizer que a organização deste encontro de duas semanas, com dois fins de semana incluídos e repletos de *programa social*, embora me tivesse traumatizado de tal modo que só a partir de 1999 me atrevi a organizar outras conferências, constitui o marco de lançamento daquilo que penso poder hoje considerar-se a “*Escola de Extremos*” em Portugal. E após um largo interregno de cerca de 15 anos, esta última década tem sido profícua na organização de conferências em Portugal na área de extremos:

1. *Workshop on Statistical Modelling — Extreme Values and Additive Laws*, Estoril, 1999.
2. *Workshop on Extremes, Risk and Resampling Techniques*, Tomar, 2003.
3. *EVA 2004: Third International Symposium on Extreme Value Analysis: Theory and Practice*, Aveiro, 2004.
4. *Extremes Day in Honor of Laurens de Haan: Extremes, Risk, Safety and the Environment*, Lisboa, FCUL, 2006.
5. *Statistical Extremes and Environmental Risk (SEER 2007)*, Lisboa, FCUL, 2007.

Com algum trabalho nesta área, embora colateral aos temas em que se centram os seus doutoramentos, temos a Eugénia Graça Martins e a Helena Iglésias Pereira, com teses de doutoramento defendidas em 1983 e 1985, respectivamente, na Universidade de Lisboa. Nomes de grande relevo na área de *Extremos* são a Margarida Brito, com tese de doutoramento defendida na Universidade de Paris VI, em 1986, sob a orientação de Paul Deheuvels, a Teresa Alpuim e a Manuela Neves. Contrariamente ao que referi em Rosado (2005), a Teresa Alpuim foi a primeira aluna a doutorar-se em Portugal na área de *Extremos*, tendo defendido a sua tese em 1989, na Universidade de Lisboa, sob minha orientação. Em 1990, doutorou-se, sob orientação de Tiago de Oliveira, e na Universidade Nova de Lisboa, a Manuela Neves. Tendo decidido referir explicitamente só as teses defendidas nesta última década (1997-2006), considero também como nomes pioneiros para o desenvolvimento daquilo que me atrevo hoje a chamar a “*Escola de Extremos em Portugal*” não só a Luísa Canto e Castro, a Fernanda Oliveira e a M. Isabel Fraga Alves, com teses de doutoramento defendidas em 1992, mas também a Isabel Barão, com

tese defendida em 1993, a Emília Athayde, a Helena Ferreira e a Teresa Themido Pereira, com teses de doutoramento defendidas em 1994, todas elas na Universidade de Lisboa. O Nuno Crato, com tese de doutoramento defendida na Universidade de Delaware, USA, em 1992, sob a orientação de Howard Taylor, interessou-se também por modelos de caudas pesadas, publicando trabalhos muito citados (Gomes, Selman and Crato, 1997; Gomes, Selman, Crato and Kautz, 1998; Crato, 2000). Finalmente, refiro, ainda por entre os da “velha guarda”, o João Gomes, aluno de doutoramento da Teresa Alpuim, e meu primeiro “neto científico”. O João Gomes defendeu a sua tese de doutoramento em 1996, na Universidade de Lisboa, e atesta no seu curriculum trabalhos relevantes na área (Gomes, 1996; Gomes and Oliveira, 1997).

Voltando novamente atrás no tempo: após o meu regresso de Sheffield em fins de 1978, e exceptuando alguns detalhes relacionados com a orientação dos doutoramentos da Isabel Fraga Alves e da Teresa Themido Pereira, a minha investigação na área de *Estatística de Extremos* estava essencialmente centrada em modelos probabilísticos e modelos estatísticos de índole paramétrica. A inflexão feita na minha carreira, para a área de inferência semi-paramétrica, foi essencialmente devida a um convite de Laurens de Haan, para proferir um Seminário em Roterdão, em finais de 1998. Andava na altura interessada na metodologia Jackknife e sua possível aplicação em extremos, e pensei que um desafio interessante seria abordar a referida metodologia em âmbito semi-paramétrico, tão querido à “*Escola de Extremos*” da Holanda, liderada por Laurens de Haan. Assim fiz, e tenho-me mantido desde essa altura na área, em estreita colaboração com Laurens de Haan, grande responsável pelo franco desenvolvimento e internacionalização da nossa “*Escola de Extremos*”, nesta última década.

Irei em seguida debruçar-me sobre alguns dos campos da *Teoria de Valores Extremos*, em que a contribuição da “*Escola de Extremos em Portugal*” tem sido importante, com trabalho reconhecido e publicado internacionalmente. Começarei por uma vertente essencialmente probabilística, onde gostaria de mencionar:

- 1. Comportamento Probabilístico de Estatísticas Ordinais, “Records” e Extremos.** Tem aqui papel de relevo o trabalho da Margarida Brito, que tem mais recentemente trabalhado na área de estimação semi-paramétrica do índice de cauda, mas que começou a sua investigação em temas puramente probabilísticos (não tivesse sido ela aluna de Paul Deheuvels), como pode ser atestado pelos artigos Brito (1986), Bacro and Brito (1991, 1992) e Barme-Delcroix and Brito (2001). Nesta área enquadro ainda o artigo de Alpuim (1986). Encontra-se parcialmente nesta área a tese de doutoramento da Sandra Mendonça (Mendonça, 2000), sob a orientação de Dinis Pestana e defendida na Universidade da Madeira.
- 2. Comportamento Pré-Assintótico e Velocidade de Convergência em Teoria de Valores Extremos.** É aqui de realçar o artigo pioneiro de Gomes (1984a), o artigo de Gomes (1986) e o artigo de Gomes and Pestana (1987a). Uma revisão crítica sobre a investigação neste tema pode ser encontrada em Gomes (1994), o primeiro artigo que escrevi em LaTeX, relacionado com uma conferência convidada proferida em Gaithersburg, 1993, na “*Conference on Extreme Value Theory and its Applications*”, organizada por Janos Galambos, entre outros. Tem também trabalho nesta área a Helena Iglésias Pereira (Iglésias Pereira, 1983). Mais recentemente aparece uma contribuição conjunta com Laurens de Haan (Gomes and de Haan, 1999), fruto do relacionamento mais estreito, já atrás referido, com este valor expoente na área de extremos. As teses de doutoramento da Luísa Canto e Castro e da Adelaide Valente de Freitas (Freitas, 1998) encontram-se parcialmente inseridas neste tema, sendo ainda de mencionar os artigos de Canto e Castro (1987) e Freitas and Hüsler (2005).
- 3. Extremos em Sucessões Dependentes e Processos Estocásticos.** Nesta área a minha contribuição foi apenas marginal e atrevo-me a mencionar unicamente o artigo Gomes, de Haan and Pestana (2004), relegando para a estimação do índice extremal alguns outros artigos publicados internacionalmente. Mas alegra-me muito ter tido a sorte de ter tido alunas de doutoramento como a Teresa Alpuim, a Helena Ferreira e a Andreia Hall (Hall, 1998), a quem consegui incutir o interesse por esta área que pessoalmente sempre achei fascinante, e que desenvolveram trabalho de

grande valor. Não posso deixar de mencionar os artigos pioneiros de Alpuim (1988, 1989, 1990), bem como já alguns anos depois Alpuim, Catkan and Hüsler (1995). E só tenho pena que a Teresa tenha abandonado a área de extremos. Mas pode ser que ainda volte.... Felizmente, a Helena Ferreira e a Andreia Hall continuam bem integradas na nossa “*Escola de Extremos*”.

A Helena Ferreira averba já no seu curriculum cerca de uma dezena de artigos individuais em revistas de elevado prestígio (Ferreira, 1993, 1994, 1995, 1996, 1998, 1999, 2000, 2003, 2006). Foi orientadora de duas alunas nesta área, também minhas “netas científicas” e mais dois valiosos elementos para a nossa “*Escola de Extremos*”, a Luísa Pereira (Pereira, 2002) e a Ana Paula Martins (Martins, 2005). Em co-autoria com elas e também com a Andreia Hall e o Manuel Scotto, encontrei através do *MathSciNet*, Pereira and Ferreira (2001, 2002, 2005), Ferreira and Scotto (2002), Scotto and Ferreira (2003), Ferreira and Martins (2003), Martins and Ferreira (2004a, 2004b, 2005a, 2005b), Hall, Scotto and Ferreira (2004).

A Andreia Hall tem também tido um contributo excelente para a manutenção do bom nome dos extremos. Refiro aqui apenas os artigos que encontrei na já referida pesquisa rápida feita através do *MathSciNet*, até porque a Andreia é uma das pessoas que contribuí com um artigo para este Boletim. Para além do artigo atrás mencionado, em co-autoria com a Helena Ferreira e o Manuel Scotto, encontrei três artigos individuais (Hall, 1996, 2001, 2003) e meia dúzia de artigos em co-autoria, Brännäs and Hall (2001), Hall and Scotto (2003, 2006), Hüsler, Cruz, Hall and Fonseca (2003), Hall and Moreira (2006), Hall and Hüsler (2006).

A Graça Temido, com contribuição fundamental na área de Max-semiestabilidade (veja-se o item 4, apresentado a seguir, e o artigo da Graça neste Boletim), tem também contribuído para o desenvolvimento desta tema, como se pode atestar pelos artigos, Temido (1999, 2000, 2003), que constituem também uma parte im portante da sua tese de doutoramento.

É também essencialmente nesta área que o trabalho do Feridun Turkman tem sido notável. Com dois alunos com doutoramento concluído nesta área, a Fernanda Oliveira e o Manuel Scotto (Scotto, 2002), o Feridun Turkman averba no seu curriculum cerca de uma vintena de artigos individuais ou em co-autoria na área de extremos de processos estocásticos, de que menciono, Turkman (1983, 1984, 2001, 2006), Turkman and Walker (1983, 1984, 1990), Turkman and Amaral Turkman (1997), Anderson and Turkman (1991, 1992, 1995), Turkman and Oliveira (1992), Oliveira and Turkman (1992), Scotto and Turkman (2002, 2005), Scotto, Turkman and Anderson (2003).

De entre os alunos do Feridun Turkman, o Manuel Scotto, cujo doutoramento foi co-orientado por Clive Anderson, tem já obra individual relevante, pós-doutoramento. Atrevo-me aqui a mencionar, mais uma vez pecando por defeito, Scotto (2003, 2005, 2006) e Pereira and Scotto (2006), este último em co-autoria com a Luísa Pereira, que também já denota alguma independência científica — veja-se o artigo individual Pereira (2004).

- 4. Extensões da Classe de Modelos Max-Estáveis.** A semelhança entre a teoria de valores extremos e o esquema de somas, interesses que o Dinis e eu partilhamos, levou-nos desde muito cedo à procura de generalizações da classe das leis max-estáveis. São aqui de mencionar os artigos Gomes and Pestana (1981), Pestana (1981), Graça Martins and Pestana (1987, 1988). Mais recentemente, a classe das f. s. d. max-semiestáveis (classe MSS) introduzida por Pancheva e Grinevich, de forma independente, tem sido muito acarinhada por alguns investigadores portugueses, como se pode ver no artigo deste Boletim elaborado pela Graça Temido. Na realidade, o envolvimento na caracterização da classe MSS foi devido aos trabalhos desenvolvidos pela Graça, para obtenção do grau de Doutor, essencialmente sob a orientação da Luísa Canto e Castro, embora o meu nome figure também como co-orientadora da tese defendida pela Graça Temido em 2000, na Universidade de Coimbra (Temido, 2000). É pois algo de muito especial a relação científica que tenho com a Graça, que é simultaneamente minha “filha e neta científica”. Também neste campo foi mais uma vez extraordinariamente relevante a contribuição de Laurens de Haan, que deu origem ao artigo Canto e Castro, de Haan and Temido (2002). Foi também feita, para sucessões estritamente estacionárias sujeitas a condições adequadas de independência assintótica, a generalização de vários resultados obtidos em contexto de max-semiestabilidade em situação iid (Temido e Canto e Castro, 2003). Em ambiente de dependência, e como seria quase inevitável, houve cooperação da Helena Ferreira (Ferreira and Temido, 2005) e da Andreia Hall (Hall and

Temido, 2005).

Passando para a vertente estatística (sempre com algum cunho probabilístico à mistura):

5. Estatística de Extremos Univariados com Base em Modelos Multivariados, Multi-dimensionais e Outros Modelos Não-Clássicos. Esta é a fase em que a estatística de extremos desenvolvida em Portugal era de índole paramétrica, com algumas incursões em modelos não-paramétricos, talvez demasiado vastos para se terem revelado interessantes. Refiro os artigos de Gomes (1981, 1984b, 1985a, 1985b) e ainda os artigos de Gomes and Alpuim (1986), Gomes and Pestana (1986, 1987b) e Athayde and Gomes (1987). Procuravam-se aqui modelos paramétricos que generalizassem o método dos máximos anuais de Gumbel (Gumbel, 1958). Neste tema ainda trabalhou a Isabel Fraga Alves, que embora tenha começado a sua investigação para doutoramento numa área de estatística paramétrica, tinha como programa de doutoramento temas de “*Inferência Paramétrica e Semi-Paramétrica em Modelos Extremais*”. A Isabel Fraga Alves, que trabalhou para doutoramento sob minha orientação, inflectiu definitivamente para a inferência semi-paramétrica com o contributo de Laurens de Haan, e após um post-doc em Roterdão, onde trabalhou de perto com a “*Escola de Extremos da Holanda*”. Apesar disso tem ainda um artigo publicado internacionalmente, em modelo puramente paramétrico (Fraga Alves, 1992).

A *Estatística de Extremos para Modelos Max-Semiestáveis* tem-se desenvolvido de forma não muito rápida, o que é natural devido à complexidade do tema, e infelizmente ainda não existe de momento a possibilidade de referir qualquer artigo, publicado em revista internacional. A Luísa Canto e Castro conseguiu no entanto uma aluna de doutoramento, a Sandra Dias, que tem trabalhado arduamente na estimação dos parâmetros da classe das leis max-semiestáveis. Os resultados recentemente apresentados parecem-me promissores (veja-se mais uma vez o artigo da Graça neste Boletim) e é de esperar que surjam em breve artigos publicados em revista internacional.

6. Escolha Estatística de Modelos Extremais (o velho Trilema de Tiago de Oliveira). Este tema começou mais uma vez o seu desenvolvimento em contexto paramétrico e Tiago de Oliveira teve papel fundamental com os trabalhos apresentados no “*NATO ASI on Statistical Distributions in Scientific Work*” (Tiago de Oliveira, 1981) e no “*NATO ASI on Statistical Extremes and Applications*” (Tiago de Oliveira, 1984b). Há ainda que mencionar os artigos de Gomes (1982, 1984c, 1987, 1989a, 1989b), bem como Fransén and Tiago de Oliveira (1984), Tiago de Oliveira and Gomes (1984), Tiago de Oliveira (1985, 1987a), van Montfort and Gomes (1985) e Gomes and van Montfort (1987). Podemos ainda referir um artigo mais recente, Brilhante (2003). O artigo de Fraga Alves and Gomes (1996), embora ainda com uma vertente paramétrica forte, é mais uma vez um artigo em que é possível notar a existência de uma inflexão para modelos semi-paramétricos. Também aqui teve uma grande responsabilidade Laurens de Haan, que me convidou para a “*Conference on Multivariate Extreme Value Estimation with Application to Economics and Finance*”, que decorreu em Roterdão, em 1994. Os contributos mais recentes para este tema têm sido desenvolvidos em ambiente semi-paramétrico e foram um dos temas em que trabalhou para doutoramento a Cláudia Neves, sob a orientação de Isabel Fraga Alves e Laurens de Haan, doutoramento que defendeu em Dezembro de 2006 (Neves, 2006). Para mais detalhes veja-se o artigo da autoria da Isabel Fraga Alves, neste trabalho conjunto sobre a nossa “*Escola de Extremos*”. Não posso no entanto deixar de mencionar os artigos Fraga Alves (1999), Neves and Fraga Alves (2006) e Neves, Picek and Fraga Alves (2006). A Cláudia é pois, de momento, a minha “neta científica” mais jovem.

7. Estatística de Extremos Multivariados. Este é um campo em que a obra de Tiago de Oliveira é pioneira, com vários artigos publicados antes de 1980. Iremos aqui mencionar alguns dos artigos de Tiago de Oliveira neste tema e posteriores a 1980: Tiago de Oliveira (1984c, 1987b, 1989a, 1989b, 1991, 1992), Deheuvels and Tiago de Oliveira (1989). Também com papel de relevo nesta área encontramos mais recentemente a Isabel Barão, aluna de doutoramento de Tiago de Oliveira, que concluiu a sua tese de doutoramento em 1993, já sob minha orientação, após a morte de Tiago de Oliveira. A Isabel Barão tem um artigo de relevo, em conjunto com Jonathan Tawn (Barão and

Tawn, 1999), e um outro mais recente (Barão, Tawn and Li, 2003). Um outro elemento desta “*Escola de Extremos*” com trabalho de relevo neste tema é a Ana Ferreira, que foi orientada para doutoramento por Laurens de Haan e John Einmhal (Ferreira, 2002b). Menciono aqui o artigo Draisma, Drees, Ferreira and de Haan (2004). Também a Alexandra Ramos, com doutoramento em Lancaster, sob a orientação de Anthony Ledford e Trevor Sweeting (Ramos, 2002), tem conseguido publicação internacional neste tema. Veja-se Ramos and Ledford (2005). A tese de doutoramento da Fátima Miguéns (Miguéns, 2005), que iniciou investigação para doutoramento sob a orientação de Tiago de Oliveira, com co-orientação final de M. Fátima Fontes de Sousa, é também nesta área de eleição de Tiago de Oliveira.

- 8. Estimção do Índice de Valores Extremos e de Outros Parâmetros de Acontecimentos Raros em Contexto Semi-Paramétrico.** Neste campo, em que não estou a incluir a estimção de viés reduzido, tem sido vasta a obra da nossa *Escola*. Mais uma vez com risco de pecar por defeito, e relativamente à estimção do *índice de cauda* ou *índice de valores extremos*, menciono Bacro e Brito (1993, 1995, 1998), Fraga Alves (1995, 2001a, 2001b), Gomes and Oliveira (2003a, 2003b, 2003c), Brito and Moreira de Freitas (2003, 2006), Martins, Gomes and Neves (2004), Drees, Ferreira and de Haan (2004). Tem também doutoramento nesta área, defendida na Universidade de Michigan, USA, em 2002, sob a orientação de Bruce Hill e George Michailidis, o Bruno Cecílio Sousa (Sousa, 2002). Na sequência deste trabalho há que referir o artigo Sousa and Michailidis (2004), onde é apresentado um método gráfico interessante de seleção do número de estatísticas ordinais de topo a incluir na estimção do índice de cauda através de um estimador como o estimador de Hill. A escolha da fração óptima de estatísticas ordinais em procedimentos clássicos em estatística de extremos foi também investigada em Neves and Fraga Alves (2004). Menciono por fim o artigo, Oliveira, Gomes and Fraga Alves (2006), publicado a título póstumo face à morte inesperada e prematura de um outro meu aluno de doutoramento, o Orlando Oliveira, pouco tempo após a defesa da sua tese de doutoramento (Oliveira, 2003).

Relativamente à estimção de *quantis elevados*, *períodos de retorno* e *probabilidades de excedência* de níveis elevados houve também contribuição de investigadores portugueses. Veja-se como exemplo, os artigos Ferreira (2002a), Ferreira, de Haan and Peng (2003).

A estimção do *índice extremal*, um parâmetro quase tão relevante como o *índice de cauda*, ou mais geralmente inferência em processos dependentes, tem também tido alguma contribuição da nossa *Escola*. Refiro os artigos, Canto e Castro (1994), Gomes (1990, 1993, 1995b) e Gomes, Hall and Miranda (2005), onde é introduzido um estimador de viés reduzido, com base na utilização da metodologia Jackknife, já de há muito utilizada com êxito na estimção do *índice de cauda*, como iremos ver de seguida.

- 9. As Metodologias Jackknife e Bootstrap em Estatística de Extremos.** Atrevo-me aqui a referir um artigo pioneiro em português (Gomes, 1995a), publicado nas *Actas do II Encontro da Sociedade Portuguesa de Estatística*. Penso ter sido a apresentação que fiz nesse encontro a responsável pela co-orientação, em conjunto com a Manuela Neves, de uma nova aluna de doutoramento, para mim na altura uma “ilustre desconhecida”, a Maria João Martins, que trabalhou de forma incansável neste tema e temas afins, relacionados com a estimção semi-paramétrica de viés reduzido, tendo concluído a sua tese em 2001 (Martins, 2001). Penso serem dignos de registar os dois artigos individuais que publiquei nesta área (Gomes, 1999a, 1999b), mas os artigos de maior interesse são em co-autoria: Martins, Gomes and Neves (1999), Gomes, Martins e Neves (2000, 2003). Mais recentemente, e na sequência de teses de Mestrado na área, a Cristina Miranda, minha aluna de doutoramento (Miranda, 2005), trabalhou também com a metodologia Jackknife. Refiro aqui as publicações Gomes, Miranda and Pereira (2005) e Gomes, Miranda and Viseu (2006). A metodologia Bootstrap também tem sido usada com sucesso em *Estatística de Extremos*, e também aqui temos a colaboração de membros da nossa *Escola*. Para além dos artigos anteriormente mencionados de Gomes em 1995 e 1999, veja-se Draisma, de Haan, Peng and Themido Pereira (1999) e ainda Gomes and Oliveira (2001).

- 10. Métodos Alternativos de Estimção Semi-paramétrica de Viés Reduzido.** Este tem sido um dos campos da *Estatística de Extremos* em que mais tenho investido nestes últimos anos, com a

colaboração de vários colegas e alunos de doutoramento, de entre os quais já referi a Maria João Martins e a Cristina Miranda (uma vez que a metodologia *Jackknife* é um dos métodos de obtenção de estimadores de viés reduzido). Não posso agora deixar de referir o Frederico Caeiro, com doutoramento concluído em Novembro de 2006 (Caeiro, 2006), a Fernanda Figueiredo, com doutoramento concluído em 2003 (Figueiredo, 2003), em *Controlo Estatístico da Qualidade*, mas que se “converteu” parcialmente à área de *Extremos*, e ainda a Lígia Rodrigues e a Clara Viseu, que começaram os trabalhos de doutoramento há pouco tempo, mas nesta área. Começo por mencionar os artigos Gomes and Martins (2001, 2002, 2004), os artigos Caeiro and Gomes (2002a, 2002b), Gomes, Figueiredo and Mendonça (2004), Gomes, Caeiro and Figueiredo (2004) e Gomes and Figueiredo (2006), onde a redução de viés é feita à custa de um aumento significativo da variância, tal como já acontecia face à utilização da metodologia *Jackknife*. Mais recentemente, a estimação dos parâmetros de segunda ordem de forma adequada permitiram-nos a manutenção da variância e a eliminação da componente dominante de viés assintótico. Detalhes sobre estes novos métodos de estimação podem ser vistos em Gomes and Pestana (2004, 2005), Caeiro, Gomes and Pestana (2005) e Gomes, Rodrigues, Vandewalle and Viseu (2006), por entre outros trabalhos ainda em via de publicação e que me abstenho de referir. Um outro tipo de estimação de viés reduzido pode ser vista em Fraga Alves (2001b). O problema de redução de viés é também abordado em de Haan and Canto e Castro (2006).

11. **Estimação Semi-paramétrica de Parâmetros de Segunda Ordem (e de Ordem Superior...).** O papel da nossa *Escola* nesta área tem também sido de relevo, particularmente devido à ênfase que tem sido dada no nosso grupo de trabalho ao desenvolvimento de estimadores de segunda-ordem de viés-reduzido. A estimação do parâmetro de “forma” de segunda ordem foi abordada em Gomes (2000), Gomes, de Haan and Peng (2002), Fraga Alves, Gomes and de Haan (2003), Fraga Alves, de Haan and Lin (2003). O parâmetro de “escala” de segunda ordem foi estimado, de forma não individualizada, em Gomes and Martins (2002) e mais recentemente em Caeiro and Gomes (2005). A necessidade de se desenvolverem técnicas de escolha adaptativa do nível óptimo para estimadores do *índice de cauda* de viés reduzido, leva-me a crer que temos ainda de avançar neste tema e afins. Um bom contributo foi já dado em Fraga Alves, de Haan and Lin (2006).
12. **Métodos Bayesianos em Estatística de Extremos.** Trata-se de uma área de desenvolvimento muito recente, onde mais uma vez tem havido contribuição relevante de elementos da nossa *Escola*. É essencialmente nesta área que se centra a tese de doutoramento da Patrícia de Zea Bermudez, desenvolvida sob a orientação de Feridun Turkman e defendida em 2003 (de Zea Bermudez, 2003). Na sequência dessa orientação, em que a Antónia Amaral Turkman teve também papel relevante, foram conseguidas as publicações internacionais de Zea Bermudez, Amaral Turkman and Turkman (2001) e de Zea Bermudez and Amaral Turkman. (2003).
13. **E outros temas diversos ...** A nossa *Escola* tem ainda grupos fortes nas vertentes de *Extremos e Modelação de Risco* (veja-se por exemplo a tese de doutoramento recente da Ana Cristina Moreira de Freitas (Moreira de Freitas, 2005), sob a orientação da Margarida Brito), *Extremos e Ambiente* (veja-se, por exemplo, Guedes Soares and Scotto, 2001, 2004, bem como Scotto and Guedes Soares, 2000, 2006 e Tobias and Scotto, 2005), *Extremos e Controlo Estatístico da Qualidade* (veja-se o artigo, Figueiredo and Gomes, 2004), *Extremos em Processos de Dimensão Infinita*, *Extremos Espaciais* (veja-se, por exemplo, o artigo de Laurens de Haan and Themido Pereira, 2006), *A Metodologia PORT* (que podia significar portuguesa, mas é acrónimo de peaks over random thresholds) em *Estatística de Extremos* (veja-se, por exemplo, o artigo, Araújo Santos, Fraga Alves and Gomes, 2006) e, a título de previsão, espero que surja em breve um grupo de *Extremos em Genética*, onde até eu gostaria de me enquadrar....

Face aos resultados apurados sou levada a crer que o dinamismo do grupo proporcionará um desenvolvimento salutar da área, com reconhecimento internacional da nossa *Escola de Extremos*, que tem crescido de forma equilibrada. Consegui contabilizar 34 teses de doutoramento na área ou em áreas muito afins, 20 das quais na última década, onde estou a incluir também a tese de doutoramento da Fátima Brilhante (Brilhante, 1999), sob orientação do Dinis Pestana e defendida na Universidade

dos Açores. A dinâmica de publicação tem sido bastante elevada, nitidamente acima dos padrões médios em Portugal, mas espero que ainda melhore nos próximos anos, uma vez que a distribuição pelos diferentes elementos da *Escola* é demasiado enviesada. Conteí 109 artigos publicados em 40 revistas internacionais de prestígio, cuja distribuição apresento em seguida:

Journal	Nº/artigos	Journal	Nº/artigos
<i>Extremes</i>	15	<i>Ann. Statist.</i>	1
<i>Satist. & Probab. Lett.</i>	10	<i>Appl. Statist.</i>	1
<i>J. Statist. Plann. and Inference</i>	9	<i>Bernoulli</i>	1
<i>Test</i>	9	<i>Comm. Statistics — Th. & Meth.</i>	1
<i>RevStat</i>	8	<i>Env. Monit. and Assess.</i>	1
<i>Portugal. Mathem.</i>	7	<i>Intern. J. Statist. & Systems</i>	1
<i>J. Statist. Comp. and Sim.</i>	4	<i>J. Amer. Statist. Assoc.</i>	1
<i>Appl. Stoch. Models in Bus. Ind.</i>	3	<i>J. Automated Research</i>	1
<i>J. Coastal Engineering</i>	3	<i>J. Comput. Graph. Statist.</i>	1
<i>Th. Probab. and Appl.</i>	3	<i>J. Multiv. Analysis</i>	1
<i>Bull. Intern. Statist. Inst.</i>	2	<i>Math. Methods in Statist.</i>	1
<i>Comm. Statistics — Sim. & Comp.</i>	2	<i>Math. Comp. Appl. Probab.</i>	1
<i>Insur.: Math. and Econom.</i>	2	<i>Ocean Engineering</i>	1
<i>J. Appl. Probab.</i>	2	<i>Publ. Inst. Math.</i>	1
<i>J. Comp. Statist. & Data Anal.</i>	2	<i>SORT</i>	1
<i>J. Time Series Analysis</i>	2	<i>Statistica Neerlandica</i>	1
<i>Nonlinear Analysis</i>	2	<i>Stoch. Models</i>	1
<i>Statistics</i>	2	<i>Stoch. Proc. And Appl.</i>	1
<i>Ann. Appl. Probab.</i>	1	<i>Th. Stoch. Processes</i>	1
<i>Ann. de l'ISUP</i>	1	<i>WSEAS Trans. Math.</i>	1

Convém ainda registar que a nossa *Escola de Extremos*, apesar da elevada contribuição a nível internacional, não tem descurado a publicação a nível nacional. Tal como referi no artigo que escrevi para o *Memorial da Sociedade Portuguesa de Estatística* (Rosado, 2005), esse contributo pode ser atestado pelas publicações de artigos em português, nas diferentes colectâneas de texto editadas pela Sociedade Portuguesa de Estatística desde 1992, onde a produção na área de Extremos tem sido, em média, de 17% (7 a 11 trabalhos/ano). Surpreendente foi a participação de elementos deste grupo no *XIV Congresso Anual da Sociedade Portuguesa de Estatística*, onde 13 das 72 comunicações orais foram proferidas por elementos da nossa *Escola*. Espero que a publicação em Actas seja condicente com esta participação elevada e permita aumentar a percentagem atrás referida. Para terminar, não posso deixar de referir o livro recente da autoria de Laurens de Haan e Ana Ferreira (Haan and Ferreira, 2006), que vem mais uma vez atestar a valiosíssima contribuição de Laurens de Haan para o estabelecimento da *Escola de Extremos em Portugal*.

Referências:

- Alpuim, M.T. (1986). Record values in populations with increasing or random dimension. *Metron* **43**:3-4, 145-155.
- Alpuim, M.T. (1988). High level exceedances in stationary sequences with extremal index. *Stochastic Processes and Applications* **30**:1, 1-16.
- Alpuim, M.T. (1989). An extremal markovian sequence. *J. Appl. Probab.* **26**:2, 219-232.
- Alpuim, M.T. and Athayde, E. (1990). On the stationary distribution of some extremal Markovian sequences. *J. Appl. Probab.* **27**:2, 291-302.
- Alpuim, M.T., Catkan, N.A. and Hüsler, J. (1995). Extremes and clustering of non-stationary max-AR(1) sequences. *Stochastic Process. Appl.* **56**:1, 171-184.
- Anderson, C.W. and Turkman, K.F. (1991). The joint limiting distribution of sums and maxima of stationary sequences. *J. Appl. Probab.* **28**:1, 33-44.
- Anderson, C.W. and Turkman, K.F. (1992). Limiting joint distributions of sums and maxima in a statistical context. *Theory Probab. Appl.* **37**: 2, 314-316.
- Anderson, C.W. and Turkman, K.F. (1995). Sums and maxima of stationary sequences with heavy tailed distributions. *Sankhyä A* **57**:1, 1-10.

- Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and high quantile estimation. *Revstat* **4**:3, 227-247.
- Athayde, E. and Gomes, M.I. (1987). Multivariate extremal models under non-classical situations. In P. Bauer et al. (eds.), *Mathematical Statistics and Probability Theory*, Vol. B, 1-9, D. Reidel, Dordrecht.
- Bacro, J.-N. and Brito, M. (1991). On Mason's extension of the Erdős-Rényi law of large numbers. *Statist. Probab. Letters* **11**:1, 43-47.
- Bacro, J.-N. and Brito, M. (1992). Comportement asymptotique presque sûr de la différence de statistiques d'ordre uniformes. *C.R. Acad. Sci. Paris, Série I – Math.* **314**:4, 317-320.
- Bacro, J.-N. and Brito, M. (1993). Strong limiting behaviour of a simple tail Pareto-index estimator. *Statistics and Decisions* **3**, 133-143.
- Bacro, J.-N. and Brito, M. (1995). Weak limiting behaviour of a simple tail Pareto-index estimator. *J. Statistical Planning and Inference* **45**:1-2, 7-19.
- Bacro, J.-N. and Brito, M. (1998). A tail bootstrap procedure for estimating the tail Pareto-index. *J. Statistical Planning and Inference* **71**:1-2, 245-260.
- Barão, M.I., de Haan, L. and Li, D. (2003). Comparison of estimators in multivariate EVT. *International J. Statistics and Systems* (aceite).
- Barão, M. I. and Tawn, J. A. (1999). Extremal analysis of short series with outliers: sea-levels and athletic records. *Applied Statistics* **48**, 469-488.
- Barme-Delcroix, M.-F. and Brito, M. (2001). Multivariate stability and strong limiting behaviour of intermediate order statistics. *J. Multivariate Analysis* **79**:2, 157-170.
- Brännäs, K. and Hall, A. (2001). Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry* **17**: 3, 277-291.
- Brilhante, M.F. (1999). *Inferência Estatística em Modelos não Gaussianos com Recurso a Spacings e Outras Funções de Estatística Ordinárias*. Tese de doutoramento, Universidade dos Açores.
- Brilhante, M.F. (2003). Exponentiality versus generalized Pareto — a resistant and robust test. *RevStat* **2**:1, 1-13.
- Brito, M. (1986). Sur l'encadrement optimal presque sûr dans un échantillon ordonné. *C.R. Acad. Sci. Paris, Série I – Math.* **303**:16, 821-824.
- Brito, M. and Moreira de Freitas, A.C. (2003). Limiting behaviour of a geometric-type estimator for tail indices. *Insurance: Mathematics and Economics* **33**:2, 211-216.
- Brito, M. and Moreira de Freitas, A.C. (2006). Weak convergence of a bootstrap geometric-type estimator with applications to risk theory. *Insur. Math. and Economics* **38**:3, 571-584.
- Caiiro, F. (2006). *Estimação de Parâmetros de Acontecimentos Raros*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.
- Caiiro, F. and Gomes, M.I. (2002a). A class of asymptotically unbiased semi-parametric estimators of the tail index. *Test* **11**:2, 345-364.
- Caiiro, F. and Gomes, M.I. (2002b). Bias reduction in the estimation of parameters of rare events. *Theory Stoch. Process.* **8**:1-2, 67-75.
- Caiiro, F. and Gomes, M.I. (2005). A new class of estimators of a "scale" second order parameter. *Extremes* (aceite).
- Caiiro, F., Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *RevStat* **3**:2, 113-136.
- Canto e Castro, L. (1987). Uniform rates of convergence in extreme-value theory — normal and gamma models. *Ann. Sc. Univ. Clermont Ferrand II, Probab. Appl.* **6**, 25-41.
- Canto e Castro, L. (1994). Estimating the extremal index, under a local dependence condition. In *Proc. of the Conference on Extreme Value Theory and its Applications*, Gaithersburgh, USA.
- Canto e Castro, L., de Haan, L. and Temido, M.G. (2002). Rarely observed sample maxima. *Theory Probab. Appl.* **45**:4, 658-661.
- Crato, N. (2000). Estimation of the maximal moment exponent with censored data. *Commun. In Statist. — Simul. and Comput.* **29**, 1239-1254.
- Deheuvels, P. and Tiago de Oliveira, J. (1989). On the nonparametric estimation of the bivariate extreme-value distributions. *Statist. Probab. Lett.* **8**:4, 315-323.
- Draisma, G., L. de Haan, L., Peng and Themido Pereira, T. (1999). A Bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2**:4, 367-404.

- Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**:2, 251-280.
- Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**:3, 1179-1201.
- Ferreira, A. (2002a). Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Planning and Inference* **104**, 83-102.
- Ferreira, A. (2002b). *Statistics of Extremes: Estimation and Optimality*. Ph.D. Thesis, University of Tilburg, Holland.
- Ferreira, A., de Haan, L. and Peng, L. (2003). On optimizing the estimation of high quantiles of a probability distribution. *Statistics* **37**, 401-434.
- Ferreira, H. (1993). Joint exceedances of high levels under a local dependence condition. *J. Appl. Probab.* **30**:1, 112-120.
- Ferreira, H. (1994). Multivariate extreme values in T-periodic random sequences under mild oscillation restrictions. *Stochastic Process. Appl.* **49**:1, 111-125.
- Ferreira, H. (1995). Extremes of a random number of variables from periodic sequences. *J. Statist. Planning and Inference* **45**:1-2, 133-141.
- Ferreira, H. (1996). Extremal behaviour of stationary sequences under local restrictions on upcrossings. *Publ. Inst. Statist. Univ. Paris* **40**:1, 57-75.
- Ferreira, H. (1998). Doubly stochastic compound Poisson processes in extreme value theory. *Portugal. Math.* **55**:4, 465-474.
- Ferreira, H. (1999). Limit distributions for point processes of exceedances of random levels. *Test* **8**:1, 191-200.
- Ferreira, H. (2000). A note on extremes of concomitants of order statistics. *Extremes* **3**:4, 385-392 (2001).
- Ferreira, H. (2003). Extremes of associated variables. *Statistics and Probability Letters* **63**, 333-338.
- Ferreira, H. (2006). A new dependence condition for time series and the extremal index of higher-order Markov chains. *RevStat* **4**:2, 143-151.
- Ferreira, H. and Martins, A. (2003). The extremal index of sub-sampled periodic sequences with strong local dependence. *RevStat* **1**, 15-24.
- Ferreira, H. and Scotto, M.G. (2002). On the asymptotic location of high values of a stationary sequence. *Statistics and Probability Letters* **60**:4, 475-482.
- Ferreira, H. and Temido, M.G. (2005). Maximum of a random number of variables attracted to a max-semistable distribution. *Ann. de l'ISUP XLIX*, I, 71-80.
- Figueiredo, F. (2003). *Controlo Estatístico da Qualidade — Métodos Robustos*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.
- Figueiredo, F. and Gomes, M.I. (2004). The total median in Statistical Quality Control. *Applied Stochastic Models in Business and Industry* **20**:4, 339-353.
- Fraga Alves, M.I. (1992). The influence of central observations on discrimination among multivariate extremal models. *Theory Probab. Appl.* **37**:2, 334-337.
- Fraga Alves, M.I. (1995). Estimation of the tail parameter in the domain of attraction of an extremal distribution. *J. Statist. Plann. Inference* **45**:1-2, 143-173.
- Fraga Alves, M.I. (1999). Asymptotic distribution of Gumbel statistic in a semi-parametric approach. *Portugaliae Mathematica* **56**:3, 282-298.
- Fraga Alves, M.I. (2001a). Weiss-Hill Estimator. *Test* **10**:1, 203-224.
- Fraga Alves, M.I. (2001b). A location invariant Hill-Type estimator. *Extremes* **4**:3, 199-217 (2002).
- Fraga Alves, M.I. and Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction — a comparative analysis. *Communications in Statistics — Theory and Methods* **25**, 789-811.
- Fraga Alves, M.I., Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:2, 193-213.
- Fraga Alves, M.I., de Haan, L. and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in Extreme value Theory. *Mathematical Methods of Statistics* **12**:2, 155-176.
- Fraga Alves, M.I., de Haan, L. and Lin, T. (2006). Third order extended regular variation. *Publications de l'Institut Mathématique* **80** (94), 109-120.
- Fransén, A. and Tiago de Oliveira, J. (1984). Statistical choice of univariate extreme models, II. In Tiago de Oliveira ed., *Statistical Extremes and Applications*, 373-394, D. Reidel, Dordrecht.

- Freitas, A.V. (1998). *Nova Classe de Aproximações em Teoria de Valores Extremos*, Tese de doutoramento, Departamento de Matemática, Universidade de Aveiro.
- Freitas, A.V. and Hüsler, J. (2005). Condition for the convergence of maxima of random triangular arrays. *Extremes* **6**:4, 381-394.
- Gomes, C.P., Selman, B., and Crato, N. (1997). Heavy-tailed probability distributions in combinatorial search. In Gert Smolka (ed.), *Principles and Practice of Constraint Programming — CP 97*, Lecture Notes in Computer Science 1330, Springer, 121-135.
- Gomes, C.P., Selman, B., Crato, N. and Kautz, H. (2000). Heavy-tailed phenomena in satisfiability and constraint satisfaction problems. *J. Automated Research* **24**, 67-100.
- Gomes, J. (1996). Extreme value theory for a thermal energy storage model. *Statist. Probab. Lett.* **30**:1, 25-31.
- Gomes, J. and Oliveira, O. (1997). Limit laws for a sequence between the maximum and the sum of independent exponentials. *Statist. Probab. Lett.* **35**:1, 25-32.
- Gomes, M.I. (1981). An i -dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, 389-410, D. Reidel.
- Gomes, M.I. (1982). A note on statistical choice of extremal models. In *Proc. IX Jornadas Mat. Hispano-Lusas, Salamanca*. 653-655.
- Gomes, M.I. (1984a). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* **36A**, 71-85.
- Gomes, M.I. (1984b). Concomitants in a multidimensional extreme model. In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, 353-364, D. Reidel.
- Gomes, M.I. (1984c). Robustness of Gumbel statistic for distribution functions in the domain of attraction of a type I distribution of largest values. In *Computational Statistics* **84**, 61-66, Physica-Verlag.
- Gomes, M.I. (1985a). Statistical theory of extremes — comparison of two approaches. *Statistics and Decision* **2**, 33-37.
- Gomes, M.I. (1985b). Concomitants and linear estimators in an i -dimensional extremal model. *Trab. Estad. Inv. Oper.* **36**, 129-140, 1985.
- Gomes, M.I. (1986). Comparison of ultimate and penultimate approximations through simulation techniques. *J. Computational Statistics and Data Analysis* **4**, 257-267.
- Gomes, M.I. (1987). Extreme value theory — statistical choice. In P. Revesz et al. (eds.), *Goodness-of-Fit*, 195-209, North-Holland, Amsterdam.
- Gomes, M.I. (1989a). Comparison of extremal models through statistical choice in multidimensional backgrounds. In Hüsler, J. and R.-D. Reiss (eds.), *Extreme Value Theory*, 191-203, Lecture Notes in Statistics 51, Springer-Verlag.
- Gomes, M.I. (1989b). Generalized Gumbel and likelihood ratio test statistics in the multivariate GEV model. *J. Computational Statistics and Data Analysis* **7**, 259-267.
- Gomes, M.I. (1990). Inference in an extremal markovian model. In Momirovic, K. and V. Mildner (eds.), *Computational Statistics* **90**, 257-262, Physica-Verlag.
- Gomes, M.I. (1993). On the estimation of parameters of rare events in environmental time series. In V. Barnett and K.F. Turkman (eds.), *Statistics for the Environment*, 225-241, Wiley.
- Gomes, M.I. (1994). Penultimate behaviour of the extremes. In J. Galambos et al. (eds.), *Extreme Value Theory and Applications*, 403-418, Kluwer Academic Publishers.
- Gomes M.I. (1995a). Metodologias Jackknife e Bootstrap em Estatística de Extremos. In Mendes-Lopes, N. et al. (eds.), *Actas da II Conferência Anual da Sociedade Portuguesa de Estatística*, 31-46, Edições SPE.
- Gomes, M.I. (1995b). The influence of the extremal index on the estimation of return periods of high levels. In *Statistical Climatology* **95**, 299-302.
- Gomes, M.I. (1999a). Generalized Jackknife Moment estimator of the tail index. *Bull. of the International Statistical Institute* **58**:1, 401-402.
- Gomes, M.I. (1999b). The Jackknife and the Bootstrap methodologies in the estimation of parameters of rare events. *Revista de Estatística* **99**, 5-23.
- Gomes, M.I. (2000). The second order framework and the modeling of rare events. In V. Núñez-Antón and E. Ferreira ed. *Statistical Modelling*, 210-215.
- Gomes, M.I. and Alpuim, M.T. (1986). Inference in multivariate generalized extreme value models. *Scandinavian J. Statistics* **13**, 291-300.
- Gomes, M.I., Caeiro, F. and Figueiredo, F. (2004). Bias reduction of a tail index estimator through an external estimation of the second order parameter. *Statistics* **38**:6, 497-510.

- Gomes, M.I. and Figueiredo, F. (2006). Bias reduction in risk modelling: semi-parametric quantile estimation. *Test* **15**:2, 375-396.
- Gomes, M.I., Figueiredo, F. and Mendonça, S. (2004). Asymptotically best linear unbiased tail estimators under a second order regular variation. *J. Statist. Planning and Inference* **134**:2, 409-433.
- Gomes, M.I. and de Haan, L. (1999). Approximations by extreme value distributions. *Extremes* **2**:1, 71-85.
- Gomes, M.I., de Haan, L. and Peng, L. (2002). Semi-Parametric Estimation of the Second Order Parameter — Asymptotic and Finite Sample Behaviour. *Extremes* **5**:4, 387-414, 2002.
- Gomes, M.I., de Haan, L. and Pestana, D. (2004). Joint exceedances of the ARCH process. *J. Applied Probab.* **41**:3, 919-926.
- Gomes, M.I., Hall, A. and Miranda, C. (2005). Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. *J. Comp. Statist. and Data Analysis* (Aceite em Dezembro, 2006).
- Gomes, M.I. and Martins, M.J. (2001). Generalizations of the Hill estimator — asymptotic versus finite sample behaviour. *J. Statist. Planning and Inference* **93**, 161-180.
- Gomes, M.I. and Martins, M.J. (2002). “Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.
- Gomes, M.I. and Martins, M.J. (2004). Bias reduction and explicit efficient estimation of the tail index. *J. Statist. Planning and Inference* **124**, 361-378.
- Gomes, M.I., Martins, M. J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology. *Extremes* **3**:3, 207-229.
- Gomes, M.I., Martins, M. J. and Neves, M. (2003). Generalized Jackknife semi-parametric estimators of the tail index. *Portugaliae Mathematica* **59**:4,393-408.
- Gomes, M.I., Miranda, C. and Pereira, H. (2005). Revisiting the role of the Jackknife methodology in the estimation of a positive tail index. *Comm. in Statistics — Theory and Methods* **34**, 1-20.
- Gomes, M.I., Miranda, C. and Viseu, C. (2006). Reduced bias tail index estimation and the Jackknife methodology. *Statistica Neerlandica* **60**:4, 1-28.
- Gomes, M.I. and van Montfort, M.A.J. (1987). Exponentiality versus Generalized Pareto — quick tests. In *Proc. III Internat. Conf. Statistical Climatology*, 185-195. (1986).
- Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes — choice of the optimal sample fraction. *Extremes* **4**:4, 331-358 (2002).
- Gomes, M.I. and Oliveira, O. (2003a). Censoring estimators of the tail index. *Statistics and Probability Letters* **65**, 147-159.
- Gomes, M.I. and Oliveira, O. (2003b). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events. *Comm. in Statist. — Simulation and Computation* **32**:4, 1005-1028.
- Gomes, M.I. and Oliveira, O. (2003c). Maximum likelihood revisited under a semi-parametric context — estimation of the tail index. *J. Statist. Computation and Simulation* **73**:4, 285-301.
- Gomes, M.I. and Pestana, D. (1981). On the domain of attraction of stable and of extreme value distributions. *Bull. Soc. Math. Grèce* **22**, 105-120.
- Gomes, M.I. and Pestana, D. (1986). Large claims — extreme value models. In Goovaerts et al (eds.), *Insurance and Risk Theory*, D. Reidel, Dordrecht.
- Gomes, M.I. and Pestana, D. (1987a). Nonstandard domains of attraction and rates of convergence. In M.L. Puri et al. (eds.), *New Perspectives in Theoretical and Applied Statistics*, 467-477, Wiley, New York.
- Gomes, M.I. and Pestana, D. (1987b). Non-classical extreme value models. In *Proc. III Internat. Conf. Statistical Climatology*, 196-200.
- Gomes, M.I. and Pestana, D. (2004). A simple second order reduced bias' tail index estimator. *J. Statist. Comput. and Simulation* (Aceite em Maio, 2005).
- Gomes, M.I. and Pestana, D. (2005). A sturdy reduced bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.* (aceite).
- Gomes, M.I., Rodrigues, L., Vandewalle, B. and Viseu, C. (2006). A heuristic adaptive choice of the threshold for bias-corrected Hill estimators. *J. Statist. Comput. and Simulation* (aceite).
- Graça Martins, M.E. and Pestana, D. (1987). Nonstable limit laws in extreme value theory. In *New Perspectives in Theoretical and Applied Statistics*, 449-457, Wiley, New York.
- Graça Martins, M.E. and Pestana, D. (1988). The extremal limit problem — extensions. In *Probability Theory and Mathematical Statistics with Applications*, 143-153, D. Reidel, Dordrecht.

- Guedes Soares, C. and Scotto, M. G. (2001). Modelling uncertainty in long-term predictions of significant wave height. *Ocean Engineering* **28**, 329-342.
- Guedes Soares, C. and Scotto, M. G. (2004). Application of the r-largest order statistics for long-term predictions of significant wave height. *Coastal Engineering*. **51**, 387-394.
- Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia Univ. Press, New York.
- de Haan, L. and Canto e Castro, L. (2006). A class of distribution functions with less bias in extreme value estimation. *Statist. Probab. Lett.* **76**:15, 1617-1624.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: an Introduction*. Springer Series in Operations Research and Financial Engineering, New York.
- de Haan, L., Peng, L. and Iglésias Pereira, H. (1999). Approximation by penultimate stable laws. *Probab. Math. Statist.* **19**:1, 105-121.
- de Haan, L., Li, D., Peng, L. and Iglésias Pereira, H. (2002). Alternative conditions for attraction to stable vectors. *Probab. Math. Statist.* **22**:2, 303-317.
- de Haan, L. and Themido Pereira, T. (1999). Estimating the index of a stable distribution. *Statist. Probab. Lett.* **41**:1, 39-55.
- de Haan, L. and Themido Pereira, T. (2006). Spatial Extremes: models for the stationary case. *Ann. Statist.* **34**, 146-168.
- Hall, A. (1996). Maximum term of a particular autoregressive sequence with discrete margins. *Comm. Statist. — Theory Methods* **25**:4, 721-736.
- Hall, A. (1998). *Extremos de Sucessões de Contagens — do Outro Lado do Espelho*. Tese de doutoramento, DEIO/FCUL, Universidade de Lisboa.
- Hall, A. (2001). Extremes of moving average models with regularly varying tails. *Extremes* **4**:3, 219-239 (2002).
- Hall, A. (2003). Extremes of integer-valued moving average models with exponential type tails. *Extremes* **6**:4, 361-379 (2005).
- Hall, A. and Hüsler, J. (2006). Extremes of stationary sequences with failures. *Stoch. Models* **22**:3, 537-557.
- Hall, A. and Moreira, O. (2006). On the extremes of a particular moving average count data model. *Statist. Probab. Lett.* **76**:2, 135-141.
- Hall, A. and Scotto, M.G. (2003). Extremes of sub-sampled integer-valued moving average models with heavy-tailed innovations. *Statist. Probab. Lett.* **63**:1, 97-105.
- Hall, A. and Scotto, M. (2006). Extremes of periodic integer-valued sequences with exponential type tails. *RevStat* **4**:3, 249-273.
- Hall, A., Scotto, M. and Ferreira, H. (2004). On the extremal behaviour of generalised periodic sub-sampled moving average models with regularly varying tails. *Extremes* **7**:2, 149-160 (2005).
- Hall, A. and Temido, M.G. (2005). On the maximum term of MA and Max-AR models with margins in Anderson's class. *Theory of Probability and its Applications* (aceite).
- Hüsler, J., Cruz, P., Hall, A. and Fonseca, C.M. (2003). On optimization and extreme value theory. *Methodol. Comput. Appl. Probab.* **5**:2, 183-195.
- Iglésias Pereira, H. (1983). Rate of convergence towards a Fréchet type limit distribution. *Ann. Sci. Univ. Clermont-Ferrand II Probab. Appl.* **1**, 67-80.
- Iglésias Pereira, H. (1990). Tests for the characteristic exponent and the scale parameter of symmetric stable distributions. *Comm. Statist. — Simulation Comput.* **19**:4, 1465-1475.
- Martins, A.P. (2005). *Coeficientes Extremais*. Tese de doutoramento, Departamento de Matemática, Universidade da Beira Interior.
- Martins, A.P. and Ferreira, H. (2004a). The extremal index of sub-sampled processes. *J. Statistical Planning and Inference* **124**:1, 145-152.
- Martins, A.P. and Ferreira, H. (2004b). Extremes of periodic moving averages of random variables with regularly varying tail probabilities. *SORT* **28**:2, 165-175.
- Martins, A.P. and Ferreira, H. (2005a). Measuring the extremal dependence. *Statist. Probab. Lett.* **73**:2, 99-103.
- Martins, A.P. and Ferreira, H. (2005b). The multivariate extremal index and the dependence structure of a multivariate extreme value distribution. *Test* **14**:2, 433-447.
- Martins, M.J. (2001). *Estimação de Caudas Pesadas — Variantes ao Estimador de Hill*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.
- Martins, M.J., Gomes, M.I. and Neves, M.M. (1999). Some results on the behaviour of Hill's estimator. *J. Statist. Comp.*

and Simulation **63**, 283-297.

Martins, M.J., Gomes, M.I. and Neves, M.M. (2004). Averages of Hill estimators. *Test* **13**:1, 1-16.

Mendonça, S. (2000). *Tópicos Sobre a Convergência Fraca de Variáveis Aleatórias*. Tese de doutoramento, Universidade da Madeira.

Miguéns, M.F. (2004). *Modelo Bivariado para Caudais Máximos*. Tese de doutoramento, FCT, Universidade Nova de Lisboa.

Miranda, M.C. (2005). *Estatística de Extremos: Estimação dos Índices de Cauda e Extremal*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.

van Montfort, M.A.J. and Gomes, M.I. (1985). Statistical choice of extremal models for complete and censored data. *J. Hydrology* **77**, 77-87.

Moreira de Freitas, A.C. (2005). *Estimação do Coeficiente de Cauda Exponencial. Aplicação à Teoria do Risco*. Tese de doutoramento, Departamento de Matemática Aplicada, Universidade do Porto.

Neves, C. (2006). *Estimation and Testing for Distributions with Light, Heavy and Super-heavy Tails*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.

Neves, C. and Fraga Alves, M.I. (2004). Reiss and Thomas' automatic selection of the number of extremes. *J. Comput. Statist. Data Anal.* **47**:4, 689-704.

Neves, C. and Fraga Alves, M.I. (2004). Semi-parametric Approach to the Hasofer-Wang and Greenwood statistics in Extremes. *Test* (aceite).

Neves, C., Picek, J. and Fraga Alves, M.I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction. *J. Statist. Plann. Inference* **136**:4, 1281-1301.

Oliveira, M.F. and Turkman, K.F. (1992). A note on the asymptotic independence of maximum and minimum of stationary sequences with extremal index. *Portugal. Math.* **49**:1, 29-36.

Oliveira, O. (2003). *In Extremis*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.

Oliveira, O., Gomes, M.I. and Fraga Alves, M.I. (2006). Improvements in the estimation of a heavy tail. *RevStat* **4**:2, 81-109.

Pereira, L. (2002). *Valores Extremos Multidimensionais de Variáveis Dependentes*. Tese de doutoramento, Universidade da Beira Interior.

Pereira, L. (2004). Extremal behaviour in models of superposition of random variables. *RevStat* **2**:2, 163-178.

Pereira, L. and Ferreira, H. (2001). Limit distribution for point processes of high local maxima. *J. Statistical Planning and Inference* **97**, 227-233.

Pereira, L. and Ferreira, H. (2002). The asymptotic locations of the maximum and minimum of stationary sequences. *J. Statistical Planning and Inference* **104**, 287-295.

Pereira, L. and Ferreira, H. (2005). Extremes of quasi-independent random fields and clustering of high values. *WSEAS Trans. Math.* **4**:4, 333-338.

Pereira, I. and Scotto, M.G. (2006). On the non-negative first-order exponential bilinear time series model. *Statist. Probab. Lett.* **76**:9, 931-938.

Pestana, D. (1981). Positive stable laws and M-scheme. *Portugal. Math.* **40**, no. 4, 433-438 (1985).

Ramos, A. (2002). *Multivariate joint Modelling and Score Tests of Independence*. Ph.D. Thesis, University of Surrey, UK.

Ramos, A. and Ledford, A.W. (2005). Regular score tests of independence in multivariate extreme values. *Extremes* **8**:1, 5-26.

Rosado, F. (ed.) (2005). *Memorial da Sociedade Portuguesa de Estatística*. Edições SPE.

Scotto, M.G. (2001). *On the Extremes of Certain Time Series*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.

Scotto, M.G. (2003). A note on the asymptotic distribution of the maxima in disaggregated time-series models. *Statist. Probab. Lett.* **65**:2, 127-137.

Scotto, M.G. (2005). Extremes of a class of deterministic sub-sampled processes with applications to stochastic difference equations. *Stochastic Process. and Appl.* **115**:3, 417-434.

Scotto, M. G. (2006). On the extremes of a class of non-linear processes with heavy tailed innovations. *Nonlinear Analysis*. (aceite).

Scotto, M. G. and Ferreira, H. (2003). Extremes of deterministic sub-sampled moving averages with heavy-tailed innovations. *Applied Stochastic Models in Business and Industry* **19**:4, 303-313.

- Scotto, M. G. and Guedes Soares, C. (2000). Modelling the long-term series of significant wave heights with non-linear threshold models. *Coastal Engineering* **40**, 313-327.
- Scotto, M.G. and Guedes Soares, C. (2006). Bayesian inference for long-term prediction of significant wave height. *Coastal Engineering*. (aceite).
- Scotto, M. G. and Turkman, K. F. (2002). On the extremal behaviour of sub-sampled solutions of stochastic difference equations. *Portugaliae Mathematica* **59**:3, 267-282.
- Scotto, M.G. and Turkman, K.F. (2005). Extremes of Volterra series expansions with heavy-tailed innovations. *Nonlinear Anal.* **63**:1, 106-122.
- Scotto, M., Turkman, K.F. and Anderson, C.W. (2003). Extremes of some sub-sampled time series. *Journal of Time Series Analysis* **24**:5, 579-590.
- Sousa, B.C. (2002). *Contribution to the Estimation of the Tail Index of Heavy-Tailed*. Ph.D. Thesis, University of Michigan, USA.
- Sousa B. and Michailidis G. (2004). A New Diagnostic Plot for the Tail Index of Heavy Tailed Distributions. *Journal of Computational and Graphical Statistics* **13**, 974-1001.
- Temido, M.G. (1999). Maxima and minima of stationary random sequences under a local dependence restriction. *Portugal. Math.* **56**:1, 59-72.
- Temido, M.G. (2000). *Classes de Leis Limites em Teoria de Valores Extremos*. Estabilidade e Semi estabilidade. Tese de doutoramento, Departamento de Matemática, Universidade de Coimbra.
- Temido, M. G. (2000). Mixture results for extremal behaviour of strongly dependent nonstationary Gaussian sequences. *Test* **9**:2, 439-453.
- Temido, M. G. (2003). New limiting distribution of maxima of independent random variables. *Portugaliae Mathematica* **60**:4, 379-388.
- Temido, M. G. and Canto e Castro, L. (2003). Max-semistable laws in extremes of stationary random sequences. *Theory Probab. and its Applications* **47**:2, 365-374.
- Themido Pereira, T. (2003). Spacial extremes: the stationary case. *Bull. Intern. Statist. Inst.* **LX**, 2:2, 249.
- Tiago de Oliveira, J. (1981). Statistical choice of univariate extreme models. In Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, 367-387, D. Reidel, Dordrecht-Boston, Mass.
- Tiago de Oliveira, J. (1982). Decision and modelling for extremes. In *Some Recent Advances in Statistics*, 101-110, Academic Press, London.
- Tiago de Oliveira (ed.) (1984a). *Statistical Extremes and Applications*. NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences, 131. D. Reidel, Dordrecht.
- Tiago de Oliveira, J. (1984b). Univariate extremes; statistical choice. In Tiago de Oliveira ed., *Statistical Extremes and Applications*, 91-107, D. Reidel, Dordrecht.
- Tiago de Oliveira, J. (1984c). Bivariate models for extremes; statistical decision. In Tiago de Oliveira ed., *Statistical Extremes and Applications*, 131-153, D. Reidel, Dordrecht.
- Tiago de Oliveira, J. (1985). Statistical choice of nonseparated one-parameter models. *Trabajos Estadist. Investigación Oper.* **36**, 138-152.
- Tiago de Oliveira, J. (1987a). Comparaison entre les modèles bivariés logistique et naturel pour les maxima et extensions. *C. R. Acad. Sci. Paris Sér. I Math.* **305**:11, 481-484.
- Tiago de Oliveira, J. (1987b). Fundamentals of finite multiple decision; application to statistical choice of models. *Metron* **45**:3-4, 159-175.
- Tiago de Oliveira, J. (1989a). Statistical decision for bivariate extremes. In *Extreme Value Theory*, 246-261, Lecture Notes in Statist., 51, Springer, New York.
- Tiago de Oliveira, J. (1989b). Intrinsic estimation of the dependence structure for bivariate extremes. *Statist. Probab. Lett.* **8**:3, 213-218.
- Tiago de Oliveira, J. (1991). Identifiability of some bivariate extremes models. *Statistics* **22**:1, 149-154.
- Tiago de Oliveira, J. (1992). Intrinsic estimation of the dependence function in bivariate extremes: a new and simpler approach. *Comm. Statist. Theory Methods* **21**:3, 599-611.
- Tiago de Oliveira, J. and Gomes, M.I. (1984). Two test statistics for choice of univariate extreme models. In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, 651-668, D. Reidel.
- Tobías, A. and Scotto, M.G. (2005). Prediction of extreme ozone levels in Barcelona, Spain. *Environmental Monitoring & Assessment*. **100**, 23-32.

- Turkman, K.F. (1983/84). Degenerate limit laws for the maxima of trigonometric polynomials with random coefficients. *Portugal. Math.* **42**:4, 355-369 (1986).
- Turkman, K.F. (1984). On the asymptotic upcrossings of a class of nonstationary sequences. In Tiago de Oliveira (ed.), *Statistical Extremes and Applications*, 669-678, D. Reidel, Dordrecht.
- Turkman, K. F. (2001). Extremal behaviour of trigonometric polynomials with random coefficients. *Nonlinear Analysis* **47**:5, 3113-3124.
- Turkman, K.F. (2006). A note on the extremal index for space-time processes. *J. Appl. Probab.* **43**:1, 114-126.
- Turkman, K.F. and Amaral Turkman, M.A. (1997). Extremes of bilinear time series models. *J. Time Ser. Anal.* **18**:3, 305-319.
- Turkman, K.F. and Oliveira, M.F. (1992). Limit laws for the maxima of chain-dependent sequences with positive extremal index. *J. Appl. Probab.* **29**:1, 222-227.
- Turkman, K.F. and Walker, A.M. (1983). Limit laws for the maxima of a class of quasistationary sequences. *J. Appl. Probab.* **20**:4, 814-821.
- Turkman, K.F. and Walker, A.M. (1984). On the asymptotic distributions of maxima of trigonometric polynomials with random coefficients. *Adv. in Appl. Probab.* **16**:4, 819-842.
- Turkman, K.F. and Walker, A.M. (1990). A stability result for the periodogram. *Ann. Probab.* **18**:4, 1765-1783.
- de Zea Bermudez, P. (2003). *Bayesian Approach to Extreme Quantile Estimation*. Tese de doutoramento, DEIO, FCUL, Universidade de Lisboa.
- de Zea Bermudez, P., Amaral Turkman, M.A. and Turkman, K.F. (2001). A predictive approach to tail probability estimation. *Extremes* **4**:4, 295-314 (2002).
- de Zea Bermudez, P. and Amaral Turkman, M.A. (2003). A Bayesian approach to parameter estimation of the generalized Pareto distribution. *Test* **12**:1, 259-277.



Ciência Estatística

• Livro

Título: Extreme Value Theory: an Introduction

Autores: Laurens de Haan e Ana Ferreira

Editora: Springer Series in Operations Research and Financial Engineering

Ano: 2006

• Artigos Científicos Publicados

Hall, A. e Husler, J. (2006) Extremes of stationary sequences with failures, *Stochastic Models*, 22, 537-557.

Hall, A. e Moreira, O. (2006) A note on the extremes of a particular moving average count data model, *Statistics and Probability Letters*, 76, 135-141.

Henriques C. e Oliveira, P. E. (2006) Convergence rates for the estimation of two-dimensional distribution functions under association and estimation of the covariance of the limit empirical process. *Journal of Nonparametric Statistics*, 18, 119--128, 2006.

Menezes, R., Garcia-Soidán, P. e Febrero-Bande, M. (2005). A comparison of approaches for valid variogram achievement. *Computational Statistics*, 20 (4), 623-42.

Salgueiro, M.F., Smith, P.W.F. e McDonald, J.W. (2006). Power of edge exclusion tests for graphical log-linear models. *Journal of Multivariate Analysis*, 97(8), p. 1691-1701

Silva, I. e Silva, M.E. (2006). Asymptotic Distribution of the Yule-Walker Estimator for INAR(p) processes. *Statistics and Probability Letters*, Vol. 76 – nº 15, pp. 1655-1663.

• Teses de Mestrado

Título: Escolha de Variáveis em Análises Estatísticas Multidimensionais

Autora: Maria Isabel Pinto Preto, isabelppreto@hotmail.com

Orientador: Francisco José Lage Campelo Calheiros

Título: Modelos Pontuais de Distribuição em Visão Computacional: Estudo, Desenvolvimento e Aplicação

Autora: Maria João Medeiros de Vasconcelos, mjmvasconcelos@gmail.com

Orientador: João Manuel Tavares

Título: Um Sistema de Informação e Modelo de Diagnóstico Diferencial no Âmbito de um Rastreamento Populacional

Autor: Rui Manuel Cerqueira Magalhães, mes99002@fe.up.pt

Orientadores: Maria Carolina Costa e Silva e Francisco José Lage Campelo Calheiros

Título: Análise de Incertezas em Ensaios de Qualidade

Autora: Cláudia Isabel Pires Moreiras, claudiamoreiras@gmail.com

Orientadores: Francisco José Lage Campelo Calheiros e Maria Elisa Abrantes da Costa

Título: Comparação Assintótica de Estimadores Semi-paramétricos do Índice de Variação Regular

Autor: Manuel Dias Pedro, mdp@netcabo.pt

Orientadora: Maria Ivette Gomes

Título: Análise conjunta baseada em preferências vs escolhas. Selecção de uma escola de gestão

Autora: Susana Tavares, sustav@netcabo.pt

Orientadores: Margarida Cardoso e José Gonçalves Dias

Título: Um Simulador de Tráfego para o Estudo do Ruído

Autora: Valquíria da Silva Fortuna, mes00003@fe.up.pt

Orientador: Francisco José Lage Campelo Calheiros

• Teses de Doutoramento

Título: Modelos Aditivos Generalizados em Análise de Sobrevida

Autora: Ana Luísa Papoila, *apapoila@hotmail.com*

Orientadora: Cristina Maria Simões Rocha

Na minha tese foi proposto um novo modelo de regressão para dados de sobrevivência, desenvolvido com base na correspondência existente entre modelos de sobrevivência e modelos de regressão binária. No sentido de resolver o problema da especificação incorrecta da função de ligação em Modelos Lineares Generalizados (GLMs) surgiram novas metodologias, as quais deram origem a modelos mais flexíveis, o que beneficiou também a Análise de Sobrevida, dada a existência da referida correspondência. De facto, ao estudarmos tempos de vida, um modelo dotado de grande flexibilidade é o modelo Gama-logit, equivalente a um GLM de resposta binária, com função de ligação desconhecida e pertencente à família de transformações assimétricas de Aranda-Ordaz. No entanto, a linearidade imposta pelos GLMs para descrever a relação funcional entre cada variável independente e a variável resposta é, frequentemente, um obstáculo, ultrapassável pelos Modelos Aditivos Generalizados (GAMs). Assim sendo, propomos a introdução de famílias paramétricas de funções de ligação em GAMs com resposta binária, em particular a referida família de Aranda-Ordaz, por forma a obter, paralelamente, uma generalização do modelo Gama-logit, que designaremos por modelo Gama-logit aditivo. A sua estimação, com base no algoritmo de scores local, baseia-se na minimização da função desvio calculada para uma grelha de valores do parâmetro (desconhecido) de que a função de ligação depende. Foram efectuados estudos de simulação, não só para avaliar a qualidade das estimativas obtidas, mas também para comparar o desempenho do modelo proposto com o do GAM logístico. Toda a inferência foi efectuada através de técnicas de bootstrap, dada a inexistência de teoria assintótica para estes modelos. A nova metodologia foi ilustrada através de uma aplicação prática com dados do estado actual, cujo objectivo foi estudar a distribuição do tempo que decorre desde o instante em que um indivíduo inicia o consumo de drogas injectáveis até à infecção pelo VIH.

Ana Luísa Papoila

Título: Desigualdades Exponenciais e Velocidades de Convergência em Problemas de Estimação para Amostras Associadas

Autora: Carla Manuela Ribeiro Henriques, *carlahenriq@mat.estv.ipv.pt*

Orientador: Paulo Eduardo Oliveira

Na minha tese estudou-se o comportamento de um estimador para a função de covariância do processo empírico limite, admitindo que as variáveis de base satisfazem uma certa estrutura de dependência positiva vulgarmente designada por associação. A importância do processo empírico para diversas aplicações estatísticas e a falta de adequação do pressuposto de independência para inúmeras situações reais serviram de motivação para este estudo. Para a situação em que as variáveis de base são associadas, a função de covariância do processo empírico limite, $\Gamma(s,t)$, é descrita à custa de uma série envolvendo as distribuições bivariadas das variáveis de base. Assim, o trabalho teve origem com o estudo da função de distribuição empírica bivariada, $\hat{F}_{k,n}(s,t)$, para o qual foram estabelecidas algumas propriedades de convergência. A partir de $\hat{F}_{k,n}(s,t)$ constrói-se o estimador para a série envolvida em $\Gamma(s,t)$ e para a própria função de covariância $\Gamma(s,t)$. Num esforço para identificar velocidades para a convergência forte dos estimadores em estudo, o trabalho seguiu um rumo na direcção das desigualdades de tipo exponencial para variáveis associadas, tendo-se obtido alguns resultados sobre velocidade de convergência daqueles estimadores e também resultados no âmbito dos grandes desvios da média empírica de variáveis aleatórias associadas. As desigualdades de tipo exponencial obtidas para o estimador $\hat{F}_{k,n}(s,t)$ são exploradas quanto à aptidão para identificar velocidades de convergência. É ainda explorado uma outra abordagem para obter velocidades de convergência que tem por base uma desigualdade para momentos do tipo Rosenthal. Finalmente, aplicam-se alguns resultados da tese à estimação dos valores próprios do operador integral de núcleo $\Gamma(s,t)$ e à determinação da lei assintótica da estatística do teste do sinal para uma amostra associada.

Carla Henriques

Título: Medidas de Desempenho para Testes de Discordância em Populações Normais

Autor: José Palma, jpalma@est.ips.pt

Orientador: Fernando Rosado

Na minha tese foram calculadas medidas de desempenho significativas para testes de discordância no quadro da distribuição normal, bem como discutidos alguns dos problemas colocados na determinação dessas medidas. Foram ainda obtidas as distribuições das estatísticas de teste, tanto na hipótese nula de ausência de contaminantes, como na hipótese alternativa de existência de um outlier na amostra. Estas distribuições constituem elementos fundamentais no cálculo das medidas.

Tradicionalmente a aproximação ao estudo dos outliers foi no sentido de os detectar através de testes. Estes, como qualquer outro teste de hipóteses, devem ter uma hipótese nula e uma outra alternativa. Assim, podemos concluir que uma observação é um outlier para uma dada hipótese, enquanto que para outra, essa mesma observação poderá ser considerada não discordante.

Como resultado da sua dependência em relação ao modelo de discordância foi proposto um vasto conjunto de testes. No entanto e no sentido de optar por um em particular é necessário obter algumas medidas do seu desempenho relativo, por exemplo a função potência. A avaliação de testes com a mesma potência deverá depender da hipótese alternativa de existência de outliers, pois requer o conhecimento do comportamento da distribuição da estatística de teste nessa hipótese e exigirá como tal outras medidas de qualidade. Por exemplo, na hipótese de existência de um outlier, são medidas significativas, entre outras: a probabilidade de rejeitar H_0 e o contaminante satisfazer o critério de rejeição; a probabilidade de rejeitar H_0 e identificar correctamente o outlier e a probabilidade de rejeitar H_0 condicionada à identificação correcta do suspeito de contaminação.

Apesar da grande variedade de testes propostos na literatura, são limitados os estudos de desempenho que os permitem comparar. Com esta tese pretendeu-se suprir essa lacuna, propondo um conjunto de metodologias e medidas de comparação de testes de discordância.

José Palma

Título: Uma Abordagem Bayesiana à Determinação de Modelos

Autora: Júlia Teles, jteles@fmh.utl.pt

Orientadores: Luís Camilo do Canto de Loura e Maria Antónia Amaral Turkman

Na minha tese faço uma abordagem bayesiana ao problema de selecção e validação de modelos, isto é, ao problema de determinação de modelos. Efectuo uma revisão de alguns métodos de selecção e validação de modelos, proponho métodos de pesquisa do “melhor” subconjunto de covariáveis em modelos de regressão e, através de estudos de simulação, comparo a performance de alguns critérios de selecção de modelos.

Relativamente aos métodos de selecção de modelos, comparo o factor Bayes com diversos critérios de verosimilhança penalizada, entre eles o critério de informação bayesiano, o critério de informação bayesiano estimado e o critério de informação da deviance. Como alternativa à selecção de modelos sugiro, em certos casos, a utilização da média ponderada de modelos bayesianos.

Nos problemas de selecção de covariáveis em modelos de regressão, o número de modelos a comparar é, de um modo geral, elevado. Neste caso particular de selecção de modelos, não sendo viável a pesquisa exaustiva, proponho dois métodos bayesianos de pesquisa do “melhor” modelo: os métodos de selecção progressiva e de eliminação regressiva de covariáveis via medidas de discrepância. Uso dois exemplos, um de regressão linear normal e outro de regressão gama, para ilustrar estes métodos de selecção de covariáveis.

Relativamente à validação de modelos, destaco a utilização do *p-value* preditivo *a posteriori* e, em particular, do *p-value* de discrepância, como medidas de adequabilidade do modelo.

Júlia Teles



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

Prémio SPE 2007

Está aberto, até **15 de Junho de 2007**, o concurso para atribuição do **Prémio SPE 2007**, de acordo com o seguinte regulamento:

Pretendendo dar destaque ao XV Congresso Anual da **SPE**, a principal reunião científica organizada pela Sociedade Portuguesa de Estatística, é instituído o **Prémio SPE 2007**.

Este prémio destina-se a estimular a actividade de estudo e investigação científica em Probabilidade e Estatística entre os jovens que trabalham nestas áreas.

O **Prémio SPE 2007** é constituído por uma quantia de 1000 euros.

Ao **Prémio SPE 2007** podem concorrer trabalhos originais sobre temas de Probabilidade e Estatística, desde que não tenham sido objecto de qualquer prémio atribuído por outra instituição.

Os candidatos ao **Prémio SPE 2007** devem ser estudantes ou investigadores em alguma instituição portuguesa ou bolseiros portugueses, não devem ter idade superior a *35 anos*, no ano a que respeita o concurso, e devem ser sócios da **SPE**. O candidato não deve ter recebido o Prémio SPE nas quatro edições anteriores.

O trabalho deve ser escrito em português e não poderá exceder 25 páginas A4.

As candidaturas, o trabalho concorrente e o curriculum vitae do candidato, deverão ser dirigidas ao Presidente da **SPE**, em carta registada, para a morada abaixo indicada. O carimbo do correio validará a data de entrega.

A admissibilidade e a apreciação dos trabalhos submetidos a concurso é da competência de um júri, cuja constituição será da responsabilidade da Direcção da **SPE**.

O júri é soberano nas suas decisões, não havendo lugar a recurso.

O trabalho galardoado com o **Prémio SPE 2007** será apresentado pelo seu autor/pela sua autora no XV Congresso Anual da **SPE** e será publicado nas respectivas Actas.

A atribuição do **Prémio SPE 2007** será anunciada logo que conhecida a decisão do júri e a sua entrega formal será feita na Sessão de Encerramento do XV Congresso Anual da **SPE**.

O **Prémio SPE 2007** poderá não ser atribuído.

Havendo mais do que um candidato para a atribuição, em igualdade de condições, terá preferência o candidato que apresentar melhor *curriculum* e, subsistindo a igualdade, preferirá o candidato mais jovem.

Sociedade Portuguesa de Estatística
Bloco C6, Piso 4 - Campo Grande - 1749-016 LISBOA



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

PRÉMIOS “ESTATÍSTICO JÚNIOR 2007”

Está aberto, até 15 de Maio de 2007, o concurso para atribuição de prémios “Estatístico Júnior 2007”, de acordo com o seguinte regulamento:

1. A atribuição de prémios “Estatístico Júnior 2007” é promovida pela Sociedade Portuguesa de Estatística (SPE), com o apoio da Porto Editora, e tem como objectivo estimular e desenvolver o interesse dos alunos do ensino básico e secundário pelas áreas da probabilidade e estatística.
 2. Os candidatos a prémios “Estatístico Júnior 2007” devem ser alunos do 3.º Ciclo do Ensino Básico ou do Ensino Secundário no ano lectivo 2006/2007.
 3. As candidaturas podem ser individuais ou em grupo com um máximo de 3 alunos. Do grupo pode ainda fazer parte um professor do ensino básico ou secundário ao qual caberá o papel de orientador.
 4. Os candidatos devem apresentar um trabalho cuja temática deve estar relacionada com a teoria da probabilidade e/ou estatística.
 5. O trabalho deverá ser constituído por um texto escrito em Português com um máximo de 10 páginas A4 dactilografadas e um poster formato A2 que resuma os principais aspectos do trabalho.
 6. Poderão ser atribuídos prémios “Estatístico Júnior 2007” a 6 trabalhos: aos três primeiros classificados de entre os trabalhos candidatos do 3.º Ciclo do Ensino Básico e aos três primeiros classificados de entre os trabalhos candidatos do Ensino Secundário. Os prémios são constituídos por produtos pedagógicos editados pela Porto Editora (à excepção de manuais escolares) no valor de 600 euros, 300 euros e 200 euros, a atribuir, respectivamente, aos grupos cujos trabalhos sejam classificados em 1.º, 2.º e 3.º lugar.
 7. Ao professor orientador do trabalho classificado em 1º lugar, em cada categoria, é ainda atribuída uma anuidade grátis como sócio da SPE, ajudas de custo para participação no XV Congresso Anual da SPE e produtos pedagógicos editados pela Porto Editora (à excepção de manuais escolares) no valor de 500 Euros.
 8. Aos grupos proponentes dos trabalhos classificados em 1º lugar será também oferecida uma ampliação do correspondente poster que será colocado na Sessão de Posters do XV Congresso Anual da SPE.
 9. O boletim de candidatura, acompanhado do trabalho concorrente, deverá ser dirigido ao Presidente da SPE para a morada abaixo indicada. O carimbo do correio validará a data de entrega.
- Sociedade Portuguesa de Estatística – Bloco C6, Piso 4 – Campo Grande – 1749-016 Lisboa**
10. A admissibilidade e apreciação dos trabalhos submetidos a concurso é da competência de um júri, cuja constituição e nomeação será da responsabilidade da Direcção da SPE.
 11. O júri é soberano nas decisões, não havendo lugar a impugnação ou recurso.
 12. A atribuição dos prémios “Estatístico Junior 2007” será anunciada logo que conhecida a decisão do júri e a sua entrega formal será realizada no XV Congresso Anual da SPE.
 13. Os prémios “Estatístico Junior 2007” poderão não ser atribuídos.

Apoio da Porto Editora