Maria de Fátima Brilhante
Dinis Pestana
José Rocha †
Maria Luísa Rocha
Sílvio Velosa

INFERÊNCIA ESTATÍSTICA SOBRE A LOCALIZAÇÃO USANDO A ESCALA

Ficha Técnica:

Título: Inferência Estatística sobre a localização usando a escala

Autores: Maria de Fátima Brilhante, Dinis Pestana, José Rocha, Maria Luísa Rocha e Sílvio Velosa

Editora: Sociedade Portuguesa de Estatística

Capa: Instituto Nacional de Estatística

Impressão: Instituto Nacional de Estatística

Tiragem: 150 exemplares

ISBN: 978-989-25-0149-9

Depósito Legal: 336729/11

Preço: €10,00 (IVA incluído à taxa em vigor)

INFERÊNCIA ESTATÍSTICA SOBRE A LOCALIZAÇÃO USANDO A ESCALA

Maria de Fátima Brilhante
Universidade dos Açores (DM) e
Centro de Estatística e Aplicações da Universidade de Lisboa

Dinis Pestana
Universidade de Lisboa (FCUL, DEIO) e
Centro de Estatística e Aplicações da Universidade de Lisboa

José Rocha †
Universidade dos Açores (DM) e
Centro de Estatística e Aplicações da Universidade de Lisboa

Maria Luísa Rocha
Universidade dos Açores (DEG) e
Centro de Estudos de Economia Aplicada do Atlântico

Sílvio Velosa
Universidade da Madeira (CCCEE) e
Centro de Estatística e Aplicações da Universidade de Lisboa

2011

Ao Professor Bento Murteira, autor de
Estatística: Inferência e Decisão,
que foi uma inspiração para todos nós.

Esta segunda edição é também dedicada
ao Professor José Carlos Andrade Rocha,
com saudade.

FCT

Este trabalho foi financiado por Fundos Nacionais através da FCT — Fundação para a Ciência e a Tecnologia no âmbito do projecto PEst-OE/MAT/UI0006/2011.

Prefácio

A presente obra inspira-se em *Inferência Estatística sobre Localização e Escala*, de Brilhante *et al.* (2001), actualizando o texto e ampliando-o, incorporando nomeadamente novos resultados sobre *ANOSp* (*Analysis of Spacings*) em populações exponenciais, observações sobre verosimilhança, suficiência e localização e escala.

No trabalho de revisão, prévio à muitas alterações a que procedemos, tropeçámos amiúde na leitura, concluindo assim que muitas das deduções tinham sido apresentadas omitindo passos que, não sendo difíceis, não são óbvios e são trabalhosos. Acrescentámos por isso detalhes que tornam a leitura mais amena e proveitosa, para além de naturalmente termos procurado melhorar o texto, expurgando-o tão exaustivamente quanto conseguimos de erros e gralhas. Mantivemos por outro lado alguma redundância propositada, que visa simplificar a leitura de partes do texto sem constantemente obrigar a consultar o índice remissivo e folhear o livro, uma vez que actualmente a edição é muito mais fácil (e, sobretudo, menos dispendiosa) do que a antiga composição tipográfica, pelo que a concisão da escrita matemática clássica deixou de ser um imperativo económico.

Os resultados interessantes que se obtêm no caso de populações parentes gaussianas, uniformes e exponenciais não são facilmente generalizáveis a outros modelos. Pareceu-nos interessante incluir algumas razões dessa limitação. Em particular, mostramos que os referidos modelos são os únicos que, sob as condições de regularidade de Koopman em famílias exponenciais, admitem uma condensação óptima da informação de uma amostra num par de estatísticas suficiente para o par de parâmetros (localização, escala). Juntámos também a caracterização da exponencial baseada na independência dos *spacings* à caracterização da gaussiana pela independência de média e variância empíricas.

A incorporação de resultados que não apareciam no anterior curso permitiu-nos também optar por um fio condutor das matérias abordadas que nos parece, actualmente, dar uma perspectiva mais coesa e interessante, e porventura mais estimulante para quem queira identificar problemas em aberto. Escolhemos também um título mais adequado. Não é portanto a reedição de um livro que envelheceu, em muitos aspectos é de facto um livro novo. Justifica-se plenamente, no entanto, a manutenção de José Rocha como co-autor.

Fátima Brilhante
Dinis Pestana
Maria Luísa Rocha
Sílvio Velosa

Ponta Delgada, Julho/Setembro de 2011

Prefácio de *Inferência Estatística Sobre Localização e Escala*

Agradecemos à Sociedade Portuguesa de Estatística o convite para realizarmos este curso. Não nos tendo sido imposto um modelo, decidimos optar por discutir alguns aspectos menos conhecidos do que quotidianamente ocupa tantos cientistas: comparar efeitos médios de tratamentos.

A solução no caso de populações parentes gaussianas com a mesma variância é bem conhecida. Noutras circunstâncias, em geral adoptam-se abordagens não paramétricas. O que procuramos é alertar os nossos auditores para as soluções que existem no caso de populações parentes gaussianas mas com variâncias diferentes (e confessamos já que transformar os dados não é do nosso agrado). Fisher, Jeffreys, Welch e Satterthwaite resolveram a questão de forma satisfatória, mas a polémica em que Fisher se envolveu com Bartlett e Welch levou-o, com o mau feitio que tinha, a preferir considerar o assunto uma curiosidade matemática sem relevância em análise de dados a ter que dar crédito a Welch (que nunca cita!). Obter um t de Student com um número de graus de liberdade fraccionário obrigou, naturalmente, a publicar tabelas de quantis críticos, ou a proceder a aproximações cuja qualidade não foi devidamente avaliada.

Scheffé iluminou a questão, mostrando que no caso de heterocedasticidade não é possível encontrar uma estatística que seja uma função invariante para permutações dos argumentos que tenha distribuição exacta t de Student. Por outro lado, construíu um teste aleatorizado que tem todas as boas propriedades requeridas, nomeadamente ter distribuição exacta t de Student, permitindo o recurso às tabelas usuais.

São questões que, contrariamente ao ponto de vista expresso por Fisher, nos parece terem cada vez mais relevância em análise de dados. De facto, a necessidade de proceder a grandes estudos de meta-análise, coloca na ordem do dia analisar dados recolhidos com metodologias diferentes, naturalmente com diferentes graus de precisão.

Transformar linearmente os dados contém em si mesmo a ideia que ou os dados originais não são gaussianos, ou os transformamos em dados não gaussianos. A inferência sobre localização em dados não gaussianos é complicada — apresentamos resultados para populações simétricas (é consolador verificar que no caso de parente gaussiana chegamos à densidade da t de Student!), e para populações uniformes e exponenciais, recorrendo a transformadas integrais inversas. É um caminho que merece ser explorado noutras populações, a dificuldade está, naturamente, em encontrar estimadores de localização e de escala que se prestem a usar o teorema de Basu, e conseguir inverter a transformada integral! A “análise da escala” que é possível fazer de forma elegante em populações exponenciais parece difícil de realizar noutros casos.

De facto, a comparação de localizações, mesmo no caso de duas amostras, é problema que não está resolvido. As dificuldades matemáticas que se põem justificam claramente, por enquanto, a opção não paramétrica.

A nossa investigação tem sido apoiada pelas nossas Universidades, pelo Centro de Estatística e Aplicações da Universidade de Lisboa, pela FCT e CITMA, instituições a quem queremos expressar a nossa gratidão.

Fátima Brilhante   Dinis Pestana
José Rocha   Sílvio Velosa

Índice