

Boletim



**SOCIEDADE PORTUGUESA
DE ESTATÍSTICA**

Publicação semestral

outono de 2013



A “Escola Bayesiana” em Portugal

Fundamentos da Inferência Bayesiana	Bento Murteira	10
Representação da Informação a priori	C. D. Paulino e M. A. Amaral Turkman	19
Metodologia Bayesiana em Aplicações Básicas	Carlos Daniel Paulino	27
Estatística Bayesiana: Aplicações em Análise de Sobrevida e Estatística Espacial	Giovani Loiola da Silva	38
Método de Monte Carlo via Cadeias de Markov com saltos reversíveis	Isabel Pereira	45
Métodos Computacionais: INLA, Integrated Nested Laplace Approximation	Isabel Natário	52
Métodos de Monte Carlo Sequenciais e Métodos Computacionais Bayesianos Aproximados	P. de Zea Bermudez	57
Um breve passeio pelo campo computacional bayesiano	Paulo Soares e Valeska Andreozzi	63
Uma pequena conversa sobre Estatística Bayesiana em Genética e Genómica	Nuno Sepúlveda	69

Editorial	1
Mensagem do Presidente	2
Notícias	4
Enigmística	9
Pós-Doc	75
Ciência Estatística	91
• Artigos Científicos Publicados	91
• Teses de Mestrado	91
• Capítulos de Livros	92
• Livros	92
• Tese de Doutoramento	93
Prémios “Estatístico Júnior 2013”	94
Edições SPE	96

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: 500 exemplares

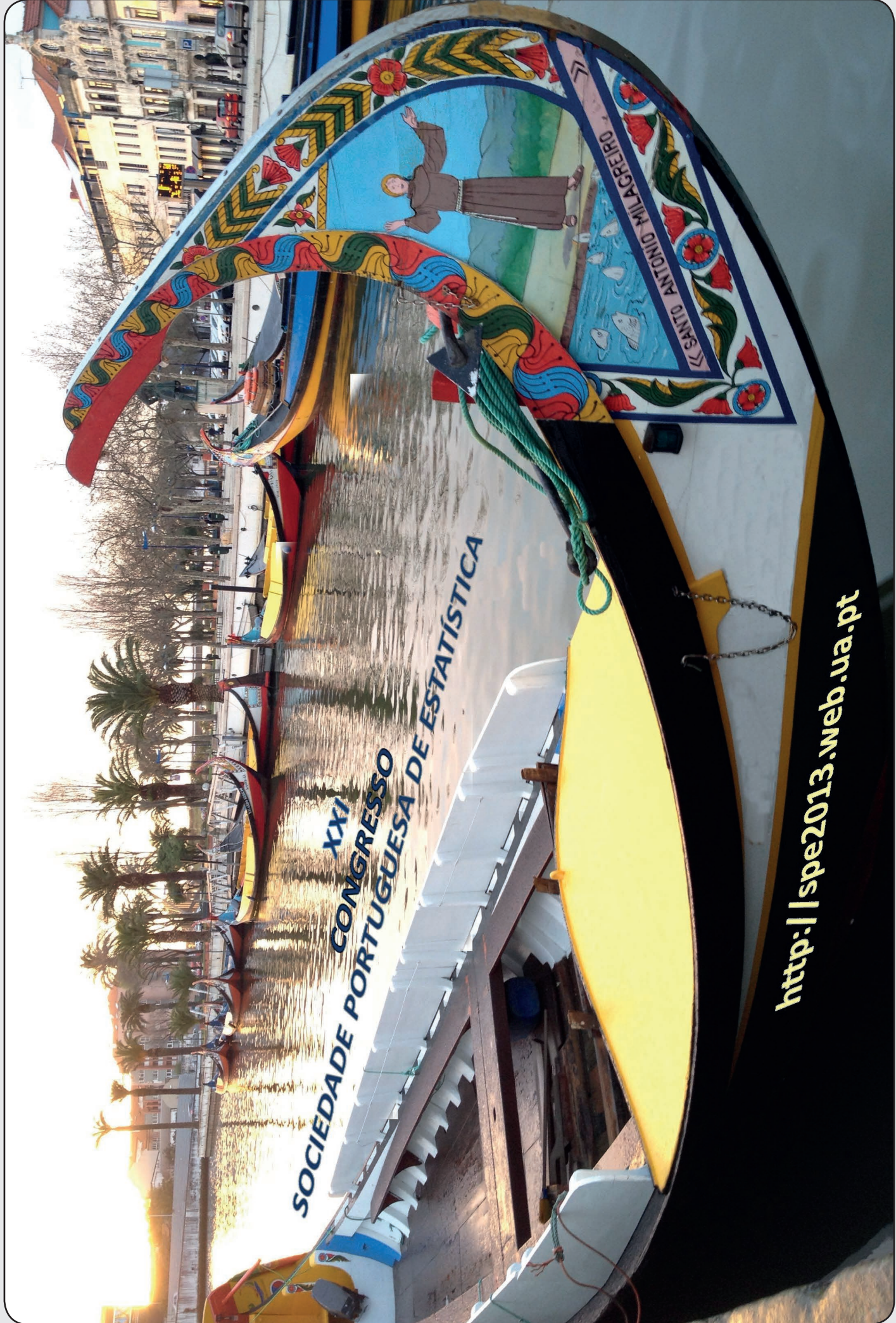
Execução Gráfica e Impressão: Gráfica Sobreireense

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

Sociedade Portuguesa de Estatística desde 1980

**XXI
CONGRESSO
SOCIETATE PORTUGUESA DE ESTATÍSTICA**

<http://spe2013.web.ua.pt>



Editorial

... onde um novo capítulo integra o Memorial da SPE...

1. A “*Escola Bayesiana*” em Portugal, como tema central desta edição, completa um ciclo de registos para memória futura no Boletim SPE - daquilo que foi o passado mas também das bases e dos alicerces, isto é, das raízes da ciência Estatística em Portugal.

Este capítulo temático, com acentuação na escola bayesiana, complementa e atualiza o que já foi inscrito no *Memorial da Sociedade Portuguesa de Estatística*. (editado em 2005).

Como em outros casos, é um tema que “registra” aqueles que na Estatística Portuguesa se realçaram no impulsionamento, implantação e desenvolvimento através de muita investigação conducente a dezenas de teses de mestrado e doutoramento – os pilares de uma continuidade científica com frutos que a todo o título se desejam.

É também um novo momento de congratulação por todo o trabalho pioneiro do Prof Bento Murteira, sócio honorário da SPE, que, reconhecidamente, na atualidade, temos o prazer e a honra de manter como autor primeiro em excelentes contribuições para o Boletim SPE e, em geral, na Sociedade Portuguesa de Estatística.

2. Esta edição do Boletim SPE de outono 2013 surge por ocasião do EVT2013 Extremes in Vimeiro Today - numa feliz (co)incidência - especialmente organizado para homenagear a Prof. Ivette Gomes e ao mesmo tempo recordar os 30 anos de uma Conferência Histórica onde ela teve um protagonismo de excelência e onde tive a oportunidade de participar. Como muitos dos que agora, em 2013, de novo se encontraram no Vimeiro, eu estava nos meus inícios científicos... e, em palavra unânime, há grande congratulação por tudo o que aquele acontecimento significou para portugueses e estrangeiros, onde o pioneirismo de Portugal foi um lema.

O texto da Fernanda Otilia que apresentamos neste Boletim é também um estímulo ao aprofundamento da matéria que ela invoca e isso pode ser feito, por exemplo, com a leitura de “*Extremistas*” num extremo da Europa – um capítulo da autoria de Ivette Gomes (in *Memorial da Sociedade Portuguesa de Estatística*, 2005, F, Rosado (editor), Edições SPE) – ou ainda no *Boletim SPE primavera de 2007* dedicado à “Escola de Extremos” em Portugal e disponível em www.spestatistica.pt. Um relato pode ser lido em <https://www.fc.ul.pt/pt/noticia/30-09-2013/reuni%C3%A3o-dos-extremos-no-vimeiro> sobre A (Re)União dos Extremos no Vimeiro, da autoria de Isabel Alves.


Na página 5 publicamos uma notícia alargada deste acontecimento muito importante: para a própria homenageada, para a comunidade dos estatísticos nacionais e internacionais e, talvez acima de tudo, pelo exemplo de pioneirismo que foi conseguido há 30 anos e que, felizmente, bem o sabemos, perdura.

Em 1983, sem a ligação *on line* da aldeia global em que vivemos, o “isolamento” produzia “outros frutos”, em todos os sentidos, era estimulador de maior disponibilidade, principalmente, para a reflexão, numa relatividade que parece fornecer mais tempo. Naqueles dias o tempo científico correu a favor de todos os participantes que viveram “quase sem sair do hotel” com uma excelente partilha e aproveitamento científico e curricular, por todos reconhecido, em todas as variáveis. Os congressos de hoje são diferentes!?

Enfim, “muita reflexão” em torno do valor de um congresso pode ser feita a propósito de uma simples efeméride que muito nos orgulha.

Parabéns Ivette!

O tema central do próximo *Boletim* será *(Um) Ano Internacional da Estatística*.



Mensagem do Presidente

Pretendo que esta mensagem tenha um cunho informativo de modo a traduzir o empenho da SPE e de muitos dos seus abnegados membros na concretização dos objetivos primários do movimento 2013 - Ano Internacional da Estatística (AIE), em cumprimento de importantes pontos do plano de atividades da atual Direção aprovado na AG de 13/03/2013. Não se aspira, precise-se, a manter preocupações seja de exaustividade na cobertura informativa, seja de concretude na referência dos dedicados colegas intervenientes, aliás já tempestivamente feita em muitos casos, como vem sendo norma nos relatórios anuais de atividade.

Eis então os aspetos que entendo dever realçar aqui pela sua íntima ligação com as linhas programáticas da Direção a que pertence e com os alvos do AIE:

1. A exposição Explorística continua a sua aventura itinerante por escolas diversas do país, vendo sobejamente reconhecida a sua utilidade por estudantes e professores.
2. O projeto Radical Estatística viu concretizada a sua 2ª edição com um aumento substancial dos estudantes envolvidos, e reconhecida internacionalmente a sua importância e o papel associado da SPE com um excelente 2º lugar do pódio no concurso *Best Cooperative Project Award 2013 in Statistical Literacy* lançado pela IASE, cujos resultados foram divulgados no 59º Congresso Mundial de Estatística do ISI realizado em Hong-Kong no fim de agosto último.
3. O ciclo de palestras em escolas prosseguiu dentro das suas possibilidades, tendo a última palestra, *Estatística aplicada em Biologia e domínios afins*, sido realizada no Museu Nacional de História Natural e da Ciência no âmbito do 31º Encontro Juvenil de Ciência realizado na 1ª semana de setembro.
4. A emissão filatélica dos CTT alusiva ao AIE, assessorada pela Direção da SPE, foi uma iniciativa lançada no fim de maio que ajudou a divulgar a nossa área científica e a ação da SPE e da empresa postal portuguesa, junto designadamente das milhares de organizações que no mundo inteiro participam na celebração do AIE.
5. O lançamento, em conjunto com o CEAUL, de um excelente conjunto de pósteres sobre a História da Estatística, exibidos originalmente na sessão de obliteração dos selos alusivos ao AIE, cuja utilidade para a divulgação da Estatística tem sido especialmente destacada por variados setores, vai merecer um meio mais digno de difusão.
6. A realização inédita de Encontros de Biometria teve assinalável êxito pelo nível científico, número de participantes, onde se incluiu um significativo contingente de companheiros brasileiros e galegos, e debate vivo e construtivo sobre a forma de inserção internacional dos biometristas portugueses.
7. A notoriedade científica de um largo leque de participantes e a diversidade dos trabalhos apresentados conferiram um considerável sucesso ao *workshop* EVT-2013, organizado para celebrar os 30 anos da 1ª conferência sobre *Statistical Extremes and Applications* e homenagear a colega Ivette Gomes pela relevância da sua atividade no desenvolvimento de tal área estatística.

8. O XXI Congresso da SPE, com o seu estatuto de reunião científica magna dos seus associados, está aí com a particularidade desta vez de marcar o encerramento oficial da celebração pela SPE do Ano Internacional da Estatística. Mais um motivo de sobra para uma significativa adesão dos sócios a tal evento, não obstante as dificuldades financeiras com que os docentes e investigadores se têm vindo a debater nos últimos tempos por força da reiterada política austeritária de quem nos governa.

9. O reconhecimento público crescente que a SPE vai granjeando tem sido materializado na solicitação oficial de pareceres sobre assuntos educativos, na sua integração em órgãos de assessoria científica e em convites de representação em diversos eventos.

As iniciativas e factos relatados devem ser motivo de satisfação e de orgulho para todos os que se reveem na SPE e, em especial, na estratégia delineada e ação empenhada da atual equipa dirigente e de muitos dos seus generosos membros. Os êxitos alcançados, apesar de todas as sérias vicissitudes externas, devem levar-nos a uma colaboração mais estreita e sintónica para que a afirmação e projeção social da Estatística sejam cada vez mais reforçadas e para que, em particular, os nossos grupos de investigação deixem definitivamente de ser obrigados a encaixarem-se em áreas científicas pouco ou nada condizentes com a sua real atividade de pesquisa na **área científica autónoma de Estatística**.

10 outubro 2013

O Presidente da SPE

Carlos Daniel Paulino

Notícias

• XXI Congresso SPE: Aveiro - 29 novembro a 2 dezembro 2013



O XXI Congresso da Sociedade Portuguesa de Estatística decorrerá este ano de 29 de Novembro a 2 de Dezembro no Hotel Meliá Ria em Aveiro.

É organizado pelo Departamento de Matemática da Universidade de Aveiro e pela Sociedade Portuguesa de Estatística (SPE).

Como tem sido habitual, o congresso da SPE será antecedido por um minicurso intitulado “**Análise de Valores Extremos: uma Introdução**” assegurado por M. Ivette Gomes – DEIO, FC-UL.

O programa científico do Congresso compreende sessões plenárias, sessões temáticas organizadas e ainda comunicações livres (orais ou posters).

Conferencistas convidados

Antónia Amaral Turkman – DEIO, FC-Universidade de Lisboa,
Adrian Bowman – University of Glasgow,
Carlos Braumann – Universidade de Évora,
Esther Ruiz – Universidad Carlos III, Madrid.

Neste congresso a SPE encerrará as atividades relativas ao Ano Internacional da Estatística, *Statistics2013*, com uma sessão especialmente dedicada ao tema. Pretende-se com esta iniciativa “*ampliar a compreensão pública do poder e impacto da Estatística em todos os aspetos da sociedade*” e ainda “*fortalecer a Estatística como uma carreira profissional, especialmente entre os jovens do ensino secundário e superior*”.

É com enorme satisfação que convidamos todos a participarem neste **Congresso em Ano Internacional da Estatística**, deslocando-se a esta cidade maravilhosa intitulada “Veneza de Portugal” para uns dias de formação, partilha de conhecimentos e de estreitamento de laços através do convívio.

A Comissão Organizadora Local

• EVT2013 - Extremes in Vimeiro Today

Foi com um enorme prazer que três amigas de longa data da Ivette Gomes, professora aposentada do Departamento de Estatística e Investigação Operacional (DEIO) da FCUL, decidiram homenageá-la no ano do seu 65º aniversário, organizando o “workshop” “EVT2013 - Extremes in Vimeiro Today”, subordinado ao tema Teoria de Valores Extremos, área de investigação principal da Ivette, na qual ela é reconhecida internacionalmente pelas suas inúmeras publicações de elevada qualidade.

O “workshop” associado a esta celebração, inserido também no âmbito do '2013 - International Year of Statistics', ocorreu de 8-11 de Setembro no grande Hotel do Vimeiro. Foi organizado por Antónia Amaral Turkman e Isabel Fraga Alves, professoras do DEIO-FCUL, e por Manuela Neves, professora do Departamento de Matemática do Instituto Superior de Agronomia, onde participaram cerca de 90 conferencistas, metade deles estrangeiros.



A abertura do evento teve início no domingo, dia 8 de Setembro, com a realização de um jantar de homenagem à Ivette, num ambiente de agradável convívio e alegria, o qual contou com a presença de muitos amigos seus para além dos participantes do congresso. É de referir a intervenção do Coro do INE na abertura do jantar, cuja presença surpreendeu a Ivette, e os discursos dos oradores Ross Leadbetter (da University of North Carolina at Chapel Hill, USA) e Jef Teugels (da Katholieke Universiteit Leuven, Belgium), que muito elogiaram o trabalho de investigação desenvolvido pela Ivette, em particular na área de valores extremos, e que continuará com certeza a desenvolver e a despertar, com o seu entusiasmo, o interesse de muitos jovens estatísticos por esta área de conhecimento. Por fim destacamos as palavras de agradecimento da professora Ivette e o pequeno relato que ela nos proporcionou sobre o seu percurso académico e de investigação, tendo a comissão organizadora oferecido à Ivette, no final do seu discurso, um livro, uma porcelana da região e um ramo de flores, como lembrança e prova de agradecimento.

O programa científico do “workshop” teve lugar de 9-11 de Setembro, tendo o encontro sido considerado um sucesso pelos participantes, não só pela elevada qualidade dos trabalhos científicos nele apresentados – 5 palestras convidadas, 27 comunicações orais distribuídas por 4 sessões temáticas organizadas e 5 sessões livres, 21 trabalhos apresentados sob a forma de poster – mas também por ter permitido o (re)encontro de vários especialistas internacionais da área de Valores Extremos, alguns dos quais estiveram presentes na primeira conferência “Statistical Extremes and Applications”, realizada neste mesmo Hotel do Vimeiro em 1983, e onde a Ivette teve papel importante na organização. Este encontro foi com certeza um ponto de partida importante para o desenvolvimento da área de Extremos em Portugal, que conta atualmente com um número considerável de investigadores.

É de referir ainda o elevado número de jovens interessados pela área que estiveram presentes no EVT2013, com trabalhos de elevado mérito, o que sem qualquer dúvida deixou a Ivette muito satisfeita e contente.

A sessão especial proferida pela homenageada no último dia do evento, a 11 de Setembro, intitulada “Penultimate Approximations: Past, Present... and Future?”, incidiu sobre aproximações pré-assintóticas em classes de distribuições de valores extremos, com uma forte aplicação a problemas de fiabilidade de sistemas complexos tipo série-paralelo e paralelo-série. É curioso referir que a Ivette no início da sua carreira de investigação começou por se dedicar ao estudo deste tipo de aproximações *penultimate*, tendo em seguida enveredado por outras linhas de investigação em valores extremos, mas de facto este tema continua a ser atual e onde é necessário investir, dadas as suas potencialidades de aplicação a situações praticas reais.



Como é desejo de todos nós, esperamos que o entusiasmo e o “know-how” da Ivette continue a fomentar o interesse dos jovens estatísticos pela área de Valores Extremos, e que os “Extremistas” possam dar a sua contribuição para a resolução de problemas importantes do mundo atual.

Fernanda Otília Figueiredo

• Conferência Optimization in the Natural Sciences

De 5 a 9 de fevereiro do próximo ano de 2014, irá realizar-se no Departamento de Matemática da Universidade de Aveiro, a conferência internacional EURO-miniconference on “Optimization in the Natural Sciences”.

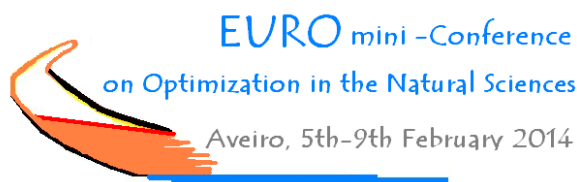
Esta conferência agrega tópicos das áreas científicas de Optimização, Estatística e Ciências da Computação dando especial relevância às aplicações na Física, Biologia, Medicina, Química, entre outras áreas das Ciências Naturais.

Na área da Estatística, o palestrante convidado é o Professor Michael Greenacre, da Universidade de Pompeu Fabra (Barcelona), reconhecido especialista na área da Estatística Multivariada.

Os trabalhos para apresentação, todos sob a forma de comunicação oral, podem ser submetidos até 1 de Novembro de 2013.

Informação detalhada sobre o congresso pode ser obtida através de envio de mensagem para DMAT-miniEURO2014@ua.pt ou consultando a website da conferência: <http://minieuro2014.web.ua.pt/index.php>.

São todos bem-vindos!



P'ela Comissão Organizadora,
Adelaide Freitas

• SPE representada no conselho científico do IAVE

O Decreto-Lei n.º 102/2013, de 25 de julho, criou o Instituto de Avaliação Educativa, I.P. (IAVE), que sucede ao Gabinete de Avaliação Educacional (GAVE). De acordo com o exposto no referido decreto, pretendeu-se criar um organismo com um estatuto de plena independência técnica, pedagógica e científica.

No nº1 do artigo 3º refere-se que “O IAVE tem por missão o planeamento, a conceção e validação dos instrumentos de avaliação externa de conhecimentos e capacidades dos alunos dos ensinos básico e secundário, o tratamento e a divulgação de informação relevante para a tomada de decisões que concorram para incrementar a qualidade, eficácia e eficiência do sistema educativo nacional, assegurar a coordenação da participação nacional em estudos internacionais de avaliação externa de alunos, bem como a elaboração de provas de certificação de conhecimentos e capacidades específicas para outros fins e outros graus de ensino, quando solicitado.”

O conselho científico é o órgão de consulta e apoio técnico-científico em matéria de avaliação e é composto por um representante de cada uma das associações e sociedades científicas e pedagógicas das disciplinas a que respeitam os instrumentos de avaliação a elaborar pelo IAVE. Assim sendo, no seguimento de diligências feitas pelo Presidente da SPE no sentido de que a Sociedade Portuguesa de Estatística integrasse o conselho científico do IAVE, é com satisfação que vos informo que a SPE passou a estar representada neste conselho, desde o passado mês de setembro.

Cristina Rocha
(Representante da SPE no conselho científico do IAVE)

• Prémio Extraordinário de Doctorado

A tese de doutoramento “*Multivariate methods to assess patterns of stability and change from a biplot perspective*”), da autoria de Susana Mendes recebeu o "*Premio Extraordinario de Doctorado correspondiente al curso 2011/12*". A cerimónia de entrega realizou-se em 8 de Março de 2013 na Universidade de Salamanca, Espanha.

O Prémio é atribuído pela *Comisión de Doctorado y Posgrado* da Universidade de Salamanca e distingue a melhor tese de doutoramento na área da Matemática, no ano a que diz respeito.

FR

• Prémios “Estatístico Júnior 2013”

No final deste Boletim divulgamos os premiados 2013.

Os Prémios serão entregues numa sessão especial organizada durante o XXI Congresso da SPE, que terá lugar no Hotel Meliá Ria em Aveiro, no dia 30 de Novembro.

Brevemente será divulgado o Regulamento do concurso para atribuição de prémios “Estatístico Júnior 2014”.

A Sociedade Portuguesa de Estatística agradece a todos os Professores e alunos que enviaram os seus trabalhos. Aos Premiados apresentamos os nossos parabéns e aos que desta vez não conseguiram ser premiados, queremos agradecer o interesse demonstrado e pedir-lhes que continuem a submeter os vossos trabalhos.

MN

Enigmística de mefqa

M O S T

análise!

No Boletim SPE primavera de 2013 (p.9):

E R A T M A O S G M

Amostragem ao acaso

quadrato carré
cuadrado
pătrat
quadrado

Quadrados latinos

Fundamentos da Inferência Bayesiana

Bento Murteira, *bjf.murteira@sapo.pt*

Professor Jubilado do Instituto Superior de Economia e Gestão, Universidade de Lisboa

1 O Problema Fundamental da Estatística

Antes de abordar os alicerces da inferência bayesiana parece conveniente fazer referência ao problema fundamental da estatística. Para O’Hagan: “*The fundamental problem towards which the study of statistics is addressed is that of inference. Some data are observed and we wish to make statements, inferences, about one or more unknown features of the physical system which gave rise to these data*”. Ao aprofundar o estudo dos fundamentos da estatística depara-se com grande número de correntes ou escolas. Sem falar nos chamados clássicos o desfile é extenso: bayesiano¹ (objectivos, subjectivos², . . .), estruturalistas, fiducialistas, verosimilancistas, . . .

A diversidade não é inesperada! As informações dos dados sobre parâmetros ou modelos enquadram-se, em geral, na indução que é um dos problemas mais controversos da filosofia. Cada escola tem princípios e procedimentos próprios cuja análise conduz aos fundamentos da inferência estatística conforme Berger descreve: “*Statistics needs a: ‘foundation’, by which I mean a framework of analysis within which any statistical investigation can theoretically be planned, performed, and meaningfully evaluated. The words ‘any’ and ‘theoretically’ are key, in that the framework should apply to any situation but may only theoretically be implementable. Practical difficulties or time limitations may prevent complete (or even partial) utilisation of such framework, but the direction in which ‘truth’ could be found would at least be known*”. Os fundamentos da inferência bayesiana são melhor compreendidos quando introduzidos em confronto com os da principal “concorrente”, a inferência clássica.

¹Até há meios-bayesianos que consideram relevante utilizar a informação *a priori* mas entendem que para a combinar com os dados a linguagem das probabilidades é inadequada e deve ser substituída por uma que atenda às relações causais.

²A que se dedica essencialmente o presente resumo.

2 O paradigma clássico

A inferência clássica procura determinar que generalizações sobre a população podem fazer-se a partir da amostra. Representando os dados estatísticos por x [ou $x = (x_1, x_2, \dots, x_n)$ onde n é a dimensão da amostra], o conjunto \mathcal{X} de amostras possíveis x designa-se por espaço-amostra. Correntemente, $\mathcal{X} \subseteq \mathfrak{R}$ ou $\mathcal{X} \subseteq \mathfrak{R}^n$. Fundamental para inferência clássica é reconhecer a variabilidade que se verifica de amostra para amostra tendo em mente que os dados observados formam apenas um dos muitos – possivelmente infinitos – conjuntos que poderiam ter sido obtidos. A interpretação dos dados depende não apenas da particular amostra observada mas também das hipóteses adoptadas acerca das possíveis amostras alternativas. Em consequência os dados consideram-se como observação de uma variável ou vector aleatório X com função de distribuição F_0 , que não é, como é óbvio, perfeitamente conhecida. No entanto, existe normalmente algum conhecimento inicial (considerações teóricas, evidência experimental, etc.) sobre a natureza do fenómeno aleatório em estudo que leva a propor ou conjecturar uma família de distribuições \mathcal{F} a que pertence F_0 e que se designa por modelo estatístico para X . A proposta de um modelo é conhecida por *especificação* e é uma fase essencial no estabelecimento de inferências. Admitindo que X é uma variável ou vector aleatório contínuo é prática corrente representar as distribuições de \mathcal{F} pelas respectivas funções densidade de probabilidade. Se estas forem rotuladas por um parâmetro θ com domínio no espaço-parâmetro Θ , o modelo estatístico pode escrever-se $\mathcal{F} = \{f(x|\theta), x \in \mathcal{X} : \theta \in \Theta\}$.

Ultrapassada a tarefa de modelação e parametrização, a inferência clássica contém inúmeros procedimentos para extrair da amostra conclusões sobre as características do modelo representativo da população e procura responder a questões como estas: (a) Os dados x suportam ou são compatíveis com a família \mathcal{F} ? (b) Supondo que a especificação está correcta e que os dados emanam de uma das distribuições da família \mathcal{F} , que conclusões podem tirar-se sobre o valor particular do parâmetro θ_0 que indexa a função de distribuição F_0 que descreve “apropriadamente” os fenómenos investigados?

Os procedimentos clássicos – também designados frequentistas – são avaliados à luz do princípio da *amostragem repetida* através do comportamento num número indefinido de hipotéticas repetições efectuadas nas mesmas condições. Uma das faces do princípio é utilização de frequências como medidas de incerteza; a outra face reside na avaliação dos métodos em termos da frequência com que produzem respostas correctas ou bons resultados. Os problemas clássicos de estimação e de ensaio de hipóteses supõem-se conhecidos e, por isso, a presente análise não se alonga.

Antes de terminar interessa referir duas opiniões. Ao fazer a defesa do pensamento frequentista, Tiago de Oliveira diz: “*E aqui está a diferença fundamental, o cerne da controvérsia: bayesianos, fiducialistas, verosimilhancistas, introduzem novos princípios (filosóficos) de inferência . . . A inferência clássica, agarrada à interpretação empirista das probabilidades, apenas tem de procurar critérios mais ajustados pois não perde o sentido do real. Nas outras inferências a escolha é a gosto num largo leque de princípios mas o real escapa-se*”. A posição de Jaynes é contrária: “*The traditional ‘frequentist’ methods which use only sampling distributions are usable and useful in many particularly simple, idealized problems; but they represent the most proscribed special cases of probability theory, because they presuppose conditions (independent repetitions of a ‘random’ experiment but no relevant prior information) that are hardly ever met in real problems. This approach is quite inadequate for the current needs of science*”.

3 O paradigma bayesiano

Para Lindley a substituição do paradigma clássico pelo paradigma bayesiano representa uma verdadeira revolução científica no sentido de Kuhn³. A semente para a abordagem bayesiana a problemas de inferência foi lançada por Richard Price quando em 1763 publicou a obra póstuma do Rev. Thomas Bayes

³Kuhn, T.S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

intitulada “*An Essay Towards Solving a Problem in the Doctrine of Chances*”. A interpretação da probabilidade como grau de credibilidade – fundamental na filosofia bayesiana – tem uma longa história. Se parece ter sido J. Bernoulli, em 1713, na sua obra *Ars Conjectandi*, um dos primeiros autores a definir probabilidade como o grau de confiança na veracidade de uma dada proposição, foi De Morgan, na *Formal Logic*, em 1847, que afirmou: (1) a probabilidade identifica-se com um grau de credibilidade; (2) os graus de credibilidade podem medir-se; (3) os graus de credibilidade podem identificar-se com um certo complexo de sentimentos. A ideia de coerência de um sistema de graus de credibilidade parece dever-se a Ramsey para quem a atitude de um indivíduo ao apostar na veracidade de uma dada proposição está associada com o grau de credibilidade que lhe atribui. Se declara que as vantagens ou possibilidades (*odds*) – a favor da veracidade contra a não veracidade – são $r : s$, então o grau de credibilidade da proposição é, para o indivíduo, $r/(r + s)$. Para Ramsey nenhum conjunto de apostas num grupo de proposições é admitido a um indivíduo coerente se conduzir a um prejuízo certo. O maior expoente do conceito de probabilidade personalista é, contudo, de Finetti. Nas ideias bayesianas e na sua aplicação à estatística tem de citar-se Harold Jeffreys que, reagindo à predominante posição clássica em meados do século, embora solitário e desapoado, conseguiu ressuscitar o bayesianismo, dar-lhe *status* lógico e avançar com soluções de problemas estatísticos que naquele tempo persistiam. A partir daí a lista de bayesianos foi engrossando sucessivamente e, na impossibilidade de citar todos, merecem realce os nomes de Good, Savage e Lindley.

O Teorema de Bayes, bem conhecido, é uma proposição sobre probabilidades condicionadas indiscutível decorrente do cálculo de probabilidades ou da axiomática de Kolmogorov. O aspecto controverso é a sua aplicação a problemas de inferência estatística. Ocupa, como é óbvio, lugar fulcral na inferência bayesiana que tem relativamente à inferência clássica uma divergência fundamental. No modelo clássico o parâmetro $\theta, \theta \in \Theta$, é um escalar ou vector desconhecido, mas fixo, *i.e.*, igual ao valor particular que indexa a distribuição da família \mathcal{F} que descreve “apropriadamente” o processo ou sistema físico que gera as observações. No modelo bayesiano o parâmetro $\theta, \theta \in \Theta$, é tomado como um escalar ou vector aleatório não observável. Segundo os bayesianos o que é desconhecido – no caso em questão, o parâmetro θ – é incerto e toda a incerteza deve ser quantificada em termos de probabilidade. Correlativamente, os bayesianos defendem que a informação inicial ou *a priori* – anterior ou externa em relação à experiência mas demasiado importante para ser ignorada pois pode até suceder que os investigadores sejam peritos na matéria – deve traduzir-se por uma distribuição de probabilidade para θ , seja $h(\theta)$, designada distribuição *a priori*. A determinação e a interpretação da distribuição *a priori* são dos pontos mais controversos da teoria bayesiana.

A família \mathcal{F} também faz parte do modelo bayesiano; quer dizer, a componente amostral ou experimental é comum aos modelos clássico e bayesiano, embora para este os elementos $f(x|\theta)$ de \mathcal{F} em geral são supostos ter, tal como $h(\theta)$, uma interpretação subjectiva.

A discussão das distribuições *a priori* ilustra alguns aspectos do confronto entre bayesianos e clássicos. Para os primeiros, Berger por exemplo, a especificação – escolha subjectiva da família \mathcal{F} – traduz muitas vezes um uso mais drástico da informação *a priori* do que o emprego de uma distribuição *a priori*. E acrescentam: na sua modelação os clássicos atendem à informação *a priori* quando muito informalmente, atitude que consideram algo limitada porquanto, para eles, a informação inicial ou *a priori* detida por um dado investigador deve traduzir-se formalmente por uma distribuição de probabilidade para a variável aleatória θ . Para os segundos, por exemplo Lehmann, há uma importante diferença entre a modelação de \mathcal{F} e a modelação de $h(\theta)$ pois no primeiro caso dispõe-se de um conjunto de observações – $x = (x_1, x_2, \dots, x_n)$ – geradas por um membro de \mathcal{F} que pode empregar-se para testar a forma da distribuição.

Para entender o ponto de vista bayesiano repare-se que um clássico *em todos* os problemas que envolvem uma variável X com distribuição binomial recorre sempre ao modelo bernoulliano em que o parâmetro θ representa a probabilidade de um “sucesso”. Para os bayesianos cada problema é *único* e tem um contexto real próprio onde θ é uma quantidade significativa acerca da qual existem, em geral, graus de conhecimento que variam de problema para problema e de investigador para investigador. Assim, a distribuição de probabilidade que capta essa variabilidade é baseada na informação *a priori* e é específica

de um dado problema e de um dado investigador. De facto, sublinham, a informação *a priori* inclui juízos ou experiências individuais da mais diversa índole, decorrentes em geral de situações não repetitivas, formalizáveis apenas em termos personalistas. No entanto, advertem, esta formalização exige que o investigador satisfaça condições de coerência ou de consistência que permitam o recurso às regras de cálculo de probabilidades estabelecidas. Assim, diferentes investigadores possuem em geral diferentes distribuições *a priori* para o mesmo parâmetro sem deixarem necessariamente de ser coerentes.

Suponha-se que se observa $X = x$; dado um qualquer $f(x|\theta) \in \mathcal{F}$ e a distribuição *a priori* $h(\theta)$, o Teorema de Bayes conduz à relação⁴,

$$h(\theta|x) = \frac{f(x|\theta)h(\theta)}{\int_{\Theta} f(x|\theta)h(\theta)d\theta}, \quad \theta \in \Theta, \quad (1)$$

onde $h(\theta|x)$ é a distribuição *a posteriori* de θ depois de conhecido $X = x$. Assim, a atitude inicial do investigador, caracterizada por $h(\theta)$, é modificada pela informação recolhida passando a traduzir-se por $h(\theta|x)$. O denominador de (1), que se representa por $f(x)$, é a distribuição marginal de X ; diz respeito à observação de X qualquer que seja θ .

O conceito de função de verosimilhança estuda-se no quadro da inferência clássica mas não é menos importante no quadro bayesiano. Na respectiva definição convém manter a distinção entre os casos discreto e contínuo [veja-se Kempthorne e Folks (1971)] mas chega-se em ambos os casos à função de θ ,

$$L(\theta|x) = Kf(x|\theta), \quad \theta \in \Theta \quad \text{ou} \quad L(\theta|x_1, \dots, x_n) = K \prod_i f(x_i|\theta), \quad \theta \in \Theta, \quad (2)$$

que para cada $\theta \in \Theta$ exprime a verosimilhança ou plausibilidade que lhe é atribuída quando se observa $X = x$ ou $(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$. A função de verosimilhança – não é uma probabilidade pois, por exemplo, não faz sentido adicionar verosimilhanças – tem importante papel na fórmula de Bayes pois representa o meio através do qual os dados, x , transformam o conhecimento *a priori* sobre θ ; quer dizer, a verosimilhança pode interpretar-se como expressão da informação sobre θ fornecida pelos dados x .

Em resumo, para os bayesianos a distribuição *a posteriori* incorpora, por via do Teorema de Bayes, toda a informação disponível sobre o parâmetro,

informação inicial + informação da experiência ou da amostra.

Daqui decorre que todos os procedimentos de inferência bayesiana são baseados em $h(\theta|x)$ [ou $h(\theta|x_1, x_2, \dots, x_n)$].

No caso em que θ representa um vector de parâmetros, *e.g.*, $\theta = (\gamma, \phi) \in \Gamma \times \Phi$, pode acontecer que o interesse inferencial se restrinja apenas a uma parte de θ , digamos γ . Nessa eventualidade, a eliminação do parâmetro perturbador ϕ no paradigma bayesiano obedece sempre à mesma via, contrariamente ao que se verifica no paradigma clássico, traduzida na marginalização da distribuição *a posteriori* conjunta, *i.e.*, na determinação de

$$h(\gamma|x) = \int_{\Phi} h(\gamma, \phi|x)d\phi = \int_{\Phi} h(\gamma|\phi, x)h(\phi|x)d\phi. \quad (3)$$

A possível dificuldade de integração analítica desaparece sempre que γ e ϕ são aprioristicamente independentes e a função de verosimilhança se factoriza em $L(\theta|x) = L_1(\gamma|x) \times L_2(\phi|x)$, resultando que $h(\gamma|x) \propto h(\gamma)L_1(\gamma|x)$.

4 Inferência bayesiana

Nos procedimentos bayesianos podem distinguir-se dois objectivos: **I** – Traçar inferências sobre o parâmetro não observável θ ; **II** – Realizar inferências sobre variáveis ainda não observadas (predição).

⁴Facilmente adaptável se x for um vector ou se o espaço-parâmetro for discreto.

4.1 Inferências paramétricas

Na óptica da inferência paramétrica há uma certa coincidência – pelo menos superficial – entre os objetivos dos clássicos e dos bayesianos mas na sua implementação as duas correntes entram em choque. Por um lado, as inferências clássicas são baseadas em probabilidades associadas com as diferentes amostras, x , que poderiam ocorrer para algum valor fixo, mas desconhecido, do parâmetro θ . Isto é, as inferências têm por base as distribuições amostrais que “ponderam”, probabilisticamente, os valores que a variável X ou a estatística $T(X)$ podem assumir percorrendo o espaço-amostra. Por outro lado, as inferências bayesianas são baseadas em probabilidades subjectivas ou credibilidades *a posteriori* associadas com diferentes valores do parâmetro θ e condicionadas pelo particular valor de x observado. O ponto x está fixo e determinado e é a variação de θ que é considerada.

Por exemplo, os bayesianos uma vez observado x e considerando a hipótese de ser $\{\theta \leq 0.5\}$, respondem à questão de forma significativa e directa calculando $P(\theta \leq 0.5|x)$ a partir de $h(\theta|x)$, *i.e.*, sem sair do cálculo de probabilidades. Em contraste, os clássicos não respondem directamente à questão e ao afirmarem, por exemplo, que a hipótese $H_0 : \theta \leq 0.5$ é rejeitada ao nível de 5% não querem afirmar que a sua probabilidade é inferior a 0.05, mas que se a hipótese H_0 for verdadeira (*i.e.*, se de facto $\theta \leq 0.5$), então a probabilidade de X pertencer a uma dada região crítica W a determinar é tal que $P(X \in W|\theta \leq 0.5) < 0.05$, e que se de facto $x \in W$, então tal hipótese deve ser rejeitada.

No dizer crítico de O’Hagan, enquanto os bayesianos podem emitir enunciados probabilísticos sobre os parâmetros, que consideram como variáveis aleatórias, isso não é possível com os clássicos cujas probabilidades dizem respeito aos dados e não ao parâmetro, embora depois reformuladas para que aparentemente digam respeito ao parâmetro. Esta questão tem correspondência na diferente atitude em relação ao espaço-amostra. Para os clássicos o conceito é fundamental pois a amostragem repetida consiste em “percorrer” o espaço-amostra. Os bayesianos começam por criticar a ideia de que é pacífico fazer repetições com n fixo; depois defendem que só interessa o resultado obtido, x , e não o conjunto ou espaço-amostra a que pertence x que pode ser absolutamente arbitrário e que contém, além de x , observações que poderiam ter sido obtidas mas que não o foram⁵.

Nos problema de estimação os clássicos consideram diferentes alternativas ou funções dos dados – estimadores – cujas propriedades amostrais investigam sob diversas ópticas (consistência, não enviesamento, etc.). Para os bayesianos há apenas *um* estimador que é precisamente a distribuição *a posteriori* $h(\theta|x)$. Pode, é claro, descrever-se esta distribuição através da moda, média, mediana ou da variância, mas isso nada tem a ver com o problema que enfrentam os clássicos quando pretendem determinar o chamado *estimador óptimo*. Para os bayesianos tal problema só existe no quadro da Teoria da Decisão, campo em que os bayesianos têm nítida vantagem sobre os clássicos. Em consonância, Savage sustenta que nas últimas décadas o problema central em face da incerteza deixou de ser *que inferências podem realizar-se* e passou a ser *que fazer ou que decisão tomar*. Como a decisão individual tem sido considerada ultrapassada por alguns filósofos, fala-se ultimamente no reforço do bayesianismo através da decisão em grupo.

Aos intervalos de confiança os bayesianos contrapõem os intervalos (ou regiões) de credibilidade. Observado x e determinada a distribuição *a posteriori*, um intervalo de credibilidade para o parâmetro θ (suposto aqui um escalar) é formado por um par de valores de Θ , sejam $[\underline{\theta}(x), \bar{\theta}(x)]$, ou mais simplesmente, $(\underline{\theta}, \bar{\theta})$, tais que,

$$P(\underline{\theta} < \theta < \bar{\theta}|x) = \int_{\underline{\theta}}^{\bar{\theta}} h(\theta|x)d\theta = 1 - \alpha, \quad (4)$$

onde $1 - \alpha$ (em geral, 0.90, 0.95 ou 0.99) é o nível de credibilidade desejado. Se $\Theta = (-\infty, +\infty)$ uma forma expedita de construir um intervalo de credibilidade (dito então central) é considerar na distribuição

⁵A relutância em relação ao espaço-amostra traduz o mesmo sentimento acerca das regras de paragem na experimentação, tema que Mayo e Kruse (2001), recordando Armitage, dizem poder levantar aos bayesianos alguns problemas.

a posteriori abas de igual credibilidade verificando

$$\int_{-\infty}^{\underline{\theta}} h(\theta|x)d\theta = \int_{\bar{\theta}}^{+\infty} h(\theta|x)d\theta = \frac{\alpha}{2}. \quad (5)$$

A definição (3) possui um inconveniente: o intervalo $(\underline{\theta}, \bar{\theta})$ não é único, podendo suceder que valores de θ contidos nesse intervalo tenham menor credibilidade que valores de θ não incluídos no mesmo intervalo. Assim, para proceder à escolha de um certo intervalo ao mesmo tempo que se minimiza a respectiva amplitude, os bayesianos preferem trabalhar com intervalos de credibilidade HPD (*highest posteriori density*) $(\theta', \theta'') = \{\theta : h(\theta|x_1, x_2, \dots, x_n) \geq k(\alpha)\}$, onde $k(\alpha)$ é o maior número real tal que $P(\theta' < \theta < \theta'') = 1 - \alpha$.

Os intervalos de credibilidade têm uma interpretação directa em termos de probabilidade. O mesmo não se passa com os intervalos de confiança em que se parte de uma probabilidade que não diz respeito a θ , mas sim aos dados, mais precisamente a um intervalo aleatório definido a partir da amostra genérica e que depois de observar uma amostra concreta se reconverte na *confiança* de cobertura do valor desconhecido de θ pelo intervalo numérico resultante, o qual não pode interpretar-se em geral como uma probabilidade ou credibilidade referente a θ . Além de outros aspectos críticos que a teoria dos intervalos (ou regiões) de confiança suscita, conhecem-se os irónicos comentários de Lindley (1990) quando diz conhecer várias axiomáticas da teoria da probabilidade – por exemplo, as devidas a Savage, de Finetti ou Kolmogorov – mas não conhece nenhuma axiomática da *confiança*.

Os bayesianos quando, por exemplo, pretendem ensaiar uma hipótese composta $H_0 : \theta \in \Theta_0$ versus uma alternativa também composta $H_1 : \theta \in \Theta_1$, com $\Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \cup \Theta_1 = \Theta$, chegam a expressões em termos de probabilidades sobre Θ . Se o investigador tem uma distribuição, $h(\theta)$, $\theta \in \Theta$, traduzindo a credibilidade inicial que atribui aos diferentes valores do parâmetro, pode determinar as probabilidades *a priori* das hipóteses em confronto,

$$P(\Theta_0) = \int_{\Theta_0} h(\theta)d\theta, \quad P(\Theta_1) = \int_{\Theta_1} h(\theta)d\theta.$$

Ao quociente $P(\Theta_0)/P(\Theta_1)$ costuma chamar-se rácio das vantagens ou razão das chances (*odds*) *a priori* de H_0 sobre H_1 . Depois de realizar a experiência, que se supõe consistir na observação de x , e de determinar $h(\theta|x)$, o estatístico bayesiano calcula as respectivas probabilidades *a posteriori*,

$$P(\Theta_0|x) = \int_{\Theta_0} h(\theta|x)d\theta, \quad P(\Theta_1|x) = \int_{\Theta_1} h(\theta|x)d\theta,$$

e também, usualmente, o rácio das vantagens *a posteriori* de H_0 sobre H_1 , seja, $P(\Theta_0|x)/P(\Theta_1|x)$. Pode talvez dizer-se que no quadro bayesiano o resultado da inferência não é tanto a aceitação ou rejeição da hipótese H_0 – como sucede na doutrina de Neyman-Pearson – mas mais a alteração das credibilidades atribuídas à hipótese e à alternativa. A análise bayesiana passa muitas vezes pela comparação do rácio das vantagens *a posteriori* com o das vantagens *a priori*,

$$B(x) = \frac{P(\Theta_0|x)/P(\Theta_1|x)}{P(\Theta_0)/P(\Theta_1)}, \quad (6)$$

que se designa por factor Bayes a favor de H_0 (ou Θ_0) e que traduz o pendor dos dados x para suportar H_0 . Evidentemente, quanto maior for o factor Bayes maior é o aumento das vantagens *a posteriori* em relação às vantagens *a priori* e maior é, portanto, o suporte que os dados dão à hipótese H_0 . O factor Bayes depende, em geral, da distribuição *a priori* e pode expressar-se como um rácio de verosimilhanças ponderadas pelas densidades *a priori* condicionais a cada hipótese sobre Θ_0 e Θ_1 (*vide* Paulino *et al.*, 2003). Neste sentido não pode dizer-se que o factor Bayes seja uma medida de suporte da hipótese H_0 baseada apenas nos dados.

Quando a hipótese conjecturada sobre θ é incisiva ao ponto de se definir como $H_0 : \theta = \theta_0$, a via do factor Bayes ou das chances *a posteriori* exige que a distribuição *a priori* seja condizente com tal conjectura em ordem a evitar uma probabilidade nula, revestindo em geral uma natureza mista. Esta implicação é tomada como natural por bayesianos como Jeffreys, com o argumento de que a distribuição *a priori* deve integrar os juízos probabilísticos inerentes à própria formulação das hipóteses em confronto que, no caso vertente, atribuem uma ordem de importância a θ_0 diferenciada da dos outros valores de θ .

Outros bayesianos como Lindley e Zellner advogaram um procedimento distinto, com uma certa analogia formal com os testes de significância clássicos, na situação em que a própria formulação de hipóteses pontuais não interfere com a distribuição *a priori*. Esse procedimento pode ser mais geralmente descrito pela quantificação da plausibilidade relativa *a posteriori* do valor θ_0 através do cálculo da medida $P = P(\theta \notin R_0(x)|x)$, onde $R_0(x) = \{\theta \in \Theta : h(\theta|x) \geq h(\theta_0|x)\}$ é a menor região de credibilidade HPD que contém θ_0 . Valores grandes (pequenos) do nível P de plausibilidade relativa *a posteriori* de H_0 apontam para uma evidência a favor de (contra) essa hipótese.

O instrumento fundamental da abordagem bayesiana e a forma de usar o modelo estatístico conjunto $M = \{f(x|\theta)h(\theta), x \in X, \theta \in \Theta\}$ na realização de inferências deixam antever que a questão de avaliar a razoabilidade do modelo conjecturado em termos absolutos não tem uma resposta popperiana (refutação/não refutação) do género daquela garantida pelos testes de ajustamento da metodologia clássica.

A via dos factores Bayes poderá ser utilizada se for possível estender o modelo M (ou parte dele) a uma família mais vasta que se admite integrar o verdadeiro modelo tido como desconhecido, o que permitirá comparar os modelos dentro dela. De outro modo, poderá recorrer-se a várias medidas de adequabilidade para uma análise relativa do desempenho do modelo de partida no quadro de uma classe a definir de modelos competidores (*vide* Paulino *et al.*, 2003). A relativa insatisfação com estas opções tem levado alguns estatísticos a defender a aplicação da abordagem bayesiana apenas quando o modelo de partida é considerado como não questionável, condição esta que Gillies (2001) apelida de *fixity of the theoretical framework*.

4.2 Inferências preditivas

Muitos bayesianos consideram que a inferência não tem de se restringir a proposições sobre parâmetros não observáveis. Afirmam, conseqüentemente, que as inferências paramétricas possuem inconvenientes na medida em que os valores dos parâmetros poucas vezes são conhecidos e portanto as conclusões a que tais inferências conduzem raramente podem ser confrontadas com a realidade. Para bayesianos, como Lindley, é mais fundamental o problema que consiste em partir de um conjunto de observações (x_1, x_2, \dots, x_n) (ontem) e inferir conclusões, em termos de probabilidade (subjectiva, claro), sobre o conjunto de variáveis ainda não observadas, $(x_{n+1}, x_{n+2}, \dots, x_{n+M})$ (amanhã).

Para facilitar aqui a exposição faz-se $M = 1$ e consideram-se as $n + 1$ variáveis aleatórias $X_1, X_2, \dots, X_n, X_{n+1}$ i.i.d. dado θ com função densidade $f(x|\theta)$ e o problema consiste em prever o comportamento da variável aleatória X_{n+1} depois de observar $(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$. Ao tentar prever a variável X_{n+1} com densidade amostral $f(x|\theta)$ encontram-se dois tipos de aleatoriedade: (a) o que se prende com o facto de a própria variável ser aleatória; (b) o derivado do desconhecimento do valor de θ . Por exemplo, quando se procede à estimação de θ e se obtém $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ pelo método da máxima verosimilhança, e se escreve, $P(a < X_{n+1} < b|x_1, x_2, \dots, x_n) \cong \int_a^b f(x|\hat{\theta})dx$, para estimar a probabilidade do acontecimento $a < X_{n+1} < b$ está-se a ignorar a aleatoriedade que envolve a substituição do parâmetro pela estimativa, porquanto ambos os tipos de aleatoriedade devem influenciar o processo preditivo. Este procedimento de enfiar uma estimativa no lugar do parâmetro na distribuição amostral (*plug-in procedure*) merece pois grandes reservas.

Embora a solução clássica do problema da predição tenha muito mais que se lhe diga [veja-se Amaral Turkman (1980)], parece poder afirmar-se que os bayesianos têm uma solução bem mais lúcida. Se dispõem apenas da informação inicial, traduzida pela distribuição *a priori* $h(\theta)$, o instrumento a aplicar é

naturalmente a já referida distribuição marginal ou distribuição preditiva *a priori* $f(x)$. Caso mais interessante é aquele em que se observa $x = (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ e se pretende prever X_{n+1} independente das observações anteriores dado θ [ou prever $(X_{n+1}, X_{n+2}, \dots, X_{n+M})$ – o problema não é muito diferente]. Utilizando raciocínio completamente probabilístico tem-se $f(x_{n+1}|x) = \int_{\Theta} f(x_{n+1}|\theta)h(\theta|x)d\theta$, onde a distribuição *a posteriori* ocupa o lugar da distribuição *a priori* por haver a informação dada pela amostra. Podem determinar-se valores típicos dessa distribuição preditiva, probabilidades de qualquer região no espaço de valores de X_{n+1} ou então valores, $a = a(x)$, $b = b(x)$, para um valor prefixado da probabilidade, $P(a < X_{n+1} < b|x) = \int_a^b f(x_{n+1}|x)dx_{n+1}$, que definem um intervalo de predição (com a característica HPD ou não).

5 Conclusão

Em resumo, segundo a perspectiva bayesiana:

- No esquema clássico a inferência estatística passa por raciocínios do tipo indutivo, como por exemplo os intervalos de confiança, que não têm interpretação directa em termos de probabilidade. A dificuldade ou impossibilidade de fazer inferências com interpretação directa em termos de probabilidade – já que o parâmetro θ não é considerado uma quantidade aleatória – é duramente criticada por Jaynes.
- No esquema bayesiano verifica-se que todas as inferências são realizadas a partir da aplicação lógica do cálculo de probabilidades. A inferência estatística bayesiana, insista-se, não recorre a resultados que não possam deduzir-se a partir das regras de cálculo de probabilidades e em particular do Teorema de Bayes. Como afirma O’Hagan: “*Probability theory is a completely self-consistent system. Any question of probabilities has one and only one answer, although there may be many ways to derive it*”.

Num plano de referência a posições não extremistas convém recordar declarações como a de Dawid (1985) que, embora confessando nítida preferência pela teoria bayesiana, declara *I believe – and believe I have proved – that no approach to statistical inference, Bayesian or not, can ever be entirely satisfactory*. O que alguns estatísticos hoje defendem não é a opção *unicamente Bayes* de autores como Savage mas a posição eclética partilhada por Wasserman (2004) ao afirmar, em síntese, que para combinar credibilidades *a priori* com dados estão indicados os métodos bayesianos, e para construir procedimentos que garantam bons resultados com a longa repetição das observações há que recorrer a métodos frequentistas.

Bibliografia

- Amaral Turkman, M.A. (1980). *Applications of Predictive Distributions*. PhD. Thesis, University of Sheffield.
- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Inference*. Springer, Berlim.
- Dawid, A.P.(1985). The impossibility of inductive inference. (Invited discussion of ‘Self-calibrating priors do not exist’, by D. Oakes.) *J. Amer. Statist. Ass.* 80, 340-341.
- Gillies, D. (2001). Bayesianism and the Fixity of the Theoretical Framework. In *Foundations of Bayesianism*, J. Corfield and J. Williamson (eds.), pp. 363–379, Kluwer Academic Publishers, Dordrecht.
- Jaynes, E.T. (2003). *Probability Theory*. Cambridge University Press.

- Kempthorn, O., Folks, L. (1971). *Probability, Statistics and Data Analysis*. The Iowa State University Press, Iowa.
- Lehmann, E.L. (1983). *The Theory of Point Estimation*. Wiley, New York.
- Lindley, D.V. (1990). The 1988 Wald Memorial Lectures, The Present Position in Bayesian Statistics. *Statistical Science*, 5, 44–89.
- Mayo, D., Kruse, M. (2001). Principles of Inference and Their Consequences. In *Foundations of Bayesianism*, J. Corfield e J. Williamson (eds.), pp. 381–403, Kluwer Academic Publishers, Dordrecht.
- O’Hagan, A. (1994). *Bayesian Inference, Kendall’s Advanced Theory of Statistics*. Volume 2B, Arnold, London.
- Tiago de Oliveira, J. (1981). Resumo das intervenções e comentários. In *Actas do II Colóquio de Estatística e Investigação Operacional*, SPEIO, Fundação.
- Wasserman, L. (2004). *All of Statistics*. Springer, New York.



Representação da Informação *a priori*

Carlos Daniel Paulino¹, dpaulino@math.ist.utl.pt
Maria Antónia Amaral Turkman², maturkman@fc.ul.pt

¹*Instituto Superior Técnico, Universidade de Lisboa e CEAUL*

²*CEAUL e Faculdade de Ciências, Universidade de Lisboa*

1 Introdução

A colocação em funcionamento da máquina inferencial exige que o insumo do seu instrumento fundamental esteja devidamente preparado. O 1º ingrediente é o modelo amostral que se supõe acomodar (com maior ou menor incerteza) os dados entretanto obtidos de algum processo experimental ou observacional que interessa analisar. Este modelo abrange um conjunto de aspetos desconhecidos sobre os quais pode haver informação apriorística que interessa incluir na análise, por mais vaga ou significativa que ela seja, o que exige saber como se deve representar ou quantificar.

O processo de representação da informação *a priori*, quantas vezes atribulado pelo envolvimento de elementos de natureza subjetiva a eliciar, vai aqui ser abordado por confinamento a duas situações:

- a primeira é aquela em que não existe informação *a priori* palpável, de natureza quer objetiva quer subjetiva (o frequentemente chamado estado de “ignorância *a priori*” ou em que o conhecimento *a priori* é pouco significativo relativamente à informação amostral (o estado de conhecimento “vago” ou “difuso”). Focar-se-ão alguns dos principais métodos advogados para o efeito conduzindo a distribuições *a priori* minimamente informativas em algum sentido e que, em regra, são apelidadas de **distribuições não-informativas**;
- A segunda está muito ligada à adoção prévia de uma forma distribucional conveniente e à escolha de um membro dessa família que se revele consentâneo com medidas-resumo cuidadosamente eliciadas para a distribuição a determinar - como um exemplo interessante deste processo eliciatório veja-se o problema médico tratado em Paulino *et al.* (2003). É neste âmbito que se inserem as denominadas **distribuições conjugadas naturais**, as quais podem igualmente servir como geradoras de distribuições não-informativas impróprias.

Para informação adicional sobre o problema de escolha da distribuição *a priori* (outros métodos geradores de distribuições vagas ou de eliciação de distribuições subjetivas, *vide* O’Hagan (2010), Kass e Wasserman (1996) e Paulino *et al.*(2003).

2 Distribuições não-informativas

Estas distribuições começaram por ser predominantemente interpretadas como representações formais de ignorância, mas há hoje uma tendência (motivada pela não aceitação de representações objetivas únicas da ignorância) para encará-las como opções convencionais de defeito a que se recorre em caso de informação *a priori* insuficiente que torne difícil eliciar uma distribuição subjetiva considerada adequada. Independentemente da interpretação, este tipo de distribuições pode desempenhar ainda um papel de referência, mesmo que se disponha de fortes crenças *a priori*, como forma de:

- deduzir as crenças *a posteriori* para quem parte de um conhecimento escasso (*i.e.*, quando a amostra fornece o grosso da informação sobre o parâmetro) e, nessa medida, se acha incapaz de determinar subjetivamente uma distribuição razoável – assim se reconhecendo a sua própria ignorância;
- permitir a comparação com os resultados da inferência clássica que “só” usa a informação amostral (no todo ou em parte);
- averiguar a influência nas inferências da distribuição *a priori* subjetiva que descreve a informação realmente existente, quando confrontada com as que resultam do uso da distribuição *a priori* de referência.

Descrevem-se sucintamente em seguida alguns dos argumentos mais usados que conduzem a este tipo de distribuições.

Método de Bayes-Laplace

Este método baseia-se na invocação do Princípio da Razão Insuficiente decorrente da escassez informativa *a priori* para adotar a ideia de equiprobabilidade. Dependendo da cardinalidade de Θ este argumento conduz às distribuições Uniforme discreta ou Uniforme contínua.

Se no caso de o número de valores de θ ser finito, e.g. $\Theta = \{\theta_1, \dots, \theta_k\}$, este argumento pode ser considerado pacífico ao levar à distribuição $h(\theta) = 1/k$, $\theta \in \Theta$, o mesmo não sucede nas outras situações. Com efeito, se Θ for infinito numerável, a distribuição decorrente é imprópria, o mesmo sucedendo no caso de Θ ser um conjunto infinito não numerável e não limitado, o que causa algum desconforto a estatísticos que não simpatizam com medidas não normalizadas (ainda que tal não ponha necessariamente em causa o seu uso no teorema de Bayes, pois a distribuição *a posteriori* – o fulcro das inferências – é muitas vezes própria).

Uma outra crítica, talvez mais séria, ao uso do argumento de que a ausência de informação, que alguns rotulam de ignorância, deve ser representada por uma distribuição uniforme, resulta do facto de esta não ser invariante relativamente a transformações não lineares, conduzindo assim a contrassensos probabilísticos. Como ilustração, tome-se o modelo $\{Ber(\theta), \theta \in (0, 1)\}$ que faz parte da família exponencial com parâmetro natural $\psi = \ln[\theta/(1-\theta)] \in \mathbb{R}$. O uso simultâneo de distribuições uniformes para θ (própria) e ψ (imprópria) é probabilisticamente inconsistente já que $\theta \sim U(]0, 1[) \equiv Be(1, 1)$ equivale para ψ a uma distribuição logística reduzida, de função densidade $h(\psi) = \frac{e^\psi}{(1+e^\psi)^2}$, $\psi \in \mathbb{R}$.

No caso geral, sendo $\psi = \psi(\theta)$ uma transformação injetiva de um parâmetro θ , que assume uma gama contínua de valores possíveis, e $h(\theta)$ uma densidade *a priori* para θ , então

$$h(\psi) = h[\theta(\psi)] \left| \frac{d\theta}{d\psi} \right| \quad (1)$$

deve ser a correspondente densidade para a reparametrização ψ , que não é uniforme quando $h(\theta)$ o é se o jacobiano depende de ψ , como acontece com transformações não lineares do género da referida no exemplo acima.

Método de Jeffreys

Entre os procedimentos que asseguram invariância sob transformações injetivas está aquele advogado por Jeffreys e que se baseia no uso da medida de informação de Fisher sobre $\theta \in \mathbb{R}$, definida por

$$I(\theta) = E \left[\left(\frac{\partial \ln f(X | \theta)}{\partial \theta} \right)^2 \middle| \theta \right].$$

Com efeito, o facto de para qualquer transformação real injetiva de $\theta \in \mathbb{R}$ se ter

$$I(\psi) = I(\theta(\psi)) \left(\frac{d\theta}{d\psi} \right)^2$$

mostra que a distribuição proposta por Jeffreys para o caso uniparamétrico, $h(\theta) \propto [I(\theta)]^{\frac{1}{2}}$, goza da referida propriedade de invariância e, deste modo, assegura a identidade das inferências qualquer que seja a transformação biunívoca no espaço paramétrico (ou seja, reparametrização) que se use.

Além disso, deve notar-se que $I(\theta)$ é tanto maior quanto maior for a taxa quadrática de variação (em média ao longo do espaço amostral) com θ de $\ln f(X | \theta)$, i.e., quanto mais diferenciado pelo modelo estiver θ de $\theta + d\theta$. Considerar então mais (menos) plausíveis *a priori* os valores de θ com maior (menor) $I(\theta)$, i.e., sobre os quais há maior (menor) informação amostral, corresponde a considerar reduzido tanto quanto possível o efeito da informação *a priori*. Daí o carácter não informativo das distribuições *a priori* geradas pela regra de Jeffreys. O seu carácter alegadamente objetivo advém de serem automaticamente obtidas do próprio modelo supostamente gerador dos dados.

Exemplo 1: Considere-se um modelo amostral cuja função de verosimilhança se pode escrever como

$$L(\theta|x, n) = k\theta^x(1 - \theta)^{n-x}, \quad \theta \in (0, 1),$$

onde k não depende de θ . Se n é fixo e $k = \binom{n}{x}$ o modelo corresponde a $X|n, \theta \sim Bi(n, \theta)$, cujo valor médio é $n\theta$ donde $I(\theta) \propto \theta^{-1}(1 - \theta)^{-1}$, pelo que a distribuição de Jeffreys é própria e definida por

$$h(\theta) \propto \theta^{-1/2}(1 - \theta)^{-1/2}, \quad \theta \in (0, 1) \Leftrightarrow \theta \sim Be(1/2, 1/2),$$

correspondendo por aplicação do argumento de transformação de variáveis à distribuição $U(0, 2\pi)$ para $\psi = \arcsin\sqrt{\theta}$.

Se alternativamente x é fixo e $k = \binom{n-1}{x-1}$ está-se na presença do modelo $N - x|x, \theta \sim BiN(x, \theta)$, de valor médio $x(1 - \theta)/\theta$ donde $I(\theta) \propto \theta^{-2}(1 - \theta)^{-1}$, implicando que a distribuição de Jeffreys é tal que

$$h(\theta) \propto \theta^{-1}(1 - \theta)^{-1/2}, \quad \theta \in (0, 1),$$

correspondendo a uma distribuição imprópria que se simboliza por “ $Be(0, 1/2)$ ” e que é consistente com uma distribuição “Uniforme” para a função $\psi = \ln \frac{1 - \sqrt{1 - \theta}}{1 + \sqrt{1 - \theta}}$. ■

Este exemplo evidencia o que era expectável, que a distribuição de Jeffreys pela sua definição depende plenamente da distribuição amostral e não apenas do seu núcleo. Esta dependência para com o espaço amostral é para uns motivo de veemente crítica, na base nomeadamente de conduzir a inferências *a posteriori* sobre o mesmo parâmetro distintas consoante a natureza da experiência efetivamente realizada (no exemplo, amostragem binomial direta ou binomial inversa), ainda que a diferença possa ser bem ligeira em amostras moderadas, como no exemplo ilustrativo em questão. Para outros, tal dependência é considerada legítima pelo argumento de que a vaguidade da informação *a priori*, que a distribuição de Jeffreys visa representar, deve ser vista em função da informação associada ao tipo de amostragem planeada e não em termos absolutos.

A aplicação da regra de Jeffreys a modelos uniparamétricos de localização, $\{f(x|\theta) = g(x - \theta), \theta \in \Theta \subseteq \mathbb{R}\}$ (e.g., modelo Normal com variância conhecida), conduz à distribuição Uniforme contínua que é invariante sob transformações lineares (i.e., face a translações) e imprópria se Θ for um conjunto não limitado. Se for aplicada a modelos uniparamétricos de escala, $\{f(x|\theta) = \frac{1}{\theta}g(x/\theta), \theta \in \Theta \subseteq \mathbb{R}_+\}$ (e.g., modelo Normal com média conhecida), conduz à distribuição imprópria $h(\theta) \propto \theta^{-1}I_{(0, +\infty)}(\theta)$, que é invariante sob potências (i.e., face a transformações de escala).

Em modelos multiparamétricos a regra de Jeffreys é baseada na raiz quadrada do determinante da matriz de informação de Fisher. Contudo, devido a implicações *a posteriori* indesejáveis, esta regra costuma ser preterida, até por sugestão do próprio Jeffreys, pela imposição prévia de independência *a priori* entre parâmetros (particularmente quando são de natureza diferente) e pelo uso das regras de Jeffreys uniparamétricas para a especificação das distribuições marginais. Por exemplo, no modelo $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\}$ a distribuição de Jeffreys é dada por

$$h(\mu, \sigma^2) \propto \sigma^{-m}, \mu \in \mathbb{R}, \sigma^2 > 0,$$

com $m = 3$ segundo a regra biparamétrica e $m = 2$ pela regra assente na referida fatorização da distribuição conjunta dos parâmetros de localização, μ , e de escala, σ^2 .

Método da entropia máxima

A noção de entropia emprestada da Física, ao estar associada a uma medida de incerteza, foi sugerida por Jaynes como um meio de chegar a uma distribuição *a priori* que possa representar um estado de ignorância pelo menos relativa. Tal distribuição deveria então corresponder ao valor máximo possível da entropia.

Tomando entropia de uma distribuição $h(\theta)$, $\theta \in \Theta$ como o valor esperado $\mathcal{E}(h(\theta)) = E_h[-\ln h(\theta)]$, facilmente se mostra no caso finito quando $\Theta = \{\theta_1, \dots, \theta_k\}$ que a distribuição de entropia máxima (i.e., com a maior incerteza) é a Uniforme discreta $h(\theta_i) = 1/k$, $i = 1, \dots, k$, à qual corresponde a entropia de $\ln k$. Basta maximizar a função lagrangiana definida por $\mathcal{E}(h(\theta))$ acrescida do termo $\lambda (\sum_{i=1}^k h(\theta_i) - 1)$, onde λ representa o multiplicador de Lagrange associado à restrição natural de uma função de probabilidade.

Se se pretender maximizar a entropia restringida à informação representada por valores preespecificados de momentos ou quantis, e.g., na forma $E(g_j(\theta)) = u_j$, $j = 1, \dots, m$, basta usar a mesma via (método dos multiplicadores de Lagrange) com introdução de tais restrições adicionais para se chegar à expressão

$$h(\theta_i) = \frac{\exp\{\sum_{j=1}^m \lambda_j g_j(\theta_i)\}}{\sum_{l=1}^k \exp\{\sum_{j=1}^m \lambda_j g_j(\theta_l)\}},$$

onde os valores dos m multiplicadores λ_j se obtêm usando as correspondentes restrições introduzidas.

Exemplo 2: No contexto discreto que vem sendo considerado, admita-se $\theta_i = i$ pelo que $\Theta = \{1, \dots, k\}$, e que a mediana é um dos valores possíveis identificado por q . Assim, tem-se uma restrição imposta com $u_1 = q$ e $g_1(\theta)$ a função indicadora do acontecimento $\theta \leq q$, ou seja, dada por $\sum_{i=1}^q h(i) = 1/2$. Pela expressão acima

$$h(i) = \begin{cases} \frac{e^{\lambda_1}}{e^{\lambda_1 q + (k-q)}}, & \text{se } i \leq q \\ \frac{1}{e^{\lambda_1 q + (k-q)}}, & \text{se } q < i \leq k \end{cases},$$

onde $e^{\lambda_1} = (k - q)/q$ pela restrição identificadora do valor da mediana. Obtém-se então

$$h(i) = \begin{cases} \frac{1}{2q}, & \text{se } i \leq q \\ \frac{1}{2(k-q)}, & \text{se } q < i \leq k \end{cases},$$

ou seja, uma distribuição uniforme por troços. ■

No caso contínuo em que Θ é um intervalo limitado da reta real, o recurso ao Cálculo de Variações permite mostrar que a distribuição de entropia máxima é Uniforme contínua que, como se sabe já, não é invariante sob toda a transformação injetiva, o que cria problemas à visão da entropia $\mathcal{E}(h(\theta))$ como uma medida absoluta de incerteza.

Baseando-se na relação entre entropia e medida de informação de Kullback-Leibler no caso discreto, Jaynes (1968) redefine entropia no caso contínuo em relação a uma distribuição de referência não informativa $h_0(\theta)$ como $\mathcal{E}(h(\theta)) = E_h \left[-\ln \frac{h(\theta)}{h_0(\theta)} \right]$.

Se se supuser a existência de informação inicial representada por restrições como se definiram acima, o uso de novo do cálculo variacional leva a que a solução do problema de maximização seja expressável por

$$h(\theta) \propto h_0(\theta) \exp \left\{ \sum_{j=1}^m \lambda_j g_j(\theta) \right\},$$

onde os multiplicadores λ_j se obtêm a partir das restrições consideradas.

Exemplo 3: Considere-se que θ é um parâmetro de localização mas que se sabe ser positivo pelo que $\Theta = (0, +\infty)$ e que o seu valor médio é u . Adotando-se como distribuição *a priori* não informativa invariante sob translações a distribuição “Uniforme” em Θ , tem-se $h(\theta) \propto \exp(\lambda_1 \theta)$, $\theta > 0$, o que implica pela restrição natural que $h(\theta) = -\lambda_1 \exp(\lambda_1 \theta) I_{(0,+\infty)}(\theta)$, com $\lambda_1 < 0$, ou seja, uma distribuição Exponencial. Tendo em conta que o seu valor médio prefixado é $-1/\lambda_1 = u$, segue-se que a distribuição de entropia máxima é $\theta \sim \text{Exp}(1/u)$. ■

Exemplo 4: Seja θ novamente um parâmetro de localização tal que $\Theta = \mathbb{R}$ e suponha-se que $E(\theta) = u_1$ e $\text{Var}(\theta) = u_2$. Deste modo, usando a mesma distribuição de referência (imprópria) referida no exemplo anterior, tem-se que $h(\theta) \propto \exp\{\lambda_1 \theta + \lambda_2 (\theta - u_1)^2\}$, $\theta \in \mathbb{R}$, onde por simples manipulação algébrica se pode reescrever

$$\lambda_1 \theta + \lambda_2 (\theta - u_1)^2 = \lambda_2 \left[\theta - \left(u_1 - \frac{\lambda_1}{2\lambda_2} \right) \right]^2 + \left[\lambda_1 u_1 - \frac{\lambda_1^2}{4\lambda_2} \right].$$

Consequentemente,

$$h(\theta) \propto \exp \left\{ \lambda_2 \left[\theta - \left(u_1 - \frac{\lambda_1}{2\lambda_2} \right) \right]^2 \right\},$$

que corresponde ao núcleo de uma distribuição gaussiana de valor médio $u_1 - \lambda_1/(2\lambda_2)$ e variância $-1/(2\lambda_2)$ com $\lambda_2 < 0$. Atendendo aos dois momentos preespecificados conclui-se que $\lambda_1 = 0$ e $\lambda_2 = -1/(2u_2)$, conduzindo à distribuição *a priori* de entropia máxima $\theta \sim N(u_1, u_2)$. ■

3 Distribuições conjugadas naturais

A família distribucional com uma dada estrutura funcional a selecionar, onde se vai procurar um membro condizente com os resumos eliciados, deve idealmente satisfazer os seguintes requisitos:

- Versatilidade para acomodar o maior número possível de crenças *a priori*;
- Acessibilidade interpretativa para facilitar o processo de sumariação dos seus membros;
- Simplicidade da derivação analítica das distribuições *a posteriori* e preditivas.

A simplicidade da operação bayesiana poderá ficar garantida se se impuser que a família de distribuições *a priori* $\mathcal{H} = \{h(\theta | a) : a \in \mathcal{A}\}$, onde \mathcal{A} denota o conjunto de valores para os índices rotuladores das várias distribuições – os denominados hiperparâmetros – seja **fechada sob amostragem** de (qualquer elemento de) $\mathcal{F} = \{f(x | \theta) : \theta \in \Theta\}$, i.e., que

$$h(\theta) \in \mathcal{H} \Rightarrow h(\theta | x) \propto h(\theta) f(x | \theta) \in \mathcal{H}.$$

Nestas condições, diz-se também que \mathcal{H} é uma **família conjugada natural** de \mathcal{F} . De outra forma, a família \mathcal{H} diz-se conjugada natural de \mathcal{F} se $L(\theta | x) \equiv f(x | \theta)$, para cada x , é proporcional a um membro de \mathcal{H} e \mathcal{H} é fechada em relação a produtos, i.e., para todo o $a_0, a_1 \in \mathcal{A}$, existe $a_2 \in \mathcal{A}$ tal que

$$h(\theta | a_0)h(\theta | a_1) \propto h(\theta | a_2).$$

Exemplo 5: Sendo $x = (x_i, i = 1, \dots, n)$ uma concretização de uma amostra aleatória do modelo bernoulliano $Ber(\theta)$, tem-se

$$f(x_1, \dots, x_n | \theta) = \theta^{\sum_i x_i} (1 - \theta)^{n - \sum_i x_i},$$

que é proporcional ao núcleo de uma distribuição $Be(\sum_i x_i + 1, n - \sum_i x_i + 1)$ para θ que é fechado relativamente a produtos. A família conjugada natural do modelo amostral de Bernoulli é então a família Beta, cuja versatilidade é bem conhecida. Por conseguinte,

$$\theta \sim Be(a, b) \Rightarrow \theta | x \sim Be(A, B), \quad A = a + \sum_i x_i, B = b + n - \sum_i x_i,$$

revelando como a informação amostral se materializa na facilmente derivável distribuição *a posteriori* através dos números de sucessos e de insucessos (ou seja, da estatística suficiente mínima), conjugados aditivamente com os hiperparâmetros representantes da informação *a priori*.

Como $a, b > 0$, a informação *a priori* anula-se em termos relativos fazendo $a, b \rightarrow 0$, pelo que a distribuição não-informativa (ou vaga) obtida da família conjugada natural é a distribuição imprópria de Haldane, “ $Be(0, 0)$ ”, definida por $h(\theta) \propto \theta^{-1} (1 - \theta)^{-1}$, $\theta \in (0, 1)$, que corresponde à distribuição “Uniforme” para $\psi = \ln[\theta/(1 - \theta)] \in \mathbb{R}$. Na sequência, a distribuição *a priori* $Be(a, b)$ é interpretável como a distribuição *a posteriori* resultante da atualização dessa distribuição não-informativa por uma pseudoamostra de tamanho $a + b$ com a sucessos. ■

O facto de a operação bayesiana de conjugação das informações apriorística e amostral na família conjugada natural se processar dentro dela conduz a que possa ser simbolicamente representada por uma transformação no espaço \mathcal{A} dos hiperparâmetros

$$a \in \mathcal{A} \xrightarrow{\mathcal{F}} A = a + (A - a) \in \mathcal{A}.$$

Esta transformação, formatável de modo a evidenciar os pesos relativos dos dois tipos de informação, acaba por consubstanciar em si a simplicidade interpretativa e analítica do mecanismo bayesiano no quadro da família conjugada natural. Na forma acima, $A - a$ exprime a influência da informação amostral na alteração da informação *a priori* materializada em a e a ilustração disso conseguiu-se claramente com o exemplo anterior.

Exemplo 6: Se no exemplo anterior a amostra aleatória fosse respeitante ao modelo $Geo(\theta)$ de função de probabilidade $f(x_i | \theta) = \theta(1 - \theta)^{x_i}$, $x_i \in \mathbb{N}_0$, a aplicação do mesmo raciocínio conduziria à mesma família conjugada natural e à mesma distribuição não-informativa. Todavia, ter-se-ia $\theta | x \sim Be(A, B)$, $A = a + n$, $B = b + \sum_i x_i$ para uma distribuição *a priori* $Be(a, b)$ que seria então visualizável como uma distribuição *a posteriori* resultante da atualização de “ $Be(0, 0)$ ” por uma amostra fictícia do modelo Geométrico de tamanho a e número total de insucessos b . ■

Exemplo 7: Sendo $x = (x_i, i = 1, \dots, n)$ uma concretização de uma amostra aleatória do modelo Erlang $Ga(m, \lambda)$, onde $m \in \mathbb{N}$ é suposto conhecido, a função densidade amostral é tal que $f(x_1, \dots, x_n | \lambda) \propto \lambda^{mn} e^{-\lambda \sum_i x_i}$, apresentando um núcleo $Ga(mn + 1, \sum_i x_i)$ para λ que é fechado relativamente a produtos. Consequentemente, a família Gama é a conjugada natural sob amostragem do modelo Erlang, sendo então $\lambda | x \sim Ga(A, B)$, com $A = a + mn$, $B = b + \sum_i x_i$, a distribuição *a posteriori* correspondente à distribuição *a priori* $Ga(a, b)$, $a, b > 0$. Esta distribuição *a priori* é então interpretável como resultante da atualização da distribuição vaga “ $Ga(0, 0)$ ”, definida por $h(\lambda) \propto \lambda^{-1}$, $\lambda > 0$, por uma pseudoamostra erlangiana de tamanho a/m e média empírica das respetivas observações mb/a . ■

Exemplo 8: Considerando-se agora uma amostra aleatória do modelo Normal de valor médio μ e precisão $1/\sigma^2$ conhecida, o núcleo da correspondente função densidade numa concretização daquela amostra $x = (x_i, i = 1, \dots, n)$ pode escrever-se como

$$f(x_1, \dots, x_n | \mu) \propto e^{-\frac{n}{2\sigma^2}(\mu - \bar{x})^2},$$

mostrando que é proporcional ao núcleo de uma distribuição gaussiana para μ de valor médio \bar{x} e variância σ^2/n conhecida. O produto de dois núcleos deste tipo ainda é um núcleo do mesmo tipo¹.

Deste modo, a família conjugada natural é gaussiana, verificando-se $\mu \sim N(a, b^2) \Rightarrow \mu | x \sim N(A, B^2)$, onde pela identidade mencionada

$$A = \frac{\frac{1}{b^2}a + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{b^2} + \frac{n}{\sigma^2}}, \quad B^2 = \left(\frac{1}{b^2} + \frac{n}{\sigma^2} \right)^{-1}.$$

Fazendo $b \rightarrow +\infty$ obtém-se a distribuição “Uniforme” em \mathbb{R} como distribuição vaga, a qual por sua vez implica a distribuição *a posteriori* $\mu | x \sim N(\bar{x}, \sigma^2/n)$. Em consonância, a distribuição *a priori* $\mu \sim N(a, b^2)$ pode ser encarada como fruto da atualização da referida distribuição vaga por uma amostra fictícia de tamanho m do modelo Normal, com média empírica a e variância conhecida mb^2 . ■

Apesar de os exemplos ilustrativos neste capítulo se reportarem a modelos amostrais uniparamétricos, o raciocínio para identificar a família conjugada natural de modelos multiparamétricos, se existente, mantém-se integralmente como aliás se patenteia em exemplos descritos no capítulo que se segue a este.

A grande diferença reside na incomparavelmente maior dificuldade em eliciar o maior número de resumos necessários para identificar a distribuição conjugada natural num contexto multivariado que, infelizmente, não possui a mesma diversidade de formas distribucionais do caso univariado. Estratégias que rodeiam ou ultrapassam estes reveses na escolha de uma distribuição *a priori* fiável incluem a especificação de independência entre subconjuntos de parâmetros e a adoção de misturas contínuas ou finitas de distribuições conjugadas naturais.

Também outros métodos particularmente delineados para modelos multiparamétricos poderão ser usados para representar a distribuição *a priori*, em especial, o método das distribuições objetivas de referência devido a Berger e Bernardo e que se encontra detalhadamente descrito em Bernardo e Smith (1994). Este método nem sempre é bem-sucedido pela impossibilidade de obter analiticamente os resultados pretendidos, mormente em modelos parametricamente complexos.

Agradecimentos

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

Referências

- Bernardo, J.M. e Smith, A.F.M. (1994). *Bayesian Theory*. Wiley, Chichester.
- Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Systems, Science and Cybernetics*, 4, 227-291.
- Kass, R. E. e Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.*, 91, 435, 1343–1370.
- O’Hagan, A. (2010). *Bayesian Inference*. 3rd. ed.. Kendall’s Advanced Theory of Statistics, Vol. 2B. Arnold, London.

¹Atenda-se à identidade algébrica $d_1(z - c_1)^2 + d_2(z - c_2)^2 = (d_1 + d_2)(z - c)^2 + \frac{d_1 d_2}{d_1 + d_2}(c_1 - c_2)^2$, onde $c = \frac{d_1 c_1 + d_2 c_2}{d_1 + d_2}$.

Paulino, C.D., Amaral Turkman, M.A. e Murteira, B. (2003). *Estatística Bayesiana*, Fundação Calouste Gulbenkian, Lisboa.

Paulino, C.D., Soares, P. and Neuhaus, J. (2003). Binomial Regression with Misclassification. *Biometrics*, 59, 670-675.

Adenda: Este texto foi redigido segundo as normas do AO 1990.



Metodologia Bayesiana em Aplicações Básicas

Carlos Daniel Paulino, *dpaulino@math.ist.utl.pt*

Instituto Superior Técnico, Universidade de Lisboa e CEAUL

1 Introdução

Tendo conhecimento das ideias fundamentais da abordagem bayesiana à Inferência Estatística e os procedimentos mais relevantes para a representação da informação *a priori* importa passar-se para uma fase ilustrativa. Julgo ser incontroverso que a ilustração básica do paradigma bayesiano deva reportar-se à sua aplicação a situações onde os resultados inferencialmente relevantes possam ser deduzidos em termos tanto quanto possível exatos, de preferência analiticamente ou quando muito através de simulação operada através de distribuições *a posteriori* perfeitamente conhecidas.

É isso que vai ser feito neste capítulo ao proceder-se à análise de uma dezena de modelos bayesianos capazes de conduzir a uma resolução potencialmente apropriada de correspondentes exercícios maioritariamente do âmbito de cursos básicos de Estatística. Sendo os modelos bayesianos definidos como a distribuição conjunta do vetor (X) de observações e dos parâmetros (θ), a sua descrição de partida vai basear-se no modelo amostral $\{f(x|\theta)\}$ (o modelo da denominada Estatística Clássica) e na distribuição *a priori* $h(\theta)$, que se indicará notacionalmente por algo como $f(x|\theta) \wedge h(\theta)$ em forma predominantemente textual.

2 Modelo Binomial \wedge Beta

Seja $x = (x_i, i = 1, \dots, n)$ uma concretização das variáveis aleatórias condicionalmente independentes $X_i|m_i, \theta \sim Bi(m_i, \theta)$, $i = 1, \dots, n$, com m_i conhecidos e o parâmetro desconhecido θ munido de uma distribuição *a priori* $Be(a, b)$ com os respetivos hiperparâmetros fixados. Então a função densidade *a posteriori* de θ apresenta o núcleo

$$\begin{aligned} h(\theta|x) &\propto \prod_{i=1}^n \left\{ \binom{m_i}{x_i} \theta^{x_i} (1-\theta)^{m_i-x_i} \right\} h(\theta|a, b) \\ &\propto \theta^{a+\sum_i x_i-1} (1-\theta)^{b+\sum_i (m_i-x_i)-1}, \end{aligned}$$

evidenciando que $\theta|x \sim Be(A, B)$, em que $A = a + \sum_i x_i$ e $B = b + \sum_i (m_i - x_i)$.

Esta distribuição *a posteriori* equivale a usar para transformações de θ de interesse em certas aplicações outros tipos distribucionais como as distribuições F e Z de Fisher,

$$(B/A) \frac{\theta}{1-\theta} \Big| x \sim F_{(2A, 2B)}; \quad \left[(1/2) \ln(B/A) + (1/2) \ln \frac{\theta}{1-\theta} \right] \Big| x \sim Z_{(2A, 2B)}.$$

Momentos mistos da distribuição *a posteriori* de θ são facilmente derivados como

$$E[\theta^{r_1} (1-\theta)^{r_2} | x] = \frac{B(A+r_1, B+r_2)}{B(A, B)},$$

o que permite calcular vários resumos dessa distribuição como a média, $E(\theta|x) = A/(A+B)$ e a variância *a posteriori*. Outra estimativa pontual relevante, a moda *a posteriori*, é bem definida quando $A, B > 1$ por $m_0 = \frac{A-1}{A+B-2}$. Quantis e probabilidades *a posteriori* de θ podem ser calculados de funções beta incompletas (que não têm uma clara expressão explícita), ou no caso de a, b inteiros, de funções de distribuição binomial atendendo a que

$$F_{Be(A,B)}(\theta_0) = 1 - F_{Bi(A+B-1, \theta_0)}(A-1).$$

Tendo em vista a realização de predições, considere-se novas v. a. independentes das já observadas, $Y_j \equiv X_{n+j}, j = 1, \dots, k \stackrel{iid|\theta}{\sim} Bi(m_{n+j}, \theta)$. Como por definição a distribuição preditiva *a posteriori* de $Y = (Y_1, \dots, Y_k)$ é a mistura da sua distribuição amostral (produto de Binomiais) pela distribuição *a posteriori* (Beta) de θ , a sua função de probabilidade preditiva é expressável por

$$p(y_1, \dots, y_k|x) = \left[\prod_{j=1}^k \binom{m_{n+j}}{y_j} \right] \frac{B(A+y_., B+m_{n+} - y_.)}{B(A, B)},$$

em que $y_ = \sum_{j=1}^k y_j$, e $m_{n+} = \sum_j m_{n+j}$, podendo ser vantajosamente definida como o produto de duas funções de probabilidade, uma condicional relativa a uma distribuição Hipergeométrica multivariada e a outra marginal *a posteriori* de tipo Binomial-Beta,

$$p(y_1, \dots, y_k|x) = PHpg(\{m_{n+j}\}, y_)(y_1, \dots, y_k|y_.) \times PBiBe(m_{n+}, A, B)(y_., |x).$$

Com efeito, esta forma elucida o tipo de dependência existente na distribuição preditiva multivariada, bem como a sua natureza univariada $BiBe(m_{n+1}, A, B)$ quando $k=1$, de valor médio $E(Y_1|x) = m_{n+1} \frac{A}{A+B}$ e variância $Var(Y_1|x) = m_{n+1} \frac{AB}{(A+B)(A+B+1)} \left(1 + \frac{m_{n+1}}{A+B}\right)$.

3 Modelo Poisson \wedge Gama

Considere-se agora que $x = (x_i, i = 1, \dots, n)$ é uma concretização de uma amostra aleatória do modelo $Poi(\theta)$ de modo que a correspondente função de probabilidade amostral $f(x_1, \dots, x_n|\theta) \propto e^{-n\theta} \theta^{x_1}$. Como esta é proporcional ao núcleo de uma $Ga(x, +1, n)$ para θ , que é fechado sob produtos, fica claro que a família conjugada natural é a família Gama pelo que

$$\theta \sim Ga(a, b) \Rightarrow \theta|x \sim Ga(A, B), A = a + x_., B = b + n \Leftrightarrow \alpha = 2B\theta \sim \chi_{2A}^2,$$

sendo $h(\theta) \propto \theta^{-1} I_{(0, +\infty)}(\theta)$ a decorrente distribuição difusa imprópria.

Probabilidades *a posteriori* de acontecimentos prefixados em $\Theta = \mathbb{R}_+$ ou níveis de plausibilidade relativa *a posteriori* de hipóteses pontuais sobre θ são calculáveis de funções gama incompletas ou de funções de distribuição Qui-quadrado. Momentos *a posteriori* são dados por

$$E(\theta^r|x) = \frac{\Gamma(A+r)}{\Gamma(A)} \frac{B^A}{B^{A+r}}.$$

Para efeitos de predição, sejam $Y_j \equiv X_{n+j}, j = 1, \dots, k \stackrel{iid|\theta}{\sim} Poi(\theta)$, independentes da amostra aleatória observada. A distribuição preditiva *a posteriori* de $Y = (Y_1, \dots, Y_k)$ é a mistura da sua distribuição amostral (produto de distribuições Poisson) pela distribuição *a posteriori* $Ga(A, B)$, com a respetiva função de probabilidade preditiva definida após integração em θ por

$$p(y_1, \dots, y_k|x) = \frac{\Gamma(A+y_.)}{\Gamma(A) \prod_j y_j!} \left(\frac{B}{B+k}\right)^A \left(\frac{1}{B+k}\right)^{y_}.$$

Como a distribuição amostral de $Y_i = \sum_j Y_j$ é $Poi(k\theta)$, um cálculo análogo ao anterior mostra que a distribuição preditiva *a posteriori* de Y_i é a mistura Poi-Ga, de função de probabilidade expressável por $p(y_1, \dots, y_k | x) \prod_j y_j! (k^{y_j}) / y_i!$, mais conhecida por distribuição Binomial Negativa generalizada com parâmetros $(A, B/(B+k))$ traduzindo o número fixado de "sucessos" e a probabilidade de ocorrência de cada um deles. Deste modo, a função de probabilidade preditiva *a posteriori* de (Y_1, \dots, Y_k) pode visualizar-se esclarecedoramente por

$$p(y_1, \dots, y_k | x) = P_{M_{k-1}(y_i, \frac{1}{k} \mathbf{1}_k)}(y_1, \dots, y_k | y_i) \times P_{BiN(A, B/(B+k))}(y_i | x),$$

como o produto de uma função de probabilidade condicional Multinomial homogênea por uma função de probabilidade marginal *a posteriori* Poi-Ga. Esta representação ilustra o tipo de dependência no seio da distribuição preditiva *a posteriori* que é do tipo Binomial Negativa quando $k=1$.

4 Modelo Rayleigh \wedge Gama

A distribuição Rayleigh com função densidade $f(x|\delta) = \delta x e^{-\delta x^2/2} I_{(0,+\infty)}(x)$ é um modelo probabilístico relevante em alguns problemas de Engenharia. Se $x = (x_i, i = 1, \dots, n)$ for uma concretização de uma amostra aleatória desse modelo amostral e se se usar uma distribuição *a priori* $\delta \sim Ga(a, b)$, facilmente se conclui que $\delta | x \sim Ga(A, B)$, $A = a + n, B = b + \sum_i x_i^2/2$. Seria previsível que a família Gama fosse a conjugada natural do modelo Rayleigh já que o núcleo da densidade deste é o de uma distribuição Gama que é fechado perante produtos.

Resumos da distribuição *a posteriori* Gama de δ podem ser obtidos da expressão dos seus momentos simples referida na secção anterior. Em especial, $E(\delta | x) = A/B$ e $Var(\delta | x) = A/B^2$. Uma outra estimativa bayesiana de δ é a moda *a posteriori* dada por $(A-1)/B$ quando $A > 1$.

Testes de hipóteses $H_0 : \delta = \delta_0$ formuladas sem impacte na distribuição *a priori* podem ser efetuados recorrendo a funções gama incompletas. Por exemplo, usando a distribuição *a priori* de Jeffreys $h(\delta) \propto \delta^{-1} I_{(0,+\infty)}(\delta)$, a distribuição *a posteriori* associada a uma observação x do modelo Rayleigh é $\delta | x \sim Exp(x^2/2)$ pelo que o nível de plausibilidade relativa *a posteriori* de H_0 é $P = P(\delta \geq \delta_0 | x) = e^{-\delta_0 x^2/2}$.

Suponha-se agora que se pretendia testar H_0 com $\delta_0 = 2$ na base da informação *a priori* de quem conjecturou tal hipótese e que crê na veracidade dela com uma probabilidade de 50%, adotando para δ sob a hipótese alternativa uma distribuição $Ga(0,02; 0,01)$, e numa amostra de tamanho 5 de medições do modelo Rayleigh tal que $\sum_i x_i^2 = 7,54$. Note-se que a distribuição *a priori* para $\delta | H_1$ é uma distribuição própria de valor médio 2 e variância 200, razoavelmente próxima da distribuição imprópria de Jeffreys ao ter uma densidade $h_1(\delta)$ bastante achatada (ou plana) na maior parte do seu suporte, à exceção da gama de menores valores positivos. De acordo com a informação referida, o fator Bayes a favor de H_0 é bem favorável a esta já que

$$B(x) = \frac{P(H_0 | x)}{P(H_1 | x)} = \frac{f(x | \delta = 2)}{\int_{\delta \neq 2} f(x | \delta) h_1(\delta) d\delta} = 2^5 e^{-7,54} \frac{\Gamma(0,02) 3,78^{5,02}}{\Gamma(5,02) 0,01^{0,02}} \approx 29,53.$$

Pretendendo-se predizer Y tal que $Y | \delta \sim Ray(\delta)$ independentemente de (X_1, \dots, X_n) , facilmente se conclui que a sua função densidade preditiva *a posteriori* é definida por

$$p(y | x) = AB^A y (B + y^2/2)^{-(A+1)} I_{(0,+\infty)}(y).$$

Resumos desta distribuição preditiva *a posteriori* obtêm-se sem dificuldade. Por exemplo, a predição pontual média é $E(Y | x) = \sqrt{\frac{B}{2}} B(1/2, A - 1/2)$ (aplique-se a propriedade sequencial da esperança condicional e prove-se via integração por partes que $E(Y | \delta) = \sqrt{\pi/2} \delta^{-1/2}$). Probabilidades preditivas *a posteriori* são de cálculo direto, e.g. $P(Y > 1 | x) = (\frac{B}{B+1/2})^A$.

5 Modelo Uniforme \wedge Pareto

Seja $x = (x_i, i = 1, \dots, n)$ uma concretização de uma amostra aleatória do modelo $U([0, \theta])$, a respetiva função densidade amostral

$$f(x_1, \dots, x_n | \theta) = \theta^{-n} I_{[t, +\infty)}(\theta), \quad t = x_{(n)} \equiv \max_{1 \leq i \leq n} x_i$$

corresponde ao núcleo de uma distribuição Pareto para θ , $Pa(n-1, t)$, que é fechado sob produtos, pelo que a família Pareto é a conjugada natural do modelo amostral Uniforme acima definido. Assim, considerando $\theta \sim Pa(a, b)$ com $a, b > 0$ tem-se $\theta | x \sim Pa(A, B)$ com $A = a + n$, $B = b + t$ com função densidade $h(\theta | x) = AB^A \theta^{-(A+1)} I_{[B, +\infty)}(\theta)$. A associada distribuição difusa é então definida pela distribuição imprópria $h(\theta) \propto \theta^{-1} I_{[0, +\infty)}(\theta)$, por vezes rotulada como $\theta \sim "Pa(0, 0)"$, passando então a distribuição *a priori* própria $\theta \sim Pa(a, b)$ a ser visualizável como atualização da distribuição difusa através de uma pseudoamostra de dimensão a e máximo amostral b .

Cálculos simples permitem chegar às estimativas pontuais bayesianas moda = B , média = $BA/(A-1)$ (se $A > 1$) e mediana = $B2^{1/A}$. A forma da densidade *a posteriori* de θ mostra que o intervalo de credibilidade HPD a $100\gamma\%$ é $(B, \bar{\theta})$ com limite superior $\bar{\theta} = B(1-\gamma)^{-1/A}$.

Para a predição de uma nova observação Y do modelo amostral supostamente independente das observações já feitas, a função densidade preditiva *a posteriori* apresenta os seguintes dois troços (sendo obviamente nula em \mathbb{R}_-):

$$p(y|x) = \begin{cases} \frac{A}{A+1} p_{U([0, B])}(y), & \text{se } 0 < y \leq B \\ \frac{1}{A+1} p_{Pa(A, B)}(y), & \text{se } y > B. \end{cases}$$

A forma desta densidade mostra que para $\gamma \geq \frac{A}{A+1} = P(0 < Y \leq B|x)$ o intervalo de predição com a máxima densidade preditiva é definido por $\{y > B : p(y|x) \geq c_\gamma\} = (0, k_\gamma)$ com limite superior $k_\gamma = B[(A+1)(1-\gamma)]^{-1/A}$.

6 Modelo Normal (com média conhecida) \wedge Gama Inversa

Seja $x = (x_i, i = 1, \dots, n)$ uma concretização do vetor de variáveis aleatórias $X_i, i = 1, \dots, n \stackrel{iid}{\sim} N(\mu_0, \sigma^2)$, com μ_0 conhecido. A correspondente função densidade, ao poder escrever-se como

$$f(x_1, \dots, x_n | \sigma^2) \propto (\sigma^2)^{-n/2} e^{-\frac{\sum_i (x_i - \mu_0)^2}{2\sigma^2}},$$

é proporcional ao núcleo de uma distribuição Gama Inversa para σ^2 , $Gal(\frac{n}{2} - 1, \frac{1}{2} \sum_i (x_i - \mu_0)^2)$, que é fechado sob produtos. A conjugada natural deste modelo amostral é então a família Gal pelo que $\sigma^2 \sim Gal(a, b) \Leftrightarrow \sigma^{2(-1)} \sim Ga(a, b)$ implica que $\sigma^2 | x \sim Gal(A, B)$ com $A = a + \frac{n}{2}$, $B = b + \frac{1}{2} \sum_i (x_i - \mu_0)^2$. A distribuição *a priori* $Gal(a, b)$ é assim interpretável como a distribuição *a posteriori* resultante da atualização da distribuição vaga "Ga(0, 0)" por uma pseudoamostra de tamanho $2a$ do correspondente modelo Normal de média zero com soma dos quadrados das observações igual a $2b$.

Inferências paramétricas sobre a dispersão ou precisão das observações amostrais obtêm-se sem grande dificuldade de distribuições Gal ou Ga (ou com elas relacionadas). Inferências preditivas sobre Y a tomar do mesmo modelo amostral independentemente de $X_i, i = 1, \dots, n$ obtêm-se da mistura de $N(\mu_0, \sigma^2)$ pela distribuição *a posteriori* $\sigma^2 | x \sim Gal(A, B)$, a qual define uma distribuição t-Student com $2A$ graus de liberdade, parâmetro de localização μ_0 e parâmetro de escala $\sqrt{B/A}$, simbolizada por $t_{(2A)}(\mu_0, B/A)$, com função densidade preditiva

$$\begin{aligned}
p(y|x) &= \frac{B^A}{\sqrt{2\pi} \Gamma(A)} \int_0^{+\infty} (\sigma^2)^{-(A+3/2)} e^{-(1/\sigma^2) \left[B + \frac{(y-\mu_0)^2}{2} \right]} d\sigma^2 \\
&= \left[B \left(\frac{2A}{2}, \frac{1}{2} \right) \right]^{-1} (\sqrt{2AB/A})^{-1} \left[1 + \frac{(y-\mu_0)^2}{2AB/A} \right]^{-\frac{2A+1}{2}}
\end{aligned}$$

7 Modelo Normal biparamétrico \wedge distribuições *a priori* marginais de Jeffreys

Denotando $x = (x_i, i = 1, \dots, n)$ os dados relativos a uma amostra aleatória ($X_i, i = 1, \dots, n$) do modelo $N(\mu, \sigma^2)$, a função de verosimilhança

$$f(x_1, \dots, x_n | \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{n}{2\sigma^2} (\mu - \bar{x})^2 - \frac{k\sigma^2}{2\sigma^2} \right\},$$

onde $k = n - 1$ e $k\sigma^2 = \sum_i (x_i - \bar{x})^2$, é o núcleo de uma distribuição conjunta Normal-Gama Inversa para (μ, σ^2) entendida como uma Normal para μ dado σ^2 e uma Gama Inversa para σ^2 , que é fechado sob produtos. A família conjugada natural é então definida por funções densidade do tipo $h(\mu, \sigma^2 | a, v, c, d) = h_{N(a, \sigma^2/v)}(\mu | \sigma^2) h_{Gal(c, d)}(\sigma^2)$.

A distribuição de Jeffreys sob independência *a priori* entre μ e σ^2 , $h(\mu, \sigma^2) \propto \sigma^{-2}$ é um caso limite de uma distribuição imprópria N-GaI. A sua atualização bayesiana pela verosimilhança indicada origina a função densidade de probabilidade

$$h(\mu, \sigma^2 | x) \propto (\sigma^2)^{-1/2} e^{-\frac{n}{2\sigma^2} (\mu - \bar{x})^2} \times (\sigma^2)^{-(\frac{n-1}{2} + 1)} e^{-\frac{k\sigma^2}{2\sigma^2}}$$

evidenciando que $\mu | \sigma^2, x \sim N(\bar{x}, \sigma^2/n)$ e $\sigma^2 | x \sim Gal(\frac{k}{2}, \frac{k\sigma^2}{2}) \Leftrightarrow \frac{k\sigma^2}{2} | x \sim \chi_{(k)}^2$.

A distribuição marginal *a posteriori* de μ é então uma mistura de Normais por uma GaI, que já se sabe ser do tipo t-Student. Com efeito, por integração de σ^2 , tendo em conta a função densidade GaI, resulta

$$h(\mu | x) = \left[B \left(\frac{k}{2}, \frac{1}{2} \right) \right]^{-1} \left(\sqrt{k\sigma^2/n} \right)^{-1} \left[1 + \frac{(\mu - \bar{x})^2}{k\sigma^2/n} \right]^{-\frac{k+1}{2}},$$

isto é, $\mu | x \sim t_{(k)}(\bar{x}, s^2/n) \Leftrightarrow \frac{\mu - \bar{x}}{s/\sqrt{n}} | x \sim t_{(k)}(0, 1) \Leftrightarrow \frac{(\mu - \bar{x})^2}{s^2/n} | x \sim F_{(1, k)}$, onde a distribuição t-Student reduzida $t_{(k)}(0, 1)$ é a conhecida t-Student da Estatística Clássica. Assim, $E(\mu | x) = \bar{x}$ (se $k > 1$) e $Var(\mu | x) = \frac{k}{k-2} \frac{s^2}{n}$ (se $k > 2$). A distribuição *a posteriori* de σ^2 condicional a μ pode então ser determinada como $\sigma^2 | \mu, x \sim Gal(\frac{k+1}{2}, \frac{k\sigma^2 + n(\mu - \bar{x})^2}{2})$.

Inferências sobre cada um dos parâmetros de localização e escala do modelo amostral obtêm-se sem dificuldade de maior com base nas distribuições t-Student e Gama Inversa (ou χ^2 por transformação apropriada). Em termos de predição, considerando-se por exemplo uma amostra aleatória futura de tamanho m do modelo cuja média \bar{Y} se pretende predizer, a correspondente distribuição preditiva *a posteriori* é a mistura das distribuições $\bar{Y} | \mu, \sigma^2 \sim N(\mu, \sigma^2/m)$ pela distribuição *a posteriori* conjunta $h(\mu, \sigma^2 | x) = h(\mu | \sigma^2, x) h(\sigma^2 | x)$. Tendo em conta a identidade algébrica da combinação linear das formas quadráticas, $\frac{1}{\sigma^2} [m(\mu - \bar{y})^2 + n(\mu - \bar{x})^2]$, essa distribuição reveste a forma t-Student, $\bar{Y} | x \sim t_{(k)}(\bar{x}, \frac{m+n}{mn} s^2)$, da qual se determinam facilmente resumos pontuais e intervalares.

8 Dois modelos Normais independentes \wedge distribuições marginais de Jeffreys

Sejam $x_j = (x_{ji}, i = 1, \dots, n_j), j = 1, 2$ concretizações de duas amostras aleatórias independentes de populações $N(\mu_j, \sigma_j^2)$ e use-se para os 4 parâmetros a distribuição *a priori* de Jeffreys comumente adotada $h(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto (\sigma_1^2 \sigma_2^2)^{-1}$ no respetivo espaço paramétrico conjunto.

Comparação de médias

Seguindo um raciocínio análogo ao da secção imediatamente anterior, conclui-se facilmente que (μ_1, σ_1^2) e (μ_2, σ_2^2) são *a posteriori* também independentes com as distribuições marginais univariadas

$$\mu_j | x_j \sim t_{(k_j)}(\bar{x}_j, s_j^2/n_j) \Leftrightarrow \mathbf{v}_j = \frac{\mu_j - \bar{x}_j}{s_j/\sqrt{n_j}} | x_j \sim t_{(k_j)}$$

$$\sigma_j^2 | x_j \sim \text{Gal}\left(\frac{k_j}{2}, \frac{k_j s_j^2}{2}\right)$$

em que $k_j = n_j - 1$ e $k_j s_j^2 = \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j)^2$.

A redução de $\lambda = \mu_1 - \mu_2$, expressa por

$$\tau = \frac{\lambda - (\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \equiv \mathbf{v}_1 \sin u + \mathbf{v}_2 \cos u,$$

em que $u = \arctan\left(\frac{s_1/\sqrt{n_1}}{s_2/\sqrt{n_2}}\right)$, é distribuída *a posteriori* como uma combinação linear de distribuições t-Student independentes, conhecida como distribuição de Behrens-Fisher e parametrizada por k_1, k_2 e u ¹. A sua função densidade, que é simétrica mas não expressável em forma fechada, costuma ser na prática aproximada por uma distribuição t-Student devida a Patil (1964), concretamente $\tau | x_1, x_2 \sim BF(k_1, k_2, u) \underset{\text{aprox}}{\sim} t_{(b)}(0, a)$, onde

$$b = 4 + c_1^2/c_2, \quad a = \sqrt{c_1(b-2)/b}$$

$$c_1 = \frac{k_1}{k_1 - 2} \sin^2 u + \frac{k_2}{k_2 - 2} \cos^2 u$$

$$c_2 = \frac{k_1^2}{(k_1 - 2)^2(k_1 - 4)} \sin^4 u + \frac{k_2^2}{(k_2 - 2)^2(k_2 - 4)} \cos^4 u.$$

Uma alternativa ao uso da aproximação de Patil consiste na geração de uma amostra da distribuição *a posteriori* de τ através de simulação a partir das distribuições *a posteriori* de \mathbf{v}_1 e \mathbf{v}_2 , com base na qual se podem calcular empiricamente estimativas pontuais e intervalares e testar hipóteses pontuais sobre a diferença de médias.

Comparação de variâncias

Tomando como parâmetro de interesse $\psi = \frac{\sigma_1^2}{\sigma_2^2}$, conclui-se das distribuições *a posteriori* Gama independentes de $\{1/\sigma_j^2\}$ que $\psi | x_1, x_2 \stackrel{d}{\equiv} \frac{s_1^2}{s_2^2} F_{(k_2, k_1)}$, o que permite realizar facilmente inferências básicas sobre ψ .

¹Note-se que esta dependência de u leva a que não haja dualidade entre as distribuições *a posteriori* e amostral de τ e, consequentemente, identidade numérica entre as inferências bayesiana e clássica sobre a diferença de médias, contrariamente ao que acontece em outras situações em que se usam distribuições *a priori* não informativas.

Comparação de médias de populações homocedásticas

Neste contexto admite-se $\sigma_1^2 = \sigma_2^2 \equiv \sigma^2$ e o uso da distribuição *a priori* de Jeffreys $h(\mu_1, \mu_2, \sigma^2) \propto \sigma^{-2}$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma^2 > 0$. Tomando de novo como referência o material da secção anterior, rapidamente se conclui que

$$\lambda = \mu_1 - \mu_2 | \sigma^2, x_1, x_2 \sim N(\bar{x}_1 - \bar{x}_2, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)); \quad \sigma^2 | x_1, x_2 \sim \text{Gal} \left(\frac{k}{2}, \frac{ks^2}{2} \right);$$

onde $k = n_1 + n_2 - 2$ e $s^2 = k^{-1} \sum_j (n_j - 1) s_j^2$ é a variância empírica combinada. Isto implica designadamente que

$$\lambda = \mu_1 - \mu_2 | x_1, x_2 \sim t_{(k)} \left(\bar{x}_1 - \bar{x}_2, s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right) \Leftrightarrow \frac{\lambda - (\bar{x}_1 - \bar{x}_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} | x_1, x_2 \sim t_{(k)},$$

que é o resultado básico para o traçado das inferências de interesse sobre a comparação das duas populações Normais.

9 Dois modelos Binomiais independentes \wedge distribuições Beta

Sejam t_j , $j = 1, 2$ contagens observadas de T_j , $j = 1, 2 | \theta_j \sim \text{Bi}(m_j, \theta_j)$, com $\{m_j\}$ conhecidos e considere-se o modelo *a priori* θ_j , $j = 1, 2 \underset{\text{ind}}{\sim} \text{Be}(a_j, b_j)$. De acordo com o que se viu para o modelo $\text{Bi} \wedge \text{Be}$, resulta que

$$\begin{aligned} \theta_j | t_j, j = 1, 2 \underset{\text{ind}}{\sim} \text{Be}(A_j, B_j), \quad A_j = a_j + t_j, \quad B_j = b_j + m_j - t_j &\Leftrightarrow (B_j/A_j) \frac{\theta_j}{1 - \theta_j} | t_j, j = 1, 2 \underset{\text{ind}}{\sim} F_{(2A_j, 2B_j)} \\ &\Leftrightarrow \left[(1/2) \ln(B_j/A_j) + (1/2) \ln \frac{\theta_j}{1 - \theta_j} \right] | t_j, j = 1, 2 \underset{\text{ind}}{\sim} Z_{(2A_j, 2B_j)}. \end{aligned}$$

Testes unilaterais exatos de comparação de proporções

Para o teste de $H_0 : \theta_1 \leq \theta_2$ contra $H_1 : \theta_1 > \theta_2$ o uso das chances *a posteriori* ou do fator Bayes requer a avaliação das quantidades

$$P(H_0 | t_1, t_2) = \int_0^1 h(\theta_1 | t_1) \left[\int_{\theta_1}^1 h(\theta_2 | t_2) d\theta_2 \right] d\theta_1$$

(simultaneamente com o seu análogo baseado em distribuições *a priori*). No caso de a_2 e b_2 serem inteiros, a relação entre as funções de distribuição Beta e Binomial, $F_{\text{Be}(A_2, B_2)}(\theta_1) = 1 - F_{\text{Bi}(A_2 + B_2 - 1, \theta_1)}(A_2 - 1)$ permite que se escreva

$$P(H_0 | t_1, t_2) = [B(A_1, B_1)]^{-1} \sum_{u=0}^{A_2-1} \binom{A_2 + B_2 - 1}{u} B(A_1 + u, B_1 + A_2 + B_2 - 1 - u),$$

com as funções beta calculáveis em termos de fatoriais se cumulativamente a_1 e b_1 também forem inteiros.

Testes de homogeneidade das Binomiais, $H_0 : \theta_1 = \theta_2$ contra $H_1 : \theta_1 \neq \theta_2$

Atendendo a que $H_0 : \pi = 0 \Leftrightarrow \ln \Delta = 0$ usando as transformações $\pi = \theta_1 - \theta_2$ e $\Delta = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}$, o recurso a simulação a partir das distribuições *a posteriori* $\text{Be}(A_j, B_j)$ – e.g., via $\theta_j = \frac{v_{A_j}}{v_{A_j} + v_{B_j}}$, com $v_{K_j}, K =$

$A, B \underset{ind}{\sim} Ga(K_j, 1)$ – de amostras de π ou de Δ (ou $\ln \Delta$) permite obter boas aproximações empíricas do nível de plausibilidade relativa *a posteriori* de H_0 ou de intervalos HPD. Obviamente que esta via é também aplicável a hipóteses unilaterais por meio do cálculo de apropriadas proporções com base nas amostras simuladas.

No caso de grandes valores observados de sucessos e insucessos, a utilização de aproximações assintóticas à distribuição Z de Fisher, por exemplo

$$Z_{(v_1, v_2)} \underset{approx}{\sim} N \left[\frac{1}{2} \ln \frac{v_1^{-1} - 1}{v_2^{-1} - 1}, \frac{1}{2} (v_1^{-1} + v_2^{-1}) \right],$$

possibilita recorrer à distribuição *a posteriori* aproximada

$$\ln \Delta | t_1, t_2 \underset{approx}{\sim} N \left[\ln \frac{(A_1 - 1/2)/(B_1 - 1/2)}{(A_2 - 1/2)/(B_2 - 1/2)}, \sum_{j=1,2} (A_j^{-1} + B_j^{-1}) \right]$$

para a construção de testes bayesianos unilaterais ou bilaterais das hipóteses em confronto.

10 Modelo Multinomial \wedge Dirichlet

Este modelo bayesiano é a versão multivariada do modelo Binomial \wedge Beta mas como é bem menos familiar do que este, opta-se por começar a descrever sucintamente as principais propriedades dele pela sua relevância no campo inferencial.

Seja $X = (X_1, \dots, X_c)$ e $\theta = (\theta_1, \dots, \theta_c)$ vetores aleatórios assumindo valores nos subespaços respectivamente $\mathcal{X} = \{x = (x_1, \dots, x_c) : x_i \in \mathbb{N}_0, x_i = \sum_{i=1}^c x_i \leq N\}$, onde N se supõe conhecido, e $\Theta = \{(\theta_1, \dots, \theta_c) : \theta_i \in (0, 1), \theta_i = \sum_{i=1}^c \theta_i < 1\}$, o chamada simplex c -dimensional \mathcal{S}_c .

Distribuição Multinomial (c -variada) para X

Função de probabilidade de $X | \theta \sim M_c(N, \theta) : f(x | \theta) = \frac{N!}{\prod_{i=1}^{c+1} x_i!} \prod_{i=1}^{c+1} \theta_i^{x_i}$, $x \in \mathcal{X}$ em que $x_{c+1} = N - x_i$.

Os seus dois primeiros momentos são definidos por (aplique-se, e.g., a técnica da função geradora de momentos (f.g.m.)):

$$\mu = E(X | \theta) = N\theta; \quad \Sigma = \text{Var}(X | \theta) = N(D_\theta - \theta\theta'),$$

onde $D_\theta = \text{diag}(\theta_1, \dots, \theta_c)$.

Denote-se por $C_k = \{j_{k-1} + 1, \dots, j_k\}$, $k = 1, \dots, s + 1$ as partes de uma partição do conjunto dos índices $\{1, 2, \dots, c, c + 1\}$ em $s + 1$ subconjuntos em que $\#C_k = d_k$, $j_0 = 0$ e $j_{s+1} = c + 1$. Em correspondência, considerem-se os desdobramentos das componentes de X definidos por

$$M_k = \sum_{i \in C_k} X_i, \quad k = 1, \dots, s + 1; \quad X^{(k)} = (X_i, i \in C_k), \quad k = 1, \dots, s + 1.$$

Então (recorra-se à via da f.g.m. e da definição de distribuição condicional)

$$M = (M_1, \dots, M_s) | \theta \sim M_s(N, \alpha), \quad \alpha = (\alpha_1, \dots, \alpha_s), \quad \alpha_k = \sum_{i \in C_k} \theta_i$$

$$X^{(k)} | M, \theta, \quad k = 1, \dots, s + 1 \underset{ind}{\sim} M_{d_k-1}(M_k, \pi_k), \quad \pi_k = (\theta_i / \alpha_k, i \in C_k - j_k).$$

Estas transformações evidenciam que distribuições marginais e condicionais apropriadas de componentes de uma distribuição Multinomial são também Multinomiais (Binomiais, no caso univariado). A sua

relevância é cabalmente esclarecida no contexto de tabelas de contingência. Por exemplo, se X representa o vetor das frequências de uma tabela de contingência bidimensional, as respectivas frequências da margem das linhas (ou das colunas) são expressas adequadamente por M e as frequências das linhas (ou das colunas) condicionadas aos totais marginais são caracterizadas pela distribuição condicional indicada Produto de Multinomiais.

Distribuição Dirichlet (c-variada) para θ

Função densidade de probabilidade de $\theta|a \sim D_c(a) : h(\theta|a) = [B(a)]^{-1} \prod_{i=1}^{c+1} \theta_i^{a_i-1}$, $\theta \in \Theta = \mathcal{S}_c$ em que $a = (a_1, \dots, a_c, a_{c+1}) \in \mathbb{R}_+^c$, $\theta_{c+1}=1-\theta$, e $B(a) = \frac{\prod_{i=1}^{c+1} \Gamma(a_i)}{\Gamma(a_\bullet)}$ é a função beta multivariada.

A distribuição Dirichlet pode ser vantajosamente definida nos planos teórico e computacional a partir de distribuições Gama independentes através da transformação $\theta_i = \frac{v_i}{\sum_{j=1}^{c+1} v_j}$, $i = 1, \dots, c$, com as variáveis aleatórias v_i , $i = 1, \dots, c+1 \underset{ind}{\sim} Ga(a_i, 1)$.

Os momentos mistos são pela sua definição dados por $E[\prod_{i=1}^{c+1} \theta_i^{r_i} | a] = \frac{B(a+r)}{B(a)}$, onde $r = (r_1, \dots, r_c, r_{c+1})$, de onde se obtém

$$E(\theta_i|a) \equiv E_i = \frac{a_i}{a_\bullet}; \text{Var}(\theta_i|a) \equiv V_i = \frac{E_i(1-E_i)}{a_\bullet + 1}; \text{cov}(\theta_i, \theta_j|a) \equiv V_{ij} = -\frac{E_i E_j}{a_\bullet + 1}, i \neq j.$$

Como $(\alpha, \pi_1, \dots, \pi_{s+1})$ é uma reparametrização de θ prova-se (use-se, e.g., a representação por distribuições Gama) que

$$\alpha|a \sim D_s(\sum_{i \in C_k} a_i, k = 1, \dots, s+1)$$

$$\pi_k|a, k = 1, \dots, s+1 \underset{ind}{\sim} D_{d_k-1}(a_i, i \in C_k)$$

Distribuição Multinomial-Dirichlet (c-variada) para X

Esta distribuição, também conhecida por distribuição de Pólya, traduz a mistura de uma Multinomial por uma Dirichlet, $X|a \sim MD_c(N, a)$, cuja função de probabilidade é dada por

$$p(x|a) = \frac{N!}{\prod_{i=1}^{c+1} x_i!} \frac{B(a_1 + x_1, \dots, a_{c+1} + x_{c+1})}{B(a_1, \dots, a_{c+1})}, x \in \mathcal{X}$$

e cujos dois primeiros momentos são expressáveis (use-se, e.g., as propriedades da esperança condicional) por

$$E(X|a) = N \frac{a}{a_\bullet}; V(X|a) = \frac{a_\bullet + N}{a_\bullet(a_\bullet + 1)} N(D_a - \frac{aa'}{a_\bullet}).$$

Aplicando os desdobramentos anteriores de X , facilmente se conclui que

$$M = (M_1, \dots, M_s)|a \sim MD_s(N; \sum_{i \in C_k} a_i, k = 1, \dots, s+1)$$

$$X^{(k)}|M, a, k = 1, \dots, s+1 \underset{ind}{\sim} MD_{d_k-1}(M_k; a_i, i \in C_k).$$

Aplicação inferencial

Sendo $x = (x_1, \dots, x_c)$ uma concretização de um vetor aleatório $X|\theta \sim M_c(N, \theta)$, como $f(x|\theta)$ é proporcional ao núcleo de uma distribuição $D_c(x_i + 1, i = 1, \dots, c+1)$ que é fechado sob produtos, conclui-se

que a família Dirichlet é a conjugada natural de uma amostragem Multinomial. Assim, se $\theta|a \sim D_c(a)$ então $\theta|a, x \sim D_c(A)$, $A = (A_i = a_i + x_i, i = 1, \dots, c + 1)$.

Estimativas bayesianas de $(\theta_i, i = 1, \dots, c + 1)$ podem ser traduzidas em particular pelas componentes da moda *a posteriori* $(A - 1_{c+1})/A$, (se $A_i > 1, \forall i$), onde 1_{c+1} representa um vetor de $c + 1$ elementos iguais a 1, ou da média *a posteriori* A/A . Note-se como esta é uma média ponderada do vetor média *a priori*, a/a , e do vetor das proporções amostrais $p = (x_i/N, i = 1, \dots, c + 1)$.

Em análise de tabelas de contingência o interesse inferencial diz muitas vezes respeito a estruturas de independência (ou a outros modelos log-lineares) nas quais desempenham um papel crucial funções paramétricas de tipo $\sum_i b_i \ln \theta_i$ com $\sum_i b_i = 0$ (*vide e.g.* Paulino e Singer, 2006). Quando as componentes de A são grandes, pode-se invocar a normalidade aproximada da sua distribuição *a posteriori* e a decorrente distribuição Qui-quadrado para apropriadas formas quadráticas, permitindo testar aquelas estruturas (para aprofundamento *vide e.g.* Cap. 6 de Paulino *et al.*, 2003). No caso especial de uma tabela 2×2 a aplicação deste procedimento à hipótese de independência, que mais não é do que a de homogeneidade de duas Binomiais, conduz à via mencionada no fim da secção anterior.

Se o objetivo for o de prever um vetor Y tal que $Y|m, \theta \sim M_c(m, \theta)$, a correspondente distribuição preditiva *a posteriori* é Multinomial-Dirichlet, $Y|m, x \sim MD_c(m, A)$, cujo resumo através dos seus dois primeiros momentos é obtível das fórmulas expostas atrás.

11 Inferência sobre populações finitas

Considere-se uma população finita de tamanho conhecido N particionada em $c \leq N$ grupos de tamanhos desconhecidos $N_i, i = 1, \dots, c$, tal que $\sum_{i=1}^c N_i = N$, da qual se selecionou aleatoriamente (sem reposição) uma amostra S de $n \leq N$ unidades com o objetivo de traçar inferências sobre o vetor de totais populacionais dos grupos $\theta = (N_1, \dots, N_c)$. Sejam $n_i, i = 1, \dots, c$ as frequências observadas dos grupos ($\sum_{i=1}^c n_i = n$) que agrupamos em $x = (n_1, \dots, n_c)$ que, pelo exposto, é uma observação da distribuição Hipergeométrica multivariada $X|N, n, \theta \sim Hpg_{c-1}(\theta, n)$ (usa-se, aqui e em seguida, por conveniência uma redundância notacional na definição de vetores aleatórios como em X).

Denotando por U_k o vetor indicador do grupo a que pertence a unidade k , cujos valores possíveis são os vetores da base ortonormal padrão de \mathbb{R}^c , o alvo inferencial é expressável por

$$\theta = \sum_{k=1}^N U_k = \sum_{k \in S} U_k + \sum_{k \notin S} U_k \equiv X + (\theta - X),$$

evidenciando particularmente que *a posteriori* apenas $\theta - X$ é desconhecido.

Admita-se uma distribuição *a priori* numa estrutura hierárquica definida por $U_1, \dots, U_N \stackrel{iid}{\sim} M_{c-1}(1, \phi)$ dado algum parâmetro subjacente $\phi = (\phi_j, j = 1, \dots, c)$, com $\sum_j \phi_j = 1$, a que se atribui num 2º nível a distribuição $\phi|a \sim D_{c-1}(a), a = (a_1, \dots, a_c) \in \mathbb{R}_+^c$.

Em termos do 1º nível da hierarquia, tem-se assim que $\theta|\phi \sim M_{c-1}(N, \phi)$ e que X e $\theta - X$, por definição, são *a priori* e condicionalmente a ϕ independentemente distribuídos segundo leis do mesmo tipo, $X|n, \phi \sim M_{c-1}(n, \phi)$ e $\theta - X|n, \phi \sim M_{c-1}(N - n, \phi)$. Observe-se ainda que a distribuição amostral Hipergeométrica de X pode ser encarada como

$$\begin{aligned} f(x|n, \theta) &= \frac{\prod_{j=1}^c \binom{N_j}{x_j}}{\binom{N}{n}} = \frac{f(x|n, \phi) \times h(\theta - x|n, \phi)}{h(\theta|\phi)} \\ &= \frac{f(x, \theta|n, \phi)}{h(\theta|\phi)} = f(x|n, \theta, \phi). \end{aligned}$$

Usando a informação do 2º nível, pode-se identificar como sendo do tipo Multinomial-Dirichlet as seguintes distribuições marginais (ou preditivas *a priori*)

$$\theta|a \sim MD_{c-1}(N, a); X|n, a \sim MD_{c-1}(n, a); \theta - X|n, a \sim MD_{c-1}(N - n, a)$$

e que a atualização por x de ϕ é tal que $\phi|x \sim D_{c-1}(a+x)$. Por outro lado, como $\theta - x|x, \phi \stackrel{d}{=} \theta - x|\phi \sim M_{c-1}(N - n, \phi)$, resulta que $\theta - x|x \sim MD_{c-1}(N - n, a+x)$.

Em suma, a distribuição *a posteriori* de $\theta - x$ sob uma amostragem Hipergeométrica é do mesmo tipo (Multinomial-Dirichlet) da respetiva distribuição *a priori* e a distribuição *a posteriori* do vetor de totais θ resulta daquela por uma translação de x , da qual se obtém trivialmente a do vetor de proporções populacionais θ/N (Basu e Pereira, 1982).

Agradecimentos

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

Referências

Basu, D. e Pereira, C.A.B. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse. *J. Statist. Plann. Infer.*, **6**, 345-362.

Bernardo, J.M. e Smith, A.F.M. (1994). *Bayesian Theory*. Wiley, Chichester.

Patil, V.H. (1964). The Behrens-Fisher problem and its Bayesian solution. *J. Indian Statist. Assoc.*, **2**, 21.

Paulino, C.D., Amaral Turkman, M.A. e Murteira, B. (2003). *Estatística Bayesiana*, Fundação Calouste Gulbenkian, Lisboa.

Paulino, C.D. e Singer, J.M. (2006). *Análise de Dados Categorizados*. Editora Edgard Blücher, São Paulo.

Adenda: Este texto foi redigido segundo as normas do AO 1990.



Estatística Bayesiana: Aplicações em Análise de Sobrevivência e Estatística Espacial

Giovani Loiola da Silva, gsilva@math.ist.utl.pt

CEAUL e Instituto Superior Técnico, Universidade de Lisboa

1 Introdução

O objetivo deste artigo é apresentar algumas aplicações de Estatística sob uma abordagem bayesiana. Dois cenários foram escolhidos para compor o artigo: i) análise de sobrevivência, ii) estatística espacial. Os métodos estatísticos desenvolvidos para estes cenários estão relacionados com as especificidades de dois conjuntos de dados: i) ocorrência de tumores de mama em ratos após um período de 60 dias expostos a um composto cancerígeno propício a esse tipo de cancro (Gail *et al.*, 1980), ii) admissão de pacientes, acima dos 25 anos, com enfarte do miocárdio em cada uma das 139 regiões de saúde da província de Quebec entre 1996 a 1999 (Silva *et al.*, 2005). Outras aplicações de Estatística Bayesiana podem ser encontradas nomeadamente nos seguintes textos: Gilks *et al.* (1996), Gelman *et al.* (2003), Paulino *et al.* (2003), Congdon (2006), Ntzoufras (2009), Lunn *et al.* (2012) e Cowles (2013).

Devido às dificuldades de implementação dos modelos subjacentes aos três cenários de aplicação, faz-se aqui o uso de métodos de simulação, sobretudo os métodos de Monte Carlo via cadeias de Markov (MCMC), implementados no *software* WinBUGS (Lunn *et al.*, 2000). Na realidade, a difusão da aplicação dos métodos bayesianos esteve limitada até aos anos 90 pelo facto de a distribuição *a posteriori* em muitas situações práticas ser analiticamente intratável. Nas últimas décadas tem sido ultrapassado devido ao aparecimento de vários métodos numéricos (vide, *e.g.*, Gelfand e Smith (1990), Chen *et al.* (2000) e Rue *et al.* (2009)). Como o algoritmo de simulação nos métodos MCMC não é muito complicado e não é restrito a um número pequeno de parâmetros, estes métodos são aplicados correntemente na resolução bayesiana de muitos problemas práticos, incluindo na simulação de distribuições *a posteriori* para obter as estimativas das quantidades de interesse nos modelos de regressão aqui em estudo.

O resto do artigo está organizado da seguinte forma. Na Secção 2 discutimos análise de sobrevivência com fragilidade tendo em conta uma estrutura aditiva quer para a função de risco quer para o termo de fragilidade. Modelos de regressão de Poisson espacial são apresentados na Secção 3 visando nomeadamente encontrar variação extra (não observada) entre regiões de saúde para o mapeamento de doenças.

2 Aplicação em Análise de sobrevivência

Gail *et al.* (1980) apresentaram um conjunto de dados de sobrevivência caracterizado pela ocorrência de um ou mais eventos de interesse nas unidades. Inicialmente injetou-se um composto cancerígeno propício ao cancro de mama em 76 ratos (dia zero) dos quais 48 ratos não apresentaram tumores de mama após um período de 60 dias. Posteriormente os 48 ratos livres de cancro foram aleatorizados em 2 grupos: controlo (25 ratos) e tratamento com *retinoid prophylaxis* (23 ratos), formando assim 48 unidades dos dados de sobrevivência. Os 48 ratos foram regularmente observados com exames feitos

duas vezes por semana, tendo-se anotado o número de novos tumores entre dois exames consecutivos. Todas as unidades (ratos) foram censuradas à direita após o 182^o dia do início da experiência. Observe-se que aqui os tempos de sobrevivência são os tempos até ao aparecimento de novos tumores entre exames consecutivos. Este conjunto de dados é um exemplo de dados de sobrevivência com eventos múltiplos, uma vez que a ocorrência de novos tumores (evento de interesse) pode ser observada mais do que uma vez em cada rato.

De acordo com Gail *et al.* (1980) foram observados 253 tempos de ocorrência de novos tumores para os 48 ratos em estudo. O total de dias observados foi 35 dias havendo sempre pelo menos um rato com novos tumores em cada um desses dias. Note-se que os tempos de ocorrência de novos tumores são encarados como medidas repetidas dos ratos e que a observação destes tempos é feita como um estudo de seguimento (*follow up*). Neste caso, apesar de haver perda de informação motivada pelo agrupamento dos dados, isto pode não afetar a questão mais importante neste cenário que é estimar a intensidade da ocorrência de novos tumores de mama.

Dados de sobrevivência com eventos múltiplos são aqui considerados como dados multivariados, visto que cada rato funciona com um bloco. Por conseguinte, um termo de fragilidade (efeito aleatório) é introduzido na função de risco ou intensidade do processos de contagem (Andersen *et al.*, 1993) de cada um dos 48 ratos levando em conta a sua heterogeneidade não observada ou a dependência entre os tempos de “sobrevivência” de um mesmo rato. Duas variáveis indicadoras z_1 e z_2 foram definidas para a única covariável observada, grupo de tratamento, designando z_1 o grupo controlo e z_2 o grupo de tratamento. Isso permite avaliar o efeito individual de cada grupo de ratos na intensidade de ocorrência de novos tumores ao longo do tempo. Neste cenário, optou-se por usar os modelos aditivos com fragilidade (Rocha, 1996) para modelar a intensidade do processo de contagem $N_i(t)$ associado ao i -ésimo rato em estudo, sob a perspectiva bayesiana (Silva e Amaral Turkman, 2004), dada por

$$I_i(t|\mathbf{w}) = Y_i(t)[z_{i1} \alpha_1(t) + z_{i2} \alpha_2(t) + w_{i1}], \quad (1)$$

onde $Y_i(t)$ indica se o rato i está em risco no instante t , $\mathbf{w} = (w_1, \dots, w_k)^T$ é o vetor de termos de fragilidade e z_{ij} é o valor da variável indicadora z_j para o rato i , $i = 1, \dots, n$, $j = 1, 2$. Observe-se que: i) na atual parametrização do modelo de fragilidade (1) a função risco subjacente é excluída, evitando assim uma sobreparametrização do modelo; ii) se omite um segundo índice para as n_i observações do rato i , por simplicidade de notação; iii) $Y_i(t) = 1$ quase sempre uma vez que o evento de interesse não é fatal.

Para os parâmetros de regressão no modelo de fragilidade (1), consideram-se usualmente processos *a priori* gama com incrementos independentes nos intervalos $B_j = (t_{j-1}, t_j]$, $j = 1, \dots, m$, para as versões discretizadas das funções de regressão cumulativas $\Omega_q(t) = \int_0^t \alpha(u) du$, $q = 1, 2$, denotadas por $\Omega_q = (\Omega_{q1}, \dots, \Omega_{qm})^T$, cuja distribuição *a priori* é dada por

$$\phi(\Omega_q) \propto \prod_{j=1}^m \frac{c_q^{\Omega_{qj}^*}}{\Gamma(c_q \Omega_{qj}^*)} \Omega_{qj}^{c_q \Omega_{qj}^* - 1} \exp(-c_q \Omega_{qj}), \quad (2)$$

onde c_q é interpretado como uma medida de precisão da conjectura inicial $\Omega_q^*(t)$ da função cumulativa $\Omega_q(t)$ e $\Omega_{qj}^* = \Omega_q^*(t_j) - \Omega_q^*(t_{j-1})$, $j = 1, \dots, m$, são as respetivas conjecturas para as componentes de Ω_q , $q = 1, 2$. Note-se que a média e a variância de Ω_{qj} , $j = 1, \dots, m$, são dadas, respetivamente, por $E(\Omega_{qj}) = \Omega_{qj}^*$ e $Var(\Omega_{qj}) = \Omega_{qj}^*/c_q$. Por vezes, assume-se que $\Omega_{qj}^* = r_q dt_j$, onde r_q é um valor proposto para a função $\alpha_q(t)$ por unidade de tempo e $dt_j = t_j - t_{j-1}$, $q = 1, 2$, $j = 1, \dots, m$. Neste caso, utilizaram-se $c_1 = c_2 = 0.001$ e $m = 35$ intervalos disjuntos B_j (35 dias de observação).

Os termos de fragilidade são supostos serem independentes e identicamente distribuídos segundo uma distribuição exponencial com hiperparâmetro de fragilidade δ (raiz quadrada da precisão dos termos de fragilidade). Portanto, a distribuição *a posteriori* do modelo (1), com processos *a priori* gama (2) e distribuição de fragilidade exponencial, denotada por $\pi(\Omega, \delta, \mathbf{w}|\mathcal{D})$, é proporcional a

$$\prod_{j=1}^m \prod_{i \in R_j} \left(I_{ij}^{N_{ij}} \exp(-I_{ij}) \right) \times \prod_{j=1}^m \prod_{q=1}^2 \left(\Omega_{qj}^{c_q r_q dt_j - 1} \exp(-c_q \Omega_{qj}) \right) \times \prod_{i=1}^n \delta \exp(-\delta w_i) \times \pi(\delta), \quad (3)$$

onde $I_{ij} \equiv I_i(t_j)dt_j = Y_{ij}(\mathbf{z}_i^T \Omega_j + w_{l_i} dt_j)$, $Y_{ij} \equiv Y_i(t_j)$, $N_{ij} \equiv dN_i(t_j)$, $dt_j = t_j - t_{j-1}$, $\Omega = (\Omega_1^T, \dots, \Omega_m^T)^T$ com $\Omega_j = (\Omega_{1j}, \Omega_{2j})^T$, $\mathbf{w} = (w_1, \dots, w_k)^T$, \mathcal{D} é o conjunto de dados, R_j é o conjunto de risco no intervalo B_j , $i = 1, \dots, n$, $j = 1, \dots, m$, $q = 1, 2$, $l_i \in \{1, \dots, k\}$ e $\pi(\delta)$ é a distribuição *a priori* do hiperparâmetro δ .

Neste contexto foram ajustados quatro modelos aditivos com fragilidade variando as constantes r_q , $q = 1, 2$, dos processos *a priori* e os parâmetros de forma ν e de escala η da distribuição gama *a priori* para o hiperparâmetro δ . Estes modelos são denotados por i) modelo M_1 com $r_1 = r_2 = 0.1$ e $\nu = \eta = 0.001$; ii) modelo M_2 com $r_1 = r_2 = 0.1$, $\nu = 6$ e $\eta = 1$; iii) modelo M_3 com $r_1 = r_2 = 1$ e $\nu = \eta = 0.001$; iv) modelo M_4 com $r_1 = r_2 = 1$, $\nu = 6$ e $\eta = 1$. Nos modelos M_1 e M_2 a conjectura inicial r_q para os parâmetros de regressão $\alpha_q(t)$ é menor do que nos modelos M_3 e M_4 , $q = 1, 2$, enquanto que ao considerar uma distribuição gama com parâmetros 6 e 1 os modelos M_2 e M_4 assumem uma maior heterogeneidade *a priori* entre os ratos do que os modelos M_1 e M_3 .

Os modelos de fragilidade associados à distribuição *a posteriori* (3) foram ajustados no *software* WinBUGS (Lunn *et al.*, 2000), onde também foram calculadas as estimativas das ordenadas preditivas condicionais

$$\pi(y_i | \mathcal{D}_{-i}) = \int \pi(y_i | \theta, \mathcal{D}_{-i}) \pi(\theta | \mathcal{D}_{-i}) d\theta, \quad (4)$$

onde \mathcal{D}_{-i} é o conjunto de dados sem a sua i -ésima observação, denotada aqui por y_i , e $\pi(y_i | \theta, \mathcal{D}_{-i})$ é a densidade preditiva de y_i , sendo igual a $\pi(y_i | \theta)$ se as observações forem condicionalmente independentes dado um vetor de parâmetros θ , $i = 1, \dots, n$. Estas estimativas foram usadas posteriormente para obter os valores *a posteriori* da medida de validação cruzada, conhecida por log-verosimilhança de validação cruzada negativa, $LCPO = -\sum_{i=1}^n \log \pi(y_i | \mathcal{D}_{-i})$, para os modelos M_1, M_2, M_3 e M_4 , dados por 703.5, 721.5, 706.4 e 725.3, respetivamente. Baseando-se nestes valores, M_1 é o modelo de fragilidade selecionado, uma vez que esse possui o menor valor de $LCPO$. Já agora, as séries simuladas para estes modelos não apresentaram problemas de convergência segundo os métodos de diagnóstico implementados na rotina R BOA (Smith, 2007).

Para o modelo de fragilidade M_1 foram calculadas as estimativas das funções de regressão cumulativas, bem como as estimativas de $\alpha_q(t)$, $q = 1, 2$, usando o método de kernel descrito em Silva e Amaral Turkman, (2004) com largura de banda $b_q = 7$ dias. Estas estimativas encontram-se na Figura 1 com a função de regressão cumulativa dos grupos controlo $\Omega_1(t)$ e tratamento $\Omega_2(t)$ e o respetivo intervalo de credibilidade a 95% nas Figuras 1(a) e 1(b), bem como as suas funções de regressão $\alpha_1(t)$ e $\alpha_2(t)$ na Figura 1(c), respetivamente. Nestas três figuras não fica claro a influência dos grupos de tratamento ao longo do tempo. Porém, verifica-se que a intensidade da ocorrência de novos tumores é quase sempre maior nos ratos do grupo controlo do que nos ratos tratados com *retinoid prophylaxis*.

Por fim, o grau de heterogeneidade não observada entre os ratos é medido aqui pelo desvio padrão da fragilidade, denotado por $\sigma_{fr} = 1/\delta$, onde δ é o hiperparâmetro de fragilidade. Para o modelo de fragilidade M_1 , a média *a posteriori* e o desvio padrão de σ_{fr} são, respetivamente, dados por 0.037 e 0.0061 com respetivo intervalo de credibilidade a 95% igual a (0.0270, 0.0509), indicando assim pouca heterogeneidade entre os 48 ratos em estudo.

3 Aplicação em Estatística espacial

Silva *et al.* (2005) analisaram o número de pacientes, acima dos 25 anos, admitidos nas emergências hospitalares com enfarte do miocárdio em 139 regiões de saúde da província de Quebeque entre 1996 a 1999. Os pacientes foram agrupados por género e, para cada região de saúde, anotaram-se os valores de dois índices sócio-económicos: índice de privação material (MDI, *Material Deprivation Index*) e índice de privação social (SDI, *Social Deprivation Index*). A privação pode ser interpretada como uma desvantagem da região local ou sociedade em geral à qual o indivíduo pertence. Privação material envolve a condição resultante do não usufruto de bens ou conveniências associados ao estilo de vida

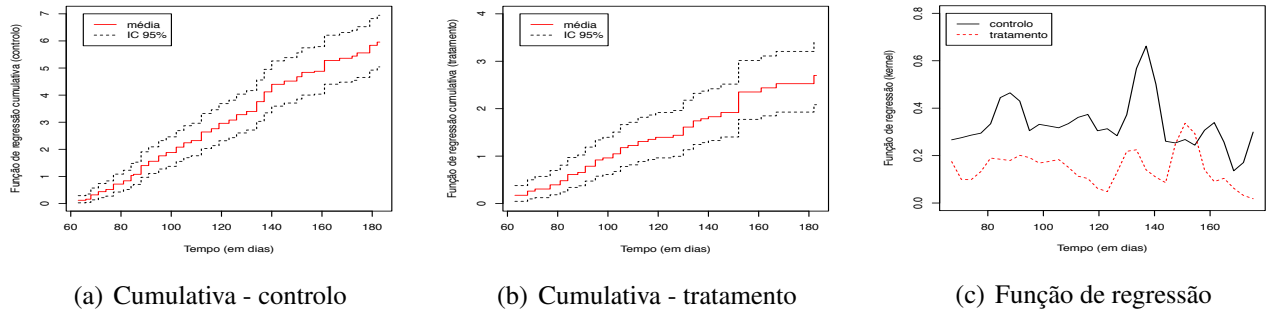


Figura 1: Funções de regressão (cumulativa ou kernel) no modelo M_1

moderno, sendo frequentemente encarado como pobreza. Privação social refere-se ao grau de ligação social entre indivíduos, tais como isolamento e coesão na família ou local de trabalho.

Este contexto aplicado induz ao ajuste de modelos de regressão de Poisson espacial, onde os dados de admissão hospitalar são dados espaciais de área ou *lattice data*, visto que o seu conjunto de localizações ($\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^r$) é formado por 139 regiões locais de saúde, com fronteiras bem definidas. Em Estatística Espacial, \mathcal{S} pode ser um subconjunto fixo de \mathbb{R}^r que contém um retângulo r -dimensional de volume positivo, e.g., $\mathbf{s} = (x, y) \in \mathbb{R}^2$, onde x e y são as coordenadas de latitude e longitude de uma localização geográfica (dados geoestatísticos), bem como um subconjunto aleatório de \mathbb{R}^r ou um processo pontual em \mathbb{R}^r (dados de padrão pontual).

Seja Y_i o número de admissões nas emergências hospitalares com enfarte do miocárdio (“sucesso”) durante um certo período numa região i e n_i o tamanho da sua população em risco, $i = 1, \dots, r$. Supondo independência entre as regiões e θ_i probabilidade de sucesso para qualquer indivíduo da região i , pode-se considerar $Y_i \sim \text{Binomial}(n_i, \theta_i)$. Todavia, em estudos epidemiológicos com doenças raras e/ou regiões pequenas e/ou períodos curtos de observação dos indivíduos, pode-se aproximar a distribuição Binomial pela distribuição Poisson, i.e., $Y_i \sim \text{Poisson}(\mu_i = n_i \theta_i = E_i \eta_i)$, onde E_i é o total esperado da população em risco e η_i é o risco relativo da ocorrência do evento para a região i , estimado tradicionalmente por $\hat{\eta}_i = y_i/n_i$ (*SMR, Standardized Mortality Ratio*), $i = 1, \dots, r$. Observando p covariáveis (z_1, \dots, z_p) suspeitas de influenciar a variável resposta Y_i , um modelo de Poisson com efeitos aleatórios espaciais para as admissões esperadas $\{\mu_i\}$ é expresso por

$$\log \mu_i = \log n_i + \beta_0 + \mathbf{z}_i^T \boldsymbol{\beta} + v_i, \quad (5)$$

onde β_0 é o risco relativo (log) global, $\boldsymbol{\beta}$ é o vetor parâmetros de regressão associado ao vetor de covariáveis $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})^T$ observado na região i , e v_i é efeito aleatório da região i , $i = 1, \dots, r$, $j = 1, \dots, p$.

Os efeitos aleatórios (fragilidades) em (5) levam em conta a eventual heterogeneidade não observada na região, sobretudo quando as covariáveis observadas não explicam completamente a incidência de uma doença. Estes podem ser não estruturados ($v_i \sim \text{Normal}(0, \sigma_v^2)$ independentes) e/ou correlacionados espacialmente. Besag *et al.* (1991) apresentaram um modelo de Poisson com efeitos aleatórios espacialmente correlacionados, $\{u_i\}$, considerando um modelo autorregressivo condicional intrínseco (ICAR, *intrinsic autoregressive*), i.e.,

$$\{u_i | u_j, j \neq i, \sigma_u^2\} \sim \text{Normal}(\bar{u}_i, \sigma_i^2), \quad (6)$$

onde $\bar{u}_i = (\sum_j u_j w_{ij}) / (\sum_j w_{ij})$, $\sigma_i^2 = \sigma_u^2 / (\sum_j w_{ij})$ e $w_{ij} = 1$, se i e j são regiões adjacentes, e $w_{ij} = 0$, no caso contrário, $i, j = 1, \dots, r$. Note-se que σ_u^2 e σ_v^2 controlam a variabilidade global dos $\{u_i\}$ e dos $\{v_i\}$, respectivamente. Note-se que, a partir das estimativas dos riscos relativos $RR_i = \exp(\beta_0 + \sum_{j=1}^p z_{ij} \beta_j + u_i + v_i)$, $i = 1, \dots, r$, se podem construir mapas com as diferentes classes de riscos relativos, usando os seus quantis como limites de classes. Além disso, são aqui supostas distribuições *a priori* vagas mas próprias, quer para os coeficientes de regressão ($\beta_j \sim \text{Normal}(0, 10^5)$, $j = 0, 1, \dots, p$), quer para as

variâncias dos efeitos aleatórios, especificamente, σ_v^2 e σ_u^2 seguem uma distribuição gama invertida com média 1 e variância 100 (grande). Para mais detalhes sobre modelos espaciais veja-se, *e.g.*, Banerjee *et al.* (2004) ou Silva e Dean (2006).

Modelos de Poisson espaciais, definidos em (5), com modelo ICAR (6), $r=139$ e $p=2$, foram ajustados aos dados em causa para homens e mulheres e em 1997, somente por simplicidade. No ajustamento dos modelos, usou-se o *software* WinBugs (Lunn *et al.*, 2000) após 10000 iterações de simulação e 5000 iterações do período de aquecimento *burn-in*. As amostras simuladas foram analisadas usando algumas técnicas de convergência implementadas no rotina R BOA (Smith, 2004), não se detetando nenhuma anomalia grave. Os resultados da análise desses modelos encontram-se na Tabela 1, que apresenta médias, desvios padrões (d.p.) e intervalos de credibilidade a 95% para as quantidades de interesse: β_0 (termo de ordenada na origem), β_1 e β_2 (coeficientes de regressão associados às covariáveis MDI e SDI), σ_u e σ_v (desvios padrões dos efeitos aleatórios espacialmente correlacionados $\{u_i\}$ e não estruturados $\{v_i\}$), bem como a medida de comparação de modelos conhecida por critério de informação da desviância (DIC, *Deviance Information Criterion*). Esta medida lida com modelos bayesianos hierárquicos de qualquer grau de complexidade, sendo calculada como a soma de dois componentes: a desviância *a posteriori* esperada, medindo a bondade do ajustamento (*goodness-of-fit*), e o número efetivo de parâmetros, medindo a complexidade do modelo, sendo expressa frequentemente por

$$DIC = 2\overline{D(\theta)} - D(\bar{\theta}), \quad (7)$$

onde $\overline{D(\theta)}$ e $\bar{\theta}$ denotam as médias *a posteriori* da desviância ($D(\theta)$, *Deviance*) e dos parâmetros do modelo (θ), respetivamente.

Tabela 1: Estimativas dos parâmetros de modelos espaciais por género em 1997.

parâm.	Homens (DIC = 1016.54)			Mulheres (DIC = 930.14)		
	média	d.p.	I.C. (95%)	média	d.p.	I.C. (95%)
β_0	-5.461	0.043	(-5.54,-5.38)	-6.31	0.058	(-6.42,-6.20)
β_1	0.003	0.106	(-0.21,0.21)	0.295	0.143	(0.01,0.57)
β_2	-0.065	0.133	(-0.32,0.20)	0.372	0.174	(0.03,0.71)
σ_u	0.231	0.036	(0.16,0.30)	0.302	0.051	(0.21,0.41)
σ_v	0.139	0.019	(0.10,0.18)	0.188	0.028	(0.13,0.24)

Com base nos resultados apresentados na Tabela 1, pode-se concluir que: i) os índices sócio-económicos MDI e SDI não influenciam os totais de homens admitidos com enfarte de miocárdio nas regiões em 1997, enquanto para as mulheres eles são influentes; ii) as estimativas da componente de variância espacialmente correlacionada (σ_u) são quase o dobro das respetivas estimativas daquela não estruturada (σ_v), indicando que a heterogeneidade não observada da admissão de pacientes com enfarte de miocárdio se deve predominantemente ao efeito espacial nas 139 regiões de saúde de Quebeque em 1997.

Na Figura 2 encontram-se os mapas dos riscos relativos (RR) para homens e mulheres em 1997. Os homens em 1997 apresentaram maior risco (relativo) de enfarte do miocárdio nas regiões com maior densidade populacional (sul de Quebeque), enquanto os maiores riscos relativos das mulheres nesse ano espalham-se quer nessas regiões quer em outras regiões menos densas em população (norte de Quebeque). A região da cidade de Montreal encontra-se em destaque nos mapas por ser a região com mais habitantes na província de Quebeque. Por fim, recorde-se que na construção desses mapas não se usou nenhuma medida de variabilidade, sendo aconselhável alguma prudência na interpretação dos mesmos.

Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação para a Ciência e a Tecnologia, projeto Pest-OE/MAT/UI0006/2011.

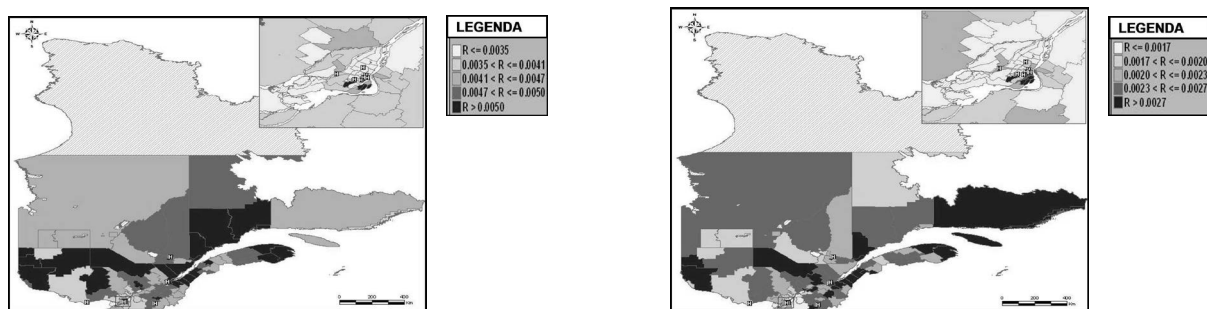


Figura 2: Mapa dos riscos relativos (RR) de homens (esquerda) e mulheres (direita) em 1997.

Bibliografia

- Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N. (1993). *Statistical Models based on Counting Processes*. Springer-Verlag, New York.
- Banerjee, S., Carlin, B.P., Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC.
- Besag, J., York, J., Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1–59.
- Chen, M.-H., Shao, Q.-H., Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.
- Congdon, P. (2006). *Bayesian Statistical Modelling*, 2nd edition. Wiley.
- Cowles, M.K. (2013). *Applied Bayesian Statistics (with R and OpenBUGS examples)*. Springer.
- Gail, M.H., Santner, T.J., Brown, C.C. (1980). An analysis of comparative carcinogenesis experiments with multiple times to tumour. *Biometrics*, 36, 255–266.
- Gelfand, A.E., Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85 (410), 398–409.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2003). *Bayesian Data Analysis*, 2nd edition. Chapman & Hall/CRC.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337.
- Lunn, D.J., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D. (2012). *The BUGS Book - A Practical Introduction to Bayesian Analysis*. CRC Press/Chapman & Hall.
- Ntzoufras, I. (2009). *Bayesian Modeling using WinBUGS*. Wiley.
- Paulino, C.D., Amaral Turkman, M.A., Murteira, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.
- Rocha, C.S. (1996). Survival Models for Heterogeneity Using the Non-Central Chi-Squared Distribution with Zero Degrees of Freedom. In Jewell, N. P.; Kimber, A. C.; Lee, M. T.; Whitmore, G. A. editors, *Lifetime Data: Models in Reliability and Survival Analysis*, pp. 275–279, Kluwer Academic Publishers.

- Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society B*, 71, 319–392.
- Silva, G.L., Amaral Turkman, M.A. (2004). Bayesian analysis of an additive survival model with frailty. *Communications in Statistics – Theory and Methods*, 33, 2517–2533.
- Silva, G.L., Dean, C.B., Courteau, J., Niyonsenga, T., Vanasse, A. (2005). Análise espacial de admissão de pacientes com enfarte do miocárdio em Quebeque. Em Carlos A. Braumann, C.A., Infante, P., Oliveira, M.M., Alpízar-Jara, R. e Rosado, F. (Editores), *Estatística Jubilar, Actas do XII Congresso Anual da SPE, Edições SPE*, pg. 717-722, Lisboa.
- Silva, G.L., Dean, C.B. (2006). *Uma Introdução à Análise de Modelos Espaço-temporais para Taxas, Proporções e Processos de Multi-estados*. Associação Brasileira de Estatística, Caxambú.
- Smith, B.J. (2007). *BOA User Manual* (version 1.1.6). Department of Biostatistics, College of Public Health, University of Iowa.

Adenda: Este texto foi redigido segundo as normas do AO 1990.



Método de Monte Carlo via Cadeias de Markov com saltos reversíveis

Isabel Pereira, *isabel.pereira@ua.pt*

CIDMA-Centro de Investigação e Desenvolvimento em Matemática e Aplicações
Departamento de Matemática da Universidade de Aveiro

1 Introdução

Os métodos de Monte Carlo via cadeias de Markov, referidos simplesmente como MCMC, constituem um dos mais notáveis avanços na estatística nos últimos 20 anos. As suas origens remontam aos trabalhos de Metropolis *et al* (1953) e Hastings (1970). No entanto, foi apenas nos finais dos anos oitenta que esta ideia começou a ser alvo de investigação. Inicialmente os métodos MCMC foram usados na estatística bayesiana, no entanto a sua utilização é mais alargada podendo ser usada na dita estatística clássica.

De uma forma geral aplicam-se os algoritmos MCMC para obter amostras da distribuição *a posteriori* dos parâmetros do modelo. Qual o procedimento a usar para comparar diferentes modelos considerando um determinado conjunto de dados? Na estatística clássica os modelos podem ser comparados usando critérios de informação, em particular o AIC (*Akaike's information criterion*) ou o BIC (*Bayesian information criterion*). Existe uma série de aproximações bayesianas para fazer o ajustamento e seleção de modelos, destacando-se o DIC (*Deviance information criterion*) e os Fatores de Bayes (ver Spiegelhalter *et al*, 2002). No entanto o algoritmo de Monte Carlo via cadeias de Markov com saltos reversíveis - RJMCMC (*reversible Jump Monte Carlo Markov Chain*)- introduzido por Green (1995), permite tratar simultaneamente a questão da seleção do modelo e da estimação dos parâmetros. Para tal contribui o facto de ser possível a variação da dimensão do espaço do parâmetro entre as diferentes iterações da cadeia de Markov. Assim, pode ser visto como uma extensão do algoritmo de Metropolis-Hastings para um espaço de estados mais geral.

2 Como funciona o RJMCMC?

O algoritmo RJMCMC move-se entre espaços de parâmetros de uma coleção de modelos, em geral em número finito e na maioria das situações são modelos encaixados. Assim o método é uma combinação do algoritmo MCMC para um dado modelo (fixo) com o passo adicional que permite a movimentação entre diferentes modelos. Por uma questão de simplificação considera-se a situação de os modelos estarem encaixados, que é a situação mais frequentemente usada. Isto significa incluir ou retirar parâmetros do modelo atual. Suponham-se K modelos possíveis, θ_k representa o conjunto de parâmetros do k -ésimo modelo, $k = 1, \dots, K$, i.e., de M_k .

- Algoritmo RJMCMC, com N iterações

1. Escolher um modelo inicial M_k , i.e., inicializar k e θ_k , na iteração $t = 1$.
2. Para cada iteração $t \geq 1$ fazer

- (a) movimento dentro do modelo: para o modelo fixo M_k - obtido no fim da iteração $t - 1$ - atualizar os parâmetros $\boldsymbol{\theta}_k$, de acordo com a metodologia MCMC;
- (b) movimento entre modelos: propôr a mudança para um novo modelo $M_{k'}$, calcular a probabilidade de aceitação para o movimento e decidir pela aceitação ou rejeição do novo modelo.

3. Incrementar a iteração $t = t + 1$; se $t < N$ voltar a 2.

- Movimento entre modelos

Veja-se o seguinte exemplo apresentado por Green (1995):

Suponha-se que se pretende fazer o movimento de um modelo 2 de parâmetros (α, β) para o modelo 1 com um parâmetro. Esse parâmetro poderia ser estabelecido como $\lambda = \frac{\alpha + \beta}{2}$. No entanto, dado o modelo 1 de parâmetro λ que valores se deveriam escolher para (α, β) no movimento contrário? Existe uma infinidade de possíveis escolhas para (α, β) que conduzem a $\lambda = \frac{\alpha + \beta}{2}$. É pois necessária uma bijeção entre os parâmetros de ambos os modelos e tal só é possível se o número de parâmetros for o mesmo. Para tal adiciona-se uma variável auxiliar u ao modelo 1, que não tenha outro papel senão o de permitir que haja uma bijeção entre os dois modelos. Seja $\alpha = \lambda + u$ e $\beta = \lambda - u$; então $\lambda = \frac{\alpha + \beta}{2}$ e $u = \frac{\alpha - \beta}{2}$.

- Probabilidade de aceitação do movimento entre modelos

De acordo com o algoritmo de Metropolis-Hastings a probabilidade de aceitação de um movimento dentro do modelo, dado \mathbf{x} e parâmetro $\boldsymbol{\theta}$ é dada por:

$$\min \left\{ 1, \frac{\pi(\boldsymbol{\theta}'|\mathbf{x})}{\pi(\boldsymbol{\theta}|\mathbf{x})} \times \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} \right\},$$

onde $\pi(\cdot|\mathbf{x})$ representa a distribuição *a posteriori* e $q(\cdot|\cdot)$ a distribuição proponente para o parâmetro.

Admitam-se K modelos possíveis e $\boldsymbol{\theta}_k$ os parâmetros do modelo k , $k = 1, \dots, K$; suponha-se que se está no modelo M_k de parâmetro $\boldsymbol{\theta}_k$ e que se pretende fazer a mudança para o modelo $M_{k'}$ de parâmetro $\boldsymbol{\theta}'_{k'}$. Sejam \mathbf{u}_k as variáveis auxiliares associadas ao modelo k , $k = 1, \dots, K$, tais que para qualquer modelo k , $(\boldsymbol{\theta}_k, \mathbf{u})$ tem sempre o mesmo número de elementos, por exemplo L . Para $1 \leq k, k' \leq K$, seja $g_{k \rightarrow k'}$ uma função invertível tal que $(\boldsymbol{\theta}'_{k'}, \mathbf{u}') = g_{k \rightarrow k'}(\boldsymbol{\theta}_k, \mathbf{u})$ com $g_{k' \rightarrow k} = g_{k \rightarrow k'}^{-1}$ e $r_k(\mathbf{u})$ a proponente para as variáveis auxiliares \mathbf{u} . A probabilidade de aceitação do movimento do modelo k com parâmetro $\boldsymbol{\theta}_k$, para o modelo k' com parâmetro $\boldsymbol{\theta}'_{k'}$, é

$$\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}'_{k'}) = \min \left\{ 1, \frac{\pi(k', \boldsymbol{\theta}'_{k'}|\mathbf{x})}{\pi(k, \boldsymbol{\theta}_k|\mathbf{x})} \times \frac{q(k'|k)}{q(k|k')} \times \frac{r_{k'}(\mathbf{u}')}{r_k(\mathbf{u})} \times J \right\},$$

onde $\pi(k', \boldsymbol{\theta}'_{k'}|\mathbf{x})$ é a distribuição *a posteriori* dos parâmetros do modelo $M_{k'}$, $q(k'|k)$ é a probabilidade do movimento $k \rightarrow k'$, $r_k(\mathbf{u})$ a distribuição proponente para as variáveis auxiliares \mathbf{u} e J o jacobiano da transformação do movimento descrito.

É necessário calcular o jacobiano da transformação sempre que se faz um movimento entre modelos com diferentes parâmetros. Este jacobiano é o determinante da matriz V de dimensão $L \times L$ definida por:

$$V = \left[\frac{\partial g_{k \rightarrow k'}(\boldsymbol{\theta}_k, \mathbf{u})}{\partial (\boldsymbol{\theta}_k, \mathbf{u})} \right].$$

Evidentemente que o movimento inverso de $M_{k'}$ para M_k é feito deterministicamente e é aceite com probabilidade

$$\alpha(\boldsymbol{\theta}'_{k'}, \boldsymbol{\theta}_k) = \left(\frac{1}{\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}'_{k'})} \right)^{-1}.$$

Admita-se que o novo modelo a propor $M_{k'}$ tem dimensão superior, i.e., $n_{k'} > n_k$. Seguem-se os seguintes passos:

- propôr o movimento $M_k \rightarrow M_{k'}$ com probabilidade $q(k'|k)$;
- gerar vetor \mathbf{u} , $n_{k'} - n_k$ dimensional, a partir de $r_k(\mathbf{u})$;
- fazer $(\boldsymbol{\theta}'_{k'}, \mathbf{u}') = g_{k \rightarrow k'}(\boldsymbol{\theta}_k, \mathbf{u})$;
- aceitar $(\boldsymbol{\theta}'_{k'}, \mathbf{u}')$ com probabilidade $\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}'_{k'})$.

3 Exemplo: Problema de mudança na localização

Considere-se o conjunto de dados obtidos por Jarrett (1979) que traduzem o número de acidentes em minas de carvão por ano, no período de 15 de março de 1851 a 22 de março de 1962. Os dados sugerem que houve uma redução no número de acidentes, por ano nesse período e que houve uma mudança no valor médio da distribuição, por volta do ano 1891.

Carlin et al (1992) sugeriram o modelo

$$\begin{aligned} X_i &\sim Po(\theta), & i = 1, \dots, k \\ X_i &\sim Po(\lambda), & i = k + 1, \dots, n (= 112). \end{aligned}$$

Este é tipicamente um problema de mudança de localização, i.e., existe uma mudança de comportamento em termos de localização, num instante desconhecido. Veja-se como é que o RJMCMC pode ser usado para analisar este problema:

Modelo 1:

$$\begin{aligned} X_i &\sim Po(\lambda), & i = 1, \dots, n \\ \text{apriori : } &\lambda \sim Gama(a, b) \end{aligned}$$

Modelo 2:

Existe $k \in \{1, 2, \dots, n - 1\}$ tal que

$$\begin{aligned} X_i &\sim Po(\theta_1), & i = 1, \dots, k \\ X_i &\sim Po(\theta_2), & i = k + 1, \dots, n \end{aligned}$$

$$\text{a priori's: } \theta_i \sim Gama(a, b), i = 1, 2; \quad k \sim U\{1, 2, \dots, n - 1\}.$$

A atualização de parâmetros em cada um dos modelos faz-se usando as condicionais completas, correspondendo ao movimento dentro do modelo. São as seguintes para cada um dos modelos:

Modelo 1:

$$\lambda | \mathbf{x} \sim Gama\left(a + \sum_{i=1}^n x_i, b + n\right).$$

Modelo 2:

$$\theta_1 | \mathbf{x}, \theta_2, k \sim \text{Gama} \left(a + \sum_{i=1}^k x_i, b + n \right);$$

$$\theta_2 | \mathbf{x}, \theta_1, k \sim \text{Gama} \left(a + \sum_{i=k+1}^n x_i, b + n \right);$$

$$p(k | \mathbf{x}, \theta_1, \theta_2) = \frac{e^{k(\theta_2 - \theta_1)} \left(\frac{\theta_1}{\theta_2} \right)^{\sum_{i=1}^k x_i}}{\sum_{j=1}^n e^{j(\theta_2 - \theta_1)} \left(\frac{\theta_1}{\theta_2} \right)^{\sum_{i=1}^j x_i}}.$$

Relativamente ao movimento entre modelos: suponha-se a passagem do modelo 1 para o modelo 2. Uma vez que o 2º modelo tem dois parâmetros a mais que o 1º são necessárias duas variáveis auxiliares. Sejam u_1 e u_2 tais que:

- $u_1 \sim N(0, \sigma^2)$
- $u_2 \sim U\{1, 2, \dots, n - 1\}$.

Usando a transformação bijetiva entre (λ, u_1, u_2) e (θ_1, θ_2, k) definida por

$$\begin{cases} \theta_1 = \lambda e^{u_1} \\ \theta_2 = \lambda e^{-u_1} \\ k = u_2 \end{cases},$$

tem-se reciprocamente

$$\begin{cases} \lambda = \frac{\sqrt{\theta_1 \theta_2}}{e^{u_1}} \\ \theta_1 = \frac{1}{2} \log(\theta_1 / \theta_2) \\ u_2 = k \end{cases}.$$

Para calcular a probabilidade de aceitação do movimento do modelo 1 para o modelo 2 determinam-se os seguintes quocientes

q.veros:

$$\frac{L(\theta_1, \theta_2, k | \mathbf{x}, M_2)}{L(\lambda | \mathbf{x}, M_1)} = \left(\frac{\theta_1}{\lambda} \right)^{\sum_{i=1}^k x_i} e^{-k(\theta_1 - \lambda)} \left(\frac{\theta_2}{\lambda} \right)^{\sum_{i=k+1}^n x_i} e^{(n-k)(\theta_2 - \lambda)};$$

q.prioris:

$$\frac{\pi(\theta_1, \theta_2, k | M_2)}{\pi(\lambda | M_1)} = \frac{b^a}{\gamma(a)} (\lambda \theta_1 \theta_2)^{a-1} e^{-b(\theta_1 + \theta_2 - \lambda)};$$

q.propon: como o movimento do modelo 2 para o modelo 1 é determinístico, apenas se tem

$$\frac{q(1|2)}{q(2|1)} = \left\{ \frac{1}{n-1} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{u_1}{2\sigma^2}\right) \right\}^{-1}.$$

jacobiano:

$$J = \begin{vmatrix} e^{u_1} & \lambda e^{u_1} & 0 \\ e^{-u_1} & -\lambda e^{-u_1} & 0 \\ 0 & 0 & 1 \end{vmatrix} = 2\lambda.$$

Probabilidade de aceitação 1 → 2:

$$\alpha(1 \rightarrow 2) = \min \{1, q.\text{veros} \times q.\text{prioris} \times q.\text{propon} \times \text{jacobiano}\}.$$

4 Áreas de aplicação

Algumas áreas onde mais se tem destacado a metodologia RJMCMC:

- Modelos com mudança na localização

Uma das primeiras aplicações do RJMCMC foi em problemas de mudança de localização bayesianos, considerando desconhecidos *a priori* o seu número e a sua localização. Green (1995) analisou também o número de acidentes de minas usando um processo de Poisson considerando os dados medidos em cada dia em vez de ser por ano, em que a intensidade do processo é definida através de um função em escada com um número desconhecido de saltos e com localização também desconhecida. Fan e Brooks (2000) aplicaram este método de saltos reversíveis para modelar a forma de túmulos pré-históricos, em que a curvatura da abóbada muda um número desconhecido de vezes.

- Modelos de mistura finita

Considerando os dados \mathbf{x} , o vetor parâmetro $\boldsymbol{\theta}_k = (\phi_1, \dots, \phi_k)$ e k componentes, os modelos de mistura são definidos por

$$f(\mathbf{x}|\boldsymbol{\theta}_k) = \sum_{i=1}^k w_i f_i(\mathbf{x}|\phi_i),$$

onde w_i representa o peso da i -ésima componente de mistura f_i e tal que $\sum_{i=1}^k w_i = 1$. O número das componentes de mistura também é desconhecido. Richardson e Green (1997) analisaram dados, de natureza multimodal, referentes à atividade enzimática no sangue de 245 indivíduos, usando um modelo de mistura de normais com o objetivo de identificar subgrupos de indivíduos com velocidade elevada ou fraca de metabolização. Tadesse et al (2005) estenderam este estudo tratando o problema da agregação em dados de natureza multidimensional (dados de *microarrays* referentes ao estudo de cancro endometrial), usando um modelo de normais multivariadas com um número desconhecido de componentes e selecionando as variáveis que melhor discriminam os diferentes grupos.

- Seleção de variáveis

O problema de seleção de variáveis surge quando se pretende modelar a relação entre a variável resposta Y e p potenciais variáveis explicativas x_1, \dots, x_p . O problema fundamental na construção do modelo de regressão múltipla é o da seleção dos preditores a incluir. Considere-se, por exemplo, um modelo de regressão com erros gaussianos

$$Y = X_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I),$$

onde $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)$ é um vetor binário que indexa o subconjunto de variáveis x_1, \dots, x_p a incluir no modelo linear, $X_{\boldsymbol{\gamma}}$ é a matriz de delineamento cujas colunas correspondem ao subconjunto indexado por $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ é o correspondente subconjunto de coeficientes de regressão. Realçam-se os trabalhos de Smith e Kohn (1996) e de Nott e Leonte (2004).

- Abordagem não paramétrica bayesiana

A metodologia RJMCMC tem sido usada com sucesso na estatística não paramétrica com o objetivo de automatizar o processo de seleção dos nós, o seu número e a sua localização, quando se usa o modelo *spline* de ordem p para o ajustamento da curva. Para estimar

$$f(x) = \alpha_0 + \sum_{j=1}^p \alpha_j x^j + \sum_{i=1}^k \eta_i (x - k_i)_+^p, \quad x \in [a, b[, \quad w_+ = \max(o, w),$$

é necessário calcular o valor desconhecido de *knots* k , as suas localizações k_i e os correspondentes coeficientes de regressão α_i e η_j , $i = 0, \dots, p$ e $j = 1, \dots, k$.

Este procedimento envolve movimentos que adicionam, retiram ou recolocam os *knots* no modelo. Vejam-se os trabalhos de Denison *et al* (1998) e DiMatteo *et al* (2001).

- Modelos de séries temporais

Existem múltiplos modelos, lineares e não lineares, sugeridos na literatura para modelarem dados x_1, \dots, x_n temporalmente dependentes. O modelo autorregressivo de ordem k , encarada como desconhecida, foi tratado na literatura por diversos autores nomeadamente Elhers e Brooks (2003), Vermaak *et al* (2004) e os modelos limiares foram tratados por Campbell (2004). Pode-se destacar o trabalho de Enciso-Mora (2008) aplicado a modelos de contagem usando o operador *thinning* binomial, designados por INARMA (Integer AutoRegressive Moving Average), que usam o salto reversível conjugado com a ampliação dados, e onde em cada passo são aumentadas ou diminuídas as ordens anteriores apenas numa unidade e é mantida fixa a média do processo.

Uma das maiores dificuldades na implementação dos saltos reversíveis tem a ver com a arbitrariedade na escolha da função $g_{k \rightarrow k'}$ e das distribuições proponentes $r_k(\mathbf{u})$ para o vetor \mathbf{u} . O bom desempenho da metodologia está intimamente relacionada com a adequabilidade da escolha de $g_{k \rightarrow k'}$ e de $r_k(\mathbf{u})$, dependendo particularmente da forma e parâmetros de $r_k(\mathbf{u})$. Uma possibilidade explorada por alguns autores consiste na utilização do movimento de nascimento e morte feito em simultâneo, mantendo fixa a dimensão do modelo; com o objetivo de reduzir a dimensionalidade/complexidade do problema faz-se a marginalização de componentes ou o aumento de dados/aumento do espaço de estados da distribuição *a posteriori* alvo. Para aprofundar esta metodologia sugere-se consultar Sisson (2005) e Fan e Sisson (2011).

Agradecimentos

Este trabalho foi parcialmente financiado por *fundos FEDER* através do *COMPETE*– Programa Operacional Factores de Competitividade e por fundos Portugueses através do *Centro de Investigação e Desenvolvimento em Matemática e Aplicações* e FCT- Fundação para a Ciência e a Tecnologia, dentro do projeto PEst-C/MAT/UI4106/2011 e FCOMP-01-0124-FEDER- 022690.

Referências

- [1] Campbell, E.P.(2004). Bayesian selection of threshold autoregressive models. *J. of Time Series Analysis* **25**, 467-482.
- [2] Carlin, B.P., Gelfand, A.E. e Smith, A.F.M. (1992) Hierarchical Bayesian Analysis of Change-point Problems. *J. of the Royal Statistical Society, C* **41**, 389-405.
- [3] Denison, D.G.T., Mallick, B.K. e Smith, A.F.M. (1998). Bayesian curve-fitting with free-knot splines. *J. of the Royal Statistical Society, B* **60**, 330-350.

- [4] DiMatteo, Genovese, C.R. e Kass, R.E. (2001). Automatic Bayesian curve fitting. *Biometrika* **88**, 1055-1071.
- [5] Enciso-Mora, V., Neal, P. e Subba Rao, T. (2008). Efficient order selection algorithms for integer-valued ARMA processes. *J. of Time Series Analysis* **30**, 1-18.
- [6] Ehlers, R.S. e Brooks, S.P. (2003). Constructing general efficient proposals for reversible jump MCMC. Technical Report, Department of Statistics, Federal University of Paraná.
- [7] Fan, Y. e Brooks, S.P. (2000). Bayesian modelling prehistoric corbelled domes. *J. of the Royal Statistical Society, D* **49**, 339-354.
- [8] Fan, Y. e Sisson, S.A. (2000). Reversible Jump MCMC. In Brooks, S., Gelman, A., Jones, C.L e Meng, X-L editors, *Handbook of Markov Chain Monte Carlo*, Taylor and Francis Group, USA, pp. 67 - 87.
- [9] Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.
- [10] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 59-109.
- [11] Jarrett, R.G. (1979). A note on the intervals between coal-mining disasters. *Biometrika* **66**, 191-1092.
- [12] Metropolis, N., Rosenbluth, A.H., Rosenbluth, M.N. e Teller, A.H. (1953). Equation of State Calculations by Fast Computing Machines. *J. of Chemical Physics* **21**, 1087-1092.
- [13] Nott, D.J. e Leonte, D. (2004). Sampling schemes for Bayesian variable selection in generalised linear models. *J. Computational and Graphical Statistics*. **13**, 362-382.
- [14] Richardson, S. e Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. of the Royal Statistical Society, B* **59**, 731-792.
- [15] Sisson, S.A. (2005). Trans-dimensional Markov chains: A decade of progress and future perspectives. *J. of American Statistical Association*. **100**, 1077-1089.
- [16] Smith, M e Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *J. of Econometrics* **75**, 317-344.
- [17] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. e van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. of the Royal Statistical Society, B* **64**, 583-639.
- [18] Tadesse, M., Sha, N. e Vannucci, M. (2005). Bayesian variable selection in clustering high-dimensional data. *J. of American Statistical Association* **100**, 602-617.
- [19] Vermaak, J., Andrieu, C., Doucet, A. e Godsill, S.J. (2004). Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *J. of Time Series Analysis* **25**, 785-809.



Métodos Computacionais: INLA, Integrated Nested Laplace Approximation

Isabel Natário¹, *icn@fct.unl.pt*

¹*CEAUL e Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*

1 Introdução

A estimação bayesiana passou a ser uma realidade com o desenvolvimento dos métodos e do poder computacionais, recorrendo a métodos aproximados ou, mais frequentemente, a métodos de simulação iterativos. Na verdade os modelos passaram a poder ser estimados por esta via mesmo que aumentados para um grau de complexidade, vantajoso em termos da modelação, mas tradicionalmente muito difíceis de estimar, como é o caso, por exemplo, de alguns modelos hierárquicos multidimensionais, modelos dinâmicos, modelos mistos, modelos espaço-temporais, modelos de processos de Cox log gaussianos, etc..

Neste contexto, têm dominado os métodos de simulação Monte Carlo via cadeias de Markov (MCMC), muito por causa de alguns softwares estatísticos que o implementam de forma quase automática, com pouca intervenção do utilizador - BUGS/WinBUGS, OpenBUGS, JAGS, ... No entanto, estes métodos, para modelos mais complicados com um grande número de parâmetros, possivelmente correlacionados, podem demorar muito tempo a fazer convergir a cadeia de Markov envolvida para a distribuição estacionária de interesse e a obter os resultados da estimação.

Muito recentemente contudo surgiu uma abordagem desenvolvida por “estatísticos computacionais” que utiliza, de forma inteligente, alguns métodos de aproximação já antigos para fazer a estimação bayesiana, a chamada Aproximação de Laplace Aninhada e Integrada, *INLA (Integrated Nested Laplace Approximation)* [3], que assenta na ideia de usar aproximações determinísticas das distribuições marginais a posteriori dos parâmetros de uma forma computacionalmente eficiente.

2 A abordagem INLA

A abordagem de estimação bayesiana INLA foi pensada para modelos que pertencem a uma subclasse especial de modelos de regressão aditivos estruturados, a classe de *modelos gaussianos latentes*. Esta classe caracteriza-se por uma variável resposta Y com distribuição na família exponencial, com valor esperado μ relacionado com um preditor linear aditivo η através de uma função ligação,

$$g(\mu) = \eta = \alpha + \sum_k \beta_k z_k + \sum_j f^{(j)}(u_j),$$

onde β_k são efeitos lineares das covariáveis \mathbf{z} e $f^{(j)}$ são funções das covariáveis \mathbf{u} . Para todos os parâmetros do preditor (as variáveis latentes α , $f^{(j)}$ e β_k) podem assumir-se distribuições a priori gaussianas, assumindo-se ainda que são condicionalmente independentes dois a dois, dados os restantes.

Esta classe de modelos é bastante mais vasta do que pode parecer à primeira vista e inclui várias aplicações para diferentes formas que as funções $f^{(j)}$ podem tomar, como se exemplifica a seguir:

- Modelos lineares generalizados: $f^{(j)}(\cdot) = 0$;
- Modelos lineares generalizados mistos, [1]: $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$;
- Modelos hierárquicos, [2]: $f^{(l)}(\cdot) \sim \mathcal{N}(\mathbf{0}, \sigma_f^2)$; $\sigma_f^2 \sim$ uma qualquer distribuição;
- Modelos espaciais, [5] e espaço-temporais, [4]:

$$\begin{aligned} f^{(1)}(\cdot) &\sim \text{CAR} \quad (\text{efeitos espaciais estruturados}); \\ f^{(2)}(\cdot) &\sim \mathcal{N}(\mathbf{0}, \sigma_{f^{(2)}}^2) \quad (\text{efeitos não espacialmente estruturados}); \\ f^{(3)}(\cdot) &\sim \text{AR}(\phi, \sigma_{f^{(3)}}^2) \quad (\text{efeitos temporais}); \\ f^{(4)}(\cdot) &\sim \mathcal{N}(\mathbf{0}, \sigma_{f^{(4)}}^2) \quad (\text{efeitos não temporalmente estruturados}). \end{aligned}$$

Seja $\mathbf{x} = (\alpha, \beta_k, f^{(j)})$ o vector de todas as variáveis latentes gaussianas, $\pi(\mathbf{x}|\boldsymbol{\theta}_1) \equiv \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_1))$, onde $\mathbf{Q}(\boldsymbol{\theta}_1)$ é a matriz de precisão. Denote \mathbf{y} o vector das observações com função densidade probabilidade $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_2)$, cujos elementos se assumem condicionalmente independentes, dados \mathbf{x} e $\boldsymbol{\theta}_2$; finalmente seja $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ o vector dos hiperparâmetros com densidade probabilidade $\pi(\boldsymbol{\theta})$ (não necessariamente gaussiana). Neste contexto, a distribuição a posteriori das quantidades desconhecidas é:

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_i \log(\pi(y_i|x_i, \boldsymbol{\theta})) \right\}.$$

A abordagem INLA recorre a alguns pressupostos (realísticos) que permite o uso de métodos numéricos, muito rápidos e eficientes, nomeadamente que \mathbf{x} (frequentemente de grandes dimensões, 100 – 100000 elementos) detém propriedades de independência condicional (*ie.* é um campo aleatório gaussiano) e que o número de hiperparâmetros é pequeno (digamos ≤ 6). A independência condicional traduz-se numa matriz de precisão para \mathbf{x} esparsa, com as simplificações óbvias que advêm para a computação, já que se verifica que $x_i \perp x_j | x_{-ij} \Leftrightarrow Q_{ij}(\boldsymbol{\theta}_1) = 0$.

O objectivo é estimar as distribuições marginais a posteriori $\pi(x_i|\mathbf{y})$ e $\pi(\boldsymbol{\theta}_j|\mathbf{y})$:

$$\begin{aligned} \pi(x_i|\mathbf{y}) &= \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\ \pi(\boldsymbol{\theta}_j|\mathbf{y}) &= \int \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j}. \end{aligned}$$

A ideia é construir para estas distribuições aproximações “aninhadas”:

$$\begin{aligned} \tilde{\pi}(x_i|\mathbf{y}) &= \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\ \tilde{\pi}(\boldsymbol{\theta}_j|\mathbf{y}) &= \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j}, \end{aligned}$$

onde $\tilde{\pi}$ é uma função densidade (condicional) aproximada.

As aproximações de $\pi(x_i|\mathbf{y})$ são obtidas aproximando $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ e $\pi(\boldsymbol{\theta}|\mathbf{y})$ e então usando integração numérica em $\boldsymbol{\theta}$ (soma finita) para o fazer desaparecer - o que é possível, naturalmente, pela baixa dimensionalidade de $\boldsymbol{\theta}$. O mesmo vale para $\pi(\boldsymbol{\theta}_j|\mathbf{y})$.

A aproximação proposta é a obtida pelo método de Laplace. Através deste método consegue aproximar-se uma função densidade $h(x)$ por uma distribuição gaussiana de média dada pela moda \hat{x} e variância dada por $\left(-\frac{\partial^2 \log h(\hat{x})}{\partial x^2}\right)^{-1}$, sendo este resultado conseguido simplesmente por se aproximar $\log h(x)$ pelos primeiros termos da sua expansão em série de Taylor (até à 2ª ordem).

Assim, aproximação Laplace proposta para $\pi(\boldsymbol{\theta}|\mathbf{y})$ é:

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}, \quad (1)$$

onde $\tilde{\pi}_G$ é a aproximação de Laplace (gaussiana) da distribuição condicional completa de \mathbf{x} e $\mathbf{x}^*(\boldsymbol{\theta})$ a sua moda, para um dado $\boldsymbol{\theta}$.

Então a aproximação INLA de $\pi(x_i|\mathbf{y}, \boldsymbol{\theta})$ é feita da seguinte forma:

- 1: Utilize-se a aproximação Laplace (1) para aproximar $\pi(\boldsymbol{\theta}|\mathbf{y})$;
- 2: Utilize-se a aproximação de Laplace de $\pi(x_i|\mathbf{y}, \boldsymbol{\theta})$ para valores seleccionados de $\boldsymbol{\theta}$ (tomando partido computacional de \mathbf{x} ser um campo aleatório gaussiano);
- 3: Combinação dos dois passos anteriores usando integração numérica (tirando partido de $\boldsymbol{\theta}$ ter m elementos, m pequeno)

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_m \tilde{\pi}(x_i|\boldsymbol{\theta}_m, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}_m|\mathbf{y}) \Delta_m \quad (2)$$

3 INLA na prática

Os autores da abordagem INLA implementaram os seus procedimentos num pacote para o *software* R-Project, o pacote R-INLA, www.r-inla.org. Na página do projecto pode-se descarregar não apenas o pacote mas também tutoriais, vários exemplos, artigos, relatórios técnicos e código de acompanhamento dos vários exemplos. Este é um pacote ainda em desenvolvimento mas que tem já capacidade de trabalhar com uma variedade bastante grande de modelos. No entanto os seus autores são muito dinâmicos e disponibilizam-se para produzir novos códigos que sejam necessários de forma a responder às necessidades dos vários utilizadores, incentivando pedidos.

Na subsecção seguinte descreve-se um exemplo em que a abordagem INLA foi usada para estimar um modelo hierárquico espacial.

3.1 Uma aplicação: modelo hierárquico espacial

No contexto de incêndios florestais [2] modelam simultaneamente a proporção de área ardida dentro de pequenas divisões do território e a probabilidade de ignição aí, para dados de incêndio registados na região norte de Portugal no ano de 2005, através de um modelo espacial hierárquico, que se detalha.

Represente $Y(i)$ a proporção de área ardida numa das pequenas divisões do território (freguesia) e seja $R(i)$ uma variável dicotómica indicando se houve um incêndio na freguesia i , em 2005. Considera-se uma transformação dos valores não nulos de Y com vista à sua simetriação. O modelo segue-se:

$$\begin{aligned}
R(i) | p(i) &\sim \text{Bernoulli}(p(i)) & \text{logit}(p(i)) &= \beta + \mathbf{X}(i)^T \boldsymbol{\eta} + V_2(i) \\
Z(i) &= \begin{cases} \log\left(\frac{Y(i)}{1-Y(i)}\right), & 0 < Y(i) < 1 \\ 0, & Y(i) = 0 \end{cases} \\
Z(i) | R(i) = 1 &\sim N(\mu(i), \sigma^2) & \mu(i) &= \alpha + \mathbf{X}(i)^T \boldsymbol{\delta} + V_1(i) \\
V_1(i) &= W_0(i) + W_1(i) & V_2(i) &= W_1(i) \\
W_0 &\sim \text{ICAR}(\tau_0) & W_1 &\sim \text{ICAR}(\tau_1)
\end{aligned}$$

onde $p(i)$ é a probabilidade de ignição na freguesia i , \mathbf{X} são covariáveis de interesse, não necessariamente as mesmas para $p(i)$ e $\mu(i)$, $V_2(i)$ são variáveis explicativas espaciais não observadas que influenciam a ignição do fogo, $V_1(i)$ são variáveis explicativas espaciais não observadas que influenciam o tamanho do fogo - $V_1(i)$ e $V_2(i)$ são efeitos aleatórios espaciais dependentes. Esta dependência é induzida pelo processo espacial latente W_1 . Consideram-se distribuições a priori autoregressivas intrínsecas para os processos espaciais latentes independentes W_0 e W_1 e distribuições a priori gaussianas para os parâmetros dos efeitos das covariáveis. Para os parâmetros da precisão escolheram-se distribuições a priori gama. Considera-se a independência a priori entre estes parâmetros.

Nesta aplicação o vector das observações \mathbf{y} é constituído pelos $R(i)$ e pelos $Z(i)$, para os quais se assumem distribuições diferentes. O vector das variáveis latentes \mathbf{x} é formado pelos parâmetros dos efeitos das covariáveis, β , $\boldsymbol{\eta}$, α , $\boldsymbol{\delta}$ e pelos efeitos dos processos espaciais latentes, W_0 e W_1 , este último partilhado pelos dois tipos de observações. O número total de freguesias consideradas é de 2654, pelo que este é um problema de grandes dimensões - o vector das variáveis latentes tem mais de 5000 elementos! O número de hiperparâmetros são apenas 3, τ_1 , τ_2 e σ^2 .

Apresenta-se de seguida o código que se utilizou para estimar este modelo no R-INLA (representado fogos2005n a base de dados no R, fogos2005n\$rr os dados da variável R , fogos2005n\$z os dados da variável Z , fogos2005n\$regiao a identificação de cada região, cov_rr os dados das covariáveis de interesse para a modelação de R e cov_zz os dados das covariáveis interesse para a modelação de Z):

```

n=dim(fogos2005n) [1]

Y=matrix(NA, 2*n, 2)
Y[1:n, 1] = fogos2005n$rr
Y[1:n+n, 2] = fogos2005n$z

regiao = c(fogos2005n$regiao, fogos2005n$regiao)
regiao2 = c(rep(NA, n), fogos2005n$regiao)

formula4 = Y ~ intercept1 + intercept2 - 1 + cov_rr + cov_zz +
f(regiao, model='besag', graph.file='ptfreg.graph', param=c(0.1, 0.1))
+ f(regiao2, model='besag', graph.file='ptfreg.graph', param=c(0.1, 0.1))

resultado4 = inla(formula4, family=c('binomial', 'gaussian'),
data=data.frame(Y, intercept1, intercept2, Region, Region2),
Ntrials=c(datanat$n, rep(0, n)), control.compute=list(dic=T), verbose=T)

```

A formula4 corresponde ao preditor linear. A instrução inla faz correr o algoritmo. Nos argumentos desta função incluem-se o preditor, as distribuições dos dados, os dados e a indicação para calcular a medida da qualidade do ajuste DIC.

O algoritmo demorou 11 minutos a correr, em oposição às 2 horas que o WinBUGS demorou para estimar o mesmo modelo, com apenas 5000 iterações após um período de aquecimento the 1000 iterações (e com

algumas paragens do programa devido a *traps*). As estimativas obtidas pela metodologia INLA e MCMC concordaram em sentido e significância mas não tanto em valor - para uma análise detalhada da aplicação e comparações entre metodologias ver [2].

4 Discussão

Apresentada a abordagem INLA fica a dúvida: mas é melhor do que o que se fazia antes, nomeadamente a abordagem MCMC? Vale a pena o esforço do investimento na sua implementação para os problemas em mãos? A resposta é: depende do problema.

Os métodos MCMC têm, por vezes, dificuldade em lidar com modelos muito complexos, apresentando problemas na convergência ou demorando muito tempo a convergir. Os métodos MCMC tradicionais tendem a não funcionar muito bem para modelos gaussianos latentes mais complicados porque, por um lado, as componentes desses modelos tendem a ser muito correlacionadas causando um impacto significativo na convergência do método MCMC e, por outro lado, quando o número de observações é grande \mathbf{x} e $\boldsymbol{\theta}$ costumam ser também muito correlacionados. O problema pode ser aliviado, mas não eliminado, fazendo MCMC por blocos.

O INLA é uma ferramenta muito eficaz para estimar modelos gaussianos latentes, com tempos de computação muito inferiores ao MCMC e com precisão das estimativas usualmente maior do que a precisão das estimativas obtidas por MCMC tradicional, como o documentam o número de trabalhos publicados nos últimos cinco anos em que se recorre ao INLA - www.r-inla.org/papers. É uma ferramenta flexível para uma grande variedade de modelos, muitos dos quais cabem nesta formulação dos modelos gaussianos latentes, podendo-se fazer então uso das boas propriedades computacionais dos campos de Markov aleatórios gaussianos e das matrizes esparsas resultantes.

Contudo, o MCMC é uma abordagem talvez um pouco mais flexível, o que é comprovado por se conseguir ajustar praticamente qualquer modelo usando o WinBUGS, por exemplo. A preparação do INLA pode ser muito especializada apesar de menos intuitiva do que a maioria dos casos com MCMC.

As complicações que podem surgir no INLA são relacionadas com modelos com muitos hiperparâmetros e observações claramente não gaussianas. A maior desvantagem do INLA é que, sendo um método aproximado, a estimação é sempre enviesada.

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

Referências

- [1] Y. Fong, H. Rue, and J. Wakefield. Bayesian inference for generalized linear mixed models. *Biostatistics*, 11:397–412, 2010.
- [2] I. Natário, M. M. Oliveira, and S. Marques. Using INLA to estimate a highly dimensional spatial model for forest fires in portugal. In A. Pacheco, R. Santos, R. Silva, and C. D. Paulino, editors, *Selected Papers of the Statistical Societies*, 2013. Accepted.
- [3] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71:319–392, 2009.
- [4] B. Schrödle and L. Held. Spatio-temporal disease mapping using INLA. *Environmetrics*, 22:725–734, 2010.
- [5] B. Schrödle and L. Held. A primer on disease mapping and ecological regression using INLA. *Computational Statistics*, 26:241–258, 2011.



Métodos de Monte Carlo Sequenciais e Métodos Computacionais Bayesianos Aproximados

P. de Zea Bermudez, *pcbermudez@fc.ul.pt*

CEAUL e Faculdade de Ciências da UL

1 Introdução

Os métodos de Monte Carlo sequenciais (SMC) e os métodos computacionais bayesianos aproximados (ABC) surgem como alternativas aos tradicionais métodos de Monte Carlo via cadeias de Markov (MCMC), fundamentalmente como metodologias para abordar os problemas computacionais que são tipicamente observados no contexto da inferência bayesiana. O facto de os modelos usados na prática serem tendencialmente mais complexos, e com frequência envolverem o processamento de grandes volumes de dados, faz com que os métodos SMC e ABC possam ser mais eficientes do que os tradicionais MCMC. Efectivamente, a representação precisa da realidade requer habitualmente a utilização de modelos não gaussianos, usualmente não lineares e com espaços paramétricos de elevada dimensão. Embora o propósito das metodologias referidas seja o mesmo, naturalmente diferem na forma de abordar o problema. Deste modo, a utilização de um algoritmo SMC ou de um algoritmo ABC depende do problema em questão.

Os métodos SMC permitem o processamento da informação “em blocos”, à medida que os dados se vão tornando disponíveis, diminuindo assim a sobrecarga computacional. São especialmente úteis em situações nas quais há dependência entre observações, tais como nas séries temporais. A estimação da volatilidade de certos produtos cotados no mercado de capitais, o acompanhamento do movimento dos aviões com base em observações fornecidas por radar e a produção fabril de produtos de risco ou de elevado custo, são exemplos de situações em que as observações são naturalmente processadas sequencialmente. Em situações como as descritas, as decisões têm de ser tomadas em cada instante, condicionalmente à informação até aí existente, sendo impossível esperar pela totalidade das observações (ver Doucet et al. [9]). Neste contexto de inferência bayesiana, e sendo a distribuição de interesse a distribuição *a posteriori* (ou a distribuição preditiva), os métodos SMC permitem simular sequencialmente valores desta distribuição usando distribuições *a posteriori* “intermédias”, no sentido em que utilizam os dados existentes até ao presente.

Por sua vez, os métodos ABC surgem na literatura como uma forma de ultrapassar os problemas computacionais associados a modelos com verosimilhanças muito complicadas, excessivamente morosas de calcular, nomeadamente quando as amostras têm uma dimensão elevada, ou mesmo intratáveis. Neste contexto, a utilização dos métodos MCMC é usualmente impraticável. Os algoritmos ABC simulam valores do modelo de interesse sem recurso à verosimilhança, sendo a informação contida nos dados resumida num conjunto de estatísticas sumárias. O sucesso do algoritmo que se implementar depende de diversos factores, nomeadamente da facilidade com a qual se simula do modelo de interesse e da capacidade de as estatísticas sumárias escolhidas representarem fidedignamente os dados em estudo. Neste contexto, obviamente impera a ausência de estatísticas suficientes. Embora as abordagens bayesianas via métodos ABC tenham surgido na modelação de dados de Genética Populacional (Beaumont et al. [2], Csilléry et al. [6]), num contexto paramétrico tipicamente discreto, actualmente são utilizados em

variados domínios científicos, tais como a arqueologia, a epidemiologia e a ecologia, alguns dos quais podem envolver espaços paramétricos contínuos (ver, por exemplo, Marin et al. [12]).

2 Notação geral e conceitos básicos

Seja θ um parâmetro, ou um vector de parâmetros, associado a um certo modelo amostral \mathcal{M} , $p(\theta)$ a distribuição que reflecte o conhecimento *a priori* sobre θ e $\mathcal{D}_t = (x_1, x_2, \dots, x_t)$ o conjunto de dados observado até ao instante t ($t \geq 1$). A distribuição *a posteriori* de θ é dada por

$$p(\theta | \mathcal{D}_t) = \frac{L(\theta | \mathcal{D}_t)p(\theta)}{\int_{\theta \in \Theta} L(\theta | \mathcal{D}_t)p(\theta)d\theta},$$

sendo usualmente reduzida à relação

$$p(\theta | \mathcal{D}_t) \propto L(\theta | \mathcal{D}_t)p(\theta),$$

ou seja, ao produto da distribuição *a priori* com a verosimilhança, $L(\theta | \mathcal{D}_t)$. Como é habitual nos problemas reais, a distribuição *a posteriori*, $p(\theta | \mathcal{D}_t)$, não tem forma fechada. Numa abordagem clássica de MCMC, usa-se habitualmente o método de amostragem de Gibbs, frequentemente incorporando um (ou mais do que um) passo do algoritmo de Metropolis-Hastings, para simular valores da distribuição de interesse (ver, por exemplo, Carlin e Louis [5] ou Paulino *et al.* [13]). Como referido na Introdução, esta opção pode não ser a mais apropriada em certas situações, podendo recorrer-se a metodologias alternativas, tais como os métodos SMC ou ABC.

Por restrições óbvias de espaço e simplicidade de exposição, abordar-se-ão apenas as versões, ditas originais, dos algoritmos SMC e ABC. No que concerne aos algoritmos ABC os dados observados são fixos e serão referidos apenas por \mathcal{D} .

3 Métodos SMC

Seja $\pi_t(\theta) = p(\theta | \mathcal{D}_t)$ a distribuição *a posteriori* de θ , baseada nos dados obtidos até ao instante t , $t \geq 1$. Suponhamos que se pretende amostrar desta distribuição sequencialmente, à medida que as observações se vão tornando acessíveis. Alternativamente, a necessidade de usar um algoritmo SMC pode dever-se ao facto de ser mais simples de amostrar de distribuições intermédias, que permitam “passar” de $\pi_1(\theta)$ para $\pi_t(\theta)$ (ver Del Moral *et al.* [7] e correspondentes referências). Os algoritmos SMC baseiam-se no método de amostragem por importância sequencial (SIS), o qual naturalmente se fundamenta no método de amostragem por importância (IS) (ver Paulino *et al.* [13]).

Os métodos SMC permitem, em determinado momento t , ter uma sequência de N partículas e pesos, $\{\theta_t^i, W_t^i\}$, $i = 1, 2, \dots, N$, sendo W_t^i o peso associado a θ_t^i . Estes pesos são normalizados, devendo satisfazer as propriedades habituais: $W_t^i > 0$, $i = 1, 2, \dots, N$ e $\sum_{i=1}^N W_t^i = 1$. Estas partículas evoluem ao longo do tempo de acordo com os pesos associados a θ_t^i . No fundo, o método IS é seguido por um processo de reamostragem que elimina as partículas mais “fracas”, no sentido de serem as que têm pesos associados menores, apenas passando para a geração seguinte as partículas mais “fortes”, as que têm pesos mais elevados. Como é de esperar, dado que o procedimento se baseia no método IS, o desempenho destes algoritmos depende, não só das características das densidades de interesse, mas também das distribuições propostas que são consideradas (ver Del Moral et al. [7]). Um dos grandes problemas destes algoritmos é a tendência que os pesos têm em degenerar. Tal situação ocorre, por exemplo, quando N não é suficientemente elevado (ver Doucet [8]). Por outro lado, considerar um N excessivamente grande naturalmente aumenta o tempo de processamento.

Basicamente, o processo SIS consiste em (usando a notação de Del Moral *et al.* [7]): aproximar a distribuição de interesse, $\pi_1(\theta)$, por η_1 , usando o método IS. No instante seguinte, a distribuição $\pi_2(\theta)$ será aproximada por η_2 , utilizando as partículas geradas na iteração anterior. No instante $t - 1$, estaremos em presença de um conjunto de N partículas com distribuição η_{t-1} . As partículas são “movimentadas” para o seu estado final através de núcleos markovianos.

Os métodos SMC têm sido muito aplicados a modelos que podem ser escritos como espaços de estados (*State Space Models*), com resultados muito interessantes (ver por exemplo Andrieu *et al.* [1]).

4 Algoritmos ABC

Os algoritmos ABC baseiam-se no sobejamente conhecido método da rejeição (ver, por exemplo, Paulino *et al.* [13]). Basicamente, consistem na execução repetida dos passos seguintes:

Passo 1. simular um valor de θ da distribuição *a priori*, $p(\cdot)$

Passo 2. simular uma amostra \mathcal{D}^* do modelo \mathcal{M}

Passo 3. aceitar o valor de θ gerado no Passo 1. se $d(S(\mathcal{D}^*), S_0) \leq \delta$

sendo $d(\cdot, \cdot)$ uma medida de distância e δ um certo nível de tolerância não negativo. O valor da estatística $S(\cdot)$ calculada com os dados observados é representado por S_0 , sendo um valor fixo, e $S(\mathcal{D}^*)$ o valor da estatística sumária calculado com os dados simulados a partir do modelo \mathcal{M} . Os dados simulados de θ constituem observações de $p(\theta \mid d(S(\mathcal{D}^*), S_0) \leq \delta)$, mas não da distribuição *a posteriori* verdadeira, a menos que δ seja nulo. Se $\delta \rightarrow \infty$ então os valores de θ simulados são simplesmente observações da distribuição *a priori*, e não da distribuição *a posteriori* como se pretende (ver Marjoram *et al.* [11] e Wilkinson [15]).

Evidentemente, o desempenho do algoritmo depende de diversos factores, tais como da escolha da função de distância, da selecção e número de estatísticas sumárias que resumem convenientemente os dados e da escolha de δ . O valor de δ deverá reflectir, dentro do possível, um equilíbrio entre a eficiência computacional e a precisão que se pretende que os resultados finais tenham.

A norma euclidiana, $d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$, $x, y \in \mathbb{R}^n$, é habitualmente a distância considerada, embora possa não ser a mais adequada em alguns casos. Uma medida de distância relativa poderá ser uma escolha mais conveniente em muitas situações, nomeadamente as que envolvem quantidades que podem tomar valores muito díspares.

A selecção do número e natureza das estatísticas sumárias a considerar depende obviamente do problema em estudo. As escolhas naturais são os momentos de ordem não superior a quatro, os quais permitem comparar a localização, a dispersão, a forma e a curtose da distribuição dos dados observados com as correspondentes características dos dados simulados a partir do modelo \mathcal{M} (condicional ao valor de θ gerado a partir da distribuição *a priori*). Por exemplo, no caso de um problema que envolva dados de cauda pesada, poderá também incluir-se uma medida da semelhança que possa existir entre as caudas dos dados observados e dos dados simulados, por exemplo o estimador dos momentos ou o estimator de Hill (Embrechts *et al.* [10]).

Relativamente a todos estas questões que envolvem os métodos ABC tem havido recentemente várias contribuições (ver, por exemplo, Beaumont *et al.* [2] e Blum *et al.* [4]). Biau *et al.* [3] propõem considerar os k_N valores de θ que estão mais próximos de S_0 para uma certa medida de proximidade. Habitualmente, k_N corresponde ao quantil de probabilidade $p = 0.90$.

Os algoritmos ABC são fáceis de implementar mas são em regra computacionalmente exigentes. Por exemplo, considerando as indicações de Blum *et al.* [3], se o vector de parâmetros tiver dimensão $p = 3$

e o número de estatísticas sumárias for $m = 7$, então ter-se-á de simular $N = 10^6$ conjuntos de dados de forma a obter $n = 1000$ valores de θ , i.e.,

$$k_N \approx N^{\frac{p+4}{m+p+4}}.$$

4.1 Exemplo ¹

O método ABC é exemplificado na estimação dos parâmetros do seguinte modelo bilinear simples:

$$Y_t = 0.6Y_{t-1} + 0.4Y_{t-1}\varepsilon_{t-1} + \varepsilon_t,$$

sendo $a = 0.6$, $b = 0.4$, ε_t v.a.'s i.i.d. $N(\mu, \sigma)$, $\mu = 0$ e $\sigma^2 = 2.0$. O modelo indicado satisfaz a condição $a^2 + b^2\sigma^2 < 1$, a qual garante que é estacionário e invertível.

Os modelos bilineares são conhecidos por terem funções de verosimilhança especialmente difíceis de manusear, o que torna a utilização da metodologia ABC uma alternativa especialmente atractiva (Turkman *et al.* [14]).

Seja \mathcal{D} um conjunto de $n = 5000$ observações (amostra fixa) geradas a partir do modelo bilinear indicado. Os valores *a priori* de $\theta = (a, b, \sigma)$ são gerados a partir das seguintes distribuições:

$$a \sim Unif[-0.90, 0.90], b \sim Unif([-0.90, -0.1] \cup [0.1, 0.90])$$

e

$$\sigma^2 | a, b \sim Unif([0.1, (1 - a^2)/b^2]).$$

Consideraram-se estatísticas sumárias que reflectem a semelhança entre a localização, a dispersão, a função de autocorrelação, o peso de cauda e o índice extremal dos dados fixos \mathcal{D} e $N = 10^6$ amostras simuladas do modelo bilinear indicado (condicional aos valores de a , b e σ^2). Considerando como função de distância a norma euclidiana e tomando o percentil 0.90 (proposta de Biau *et al.* [3]) das $N = 1000000$ distâncias calculadas, obteve-se uma amostra de dimensão $k_N = 1000$. Os resultados obtidos (da distribuição *a posteriori* aproximada), a partir da amostra dos "melhores" 1000 valores de a , b e σ , são apresentados na Tabela 1. As estimativas das densidades, usando um núcleo gaussiano, constam da Figura 1. A quase coincidência entre os valores de a , b e σ^2 e as estimativas obtidas claramente revelam os benefícios de usar um algoritmo ABC no contexto de modelos particularmente complexos como os modelos bilineares.

Parâmetro	Mínimo	$Q_{0.25}$	Mediana	Média	$Q_{0.75}$	Máximo	s
a	0.3615	0.5325	0.6001	0.5976	0.6623	0.8271	0.09
b	0.1605	0.3362	0.4215	0.4190	0.4988	0.6941	0.11
σ^2	0.8264	1.7320	2.1110	2.2110	2.6380	4.2670	0.65

Tabela 1: Estatísticas dos dados simulados ($n = 1000$) - Modelo $Y_t = 0.6Y_{t-1} + 0.4Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$, ε_t são v.a. i.i.d. $N(0,2)$.

5 Comentários finais

Tanto os métodos SMC, como os ABC, são muito úteis para simular valores das distribuições de interesse, tais como as distribuições *a posteriori* e preditiva, constituindo excelentes alternativas aos habituais algoritmos MCMC. Fica patente pela (muito) breve exposição realizada que os algoritmos SMC são

¹Elaborado por de Zea Bermudez, P., Amaral Turkman, M. A. e Turkman, K. F..

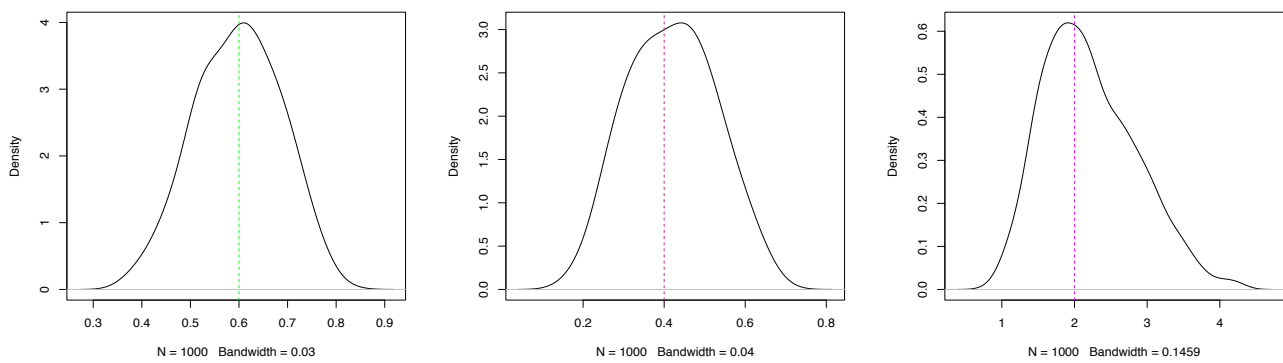


Figura 1: Estimativas das densidades - Modelo $Y_t = 0.6Y_{t-1} + 0.4Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$, ε_t i.i.d. $N(0, 2)$ - a (esquerda), b (centro) e σ^2 (direita)

mais complexos que os ABC, podendo os seus utilizadores ser confrontados com algumas questões de abordagem difícil, tais como a degeneração das partículas que vão sendo criadas à medida que o algoritmo SMC que está a ser aplicado vai evoluindo. A metodologia ABC é muito simples de implementar, embora possa ser “pesada” em termos computacionais. Como se verifica no exemplo apresentado, pode conduzir a resultados muito satisfatórios, basicamente no caso de ser fácil simular observações do modelo em questão e se as estatísticas que resumem os dados, assim como a função distância, forem adequadamente seleccionadas.

Agradecimentos: Este trabalho foi parcialmente financiado pela Fundação para a Ciência e Tecnologia, Portugal, através dos projectos PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010.

Referências

- [1] Andrieu C., Doucet A. and Holenstein R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B*, 72, 269-342.
- [2] Beaumont, M. A., Zhang, W. e Balding D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162, 2015-2035.
- [3] Biau, G., Cérou, F. e Guyader, A. (2012). New Insights into Approximate Bayesian Computation. *Submitted*.
- [4] Blum, M. G. B. e François, O. (2008). Highly tolerant likelihood-free Bayesian inference: an adaptive non-linear heterocedastic model. *Available at arXiv:0810.0896v1*.
- [5] Carlin, B. P. e Louis, T. A. (2009). Bayesian methods for data analysis, 3rd Edition, CRC Press, Chapman and Hall, Boca Raton.
- [6] Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. e François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *doi:10.1016/j.tree.2010.04.001*.
- [7] Del Moral, P., Doucet, A. e Jasra, A. (2006). Sequential Monte Carlo Samplers. *J. R. Statist. Soc. B*, 68, 411-436.
- [8] Doucet, A. (2010). Sequential Importance Sampling Resampling. Lecture Notes 3, SAMSI, North Carolina, USA.

- [9] Doucet, A., Freitas, N. e Gordon, N. (2001). An introduction to Sequential Monte Carlo Methods, 3-13, in *Sequential Monte Carlo Methods in Practice*, Editors Doucet, A., Freitas, N. e Gordon, N., Springer.
- [10] Embrechts, P., Klüppelberg, C. e Mikosch, T., 1997. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.
- [11] Marjoram, P., Molitor, J., Plagnol, V., e Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A*, 100 (26), 15324 - 15328.
- [12] Marin, J.M., Pudlo, P., Robert, C.P. e Ryder, R. (2012). Approximate Bayesian Computational methods. *Statistics and Computing*, in press.
- [13] Paulino, C. D., Amaral Turkman, M. A. e Murteira, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.
- [14] Turkman, K. F., Scotto, M. G. e de Zea Bermudez, P. (2013). Nonlinear time models. Accepted for publication in Springer.
- [15] Wilkinson, R. D. (2008). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. arXiv:0811.3355v2.[stat.CO].



Um breve passeio pelo campo computacional bayesiano

Paulo Soares¹, psouares@math.ist.utl.pt
Valeska Andreozzi², valeska.andreozzi@exigoconsultores.com

¹CEAUL e Instituto Superior Técnico

²CEAUL, Exigo Consultores e Faculdade de Ciências Médicas da Universidade Nova de Lisboa

1 Introdução

Há já muito tempo que a estatística não pode passar sem o auxílio de poderosos meios computacionais. A análise bayesiana de grandes conjuntos de dados recorrendo a modelos cada vez mais complexos vem criando necessidades de cálculo que não param de crescer. O desenvolvimento teórico de metodologias ou algoritmos que permitem analisar distribuições *a posteriori* arbitrárias tem acompanhado essas necessidades e, desde o início da década de 90 do século passado, o progresso da estatística bayesiana segue a par de uma generosa produção de software, gratuito e frequentemente de código aberto, que muito tem contribuído para a divulgação e afirmação dessa visão da estatística.

Neste trabalho apresentaremos, de uma forma necessariamente breve e incompleta, as principais estrelas da rica paisagem computacional bayesiana. Mas, antes de sequer referirmos qualquer software especificamente bayesiano, parece-nos importante destacar o papel central do software R [7], que se vem afirmando como a língua franca da estatística computacional e que possui um grande número de pacotes relacionados com a inferência bayesiana, como se pode ver no catálogo de uma das *Task Views* do R dedicada a esse tema. Com a contribuição de alguns de esses pacotes, o R pode mesmo servir de interface para todas as aplicações que iremos abordar sendo possível ter assim numa plataforma única o pré-processamento de dados, o uso dos diferentes softwares bayesianos e o diagnóstico e análise de resultados.

A título de ilustração iremos recorrer a um simples modelo de regressão de Poisson definido por:

$$Y_i \sim \text{Poisson}(\mu_i), \log \mu_i = \beta_1 + \beta_2 \log n_i + \beta_3 \log d_i, \quad i = 1, \dots, N,$$

para as respostas Y_i e os valores observados das covariáveis n_i e d_i . Numa perspetiva bayesiana é ainda necessário especificar distribuições *a priori* para os parâmetros β_1 , β_2 e β_3 , como, por exemplo, $\beta_i \sim \text{normal}(a_i, b_i)$, em que a_i e b_i são constantes para $i = 1, 2, 3$.

Nas próximas secções usaremos o modelo anterior para exemplificar apenas as diferentes formas de o representar nas linguagens de modelação fornecidas pelos vários softwares. O leitor interessado pode descarregar do endereço <http://goo.gl/FXInt6> um ficheiro comprimido contendo um curto *script* do R que permite analisar¹ um conjunto de dados recorrendo a todos os softwares que serão abordados nas secções seguintes. Os resultados que se podem obter (os mais imediatos estão incluídos no ficheiro *resultados.txt*) são razoavelmente próximos mas naturalmente distintos uma vez que são obtidos por simulação ou, num só caso, recorrendo a aproximações. Nesse *script* podem encontrar-se ainda os endereços das páginas *web* de cada software.

¹Desde que todo o software necessário esteja instalado, é claro.

2 A ilustre família BUGS

É curioso reparar que quando foi publicado um dos primeiros trabalhos de aplicação da simulação de Monte Carlo com recurso a cadeias de Markov (MCMC) em estatística bayesiana, o artigo clássico de Alan Gelfand e Adrian Smith [1], já estava a ser desenvolvido na *MRC Biostatistics Unit* em Cambridge o primeiro software que permitiu a análise generalizada de modelos bayesianos por amostragem de Gibbs, o BUGS [9] – *Bayesian inference Using Gibbs Sampling*. A primeira apresentação pública deste software ocorreu no *IV Valencia Meeting* em 1991, mas só em 1993 foi lançada a sua primeira versão a que se seguiram várias outras até à versão final (0.6) em 1996.

Para além do papel decisivo na popularização dos métodos de MCMC, o fator que mais contribuiu para o sucesso do projeto BUGS foi a definição de uma linguagem para a especificação textual de modelos bayesianos. A linguagem BUGS dá acesso a um generoso número de distribuições de probabilidade e a outras funções matemáticas, e inclui ainda algumas estruturas típicas de linguagens de programação, tais como ciclos, permitindo representar de uma forma simples as relações estocásticas (com o símbolo \sim) ou lógicas/determinísticas (com os símbolos $< -$) estabelecidas entre as quantidades presentes em modelos estatísticos de complexidade (quase) arbitrária.

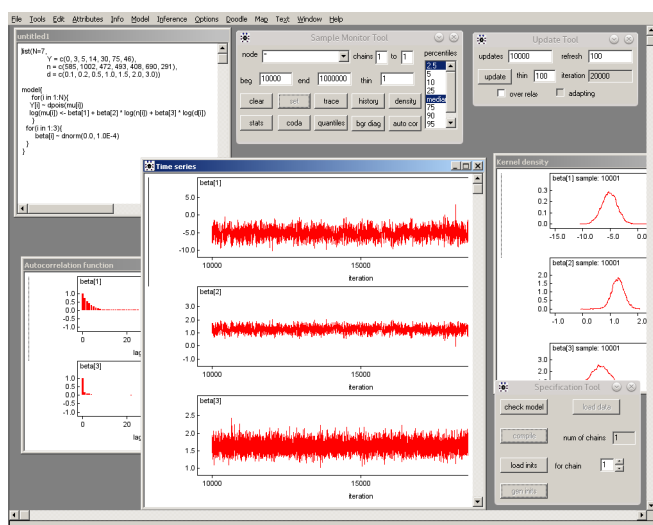
Para ilustrar a simplicidade e flexibilidade da linguagem BUGS retomemos o modelo de regressão de Poisson atrás descrito para o qual bastariam não mais do que 10 linhas de código. Em vez de esse modelo mais simples, apresenta-se abaixo uma versão ligeiramente mais extensa em que as covariáveis são centradas nos respetivos valores médios. Esta estratégia obriga a transformar os dados mas conduz a um ganho de eficiência na simulação ao reduzir substancialmente a autocorrelação nas cadeias de valores gerados.

```
# Modelo BUGS -----
model {
  for(i in 1:N){
    n_t[i] <- log(n[i])
    d_t[i] <- log(d[i])
  }
  n_t.m <- mean(n_t[])
  d_t.m <- mean(d_t[])
  for(i in 1:N){
    Y[i] ~ dpois(mu[i])
    log(mu[i]) <- beta[1] + beta[2] * (n_t[i] - n_t.m) + beta[3] * (d_t[i] - d_t.m)
  }
  beta_1 <- beta[1] - beta[2] * n_t.m - beta[3] * d_t.m
  for(i in 1:3){
    beta[i] ~ dnorm(a[i], b[i])
  }
}
```

O código anterior é uma representação próxima do modelo estatístico mas merece, ainda assim, alguns esclarecimentos. Em primeiro lugar, a linguagem BUGS é uma linguagem declarativa, ou seja, a ordem das instruções é irrelevante podendo ser feitas referências a variáveis definidas posteriormente. Esta característica permite a escrita de código mais simples e fácil de ler mas, em contrapartida, impede o uso de expressões condicionais (*if-then-else*), o que representa uma importante limitação da linguagem que, ainda assim, pode ser contornada nalgumas situações. Em seguida, a distribuição normal está parametrizada no BUGS de uma forma pouco usual em termos da precisão, ou seja, o valor atribuído a $b[i]$ não deve ser o da variância *a priori* mas sim o seu inverso. Note-se ainda que o conjunto de dados e os valores iniciais dos parâmetros β_1 , β_2 e β_3 devem ser fornecidos em dois outros ficheiros de texto separados do modelo e que, para delinear e controlar a simulação, é necessário ainda um quarto ficheiro de texto contendo algumas instruções.

Apesar de limitado, o BUGS foi crucial na afirmação da estatística bayesiana, dentro e fora do meio académico, e desencadeou o surgimento de um grande número de aplicações bayesianas em todas as áreas de aplicação da estatística. Em meados da década de 1990, as limitações do BUGS foram-se tornando crescentemente difíceis de suportar pela conjugação de três fatores: o aumento da complexidade dos modelos estatísticos analisados, um grande desenvolvimento dos métodos de MCMC e a enorme popularidade que o Microsoft Windows atingiu.

Em 1996 o projeto BUGS foi deslocado para o *Imperial College* de Londres e o desenvolvimento do software acelerou fortemente conduzindo a uma nova versão, agora em exclusivo para o *Microsoft Windows*, apropriadamente chamada de WinBUGS [3]. O WinBUGS apresentou-se com o novo interface gráfico, novos algoritmos de MCMC (Metropolis-Hastings e *slice sampling*, entre outros) e diversas ferramentas de diagnóstico e análise de resultados.



A capacidade de análise do WinBUGS foi também sendo expandida pela criação de módulos especializados: o PKBugs para a análise de modelos farmacocinéticos, o GeoBugs para modelação espacial, o WBDiff que permite lidar com sistemas de equações diferenciais ordinárias, o WinBUGS Jump Interface que implementa o algoritmo de MCMC com saltos reversíveis e ainda o WBDev, um interface de desenvolvimento que permite adicionar novos componentes ao WinBUGS. Com a introdução de todas estas inovações, o WinBUGS tornou-se rapidamente um dos mais populares softwares estatísticos.

Em 2007, o WinBUGS atingiu a versão 1.4.3 e o projeto BUGS, após 18 anos de desenvolvimento,

chegou ao fim. No entanto, o código-fonte do WinBUGS foi entregue a um novo projeto, o OpenBUGS, que se encarregou de o abrir à comunidade, terminou com a exclusividade do Microsoft Windows e continua ainda o seu desenvolvimento. Atualmente, o OpenBUGS já é mais sofisticado do que o WinBUGS em múltiplos aspetos como, por exemplo, a seleção dos algoritmos de MCMC empregues na simulação de cada modelo, mas mantém ainda um carácter experimental e o seu desenvolvimento tem sido muito lento. De fato, o futuro do OpenBUGS não parece particularmente risonho uma vez que a linguagem de programação escolhida para a criação do WinBUGS – uma variante do Pascal chamada Component Pascal – se tornou rapidamente obsoleta e, por isso, apesar de o código-fonte do OpenBUGS ser público, o projeto não tem conseguido atrair novos programadores.

Como alternativa ao OpenBUGS, surgiu em 2002 o JAGS [6] – *Just Another Gibbs Sampler* – criado e presentemente desenvolvido por Martyn Plummer da *International Agency for Research on Cancer*. O JAGS é largamente inspirado no BUGS e, por isso, os modelos escritos para o BUGS/WinBUGS podem ser usados no JAGS com poucas ou nenhuma alteração. A principal vantagem do JAGS decorre do fato de ser escrito em C++ o que permite o seu uso num grande número de sistemas operativos. Para além disso, o JAGS promove o uso mais alargado de funções vectoriais ou matriciais, que evitam o uso de ciclos, e introduz diversas alterações na linguagem BUGS que a ampliam ou melhoram.

```
# Modelo JAGS -----
data {
  n_t <- log(n)
  d_t <- log(d)
  n_t.m <- mean(n_t)
  d_t.m <- mean(d_t)
}

model {
  for(i in 1:N){
    Y[i] ~ dpois(mu[i])
    log(mu[i]) <- beta[1] + beta[2] * (n_t[i] - n_t.m) + beta[3] * (d_t[i] - d_t.m)
  }
  beta_1 <- beta[1] - beta[2] * n_t.m - beta[3] * d_t.m
  for(i in 1:3){
    beta[i] ~ dnorm(a[i], b[i])
  }
}
```

Uma das inovações do JAGS, acima ilustrada, é a introdução de um novo bloco (`data{}`) que pode ser usado para a transformação de dados e que pode incluir relações estocásticas, podendo assim ser usado para simular conjuntos de dados, por exemplo.

3 O contra-ataque de Laplace (R-INLA)

O domínio quase absoluto da simulação de MCMC na análise de modelos hierárquicos bayesianos, iniciado pelo BUGS, durou quase 20 anos até que, em 2009, Håvard Rue e os seus colaboradores propõem uma nova abordagem. Em sua publicação [8], os autores descrevem o método INLA (*Integrated Nested Laplace Approximation*) que veio revolucionar a maneira como modelos hierárquicos bayesianos, por natureza complexos, podem ser agora analisados com extrema facilidade e principalmente, com uma rapidez que antes não era possível. Com esta nova metodologia, baseada na teoria de aproximação de Laplace, não é mais necessário se preocupar com o problema da convergência inerente ao método de simulação MCMC, muitas vezes desprezado pelos utilizadores mais inexperientes. Além disso, o custo computacional sofre um drástico decréscimo, pois sua abordagem baseia-se na aproximação de forma direta das distribuições *a posteriori* de interesse. Contudo, nada é graça. O método de aproximação INLA só pode ser aplicado a uma classe muito abrangente de modelos denominados modelos latentes gaussianos (*latent Gaussian Markov Random Field (GMRF) models*). Por exemplo, fazem parte desta lista os modelos lineares generalizados mistos, passando pelos modelos aditivos, os modelos espaciais e espaço-temporais, até aos modelos de sobrevivência.

Na mesma altura da publicação do artigo de Rue e colaboradores [8], foi disponibilizado o pacote INLA [4] para o software R que até à data engloba uma grande parte dos modelos GMRF mais utilizados na literatura e dos quais se indicam, na tabela abaixo, os mais conhecidos:

Modelo	Nome no R-INLA
Independent random variables	iid
Linear	linear
Random walk of order 1 and 2	rw1, rw2
Continuous random walk of order 2	crw2
Model for seasonal variation	seasonal
Model for spatial effect	besag, besagproper
Model for weighted spatial effects	besag2
Model for spatial effect + random effect	bym
Autoregressive model of order 1 and p	ar1, ar
Classical random effect model	z
Random walk of 2nd order on a lattice	rw2d
Gaussian field with Matern covariance function	matern2d
Classical measurement error model	mec

A sintaxe do R-INLA baseia-se na estrutura do modelo linear generalizado do R [5]. Primeiro definimos o preditor linear através da fórmula:

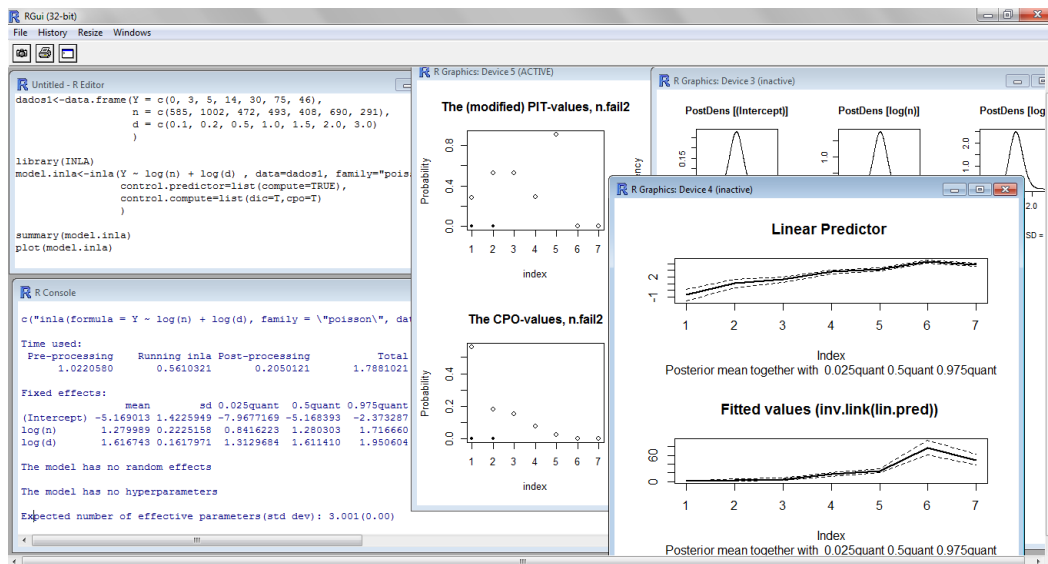
```
formula = y ~ a + b + c * d + f(grf, modelo)
```

em que o termo adicional `f()` serve para especificar os componentes dos efeitos aleatórios gaussianos (`grf`) dos modelos da tabela anterior. Depois de definido o preditor linear, o modelo é ajustado através da função `inla()`, como se mostra a seguir:

```
mod <- inla(formula, data = data.frame(a, b, c, d, graf), family = "poisson")
```

O modelo de regressão de Poisson pode ser analisado através de aproximação bayesiana utilizando o método INLA no R, da seguinte forma:

```
# Modelo R-INLA -----  
model.inla <- inla(formula = Y ~ log(n) + log(d), data = dados, family = "poisson",  
+ control.predictor = list(compute = TRUE), control.compute = list(dic = T, cpo = T))
```



A figura acima ilustra os resultados obtidos através das funções `summary()` e `plot()`.

O pacote INLA pode ser descarregado no site <http://www.r-inla.org>, onde também encontrará informações adicionais, como modelos que estão disponíveis para estimação e seus argumentos avançados, artigos e diversos exemplos ilustrados.

4 O novo miúdo em campo (STAN)

Nos anos mais recentes, o uso de modelos hierárquicos com centenas ou milhares de parâmetros esbarrou com a dificuldade do amostrador de Gibbs em lidar com parâmetros altamente correlacionados *a posteriori*. Esta dificuldade serviu de motivação para que, em 2001, uma equipa liderada por Andrew Gelman da *Columbia University* em Nova Iorque tenha iniciado o desenvolvimento de um novo software de MCMC – o STAN² [10]. Este software não recorre ao amostrador de Gibbs mas sim a uma variante de Monte Carlo Hamiltoniano que permite uma simulação rápida e particularmente eficiente de modelos complexos para os quais se observa frequentemente que o amostrador de Gibbs não converge num tempo razoável. Com o algoritmo escolhido (o *no-U-turn sampler* [2]) todos os parâmetros de um modelo são simulados em bloco e, desta forma, os problemas de convergência são substancialmente atenuados.

A descrição de modelos no STAN é ainda baseada na linguagem BUGS mas a equipa de desenvolvimento aproveitou a oportunidade para a repensar profundamente, corrigiu alguns erros ou inconsistências e conseguiu assim um verdadeiro salto tecnológico. A linguagem do STAN que, ao contrário do BUGS, não é declarativa mas sim imperativa, expande em muitos sentidos as capacidades de modelação do OpenBUGS. Como exemplos, salienta-se a possibilidade de usar variáveis locais e expressões condicionais, um vasto conjunto de funções que inclui as funções de distribuição e de probabilidade de muitas das mais usuais distribuições e, principalmente, o acesso direto ao logaritmo da distribuição *a posteriori* que evita o recurso a alguns truques que se tornaram populares no BUGS.

Como a maior poder deve corresponder maior responsabilidade, o STAN exige do utilizador uma programação mais rigorosa sendo obrigatório declarar todas as variáveis utilizadas, os seus tipos e, opcionalmente, os seus domínios. Outra característica nova é o desdobramento do usual bloco `model{}` do BUGS em seis possíveis blocos cujos nomes são esclarecedores: *data*, *transformed data*, *parameters*, *transformed parameters*, *model* e *generated quantities*. Esta aparente complicação visa eliminar a possibilidade de se incluir factores indevidos na distribuição *a posteriori*, um erro que é feito por vezes no BUGS quando este é usado de uma forma desleixada ou pouco informada.

²O nome do software é uma homenagem a Stanislaw Ulam, um dos criadores do método de Monte-Carlo.

Voltando ao modelo de regressão de Poisson que tem servido de exemplo, note-se que no STAN deixa de ser necessário centrar as covariáveis nos seus valores médios. O modelo pode ser representado no STAN com o código que se segue, mais extenso do que as versões anteriores mas também bem melhor organizado e, por isso, menos propenso a erros.

```
# Modelo STAN -----
data {
  int N;
  int Y[N];
  int n[N];
  real d[N];
  real a[3];
  real b[3];
}
parameters {
  real beta[3];
}
transformed parameters {
  real mu[N];
  for (i in 1:N) {
    mu[i] <- exp(beta[1] + beta[2] * log(n[i]) + beta[3] * log(d[i]));
  }
}
model {
  for (i in 1:N) {
    Y[i] ~ poisson(mu[i]);
  }
  for (i in 1:3) {
    beta[i] ~ normal(a[i], b[i]);
  }
}
```

Agradecimentos: este trabalho foi financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia no âmbito dos projetos PTDC/MAT/118335/2010 e PEst-OE/MAT/UI0006/2011.

Referências

- [1] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.
- [2] Matthew D. Hoffman and Andrew Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, In press.
- [3] David J. Lunn, Andrew Thomas, Nicky Best, and David Spiegelhalter. Winbugs — a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337, 2000.
- [4] S. Martino and H. Rue. *INLA: Functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. R package version 0.0.*, 2009.
- [5] Thiago G. Martins, Daniel Simpson, Finn Lindgren, and Håvard Rue. Bayesian computing with inla: New features. *Comput. Stat. Data Anal.*, 67:68–83, 2013.
- [6] Martyn Plummer. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. March, pages 20–22, 2003.
- [7] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [8] H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392, 2009.
- [9] D. Spiegelhalter, A. Thomas, N. Best, and W. Gilks. *BUGS – Bayesian inference using Gibbs sampling: Version 0.6*. MRC Biostatistics Unit, Cambridge, 1997.
- [10] Stan Development Team. *Stan Modeling Language User’s Guide and Reference Manual, Version 1.3*, 2013.



Uma pequena conversa sobre Estatística Bayesiana em Genética e Genômica

Nuno Sepúlveda, *nuno.sepulveda@lshtm.ac.uk*

London School of Hygiene and Tropical Medicine e Centro de Estatística e Aplicações da Universidade de Lisboa

1 Introdução

A Estatística Bayesiana tem vindo a introduzir-se paulatinamente em diversas áreas científicas, nomeadamente em Genética e Genômica, tornando-se numa ferramenta incontornável de análise de dados. Para além da sua elegância e coerência teórica, a Estatística Bayesiana mostra a vantagem inegável de poder incorporar na análise toda a incerteza associada ao desconhecimento dos parâmetros, a recolha e processamento de dados, e sobre todos os mecanismos hipotéticos para o fenómeno em estudo. Em termos gerais, o desbloqueamento inferencial de muitos problemas científicos tem vindo a ser conseguido através de formulações bayesianas gerais, aliadas a poderosos algoritmos de simulação estocástica, tal como os métodos Monte Carlo baseados em Cadeias de Markov (MCMC), que permitem obter as distribuições *a posteriori* dos parâmetros de interesse. Este trabalho tem, então, como objetivo dar uma panorâmica de algumas contribuições bayesianas ao campo da Genética e Genômica. Para o leitor mais interessado nestas matérias, deixa-se aqui a referência a um excelente artigo de revisão escrito numa linguagem muito acessível que, embora não seja muito recente, resume o grande interesse que a comunidade genética tem prestado à Estatística Bayesiana (Beaumont e Rannala, 2004).

2 Estatística Bayesiana em Genética

Genética é a ciência que estuda as leis da hereditariedade. É também a ciência que estuda os genes, as suas variantes e as suas interações na construção de uma determinada característica biológica — o fenótipo —, tal como a presença ou ausência de uma determinada doença num indivíduo. Em termos de Biologia Molecular, os genes são sequências de nucleótidos (A, C, G e T) de ADN que codificam uma determinada proteína com alguma função fisiológica no organismo. As diferentes variantes do mesmo gene são denominadas de alelos. Num organismo haploide, tal como é o caso do parasita *Plasmodium falciparum* que provoca o paludismo, existe apenas uma cópia de um determinado gene, enquanto num organismo diploide, tal como é a espécie humana, existem duas cópias do mesmo gene, cada uma herdada de cada progenitor, definindo o que se chama de genótipo. Por simplicidade, o genoma é aqui definido pelo conjunto de todos os genes de um organismo, ordenados em estruturas moleculares a que se dá o nome de cromossomas. Todos os processos moleculares, fisiológicos e celulares que vão desde a expressão de um ou mais genes até a observação do fenótipo formam o "plano de construção" ou a "arquitetura" genética do fenótipo. Por facilidade de análise, estuda-se a correspondência direta entre genótipo e fenótipo, não considerando os inúmeros processos biológicos intermédios.

Um dos principais objetivos inferenciais em Genética prende-se com a identificação dos genes que fazem parte da arquitetura genética de um determinado fenótipo. Com esse objetivo recorre-se frequentemente a variações em sequências de ADN conhecidas e localizadas em posições específicas do genoma. Estas

variações são denominadas por marcadores genéticos. Os mais conhecidos e usados na prática são os polimorfismos num único nucleótido (SNP - *single nucleotide polymorphism*), onde uma determinada região do genoma apresenta variações na população num nucleótido específico. Os chamados estudos de associação em todo o genoma (GWA - *genome-wide association studies*) usam milhares, e por vezes milhões, destes SNP com o objetivo de detetar as regiões genómicas mais associadas com o fenótipo. Tipicamente, a análise é conduzida em cada SNP individualmente através de modelos de regressão (*e.g.*, modelos de regressão logística para fenótipos binários), nos quais se descrevem diversos efeitos genéticos do SNP em causa sobre o fenótipo, e se incluem covariáveis que ajustam a análise para eventuais efeitos de confundimento, ou efeitos contraditórios provenientes de subestruturas populacionais latentes nos dados. A associação é então medida por $-\log_{10}$ (valor-p) estando relacionada com a significância do efeito genético do SNP no modelo. Como os estudos GWA visam avaliar esta associação estatística ao longo de todo o genoma, há a necessidade em termos frequentistas de corrigir o nível de significância de cada teste individual - por vezes, pelo método de Bonferroni - de forma a garantir um determinado nível de significância global da análise. Esta correção é normalmente definida em função do número de testes a executar, ou por outras palavras, em função do número de SNP avaliados. Como consequência prática, os tamanhos amostrais usados neste tipo de estudos rondam a ordem dos milhares de indivíduos para fazer face ao limiar extremamente exigente imposto para a significância estatística dos resultados ($\approx 5 \times 10^{-8}$). Para além desta dificuldade amostral, existem essencialmente mais dois problemas de análise, um de cariz prático e outro de foro mais teórico. Por um lado, mesmo no caso de um estudo bem delineado, a potência estatística pode diminuir drasticamente com a presença de dados omissos em marcadores testados, ou de não se ter considerado marcadores genéticos suficientemente próximos do gene causal (marcadores não-testados). Por outro lado, é sabido que os fenótipos de grande interesse para a saúde pública seguem uma arquitetura genética complexa, envolvendo interações de múltiplos genes e de fatores ambientes. Mais alarmante é o facto de se conhecerem situações (teóricas) de interações genéticas onde o aumento do tamanho da amostra não implica necessariamente uma maior probabilidade de deteção dos genes envolvidos (Moore e Williams, 2005). Deste modo, uma análise de SNP em SNP, embora muito popular, está condenada a fornecer uma imagem parcial da verdadeira arquitetura genética subjacente ao fenótipo.

Neste contexto, a Estatística Bayesiana mostra inúmeras vantagens em relação à sua contrapartida frequentista. Mesmo no caso simples em que existe apenas um gene a atuar sobre o fenótipo, como acontece nas doenças mendelianas, a Estatística Bayesiana permite confrontar a hipótese de ausência de associação contra diferentes modelos de associação alternativos, e combinar os respetivos resultados num único fator de Bayes por um processo de ponderação das inferências produzidas pelos vários modelos candidatos (BMA - Bayes model averaging), tal como discutido em Stephens e Balding (2009). Este procedimento tem a desejada propriedade da probabilidade de falsos positivos variar em função da proporção de testes que favorecem a hipótese nula e não do número de testes executados (Stephens e Balding, 2009).

No caso da análise de dados omissos em marcadores testados ou não-testados, é usual aplicar-se métodos de imputação onde os genótipos em falta são substituídos por um valor plausível, derivado a partir da forte estrutura de correlação existente entre marcadores próximos, e do uso simultâneo de uma amostra adicional proveniente de uma população de referência (tradicionalmente, uma das populações estudadas no projeto Hapmap — *vide* www.hapmap.org). Embora haja diversos métodos frequentistas a exibirem bons desempenhos em termos de precisão dos genótipos imputados, as suas contrapartidas bayesianas trazem consigo a quantificação correta da incerteza associada ao processo de imputação, assim como a possibilidade de se desenvolverem abordagens integradas com o objetivo principal da análise, o de deteção da associação genética, tal como proposto por Zhang (2011).

No que diz respeito a modelação da interação genética num contexto de estudos de associação, atuais métodos bayesianos permitem explorar o universo das interações muito mais eficientemente do que os seus rivais frequentistas. Estes métodos vão desde o uso de redes ou grafos bayesianos (Kozyryev e Zhang, 2012) à utilização de modelos de regressão acoplados a procedimentos do tipo BMA (Ferreira e Marchini, 2011). Mesmo em situações onde é possível aplicar ambos os princípios inferenciais, tal

como é o caso da formulação de diferentes mecanismos genéticos baseados na probabilidade de um determinado alelo expressar o fenótipo (Sepúlveda, 2009), a análise frequencista pode produzir resultados inferenciais pouco fidedignos quando a probabilidade de expressão alélica está próxima de 0 ou 1, enquanto na análise bayesiana essa situação pode ser descrita explicitamente na modelação.

Outro dos principais objetivos em Genética centra-se no conhecimento da composição e estrutura genética de uma população, e na compreensão da sua evolução por mecanismos de seleção natural, migração ou alterações ambientais. Um exemplo muito popular em genética humana é uma mutação — neste caso um SNP — no gene da hemoglobina β , que está na origem da anemia falciforme, mas que tem vindo a ser selecionada em África, devido ao seu cariz protetor contra o paludismo (Ashley-Koch *et al.*, 2000). Os métodos bayesianos proliferam nesta área de genética de populações, nomeadamente em aplicações que tentam reconstruir os eventos genéticos que conduziram a uma determinada composição genética da população (Wilson *et al.*, 2003). Neste contexto específico a Estatística Frequencista falha em termos do seu fundamento principal, pois a história de uma população não é conceptualmente equivalente a uma experiência infinitamente repetível. Assim, todas as inferências produzidas por esta abordagem são inevitavelmente limitadas e conceptualmente incoerentes. Por seu lado, a Estatística Bayesiana, usando muitas vezes árvores filogenéticas, é muito apelativa não só pela sua coerência teórica, mas também pela sua vertente prática. Vários modelos demográficos podem ser propostos e comparados diretamente através de fatores de Bayes. O uso de distribuições *a priori* permite incluir na análise informação externa aos dados, tal como registos fósseis, históricos ou de outra natureza, refletindo assim mais cabalmente o *status quo* sobre o que se conhece sobre a população em estudo. Por fim, os métodos MCMC ou de computação bayesiana aproximada (ABC) permitem executar inferências *a posteriori* com boa precisão em modelos demográficos tão complexos, que seriam impensáveis do ponto de vista frequencista. Outras aplicações de sucesso na área da genética de populações podem ser encontradas em Beaumont e Rannala (2004).

3 Estatística Bayesiana em Genómica

Sendo tipicamente classificada como um ramo da Biologia Molecular, a Genómica é a ciência que estuda a estrutura, a função e a evolução de genomas. É uma área científica onde abunda grande volume de dados produzidos por tecnologias experimentais de ponta, tais como diferentes tipos de microarranjos (*microarrays*) para expressão genética. Atualmente as chamadas tecnologias de sequenciação de nova geração estão muito em voga devido à sua capacidade de estudar um grande número de genomas a baixo custo e com alta resolução, dependendo do tamanho do genoma em questão. Estas tecnologias estão baseadas na geração de milhões de pequenos fragmentos de ADN (*reads*) provenientes de um genoma alvo. As sequências desses fragmentos de ADN são determinadas e, quando possível, alinhadas (ou mapeadas) a um genoma de referência — o da estirpe 3D7 no caso de estudos genómicos do parasita *Plasmodium falciparum* (Sepúlveda *et al.*, 2013). Após este processo de alinhamento prossegue-se para a análise da estrutura do genoma propriamente dita.

Os objetivos de análise podem ser diversos, mas atualmente há um enorme interesse em identificar e catalogar toda a variação genómica presente numa determinada população de indivíduos. Este interesse vem da observação de que os genes encontrados em inúmeros estudos GWA apenas explicam uma fração reduzida da variação fenotípica observada. Em termos técnicos, a variação genómica pode ser de várias ordens, incluindo os populares SNP, as inserções-remoções (*indels*), alterações do número de cópias, inversões, entre outras. Tal como explicado anteriormente, os SNP são variações numa sequência de referência em apenas um único nucleótido. As variações *indels* são definidas por inserções de novos nucleótidos ou remoções de nucleótidos a uma sequência de referência. Alterações do número de cópias são várias repetições consecutivas da mesma sequência de referência. Por sua vez, as inversões são sequências que estão invertidas em relação às encontradas no genoma de referência.

Estes diferentes tipos de variação genómica foram encontrados em vários estudos de grande relevância para a saúde pública. Tal como mencionado anteriormente, um SNP no gene da hemoglobina β parece

ser importante na proteção das populações africanas contra o paludismo. Outro exemplo é o aumento do número de cópias do gene *pfMDR* em relação ao genoma 3D7 do parasita *Plasmodium falciparum*, aumento esse associado à resistência deste parasita a diversos medicamentos antipaludismo no Sudoeste Asiático (Price et al, 2004). Tendo sido identificada e catalogada toda esta variação genómica, o ultimo passo da análise consiste em estudá-la do ponto de vista da genética de populações, ou em termos de associação com um fenótipo de interesse.

As várias etapas de uma análise genómica baseada em dados gerados por estas novas tecnologias de sequenciação têm um grande pendor computacional mas também estatístico. Um dos principais fatores para o uso de ideias bayesianas reside no facto de se poder incorporar mais cabalmente toda a incerteza associada à probabilidade de ocorrência de erros experimentais ao longo do processamento de dados, tais como erros de sequenciação ou de mapeamento. Outro fator importante para o uso de metodologia bayesiana consiste na possibilidade de trazer para a análise informação *a priori* específica sobre regiões genómicas que tendem a mostrar variação genómica, ou a exibir erros de sequenciação ou de mapeamento. Estes fatores estão na génese de algumas metodologias bayesianas de determinação de sequências, onde se incorporam explicitamente as diferentes fontes de erro durante o processo de sequenciação propriamente dito (Li et al., 2008; Kao et al., 2009). O mesmo se passa quando se procura determinar o alinhamento mais provável de uma determinada sequência em relação à de referência. Os atuais algoritmos bayesianos de alinhamento permitem modelar corretamente a incerteza associada à magnitude, variabilidade e localização do erro de alinhamento ao longo do genoma, tarefa passível de muita crítica quando realizada por métodos frequentistas. Na deteção de variações em número de cópias é usual analisar-se dados de cobertura genómica definida pelo número de fragmentos mapeados para cada uma das regiões associadas a uma partição do genoma em fragmentos de um determinado tamanho. Em teoria, procuram-se detetar regiões onde a cobertura é muito menor do que a esperada sob a hipótese nula de um número "normal" de cópias, indicando possivelmente um decréscimo no número de cópias. Procuram-se também identificar regiões onde a cobertura é muito maior do que a esperada sob a hipótese nula, sugerindo um aumento do número de cópias. Vários trabalhos mostraram que este tipo de dados pode ser descrito por várias distribuições de probabilidade, tal como a distribuição Normal, Poisson, Poisson-Gama, ou Poisson-Lognormal (Sepúlveda et al., 2013). Neste contexto, a análise bayesiana permite combinar os resultados produzidos pelos vários modelos candidatos através de um procedimento BMA e, assim, produzir inferências sobre regiões com alteração do número de cópias mais flexíveis e com menor proporção de falsos positivos.

A aplicação da metodologia bayesiana à Genómica não se esgota com a análise de dados provenientes das modernas máquinas de sequenciação. De facto, existe uma panóplia de métodos bayesianos aplicados a diferentes tecnologias de grande volume de dados genómicos. Uma aplicação recente é o uso de metodologia bayesiana semiparamétrica a dados gerados pela tecnologia SAGE (*Serial Analysis of Gene Expression*) para comparar a abundância de um elevado número de tipos de ARN mensageiro - moléculas derivadas do ADN de um determinado gene e que transportam toda a informação genética necessária para a síntese da correspondente proteína pelo ribossoma - sob diferentes condições biológicas ou experimentais (Guindani et al., 2013). Partindo do pressuposto de que o número de cópias de um determinado tipo de ARN mensageiro segue uma distribuição Poisson com um valor médio variável na população, o grande avanço desta metodologia em relação às suas contrapartidas frequentistas reside na modelação explícita de toda a incerteza associada à própria distribuição de abundância média de cada tipo de ARN mensageiro. Neste contexto, a distribuição das abundâncias médias é vista como uma função de probabilidade desconhecida e, assim, passível de ser modelada *a priori* - neste caso por uma mistura por Processo Dirichlet. Deste modo, as inferências captam mais cabalmente toda a incerteza inerente aos dados. Um outro exemplo refere-se à deteção de alterações de número de cópias através da tecnologia aCGH (*array Comparative Genomic Hybridization*). Esta tecnologia segue um princípio experimental semelhante à dos microarranjos: fragmentos de ADN de um genoma alvo conjugam-se com os de um genoma de referência presente num microarranjo, conjugações essas convertidas posteriormente em intensidades normalizadas de luz. Tendo calculado os perfis de intensidade ao longo de todo o genoma, o objetivo consiste em detetar regiões genómicas com intensidades extremamente baixas

ou altas, sugerindo potenciais decréscimos ou acréscimos, respetivamente, em relação ao número de cópias encontrado no genoma de referência. Uma abordagem estatística muito popular para este problema consiste em segmentar os perfis de intensidade em blocos de intensidade estatisticamente equivalente de forma a tentar descobrir os pontos de mudança do número de cópias. Muitas propostas frequencistas existem para executar tal tarefa, mas nenhuma delas consegue modelar toda a incerteza associada ao processo de segmentação, nomeadamente na determinação dos pontos de mudança no número de cópias, tal como por metodologia bayesiana (Baladandayuthapani *et al.*, 2010). Os métodos bayesianos são ainda capazes de produzir resultados mais intuitivos para os seus utilizadores do que os seus congéneres frequencistas, tal como associar a cada segmento obtido uma probabilidade *a posteriori* de conter uma alteração do número de cópias.

4 Conclusões finais

Em suma, pensa-se ter demonstrado que a Genética e Genómica estão recheadas de questões estatísticas onde a incerteza precisa de ser quantificada adequadamente. No campo da Genética, que assenta num edifício teórico sólido, interessa estudar a plausibilidade relativa de várias hipóteses ou modelos competidores para os dados. No campo da Genómica, que se baseia fortemente em grande volume de dados, interessa produzir inferências que captam não só os princípios de amostragem inerentes, mas também eventuais erros ou enviesamentos experimentais que possam ocorrer durante a geração e processamento de dados. Neste cenário, a Estatística Bayesiana oferece uma abordagem estatisticamente sólida e coerente, mas suficientemente flexível para dar resposta aos inúmeros problemas que estas duas disciplinas enfrentam atualmente. Com o constante desenvolvimento tecnológico que se tem vindo a testemunhar na última década, prevê-se que o volume de dados disponível aumente drasticamente, assim como em termos de complexidade. Assim, o desafio futuro da Estatística Bayesiana, bem como da Estatística em geral, passa por continuar a ser exequível em tempo real. É certo que a capacidade computacional continuará a aumentar, nem que seja por força da computação paralela atualmente em voga. Contudo, resta a nós, estatísticos, a tarefa de nos prepararmos para afiar a navalha de Occam e, assim, encontrarmos modelos parcimoniosos para os complexos problemas genéticos e genómicos que o futuro nos reserva!

Agradecimentos

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

Referências

- [1] Ashley-Koch, A., Yang, Q., e Olney, R. S. (2000). Sickle hemoglobin (HbS) allele and sickle cell disease: a HuGE review. *Am. J. Epid.* 151:839-845.
- [2] Baladandayuthapani, V., Ji, Y., Talluri, R., Nieto-Barajas, L. E., e Morris, J. S. (2010). Bayesian Random Segmentation Models to Identify Shared Copy Number Aberrations for Array CGH Data. *J. Am. Stat. Assoc.* 105:1358-1375.
- [3] Beaumont, M. e Rannala, M. (2004). The Bayesian Revolution in Genetics. *Nat. Rev Genet.* 5:251-261.
- [4] Ferreira, T. e Marchini, J. (2011). Modeling interactions with known risk loci - a Bayesian model averaging approach. *Ann. Hum Genet.* 75, 1-9.
- [5] Guindani, M., Sepúlveda, N., Paulino, C. D., e Müller, P. (2013). A Bayesian Semi-parametric Approach for the Differential Analysis of Sequence Counts Data. *App. Stat.* (no prelo).

- [6] Kao, W.-C., Stevens, K., e Song Y.-S. (2009). BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing. *Genome Res.* 19:1884-1895.
- [7] Kozyryev I., e Zhang, J. (2012). Bayesian Exploration of Multilocus Interactions on the Genome-Wide Scale. *Am. J. Bioinform.* 1:70-78.
- [8] Li, H., Ruan, J., e Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18:1851-1858.
- [9] Moore, J. H., e Williams, S. M. (2005). Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *Bioessays* 27:637-646.
- [10] Price, R. N., Uhlemann, A. C., Brockman Price, R. N., Uhlemann, A. C., Brockman A., McGready, R., Ashley, E., Phaipun, L., Patel, R., Laing, K., Looareesuwan, S., White, N. J., Nosten, F., e Krishna, S. (2004). Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* 364:438-447.
- [11] Sepúlveda N. (2009). Abordagem por penetrâncias alélicas: uma nova modelação estatística para a análise de fenótipos binários complexos. Em *Estatística: Arte de explicar o acaso* (Oliveira I, Correia E, Ferreira F, Dias S, Braumann C, eds.), Edições SPE, 109-129.
- [12] Sepúlveda, N., Campino, S. G., Assefa, S. A., Sutherland, C. J., Pain, A., e Clark, T. G. (2013). A Poisson hierarchical modelling approach to detecting copy number variation in sequence coverage data. *BMC Genomics* 14:128.
- [13] Stephens, M. e Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* 10:681-690.
- [14] Wilson, I. J., Weale, M. E., e Balding, D. J. (2003). Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. Roy Stat. Soc. A* 166: 155-201.
- [15] Zhang, Y. (2011). Bayesian epistasis association mapping via SNP imputation. *Biostatistics* 12: 211-222.

Adenda: Este texto foi redigido segundo as normas do AO 1990.



Tese de Doutoramento: META-ANÁLISE – Harmonização de testes usando os valores de prova

(Boletim SPE outono de 2009, p.96)

Fernando Sequeira, fjsequeira@fc.ul.pt

*Faculdade de Ciências da Universidade de Lisboa
Centro de Estatística e Aplicações da Universidade de Lisboa*

Doutoramento, e o que vem depois

Conta-se que Florence Nightingale David, quando perguntaram que diferença tinha feito o seu doutoramento, respondeu “vinte libras na carteira”. Sem me poder comparar a essa investigadora de exceção, posso contudo comungar da ideia que o doutoramento foi uma prova, necessária para me manter a fazer o que gosto — levar alguns alunos a gostar de Probabilidade e Estatística —, que me obrigou a ganhar conhecimentos e desenvolver algumas novas metodologias em meta análise, gastar tempo e porventura um pouco mais do que 20 libras, ter muitas vantagens, como passar de convidado a docente de carreira, e numa posição melhor. Mas, substancialmente, tenho que fazer *mea culpa*, não mudou de facto, nada, pois continuo a trabalhar com os meus seniores, Dinis Pestana e Maria de Fátima Brilhante, com alguma habilidade para resolver com alguma elegância e rigor problemas matemáticos, mas muito dependente deles no que toca a um ponto essencial, que é a identificação de problemas que vale a pena resolver.

No meu doutoramento estabeleci alguns resultados sobre maus resultados inesperados que advinham do aumento computacional de amostras, quando se esperava que trabalhar com amostras maiores seria sempre benéfico. Ter explicado rigorosamente porque tal pode acontecer é um resultado que me parece interessante, e que nos pós-doutoramento descobri que tinha consequências importantes na combinação de valores de prova generalizados. Isso trouxe-me a novos desenvolvimentos na investigação de questões de viés de publicação em meta análise, tema que dominara o último capítulo da minha tese. Isso permitiu-me orientar uma aluna de mestrado nesse problema, tal como a questão de atração para a uniformidade na ampliação computacional de amostras me levou a orientar um aluno de mestrado na área de informação e entropia, e seu papel em caracterização de leis limites, em algum sentido.

Os temas atrás descritos foram objecto de publicações indexadas.

Recentemente houve mais uma ampliação dos meus interesses, de novo por influência de Dinis Pestana: por um lado, colaboração no trabalho de tradução de textos contendo componente estatística, por outro lado um interesse ainda muito incipiente pela relevância da Estatística na transformação de informação em conhecimento, nas ciências de factos.

Recentemente estreei-me também na orientação de um doutoramento, pois a passagem de testemunho a gerações vindouras é decerto uma das responsabilidades da carreira. Realço, de novo, que esta estreia é feita com os meus mentores de sempre a amparar estes primeiros passos, e sinto que vou ter que aprender tanto ou mais do que o nosso orientado.

O meu doutoramento foi tardio, pois pelo caminho tropecei muitas vezes, algumas delas por culpa minha, outras por necessidade da instituição, que me pediu que interrompesse a dispensa de serviço docente para reger cadeiras, quando a ida de professores para sabática deixou áreas como Amostragem a descoberto. Quero no entanto deixar claro que fui fortemente apoiado pela hierarquia do meu departamento nomeadamente pelos Professores Doutores Maria Antónia Turkman e Kamil Feridun Turkman, que presidiam aos destinos do DEIO em períodos cruciais para a renovação dos meus contratos. De qualquer modo, a minha experiência leva-me a dizer aos vindouros que, se pretendem fazer doutoramento, embarquem de alma e coração o mais cedo possível na tarefa, e despachem-se a ficar com menos 20 libras no bolso (ou o que a inflação ditar).

Fernando Sequeira



Marta Ferreira, *msferreira@math.uminho.pt*

Departamento de Matemática e Aplicações e Centro de Matemática da Universidade do Minho

O grau de doutorada foi adquirido a 12 de Dezembro de 2008, na Universidade de Lisboa, ao final da manhã. Um respirar fundo, num lindo dia de sol, a descomprimir os nervos e a ansiedade do momento, ou talvez acumulados ao longo dos anos dedicados a este projeto. O fechar de um duradouro ciclo (desde de 2002) de visitas regulares à minha querida professora Luísa Canto e Castro, minha orientadora de mestrado e de doutoramento, portanto, minha “mãe” de investigação. Aliás, “mãe” é uma boa palavra para definir a professora Luísa, dada a sua infinita generosidade e um imenso colo pronto a acolher o mundo. Muito lhe devo do que sei e um pouquinho do que sou, pois não deixa de ser para mim uma referência presente. Acredito também que a sua intuição (aquela das mães) a levou a colocar a professora Helena Ferreira da Universidade da Beira Interior no meu júri de doutoramento, a semente do meu pós-doc. Pois bem, a professora Helena, de uma imaginação prodigiosa sempre a fervilhar novas ideias, contactou-me no sentido de estender conceitos abordados na minha tese de doutoramento. A partir daqui, o epíteto “professora” caiu, pois, a par de um intenso trabalho de colaboração científica, inicia-se uma amizade, para mim, inestimável. E mais um modelo de verdadeira generosidade aparece no meu caminho, numa espécie de sinal e de responsabilidade no perfil de conduta a seguir. Estou pois muito grata pela experiência que a carreira académica, desenvolvida na Universidade do Minho, hoje professora auxiliar, me tem proporcionado, quer ao nível de exigência intelectual, quer ao nível de riqueza pessoal.

Pelas duas pessoas que referi, e para quem conhece um pouco do meio, alguns já deduziram que a minha área de investigação é a teoria de valores extremos. O tema da minha tese insere-se no estudo do comportamento extremal de séries temporais auto-regressivas de máximos (e.g., considerar os modelos ARMA, substituindo o operador “soma” pelo operador “máximo”). O estudo de medidas de dependência ao nível da cauda impulsionou a criação de novos modelos contemplando independência assintótica, situação não pouco usual em dados reais, mas pouco explorada na literatura ao nível de modelos de séries temporais. Uma metodologia de ajustamento dos modelos criados foi implementada e aplicada com sucesso a uma série de dados de índices de bolsa. Numa primeira fase pós-doc, e por conta própria, introduzi um novo estimador para o parâmetro de um dos modelos desenvolvidos na tese e explorei um pouco mais a dependência serial de cauda dos mesmos.

As medidas de dependência de cauda são usualmente definidas a partir de eventos extremos do tipo “excedência” de um valor elevado ($X > x$, para x elevado). Contudo, outros eventos extremos são também relevantes num contexto de risco, nomeadamente, o “cruzamento” de um valor elevado ($X_i \leq x < X_{i+1}$, para x elevado). Por sugestão da Helena, desenvolvemos novas medidas de dependência de cauda, adaptando o conceito de cruzamentos ao das medidas usuais, estabelecendo também relações entre os dois e, sob condições de dependência local convenientes, relações com

outros coeficientes de dependência extremal da literatura, como o índice extremal e o índice de cruzamentos. Sempre sob a batuta da Helena, outras medidas de dependência de cauda foram implementadas, permitindo avaliar situações como, até que ponto o k-ésimo melhor (s-ésimo pior) desempenho num sistema é atraído pelo s-ésimo pior (k-ésimo melhor), dado que este último tem muito mau (bom) desempenho, em que medida a queda de pelo menos um mercado bolsista europeu influencia também a ocorrência de pelo menos um *crash* num mercado norte-americano, ou ainda, avaliar a fragilidade de um sistema dividido em blocos de componentes, no sentido de apurar a sua estabilidade sistémica com base no número esperado de blocos onde ocorre pelo menos uma excedência de níveis elevados, sabendo que ocorreu pelo menos uma excedência nalgum dos blocos. Mais recentemente, temo-nos focado em séries temporais multivariadas, nomeadamente, na implementação de novos modelos e respetivo estudo do seu comportamento extremal, tendo em vista a modelação da cauda de séries de dados reais multivariados. Aqui incluem-se versões multivariadas dos modelos auto-regressivos de máximos estudados na minha tese.

Entretanto, outras colaborações pontuais surgiram. Na sequência de um seminário e de um mini-curso lecionado pelo professor Víctor Leiva (Universidade de Valparaíso, Chile), na minha universidade, travei conhecimento com os modelos de Birnbaum-Saunders, membros da grande família dos modelos de vida, com vasta aplicação sobretudo na área ambiental. Acontece que estes modelos apresentam alguma dificuldade justamente na modelação das caudas. Foi assim que, contando com a preciosa colaboração da professora Maria Ivette Gomes (conhecida demais para dispensar afiliações), desenvolvemos uma nova versão dos mesmos, baseada numa distribuição generalizada de valores extremos. Isto levou à introdução de um novo parâmetro no modelo, o índice de cauda, o qual permite medir o peso da cauda e, conseqüentemente, um melhor ajustamento da mesma.

A categoria de professora auxiliar acarreta, entre outras coisas, a orientação de alunos. Na Universidade do Minho, no campus de Azurém (Guimarães), temos a funcionar uma licenciatura em Estatística Aplicada, cujo último semestre inclui a realização de um estágio profissional. Cada aluno tem necessariamente um orientador proveniente do local de estágio e um orientador da universidade. Orientei alguns estágios neste âmbito, entre eles, um no INE com a colaboração do professor Pedro Campos, na temática da estimação em pequenos domínios. Devo confessar que foi a primeira vez que me cruzei com este assunto que é de grande interesse e aplicabilidade na produção de estatísticas oficiais. Na prática, e devido à crescente necessidade de informação precisa e atualizada, torna-se necessário produzir estatísticas também para frações muito pequenas da população, cuja reduzida dimensão amostral compromete a fiabilidade das mesmas. Daí a necessidade de implementação de métodos indiretos, normalmente, recorrendo a modelos de regressão e informação temporal/espacial disponível, com vista à produção de estimativas com precisão aceitável. Entretanto, temos contado também com a sábia colaboração de Inês Sousa, uma minha colega de departamento.

Além disso, encontro-me a orientar uma tese de mestrado sobre a estimação de medidas de dependência da cauda, fruto dos estudos da minha preciosa colaboração com a Helena, com a qual aprendo sempre e espero poder continuar a sorver cada pedacinho do seu imenso saber, que a sua infinita generosidade permite partilhar comigo.

No meio de tudo isto, a experiência da maternidade, uma nova aventura ... a melhor de todas.

Marta Ferreira



Tese de Doutorado: Extremum Estimators and Stochastic Optimization Methods

(Boletim SPE outono de 2009, p. 100)

Miguel de Carvalho, *MdeCarvalho@mat.puc.cl*

*Department of Statistics
Pontificia Universidad Catolica de Chile*

Caros colegas e amigos,

Realizei as minhas provas públicas de Doutorado na Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa (UNL), no dia 20 de maio de 2009. Foi um dia memorável: separa aquilo que chamo por vezes, em jeito de piada, a época A.D. (*Antes do Doutorado*) e D.D (*Depois do Doutorado*)! A época D.D. será o foco desta pequena nota, mas vou começar por fazer um pequeno apanhado sobre a época A.D., que na verdade foi bastante turbulenta!

Época A.D.

Defendi a minha tese de Mestrado em Econometria na Faculdade de Economia, Universidade Nova de Lisboa, no dia 20 de março de 2009 - e portanto somente dois meses antes das minhas provas públicas de Doutorado. A minha tese em Econometria foi sobre Análise de Duração, que é essencialmente a designação que os econometristas dão à Análise de Sobrevivência. Foi realmente uma época altamente conturbada, mas nada que uns Capuccinos da Nestlé adoçados com a motivação certa não permitam resolver!

Depois de terminar a minha licenciatura em Matemática Aplicada, candidatei-me ao PDM (Programa de Doutorado) em Economia da atual *NOVA School of Business and Economics* onde fui aceite em 2004-2005. Dada a minha formação base, ansiava por aplicar conceitos matemáticos a modelos económicos e recordo hoje com saudade as cadeiras de Macroeconomia, Microeconomia, Teoria de Jogos, entre outras. Era um desafio competir com colegas licenciados e mestres em Economia, mas com dedicação e esforço acabei por concluir ano e meio de parte curricular, até que percebi que o caminho que gostaria de percorrer seria outro. A verdade é que aprendi a gostar de Estatística nas aulas de Econometria do PDM, com o Prof. José Machado - que mais tarde foi meu orientador de mestrado. As aulas de Econometria dos Profs. Catela Nunes e Pedro Portugal cimentaram a certeza de que a Estatística era a minha verdadeira paixão e que portanto devia antes apostar num doutorado em Estatística e não num doutorado em Economia. Dos meus orientadores de doutorado - Profs. Manuel Esquível e Tiago Mexia - herdei o gosto por modelação probabilística; dos meus orientadores de mestrado - Profs. José Machado e Pedro Portugal - herdei o gosto por modelação econométrica.

Conheci a minha esposa - Vanda de Carvalho - durante a época A.D. O Prof. Mexia dizia, em jeito de piada (ou não!), que a Vanda fez de mim um rapaz ajuizado. E tinha toda a razão, há definitivamente uma época A.V. (*Antes de Vanda*) e D.V. (*Depois de Vanda*)...

A minha tese de doutoramento incidiu sobre o desenvolvimento de métodos de otimização estocástica para a classe de estimadores extremos (*extremum estimators*); para uma introdução aos métodos de pesquisa e otimização estocástica ver Spall (2003). A classe de estimadores extremos inclui todos os estimadores que podem ser definidos através de um problema de otimização e portanto inclui por exemplo máxima verosimilhança, mínimos quadrados, GMM (Generalized Method of Moments), entre muitos outros. Enquanto desenvolvia a minha tese fui sempre acompanhando com curiosidade o movimento da 'escola de extremos' em Portugal, particularmente dos Profs. Frederico Caeiro, Isabel Fraga Alves, Helena Ferreira, Ivette Gomes, Cláudia Neves, Manuel Scotto, Feridun Turkman, entre muitos, muitos outros; já durante essa fase tive sempre muita curiosidade pelo trabalho do Prof. Tiago de Oliveira, que nunca tive a oportunidade de conhecer. Desde que iniciei o meu doutoramento que mantive a suspeita de que existiria uma relação entre otimização estocástica e a estatística de valores extremos - uma vez que ambas as disciplinas se preocupavam com a modelação de máximos; vim a descobrir mais tarde que Laurens de Haan já tinha formalizado tudo isto, mais ou menos na mesma altura em que eu começava a dizer as minhas primeiras palavras (de Haan, 1981)!

Época D.D.

"A quem muito foi dado, muito será exigido; e a quem muito foi confiado, muito mais será pedido." (Lucas 12, 42-48)

A transição de A.D. para D.D. é fascinante, mas envolve grande responsabilidade. Apesar de todas as mudanças envolvidas durante esta transição, aprendi que há coisas que não devem mudar nunca, incluindo: a busca pelo conhecimento e pela verdade. A minha família tem sido sempre muito compreensiva e tenho feito muitos amigos ao longo de toda esta aprendizagem.

Durante esta fase comecei também a trabalhar em tópicos diferentes daquele no qual desenvolvi a minha tese de doutoramento, tais como a estatística de valores extremos e a modelação de dados de diagnóstico.

A trajetória da minha aventura durante esta fase decorreu mais ou menos do seguinte modo:

- **Banco de Portugal:** A época D.D. começou essencialmente no Banco de Portugal, onde me tinha candidatado a uma posição de investigador visitante por um período de 6 meses. Fazer investigação fora do âmbito académico foi uma experiência fantástica e mudou por completo o meu modo de abordar problemas. A investigação a ser desenvolvida deveria estar enquadrada nos interesses estratégicos do Banco de Portugal - o que excluía investigação puramente teórica - e durante a minha estadia tive a oportunidade de contactar com outros econométricos tais como o Paulo M. Rodrigues, António Rua, entre muitos outros. Entendi que numa instituição como o Banco de Portugal - onde é necessário trabalhar com técnicos com diversos backgrounds - a comunicação desempenha um papel fundamental. Parte da investigação que desenvolvi no Banco envolvia o que gosto de chamar de Econometria de Valores Extremos, que permite, por exemplo, avaliar como se movem conjuntamente várias economias durante períodos de crise. Que acontece, por exemplo, à economia Portuguesa quando a economia Alemã entra em recessão? Para mais detalhes ver de Carvalho e Rua (2013). Tudo isto se passou em 2009-2010, altura em que já se ouvia falar nos corredores do Banco sobre a bancarrota para que caminhava a nossa economia. Isto motivou a minha partida em meados de 2010; senti-me triste de partir de Portugal nesta altura, mas abriram-se também várias oportunidades com a minha partida.

- **Ecole Polytechnique Fédérale de Lausanne (EPFL):** Cheguei a Lausanne para visitar o Prof. Anthony Davison, em meados de 2010. Esta fase da minha carreira foi extremamente instrutiva, e a isso muito devo ao meu mentor Prof. Anthony Davison, que me acolheu de uma forma extraordinária. Inicialmente tínhamos combinado uma visita de dois meses, que acabou por se converter num pós-doc de cerca de dois anos e meio. Mudar de país envolveu vários desafios, mas foi uma experiência fantástica que abracei com grande entusiasmo desde o início. Tive também a possibilidade de orientar projetos de investigação de alunos de mestrado e de lecionar cursos do Mestrado em Matemática Aplicada e do Mestrado em Finanças. Lecionei os cursos de *Time Series* e *Quantitative Risk Management* e criei ainda um outro curso, *Applied Statistics*. Este último foi alvo de grande interesse por parte de vários alunos, que tinham muita curiosidade por modelação estatística com dados reais; uma versão deste curso continua ainda a ser ensinada na EPFL por alguns dos meus ex-colegas, o que me deixa bastante feliz! Entretanto a Vanda (de Carvalho) terminou o Doutoramento, casámos (!) e partimos para o outro lado do mundo onde encontrámos duas posições académicas na mesma Universidade. O Chile foi o país que nos acolheu!
- **Pontificia Universidad Católica de Chile:** Chegámos há cerca de um ano e tem sido uma experiência fantástica, convido-vos sinceramente a fazerem-nos uma visita! Quando chegámos tínhamos uma mala de viagem e 15 dias para tratar de papelada e encontrar um apartamento. Temos colegas excepcionais tais como o Guido del Pino e o Alejandro Jara que nos ajudaram bastante durante a fase de integração. Ganhei também um projeto de investigação da Fundação Chilena de cerca de 80.000€ (2012–2015) e tenho entretanto os meus primeiros alunos de doutoramento: sinto a responsabilidade de os educar cientificamente e eticamente, para que também eles possam vir a ter produção científica independente e a orientar outros alunos no futuro.

Miguel de Carvalho

Referências

- de Carvalho, M., Rua, A. (2013) "Extremal Dependence in International Output Growth: Tales from the Tails" *Oxford Bulletin of Economics and Statistics* (in press).
- de Haan, L. (1981) "Estimation of the Minimum of a Function using Order Statistics," *Journal of the American Statistical Association*, Vol. 76, pp. 467–469.
- Spall, J.C. (2003) *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. New York. Wiley.



Tese de Doutoramento: Modelação Estocástica com Misturas e Pseudo-Misturas

(Boletim SPE outono de 2009, p.97)

Miguel Felgueiras, mfelg@ipleiria.pt

*Escola Superior de Tecnologia e Gestão – Instituto Politécnico de Leiria
Centro de Estatística e Aplicações da Universidade de Lisboa*

Caros colegas e amigos.

Confesso que, contrariamente à maioria dos colegas, não foi o gosto pela investigação que me levou a ingressar no doutoramento. Foi, apenas, a noção de que este grau viria a ser imprescindível para o exercício da lecionação no ensino superior, como efetivamente veio a acontecer, que me levou a inscrever no doutoramento em Probabilidades e Estatística, na minha casa de sempre: a FCUL. Sem falsas modéstias, penso que as condições em que realizei o doutoramento foram mais “duras” que as da maioria. As dispensas de serviço durante três ou mais anos, de que a maior parte dos colegas anteriormente beneficiou, já não eram atribuídas, na instituição em que leciono, na altura em que ingressei no doutoramento. Beneficiei somente de dois semestres (e não consecutivos) de dispensa de serviço, sendo que nos restantes semestres em que realizei o doutoramento acumulei a investigação com a lecionação de 12 horas semanais na Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria (pois é, nos politécnicos é esta a nossa carga letiva habitual, para mais e não para menos). Saliento ainda que as funções que desempenhava nessa altura (equiparado a professor adjunto) implicavam a regência de várias unidades curriculares. Isto talvez faça alguma confusão aos colegas do ensino universitário, mas antes da revisão da carreira docente era possível aceder à categoria de professor adjunto (similar à de professor auxiliar) com o grau de mestre, o que foi o meu caso. Este foi realmente um período complicado da minha vida, especialmente na parte curricular do doutoramento (penso que fui dos primeiros a ter de a realizar). Nessa altura tinha um filho com 6 meses, e precisei de conciliar a paternidade com as aulas em Leiria e com as aulas em Lisboa. Talvez acreditem se vos disser que cheguei a adormecer, dobrado nas grades do berço, enquanto dava palmadinhas no rabo do meu filho para o tentar adormecer...

A escrita da tese, com a investigação a esta associada, foi uma experiência bastante enriquecedora. Comecei por trabalhar em tabelas de mortalidade e modelos geracionais, e acabei a trabalhar em modelação de dados com recurso a misturas de distribuições, num exemplo concreto de que a investigação não é um processo linear. Outra recordação que guardo desse tempo tem a ver com a evolução do trabalho: por vezes passava dias e dias a ler e a reler artigos, sem que, aparentemente, conseguisse alcançar qualquer progresso; outras vezes acordava a meio da noite e rabiscava algumas ideias que, às três ou quatro da manhã, me pareciam prometedoras (claro que de manhã grande parte dessas ideias eram abandonadas por motivos vários). Fiquei com a sensação de que, pelo menos no meu caso, é preferível às vezes mudar temporariamente de tarefa, para assentar ideias, a insistir continuamente num mesmo assunto.

Realizei as minhas provas de doutoramento em 01 de Abril de 2009 (e não é mentira!...). Após a obtenção do grau, a minha situação laboral manteve-se exatamente igual - as mesmas funções e o mesmo salário, como equiparado a professor adjunto, na Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria. Neste período de algum marasmo (e mais descanso...) aproveitei para organizar e preparar trabalhos com base nos principais temas focados na tese. Talvez o mais relevante, nesse ano, tenha sido a procura de independência científica. Procurei explorar algumas ideias (grande parte delas absolutamente inúteis) num processo de crescimento e independência instigado pelo meu orientador de doutoramento, o “eterno” Professor Doutor Dinis Pestana, que me incentivou a trabalhar

naquilo que eu pretendia, procurando apoiar-me sempre que eu precisava. Acho que é similar ao que se passa quando se sai de casa dos pais: queremos viver sozinhos mas recorremos a eles quando os problemas surgem...

A lei 7 de 2010, que estabeleceu um regime transitório para os docentes do ensino superior que já estavam no sistema, veio contribuir decisivamente para alterar a minha situação laboral. Não que os atuais contratos a tempo indeterminados sejam comparáveis aos antigos lugares de quadro, mas pelo menos dão-nos algumas garantias de manutenção do posto de trabalho que os contratos de equiparado não davam. Mas mais importante, a entrada na carreira permitiu-me aceder a órgãos de gestão e outras funções que me estavam (quase) vetadas como equiparado. Com alguma surpresa minha, fui convidado para fazer parte de uma lista de professores candidata ao Conselho Pedagógico da minha escola. Com ainda maior surpresa, acabei por ser eleito Presidente deste Conselho, funções que exerço desde abril de 2011 e que previsivelmente deixarei (num espírito de salutar alternância nos órgãos de gestão) deixar de exercer no final de 2014, quando concluir o segundo mandato. Também a minha investigação progrediu bastante entre 2010 e 2011, com mais apresentações e publicações, e com uma nova colaboração: a integração como membro do Centro de Investigação em Informática e Comunicações de Leiria (CIIC), que não tendo fundos tinha vários projetos interessantes em que era necessário, basicamente, tratar convenientemente a informação disponível (neste aspeto, permitam-me um aparte: os investigadores que não estejam associados a um centro de investigação forte, como o CEAUL de que faço parte, têm condições muito difíceis de trabalho). Finalmente, estreei-me na organização de eventos, fazendo parte da comissão organizadora do XIX congresso da SPE (Nazaré 2011). Talvez já não se recordem das conferências, mas espero que recordem com prazer o magnífico jantar realizado na quinta do Fidalgo, na Batalha.

Os anos de 2012 e 2013 foram, sem sombra de dúvida, os mais profícuos em termos de investigação. Reconheço que o gosto pela investigação foi começando a fazer parte de mim, e que após a “obrigação” inicial que foi obter o grau de Doutor, comecei a pensar com prazer em diversos problemas na área das probabilidades e estatística, e suas aplicações. Posteriormente, surgiu nesta fase uma colaboração natural que, nem sei bem porquê, ainda não tinha existido. Além de mim, Leiria tem ainda dois investigadores, igualmente delfins do professor Dinis Pestana, os meus colegas João Martins e Rui Santos. Em 2012, resolvemos formar o “cluster” de Leiria e, sem descurarmos os nossos trabalhos individuais e outras colaborações, começámos igualmente a realizar trabalhos conjuntos. Esta parceria compele-me, mais do que qualquer outra, a trabalhar afincadamente. Quem tem coragem de dizer ao colega da frente e ao colega do lado (o Rui trabalha à minha frente, e o João no gabinete ao lado) que não conseguiu terminar um artigo porque o filme que passava na televisão era muito mais interessante? Têm sido uns anos bastante produtivos! Por outro lado, não me lembro de ir ao cinema ver um filme que não seja para crianças e falado em português...

Talvez a diversidade de temas investigados tenha contribuído para desenvolver o meu gosto pela investigação. Sozinho, estudei aplicações de distribuições de caudas pesadas na economia e na libertação de energia nos grandes sismos, e alguns desenvolvimentos teóricos. Com os colegas da área da informática, membros do CIIC, abarquei um projeto que, através de um sensor tri-axial e um acelerómetro, procura detetar quedas em idosos. Com os colegas Rui Santos e João Martins, a atenção tem sido dada aos testes conjuntos e suas aplicações, principalmente em medicina. O que gostava de fazer no futuro? Essencialmente, começar a concentrar as minhas publicações em revistas de maior renome, procurando realizar “poucos mas bons” trabalhos. Não sei até que ponto este meu objetivo não é pouco tangível: por um lado as revistas de maior qualidade são muito seletas nas suas publicações, o que aumenta a taxa de recusa dos trabalhos, por vezes vários meses após a submissão (e isso é tão frustrante!...); por outro lado, os atuais regulamentos de avaliação de desempenho de docentes pesam as diversas tarefas de forma no mínimo discutível, com recurso a uma grelha de pontos que premeia a quantidade e a diversidade, mas não a qualidade.

Não posso ainda deixar de exprimir a minha preocupação com o futuro da investigação, mas também da lecionação no ensino superior, neste nosso Portugal. Os centros de investigação sofrem cortes atrás de cortes, e as novas regras de financiamento destes levam ainda a que muitos investigadores tenham de ser, e perdoem-me a expressão, erradicados dos centros por não cumprirem certas métricas. É óbvio que a qualidade tem de ser premiada, mas mais uma vez temos um sistema acéfalo que decide com base em quantidades. Por outro lado, a qualidade das aulas lecionadas tem muito pouco peso na avaliação de um docente. Esta deveria ser uma parte importante do nosso trabalho, mas considero que

é muitas vezes desvalorizada. Finalmente, qual é a motivação que um bom aluno tem para seguir a carreira académica? Após a licenciatura, as bolsas para doutoramento são escassas. E começar como assistente, com um salário razoável e perspectivas de futuro, já não é possível. A precarização, incerteza e aleatoriedade abundam.

Esperemos que o futuro nos traga melhores dias.

Saudações académicas!

Miguel Felgueiras



Tese de Doutoramento: Métodos Analíticos em Probabilidades e Métodos Probabilísticos em Análise: Fractalidade Associada aos Modelos Beta(p,q), Evolução de Populações e Dimensões de Hausdorff.

(Boletim SPE Outono de 2009, p.101)

Sandra Maria da Silva Figueiredo Aleixo, *sandra.aleixo@adm.isel.pt*

*Área Departamental de Matemática,
Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa*

Como o tempo voa!! Já passaram quase cinco anos após a apresentação da minha dissertação de doutoramento, orientada pelo Professor Doutor José Leonel Rocha, investigador na área de Sistemas Dinâmicos, e pelo Professor Doutor Dinis Pestana, investigador no ramo de Probabilidade e Estatística. O facto de ter sido orientada por professores destas duas áreas, fez com que a minha dissertação, e a pesquisa que tenho desenvolvido, resulte da interação entre tópicos dos referidos domínios de investigação. Ao longo desse período, foram estudados temas no âmbito de diversas matérias, nomeadamente: modelos de crescimento populacional (muitos deles associados a modelos probabilísticos), sincronização em sistemas caóticos associados a modelos de crescimento populacional, conjuntos de Cantor modificados (deterministas e aleatórios) e respetivas dimensões de Hausdorff. Neste artigo, são apresentados os tópicos predominantes da minha atividade de investigação, enquadrando os vários trabalhos publicados relativamente a cada tema.

Modelos de crescimento populacional

O estudo do comportamento de processos de crescimento através de modelos adequados é um tema de extrema relevância na sociedade e na comunidade científica atual. Entender os aspetos estruturais da dinâmica de populações de modo a prever a sua evolução futura é um instrumento de decisão importante no planeamento económico e demográfico, e essencial em muitos ramos da biologia tais como: a preservação do ambiente, a gestão da vida selvagem, as restrições nas pescas e na agricultura ou o controle de pestes e epidemias. Um dos aspetos importantes da dinâmica de populações prende-se com a modelação do crescimento populacional, de modo a avaliar o número de indivíduos num dado instante. A aplicação logística, também designada por modelo de Verhulst, tem sido usada com sucesso para modelar o crescimento de algumas espécies limitadas por recursos finitos. Esta aplicação incorpora nos seus parâmetros: a taxa de crescimento malthusiana e a retroação devida à limitação de recursos naturais. No entanto, para outras espécies o recurso a tais técnicas revela-se inadequado. O modelo de Verhulst é proporcional à densidade Beta (2,2). Outros modelos proporcionais a densidades Beta ($p,2$), com $p > 1$, são interessantes para modelar crescimentos populacionais relativos a populações com uma evolução à qual o modelo de Verhulst não se ajuste de forma adequada, podendo ser plausíveis quando o paradigma da reprodução da população é mais geral. Assim, na tese de doutoramento [1], foram investigados os comportamentos fractal e caótico destes modelos, cuja complexidade dinâmica depende do seu parâmetro de forma p mas também do parâmetro malthusiano. Este trabalho deu origem a uma publicação em 2009 [2].

O efeito de Allee é um dos principais fenómenos ecológicos, o qual foi originalmente descrito em 1931. A partir da sua investigação Allee concluiu que a agregação de indivíduos de uma dada espécie pode melhorar a sua taxa de sobrevivência. A visão clássica da dinâmica populacional afirma que, devido à competição por recursos, e para densidades elevadas, a população apresenta uma taxa de crescimento global reduzida, enquanto que para baixas densidades a população tem uma maior taxa de crescimento. No entanto, o conceito do efeito de Allee introduziu a ideia de que o inverso é verdadeiro quando a densidade populacional é baixa; geralmente os indivíduos de uma espécie necessitam do auxílio de outros indivíduos da mesma espécie para que possam persistir. Entre os comportamentos (mais citados na literatura) que contribuem para o efeito Allee incluem-se: limitação de encontrar companheiro para reprodução, defesa em colaboração com indivíduos da mesma espécie, cooperativismo na alimentação e alteração ambiental provocada por um conjunto de indivíduos de uma dada espécie a fim de melhorar o ambiente em benefício da sua espécie. O efeito de Allee pode ser classificado como forte ou fraco. O efeito de Allee forte introduz um limite populacional: a população precisa ultrapassar este limite para crescer; abaixo dele irá extinguir-se. Para o efeito de Allee fraco não há limiar de extinção. A população mostra um fator de crescimento pequeno para baixas densidades populacionais, o que pode dever-se a uma baixa eficiência na procura de alimentos ou parceiros.

Uma das críticas apontadas ao modelo logístico na modelação do crescimento populacional relaciona-se com o facto de não ter em conta o fenómeno de extinção, designado por efeito de Allee, como referem por exemplo: Wilson e Bossert (1971), Lewis e Kareiva (1993), Gruntfest et al (1997), Amarasekare (1998a, 1998b), Courchamp et al (1999a, 1999b, 2000), Keitt et al (2001), e Brassil (2001), referenciados em [3]. Estes autores apresentaram vários modelos logísticos corrigidos de modo a considerarem o efeito de Allee, os quais foram sumarizados por Boukal e Berec (2002), citado em [3]. Ainda na dissertação de doutoramento, o efeito de Allee foi analisado nos modelos proporcionais a densidades Beta ($p,2$), tendo-se verificado que, para $p > 2$, eles exibem uma dinâmica populacional onde o efeito de Allee surge naturalmente; no entanto, no caso de $1 < p \leq 2$, estes modelos não incluem este efeito. Para o invocar, em [3] foram apresentados alguns modelos alternativos, para os quais se estudou o comportamento dinâmico. Num outro trabalho, em 2011, foram formalizados resultados importantes sobre a caracterização dinâmica no espaço de parâmetros, através do estudo da bifurcação com recurso à dinâmica simbólica, sobre os modelos corrigidos com efeito de Allee [4].

Em 2012, em [5] e [6] generalizam-se os modelos apresentados anteriormente. Em [5] foram apresentados e estudados modelos de crescimento populacional generalizados com efeito Allee, ajustados a partir de densidades Beta ($p,2$). Nestes modelos, a utilização de um parâmetro positivo conduziu à generalização apresentada, a qual produziu alguns modelos mais flexíveis com taxas de extinção variáveis. Foi incorporado um limite de Allee de modo que os modelos estudados contemplassem efeito de Allee forte. Neste trabalho “efeito de Allee” tem o significado estrito de que as taxas de crescimento *per capita* são superiores à taxa de crescimento da população no momento inicial. As definições de efeitos Allee forte e fraco são utilizadas no sentido de Clark 1990, Schreiber 2003, Wang et al 2011, referenciados em [5]. Em [6] foi apresentada uma abordagem dinâmica para estudar o comportamento de modelos de crescimento populacional generalizados das densidades Beta ($p,2$) com efeito de Allee forte. Foi feita a análise dinâmica das respetivas aplicações unimodais utilizando técnicas de dinâmica simbólica. A complexidade dos sistemas dinâmicos discretos correspondentes foi medida em termos de entropia topológica. Foram obtidos regimes dinâmicos populacionais distintos quando as taxas de crescimento intrínsecas sofreram alterações: extinção, bi-estabilidade, semi-estabilidade, caótico e extinção essencial, sendo pois o espaço paramétrico dividido em diferentes regiões.

Já em 2013 foram desenvolvidos e investigados novos modelos generalizados de von Bertalanffy corrigidos com efeitos Allee [7]. O modelo de von Bertalanffy é uma das equações diferenciais mais

usadas para estudar o aumento do comprimento ou peso médio de espécies de peixes. No entanto, este modelo não inclui efeito Allee demográfico. Este fenómeno surge quando as populações diminuem rapidamente a baixas densidades populacionais. A generalização proposta neste trabalho resulta de considerar fatores de correção, um de tipo racional e os outros de tipo polinomial, onde são considerados dois parâmetros. Foi analisada e demonstrada a transição do efeito de Allee forte à inexistência desse efeito, através de um “enfraquecimento” do efeito de Allee, dependendo dos parâmetros introduzidos. Finalmente, foi discutida a flexibilidade de curvas de crescimento sigmóides de von Bertalanffy corrigidas, verificando-se que os correspondentes pontos de inflexão são variáveis. Este facto levou a concluir que a quantidade existente de peixes, quando a taxa de crescimento é máxima, pode ser variável.

Ainda em 2013, na continuação do tema de investigação relativo a modelos de crescimento populacional associados à distribuição Beta (p,q) , foi implementada uma generalização dos trabalhos anteriores [8]. Neste artigo foi apresentada uma nova abordagem dinâmica para a equação de Blumberg, considerando uma família de aplicações unimodais. Estas aplicações são proporcionais a funções de densidade de probabilidade Beta (p,q) . Usando a simetria da distribuição Beta (p,q) e técnicas de dinâmica simbólica, foi definido um novo conceito de simetria espelho para esta família de aplicações. A teoria do *kneading* foi utilizada para analisar e demonstrar o efeito de tal simetria nos modelos apresentados. O principal resultado prova que duas aplicações unimodais simétricas espelho têm a mesma entropia topológica. Ao alterar os valores da taxa de crescimento intrínseca foram identificados diferentes regimes de dinâmica: extinções, estabilidades, duplicação de período, caos e efeito Allee. Para ilustrar os resultados, foi apresentada uma análise numérica, onde foram demonstradas: monotonicidade da entropia topológica com a variação da taxa de crescimento intrínseca e a existência de conjuntos isentrópicos no espaço de parâmetros e simetria de espelho.

Por outro lado, em 2011 num outro trabalho foi introduzida uma nova família de funções de densidade de probabilidade ligada à família clássica Beta (p,q) , a qual foi designada por Beta* (p,q) [9]. A investigação foi alargada para outros modelos de extremos estáveis, nomeadamente do tipos Fréchet e Weibull. Foi apresentado um esboço do comportamento dinâmico de alguns modelos baseados em modelos Beta* $(p,2)$, evidenciando o caso $p = q = 2$. Foi ainda demonstrado que estes modelos Beta* (p,q) estão ligados a misturas generalizadas de funções potência.

Num outro trabalho, em 2013 foram considerados modelos populacionais proporcionais a densidades Beta* (p,q) [10]. Foi proposta uma nova abordagem probabilística e dinâmica para uma extensão da lei de Gompertz, a qual é usualmente aplicada para modelar crescimentos de tumores malignos. Foi estudada a família generalizada de funções densidade de probabilidade, já designada por Beta* (p,q) , a qual é proporcional ao lado direito da equação diferencial ordinária que representa o modelo Tsoularis-Wallace-Schaefer. Em particular, para $p = 2$, a investigação estendeu-se aos modelos de valores extremos de tipo Weibull e Fréchet. Estes modelos, descritos por equações diferenciais, são proporcionais ao modelo de crescimento híper-Gompertz. Foi demonstrado que as densidades Beta* $(2,q)$ são uma potência de misturas de betas, e que as suas dinâmicas são determinadas por um acoplamento não-linear de probabilidades. Foi realizada uma análise dinâmica utilizando técnicas de dinâmica simbólica e a complexidade do sistema foi medida usando a entropia topológica. Utilizando diagramas de bifurcação, refletiu-se a história natural de um tumor maligno, identificando as regiões de regressão, estabilidade, duplicação de período, caos e *terminus*.

Sincronização em sistemas caóticos associados a modelos de crescimento populacional

A aplicação logística, o modelo de Gompertz e os modelos de crescimento de von Bertalanffy têm sido usados para descrever o crescimento de mamíferos marinhos, aves marinhas, tartarugas marinhas, peixes, invertebrados e répteis. A sincronização é um fenômeno de grande interesse no mundo real. O comportamento dos bandos de pássaros, dos cardumes de peixes e outros grupos de animais, é um fenômeno importante caracterizado pelo movimento sincronizado dos indivíduos. Recentemente, em [11] foi estudada a família generalizada de aplicações de von Bertalanffy, a qual é proporcional ao lado direito da equação de crescimento de von Bertalanffy. Foi proposta uma abordagem dinâmica destes modelos, onde a complexidade do sistema foi medido usando expoentes Lyapunov, que dependem de dois parâmetros biológicos: taxa de crescimento de von Bertalanffy e o peso assintótico. Neste trabalho, foram consideradas redes, nas quais se considera que a dinâmica em cada nó é governada por um modelo de von Bertalanffy. Foi estudado o intervalo de sincronização dessas redes em função dos dois parâmetros biológicos, tendo sido apresentados resultados de simulação numérica. Em [12] foram estabelecidos novos sistemas dinâmicos discretos unidimensionais: uma família de aplicações unimodais que é proporcional ao lado direito da equação de crescimento de Richards. Foi investigada em detalhe a estrutura de bifurcação das funções de Richards, no espaço de parâmetros bidimensional (β, r) , onde β é o parâmetro de forma, relacionado com o fenômeno de crescimento-retardação, e r representa a taxa de crescimento intrínseca. Foram estabelecidas condições suficientes para a ocorrência de extinção, estabilidade, duplicação de período, caos e não admissibilidade da dinâmica de Richards. Foram consideradas redes tendo em cada nó uma função de Richards para descrever a dinâmica local. Foram provados alguns resultados sobre o nível de sincronização quando se fixa a topologia da rede e se alteram as dinâmicas locais expressas pelos expoentes de Lyapunov (o que depende dos parâmetros (β, r)). Além disso, foram fixadas as dinâmicas locais e provados resultados sobre a sincronização quando se altera a topologia de rede (para algum tipo de redes), ou seja quando se altera a dinâmica global. Finalmente, usando simulações numéricas, foram calculados os expoentes de Lyapunov para medir a complexidade do sistema, obtiveram-se intervalos de sincronização dessas redes, e foi discutida a evolução do nível de sincronização em termos dos parâmetros da função de Richards.

Conjuntos de Cantor modificados (deterministas e aleatórios) e respectivas dimensões de Hausdorff

Durante anos os matemáticos preocuparam-se fundamentalmente com conjuntos e funções às quais podem ser aplicados os métodos de cálculo clássicos. Os conjuntos e funções que não são suficientemente regulares tendiam a ser ignorados, sendo considerados “patológicos” e sem interesse de estudo; eram vistos como curiosidades pontuais e só raramente se pensava neles como uma classe à qual se poderia aplicar uma teoria geral. Mais recentemente esta atitude alterou-se. Alguns matemáticos perceberam que muito pode e merece ser dito acerca da matemática de “conjuntos irregulares”. Muitos destes conjuntos fornecem uma representação muito melhor de muitos fenômenos naturais do que as figuras da geometria clássica. A geometria fractal fornece uma estrutura geral para o estudo dos referidos conjuntos irregulares, entre os quais se encontram os conjuntos de Cantor.

Uma das questões relevantes na investigação relativa aos modelos de crescimento populacional, estudados como aplicações unimodais, é a caracterização do seu comportamento fractal e caótico em função dos respetivos parâmetros. Na região do espaço de parâmetros em que o gráfico de qualquer aplicação unimodal, com parâmetros pertencentes a essa região, deixa de estar totalmente contido no conjunto invariante $[0,1] \times [0,1]$ surgem os chamados conjuntos de Cantor. O estudo de conjuntos deste tipo, bem como a determinação da sua dimensão de Hausdorff, foi tema da última parte da minha dissertação de doutoramento e deu origem à elaboração de três trabalhos [13], [14] e [15], desenvolvidos entre 2009 e 2011. Em [13] foi definido o conceito de conjunto de Cantor aleatório, um

fractal que é construído pela eliminação recursiva do espaçamento central gerado pelo mínimo e máximo de duas observações de uma variável aleatória com distribuição uniforme no intervalo resultante na iteração anterior. Foi calculada a respetiva dimensão de Hausdorff (em termos intuitivos, um parâmetro que avalia até que ponto um conjunto é denso), verificando-se que: apesar de o conjunto de Cantor ser o “fractal médio” do conjunto de Cantor aleatório, é mais denso, pois, a dimensão de Hausdorff do conjunto de Cantor é superior à dimensão de Hausdorff do conjunto de Cantor aleatório. Este facto pode ser explicado pela estrutura de dependência inerente às estatísticas ordinais. Em [14] as ideias do artigo anterior foram formalizadas de forma cuidada. A partir do extensão aleatória do conjunto de Cantor definido em $[0,1]$, removendo iterativamente o espaçamento uniforme central a partir dos intervalos remanescentes na etapa anterior, foram definidos conjuntos Beta (p, I) -Cantor aleatórios, e foi calculada a sua dimensão de Hausdorff. Em seguida, foi definida uma contrapartida determinista, removendo iterativamente o valor esperado do espaçamento definido pelas estatísticas ordinais apropriadas da distribuição Beta (p, I) . Foram investigadas as razões pelas quais a dimensão Hausdorff deste fractal determinista é maior do que a dimensão Hausdorff dos fractais aleatórios correspondentes. Neste estudo, a pesquisa teve como objetivo definir e caracterizar uma nova estrutura de conjuntos de Cantor aleatórios modificados, quando os conjuntos centrais removidos em cada passo seguem uma distribuição Beta (p, I) . Estas ideias foram generalizadas para todos os outros valores do parâmetro $q > 0$ em [15], tendo sido formalmente definidos conjuntos Beta (p, q) -Cantor aleatórios e deterministas, e tendo sido desenvolvidos métodos para determinar a dimensão de Hausdorff destes novos conjuntos. Foram ainda apresentados alguns resultados numéricos, os quais mostraram que a dimensão de Hausdorff de um determinado conjunto Beta (p, q) -Cantor aleatório é menor do que a dimensão de Hausdorff do conjunto determinista correspondente: o conjunto Beta (p, q) -Cantor, facto que é explicado pela desigualdade de Jensen.

Ainda em 2011, foram desenvolvidos outros dois trabalhos com conjuntos de Cantor modificados deduzidos por um processo de “gaguejamento” [16] e [17]. Em [16] obtiveram-se conjuntos de Cantor aleatórios de “gaguejamento”, utilizando segmentos aleatórios uniformes tanto na fase de dano (eliminação) como na fase de reparação, em cada passo do processo iterativo. Consideraram-se duas operações complementares em cada etapa do processo de construção do conjunto de Cantor: na primeira fase (dano) o segmento do meio de cada segmento restante da etapa anterior foi suprimido, e na segunda fase (reparação aleatória) um segmento aleatório uniforme foi unido ao que permanecia após a eliminação. Foi também calculada a dimensão de Hausdorff do fractal limite obtido pela interseção dos conjuntos obtidos na repetição *ad infinitum* deste procedimento iterativo de “gaguejamento” que, como esperado, é maior do que a dimensão de Hausdorff do clássico conjunto de Cantor definido sem qualquer reparação. Foi também discutida a utilização de segmentos aleatórios Beta (p, q) , com $p, q > 0$, na construção de conjuntos de Cantor aleatórios de “gaguejamento”. Este processo foi testado e analisado num outro trabalho [17]. Partindo do intervalo $[0,1]$, cada passo do processo iterativo da construção do conjunto de Cantor aleatório modificado era composto por dois processos: dano aleatório – exclusão de $(X_{1:2}, X_{2:2})$ do segmento restante da etapa anterior, seguido de reparação aleatória, ou seja, a união de $(Y_{1:2}, Y_{2:2})$ ao conjunto danificado, em que X e Y são variáveis aleatórias independentes, e $X_{i:j}$ representa a i -ésima estatística ordinal de uma amostra de dimensão j associada à variável aleatória X . A dimensão de Hausdorff foi calculada sob vários padrões de aleatoriedade que representam vários graus de redundância, nomeadamente usando variáveis aleatórias com distribuição Beta (p, q) e variáveis aleatórias com distribuição Beta $^*(p, q)$.

Graças à investigação científica que tenho vindo a desenvolver e à obtenção do grau de doutor na especialidade de Probabilidade e Estatística, sou atualmente professora coordenadora do Instituto Superior de Engenharia de Lisboa, na Área Departamental de Matemática, onde exerço funções docentes desde Setembro de 1998.

Sandra Aleixo

- [1] “Métodos Analíticos em Probabilidades e Métodos Probabilísticos em Análise: Fractalidade Associada aos Modelos Beta (p,q), Evolução de Populações e Dimensões de Hausdorff”, Sandra M. Aleixo, Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa, 2008.
- [2] “Dynamical Behaviour on the Parameter Space: New Populational Growth Models Proportional to Beta Densities”, Sandra M. Aleixo, J. Leonel Rocha and Dinis D. Pestana, In Proceedings of the ITI 2009, 31th International Conference on Information Technology Interfaces, 2009, 213-218. (DOI:10.1109/ITI.2009.5196082)
- [3] “Populational Growth Models Proportional to Beta Densities with Allee Effect”, Sandra M. Aleixo and J. Leonel Rocha and Dinis D. Pestana, In Mathematical Models in Engineering, Biology and Medicine, A. Cabada, E. Liz, J. Nieto. (Eds): American Inst. of Physics, 2009, vol. 1124, 3-12. (DOI: 10.1063/1.3142952)
- [4] “Probabilistic Methods in Dynamical Analysis: Populations Growths Associated to Models Beta (p,q) with Allee Effect”, Sandra M. Aleixo, J. Leonel Rocha and Dinis D. Pestana, In Dynamics, Games and Science II, M. M. Peixoto, A. A. Pinto, D. A. J. Rand (Eds): Springer Proceedings in Mathematics 2, Springer-Verlag Berlin Heidelberg, 2011, Vol. 2, Chap. 5, 79-95. (DOI: 10.1007/978-3-642-14788-3-5)
- [5] “Modeling Allee Effect from Beta ($p,2$) Densities”, J. Leonel Rocha and Sandra M. Aleixo, In Proceedings of the ITI 2012, 34th International Conference on Information Technology Interfaces, Eds. V. Luzar-Stiffler, I. Jarec, Z. Bekic, 2012, 461-466. (DOI: 10.2498/ITI.2012.0420)
- [6] “Generalized Models from Beta ($p,2$) Densities with Strong Allee Effect: Dynamical Approach”, Sandra M. Aleixo and J. Leonel Rocha, Journal of Computing and Information Technology, 2012, Vol. 20, No. 3, 201-207. (DOI: 10.2498/cit.1002098)
- [7] “Von Bertalanffy' Growth Dynamics with Strong Allee Effect”, J. Leonel Rocha and Sandra M. Aleixo, *Discussiones Mathematicae, Probability and Statistics*, 2013, 32, p. 35-45. (DOI: 10.7151/dmps.1142)
- [8] “Dynamical Analysis in Growth Models: Blumberg's Equation”, J. Leonel Rocha and Sandra M. Aleixo, *Discrete & Continuous Dynamical Systems - Series B*, 2013, Vol. 18, No. 3, 783-795. (DOI: 10.3934/dcdsb.2013.18.783)
- [9] “Regular Variation, Paretian Distributions, and the Interplay of Light and Heavy Tails in the Fractality of Asymptotic Models”, Dinis D. Pestana, Sandra M. Aleixo and J. Leonel Rocha, In *Chaos Theory: Modeling, Simulation and Applications*, Skiadas, C. H., Dimotikalis, Y. and Skiadas, C. (Eds), World Scientific Publishing Co, 2011, 309-316. (DOI: 10.1142/9789814350341-036)
- [10] “An Extension of Gompertzian Growth Dynamics: Weibull and Fréchet Models”, J. Leonel Rocha and Sandra M. Aleixo, *Mathematical Biosciences and Engineering*, 2013, Vol. 10, No. 2, 379-398. (DOI 10.3934/mbe.2013.10.379)
- [11] “Synchronization in von Bertalanffy's models”, J. Leonel Rocha, Sandra M. Aleixo and Acilina Caneco, *CMSIM - Chaotic Modeling and Simulation Journal*, a aparecer.
- [12] “Synchronization in Richards' chaotic systems”, J. Leonel Rocha, Sandra M. Aleixo and Acilina Caneco, *Journal of Applied Nonlinear Dynamics*, a aparecer.
- [13] “Hausdorff Dimension of the Random Middle Third Cantor Set”, Dinis D. Pestana, Sandra M. Aleixo and J. Leonel Rocha, In Proceedings of the ITI 2009, 31th International Conference on Information Technology Interfaces, 2009, 279-284. (DOI: 10.1109/ITI.2009.5196094)
- [14] “The Beta ($p,1$) Extensions of the Random (Uniform) Cantor Sets”, Dinis D. Pestana, Sandra M. Aleixo and J. Leonel Rocha, *Discussiones Mathematicae, Probability and Statistics*, 2009, 29, 199-221. (DOI: 10.7151/dmps.1115)
- [15] “Beta (p,q)-Cantor Sets: Determinism and Randomness”, J. Leonel Rocha, Sandra M. Aleixo and Dinis D. Pestana, In *Chaos Theory: Modeling, Simulation and Applications*, C. H. Skiadas, Y. Dimotikalis, C. Skiadas (Eds), World Scientific Publishing Co, 2011, 333-340. (DOI: 10.1142/9789814350341-039)
- [16] “Stuttering Cantor-Like Random Sets”, Dinis D. Pestana and Sandra M. Aleixo, In Proceedings of the ITI 2011, 33th International Conference on Information Technology Interfaces, Eds. V. Luzar-Stiffler, I.Jarec, Z.Bekic, 2011, 29-34. (IEEE Catalog Number CFP11498-PRT)
- [17] “General stuttering Beta (p,q) Cantor-like random sets”, Sandra M. Aleixo, Dinis D. Pestana and F. Brilhante, *Electronic Publication of the Proceedings of the 58th Session of the International Conference of the International Statistical Institute, ISI 2011, CPS008: 1-16.*

• Artigos Científicos Publicados

- Ferreira, L.N.; Ferreira, P.L.; Pereira, L.N.; e Rowen, D. (2012). Reducing the Floor Effect in the SF-6D: A Feasibility Study. *Applied Research in Quality of Life*, 7:2, 193-208.
- Filipe, P.A., Braumann, C.A. e Roquete, C.J. (2012). Multiphasic individual growth models in random environments. *Methodology and Computing in Applied Probability*, 14 (1): 49-56.
- Leonel Rocha, J. and Sandra M. Aleixo (2013). Dynamical Analysis in Growth Models: Blumberg's Equation, *Discrete & Continuous Dynamical Systems - Series B*, Vol. 18, No. 3., p. 783-795.
- Leonel Rocha, J. and Sandra M. Aleixo (2013). An Extension of Gompertzian Growth Dynamics: Weibull and Fréchet Models, *Mathematical Biosciences and Engineering*, Vol. 10, No. 2, p. 379-398.
- Leonel Rocha, J. and Sandra M. Aleixo (2013). Von Bertalanffy' Growth Dynamics with Strong Allee Effect, *Discussiones Mathematicae, Probability and Statistics*, 32, p. 35-45.
- Moreira, E. E., João T. Mexia, Christoph E. Minder (2013). F tests with random sample size. *Theory and applications. Statistics & Probability Letters*, 83 (6), 1520–1526.
- Pereira, L.N. e Coelho, P.S. (2013). An applied comparison of area-level linear mixed models in small area estimation. *Communications in Statistics – Simulation and Computation*, 42:3, 671-685.
- Pereira, L.N. e Coelho, P.S. (2012). Small area estimation using a spatio-temporal linear mixed model. *REVSTAT – Statistical Journal*, 10:3, 285-308.

• Teses de Mestrado

Título: *Contributo das Técnicas de Análise Fatorial para o Estudo do Programa “Ocupação Científica de Jovens nas Férias”*

Autora: Domingas Portela, domingas.portela@gmail.com

Orientadores: Maria do Rosário Ramos e Carlos Catalão Alves

Título: *Modelos Estatísticos para a Previsão de Inatividade de Pré-Pagos*

Autora: Marta Gonçalves Cruces Simão Portugal, marta.goncalves.portugal@gmail.com

Orientadoras: Marília Antunes e Paula Figueiredo Mestre

Título: *Maternal breastfeeding in Portugal: factors associated with the duration and premature cessation*

Autora: Sílvia Cristina Martins Mendonça, silvia6m@gmail.com

Orientadora: Marília Antunes

• Capítulos de Livros

- Larguinho, M., Dias, J.C. and Braumann, C.A. (2013). Absolute Diffusion Process: Sensitivity Measures. Em *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications*, da Silva, J.L.; Caeiro, F.; Natário, I.; Braumann, C.A.; Esquível, M.L.; Mexia, J. (Eds.), Springer, ISBN 978-3-642-34903-4, p. 249-257.
- Carlos, C., Braumann, C.A e Filipe, P.A. (2013). Models of individual growth in a random environment: study and application of first passage times. Em *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications*, da Silva, J.L.; Caeiro, F.; Natário, I.; Braumann, C.A.; Esquível, M.L.; Mexia, J. (Eds.), Springer, ISBN 978-3-642-34903-4, p. 103-111.
- Filipe, P.A.; Braumann, C.A.; Brites, N.M.; Roquete, C.J. (2013). Prediction for individual growth in a random environment. Em *Recent Developments in Modeling and Applications in Statistics*. Oliveira, P.E.; Graça Temido, M.; Henriques, C.; Vichi, M. (Eds.), Springer, ISBN 978-3-642-32419-2, p. 193-201.
- Larguinho, M.; Dias, J.C.; Braumann, C.A. (2013). A note on (dis)investment options and perpetuities under CIR interest rates. Em *Recent Developments in Modeling and Applications in Statistics*. Oliveira, P.E.; Graça Temido, M.; Henriques, C.; Vichi, M. (Eds.), Springer, ISBN 978-3-642-32419-2, p. 203-211.
- Larguinho, M.; Dias, J. C.; Braumann, C. A. (2012). Bond Options, Sensitivity Measures, and Sinking-Fund Bonds under the CIR Framework. Em "*7th Finance Conference of the Portuguese Finance Network*" (electronic proceedings). ISBN: 978-972-789-362-1.
- Braumann, C. A.; Carlos; C. (2013). Allee effects in randomly varying environments. Em *Proceedings of the 2013 International Conference on Computational and Mathematical Methods in Science and Engineering*. Hamilton, I. P.; Vigo-Aguiar, J. (Eds.), ISBN 978-84-616-2723-3, p. 304-307.

• Livros

Título: *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications*
Editores: da Silva, J.L.; Caeiro, F.; Natário, I.; Braumann, C.A.; Esquível, M.L.; Mexia, J.
Ano: 2013. Springer Series: Studies in Theoretical and Applied Statistics, Subseries: Selected Papers of the Statistical Societies. ISBN 978-3-642-34903-4.

• Tese de Doutoramento

Título: Testes Conjuntos - Extensões da Teoria de Dorfman

Autor: Ricardo Sousa, *ricardosasousa@sapo.pt*

Orientadores: Rui Santos e Dinis Pestana

Na minha tese são apresentados resultados inovadores na área de análises combinadas e sua aplicação em saúde pública, nomeadamente em análises clínicas, bem como em amostragem para aceitação no controlo de qualidade.

O recurso a análises conjuntas ao sangue, proposto por Dorfman durante a segunda grande guerra, permitiu uma gestão mais eficiente de recursos na deteção dos soldados infetados com sífilis, e tornou-se um paradigma, que se pode tornar mais realista e aplicável se considerarmos que os testes de diagnósticos podem estar sujeitos a erros de classificação.

Assim, começamos por estender os conceitos de sensibilidade e especificidade para a realização de testes conjuntos, tendo em consideração o problema da diluição e conseqüente rarefação da substância em análise.

Numa segunda fase analisamos, via simulação, as performances de alguns estimadores para a taxa de prevalência baseados em testes conjuntos, quer na ausência quer na presença de erros de classificação, comparando-as com a dos principais estimadores baseados em testes individuais.

No que toca a testes quantitativos discretos, estendemos os cálculos da sensibilidade e da especificidade para o modelo de Poisson a populações mais dispersas, nomeadamente binomiais negativas, dando especial relevo ao caso mais tratável de população geométrica.

Por fim, relativamente a testes quantitativos contínuos, investigamos a informação da média sobre o máximo (ou sobre o mínimo) de cada grupo, de forma a, com base num resultado conjunto, ser possível identificar os grupos infetados (nos quais há pelo menos um indivíduo infetado), sem correremos o risco de termos uma probabilidade elevada de ocorrência de má classificação.

Ricardo Sousa



PRÉMIO ESTATÍSTICO JÚNIOR 2013



CONTACTOS

Sociedade Portuguesa de Estatística
Bloco C6, Piso 4 – Campo Grande
1749-016 Lisboa
Telef./Fax 21 750 01 20

www.spestatistica.pt
spe@fc.ul.pt

Com o apoio:



Prémios Estatístico Júnior 2013

Trabalho classificado em 1º lugar (3º ciclo do Ensino Básico)

Título: “*Os Jovens e a sua Imagem*”

Autores: Andreia Filipa Lopes Sousa, Carolina Cabeleira Felgueiras e Inês da Costa Delgado

Professora orientadora: Maria Alice da Silva Martins.

Estabelecimento de Ensino: Agrupamento de Escolas Artur Gonçalves, Torres Novas

Trabalho classificado em 2º lugar (3º ciclo do Ensino Básico)

Título: “*A nossa turma - 8º C*”

Autores: Camila Sofia Fragoso Marcos e Daniela Marques Silva

Professor orientador: Carlos André Pimentel Lameirinhas

Estabelecimento de Ensino: Colégio Senhor dos Milagres, Leiria.

Trabalho classificado em 3º lugar (3º ciclo do Ensino Básico)

Título: “*3º ciclo: Quem somos?*”

Autores: Joana Alves Domingos, Carolina dos Santos Elias

Professor orientador: Susana Patrícia Varela Alves.

Estabelecimento de Ensino: Externato de Penafirme, A-dos-Cunhados

Trabalho classificado em 2º lugar (Ensino Secundário)

Título: “*O Estado da Nação visto pelos Jovens*”

Autores: Carlos Moura P. Lucas Teixeira

Professora orientadora: José Alberto Mesquita da Silva

Estabelecimento de Ensino: Escola Secundária 3EB Dr. Jorge Correia, Tavira

Trabalho classificado em 3º lugar (Ensino Secundário)

Título: “*Ensino Profissional e o futuro*”

Autor: Sara Raquel C. Pereira, Cláudia Batista C. Monteiro e Susana Isabel A. D. Silva

Professor orientador: Rui Pedro Ferreira Januário

Estabelecimento de Ensino: Escola Profissional Gustave Eiffel

Nota: Não foi atribuído o 1º lugar (Ensino Secundário)

Edições SPE - Minicursos

Título: Modelos com Equações Estruturais

Autora: Maria de Fátima Salgueiro

Ano: 2012.

Título: Análise de Dados Longitudinais

Autoras: Maria Salomé Cabral e Maria Helena Gonçalves

Ano: 2011

Título: Uma Introdução à Estimação Não-Paramétrica da Densidade

Autor: Carlos Tenreiro

Ano: 2010

Título: Análise de Sobrevivência

Autoras: Cristina Rocha e Ana Luísa Papoila

Ano: 2009

Título: Análise de Dados Espaciais

Autoras: M. Lucília de Carvalho e Isabel C. Natário

Ano: 2008

Título: Introdução aos Métodos Estatísticos Robustos

Autores: Ana M. Pires e João A. Branco

Ano: 2007

Título: Outliers em Dados Estatísticos

Autor: Fernando Rosado

Ano: 2006

Título: Introdução às Equações Diferenciais Estocásticas e Aplicações

Autor: Carlos Braumann

Ano: 2005

Título: Uma Introdução à Análise de Clusters

Autor: João A. Branco

Ano: 2004

Título: Séries Temporais – Modelações lineares e não lineares

Autoras: Esmeralda Gonçalves e Nazaré Mendes Lopes

Ano: 2003 (2ª Edição em 2008)

Título: Modelos Heterocedásticos. Aplicações com o software Eviews

Autor: Daniel Muller

Ano: 2002

Título: Inferência sobre Localização e Escala

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

Ano: 2001

Título: Modelos Lineares Generalizados – da teoria à prática

Autores: M. Antónia Amaral Turkman e Giovani Silva

Ano: 2000

Título: Controlo Estatístico de Qualidade

Autoras: M. Ivette Gomes e M. Isabel Barão

Ano: 1999

Título: Tópicos de Sondagens

Autor: Paulo Gomes

Ano: 1998



STATISTICS2013.ORG

INTERNATIONAL YEAR OF STATISTICS

PARTICIPATING ORGANIZATION



Índice

Editorial	1
Mensagem do Presidente	2
Notícias	4
<i>Enigmística</i>	9

A “Escola Bayesiana” em Portugal

Fundamentos da Inferência Bayesiana	<i>Bento Murteira</i>	10
Representação da Informação <i>a priori</i>	<i>C. D. Paulino e M. A. Amaral Turkman</i>	19
Metodologia Bayesiana em Aplicações Básicas	<i>Carlos Daniel Paulino</i>	27
Estatística Bayesiana: Aplicações em Análise de Sobrevivência e Estatística Espacial	<i>Giovani Loiola da Silva</i>	38
Método de Monte Carlo via Cadeias de Markov com saltos reversíveis	<i>Isabel Pereira</i>	45
Métodos Computacionais: INLA, Integrated Nested Laplace Approximation	<i>Isabel Natário</i>	52
Métodos de Monte Carlo Sequenciais e Métodos Computacionais Bayesianos Aproximados	<i>P. de Zea Bermudez</i>	57
Um breve passeio pelo campo computacional bayesiano	<i>Paulo Soares e Valeska Andreozzi</i>	63
Uma pequena conversa sobre Estatística Bayesiana em Genética e Genómica	<i>Nuno Sepúlveda</i>	69

Pós - Doc

<i>Fernando Sequeira</i>	75
<i>Marta Ferreira</i>	77
<i>Miguel de Carvalho</i>	79
<i>Miguel Felgueiras</i>	82
<i>Sandra Aleixo</i>	85

Ciência Estatística

<i>Artigos Científicos Publicados</i>	91
<i>Teses de Mestrado</i>	91
<i>Capítulos de Livros</i>	92
<i>Livros</i>	92
<i>Tese de Doutoramento</i>	93

Prémios “Estatístico Júnior 2013”	94
---	----