Análise de Valores Extremos: Uma Introdução

M. Ivette Gomes
C.E.A.U.L. e D.E.I.O., F.C.U.L., Universidade de Lisboa,
Instituto de Investigação Científica Bento da Rocha Cabral

M. Isabel Fraga Alves
D.E.I.O., F.C.U.L. e C.E.A.U.L., Universidade de Lisboa

Cláudia Neves
C.E.A.U.L., Universidade de Lisboa,
DMat, Universidade de Aveiro

Edições SPE

Ficha Técnica:

Análise de Valores Extremos: Uma Introdução[1]

M. Ivette Gomes
C.E.A.U.L. e D.E.I.O., F.C.U.L., Universidade de Lisboa,
Instituto de Investigação Científica Bento da Rocha Cabral

M. Isabel Fraga Alves
D.E.I.O., F.C.U.L. e C.E.A.U.L., Universidade de Lisboa

Cláudia Neves
C.E.A.U.L., Universidade de Lisboa,
DMat, Universidade de Aveiro

Editora: Sociedade Portuguesa de Estatística

Capa: Carina Sousa

Impressão: Instituto Nacional de Estatística

Tiragem: 200 exemplares

ISBN: 978-972-8890-30-8

Depósito Legal: 366446/13

[1]Investigação parcialmente financiada pelos fundos nacionais da **FCT**—Fundação para a Ciência e a Tecnologia, projecto PEst-OE/MAT/UI0006/2011, EXTREMA, PTDC/MAT/101736/2008 e PTDC/MAT/112770/2009: EXTREMES IN SPACE.

Conteúdo

Prefácio

Neste texto procedemos em grande parte a uma compilação do material leccionado em cadeiras das áreas de *Estatísticas Ordinais*, de *Teoria de Valores Extremos*, de *Estatística de Extremos* e de *Modelação de Acontecimentos Raros*, ampliado com alguns desenvolvimentos recentes.

Trata-se de um manual de trabalho, ainda em fase embrionária, em que se procurou encontrar um compromisso entre o rigor teórico e uma abordagem intuitiva às áreas em estudo, disseminando técnicas simples, mas poderosas da área de *Estatística de Extremos*, que têm sido largamente utilizadas nos mais variados campos, entre os quais destacamos *Ciências Ambientais*, *Finanças* e *Seguros*.

Começamos por apresentar no Capítulo 2 alguma *Motivação* para a necessidade da *Teoria de Valores Extremos* (TVE), muito frequentemente denotada EVT, do ingês '*Extreme Value Theory*'. No Capítulo 3 avançamos com algumas *Técnicas Gráficas* usadas na análise preliminar de qualquer tipo de dados, tais como os QQ-plots e os PP-plots, e *Técnicas Gráficas* específicas da área de valores extremos, como os ME-plots e os W-plots. No Capítulo 4, através de alguns exemplos de aplicação a dados univariados, tentamos responder à pergunta *Porquê a Teoria de Valores Extremos*? Mas em EVT, e mais geralmente, em quase todos as áreas da *Estatística*, a ordenação de uma amostra aleatória univariada, como base para uma representação clara do conteúdo dessa amostra, é crucial. Tal justifica a consideração dos Capítulos 5 e 6, respectivamente sobre o *Comportamento Distribucional Exacto* e o *Comportamento Distribucional Assintótico* das estatísticas ordinais. Finalmente, nos Capítulo 7, 8 e 9, debruçamo-nos sobre *Estatística de Extremos*, área de grande utilidade em aplicações quando se pretende inferir na cauda de um modelo, estimando parâmetros de acontecimentos raros, como por exemplo quantis elevados ou períodos de retorno de níveis elevados. No Capítulo 7, abordamos as perspectivas paramétricas de inferência estatística em acontecimentos raros. O Capítulo 8 é dedicado a alguns métodos de inferência semi-paramétrica. Finalmente, no Capítulo 9, procedemos à análise de três casos de estudo.

O texto é, como convém, consideravelmente mais ambicioso do que será o

curso breve no XXI *Congresso Anual da Sociedade Portuguesa de Estatística*. Fica no entanto como elemento de referência para os interessados, enquanto num curso de algumas horas, mesmo intensivas e com a celeridade que uma audiência conhecedora impõe, apenas os tópicos mais relevantes podem ser abordados. Qualquer curso é um compromisso, procurando um equilíbrio pessoal (neste caso de uma trindade geracional) entre o que é reconhecidamente fundamental e imprescindível, e os gostos e interesses de quem o escreve. Assim, ficaram naturalmente de fora questões muito importantes mas que não serviam o nosso *puzzle*, tal como ficaram de fora questões que estão entre os nossos interesses directos de investigação, tais como como estimação de viés reduzido, utilização de metodologias de re-amostragem, como o bootstrap e o jackknife em *Estatística de Extremos*, entre outros, mas que certamente iriam desequilibrar a dinâmica do texto.

Não é decerto por ingratidão, mas em todo o rol de agradecimentos há esquecimentos. Por isso preferimos os *clichés*: às nossas famílias e aos nossos amigos, aos nossos mestres, aos nossos colegas, aos nossos alunos.

Agradecemos por outro lado à *Sociedade Portuguesa de Estatística* (SPE) e aos organizadores do XXI *Congresso Anual da* SPE esta honra que nos conferiram. Agradecemos também o apoio institucional do CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa. E, claro, as palavras mágicas e o logotipo: Esta investigação foi parcialmente subsidiada por **FCT** — Fundação para a Ciência e a Tecnologia, projectos PEst-OE/MAT/UI0006/2011, EXTREMA, PTDC/MAT/101736/2008 e PTDC/MAT/112770/2009: EXTREMES IN SPACE.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

M. Ivette Gomes
M. Isabel Fraga Alves
Cláudia Neves

Capítulo 1

Introdução e Comentários Bibliográficos

Em *Teoria de Valores Extremos* (TVE), muito frequentemente denotada EVT, do ingês '*Extreme Value Theory*' a ordenação da amostra é primordial. Mais geralmente, e em quase todos as áreas da Estatística, a ordenação de uma amostra aleatória univariada, como base para uma representação clara do conteúdo dessa amostra, foi desde há muito considerada importante. Tal importância permitiu chegar ao patamar em que estamos hoje — uma vasta metodologia estatística e associada teoria distribucional relativas a amostras ordenadas — tal como se pode ver nos livros de Sarhan & Greenberg[1] (1962), David[2] (1981), Arnold & Balakrishnan[3] (1989), Reiss[4] (1989), Arnold *et al.*[5] (1992; 2008) e David & Nagaraja[6] (2003), sobre estatísticas ordinais (e.o.'s),

[1] Sarhan, A.E. & Greenberg, B.G. (1962). *Contributions to Order Statistics*. Wiley.

[2] David, H.A. (1981). *Order Statistics*, 2nd Ed., Wiley.

[3] Arnold, B.C. & Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics*. Springer-Verlag.

[4] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer-Verlag.

[5] Arnold, B., Balakhrishna, N. & Nagaraja, H. N. (1992; 2008). *A First Course in Order Statistics*. 1st Ed., Wiley; 2nd Ed., SIAM.

[6] David, H.A. & Nagaraja, H.N. (2003). *Order Statistics*. 3rd. Ed., Wiley.

e nos livros de Leadbetter *et al.*[7] (1984), Galambos[8] (1987), Resnick[9] (1987) e Falk *et al.*[10] (1994; 2005; 2010), sobre e.o.'s extremais.

Temos ainda de referir o clássico livro de Gumbel[11] (1958; 2004), livro pioneiro em *Estatística de Extremos*, e os livros de Beirlant *et al.*[12] (1996), Tiago de Oliveira[13] (1997), Reiss & Thomas[14] (1997; 2007), Embrechts *et al.*[15] (1998), Kotz & Nadarajah[16] (2000), Coles[17] (2001), Beirlant *et al.*[18] (2004), Castillo *et al.*[19] (2004), de Haan & Ferreira[20] (2006), Resnick[21] (2007) e Markovich[22] (2007).

Existe, por um lado, um *interesse natural pela ordenação*:

- Os valores extremos são crucialmente importantes como expressão do pior ou do melhor que pode ser encontrado numa amostra (temperaturas mínimas, níveis máximos de barragens, tempos de vida mínimos

[7] Leadbetter, R., Lindgren, G. & Rootzén, H. (1984). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag.

[8] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Krieger.

[9] Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag.

[10] Falk, M., Hüsler, J. & Reiss, R.-D. (1994; 2005; 2010). *Laws of Small Numbers: Extremes and Rare Events*. Birkhäuser.

[11] Gumbel, E.J. (1958; 2004). *Statistics of Extremes*. Columbia University Press, Dover Publications, Inc., New York.

[12] Beirlant, J., Teugels, J.L. & Vynckier, P. (1996). *Practical Analysis of Extremes*. Leuven University Press.

[13] Tiago de Oliveira, J. (1997). *Statistical Analysis of Extremes*. Pendor.

[14] Reiss, R.-D. & Thomas, M. (1997; 2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. 2nd Ed.; 3rd. Ed., Birkhaüser Verlag, Berlin.

[15] Embrechts, P., Klüppelberg, C. & Mickosh, T. (1998). *Modelling Extremal Events for Insurance and Finance*. Springer Verlag.

[16] Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions – Theory and Applications*. Imperial College Press, London.

[17] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

[18] Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes. Theory and Applications*. Wiley.

[19] Castillo E., Hadi A.S., Balakrishnan, N. & Sarabia, J.M. (2004). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley.

[20] de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: an Introduction*. Springer Science+Business Media, LLC, New York.

[21] Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Verlag.

[22] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*. Wiley.

em teoria da fiabilidade). Tais valores podem ainda dar-nos indicações sobre 'outliers', indicando influências estranhas ou erros no processo de colecção de dados.

- A amplitude da amostra é uma medida importante de escala.
- A mediana da amostra é uma medida importante de localização.
- A própria forma de um problema pode censurar um conjunto de dados, e somos muitas vezes forçados a trabalhar com um conjunto de e.o.'s superiores ou e.o.'s inferiores.

Tudo isto é *inevitável e natural*.

Alternativamente, um conjunto de observações pode ser *deliberadamente ordenado*, de forma a facilitar a análise estatística pretendida. Por exemplo, podemos estar interessados em:

- Estimadores centrados, de variância mínima, que sejam combinações lineares das e.o.'s.
- Métodos rápidos para estimação de parâmetros ou para testes de significância baseados em estatísticas sistemáticas, como a amplitude e a semi-amplitude.
- Eliminar valores extremos de forma a aumentar a robustez de um estimador, com a consideração de estimadores como a média *Winsorizada* e a *trimmed mean*.
- Usar a técnica de papel de probabilidade (PP) ou outras alternativas gráficas como o QQ-plot para validação de um modelo e estimação dos parâmetrosdesconhecidos, onde QQ é o acrónimo de *quantil versus quantil*.
- Usar a técnica dos resíduos ordenados em *Análise de Variância*.
- Metodologias de índole não paramétrica, onde a noção de ordem é fundamental.

De qualquer forma, e qualquer que seja a finalidade da ordenação dos dados, necessitamos obviamente do conhecimento prévio da teoria distribucional exacta, e frequentemente da teoria distribucional assintótica das e.o.'s.

No que se segue, vamos usar a notação usual X para uma variável aleatória (v.a.) genérica com função de distribuição (f.d.) F eventualmente dependente

de parâmetros desconhecidos de localização, $\lambda \in \mathbb{R}$, e de escala, $\delta \in \mathbb{R}^+$. Em situação univariada, a amostra original $(X_1, \ldots, X_n)$ é imediatamente ordenada ascendentemente e denotada $(X_{1:n} \leq \cdots \leq X_{n:n})$, com

$$X_{1:n} := \min_{1 \leq i \leq n} X_i \quad \text{e} \quad X_{n:n} := \max_{1 \leq i \leq n} X_i.$$

Usamos por vezes a notação $M_n^{(i)} := X_{n-i+1:n},\ 1 \leq i \leq n$, quando estamos interessados em e.o.'s superiores.

A notação Z é frequentemente usada para a v.a. standardizada, $Z = (X - \lambda)/\delta$, com a qual trabalhamos na maior parte das situações, por simplicidade e sem perda de generalidade, uma vez que $X_{i:n} = \lambda + \delta Z_{i:n}$, $1 \leq i \leq n$. Modelos de especial interesse no contexto das e.o.'s são o modelo Uniforme, o Exponencial e o Pareto. Usamos a notação óbvia para representar variáveis aleatórias (v.a.'s) provenientes desses modelos, U, E e P, respectivamente.

1.1 Tópicos a abordar

Neste livro, começamos por apresentar no Capítulo 2 alguma *Motivação* para a necessidade da EVT. No Capítulo 3 avançamos com algumas *Técnicas Gráficas* usadas na análise de valores extremos, tais como os QQ-plots, os PP-plots, os W-plots e os ME-plots. No Capítulo 4, através de alguns exemplos de aplicação a dados univariados nas áreas de ambiente, hidrologia, meteorologia e seguros, tentamos responder à pergunta *Porquê a Teoria de Valores Extremos*? No Capítulo 5 abordamos alguns resultados sobre a *Teoria Distribucional Exacta*, colocando-nos pois numa perspectiva probabilística. Referimos neste capítulo as distribuições exactas de e.o.'s, o cálculo dos seus momentos, a importância da sua estrutura markoviana, dando ainda algumas indicações sobre estatísticas sistemáticas e aproximações para os momentos. O Capítulo 6 incide sobre a *Teoria Distribucional Assintótica*, onde referimos as leis limite das e.o.'s centrais, extremais e intermédias, as leis limite estáveis para max e min–domínios de atracção, as condições necessárias e suficientes a impor nas caudas dos modelos subjacentes às amostras em estudo e os POT–domínios de atracção. São introduzidas neste capítulo a distribuição (geral) de valores extremos (GEV, do inglês '*general extreme value*' ou '*generalized extreme value*')

e a distribuição generalizada de Pareto (GP), bem como o índice de valores extremos (denotado EVI, do inglês '*extreme value index*') e a noção de peso de cauda, fortemente relacionada com a teoria das funções de variação regular (Bingham *et al.*[23], 1987). Finalmente, nos Capítulos 7, 8 e 9, debruçamo-nos sobre *Estatística de Extremos*, área de grande utilidade em aplicações quando se pretende inferir na cauda de um modelo, estimando parâmetros de acontecimentos raros, como por exemplo quantis elevados ou períodos de retorno de níveis elevados. No Capítulo 7, abordamos as perspectivas paramétricas de inferência estatística em acontecimentos raros e a escolha estatística de modelos extremais e de max-domínios de atracção, o chamado método dos máximos anuais (MMA), e entre outras, as metodologias POT (do inglês '*peaks over threshold*') e PORT (do inglês '*peaks over random threshold*'), muito úteis na inferência de acontecimentos extremos. O Capítulo 8 é dedicado a alguns métodos de inferência semi-paramétrica. No Capítulo 9, procedemos à análise de vários casos de estudo.

[23]Bingham, N., Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation.* Cambridge Univ. Press, Cambridge.

Capítulo 2

Motivação

É perfeitamente natural perguntar qual o porquê da EVT. Para motivar o interesse por este tema, damos em seguida alguns exemplos recentes de grande relevância para a sociedade, e que envolvem esta teoria.

2.1 Katrina: Um desastre (não) natural?

Nova Orleães encontra-se situada abaixo do nível do mar, no meio de dois lagos, a Norte e a Este, e do rio Mississipi a sul. De acordo com as informações divulgadas pelas autoridades locais, a inundação provocada pelo Katrina deveu-se, sobretudo, a uma brecha de 60 metros num dique junto ao lago Pontchartrain.

Traduzimos de forma livre parte de uma notícia do *New York Times, Sept'05*, intitulada '*New Orleans After Hurricane Katrina: An Unnatural Disaster*?' Dizia o redator que o que teriam de fazer em seguida era construir um sistema de diques adequado, para o que necessitariam de engenheiros holandeses, capazes de desenhar essas estruturas. A primeira estrutura deveria ser uma barragem com pelo menos 40-50 pés de altura, construída ao longo do lago e de cada canal com ligação ao lago. Tratar-se-ia de um plano que custaria biliões, mas conseguir-se-ia assim que Nova Orleães NUNCA tornasse a en-

Figura 2.1: Nova Orleães após o furacão Katrina

frentar semelhante tragédia. E terminava esperando que se aprendesse a lição, de modo a não se ter uma repetição dentro dos próximos 20 anos.

Parece óbvio que não só este desastre, mas também cheias históricas, como a que aconteceu no Mar do Norte às primeiras horas da manhã do dia 1 de Fevereiro de 1953, podem servir de guia. Nesse dia, o nível das águas excedeu então os 5.6 metros acima do nível do mar, destruiu as defesas marítimas, tendo inundado áreas na Holanda, Inglaterra, Bélgica, Dinamarca e França e cerca de 2500 pessoas morreram. Como resultado, o governo holandês, constituiu uma comissão, designada '*Delta Committee*'. O governo decretou que os diques devem ser construídos com uma altura tal que

- **a probabilidade de uma inundação num determinado ano é de 1 em 10.000**.

Ora o período de observação dos dados é muitíssimo mais curto!... É então necessário proceder a uma ***extrapolação*** para além dos dados observados!! ...E a EVT consegue dar respostas fidedignas sobre a altura da referida barragem, entrando em linha de conta com aquilo a que chamamos período de retorno (conceito a ser definido mais adiante) de um acontecimento extremo, como o furacão Katrina.

Figura 2.2: A cheia no Mar do Norte a 1 de Fevereiro de 1953

2.2 Extremos no mercado financeiro

O Comité de Basileia sobre o controlo bancário formula normas e directrizes de supervisão e recomenda boas práticas para as instituições financeiras. Entre outras medidas de risco, essa regulamentação envolve a estimação de uma quantidade denominada *Value-at-Risk* (VaR), que não é mais do que um quantil extremo da distribuição de perdas e ganhos. Como poderá ser estimado o VaR a partir da série de retornos diários R_t (em percentagem), definidos por

$$R_t = 100 \log(P_t/P_{t-1}),$$

sendo P_t o preço de fecho no dia t? Para o PSI20, apresentamos na Figura 2.3 os valores P_t (*esquerda*) e R_t (*direita*).

A simulação histórica é muito pobre! E existem contribuições positivas da EVT, como se ilustra nas duas figuras seguintes, 2.4 (veja-se Araújo Santos[1], 2011) e 2.5. Um breve texto crítico sobre o cálculo do VaR feito através

[1] Araújo Santos, P. (2011). *Excesses, Durations and Forecasting Value-at-Risk.* PhD Thesis, University Lisbon.

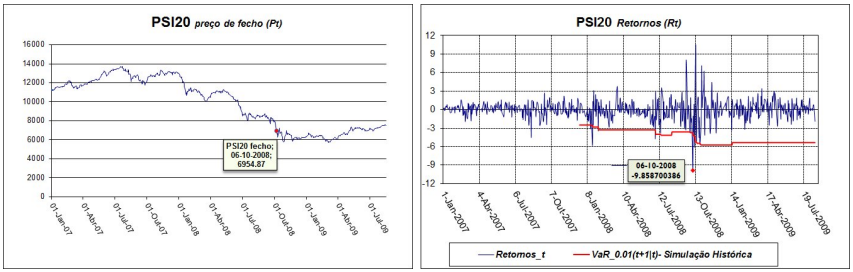


Figura 2.3: Preços de fecho (*esquerda*) e log-retornos diários (*direita*) do PSI20

das metodologias tradicionais (Normal-VaR) versus a EVT (EV-VaR), de que apresentamos a Figura 2.5, ilustrativa da comparação das duas metodologias, pode ser consultado em Aragonés *et al.*[2] (2000), um artigo introdutório em EVT.

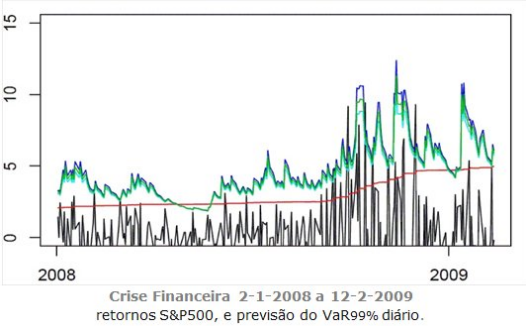


Figura 2.4: Previsão para o VaR99% diário do S&P500

As principais questões a ter em consideração são essencialmente as seguintes:

- Usualmente existem poucas observações na cauda da distribuição.
- São requeridas estimativas muito para além do máximo observado.
- Necessitamos de recorrer a modelos para a cauda baseados em resultados assintóticos.
- Será sensato usar esses modelos em todas as situações reais envolvendo acontecimentos raros?

[2] Aragonés, J., Blanco, C. & Dowd, K. (2000). The Learning Curve: Extreme Value Theory for VaR (`http://www.fea.com/resources/pdf/a_evt_1.pdf`Part 1 & `http://www.fea.com/resources/pdf/a_evt_2.pdf`Part 2).

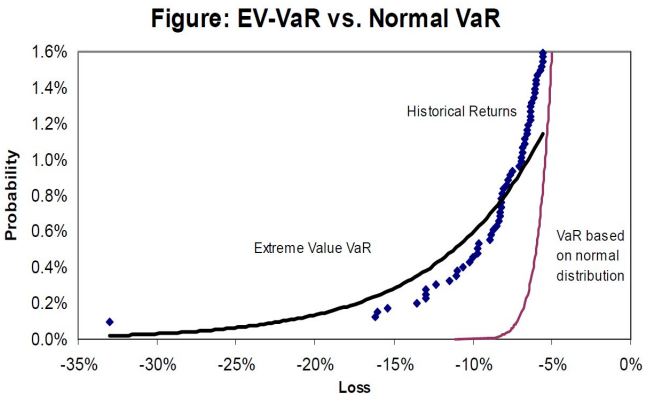


Figura 2.5: EV-VaR vs Normal-VaR

- É preciso não esquecer, parafraseando George Box (1919–2013), genro de Sir Ronald Fisher, que *'... all models are wrong but some models are useful'* (Box & Draper[3], 1987, p. 424).

Note-se desde já que as áreas de aplicação da EVT na análise de acontecimentos raros são tão diversas como o Ambiente, as Finanças, os Seguros, a Resistência de Materiais, o Desporto e a Sismologia, entre outras.

2.3 EVT: porque nem tudo é normal!

- De que altura deverá ser projectada uma barragem de aterro, de tal forma que o mar só atinja este nível uma vez em 1000 anos?
- Qual a probabilidade de rotura de determinado dique marítimo?
- Que ordem de grandeza poderá vir a atingir um '*crash*' bolsista amanhã?
- Qual a probabilidade de ser ultrapassada a melhor marca de 8.95m em salto em comprimento, dado o actual '*state of the art*'?

Muitas questões da vida real requerem a estimação sobre acontecimentos acerca dos quais os dados são inexistentes ou se existem são escassos — são

[3]Box, G.E.P. & Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces.* Wiley.

os designados *acontecimentos extremos ou raros*. A EVT é um ramo probabilístico de suporte à Estatística que lida exactamente com tais situaçõoes, ajudando a descrever e a quantificar os ditos acontecimentos raros. Em particular, permite a estimação de probabilidades de acontecimentos que não contêm dados, ou como usualmente dizemos, permite *extrapolar para além da amostra*.

Quantidades relevantes são, entre outras, um *quantil extremo*, noção já atrás referida, e o *período de retorno*, que não é mais do que o intervalo de tempo médio entre ocorrências de um determinado valor extremo.

Na análise de dados clássica os extremos podem vir a ser rotulados de '*outliers*', chegando por vezes mesmo a ser ignorados no estudo, uma vez que se afastam do modelo 'ajustado'. Se o objectivo for inferir acerca de acontecimentos do dia-a-dia, realmente poderá ser irrelevante suprimir tais dados das pontas, mas se a questão fulcral residir em eventos que não ocorrem com muita frequência então dever-se-á aplicar o contexto EVT, dando relevância exactamente a esses valores extremos.

Existirá um padrão escondido subjacente a todo o tipo de eventos?

Se medirmos as alturas de muitas pessoas de um mesmo estrato homogéneo e as representarmos por um simples histograma, facilmente descobrimos uma mesma regra, a famosa curva de Gauss, por vezes também denominada distribuição em forma de sino, que não é mais do que a constatação de que o modelo Normal como que 'regula' a característica em causa. Surpreendentemente (*ou talvez não ...*) muitos dos dados da vida real seguem a distribuição Normal e suas congéneres.

Metodologias estatísticas mais comuns assentam no pressuposto de que os dados disponíveis correspondem a realizações independentes de v.a.'s provenientes de uma população com distribuição Normal. É o caso, por exemplo, do teste-t para comparação de valores médios. Outras abordagens usuais são essencialmente motivadas pela concepção mais ou menos consensual de que qualquer fenómeno que dê origem a um grande número de observações independentes, em que nenhuma delas é dominante, pode ser convenientemente modelado por uma distribuição Normal. A manifesta vantagem que daqui deriva é a da possibilidade de simplificar um elevado número de situações, decorrente de propriedades que surgem como apanágio da distribuição Normal,

nomeadamente a que resulta da aplicação do Teorema Limite Central (TLC) para somas ou a que deriva desta distribuição poder ser completamente especificada à custa dos seus momentos.

Contudo, quando nos focamos nos extremos, localizados nas *caudas das distribuições*, esta deixa de ser uma verdade irrefutável.

Contrariamente à condição de normalidade acima descrita, não é difícil deparar, no decurso da vida quotidiana, com situações em que uma única observação que se afasta da tendência central dos dados poderá, pela sua magnitude, ser comparável à acumulação de todas as outras não dominantes. É também neste sentido que as e.o.'s extremais têm protagonizado tão grande número de situações práticas em áreas tão diversas quanto Seguros, Finanças, Hidrologia, Biologia, Controlo de Qualidade, Telecomunicações ou Teletráfego, ao ponto de justificar o constante desenvolvimento da *Teoria de Valores Extremos*.

Por exemplo, no campo financeiro, e em particular nas distribuições associadas aos retornos, é habitual encontrar caudas mais pesadas do que as abordagens clássicas consideram. Isto quer basicamente dizer o seguinte: os acontecimentos extremos, embora improváveis por hipótese, são mais frequentes do que seria de esperar segundo o modelo gaussiano, um modelo com caudas leves, de tipo Exponencial.

Existem situações onde a abordagem EVT é primordial. A distribuição associada às maiores observações para aplicações a dados ou temperaturas anuais de pico, por exemplo. Por outro lado, a distribuição das menores observações é aplicada a problemas de resistência de materiais, onde o princípio do elo mais fraco impera, ou ainda a fenómenos como a duração da vida humana, com limite superior de suporte necessariamente finito.

Em *Estatística*, o *Teorema de Fisher-Tippett-Gnedenko* (o teorema fulcral dos tipos em valores extremos) é um resultado acerca da distribuição assintótica das e.o.'s extremais. O teorema dos tipos extremais desempenha um papel análogo ao tão famoso TLC para as médias (somas). Basicamente, estabelece que *o máximo amostral convenientemente normalizado converge para uma de* 3 *distribuições possíveis, a* Gumbel *a* Fréchet *ou a* Max-Weibull, a serem estudadas mais adiante, no Capítulo 7, mas abordadas em situações várias ao longo deste livro. Independentemente da forma do centro de distribuição, a *cauda assume formas sempre muito especiais* quando estamos suficiente-

mente longe na cauda. O crédito deste resultado é devido essencialmente a Gnedenko[4] (1943), embora versões anteriores tivessem sido estabelecidas por Fréchet[5] (1927) e Fisher & Tippett[6] (1928).

2.4 Estatísticos históricos na área de extremos

Começamos por referir Fisher e Tippet, na Figura 2.6, e em seguida Weibull, Gumbel e Fréchet, na Figura 2.7. Finalmente, na Figura 2.8, referimos von Mises (veja-se von Mises[7], 1936), outro dos pioneiros na área.

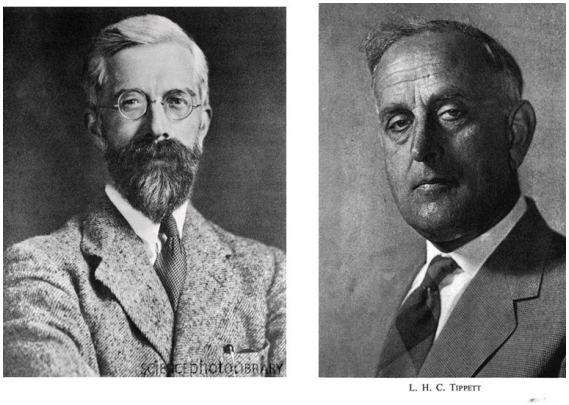


Figura 2.6: Sir Ronald Alymer Fisher (1890-1962) e Leonard Henry Caleb Tippett (1902-1985)

Referimos em seguida algumas das frases célebres de Emil Gumbel, um dos nome sonantes e pioneiros na área de *Estatística de Extremos*:

[4]Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**:6, 423–453.

[5]Fréchet, M. (1927). Sur le loi de probabilité de l'écart maximum. *Ann. Société Polonaise de Mathématique* **6**, 93–116.

[6]Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings Cambridge Philosophical Society* **24**, 180–190.

[7]Mises, R. von (1936). La distribution de la plus grande de n valeurs. *Revue Math. Union Interbalcanique* **1**, 141–160. Reprinted in *Selected Papers of Richard von Mises*, Amer. Math. Soc. **2** (1954), 271–294.

Figura 2.7: Ernst Hjalmar Waloddi Weibull (1887-1979), Emil Julius Gumbel (1891-1966), e Maurice René Fréchet (1878-1973)

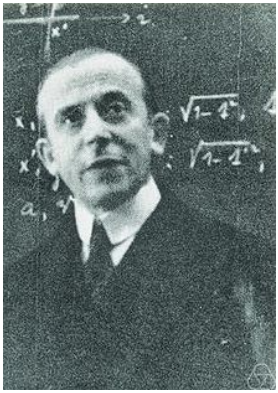


Figura 2.8: Richard Edler von Mises (1883-1953)

'It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.'
'Il est impossible que l'improbable n'arrive jamais.'
'Il y aura toujours une valeur qui dépassera toutes les autres.'

Existem hoje em dia vários 'R-*Packages for Extreme Values*', tais como **evd**, **evdbayes**, **evir**, **ismev**, **extRemes**, **extremevalues**, **fExtremes**, **lmom**, **lmomRFA**, **lmomco**, **POT**, **SpatialExtremes**, alguns dos quais a serem usados neste livro.

Capítulo 3

Metodologias Gráficas para Análise Preliminar de Valores Extremos (APVE)

É óbvio que a *linearidade num gráfico* pode ser facilmente constatada por observação directa de uma *nuvem de pontos*, e quantificada em termos do *coeficiente de correlação*. A ideia subjacente aos PP-plots, ou equivalentemente aos actuais QQ-plots, existentes em quase todos os 'packages' estatísticos, e a estudar mais em pormenor nas secções 3.1 e 3.2, respectivamente, surgiu da necessidade de responder à pergunta:

Será que um determinado modelo probabilístico fornece um ajustamento sensato à distribuição subjacente aos dados em causa?

O método gráfico mais antigo para selecção de modelos é a técnica do *papel de probabilidade*, que introduzimos em seguida, e que pode ser visto com mais detalhe em Gomes *et al.*[1] (2010).

[1] Gomes, M.I., Figueiredo, F. & Barão, M.I. (2010). *Controlo Estatístico da Qualidade*. Edições INE.

3.1 Método Gráfico Clássico de Selecção de Modelos — Papel de Probabilidade (PP-plot)

A técnica do *papel de probabilidade* que, com modificações convenientes, pode ser usada para dados contínuos ou discretos, completos ou censurados, tem sido usada nas mais variadas formas, desde que Hazen[2] (1914) (veja-se também Hazen[3], 1930) sugeriu o princípio de linearização da f.d. Normal, num estudo de cheias, mas a sua principal aplicação tem sido na *obtenção de uma confirmação visual rápida do ajustamento de determinado modelo probabilístico, sugerido por exemplo pelo histograma, a dados* $(x_1, \ldots, x_n)$*, permitindo ainda a estimação grosseira de parâmetros.*

O *papel de probabilidade* é frequentemente usado quando os dados, $(x_1, \ldots, x_n)$, podem ser considerados observações independentes de uma v.a. X com f.d. do tipo $F((x-\lambda)/\delta)$, λ e δ parâmetros de localização e escala, respectivamente. Trata-se de um método de linearização da f.d.: face à amostra ordenada $(x_{1:n} \leq \cdots \leq x_{n:n})$, e para um modelo $F(x) = F((x-\lambda)/\delta)$, represente-se graficamente a nuvem de pontos:

$$(x_{i:n},\ y_i := F^{\leftarrow}(p_i)), \quad p_i := i/(n+1),\ 1 \leq i \leq n, \tag{3.1}$$

onde $F^{\leftarrow}$ denota a inversa generalizada de F, i.e.,

$$F^{\leftarrow}(x) := \inf\{y : F(y) \geq x\},\ 0 \leq x \leq 1. \tag{3.2}$$

Se o gráfico resultante mostrar que existe uma relação linear entre $x_{i:n}$ e y_i temos uma validação informal do modelo $F(\cdot)$, postulado. A intersecção com o eixo das abcissas e a inclinação da recta fornecem-nos então estimativas grosseiras de λ e δ.

Na realidade, admitindo que $F^{-1}(\cdot)$ existe, sendo $F^{-1}(x)$ o valor de y tal que $F(y) = x$, e escrevendo

$$p_i = F((x_{i:n} - \lambda)/\delta), \quad 1 \leq i \leq n,$$

[2] Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Amer. Soc. Civil Engrs.* **77**, 1539-1659.

[3] Hazen, A. (1930). *Flood Flows. A Study of Frequencies and Magnitudes.* Wiley.

tem-se

$$y_i = F^{-1}(p_i) = x_{i:n}/\delta - \lambda/\delta \iff x_{i:n} = \lambda + \delta y_i, \quad 1 \leq i \leq n,$$

i.e. existe uma relação linear entre $x_{i:n}$ e $y_i = F^{-1}(p_i)$, devendo p_i ser qualquer estimativa plausível de $F((X_{i:n} - \lambda)/\delta)$. Uma escolha possível para os valores de p_i, $1 \leq i \leq n$, as chamadas '*plotting positions*', foi dada em Weibull[4] (1939). Trata-se dos valores $p_i = i/(n+1)$, $1 \leq i \leq n$, já definidos em (3.1), os valores de $\mathbb{E}\big(F((X_{i:n} - \lambda)/\delta)\big)$, $\forall F(\cdot)$ absolutamente contínua, uma vez que então

$$F\Big(\frac{X_{i:n} - \lambda}{\delta}\Big) \stackrel{d}{=} B_{i,n-i+1},$$

onde $B_{p,q}$ denota uma v.a. Beta de parâmetros p e q, i.e., uma v.a. com f.d.p.,

$$f(z; p, q) = \frac{1}{B(p,q)}\, z^{p-1}(1-z)^{q-1}, \quad 0 \leq z \leq 1,$$

com $B(\cdot, \cdot)$ a função Beta completa.

Observação 3.1.1. *A função* Beta completa *é o integral,*

$$B(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \alpha, \beta \in \mathbb{R}^+, \tag{3.3}$$

com $\Gamma(\cdot)$ *a função* Gama (factorial) completa, *i.e.,*

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}e^{-x}dx, \quad \alpha \in \mathbb{R}^+. \tag{3.4}$$

A função Gama *é a extensão da função* factorial *a qualquer real positivo* (*na realidade, a função* Gama *pode mais geralmente ser definida para qualquer complexo* z*, cuja parte real seja positiva — para detalhes, veja-se por exemplo o* Capítulo 6 *de* Abramowitz & Stegun[5]*,* 1972). *Para valores de* $n \in \mathbb{N}_0$*, inteiro não negativo, temos*

$$\Gamma(n+1) = 1 \times 2 \times 3 \times \cdots \times (n-1) \times n = n!, \quad (0! \equiv 1).$$

Não podemos aqui deixar de referir uma relação de recorrência relativa à função Gama *frequentemente utilizada ao longo deste livro,*

$$\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha), \quad \alpha > 0.$$

[4]Weibull, W. (1939). *A Statistical Theory of Strength of Materials.* Ing. Vet. A.K, Handl., 151, Genelstabens Litografiska Anstals Forlg Stocklholm, Sweden.

[5]Abramowitz, M. & Stegun, I.A. (1992). *Handbook of Mathematical Functions.* Dover, New York.

Valores particulares importantes são

$$\Gamma\left(\tfrac{1}{2}\right) = 2\int_0^\infty e^{-t^2}dt = \sqrt{\pi} = 1.77245\,38509\ldots, \quad \Gamma(\tfrac{3}{4}) = 1.22541\,67024\ \ldots.$$

Podemos ainda dizer que em modelo F absolutamente contínuo se tem

$$\mathbb{E}\left(F\Big(\frac{X_{i:n}-\lambda}{\delta}\Big)\right) = \mathbb{E}(B_{i,n-i+1}) = \mathbb{E}(U_{i:n}) = \frac{i}{n+1}, \quad 1 \le i \le n.$$

Para outras possíveis escolhas de *plotting positions* em papel de probabilidade veja-se, por exemplo, Barnett[6] (1975).

Tal como já foi referido, se o gráfico resultante mostrar que existe uma relação linear entre $x_{i:n}$ e $y_i = F^{-1}(i/(n+1)) =: Q(i/(n+1))$, com $Q(\cdot)$ a função quantil, temos uma validação informal da forma da distribuição $F(\cdot)$, postulada. A intersecção com o eixo das abcissas e a inclinação da recta fornecem-nos então estimativas grosseiras de λ e δ. Caso exista linearidade, a estimação dos parâmetros pode então ser feita através do módulo de regressão de qualquer *package* estatístico.

3.1.1 Referência histórica aos papéis de probabilidade

Como o processo pretendia ser um método visual rápido, tornava-se importante a facilidade da sua aplicação, tendo sido produzidos tipos especiais variados de *papel de probabilidade*, com uma escala funcional que mede convenientemente $F^{-1}(p)$, mas que é graduada em p. Torna-se então unicamente necessário representar graficamente $x_{i:n}$ versus p_i na(s) escala(s) transformada(s). Com a acessibilidade a algoritmos computacionais de cálculo de $F^{-1}(\cdot)$ para uma grande variedade de modelos, estes *papéis de probabilidade* têm hoje em dia apenas **interesse histórico**.

O exemplo mais vulgar, ainda muito usado em aplicações diversas, particularmente em áreas de *Hidrologia* e *Climatologia Estatística*, é o *papel de probabilidade Normal* (Chernof & Lieberman[7], 1954), acessível em qualquer papelaria do Reino Unido pelo menos até meados dos anos 80, e representado graficamente na Figura 3.1.

[6] Barnett, V. (1975). Probability plotting methods and order statistics. *Applied Statistics* **24**, 95–108.

[7] Chernoff, H. & Lieberman, G.J. (1954). Use of normal probability paper. *J. Amer. Statist. Assoc.* **49**, 778–785.

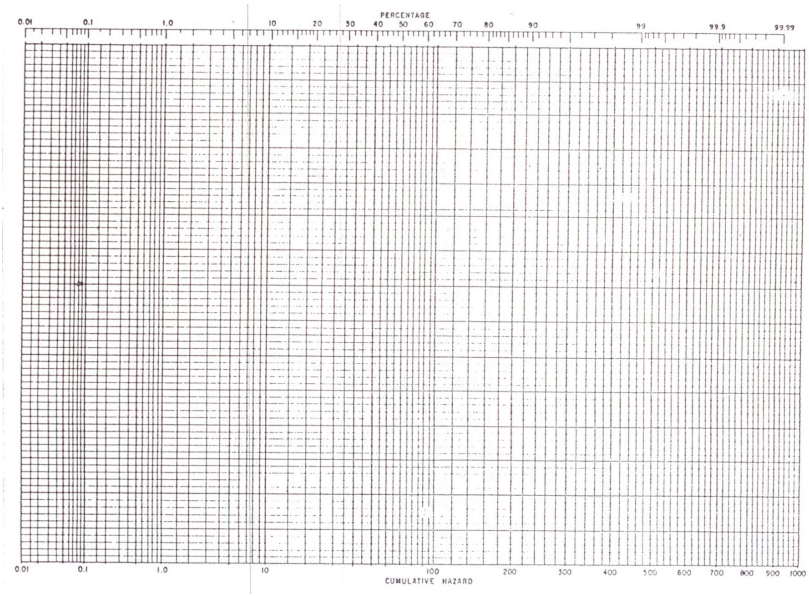


Figura 3.1: Papel de probabilidade Normal

No *papel de probabilidade Normal*, ilustrado na Figura 3.1, uma das escalas (neste caso, a das ordenadas) é uma escala aritmética, em que se marcam as observações ordenadas. A outra escala (a das abcissas) é uma escala probabilística, graduada em $\Phi^{-1}(p)$, com $\Phi(\cdot)$ a f.d. da Normal reduzida, $\mathcal{N}(0,1)$, mas em que se marca p (ou $100 \times p$). Esta escala funcional aparece ilustrada na Figura 3.2, e tem a seguinte tabela associada:

p	0.001	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	...
$\Phi^{-1}(p)$	-3.09	-2.33	-1.28	-0.84	-0.52	-0.25	0.0	0.25	0.52	...

Se as observações recolhidas forem efectivamente $\mathcal{N}(\lambda, \delta)$ teremos um gráfico do tipo do ilustrado na Figura 3.3.

Ao ajustarmos uma recta aos pontos marcados em papel de probabilidade Normal, obtemos facilmente estimativas de λ e δ, dadas por

$$\begin{aligned} \lambda^* &= \text{abcissa no ponto } 50\%, \\ \delta^* &= \tfrac{1}{2} \text{ (diferença entre as abcissas dos pontos } 84\% \text{ e } 16\%). \end{aligned}$$

Note-se no entanto que se para dados normais não usarmos *papel de probabi-*

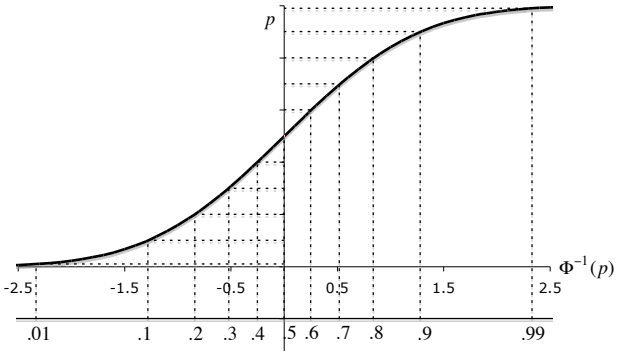


Figura 3.2: Escala funcional de um papel de probabilidade Normal

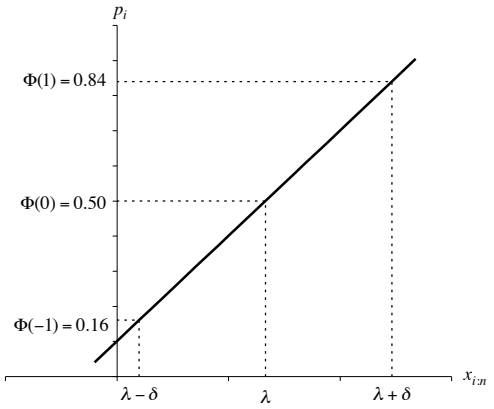


Figura 3.3: Gráfico em papel de probabilidade Normal

lidade, mas fizermos um gráfico de $\left(\Phi^{-1}(i/(n+1)), x_{i:n}\right)$, $1 \leq i \leq n$, com Φ a f.d. da $\mathcal{N}(0,1)$ (o actualmente chamado QQ-plot), como se tem

$$x_{i:n} = \lambda + \delta\ \Phi^{-1}\Big(\frac{i}{n+1}\Big),$$

obtemos as estimativas:

λ^{**} = intersecção com o eixo dos $x_{i:n}$ (ordenadas),

δ^{**} = inclinação da recta ajustada.

Exemplo 3.1.1. *Utilizando o* `package R`*, gerámos* 250 *observações de um modelo* $\mathcal{N}(3,1)$*. Temos para isso disponível a função* `rnorm`*. O gráfico da* Figura 3.4*, à esquerda, está associado às instruções:*

```
> x <- rnorm(250,3,1)
> x_in <- sort(x)
> n <- length(x)
> p_i <- (1:n)/(n+1)
> y_i <- qnorm(p_i)
> plot(y_i, x_in, col="blue", xlab=expression(y[i]), ylab=expression(x[i:n]),cex.lab=1.2)
> res1 <- lm(x_in ~ y_i)
> abline(res1, col="red",lty=1,lwd=2)
> legend(0,1,c("lmline"), col="red", lty=1,lwd=2, bty="n", cex=1)
```

A recta `abline` *está relacionada com a regressão linear, e o método de mínimos quadrados.*

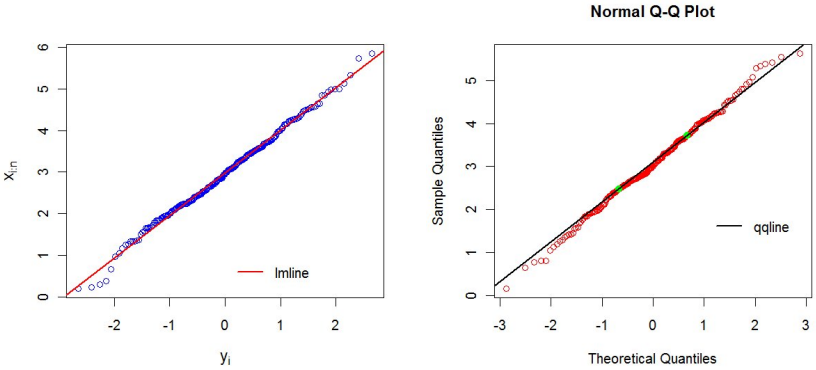


Figura 3.4: Papel de probabilidade (*esquerda*) e QQ-plot (*direita*) normais

Obtemos um gráfico semelhante, com a função `qqnorm`*, que permite o traçado do chamado* QQ-plot Normal*, ao qual ajustámos também a recta* `qqline`*. No caso do* QQ-plot Normal*, a função* `qqline` *permite o ajustamento de uma recta que passa pelo* 1^o *e* 3^o *quartis.*

O gráfico da Figura 3.4*, à direita, foi obtido através das instruções:*

```
> qqnorm(x, col="red")
> points(c(qnorm(0.25),qnorm(0.75)),
+ quantile(x,c(.25,.75)),col="green",cex=1.1,bg="green",pch=21)
```

```
> qqline(x, lwd=1.8)
> legend(1,2,c("qqline"), col="black", lty=1,lwd=2, bty="n", cex=1)
```

Foi ainda feita uma análise dos resíduos associados à regressão linear, que nos fornece estimativas para λ e δ. O 'output' foi o seguinte:

```
>  summary(res1)
Call: lm(formula = x_in ~ y_i)
Residuals:
 Min                    1Q            Median          3Q              Max
-0.725256   -0.037243   0.006295   0.043307   0.208526
Coefficients:
                      Estimate     Std. Error   t value   Pr(>|t|)
(Intercept)   2.983100=: lambda**   0.005704    523.0     <2e-16     ***
y_i           1.031064=: delta**    0.005814     77.3     <2e-16     ***
---
Signif. codes:  0'***' 0.001'**' 0.01'*' 0.05'.' 0.1' ' 1

Residual standard error: 0.09019 on 248 degrees of freedom
Multiple R-squared: 0.9922,Adjusted R-squared: 0.9921
F-statistic: 3.145e+04 on 1 and 248 DF,  p-value: < 2.2e-16
```

Outro exemplo simples é o *papel de probabilidade* Gumbel, um modelo muito usual em *Estatística de Extremos*, como veremos mais adiante.

Exemplo 3.1.2. *Se $F \equiv \Lambda$, a f.d.* Gumbel*, dada por:*

$$\Lambda(x;\lambda,\delta) = e^{-e^{-(x-\lambda)/\delta}},\ x \in \mathbb{R} \quad \Longrightarrow \quad x_{i:n} = \lambda + \delta(-\log(-\log(p_i))).$$

Consequentemente, o papel de probabilidade Gumbel *terá uma escala aritmética (onde marcamos as observações ordenadas ascendentemente, $x_{i:n}$, $1 \leq i \leq n$), versus uma escala duplamente logarítmica (onde marcamos as 'plotting positions', $p_i = i/(n+1),\ 1 \leq i \leq n$).*

Do ponto de vista conceptual, é óbvio que também podemos marcar $x_{i:n}$ versus $y_i = -\log(-\log(i/(n+1)))$ (ou $y_i = -\log(-\log(i/(n+1)))$ versus $x_{i:n}$), $1 \leq i \leq n$, num papel milimétrico usual, ou utilizar um QQ-plot*. É aliás isto que se faz, quando possuímos facilidades computacionais, como o package* `R`*, entre outros, a situação usual nos dias de hoje.*

A geração de observações ou números pseudo-aleatórios (NPA's) Gumbel *é simples. Procedemos pois à geração de* 250 NPA's Gumbel(0,1)*, colocados no vector* `gumb`*. O gráfico da* Figura 3.5 *(esquerda) foi obtido através dos comandos:*

```
> qqnorm(gumb, col="red")
> points(c(qnorm(0.25),qnorm(0.75)),
+ quantile(gumb,c(.25,.75)),col="green",cex=1.1,bg="green",pch=21)
> qqline(gumb, lwd=1.8)
> legend(1,0,c("qqline"), col="black", lty=1,lwd=2, bty="n", cex=1)
```

Trata-se pois de um QQ-plot Normal*, que fornece indicação imediata da não-normalidade dos dados. O traçado da nuvem de pontos* $\big(y_i = -\log(-\log(i/(n+1))), x_{i:n}\big)$*,* $1 \leq i \leq n$*, forneceu o gráfico da* Figura 3.5 (*direita*)*, e foi obtido de forma análoga ao que fizemos anteriormente para os dados de uma* $\mathcal{N}(3,1)$*.*

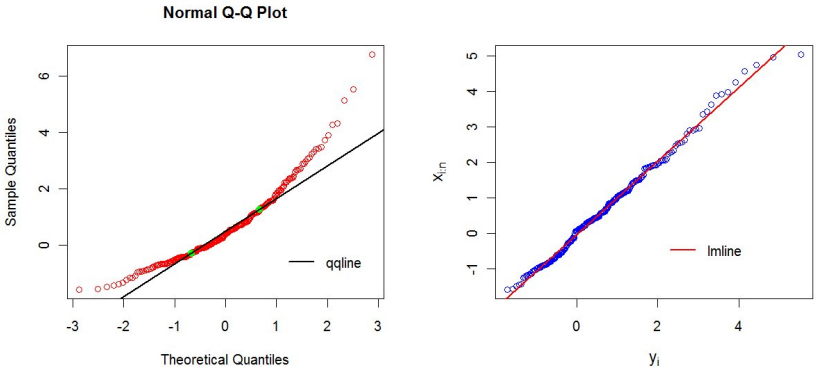


Figura 3.5: Dados Gumbel em '*papel de probabilidade*' Normal (*esquerda*) e Gumbel (*direita*)

A recta dos mínimos quadrados fornece-nos as estimativas,

$$\lambda^{**} = -0.102893, \qquad \delta^{**} = 0.993790.$$

Quando o gráfico em papel de probabilidade é nitidamente não linear, resultando consequentemente a rejeição do modelo postulado, $F(\cdot)$, podemos obter informação adicional a partir do gráfico (para mais detalhes veja-se Bury[8], 1975; Gomes *et al.*, 2010).

[8] Bury, K.V. (1975). *Statistical Models in Applied Science*. Wiley.

3.2 QQ-plots: outra perspectiva equivalente

3.2.1 QQ–plot: modelo Exponencial

Em *Estatística de Extremos*, e como veremos mais adiante nos Capítulos 6, 7–9, o modelo Exponencial de parâmetro $\lambda > 0$, denotado por $\mathcal{E}(\lambda)$, desempenha um papel bem mais importante do que o modelo Normal. A função de sobrevivência (ou cauda) é para a $\mathcal{E}(\lambda)$,

$$1 - F_\lambda(x) := \exp(-\lambda x),\ x > 0,$$

e para o caso da Exponencial standard ou reduzida,

$$1 - F_1(x) := \exp(-x),\ x > 0.$$

Será que a distribuição subjacente às observações $x_1, \ldots, x_n$ pertence a esta família $\mathcal{E}(\lambda)$? Para este modelo, temos a *função quantil*,

$$Q_\lambda(p) = F^{\leftarrow}(p) = -\frac{1}{\lambda}\log(1-p),\ p \in (0,1),$$

e para a $\mathcal{E}(1)$ temos pois a função quantil,

$$Q_1(p) = -\log(1-p),\ p \in (0,1).$$

Existe então uma relação linear

$$Q_\lambda(p) = \frac{1}{\lambda}Q_1(p) = \frac{1}{\lambda}(-\log(1-p)),\ p \in (0,1).$$

Dada uma amostra de observações $(x_1, \ldots, x_n)$, substitua-se $Q(p)$ pela contrapartida empírica $\widehat{Q}_n(p)$, e represente-se num sistema de eixos ortogonais a nuvem de pontos

$$(-\log(1-p), \widehat{Q}_n(p)),\ \text{para valores de}\ \ p \in (0,1).$$

Se o modelo Exponencial for bem ajustado, espera-se que a nuvem de pontos se distribua ao longo de uma recta. O declive dessa recta pode ser identificado com $1/\lambda$ e usado para obtenção de uma estimativa preliminar de λ. Note-se que a ordenada na origem deverá ser nula, já que $Q(0) = 0$.

De um modo geral, e denotando por $x_{i:n}$ o *i-ésimo* valor amostral, a nuvem de pontos

$$\widehat{Q}_n(p) = x_{i:n}, \text{ para } \frac{i-1}{n} < p \leq \frac{i}{n},$$

deve ser aproximadamente linear. Tal como mencionámos atrás, são possíveis várias escolhas de p (*plotting positions*), de entre as quais referimos:

$$p \in \left\{ \tfrac{1}{n}, \tfrac{2}{n}, \ldots, \tfrac{n-1}{n}, 1 \right\},$$

$$p \in \left\{ \tfrac{1-.5}{n}, \tfrac{2-.5}{n}, \ldots, \tfrac{n-1-.5}{n}, \tfrac{n-.5}{n} \right\},$$

$$p \in \left\{ \tfrac{1}{n+1}, \tfrac{2}{n+1}, \ldots, \tfrac{n-1}{n+1}, \tfrac{n}{n+1} \right\},$$

sendo usual considerar as *plotting positions* definidas em (3.1), i.e., $p_i = i/(n+1),\ 1 \leq i \leq n$, que, tal como já foi dito anteriormente, podem ser consideradas como o posicionamento dado pelo valor médio das e.o.'s associadas ao modelo Uniforme, $\mathcal{U}(0,1)$, uma vez que $F(X_{i:n}) \stackrel{d}{=} U_{i:n},$ para $i = 1, \ldots, n$, e se tem $U_{i:n} \frown \text{Beta}(i, n-i+1)$ de valor médio $\mathbb{E}[U_{i:n}] = i/(n+1) =: p_i$.

Voltemos ao caso Exponencial. A recta (declive=a; ordenada na origem=0) ajustada à nuvem de pontos pelo *método dos mínimos-quadrados*, é obtida pela minimização de

$$\sum_{i=1}^{n} \left(x_{i:n} + a \, \log(1-p_i)\right)^2 ,$$

vindo

$$\hat{a} = \frac{\sum_{i=1}^{n} x_{i:n} q_i}{\sum_{i=1}^{n} q_i^2}, \quad \text{com } q_i := -\log(1-p_i), \quad i = 1, \ldots, n\,.$$

Para um conjunto de 1000 NPA's, $\mathcal{E}(\lambda)$, com $\lambda = 2$, gerados no `R`, o gráfico é o apresentado na Figura 3.6. O declive da recta, com ordenada na origem nula, é de $0.5091 = 1/\hat{\lambda}$ e consequentemente, lembrando a relação teórica $Q_\lambda(p) = \frac{1}{\lambda}(-\log(1-p))$, temos uma estimativa preliminar dada por $\hat{\lambda} = 1.9643$.

Vejamos uma outra interpretação: A função que se está a aproximar ao marcar a nuvem de pontos $x_{i:n} \leftrightsquigarrow -\log(1-p_i),\ i = 1, \ldots, n$, é $x \mapsto -\log(1-F(x))$. Esta é exactamente a transformação que converte qualquer v.a. X, com f.d.

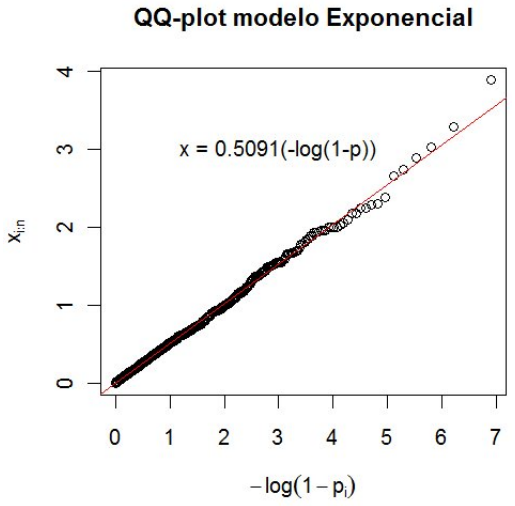


Figura 3.6: QQ-plot Exponencial

contínua F, na $\mathcal{E}(1)$. Realmente

$$\begin{aligned}\mathbb{P}[-\log(1-F(X)) \leq x] &= \mathbb{P}[X \leq Q(1-\exp(-x))] \\ &= F(Q(1-\exp(-x)) = 1-\exp(-x)\end{aligned}$$

ou seja,

$$-\log(1-F(X)) \frown \mathcal{E}(1).$$

Pensemos agora nas observações acima de um nível t (método POT, do inglês *peaks over threshold*). Na realidade, muitas vezes os dados só estão disponíveis acima de um nível t. Por exemplo, uma Resseguradora pode só receber informação acerca de pedidos de indemnização acima de um nível/franquia t, elevado. Abordemos o caso de X ser $\mathcal{E}(\lambda)$, e condicionemos no acontecimento $\{X > t\}$,

$$\mathbb{P}[X > x | X > t] = \frac{\mathbb{P}[X > x]}{\mathbb{P}[X > t]} = \exp(-\lambda(x-t)), \text{ para } x > t,$$

pelo que a correspondente *função quantil* é

$$Q(p) = t - \frac{1}{\lambda}\log(1-p),\ 0 < p < 1\,.$$

Então o QQ-plot tem ordenada na origem igual a t.

Como estimar um **quantil extremal**

$$q_p := Q(1-p), \text{ com } p \text{ pequeno ?}$$

Se pelo QQ-plot é sensato assumir o modelo Exponencial, então

$$\hat{q}_p = t - \frac{1}{\hat{\lambda}} \log(p).$$

Inversamente, uma **probabilidade de excedência pequena**

$$p \equiv p_x := \mathbb{P}[X > x | X > t]$$

pode ser estimada por

$$\hat{p}_x = \exp\left(-\hat{\lambda}(x-t)\right).$$

Estimação preliminar de λ: Poder-se-á estimar λ a partir do QQ-plot, através do método dos *mínimos quadrados*, ou alternativamente, considerar o estimador de *máxima verosimilhança* (ML, do inglês '*maximum likelihood*'),

$$\hat{\lambda} = 1/\big(\overline{x} - t\big).$$

3.2.2 QQ–plot: caso geral

No caso geral, seja Q_s a função quantil para o modelo standard de uma determinada família. De forma a aceitarmos um modelo como plausível para a população subjacente à amostra:

1. Deve existir uma relação linear entre os quantis teóricos $Q(p)$ e $Q_s(p)$.
2. Os quantis teóricos $Q(p)$, desconhecidos, devem ser substituídos pelos quantis empíricos $\widehat{Q}_n(p)$.
3. Devemos pois representar graficamente a nuvem de pontos

$$\left\{\left(Q_s(\tfrac{i}{n+1}), \widehat{Q}_n(\tfrac{i}{n+1})\right) = (Q_s(p_i), x_{i:n}) : i = 1, \ldots, n\right\}.$$

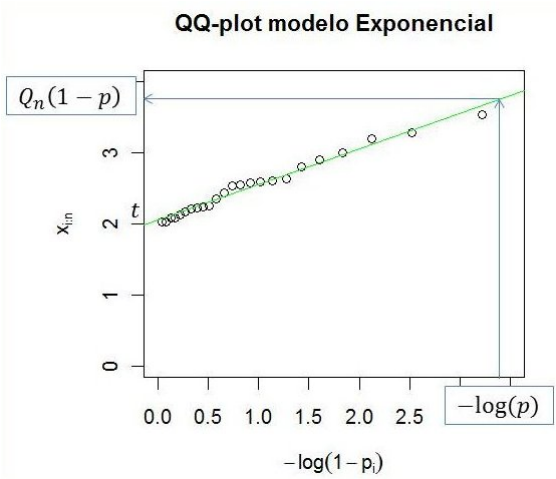


Figura 3.7: QQ-plot para dados $\mathcal{E}(\lambda)$ e estimação de quantis

4. Finalmente, devemos investigar a linearidade, levando a cabo uma regressão linear no QQ-plot, por exemplo.

Os quantis e períodos de retorno podem ser estimados através da aceitação de um relação linear no QQ-plot, $y = \hat{b} + \hat{a}x$, com

$$\overline{q} = \frac{1}{n}\sum_{i=1}^{n} Q_s(p_i), \quad \hat{a} = \frac{\sum_{i=1}^{n}(x_{i:n} - \overline{x})Q_s(p_i)}{\sum_{i=1}^{n}(Q_s(p_i) - \overline{q})^2} \quad \text{e} \quad \hat{b} = \overline{x} - \hat{a}\overline{q}\,.$$

Com F_s a f.d. reduzida e $Q_s = F_s^{\leftarrow}$ a inversa generalizada de F_s, $q_p = Q(1-p)$ e $p_x = \mathbb{P}[X > x]$, tem-se:

- **Estimação de quantis extremais:** $\hat{q}_p = \hat{b} + \hat{a}Q_s(1-p)$.
- **Probabilidades de excedência:** $\hat{p}_x = \overline{F}_s((x-\hat{b})/\hat{a})$, $\overline{F}_s := 1 - F_s$.

3.2.3 QQ–plots para modelos Normal e Log-Normal

Uma vez que os quantis da Normal, $\mathcal{N}(\mu, \sigma)$, se relacionam com os quantis da Normal *standard*, $\mathcal{N}(0,1)$, através de

$$Q(p) = \mu + \sigma\Phi^{-1}(p), \;\; \Phi^{-1} \text{ função quantil da } \mathcal{N}(0,1)\,,$$

as coordenadas do QQ-plot para este modelo são

$$\left(\Phi^{-1}(p_i), x_{i:n}\right),\ i = 1, \ldots, n\,.$$

Uma vez que a transformação logarítmica da Log-Normal é uma Normal, as coordenadas do QQ-plot para o modelo Log-Normal são obtidos pela transformação logarítmica dos dados

$$\left(\Phi^{-1}(p_i), \log x_{i:n}\right),\ i = 1, \ldots, n\,.$$

3.2.4 QQ–plot: Tabela de distribuições

A Tabela seguinte foi retirada de Beirlant *et al.* (2004):

Table 1.1 QQ-plot coordinates for some distributions.

Distribution	$F(x)$	Coordinates
Normal	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) \mathrm{d}u$ $x \in \mathbb{R}; \mu \in \mathbb{R}, \sigma > 0$	$(\Phi^{-1}(p_{i,n}), x_{i,n})$
Log-normal	$\int_{0}^{x} \frac{1}{\sqrt{2\pi}\sigma u} \exp\left(-\frac{(\log u-\mu)^2}{2\sigma^2}\right) \mathrm{d}u$ $x > 0; \mu \in \mathbb{R}, \sigma > 0$	$(\Phi^{-1}(p_{i,n}), \log x_{i,n})$
Exponential	$1 - \exp(-\lambda x)$ $x > 0; \lambda > 0$	$(-\log(1 - p_{i,n}), x_{i,n})$
Pareto	$1 - x^{-\alpha}$ $x > 1; \alpha > 0$	$(-\log(1 - p_{i,n}), \log x_{i,n})$
Weibull	$1 - \exp(-\lambda x^{\tau})$ $x > 0; \lambda, \tau > 0$	$(\log(-\log(1 - p_{i,n})),$ $\log x_{i,n})$

3.3 QQ-plots e PP-plots: caso geral $F(\cdot|\theta)$

Até agora a maior parte dos modelos considerados permitiram a construção dos QQ-plots *sem qualquer conhecimento dos valores exactos dos parâmetros*. Aliás, *estimativas preliminares* desses valores puderam ser obtidas como resultado colateral do QQ-plot. Esta situação está essencialmente relacionada com

o caso de estarmos a lidar com modelos com *localização/escala*, para os quais a ordenada na origem representa a *localização*, estando o declive relacionado com a *escala*.

Mais geralmente, o *papel de probabilidade* usa-se quando $F(x_{i:n}, \underline{\theta})$, com $x_{i:n}$ a i-ésima estatística ordinal (e.o.) ascendente associada à amostra $(x_1, \ldots, x_n)$, e $\underline{\theta}$ vector de parâmetros desconhecidos, pode ser transformada numa relação linear, i.e. existem funções $g_i(\cdot),\ i = 1, 2, 3, 4$ tais que

$$g_1[F(x_{i:n}, \underline{\theta})] = g_2(\underline{\theta}) + g_3(\underline{\theta})\ g_4(x_{i:n}),$$

onde $F(x_{i:n}, \underline{\theta})$, desconhecido, é substituído por uma sua estimativa plausível, como por exemplo $p_i = i/(n+1)$, sempre que $F(\cdot)$ for absolutamente contínua (Chernoff & Lieberman[9], 1956). Têm-se pois gráficos do tipo do apresentado na Figura 3.8.

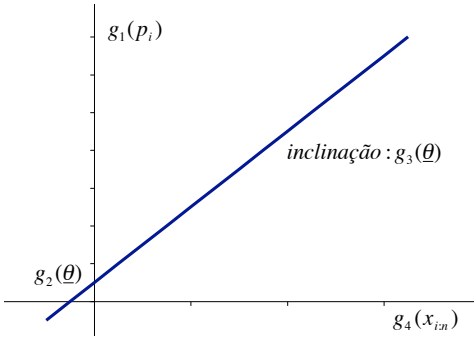


Figura 3.8: Aspecto possível de um papel de probabilidade genérico

Isto significa que o *papel de probabilidade* pode ser facilmente usado quando queremos testar informalmente a validade de uma população subjacente que, não dependendo apenas de parâmetros de localização e escala, pode ser transformada numa população com essas características, e o aspecto possível é o indicado na Figura 3.8. Vejamos alguns exemplos.

Exemplo 3.3.1. *Se estivermos interessados em validar informalmente um outro modelo muito comum em Estatística de Extremos, o modelo* Fréchet (*de*

[9] Chernoff, H. & Lieberman, G.J. (1956). The use of generalized probability paper for continuous distributions. *Ann. Math. Statist.* **27**, 806–818.

máximos) *com localização* $\lambda = 0$*, com f.d.*

$$F(x; 0, \delta, \alpha) = \exp(-(x/\delta)^{-\alpha}),\ x \geq 0,$$

temos

$$-\log(-\log(p_i)) = -\alpha \log \delta + \alpha \log x_{i:n},\ 1 \leq i \leq n,$$

i.e. um papel de probabilidade para esta população terá uma escala logarítmica (*onde marcamos* $x_{i:n}$)*, sendo a outra escala duplamente logarítmica, a escala funcional* Gumbel*.*

Observação 3.3.1. *Na realidade o logaritmo de uma v.a.* Fréchet (*de máximos*) *tem distribuição* Gumbel (*de máximos*)*.*

Exemplo 3.3.2. *Acontece algo semelhante para* X Log-normal*, com localização* $\theta = 0$*, pois então* $\log X$ *é* Normal*, tal como já foi referido. Um papel de probabilidade* Log-normal *terá uma escala logarítmica* (*onde marcamos* $x_{i:n}$)*, sendo a outra escala a escala funcional* Normal*. Na situação actual, temos simplesmente de marcar a nuvem de pontos* $\left(\log x_{i:n}, \Phi^{-1}(i/(n+1))\right)$*,* $1 \leq i \leq n$*, ou, em* `R` *usar o* `qqnorm(y)` *com* $y = \log x$*.*

Neste caso geral, e mesmo que não seja viável proceder a uma transformação simples dos dados, os QQ-plots comparam os *dados ordenados* $X_{i:n}$ com os correspondentes *quantis da distribuição a ajustar*. Seja X uma v.a. com f.d. F_θ, com θ vector de parâmetros desconhecidos. Seja $(X_1, \ldots, X_n)$ uma amostra aleatória (a.a.) associada a X. Denotemos $\hat{\theta} \equiv \hat{\theta}(X_1, \ldots, X_n)$, um estimador consistente de θ. Então, podemos conceber um QQ-plot, em que se marcam os pontos $\left(F^{\leftarrow}_{\hat{\theta}}(p_i), X_{i:n}\right)$, $i = 1, \ldots, n$. O QQ-plot e o PP-plot são muito frequentemente definidos da mesma forma. Existe no entanto quem considere no PP-plot a marcação dos pontos $\left(F_{\hat{\theta}}(X_{i:n}), p_i\right)$, $i = 1, \ldots, n$.

3.4 W-plots: caso geral $F(\cdot|\theta)$

Tal como referido anteriormente, o princípio em que se baseia o PP-plot e o QQ-plot é a identificação

$$F_\theta(X_{i:n}) \stackrel{d}{=} U_{i:n}, \qquad i = 1, \ldots, n,$$

com $U_{i:n},\ i = 1, \ldots, n$ as e.o.'s associadas a uma a.a. da $\mathcal{U}(0,1)$. Então

$$-\log(1 - F_\theta(X_{i:n})) \stackrel{d}{=} E_{i:n}, \quad i = 1, \ldots, n,$$

com $E_{i:n}$ as e.o.'s associadas a uma a.a. de dimensão n da $\mathcal{E}(1)$.

O W-plot é outra representação gráfica, que decorre dos quantis da Exponencial. Marcamos

$$\big(-\log(1 - p_i), -\log(1 - F_{\hat{\theta}}(X_{i:n}))\big) \quad i = 1, \ldots, n\,,$$

e analisamos se a nuvem de pontos está razoavelmente perto da diagonal.

3.5 Função de excesso médio e ME-plot

Na prática actuarial, o condicionamento no acontecimento $\{X > t\}$ assume a maior importância, especialmente no *Resseguro*. Considere-se um tratado *excess-of-loss* com retenção t, para qualquer indemnização da carteira. O *Ressegurador* terá de pagar um montante aleatório $X - t$, *o excesso acima de* t, mas só se $X > t$. Tendo em vista o *cálculo do prémio*, o actuário pretende estabelecer uma *franquia* ou *nível* t, o chamado *t-threshold*, da metodologia POT, procedendo ao cálculo do *montante esperado a ser pago por cliente*, quando um dado nível t é escolhido.

Por exemplo, o actuário calcula a *função de excesso médio* (em inglês, '*mean excess function*'),

$$e(t) := \mathbb{E}\left[X - t | X > t\right], \tag{3.5}$$

assumindo que $\mathbb{E}[X] < \infty$.

3.5.1 ME-plots — *mean excess plots*

Na prática, a função de *excesso médio*, $e(\cdot)$, em (3.5), é substituída pela sua contrapartida empírica, $\hat{e}_n(\cdot)$, com base na amostra de dados observados $x_1, \ldots, x_n$, e definida por

$$\hat{e}_n(t) := \frac{\sum_{i=1}^n x_i I_{(t,+\infty)}(x_i)}{\sum_{i=1}^n I_{(t,+\infty)}(x_i)} - t\,, \text{ com } I_{(t,+\infty)}(x_i) := \begin{cases} 1 & x_i > t \\ 0 & \text{caso contrário.} \end{cases}$$

Usualmente, $\hat{e}_n$ é representada nos valores

$$t = x_{n-k:n},\ k = 1, \ldots, n-1,$$

ou seja utiliza-se a chamada metodologia PORT. Tem-se então,

$$\sum_{i=1}^{n} x_i I_{(t,+\infty)}(x_i) = \sum_{j=1}^{k} x_{n-j+1:n},$$

com $k \equiv \#\ x_i : x_i > t$ e as *estimativas dos excessos médios* dadas por

$$e_{k,n} := \hat{e}_n(x_{n-k:n}) = \frac{1}{k}\sum_{j=1}^{k} x_{n-j+1:n} - x_{n-k:n}.$$

3.5.2 Padrões das funções de excesso médio

Vamos em seguida investigar os padrões das *funções de excesso médio* associados a determinados modelos, i.e.

$$e(t) = \mathbb{E}\left[X - t | X > t\right] = \frac{\int_t^{x^F} (1 - F(u))du}{1 - F(t)}, \tag{3.6}$$

com $x^F := \sup\{x : F(x) < 1\} \leq \infty$, limite superior do suporte ou '*right endpoint*' de F.

Observação 3.5.1. *Note-se que, o numerador de $e(t)$, em (3.6), é obtido pela inversão da ordem de integração* (Teorema de Fubini)

$$\begin{aligned}\int_t^{x^F} (x-t)dF(x) &= \int_t^{x^F} \left(\int_t^x du\right) dF(x) \\ &= \int_t^{x^F} \left(\int_u^{x^F} dF(x)\right) du = \int_t^{x^F} (1 - F(u))du.\end{aligned}$$

Mais uma vez a distribuição **Exponencial** desempenha um papel fulcral, devido por exemplo à ***propriedade de falta de memória da Exponencial***:

$$\begin{aligned}e(t) := \mathbb{E}[X - t | X > t] = \frac{\int_t^{x^F} (1 - F(u))du}{1 - F(t)} &= \frac{\int_t^{+\infty} e^{-\lambda u} du}{e^{-\lambda t}} \\ &= \frac{1}{\lambda}, \quad \forall t > 0,\end{aligned}$$

i.e., é *irrelevante o condicionamento em* $\{X > t\}$.

Genericamente, a ***forma*** *de* $e(\cdot)$, em (3.5) ou em (3.6), dá informação acerca de *caudas mais pesadas do que a da* Exponencial ou *mais leves que a da* Exponencial. Quando a distribuição de X tem cauda mais pesada do que a da Exponencial, a função de excesso médio $e(t)$ tende a exibir monotonia crescente para valores elevados de t. Na presença de caudas mais leves, a tendência desta função é no sentido decrescente.

3.5.3 Funções de excesso médio — modelo Weibull

O modelo Min-Weibull, ou simplesmente Weibull, tem função de sobrevivência,

$$\overline{F}(x) = 1 - F(x) = \exp(-\lambda x^{\tau}),\ x > 0,\ \tau > 0.$$

É relativamente fácil mostrar que

$$e(t) = \frac{t^{1-\tau}}{\lambda\tau}(1 + o(1))\,,$$

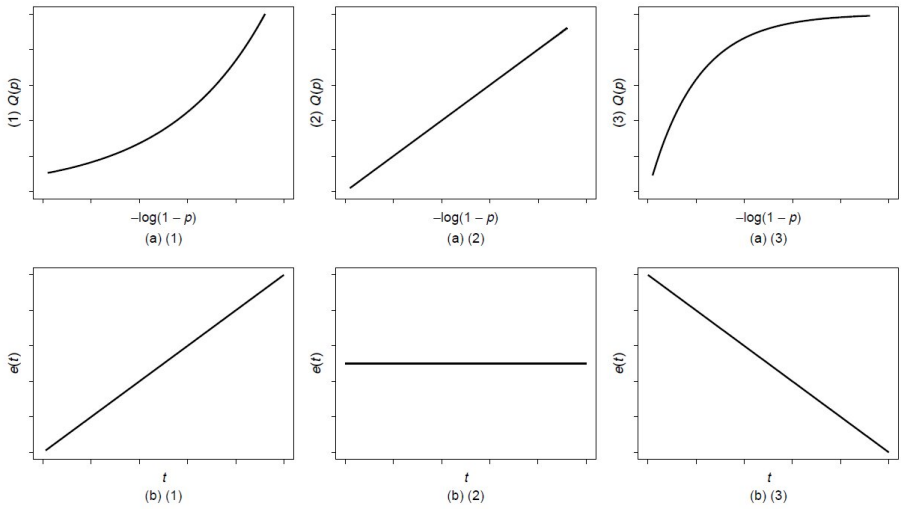
pelo que para grandes valores de t, $e(t)$ é crescente para $\tau < 1$, e decrescente para $\tau > 1$. Note-se que $\tau = 1$ **corresponde ao modelo Exponencial**, para o qual $e(t)$ é constante.

3.6 Caudas mais pesadas/leves (HTE/LTE) do que a Exponencial e caudas exponenciais

Na Figura 3.9 exibimos o aspecto dos QQ-plots e ME-plots para caudas HTE, do inglês '*heavier than exponential*', Exponencial e LTE, do inglês '*lighter than exponential*'.

3.7 Dados hidrológicos — parâmetros de interesse

O objectivo último na análise da frequência de cheias é a estimação do chamado 'T**-year flood discharge** (water level)', i.e., o nível das águas do caudal



(Beirlant *et al.*, 2004)

Figura 3.9: (a) QQ-plot Exponencial; (b) ME-plot para exemplos de distribuições com caudas tipo: (1) HTE, (2) Exponencial, e (3) LTE

do rio ultrapassado todos os T anos, em média.

Usualmente, toma-se para horizonte temporal $T = 100$, mas a estimação é levada a cabo tendo por base dados das descargas fluviais num período inferior.

Nos Países Baixos (Holanda e Bélgica), por exemplo, a exigência legal para a construção de diques exige que a sua altura é a que decorre de $T = 10^4$ para horizonte temporal de inundação, i.e., o nível ultrapassado apenas cada 10^4-anos, em média.

Outro fenómeno hidrológico em que tem interesse especial o estudo da cauda da distribuição é a intensidade da precipitação, dando atenção especial aos níveis de pluviosidade mais extremos.

3.7.1 Dados de máximos anuais

Muito frequentemente os dados disponíveis são de natureza periódica, em especial **Dados de Máximos Anuais**.

Alternativamente aos níveis T-anual (T–year water levels) a *análise de valores extremos* pode ser abordada em termos recíprocos através dos denominados **Períodos de Retorno**:

$$T(x) := \frac{1}{\mathbb{P}[Y > x]}\,,$$

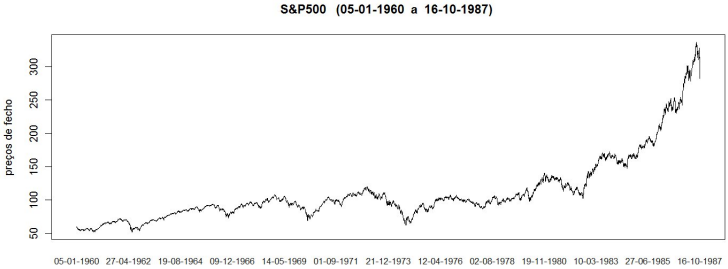
com Y a v.a. associada ao máximo periódico anual, por exemplo.

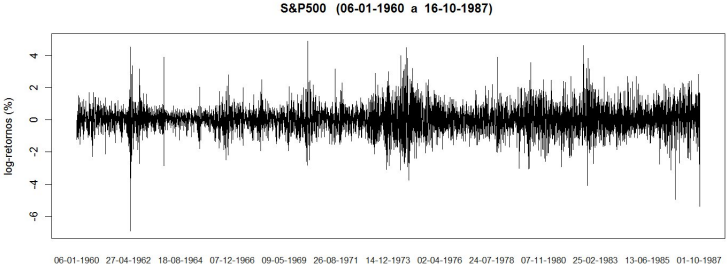
3.8 Dados financeiros

Séries temporais financeiras consistem em preços especulativos dos activos, como stocks, moeda estrangeira ou *commodities*. A gestão do risco num banco comercial destina-se a protecção contra os riscos de perda, devido a quedas nos preços dos activos financeiros, detidos ou emitidos pelo banco. Assim, as diferenças relativas de preços consecutivos, ou os 'log-retornos' são quantidades adequadas para ser investigados.

O VaR de um portfólio é essencialmente o nível abaixo do qual o portfólio futuro vai cair com apenas uma pequena probabilidade. O VaR é uma das importantes medidas de risco que têm sido utilizadas por investidores ou gestores de fundos para tentar avaliar ou prever o impacto de eventos desfavoráveis que podem ser piores do que o que foi observado durante o período para o qual estão disponíveis dados relevantes. O VaR é pois um *quantil extremal*.

Sem entrar ainda em detalhes de estimação de parâmetros, apresentamos em seguida os dados 'SP500.txt' (veja-se Beirlant *et al.*, 2004; `http://lstat.kuleuven.be/Wiley/`): Standard&Poors 500, com os preços de fecho (n=6985) e os log-retornos diários (n=6984), de 5 de Janeiro 1960 a 16 Outubro de 1987 (dia de mercado anterior ao *Big Crash Black Monday 19-Out-87*) (Beirlant et al., 2004).


S&P500 (05-01-1960 a 16-10-1987)
preços de fecho
50
100
150
200
250
300
05-01-1960
27-04-1962
19-08-1964
09-12-1966
14-05-1969
01-09-1971
21-12-1973
12-04-1976
02-08-1978
19-11-1980
10-03-1983
27-06-1985
16-10-1987


S&P500 (06-01-1960 a 16-10-1987)
log-retornos (%)
-6
-4
-2
0
2
4
06-01-1960
27-04-1962
18-08-1964
07-12-1966
09-05-1969
26-08-1971
14-12-1973
02-04-1976
24-07-1978
07-11-1980
25-02-1983
13-06-1985
01-10-1987

Capítulo 4

Análise Preliminar de Valores Extremos — O Porquê da EVT

4.1 Problemas simples em valores extremos

Consideremos que os dados em análise são observações de uma amostra $(X_1, \ldots, X_n)$ de v.a.'s independentes e identicamente distribuídas (i.i.d.) com f.d. comum $F(.)$. Denotemos as correspondentes e.o.'s associadas à amostra ordenada por $(X_{1:n} \leq \cdots \leq X_{n:n})$, onde $X_{1:n}$ denota o mínimo, $X_{n:n}$ o máximo, sendo $X_{i:n}$ a i-ésima e.o. ascendente, $1 \leq i \leq n$. Os dados amostrais são usados tipicamente para estudar as propriedades da f.d.

$$F(x) := \mathbb{P}(X \leq x), \tag{4.1}$$

ou da sua inversa, a *função quantil*

$$Q(p) := \inf\{x : F(x) \geq p\} =: F^{\leftarrow}(p),$$

onde $F^{\leftarrow}$ denota a inversa generalizada de F, já definida em (3.2).

4.1.1 Escassez de dados nas caudas

Suponhamos que estamos interessados em estimar as caudas de F. A Figura 4.1 ilustra as dificuldades em estimar essas caudas de forma precisa, uma vez que a maioria dos dados está concentrada no centro da distribuição. As observações nas caudas são escassas, e frequentemente é requerida a estimação para além do máximo e/ou do mínimo amostrais.

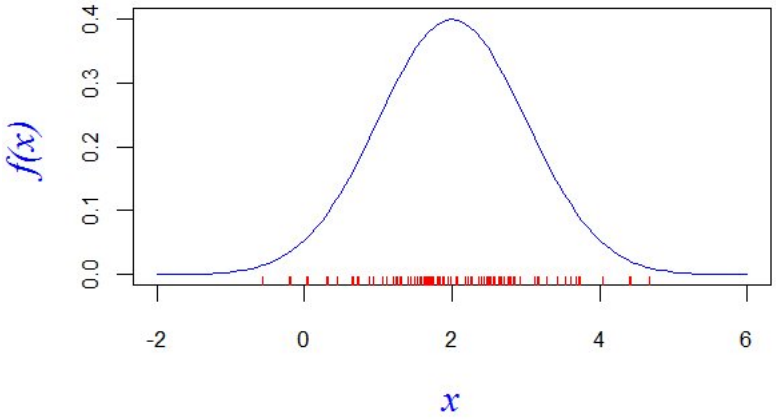


Figura 4.1: Escassez das observações na cauda

4.1.2 Metodologias tradicionais inadequadas

Poderíamos ajustar um modelo probabilístico, por exemplo o modelo Normal, $\mathcal{N}(\mu, \sigma^2)$, a toda a amostra disponível, e usar esse modelo para estimar as probabilidades de cauda, ou seja, considerar a estimativa

$$\widehat{\mathbb{P}}(X > x) = 1 - \widehat{F}(x) = 1 - \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right).$$

Os problemas associados são essencialmente os seguintes:

- O ajustamento do modelo é essencialmente ditado pelas observações centrais.
- Modelos diferentes que ajustam bem o corpo dos dados podem conduzir a diferentes extrapolações na caudas.
- Por outro lado, se o interesse reside nas caudas,

– porquê comprometer o ajustamento nas pontas ao forçar simultaneamente que o modelo ajuste bem a parte central dos dados?

Em EVT pretendemos frequentemente estimar a probabilidade de excedência

$$p \equiv p_x := \mathbb{P}[X > x], \quad \text{onde} \quad x(> x_{n:n}) \tag{4.2}$$

é pois um valor elevado. A *função de distribuição empírica* (f.d.e.) definida por

$$\widehat{F}_n(x) = \begin{cases} 0 & \text{se} \quad x < x_{1:n}, \\ i/n & \text{se} \quad x_{i:n} \leq x < x_{i+1:n},\ 1 \leq i < n, \\ 1 & \text{se} \quad x \geq x_{n:n}, \end{cases} \tag{4.3}$$

onde $x_{i:n}$ é o *i-ésimo* valor amostral, $1 \leq i \leq n$, não gera pois qualquer informação útil. Inversamente, em termos da função de quantil empírica,

$$\widehat{Q}_n(p) := \inf\{x : \widehat{F}_n(x) \geq p\},$$

surgem problemas quando consideramos quantis extremais

$$\widehat{Q}_n(1-p), \quad \text{com} \quad p < 1/n\,.$$

Realmente quando o problema alvo é a estimação de quantis extremais ou pequenas probabilidades de cauda torna-se necessário providenciar técnicas especialmente dirigidas a valores extremos, em consonância com a respectiva EVT. Na análise prática de acontecimentos raros estas questões são de interesse fulcral.

4.2 Um primeiro exemplo: velocidade máxima do vento em Albuquerque

Suponhamos que estamos interessados na base de dados 'albuq.txt', de dimensão $n = 6939$, acessível em Beirlant *et al.* (2004), `http://lstat.kuleuven.be/Wiley/`, cuja caixa-com-bigodes é apresentada na Figura 4.2.

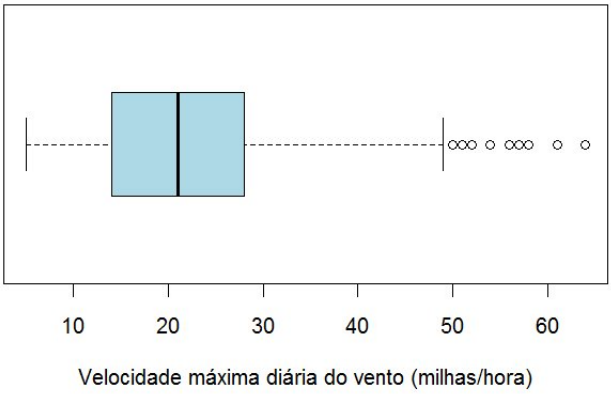


Figura 4.2: Caixa-com-bigodes associada aos dados 'albuq.txt'

Na teoria clássica, estamos muitas vezes interessados no comportamento típico da característica em estudo, descrito através do valor médio da distribuição, denotado $\mathbb{E}[X]$. Apoiados pela *Lei Fraca dos Grandes Números* (LFGN), usamos a média empírica $\overline{X}$ como estimador consistente de $\mathbb{E}[X]$. Além disso, o TLC caracteriza o comportamento assintoticamente Normal da média amostral, caso existam segundos momentos finitos. Este resultado pode ser usado para fornecer um intervalo de confiança para $\mathbb{E}(X)$ no caso da dimensão da amostra ser suficientemente elevada, uma condição necessária quando invocamos o TLC. Para os dados da velocidade de vento em Albuquerque, estas técnicas conduzem a um valor máximo diário médio da velocidade do vento de 21.658 *milhas/h*, constituindo $(21.4369, 21.8797)$ um intervalo de confiança a 95% para $\mathbb{E}(X)$.

Mas no caso de velocidades de vento, pode ser mais importante estimar probabilidades de cauda. Suponhamos que uma estrutura se quebra se a velocidade do vento for superior a $30\,Km/hora$ ... então o valor de interesse é a estimativa da probabilidade de cauda

$$p = \mathbb{P}[X > 30].$$

Para o efeito, pode-se usar a f.d.e., definida em (4.3). Para os dados 'albuq.txt, a f.d.e. associada está representada na Figura 4.3, e tal procedimento conduz-

nos ao valor de

$$\hat{p} = 1 - \widehat{F}_n(30) = 0.18\,.$$

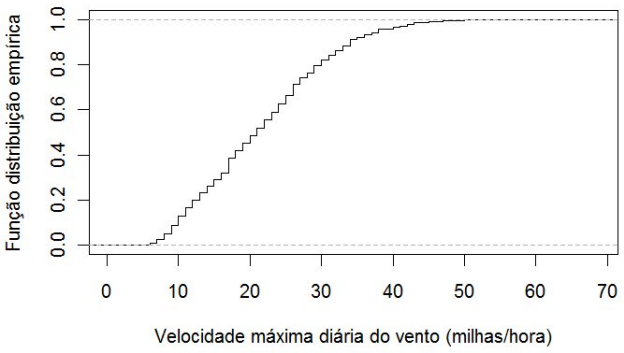


Figura 4.3: F.d.e. associada aos dados 'albuq.txt'

No entanto ...se o segundo momento $\mathbb{E}[X^2]$ ou até mesmo a média $\mathbb{E}[X]$ não são finitos, então o TLC não se aplica e a teoria clássica, dominada pela distribuição Normal, já não é pertinente. Ou seja, se se quer estimar p_x, em (4.2), a estimativa $\hat{p}$, definida à custa da f.d.e. produz o valor 0 ...

Este tipo de questões associadas a acontecimentos raros são muito importantes, uma vez que os danos causados por velocidades extremas de vento podem revelar-se catastróficos. Claramente, nós não podemos simplesmente assumir que esses valores de x são impossíveis, pois tal é totalmente irrealista.

4.3 Um segundo exemplo: velocidade máxima do vento em Zaventem

Procedemos em seguida à aplicação de alguns dos métodos do Capítulo 3 aos dados 'zaventem.txt', velocidade máxima diária do vento em Zaventem (1985-1992), de dimensão n=3225 (Beirlant *et al.*, 2004, `http://lstat.kuleuven.be/Wiley/`). Temos 74 Velocidades acima de $t = 82Km/h$, a que estão associados os histogramas apresentados na Figura 4.4.

A função densidade de probabilidade (f.d.p.) condicional ajustada a estas 74

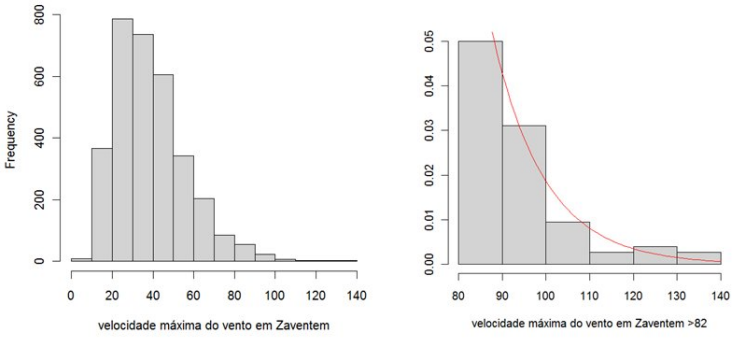


Figura 4.4: Histograma associado à amostra de velocidades máximas diárias do vento em Zaventem (*esquerda*) e à sub-amostra das velocidades acima de 82 Km/h (*direita*)

velocidades é

$$f(x) = \hat{\lambda}\exp(-\hat{\lambda}(x-t)),\ \text{com } t = 82, \quad \hat{\lambda} = \frac{1}{\overline{x} - t} = 0.0829\,.$$

Utilizando o `R` para ajustar a recta

$$\widehat{Q}_n(p) = t - \frac{1}{\hat{\lambda}}\log(1-p),$$

obtemos através do QQ-plot apresentado na Figura 4.5, o valor 80.86 para ordenada na origem, valor que difere um pouco do valor exacto $t = 82$.

Considerando valor exacto de $t = 82$ para ordenada na origem da recta

$$\widehat{Q}_n(p) = t - \frac{1}{\hat{\lambda}}\log(1-p),$$

a recta (declive=a; ordenada na origem=t=82) ajustada à nuvem de pontos pelo *método dos mínimos-quadrados*, é obtida pela minimização de

$$\sum_{i=1}^{n} \left(x_{i:n} - t + a\,\log(1-p_i)\right)^2,$$

vindo

$$\hat{a} = \frac{\sum_{i=1}^{n} x_{i:n} q_i - t\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} q_i^2}\,,\ \text{com } q_i := -\log(1-p_i), \quad i = 1,\ldots,n\,.$$

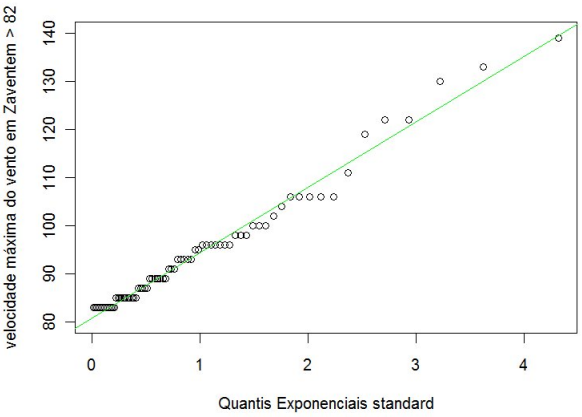
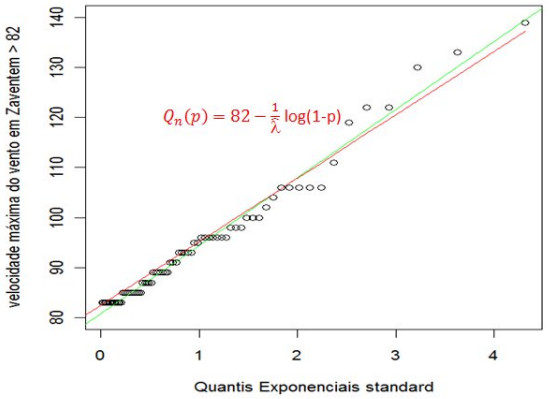


Figura 4.5: QQ-plot para as velocidades de vento acima de 82 Km/h

Para os dados 'zaventem.txt',

$$\hat{a} = 12.96368 = \frac{1}{\hat{\lambda}}, \quad \hat{\lambda} = 0.07713861.$$



Temos então o coeficiente de correlação amostral,

$$r_Q = \frac{\sum_{i=1}^{n}(x_{i:n} - \overline{x})(q_i - \overline{q})}{\sqrt{\sum_{i=1}^{n}(x_{i:n} - \overline{x})^2 \sum_{i=1}^{n}(q_i - \overline{q})^2}},$$

com

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_{i:n}, \quad \overline{q} = \frac{1}{n}\sum_{i=1}^{n} q_i \qquad \text{e} \qquad q_i := -\log(1-p_i) = -\log(1-i/(n+1)).$$

Note-se que:

- $0 \leq r_Q \leq 1$, pois os valores estão ordenados por ordem crescente.
- $r_Q = 1$ se e só se os pontos estão todos sobre a recta.
- r_Q pode ser usado como medida de ajustamento do modelo Exponencial aos dados.
- Um teste baseado na estatística r_Q rejeita a hipótese de Exponencialidade quando r_Q difere significativamente de 1.

Para os dados em 'zaventem.txt', consideraremos dois gráficos distintos, na Figuras 4.6, onde representamos ME-plots em que $e_{k,n}$ é representado versus k e versus $x_{n-k:n}$, respectivamente.

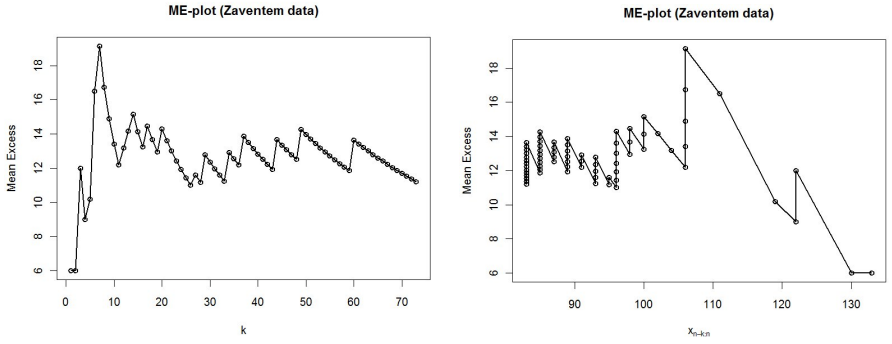


Figura 4.6: ME-plot, $e_{k,n}$ versus k (*esquerda*) e $e_{k,n}$ versus $x_{n-k:n}$ (*direita*) das velocidade máxima diária do vento ($> 82\,km/h$) em Zaventem

O padrão 'aproximadamente' constante dá-nos a indicação de cauda Exponencial.

4.4 Um terceiro exemplo: dados de seguros de incêndios

Consideremos em seguida uma aplicação aos dados 'norwegianfire.txt', relativos a seguro de incêndios, ou mais especificamente a valores de indemnizações acima das 500($\times$1000 *Coroa*), ano, no período 1972-1992, com dimensão n=9181 (Beirlant *et al.*, 2004, `http://lstat.kuleuven.be/Wiley/`). Os dados e o histograma associado estão representados na Figura 4.7.

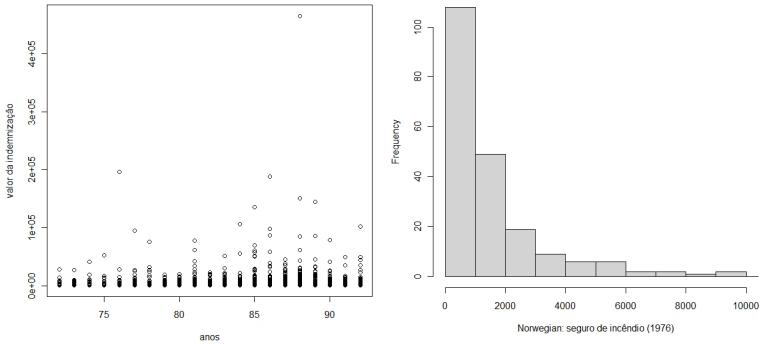


Figura 4.7: Indemnizações (*esquerda*) e histograma associado (*direita*)

De forma a ter acesso às propriedades distribucionais para os dados foram construídos os QQ-plots Exponencial e Pareto, apresentados na Figura 4.8.

Note-se que o QQ-plot Exponencial *curva para cima e tem um padrão convexo*, o que leva a considerar que a distribuição subjacente aos dados das indemnizações tem *uma cauda direita mais pesada do que o esperado com a* Exponencial.

À excepção dos últimos pontos, o QQ-plot Pareto é mais ou menos linear, indicando um *ajustamento razoável da distribuição Pareto à cauda* dos valores de indemnização.

De forma semelhante ao que fizemos na Secção 4.3 para os dados em 'zaventem.txt', passamos em seguida aos ME-plots associados aos dados 'norwegi-

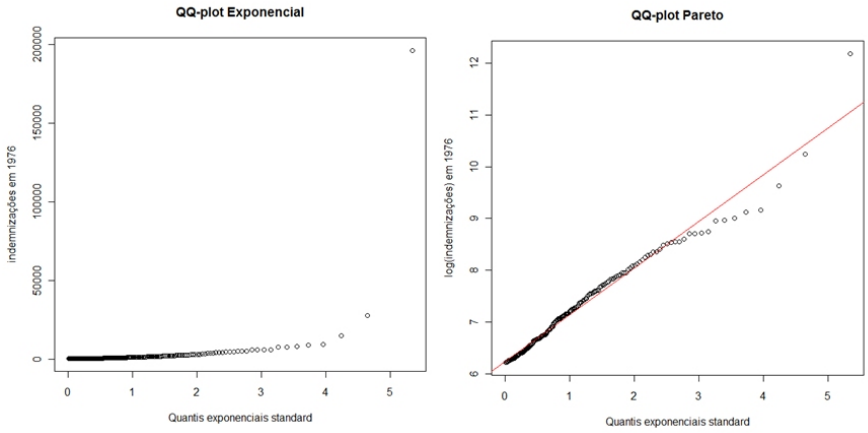


Figura 4.8: QQ-plot Exponencial (*esquerda*) e Pareto (*direita*) para os dados 'Norwegian'

anfire.txt'. A Figura 4.9 é pois semelhante à Figura 4.6, mas para esta nova base de dados.

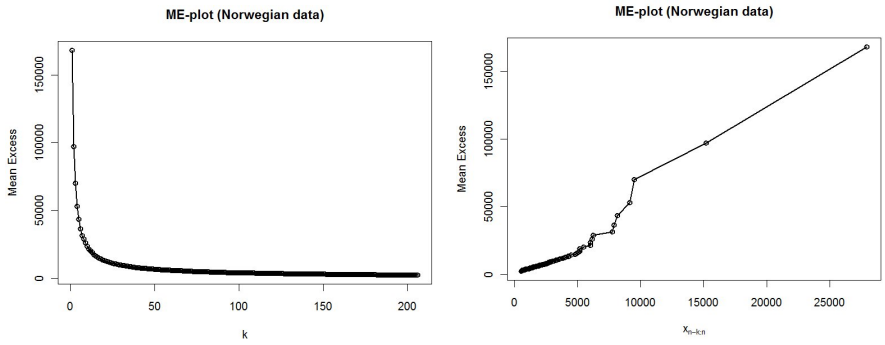


Figura 4.9: **Dados:** 'Norwegian' seguro de incêndio: indemnização > 500 ($\times 1000\, Coroa$), 1976. ME-plot, $e_{k,n}$ versus k (*esquerda*), e $e_{k,n}$ versus $x_{n-k,n}$ (*direita*)

Neste caso, detecta-se uma cauda mais pesada do que a Exponencial (possivelmente Pareto).

4.5 Um último exemplo: descargas máximas anuais do rio Meuse

Passamos à análise dos dados 'maasmax.txt', relativos a descargas anuais máximas do rio Meuse, em Borgharen (NL). As variáveis a que temos acesso são as descarga máximas anuais, no período 1911-1995, num total de n=85 observações (Beirlant *et al.*, 2004, `http://lstat.kuleuven.be/Wiley/`). Esses dados são apresentados na Figura 4.10, onde também apresentamos o histograma associado.

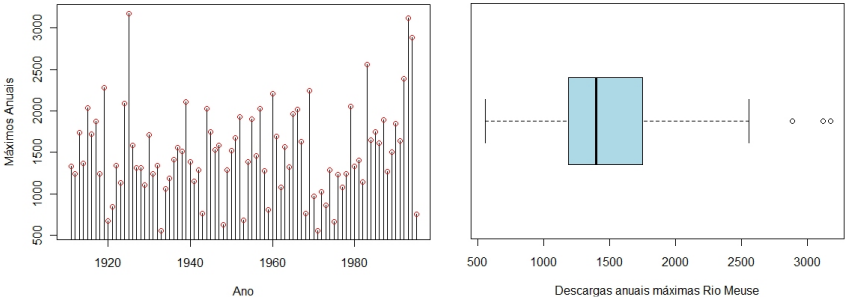


Figura 4.10: Máximos anuais do rio Meuse em 1911-1995 (*esquerda*) e histograma associado (*direita*)

Embora se verifique uma assimetria à direita, vejamos o que se passaria se se seguisse uma análise estatística tradicional, fazendo um ajustamento do *máximo anual* ao modelo Normal, para parâmetros convenientes. Na Figura 4.11, apresentamos o QQ-plot Normal associado a estes dados.

Ajustando uma recta de mínimos quadrados, obtem-se para $\hat{\mu} = 1495.962$ e $\hat{\sigma} = 551.0057$, a que corresponde um coeficiente de correlação amostral de $r_Q = 0.9788504$. Sobrepondo ao histograma dos dados (máximos anuais) a Normal estimada $Y \frown \mathcal{N}(\hat{\mu} = 1495.962, \hat{\sigma} = 551.0057)$, obtém-se o resultado apresentado na Figura 4.12.

Dada a natureza dos dados, a importância do modelo Gumbel em EVT, como teremos oportunidade de ver mais adiante, e ainda ao facto de se verificar uma assimetria à direita, vejamos o que se passaria se se considerasse um

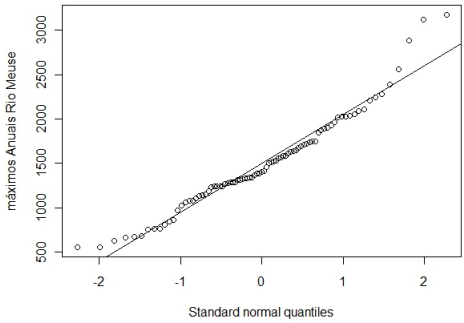


Figura 4.11: QQ-plot Normal: $\{(\Phi^{\leftarrow}(i/(n+1)), y_{i:n}) : i = 1, \ldots, n\}$

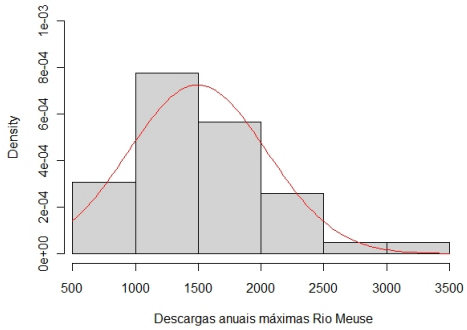


Figura 4.12: Histograma e Normal ajustada

ajustamento do *máximo anual* ao modelo Gumbel, para parâmetros convenientes. Representamos na Figura 4.13 o QQ-plot Gumbel associado aos dados. Ajustando uma recta de mínimos quadrados, obtem-se para parâmetros de localização e escala, repectivamente, $\hat{b} = 1247.363$ e $\hat{a} = 445.6884$, a que corresponde um coeficiente de correlação amostral $r_Q = 0.9924606$ (superior ao encontrado no caso do ajustamento à Normal).

Sobrepondo ao histograma dos dados (máximos anuais) a f.d.p. Gumbel estimada, $Y \frown \Lambda(\hat{b} = 1247.363, \hat{a} = 445.6884)$, obtem-se a Figura 4.14, onde o ajustamento parece bem mais fidedigno do que o obtido na Figura 4.12.

De forma a ter uma ideia do comportamento de cauda da distribuição do Má-

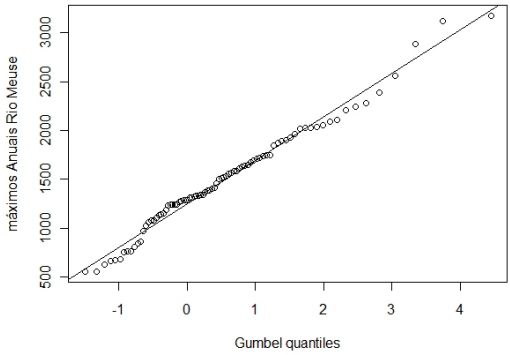


Figura 4.13: QQ-plot Gumbel: $\{(\Lambda^{\leftarrow}(i/(m+1)), y_{i:m}) : i = 1, \ldots, m\}$

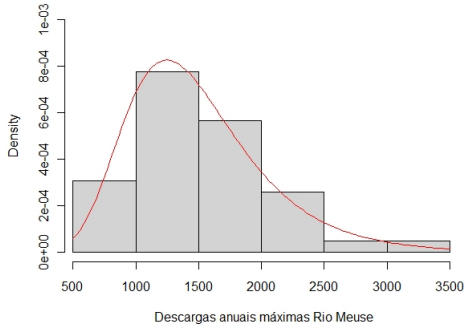


Figura 4.14: Histograma e Gumbel ajustada

ximo Anual, construimos ainda o QQ-plot Exponencial, apresentado na Figura 4.15, à esquerda. Claramente a cauda da distribuição do Máximo Anual não será mais pesada do que a Exponencial. Esta conclusão é confirmada ainda pelo ME-plot apresentado na mesma figura, à direita.

Em Beirlant *et al.* (2004), Coles (2001) e Castillo *et al.* (2004) são tratados outros casos de estudo, associados a bases de dados, num leque variado de áreas de aplicação de Modelação de Acontecimentos Raros (MAR).

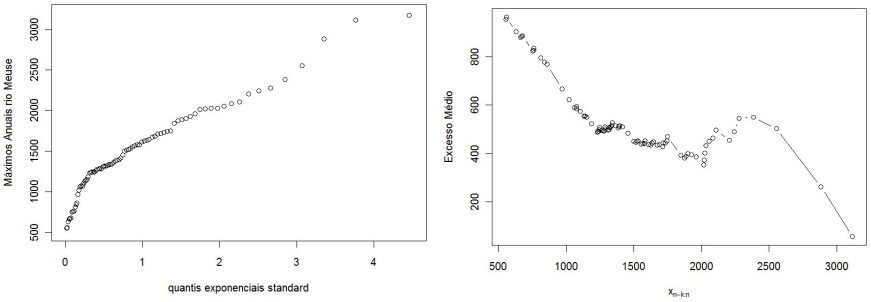


Figura 4.15: QQ-plot Exponencial standard (*esquerda*) e ME-plot (*direita*) para os máximos anuais do rio Meuse

Capítulo 5

Estatísticas Ordinais – Teoria Distribucional Exacta

Associada à amostra aleatória $(Z_1, \ldots, Z_n)$, proveniente de uma população com f.d. F, consideremos a amostra das e.o.'s ascendentes $(Z_{1:n} \leq \cdots \leq Z_{n:n})$. Ao leitor mais interessado no estudo elementar da teoria distribucional de e.o.'s recomenda-se Arnold *et al.* (1992). Para estudos mais avançados, pode-se consultar David & Nagaraja (2003) e Reiss (1989).

5.1 Comportamento distribucional de uma estatística ordinal

Teorema 5.1.1 (f.d. e f.d.p. exactas de $Z_{k:n}$)**.** *Sejam $Z_1, \ldots, Z_n$ i.i.d. a Z com f.d. $F(\cdot)$ e f.d.p. $f(\cdot)$. Então,*

$$\begin{aligned} F_{k:n}(z) &:= \mathbb{P}[Z_{k:n} \leq z] \\ &= \mathbb{P}[\textit{pelo menos } k \textit{ das } n \textit{ v.a's } Z_i \leq z] \\ &= \sum_{i=k}^{n} \binom{n}{i} F^i(z)[1 - F(z)]^{n-i} \end{aligned}$$

e

$$f_{k:n}(z) = \frac{n!}{(k-1)!(n-k)!} F(z)^{k-1}[1-F(z)]^{n-k} f(z)\,.$$

Se Z for absolutamente contínua, com f.d.p. $f(z) = F'(z)$, $Z_{i:n}, 1 \leq i \leq n$, tem f.d.p.

$$f_{i:n}(z) = \frac{1}{B(i, n-i+1)} F^{i-1}(z)[1-F(z)]^{n-i} f(z), \quad z \in \mathbb{R}, \tag{5.1}$$

onde $B(p,q)$ denota a função Beta completa, em (3.3). Na realidade, como é possível ter $Z_{i:n} \in (z, z+dz]$?

- Dos n $ZZ's$, $(i-1)$ arbitrários têm de ser $\leq z$ — acontecimento com probabilidade $\binom{n}{i-1} F^{i-1}(z)$.
- Dos $(n-i+1)$ restantes, um arbitrariamente deve pertencer ao intervalo $(z, z+dz]$ — acontecimento a que está associada uma probabilidade dada por $(n-i+1)(F(z+dz) - F(z))$.
- Os restantes $(n-i)$ elementos da amostra têm de ser superiores a $z+dz$ — acontecimento com probabilidade $[1-F(z+dz)]^{n-i}$.

A independência dos acontecimentos anteriores, juntamente com o facto de se ter

$$f_{i:n}(z) = \lim_{dz \to 0} \frac{\mathbb{P}\left(Z_{i:n} \in (z, z+dz]\right)}{dz},$$

e de $f(z) = \frac{d}{dz} F(z) = \lim_{dz \to 0} \frac{F(z+dz) - F(z)}{dz}$, conduz-nos de imediato à expressão (5.1).

5.1.1 Relação com os modelos Binomial e Beta

Por definição de e.o., $Z_{i:n}$ é menor ou igual a z se e só se existirem pelo menos i observações, de entre as n, que sejam menores ou iguais a z. Consequentemente, e tal como vimos no Teorem 5.1.1, para modelos $F(\cdot)$ contínuos, a f.d. de $Z_{i:n}, 1 \leq i \leq n$, é, para F discreta ou contínua,

$$F_{i:n}(z) = \sum_{k=i}^{n} \binom{n}{k} F^k(z)(1-F(z))^{n-k}, \quad 1 \leq i \leq n. \tag{5.2}$$

Para melhor compreendermos a fórmula (5.2), atentemos na seguinte relação entre a distribuição de uma e.o. e o modelo Binomial. Consideremos as seguintes variáveis de contagem:

$$S_z := \#\{i : Z_i > z,\ 1 \leq i \leq n\} = \sum_{i=1}^{n} I_{[Z_i > z]},$$

$$S_z^* := \#\{i : Z_i \leq z,\ 1 \leq i \leq n\} = \sum_{i=1}^{n} I_{[Z_i \leq z]},$$

onde I_A é a função indicatriz do acontecimento A, i.e.

$$I_A = \begin{cases} 1 & \text{se A ocorre} \\ 0 & \text{caso contrário.} \end{cases}$$

Tem-se obviamente $S_z + S_z^* = n$, a v.a. S_z é Binomial$(n, 1 - F(z))$ e a v.a. S_z^* é também Binomial$(n, F(z))$. Como

$$Z_{i:n} \leq z \quad \text{se e só se} \quad S_z < n - i + 1 \quad \text{se e só se} \quad S_z^* \geq i$$

tem-se,

$$\begin{aligned} F_{i:n}(z) &= \mathbb{P}[\text{Binomial}(n, 1 - F(z)) < n - i + 1] && (5.3) \\ &= \mathbb{P}[\text{Binomial}(n, F(z)) \geq i]. && (5.4) \end{aligned}$$

Da igualdade (5.4) segue (5.2), de forma imediata.

A utilização da relação (5.3) conduzir-nos-ia à expressão

$$F_{i:n}(z) = \sum_{k=0}^{n-i} \binom{n}{k} (1 - F(z))^k \ F^{n-k}(z),$$

que também podia ter sido obtida a partir de (5.2) por utilização de uma mudança no índice de soma, passando a somar em $k_1 = n - k$.

Note-se ainda que o somatório em (5.2) se pode escrever como

$$F_{i:n}(z) = \frac{1}{B(i, n - i + 1)} \int_0^{F(z)} t^{i-1}(1 - t)^{n-i} dt,$$

que é a função Beta incompleta com parâmetros i e $n - i + 1$, calculada em $F(z)$, frequentemente denotada por $\mathcal{I}_{F(z)}(i, n - i + 1)$. Tal função encontra-se extensivamente tabelada na Tabela 6 de Pearson & Hartley[1] (1970).

[1] Pearson, E.S. & Hartley, H.O. (1970). *Biometrika Tables for Statisticians.* Cambridge Univ. Press.

Denotando $B_{p,q}$ uma v.a. Beta de parâmetros p e q, temos pois

$$F_{i:n}(z) = \mathbb{P}\left[B_{i,n-i+1} \leq F(z)\right]. \tag{5.5}$$

É por esta razão que a f.d. de uma e.o. (e frequentemente a própria e.o.) é usualmente designada Beta transformada. Tem-se na realidade $F(Z_{i:n}) \stackrel{d}{=} B_{i,n-i+1}$.

Para F discreta, com suporte $\mathcal{S}$ constituído por inteiros consecutivos,

$$\begin{aligned} f_{i:n}(z) &= \mathbb{P}[Z_{i:n} = z] = F_{i:n}(z) - F_{i:n}(z-1) \\ &= \mathcal{I}_{F(z)}(i, n-i+1) - \mathcal{I}_{F(z-1)}(i, n-i+1),\ z \in \mathcal{S}. \end{aligned}$$

Casos particulares importantes de e.o.'s são o *mínimo*, para $i = 1$, em (5.2), e o *máximo*, para $i = n$, em (5.2). Tem-se então

$$F_{1:n}(z) = F_{Z_{1:n}}(z) = 1 - (1 - F(z))^n \quad \text{e} \quad F_{n:n}(z) = F_{Z_{n:n}}(z) = F^n(z).$$

Exemplo de aplicação. Em Fiabilidade é usual trabalhar com estruturas cujo tempo de vida é função exclusiva dos tempos de vida das suas n componentes. O número n de componentes é usualmente designado por *ordem* da estrutura, e uma estrutura *coerente*[2] pode sempre ser simultaneamente decomposta como uma estrutura em série de componentes em paralelo, e como uma estrutura em paralelo de componentes em série. As estruturas fundamentais são pois as estruturas em série, que descrevemos esquematicamente na Figura 5.1, e as estruturas em paralelo, descritas esquematicamente na Figura 5.2.

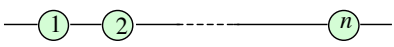


Figura 5.1: Esquema de um circuito em série com n componentes

Uma *estrutura em série* funciona se e só se funcionarem todas as suas componentes e uma *estrutura em paralelo* funciona se e só se funcionar pelo menos uma das suas componentes.

[2] A formalização do conceito de estrutura coerente pode ser, por exemplo, encontrada no livro de Barlow R.E. & Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart & Winston

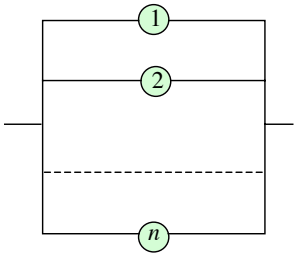


Figura 5.2: Esquema de um circuito em paralelo com n componentes

Note-se que, se designarmos por T_i o tempos de vida da componente i, $1 \leq i \leq n$, o tempo de vida $T_{_S}$ do circuito em série é $T_{_S} = \min\limits_{1\leq i\leq n} T_i$ e o tempo de vida do circuito em paralelo é $T_{_P} = \max\limits_{1\leq i\leq n} T_i$. Tem-se pois, caso as componentes sejam indepedentes e T_i tenha modelo de vida F_i, $1 \leq i \leq n$,

$$F_{T_{_S}}(t) = 1 - \prod_{i=1}^{n}(1 - F_i(t)), \qquad F_{T_{_P}}(t) = \prod_{i=1}^{n} F_i(t).$$

Estruturas i-de-n são também frequentes em fiabilidade. São estruturas que funcionam se e só se funcionarem pelo menos i das suas componentes, para $1 \leq i \leq n$. Então $T = T_{n-i+1:n}$, e consequentemente, se as componentes forem identicamente distribuídas (i.d.), $F_i \equiv F,\ 1 \leq i \leq n$,

$$\begin{aligned} F_T(t) = F_{n-i+1:n}(t) &= \sum_{k=n-i+1}^{n} \binom{n}{k} F^k(t)\,(1 - F(t))^{n-k} \\ &= \frac{1}{B(i, n-i+1)} \int_0^{F(t)} u^{n-i}(1-u)^{i-1} du. \end{aligned}$$

A situação de componentes não i.d. torna-se formalmente mais complexa nesta situação geral, mas pode perfeitamente ser explicitada.

Se as componentes constituintes dos circuitos forem dependentes — situação frequente na prática — as expressões das funções de distribuição (ou funções densidade de probabilidade dos tempos de vida destes circuitos elementares, que continuam a ser e.o.'s, variam consoante variar a estrutura de dependência.

5.2 Distribuição conjunta de duas ou mais estatísticas ordinais

Se nos colocarmos mais uma vez no contexto de Z absolutamente contínua, o par de v.a.'s $(Z_{i:n}, Z_{j:n}),\ 1 \leq i < j \leq n$, tem f.d.p. conjunta dada por

$$f_{i,j:n}(z_1, z_2) = \frac{1}{B(i, j-i)B(j, n-j+1)} F^{i-1}(z_1)[F(z_2) - F(z_1)]^{j-i-1} \\ [1 - F(z_2)]^{n-j} f(z_1) f(z_2), \qquad z_1 < z_2.$$

Generalizações para um número qualquer de e.o.'s são óbvias:

O vector de v.a.'s $(Z_{n_1:n}, Z_{n_2:n}, \ldots, Z_{n_k:n})$, $1 \leq n_1 < n_2 < ... < n_k \leq n$, terá f.d.p. dada por

$$f_{n_1,n_2,...,n_k:n}(z_1, z_2, ..., z_k) = n! \prod_{j=0}^{k} \frac{[F(z_{j+1}) - F(z_j)]^{n_{j+1} - n_j - 1}}{(n_{j+1} - n_j - 1)!} \prod_{j=1}^{k} f(z_j) \\ \text{se } z_1 < z_2 < . \cdots < z_k, \quad (5.6)$$

com a notação $z_0 = -\infty, z_{k+1} = +\infty, n_0 = 0,\ n_{k+1} = n + 1$.

Para f.d.p. conjunta de todas as e.o.'s, $(Z_{1:n} \leq \cdots \leq Z_{n:n})$, tem-se

$$f_{1,\ldots,n:n}(z_1, \ldots, z_n) = n! \prod_{j=1}^{n} f(z_j), \quad \text{se} \quad z_1 < \cdots < z_n, \qquad (5.7)$$

caso particular de (5.6) com $k = n$, e $n_j = j,\ 1 \leq j \leq n$.

A f.d. conjunta, por exemplo de duas e.o.'s, pode também ser expressa à custa de v.a.'s Binomiais dependentes, e obtem-se, para F discreta ou contínua,

$$\begin{aligned} F_{i,j:n}(x, y) &= \mathbb{P}\,[\text{pelo menos } i \text{ dos } ZZ\text{'s serem} \leq x, \\ &\qquad \text{e pelo menos } j \text{ dos } ZZ\text{'s serem} \leq y] \\ &= \begin{cases} \mathbb{P}\,[\text{termos exactamente } k\ ZZ\text{'s} \leq y, k \geq j, \text{ e} \\ \qquad \text{exactamente } l\ ZZ\text{'s} \leq x, i \leq l \leq k] & \text{se} \quad x < y \\ \mathbb{P}\,[\text{pelo menos } j \text{ dos } ZZ\text{'s serem} \leq y] & \text{se} \quad x \geq y \end{cases} \\ &= \begin{cases} \sum_{k=j}^{n} \sum_{l=i}^{k} \frac{n!}{l!(k-l)!(n-k)!} \\ \quad F^l(x)[F(y) - F(x)]^{k-l}[1 - F(y)]^{n-k} & \text{se} \quad x < y \\ \\ F_{j:n}(y) & \text{se} \quad x \geq y. \end{cases} \end{aligned}$$

Este argumento directo é também válido no caso discreto, e para $(x, y) \in \mathcal{S}$, tem-se

$$\begin{aligned} f_{ij:n}(x,y) &= \mathbb{P}[Z_{i:n} = x, Z_{j:n} = y] \\ &= F_{i,j:n}(x,y) - F_{i,j:n}(x-1,y) - F_{i,j:n}(x,y-1) \\ &\qquad + F_{i,j:n}(x-1,y-1). \end{aligned}$$

5.2.1 Estatísticas ordinais em modelo Uniforme

Designemos genericamente por U_i, $1 \leq i \leq n$, uma amostra aleatória de dimensão n proveniente de modelo Uniforme em (0,1), $\mathcal{U}(0,1)$, com f.d. $F(z) = z$, para $0 \leq z \leq 1$. Denotemos $U_{1:n} \leq \cdots \leq U_{n:n}$ a amostra das e.o.'s ascendentes associadas.

Da relação (5.5) segue imediatamente que $U_{i:n} \stackrel{d}{=} B_{i,n-i+1}$, i.e.

$$f_{U_{i:n}}(z) = \frac{1}{B(i, n-i+1)}\, z^{i-1}\, (1-z)^{n-i}, \quad 0 \leq z \leq 1.$$

Apresentamos em seguida, na Figura 5.3, para ilustração do comportamento das e.o.'s em modelo Uniforme, as f.d.p. e as f.d. de U, $U_{1:3}$, $U_{2:3}$ e $U_{3:3}$.

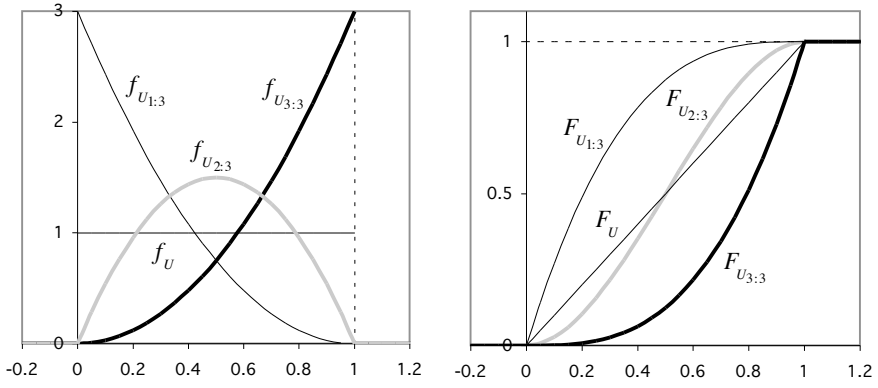


Figura 5.3: Funções densidade de probabilidade (*esquerda*) e funções de distribuição (*direita*) de uma v.a. $\mathcal{U}(0,1)$, U, de $U_{1:3}$, de $U_{2:3}$ e de $U_{3:3}$

Observação 5.2.1. *Note-se o comportamento simétrico da e.o., $U_{2:3}$, e as caudas semelhantes e 'leves', devido ao suporte ser limitado à esquerda e à direita, com $F(x)$ e $1-F(x)$ a aproximarem-se de 0 (quando $x \to 0$) e de 1 (quando $x \to 1$), respectivamente, de modo semelhante, o que irá fornecer comportamentos, simetrizados, mas do mesmo tipo, para o mínimo e para o máximo.*

Propriedades interessantes das e.o.'s em modelo Uniforme são a da independência entre quocientes de tais e.o.'s e a 'resistência' do modelo Beta face a operações diversas efectuadas às e.o.'s em modelo Uniforme. Tem-se o seguinte resultado.

Teorema 5.2.1. *Em modelo* Uniforme,

$$\left(U_{i:n}/U_{j:n},\ U_{j:n}\right),\ 1 \leq i < j \leq n,$$

é um par de v.a.'s independentes e provenientes de modelos Beta, *com*

$$U_{i:n}/U_{j:n} \overset{d}{=} B_{i,j-i} \overset{d}{=} U_{i:j-1}, \quad 1 \leq i < j \leq n.$$

Para $j = n+1$ e com a notação óbvia $U_{n+1:n} = 1$, obtemos o resultado já atrás referido, $U_{i:n} \overset{d}{=} B_{i,n-i+1}$.

Demonstração. Se na realidade pensarmos na transformação

$$\begin{cases} Y_1 &= U_{i:n}/U_{j:n} \\ Y_2 &= U_{j:n}, \end{cases}$$

cuja inversa

$$\begin{cases} U_{i:n} &= Y_1 Y_2 \\ U_{j:n} &= Y_2 \end{cases}$$

tem Jacobiano y_2, temos para $f_{_{Y_1,Y_2}}(y_1, y_2)$, na região $0 \leq y_1, y_2 \leq 1$,

$$\begin{aligned} &\frac{n!}{(i-1)!(j-i-1)!(n-j)!}\ (y_1 y_2)^{i-1}\,(y_2(1-y_1))^{j-i-1}\,(1-y_2)^{n-j} y_2 \\ &\quad = \frac{1}{B(i, j-i)}\ y_1^{i-1}(1-y_1)^{j-i-1}\ \frac{1}{B(j, n-j+1)}\ y_2^{j-1}(1-y_2)^{n-j}, \end{aligned}$$

donde segue o resultado pretendido. □

O Teorema anterior admite a seguinte generalização.

Teorema 5.2.2. *Em modelo* Uniforme, *as v.a.'s*

$$(U_{1:n}/U_{2:n},\ U_{2:n}/U_{3:n},\ \ldots, U_{n-1:n}/U_{n:n},\ U_{n:n})$$

são independentes. Para qualquer inteiro $i \in [1, n]$,

$$U_{i:n}/U_{i+1:n} \stackrel{d}{=} B_{i,1} \stackrel{d}{=} U_{i:i}.$$

Demonstração. Temos agora em jogo a transformação n-variada

$$Y_j = U_{j:n}/U_{j+1:n}, \quad 1 \leq j \leq n \quad (U_{n+1:n} \equiv 1),$$

cuja inversa é

$$U_{j:n} = \prod_{i=j}^{n} Y_i, \quad 1 \leq j \leq n.$$

O Jacobiano da transformação inversa é

$$\begin{vmatrix} y_2\ y_3 \ldots\ y_n & y_1\ y_3 \ldots\ y_n & \ldots & \ldots & 1 \\ 0 & y_3 \ldots\ y_n & \ldots & \ldots & 1 \\ 0 & 0 & \ldots & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots & 1 \\ 0 & 0 & \ldots & \ldots & 1 \end{vmatrix} = y_2\ y_3^2\ \ldots\ y_{n-1}^{n-2}\ y_n^{n-1}.$$

Note-se ainda que ao considerarmos os quocientes $Y_j = U_{j:n}/U_{j+1:n}$ estamos a destruir a ordenação inicial, e como $0 \leq U_{j:n} \leq U_{j+1:n} \leq 1$ temos $0 \leq Y_j \leq 1$, $1 \leq j \leq n$. Como conhecemos a f.d.p. conjunta das n e.o.'s em modelo Uniforme, que é dada por

$$f_{U_{1:n},\ldots,U_{n:n}}(u_1, \ldots, u_n) = n!, \quad 0 \leq u_1 < \cdots < u_n \leq 1,$$

caso particular de (5.7), obtem-se de imediato a f.d.p. conjunta de $(Y_1, \ldots, Y_n)$, que é dada por

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = n!\ y_2\ y_3^2\ \ldots\ y_{n-1}^{n-2}\ y_n^{n-1}, \quad 0 \leq y_j \leq 1,\ 1 \leq j \leq n,$$

cujos termos se podem rearranjar, escrevendo

$$y_2 \times \left(2y_3^2\right) \times \left(3y_4^3\right) \cdots \times \left((n-1)y_{n-1}^{n-2}\right) \times \left(ny_n^{n-1}\right),$$

o que para além de evidenciar a independência das v.a.'s $Y_j,\ 1 \leq j \leq n$, nos permite identificar a sua distribuição: Y_j é uma v.a. Beta com parâmetros $(j, 1)$, i.e.

$$U_{j:n}/U_{j+1:n} \overset{d}{=} B_{j,1} \overset{d}{=} U_{j:j},\ 1 \leq j \leq n,$$

o que é um caso particular do resultado obtido no Teorema 5.2.1, com i substituído por j e j substituído por $j+1$

□

Observação 5.2.2. *Note-se que a independência entre $U_{i:n}/U_{j:n}$ e $U_{j:n}$, demonstrada no* Teorema 5.2.1 *segue directamente do* Teorema 5.2.2*, uma vez que podemos escrever*

$$U_{i:n}/U_{j:n} = (U_{i:n}/U_{i+1:n}) \times (U_{i+1:n}/U_{i+2:n}) \times \cdots \times (U_{j-1:n}/U_{j:n})$$

e

$$U_{j:n} = (U_{j:n}/U_{j+1:n}) \times (U_{j+1:n}/U_{j+2:n}) \times \cdots \times (U_{n-1:n}/U_{n:n})\, U_{n:n}.$$

O papel da transformação uniformizante na teoria distribucioanl exacta de e.o.'s

Dada a v.a. Z, absolutamente contínua, proveniente de um modelo F, a nova v.a. $U := F(Z)$ é $\mathcal{U}(0,1)$. Como $U \overset{d}{=} 1-U$, também $1-F(Z)$ é $\mathcal{U}(0,1)$. No entanto, enquanto a primeira transformação, F, é não decrescente, a segunda transformação $1-F$ é não crescente, ou seja, a primeira transformação preserva a ordem, enquanto a segunda transformação reverte a ordem. Formalmente, temos

$$\begin{aligned} F(Z_{i:n}) &\overset{d}{=} U_{i:n}, \quad 1 \leq i \leq n, \\ 1 - F(Z_{i:n}) &\overset{d}{=} U_{n-i+1:n}, \quad 1 \leq i \leq n. \end{aligned}$$

Mais geralmente, para modelo F discreto ou contínuo, podemos escrever

$$Z_{i:n} \overset{d}{=} F^{\leftarrow}(U_{i:n}) \overset{d}{=} F^{\leftarrow}(1 - U_{n-i+1:n}), \quad 1 \leq i \leq n,$$

onde $F^{\leftarrow}$ denota a inversa generalizada de F, i.e., a função em (3.2). Veremos adiante a importância das representações distribucionais anteriores na derivação de propriedades das e.o.'s em modelos diferentes do modelo Uniforme, e com base no comportamento das e.o.'s em modelo Uniforme.

5.2.2 Estatísticas ordinais em modelo Exponencial

Designemos agora genericamente por $E_i,\ 1 \leq i \leq n$ uma amostra de dimensão n proveniente de modelo Exponencial standard, com f.d.p. $f(x) = e^{-x} I_{[0,\infty)}$. Denotemos $E_{1:n} \leq \cdots \leq E_{n:n}$ a amostra das e.o.'s ascendentes.

Na Figura 5.4 exibimos as funções de densidade e de distribuição do modelo original E, de que são provenientes as n observações i.i.d., de $E_{1:n}$ e de $E_{n:n}$, para $n = 3$.

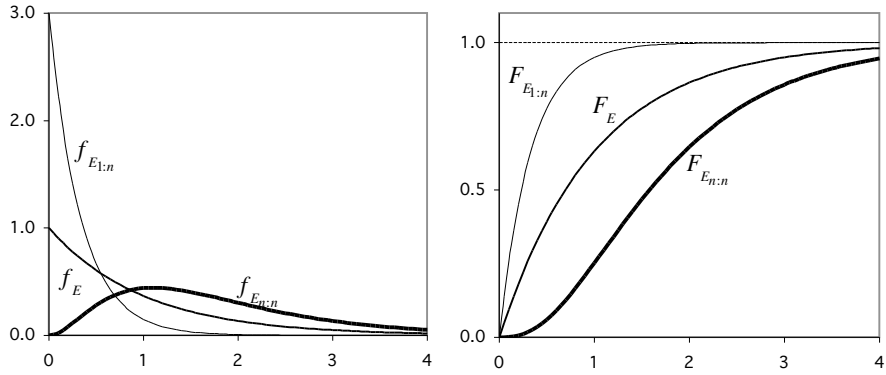


Figura 5.4: Funções densidade de probabilidade (*esquerda*) e funções de distribuição (*direita*) de uma v.a. Exponencial unitária, E, de $E_{1:3}$ e de $E_{3:3}$

Observação 5.2.3. *De momento limitar-nos-emos a realçar o comportamento totalmente diferenciado das cauda esquerda e direita em modelo* Exponencial.

Independência dos espaçamentos em modelo Exponencial

Defina-se

$$S_i^E := E_{i:n} - E_{i-1:n}, \quad 1 \leq i \leq n \qquad (E_0 = 0), \tag{5.8}$$

os *espaçamentos* entre e.o.'s consecutivas em modelo Exponencial. *Abordagem de Rényi* (Rényi[3], 1953). Interpretemos as v.a.'s $(E_1, \ldots, E_n)$ como

[3]Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4**, 191–231.

os tempos de vida (potenciais), agora denotadas $(T_1, \ldots, T_n)$, de n aparelhos postos em funcionamento num instante inicial $t = 0$. O primeiro a avariar-se terá um tempo de vida $T_{1:n}$, e como facilmente se obtém, tem-se para $t \geq 0$,

$$\begin{aligned} F_{T_{1:n}}(t) = \mathbb{P}(T_{1:n} \leq t) = 1 - \mathbb{P}(T_{1:n} > t) &= 1 - \mathbb{P}\Big(\bigcap_{i=1}^{n} \{T_i > t\} \Big) \\ &= 1 - e^{-nt}, \end{aligned}$$

i.e., $T_{1:n}$ continua a ser Exponencial, de valor médio $1/n$. E devido à falta de memória da Exponencial, tudo agora recomeça em $T_{1:n}$, independentemente do que se passou para trás, e com $n-1$ unidades. Consequentemente, o tempo que decorre até à próxima avaria, i.e., $T_{2:n} - T_{1:n}$ é Exponencial, com valor médio $1/(n-1)$, ou seja $T_{2:n} - T_{1:n} \stackrel{d}{=} T_{1:n-1}$. E o processo recomeça em $T_{2:n}$, com $n-2$ unidades, i.e., $T_{3:n} - T_{2:n}$ é independente de tudo o que se passa para trás, e $T_{3:n} - T_{2:n} \stackrel{d}{=} T_{1:n-2}$ é Exponencial de valor médio $1/(n-2)$. De modo intuitivo, cuja demonstração formalizaremos adiante, e através de um *teste de vida* em aparelhos com tempo de vida Exponencial, demonstrámos pois a validade do teorema seguinte.

Teorema 5.2.3. *Seja $(E_1, \ldots, E_n)$ uma amostra aleatória proveniente de um modelo* Exponencial *unitário, e sejam $(E_{1:n} \leq \cdots \leq E_{n:n})$ as e.o.'s ascendentes associadas a essa amostra.*
Então os espaçamentos S_i^E em (5.8) são mutuamente independentes e exponenciais, de valor médio $1/(n-i+1)$, $1 \leq i \leq n$, i.e., S_i^E tem f.d.p.

$$f_{S_i^E}(s) = (n-i+1)e^{-(n-i+1)s} I_{[0,+\infty)}, \quad 1 \leq i \leq n.$$

Demonstração. Comecemos por uma demonstração clássica, com recurso ao método de transformação de variáveis aleatórias. Notemos em primeiro lugar que, ao escrever a transformação em (5.8) estamos a destruir a ligação de ordem entre as e.o.'s, e como $0 \leq E_{i:n} \leq E_{i+1:n}$, tem-se $S_i^E \geq 0$, $1 \leq i \leq n$. A transformação inversa da transformação em (5.8) é

$$E_{i:n} = \sum_{j=1}^{i} S_j^E, \ 1 \leq i \leq n,$$

com Jacobiano unitário, e a f.d.p. conjunta das n e.o.'s em modelo Exponencial é dada por

$$f_{E_{1:n},\ldots,E_{n:n}}(e_1,\ldots,e_n) = n!\; e^{-\sum_{i=1}^{n} e_i}, \quad 0 \le e_1 \le \cdots \le e_n.$$

Consequentemente, a f.d.p. conjunta de $(S_1^E,\ldots,S_n^E)$ é dada por

$$f_{S_1^E,\ldots,S_n^E}(s_1,\ldots,s_n) = n!\; e^{-\sum_{i=1}^{n}\sum_{j=1}^{i} s_j}, \quad s_j \ge 0,\ 1 \le j \le n.$$

Como

$$\sum_{i=1}^{n}\sum_{j=1}^{i} s_j = \sum_{j=1}^{n}\sum_{i=j}^{n} s_j = \sum_{j=1}^{n}(n-j+1)s_j$$

e $n! = \prod_{j=1}^{n}(n-j+1)$, conseguimos a decomposição

$$f_{S_1^E,\ldots,S_n^E}(s_1,\ldots,s_n) = \prod_{j=1}^{n}(n-j+1)e^{-(n-j+1)s_j}, \quad s_j \ge 0,\ 1 \le j \le n,$$

o que nos permite garantir que os n espaçamentos S_j^E são mutuamente independentes, com S_j^E Exponencial de valor médio $1/(n-j+1)$, $1 \le j \le n$.

Como demonstração alternativa podíamos ter recorrido à transformação uniformizante e ao Teorema 5.2.2. Na realidade, o facto da v.a. E ter f.d. $F(x) = \{1-e^{-x}\}I_{[0,+\infty)}$, permite-nos obter $1-e^{-E} \overset{d}{=} U$, $U \frown \mathcal{U}(0,1)$, e podemos pois escrever $E \overset{d}{=} -\log(1-U)$ ou $E \overset{d}{=} -\log U$. Em termos de e.o.'s, e face ao facto de a transformação $-\log$ ser decrescente, tem-se

$$E_{i:n} \overset{d}{=} -\log U_{n-i+1:n}.$$

Consequentemente, para $1 \le i \le n$, e com a notação usual $U_{n+1:n} \equiv 1$,

$$S_i^E = E_{i:n} - E_{i-1:n} \overset{d}{=} -\log U_{n-i+1:n} - (-\log U_{n-i+2:n}) = -\log\frac{U_{n-i+1:n}}{U_{n-i+2:n}},$$

que é ainda distribucionalmente igual a $-\log B_{n-i+1:1} \overset{d}{=} -\log U_{n-i+1:n-i+1} \overset{d}{=} E_{1:n-i+1} \overset{d}{=} E/(n-i+1)$, i.e., Exponencial de valor médio $1/(n-i+1)$. Conjuntamente, tem-se

$$\left(S_1^E, S_2^E, \ldots, S_n^E\right) \overset{d}{=} \left(-\log U_{n:n}, -\log\frac{U_{n-1:n}}{U_{n:n}}, \ldots, -\log\frac{U_{1:n}}{U_{2:n}}\right),$$

donde segue a independência dos n espaçamentos em modelo Exponencial. □

Observação 5.2.4. *Note-se que se o modelo fosse* Exponencial *de valor médio σ, i.e., se estivéssemos a trabalhar com X, v.a. com f.d.p. $f(x) = \frac{1}{\sigma}e^{-x/\sigma}I_{[0,+\infty)}$, X/σ seria* Exponencial *unitária ou padrão, e tudo o que foi dito anteriormente sofreria uma adaptação trivial.*

Representação de Rényi para os espaçamentos de e.o.'s em modelo Exponencial

Comecemos por notar que qualquer e.o. se pode escrever como soma dos espaçamentos entre e.o.'s consecutivas, até chegarmos a essa estatística ordinal. Mais precisamente, e no contexto de amostra Exponencial em que nos colocámos, temos

$$E_{i:n} = \sum_{j=1}^{i}\{E_{j:n} - E_{j-1:n}\}, \quad 1 \leq i \leq n, \quad (E_0 \equiv 0).$$

Como, por outro lado, pelo Teorema 5.2.3,

$$E_{j:n} - E_{j-1:n} \stackrel{d}{=} E_{1:j} \stackrel{d}{=} \frac{E_j}{n-j+1}, \ 1 \leq j \leq n,$$

segue-se que:

Teorema 5.2.4. (Representação de Rényi). *No contexto do* Teorema 5.2.3*, temos a validade da representação distribucional,*

$$E_{i:n} \stackrel{d}{=} \sum_{j=1}^{i} \frac{E_j}{n-j+1}, \ 1 \leq i \leq n, \tag{5.9}$$

com $\{E_j\}_{j\geq 1}$ sucessão de v.a.'s i.i.d., exponenciais unitárias.

5.2.3 Estatísticas ordinais em modelo Pareto

O modelo Pareto com parâmetro de forma unitário, ou seja, uma v.a. P com f.d. dada por $F_P(y) = 1 - y^{-1}, \ y \geq 1$, goza de papel fundamental em *Estatística de Extremos*, aquando da obtenção das propriedades de estimadores de parâmetros de acontecimentos raros, em contexto semi-paramétrico, a precisar mais adiante.

Temos $1 - 1/P \stackrel{d}{=} U$, e consequentemente não só $P \stackrel{d}{=} 1/U$, uma vez que $U \stackrel{d}{=} 1 - U$, mas também $\log P \stackrel{d}{=} -\log(1 - U)$. Por outro lado, como a transformação uniformizante nos permite garantir que para uma v.a. Exponencial unitária, E, se tem $E \stackrel{d}{=} -\log(1 - U)$, temos $\log P \stackrel{d}{=} E$. Condensando toda a informação anterior podemos pois escrever

$$P \stackrel{d}{=} 1/U \stackrel{d}{=} e^E,$$

onde a primeira transformação é decrescente, e a segunda crescente.

As propriedades das e.o.'s em modelo Pareto podem facilmente deduzir-se a partir de propriedades de e.o's quer em modelo Uniforme, quer em modelo Exponencial. Tem-se, na realidade, e com base no que foi anteriormente referido

$$P_{i:n} \stackrel{d}{=} 1/U_{n-i+1:n} \stackrel{d}{=} e^{E_{i:n}}, \quad 1 \leq i \leq n,$$

que por vezes nos convem escrever na forma

$$\log P_{i:n} \stackrel{d}{=} -\log U_{n-i+1:n} \stackrel{d}{=} E_{i:n}, \quad 1 \leq i \leq n.$$

Na Figura 5.5 ilustramos o comportamento das e.o.'s em modelo Pareto.

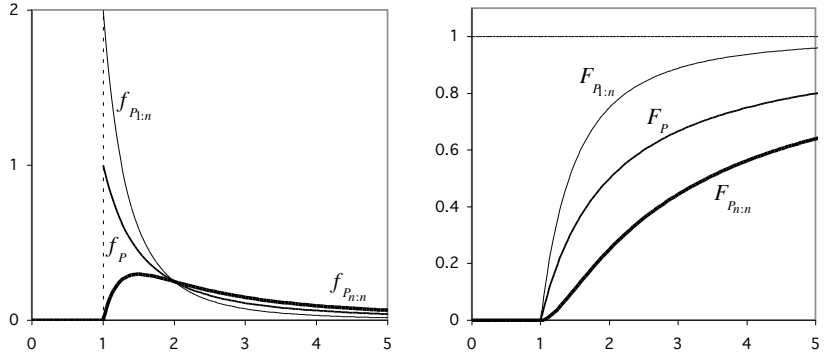


Figura 5.5: Funções densidade de probabilidade (*esquerda*) e funções de distribuição (*direita*) de uma v.a. Pareto unitária, P, de $P_{1:3}$ e de $P_{3:3}$

Observação 5.2.5. *Limitemo-nos a atentar no 'peso elevado' da cauda direita do modelo* Pareto. *Na realidade o modelo* Pareto *é um dos modelos de*

cauda pesada mais simples, frequentemente usado na prática. O conceito de 'peso de cauda' será precisado mais adiante, aquando do estudo do comportamento assintótico de e.o.'s.

O Teorema 5.2.2 pode ser refraseado em termos do modelo Pareto, e temos

Teorema 5.2.5. *Em modelo* Pareto *unitário os quocientes entre e.o.'s consecutivas são independentes, i.e., as v.a.'s*

$$(1/P_{1:n},\ P_{1:n}/P_{2:n},\ \ldots,\ P_{n-2:n}/P_{n-1:n},\ P_{n-1:n}/P_{n:n})$$

são independentes. Para qualquer inteiro $i \in [1, n]$,

$$P_{i-1:n}/P_{i:n} \stackrel{d}{=} B_{n-i+1,1} \quad (P_{0:n} \equiv 1).$$

Uma representação do tipo da representação de Rényi, dada no Teorema 5.2.4, é também frequentemente usada para e.o.'s em modelo Pareto. Tem-se

Teorema 5.2.6. *Em modelo* Pareto *temos a validade da representação distribucional,*

$$P_{i:n} \stackrel{d}{=} \prod_{j=1}^{i} e^{\frac{E_j}{n-j+1}}, \qquad 1 \le i \le n,$$

com $\{E_j\}_{j\ge1}$ *sucessão de v.a.'s i.i.d., exponenciais unitárias. Tem-se ainda,*

$$P_{j:n}/P_{i:n} \stackrel{d}{=} P_{j-i:n-i}, \quad 1 \le i < j \le n.$$

Demonstração. Já vimos anteriormente que $P_{i:n} \stackrel{d}{=} e^{E_{i:n}}$. A utilização da representação de Rényi para $E_{i:n}$ permite-nos então escrever

$$P_{i:n} \stackrel{d}{=} e^{E_{i:n}} \stackrel{d}{=} e^{\sum_{j=1}^{i} \frac{E_j}{n-j+1}} = \prod_{j=1}^{i} e^{\frac{E_j}{n-j+1}},$$

como pretendíamos demonstrar. Esta representação permite-nos escrever, para $j > i$,

$$\frac{P_{j:n}}{P_{i:n}} \stackrel{d}{=} \frac{\prod_{k=1}^{j} e^{\frac{E_k}{n-k+1}}}{\prod_{k=1}^{i} e^{\frac{E_k}{n-k+1}}} = \prod_{k=i+1}^{j} e^{\frac{E_k}{n-k+1}} = \prod_{l=1}^{j-i} e^{\frac{E_l}{(n-i)-l+1}} \stackrel{d}{=} P_{j-i:n-i}.$$

□

5.3 Momentos de estatísticas ordinais

A estrutura de segunda ordem das e.o.'s fica completamente definida se conhecermos o seu vector coluna de valores médios, e a sua matriz de covariâncias,

$$\underline{\mu} = [\mu_{i:n}]_{1\times n}\,, \quad \mu_{i:n} := \mathbb{E}(Z_{i:n}), \quad 1 \le i \le n,$$
$$\Sigma = [\sigma_{i,j:n}]_{n\times n}\,,\ \sigma_{i,j:n} := Cov(Z_{i:n}, Z_{j:n}) = \sigma_{j,i:n},\ 1 \le i \le j \le n.$$

Como obter μ e Σ? Tudo depende do modelo F. Formalmente:

$$\mu_{i:n} = \mathbb{E}(Z_{i:n}) = \frac{1}{B(i, n-i+1)} \int_0^1 F^{\leftarrow}(u) u^{i-1}(1-u)^{n-i} du, \qquad (5.10)$$

onde $F^{\leftarrow}$ é a função inversa generalizada, definida em (3.2). Mais geralmente, para $\mathbb{E}[Z_{i:n}^k]$ tem-se

$$\mu_{i:n}^{(k)} := \mathbb{E}(Z_{i:n}^k) = \frac{1}{B(i, n-i+1)} \int_0^1 \left[F^{\leftarrow}(u)\right]^k u^{i-1}(1-u)^{n-i} du. \quad (5.11)$$

Para a estrutura bivariada,

$$\mu_{i,j:n} := \mathbb{E}(Z_{i:n} Z_{j:n}) = \frac{1}{B(i, j-i) B(j, n-j+1)} \iint\limits_{0\le u<v\le 1} F^{\leftarrow}(u) F^{\leftarrow}(v) u^{i-1}(v-u)^{j-i-1}(1-v)^{n-j} du dv,$$

e

$$\sigma_{i,j:n} = Cov(Z_{i:n}, Z_{j:n}) = \mu_{i,j:n} - \mu_{i:n}\ \mu_{j:n},\ 1 \le i, j \le n. \qquad (5.12)$$

Do ponto de vista de cálculo é usual recorrer, em situações complicadas, a relações de recorrência, pois o cálculo integral directo provoca frequentemente erros de arredondamento. Note-se que para o cálculo de momentos de e.o.'s é usualmente necessário utilizar métodos de integração numérica, e daí a existência de tabelas razoavelmente exaustivas para alguns modelos, entre os quais incluímos o modelo Normal, Gama, Logístico e Gumbel (veja-se, por exemplo, Sarhan & Greenberg, 1962).

O processo directo mais usual para o cálculo de $\mu_{i:n}$ consiste em desenvolver $(1-u)^{n-i}$ em binómio de Newton, na expressão (5.10), obtendo-se

$$\begin{aligned}\mu_{i:n} &= \frac{1}{B(i,n-i+1)}\int_0^1 F^{\leftarrow}(u)u^{i-1}(1-u)^{n-i}du \\ &= \sum_{k=0}^{n-i}\binom{n-i}{k}(-1)^{n-i-k}\frac{1}{B(i,n-i+1)}\int_0^1 F^{\leftarrow}(u)u^{n-k-1}du. \quad (5.13)\end{aligned}$$

Este último integral $\int_0^1 F^{\leftarrow}(u)u^{n-k-1}du$ é usualmente fácil de calcular, mas mesmo assim temos em (5.13) uma soma de termos alternadamente positivos e negativos, e eventualmente de grandeza elevada especialmente quando n for elevado e $i \approx [n/2]$. Tal ocasiona erros de arredondamento, que podem ser graves, e que se agravam quando passamos ao cálculo directo dos elementos da matriz de covariâncias Σ com utilização do mesmo processo, que, consequentemente, não aconselhamos.

Surge pois a necessidade da utilização de relações de recorrência, que vão permitir reduzir o tempo de computação, o número de cálculos independentes e os erros de arredondamento.

A estrutura distribucional das e.o.'s permite facilmente a obtenção dessas relações de recorrência, que servem para controlar, simplificar ou meramente facilitar o cálculo efectivo dos momentos.

5.3.1 Relações de controlo

Tratam-se de relações de controlo global (nunca individual), baseadas no facto de se ter, para qualquer $k \geq 1$ e $m \geq 1$,

$$\left[\sum_{i=1}^{n} Z_{i:n}^{k}\right]^{m} = \left[\sum_{i=1}^{n} Z_{i}^{k}\right]^{m}. \quad (5.14)$$

Teorema 5.3.1. *Com $\mu = \mathbb{E}(Z)$ e $\sigma^2 = \mathbb{V}ar(Z)$, tem-se*

$$\sum_{i=1}^{n} \mu_{i:n} = n\mu, \tag{5.15}$$

$$\sum_{i=1}^{n} \mu_{i,i:n} = n(\mu^2 + \sigma^2), \tag{5.16}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{i,j:n} = n(\mu^2 + \sigma^2) + n(n-1)\mu^2, \tag{5.17}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{i,j:n} = n\sigma^2. \tag{5.18}$$

Demonstração. As relações (5.15), (5.16) e (5.17) obtêm-se de (5.14) fazendo $(k,m) = (1,1)$, $(k,m) = (2,1)$ e $(k,m) = (1,2)$, respectivamente. A relação (5.18) de controlo de covariâncias obtém-se também facilmente atendendo a que

$$\Big[\sum_{i=1}^{n}(Z_{i:n} - \mu_{i:n})\Big]^2 = \Big[\sum_{i=1}^{n}(Z_i - \mu)\Big]^2.$$

□

5.3.2 Relações simplificativas

No cálculo efectivo de momentos de e.o.'s é importante ter em linha de conta as seguintes relações simplificativas para populações simétricas e em particular para a população Normal. Tais relações permitem diminuir drasticamente o número de operações a efectuar.

Teorema 5.3.2. *Seja F um modelo simétrico (em torno de 0, sem perda de generalidade). Então*

$$\mu_{i:n}^{(k)} = (-1)^k \mu_{n-i+1:n}^{(k)},\ 1 \le i \le n, \tag{5.19}$$

$$\sigma_{i,j:n} = \sigma_{n-j+1,n-i+1:n},\ 1 \le i < j \le n. \tag{5.20}$$

Demonstração. Em modelo simétrico, $F(x) = 1 - F(-x),\ \forall x \in \mathbb{R}$, o que em termos da função quantil $F^{\leftarrow}$ se pode escrever como $F^{\leftarrow}(t) = -F^{\leftarrow}(1-t)$, para qualquer t, $0 \le t \le 1$. A mudança de variável $v = 1 - u$ em (5.11)

conduz de imediato á relação (5.19). A demonstração da relação (5.20) é análoga, usando a mudança de variável $x = 1 - u,\ y = 1 - v$ em (5.12). □

Teorema 5.3.3. *A matriz de covariâncias das e.o.'s em modelo* Normal *padrão é duplamente estocástica, i.e.,*

$$\sum_{j=1}^{n} \sigma_{i,j:n} = \sum_{i=1}^{n} \sigma_{i,j:n} = 1.$$

Demonstração. Esta propriedade (característica das e.o.'s em modelo Normal) decorre imediatamente da independência entre $\overline{Z}$ e $\left\{Z_{i:n} - \overline{Z}\right\}$, $1 \leq i \leq n$. Desta independência entre $\overline{Z}$ e $\left\{Z_{i:n} - \overline{Z}\right\}$ segue que

$$\mathbb{E}\left[(Z_{i:n} - \overline{Z})\overline{Z}\right] = \mathbb{E}\left[(Z_{i:n} - \overline{Z})\right] \times \mathbb{E}\left[\overline{Z}\right] = 0.$$

Por outro lado, podemos escrever

$$\begin{aligned} 0 \equiv \mathbb{E}\left[(Z_{i:n} - \overline{Z})\overline{Z}\right] &= \mathbb{E}[Z_{i:n}\overline{Z}] - \mathbb{E}[\overline{Z}^2] \\ &= \mathbb{E}\Big[Z_{i:n}\Big(\frac{1}{n}\sum_{j=1}^{n} Z_{j:n}\Big)\Big] - \frac{1}{n} = \frac{1}{n}\sum_{j=1}^{n} \mu_{i,j:n} - \frac{1}{n}, \end{aligned}$$

e consequentemente $\sum_{j=1}^{n} \mu_{i,j:n} = 1$. Então

$$\sum_{j=1}^{n} \sigma_{i,j:n} = \sum_{j=1}^{n} \mu_{i,j:n} - \mu_{i:n} \sum_{j=1}^{n} \mu_{j:n} = 1,$$

uma vez que $\sum_{j=1}^{n} \mu_{j:n} = n\ \mathbb{E}[X] = 0$.
De forma análoga se demonstra a segunda identidade. □

5.3.3 Relações de cálculo efectivo

As relações mais importantes para o cálculo efectivo do vector coluna $\underline{\mu}$ de valores médios e da matriz Σ de covariâncias das e.o.'s são

$$\mu_{i+1:n} = \frac{n\mu_{i:n-1} - (n-i)\mu_{i:n}}{i},\ 1 \leq i \leq n-1, \tag{5.21}$$

e para $1 < i < j \leq n$,

$$\mu_{i,j:n} = \frac{n\mu_{i-1,j-1:n-1} - (j-i)\mu_{i-1,j:n} - (n-j+1)\mu_{i-1,j-1:n}}{i-1}. \quad (5.22)$$

A relação de recorrência (5.21) foi derivada, em contexto mais geral, por Srikantan[4] (1962):

Teorema 5.3.4. *Para uma dada função g, para a qual existe $\mathbb{E}[g(Z_{i:n})]$, $1 \leq i \leq n$, tem-se*

$$(n-i)\mathbb{E}[g(Z_{i:n})] + i\mathbb{E}[g(Z_{i+1:n})] = n\mathbb{E}[g(Z_{i:n-1})], \ 1 \leq i \leq n-1,$$

i.e.,

$$\mathbb{E}[g(Z_{i+1:n})] = \frac{n\mathbb{E}[g(Z_{i:n-1})] - (n-i)\mathbb{E}[g(Z_{i:n})]}{i}, \ 1 \leq i \leq n-1. \quad (5.23)$$

Demonstração. A demonstração deste resultado decorre facilmente se introduzirmos o factor artificial $u+(1-u) \equiv 1$ na expressão integral de $\mathbb{E}[g(Z_{i:n})]$, i.e., se repararmos que

$$\begin{aligned}
\mathbb{E}[g(Z_{i:n})] &= \frac{1}{B(i,n-i+1)} \int_0^1 g\left(F^{\leftarrow}(u)\right) u^{i-1}(1-u)^{n-i} du \\
&= \frac{1}{B(i,n-i+1)} \int_0^1 g\left(F^{\leftarrow}(u)\right) u^{i-1}(1-u)^{n-i} \{u+(1-u)\} du \\
&= \frac{1}{B(i,n-i+1)} \int_0^1 g\left(F^{\leftarrow}(u)\right) u^{i}(1-u)^{n-i} du \\
&+ \frac{1}{B(i,n-i+1)} \int_0^1 g\left(F^{\leftarrow}(u)\right) u^{i-1}(1-u)^{n-i+1} du \\
&= \frac{B(i+1,n-i+1)}{B(i,n-i+1)} \mathbb{E}[g(X_{i+1:n+1})] + \frac{B(i,n-i+2)}{B(i,n-i+1)} \mathbb{E}[g(Z_{i:n+1})].
\end{aligned}$$

□

A relação (5.21) é pois válida para momentos de qualquer ordem k, i.e., para $g(x) = x^k$, tal foi demonstrado por Cole[5] (1951) para modelos contínuos,

[4] Srikantan, K.S. (1962). Recurrence relations between the PDF's of order statistics and some applications. *Ann. Math. Statist.* **42**, 35–45.

[5] Cole, R.H. (1951). Relations between moments of order statistics. *Ann. Math. Statist.* **22**, 308–310.

podendo ser facilmente derivado para modelos discretos. Mais genericamente que (5.21):

Corolário 5.3.1. *Com* $\mu^{(k)}_{i+1:n} = \mathbb{E}\left[Z^k_{i:n}\right]$, $1 \leq i \leq n$, *tem-se*

$$\mu^{(k)}_{i+1:n} = \frac{n\mu^{(k)}_{i:n-1} - (n-i)\mu^{(k)}_{i:n}}{i}, \quad 1 \leq i \leq n-1, \quad k \geq 1. \tag{5.24}$$

Note-se pois que basta obter valores médios de mínimos $\mu^{(k)}_{1:j}$, $1 \leq j \leq n$ (ou valores médios de máximos, $\mu^{(k)}_{j:j}$, $1 \leq j \leq n$), para imediatamente termos toda a estrutura $\mu^{(k)}_{j:n}$, $1 \leq j \leq n$, para a nossa dimensão de amostra.

Tem-se ainda

Corolário 5.3.2. *A solução da relação de recorrência* (5.23) *é*

$$\mathbb{E}\left[g(Z_{i:k})\right] = (-1)^{k-i+1} \sum_{j=n-i+1}^{k} (-1)^j \binom{k}{j} \binom{j-1}{k-i} \mathbb{E}\left[g(Z_{1:j})\right], \tag{5.25}$$

para $1 \leq i \leq k \leq n$.

Corolário 5.3.3. *Outra solução possível da relação de recorrência* (5.23) *vem expressa em termos das e.o.'s superiores e é dada por*

$$\mathbb{E}\left[g(Z_{i:k})\right] = (-1)^i \sum_{j=i}^{k} (-1)^j \binom{k}{j} \binom{j-1}{i-1} \mathbb{E}\left[g(Z_{j:j})\right], \tag{5.26}$$

para $1 \leq i \leq k \leq n$.

Observações.

1. (5.26) obtém-se de (5.25) passando para $Y = -Z$ em vez de Z e para $g(-y)$ em vez de $g(z)$.

2. O uso de qualquer das fórmulas (5.25) ou (5.26) torna os erros de arredondamente importantes, quando $k - i$ cresce.

3. A fórmula (5.25) pode-se obter directamente, colocando em $\mathbb{E}\left[g(Z_{i:n})\right] = \frac{1}{B(i,n-i+1)} \int_0^1 g\left(F^{\leftarrow}(u)\right) u^{i-1}(1-u)^{n-i} du$, $u^{i-1} = (1-(1-u))^{i-1}$, e desenvolvendo $(1-(1-u))^{i-1}$ em binómio de Newton.

4. Analogamente a expressão (5.26) obtém-se directamente de $\mathbb{E}\left[g(Z_{i:n})\right] = \frac{1}{B(i,n-i+1)} \int_0^1 g\left(F^{\leftarrow}(u)\right) u^{i-1}(1-u)^{n-i} du$, desenvolvendo $(1-u)^{n-i}$ em binómio de Newton.

Fazendo $i = n/2$ em (5.24), têm-se as seguintes consequências imediatas.

Corolário 5.3.4.

$$\mu^{(k)}_{n/2:n-1} = \frac{1}{2}\left\{\mu^{(k)}_{n/2+1:n} + \mu^{(k)}_{n/2:n}\right\}. \tag{5.27}$$

Adicionalmente, como caso particular da relação anterior, surge o resultado:

Corolário 5.3.5. *Se a f.d. F for simétrica em torno da origem,*

$$\mu^{(k)}_{n/2:n-1} = \begin{cases} 0 & \text{se } k \text{ ímpar,} \\ \mu^{(k)}_{n/2:n} & \text{se } k \text{ par.} \end{cases}$$

Demonstração. Basta usar (5.27) e o facto de se ter, por (5.19), $\mu^{(k)}_{n/2+1:n} = (-1)^k \mu^{(k)}_{n-n/2-1+1:n} = (-1)^k \mu^{(k)}_{n/2:n}$. □

Teorema 5.3.5. *Para F arbitrária, e qualquer que seja a função $\Psi(x,y)$ tal que $\mathbb{E}\left[\Psi(Z_{i:n}, Z_{j:n})\right]$ exista para $1 \leq i < j \leq n$, tem-se*

$$\begin{aligned}\mathbb{E}\left[\Psi(Z_{i+1:n}, Z_{j+1:n})\right] = \{&n\mathbb{E}\left[\Psi(Z_{i:n-1}, Z_{j:n-1})\right] \\ &- (n-j)\mathbb{E}\left[\Psi(Z_{i:n}, Z_{j:n})\right] - (j-i)\mathbb{E}\left[\Psi(Z_{i:n}, Z_{j+1:n})\right]\}/i,\end{aligned} \tag{5.28}$$

para $1 \leq i < j \leq n$.

Demonstração. A demonstração é absolutamente análoga à demonstração do Teorema 5.3.4 por introdução do factor $[u + (v-u) + (1-v)] \equiv 1$ na função integranda de $\mathbb{E}\left[\Psi(Z_{i:n}, Z_{j:n})\right]$. □

Corolário 5.3.6. *Para F arbitrária e $1 \leq i < j \leq n$, tem-se*

$$(i-1)\mu_{i,j:n} + (j-i)\mu_{i-1,j:n} + (n-j+1)\mu_{i-1,j-1:n} = n\mu_{i-1,j-1:n-1}. \tag{5.29}$$

Note-se que a relação (5.22) é a relação (5.29), após explicitação do termo de maior ordem em amostra de maior dimensão em termos dos outros.

Corolário 5.3.7. *A solução da relação de recorrência* (5.28) *é*

$$\mathbb{E}\left[\Psi(Z_{i:k}, Z_{j:k})\right] = (-1)^{i-1} \sum_{l=0}^{i-1} (-1)^l \binom{k}{l} \sum_{m=l}^{i-1} \binom{k+j+m-l}{n-l} \times \\ \times \binom{j-2-m}{i-1-m} \mathbb{E}\left[\Psi(Z_{1:k-l}, Z_{j-m:k-l})\right],$$

para $1 \leq i < j \leq k \leq n$.

As vantagens da relação de recorrência (5.28) é a de permitir obter os $n(n^2-1)/6$ valores esperados $\mathbb{E}\left[\Psi(Z_{i:k}, Z_{j:k})\right]$, $1 \leq i < j \leq k \leq n$, em termos dos $n(n-1)/2$ valores esperados, $\mathbb{E}\left[\Psi(Z_{1:k}, Z_{j:k})\right]$, $1 < j \leq k \leq n$.

5.3.4 Momentos em modelo Uniforme

Já vimos anteriormente que, em modelo Uniforme, $U_{i:n} \overset{d}{=} B_{i,n-i+1}$, ou seja

$$f_{U_{i:n}}(u) = \frac{n!}{(i-1)!(n-i)!} u^{i-1}(1-u)^{n-i}, \ 0 \leq u \leq 1, \ 1 \leq i \leq n.$$

Consequentemente, tem-se

$$\mathbb{E}\left[U_{i:n}^{\alpha}\right] = \frac{B(i+\alpha, n-i+1)}{B(i, n-i+1)}, \quad 1 \leq i \leq n, \quad \alpha \in \mathbb{R}^+, \tag{5.30}$$

e como casos particulares importantes,

$$\mathbb{E}[U_{i:n}] = \frac{i}{n+1}, \quad 1 \leq i \leq n, \tag{5.31}$$
$$\mathbb{V}ar[U_{i:n}] = \frac{i(n-i+1)}{(n+1)^2(n+2)}, \quad 1 \leq i \leq n.$$

A relação (5.31), já sobejamente utilizada no Capítulo 4, permite-nos dizer que as e.o.'s em modelo Uniforme dividem a área abaixo da curva $y = f(x)$ em $n+1$ partes, cada uma com valor médio $1/(n+1)$, i.e., fazem a *divisão aleatória* do intervalo $(0,1)$.

Uma propriedade importante das e.o.'s em modelo Uniforme, que permite simplificar o cálculo da estrutura dos seus momentos, e que enunciámos no Teorema 5.2.1, garante-nos que quaisquer que sejam os inteiros $1 \leq i < j \leq n$,

as componentes do par aleatório $(U_{i:n}/U_{j:n}, U_{j:n})$ são independentes, ambas com distribuição Beta, onde $U_{i:n}/U_{j:n}$ é uma v.a. $B_{i,j-i}$.

Mais geralmente, para $1 \leq i < j < k < l \leq n$, as v.a.'s

$$\begin{cases} Y_1 &= U_{i:n}/U_{j:n} \\ Y_2 &= U_{j:n}/U_{k:n} \\ Y_3 &= U_{k:n}/U_{l:n} \\ Y_4 &= U_{l:n}, \end{cases}$$

são independentes, e podemos escrever

$$\mathbb{E}\left[U_{i:n}^{\alpha} U_{j:n}^{\beta}\ U_{k:n}^{\gamma}\ U_{l:n}^{\delta}\right] = \mathbb{E}\left[Y_1^{\alpha}\ Y_2^{\alpha+\beta}\ Y_3^{\alpha+\beta+\gamma}\ Y_4^{\alpha+\beta+\gamma+\delta}\right],$$

i.e., podemos exprimir o valor médio do produto de v.a.'s dependentes como o valor médio do produto de v.a.'s independentes, que é consequentemente igual ao produto dos valores médios dessas v.a.'s, valores fáceis de calcular, devido ao facto de os Y_i serem Beta. Consequentemente, e recorrendo a (5.30) obtém-se para $\mathbb{E}\left[U_{i:n}^{\alpha}\ U_{j:n}^{\beta}\ U_{k:n}^{\gamma}\ U_{l:n}^{\delta}\right]$, $1 \leq i < j < k < l \leq n$, o valor

$$\begin{aligned} &\frac{B(i+\alpha, j-i)}{B(i, j-i)} \times \frac{B(j+\alpha+\beta, k-j)}{B(j, k-j)} \times \frac{B(k+\alpha+\beta+\gamma, l-k)}{B(k, l-k)} \\ &\qquad \times \frac{B(l+\alpha+\beta+\gamma+\delta, n-l+1)}{B(l, n-l+1)} \\ &= \frac{(i-1+\alpha)!(j-1+\alpha+\beta)!(k-1+\alpha+\beta+\gamma)!(l-1+\alpha+\beta+\gamma+\delta)!n!}{(i-1)!(j-1+\alpha)!(k-1+\alpha+\beta)!(l-1+\alpha+\beta+\gamma)!(n+\alpha+\beta+\gamma+\delta)!}. \end{aligned}$$

Em particular, pondo

$$\mu_{i:n} = \frac{i}{n+1} =: p_i, \qquad q_i := 1 - p_i,\ 1 \leq i \leq n,$$

tem-se

$$\sigma_{i,j:n} = \frac{i!(j+1)!n!}{(i-1)!j!(n+2)!} - \frac{i}{n+1}\ \frac{j}{n+1} = \frac{p_i q_j}{n+2},\ 1 \leq i \leq j \leq n.$$

Para a variância temos

$$\sigma_{i:n}^2 = \sigma_{i,i:n} = \frac{p_i q_i}{n+1} = \frac{i(n-i+1)}{(n+1)^2(n+1)}, \quad 1 \leq i \leq n.$$

De interesse mais adiante, aquando da obtenção de aproximações para os momentos de e.o.'s, referimos ainda que, para $1 \leq i \leq j \leq k \leq n$,

$$\mathbb{E}\left[(U_{i:n} - \mu_{i:n})(U_{j:n} - \mu_{j:n})(U_{k:n} - \mu_{k:n})\right] = \frac{2p_i(q_j - p_j)q_k}{(n+2)(n+3)},$$

e, para $1 \leq i \leq n$,

$$\mathbb{E}\left[(U_{i:n} - \mu_{i:n})^4\right] = \frac{3p_i^2 q_i^2}{(n+2)^2} + \frac{6p_i q_i}{(n+2)(n+3)(n+4)}\left[(q_i - p_i)^2 - \frac{(n+3)p_i q_i}{n+2}\right].$$

5.3.5 Momentos em modelo Exponencial

Começaremos por obter os momentos das e.o.'s em modelo Exponencial, com base na representação de Rényi, dada em (5.9),

$$E_{i:n} = \sum_{j=1}^{i} \frac{E_j}{n-j+1},\ 1 \leq i \leq n,$$

com $\{E_j\}_{j\geq 1}$ sucessão de v.a.'s independentes e exponenciais unitárias. Após essa derivação, que como veremos é bastante expedita, iremos esboçar o cálculo directo, a fim de ilustrar, por um lado, a potencialidade da representação de Rényi, e por outro lado para apresentar algumas funções clássicas da análise, importantes em contextos diversos da área da *Probabilidade* e da *Estatística*, incluindo a área das *Estatísticas Ordinais*.

Cálculo com base na representação de Rényi

De forma imediata, utilizando a representação de Rényi, obtemos

$$\begin{aligned}\mu_i^E = \mathbb{E}\left[E_{i:n}\right] = \sum_{j=1}^{i} \mathbb{E}\left\{\frac{E_j}{n-j+1}\right\} &= \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-i+1} \\ &= \psi(n+1) - \psi(n-i+1),\end{aligned}$$

onde ψ é a chamada função digama, a derivada do logaritmo da função Gama em (3.4).

Observação 5.3.1. *Mais precisamente tem-se*

$$\psi(\alpha) = \frac{d}{d\alpha}\log\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{\int_0^{+\infty} \log x\ x^{\alpha-1}\ e^{-x}\ dx}{\int_0^{+\infty} x^{\alpha-1}\ e^{-x}\ dx}, \quad \alpha \in \mathbb{R}^+. \tag{5.32}$$

Esta função encontra-se extensivamente tabelada no livro já anteriormente referido de Abramowitz & Stegun (1992)*. Para valores de α inteiros positivos, tem-se*

$$\psi(1) = -\epsilon, \quad \psi(n) = -\epsilon + \sum_{j=1}^{n-1} \frac{1}{j}, \quad n \geq 2, \quad \epsilon = 0.57721\ 56649\ \ldots.$$

A constante ϵ é a chamada constante de Euler*. A série harmónica, $\sum_{j\geq 1} 1/j$, é divergente, com um comportamento assintótico análogo ao da função logarítmica, e tem-se*

$$\epsilon = \lim_{m\to\infty} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m} - \log m\right] = \lim_{m\to\infty} \left[\psi(m+1) - \log m\right].$$

Também temos para a função digama*, em* (5.32)*, uma relação de recorrência importante,*

$$\psi(\alpha + 1) = \psi(\alpha) + 1/\alpha,$$

e um valor particular importante é

$$\psi\left(\tfrac{1}{2}\right) = -\epsilon - 2\log 2 = -1.96351\ 00260\ \ldots$$

Passemos agora ao cálculo da variância de e.o.'s em modelo Exponencial, mais uma vez recorrendo à representação de Rényi. A independência dos termos dessa representação permite-nos escrever

$$\begin{aligned}
\left(\sigma_i^E\right)^2 = \mathbb{V}ar\left[E_{i:n}\right] &= \sum_{j=1}^{i} \mathbb{V}ar\left\{\frac{E_j}{n-j+1}\right\} \\
&= \frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{(n-i+1)^2} \\
&= \psi'(n-i+1) - \psi'(n+1),
\end{aligned}$$

onde ψ' é a chamada *função trigama*, a derivada da *função digama* em (5.32).

Observação 5.3.2. *A função* trigama *é um caso particular das chamadas funções* poligama*, que são as derivadas sucessivas da função* Gama*,*

$$\psi^{(m)}(\alpha) = \tfrac{d^m}{d\alpha^m}\ \psi(\alpha) = \tfrac{d^{m+1}}{d\alpha^{m+1}} \log\Gamma(\alpha), \quad \alpha \in \mathbb{R}^+, \quad m = 1, 2, 3, \ldots. \quad (5.33)$$

Estas funções encontram-se mais uma vez extensivamente tabeladas no livro de Abramowitz & Stegun*, e estão fortemente relacionadas com uma outra função clássica da Análise, a função* zeta de Riemann, $\xi(m)$. *Para valores de* α *inteiros positivos, tem-se*

$$\psi^{(m)}(1) = (-1)^{m+1}\ n!\ \xi(m+1),$$

e

$$\psi^{(m)}(n+1) = (-1)^m\ m!\ \left[-\xi(m+1) + 1 + \frac{1}{2^{m+1}} + \cdots + \frac{1}{n^{m+1}}\right].$$

O valor de $\xi(m)$ *está relacionado com somas de potências* m *de recíprocos de inteiros. Mais especificamente*

$$\xi(m) = \sum_{k=1}^{\infty} \frac{1}{k^m}, \quad m > 1, \ \textit{inteiro.}$$

Alguns valores especiais são

$$\xi(2) = 1 + \tfrac{1}{2^2} + \tfrac{1}{3^2} + \cdots = \tfrac{\pi^2}{6}, \qquad \xi(4) = 1 + \tfrac{1}{2^4} + \tfrac{1}{3^4} + \cdots = \tfrac{\pi^4}{90}.$$

Também temos para a função poligama *uma relação de recorrência importante,*

$$\psi^{(n)}(\alpha+1) = \psi^{(n)}(\alpha) + (-1)^n\ n!\ \alpha^{-n-1},$$

e um valor particular importante, de argumento não inteiro, é

$$\psi^{(n)}\left(\tfrac{1}{2}\right) = (-1)^{n+1}\ n!\ (2^{n+1}-1)\ \xi(n+1), \quad n = 1, 2, \ldots$$

Passemos finalmente ao cálculo da covariância de e.o.'s em modelo Exponencial, mais uma vez recorrendo à representação de Rényi. Tem-se, para $1 \leq i \leq j \leq n$,

$$\begin{aligned} \sigma^E_{i,j} &= \mathbb{C}ov\left[E_{i:n}, E_{j:n}\right] = \mathbb{C}ov\Big(\sum_{k=1}^{i} \frac{E_k}{n-k+1}, \sum_{l=1}^{j} \frac{E_l}{n-l+1}\Big) \\ &= \mathbb{C}ov\Big(\sum_{k=1}^{i} \frac{E_k}{n-k+1}, \sum_{l=1}^{i} \frac{E_l}{n-l+1} + \sum_{l=i+1}^{j} \frac{E_l}{n-l+1}\Big). \end{aligned}$$

A bilinearidade da covariância e o facto de $E_{i+1}, \ldots, E_j$ serem independentes de $E_1, \ldots, E_i$, leva-nos a poder escrever, para $1 \leq i \leq j \leq n$,

$$\sigma^E_{i,j} = \mathbb{V}ar\Big(\sum_{k=1}^{i} \frac{E_k}{n-k+1}\Big) = \mathbb{V}ar\left(E_{i:n}\right) = \psi'(n-i+1) - \psi(n+1).$$

Cálculo directo

O cálculo directo recorre às *funções* poligama anteriormente referidas, em (5.33), e a funções derivadas da *função* Beta. Comecemos por relembrar que a f.d.p. de $E_{i:n}$ é

$$f^E_{i:n}(z) = \frac{1}{B(i, n-i+1)} \left(1 - e^{-z}\right)^{i-1} \; e^{-(n-i+1)z}, \quad z \geq 0,$$

que segue directamente de (5.1). Consequentemente, e usando a transformação $e^{-z} = t$, segue-se que

$$\begin{aligned} \mu^E_{i:n} = \int_0^\infty z \; f_{i:n}(z) dz &= -\tfrac{1}{B(i,n-i+1)} \int_0^1 \log t \; t^{n-i} \; (1-t)^{i-1} \; dt \\ &= -\tfrac{1}{B(i,n-i+1)} \left[\tfrac{\partial}{\partial \alpha} \; B(\alpha, \beta)\right]_{\alpha=n-i+1, \beta=i}. \end{aligned}$$

Precisamos agora de obter informação sobre $\frac{\partial}{\partial \alpha} \; B(\alpha, \beta)$. Tem-se

$$\begin{aligned} \tfrac{\partial}{\partial \alpha} \; B(\alpha, \beta) = \tfrac{\partial}{\partial \alpha} \; \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} &= \frac{\Gamma'(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta) - \Gamma(\alpha)\Gamma(\beta)\Gamma'(\alpha+\beta)}{\Gamma^2(\alpha+\beta)} \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)}\right] \\ &= B(\alpha, \beta) \left[\psi(\alpha) - \psi(\alpha+\beta)\right]. \end{aligned}$$

Voltemos pois ao cálculo do valor médio da i-ésima e.o. em modelo Exponencial,

$$\begin{aligned} \mu^E_{i:n} &= -\tfrac{1}{B(i,n-i+1)} \left[\tfrac{\partial}{\partial \alpha} \; B(\alpha, \beta)\right]_{\alpha=n-i+1, \beta=i} \\ &= -\tfrac{1}{B(i,n-i+1)} \; B(n-i+1, i) \left[\psi(n-i+1) - \psi(n+1)\right] \\ &= \psi(n+1) - \psi(n-i+1), \end{aligned}$$

tal como tínhamos obtido através da representação de Rényi.

Embora de forma mais pesada, do ponto de vista de cálculo, a obtenção da estrutura de segunda ordem das e.o.'s em modelo Exponencial pode também ser obtida de forma directa. Irão obviamente surgir segundas derivadas da função Beta, para as quais se podem obter expressões simples relacionadas agora com a função digama. Deixamos esses cálculos ao cuidado do leitor mais persistente.

5.3.6 Momentos em modelo Pareto

A estrutura distribucional de segunda ordem das e.o.'s em modelo Pareto pode ser facilmente obtida através dessa mesma estrutura em modelo Uniforme, uma vez que $P_{i:n} \stackrel{d}{=} U^{-1}_{n-i+1:n}$, $1 \leq i \leq n$.

5.4 Estrutura markoviana das estatísticas ordinais

5.4.1 Estatísticas ordinais e processo de Poisson

Consideremos uma amostra aleatória $Z_j, 1 \leq j \leq n$, proveniente de um modelo com f.d contínua F, estritamente crescente, e a correspondente amostra ordenada, $Z_{j:n}, 1 \leq j \leq n$, com $Z_{1:n} \leq \cdots \leq Z_{n:n}$.

Sem perda de generalidade, comecemos por fazer mais uma vez uma simplificação importante, induzindo nas nossas v.a.'s a transformação uniformizante, i.e., consideremos

$$U_i = F(Z_i) \ \frown \ \mathcal{U}(0,1), \ 1 \leq i \leq n.$$

Mediante transformação não decrescente, como é o caso em estudo, a relação de ordem é preservada e portanto tem-se

$$U_{i:n} = F(Z_{i:n}), \quad 1 \leq i \leq n,$$

e como já vimos (*cf.* (5.7)) ter-se-á a densidade conjunta

$$f_{U_{1:n},\ldots,U_{n:n}}(u_1,\ldots,u_n) = n! \quad \text{se} \quad 0 \leq u_1 < \cdots < u_n \leq 1.$$

No que se segue, iremos verificar que esta densidade é coincidente com a densidade de dimensão finita de um processo de Poisson, dado que se registaram n ocorrências no intervalo $[0,1]$.

A fim de estabelecer a relação entre as e.o's e o processo de Poisson $\{N_t\}_{t\geq 0}$, com $N(0)=0$, relembremos a seguinte propriedade do processo de Poisson: um processo de contagem $\{N_t\}_{t\geq 0}$ é um processo de Poisson de intensidade λ se e só se os intervalos de tempo entre ocorrências consecutivas desse processo forem v.a.'s independentes e identicamente distribuídas, exponenciais, com valor médio $1/\lambda$. Significa então que, num processo de Poisson, as variáveis $W_j = T_j - T_{j-1}$, $j \geq 1$ ($T_0 = 0$), com $\{T_j\}_{j\geq 1}$ instantes de ocorrência dos acontecimentos de Poisson, são independentes e exponenciais, conferindo um carácter Erlangiano às variáveis absolutas T_j, $j \geq 1$. No entanto, as variáveis $(T_1, \ldots, T_n)$ condicionais a ter n ocorrências de Poisson no intervalo de amplitude unitária, têm uma distribuição conjunta que coincide com a distribuição das n e.o.'s em modelo $\mathcal{U}(0,1)$, i.e.,

$$f_{T_1,\ldots,T_n|N_1=n}(t_1,\ldots,t_n) = n! \quad \text{se} \quad t_1 < t_2 < \cdots < t_n,$$

ou seja

$$U_{1:n}, U_{2:n}, \ldots, U_{n:n} \stackrel{d}{=} T_1, T_2, \ldots, T_n | N_1 = n. \tag{5.34}$$

Assim, podemos proceder a uma *identificação distribucional entre as e.o.'s em modelo* Uniforme *e a ocorrência condicionada de acontecimentos de Poisson.* Note-se que, não são os intervalos entre ocorrências, mas sim os instantes das ocorrências, que são completamente aleatórios e portanto identicamente distribuídos às e.o.'s de n uniformes em $[0,1]$. Esta identificação, a qual não iremos aprofundar neste momento, mas que pode ser consultada em Durbin[6] (1973) ou Gross *et al.*[7] (2008), permite frequentemente simplificar a derivação de várias propriedades distribucionais das e.o.'s e de estatísticas relacionadas. Um exemplo pertinente refere-se ao tratamento da f.d.e. $\{\widehat{F}_n(t)\}$, definida em (4.3), como um processo estocástico a tempo contínuo em $[0,1]$. Da identificação distribucional anterior resulta que o processo estocástico $\{\widehat{F}_n(t)\}_{t\in[0,1]}$

[6] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* CMBS-NSF 9, SIAM.

[7] Gross, D., Shortle, J.F., Thompson, J.M. & Harris, C. (2008). *Fundamentals of Queueing Theory.* 4th Ed., Wiley.

tem distribuição coincidente com a de um processo de Poisson $\{P_n(t)\}_{t\in[0,1]}$ com taxa de ocorrências n, condicional a $P_n(1) = 1$. Esta representação da f.d.e. foi utilizada no trabalho fundamental de Kolmogorov[8] (1933).

Mais geralmente, por simples 'mudança de relógio',

$$(T_1, \ldots, T_n | N_{t_0} = n) \stackrel{d}{=} (t_0 U_{1:n}, \ldots, t_0 U_{n:n}),$$

as n e.o.'s associadas a uma amostra de dimensão n de um modelo $\mathcal{U}(0, t_0)$.

5.4.2 Estatísticas ordinais como processo de Markov

Pensemos no vector condicional

$$(U_{1:n}, \ldots, U_{k-1:n} |\ U_{k:n} = u_k, \ldots, U_{n:n} = u_n). \tag{5.35}$$

A identificação anteriormente estudada, e o facto de o processo de Poisson ser um processo estocástico com incrementos independentes, permite-nos identificar distribucionalmente o vector aleatório em (5.35) com os $k-1$ instantes de chegada dos $k-1$ acontecimentos de Poisson que sabemos terem ocorrido em $(0, u_k)$, que podem por sua vez ser identificados distribucionalmente com as $k-1$ e.o.'s em modelo $\mathcal{U}(0, u_k)$, ou seja

$$\begin{aligned}
&(U_{1:n}, \ldots, U_{k-1:n} |\ U_{k:n} = u_k, \ldots, U_{n:n} = u_n) \\
&\stackrel{d}{=} (T_1, \ldots, T_{k-1} |\ N_{u_k} = k-1) \\
&\stackrel{d}{=} (\widetilde{Z}_{1:k-1}, \ldots, \widetilde{Z}_{k-1:k-1}), \text{ com } \left\{\widetilde{Z}_j\right\}_{j\geq 1} \text{ i.i.d. uniformes em } (0, u_k) \\
&\qquad\qquad \stackrel{d}{=} (U_{1:n}, \ldots, U_{k-1:n} |\ U_{k:n} = u_k),
\end{aligned}$$

i.e., as e.o.'s $(U_{n:n}, \ldots, U_{1:n})$ formam um processo de Markov. Mais ainda,

$$\begin{aligned}
&f_{U_{1:n},\ldots,U_{k-1:n}|U_{k:n},\ldots,U_{n:n}}(u_1, \ldots, u_{k-1}|u_k, \ldots, u_n) \\
&= f_{U_{1:n},\ldots,U_{k-1:n}|U_{k:n}}(u_1, \ldots, u_{k-1}|u_k) \\
&\qquad\qquad = \frac{(k-1)!}{u_k^{k-1}} \quad \text{se} \quad 0 \leq u_1 < \cdots < u_k \leq 1.
\end{aligned}$$

[8] Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari* **4**, 83–91.

De forma análoga

$$\begin{aligned}&(U_{k+1:n},\ldots,U_{n:n}|U_{1:n}=u_1,\ldots,U_{k:n}=u_k)\\&\overset{d}{=}(Z^*_{1:n-k},\ldots,Z^*_{n-k:n-k}),\text{ com }\left\{Z^*_j\right\}_{j\geq 1}\text{ i.i.d. uniformes em }(u_k,1)\\&\qquad\overset{d}{=}(U_{1:n},\ldots,U_{k:n}|U_{k:n}=u_k).\quad(5.36)\end{aligned}$$

As e.o.'s $(U_{1:n},\ldots,U_{n:n})$ constituem também um processo de Markov, ou seja, as e.o.'s possuem uma estrutura markoviana *reversível* no tempo — não só o futuro, condicional ao presente e ao passado depende apenas do presente, mas também o passado, condicional ao presente e ao futuro, depende unicamente do presente.

Da relação (5.36) segue-se que

$$\begin{aligned}&f_{U_{k+1:n},\ldots,U_{n:n}|U_{1:n},\ldots,U_{k:n}}(u_{k+1},\ldots,u_n|u_1,\ldots,u_k)\\&=f_{U_{k+1:n},\ldots,U_{n:n}|U_{k:n}}(u_{k+1},\ldots,u_n|u_k)=\frac{(n-k)!}{(1-u_k)^{n-k}},\\&\qquad\qquad\text{se}\quad u_k\leq u_{k+1}<\cdots<u_n\leq 1.\end{aligned}$$

Mais ainda

$$U_{1:n},\ldots,U_{k-1:n}|\ U_{k:n}\quad\text{é independente de}\quad U_{k+1:n},\ldots,U_{n:n}|\ U_{k:n},$$

i.e., dado $U_{k:n}$, $(U_{1:n},\ldots,U_{k-1:n})$ e $(U_{k+1:n},\ldots,U_{n:n})$ são condicionalmente independentes.

Mais geralmente, para $i<k$, dado o vector $(U_{i+1:n},\ldots,U_{k:n})$, os vectores $(U_{1:n},\ldots,U_{i:n})$ e $(U_{k+1:n},\ldots,U_{n:n})$ são condicionalmente independentes. Usando mais uma vez a propriedade markoviana das e.o.'s, obtem-se com facilidade a f.d.p. conjunta de um número qualquer de e.o.'s consecutivas (que também se obtinha com facilidade por raciocínio directo). Na realidade, para $i<k$, tem-se por um lado,

$$\begin{aligned}&f_{U_{1:n},\ldots,U_{i:n},U_{k+1:n},\ldots,U_{n:n}|U_{i+1:n},\ldots,U_{k:n}}(u_1,\ldots,u_i,u_{k+1},\ldots,u_n|u_{i+1},\ldots,u_k)\\&\qquad=\frac{f_{U_{1:n},\ldots,U_{n:n}}(u_1,\ldots,u_n)}{f_{U_{i+1:n},\ldots,U_{k:n}}(u_{i+1},\ldots,u_k)}.\quad(5.37)\end{aligned}$$

Mas como temos independência condicional de $(U_{1:n}, \ldots, U_{i:n})$ e $(U_{k+1:n}, \ldots, U_{n:n})$, dado $(U_{i+1:n}, \ldots, U_{k:n})$, e estrutura markoviana para as e.o.'s, podemos factorizar a f.d.p. condicional em (5.37), e escrever

$$\begin{aligned} &f_{U_{1:n},\ldots,U_{i:n},U_{k+1:n},\ldots,U_{n:n}|U_{i+1:n},\ldots,U_{k:n}}(u_1, \ldots, u_i, u_{k+1}, \ldots, u_n|u_{i+1}, \ldots, u_k) \\ &= f_{U_{1:n},\ldots,U_{i:n}|U_{i+1:n}}(u_1, \ldots, u_i|u_{i+1}) \times f_{U_{k+1:n},\ldots,U_{n:n}|U_{k:n}}(u_{k+1}, \ldots, u_n|u_k) \\ &= \frac{i!}{u_{i+1}^i} \frac{(n-k)!}{(1-u_k)^{n-k}}, \end{aligned}$$

se $0 \leq u_1 < \cdots < u_i \leq u_{i+1}$, e $u_k \leq u_{k+1} < \cdots < u_n \leq 1$, podendo finalmente obter-se

$$f(u_{i+1}, \ldots, u_k) = \frac{n!}{i!(n-k)!} u_{i+1}^i (1-u_k)^{n-k}, \ 0 \leq u_{i+1} < \cdots < u_k \leq 1.$$

Em particular,

$$f_{U_{i:n}}(u) = \frac{n!}{(i-1)!(n-i)!} \ u^{i-1}(1-u)^{n-i}, \quad 0 \leq u \leq 1,$$

i.e. a i-ésima e.o. em modelo Uniforme tem uma distribuição Beta com parâmetros i e $n-i+1$, como já vimos anteriormente.

Em geral

$$f_{Z_{i:n}}(z) = \frac{n!}{(i-1)!(n-i)!} \ F^{i-1}(z)(1-F(z))^{n-i}, \quad z \in \mathbb{R},$$

é a chamada Beta transformada, como também já vimos.

Distribuição condicional

A propriedade markoviana das e.o.'s permite-nos garantir facilmente que a distribuição condicional de $Z_{j:n}|Z_{i:n} = x$, para $1 \leq i < j \leq n$, é a da $(j-i)$-ésima e.o. de uma amostra de dimensão $n-i$ de um modelo truncado, $Y = Z|Z \geq x$, com f.d.p. dada por

$$h(y) = \frac{f(y)}{1-F(x)}, \quad y \geq x, \tag{5.38}$$

e consequentemente com f.d.

$$H(y) = \frac{F(y)-F(x)}{1-F(x)}, \quad y \geq x,$$

i.e., para $1 \leq i < j \leq n$,

$$F_{Z_{j:n}|Z_{i:n}=x}(y|x) = \frac{(n-i)!}{(j-i-1)!(n-j)!} \frac{[F(y)-F(x)]^{j-i-1}\,[1-F(y)]^{n-j}}{[1-F(x)]^{n-i}}\, f(y), \quad y \geq x. \quad (5.39)$$

Na realidade, a identificação distribucional entre as e.o.'s em modelo Uniforme e a ocorrência condicionada de acontecimentos de Poisson permite-nos escrever

$$(U_{i+1:n}, \ldots, U_{n:n} |\, U_{i:n} = x) \stackrel{d}{=} (\widetilde{Z}_{1:n-i}, \ldots, \widetilde{Z}_{n-i:n-i}),$$

com $\left\{\widetilde{Z}_j\right\}_{j\geq 1}$ i.i.d. e $\mathcal{U}(x,1)$.

Consequentemente, do facto de se ter $Z_{i:n} = F^{\leftarrow}(U_{i:n})$, $1 \leq i \leq n$, obtém-se facilmente

$$(Z_{i+1:n}, \ldots, Z_{n:n} |\, Z_{i:n} = x) \stackrel{d}{=} (\widetilde{Z}^*_{1:n-i}, \ldots, \widetilde{Z}^*_{n-i:n-i}),$$

com $\left\{\widetilde{Z}^*_j\right\}_{j\geq 1}$ i.i.d. com f.d.p. truncada h em (5.38). Segue-se pois (5.39).

5.4.3 Uma cadeia de Markov aditiva

Relembremos a representação de Rényi para as e.o.'s em modelo Exponencial, dada em (5.9), e que nos permite escrever

$$E_{k:n} = \frac{E_1}{n} + \frac{E_2}{n-1} + \cdots + \frac{E_k}{n-k+1}, \quad 1 \leq k \leq n,$$

com $\{E_j\}_{j\geq 1}$ v.a.'s i.i.d. exponenciais unitárias. Os resultados derivados nos parágrafos anteriores e a representação de Rényi, permitem-nos pois garantir que as e.o.'s em modelo Exponencial formam um *processo de Markov aditivo*.

Pensemos em seguida no caso geral da amostra $(Z_1, \ldots, Z_n)$ ser proveniente de um modelo F, com F contínua e estritamente crescente. Ao induzir em Z a transformação $-\log F$ obtemos uma amostra Exponencial unitária, i.e.

$$(E_1, \ldots, E_n) \stackrel{d}{=} (-\log F(Z_1), \ldots, -\log F(Z_n)),$$

e como a transformação $-\log F$ é decrescente, tem-se $-\log F(Z_{k:n}) = E_{n-k+1:n}$, $1 \leq k \leq n$, e consequentemente,

$$Z_{k:n} = F^{\leftarrow}\left\{e^{-Z_{n-k+1:n}}\right\} = F^{\leftarrow}\left\{e^{-\left(\frac{E_1}{n}+\cdots+\frac{E_{n-k+1}}{k}\right)}\right\}.$$

De modo análogo podemos escrever

$$\begin{aligned} Z_{n-k:n} &= F^{\leftarrow}\left\{e^{-E_{k+1:n}}\right\} = F^{\leftarrow}\left\{e^{-\left(\frac{E_1}{n}+\cdots+\frac{E_{k+1}}{n-k}\right)}\right\} \\ &= F^{\leftarrow}\left\{e^{-\frac{E_{k+1}}{n-k}}\ F(Z_{n-k+1:n})\right\} \\ &= F^{\leftarrow}\left\{e^{\log F(Z_{n-k+1:n})-\frac{E_{k+1}}{n-k}}\right\}, \end{aligned} \tag{5.40}$$

o que realça mais uma vez o o facto de $Z_{n:n},\ldots,Z_{1:n}$ ser um processo de Markov, o mesmo acontecendo, como já se viu atrás, a $(Z_{1:n},\ldots,Z_{n:n})$.

Mais adiante recorreremos a esta mesma linha de raciocínio para derivação de vários resultados assintóticos.

Observação 5.4.1. *Um problema de índole geral, colocado por* Rényi*, e que tanto quanto sabemos continua em aberto, é o seguinte*: *dado um processo de* Markov $\{X_t\}_{t\in T}$*, para que funções* $G_t(\cdot)$ *se tem ainda a propriedade markoviana para* $\{Y_t = G_t(X_t)\}_{t\in T}$*? Esperamos que algum dos leitores mais interessado possa dar uma resposta a esta questão.*

5.5 Estatísticas sistemáticas: Espaçamentos e Amplitude

Definição 5.5.1. *Estatísticas sistemáticas são estatísticas que são combinações lineares de e.o.'s.*

A distribuição destas estatísticas pode ser complicada, e voltaremos a este assunto mais adiante. Nesta Secção iremos considerar alguns casos simples, como os já considerados na Secção 5.2.2 para o modelo Exponencial, do tipo

$$S_{ij} := X_{j:n} - X_{i:n},\ 1 \le i < j \le n. \tag{5.41}$$

Um caso particular é o da *amplitude amostral*, que se obtém para $i = 1$ e $j = n$. Na realidade, uma das estatísticas sistemáticas mais importantes em várias áreas da *Estatística Teórica* e *Aplicada*, e muito particularmente na área de *Controlo de Qualidade*, é a *amplitude amostral*,

$$R_n := S_{1n} = X_{n:n} - X_{1:n},$$

com S_{ij} definido em (5.41).

5.5.1 Distribuição de amostragem da amplitude e estatísticas similares

A distribuição de amostragem de R_n vai depender fortemente de F, mas pode facilmente ser escrita em forma integral:

$$f_{R_n}(r) = n(n-1)\int_{\mathbb{R}} f(x)[F(x+r) - F(x)]^{n-2} f(x+r)dx, \tag{5.42}$$

ou, equivalentemente,

$$F_{R_n}(r) = n\int_{\mathbb{R}} f(x)[F(x+r) - F(x)]^{n-1} dx, \tag{5.43}$$

expressão que deve ser interpretada do modo seguinte: dado x, que pode variar na recta real $\mathbb{R}$, a função integranda é a probabilidade de um dos XX's, arbitrariamente de entre n, cair em $(x - dx, x]$ e os restantes $n-1$ cairem em $(x, x+r]$.

Exemplo 5.5.1. Amplitude em modelo Exponencial. *Em modelo* Exponencial *unitário, a utilização de* (5.43) *leva-nos a*

$$\begin{aligned} F_{E_{n:n}-E_{1:n}}(r) &= n\int_0^\infty e^{-x}\left(e^{-x} - e^{-(x+r)}\right)^{n-1} dx \\ &= n\left(1-e^{-r}\right)^{n-1}\int_0^\infty e^{-nx}dx = \left(1-e^{-r}\right)^{n-1}, \end{aligned}$$

sempre que $r > 0$, *i.e.,*

$$R_n = E_{n:n} - E_{1:n} \stackrel{d}{=} E_{n-1:n-1}. \tag{5.44}$$

Note-se que o resultado obtido em (5.44) é o esperado desde que repensemos na abordagem de Rényi para a derivação do comportamento distribucional dos espaçamentos em modelo Exponencial. Na realidade, interpretando E_i, $1 \leq i \leq n$, como os tempos de vida exponenciais de n unidades independentes, a falta de memória da Exponencial leva a que tudo recomece de novo em $E_{1:n}$, mas com $n-1$ unidades em prova. Consequentemente, o tempo de espera até à avaria de todas unidades, dado por $E_{n:n} - E_{1:n}$, deve ser distribucionalmente equivalente ao máximo de $n-1$ exponencias, i.e., a $E_{n-1:n-1}$.

Mais geralmente, argumentos probabilísticos directos levam-nos a poder escrever com facilidade a f.d.p. de S_{ij}, definida em (5.41). Para se ter S_{ij} em $[s-ds, s+ds)$, das n observações na amostra temos de ter $i-1$ arbitrárias inferiores a um x, também arbitrário, que pode variar em $\mathbb{R}$, uma arbitrária terá de estar em $[x, x+dx)$, $j-i-1$, também arbitrariamente de entre as $n-i$ restantes terão de estar entre $x+dx$ e $x+s-ds$, uma terá de estar num intervalo infinitésimo centrado em $x+s$ e as restantes $n-j$ terão de ser superiores a $x+s+ds$. Em suma, teremos

$$(i-1)!(j-i-1)!(n-j)!f_{S_{ij}}(s)/n! = \int_{\mathbb{R}} F^{i-1}(x)\left[F(x+s)-F(x)\right]^{j-i-1}\left[1-F(x+s)\right]^{n-j} f(x)f(x+s)dx \tag{5.45}$$

Como caso particular de (5.45), para $i=1$ e $j=n$, obtemos (5.42).

5.5.2 A Amplitude como Estimador de um Parâmetro de Escala.

Admitamos agora que $(X_1, \ldots, X_n)$ é uma amostra aleatória proveniente de uma população com f.d. $F((x-\lambda)/\delta)$, $\lambda \in \mathbb{R}$ e $\delta \in \mathbb{R}^+$ parâmetros desconhecidos de localização e escala, respectivamente. Consideremos as e.o.'s ascendentes $(X_{1:n} \leq \cdots \leq X_{n:n})$ e designemos, mais uma vez, por $Z_{i:n} = (X_{i:n}-\lambda)/\delta$ as e.o.'s correspondentes ao modelo F com localização 0 e dispersão 1.

A amplitude $R_n = X_{n:n} - X_{1:n}$ fornece uma medida de dispersão simples, que é facilmente convertida num estimador centrado de δ. Na realidade

$$\mathbb{E}[R_n] = \mathbb{E}(X_{n:n}) - \mathbb{E}(X_{1:n}) = \delta\{\mu_{n:n} - \mu_{1:n}\},$$

e

$$\mathbb{V}ar[R_n] = \delta^2\left[\sigma_{n,n:n} + \sigma_{1,1:n} - 2\sigma_{1,n:n}\right],$$

com $\mu_{i:n} = \mathbb{E}[Z_{i:n}]$, $1 \leq i \leq n$, e $\sigma_{i,j:n} = \mathbb{C}ov(Z_{i:n}, Z_{j:n})$, $1 \leq i < j \leq n$, independentes de parâmetros desconhecidos.

Tem-se então que

$$T_n = R_n/(\mu_{n:n} - \mu_{1:n}) \tag{5.46}$$

é um estimador centrado de δ, consistente numa classe vasta de modelos subjacentes. Este estimador tem usualmente eficiência elevada para amostras de dimensão pequena, digamos para $n \leq 12$, e para uma grande variedade de modelos, incluindo o modelo Normal. Trata-se, além disso, de um estimador que, para pequenas amostras, é mais 'robusto' ou 'resistente', no sentido de ser menos sensível a mudanças no modelo subjacente aos dados, que o desvio padrão empírico. As propriedades já anteriormente apontadas, aliadas ao facto do cálculo de R_n ser imediato para amostras de dimensão pequena, usuais na área de *Controlo de Qualidade* em linha de produção, onde os subgrupos racionais recolhidos ao longo do tempo têm usualmente dimensão $n = 5$ (veja-se, por exemplo, Montgomery[9], 1991 ou Gomes *et al.*, 2010), levam a que o estimador T_n em (5.46) seja preferido relativamente ao desvio padrão empírico, na estimação de um parâmetro de escala δ.

Para podermos utilizar T_n, como estimador de δ precisamos de saber calcular, ou ter acesso a tabelas de

$$d_{2,n} = \mu_{n:n} - \mu_{1:n},$$

que, como já vimos, são valores que dependem intrinsecamente do modelo F, e da estrutura probabilística das e.o.'s. Para alguns modelos, de entre os usuais em aplicações, é possível obter expressões explícitas para $\mu_{i:n}$. Entre esses encontram-se os modelos seguintes:

$$\begin{aligned}
\text{Modelo Uniforme:} \quad \mu_{i:n} &= \frac{i}{n+1},\ 1 \leq i \leq n. \\
\text{Modelo Exponencial:} \quad \mu_{i:n} &= \sum_{j=1}^{i} \frac{1}{n-j+1} \\
&= \psi(n+1) - \psi(n-i+1),\ 1 \leq i \leq n. \\
\text{Modelo Logístico:} \quad \mu_{i:n} &= \psi(i) - \psi(n-i+1),\ 1 \leq i \leq n,
\end{aligned}$$

onde $\psi(\cdot)$ denota novamente a função *digama*, definida em (5.32). Por exemplo, para $n = 5$, dimensão usual em Controlo de Qualidade,

[9] Montgomery, D.C. (1991). *Introduction to Statistical Quality Control.* 2nd Ed. Wiley.

Modelo	$\mu_{1:5}$	$\mu_{5:5}$	$d_{2,5}$
Exponencial	0.200	2.28(3)	2.08(3)
Normal	-1.163	1.163	2.326
Gama(2)	0.702	3.808	3.106
Gama(5)	2.722	7.803	5.081

5.5.3 Espaçamentos de estatísticas ordinais

Iremos agora aflorar o estudo dos espaçamentos (*spacings*) de e.o.'s, no caso geral, de grande importância em grande parte da metodologia estatística de índole não-paramétrica. Comecemos pela situação de uma amostra aleatória proveniente de modelo $\mathcal{U}(0,1)$, $(U_1,\ldots,U_n)$. Consideremos a amostra de e.o.'s ascendentes associada, $(U_{1:n},\ldots,U_{n:n})$, e os espaçamentos ou *spacings*,

$$S_j := U_{j:n} - U_{j-1:n}, \quad 1 \le j \le n+1 \quad (U_{0:n} \equiv 0,\ U_{n+1:n} \equiv 1), \quad (5.47)$$

que constituem aquilo a que muito frequentemente em Estatística se chama a *divisão aleatória do intervalo*.

As v.a.'s $(S_1,\ldots,S_{n+1})$ são obviamente não independentes, pois estão sujeita a uma ligação forte, $\sum_{i=1}^{n+1} S_i = 1$. Se pensarmos unicamente em $(S_1,\ldots,S_n)$, como a f.d.p. conjunta de $(U_{1:n},\ldots,U_{n:n})$ é $n!$ na região $0 \le u_1 < \cdots < u_n \le 1$, e o Jacobiano da transformação em (5.47) é unitário, tem-se

$$g_{S_1,\ldots,S_n}(s_1,\ldots,s_n) = n!, \quad s_i \ge 0, \quad 1 \le i \le n, \quad \sum_{i=1}^{n} s_i \le 1,$$

i.e., $(S_1,\ldots,S_n)$ estão uniformemente distribuídos na região

$$s_i \ge 0, \quad 1 \le i \le n, \quad \sum_{i=1}^{n} s_i \le 1,$$

o que determina a distribuição das v.a.'s $(S_1,\ldots,S_{n+1})$ na região

$$s_i \ge 0, \quad 1 \le i \le n+1, \quad \sum_{i=1}^{n+1} s_i = 1,$$

distribuição *completamente simétrica* nos s_i, $1 \le i \le n+1$. O ser completamente simétrica significa que quaiquer que sejam os k inteiros distintos,

$1 \leq n_1, n_2, \ldots, n_k \leq n+1$,

$$(S_{n_1}, \ldots, S_{n_k}) \stackrel{d}{=} (S_1, \ldots, S_k).$$

Este resultado é interessante, no sentido de poder simplificar cálculos. Tem-se por exemplo que

$$\forall\ 1 \leq n_1, n_2, \ldots, n_k \leq n+1,\ (S_{n_1} + \cdots + S_{n_k}) \stackrel{d}{=} (S_1 + \cdots + S_k) = U_{k:n},$$

que se sabe ter distribuição Beta$(k, n-k+1)$.

Coberturas elementares e intervalos de confiança para quantis

Definição 5.5.2. *As v.a.'s* $S_j = F(X_{j:n}) - F(X_{j-1:n})$, $1 \leq j \leq n$, $(F(X_{0:n}) \equiv 0)$, *são frequentemente designadas coberturas elementares.*

Esta terminoloia foi introduzida em Wilks[10] (1948) e gozam, como se disse logo de início, de papel importantíssimo no desenvolvimento de modelos não paramétricos em Estatística.

Um par de e.o.'s $(X_{r:n}, X_{s:n})$, $r < s$, pode fornecer com relativa facilidade um intervalo de confiança a $100 \times (1-\alpha)\%$ para qualquer quantil teórico $\chi_p = F^{\leftarrow}(p)$. Na realidade,

$$\begin{aligned}
\mathbb{P}\left(\chi_p \in (X_{r:n}, X_{s:n})\right) &= \mathbb{P}\left(F(\chi_p) \in (F(X_{r:n}), F(X_{s:n}))\right) \\
&= \mathbb{P}(U_{r:n} < p < U_{s:n}) \\
&= \sum_{i=r}^{s-1} \binom{n}{i} p^i\ (1-p)^{n-i} =: \pi(r, s; n, p).
\end{aligned}$$

Se utilizarmos a aproximação da Binomial pela Normal e fizermos igual a $(1-\alpha)$ o valor aproximado de $\pi(r, s; n, p)$, obtemos como intervalo de confiança a $100 \times (1-\alpha)\%$ para χ_p, o intervalo $(X_{r:n}, X_{s:n})$, com

$$r = \left\lfloor np - \sqrt{np(1-p)}\ z_{1-\alpha/2} + \tfrac{1}{2} \right\rfloor \quad s = \left\lfloor np + \sqrt{np(1-p)}\ z_{1-\alpha/2} + \tfrac{1}{2} \right\rfloor,$$

onde $\lfloor z \rfloor$ denota a parte inteira de z, sendo $z_{1-\alpha/2}$ o quantil de probabilidade $1-\alpha/2$ de uma $\mathcal{N}(0,1)$.

[10] Wilks, S.S. (1948). Order Statistics. *Bull. Amer. Math. Soc.* **54**:1, Part 1 , 6–50.

A distribuição de Dirichlet

A distribuição de $(S_1, \ldots, S_n)$ é um caso particular da célebre distribuição de Dirichlet, $D_{\nu_1,\ldots,\nu_{n+1}}$, quando os parâmetros são todos unitários, i.e., $(\nu_1, \ldots, \nu_{n+1}) = (1, \ldots, 1)$.

A f.d.p. de Dirichlet é uma genaralização ao caso multivariado de uma Beta(ν_1, ν_2), e tem a expressão funcional,

$$d_{\nu_1,\ldots,\nu_{n+1}}(x_1, \ldots, x_n) = \frac{\Gamma(\nu_1 + \cdots + \nu_{n+1})}{\Gamma(\nu_1) \ldots \Gamma(\nu_{n+1})}\ x_1^{\nu_1 - 1} \ldots x_n^{\nu_n - 1} \Big(1 - \sum_{i=1}^{n} x_i\Big)^{\nu_{n+1} - 1},$$

em qualquer ponto do simplex

$$\mathcal{S}_n = \Big\{(x_1, \ldots, x_n) : x_i \geq 0,\ 1 \leq i \leq n,\ \sum_{i=1}^{n} x_i \leq 1\Big\}.$$

Realce-se o facto de a f.d.p. de Dirichlet ser uma f.d.p. básica não só na teoria distribucional das e.o.'s, mas também em campos diversos de *Estatística Multivariada*.

Um método de transformação interessante, útil na geração de NPA's com f.d.p. de Dirichlet, e de demonstração imediata, é o seguinte:

Proposição 5.5.1. *Se $Y_1, \ldots, Y_{n+1}$ forem v.a.'s independentes, e se Y_j for proveniente de um modelo* Gama(ν_j), $1 \leq j \leq n+1$, *então*

$$X_i = \frac{Y_i}{\sum_{j=1}^{n+1} Y_j},\ 1 \leq i \leq n,$$

tem distribuição Dirichlet$(\nu_1, \nu_2, \ldots, \nu_{n+1})$.

Face ao resultado anterior podemos escrever os espaçamentos S_j em (5.47) em termos de v.a.'s exponenciais unitárias, $\{E_j\}_{j \geq 1}$:

$$(S_1, \ldots, S_{n+1}) \stackrel{d}{=} \left(\frac{E_1}{\sum_{j=1}^{n+1} E_j}, \ldots, \frac{E_{n+1}}{\sum_{j=1}^{n+1} E_j}\right).$$

Note-se que este mesmo resultado pode ser obtido por associações com a ocorrência condicionada de acontecimentos de Poisson, na linha do estudo desenvolvido na Secção 5.4.

5.5.4 O método de Steutel

Com base na identificação entre as e.o.'s e a ocorrência condicionada de acontecimentos de Poisson em (5.34),

$$U_{1:n}, \ldots, U_{n:n} \stackrel{d}{=} T_1, \ldots, T_n | N_1 = n.$$

Steutel[11] (1967) derivou o resultado seguinte:

Considerem-se v.a.'s U_j, $\mathcal{U}(0,1)$, e efectue-se a mudança de relógio, $U_j(t) = t \times U_j$, $1 \leq j \leq n+1$. Seja $f(\cdot, \ldots, \cdot)$ uma função Borel-mensurável tal que $\int_0^\infty \mathbb{E}\left[f(U_1(t), \ldots, U_{n+1}(t))\right] t^n e^{-\lambda t} dt < \infty$, $\forall \lambda > 0$. Então

$$\int_0^\infty \mathbb{E}\left[f(U_1(t), \ldots, U_{n+1}(t))\right] t^n e^{-\lambda t} dt = \frac{n!}{\lambda^{n+1}} \mathbb{E}\left[f(E_1, \ldots, E_{n+1})\right] \quad (5.48)$$

com $\{E_i\}_{i \geq 1}$ sucessão de v.a.'s i.i.d. $\mathcal{E}(\lambda)$. A relação (5.48) permite-nos obter probabilidades de vários acontecimentos com interesse, expressas em termos de $\mathbb{E}\left[f(U_1(t), \ldots, U_{n+1}(t)\right]$, por simples inversão de uma transformada de Laplace.

Sem entramos em detalhes de demonstração do argumento de Steutel, vejamos um exemplo de aplicação do referido argumento, com o cálculo da probabilidade, $\mathbb{P}\left[S_1 > x_1, \ldots, S_{n+1} > x_{n+1}\right]$. A probabilidade de um acontecimento A é sempre o valor médio da v.a. indicatriz de A, i.e.,

$$\mathbb{P}\left[S_1 > x_1, \ldots, S_{n+1} > x_{n+1}\right] = \mathbb{E}\left[f(U_1(t), \ldots, U_{n+1}(t)\right],$$

com $f(y_1, \ldots, y_{n+1}) = \begin{cases} 1 & y_1 > x_1, \ldots, y_{n+1} > x_{n+1} \\ 0 & \text{caso contrário.} \end{cases}$

Podemos pois escrever

$$\mathbb{P}\left[S_1 > x_1, \ldots, S_{n+1} > x_{n+1}\right] = \mathbb{E}\Big\{ \prod_{j=1}^{n+1} \left(1 - I(x_j - U_j(t))\right) \Big\},$$

com $I(t) = \begin{cases} 1 & \text{se} \quad t \geq 0 \\ 0 & \text{se} \quad t < 0 \end{cases}$, e consequentemente,

[11] Steutel, F.W. (1967). Random division of an interval. *Statistica Neerlandica* **21**, 231–244.

$$\int_0^\infty \mathbb{P}\left[U_1(t) > x_1, \ldots, U_n(t) > x_n, U_{n+1}(t) > x_{n+1}\right]\ t^n\ e^{-\lambda t} dt$$
$$= \int_0^\infty \mathbb{E}\Big\{ \prod_{j=1}^{n+1} \left(1 - I(x_j - U_j(t))\right) \Big\}\ t^n\ e^{-\lambda t} dt$$
$$= \frac{n!}{\lambda^{n+1}}\ \mathbb{E}\Big\{ \prod_{j=1}^{n+1} \left(1 - I(x_j - E_j)\right) \Big\} = \frac{n!}{\lambda^{n+1}}\ e^{-\lambda(x_1 + \cdots + x_{n+1})},\ \sum_{j=1}^{n+1} x_j \leq t.$$

Por inversão da transformada de Laplace obtém-se então

$$\mathbb{P}\left[U_1(t) > x_1, \ldots, U_n(t) > x_n, U_{n+1}(t) > x_{n+1}\right]\ t^n = \Big(t - \sum_{j=1}^{n+1} x_j\Big)^n,$$

e consequentemente

$$\mathbb{P}\left[U_1 > x_1, \ldots, U_n > x_n, U_{n+1} > x_{n+1}\right] = \Big(1 - \sum_{j=1}^{n+1} x_j\Big)^n, \quad \sum_{j=1}^{n+1} x_j \leq 1.$$

O mesmo processo serve obviamente para obter a f.d. conjunta de $(U_1(t), \ldots, U_{n+1}(t))$, i.e., $\mathbb{P}\left[U_1(t) \leq x_1, \ldots, U_{n+1}(t) \leq x_{n+1}\right]$, embora a expressão seja muito mais complicada.

Outro exemplo: Cálculo de $\mathbb{P}\left[S_{n-k+1:n} \leq z\right]$. Tem-se

$$\int_0^\infty \mathbb{P}\left[S_{n-k+1:n} \leq z\right]\ t^n e^{-\lambda t} dt = \int_0^\infty \mathbb{E}\left\{I(z - U_{n-k+1:n}(t)\right\}\ t^n e^{-\lambda t} dt$$
$$= \frac{n!}{\lambda^{n+1}} \mathbb{P}\left[E_{n-k+1:n} \leq z\right] = \frac{n!}{\lambda^{n+1}} \sum_{i=n-k+1}^{n+1} \binom{n+1}{i} F^i(z)(1 - F(z))^{n+1-i}$$
$$= \frac{n!}{\lambda^{n+1}} \sum_{j=0}^{k} \binom{n+1}{j} e^{-\lambda j z} \left(1 - e^{-\lambda z}\right)^{n+1-j}.$$

Como $\frac{1}{\Gamma(p)} \int_0^\infty (t-\alpha)^{p-1} I(t-\alpha) e^{-\lambda t} dt = \frac{e^{-\lambda\alpha}}{\lambda^p}$, $\alpha \geq 0$, $p \geq 0$, segue-se que

$$\mathbb{P}\left[S_{n-k+1:n} \leq z\right] = \sum_{j=0}^{k} \sum_{l=0}^{n+1-j} \binom{n+1}{j} \binom{n+1-j}{l} (-1)^l \left\{1 - (j+l)\frac{z}{t}\right\}^n I(t-(j+l)z). \tag{5.49}$$

O resultado (5.49), para $k = 1$, foi obtido por Fisher[12] (1929), através de argumentos geométricos demorados.

5.6 Enquadramentos e aproximações para momentos de estatísticas ordinais

As dificuldades inerentes ao cálculo de momentos de e.o.'s levam-nos à pesquisa de (e ao eventual recurso a) enquadramentos, bem como aproximações para os momentos de e.o.'s. Começaremos por ver enquadramentos independentes do modelo subjacente F, passando em seguida a aproximações para os momentos em termos da inversa de F e das suas derivadas.

5.6.1 Enquadramentos 'distribution-free'

Comecemos por trabalhar com o valor máximo $X_{n:n}$ associado a uma amostra aleatória $(X_1, \ldots, X_n)$ proveniente de uma população com f.d. F, contínua e estritamente crescente. Como vimos anteriormente, podemos escrever

$$\mathbb{E}[X_{n:n}] = n \int_0^1 F^{\leftarrow}(u)\ u^{n-1} du. \tag{5.50}$$

Admitamos, sem perda de generalidade desde que existam segundos momentos finitos, que $\mathbb{E}(X) = 0$ e $\mathbb{V}ar(X) = 1$, i.e.,

$$\int_0^1 F^{\leftarrow}(u) du = 0, \qquad \int_0^1 \left(F^{\leftarrow}(u)\right)^2 du = 1. \tag{5.51}$$

[12] Fisher, R.A. (1929). Tests of Significance in Harmonic Analysis. *Proc. Royal Statist. Soc. A* **125**:796, 54–59.

O cálculo das variações leva-nos facilmente aos valores estacionários de (5.50) condicionados à validade de (5.51) e tem-se

Teorema 5.6.1. *Em modelo genérico F, com segundos momentos finitos, e com $\mu = \mathbb{E}(X)$, $\sigma^2 = \mathbb{V}ar(X)$, tem-se*

$$\mathbb{E}\left[X_{n:n}\right] \leq \mu + \sigma \frac{n-1}{\sqrt{2n-1}}, \quad \mathbb{E}\left[X_{1:n}\right] \geq \mu - \sigma \frac{n-1}{\sqrt{2n-1}}. \tag{5.52}$$

No caso particular de modelo F simétrico, podemos melhorar o resultado anterior, e escrever

$$\mathbb{E}\left[X_{n:n}\right] \leq \mu + \sigma \frac{n}{2} \sqrt{\frac{2\left[1 - 1/\binom{2n-2}{n-1}\right]}{2n-1}},$$

$$\mathbb{E}\left[X_{1:n}\right] \geq \mu - \sigma \frac{n}{2} \sqrt{\frac{2\left[1 - 1/\binom{2n-2}{n-1}\right]}{2n-1}}.$$

Demonstração. Na realidade, ao procurarmos os valores estacionários de (5.50) condicionados à validade de (5.51) somos levados à consideração do funcional

$$H\left(F^{\leftarrow}(u), \alpha, \beta\right) = \int_0^1 \left[nF^{\leftarrow}(u)\ u^{n-1} - \alpha F^{\leftarrow}(u) - \beta\left((F^{\leftarrow}(u))^2 - 1\right)\right] du,$$

cuja solução estacionária é dada por

$$n\ u^{n-1} - \alpha - 2\beta\ F^{\leftarrow}(u) = 0 \quad \text{sse} \quad F^{\leftarrow}(u) = \frac{n\ u^{n-1} - \alpha}{2\beta},$$

com α e β tais que $\int_0^1 (n\ u^{n-1} - \alpha) du = 0$, ou seja $\alpha = 1$, e ainda $\beta^2 = \int_0^1 (n\ u^{n-1} - 1)^2 du$, ou seja $\beta = (n-1)/\sqrt{2n-1}$.

Tem-se pois

$$F^{\leftarrow}(u) = \frac{\sqrt{2n-1}\left(n\ u^{n-1} - 1\right)}{2(n-1)}, \tag{5.53}$$

que, pela desigualdade de Schwarz

$$\int f g\, du \leq \Big(\int f^2 du \int g^2 du \Big)^{1/2},$$

aplicada a $f(u) = F^{\leftarrow}(u)$ e $g(u) = n\ u^{n-1} - 1$, é a função que fornece mesmo um máximo para $\mathbb{E}[X_{n:n}]$, o qual é então dado por $\mathbb{E}[X_{n:n}] = (n-1)/\sqrt{2n-1}$. Tem-se então

$$\mathbb{E}[X_{n:n}] \leq \frac{n-1}{\sqrt{2n-1}}. \tag{5.54}$$

Se $\mathbb{E}(X) = \mu$ e $\mathbb{V}ar(X) = \sigma^2$ tem-se obviamente a primeira desigualdade em (5.52). A segunda desigualdade obtem-se facilmente, atendendo ao facto de se ter $\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$.
Se F for simétrica, podemos na realidade melhorar os valores anteriormente obtidos, devido ao facto de, por termos em modelo standardizado ($\mu = 0,\ \sigma = 1$) $F(x) = 1 - F(-x)$, podermos restringir o integral que nos fornece $\mathbb{E}[X_{n:n}]$ à região $(1/2, 1)$. Na realidade

$$\begin{aligned}\mathbb{E}[X_{n:n}] &= n \int_0^{\infty} x \left\{[F(x)]^{n-1} - [1-F(x)]^{n-1}\right\} dF(x) \\ &= n \int_{1/2}^{1} F^{\leftarrow}(u) \left\{u^{n-1} - (1-u)^{n-1}\right\} du,\end{aligned}$$

o que fornece o ponto de estacionaridade

$$F^{\leftarrow}(u) = \left\{\frac{2n-1}{2\left[1 - 1/\binom{2n-2}{n-1}\right]}\right\}^{1/2} \left[u^{n-1} - (1-u)^{n-1}\right],$$

e consequentemente

$$\mathbb{E}[X_{n:n}] \leq \frac{n}{2}\left\{\frac{2\left[1 - 1/\binom{2n-2}{n-1}\right]}{2n-1}\right\}^{1/2}.$$

□

Corolário 5.6.1. *A igualdade em* (5.54) *ocorre quando estamos a trabalhar em modelo com f.d.*

$$F(x) = \left\{\frac{1 + (n-1)\ x/\sqrt{2n-1}}{n}\right\}^{1/(n-1)}, \qquad -\frac{\sqrt{2n-1}}{n-1} \leq x \leq \sqrt{2n-1},$$

ao qual corresponde a f.d.p.

$$f(x) = \frac{1}{n\sqrt{2n-1}} \Big\{ \frac{1 + (n-1)\ x/\sqrt{2n-1}}{n} \Big\}^{1/(n-1)-1}, \\ -\frac{\sqrt{2n-1}}{n-1} \leq x \leq \sqrt{2n-1}.$$

Demonstração. Basta inverter a função quantil dada em (5.53). □

Corolário 5.6.2. *Para a amplitude amostral, $R_n = X_{n:n} - X_{1:n}$ tem-se*

$$\mathbb{E}[R_n] \leq n \Big\{ \frac{2\big[1 - 1/\binom{2n-2}{n-1}\big]}{2n-1} \Big\}^{1/2}.$$

Demonstração. Como se pode escrever

$$\mathbb{E}[R_n] = \mathbb{E}[X_{n:n}] - \mathbb{E}[X_{1:n}] = n \int_0^1 F^{\leftarrow}(u) \left[u^{n-1} - (1-u)^{n-1}\right] du,$$

o resultado segue por analogia com o que se fez para $\mathbb{E}[X_{n:n}]$ em modelo F simétrico em torno de 0. □

Resultados semelhantes podem ser derivados para $\mathbb{E}[X_{i:n}]$,

$$\Big|\mathbb{E}[X_{i:n}]\Big| \leq \Big\{ n\ \frac{\binom{2n-2i}{n-i}\binom{2i-2}{i-1}}{\binom{2n-1}{n-1}} - 1 \Big\}^{1/2}, \quad 1 \leq i \leq n,$$

mas este limite superior só é bom quando $i = n$ ou $i = 1$.

5.6.2 Aproximações para os momentos

Recorreremos aqui ao facto de se poder escrever, para qualquer modelo F e $(X_1, \ldots, X_n)$ proveniente desse modelo,

$$X_{i:n} = F^{\leftarrow}(U_{i:n}) =: G(U_{i:n}), \quad 1 \leq i \leq n,$$

com $U_{i:n},\ 1 \leq i \leq n$, as n e.o's associadas a uma amostra aleatória de dimensão n, $\mathcal{U}(0,1)$.

O desenvolvimento de $G(U_{i:n})$ em série de Taylor em torno de $\mathbb{E}[U_{i:n}] = p_i = i/(n+1),\ 1 \leq i \leq n$, permite-nos obter

$$G(U_{i:n}) = G(p_i) + (Y_{i:n} - p_i)G'(p_i) + \frac{1}{2}(Y_{i:n} - p_i)^2 G''(p_i) + \ldots,$$

o que nos permite obter, com a notação atrás referida, $q_i = 1 - p_i,\ 1 \leq i \leq n$, aproximações tão precisas quanto se deseje,

$$\begin{aligned} \mathbb{E}[X_{i:n}] &= G(p_i) + R_{1,i}, \\ \mathbb{E}[X_{i:n}] &= G(p_i) + \frac{1}{2}\frac{p_i\ q_i}{n+2}G''(p_i) + R_{2,i}, \end{aligned} \tag{5.55}$$

onde, para uma classe vasta de funções de distribuição (f.d.'s) $R_{j,i} = o(1/n^j)$. Condições suficientes para que tal aconteça podem-se encontrar em Blom[13] (1958) — basta ter-se G limitada e contínua e G' limitada.

Para a estrutura de segunda ordem tem-se

$$\mathbb{C}ov(X_{i:n}, X_{j:n}) = \frac{p_i\ q_j}{n+1}G'(p_i)G'(p_j) + R_{ij}(n), \quad i \leq j. \tag{5.56}$$

Sob condições análogas às anteriormente impostas (G, G' e G'' limitadas e contínuas, e G''' limitada) tem-se $R_{ij}(n) = o(1/n^2)$.

Para n suficientemente grande, as aproximações mais usuais são

$$\begin{aligned} \mathbb{E}[X_{i:n}] &\approx F^{\leftarrow}\left(\frac{i}{n+1}\right), \\ \mathbb{C}ov(X_{i:n}, X_{j:n}) &\approx \frac{i(n+1-j)}{(n+1)^2(n+2)\ f\left(F^{\leftarrow}\left(\frac{i}{n+1}\right)\right)\ f\left(F^{\leftarrow}\left(\frac{j}{n+1}\right)\right)}. \end{aligned}$$

Blom (1958) introduz ainda aquilo a que chama uma correcção (α, β), i.e., considera uma generalização de (5.55) e de (5.56) do tipo

$$\mathbb{E}[X_{i:n}] = G(\pi_{in}) + R_i^{'}, \quad \pi_{in} = \frac{i - \alpha_{in}}{n - \alpha_{in} - \beta_{in} + 1},$$

$$\mathbb{C}ov[X_{i:n}, X_{j:n}] = \frac{\pi_{in}(1 - \pi_{jn})}{n - \alpha_{ijn} - \beta_{ijn} + 2}G'(\pi_{in})\ G'(\pi_{jn}) + R_{ij}^{'},$$

$$\pi_{\nu n} = \frac{\nu - \alpha_{ijn}}{n - \alpha_{ijn} - \beta_{ijn} + 1}, \quad \nu = i, j,$$

[13] Blom, G. (1958). *Transformed Beta Variates.* Wiley.

e determina quais as constantes possíveis e a ordem de grandeza dos restos associados. Blom determina ainda quais os pares de constantes (α_1, β_1) e (α_2, β_2), que permitem a validade da desigualdade

$$G(\pi_{i_1}) \leq \mathbb{E}[X_{i:n}] \leq G(\pi_{i_2}), \quad \pi_{i\nu} = \frac{i - \alpha_\nu}{n - \alpha_\nu - \beta_\nu + 1},\ \nu = 1, 2,$$

para todo o i, $1 \leq i \leq n$, e quando $n \to \infty$.

Apresentamos, em seguida, os valores obtidos para três modelos, Normal, Cauchy e Gumbel. Em modelo Normal tem-se

$$\Phi^{-1}\Big(\frac{i - 0.39}{n + 0.22}\Big) \leq \mathbb{E}[X_{i:n}] \leq \Phi^{-1}\Big(\frac{i - 0.5}{n}\Big), \quad i > \frac{n+1}{2},$$
$$\Phi^{-1}\Big(\frac{i - 0.5}{n}\Big) \leq \mathbb{E}[X_{i:n}] \leq \Phi^{-1}\Big(\frac{i - 0.39}{n + 0.22}\Big), \quad i \leq \frac{n+1}{2}.$$

Para o modelo Cauchy tem-se, para $2 \leq i \leq n + 1/2$,

$$\tan\Big[\pi\Big(\frac{i-1}{n-1} - \frac{1}{2}\Big)\Big] \leq \mathbb{E}[X_{i:n}] \leq \tan\Big[\pi\Big(\frac{i - 1.23}{n - 1.46} - \frac{1}{2}\Big)\Big],$$

e para $(n+1)/2 \leq i \leq n - 1$,

$$\tan\Big[\pi\Big(\frac{i - 1.23}{n - 1.46} - \frac{1}{2}\Big)\Big] \leq \mathbb{E}[X_{i:n}] \leq \tan\Big[\pi\Big(\frac{i-1}{n-1} - \frac{1}{2}\Big)\Big].$$

Finalmente, para o modelo Gumbel,

$$-\log\Big(-\log\Big(\frac{i - 0.5}{n}\Big)\Big) \leq \mathbb{E}[X_{i:n}] \leq -\log\Big(-\log\Big(\frac{i - 0.25}{n + 0.25}\Big)\Big).$$

Foram exactamente algumas destas desigualdades que levaram às escolhas de 'plotting positions' diferentes de $i/(n+1)$, tal como foi anteriormente referido no Capítulo 4.

5.7 O Teorema de Malmquist e a simulação de estatísticas ordinais

O teorema de Malmquist (Malmquist[14], 1950), de grande utilidade na simulação de e.o.'s de topo, garante-nos que, no mesmo contexto em que nos temos

[14] Malmquist, S. (1950). On a property of order statistics from a rectangular distribution. *Skand. Aktuar.* **33**, 214–222.

colocado, se tem

$$\left(\frac{F(Z_{k:n})}{F(Z_{k+1:n})}\right)^k \stackrel{d}{=} e^{-E_{n-k+1}}, \quad 1 \le k \le n, \tag{5.57}$$

i.e., as v.a.'s $(F(Z_{k:n})/F(Z_{k+1:n}))^k = (U_{k:n}/U_{k+1:n})^k$ são mutuamente independentes e $\mathcal{U}(0,1)$.

Na realidade,

$$-\log \frac{F(Z_{k:n})}{F(Z_{k+1:n})} \stackrel{d}{=} E_{n-k+1:n} - E_{n-k:n} \stackrel{d}{=} \frac{E_{n-k+1}}{k},$$

donde segue (5.57).

A aplicação do teorema de Malmquist em simulação foi efectuada por Schucany[15] (1972), e permite a geração de, por exemplo, k e.o.'s superiores (ou inferiores), k não demasiado elevado, de forma rápida e expedita, usando apenas a geração de k NPA's, $\mathcal{U}(0,1)$, U_j, $1 \le j \le k$, sem recorrer à geração de toda a amostra de dimensão n e consequente ordenação.

Na realidade,

$$U_{n:n}^n = U_1 \quad \Longrightarrow \quad U_{n:n} = U_1^{1/n}.$$

Em seguida

$$\frac{U_{n-1:n}}{U_{n:n}} = U_2^{1/(n-1)} \quad \Longrightarrow \quad U_{n-1:n} = U_1^{1/n} U_2^{1/(n-1)}.$$

Genericamente

$$U_{n-k:n} = U_1^{1/n} U_2^{1/(n-1)} \dots U_{k+1}^{1/(n-k)}, \quad X_{n-k:n} = F^{\leftarrow}(U_{n-k:n}),$$

para k fixo, pequeno relativamente a n.

Do facto da simetria de U em torno de $1/2$, ou seja, do facto de se ter $U \stackrel{d}{=} 1-U$, tem-se também

$$U_{k+1:n} = 1 - U_1^{1/n} U_2^{1/(n-1)} \dots U_{k+1}^{1/(n-k)}, \quad X_{k+1:n} = F^{\leftarrow}(U_{k+1:n}),$$

que permite a simulação de e.o.'s inferiores, para k pequeno.

[15] Schucany, R.W. (1972). Order statistics in simulation. *J. Statist. Comp. and Simul.* **1**, 281–286.

Capítulo 6

Teoria Distribucional Assintótica

6.1 Introdução

Seja $\{X_n\}_{n\geq 1}$ uma sucessão de v.a.'s i.i.d. $\frown F$, e defina-se

$$M_n := X_{n:n} = \max_{1\leq i\leq n} X_i, \quad n \geq 1. \tag{6.1}$$

Tal como vimos no Capítulo 5, a f.d. de M_n é dada por

$$\mathbb{P}\{M_n \leq x\} = F^n(x).$$

Sendo F desconhecida, esta expressão não tem grande utilidade. Como estamos muitas vezes interessados no máximo de um grande número de variáveis modela-se o máximo usando um argumento assintótico. Em particular, esperamos que a distribuição de M_n, quando $n \to \infty$, não dependa fortemente de F. Comecemos por relembrar a definição de convergência em distribuição:

Definição 6.1.1 (Convergência em distribuição)**.** *Sejam $X, X_1, X_2, \ldots$ v.a.'s com f.d.'s $F, F_1, F_2, \ldots$ Diz-se que X_n converge em distribuição para X, quando $n \to \infty$, e usa-se a notação $X_n \xrightarrow[n\to\infty]{d} X$, se*

$$\lim_{n\to\infty} F_n(x) = F(x), \text{ para todo o } x, \text{ ponto de continuidade de } F(x).$$

Que distribuições podem surgir para a v.a. limite de M_n? Quando $n \to \infty$, com $x^F := \sup\{x : F(x) < 1\} \leq \infty$, o limite superior do suporte de F,

$$F^n(x) \to \begin{cases} 0, & \text{se} \quad F(x) < 1 \\ 1, & \text{se} \quad F(x) = 1, \end{cases} \tag{6.2}$$

pelo que $M_n \xrightarrow[n\to\infty]{d} x^F$. A convergência em distribuição anterior implica, por sua vez, a convergência em probabilidade para a mesma constante, i.e., $M_n \xrightarrow[n\to\infty]{p} x^F$, mesmo que $x^F = \infty$. A distribuição assintótica de M_n é portanto degenerada. Além disso, a propriedade ascendente das e.o's em causa, combinada com esta convergência em probabilidade, determina a convergência quase certa, $M_n \xrightarrow[n\to\infty]{q.c.} x^F$.

Uma vez que a distribuição limite de M_n é degenerada há que recorrer a uma normalização de modo a obter uma lei limite não-degenerada, de forma semelhante ao que se passa no caso da teoria assintótica para somas, que relembramos:

Teorema 6.1.1 (TLC e LFGN)**.** *Sob condições de regularidade e sendo $X, X_1, \ldots, X_n$ i.i.d. a X, com*

$$\overline{X}_n := \sum_{i=1}^{n} X_i/n,$$

se existir $\sigma_n^2 = \mathbb{V}ar[\overline{X}_n] = \mathbb{V}ar[X]/n = \sigma^2/n$, e consequentemente, $\mu_n = \mathbb{E}[\overline{X}_n] = \mathbb{E}[X] = \mu$, tem-se, quando $n \to \infty$,

- $\overline{X}_n \xrightarrow[n\to\infty]{p} \mu$ (LFGN)
- $(\overline{X}_n - \mu_n)/\sigma_n \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$ (TLC).

Analogamente às somas, parece pois ser sensato investigar as leis limite para o máximo normalizado $(M_n - b_n)/a_n$, para sequências reais convenientes $a_n > 0$ e $b_n \in \mathbb{R}$, tema a abordar com toda a generalidade na Secção 6.4, mas com alguns casos particulares discutidos na Secção 6.2.

Mais geralmente do que considerarmos a sucessão de máximos parciais, M_n, em (6.1), iremos considerar o comportamento assintótico de três tipos de e.o.'s:

1. **Estatísticas ordinais centrais**. Trata-se de e.o.'s $X_{k:n}$ onde a **ordem** $k = k_n \to \infty$, mas $k/n \to \lambda$, $0 < \lambda < 1$, quando $n \to \infty$. Usualmente

$k = \lfloor n\lambda \rfloor + 1$, $0 < \lambda < 1$, e $X_{k:n}$ é o **quantil empírico de ordem** λ, que vai ser, como veremos adiante, assintoticamente Normal, desde que o modelo F subjacente satisfaça determinadas condições de regularidade, pouco restrictivas.

2. **Estatísticas ordinais extremais**. Neste caso estamos a considerar $X_{k:n}$ ou $X_{n-k:n}$ com k inteiro fixo. O modelo limite é então muito diferente do modelo Normal.
3. **Estatísticas ordinais intermédias**. Tem-se tal como em 1., $k = k_n \to \infty$, mas $k/n \to 0$ ou $k/n \to 1$, quando $n \to \infty$. Sob condições adequadas de regularidade, voltamos então a ter um comportamento assintoticamente Normal.

6.2 Alguns resultados parciais sobre a teoria assintótica de estatísticas ordinais. Método de Rényi

6.2.1 O modelo Exponencial, $\mathcal{E}(1)$

Consideremos $E \frown \mathcal{E}(1)$, i.e.

$$F(x) = F_{_E}(x) = 1 - \exp(-x), \text{ para } x > 0.$$

Teorema 6.2.1. *Seja $k \geq 1$ um inteiro fixo e positivo. Então*

$$\lim_{n\to\infty} \mathbb{P}\left[nE_{k:n} \leq x\right] = \int_0^x \frac{t^{k-1}e^{-t}}{\Gamma(k)}dt, \quad x \geq 0, \tag{6.3}$$

i.e., $nE_{k:n}$ é assintoticamente uma v.a. Gama(k).

Demonstração. Podemos derivar facilmente este resultado através da expressão da f.d.p. de uma e.o., apresentada em (5.1). Então

$$\begin{aligned} f_{nE_{k:n}}(x) &= \frac{1}{n} f_{E_{k:n}}(x/n) = \binom{n-1}{k-1} e^{-x} \left(e^{x/n} - 1\right)^{k-1} \\ &= \frac{e^{-x}}{(k-1)!} \prod_{j=1}^{k-1} (n-j) \left(e^{x/n} - 1\right). \end{aligned}$$

Como $\lim_{n\to\infty}(n-j)\left(e^{x/n}-1\right) = x$, $\lim_{n\to\infty} f_{nE_{k:n}}(x) = e^{-x}x^{k-1}/\Gamma(k)$. Pelo teorema de Scheffé, se $f_n(x) \to g(x)$, então $F_n(x) \to G(x)$ (não sendo a recíproca obviamente verdadeira). Tem-se pois

$$n\ E_{k:n} \xrightarrow[n\to\infty]{d} \text{Gama}(k),$$

ou seja, a validade de (6.3). □

Vejamos outra demonstração mais natural do Teorema 6.2.1:

Demonstração. Da representação de Rényi, $E_{k:n} \stackrel{d}{=} \sum_{j=1}^{k} E_j/(n-j+1)$, deduz-se que

$$nE_{k:n} \stackrel{d}{=} \sum_{j=1}^{k}\left(1+\frac{j-1}{n-j+1}\right)E_j = \sum_{j=1}^{k} E_j + \sum_{j=1}^{k}\frac{j-1}{n-j+1}E_j.$$

Como a primeira componente desta soma, $\sum_{j=1}^{k} E_j \stackrel{d}{=} Z \frown \text{Gama}(k)$, e a segunda componente $\xrightarrow[n\to\infty]{p} 0$, $nE_{k:n} \stackrel{d}{=} Z + o_p(1) \xrightarrow[n\to\infty]{d} \text{Gama}(k)$. □

Observação 6.2.1. *Note-se que* $E_{k:n} \xrightarrow[n\to\infty]{p} 0$, *uma vez que* $n\,E_{k:n} = Z + o_p(1)$, *i.e.,* $E_{k:n} = Z/n + o_p(1/n) = \mathcal{O}_p(1/n) + o_p(1/n)$, *donde* $E_{k:n} = o_p(1)$.

No caso de termos $k = n$, ou seja, de estarmos interessados no máximo de n v.a.'s $\mathcal{E}(1)$, podemos desde logo perguntar se será fácil encontrar uma normalização de $E_{n:n}$ de forma a se obter uma distribuição limite não-degenerada. Consideremos as sucessões $a_n = 1$ e $b_n = \log n$. Então

$$\begin{aligned}
\mathbb{P}[E_{n:n} - \log n \le x] &= \mathbb{P}[E_{n:n} \le x + \log n] = \{F(x+\log n)\}^n \\
&= \{1-\exp(-x-\log n)\}^n, \text{ para } x > -\log n \\
&= \{1-\exp(-x)/n\}^n, \\
&\to \exp[-\exp(-x)], \text{ quando } n\to\infty \text{ para } x\in\mathbb{R}.
\end{aligned}$$

Na Figura 6.1 ilustramos, *à direita*, a velocidade de convergência de $E_{n:n} - \log n$ para uma v.a. Gumbel. Como se vê a convergência é neste caso muito rápida, o que nem sempre acontece em EVT, como se pode ver na Figura 6.2, onde, denotando N uma v.a. $\mathcal{N}(0,1)$, se representa $(N_{n:n}-b_n)/a_n$, com $a_n = (2\log n)^{-0.5}$, $b_n = (2\log n)^{0.5} - 0.5(2\log n)^{-0.5}(\log\log n + \log 4\pi)$, valores estes a serem justificados mais adiante, na Secção 6.4.6.

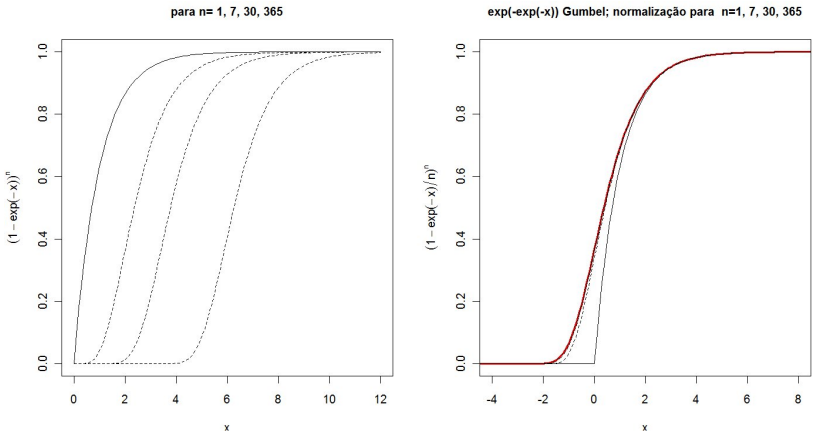


Figura 6.1: Distribuições de $E_{n:n}$ (*esquerda*) e de $E_{n:n} - \log n$ (*direita*), para $n = 1, 7$ (*semanal*), 30 (*mensal*) e 365 (*anual*), comparativamente à lei limite Gumbel

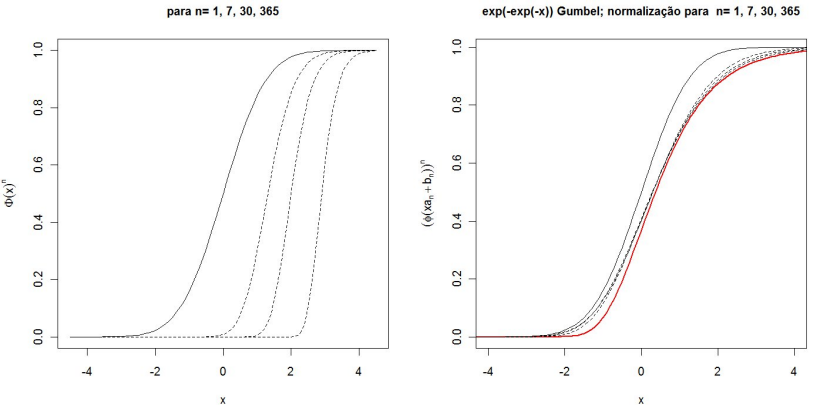


Figura 6.2: Distribuições de N_n (*esquerda*) e de $(N_n - b_n)/a_n$ (*direita*), para $n = 1, 7$ (*semanal*), 30 (*mensal*) e 365 (*anual*) para variáveis $\mathcal{N}(0, 1)$, comparativamente à lei limite Gumbel

6.2.2 O modelo Uniforme, $\mathcal{U}(0, 1)$

Comecemos por recordar o seguintes resultado:

Teorema 6.2.2. *Seja* $Z_n = X_n + Y_n$, $n \geq 1$, *onde* $\{X_n\}_{n\geq 1}$ *e* $\{Y_n\}_{n\geq 1}$ *são sucessões de v.a.'s i.i.d. Designemos por* $F_n(x)$ *a f.d. de* X_n, $n \geq 1$, *e admitamos que*

$$\left.\begin{array}{l} \mathbb{E}[Y_n] \underset{n\to\infty}{\longrightarrow} 0 \\ \mathbb{V}ar[Y_n] \underset{n\to\infty}{\longrightarrow} 0 \end{array}\right\} \Longrightarrow Y_n \xrightarrow[n\to\infty]{p} 0.$$

Então, se $X_n \xrightarrow[n\to\infty]{d} X$, *com f.d.* $F(x)$, *então* $Z_n \xrightarrow[n\to\infty]{d} X$.
Se U_n *com f.d.* $H_n(x)$ *for independente de* X_n, $n \geq 1$, *e* $U_n \xrightarrow[n\to\infty]{d} U$, *com f.d.* $H(x)$, *então*

$$(Z_n, U_n) \xrightarrow[n\to\infty]{d} (Z, U), \quad G_{_{Z,U}}(z, u) = F(z)H(u),$$

i.e., Z_n *e* U_n *são assintoticamente independentes.*

Teorema 6.2.3. *Seja* $\{U_j\}_{j\geq 1}$ *uma sucessão de v.a.'s i.i.d.,* $\mathcal{U}(0,1)$. *Então, para todo o* $k \geq 1$,

$$n\, U_{k:n} \overset{d}{=} n\,(1 - U_{n-k+1:n}) \xrightarrow[n\to\infty]{d} \text{Gama}(k).$$

Consequentemente,

$$\mathbb{P}\left[E_{n-k+1:n} - \log n \leq y\right] \underset{n\to\infty}{\longrightarrow} 1 - \int_0^{e^{-y}} \frac{t^{k-1}e^{-t}}{\Gamma(k)} dt.$$

Tem-se ainda que para j *e* k *inteiros fixos,* $U_{k:n}$ *e* $U_{n-j+1:n}$ *são estatisticamente independentes, i.e.,*

$$\lim_{n\to\infty} P\left[U_{k:n} \leq x/n, 1 - U_{n-j+1:n} < y/n\right] = \int_0^x \int_0^y \frac{u^{k-1}v^{j-k}e^{-(u+v)}}{\Gamma(k)\Gamma(j)} du dv, \; x > 0, y > 0.$$

Demonstração. Pensemos então em $\{U_j\}_{j\geq 1}$, v.a.'s i.i.d., $\mathcal{U}(0,1)$. O método da transformação uniformizante permite-nos escrever:

$$U_{n-k+1:n} = e^{-E_{k:n}} \quad \text{i.e.} \quad E_{k:n} = -\log U_{n-k+1:n}.$$

Consequentemente, como $E_{k:n} \xrightarrow{p} 0$, quando $n \to \infty$,

$$1 - U_{n-k+1:n} = E_{k:n} + \sum_{j \geq 2} E_{k:n} \ \frac{E_{k:n}^{j-1}(-1)^{j-1}}{j!}.$$

Na soma anterior, o primeiro termo converge fracamente para uma v.a. Gama(k) e o segundo converge em probabilidade para 0, quando $n \to \infty$. Temos consequentemente

$$n\,(1 - U_{n-k+1:n}) \xrightarrow[n\to\infty]{d} \text{Gama}(k),$$

resultado este a que se poderia ter chegado directamente, pensando que $\mathbb{P}\,(n\,(1 - U_{n-k+1:n}) \leq x) = \mathbb{P}\,(E_{k:n} \leq -\log\,(1 - x/n))$, tendo $-\log(1 - x/n)$ um comportamento assintótico análogo ao de x/n, quando $n \to \infty$, o que juntamente com a continuidade da função Gama forneceria o mesmo resultado. Como $U_k \overset{d}{=} 1 - U_k \ \frown \ \mathcal{U}(0,1)$, tem-se $U_{k:n} \overset{d}{=} 1 - U_{n-k+1:n}$, $1 \leq k \leq n$, e consequentemente

$$n\ U_{k:n} \xrightarrow[n\to\infty]{d} \text{Gama}(k).$$

□

Corolário 6.2.1. *Em modelo* Exponencial, $\mathcal{E}(1)$

$$\mathbb{P}\,[E_{n-k+1:n} - \log n \leq y] \underset{n\to\infty}{\longrightarrow} 1 - \int_0^{e^{-y}} \frac{t^{k-1}e^{-t}}{\Gamma(k)} dt.$$

Demonstração. Decorre directamente do facto de se ter

$$\begin{aligned} \mathbb{P}[nU_{k:n} \leq x] &= \mathbb{P}[-\log n - \log U_{k:n} \geq -\log x] \\ &= 1 - \mathbb{P}[E_{n-k+1:n} - \log n \leq -\log x] \\ &\underset{n\to\infty}{\longrightarrow} \int_0^x \frac{t^{k-1}e^{-t}}{\Gamma(k)} dt. \end{aligned}$$

□

Corolário 6.2.2. *Em modelo* Uniforme, $\mathcal{U}(0,1)$, *e para j e k inteiros fixos, $U_{k:n}$ e $U_{n-j+1:n}$ são estatisticamente independentes, i.e.,*

$$\begin{aligned} &\lim_{n\to\infty} P\,[U_{k:n} \leq x/n, 1 - U_{n-j+1:n} < y/n] \\ &\qquad = \int_0^x \int_0^y \frac{u^{k-1}v^{j-k}e^{-(u+v)}}{\Gamma(k)\Gamma(j)} du dv, \ x > 0, y > 0. \end{aligned}$$

Corolário 6.2.3. *A amplitude amostral em modelo* Uniforme, $R_n := U_{n:n} - U_{1:n}$, *tem, após normalização adequada, distribuição limite* Gama. *Mais especificamente*

$$n\,[1 - R_n] = n\,[1 - (U_{n:n} - U_{1:n})] \xrightarrow[n\to\infty]{d} \text{Gama}(2).$$

Demonstração. O Teorema 6.2.3 permite-nos concluir que $U_{n:n} \xrightarrow{p} 1$ e $U_{1:n} \xrightarrow{p} 0$, e consequentemente $R_n = U_{n:n} - U_{1:n} \xrightarrow{p} 1$. Mas $nU_{1:n}$ e $n(1-U_{n:n})$ convergem para leis não degeneradas Gama(1) e são assintoticamente independentes. É então lógico considerar $n\,[1 - (U_{n:n} - U_{1:n})] = n(1-U_{n:n}) + nU_{1:n}$, donde de imediato segue o resultado. □

Por intermédio das distribuições limites obtidas para $U_{k:n}$ (e para $U_{n-k+1:n}$) podem determinar-se as distribuições limite das v.a.'s $X_{k:n}$ (e de $X_{n-k+1:n}$), as chamadas **e.o.'s extremais**, mas voltaremos a este assunto mais tarde, na Secção 6.4.

6.3 Comportamento limite de estatísticas ordinais centrais (quantis)

Temos a validade do resultado seguinte:

Teorema 6.3.1 (Teorema de Rényi)**.** *Seja* $\{X_n\}_{n\geq 1}$ *uma sucessão de v.a.'s i.i.d. com f.d.* F *absolutamente contínua. Seja* $f(x) = F'(x)$ *e admitamos que* $f(x)$ *é contínua e positiva no intervalo* $a < x < b$. *Se* $0 < F(a) < \lambda < F(b) < 1$ *e se* $|k_n - n\lambda| = o(\sqrt{n})$ (*o que automaticamente* $\Longrightarrow k/n \to \lambda$, *quando* $n \to \infty$, *então*

$$\frac{X_{k:n} - F^{\leftarrow}(\lambda)}{\frac{1}{f(F^{\leftarrow}(\lambda))}\sqrt{\frac{\lambda(1-\lambda)}{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1), \tag{6.4}$$

ou, equivalentemente, com $\chi_\lambda := F^{\leftarrow}(\lambda)$

$$\frac{\sqrt{n} f(\chi_\lambda)}{\sqrt{\lambda(1-\lambda)}}(X_{k:n} - \chi_\lambda) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)\,.$$

Demonstração. Comecemos por ver a situação Exponencial, $\mathcal{E}(1)$. Como

$$E_{n-k_n+1:n} \stackrel{d}{=} \sum_{j=1}^{n-k_n+1} \frac{E_j}{n-j+1} =: \sum_{j=1}^{n-k_n+1} Z_j,$$

sendo as v.a.'s Z_j independentes e tais que

$$\mathbb{E}[Z_j] = \frac{1}{n-j+1}; \qquad \mathbb{V}ar[Z_j] = \frac{1}{(n-j+1)^2},$$

tem-se

$$\begin{aligned} \mu_n &= \mathbb{E}[E_{n-k_n+1:n}] = \sum_{j=1}^{n-k_n+1} \frac{1}{n-j+1}; \\ \sigma_n^2 &= \mathbb{V}ar[E_{n-k_n+1:n}] = \sum_{j=1}^{n-k_n+1} \frac{1}{(n-j+1)^2}. \end{aligned}$$

A validade da condição de Lindberg-Feller implica então a validade do TLC para $(E_{n-k_n+1:n} - \mu_n)/\sigma_n$.

Como $\psi(n) = \log n - \frac{1}{2n}(1+o(1))$ e $\psi'(n) = \frac{1}{n} + \frac{1}{2n^2}(1+o(1))$, quando $n \to \infty$

$$\mu_n = \sum_{j=1}^{n-k_n+1} \frac{1}{n-j+1} = \psi(n+1) - \psi(k_n) = \log\left(\frac{n}{k_n}\right) + O\left(\frac{1}{k_n}\right)$$

e

$$\sigma_n^2 = -\psi'(n+1) - \psi'(k_n) = \frac{n-k_n}{nk_n} + O\left(\frac{1}{k_n^2}\right).$$

Consequentemente,

$$\mathbb{P}\left[\frac{E_{n-k_n+1:n} - \log(n/k_n)}{\sqrt{\frac{n-k_n}{nk_n}}} \le x\right] \underset{n\to\infty}{\longrightarrow} \Phi(x),$$

a f.d. de uma v.a. $\mathcal{N}(0,1)$.

O facto de se ter $\lim_{n\to\infty} k_n/n = \lambda$, implica a possibilidade de se poderem substituir os coeficientes de atracção de localização e de escala por $-\log\lambda$ e $\sqrt{(1-\lambda)/(n\lambda)}$, i.e., tem-se

$$\mathbb{P}\left[\frac{E_{n-k_n+1:n} + \log\lambda}{\sqrt{(1-\lambda)/(n\lambda)}} \le x\right] \underset{n\to\infty}{\longrightarrow} \Phi(x),$$

Passemos agora ao caso geral: Tem-se $E_{n-k_n+1:n} = -\log F(X_{k_n:n})$, e consequentemente,

$$P\left[\frac{-\log F(X_{k_n:n}) - \log(1/\lambda)}{\sqrt{(1-\lambda)/(n\lambda)}} \leq x\right] = P\left[X_{k_n:n} \geq F^{\leftarrow}\left(\lambda \mathrm{e}^{-x\sqrt{\frac{1-\lambda}{n\lambda}}}\right)\right] \underset{n\to\infty}{\longrightarrow} \Phi(x).$$

Como, com θ_n a convergir para 1,

$$F^{\leftarrow}\left(\lambda \mathrm{e}^{-x\sqrt{\frac{1-\lambda}{n\lambda}}}\right) = F^{\leftarrow}(\lambda) + \frac{\lambda\left(\mathrm{e}^{-x\sqrt{\frac{1-\lambda}{n\lambda}}} - 1\right)}{f\left(F^{\leftarrow}(\lambda)\theta_n\right)} = F^{\leftarrow}(\lambda) + O\left(\frac{1}{\sqrt{n}}\right),$$

tem-se a validade de (6.4). □

Mais geralmente temos a validade do resultado multivariado seguinte:

Teorema 6.3.2. *Seja $\{X_n\}_{n\geq 1}$ uma sucessão de v.a.'s i.i.d. com f.d.p. $f(\cdot)$. Consideremos as e.o.'s $(X_{n_1:n}, \ldots, X_{n_k:n})$, $n_j = \lfloor n\lambda_j \rfloor + 1$, $0 < \lambda_1 < \cdots < \lambda_k < 1$. Designemos por χ_{λ_j} o quantil populacional correspondente a λ_j, $1 \leq j \leq k$. Se $0 < f(\chi_{\lambda_j}) < \infty$, $1 \leq j \leq k$, então a distribuição assintótica conjunta de*

$$\sqrt{n}\left(X_{n_1:n} - \chi_{\lambda_1}\right), \ldots, \sqrt{n}\left(X_{n_k:n} - \chi_{\lambda_k}\right)$$

é a multinormal com valor médio nulo e matriz de covariâncias

$$\Sigma = [\sigma_{ij}], \qquad \sigma_{ij} = \frac{\lambda_i(1-\lambda_j)}{f(\chi_{\lambda_i})f(\chi_{\lambda_j})}, \quad i \leq j.$$

Demonstração. Faremos a demonstração com base na importante representação de Bahadur[1] (1966) de e.o.'s em termos da f.d. empírica,

$$X_{n_j:n} = \chi_{\lambda_j} - \frac{\widehat{F}_n(\chi_{\lambda_j}) - \lambda_j}{f(\chi_{\lambda_j})} + R_n(j), \; R_n(j) = o_p\left(\frac{1}{\sqrt{n}}\right).$$

[1] Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580.

Tem-se então

$$\sqrt{n}\,(X_{n_1:n}-\chi_{\lambda_1}),\ldots,\sqrt{n}\,(X_{n_k:n}-\chi_{\lambda_k}) \\ \stackrel{d}{=} \sqrt{n}\,\frac{\lambda_1-\widehat{F}_n(\chi_{\lambda_1})}{f(\chi_{\lambda_1})},\ldots,\sqrt{n}\,\frac{\lambda_k-\widehat{F}_n(\chi_{\lambda_k})}{f(\chi_{\lambda_k})}.$$

Mas

$$\widehat{F}_n(\chi_{\lambda_j})=\frac{1}{n}\sum_{j=1}^{n}I(\chi_{\lambda_j}),\ \text{com}\ I(\chi_{\lambda_j})=\left\{\begin{array}{ll}1 & \lambda_j\\ 0 & 1-\lambda_j.\end{array}\right.$$

O resultado segue-se portanto por aplicação do TLC multidimensional. □

Observação 6.3.1. *Como consequência do* Teorema 6.3.2*, vemos que, mesmo assintoticamente, duas e.o.'s centrais são dependentes. Isso contrasta com o caso extremal, em que se constata a independência assintótica das e.o.'s extremais inferiores das e.o's extremais superiores, como vimos no Teorema* 6.2.3*, para o modelo* $\mathcal{U}(0,1)$*, mas mais geralmente válido para qualquer modelo* F*.*

6.4 Teoria assintótica de valores extremos

Estamos aqui interessados em alguns aspectos da TVE (ou EVT) assintótica em sucessões de v.a.'s i.i.d., e algumas das suas generalizações a sucessões estacionárias.

Consideraremos aqui os seguintes tópicos de teoria de valores extremos: Começamos por enunciar e esboçar a demonstração do teorema de Gnedenko, onde surgem os três tipos possíveis de distribuição assintótica não degenerada de M_n, em (6.1), convenientemente normalizado. Introduzimos em seguida uma forma unificada dos três tipos, e procedemos à caracterização dos chamados max-domínios de atracção e às escolhas possíveis dos chamados coeficientes de atracção. Após a referência às condições suficientes de von Mises, referimos a equivalência entre as relações $\lim_{n\to\infty}\mathbb{P}[M_n\leq u_n(\xi)]=\exp(-\xi)$ e $n\mathbb{P}[X_i>u_n(\xi)]=\xi(1+o(1))$, quando $n\to\infty$. Derivamos em seguida a forma da distribuição limite do máximo em modelos normais e a escolha adequada de constantes de atracção. Dedicamo-nos ainda ao estudo do carácter

poissoniano do número de excedências de $u_n(\xi)$ por $X_1, \ldots, X_n$, ao cálculo da distribuição assintótica do k-ésimo máximo, para k fixo, e das k maiores e.o.'s associadas a uma amostra de dimensão n. Faremos ainda uma breve referência à robustez dos problemas apresentados.

Note-se desde já que os resultados que aqui iremos obter para máximos, podem ser de imediato convertidos em resultados para mínimos, pois

$$\min_{1\leq i\leq n} X_i = -\max_{1\leq i\leq n}(-X_i). \tag{6.5}$$

6.4.1 O teorema de Gnedenko

Tal como referido anteriormante, a f.d. de M_n, em (6.1) é dada por

$$F_{M_n}(x) = \mathbb{P}[M_n \leq x] = F^n(x),$$

e fazendo $n \to \infty$, obtém-se (6.2), i.e., a distribuição limite, mesmo quando é própria, é degenerada e consequentemente de interesse muito limitado. É então sensato colocar a pergunta: será possível encontrar constantes reais $\{a_n\}_{n\geq 1}$ $(a_n > 0)$ e $\{b_n\}_{n\geq 1}$ tais que

$$\frac{M_n - b_n}{a_n} \xrightarrow[n\to\infty]{d} Y, \text{ não degenerada}, \tag{6.6}$$

i.e., tais que

$$F^n(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} G(x), \tag{6.7}$$

com $G(x)$ f.d. não degenerada?

Definição 6.4.1. *Se existirem sucessões de constantes* $\{a_n\}_{n\geq 1}$ $(a_n > 0)$ *e* $\{b_n\}_{n\geq 1}$ *tais que se verifica* (6.7)*, tais sucessões de constantes são designadas por* **coeficientes de atracção** *de F para G, e diremos que F pertence ao* **max-domínio de atracção** *da lei G, o que será denotado por $F \in \mathcal{D}_{\mathcal{M}}(G)$.*

São pois de interesse os dois problemas a seguir apresentados, formulados explicitamente a primeira vez por Gnedenko (1943):

(i) Identificação das possíveis formas limites G da f.d. de $(M_n - b_n)/a_n$, para sucessões de constantes convenientes, $\{a_n\}_{n\geq 1}$ $(a_n > 0)$ e $\{b_n\}_{n\geq 1}$.

(ii) Caracterização do max-domínio de atracção das possíveis leis limites para máximos, i.e., caracterização de

$$\mathcal{D}_{\mathcal{M}}(G) := \Big\{ F : \exists \{a_n\}_{n\geq 1} \, (a_n > 0) \text{ e } \{b_n\}_{n\geq 1} \text{ para as quais } F^n(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} G(x), \forall x \in \mathcal{C}(G) \Big\},$$

denotando $\mathcal{C}(G)$ o conjunto dos pontos de continuidade de G.

Começamos por introduzir o conceito de **tipo** (de Khinchine):

Lema 6.4.1 (convergência de tipos de Khinchine)**.** *Sejam $U_1(x)$ e $U_2(x)$ duas f.d.'s não degeneradas. Se para uma sucessão $\{F_n\}_{n\geq 1}$ de f.d.'s existem sucessões de números reais $\{a_n\}_{n\geq 1}$ $(a_n > 0)$, $\{b_n\}_{n\geq 1}$, $\{a'_n\}_{n\geq 1}$ $(a'_n > 0)$ e $\{b'_n\}_{n\geq 1}$ tais que*

$$\lim_{n\to\infty} F_n(a_n x + b_n) = U_1(x) \quad e \quad \lim_{n\to\infty} F_n(a'_n x + b'_n) = U_2(x) \tag{6.8}$$

então

$$a'_n/a_n \underset{n\to\infty}{\longrightarrow} A \ (A > 0), \quad (b'_n - b_n)/a_n \underset{n\to\infty}{\longrightarrow} B, \tag{6.9}$$

e

$$U_2(x) = U_1(Ax + B), \quad \forall x \in \mathbb{R}. \tag{6.10}$$

Inversamente, se (6.9) *for válido, qualquer das relações em* (6.8) *implica a outra e* (6.10)*.*

Demonstração. Para uma boa demonstração deste lema veja-se Feller[2], Vol. II, página 246. □

Este lema conduz obviamente à seguinte definição formal do conceito de tipo:

Definição 6.4.2. *Duas f.d.'s $U_1(x)$ e $U_2(x)$ são do mesmo tipo se existem constantes reais $A > 0$ e B tais que $U_2(x) = U_1(Ax + B)$, para todo o $x \in \mathbb{R}$.*

Enunciaremos sem demonstração o célebre teorema de Gnedenko.

[2]Feller, W. (1966). *An Introduction to Probability Theory and Its Applications.* Vol. 2, Wiley.

Teorema 6.4.1 (Teorema dos tipos extremais: Gnedenko, 1943)**.** *No contexto acima referido, admitamos que existem sucessões de constantes* $\{a_n\}_{n\geq 1}$ $(a_n > 0)$, $\{b_n\}_{n\geq 1}$ *e uma f.d. não degenerada* G *tais que*

$$\lim_{n\to\infty} P\left[\frac{M_n - b_n}{b_n} \leq x\right] = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x),\ \forall x \in \mathcal{C}(G). \quad (6.11)$$

Então G *pertence a um dos três tipos de distribuições de Valores Extremos:*

$$\text{Tipo I}: \quad \Lambda(x) = \mathrm{e}^{-\mathrm{e}^{-x}},\ x \in \mathbb{R} \quad [\text{Gumbel}], \quad (6.12)$$

$$\text{Tipo II}: \quad \Phi_\alpha(x) = \mathrm{e}^{-x^{-\alpha}},\ x > 0,\ \alpha > 0 \quad [\text{Fréchet}], \quad (6.13)$$

$$\text{Tipo III}: \quad \Psi_\alpha(x) = \mathrm{e}^{-(-x)^{\alpha}},\ x < 0,\ \alpha > 0 \quad [\text{Max} - \text{Weibull}]. \quad (6.14)$$

Para ilustrar como é que a cauda $\overline{F}(x) := 1 - F(x)$ decai para 0, quando $x \to x^F := \sup\{x : F(x) < 1\}$, apresentamos a Figura 6.3.

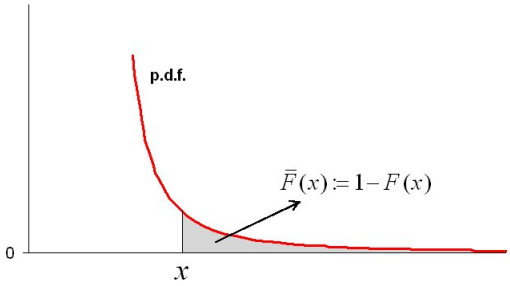


Figura 6.3: Ilustração de $\overline{F}(x)$=área assinalada a cinzento

Na Figura 6.4, ilustramos o comportamento dos três tipos de distribuições de valores extremos.

Note-se que o Teorema de Gnedenko **admite** a validade de (6.6), ou equivalentemente de (6.11). É fácil ver que existem sucessões de v.a.'s $\{X_n\}_{n\geq 1}$ para as quais não é possível encontrar sucessões de constantes $\{a_n\}_{n\geq 1}$ $(a_n > 0)$ e $\{b_n\}_{n\geq 1}$ tais que tal aconteça. Basta que exista $x^F < \infty$: $\mathbb{P}(X < x^F) < 1$ e $\mathbb{P}(X \leq x^F) = 1$, i.e., basta que F possua um limite superior $x^F = \sup\{x : F(x) < 1\} < \infty$ e $\mathbb{P}(X = x^F) \neq 0$. Na realidade:

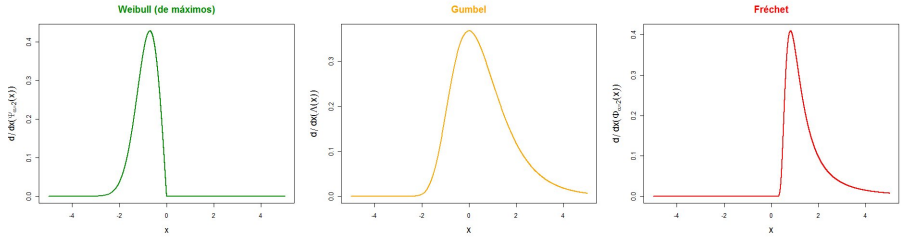


Figura 6.4: F.d.p. Max-Weibull, associada à f.d. $\Psi_{\alpha=2}$, limitada superiormente e com cauda curta (*esquerda*), f.d.p. Gumbel com suporte no eixo real (*centro*), e f.d.p. Fréchet, associada à f.d. $\Phi_{\alpha=2}$ (limitada inferiormente), não-limitada superiormente, com cauda pesada (*direita*)

Teorema 6.4.2. *Admitamos que* $\{X_n\}_{n\geq 1}$ *são v.a.'s i.i.d. com f.d. comum* $F(.)$ *tal que* $F(x^F-) < 1 = F(x^F)$, $x^F < \infty$. *Então não é possível normalizar* M_n *de modo a obtermos convergência fraca para uma v.a. não degenerada.*

Demonstração. Admitamos que $\mathbb{P}[M_n \leq a_n x + b_n] \underset{n\to\infty}{\longrightarrow} G(x)$. Ponhamos $u_n := a_n x + b_n$. Se $u_n < x^F$ teremos $\mathbb{P}[M_n \leq u_n] \leq F^n(x^F-) \underset{n\to\infty}{\longrightarrow} 0$. Logo, se $G(x) > 0$, tem-se $u_n \geq x^F$ para todo o n suficientemente elevado. Mas então $F(u_n) = 1$ e $\mathbb{P}[M_n \leq u_n] \underset{n\to\infty}{\longrightarrow} 1$, i.e., G é degenerada. □

Apresentamos ainda, sem demonstração, um outro resultado da mesma índole:

Teorema 6.4.3. *Seja* $F(x)$ *uma f.d. tal que para uma sucessão* $x_1 < x_2 < \cdots < x_n < \ldots$ *de reais com limite finito ou infinito*

$$\frac{1 - F(x_n-)}{1 - F(x_n)} \geq 1 + \beta, \ \beta > 0.$$

Então F *não pode pertencer ao domínio de atracção de nenhuma das leis limites de máximos.*

Temos vários exemplos ilustrativos do teorema anterior, de entre os quais destacamos:

Exemplo 6.4.1. *Não é possível normalizar o máximo de v.a.'s* Poisson *de modo a obter convergência para uma lei não degenerada. Na realidade, se*

tivermos uma v.a. X de Poisson

$$\frac{1-F(n-)}{1-F(n)} \approx \frac{\lambda^n \mathrm{e}^{-\lambda}}{n!} \ \frac{(n+1)!}{\lambda^{n+1}\mathrm{e}^{-\lambda}} = \frac{n+1}{\lambda} \underset{n\to\infty}{\longrightarrow} \infty.$$

Embora não tencionemos fazer aqui a demonstração do teorema de Gnedenko, iremos apontar as ideias principais da demonstração. A ideia fundamental da demonstração dada por Gnedenko, pode ser sumarizada do modo seguinte: A validade de (6.11) implica

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}[M_n \le a_{nk}x + b_{nk}] &= \lim_{n\to\infty} F^n(a_{nk}x + b_{nk}) \\ &= \lim_{n\to\infty} \left\{F^{nk}(a_{nk}x + b_{nk})\right\}^{1/k} \\ &= G^{1/k}(x), \ \forall k \ge 1, \end{aligned}$$

e consequentemente:

Lema 6.4.2. *Seja $\{X_n\}_{n\ge1}$ uma sucessão de v.a.'s i.i.d. com f.d. comum nas condições do* Teorema 6.4.1. *Então, tem-se para todo $k \ge 1$,*

$$\lim_{n\to\infty} \mathbb{P}\left[M_n \le a_{nk}x + b_{nk}\right] = G^{1/k}(x), \tag{6.15}$$

i.e. a validade de (6.15) *para $k = 1$ implica a sua validade para todo o $k > 1$.*

Consequentemente, e no contexto do Teorema de Gnedenko, tem-se

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$$

e

$$\lim_{n\to\infty} F^n(a_{nk}x + b_{nk}) = G^{1/k}(x), \ \forall k > 1.$$

A aplicação directa do lema de Khinchine leva-nos então a concluir que G e $G^{1/k}$ são do mesmo tipo para todo o $k \ge 1$, i.e.,

Lema 6.4.3. *Se* (6.15) *for válido para todo o $k \ge 1$, sendo G uma f.d. não degenerada, existem constantes $\alpha_k > 0$ e β_k, $k \ge 1$, tais que*

$$G^k(\alpha_k x + \beta_k) = G(x), \ \forall x \in \mathbb{R}. \tag{6.16}$$

Convém pois avançar com a definição de **max-estabilidade**:

Definição 6.4.3 (max-estabilidade)**.** *Uma distribuição G diz-se* max-estável *se existem constantes reais $A_k > 0$ e B_k tais que*

$$G^k(x) = G(A_k x + B_k), \text{ para todo } k.$$

Observação 6.4.1. Se existe uma distribuição limite para o máximo normalizado, então essa distribuição limite terá de ser max-estável.

O lema seguinte completa a demonstração do teorema de Gnedenko e a sua demonstração é na realidade a parte fundamentel da derivação de Gnedenko — obtenção da solução completa da equação funcional em jogo, a equação (6.16).

Lema 6.4.4. *Se G é uma f.d. não degenerada satisfazendo* (6.16)*, $\forall k \geq 1$, então G é uma f.d. de valores extremos de um dos tipos, tipo* I, II *ou* III*, dados em* (6.12)*,* (6.13) *e* (6.14)*, respectivamente.*

Voltando um pouco atrás, convém ainda notar que a mesma demonstração continua válida desde que a validade de (6.15) para $k = 1$ implique a validade de (6.15) para qualquer $k > 1$.

Observações:

1. A resolução da equação funcional (6.16) depende essencialmente de qual das três condições seguintes é a válida.

   (i) $\exists\, k > 1 : \alpha_k < 1 \Longrightarrow \alpha_k < 1,\ \forall k > 1$. Então $\overline{G}(x) = G(x+\beta_k/(1-\alpha_k))$, $\overline{G}(x) = 1,\ x \geq 0$ e $\overline{G}^k(\alpha_k x) = \overline{G}(x),\ x < 0$, tendo-se uma solução de tipo III; ou

   (ii) $\exists\, k > 1 : \alpha_k > 1 \Longrightarrow \alpha_k > 1,\ \forall k > 1$. Então $\overline{G}(x) = G(x+\beta_k/(1-\alpha_k))$, $\overline{G}(x) = 0,\ x \leq 0$ e $\overline{G}^k(\alpha_k x) = \overline{G}(x),\ x > 0$, tendo-se uma solução de tipo II; ou

   (iii) $\exists\ k > 1 : \alpha_k = 1 \Longrightarrow \alpha_k = 1,\ \forall k > 1$. Então $\overline{G}^k(x + \beta_k) = \overline{G}(x),\ x \in \mathbb{R}$, tendo-se uma solução de tipo I.

2. A equação de estabilidade (6.16) foi postulada de modo menos formal por Fréchet (1927), ao observar que o máximo dos $m \times n$ valores $X_1, X_2, \ldots, X_{m\times n}$ é também o máximo dos n valores máximos de $X_{(i-1)m+1}, \ldots, X_{im},\ 1 \leq i \leq n$.

O postulado de estabilidade foi ainda usado por Fisher e Tippett, no seu importante trabalho de 1928, já atrás referido, onde podemos dizer estarem reunidas pela primeira vez as ideias e os problemas mais importantes da Teoria de Valores Extremos clássica.

Para uma demonstração completa do Teorema de Gnedenko veja-se o próprio trabalho de Gnedenko (1943) ou o livro de Galambos (1987).

6.4.2 A distribuição de Valores Extremos e índice de valores extremos

Von Mises (1936; 1954) e Jenkinson[3] (1955) unificaram as três famílias, Gumbel, Fréchet e Max-Weibull, considerando a f.d. geral de valores extremos (GEV, do inglês '*generalized extreme value*' ou '*general extreme value*'), com a expressão funcional,

$$G_\gamma(x) = \exp\left\{-\left[1+\gamma x\right]_+^{-1/\gamma}\right\}, \tag{6.17}$$

com $x_+ := \max(0, x)$ e sendo γ, o EVI, o parâmetro fundamental em *Estatística de Extremos*.

Observação 6.4.2. *Se* $\gamma = 0$, *em* (6.17), *obtemos a f.d.* Gumbel,

$$G_0(x) = \lim_{\gamma\uparrow 0^-} G_\gamma(x) = \lim_{\gamma\downarrow 0_+} G_\gamma(x) = \exp[-\exp(-x)] = \Lambda(x),\ x \in \mathbb{R}.$$

Se $\gamma > 0$, *em* (6.17), *obtemos a f.d.* Fréchet,

$$G_\gamma(x) = \begin{cases} 0, & x < -1/\gamma, \\ \exp\left\{-\left[1+\gamma x\right]^{-1/\gamma}\right\}, & x \geq -1/\gamma. \end{cases}$$

Se $\gamma < 0$, *em* (6.17), *obtemos a f.d.* Max-Weibull,

$$G_\gamma(x) = \begin{cases} \exp\left\{-\left[1+\gamma x\right]^{-1/\gamma}\right\}, & x \leq -1/\gamma, \\ 1, & x > -1/\gamma. \end{cases}$$

[3]Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Royal Meteorol. Soc.* **81**, 158–171.

Observação 6.4.3 (Relação da GEV com os 3 tipos extremais)**.** *Tal como vimos na* Observação 6.4.2*, a* Gumbel *corresponde a* $\gamma = 0$ (*limite por continuidade* $\gamma \uparrow 0^- \ \gamma \downarrow 0_+$)*,* $\Lambda(x) = G_0(x; 0, 1)$*. A* Fréchet *corresponde a* $\gamma > 0$*,* $\alpha = 1/\gamma$*,* $\Phi_\alpha(x) = G_{1/\alpha}(x; 1, 1/\alpha)$*. Finalmente, a* Weibull *corresponde a* $\gamma < 0$*,* $\alpha = -1/\gamma$*,* $\Psi_\alpha(x) = G_{-1/\alpha}(x; -1, -1/\alpha)$*.*

Podemos pois escrever o Teorema de Gnedenko (Teorema 6.4.1), do modo seguinte:

Teorema 6.4.4 (Teorema unificado dos tipos extremais)**.** *Se existem sucessões* $a_n > 0$ *e* b_n*, tais que, quando* $n \to \infty$

$$\mathbb{P}\left[\frac{M_n - b_n}{a_n} \leq x\right] \longrightarrow G(x) \text{ para alguma f.d. } G \text{ não-degenerada,}$$

então G *é do mesmo tipo da distribuição* GEV*, em* (6.17)*, para algum* $\gamma \in \mathbb{R}$*.*

De forma natural, diz-se então que F pertence ao domínio de atracção de G_γ, e escreve-se $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$.

- Se $\gamma < 0$, a cauda curta, $x_F < \infty$.
- Se $\gamma = 0$, temos uma cauda Exponencial , $x_F < \infty$ ou $x_F = \infty$.
- Se $\gamma > 0$, temos uma cauda pesada, $x_F = \infty$.

Observação 6.4.4. *O* Teorema de Gnedenko *não garante a existência de uma lei limite não-degenerada, e não diz qual é o tipo de distribuição limite, i.e., qual é o* γ*. E isto contrariamente com o que se passa com o* TLC*.*

6.4.3 Breve referência ao teorema unificado dos tipos extremais para mínimos

A teoria assintótica para **mínimos**,

$$m_n := \min(X_1, \ldots, X_n)$$

é, devido à Eq. (6.5), uma consequência da teoria assintótica para máximos, tal como referido anteriormente.

Teorema 6.4.5 (tipos extremais para *mínimos*). *Se existem sucessões* $a_n^* > 0$ *e* b_n^*, *tais que, quando* $n \to \infty$

$$\mathbb{P}\left[\frac{m_n - b_n^*}{a_n^*} \leq x\right] \longrightarrow G^*(x) \text{ para alguma f.d. } G^* \text{ não-degenerada,}$$

então G^* *é do mesmo tipo da distribuição* GEV*,

$$G^*_{\gamma^*}(x) = 1 - G_{\gamma^*}(-x) = 1 - \exp\left\{-\left[1 - \gamma^* x\right]_+^{-1/\gamma^*}\right\},$$

para algum $\gamma^* \in \mathbb{R}$.

Os modelos clássicos min-estáveis são pois:

$$\begin{array}{rrl} \text{Tipo I}^*: & \Lambda^*(x) = 1 - \mathrm{e}^{-\mathrm{e}^x},\ x \in \mathbb{R} & \text{[Min-Gumbel]} \\ \text{Tipo II}^*: & \Phi^*_{\alpha^*}(x) = 1 - \mathrm{e}^{x^{-\alpha^*}},\ x < 0,\ \alpha^* > 0 & \text{[Min-Fréchet]} \\ \text{Tipo III}^*: & \Psi^*_{\alpha^*}(x) = 1 - \mathrm{e}^{-x^{\alpha^*}},\ x > 0,\ \alpha^* > 0 & \text{[Weibull].} \end{array}$$

6.4.4 Caracterização de max-domínios de atracção e coeficientes de atracção

Um dos temas de investigação em EVT é a caracterização dos domínios de atracção, i.e., dada uma distribuição limite, caracterizar o conjunto de distribuições F para as quais o máximo convenientemente normalizado converge para esse limite, ou equivalentemente, dada F, encontrar b_n e a_n tais que se verifique a convergência para uma lei não-degenerada, e determinar esse limite. O problema de uma forma geral pode ser complexo. Iremos focar apenas os domínios de atracção para distribuições absolutamente contínuas.

Se $F \in \mathcal{D}_{\mathcal{M}}(G)$, os **coeficientes de atracção** $\{a_n\}_{n\geq 1}$ $(a_n > 0)$ e $\{b_n\}_{n\geq 1}$ de F para a lei limite G, podem ser escolhidos de forma precisa, e o teorema de tipos de Khichine assegura-nos que qualquer outra escolha $\{a'_n\}_{n\geq 1}$ $(a_n > 0)$, $\{b'_n\}_{n\geq 1}$ de coeficientes de atracção será assintoticamente equivalente à escolha inicial se e só se

$$a_n/a'_n \underset{n\to\infty}{\longrightarrow} 1 \qquad \text{e} \qquad \frac{b'_n - b_n}{a_n} \underset{n\to\infty}{\longrightarrow} 0.$$

No que diz respeito à caracterização dos max-domínios de atracção de cada uma das leis limites de máximos, enunciaremos, por ordem crescente de dificuldade, algumas condições necessárias e suficientes para que $F \in \mathcal{D}$(Tipo i), $i = \text{I}, \text{II}$ ou III.

Teorema 6.4.6 (Gnedenko, 1943)**.** *É condição necessária e suficiente para que* $F \in \mathcal{D}_{\mathcal{M}}(\Phi_\alpha)$ *que*

$$\forall x > 0, \quad \frac{1 - F(tx)}{1 - F(t)} \underset{t \to \infty}{\longrightarrow} x^{-\alpha},$$

i.e. $\overline{F} := 1 - F$ *é uma função de variação regular no* ∞*, com expoente* $-\alpha$*, que denotaremos* $1 - F \in RV_{-\alpha}$*. Pode então escolher-se*

$$b_n = 0; \qquad a_n = \inf\{x : F(x) \geq 1 - 1/n\}.$$

Teorema 6.4.7 (Gnedenko, 1943)**.** *É condição necessária e suficiente para que* $F \in \mathcal{D}_{\mathcal{M}}(\Psi_\alpha)$ *que*

$$x^F = \sup\{x : F(x) < 1\} < \infty, \ \text{e } F_0(x) = F(x^F - 1/x) \in \mathcal{D}_{\mathcal{M}}(\Phi_\alpha),$$
$$\text{i.e. } \forall k > 0, \quad \frac{1 - F(x^F - kh)}{1 - F(x^F - h)} \underset{h \downarrow 0}{\longrightarrow} k^\alpha.$$

Pode então escolher-se

$$b_n = x^F; \qquad a_n = \frac{1}{x^F - \inf\{x : F(x) \geq 1 - 1/n\}}.$$

Tem-se ainda

Teorema 6.4.8 (Gnedenko, 1943)**.** $F \in \mathcal{D}_{\mathcal{M}}(\Lambda)$ *se e só se* $x^F < \infty$ *ou* $x^F = \infty$ *e se tem*

$$\lim_{t \uparrow x^F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x},$$

para todo x*, com* $\int_t^{x^F} (1 - F(s))ds < \infty$*, sendo uma escolha possível*

$$g(t) := \frac{\int_t^{x^F} (1 - F(s))ds}{1 - F(t)} = \mathbb{E}[X - t | X > t], \quad \text{para } \ t < x^F.$$

Observação 6.4.5. *À função $g(t)$ chama-se função de excesso médio, como já vimos na* Secção 3.5.

Finalmente:

Teorema 6.4.9 (de Haan[4], 1970)**.** *É condição necessária e suficiente para que $F \in \mathcal{D}_{\mathcal{M}}(\Lambda)$ que*

$$\exists f : \mathbb{R} \to \mathbb{R}^+ \text{ tal que } \lim_{x \to x^F} f(x) = 0 \quad e$$
$$\lim_{x \uparrow x^F} \frac{1 - F(x(1 + hf(x)))}{1 - F(x)} = e^{-h}. \quad (6.18)$$

É útil comentar brevemente estes resultados:

1. Se $x^F < \infty$ então $M_n \xrightarrow{p} x^F$. Isto não impede $(M_n - b_n)/a_n$ de poder ter uma lei limite não degenerada — a não ser, como vimos, se $F(x^F-) < 1$. Na realidade, sendo $x^F < \infty$, $(M_n - b_n)/a_n$ pode convergir para uma lei de valores extremos de tipo I ou de tipo III.

2. As f.d.'s F atraídas para uma lei de valores extremos de tipo II têm sempre $x^F = \infty$ e a sua ponta $1 - F$ tem de ser de variação regular, quando $x \to \infty$, i.e., $1 - F(x) = x^{-\alpha} L(x)$, com $L(\cdot)$ função de variação lenta, i.e. $L \in RV_0$. A condição de atracção para uma lei de tipo III envolve considerações semelhantes na vizinhança de $x^F < \infty$.

Condição necessária e suficiente para $F \in \mathcal{D}(G_\gamma)$, $\gamma \in \mathbb{R}$, em termos de U

Vamos em seguida exprimir, em termos da *função quantil de cauda*

$$U(t) := F^{\leftarrow}(1 - 1/t), \quad F^{\leftarrow}(x) := \inf\{y : F(y) \geq x\}, \quad (6.19)$$

as condições necessárias e suficientes anteriores.

[4] de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes.* Thesis, University of Amsterdam / Mathematical Centre Tract 32.

Teorema 6.4.10 (Condição de primeira ordem para $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, $\gamma \in \mathbb{R}$)**.**

$$F \in \mathcal{D}_{\mathcal{M}}(G_\gamma),\ \gamma \in \mathbb{R} \quad \textit{se e só se} \quad \lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{6.20}$$

para alguma função positiva $a(.)$ e com $x > 0$. O segundo membro é interpretado como $\log x$, para $\gamma = 0$, correspondendo ao limite por continuidade.

Teorema 6.4.11 (Condições necessárias para $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$ $\gamma \in \mathbb{R}$)**.** *Se*

$$F \in \mathcal{D}(G_\gamma), \quad \gamma \in \mathbb{R},\ \textit{i.e., se } \lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad \textit{então:}$$

1. *Se $\gamma > 0$, então $U(\infty) = \infty$ e $\lim_{t\to\infty} \frac{U(t)}{a(t)} = \frac{1}{\gamma}$.*
2. *Se $\gamma < 0$, então $U(\infty) < \infty$ e $\lim_{t\to\infty} \frac{U(\infty)-U(t)}{a(t)} = -\frac{1}{\gamma}$.*
3. *Se $\gamma = 0$, então $\lim_{t\to\infty} \frac{U(tx)}{U(t)} = 1$, para todo $x > 0$ e $\lim_{t\to\infty} \frac{a(t)}{U(t)} = 0$. Além disso, se $U(\infty) < \infty$, $\lim_{t\to\infty} \frac{U(\infty)-U(tx)}{U(\infty)-U(t)} = 1$, para $x > 0$ e $\lim_{t\to\infty} \frac{a(t)}{U(\infty)-U(t)} = 0$. Tem-se ainda que $\lim_{t\to\infty} a(tx)/a(t) = 1$, para $x > 0$.*

Teorema 6.4.12 (Condições necessárias e suficientes para $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma), \gamma \neq 0$)**.** *Podemos garantir as condições seguintes:*

1. *Se $\gamma > 0$, então*

$$F \in \mathcal{D}_{\mathcal{M}}(G_{\gamma>0}) \iff \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad \textit{para } x > 0.$$

2. *Se $\gamma < 0$, tem-se*

$$F \in \mathcal{D}_{\mathcal{M}}(G_{\gamma<0}) \iff U(\infty) < \infty \text{ e } \lim_{t\to\infty} \frac{U(\infty)-U(tx)}{U(\infty)-U(t)} = \lim_{t\to\infty} \frac{x^F - U(tx)}{x^F - U(t)} = x^\gamma, \textit{ para } x > 0.$$

Observação 6.4.6. *Recordemos que quando U verifica a condição* $\lim_{t\to\infty} U(tx)/U(t) = x^\gamma$, *para* $x > 0$ *dizemos que U é de* variação regular *de índice* γ (*no infinito*), *e usamos a notação,* $U \in \mathrm{RV}_\gamma$. *Já no caso de* $a(\cdot)$ *verificar a condição* $\lim_{t\to\infty} a(tx)/a(t) = 1$, *para* $x > 0$, *dizemos que a é de* variação lenta (*no infinito*), *e escrevemos* $a \in RV_0$.

Exemplo 6.4.2 (Distribuições no Domínio Fréchet, $F \in \mathcal{D}(G_\gamma)$, $\gamma > 0$)**.** *Entre outros, apresentamos os seguintes exemplos:*

- Pareto, Pa(α) : $F(x) = 1 - x^{-\alpha},\ x > 1; \alpha > 0$; EVI: $\gamma = 1/\alpha$;
- Generalizada Pareto, GP(σ,γ) : $F(x) = 1 - \left(1 + \gamma\frac{x}{\sigma}\right)^{-\frac{1}{\gamma}},\ x > 0;$ $\sigma, \gamma > 0$; EVI: γ;
- Burr(η,τ,λ) : $F(x) = 1 - \left(\frac{\eta}{\eta+x^\tau}\right)^\lambda,\ x > 0; \quad \eta,\tau,\lambda > 0;$ EVI: $\gamma = 1/(\lambda\tau)$;
- Fréchet(α) : $F(x) = \exp(-x^{-\alpha}),\ x > 0;\ \alpha > 0$; EVI: $\gamma = 1/\alpha$;
- t-Student *com* n g.l.; EVI: $\gamma = 1/n$;
- Cauchy: $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x,\ x \in \mathbb{R}$; EVI: $\gamma = 1$;
- Log-Gama(α,λ) : $F(x) = \int_1^x \frac{\lambda^\alpha}{\Gamma(\alpha)}(\log t)^{\alpha-1}t^{-\lambda-1}dt,\ x \geq 0$; EVI: $\gamma = 1/\lambda$.

Exemplo 6.4.3 (Distribuições no Domínio Max-Weibull, $F \in \mathcal{D}_\mathcal{M}(G_\gamma)$, $\gamma < 0$)**.** *Exemplos clássicos de modelos neste max-domínio, são:*

- Uniforme, $\mathcal{U}(0,1)$: $F(x) = x,\ 0 < x < 1$; EVI: $\gamma = -1$;
- Beta(p,q) : $F(x) = \int_0^x \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}u^{p-1}(1-u)^{q-1}du,\ 0 < x < 1;\ p,q > 0$; EVI: $\gamma = -1/q$;
- Reversed Burr(β,τ,λ) : $F(x) = 1 - \left(\frac{\beta}{\beta+(-x)^{-\tau}}\right)^\lambda;\ x < 0\ ;\ \beta,\tau,\lambda > 0;$ EVI: $\gamma = -1/(\lambda\tau)$;
- Max-Weibull: $F(x) = \exp(-(-x)^\alpha),\ x < 0;\ \alpha > 0$; EVI: $\gamma = -1/\alpha$.

Exemplo 6.4.4 (Distribuições no Domínio Gumbel, $F \in \mathcal{D}_\mathcal{M}(G_\gamma)$, $\gamma = 0$)**.** *Neste caso estão os modelos seguintes:*

- *Exponencial,* $\mathcal{E}(1) : F(x) = 1 - \exp(-x),\ x > 0;$
- Weibull (de mínimos): $F(x) = 1 - \exp(-\lambda x^{\tau}),\ x > 0;\ \lambda, \tau > 0;$
- Logística: $F(x) = 1 - \frac{1}{1+\exp(x)},\ x \in \mathbb{R};$
- Gumbel: $\Lambda(x) = \exp(-\exp(-x)),\ x \in \mathbb{R};$
- Normal: $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt,\ x \in \mathbb{R};$
- Log-normal: $\log X \ \frown$ *Normal;*
- Gama$(\alpha, \beta) : F(x) = \int_0^x \frac{\alpha^{\beta}}{\Gamma(\alpha)} (\log t)^{\beta-1} t^{-\alpha-1} dt, x > 0;$
- Min-Fréchet: $\Phi_{\alpha}^{*}(x) = 1 - \Phi_{\alpha}(-x) = 1 - \exp(-(-x)^{-\alpha}),\ x < 0;\ \alpha > 0.$

Apresentamos em seguida um outro exemplo, que ilustra a existência de modelos absolutamente contínuos, para os quais não é possível normalizar a sucessão de máximos parciais, de modo a obter um comportamento limite não-degenerado.

Exemplo 6.4.5 (Log-Pareto)**.** *Considere-se*

$$F(x) = 1 - 1/\log x, \quad \textit{para} \quad x > e.$$

Para este modelo, vejamos que não existe possibilidade de linearização do máximo de forma a obter uma distribuição limite não-degenerada. Na realidade, uma vez que $x^F = \infty$*, o domínio* Max-Weibull *está descartado. Relativamente à condição necessária e suficiente para o domínio* Fréchet*, tem-se*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t\to\infty} \frac{\log t}{\log(tx)} = 1$$

pelo que falha a condição $\overline{F} \in RV_{-1/\gamma}$*, para algum* $\gamma > 0$*. Resta o domínio Gumbel. Vejamos que no entanto falha a condição* $\int_t^{x^F} (1 - F(s))ds < \infty$*, para todo* $t > e$*. Realmente,* $\int_t^{x^F} (1 - F(s))ds = \int_t^{+\infty} \frac{1}{\log s} ds > \int_t^{+\infty} \frac{1}{s} ds$*, que é divergente. Note-se que se trata de uma cauda direita* **super pesada***, resultado válido para qualquer f.d. com cauda de variação lenta, i.e., tal que* $\overline{F} \in RV_0$*.*

6.4.5 Condições suficientes de von Mises para $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$

As condições de von Mises, a apresentar em seguida, são condições suficientes, mas não necessárias, como são as anteriormente referidas nos teoremas 6.4.6, 6.4.7, 6.4.8 e 6.4.9.

Teorema 6.4.13 (condição suficiente de von Mises)**.** *Para uma distribuição F absolutamente contínua, e admitindo a existência de $f = F'(x)$ e $F''(x)$, defina-se a* função hazard *ou* função de mortalidade instantânea *e o seu* recíproco*, dadas respectivamente, por*

$$h(x) := \frac{f(x)}{1 - F(x)} \qquad e \qquad r(x) := \frac{1 - F(x)}{f(x)}.$$

Se

$$\lim_{x \to x^F} r'(x) = \gamma, \quad \textit{então} \quad F \in \mathcal{D}_{\mathcal{M}}(G_\gamma),$$

com

$$b_n = F^{\leftarrow}(1 - 1/n) = U(n), \qquad a_n = r(b_n) = \frac{1}{nf(b_n)} = nU'(n).$$

Exemplo 6.4.6 (Exponencial, $\mathcal{E}(1)$)**.** *Tem-se $F(x) = 1 - \exp(-x)$, $f(x) = \exp(-x)$, $r(x) = (1 - F(x))/f(x) = 1$, e $r'(x) = 0$ para todo $x > 0$. Então, $F \in \mathcal{D}(G_0)$. O modelo* Exponencial *pertence pois ao max-domínio da* Gumbel*. Quanto às constantes normalizadores, como*

$$b_n = F^{\leftarrow}(1 - 1/n) \quad \Longleftrightarrow \quad 1 - F(b_n) = 1/n$$

tem-se

$$\exp(-b_n) = 1/n \Longleftrightarrow b_n = \log n \quad e \quad a_n = r(b_n) = 1,$$

pelo que

$$(M_n - b_n)/a_n = M_n - \log n$$

converge para a distribuição Gumbel*, como vimos anteriormente, na* Secção 6.2.1.

Exemplo 6.4.7 (Normal, $\mathcal{N}(0,1)$)**.** *Neste caso*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2),$$
$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt, x \in \mathbb{R}, \quad \textit{pelo que} \quad r(x) := \frac{1-\Phi(x)}{\phi(x)}.$$

Comecemos por notar que

$$\phi'(x) = -x\phi(x) \quad e \quad \lim_{x\to\infty} \frac{1-\Phi(x)}{\phi(x)} = \lim_{x\to\infty} \frac{-\phi(x)}{\phi'(x)} = \lim_{x\to\infty} \frac{1}{x}.$$

Então

$$r'(x) = -1 - \frac{(1-\Phi(x))\phi'(x)}{\phi^2(x)} = -1 - \frac{(1-\Phi(x))(-x\phi(x))}{\phi^2(x)} = -1 + \frac{1-\Phi(x)}{\phi(x)}\, x$$

Assim

$$\lim_{x\to\infty} r'(x) = -1 + \lim_{x\to\infty} \frac{1-\Phi(x)}{\phi(x)}\, x = 0,$$

e consequentemente, $F \in \mathcal{D}(G_0)$*, i.e., o modelo* Normal *pertence ao domínio da* Gumbel*. Quanto às constantes normalizadores, embora não trivialmente, mostra-se que se podem considerar*

$$b_n = \sqrt{2\log n} - \frac{\frac{1}{2}\log(4\pi\log n)}{\sqrt{2\log n}} \quad e \quad a_n = \frac{1}{n\phi(b_n)} \sim \frac{1}{b_n} \sim \frac{1}{\sqrt{2\log n}},$$

como referimos anteriormente.

Observação 6.4.7 (Convergência lenta da Normal para a Gumbel)**.** *No modelo* Normal*, a convergência do máximo normalizado para a* Gumbel *é muito lenta. Esta convergência lenta de* $\Phi^n(a_n x + b_n)$ *para* $\Lambda(x)$ *foi observada por* Fisher & Tippett (1928)*, que mostraram que os* 4 *primeiros momentos de* $\Phi^n(a_n x + b_n)$ *estão mais perto dos correspondentes momentos de uma* Max-Weibull Ψ_α*, para um* α *conveniente. Este tipo de estudo está ligado aos chamados comportamentos* pré assintótico *ou* penultimate *dos extremos* (*veja-se* Gomes[5]*,* 1994*, e* Beirlant *et al.*[6]*,* 2012*, para recensões críticas sobre o assunto*).

[5]Gomes, M.I. (1994). Penultimate behaviour of the extremes. In J. Galambos et al. (eds.), *Extreme Value Theory and Applications*, 403-418, Kluwer Academic Publishers.

[6]Beirlant, J., Caeiro, F. & Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* **10**:1, 1–31.

As condições suficientes de von Mises mais simples, são as que são válidas para $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, com $\gamma \neq 0$, que apresentamos em seguida.

Teorema 6.4.14 ($\gamma > 0$)**.** *Suponhamos que $x^F = \infty$ e que $F' = f$ existe. Se para algum γ positivo*

$$\lim_{t\to\infty} \frac{tF'(t)}{1-F(t)} = \lim_{t\to\infty} t\,h(t) = \frac{1}{\gamma}, \quad \textit{então} \quad F \in \mathcal{D}_{\mathcal{M}}(G_\gamma).$$

Teorema 6.4.15 ($\gamma < 0$)**.** *Suponhamos que $x^F < \infty$ e que $F' = f$ existe. Se para algum γ negativo*

$$\lim_{t\uparrow x^F} \frac{(x^F - t)F'(t)}{1-F(t)} = \lim_{t\uparrow x^F} (x^F - t)h(t) = -\frac{1}{\gamma}, \quad \textit{então} \quad F \in \mathcal{D}_{\mathcal{M}}(G_\gamma).$$

Teorema 6.4.16 (constantes normalizadoras em modelo GEV)**.** *Suponhamos que $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$. Então:*

1. *Para $\gamma > 0$:*

$$\lim_{n\to\infty} F^n(a_n x) = \exp\left(-x^{-1/\gamma}\right) = \Phi_{1/\gamma}(x),$$

   para $x > 0$, com $a_n = U(n)$.

2. *Para $\gamma < 0$:*

$$\lim_{n\to\infty} F^n(a_n x + x^F) = \exp\left(-(-x)^{-1/\gamma}\right) = \Psi_{-1/\gamma}(x),$$

   para $x < 0$, com $a_n = x^F - U(n)$.

3. *Para $\gamma = 0$:*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \exp\left(-e^{-x}\right) = \Lambda(x),$$

   para todo o x, com $a_n = g(U(n))$ e $b_n = U(n)$, sendo $g(t) = \mathbb{E}[X - t|X > t]$ a função de excesso médio.

Exemplo 6.4.8 (Uniforme, $\mathcal{U}_{(0,1)}$). *Já que* $x^F = 1 < \infty$ *o domínio* Fréchet *está descartado. Vejamos que este modelo verifica a condição necessária e suficiente para o domínio* Max-Weibull. *Realmente, tem-se que*

$$\lim_{t\downarrow 0} \frac{1-F(x^F - xt)}{1-F(x^F - t)} = \lim_{t\downarrow 0} \frac{1-(1-xt)}{1-(1-t)} = \lim_{t\downarrow 0} \frac{xt}{t} = x,$$

ou seja, $\lim_{t\downarrow 0} \frac{1-F(x^F - xt)}{1-F(x^F - t)} = x^{-\frac{1}{\gamma}}$, *com* $\gamma = -1$, *pelo que* $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, *para* $\gamma = -1$. *Como possíveis constantes normalizadoras temos* $b_n = x^F = 1$ *e* $a_n = x^F - U(n) = 1 - U(n)$. *Como* $U(t) = 1 - 1/t$, *então* $a_n = 1 - (1 - 1/n) = 1/n$, *e* $n(M_n - 1)$ *converge para uma v.a. com distribuição* Max-Weibull, $\Psi_1(x)$.

Exemplo 6.4.9 (Pareto). *Tem-se* $F(x) = 1 - x^{-\alpha}$, *para* $x > 1$ *e* $\alpha > 0$. *Já que* $x^F = \infty$, *o domínio* Max-Weibull *está neste caso descartado. Vejamos que verifica a condição necessária e suficiente para o domínio* Fréchet. *Realmente, tem-se que*

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = \lim_{t\to\infty} \frac{(tx)^{-\alpha}}{t^{-\alpha}} = x^{-\alpha} \Leftrightarrow \overline{F} \in \mathrm{RV}_{-\alpha},$$

pelo que $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, *para* $\gamma = 1/\alpha$. *As constantes normalizadoras podem ser escolhidas como* $b_n = 0$ *e* $a_n = U(n)$. *Como* $U(t) = (1/t)^{-1/\alpha}$, *então* $a_n = n^{1/\alpha}$, *resultando que* $n^{-1/\alpha} M_n$ *converge para uma v.a. com distribuição* Fréchet, $\Phi_\alpha(x)$.

6.4.6 Níveis normalizados e a distribuição limite no modelo Normal

Começamos por enunciar um resultado auxiliar:

Lema 6.4.5. *Seja* $\{\tau_n,\ n \geq 1\}$ *uma sucessão de números reais. Então:*

$$\tau_n \underset{n\to\infty}{\longrightarrow} \tau \quad \Longleftrightarrow \quad \left(1 - \frac{\tau_n}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\tau}.$$

Demonstração. ($\Rightarrow$) Tem-se $\tau_n \underset{n\to\infty}{\longrightarrow} \tau \ \Leftrightarrow \ \tau_n = \tau + o(1), n \to \infty$. Consequentemente,

$$\left(1 - \frac{\tau_n}{n}\right)^n = \left[1 - \frac{\tau}{n} + o(\tfrac{1}{n})\right]^n \underset{n\to\infty}{\longrightarrow} e^{-\tau}.$$

($\Leftarrow$) Tem-se $\left(1-\frac{\tau_n}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\tau} \Rightarrow n\log\left(1-\frac{\tau_n}{n}\right) \underset{n\to\infty}{\longrightarrow} -\tau$. Usando a aproximação $\log(1-x)\sim -x$, com $x\to 0$, tem-se $n(-\tau_n/n) = \underset{n\to\infty}{\longrightarrow} -\tau$, pelo que $\tau_n \underset{n\to\infty}{\longrightarrow} \tau$. □

Seja mais uma vez $\{X_n\}_{n\geq 1}$ uma sucessão de v.a.'s i.i.d. com f.d. comum F.

Definição 6.4.4. *Uma sucessão de números reais $\{u_n(\xi)\}_{n\geq 1}$ diz-se uma sucessão de* **níveis normalizados** *se*

$$1-F(u_n(\xi)) \leq \xi/n \leq 1-F(u_n(\xi)-),$$

i.e.,

$$u_n(\xi) := \inf\{x : 1-F(x) < \xi/n\}. \tag{6.21}$$

Tem-se então o seguinte resultado, devido a O'Brien[7] (1974):

Teorema 6.4.17. *O conjunto de pontos limite de $\mathbb{P}[M_n \leq u_n(\xi)] = F^n(u_n(\xi))$, com $u_n(\xi)$ nível normalizado definido em* (6.21)*, é o intervalo*

$$[e^{-\xi}, e^{-\xi/\delta}], \quad \delta = \lim_{x\to x^F-} \frac{1-F(x-)}{1-F(x)},\ x^F = \sup\{x : F(x) < 1\}.$$

Veremos unicamente um caso particular:

Teorema 6.4.18. *Se para uma sucessão de v.a.'s $\{X_n\}_{n\geq 1}$, for possível escolher $u_n = u_n(\xi),\ \xi > 0,\ n \geq 1$, tal que*

$$1-F(u_n(\xi)) = \xi/n + o(1/n),\ \textit{quando } n\to\infty, \tag{6.22}$$

então, com $M_n = \max_{1\leq i\leq n} X_i$,

$$\lim_{n\to\infty} \mathbb{P}[M_n \leq u_n(\xi)] = e^{-\xi}. \tag{6.23}$$

Reciprocamente, se para uma sucessão $u_n = u_n(\xi)$, $n \geq 1$, (6.23) *for válida, então* (6.22) *também é válida., i.e.*

$$\mathbb{P}[M_n \leq u_n(\xi)] \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\xi} \quad \textit{sse} \quad n(1-F(u_n(\xi)) \underset{n\to\infty}{\longrightarrow} \xi.$$

[7] O'Brien, G. (1974). Limit Theorems for the Maximum Term of a Stationary Process. *Ann. Probab.* **2**(3), 540–545.

Observação 6.4.8. *Note-se que* (6.22) *é possível se e só se* $\delta = 1$, *com* δ *definido no* Teorema 6.4.17. *Se, por exemplo,* $F(j) = 1 - 1/2^j,\ j \geq 1$, *tem-se* $p_j = F(j) - F(j-1) = 1/2^j,\ j \geq 1$ *e* $(1 - F(n-1)/1 - F(n)) = 2$, *i.e.,* $\delta = 2$. *Consequentemente*

$$u_{2^n}(1) = u_{2^n-1}(1) = \cdots = u_{2^{n-1}+1}(1) = n.$$

Tomando a subsucessão 2^n, $2^n\left(1 - F\left(u_{2^n}(1)\right)\right) \underset{n\to\infty}{\longrightarrow} 1$.
Por outro lado, tomando a sucessão $2^{n-1}+1$, *tem-se que*
$(2^{n-1}+1)\left(1 - F\left(u_{2^{n-1}+1}(1)\right)\right) = (2^{n-1}+1)/2^n \underset{n\to\infty}{\longrightarrow} 1/2$. *Consequentemente* $F^n\left(u_{2^n}(1)\right) \underset{n\to\infty}{\longrightarrow} e^{-1}$ *e* $F^n\left(u_{2^{n-1}+1}(1)\right) \underset{n\to\infty}{\longrightarrow} e^{-1/2}$.

Demonstração. Tem-se

$$\mathbb{P}[M_n \leq u_n(\xi)] = [1 - (1 - F(u_n(\xi))]^n = [1 - \xi/n + o(1/n)]^n \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\xi}.$$

Inversamente, se $\mathbb{P}[M_n \leq u_n(\xi)] \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\xi}$, $F^n(u_n(\xi)) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\xi}$ e consequentemente $\mathbb{P}\{\log[1 - (1 - F(u_n(\xi))]\} \underset{n\to\infty}{\longrightarrow} -\xi$. Mas $\log(1-t) = -t(1+o(1))$, quando $t \to 0$, e consequentemente $n[1 - F(u_n(\xi))] \underset{n\to\infty}{\longrightarrow} \xi$. □

Seja agora $\{X_n\}_{n\geq 1}$ uma sucessão de v.a.'s i.i.d. com f.d. comum $\mathcal{N}(0,1)$. Podemos então escolher $u_n : 1 - \Phi(u_n) = \xi/n$, para $\xi > 0$, onde $\Phi(\cdot)$ é a f.d. Normal padrão. Em modelo Normal

$$1 - \Phi(x) = \varphi(x)(1 + o(1)), \quad \text{quando } x \to \infty.$$

e consequentemente

$$\frac{n\ \varphi(u_n)}{u_n} \underset{n\to\infty}{\longrightarrow} \xi.$$

Tomando logaritmos,

$$u_n^2 = 2\log n - 2\log u_n - \log 2\pi - 2\log\xi + o(1).$$

Dividindo por u_n^2, obtemos que

$$\frac{2\ \log n}{u_n^2} \underset{n\to\infty}{\longrightarrow} 1, \quad \text{i.e.} \quad u_n^2 = 2\ \log n(1 + o(1)).$$

Tomando novamente logaritmos:

$$\log 2 + \log\log n - 2\log u_n = o(1), \quad \text{i.e.} \quad \log u_n = \frac{1}{2}\log 2 + \frac{1}{2}\log\log n + o(1),$$

e consequentemente,

$$\begin{aligned} u_n &= \sqrt{2\ \log n - \log\log n - \log 4\pi - 2\log\xi + o(1)} \\ &= (2\ \log n)^{1/2}\left[1 - \frac{\log\log n + \log 4\pi + 2\ \log\xi}{2\ \log n} + o\left(\frac{1}{\log n}\right)\right]^{1/2}. \end{aligned}$$

Por aplicação directa do teorema binomial, que nos garante que $(1+x)^\alpha = 1 + \alpha\ x(1+o(1))$, quando $x \to 0$, e fazendo $\xi = \mathrm{e}^{-x}$, obtemos

$$\begin{aligned} u_n &= (2\ \log n)^{1/2}\left[1 - \frac{1}{2}\frac{\log\log n + \log 4\pi}{2\ \log n} + \frac{x}{2\ \log n} + O\left\{\left(\frac{\log\log n}{\log n}\right)^2\right\}\right] \\ &= \frac{x}{\sqrt{2\ \log n}} + \sqrt{2\ \log n} - \frac{1}{2}(2\ \log n)^{-1/2}[\log\log n + \log 4\pi] + o\left((\log n)^{-1/2}\right). \end{aligned}$$

Podemos pois escrever

$$u_n = a_n x + b_n, \begin{cases} a_n = 1/\sqrt{2\ \log n} \\ b_n = \sqrt{2\ \log n} - \frac{1}{2}(2\ \log n)^{-1/2}[\log\log n + \log 4\pi], \end{cases}$$

tal como mencionámos no final da Secção 6.2.1 e no Exemplo 6.4.7, e tem-se $\mathbb{P}[M_n \le a_n x + b_n + o(a_n)] \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\mathrm{e}^{-x}}$, ou seja

$$\mathbb{P}[M_n \le a_n x + b_n] \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\mathrm{e}^{-x}}$$

Convém realçar que esta técnica de utilização dos níveis normalizados pode ser usada com êxito para encontrar, para qualquer modelo F a lei limite do máximo convenientemente normalizado, que será de tipo I, II ou III dependente de se ter de substituir ξ por e^{-x}, por $x^{-\alpha}$ ou por $(-x)^\alpha$ de modo a obter $u_n = a_n x + b_n + o(a_n)$, seguindo-se de imediato quais as constantes de atracção necessárias.

6.4.7 Carácter poissoniano de excedências de níveis elevados

O Teorema 6.4.18 pode ser generalizado de forma extraordinariamente útil. Na realidade, a dedução da distribuição assintótica de uma estatística extremal pode passar por alguns resultados interessantes, envolvendo o comportamento limite Poisson para o número de excedências de nível e que passaremos a expôr.

Seja mais uma vez $\{X_n\}_{n\geq 1}$ uma sucessão de v.a.'s i.i.d. com f.d. F e admitamos que se pode escolher a sucessão de níveis normalizados $\{u_n = u_n(\xi)\}_{n\geq 1}$ tal que $1 - F(u_n(\xi)) = \xi/n + o(1/n)$, quando $n \to \infty$. Designemos por S_n o número de excedências de u_n por $X_j,\ 1 \leq j \leq n$, i.e.

$$S_n = \#\{j :\ 1 \leq j \leq n,\ X_j > u_n\}.$$

Tem-se obviamente

$$S_n = \sum_{j=1}^{n} I(X_j - u_n), \quad \text{com} \quad I(t) = \begin{cases} 1 & \text{se} \quad t > 0 \\ 0 & \text{se} \quad t \leq 0, \end{cases}$$

ou seja

$$S_n \frown \text{Binomial}(n, 1 - F(u_n)).$$

Como $n(1 - F(u_n)) \underset{n\to\infty}{\longrightarrow} \xi$, tem-se que S_n é assintoticamente uma v.a. Poisson(ξ). Consequentemente,

Teorema 6.4.19. *Se S_n denota o número de excedências da sucessão de níveis normalizados $u_n(\xi)$, definidos em (6.22), por X_j $1 \leq j \leq n$, com f.d. F, então*

$$\lim_{n\to\infty} \mathbb{P}[S_n = k] = e^{-\xi}\ \xi^k/k!,\ k = 0, 1, 2, \ldots$$

Consequentemente,

Corolário 6.4.1. *Qualquer que seja a forma de F, e desde que*

$$\lim_{x\uparrow x^F} \frac{1 - F(x-)}{1 - F(x)} = 1$$

as excedências de níveis normalizados elevados têm assintoticamente carácter poissoniano.

6.4.8 Distribuição assintótica de $X_{k:n}$ e $X_{n-k+1:n}$, k fixo

Os casos anteriormente estudados do mínimo, $X_{k:n} = X_{1:n}$, $k = 1$, e do máximo, $X_{k:n} = X_{n:n}$, $n - k = 0$, ou, $X_{n-k+1:n} = X_{n:n}$, $k = 1$, são casos particulares de e.o.'s extremais. Designemos por

$$M_n^{(k)} = X_{n-k+1:n} \quad (M_n^{(1)} \equiv M_n)$$

o k-ésimo máximo de $(X_1, \ldots, X_n)$, $n \geq k$. Como relacionar a f.d. de $M_n^{(k)}$ com a de S_n? Tem-se obviamente

$$P\left[M_n^{(k)} \leq u_n\right] = P\left[S_n < k\right],$$

e consequentemente,

$$P\left[M_n^{(k)} \leq u_n\right] \underset{n\to\infty}{\longrightarrow} e^{-\xi} \sum_{s=0}^{k-1} \xi^s/s!,$$

ou seja

Teorema 6.4.20. *Se $\{u_n\}_{n\geq 1}$ é uma sucessão de números reais tal que*

$$P\left[M_n \leq u_n\right] \underset{n\to\infty}{\longrightarrow} e^{-\xi}$$

então

$$P\left[M_n^{(k)} \leq u_n\right] \underset{n\to\infty}{\longrightarrow} e^{-\xi} \sum_{s=0}^{k-1} \xi^s/s!, \ \forall k > 1, \ \textit{inteiro fixo.}$$

Podemos agora levar um pouco mais adiante esta linha de raciocínio de forma a obter a distribuição assintótica de $M_n^{(k)}$, $k > 1$, a partir da de M_n. Especificamente, se existem coeficientes de atracção $\{a_n\}_{n\geq 1}$ ($a_n > 0$) e $\{b_n\}_{n\geq 1}$ e $G(\cdot)$ não degenerada, tais que

$$\lim_{n\to\infty} \mathbb{P}[M_n \leq a_n x + b_n] = G(x),$$

e se $0 < G(x) < 1$, denotando $a_n x + b_n =: u_n(\xi)$, $\xi = -\log G(x)$, tem-se que, pelo Teorema 6.4.20,

$$P\left[M_n^{(k)} \leq a_n x + b_n\right] \underset{n\to\infty}{\longrightarrow} G(x) \sum_{s=0}^{k-1} [-\log G(x)]^s/s!, \ \forall k > 1, \ \text{inteiro fixo.}$$

Mas, como G é contínua, o limite é 0 se $G(x) = 0$ e é 1 se $G(x) = 1$, i.e., temos a validade do

Teorema 6.4.21 (distribuição assintótica da k-ésima e.o. superior)**.** *Considere-se $M_n^{(k)} = X_{n-k+1:n}$ ($M_n^{(1)} \equiv M_n$), k fixo. Então, se existem $a_n > 0$ e b_n tais que $F \in \mathcal{D}_{\mathcal{M}}(G)$, i.e.,*

$$\lim_{n\to\infty} \mathbb{P}[M_n \leq a_n x + b_n] = G(x) \ \textit{não-degenerada}$$

$$\Longleftrightarrow$$

$$\lim_{n\to\infty} \mathbb{P}[M_n^{(k)} \leq a_n x + b_n] = G(x) \sum_{i=0}^{k-1} \frac{[-\log G(x)]^i}{i!}.$$

Observação 6.4.9. *As constantes normalizadoras para* $M_n^{(k)} \equiv X_{n-k+1:n}$ *são as mesmas que as obtidas para o máximo* (*ou assintoticamente equivalentes*). *A f.d.* G *é a que figura no* max-domínio de atracção. *A f.d.p. limite correspondente, obtem-se por derivação da f.d., sendo dada por*

$$\frac{\{-\log G(x)\}^{k-1}}{(k-1)!} g(x)\,.$$

A concretização da f.d. limite para a extremal superior passa, evidentemente, pela substituição de G *por* Λ, Φ_α *ou* Ψ_α. *Alternativamente,* G *pode ser substituída pela* GEV, G_γ.

Embora estejamos mais focados nos grandes valores, para o caso da extremal inferior, $X_{k:n}$, k fixo, temos os resultados correspondentes:

Teorema 6.4.22 (distribuição assintótica da k-ésima e.o. inferior)**.** *Considere-se* $X_{k:n}$, k *fixo. Então, se existem* $a_n^* > 0$ *e* b_n^* *tais que* $F \in \mathcal{D}_{\rm m}(G^*)$, *i.e.,*

$$\lim_{n\to\infty} \mathbb{P}[X_{1:n} \leq a_n^* x + b_n^*] = G^*(x) \text{ não-degenerada}$$

$$\Longleftrightarrow$$

$$\lim_{n\to\infty} \mathbb{P}[X_{k:n} > a_n^* x + b_n^*] = (1 - G^*(x)) \sum_{i=0}^{k-1} \frac{[-\log(1 - G^*(x))]^i}{i!}\,,$$

sendo a f.d.p. limite dada por

$$\frac{\{-\log(1 - G^*(x))\}^{k-1}}{(k-1)!} g^*(x)\,.$$

6.4.9 Distribuição assintótica conjunta de estatísticas ordinais superiores e inferiores

Teorema 6.4.23 (distribuição assintótica das k e.o.'s superiores)**.**

$$F \in \mathcal{D}_{\mathcal{M}}(G),\ \textit{para constantes } b_n, a_n > 0,\ \textit{com } G' = g$$

se e só se para k fixo, o k-vector

$$\left(\frac{X_{n:n} - b_n}{a_n}, \ldots, \frac{X_{n-k+1:n} - b_n}{a_n} \right)$$

tem distribuição limite não-degenerada, com f.d.p. conjunta

$$g_{1,\ldots,k}(w_1, \ldots, w_k) := G(w_k) \prod_{i=1}^{k} \frac{g(w_i)}{G(w_i)},\ \textit{para } w_1 > \cdots > w_k. \tag{6.24}$$

Teorema 6.4.24 (distribuição assintótica das k e.o.'s inferiores)**.**

$$F \in \mathcal{D}_{\mathrm{m}}(G^*),\ \textit{para constantes } {b_n}^*, {a_n}^* > 0,\ \textit{com } {G^*}' = g^*$$

se e só se para k fixo, o k-vector

$$\left(\frac{X_{1:n} - b_n^*}{a_n^*}, \ldots, \frac{X_{k:n} - b_n^*}{a_n^*} \right)$$

tem distribuição limite com f.d.p. conjunta

$$g^*_{1,\ldots,k}(w_1, \ldots, w_k) := \{1 - G^*(w_k)\} \prod_{i=1}^{k} \frac{g^*(w_i)}{1 - G^*(w_i)},\ \textit{para } w_1 < \cdots < w_k\ .$$

Exemplo 6.4.10 ($X \frown \text{Pareto}(\alpha)$)**.** *Considere-se a f.d. de Pareto $F(x) = 1 - x^{-\alpha},\ x \geq 1, \alpha > 0$. Então, para k fixo, com $x > 0$,*

$$\mathbb{P}[X_{n-k+1} \leq n^{1/\alpha} x] \approx \exp\{-x^{-\alpha}\} \sum_{j=0}^{k-1} \frac{x^{-j\alpha}}{j!}.$$

O vector de observações extremais superiores

$$\left(\frac{X_{n:n}}{n^{1/\alpha}}, \ldots, \frac{X_{n-k+1:n}}{n^{1/\alpha}} \right)$$

tem f.d.p. conjunta limite

$$g_{1,\ldots,k}(w_1,\ldots,w_k) = \exp\{-w_k^{-\alpha}\}\prod_{i=1}^{k}\frac{\alpha}{w_i^{\alpha+1}}\,,$$

para $w_1 > \cdots > w_k > 0$.

Exemplo 6.4.11 ($X \frown \mathcal{E}(1)$)**.** *Considere-se a f.d. da Exponencial reduzida* $F(x) = 1-\exp(-x)$, $x \geq 0$. *Vimos anteriormente que, quando* $n \to \infty$, $\mathbb{P}[X_{n:n} - \log n \leq x]$ *converge para* $\Lambda(x) = \exp[-\exp(-x)]$, *para todo o real* x. *Por outro lado, para o mínimo é verificada a estabilidade* $\mathbb{P}[nX_{1:n} \leq x] = 1-\exp(-x)$, $x > 0$. *Então, para* k *fixo, e para* $x \geq 0$, $\mathbb{P}[X_{n-k+1:n} - \log n \leq x]$ *converge para*

$$\exp[-\exp(-x)]\sum_{i=0}^{k-1}\frac{\exp(-ix)}{i!}\,.$$

Analogamente, $\mathbb{P}[nX_{k:n} \leq x]$ *converge para*

$$1-\exp(-x)\sum_{i=0}^{k-1}\frac{x^i}{i!}\,.$$

Exemplo 6.4.12 ($U \frown \mathcal{U}(0,1)$)**.** *Neste modelo, considerando* $\{U_i\}_{i=1}^n$ *i.i.d. a* U, *para* $k \geq 1$, *fixo, tem-se*

$$n\,U_{k:n} \overset{d}{=} n\,(1-U_{n-k+1:n}) \xrightarrow[n\to\infty]{d} \text{Gama}(k)\,.$$

Realmente, pela propriedade uniformizante,

$$1-\exp(-E) \overset{d}{=} U, \quad E \frown \mathcal{E}(1).$$

Por outro lado,

$$1-U \overset{d}{=} U,$$

pelo que

$$\exp(-E_{k:n}) \overset{d}{=} 1-U_{k:n} \overset{d}{=} U_{n-k+1:n}$$

ou

$$E_{k:n} \overset{d}{=} -\log U_{n-k+1:n}.$$

$$\begin{aligned}\mathbb{P}[n(1-U_{n-k+1:n}) \le x] &= \mathbb{P}[-\log U_{n-k+1:n} \le -\log(1-x/n)] \\ &= \mathbb{P}[E_{k:n} \le -\log(1-x/n)] \\ &= \mathbb{P}[E_{k:n} \le x/n] + o(1)\end{aligned}$$

pela continuidade da f.d. Gama *e uma vez que* $-\log(1-x/n) \sim x/n, n \to \infty$. *Assim*

$$\mathbb{P}[n(1-U_{n-k+1:n}) \le x] = \mathbb{P}[n\,E_{k:n} \le x] + o(1) = \int_0^x \frac{e^{-t}t^{k-1}}{\Gamma(k)}dt + o(1),$$

verificando-se então

$$n\,U_{k:n} \stackrel{d}{=} n\,(1-U_{n-k+1:n}) \xrightarrow[n\to\infty]{d} \text{Gama}(k)\,.$$

6.4.10 Teorema Pickands-Balkema-de Haan

Pickands[8] (1975) e Balkema & de Haan[9] (1974), estabeleceram a dualidade entre a GEV(γ) e a GP(γ), propondo para efeitos de aplicação prática, e para níveis elevados u, a aproximação

$$F_u(y) \approx H_\gamma(y;\sigma_u),$$

para $y \in [0, x^F - u]$, se $\gamma \ge 0$, e para $y \in [0, -\sigma_u/\gamma]$, se $\gamma < 0$. A f.d. F_u descreve a distribuição dos excessos acima de um nível u, dado que u é excedido, i.e., da v.a. $Y|Y > 0$, com $Y = X - u$.

Teorema 6.4.25 (Teorema Pickands-Balkema-de Haan)**.**

$$F \in \mathcal{D}(G_\gamma),\, \gamma \in \mathbb{R} \quad \Longleftrightarrow \quad \lim_{u\to x^F} \sup_{0<y<x^F-u} |F_u(y) - H_\gamma(y;\sigma_u)| = 0.$$

[8] Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.

[9] Balkema, A.A. & de Haan, L. (1974). Residual life time at great age. *Annals of Probability* **2**, 792–804.

6.5 Comportamento limite de estatísticas ordinais intermédias

Consideremos agora o caso de uma sucessão intermédia em que se permite também que

$$k \equiv k(n) \to \infty$$

mas com uma *velocidade inferior a* n, $k = o(n)$, i.e., com $n \to \infty$,

$$\frac{k}{n} \to 0\,.$$

Dizemos que $X_{n-k+1:n}$ é uma e.o. intermédia superior e $X_{k:n}$ é uma e.o. intermédia inferior.

Teorema 6.5.1 (distribuição assintótica de uma e.o. intermédia)**.** *Suponhamos que as condições de* von Mises*, enunciadas na* Secção 6.4.5*, se verificam para algum* domínio de atracção*. Denote-se* $f := F'$*. Então, se* $k \equiv k(n) \to \infty$*, com* $k/n \to 0$ *quando* $n \to \infty$*,*

$$\frac{X_{n-k+1:n} - b_n}{a_n} \xrightarrow[n\to\infty]{d} Z \frown \mathcal{N}(0,1)\,,$$

com constantes normalizadoras dadas por

$$b_n = F^{\leftarrow}(1 - k/n) = U\left(\frac{n}{k}\right) \quad e \quad a_n = \frac{\sqrt{k}}{n\, f(b_n)} = \frac{\frac{n}{k} U'\left(\frac{n}{k}\right)}{\sqrt{k}}$$

com $U(\cdot)$ *a função em* (6.19) *e* U' *a sua derivada.*

6.6 Breve referência a generalizações a esquemas originais não i.i.d.

É natural perguntar qual a robustez dos resultados limite anteriormente apresentados, e como resposta podemos afirmar que tal robustez, embora não total, é razoável de um ponto de vista de aplicações. Admitamos, por exemplo, que $\{X_n\}_{n\geq 1}$ é uma sucessão estacionária, i.e.,

$$\forall n, \{j_1, j_2, \ldots, j_n\}\,, \text{ e } m \geq 1,$$
$$(X_{j_1}, \ldots, X_{j_n}) \stackrel{d}{=} (X_{j_1+m}, \ldots, X_{j_n+m})\,. \quad (6.25)$$

Em que condições iremos obter as mesmas leis limites para M_n? É razoável admitir que para que tal aconteça a dependência entre X_i e X_j deve decrescer quando $|i-j|$ aumenta.

O exemplo mais simples de tal situação é o da m-dependência, adoptada no trabalho de Watson[10] (1954).

Definição 6.6.1. *A sucessão $\{X_n\}_{n\geq 1}$ é m-dependente se e só se X_i e X_j são independentes quando $|i-j|>m$.*

Um outro exemplo usual, que foi abordado pela primeira vez no contexto de valores extremos por Loynes[11] (1965), é o de sucessões de **mistura-forte**, tradução directa da terminologia inglesa *strong-mixing*.

Definição 6.6.2. *A sucessão $\{X_n\}_{n\geq 1}$ é de mistura-forte (strong-mixing) se e só se existe uma função de mistura $g(k) \underset{k\to\infty}{\longrightarrow} 0$, e $\forall m \geq 1$, $k \geq 1$, $A \in \mathcal{B}(X_1,\ldots,X_m)$, $B \in \mathcal{B}(X_{m+k},\ldots)$,*

$$|P(A\cap B) - P(A)P(B)| \leq g(k).$$

Um outro exemplo ainda, hoje em dia clássico, foi o introduzido por Leadbetter[12] (1973), e usualmente conhecido por D-dependência. Trata-se de uma condição mais fraca que a condição de mistura-forte, e que é suficiente para obter a mesma equação funcional (6.16) para a distribuição limite do máximo convenientemente normalizado.

Definição 6.6.3. *A sucessão $\{X_n\}_{n\geq 1}$ satisfaz a condição de D-dependência se e só se existe uma função $g(k) \underset{k\to\infty}{\longrightarrow} 0$, e*
$\forall \{i_1 < i_2 < \cdots < i_n < j_1 < \cdots < j_m\}$ tais que $j_1 - i_n \geq k$, e $\forall u \in \mathbb{R}$

$$|F_{i_1,\ldots,i_n,j_1,\ldots,j_m}(u,\ldots,u) - F_{i_1,\ldots,i_n}(u,\ldots,u)\ F_{j_1,\ldots,j_m}(u,\ldots,u)| \leq g(k).$$

[10] Watson, G.S. (1954). Extreme values in samples from m-dependent stationary stochastic processes. *Ann. Math. Statist.* **25**:4, 798–800.

[11] Loynes, R.M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36**:3, 993–999.

[12] Leadbetter, M.R. (1973). On extreme values in stationary sequences. *Z. Wahrsch. und Verw. Gebiete* **28**, 289–303.

Tem-se obviamente

$$\{X_n\}_{n\geq 1} \text{ é } m\text{-dependente} \Longrightarrow \{X_n\}_{n\geq 1} \text{ é de mistura-forte,}$$

e

$$\{X_n\}_{n\geq 1} \text{ é de mistura-forte} \Longrightarrow \{X_n\}_{n\geq 1} \text{ é } D\text{-dependente.}$$

É possível demonstrar que, para sucessões estacionárias verificando qualquer uma destas condições de dependência fraca, o teorema de Gnedenko continua válido — i.e., chegamos à mesma equação funcional para a lei limite G, do máximo linearmente normalizado.

Impondo uma condição adicional na estrutura bivariada do processo, por exemplo, a condição usualmente designada por condição-D', i.e., admitindo que

$$\lim_{n\to\infty} \sup n \sum_{j=2}^{n} P\left(X_1 > u_{nk}, X_j > u_{nk}\right) = o(1/k), \text{ quando } k \to \infty,$$

para uma sucessão conveniente $\{u_n\}_{n\geq 1}$, garante-se que a distribuição limite de M_n, convenientemente normalizado, é exactamente a mesma que surgiria se considerássemos o máximo, M_n^* da sucessão i.i.d. associada, i.e., da sucessão $\{Y_n\}_{n\geq 1}$ de v.a.'s i.i.d. com f.d. F, a marginal de X_n, $\forall n \geq 1$. As constantes de atracção são as mesmas num e noutro caso.

Em situações de esquema original não identicamente distribuído, mas independente, por exemplo T-periodicidade, continuam frequentemente válidas as distribuições limite anteriormente apresentadas.

Sob condições convenientes de dependência fraca aparecem também os mesmos tipos de leis limite (*max-estáveis*) para o máximo convenientemente normalizado, mas surge um parâmetro adicional com interesse — o chamado *índice extremal*, fácil de relacionar com as excedências de níveis elevados $u_n : \ F(u_n) = 1 - \xi/n + o(1/n)$, quando $n \to \infty$, já definidos em (6.22). Qual o comportamento assintótico das excedências de níveis elevados? Seja $\{X_n\}_{n\geq 1}$ uma sucessão de v.a.'s i.i.d. Então, tal como vimos na Secção 6.4.7, qualquer processo pontual limite de excedências de níveis elevados, após normalização no tempo, é um **Processo de Poisson**.

Designemos então por $\{Y_n\}_{n\geq 1}$ uma sucessão de v.a.'s estacionária e dependente, verificando condições convenientes e gerais de independência local e

assintótica. Então, qualquer processo pontual limite de excedências de níveis elevados, após normalização no tempo, é um **Processo de Poisson composto**, i.e. existe um agrupamento de excedências elevadas: as posições dos grupos continuam a ser representadas por um processo de Poisson, mas a esses pontos estão associadas multiplicidades, que correspondem aos tamanhos dos grupos. Para essa classe de sucessões estacionárias existe e está bem definido o chamado *índice extremal*, θ, $0 \leq \theta \leq 1$, que goza de um papel importante na determinação dos tamanhos dos grupos de excedências, pois em muitas situações interessantes do ponto de vista prático (não em geral!) o *índice extremal* é o recíproco do limite do tamanho médio dos grupos de excedências.

Temos obviamente $\theta = 1$ para sucessões i.i.d., i.e. as excedências de níveis elevados aparecem individualmente, sendo $\theta > 0$ para 'quase todos' os casos de interesse. Atentemos nas trajectórias seguintes, apresentadas na Figura 6.5, associadas às seguintes sucessões:

(a) $\{X_n\}_{n\geq 1}$ é uma sucessão de v.a.'s i.i.d. provenientes de um modelo Exponencial: $F(x) = (1 - \exp(-x))^2$, $x \geq 0$,

(b) $\{Y_n\}_{n\geq 1}$ é uma sucessão 2-dependente, $Y_n = \max(Z_{n-1}, Z_n)$, $n \geq 1$, onde $\{Z_n\}_{n\geq 1}$ v.a.'s i.i.d. provenientes de $H(z) = 1 - \exp(-z)$, $z \geq 0$, sendo consequentemente $F(y)$ também $(1 - \exp(-y))^2$, $y \geq 0$.

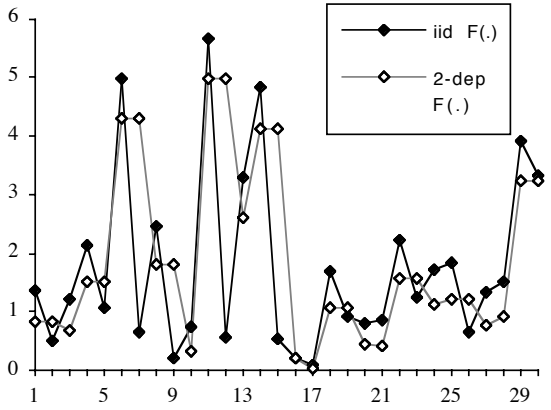


Figura 6.5: Trajectórias amostrais de um processo i.i.d. e de outro 2-dependente

É evidente o tamanho 2 dos grupos de excedências de níveis elevados, o que vai implicar um índice extremal igual a 1/2 para esta sucessão. É também evidente a redução sofrida pelos maiores valores quando passamos da sucessão i.i.d para a sucessão 2-dependente, proveniente do mesmo modelo F.

E qual a influência do *índice extremal* no comportamento limite das e.o.'s superiores?

Considere-se um nível elevado, tal que:

$$\lim_{n\to\infty} n(1-F(u_n(\xi))) = \xi \quad \text{sse} \quad F(u_n) = 1 - \frac{\xi}{n} + o(1/n), \text{ quando } n \to \infty.$$

Seja

$$\begin{aligned} N_n(B) &= \sum_{i=1}^{n} \epsilon_{i/n}(B) I_{[X_i > u_n]} \\ &= \#\{i/n, 1 \le i \le n : i/n \in B \text{ e } X_i > u_n\}, \ B \in \mathcal{B}(0,1) \end{aligned} \quad (6.26)$$

o *processo pontual das excedências*, e

$$\widetilde{N}_n(B) = \sum_{i=1}^{n} \epsilon_{i/n}(B) I_{[X_i \le u_n \le X_{i+1}]}, \ B \in \mathcal{B}, \quad (6.27)$$

o *processo pontual dos cruzamentos*. Então, em situação i.i.d.,

$$N_n \to N, \text{ Processo de Poisson}(\xi) \quad \text{e} \quad \widetilde{N}_n \to \widetilde{N}, \text{ Processo de Poisson}(\xi).$$

e

$$\widetilde{N}_n \to \widetilde{N}, \quad \text{Processo de Poisson}(\xi).$$

Em situação de dependência fraca,

$$\widetilde{N}_n \to \widetilde{N}, \text{ Processo de Poisson}(\theta\xi) \quad \text{e} \quad N_n \to N = \sum_{i=1}^{\widetilde{N}} Z_i.$$

Seja $\{X_n\}_{n\ge1}$ uma sucessão i.i.d. proveniente de um modelo $F(\cdot)$ e $\{Y_n\}_{n\ge1}$ uma sucessão dependente estacionária, sob a validade de restrições gerais de dependência, que nos permitem assegurar a existência de um índice extremal θ. Seja também $F(\cdot)$ a f.d. marginal de Y_n, $n \ge 1$. Consideremos a notação,

$$M_n^X = M_{n,X}^{(1)} = \max_{1\le i\le n} X_i, \qquad M_n^Y = M_{n,Y}^{(1)} = \max_{1\le i\le n} Y_i.$$

Para qualquer $\xi > 0$, seja $u_n = u_n(\xi)$, $n \geq 1$, o quantil de probabilidade $(1 - \xi/n)$ de $F(\cdot)$, i.e., um nível tal que $\lim_{n\to\infty} n(1 - F(u_n)) = \xi$.

Seja $S_n^X = \sum_{j=1}^n I_{[Xj>u_n]}$ e $S_n^Y = \sum_{j=1}^n I_{[Yj>u_n]}$. Então S_n^X é Binomial e converge para uma v.a. Poisson, quando $n \to \infty$, enquanto S_n^Y vai convergir para uma v.a. Poisson composta. Em particular:

$$\mathbb{P}[M_n^X \leq u_n] = \mathbb{P}[S_n^X = 0] \to \exp(-\xi),$$

onde ξ é a intensidade das posições limites de Poisson, e

$$\mathbb{P}[M_n^Y \leq u_n] = \mathbb{P}[S_n^Y = 0] \to \exp(-\theta\xi),$$

i.e. a intensidade passa a ser $\theta\xi < \xi$.

Consequentemente, se existirem $\{a_n\}_{n\geq 1}$ $(a_n > 0)$, $\{b_n\}_{n\geq 1}$, e uma v.a. não degenerada L tal que

$$\frac{M_n^X - b_n}{a_n} \to L, \quad \text{quando} \quad n \to \infty,$$

então L tem f.d. GEV, $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$, e

$$\frac{M_n^Y - b_n}{a_n} \to L^*, \quad \text{quando} \quad n \to \infty,$$

tendo L^* f.d. $G_\gamma^\theta(x)$.

Isto significa que para estruturas dependentes temos um '*shrinkage*' de valores máximos, não sendo no entanto modificado o tipo da lei limite, pois $G_\gamma(x)$ é estável para máximos, e consequentemente,

$$G_\gamma^\theta(x) = G_\gamma\Big(\frac{x - \lambda_\theta}{\delta_\theta}\Big), \quad \lambda_\theta = \frac{\theta^{-\gamma} - 1}{\gamma}, \quad \delta_\theta = \theta^{-\gamma}.$$

Contudo, o mesmo não acontece para outras e.o.'s superiores. Se denotarmos $M_{n,X}^{(k)}$ a k-ésima maior e.o. por entre $(X_1, \ldots, X_n)$ e por $M_{n,Y}^{(k)}$ a k-ésima maior e.o. por entre $(Y_1, \ldots, Y_n)$,

$$\mathbb{P}[M_{n,X}^{(k)} \leq u_n] = \mathbb{P}[S_n^X < k] \to e^{-\xi} \sum_{j=0}^{k-1} \frac{\xi^j}{j!},$$

enquanto

$$\mathbb{P}[M_{n,Y}^{(k)} \leq u_n] = \mathbb{P}[S_n^Y < k] \to e^{-\theta\xi} \sum_{j=0}^{k-1} \frac{(\theta\xi)^j}{j!} \sum_{i=j}^{k-1} \pi^{*j}(i),$$

onde $\pi^{*j}(\cdot)$ é a j-ésima convolução de uma f.m.p. $\pi(.)$, a f.m.p. limite do tamanho dos grupos $\left(\pi^{*0}(i) = 1, \forall i \geq 1\right)$, i.e.

$$\pi(j) = \lim_{n\to\infty} \pi_n(j), \text{ com}$$
$$\pi_n(j) = \mathbb{P}\left[\sum_{i=1}^{r_n} I[Y_i > u_n] = j \middle| \sum_{i=1}^{r_n} I[Y_i > u_n] > 0\right]$$

(distribuição do número de acontecimentos num grupo de excedência condicional a existir pelo menos uma excedência).

Este resultado dá origem às seguintes leis limite para a k-ésima maior e.o., em esquema i.i.d.,

$$G^{(k)}_{\gamma,X}(x) = G_\gamma(x) \sum_{j=0}^{k-1} \frac{(-\log G_\gamma(x))^j}{j!}, \tag{6.28}$$

enquanto em esquema de dependência fraca,

$$G^{(k)}_{\gamma,Y}(x) = G^\theta_\gamma(x) \sum_{j=0}^{k-1} \frac{(-\theta \log G_\gamma(x))^j}{j!} p_{j,k}. \tag{6.29}$$

Repare-se que a f.d. limite de $M^{(k)}_{n,Y}$ pode ser expressa como uma mistura das f.'s d. limite das i maiores e.o.'s, $1 \leq i \leq k$, numa estrutura i.i.d. em que o valor máximo é atraído para $G^\theta_\gamma(x) = G_\gamma\left(\frac{x-\lambda_\theta}{\delta_\theta}\right)$, i.e., pode ser expressa como

$$\sum_{j=1}^{k} \alpha_j G^{(j)}_{\gamma,X}\left(\frac{x-\lambda}{\delta}\right), \alpha_j \geq 0, \sum_{j=1}^{k} \alpha_j = 1,\ \lambda_\theta = \frac{\theta^{-\gamma}-1}{\gamma},\ \delta_\theta = \theta^{-\gamma},$$
$$\alpha_1 = 1 - p_{1,k} = 1 - (\pi_1 + \cdots + \pi_{k-1}),\ \alpha_j = p_{j-1,k} - p_{j,k},$$
$$\text{para} \quad 2 \leq j \leq k-1,\ \alpha_k = p_{k-1,k}. \tag{6.30}$$

A diferença é ainda mais acentuada quando consideramos a distribuição conjunta de $(M^{(1)}_n, \ldots, M^{(k)}_n)$ em ambas as situações, i.e., a presença de agrupamentos de excedências de níveis elevados não afecta o tipo da distribuição do máximo, mas afecta a distribuição das outras e.o.'s, em virtude do facto de o segundo máximo poder ocorrer no mesmo grupo que o máximo.

Em *Estatística de Extremos*, como facilmente se depreende do que anteriormente foi dito, só a abordagem clássica de Gumbel não é, numa fase inicial, drasticamente afectada pela existência de um índice extremal diferente de 1. Todos os outros modelos são seriamente perturbados se a estrutura subjacente for dependente, tal como foi mencionado atrás. É então necessário ter boas estimativas do índice extremal, e mesmo das probabilidades $\pi_i,\ i \geq 1$, para prosseguir qualquer inferência estatística.

6.7 Breve referência à distribuição assintótica de funções lineares de estatísticas ordinais

Como se viu anteriormente, se $k = k_n \to \infty$ e $k_n/n \to \lambda,\ 0 < \lambda < 1$, então sob condições de regularidade pouco restrictivas, $X_{k_n:n} \overset{a}{\frown} \text{Normal}$. Consequentemente, qualquer combinação linear de um número finito de tais quantis empíricos é ainda assintoticamente Normal.

Por outro lado, para i, inteiro fixo, $X_{i:n}$ e $X_{n-i+1:n}$ têm distribuições limite não-normais. No entanto,

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} X_{i:n} \overset{a}{\frown} \text{Normal},$$

e $\overline{X}_n$ envolve e.o.'s extremais com peso igual ao das e.o.'s centrais.

Coloca-se pois de imediato a pergunta: em que condições uma **estatística sistemática** ou **estatística-L**,

$$T_n = \sum_{i=1}^{n} a_{in} X_{i:n} \tag{6.31}$$

é assintoticamente Normal?

Diferentes conjuntos de condições têm sido impostos de forma a garantir a normalidade assintótica de T_n em (6.31) — uns dão maior importância aos $a_{in},\ 1 \leq i \leq n$, outros a $F(x)$.

O resultado de normalidade assintótica de estatísticas-L mais manejável é

talvez o devido a Stiegler[13] (1974). Para a derivação do resultado, a qual não faremos, torna-se necessário exprimir T_n do seguinte modo:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} J(i/n)X_{i:n} = \int_{\mathbb{R}} xJ(\widehat{F}_n(x))d\widehat{F}_n(x),$$

onde $J(u)$ é uma função de u, $0 \leq u \leq 1$, tal que $J(i/n) = n\ a_{in}$, $1 \leq i \leq n$. Tem-se então:

Teorema 6.7.1. *Se $E(X^2) < \infty$ e $J(u)$ é limitada e contínua a.e. $F^{\leftarrow}$, e se além disso*

$$\lim_{n\to\infty}\ n\mathbb{V}ar[T_n] = 2\iint\limits_{x<y} J(F(x))\ J(F(y))\ F(x)\ (1-F(y))dxdy > 0$$

então

$$\frac{T_n - \int_{\mathbb{R}} xJ(F(x))dF(x)}{\sqrt{\mathbb{V}ar(T_n)}} \xrightarrow{d} Z \frown \text{Normal}(0,1).$$

Outras condições razoavelmente manejáveis que garantem a normalidade assintótica de uma estatística-L foram dadas por Moore[14] (1968) e são essencialmente as seguintes:

1. $E[|X|] = \int_0^1 |F^{\leftarrow}(u)|du < \infty$;
2. $J(\cdot)$ contínua em [0,1], excepto em pontos de salto $\alpha_1, \ldots, \alpha_m$;
3. $J'(\cdot)$ contínua e de variação limitada em $[0,1] - \{\alpha_1, \ldots, \alpha_m\}$.

[13] Stiegler, S.M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2**, 676–693.

[14] Moore, D.S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.* **39**, 263–265.

Capítulo 7

Estatística de Extremos: Abordagens Paramétricas

7.1 Parâmetros de acontecimentos extremos

Um dos parâmetros cruciais em *Estatística de Extremos* é o EVI, γ, definido na Secção 6.4.2. Com elevado interesse em situações de amostras dependentes, temos o chamado índice extremal θ, definido na Secção 6.6. Para além destes parâmetros fundamentais, convém começar por relembrar as definições de *quantil extremal* e de *período de retorno* de um nível t, elevado.

Definição 7.1.1 (função quantil de cauda, quantil extremal e período de retorno)**.** *Seja F uma f.d. contínua com inversa generalizada*

$$F^{\leftarrow}(u) = \inf\{x : F(x) \geq u\},$$

já definida em (3.2)*. A correspondente função quantil de cauda, também já definida em* (6.19)*, é dada por*

$$U(t) = F^{\leftarrow}(1 - 1/t), \qquad t \in [1, \infty].$$

Um quantil extremal é, para p pequeno (usualmente inferior a $1/n$, com n a dimensão da amostra), o valor χ_p tal que

$$\chi_p := F^{\leftarrow}(1 - p) = U(1/p).$$

O valor $u := U(t)$ *é usualmente designado por nível de retorno do valor* t*, e dado* u*, o valor* $T(u)$ *tal que* $u = U(T(u))$ *é o chamado período de retorno do nível* u*. Tem-se pois*

$$T(u) = \frac{1}{1 - F(u)},$$

o númerio médio de excedências do nível u *em esquema i.i.d.*

Observação 7.1.1. *Suponhamos que dispomos de observações independentes diárias. Ter-se* $F(U(t)) = 1 - 1/t$ *significa que o nível* $U(t)$*, o chamado* nível de retorno*, é excedido em média uma vez em cada* t *dias* (período de retorno)*. A função* $U(t)$ *é monótona e não-decrescente. Para além disto,* $U(1) = \inf\{x : F(x) \geq 0\} = x_{_F}$ *limite inferior do suporte de* F*, e* $U(\infty) = \inf\{x : F(x) \geq 1\} = \sup\{x : F(x) < 1\} = x^F$*, limite superior do suporte de* F*.*

Na área financeira, uma quantidade de grande interesse é o chamdo VaR, do inglês '*Value-at-Risk*'.

Definição 7.1.2 (*Value-at-Risk*, VaR)**.** *Seja* X *a v.a. associada aos Ganhos e Perdas* (P&L*, do inglês, 'profits and losses'*) *ou aos retornos de um produto financeiro, num certo horizonte temporal. O* VaR *não é mais do que um percentil extremal da f.d.* F *associada a* X*,*

$$\mathrm{VaR}_p := F^{\leftarrow}(1 - p),$$

ou ainda o valor que é ultrapassado com probabilidade p *muito pequena,*

$$\mathrm{VaR}_p : \mathbb{P}[X > \mathrm{VaR}_p] = p.$$

Observação 7.1.2. *Os valores de* p *mais usuais na prática são do tipo* $p = 0.01,\ 0.001,\ 0.0001$*.*

Definição 7.1.3 (Excedência, Probabilidade de Cauda)**.** *Designamos por* excedência (*do nível* c) *o acontecimento em que uma observação de um população* X *excede um dado nível* c*.* Probabilidade de cauda *é a probabilidade de um nível elevado* c *ser excedido* (*ou de um nível muito baixo* c *não ser atingido*)*,*

$$p^c := \mathbb{P}[X > c] \quad ou \quad p_c := \mathbb{P}[X < c],$$

Observação 7.1.3. *Observamos que, na área de Seguros, esta última probabilidade está relacionada com a* probabilidade de ruína.

Definição 7.1.4 (Valor Esperado de Cauda Condicional, CTE)**.** *O* CTE*, do inglês 'condicional tail expectation' é o valor médio condicional a* $100p\%$ *do topo da população. Denotando por* χ_p *o quantil de excedência* p*, i.e., o valor* χ_p *tal que* $\mathbb{P}[X > \chi_p] = p$*, tem-se*

$$\text{CTE}_p \equiv \mu_p := \mathbb{E}[X|X > \chi_p] = \frac{1}{p}\int_{\chi_p}^{\infty} x\, f(x)dx\,.$$

Definição 7.1.5 (Função de Excesso Médio, MEF)**.** *A* MEF*, do inglês 'mean excess function', é o valor médio condicional dos excessos acima de um nível* c*.*

$$\begin{aligned} m(c) &:= \mathbb{E}[X-c|X>c] = \frac{1}{1-F(c)}\int_c^{\infty}(x-c)f(x)dx \\ &= \frac{1}{1-F(c)}\int_c^{\infty}[1-F(x)]dx\,. \end{aligned}$$

Observação 7.1.4. *Os valores do* CTE*,* μ_p*, e da* MEF*,* $m(c)$*, estão relacionados, uma vez que*

$$\mathbb{E}[X|X > c] = c + m(c).$$

7.2 Abordagem clássica de Gumbel à Estatística de Extremos

Uma dificuldade implícita a qualquer análise de valores extremos é a limitada quantidade de dados disponíveis em grande parte das situações, com consequências na precisão da estimação associada. Realmente estes acontecimentos são raros, uma vez que os extremos também o são. No caso de estudo relacionado com os dados `maasmax.txt`, relativo às descargas anuais máximas do Rio Meuse (1911-1995), estão disponíveis só os máximos anuais ($m = 85$ anos), que já começámos a estudar na Secção 4.5. Mesmo mais geralmente, admitamos que temos, por exemplo, acesso aos níveis ou descargas

médias diárias de um rio em determinado local, $(x_1, x_2, \ldots, x_n)$, e estamos interessados em cheias (expressas em termos de $\max_{1\leq i\leq n} x_i$) ou em secas (expressas em termos de $\min_{1\leq i\leq n} x_i$) desse rio nesse mesmo local. É então sensato trabalhar com esses máximos (ou mínimos) de um número relativamente elevado de observações. Como o comportamento distribucional exacto dos maiores (ou dos menores) valores numa amostra é usualmente difícil, e como podemos trabalhar com o máximo (ou mínimo) de um grande número de observações recorremos, em *Estatística de Extremos*, a resultados assintóticos para esses valores de cauda, cuja distribuição é conhecida a menos de parâmetros desconhecidos.

Na realidade, no Capítulo 6 referimos o resultado limite fundamental em EVT que apresenta as distribuições *max-estáveis* como sendo as únicas formas limite para o máximo de uma amostra aleatória, devidamente normalizado, unificadas na forma da distribuição GEV,

$$G_\gamma(x;\lambda,\delta) = \exp\left\{-\left[1+\gamma\left(\frac{x-\lambda}{\delta}\right)\right]_+^{-1/\gamma}\right\}, \tag{7.1}$$

com $(x)_+ := \max(0, x)$. Realçámos ainda que o *índice de valores extremos* ou EVI, γ, assume aqui papel fundamental, pois dá a forma diferenciada do tipo de distribuição limite:

- Cauda Pesada (Domínio Fréchet).
- Cauda Exponencial (Domínio Gumbel).
- Cauda Curta (Domínio Max-Weibull).

O modelo GEV engloba realmente os três tipos clássicos: **Tipo I**, Gumbel:

$$\Lambda(x;\lambda,\delta) = \exp(-\exp(-(x-\lambda)/\delta)), \quad x\in\mathbb{R} \quad (\gamma = 0). \tag{7.2}$$

Tipo II, Fréchet:

$$\Phi_\alpha(x;\lambda,\delta) = \exp(-((x-\lambda)/\delta)^{-\alpha}),\ x>\lambda \quad (\alpha>0) \quad (\gamma = 1/\alpha > 0). \tag{7.3}$$

Tipo III, Max-Weibull:

$$\Psi_\alpha(x;\lambda,\delta) = \exp(-(-(x-\lambda)/\delta)^{\alpha}),\ x<\lambda \quad (\alpha>0)\ (\gamma = -1/\alpha < 0). \tag{7.4}$$

Este resultado limite levou Gumbel a sugerir o primeiro modelo em *Estatística de Extremos*, frequentemente designado por *modelo dos máximos anuais* (MMA), ou *modelo* GEV *univariado* ou ainda *modelo de* Gumbel. De acordo com este modelo, dividimos os N dados, $(X_1, \ldots, X_N)$, em m sub-amostras (usualmente correspondentes a m anos) de dimensão n ($N = nm$) e ajustamos um dos modelos extremais (Tipo I, II ou III) ou o modelo GEV à amostra formada pelos m máximos de cada sub-amostra, associados pois à v.a. $Y = \max(X_1, \ldots, X_n)$.

Associados ao ajustamento GEV para Y estão alguns parâmetros de acontecimentos raros, como por exemplo:

- *Probabilidade de excedência* de níveis elevados u (para Y),

$$1 - G_\gamma(u; \lambda, \delta) = \begin{cases} 1 - \exp\left\{-\left[1 + \gamma\left(\frac{u-\lambda}{\delta}\right)\right]^{-1/\gamma}\right\}, & \text{se} \quad \gamma \neq 0, \\ 1 - \exp\left\{-\exp\left[-\frac{u-\lambda}{\delta}\right]\right\}, & \text{se} \quad \gamma = 0, \end{cases}$$

  para u elevado.

- *Período de retorno* do nível u (para Y), $T_u = \dfrac{1}{1 - G_\gamma(u; \lambda, \delta)}$.

- *Nível de retorno* a T-anos (para Y),

$$\begin{aligned} U(T) &= G_\gamma^{\leftarrow}\left(1 - \tfrac{1}{T}; \lambda, \delta\right) \\ &= q_{_{Y,p}} = \begin{cases} \lambda + \frac{\delta}{\gamma}\left[(-\log(1-p))^{-\gamma} - 1\right], & \text{se} \quad \gamma \neq 0, \\ \lambda - \delta \log(-\log(1-p)), & \text{se} \quad \gamma = 0, \end{cases} \end{aligned}$$

  para $p = 1/T$.

- E, no caso de se ter um parâmetro de forma negativo ($\gamma < 0$), o *limite superior do suporte* (para Y),

$$x^F = q_{_{Y,0}} = \lambda - \delta/\gamma\,.$$

Evidentemente, caso se pretenda inferir sobre parâmetros análogos, mas para a população $X \frown F$, há que recorrer à propriedade de max-estabilidade, ou mais simplesmente, ao facto de

$$Y := \max_{1 \leq i \leq n} X_i = X_{n:n}$$

possuir distribuição

$$F_{_Y} \equiv F_{_{X_{n:n}}} = F^n \approx G_\gamma$$

pelo que os parâmetros para F se deduzem por esta relação.

Por exemplo, para blocos de tamanho n, os quantis-$(1-p)$ ajustados para X, $q_{_{X,p}}$, tais que $F(q_{_{X,p}}) = 1-p$ são tais que

$$F^n(q_{_{X,p}}) = (1-p)^n \approx G_\gamma\left(q_{_{X,p}}; \lambda, \delta\right),$$

ou seja,

$$\begin{aligned} q_{_{X,p}} &= G_\gamma^{\leftarrow}\left((1-p)^n; \lambda, \delta\right) \\ &= \begin{cases} \lambda + \delta\left[(-n\log(1-p))^{-\gamma} - 1\right]/\gamma, & \text{se} \quad \gamma \neq 0, \\ \lambda - \delta\log(-n\log(1-p)), & \text{se} \quad \gamma = 0. \end{cases} \end{aligned} \tag{7.5}$$

Toda a inferência estatística se resume pois à inferência associada aos modelos em questão, sendo fundamental a estimação de λ, δ, e γ (ou α), a partir da qual se estimarão todos os outros parâmetros relevantes. A estimação do EVI, γ, é feita conjuntamente com a estimação de (λ, δ). Como a utilidade prática de uma distribuição depende parcialmente da existência de bons métodos para estimação dos seus parâmetros, e tal estimação não é sempre fácil para a f.d. GEV, é usual, quando estamos face a uma amostra de máximos, tentar o ajustamento de uma das três distribuições, Gumbel (a mais simples), Fréchet ou Max-Weibull, fazendo primeiro um teste de escolha estatística de um dos três modelos, que pode ser tão simples como um gráfico em papel de probabilidade ou um QQ-plot, como especificado no Capítulo 3. No Capítulo 4, Secção 4.5, conduzimos uma abordagem preliminar, escolhendo o valor de γ que maximiza a correlação no QQ-Plot, e em seguida os parâmetros de localização/escala, (λ, δ), foram obtidos pelo *método dos mínimos quadrados* (relembre-se o caso de estudo, `maasmax.txt`, relacionado com descargas máximas anuais do rio Meuse (1911-1995), num total d $m = 85$ máximos). Neste Capítulo focamos outros métodos de estimação. Mais especificamente, abordaremos essencialmente as metodologias:

1. *Máxima-verosimilhança* (ML, do inglês '*maximum likelihood*').
2. *Momentos ponderados de probabilidade* (PWM, do inglês '*probability weighted moments*').

7.2.1 Modelos Gumbel, Fréchet, Max-Weibull e GEV: principais características

Para a f.d. Gumbel, em (7.2), cálculos simples permitem-nos concluir que as principais características populacionais deste modelo são:

$$\begin{aligned} \mathbb{E}(X) &= \lambda + \epsilon\,\delta, \\ \mathbb{V}ar(X) &= (\pi\delta)^2/6, \\ \mathbb{M}oda &= \lambda, \\ \mathbb{M}ediana &= \epsilon - \delta \log\log 2 \end{aligned}$$

onde $\epsilon = -\Gamma'(1) = 0.57721\ 56649\ldots$ é a chamada *constante de Euler*. A f.d. Gumbel ($\gamma = 0$ em modelo GEV) é sem dúvida a distribuição mais frequentemente ajustada a dados que sejam valores máximos de outras grandezas aleatórias. A principal razão de tal escolha é devida ao facto de a inferência para tais distribuições ser muito mais simples do que para distribuições $\Phi_\alpha(\cdot)$ ou $\Psi_\alpha(\cdot)$, como veremos na secção 7.2.2. A f.d. Gumbel goza, como f.d. limite de máximos, convenientemente normalizados, do mesmo papel que a f.d. Normal goza, como f.d. limite de somas convenientemente normalizadas, e a Log-Normal goza, como f.d. limite de produtos convenientemente normalizados.

A f.d. Fréchet, em (7.3), tem como principais características,

$$\begin{aligned} \mathbb{E}(X) &= \lambda + \delta\Gamma(1-1/\alpha) \text{ (só existe se } \alpha > 1), \\ \mathbb{V}ar(X) &= \delta^2 \left\{\Gamma(1-2/\alpha) - \Gamma^2(1-1/\alpha)\right\} \text{ (só existe se } \alpha > 2), \\ \mathbb{M}oda &= \lambda + \delta(1+1/\alpha)^{-1/\alpha}, \\ \mathbb{M}ediana &= \lambda + \delta(\log 2)^{-1/\alpha}. \end{aligned}$$

Para o modelo Max-Weibull, com f.d. dada em (7.4), tem-se

$$\begin{aligned} \mathbb{E}(X) &= \lambda + \delta\Gamma(1+1/\alpha), \\ \mathbb{V}ar(X) &= \delta^2 \left\{\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)\right\}, \\ \mathbb{M}oda &= \lambda + \delta(1-1/\alpha)^{1/\alpha}, \\ \mathbb{M}ediana &= \lambda + \delta(\log 2)^{1/\alpha}. \end{aligned}$$

Finalmente, para o modelo GEV$(\gamma;\lambda,\delta)$, em (7.1), tem-se

$$\begin{aligned}
\mathbb{E}(X) &= \begin{cases} \lambda+\frac{\delta}{\gamma}[\Gamma(1-\gamma)-1], & \text{se} \quad \gamma<1,\ \gamma\neq 0, \\ \lambda+\delta\epsilon, & \text{se} \quad \gamma=0, \end{cases} \\
\mathbb{V}ar(X) &= \begin{cases} (\delta/\gamma)^2[\Gamma(1-2\gamma)-\Gamma^2(1-\gamma)], & \text{se} \quad \gamma<1/2,\ \gamma\neq 0, \\ \delta^2\frac{\pi^2}{6}, & \text{se} \quad \gamma=0, \end{cases} \\
\beta(X) &= \begin{cases} -\text{sign}(\gamma)\frac{\Gamma(1-3\gamma)-3\Gamma(1-2\gamma)\Gamma(1-\gamma)+2\Gamma^3(1-\gamma)}{[\Gamma(1-2\gamma)-\Gamma^2(1-\gamma)]^{3/2}}, & \\ & \text{se} \quad \gamma<1/3,\ \gamma\neq 0, \\ \frac{12\sqrt{6}}{\pi^3}\zeta(3), & \text{se} \quad \gamma=0, \end{cases}
\end{aligned}$$

com $\epsilon = 0.57721$ a *constante de Euler* e $\zeta(\cdot)$ a denotar a *função zeta de Riemann*. O modelo GEV tem assimetria positiva para $\gamma > -0.28$.

Se a amostra for de máximos, e caso nos decidamos a tal, como escolher entre os modelos Gumbel, Fréchet e Max-Weibull? Podemos por exemplo utilizar a técnica do papel de probabilidade ou QQ-plot. Como já vimos anteriormente, também no Capítulo 4, se o modelo subjacente aos dados for Gumbel, deverá existir uma relação linear entre as observações ordenadas ascendentemente, $x_{i:n}$, e $-\log(-\log(p_i))$, com $p_i = i/(n+1)$. Essa relação linear dá-nos uma validação informal do modelo Gumbel. Mais do que isso, curvas $(x_{i:n}, -\log(-\log(p_i)))$, $1 \leq i \leq n$, com a concavidade voltada para baixo fornecem-nos uma validação informal de um modelo Fréchet, enquanto curvas $(x_{i:n}, -\log(-\log(p_i)))$, $1 \leq i \leq n$ com a concavidade voltada para cima nos fornecem uma validação informal de um modelo Max-Weibull.

Podemos ainda fazer um teste de escolha estatística de modelos extremais. Devido à facilidade relativa da inferência estatística associada a populações Gumbel, o primeiro passo a dar será por exemplo o de testar a hipótese H_0 : $\gamma = 0$ versus $H_1 : \gamma \neq 0$ (ou $\gamma > 0$) (ou $\gamma < 0$) no modelo GEV(γ). Se tal hipótese H_0 não for rejeitada, ao nível de significância que nos parecer conveniente, devemos enveredar pelo ajustamento de um modelo Gumbel. Caso contrário, o valor observado da estatística de teste irá indicar a possível escolha, Fréchet ou Max-Weibull. Dada a amostra $(X_1, \ldots, X_n)$, uma das estatística de teste frequentemente usada, devido à sua simplicidade, é:

$$W_n = \frac{X_{n:n} - X_{\lfloor n/2 \rfloor + 1:n}}{X_{\lfloor n/2 \rfloor + 1:n} - X_{1:n}}.$$

A distribuição assintótica de W_n, sob a validade de H_0, e após normalizção conveniente, é uma v.a. Gumbel, i.e., quando $n \to \infty$,

$$W_n^* = \log\log n \left\{ W_n - \frac{\log n + \log\log 2}{\log\log n - \log\log 2} \right\} \xrightarrow{d} Z \frown \Lambda,$$

com f.d. $\Lambda(\cdot)$ (veja-se Tiago de Oliveira & Gomes[1], 1984). Para valores pequenos de n, temos os seguintes pontos críticos para W_n^*.

Tabela 7.1: Pontos críticos de W_n^*

α	0.025	0.05	0.10	0.90	0.95	0.975
n						
10	-1.06	-0.98	-0.87	1.42	2.12	2.83
50	-1.30	-1.15	-0.94	1.80	2.47	3.11
100	-1.35	-1.16	-0.95	1.77	2.34	3.04
∞	-1.31	-1.10	-0.83	2.25	2.97	3.68

Para uma resenha acerca de testes sobre Modelos Extremais veja-se Neves & Fraga Alves[2] (2008) e Hüsler & Peng[3] (2008).

7.2.2 Estimação dos parâmetros em modelos extremais clássicos

Admitamos ter acesso à amostra $(y_1, y_2, \cdots, y_m)$, proveniente de um modelo extremal.

Gumbel(λ, δ)**.** As estimativas de *máxima verosimilhança*, ou ML, dos parâmetros desconhecidos (λ, δ) obtém-se resolvendo primeiro numericamente a

[1] Tiago de Oliveira, J. & Gomes, M. I. (1984). Two test statistics for choice of univariate extreme models. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*. D. Reidel, Dordrecht, Holland, 651– 668.

[2] Neves, C. & Fraga Alves, M.I. (2008). Testing extreme value conditions — an overview and recent approaches. *Revstat* **6**:1, 83–100. Special issue on "*Statistics of Extremes and Related Fields*" edited by J. Beirlant, I. Fraga Alves & R. Leadbetter.

[3] Hüsler, J. & Peng, L. (2008). Review of testing issues in extremes: in honor of Professor Laurens de Haan. *Extremes*, **11**:1, 99– 111.

equação

$$\hat{\delta} = \sum_{i=1}^{m} y_i/m - \frac{\sum_{i=1}^{m} y_i \exp\big(-y_i/\hat{\delta}\big)}{\sum_{i=1}^{m} \exp\big(-y_i/\hat{\delta}\big)},$$

através por exemplo do método do ponto fixo, e obtendo em seguida

$$\hat{\lambda} = -\hat{\delta}\ \log\Big\{\sum_{i=1}^{m} \exp\big(-y_i/\hat{\delta}\big)\Big\}/m.$$

Fréchet$(\lambda, \delta, \alpha)$. Em modelo Fréchet, o estimador ML de δ obtém-se explicitamente à custa dos estimadores ML de λ e α, através da equação,

$$\hat{\delta} = \Big\{m/\sum_{i=1}^{m}(y_i - \hat{\lambda})^{\hat{\alpha}}\Big\}^{1/\hat{\alpha}}.$$

Os estimadores ML de λ e α podem obter-se por resolução numérica do sistema de duas equações a duas incógnitas:

$$\begin{cases} \frac{\hat{\alpha}+1}{m}\sum_{i=1}^{m}(y_i-\hat{\lambda})^{-1} - \frac{\hat{\alpha}\sum_{i=1}^{m}(y_i-\hat{\lambda})^{-\hat{\alpha}-1}}{\sum_{i=1}^{m}(y_i-\hat{\lambda})^{-\alpha}} = 0 \\ \\ \frac{1}{\hat{\alpha}} + \frac{\sum_{i=1}^{m}(y_i-\hat{\lambda})^{-\alpha}\log(y_i-\lambda)}{\sum_{i=1}^{m}(y_i-\hat{\lambda})^{-\alpha}} - \sum_{i=1}^{m}\log(y_i-\hat{\lambda})/m = 0. \end{cases}$$

Para obtenção das estimativas ML de λ e α já é então necessário utilizar, por exemplo, o algoritmo de Newton-Raphson.

Max-Weibul$(\lambda, \delta, \alpha)$. A situação não é muito diferente da anterior, com ligeiras alterações nas equações de máxima verosimilhança. Em modelo Weibull, o estimador ML de δ também se obtem explicitamente à custa dos estimadores ML de λ e α, através da equação,

$$\hat{\delta} = \Big\{m/\sum_{i=1}^{m}(\hat{\lambda} - y_i)^{\hat{\alpha}}\Big\}^{-1/\hat{\alpha}}.$$

Os estimadores ML de λ e α podem obter-se por resolução numérica do sistema de duas equações a duas incógnitas:

$$\begin{cases} \frac{\hat{\alpha}-1}{m}\sum_{i=1}^{m}(y_i-\hat{\lambda})^{-1} - \frac{\hat{\alpha}\sum_{i=1}^{m}(\hat{\lambda}-y_i)^{\hat{\alpha}-1}}{\sum_{i=1}^{m}(\hat{\lambda}-y_i)^{\alpha}} = 0 \\ \\ -\frac{1}{\hat{\alpha}} + \frac{\sum_{i=1}^{m}(\hat{\lambda}-y_i)^{\alpha}\log(\hat{\lambda}-y_i)}{\sum_{i=1}^{m}(\hat{\lambda}-y_i)^{\alpha}} - \sum_{i=1}^{m}\log(\hat{\lambda}-y_i)/m = 0. \end{cases}$$

Aqui, para resolução deste sistema, torna-se por vezes necessário utilizar algoritmos mais sofisticados que o de Newton-Raphson (veja-se Wingo[4],1972).

7.2.3 Modelo GEV: Método ML

O método ML, de máxima verosimilhança, é um dos métodos possíveis, que exige cálculo computacional pesado, mas existem hoje em dia formas fáceis de obter essas estimativas em `R`, como veremos adiante no Capítulo 9. A log-verosimilhança para a amostra observada, $y_1, \ldots, y_m$, i.i.d. e GEV, é para $\gamma \neq 0$ e $1 + \gamma(y_i - \lambda)/\delta > 0$,

$$\begin{aligned}&\log L(\gamma, \lambda, \delta)\\ &= -m \log \delta - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{m} \log\left(1 + \gamma \frac{y_i - \lambda}{\delta}\right) - \sum_{i=1}^{m} \left(1 + \gamma \frac{y_i - \lambda}{\delta}\right)^{-\frac{1}{\gamma}}.\end{aligned}$$

Se $\gamma = 0$, temos

$$\log L(0, \lambda, \delta) = -m \log \delta - \sum_{i=1}^{m} \exp\left(-\frac{y_i - \lambda}{\delta}\right) - \sum_{i=1}^{m} \frac{y_i - \lambda}{\delta}.$$

Os estimadores ML serão os valores $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$ e $(\hat{\lambda}, \hat{\delta})$, respectivamente, que maximizam a log-verosimilhança. Os estimadores ML, para $\gamma > -0.5$, quando $m \to \infty$, são consistentes, assintoticamente eficientes e

$$\sqrt{m}\big((\hat{\gamma}, \hat{\lambda}, \hat{\delta}) - (\gamma, \lambda, \delta)\big)$$

é assintoticamente Normal de vector de valores médios nulo e matriz de covariâncias dada pela inversa da Matriz de Informação de Fisher (vejam-se detalhes em Beirlant *et al.*, 2004).

Por vezes, podem surgir computacionalmente casos de falta de convergência dos métodos numéricos necessários para a sua obtenção. No Capítulo 9 iremos fazer uso de alguns packages de `R` para ajustamento das max-estáveis a algumas amostras de dados reais.

[4]Wingo, D.R. (1972). Maximum likelihood estimation of the parameters of the Weibull distribution by modified quasilinearization, *IEEE Transactions on Reliability* **21**, 89–93.

7.2.4 Modelo GEV: Método PWM

Os momentos ponderados de probabilidade para uma v.a. Y, com f.d. F foram apresentados por Greenwood *et al.*[5] (1979), e são os momentos seguintes:

$$M_{p,r,s} = E\left\{Y^p[F(Y)]^r[1-F(Y)]^s\right\}, \quad p,r,s \in \mathbb{R}\ .$$

Um método alternativo, muito utilizado em aplicações, é o método dos *momentos ponderados de probabilidade* ou método PWM, introduzido em Hosking[6] (1985), e aprofundado em Hosking *et al.*[7] (1985). No caso de $\gamma = 0$, i.e. a distribuição Gumbel, tem-se para $p = 1, r = 0, 1, 2, \cdots$ e $s = 0$,

$$M_{1,r,0} = E\left\{Y[F(Y)]^r\right\} = \frac{1}{r+1}[\lambda + \delta\{\epsilon + \log(1+r)\}],$$

com $\epsilon = 0.57721$ a *constante de Euler*. Verifica-se facilmente que

$$\delta = \frac{2M_{1,1,0} - M_{1,0,0}}{\log 2} \qquad \text{e} \qquad \lambda = M_{1,0,0} - \delta\epsilon.$$

Suponhamos que temos disponível a a.a. $Y_1, \ldots, Y_m$ proveniente de uma população GEV. Então um estimador centrado para $M_{1,r,0}$ é

$$\widehat{M}_{1,r,0} = \frac{1}{m}\sum_{i=1}^{m}\left(\prod_{k=1}^{r}\frac{(i-k)}{(m-k)}\right)Y_{i:m}\ .$$

Substituindo $M_{1,0,0}$ e $M_{1,1,0}$ pelos estimadores

$$\widehat{M}_{1,0,0} = \frac{1}{m}\sum_{i=1}^{m}Y_{i:m} \qquad \text{e} \qquad \widehat{M}_{1,1,0} = \frac{1}{m}\sum_{i=1}^{m}\frac{i-1}{m-1}Y_{i:m},$$

respectivamente, temos os estimadores para a localização e escala

$$\hat{\delta} = \frac{2\widehat{M}_{1,1,0} - \widehat{M}_{1,0,0}}{\log 2} \qquad \text{e} \qquad \hat{\lambda} = \widehat{M}_{1,0,0} - \epsilon\,\hat{\delta}$$

[5] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research* **15**:5, 1049–1054.

[6] Hosking, J.R.M. (1985). Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Appl. Statist.* **34**, 301–310.

[7] Hosking, J.R.M., Wallis, J.R. & Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261.

No caso de $\gamma \neq 0$, i.e., no caso da distribuição GEV, tem-se para $p = 1, r = 0, 1, 2, \cdots$ e $s = 0$, para $\gamma < 1$,

$$M_{1,r,0} = E\left\{Y[F(Y)]^r\right\} = \frac{1}{r+1}\left\{\lambda - \frac{\delta}{\gamma}[1-(r+1)^{\gamma}\Gamma(1-\gamma)]\right\},$$

verificando-se

$$\frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} = \frac{3^{\gamma}-1}{2^{\gamma}-1}, \quad \delta = \frac{\gamma(2M_{1,1,0} - M_{1,0,0})}{\Gamma(1-\gamma)(2^{\gamma}-1)},$$

$$\lambda = M_{1,0,0} + \frac{\delta}{\gamma}\left(1 - \Gamma(1-\gamma)\right).$$

Substituindo $M_{1,0,0}$ e $M_{1,1,0}$ pelos estimadores $\widehat{M}_{1,0,0}$ e $\widehat{M}_{1,1,0}$ e estimando $M_{1,2,0}$ através de

$$\widehat{M}_{1,2,0} = \frac{1}{m}\sum_{i=1}^{m} \frac{(i-1)(i-2)}{(m-1)(m-2)} Y_{i:m}$$

os estimadores para a forma, escala e localização são obtidos a partir de

$$\frac{3\widehat{M}_{1,2,0} - \widehat{M}_{1,0,0}}{2\widehat{M}_{1,1,0} - \widehat{M}_{1,0,0}} = \frac{3^{\hat{\gamma}}-1}{2^{\hat{\gamma}}-1},$$

$$\hat{\delta} = \frac{\hat{\gamma}(2\widehat{M}_{1,1,0} - \widehat{M}_{1,0,0})}{\Gamma(1-\hat{\gamma})(2^{\hat{\gamma}}-1)}, \qquad \hat{\lambda} = \widehat{M}_{1,0,0} + \frac{\hat{\delta}}{\hat{\gamma}}\left(1 - \Gamma(1-\hat{\gamma})\right).$$

Para os estimadores baseados nos método PWM, e para $\gamma < 1$, quando $m \to \infty$, verifica-se que

$$\sqrt{m}\big((\hat{\gamma}, \hat{\lambda}, \hat{\delta}) - (\gamma, \lambda, \delta)\big)$$

é assintoticamente Normal de vector de valores médios nulo (mais uma vez, vejam-se detalhes em Beirlant *et al.*, 2004)

7.2.5 Intervalos de confiança para os parâmetros da GEV

Os IC's para os parâmetros da GEV$(\gamma, \lambda, \delta)$, em (7.1), decorrem da aproximação à Normal dos estimadores ML e PWM. Em Beirlant *et al.* (2004)

podemos consultar a construção de IC's quer para o parâmetro de forma γ, quer para quantis $q_{Y,p}$, baseados na normalidade assintótica dos respectivos estimadores ML ou PWM. É claro que se tratam de IC's centrados nas respectivas estimativas pontuais, uma vez que a Normal é simétrica. Em geral, podem ser obtidas estimativas intervalares de melhor qualidade, que não são obrigatoriamente centradas na estimativa ML, usando a função de *profile log-likelihood*.

Definição 7.2.1 (Profile log-Likelihood para γ)**.** *Para cada valor de γ, o profile log-likelihood é a log-verosimilhança maximizada relativamente aos outros parâmetros, (λ, δ), i.e.*

$$\log L_p(\gamma) := \max_{(\lambda,\delta)|\gamma} \log L(\gamma, \lambda, \delta).$$

A estatística de *razão de verosimilhanças* (RV ou LR, do inglês, '*likelihood ratio*'),

$$\Lambda := \frac{L_p(\gamma_0)}{L_p(\hat{\gamma})}$$

do teste de RV para testar $H_0 : \gamma = \gamma_0$ *versus* $H_1 : \gamma \neq \gamma_0$ tem por distribuição assintótica χ^2_1, quando $m \to \infty$, após logaritmização e mudança de escala conveniente, i.e.,

$$-2 \log \Lambda \xrightarrow[n\to\infty]{d} \chi^2_1 .$$

O teste de nível assintótico α rejeita H_0 se $-2 \log \Lambda > \chi^2_1(1-\alpha)$. Consequentemente, o IC para γ baseado no profile-likelihood com grau de confiança $100(1-\alpha)\%$ é dado por

$$\mathrm{IC}_\gamma = \left\{ \gamma : -2 \log \Lambda \leq \chi^2_1(1-\alpha) \right\},$$

ou, equivalentemente,

$$\mathrm{IC}_\gamma = \left\{ \gamma : \log L_p(\gamma) \geq \log L_p(\hat{\gamma}) - \frac{\chi^2_1(1-\alpha)}{2} \right\}.$$

Estimação de quantis elevados e períodos de retorno em situações de dependência. Temos $N = n \times m$ observações originais. Consideramos m máximos de amostras de dimensão n. Na abordagem de Gumbel temos a aproximação,

$$F^n(x) \;\approx\; G_\gamma\Big(\frac{x-\lambda}{\delta}\Big),$$

que nos permite facilmente estimar quantis elevados

$$q_{X,p} : F(\chi_p) = 1 - p, \quad p \text{ pequeno.}$$

Tem-se pois (7.5), ou, considerando a aproximação $-\log(1-p) \sim p$, para $p \downarrow 0$,

$$q_{X,p} = \begin{cases} \lambda - \delta \ \left(1 - (np)^{-\gamma}\right)/\gamma, & \text{se} \quad \gamma \neq 0, \\ \lambda - \delta \ \log(np), & \text{se} \quad \gamma = 0, \end{cases}$$

que permite a estimação de χ_p a partir da estimação de λ, δ e γ, tendo por base a a.a. de dimensão m de $Y := X_{n:n}$.

Para estimação do período de retorno, utilizamos a relação:

$$T_X(u) = \frac{1}{1 - F(u)} \approx \begin{cases} n\,(1 + \gamma(u-\lambda)/\delta)^{1/\gamma}, & \text{se} \quad \gamma \neq 0, \\ n \ \exp\big((u-\lambda)/\delta\big), & \text{se} \quad \gamma = 0. \end{cases}$$

A existência de um índice extremal θ exigirá as alterações, uma vez que passamos a ter a relação

$$F^{n\theta}(x) \ \approx \ G_\gamma\left(\frac{x-\lambda}{\delta}\right),$$

donde seguem as relações:

$$q_{X,p} = \begin{cases} \lambda - \delta\big(1 - (np\theta)^{-\gamma}\big)/\gamma, & \text{se} \quad \gamma \neq 0, \\ \lambda - \delta \log(np\theta), & \text{se} \quad \gamma = 0, \end{cases}$$

e

$$\begin{aligned} T_X(u) \ &= \ \frac{1}{\theta} \ \frac{1}{1 - F(u)} - \frac{1}{\theta} + 1 \\ &\approx \ \begin{cases} n\,(1 + \gamma(u-\lambda)/\delta)^{1/\gamma} - \frac{1}{\theta} + 1, & \text{se} \quad \gamma \neq 0 \\ n \ e^{\frac{u-\lambda}{\delta}} - \frac{1}{\theta} + 1, & \text{se} \quad \gamma = 0. \end{cases} \end{aligned}$$

Observação 7.2.1. *No caso de índice extremal θ, o período de retorno*

$$T_X(u) := \textit{número médio de observações até ultrapassar o nível } u$$

tem de contemplar o número médio de de excedências por clusters, $1/\theta$,

quando $u_n \to x^F$, *com* $n \to \infty$. *Para isso considere-se que*

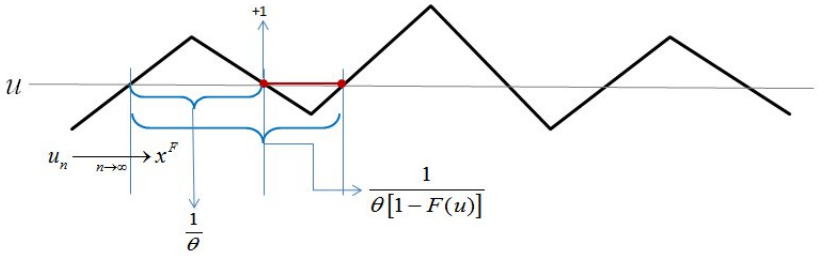
$$\begin{aligned}
\mathbb{P}[\textit{sucesso}] &= \mathbb{P}[\textit{cruzamento descendente}] = \mathbb{P}[X_i > u \ \wedge \ X_{i+1} < u] \\
&= \mathbb{P}[X_i > u] \cdot \mathbb{P}[X_{i+1} < u \,|\, X_i > u] \\
&= [1 - F(u)] \cdot \mathbb{P}[X_{i+1} < u \,|\, X_i > u] \\
&\longrightarrow [1 - F(u)]\theta, \ \textit{quando } u_n \to x^F, \ \textit{com } n \to \infty.
\end{aligned}$$

Denotando

$N := \#$ *médio de observações necessárias para um cruzamento descendente,*

então N *é uma Geométrica de parâmetro* $[1 - F(u)]\theta$, *pelo que*

$$\mathbb{E}[N] = \frac{1}{[1 - F(u)]\theta}.$$



Justifica-se assim a expressão

$$T_X(u) = \frac{1}{[1 - F(u)]\theta} - \frac{1}{\theta} + 1$$

que abrange igualmente o caso de independência, i.e., quando $\theta = 1$.

7.3 Abordagens não-clássicas: maiores observações e excessos de nível

No caso de estudo relativo aos dados `maasmax.txt`, já abordado na Secção 4.5, estão apenas disponíveis as *descargas anuais máximas* do Rio Meuse, no período $1911 - 1995$. Justifica-se então a utilização do *método de Gumbel*, tal

como faremos na Secção 9.1. Esse tipo de situação é ilustrada na Figura 7.1, em contrapartida com a Figura 7.2, um caso em que a natureza dos dados é mais rica, com informação sobre as k maiores observações em cada um dos m anos.

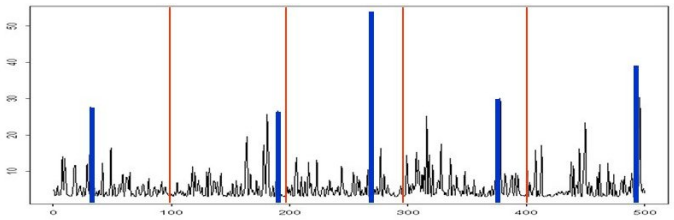


Figura 7.1: Máximos anuais – MMA

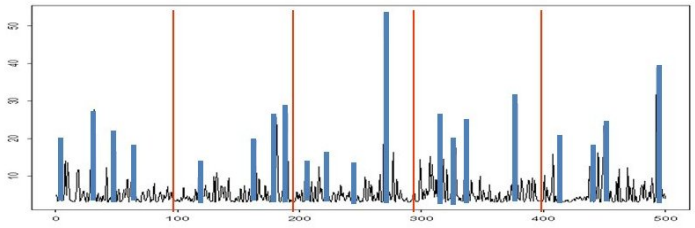


Figura 7.2: Maiores observações anuais – MMO

Mas em outras bases de dados, como por exemplo a base de dados `soa.txt` do *Group Medical Insurance*, relativa a Grandes Indemnizações, temos acesso às $k = 75789$ excedências acima dos 25 000 USD, no ano 1991, como ilustrado na Figura 7.3.

Podemos então trabalhar com essas k excedências, ou, como também é usual em *Estatística de Extremos*, trabalhar com os *excessos* acima de determinado nível extremo u, $X_i - u$, utilizando a chamada metodologia POT, do inglês '*peaks over threshold*'. Outra das metodologias relevantes nesta área é a que se foca nas k maiores observações da amostra, o que de certo modo corresponde a considerar um nível aleatório, $X_{n-k:n}$, e a trabalhar com as k observações que excedem esse nível, $X_{n-i+1:n}$, $1 \leq i \leq k$, ou com os excessos $X_i - X_{n-k:n} | X_i >$

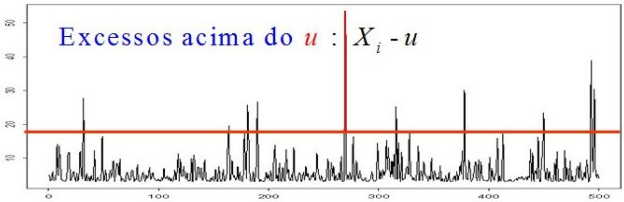


Figura 7.3: Excessos acima do nível u – POT

$X_{n-k+1:n}$. Esta última abordagem é a chamada metodologia PORT, do inglês *peaks over random threshold*', terminologia introduzida em Araújo Santos *et al.*[8] (2006).

7.3.1 Modelo GEV multivariado e multidimensional

Em várias áreas de aplicação, não existe uma sazonalidade natural nos dados, parecendo de certo modo artificial e subjectivo o método das sub-amostras. E se se considerasse um número reduzido de observações de topo da colecção de dados original? Na realidade, se temos dados diários, certos anos podem conter alguns de entre esses maiores valores (certamente relevantes para inferir sobre a cauda de F), enquanto que outros anos podem não conter nenhuns deles. Podemos de certo modo dizer que este novo tipo de abordagem repõe alguma informação acerca da amostra inicial, que o método tradicional parece desperdiçar. Esta abordagem, introduzida em Gomes[9] (1981) vai certamente depender do comportamento distribucional conjunto das maiores e.o.'s, que referimos em seguida, e que já foi abordado no Capítulo 6.

Em inferência estatística para acontecimentos raros, uma abordagem paramétrica possível é modelar as maiores observações disponíveis da amostra através do comportamento conjunto estabelecido no Teorema 6.4.23 com a expressão

[8] Araújo Santos, P., Fraga Alves, M.I. & Gomes, M.I. (2006). Peaks over random threshold methododlogy for tail index and high quantile estimation. *Revstat* **4**:3, 227–247.

[9] Gomes, M.I. (1981). An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, D. Reidel, 389–410.

(6.24). Quando $N \to \infty$, k fixo

$$\mathbb{P}[X_{N:N} \leq x_1, \ldots, X_{N-k+1:N} \leq x_k] \approx H_\gamma\Big(\frac{x_1-\lambda}{\delta}, \ldots, \frac{x_k-\lambda}{\delta}\Big) \quad (7.6)$$

onde, com G_γ a distribuição GEV univariada, e g_γ a f.d.p. associada, se tem

$$\begin{aligned} h_\gamma(x_1, \ldots, x_k) &= \frac{\partial^k}{\partial x_1 \ldots \partial x_k} H_\gamma(x_1, \ldots, x_k) \\ &= \frac{1}{\delta^k}\ g_\gamma\Big(\frac{x_k-\lambda}{\delta}\Big) \prod_{j=1}^{k-1} \frac{g_\gamma\big(\frac{x_j-\lambda}{\delta}\big)}{G_\gamma\big(\frac{x_j-\lambda}{\delta}\big)}, \quad (7.7) \end{aligned}$$

a chamada *distribuição* GEV *multivariada*. Assim, a inferência sobre o máximo e/ou sobre acontecimentos raros de uma forma geral, pode basear-se na informação das k maiores observações referentes aos m blocos, usualmente designados por *anos*. Surge assim a necessidade de estimação dos parâmetros de forma, localização e escala, para posterior inferência sobre parâmetros de acontecimentos raros de interesse.

Temos pois um segundo modelo em *Estatística de Extremos*, o designado *Modelo* GEV *multivariado* ou *modelo das maiores observações* (MMO). Neste modelo é mais fácil aumentar a dimensão da amostra, não sendo pois necessário um tão grande número de observações originais.

Note-se que para uma comparação fidedigna com o MMA, podemos considerar $k = m$, com m o número de anos, e trabalhar com as m maiores observações de entre as N observações originais. Note-se ainda que se pode facilmente combinar estas duas abordagens considerando que em cada uma das sub-amostras podemos recolher algumas estatísticas de topo, as quais são modeladas pelo modelo GEV Multivariado. Temos então o chamado *Modelo* GEV *Multidimensional*, em que temos acessibilidade a uma amostra multivariada

$$(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_r), \text{ onde } \underline{X}_j = (X_{1j}, \ldots, X_{i_j j}), 1 \leq j \leq r,$$

são vectores extremais multivariados.

Durante m anos recolhemos, por exemplo, as k maiores observações em cada ano. Temos consequentemente acesso a uma amostra

$$(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m), \text{ com } \underline{X}_j = (X_{1j} > X_{2j} > \cdots > X_{kj}),\ 1 \leq j \leq m,$$

proveniente, para $\gamma = 0$, de um modelo com f.d.p.,

$$f(x_1, \ldots, x_k) = \delta^{-k} \exp\Big[- \exp\big(- \frac{x_k - \lambda}{\delta}\big) - \sum_{j=1}^{k} \frac{x_j - \lambda}{\delta}\Big], \quad \text{se } x_1 \geq x_2 \geq \cdots \geq x_k, \quad (7.8)$$

ou, para $\gamma \neq 0$, de um modelo com f.d.p.,

$$\delta^{-k} \exp\Big(- \big[1 + \gamma\big(\frac{x_k - \lambda}{\delta}\big)\big]^{-1/\gamma} - (1/\gamma + 1) \sum_{j=1}^{k} \log[1 + \gamma\big(\frac{x_j - \lambda}{\delta}\big)]\Big)$$
$$\text{se } x_1 \geq x_2 \geq \cdots \geq x_k,\ 1 + \gamma(x_j - \lambda)/\delta > 0,\ 1 \leq j \leq k.$$

Sob a validade de um modelo Gumbel temos os seguintes estimadores de λ e δ (veja-se Gomes, 1981):

$$\hat{\lambda} = -\hat{\delta} \log\Big\{ \sum_{j=1}^{m} \exp(-X_{rj}/\hat{\delta}\Big\}/(km), \quad (7.9)$$

e

$$\hat{\delta} = \sum_{j=1}^{m} \sum_{i=1}^{k} \frac{X_{ij}}{km} - \frac{\sum_{j=1}^{m} X_{kj} \exp(-X_{kj}/\hat{\delta})}{\sum_{j=1}^{m} \exp(-X_{kj}/\hat{\delta})}. \quad (7.10)$$

Os estimadores complicam-se um pouco quando passamos para os modelos Fréchet, max-Weibull, ou, mais geralmente, para a GEV como distribuição do máximo, mas são generalizações relativamente simples das obtidas anteriormente para $k = 1$.

Se considerarmos $m = 1$ (i.e., se estivermos a trabalhar só com as k maiores observações, que serão obviamente não de um único ano, mas sim de um conjunto de anos, os estimadores são muito simples. Também para o modelo Gumbel obtém-se

$$\hat{\delta} = \frac{1}{r} \sum_{j=1}^{k} (X_{j1} - X_{k1}), \quad \hat{\lambda} = \hat{\delta} \log k + X_{k1}.$$

7.3.2 A metodologia POT e o modelo GP

Uma outra perspectiva paramétrica equivalente ao Modelo GEV Multivariado é aquela em que restringimos a nossa atenção às observações que excedem um

certo nível ou '*threshold*', ajustando modelos estatísticos apropriados quer às **excedências** quer aos **picos** (valor máximo de uma sequência de observações consecutivas acima do nível) acima desse nível. Qual o modelo adequado? Vimos no Teorema 6.4.25 que a função de distribuição condicional das excessos de um nível elevado u, $X-u|X>u$, (ou com $Y:=X-u$ acima do nível u, a distribuição de $Y|Y>0$) pode ser bem aproximada por uma generalizada de Pareto (GP),

$$F_u(y)\approx H_\gamma(y;\sigma_u=\sigma):=\begin{cases}1-\left(1+\frac{\gamma y}{\sigma}\right)^{-1/\gamma}, & \text{se} \quad y\in(0,\infty),\gamma>0,\\ 1-\exp\left(-\frac{y}{\sigma}\right), & \text{se} \quad y\in(0,\infty),\gamma=0,\\ 1-\left(1+\frac{\gamma y}{\sigma}\right)^{-1/\gamma}, & \text{se} \quad y\in(0,-\frac{\sigma}{\gamma}),\gamma<0.\end{cases}$$

Como se sabe, existe obviamente uma dualidade forte entre a distribição GP e a distribuição GEV. De modo específico, se o máximo converge para uma GEV(γ), então $(X-u)|X>u$ converge para uma GP(γ). A distribuição GP desempenha um papel muito importante na modelação de acontecimentos raros. Dada uma v.a. $X\frown F$, interessa caracterizar a distribuição dos excessos acima de um nível u, F_u, a chamada *distribuição dos excessos condicional*

$$F_u(y)=\mathbb{P}[X-u\le y|X>u],\qquad 0\le y\le x^F-u,$$

com

$$F_u(y)=\frac{F(u+y)-F(u)}{1-F(u)}.$$

Quando $u\to x^F$, limite superior do suporte de F, tem-se

$$\mathbb{P}\left[Y-u\le t|Y>u\right]\ \approx\ H_\gamma(t/\sigma),$$

onde

$$H_\gamma(t)\equiv H_\gamma(t;1)=1-(1+\gamma t)^{-1/\gamma},\quad \text{se}\quad 1+\gamma t>0,\ t\ge 0,$$

é a distribuição GP *univariada*.

Sendo x definido por $x=u+y$, a GP pode ser também expressa em termos de 3 parâmetros, um parâmetro de forma(γ), outro de localização(u) e um último de escala(σ), para valores $x>u$:

$$H_\gamma(x;u,\sigma):=\begin{cases}1-\left(1+\gamma\frac{x-u}{\sigma}\right)_+^{-1/\gamma}, & \gamma\neq 0\\ 1-\exp\left(-\frac{x-u}{\sigma}\right), & \gamma=0,\end{cases}$$

com $(z)_+ := \max(0, z)$.

É pois natural esta outra abordagem à *Estatística de Extremos*, em que se considera um nível elevado, e se trabalha com os excessos (diferença entre as observações que excedem o nível e o próprio nível), os quais são modelados por uma distribuição de Pareto generalizada. Este modelo é frequentemente designado por *Modelo Paretiano de Excessos* ou *Modelo* POT, do inglês, '*peaks over thresholds*').

Considere-se a amostra original, $X_1, \ldots, X_n$. Seja u um nível elevado. E denotemos N_u, o número de excedências (observações da amostra original que excedem u). Seja

$$Y_j := X_i - u | X_i > u, \ j = 1, 2, \ldots, N_u \quad \text{(excessos)}. \tag{7.11}$$

Somos agora confrontados com o problema de estimação de γ e σ, baseadas na amostra observada dos excessos, em (7.11). Iremos mais uma vez considerar as metodologias ML e PWM.

Observação 7.3.1. *O nível u é usualmente uma das observações, i.e., $u = X_{n-k:n}$. Neste caso, os excessos* (ordenados) *são*

$$Y_{j:k} := X_{n-k+j:n} - X_{n-k:n}, \qquad j = 1, \ldots, k. \tag{7.12}$$

Inferência em modelo Generalizado de Pareto

O modelo Generalizado de Pareto tem f.d.

$$H_\gamma(x) = 1 - \Big(1 + \gamma \ y/\sigma\Big)^{-1/\gamma}, \quad 1 + \gamma \ y/\sigma > 0, \quad y \geq 0.$$

A densidade de Pareto, fornece um modelo Exponencial para $\gamma = 0$ e um modelo Uniforme para $\gamma = -1$.

Têm-se as seguintes características populacionais,

$$\begin{aligned} \mathbb{E}(Y) &= \frac{\sigma}{1-\gamma}, \quad \text{se } \gamma < 1, \\ \mathbb{V}ar(Y) &= \frac{\beta^2}{(1-\sigma)^2(1-2\gamma)}, \quad \text{se } \gamma < 1/2. \\ \beta(Y) &= \frac{2(1+\gamma)\sqrt{1-2\gamma}}{1-3\gamma}, \quad \text{se } \gamma < 1/3. \end{aligned}$$

Tal como anteriormente podemos enveredar inicialmente por um teste de Exponencialidade ($\gamma = 0$) dos excessos acima de um nível elevado. Esse teste pode mais uma vez ser um teste gráfico em papel de probabilidade Exponencial (marcação de $y_{i:n}$ versus $-\log(i/(n+1), 1 \leq i \leq n)$.

Podemos ainda usar os estimadores de momentos, válidos para $\gamma > -1/2$, e dados por

$$\beta^* = \frac{1}{2}\,\overline{Y}\left(\overline{Y}^2/S^2 + 1\right), \quad \gamma^* = \frac{1}{2}\left(1 - \overline{Y}^2/S^2\right),$$

para $\gamma \neq 0$, e, para $\gamma = 0$, por $\beta^* = \overline{Y}$, que também é estimador de máxima verosimillança. Quando $\gamma \leq -1/4$ estes estimadores têm 'péssimo' comportamento, tornando-se necessário utilizar alternativas, de entre as quais destacamos os estimadores PWM, válidos para $\gamma > -1$, ou os estimadores ML. A obtenção de estimadores ML traz problemas análogos aos que surgem com a estimação em modelo GEV, mas pode ser feita numericamente, por qualquer processo de optimização da log-verosimilhança, processo esse hoje em dia já 'standard' no 'package' `R`.

Método ML. A log-verosimilhança para $Y_1, \ldots, Y_{N_u}$, i.i.d. a Y, e GP, com f.d. $H_\gamma(y;\sigma)$ é, para $\gamma \neq 0$, e para $1 + \gamma Y_i/\sigma > 0,\ i = 1, \ldots, N_u$,

$$\log L(\sigma, \gamma) = -N_u \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \log\left(1 + \frac{\gamma Y_i}{\sigma}\right).$$

Se $\gamma = 0$, temos

$$\log L(\sigma, 0) = -N_u \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{N_u} Y_i.$$

Observação 7.3.2. *Por vezes, em termos computacionais, é vantajoso proceder à reparametrização*

$$(\sigma, \gamma) \rightsquigarrow (\tau, \gamma) \quad \textit{com} \quad \tau := \gamma/\sigma$$

resultando em

$$\log L(\tau, \gamma) = -N_u \log \gamma + N_u \log \tau - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \log(1 + \tau Y_i).$$

Os estimadores ML *de* τ *e* γ *verificam*

$$\frac{1}{\hat{\tau}^{ML}} - \left(\frac{1}{\hat{\gamma}^{ML}} + 1\right) \frac{1}{N_u} \sum_{i=1}^{N_u} \frac{Y_i}{1 + \hat{\tau}^{ML} Y_i} = 0,$$

$$com \quad \hat{\gamma}^{ML} = \frac{1}{N_u} \sum_{i=1}^{N_u} \log(1 + \tau Y_i).$$

Método PWM. O método dos momentos ponderados de probabilidade para a distribuição GP foram trabalhados por Hosking & Wallis[10] (1987). Note-se que os momentos da GP só existem para ordens r tais que $\gamma < 1/r$. Para a GP é conveniente considerar

$$M_{p,r,s} = E\left\{Y^p [F(Y)]^r [1 - F(Y)]^s\right\}, \quad p, r, s \in \mathbb{R},$$

para valores de $p = 1$, $r = 0$, $s = 0, 1, 2, \ldots$, o que conduziu a

$$M_{1,0,s} = \frac{\sigma}{(s+1)(s+1-\gamma)}, \ \gamma < 1.$$

Suponhamos que temos disponível a a.a., $Y_1, Y_2, \ldots, Y_{N_u}$, proveniente de uma população GP. Então substituindo $M_{1,0,s}$ pela sua contrapartida empírica, foi obtido

$$\widehat{M}_{1,0,s} = \frac{1}{N_u} \sum_{i=1}^{N_u} \left(\prod_{k=1}^{s} \frac{(N_u - i - k + 1)}{(N_u - k)} \right) Y_{i:N_u}.$$

Os estimadores para a forma e escala foram então obtidos para $s = 0$ e $s = 1$,

$$\hat{\gamma} = 2 - \frac{\widehat{M}_{1,0,0}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}} \quad e \quad \hat{\sigma} = \frac{2\widehat{M}_{1,0,0}\widehat{M}_{1,0,1}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}}.$$

Observação 7.3.3. *A aplicação do método* PWM *não é isenta de problemas. No caso de* $\gamma \geq 1$ *os estimadores* PWM *não existem. Por outro lado, a sua implementação prática pode conduzir a valores inadmissíveis*; *por exemplo, quando* $\gamma < 0$ (*caso em que o limite superior do suporte,* x^F*, é finito*), *pode conduzir a inferências para além de* x^F.

[10] Hosking, J.R.M. & Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.

Escolha do nível u. A escolha do nível u é um problema controverso e em aberto. É o problema paralelo à escolha de k quando se considera a o nível aleatório $X_{n-k:n}$, e os excessos em (7.12), denotando então esta metodologia por metodologia PORT, do inglês '*peaks over random threshold*', terminologia introuzida em Araújo Santos *et al.* (2006), tal como referido no início da Secção 7.3. A sua escolha envolve um compromisso entre valores elevados do nível u, onde o viés dos estimadores é menor, e valores pequenos para u, onde a variância é menor.

Davison & Smith[11] (1990) propuseram para a escolha do valor conveniente de u, o estudo da *função de excesso médio* (*mean excess function*), que no caso GP assume a expressão

$$e(t) := \mathbb{E}\left[X - t | X > t\right] = \mathbb{E}\left[Y | Y > 0\right] = \frac{\sigma + \gamma t}{1 - \gamma}, \quad \text{se} \quad \gamma < 1.$$

Na prática, trata-se de verificar uma *linearidade* à direita de u para o plot de $\hat{e}_n$, com base na amostra de dados observados $y_1, \cdots, y_{N_t}$, $\hat{e}_n(t) := \sum_{j=1}^{N_t} y_j / N_t$, i.e.

$$\hat{e}_n(t) := \frac{\sum_{i=1}^{n} x_i I_{(t,+\infty)}(x_i)}{\sum_{i=1}^{n} I_{(t,+\infty)}(x_i)} - t$$

ou em alternativa

$$\hat{e}_n(x_{n-k:n}) = \frac{1}{k} \sum_{j=1}^{k} x_{n-j+1:n} - x_{n-k:n}.$$

Estimação de outros parâmetros de acontecimentos raros. Outro problema com interesse é o da estimação de *quantis extremais* e do *limite superior do suporte*. Invertendo a f.d. da GP, os quantis extremais de probabilidade de excedência p, i.e.,

$$U_{H_\gamma}(1/p) = H_\gamma^{\leftarrow}(1 - p)$$

são dados por

$$U_{H_\gamma}(1/p) = \begin{cases} \sigma(p^{-\gamma} - 1)/\gamma, & \text{se} \quad \gamma \neq 0, \\ -\sigma \log p, & \text{se} \quad \gamma = 0. \end{cases}$$

[11] Davison, A.C. & Smith, R.L. (1990). Models for exceedances over high thresholds. *J. Royal Statist. Soc.* **B 52**, 393–442.

Para $\gamma < 0$, o limite superior do suporte de H_γ é finito e dado por

$$U_{H_\gamma}(\infty) = H_\gamma^{\leftarrow}(1) = -\sigma/\gamma\,.$$

Substituindo o vector de parâmetros desconhecidos, (γ, σ), por estimativas teremos a estimativa para os quantis extremais e para o limite superior do suporte da GP, H_γ.

Considerando que $X \frown F$, a distribuição condicional dos excessos acima de u,

$$F_u(y) = \mathbb{P}[X - u \leq y | X > u] = \frac{F(u+y) - F(u)}{1 - F(u)}$$

é aproximada pela GP

$$F_u(y) \approx H_\gamma(y; \sigma).$$

Então, como $F(u+y) = F_u(y)\{1 - F(u)\} + F(u)$, ou seja, com $x = u + y$

$$1 - F(x) = \{1 - F(u)\}[1 - F_u(x-u)] \Leftrightarrow \overline{F}(x) = \overline{F}(u)[1 - F_u(x-u)],$$

tem-se que

$$\overline{F}(x) \approx \overline{F}(u)[1 - H_\gamma(x - u; \sigma)].$$

Para estimar então a *probabilidade de excedência* de x elevado, $\overline{F}(x)$, considera-se a aproximação

$$\overline{F}(x) \approx \overline{F}(u)\,(1 + \gamma(x-u)/\sigma)^{-1/\gamma}\,.$$

Estimando $\overline{F}(u)$ pela frequência relativa N_u/n das observações que excedem u na amostra original de dimensão n, $X_1, \ldots, X_n$, vem que

$$\widehat{\overline{F}}(x) = \frac{N_u}{n}\,(1 + \hat{\gamma}(x-u)/\hat{\sigma})^{-1/\hat{\gamma}}\,.$$

O estimador de *quantis elevados* de F, $U\,(1/p) = F^{\leftarrow}(1-p)$, é então dado por

$$\widehat{U}\,(1/p) = u + \frac{\hat{\sigma}}{\hat{\gamma}}\left(\left(\frac{np}{N_u}\right)^{-\hat{\gamma}} - 1\right)$$

(*basta considerar* $F(x) = 1 - p$ *na expressão do estimador da probabilidade de excedência e inverter*). Para $\gamma < 0$, o estimador do *limite superior do suporte* para F é

$$\hat{x}^F = \widehat{U}(\infty) = u - \hat{\sigma}/\hat{\gamma}\,.$$

Intervalos de Confiança (IC). Os IC's para os parâmetros da GP(γ, σ), na abordagem POT, decorrem da aproximação à Normal dos estimadores ML, apoiada pela normalidade assintótica de valor médio nulo para

$$\sqrt{N_u}(\hat{\gamma} - \gamma) \quad \text{e} \quad \sqrt{N_u}\left(\widehat{U}(1/p) - U(1/p)\right).$$

Em Beirlant *et al.* (2004) podemos consultar a construção de IC's quer para o parâmetro de forma γ, quer para quantis elevados $U(1/p)$, baseados na normalidade assintótica dos respectivos estimadores ML. Claro que se tratam de IC's centrados nas respectivas estimativas pontuais, uma vez que a Normal é simétrica. Tal como na abordagem MMA, podem ser obtidas para a abordagem POT estimativas intervalares de melhor qualidade, não obrigatoriamente centradas na estimativa ML, usando a função de *profile log-likelihood* para γ. O traçado de

$$\log L_p(\gamma) := \max_{\sigma|\gamma} \log L(\gamma, \sigma)$$

conduz ao IC para γ com grau de confiança $100(1-\alpha)\%$

$$IC_\gamma = \left\{\gamma : \log L_p(\gamma) \geq \log L_p(\hat{\gamma}) - \frac{\chi^2_1(1-\alpha)}{2}\right\}.$$

7.4 Breve referência à estimação do índice extremal

Muito frequentemente usamos, para estimar θ, o *índice extremal*, introduzido na Secção 6.6, a validade das seguintes relações:

$$\theta \quad = \quad \frac{1}{\text{limite tamanho médio dos grupos}} = \lim_{n\to\infty} \mathbb{P}[X_2 \leq u_n | X_1 > u_n].$$

Consideraremos dois estimadores alternativos.

I Considere-se $u = X_{n-\tau:n}, \quad \lfloor n/50 \rfloor \leq \tau \leq \lfloor n/5 \rfloor$.
Calcule-se para cada um dos valores de u

$$\hat{\theta}_{n,u}(1) \quad := \quad \frac{\sum_{j=1}^{n-1} I_{[X_j < u \leq X_{j+1}]}}{\sum_{j=1}^{n} I_{[X_j > u]}} = \frac{\text{número cruzamentos}}{\text{número excedências}}.$$

Seleccione-se o valor da estimativa correspondente, por exemplo, ao ponto médio da zona de estabilidade.

II Face a k_n blocos de dimensão r_n escolhidos arbitrariamente, calcule-se

$$\begin{aligned}\hat{\theta}_{n,u_i,1\leq i\leq k_n}(2) &= \frac{k_n}{\sum_{i=1}^{k_n}\sum_{j=(i-1)r_n+1}^{ir_n} I_{[X_j>u_i]}} \\ &= \frac{\text{número de blocos}}{\text{número total de excedências}}\end{aligned}$$

A escolha de $u_i, 1 \leq i \leq k_n$ não é problemática, pois é baseada no facto de dever ser 1 o número médio de cruzamentos em cada um dos k_n blocos de dimensão r_n, $k_n r_n = n$ (o segundo máximo local em cada bloco é consequentemente um candidato elegível!). Pode por exemplo tomar-se $k_n =$ menor inteiro que divide n e é maior que $\sqrt[3]{n}$.

Que fazer em seguida em qualquer das abordagens não clássicas? Escolham-se representantes dos grupos de excedências e trabalhe-se com eles e com as estruturas definidas em esquema i.i.d. Como escolher esses representantes? A estimativa de θ obtida fornece-nos indicações sobre:

- O número de observações de topo a considerar (aproximadamente $k\theta$) para obter uma amostra de dimensão k proveniente do modelo GEV *multivariado*, ou equivalentemente,
- O nível elevado a considerar para obter k excessos independentes e Paretianos.

7.5 Estimação do CTE

Suponhamos que estamos interessados em estimar o CTE, apresentado na Definição 7.1.4, parâmetro muito usado no âmbito da área financeira. Tem-se

$$\text{CTE}_p \equiv \mu_p = E[X|X > \chi_p],$$

onde χ_p é o quantil-p superior ($p = 5\%$, por exemplo), i.e., $\mathbb{P}[X > \chi_p] = p$. Considerando o quantil empírico dado pela e.o. central

$$X_{n-k+1:n}, \text{ com } k = \lfloor np \rfloor + 1,$$

ou mais geralmente, com k tal que $k/n \underset{n\to\infty}{\longrightarrow} p$, um possível estimador do CTE é dado pela média das observações acima desse quantil empírico, ou seja

$$D_k := \frac{1}{k} \sum_{i=n-k+1}^{n} X_{i:n} \, .$$

É possvél mostrar a validade do seguinte resultado:

Teorema 7.5.1 (Média de e.o.'s superiores)**.** *Tem-se*

$$\sqrt{k}(D_k - \mu_p) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Z \frown \mathcal{N}(0, \sigma_p^2 + p(\chi_p - \mu_p)^2),$$

com $\sigma_p^2 := \mathbb{V}ar[X|X > \chi_p]$.

Observação 7.5.1. *Os parâmetros* χ_p, μ_p *e* σ_p^2 *associados à variância assintótica podem ser estimados, respectivamente, por*

$$X_{n-k+1:n}, \qquad D_k \quad e \quad S_D^2 = \frac{1}{k-1} \sum_{i=n-k+1}^{n} (X_{i:n} - D_k)^2 \, .$$

7.6 Breve referência a extremos bivariados

Para extremos bivariados, ou mesmo multivariados, a situação é idêntica, embora levemente mais complicada. Admitamos por exemplo que temos as descargas diárias do mesmo rio, a determinada hora, mas em locais diferentes desse rio. É então evidente que essas medições estão positivamente correlacionadas, sendo a correlação tanto mais forte quanto mais próximos estiverem os locais. Da série de pares de observações diárias podemos recolher, em cada ano, os pares de máximos observados, e trabalhar com essa amostra de pares de máximos. Qual o modelo a utilizar em inferência subsequente?

Mais uma vez vamos repousar na teoria assintótica, que nos vai fornecer modelos, agora dependentes não só de parâmetros desconhecidos, mas também de uma função de dependência, também a ser estimada com base na amostra de pares de máximos.

Consideremos, por exemplo, o caso particular de margens **Gumbel**. Dada a amostra de pares aleatórios

$$(X_1, Y_1), \ldots, (X_n, Y_n),$$

independentes com f.d. $F(x, y)$, a distribuição limite do par

$$\Big(\max_{1\leq i\leq n} X_i, \max_{1\leq i\leq n} Y_i \Big),$$

convenientemente normalizado, é,

$$\Lambda(x, y) = [\Lambda(x)\ \Lambda(y)]^{k(y-x)}$$

se ambas as sucessões de valores máximos forem atraídas para a lei Gumbel $\Lambda(x) = \exp(-\exp(-x)),\ x \in \mathbb{R}$. A função de dependência $k(\cdot)$ satisfaz condições específicas mas vastas (veja-se Tiago de Oliveira[12], 1984), que permitem a consideração de diferentes modelos, de entre os quais referimos.

Modelo Logístico: $k(w|\theta) = \left(\frac{(1+e^{-w/(1-\theta)})^{1-\theta}}{1+e^{-w}} \right)$, $0 \leq \theta \leq 1$.

Modelo Misto: $k(w|\theta) = 1 - \theta \frac{e^w}{(1+e^w)^2}$, $0 \leq \theta \leq 1$.

Modelo de Gumbel: $k(w|\theta) = 1 - \theta \frac{min(1,e^w)}{1+e^w}$, $0 \leq \theta \leq 1$.

Modelo Biextremal: $k(w|\theta) = 1 - \frac{\min(\theta,e^w)}{1+e^w}$, $0 \leq \theta \leq 1$.

7.7 Resumo

Mais recentemente quer o método POT, quer o método MMO têm vindo a ser abordados sob um ponto de vista semi-paramétrico. O tipo de ajustamento utilizado para as maiores observações não se identifica então com uma forma paramétrica dependente de parâmetros de localização λ, de dispersão δ e de forma γ. Pressupõe-se apenas que F está no domínio de atracção para máximos de G_γ, sendo γ o único parâmetro a estimar, com base em algumas observações de topo, e de acordo com metodologia adequada. Será essa a abordagem a considerar no Capítulo 8. Mas antes disso procedemos a um breve resumo das abordagens mais frequentes em *Estatística de Extremos*.

Abordagens paramétricas

[12]Tiago de Oliveira, J. (1984) Bivariate models for extremes: statistical decision. In J. Tiago de Oliveira (ed.), *Statistical Extremes and Applications*, D. Reidel, 131–153.

I Modelo GEV univariado (para os m máximos de sub-amostras de dimensão n, $N = nm$.) (Abordagem clássica de GUMBEL)

II Modelo GEV multivariado (para as m maiores e.o.'s associadas à totalidade N das observações.)

III Modelo GEV multi-dimensional (modelo GEV multivariado para as i_j maiores observações, $j = 1, 2, \ldots, r$, em sub-amostras de dimensão m', $rm' = N$.)
$r = m$ ($m' = n$) e $i_j = 1$ para $1 \leq j \leq r$ origina I;
$m' = N$ ($r = 1$) e $i_1 = m$ origina II.

IV Modelo Paretiano (para os excessos $Y_j - u$, de um nível elevado u, convenientemente escolhido) [Abordagem POT ('Peaks Over Thresholds']

Abordagens semi Paramétricas

V Trabalha-se com as k maiores e.o.'s associadas à totalidade das N observações, não considerando um modelo paramétrico, mas admitindo apenas que $F \in D_M(G_\gamma)$, situação a ser abordada no Capítulo 8, e que pode ser encontrada com maior detalhe em de Haan & Ferreira (2006). Para recensões críticas recentes veja-se Gomes *et al.*[13] (2008) e Beirlant *et al.* (2012).

[13]Gomes, M.I., Canto e Castro, L., Fraga Alves, M.I. & Pestana, D.D. (2008). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. *Extremes* **11**:1, 3–34.

Capítulo 8

Estatística de Extremos: Abordagem Semi-Paramétrica

Tal como mencionámos anteriormente, em contexto semi-paramétrico não se supõe qualquer modelo paramétrico subjacente à amostra $(X_1, \ldots, X_n)$ de observações i.i.d. provenientes de $F(.)$. Limitamo-nos a admitir que

1. $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, para algum $\gamma \in \mathbb{R}$, com G_γ definida em (6.17),

2. e as $k+1$ observações de topo

$$X_{n:n} \geq X_{n-1:n} \geq \cdots \geq X_{n-k:n}$$

são consideradas acima de um *threshold* aleatório $X_{n-k:n}$ que se entende como uma e.o. *intermédia superior* ao admitir $k \equiv k_n$ tal que

$$k \to \infty \quad \text{e} \quad k/n \to 0, \quad \text{quando} \quad n \to \infty\,. \tag{8.1}$$

Baseamos pois a estimação de γ nas k e.o.'s de topo da amostra, com k sucessão intermédia de inteiros, i.e., inteiros entre 1 e n, a dimensão da amostra. Estes estimadores, em conjunto com estimadores semi-paramétricos de localização e escala (veja-se, por exemplo, de Haan & Ferreira, 2006), podem ser

usados para a estimação de quantis elevados, períodos de retorno de níveis elevados, probabilidades de excedências de níveis elevados e outros parâmetros de acontecimehtos extremos, sempre em contexto semi-paramétrico.

8.1 Condições de segunda ordem e de ordem superior

Em contexto semi-paramétrico, para além da condição de primeira ordem em (6.20), é frequente admitir a validade de uma condição de segunda ordem, que especifica a velocidade de convergência em (6.20). É então usual admitir a existência de uma função A, convergente para zero quando $t \to \infty$, tal que

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma,\rho}(x) := \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right), \quad (8.2)$$

$\forall x > 0$, onde $\rho \leq 0$ é um parâmetro de segunda ordem que controla a velocidade de convergência dos valores máximos, linearmente normalizados, para a lei limite em (6.17). Então $\lim_{t\to\infty} A(tx)/A(t) = x^\rho$, $\forall\ x > 0$, i.e. $|A| \in RV_\rho$ (de Haan & Stadtmüller[1], 1996).

De modo paralelo, e sempre que necessário, podemos pensar em condições de terceira ordem, que especificam a velocidade de convergência em (8.2). Para detalhes sobre uma condição geral de terceira ordem, veja-se Fraga Alves *et al.*[2] (2003, Appendix) e Fraga Alves *et al.*[3] (2006). A utilidade destas condições pode ser vista em artigos variados sobre estimação semi-paramétrica de viés reduzido, fora do âmbito deste livro.

[1] de Haan, L. & Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc.* **A61**, 381–395.

[2] Fraga Alves, M.I., de Haan, L. & Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods of Statist.* **12**, 155–176.

[3] Fraga Alves, M.I., de Haan, L. & Lin, T. (2006). Third order extended regular variation. *Publications de l'Institut Mathématique* **80** (94), 109–120.

8.2 Estimação semi-paramétrica do EVI

O número de estimadores semi-paramétricos do EVI é elevado. Limitamo-nos a referir quatros dos estimadores clássicos deste parâmetro crucial de acontecimentos extremos ou mesmo raros.

8.2.1 O estimador de Hill (H)

Para caudas pesadas, i.e. para $\gamma > 0$, um dos estimadores mais simples de γ foi o proposto por Hill[4] (1975). As suas propriedades têm sido estudadas por vários autores, de entre os quais mencionamos de Haan & Peng[5] (1998). O estimador de Hill tem a forma funcional

$$\widehat{\gamma}_{k,n}^{H} := \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n} =: M_{k,n}^{(1)}. \tag{8.3}$$

A consistência fraca do estimador em (8.3) é alcançada em $\mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$, sempre que for válida a condição de primeira ordem, em (6.20), e desde que k seja um valor intermédio, verificando pois as condições em (8.1).

8.2.2 O estimador de Pickands (P)

Por facilidade de exposição, denotemos novamente a i-ésima maior observação por $M_n^{(i)} = X_{n-i+1}$, $i = 1, \ldots, n$. Este estimador, introduzido em Pickands (1975) envolve para cada $k(= 4m)$ apenas 3 observações de topo

$$\cdots \geq M_n^{(m)} \geq \cdots\cdots \geq M_n^{(2m)} \geq \cdots\cdots\cdots\cdots \geq M_n^{(4m)}$$

e é definido por

$$\hat{\gamma}_{m,n}^{P} = \frac{1}{\log 2} \log\left[\frac{M_n^{(m)} - M_n^{(2m)}}{M_n^{(2m)} - M_n^{(4m)}}\right], \qquad m = 1, 2, \ldots, \lfloor n/4 \rfloor,$$

onde $\lfloor a \rfloor$ designa novamente a parte inteira de $a > 0$. Por uma questão comparativa com os estimadores precedentes, é usual considerar-se $k = 4m$, e

[4] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

[5] de Haan, L. & Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60–70.

a forma funcional

$$\widehat{\gamma}^{P}_{k,n} := \frac{1}{\log 2} \log \frac{X_{n-\lfloor k/4 \rfloor+1:n} - X_{n-\lfloor k/2 \rfloor+1:n}}{X_{n-\lfloor k/2 \rfloor+1:n} - X_{n-k+1:n}}. \tag{8.4}$$

A consistência fraca deste estimador do EVI, sob a validade de uma condição de primeira ordem do tipo da considerada em (6.20) e para k intermédia, foi demonstrada por Pickands (1975) para qualquer $\gamma \in \mathbb{R}$. Um estudo aprofundado deste estimador, como estimador baseado em estatísticas intermédias e extremais, pode ser visto em Dekkers & de Haan[6] (1989).

8.2.3 O estimador dos Momentos (M)

O estimador dos Momentos aplica-se no caso geral de $\gamma \in \mathbb{R}$.

Definam-se

$$M^{(r)}_{k,n} := \frac{1}{k} \sum_{i=1}^{k} (\log X_{n-i+1:n} - \log X_{n-k:n})^r, \quad r = 1, 2\,,$$

e ainda

$$\hat{\gamma}^{+}_{k,n} = M^{(1)}_{k,n} = \hat{\gamma}^{H}_{k,n}, \qquad \hat{\gamma}^{-}_{k,n} = 1 - \frac{1}{2} \left\{ 1 - \frac{(M^{(1)}_{k,n})^2}{M^{(2)}_{k,n}} \right\}^{-1}. \tag{8.5}$$

O estimador dos Momentos (Dekkers *et al.*[7], 1989) de

$$\gamma = \gamma_- + \gamma_+ \qquad (\gamma_- := \min(0, \gamma),\ \gamma_+ := \max(0, \gamma))$$

é

$$\hat{\gamma}^{M}_{k,n} = \hat{\gamma}^{-}_{k,n} + \hat{\gamma}^{+}_{k,n} \equiv M^{(1)}_{k,n} + \tfrac{1}{2} \Big\{ 1 - \big(M^{(2)}_{k,n} / [M^{(1)}_{k,n}]^2 - 1 \big)^{-1} \Big\}. \tag{8.6}$$

Mais uma vez, consegue-se demonstrar a consistência fraca destes estimadores $\forall\, \gamma \in \mathbb{R}$ sempre que se tiver a validade de (6.20) e de (8.1).

[6] Dekkers, A.L.M. & de Haan, L. de (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics* **17**, 1795–1832.

[7] Dekkers, A.L.M., Einmahl, J.H.J. & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833–1855.

8.2.4 O estimador POT-ML (ML)

Trata-se de um estimador levemente mais complicado que os anteriormente referidos, mas com propriedades interessantes, e que não queremos deixar de referir nesta breve introdução à estimação do EVI. Tal como referido em de Haan & Ferreira (2006), a classe de funções de distribuição $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, para algum $\gamma \in \mathbb{R}$, não pode ser parametrizada num número finito de parâmetros, e consequentemente não existe um estimador ML para γ nessa classe vasta de modelos. Existe no entanto um estimador, introduzido em Smith[8] (1987) usualmente denotado estimador ML. Esse estimador está associado à aproximação GP para os excessos acima de uma observação elevada. Os excessos $V_{ik} := X_{n-i+1,n} - X_{n-k,n}$, $1 \le i \le k$, são aproximadamente as k e.o.'s de topo associadas a uma amostra de dimensão k de uma GP com f.d. $H_\gamma(\tau x/\gamma) = 1 - (1+\tau x)_+^{-1/\gamma}$, com a reparametrização $\tau = \gamma/\sigma$ (ver também secção 7.3.2 e observação 7.3.2).

A solução das equações ML equations associadas (Davison[9], 1984) dá origem a um estimador explícito do EVI,

$$\widehat{\gamma}_{k,n}^{ML} := \frac{1}{k}\sum_{i=1}^{k} \log(1+\hat{\alpha}\ V_{ik}), \tag{8.7}$$

onde $\hat{\alpha}$ é o estimador (implícito) ML do parâmetro de escala α, também desconhecido. Um estudo exaustivo das propriedades assintóticas do estimador ML em (8.7) foi feito em Drees *et al.*[10] (2004). Conseguimos mais uma vez consistência fraca sempre que temos a validade de (6.20) e de (8.1), com $\gamma > -1$ (para a região $-1 < \gamma \le 1/2$, veja-se Zhou[11], 2009).

[8] Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–1207.

[9] Davison, A. (1984). Modeling excesses over high threshold with an application. In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, D. Reidel, 461–482.

[10] Drees, H., Ferreira, A. & de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**, 1179–1201.

[11] Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Analysis* **100**:4, 794–815.

8.2.5 Normalidade assintótica dos estimadores

Face à validade da condição de segunda ordem, em (8.2), é possível garantir a normalidade assintótica dos estimadores atrás referidos. Mais precisamente, denotemos genericamente E qualquer dos estimadores H, P, M e ML. É então possível garantir para $\gamma \in C_{_E} \subset \mathbb{R}$, a existência de constantes reais $\sigma_{_E} > 0$, tais que:

$$\widehat{\gamma}^{E}_{k,n} \stackrel{d}{=} \gamma + \sigma_{_E} P^{E}_{k}/\sqrt{k} + O_p(A(n/k)), \tag{8.8}$$

com P^{E}_{k} assintoticamente normal padrão, $C_{_H} = \mathbb{R}^+$, $C_{_P} = C_{_M} = \mathbb{R}$ e $C_{_{ML}} = (-1, +\infty)$ (para a região $(-1, 1/2]$, veja-se Zhou[12] 2010). Consequentemente, para valores de k tais que $\sqrt{k}\, A(n/k) \to 0$, quando $n \to \infty$,

$$\sqrt{k}\left(\hat{\gamma}^{E}_{k,n} - \gamma\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2_{_E}). \tag{8.9}$$

Os valores $\sigma^2_{_E}$ são usualmente denotados como *variância assintótica* de $\hat{\gamma}^{E}_{k,n}$. Essas variâncias assintóticas são dadas por

$$\sigma^2_{_H} = \gamma^2, \qquad \sigma^2_{_P} = \begin{cases} \frac{\gamma^2(2^{2\gamma+1}+1)}{(\log 2)^2(2^\gamma-1)^2}, & \text{se } \gamma \neq 0 \\ \frac{3}{(\log 2)^4}, & \text{se } \gamma = 0, \end{cases}$$

$$\sigma^2_{_{ML}} = (1+\gamma)^2, \quad \sigma^2_{_M} = \begin{cases} 1+\gamma^2 & \text{se } \gamma \geq 0 \\ \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)} & \text{se } \gamma < 0. \end{cases}$$

Na Figura 8.1, procedemos à comparação das variâncias assintóticas para os estimadores do EVI atrás referidos.

8.2.6 ICs semi-paramétricos e assintóticos para o EVI

Então, para as condições convenientes sobre k_n e sobre a cauda $\overline{F}$ atrás mencionadas e para qualquer dos estimadores considerados — Hill (H), Pickands (P), Momentos (M) e ML — tem-se, como vimos, o seguinte comportamento assintótico:

$$\sqrt{k}\left(\hat{\gamma}^{E}_{k,n} - \gamma\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2_{_E}),$$

[12] Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *J. Multivariate Analysis* **101**:4, 971–983.

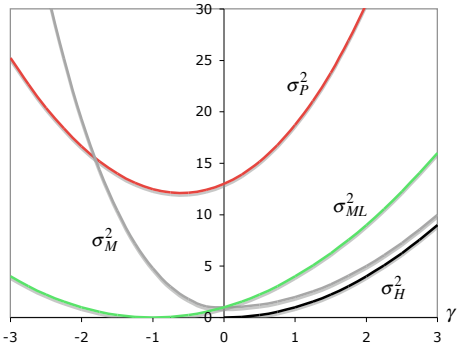


Figura 8.1: Comparação das variâncias assintóticas dos estimadores de Hill, Pickands, Momentos e POT-ML

com

$$E = H,\, M,\, P,\, ML, \quad \text{e em que } \sigma_E^2 \equiv \sigma_E^2(\gamma).$$

Esta normalidade assintótica permite calcular ICs aproximados a $100 \times (1 - \alpha)\%$ para γ, dados por

$$\hat{\gamma}_{k,n}^E \pm z_{1-\alpha/2}\sqrt{\hat{\sigma}_E^2/k},$$

com $\hat{\sigma}_E^2 := \sigma_E^2(\hat{\gamma})$ e denotando z_p o quantil de probabilidade p de uma $\mathcal{N}(0,1)$.

8.2.7 Observações adicionais

1. É de realçar que estes estimadores são assintóticamente centrados apenas para valores de níveis $k = k_n$ convenientes que estejam nas condições suplementares atrás referidas.

2. Para valores maiores de $k = k_n$ este tipo de estimadores apresenta um viés que pode ser considerável.

3. De um modo geral,
   - para valores pequenos de k, ou equivalentemente para níveis elevados, estes estimadores apresentam uma grande variabilidade,
   - para valores grandes de k, ou equivalentemente para níveis mais baixos, estes estimadores apresentam um grande viés.

A escolha conveniente de k para a estimativa a considerar para γ deve ter em conta um balancear entre o viés e a variância, e têm recentemente sido considerados estimadores de viés reduzido, menos sensíveis à escolha de k. Este estudo sai fora do âmbito deste livro, mas podem encontrar-se detalhes sobre este tipo de estimadores e sobre a escolha da fração óptima, em Gomes *et al.*[13] (2007), e nas recensões críticas em Gomes *et al.* (2008) e Beirlant *et al.* (2012).

8.3 Estimação semi-paramétrica de outros parâmetros de acontecimentos extremos

Quantis elevados, ou equivalentemete na área financeira o *Value-at-Risk* no nível p, VaR_p, são talvez os parâmetros de acontecimentos extremos de maior relevância, funções do EVI e de parâmetros de localização e escala. A motivação para estes estimadores reside na condição de primeira ordem em (6.20), que reescrevemos:

$$F \in \mathcal{D}_{\mathcal{M}}(G_\gamma), \quad \gamma \in \mathbb{R}$$
$$\iff \lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \text{se} \quad \gamma \neq 0, \\ \log x, & \text{se} \quad \gamma = 0, \end{cases}$$

para alguma função positiva $a(.)$ e com $x > 0$. Pode-se ainda escrever

$$F \in \mathcal{D}_{\mathcal{M}}(G_\gamma),$$
$$\iff \lim_{t\uparrow x^F} \frac{1-F(t+xg(t))}{1-F(t)} = \begin{cases} (1+\gamma x)^{-1/\gamma}, & \text{se} \quad \gamma \neq 0, \\ \exp(-x), & \text{se} \quad \gamma = 0, \end{cases}$$

tendo-se $g(t) = a\left(1/(1-F(t))\right)$.

[13] Gomes, M.I., Reiss, R.-D.& Thomas, M. (2007). Reduced-bias estimation. In Reiss, R.-D. & Thomas, M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd Ed., Chapter 6, 189–204, Birkhäuser Verlag, Basel-Boston-Berlin.

8.3.1 Estimação de quantis extremais

Em contexto semi-paramétrico, o estimador mais usual de um quantil extremal

$$\chi_{1-p} := U(1/p) = F^{\leftarrow}(1-p),$$

com p próximo de zero (*ou teoricamente* $p \equiv p_n \to 0$*, com* $p_n \ll k/n$*, quando* $n \to \infty$), pode ser facilmente derivado de (6.20), que fornece a aproximação

$$U(tx) \approx U(t) + a(t)(x^{\gamma}-1)/\gamma.$$

O facto de se ter $X_{n-k+1:n} \overset{p}{\sim} U(n/k)$ permite-nos considerar $t = n/k$, $x = k/(np)$ e estimar χ_{1-p} com base nesta aproximação e em estimativas adequadas do parâmetro de forma γ e do parâmetro de escala $a(n/k)$, uma vez que

$$U(1/p) \approx U(n/k) + a(n/k)\,\frac{(k/(np))^{\gamma}-1}{\gamma}.$$

Prova-se (de Haan & Ferreira, 2006, Teorema 4.2.1) que um estimador consistente de $a(n/k)$, no sentido de que $\hat{a}(n/k)/a(n/k) \xrightarrow[n\to\infty]{p} 1$, é dado por

$$\hat{a}(n/k) = X_{n-k:n} M_{k,n}^{(1)}(1-\hat{\gamma}_{k,n}^{-}),$$

vindo assim, para $\gamma \neq 0$ real, o estimador do *quantil extremal*,

$$\widehat{U}(1/p) = X_{n-k:n} + \hat{a}(n/k)\,\frac{(k/(np))^{\hat{\gamma}}-1}{\hat{\gamma}}.$$

com

$$\hat{a}(n/k) = X_{n-k:n} M_{k,n}^{(1)}(1-\hat{\gamma}_{k,n}^{-}) \quad \text{e} \quad \hat{\gamma} = \hat{\gamma}_{k,n}^{E},\ E = H,\ M,\ P,\ ML,$$

com $\hat{\gamma} = \hat{\gamma}_{k,n}^{E},\ E = H,\ M,\ P,\ ML$ e $\hat{\gamma}_{k,n}^{-}$ definido em (8.5).

Para $\gamma = 0$, o *quantil extremal* é estimado por

$$\widehat{U}(1/p) = X_{n-k:n} + \hat{a}\left(\frac{n}{k}\right)\log\left(\frac{k}{np}\right).$$

Para γ positivo, a aproximação $U(tx) \approx U(t)x^{\gamma}$, permite-nos obter o estimador simplificado,

$$\widehat{U}(1/p) = X_{n-k:n}\left(\frac{k}{np}\right)^{\hat{\gamma}},$$

por vezes designado por estimador de Weissman (Weissman[14], 1978), mais uma vez com $\hat{\gamma}$ qualquer estimador consistente do EVI.

Nenhum dos estimadores de quantis elevados atrás mencionados reage adequadamente a mudanças de localização nos dados. Araújo Santos *et al.* (2006) estudaram uma classe de estimadores semi-paramétricos de um *quantil elevado* em concordância com uma propriedade desejável para quantis face à presença de transformações lineares nos dados. Essa propriedade tem a ver com a linearidade teórica de um quantil, χ_p, i.e., com o facto de se ter $\chi_p(\delta X + \lambda) = \delta\chi_p(X) + \lambda$, $\forall\ \lambda \in \mathbb{R}$ e $\delta > 0$. Essa classe de estimadores é baseada na amostra dos excessos acima de um limiar aleatório, i.e., é baseada na metodologia PORT, e fornece-nos as propriedades desejáveis para as medidas de risco em finanças: equivariância para as translações e homogeneidade positiva.

8.3.2 Estimação semi-paramétrica do limite superior do suporte

Uma vez que $U(t) = F^{\leftarrow}(1 - 1/t)$, o *limite superior do suporte*, pode ser escrito como $x^F = F^{\leftarrow}(1)$, i.e., pode ser expresso como função de U fazendo na expressão do quantil extremal $p = 0$, i.e., $U(\infty) = F^{\leftarrow}(1) = x^F$.

Para o caso de $\gamma < 0$, i.e., caudas curtas de f.d.s no max-domínio da Max-Weibull, um estimador do *limite superior do suporte* (que é *finito*) é obtido a partir do estimador do *quantil extremal* fazendo $p = 0$. Obtemos

$$\hat{x}^F = X_{n-k:n} - \hat{a}\left(\frac{n}{k}\right)/\hat{\gamma},$$

com $\hat{\gamma} = \hat{\gamma}^E_{k,n}, \quad E = M,\ P,\ ML,$ ou ainda $\hat{\gamma} = \hat{\gamma}^-_{k,n}$, em (8.5). A estimação do limite superior do suporte em contexto semi-paramétrico foi iniciado em Hall[15] (1982), Csörgő & Mason[16] (1989) e Aarssen & de Haan[17] (1994), por

[14] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812–815.

[15] Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10**, 556–568.

[16] Csörgő, S., & Mason, D.M. (1989). Simple estimators of the endpoint of a distribution. In Hüsler, J., & Reiss, R.-D. (1989), *Extreme Value Theory, Proceedings Oberwolfach* 1987, 132–147, Springer-Verlag, Berlin, Heidelberg.

[17] Aarssen, K., & de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* **4**:4, 259–281.

entre outros autores.

Observação 8.3.1. *Na prática a implementação do estimador do 'endpoint' deverá ter em consideração o facto de que se deve ter* $\hat{x}^F = \widehat{U}(\infty) \geq X_{n:n}$.

8.3.3 Estimação semi-paramétrica da probabilidade de excedência

O problema dual do da estimação de quantis elevados, i.e. a estimação da probabilidade de excedência de um nível fixo elevado, foi inicialmente abordado em contexto semi-paramétrico por Dijk & de Haan[18] (1992) e Ferreira[19] (2002), por entre outros autores. Trata-se do problema de estimar a *probabilidade de ultrapassagem* de um nível elevado x, i.e., $p = 1 - F(x)$. Teoricamente, os resultados são obtidos para níveis elevados

$$x = x_n, \quad \text{tais que} \quad 1 - F(x_n) =: p_n \to 0, \quad \text{quando} \quad n \to \infty.$$

Mostra-se que um estimador consistente de

$$p_n = 1 - F(x_n),$$

no sentido de que $\hat{p}_n / p_n \xrightarrow[n\to\infty]{p} 1$, para $\gamma \neq 0$, é dado por

$$\hat{p}_n = \frac{k}{n} \left\{ \max\left(0, 1 + \hat{\gamma} \frac{x_n - X_{n-k:n}}{\hat{a}\left(\frac{n}{k}\right)}\right) \right\}^{-1/\hat{\gamma}} \tag{8.10}$$

e para $\gamma = 0$ é dado por continuidade,

$$\hat{p}_n = \frac{k}{n} \exp\left\{ -\frac{x_n - X_{n-k:n}}{\hat{a}\left(\frac{n}{k}\right)} \right\}.$$

No caso de um EVI positivo, $\gamma > 0$, um estimador consistente mais simples de $p_n = 1 - F(x_n)$, é dado por

$$\hat{p}_n = \frac{k}{n} \left\{ \frac{x_n}{X_{n-k:n}} \right\}^{-1/\hat{\gamma}}. \tag{8.11}$$

[18] Dijk, V. & de Haan, L. (1992). On the estimation of the exceedance probability of a high level. Order statistics and nonparametrics: theory and applications. In Sen, P. K., & Salama, I. A. (eds.), 79–92, Elsevier, Amsterdam.

[19] Ferreira, A. (2002). Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Plannning and Inference* **104**, 83–102.

Relembremos que uma vez que existe g, por exemplo, $g(t) = a\,(1/(1-F(t)))$, tal que

$$\lim_{t\uparrow x^F} \frac{1-F(t+xg(t))}{1-F(t)} = (1+\gamma x)^{-1/\gamma},$$

para todo o $x : 1+\gamma x > 0$, então isso equivale a dizer que

$$\lim_{t\uparrow x^F} \mathbb{P}\left[\frac{X-t}{g(t)} > x \Big| X > t\right] = (1+\gamma x)^{-1/\gamma} = 1 - H_\gamma(x),$$

i.e., a distribuição de $(X-t)/g(t)$ condicional a $X > t$ tem por distribuição limite, quando $t \uparrow x^F$, a generalizada Pareto, GP(γ). Assim, a partir de um nível elevado t, (i.e., $X > t$),

$$\mathbb{P}\left[X > t + g(t)x\right] \approx \mathbb{P}[X > t](1 - H_\gamma(x)),$$

ou ainda

$$\mathbb{P}\left[X > x\right] = \mathbb{P}\left[X > t + g(t)\tfrac{x-t}{g(t)}\right] \approx \mathbb{P}[X > t]\left\{1 - H_\gamma(\tfrac{x-t}{g(t)})\right\},\ \ x > t,$$

ou seja,

$$1 - F(x) \approx (1-F(t))\left\{1 - H_\gamma(\tfrac{x-t}{g(t)})\right\},\ x > t.$$

Ora, fazendo $t := X_{n-k:n}$, nível aleatório intermédio,

$$1 - F(x) \approx (1 - F(X_{n-k:n}))\left\{1 - H_\gamma\left(\tfrac{x-X_{n-k:n}}{g(X_{n-k:n})}\right)\right\}.$$

Relembrando que $g(t) = a\,(1/(1-F(t)))$, tem-se

$$g(X_{n-k:n}) = a\left(\frac{1}{1-F(X_{n-k:n})}\right) \approx a\left(\tfrac{n}{k}\right),$$

uma vez que $1 - F(X_{n-k:n}) \approx k/n$.

Vem por fim o estimador da probabilidade de excedência $p = 1 - F(x)$

$$\hat{p} = \frac{k}{n}\left\{1 - H_{\hat\gamma}\left(\tfrac{x-X_{n-k:n}}{\hat a(\frac{n}{k})}\right)\right\},$$

e consequentemente, o estimador em (8.10).

Para o caso de $\gamma > 0$ o estimador simplificado da probabilidade de excedência é motivado pelo facto de se poder considerar

$$\hat a\left(\frac{n}{k}\right) = X_{n-k:n} M_{k,n}^{(1)},$$

já que $\hat{\gamma}^{-}_{k,n} \xrightarrow[n\to\infty]{p} 0$. Assim, como $\hat{\gamma}/M^{(1)} \xrightarrow[n\to\infty]{p} 1$, verifica-se

$$1+\hat{\gamma}\frac{x-X_{n-k:n}}{\hat{a}\left(\frac{n}{k}\right)} = 1+\hat{\gamma}\frac{x-X_{n-k:n}}{X_{n-k:n}M^{(1)}_{k,n}} \approx 1+\frac{x-X_{n-k:n}}{X_{n-k:n}} = \frac{x}{X_{n-k:n}}$$

vindo o estimador em (8.11).

8.4 Invariância versus não-invariância

Muitos dos estimadores do EVI são baseados nos excessos das log-observações, e consequentemente, não são invariantes a mudanças na localização.

A propriedade de invariância dos estimadores do EVI face a mudanças de localização e escala é estatisticamente apelativa. Contudo, com estimadores fortemente dependentes de k, o número de e.o.'s usadas, temos de pensar que até é sensato considerar, tal como sugerido em Gomes & Oliveira[20] (2003), a introdução de um *parâmetro de controlo* que é simplesmente uma mudança de localização determinística, τ. Então, os estimadores, funções de $X+\tau$, em vez de X, nem sequer são invariantes para mudanças de escala, mas podemos jogar com τ, procurando qual o valor de τ que, através de qualquer critério de estabilidade sensato, produz a mais elevada estabilidade em torno de um poto alvo, que deverá ser o parâmetro γ, desde que estejamos a trabalhar com estimadores consistentes de γ. Este truque é simples, fornece resultados interessantes com elevada frequência, mas deve ser usado com algum cuidado. Para uma discussão mais aprofundada sobre este assunto, recomendamos a leitura de Araújo Santos *et al.* (2006), onde quer o estimador de Hill quer o estimador de Momentos, respectivamente em (8.3) e (8.6), são transformados em estimadores invariantes para mudanças de localização e escala através do uso da amostra dos excessos,

$$X^*_i := X_i - X_{\lfloor np\rfloor+1:n}, \quad 0<p<1, \quad 1\le i\le n. \tag{8.12}$$

[20] Gomes, M.I. & Oliveira, O. (2003). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events. *Comm. in Statist.—Simulation and Computation* **32**:4, 1005–1028.

Processo semelhante foi usado em Fraga Alves *et al.*[21] (2009), onde é proposto um novo estimador do EVI, o chamado estimador de *momentos mistos*, sendo introduzidas classes alternativas invariantes para mudanças de localização e escala, também dependentes de um *parâmetro de controlo* p, $0 < p < 1$. Qualquer desses estimadores tem a mesma forma funcional do estimador original E, digamos, mas as observações originais X_i são substituídas por X_i^* in (8.12), $1 \leq i \leq n$. É então sensato definir para qualquer *parâmetro de controlo* p, $0 < p < 1$

$$\hat{\gamma}_{k,n}^E(p) := \hat{\gamma}_{k,n}^E(X_{n-j+1:n} - X_{[np]+1:n}, 1 \leq j \leq k+1). \tag{8.13}$$

As versões invariantes para a localização, em (8.13), têm propriedades semelhantes às do estimador original $\hat{\gamma}_{k,n}^E(X_{n-j+1:n}, 1 \leq j \leq k+1)$, desde que escolhamos os níveis k e o *parâmetro de controlo* p adequados.

[21] Fraga Alves, M.I., Gomes, M.I., de Haan, L. & Neves, C. (2009). Mixed moment estimators and location invariant alternatives. *Extremes* **12**, 149–185.

Capítulo 9

Estatística de Extremos: Casos de Estudo

9.1 Dados 'maasmax.txt'

Trata-se de um conjunto de dados, $Y_1, \ldots, Y_m$, de descargas anuais máximas do rio Meuse, em m^3/s, no período 1911–1995, num total de $m = 85$ anos (Beirlant *et al.*, 2004; `http://lstat.kuleuven.be/Wiley/`), que começámos a analisar na Secção 4.5. Tratam-se de **réplicas da v.a.** $Y \equiv M_n$, com $M_n :=$ máximo anual, i.e.,

$$M_n = \max(X_1, \ldots, X_n), \qquad n = 365\,(dias).$$

Caso decidissemos avançar com uma análise estatística tradicional, fazendo um ajustamento do *máximo anual* a um modelo Normal, $\mathcal{N}(\mu, \sigma)$, o QQ-plot e o *método dos mínimos quadrados*, levaram-nos na Secção 4.5, às estimativas $\hat{\mu} = 1495.962$ e $\hat{\sigma} = 551.0057$, a que corresponde um coeficiente de correlação de $r = 0.9788504$. Então, o nível de retorno a T-anos,

$$U(T) = F_Y^{\leftarrow}\,(1 - 1/T)\,,$$

i.e., o nível das descargas ultrapassado pelo máximo anual todos os $T = 100$

anos, em média, seria estimado por

$$\begin{aligned}\widehat{U}(100) &= \hat{\mu} + \hat{\sigma}\,\Phi^{\leftarrow}(1-1/100)\\ &= 1495.962 + 551.0057\,\Phi^{\leftarrow}(0.99) = 2777.793.\end{aligned}$$

Se pensarmos no período de retorno do nível $y_{_T}$, com

$$T = \frac{1}{\mathbb{P}[Y > y_{_T}]} = \frac{1}{1 - F_Y(y_{_T})},$$

i.e., período T (em média) em que é ultrapassado o nível $y_{_T} = 3175$ pelo máximo anual, obtemos a estimativa

$$\hat{T} = \frac{1}{1 - \Phi\left(\frac{3175-\hat{\mu}}{\hat{\sigma}}\right)} \simeq 866 \text{ anos.}$$

Observação 9.1.1. *Note-se que o valor de $y_{_T}$ escolhido aqui corresponde ao máximo amostral observado para a amostra de máximos anuais ao longo dos* 85 *anos, i.e. escolhemos $y_{_T} = y_{85:85} = 3175$.*

Dada a natureza dos dados e o facto de se verificar uma assimetria à direita, justifica-se uma análise estatística que leve em consideração os modelos extremais, começando por um ajustamento do *máximo anual* ao modelo Gumbel, tal como também fizemos na Secção 4.5. Fomos então levados à Gumbel estimada, $Y \frown \Lambda(\hat{b} = 1247.363, \hat{a} = 445.6884)$.

Pensemos em seguida no nível de retorno a T-anos, $U(T) = F_Y^{\leftarrow}(1 - 1/T)$. Esse nível de retorno é estimado por

$$\begin{aligned}\widehat{U}(100) &= \hat{b} + \hat{a}\,\Lambda^{\leftarrow}(1-1/100)\\ &= 1247.363 + 445.6884\,\Lambda^{\leftarrow}(0.99) = 3297.596,\end{aligned}$$

acima do valor estimado face a um ajustamento ao modelo Normal. *É pois de notar que a normal sub-estima o nível de retorno de $(T = 100)$-anos.*

Se pensarmos no período de retorno do nível $y_{_T} = 3175$,

$$T = \frac{1}{\mathbb{P}[Y > y_{_T}]} = \frac{1}{1 - F_Y(y_{_T})},$$

obtemos a estimativa

$$\hat{T} = \frac{1}{1 - \Lambda\left(\frac{3175-\hat{b}}{\hat{a}}\right)} \simeq 76 \ \textit{anos},$$

valor este consideravelmente inferior ao encontrado no caso do tratamento com a normal.

Consideremos agora de forma mais alargada o modelo GEV, em (7.1), como candidato a modelar o máximo anual Y.

Estudemos para que valor de γ se obtem a correlação máxima nos QQ-plot respectivos, quando $\gamma \in [-0.5, 0.5]$.

Essa correlação é reproduzida na Figura 9.1.

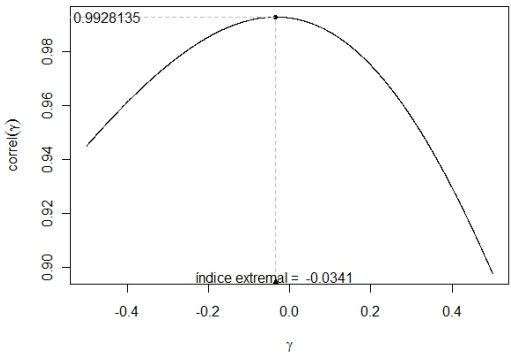


Figura 9.1: Estimativa de γ associada a correlação máxima

Obtém-se então $\hat{\gamma} = -0.03419188$, e o QQ-plot GEV associado a este valor de γ é representado na Figura 9.2.

Ajustando uma recta de mínimos quadrados, obtem-se para parâmetros de localização e escala, respectivamente $\hat{b} = 1251.887$ e $\hat{a} = 461.6413$, a que corresponde um coeficiente de correlação de $r = 0.9928135$, superior ao encontrado no caso do ajustamento à Gumbel feito na Seção 4.5.

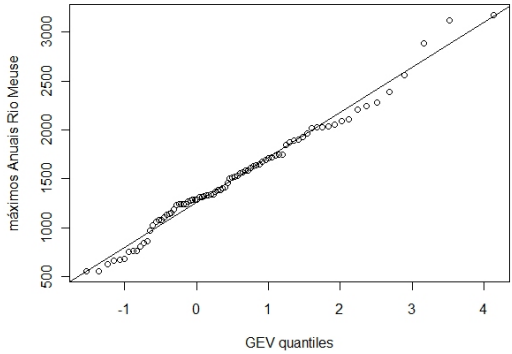


Figura 9.2: QQ-plot GEV ($\hat{\gamma}$): $\left\{(G^{\leftarrow}_{\hat{\gamma}}(i/(n+1)), y_{i:m}) : i = 1, \ldots, m\right\}$

Sobrepondo ao histograma dos dados (máximos anuais) a f.d.p. GEV estimada, $Y \sim G_{\hat{\gamma}}(\hat{b} = 1251.887, \hat{a} = 461.6413)$, $\hat{\gamma} = -0.03419188$, obtem-se a Figura 9.3.

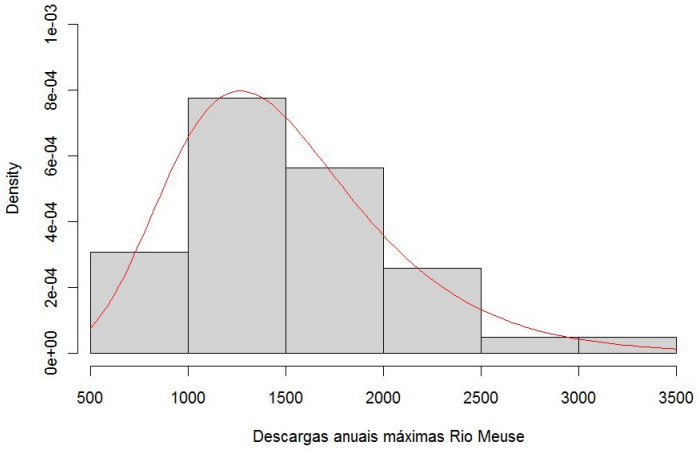


Figura 9.3: Histograma e GEV ajustada

O nível de retorno a T-anos, $U(T) = F_{Y}^{\leftarrow}(1 - 1/T)$, i.e., o nível das descargas ultrapassado pelo máximo anual todos os $T = 100$ anos, em média, é estimado por

$$\begin{aligned}\hat{U}(100) &= \hat{b} + \hat{a}\, G_{\hat{\gamma}}^{\leftarrow}(1 - 1/100) \\ &= 1251.887 + 461.6413\, G_{\hat{\gamma}}^{\leftarrow}(0.99) = 3216.919,\end{aligned}$$

com $\hat{\gamma} = -0.03419188$. *De notar que a* Gumbel *sobre-estima o nível de retorno de* $(T = 100)$*-anos.*

O período de retorno do nível y_{T},

$$T = \frac{1}{\mathbb{P}[Y > y_{T}]} = \frac{1}{1 - F_{Y}(y_{T})},$$

i.e., o período T (em média) em que é ultrapassado o nível $y_{T} = 3175$ pelo máximo anual é estimado por

$$\hat{T} = \frac{1}{1 - G_{\hat{\gamma}}\left(\frac{3175 - \hat{b}}{\hat{a}}\right)} \simeq 90 \text{ anos},$$

valor este superior ao encontrado no caso da aproximação Gumbel*, mas bastante inferior ao obtido pela abordagem clássica normal.*

Fez-se neste caso o ajustamento da $\text{GEV}(\gamma; \lambda, \delta)$ à v.a. $Y := \max(X_1, \ldots, X_n)$, tendo por base uma amostra de dimensão m de máximos anuais $Y_1, Y_2, \cdots, Y_m$.

Tal como no Capítulo 4, conduzimos uma abordagem preliminar, escolhendo o γ que maximiza a correlação no QQ-Plot, e em seguida os parâmetros de localização/escala, λ, δ, pelos *método dos mínimos quadrados.*

Vamos agora utilizar o método ML na estimação dos parâmetros desconhecidos, e tal como referimos nos Capítulos 2 e 7, vamos fazer uso de alguns packages de `R` para ajustamento das max-estáveis, relativamente a este caso de estudo.

Método ML

```
library(ismev)
```

```
#####################################################
#      Max Anuais   Máxima-Verosimilhança           #
#####################################################
maas<-read.table('maasmax.txt', header=FALSE,dec=".")
y<-maas$V1       ## DADOS ##
#####################################################
#            GUMBEL - ML     library(ismev)       #
#####################################################
library(ismev)
gum.fit(y)          # library(ismev)
dGumbel <- function(x,a,b) 1/a*exp((b-x)/a)*exp(-exp((b-x)/a))
gumbel_fit<-gum.fit(y)   # library(ismev)
parameters <-gumbel_fit$mle
parameters
b=parameters[1];a=parameters[2]
#####################################################
#           histograma +  as 3 curvas               #
#####################################################
hist(y,xlab="Descargas anuais máximas Rio Meuse",freq=F,
main="",breaks=6,col="lightgrey",xlim=c(min(y),3500),ylim=c(0,0.001))
lines(density(y),col="green",lwd=2) # densidade estimada
curve(dnorm(x,mean=mean(y),sd=sd(y)),add=TRUE, lty=1,lwd=2,col="red")#Normal estimada
curve(dGumbel(x, a=a, b=b), add=TRUE, lty=1,lwd=2,col="cyan") # Gumbel estimada
temp <- legend("topright", legend = c(" "," "," "),
          text.width = strwidth("densidade estimada"),
          lty =c(1,1,1),lwd=c(2,2,2),col=c("green","red","cyan"), xjust=1, yjust=1)
text(temp$rect$left + temp$rect$w, temp$text$y,
    c("densidade estimada","NORMAL estimada","GUMBEL estimada"), pos=2)
> # --  parâmetros EML Gumbel: localização=b; escala=a -- #
> cat("localização=b=",b,"escala=a=",a,"\n")
localização=b= 1243.567 escala=a= 456.454
```

Obtiveram-se as estimativas ML para um ajustamento Gumbel, $\hat{b} = 1243.567$ e $\hat{a} = 456.454$ (*relembre-se que as estimativas preliminares para o ajustamento Gumbel eram* $\hat{b} = 1247.363$ e $\hat{a} = 445.688$).

Apresenta-se na Figura 9.4, as três densidades estimadas e o histograma associado aos dados em estudo.

Em alternativa podemos utilizar outra biblioteca do `R`:

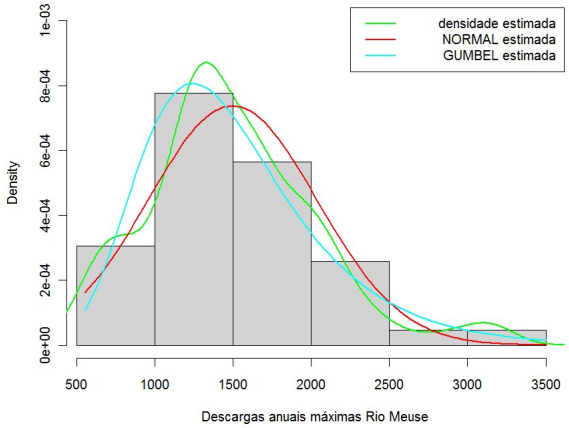


Figura 9.4: Densidades estimadas associadas ao caudais anuais máximos do rio Meuse

library(fitdistrplus)

```
#####################################################
#      GUMBEL - ML    library(fitdistrplus)       #
#####################################################
library(fitdistrplus)
dGumbel <- function(x,a,b) 1/a*exp((b-x)/a)*exp(-exp((b-x)/a))
pGumbel <- function(q,a,b) exp(-exp((b-q)/a))
qGumbel <- function(p,a,b) b-a*log(-log(p))
a_Gumbel=445.6884; b_Gumbel=1247.363 # valores iniciais
fGumb<-fitdist(y,"Gumbel",start=list(a=a_Gumbel,b=b_Gumbel)) #library(fitdistrplus)

> print(fGumb)
Fitting of the distribution ' Gumbel ' by maximum likelihood
Parameters:
   estimate Std. Error
a  456.5983   37.69362
b 1243.9398   52.31971


> par(oma=c(0,0,2,0)) #Add space for main title
> plot(fGumb)
> mtext("RIO MEUSE - Gumbel ??" , side=3, outer=T,cex= 1.,col="red") #Add main title
> gofstat(fGumb,print.test=TRUE) #library(fitdistrplus)
Kolmogorov-Smirnov statistic:  0.09702111
Kolmogorov-Smirnov test:  not rejected
   The result of this test may be too conservative as it
```

```
   assumes that the distribution parameters are known
Cramer-von Mises statistic:  0.09497129
Crame-von Mises test: not calculated
Anderson-Darling statistic:  0.6315043
Anderson-Darling test: not calculated
```

Os resultados são apresentados na Figura 9.5.

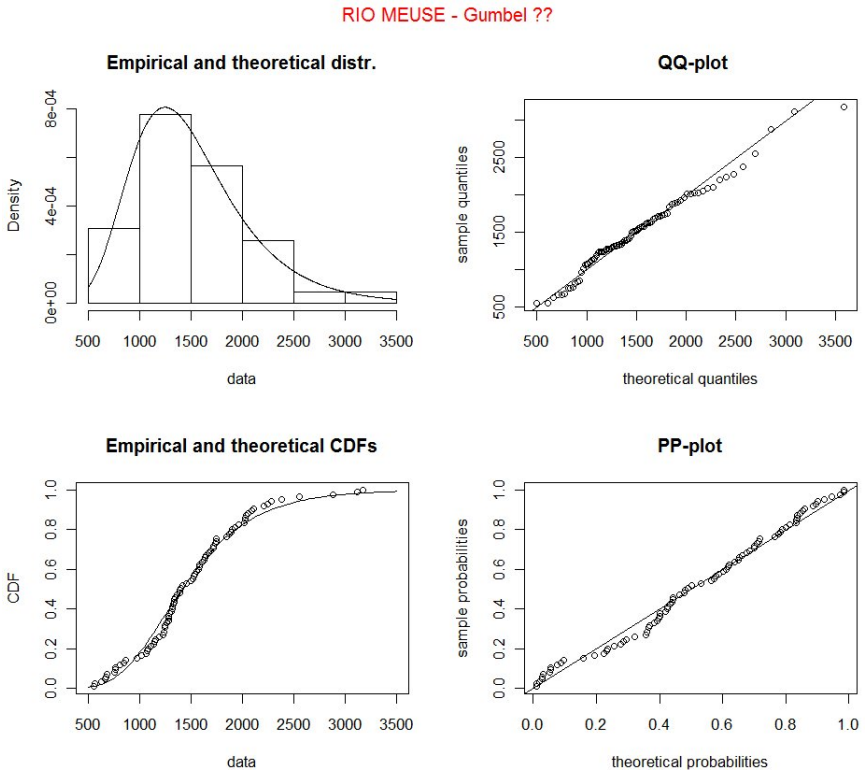


Figura 9.5: Histograma e p.d.f. Gumbel ajustada (*cima, esquerda*), QQplot Gumbel (*cima, direita*), f.d.e. e f.d. Gumbel ajustada (*baixo, esquerda*) e PP-plot Gumbel (*baixo, direita*) para os dados em estudo

A biblioteca `evir` é outra possibilidade e pode ser interpretada como um *update* do `evir`, fornecendo estimativas semelhantes às do `ismev`.

```
library(evir)
```

```
##################################################
#           GUMBEL - ML    library(evir)          #
##################################################
library(evir)
# Fit Gumbel
out <- gumbel(y) # library(evir)
> out$par.ests
   sigma       mu
 456.454 1243.567
```

No que diz respeito aos períodos e níveis de retorno exemplificamos com os valores de saída do `evir`:

```
#-------------------- período de retorno  ------------------------##
# ----período (em média) em que é ultrapassado o nível x ---------##
# ------------   PERIODO DE RETORNO = 1/(P[X>x]) -----------------##
pGumbel<-function(x,a,b){exp(-exp(-(x-b)/a))}
b=   1243.567; a=  456.454  # estimados pelo evir

>   max(y) #  3175
[1] 3175

Periodo_Retorno  <- 1/(1-pGumbel(3175,a=a,b=b))
>   Periodo_Retorno
[1] 69.31374

## ------------------- nível de retorno 100-anos  ----------------------##
## - nível das descargas ultrapassado todos os T=100 anos, em média ----##
## ----------------  NIVEL DE RETORNO = F^(-1)(1-1/T) ------------------##
qGumbel<-function(x,a,b){b-a*log(-log(x))}
nivel<-qGumbel(1-1/100,b = b, a = a)
> nivel
[1] 3343.324
```

outras alternativas possíveis seriam

```
library(evd)
```

```
##############################
library(evd)
fgev(y, shape = 0)    #    library(evd)  fit Gumbel

... ... ...
Estimates
   loc   scale
1240.3   455.2
```

```
library(fExtremes)
```

```
##############################
library(fExtremes)
 gumbelFit(y)    #     library(fExtremes)

... ... ...
Estimated Parameters:
       mu      beta
1243.5833  456.4817
```

Na Tabela 9.1 sumarizamos os resultados obtidos.

library	$\hat{b}$ (localização)	$\hat{a}$ (escala)	nível de retorno a 100-anos
`ismev`	1243.6	456.5	3343.32
`fitdistrplus`	1243.9	456.6	3344.36
`evir`	1243.6	456.5	3343.32
`evd`	1240.3	455.2	3334.29
`fExtremes`	1243.6	456.5	3343.47

Tabela 9.1: **Ajustamento à Gumbel**: estimação dos parâmetros de localização e escala e nível de retorno a 100-anos

Observação 9.1.2. *Este ajustamento à distribuição* Gumbel *justifica-se pela análise preliminar de dados em que apenas foi utilizado o* QQ-plot *e a estimação dos parâmetros pelos mínimos quadrados. Obtivémos então* $\hat{b} = 1247.363 \quad \hat{a} = 445.688 \quad nível = 3297.596$.

O Ajustamento à GEV utilizando os *packages* atrás mencionados e partindo dos valores iniciais estimados pelos mínimos quadrados no QQ-plot

$$\hat{b} = 1251.887; \quad \hat{a} = 461.6413; \quad \hat{\gamma} = -0.03419188$$

conduz ao quadro apresentado na Tabela 9.2.

Observe-se os valores inferiores obtidos para o nível de retorno a 100-anos com o ajustamento GEV, comparativamente aos obtidos com o ajustamento à Gumbel.

library	$\hat{\gamma}$	$\hat{b}$ (localização)	$\hat{a}$ (escala)	nível de retorno a 100-anos
ismev	-0.09243	1267.22750	466.79293	3016.41
fitdistrplus	-0.09220	1266.97383	466.43461	3015.67
evir	-0.09243	1267.22750	466.79293	3016.41
evd	-0.08944	1259.08675	466.12376	3016.97
fExtremes	-0.09232	1266.88569	466.45138	3015.17

Tabela 9.2: **Ajustamento à GEV**: estimação dos parâmetros de forma, localização e escala e nível de retorno 100-anos

```
library(fitdistrplus)
```

```
#####################################################
#         GEV - ML    library(fitdistrplus)       #
#####################################################
##------ b= localização   a= escala -------------##
dGev<-function(x,g,a,b){ exp(-(1+g*(x-b)/a)^(-1/g))* (1+g*(x-b)/a)^(-1/g-1)/a}
pGev<-function(q,g,a,b){exp(-(1+g*(q-b)/a)^(-1/g))}
qGev<-function(p,g,a,b){b+a*((-log(p))^(-g)-1)/g}
b_Gev= 1251.887; a_Gev=461.6413 ;c_Gev=-0.03419188 #papel de probabilidades
fGev <- fitdist(y,"Gev",start=list(a=a_Gev,b=b_Gev,g=c_Gev))#library(fitdistrplus)

> print(fGev)
Fitting of the distribution ' Gev ' by maximum likelihood
Parameters:
      estimate  Std. Error
a  466.4346074 39.43540289
b 1266.9738306 56.21115224
g   -0.0921989  0.06981646

par(oma=c(0,0,2,0))  #Add space for main title
plot(fGev)
mtext("RIO MEUSE - GEV ??", side=3, outer=T,cex= 1.,col="red") #Add main title
```

Os resultados do ajustamento obtido pelo modelo GEV, em (7.1), são apresentados na Figura 9.6.

Procedemos em seguida à obtenção dos valores observados de diversas estatística de ajustamento, as estatísticas de Kolmogorov-Smirnov, de Cramer-von Mises e de Anderson Darling.

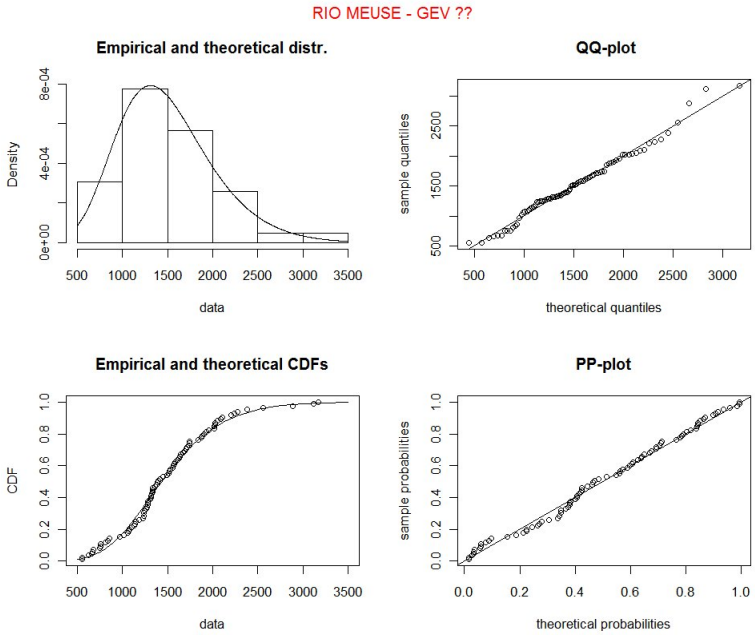


Figura 9.6: Histograma e p.d.f. GEV ajustada (*cima, esquerda*), QQ-plot GEV (*cima, direita*), f.d.e. e f.d. GEV ajustada (*baixo, esquerda*) e PP-plot GEV (*baixo, direita*) para os dados em estudo

```
library(fitdistrplus)
```

```
> gofstat(fGev,print.test=TRUE) #library(fitdistrplus)
Kolmogorov-Smirnov statistic:  0.07925095
Kolmogorov-Smirnov test:  not rejected
   The result of this test may be too conservative as it
   assumes that the distribution parameters are known
Cramer-von Mises statistic:  0.0580412
Crame-von Mises test: not calculated
Anderson-Darling statistic:  0.4138517
Anderson-Darling test: not calculated
```

É de observar que no caso do ajustamento à GEV as estatísticas de ajustamento assumem valores inferiores do que no ajustamento à Gumbel.

Passamos em seguida à utilização do **método PWM**:

```
## -----  PWM para a Gumbel ----- ##
n<-length(y)
M100=mean(y)
y<-sort(y)
yy<-c()
for(i in 1:n) {
    yy[i]=(i-1)/(n-1)*y[i]
}
M110=mean(yy)
a=(2*M110-M100)/log(2)
euler=0.57721 # Const Euler
b=M100-euler*a

> a # escala
[1] 430.8378
> b # localização
[1] 1247.278



## -----  PWM para a GEV ----- ##
n<-length(y);
y<-sort(y); yy<-c();yyy<-c()
for(i in 1:n) {
    yy[i]=(i-1)/(n-1)*y[i]
    yyy[i]=(i-1)*(i-2)/((n-1)*(n-2))*y[i]
}
M100=mean(y); M110=mean(yy); M120=mean(yyy)
h <- function(g) {
 (3*M120-M100)/ (2*M110-M100) - (3^g-1)/(2^g-1)
}
g<-uniroot(h,lower=-0.1,upper=0.1)$root
> g  # forma
[1] -0.09891542
> a=g*(2*M110-M100)/(gamma(1-g)*(2^g-1))
> a  # escala
[1] 468.3563
> b<-M100+a*(1-gamma(1-g))/g
> b  # localização
[1] 1267.686
```

Alternativamente, podemos usar

```
library(fExtremes)

out <- gevFit(y,type="pwm")  #library(fExtremes)
> out        # show(out) \'e igual a   out
Title:
 GEV Parameter Estimation
```

```
Call:
 gevFit(x = y, type = "pwm")
Estimation Type:
  gev pwm
Estimated Parameters:
           xi            mu          beta
  -0.09895205 1267.69438205  468.36950596
```

No caso em que o parâmetro de forma é negativo, i.e., $\gamma < 0$, o *limite superior do suporte da* GEV *é finito* e pode ser estimado por

$$\hat{x}^F = \hat{\lambda} - \hat{\delta}/\hat{\gamma}\,.$$

Neste caso de estudo, podemos adiantar que o *máximo anual* Y terá um limite superior do suporte a rondar o valor 6000.

```
################## endpoint ###################
a_mv = 466.79293305 ;  b_mv = 1267.22749758 ; g_mv = -0.09242819
a_pwm = 468.36950596; b_pwm = 1267.69438205 ; g_pwm =-0.09895205
xF_pwm = b_pwm-a_pwm/g_pwm
xF_mv = b_mv-a_mv/g_mv


> xF_mv;
[1] 6317.558
> xF_pwm
[1] 6000.992
```

Suponhamos que havia interesse em estimar o nível de retorno a $T = 100$ meses. Podemos identificar o *máximo anual* Y como o máximo de $n = 12$ máximos mensais, *supostamente i.i.d.* a X (*máximo mensal*), i.e., $Y = X_{n:n} = \max(X_1, \ldots, X_{12})$. Para blocos de tamanho $n = 12$, os quantis-$(1-p) = 1 - 1/100$ ajustados para X, são estimados por

$$\hat{q}^*_{X,p} = G^{\leftarrow}_{\hat{\gamma}}\Big((1-1/100)^{12}; \hat{\lambda}, \hat{\delta}\Big),$$

pelo que

```
## ------- nível de retorno 100-meses  - (GEV) --------------------------##
## -- nível das descargas ultrapassado todos os T=100 meses, em média ----##
b_Gev=  1267.22749758; a_Gev=466.79293305 ;c_Gev=-0.09242819 # ismev
qGev<-function(x,g,a,b){b+a*((-log(x))^(-g)-1)/g}
nivel <- qGev((1-1/100)^12,g= c_Gev, b = b_Gev, a = a_Gev )

> nivel
[1] 2164.082
```

Vamos em seguida ver como obter **intervalos de confiança (IC).** O IC para γ baseado no *profile-likelihood* com grau de confiança $100(1-\alpha)\%$ *aproximadamente* pode ser obtido do modo seguinte:

```
library(ismev)
```

```
##############################
library(ismev)
> ppfit <-gev.fit(y)        # library(ismev)
$conv
[1] 0

$nllh
[1] 651.7391

$mle
[1] 1267.22749758  466.79293305   -0.09242819

$se
[1] 56.23057065 39.49955175  0.06987208
```

Conclui-se daqui que a estimativa ML para γ é $\hat{\gamma} = -0.09242819$, para a qual a log-verosimilhança é -651.7391.

O IC para γ baseado no *profile-likelihood* com grau de confiança 95% *aproximadamente* pode ser obtido do modo a seguir apresentado, e é ilustrado na Figura 9.7.

```
library(ismev)
```

```
##############################
gev.profxi(ppfit, conf = 0.95,xlow=-0.23, xup=0.1)
```

As 2 rectas horizontais no gráfico da Figura 9.7, correspondem a uma ordenada igual ao máximo da log-verosimilhança, e igual ao máximo da log-verosimilhança subtraído de $\chi^2_1(0.95)/2$ – ver Figura 9.8 (*esquerda*).

```
abline(h=-651.7391,col="red")
q=qchisq(.95, df=1)   # 1 degrees of freedom
abline(h=-651.7391-0.5*q,col="green")
```

As 2 rectas verticais da Figura 9.8 (*direita*) correspondem aos limites do IC $[-0.212, 0.066]$.

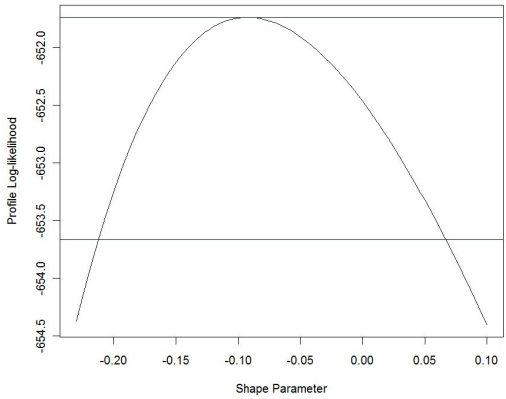


Figura 9.7: ismev: *Profile-loglikelihood* em função de γ

```
abline(v=-0.09242819,col="grey")
abline(v=-0.212,col="blue")
abline(v=0.066,col="blue")
```

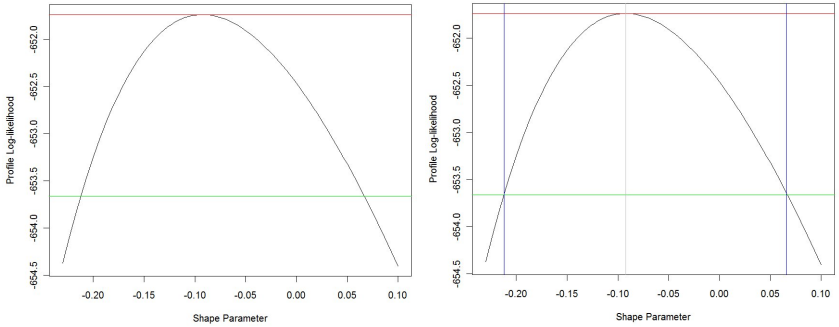


Figura 9.8: ismev: *Profile-likelihood* em função de γ: confirmação das rectas horizontais (*esquerda*); confirmação das rectas verticais (*direita*).

Os IC's para γ, λ, δ baseados no *profile-likelihood* com grau de confiança 95% *aproximadamente* podem ser obtidos também a partir do seguinte `script`:

```
library(evd)
```

```
library(evd)
> library(evd)
> confint(profile(fgev(y),conf = 0.95)) #library(evd)
[1] "profiling loc"
[1] "profiling scale"
[1] "profiling shape"
             lower        upper
loc   1157.4339835 1.379016e+03
scale  398.0361112 5.551471e+02
shape   -0.2120359 6.683597e-02
```

O seguinte `script` dá origem à Figura 9.9.

```
fGev<-fgev(y)
shapgev<-as.vector(fGev$param[3])  # ML shape
locgev<-as.vector(fGev$param[1])   # ML loc
scalgev<-as.vector(fGev$param[2])  # ML scale
## ----  IC 95% profile-likelihood ---- ##
gCI<-confint(profile(fGev,which="shape"))
ginf<-gCI[1]; gsup<-gCI[2]         # ML IC-shape
lCI<-confint(profile(fGev,which="loc"))
linf<-lCI[1]; lsup<-lCI[2]         # ML IC-loc
sCI<-confint(profile(fGev,which="scale"))
sinf<-sCI[1]; ssup<-sCI[2]         # ML IC-scale
# ---  profile log-likelihood plot ---- ##
par(mfrow=c(2,2))
## ----  shape ----- ##
plot(profile(fGev,which="shape",conf = 0.95))
abline(v=ginf,col="blue"); abline(v=gsup,col="blue"); abline(v=shapgev,col="red")
text( ginf,-653.5,paste(round(ginf,3)) ); text( gsup,-653.5,paste(round(gsup,3)) )
text(shapgev,-653.5,paste(round(shapgev,3)) )
## -----  loc  ----- ##
plot(profile(fGev,which="loc",conf = 0.95))
abline(v=linf,col="blue"); abline(v=lsup,col="blue"); abline(v=locgev,col="red")
text( linf,-654,paste(round(linf,3)) ); text( lsup,-654,paste(round(lsup,3)) )
text(locgev,-654,paste(round(locgev,3)) )
## ---  scale -----##
plot(profile(fGev,which="scale",conf = 0.95))
abline(v=sinf,col="blue"); abline(v=ssup,col="blue"); abline(v=scalgev,col="red")
text( sinf,-653.5,paste(round(sinf,3)) ); text( ssup,-653.5,paste(round(ssup,3)) )
text(scalgev,-653.5,paste(round(scalgev,3)) )
```

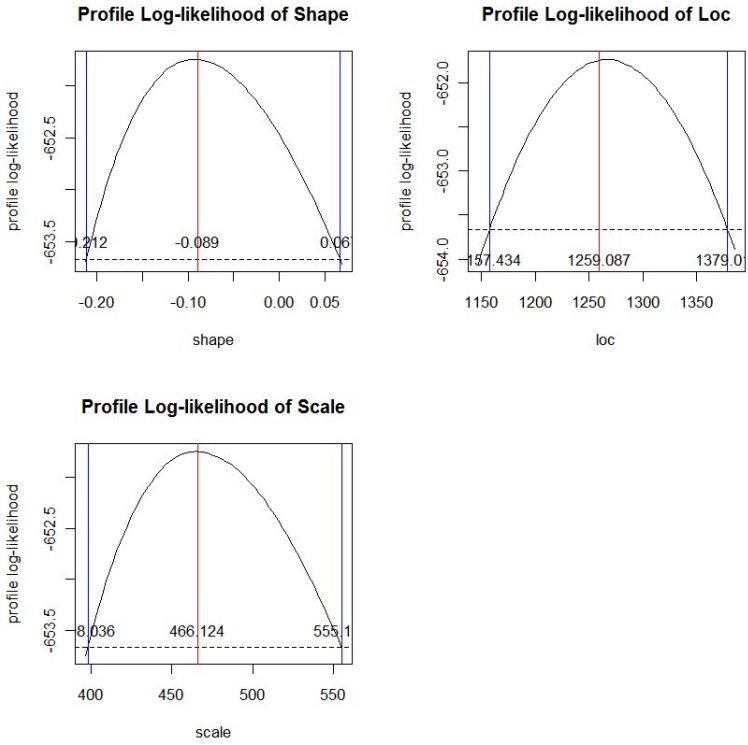


Figura 9.9: evd: *Profile log-likelihood* em função de γ (forma), λ (localização) e δ (escala)

O IC para o nível de retorno 100-anos, baseado no *profile-likelihood* com grau de confiança 95% *aproximadamente* pode ser obtido por exemplo a partir de (ver Figura 9.10):

```
library(ismev)
```

```
ppfit <-gev.fit(y)                    # library(ismev)
gev.prof(ppfit, m = 100, 2600, 4000) # m=100 é T=100
abline(v=2685,col="blue")
abline(v=3016) # a EML é  3016.41
abline(v=3785,col="blue")
```

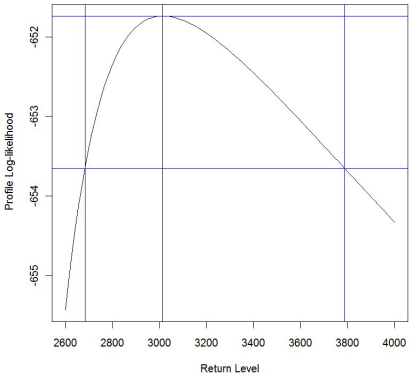


Figura 9.10: ismev : *Profile log-likelihood* para nível de retorno 100-anos

Alternativamente, o IC para o nível de retorno 100-anos também pode ser obtido a partir de (ver Figura 9.11):

```
library(evir)
```

```
library(evir)
out <- gev(y)    # Fit GEV  library(evir)
rlevel.gev(out, k.blocks = 100,add=F)
rl<-rlevel.gev(out, 100,add=T)
abline(v=rl[1],col="blue")
abline(v=rl[3],col="blue")
cat("A estimativa ML do nível de Retorno de 100-anos é =", rl[2],
"com Intervalo de Confiança a 95% =(", rl[1],",",rl[3],")","\n")

A estimativa ML do nível de Retorno de 100-anos é = 3016.41
com Intervalo de Confiança a 95% =( 2686.727 , 3778.092 )
```

Os IC's para o nível de retorno 100-anos, baseados no *profile-likelihood* com graus de confiança $90\%, 95\%, 99\%$ *aproximadamente* podem ser obtidos por exemplo a partir de (ver Figura 9.12):

```
library(fExtremes)
```

```
library(fExtremes)
out <- gevFit(y)  # out <- gevFit(y,type="mle") #library(fExtremes)
gevrlevelPlot(out,kBlocks=100) # 100-Years return level & 90% CI
```

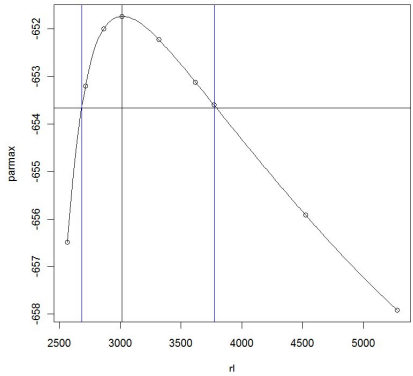


Figura 9.11: evir : *Profile log-likelihood* para nível de retorno 100-anos

```
                      min        v     max   kBlocks
GEV Return Level 2726.536 3015.175 3599.27       100

abline(v= 2726.536,col="blue") # IC 90% inferior
abline(v=3599.27,col="blue")   # IC 90% superior
abline(v=3015.175)             # EML para nível de retorno 100-anos
```

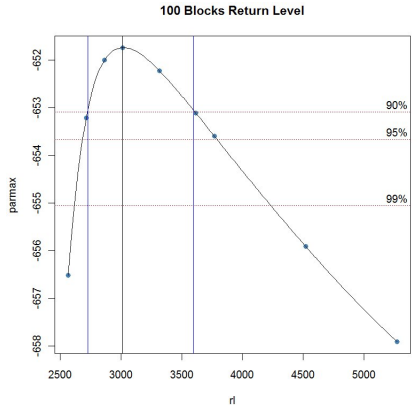


Figura 9.12: fExtremes : *Profile log-likelihood* para nível de retorno 100-anos

9.2 Caso de Estudo: 'venice, library(ismev)'

Iremos aqui considerar os dados `venice` do package `ismev` do `R`. Trata-se de uma base de dados relativos às alturas, em *cm*, do nível do mar em Veneza entre 1931 e 1981. A *data frame* `venice` tem 51 linhas e 11 colunas, sendo as últimas 10 relativas às maiores observações anuais (1931-1981) e a coluna relativa aos anos respectivos (*excepção de* 1935 *só com os* 6 *maiores valores*).

library(ismev)

```
library(ismev); data(venice)

> head(venice)
  Year  r1  r2  r3  r4  r5 r6 r7 r8 r9 r10
1 1931 103  99  98  96  94 89 86 85 84  79
2 1932  78  78  74  73  73 72 71 70 70  69
3 1933 121 113 106 105 102 89 89 88 86  85
4 1934 116 113  91  91  91 89 88 88 86  81
5 1935 115 107 105 101  93 91 NA NA NA  NA
6 1936 147 106  93  90  87 87 87 84 82  81
> head(venice[,-1])
   r1  r2  r3  r4  r5 r6 r7 r8 r9 r10
1 103  99  98  96  94 89 86 85 84  79
2  78  78  74  73  73 72 71 70 70  69
3 121 113 106 105 102 89 89 88 86  85
4 116 113  91  91  91 89 88 88 86  81
5 115 107 105 101  93 91 NA NA NA  NA
6 147 106  93  90  87 87 87 84 82  81
```

Estes dados são usados em Coles (2001), e foram analisados originalmente em Smith[1] (1986). Na Figura 9.13, representamos essas observações.

Uma abordagem paramétrica de modelação possível consiste em considerar o ajustamento ao modelo limite das k (com k fixo) maiores e.o.'s referido anteriormente, na Secção 7.3.1, e para o qual temos m blocos – *Modelo GEV multidimensional*.

Neste caso de estudo os blocos correspondem aos $m = 51$ anos, com $r = 10$ maiores observações disponíveis (com excepção do ano 1935 em que $r = 6$) e $1 \leq k \leq 10$.

[1] Smith, R.L. (1986). Extreme value theory based on the r largest annual events. *J. Hydrology* **86**, 27–43.

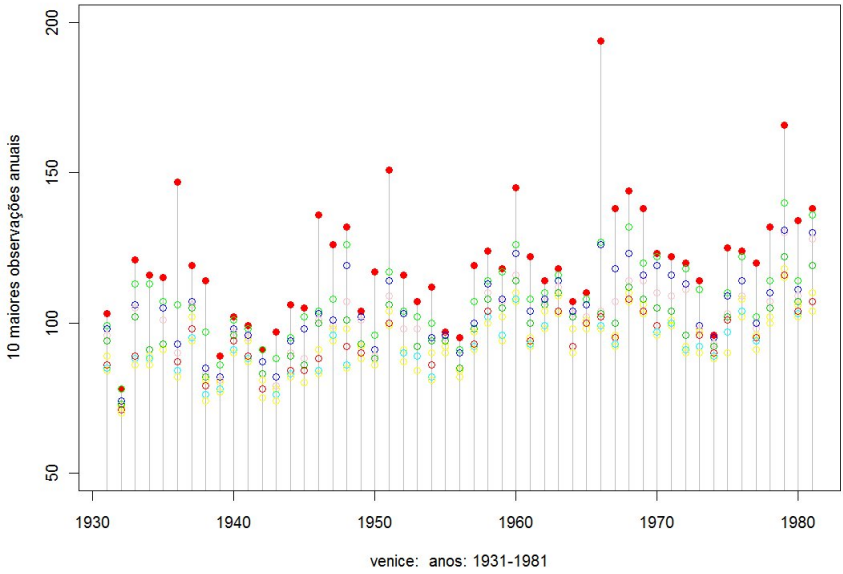


Figura 9.13: venice: 10 maiores níveis do mar em Veneza (1931-1981)

Usaremos o `ismev` para proceder à estimação ML dos parâmetros do modelo.

A maximização de log-verosimilhança, $\log L$, e as estimativas dos parâmetros da GEV (*erros padrão entre parêntesis*) ajustada ao máximo anual, baseadas no modelo limite das k maiores observações, ajustados aos dados do nível do mar em Veneza, para diferentes valores de k (1,5,10), (ver Tabela 9.3), podem ser obtidos através da instrução

```
rlarg.fit
```

```
## ----  r-largest observations (r=1 (MMA), r=5, r=10) EML ---- ##
rlarg.fit(venice[,-1],r=1)
rlarg.fit(venice[,-1],r=5)
rlarg.fit(venice[,-1],r=10)
```

k	$\log L$	$\hat{\lambda}$	$\hat{\delta}$	$\hat{\gamma}$
1	-222.7	111.1 (2.6)	17.2 (1.8)	-0.077 (0.074)
5	-732.0	118.6 (1.6)	13.7 (0.8)	-0.088 (0.033)
10	-1139.1	120.5 (1.4)	12.8 (0.5)	-0.113 (0.020)

Tabela 9.3: venice: estimativas dos parâmetros da GEV (*erros padrão entre parêntesis*) ajustada ao máximo anual, baseadas no modelo limite das k maiores observações

Observação 9.2.1. *Para valores crescentes de k observa-se decréscimo dos erros padrão. Contudo, existe falta de estabilidade para as estimativas, para diferentes k. A validade do modelo é pois questionável* (assunto a explorar seguidamente ...)

Investiguemos agora o ajustamento para cada uma das k maiores observações.

Exemplificando com o modelo para as 5 *maiores observações*, o ajustamento para o máximo anual dos dados do nível do mar de Veneza pode ser investigado através do plot de diagnóstico obtido através de instrução

`rlarg.diag`

```
venfit <- rlarg.fit(venice[,-1],r=5)
rlarg.diag(venfit)
```

O resultado corresponde aos gráficos apresentados na Figura 9.14.

Considerando o modelo para as 5 *maiores observações*, o ajustamento da marginal para as maiores $k = 1, 2, 3, 4, 5$ observações do mar de Veneza pode também ser investigado através dos *PP-Plots* e dos *QQ-Plots*, pelas sucessivas saídas do diagnóstico obtido utilizando a instrução `rlarg.diag`.

Estes plots parecem indicar um *fraco ajustamento do modelo*, tal como de pode ver na Figura 9.15.

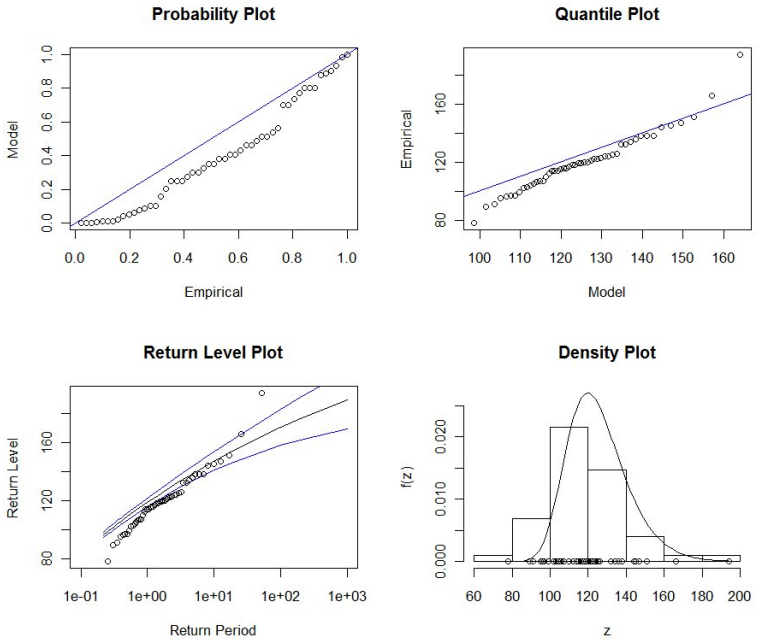


Figura 9.14: venice: ajustamento para o máximo anual, baseado nas 5 *maiores observações*.

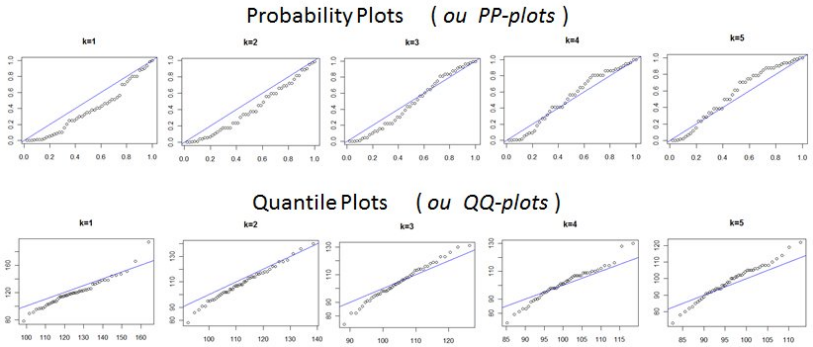


Figura 9.15: venice: ajustamento da marginal do modelo para as maiores $k = 1, 2, 3, 4, 5$ baseado nas 5 *maiores observações*.

Consideremos agora o *método dos máximos anuais*, aqui como caso particular do modelo para as r *maiores observações*, com $r = 1$. O ajustamento para o máximo anual do mar de Veneza pode ser investigado através de diagnóstico análogo ao anterior, obtendo-se a Figura 9.16.

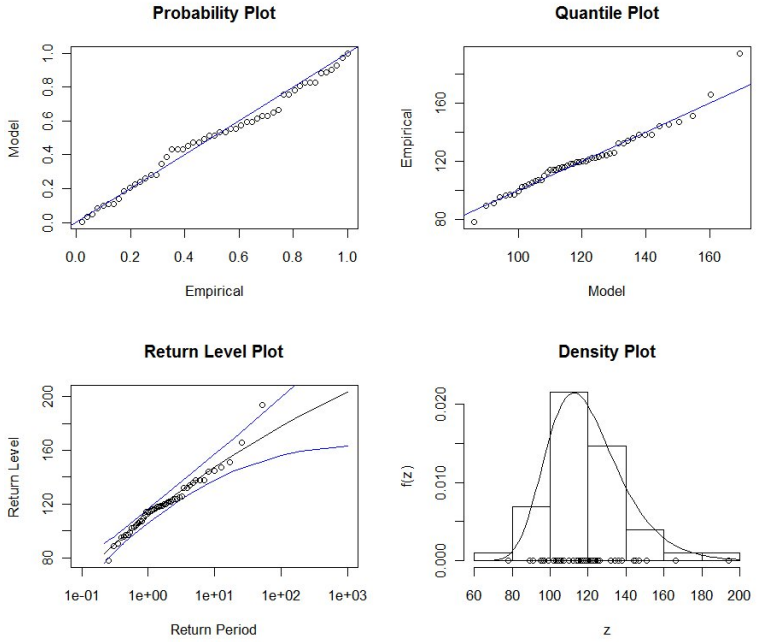


Figura 9.16: venice: ajustamento para o máximo anual do mar de Veneza quando se considera apenas a maior observação em cada um dos 51 anos.

Parece existir um *melhor ajustamento do modelo*, em alternativa a tomar mais observações por cada ano, além do máximo anual.

Note-se no entanto que tal é devido a problemas de não estacionariedade dos dados, tal como detectado em Smith (1986). O *plot* dos dados apresentado inicialmente na Figura 9.13, deixa transparecer visualmente a presença de uma tendência (*trend*) nesses dados. Essa tendência pode ser estabelecida quer apenas em termos dos 51 *máximos anuais*, quer em relação às 10 *maiores observações* tomadas em cada um dos 51 anos.

Ajustamento Linear à nuvem de máximos anuais:

```
library(ismev)
data(venice)
attach(venice)
y <-r1
x <- Year-1930
lm(y~x)
intercept=lm(formula = y ~ x)$coefficients[1] # intercept=104.8666667
slope=lm(formula = y ~ x)$coefficients[2]     # slope=0.5669683
plot(x+1930,y,ylab="Sea level (cm)",xlab="Year")
curve(intercept+slope*(x-1930),col="red",lwd=2,add=T)
```

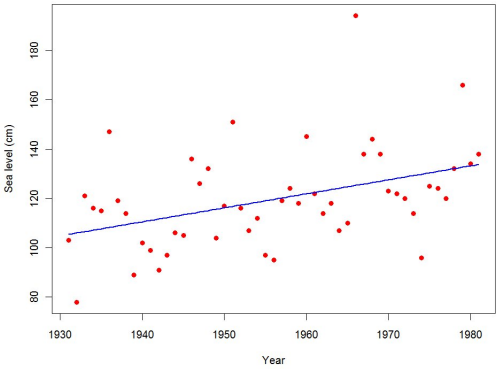


Figura 9.17: Ajustamento Linear à núvem de máximos anuais

Procedamos à estimação ML da tendência linear para λ,

$$\lambda = \lambda(t) = \beta_0 + \beta_1 t, \qquad t = 1, 2, \cdots, 51$$

no modelo ajustado às 10 *maiores observações anuais*.

```
####################################################################
## -----------  estimação de beta0 e beta1 (r=10) ----------- ##
####################################################################
fit <- rlarg.fit(venice[,-1],ydat=venice-1930,mul=c(1),r=10)
beta0 <- fit$mle[1]
beta1<- fit$mle[2]
data.vec<-as.vector(t(as.matrix(venice[,2:11])))
time.vec<-as.vector(t(matrix(rep(1931:1981,10),ncol=10)))
plot(time.vec,data.vec,xlab="Year",ylab="Sea level(cm)")
#  mle(r=10)  beta0=intercept=104.51265767   beta1=slope=0.48173120
curve(beta0+beta1*(x-1930),col="red",lwd=2,add=T)
```

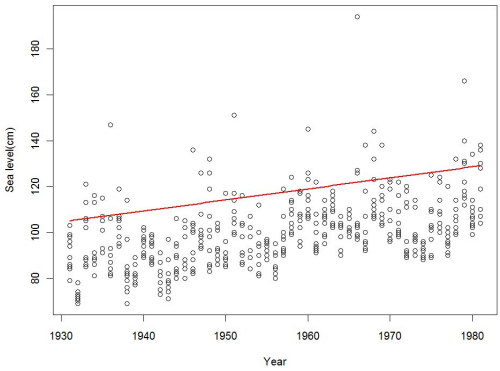


Figura 9.18: Recta de tendência sobreposta à nuvem global de pontos

A maximização de log-verosimilhança, $\log L$, conduz-os às estimativas ML para os parâmetros (*erros padrão entre parêntesis*) para o modelo limite das k maiores observações, ajustados aos dados do nível do mar em Veneza, para diferentes valores de k (1,5,10), pela instrução `rlarg.fit`

```
rlarg.fit(venice[,-1],ydat=venice-1930,mul=c(1),r=1)
rlarg.fit(venice[,-1],ydat=venice-1930,mul=c(1),r=5)
rlarg.fit(venice[,-1],ydat=venice-1930,mul=c(1),r=10)
```

k	$\log L$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}$	$\hat{\gamma}$
1	-216.1	97.0 (4.2)	0.56 (0.14	14.6 (1.6)	-0.027 (0.083)
5	-704.8	104.2 (2.0)	0.46 (0.06)	12.3 (0.8)	-0.037 (0.042)
10	-1084.1	104.5 (1.7)	0.48 (0.04)	11.7 (0.6)	-0.065 (0.027)

Tabela 9.4: venice: estimação ML de β_0, β_1, δ e γ no modelo ajustado às 10 *maiores observações anuais*, com $\lambda = \lambda(t) = \beta_0 + \beta_1 t$.

Observação 9.2.2. *Comparação das Abordagens: Estacionariedade vs Não-Estacionariedade, com* $\lambda = \lambda(t) = \beta_0 + \beta_1 t$*:*

1. *Comparando as log-verosimilhanças nas* 2 *tabelas* 9.3 *e* 9.4 *verifica-se para cada* k *que a introdução no modelo de uma* tendência linear *para*

o parâmetro de localização $\lambda(t) = \beta_0 + \beta_1 t$*, com* $\beta_1 \simeq 0.5$*, é sem dúvida decisiva para um melhor ajuste.*

2. *Existe uma maior estabilidade nas estimativas dos parâmetros, através dos diferentes* k*, na abordagem não-estacionária* (Tabela 9.4) *comparativamente aos correspondentes valores da tabela da abordagem estacionária* (Tabela 9.3)*.*
3. *Por outro lado, os* PP-Plots *e os* QQ-Plots *exibem um melhor ajustamento para as* k *maiores observações relativamente ao modelo limite para cada uma das* k *maior e.o.'s.*

Estimação de Parâmetros de Acontecimentos Raros

Face ao ajustamento da GEV ao máximo anual do nível do mar em Veneza, quando são consideradas 1 (MMA), 5 ou 10 maiores observações anuais, iremos estimar alguns parâmetros de acontecimentos raros.

Figura 9.19: A maior cheia em Veneza verificou-se até hoje (desde que há registo) no dia 4 de Novembro de 1966 e correspondeu a uma altura de 194 cm – na figura está assinalada a cheia de 1966, no exterior de uma loja em Veneza.

Para um determinado ano, serão alvo de estimação:

- níveis ultrapassados com uma probabilidade p pequena, ou o problema traduzido em níveis de retorno de T-anos;
- probabilidades de excedência de níveis elevados, ou equivalentemente períodos de retorno T para níveis elevados u.

As `R-functions` abaixo respondem às questões referidas, e serão usadas para obter estimativas para os anos 1990, 2000 e 2013. O ano especial de 1966 integrará igualmente a análise pela observação excepcional de 194 cm, a que corresponde uma área de Veneza submergida de 82.39%.

```
### ----  GEV(g,a,b)  b= localização   a= escala  g=forma ---- ###
dGev<-function(x,g,a,b){ exp(-(1+g*(x-b)/a)^(-1/g))* (1+g*(x-b)/a)^(-1/g-1)/a}
pGev<-function(q,g,a,b){exp(-(1+g*(q-b)/a)^(-1/g))}
qGev<-function(p,g,a,b){b+a*((-log(p))^(-g)-1)/g}
### --  nivel ultrapassado com probabilidaden p no ano year -- ###
level <- function(beta0,beta1,gama,delta,p,year)
{
t1 = 1931;  t = year-t1+1;   lambda <- beta0 + beta1*t;
nivel<-qGev(1-p,gama,delta,lambda)
return(nivel)
}
### ---------- nivel de retorno T-year no ano 'year' ---------- ###
level_Tyear <- function(beta0,beta1,gama,delta,T,year)
{
t1 = 1931;  t = year-t1+1;   lambda <- beta0 + beta1*t;
nivel<-qGev(1-1/T,gama,delta,lambda)
return(nivel)
}
### ---  probabilidade de excedência do 'nivel' no ano 'year' --- ###
prob <- function(beta0,beta1,gama,delta,nivel,year)
{
t1 = 1931;  t = year-t1+1;   lambda <- beta0 + beta1*t;
prob <- 1-pGev(nivel,gama,delta,lambda)
return(prob)
}
### ------  período de retorno do 'nivel' no ano 'year'   ------- ###
retorno <- function(beta0,beta1,gama,delta,nivel,year)
{
t1 = 1931;  t = year-t1+1;   lambda <- beta0 + beta1*t;
periodo_retorno <- 1/(1-pGev(nivel,gama,delta,lambda))
return(periodo_retorno)
}
```

Considerando o modelo GEV com tendência linear, ajustado ao nível máximo anual do mar em Veneza, baseado apenas nos 51 máximos anuais, obtêm-se os quantis estimados ($p = 0.5, 0.1, 0.01, 0.001$):

```
> gama= -0.027; delta=14.6; beta0= 97.0 ; beta1= 0.56  # EML  (r=1)
> p=c(0.5,0.1,0.01,0.001)
> ### ------ máximo=194cm observado no ano 1966  (1931-1981) ------ ###
> year=1966; level(beta0,beta1,gama,delta,p,year)
[1] 122.4847 149.0371 180.3187 209.1606
> year=1990; level(beta0,beta1,gama,delta,p,year)
[1] 135.9247 162.4771 193.7587 222.6006
> year=2000; level(beta0,beta1,gama,delta,p,year)
[1] 141.5247 168.0771 199.3587 228.2006
> year=2013; level(beta0,beta1,gama,delta,p,year)
[1] 148.8047 175.3571 206.6387 235.4806
```

Na tabela 9.5 estão sumarizados os resultados quando se consideram 1,5, ou

10 maiores observações anuais.

	$ano \setminus p$	0.5	0.1	0.01	0.001
(MMA)	1966	122.4847	149.0371	180.3187	209.1606
$k = 1$	1990	135.9247	162.4771	193.7587	222.6006
	2000	141.5247	168.0771	199.3587	228.2006
	2013	148.8047	175.3571	206.6387	235.4806
	1966	125.2377	147.3185	172.7885	195.7314
$k = 5$	1990	136.2777	158.3585	183.8285	206.7714
	2000	140.8777	162.9585	188.4285	211.3714
	2013	146.8577	168.9385	194.4085	217.3514
	1966	126.0175	146.2742	168.3006	186.8888
k=10	1990	137.5375	157.7942	179.8206	198.4088
	2000	142.3375	162.5942	184.6206	203.2088
	2013	148.5775	168.8342	190.8606	209.4488

Tabela 9.5: venice: estimação de niveis com probabilidade p de serem excedidos.

Foquemo-nos agora no nível de 194 cm e analisemos a probabilidade de excedência deste nível e o período de retorno, para os anos de 1966 e 2013, quando se considera a abordagem MMA com tendência:

```
> gama= -0.027; delta=14.6; beta0= 97.0 ; beta1= 0.56  # EML  (r=1)
> ### ------------- P[máximo>194cm] no ano 1966  ----------------- ###
> nivel= 194; year=1966; prob(beta0,beta1,gama,delta,nivel,year)
[1] 0.003419182
> ### ---------------- P[máximo>194cm] no ano  2013 -------------- ###
> nivel= 194; year=2013; prob(beta0,beta1,gama,delta,nivel,year)
[1] 0.02609689
> ### ---------- Período de Retorno do nível=194cm em 1966 ------- ###
> nivel= 194; year=1966; retorno(beta0,beta1,gama,delta,nivel,year)
[1] 292.4676
> ### ---------- Período de Retorno do nível=194cm em 2013 ------- ###
> nivel= 194; year=2013; retorno(beta0,beta1,gama,delta,nivel,year)
[1] 38.31875
```

A probabilidade de excedência de 194 cm e o período de retorno, para o ano de 2013, quando se consideram as maiores 5 observações anuais para o modelo com tendência, resulta do `output`

```
> ### ---------- Período de Retorno do nível=194cm em 2013 ------- ###
> nivel= 194; year=2013; retorno(beta0,beta1,gama,delta,nivel,year)
[1] 96.16106
> ### ---------------- P[máximo>194cm] no ano  2013 -------------- ###
> nivel= 194; year=2013; prob(beta0,beta1,gama,delta,nivel,year)
[1] 0.01039922
```

É de salientar, adoptando a o modelo MMA com tendência, que o período de retorno do nível $194\,cm$ para o ano de 1966 andaria à volta dos $T = 292\,anos$, enquanto que para o ano de 2013 o correspondente período de retorno é de apenas $T = 38\,anos$. Já se adoptarmos um modelo multidimensional para as 5 maiores observações anuais nos 51 anos da base de dados, o período de retorno do nível $194\,cm$ para o ano 2013 é de $T = 96$ anos.

Observação 9.2.3. *Embora esta base de dados* `venice` *diga respeito ao período dos anos* 1931–1981*, em termos históricos os dados foram inicialmente coligidos por* Pirazzoli[2] (1982)*, tendo recolhido os dez maiores valores do nível do mar em* Veneza (*com poucas excepções*) *para o período dos anos* 1887–1981*. Estes dados têm sido alvo de estudo segundo diversas abordagens, entre as quais salientamos* Smith (1986) *e* Ferreira[3] (1997).

Inspirado nessas abordagens, e atendendo à trajectória do máximo anual do nível do mar em Veneza *revelar a existência de uma tendência crescente nos dados, razoavelmente bem modelada de forma linear, poderíamos considerar num modelo* GEV *multidimensional, por exemplo, com localização e escala do tipo*:

$$\lambda_t = \alpha + \beta t/m; \qquad \delta_t = \delta, \quad 1 \leq t \leq m \qquad (m = 51).$$

A estimação dos três parâmetros α, β e δ, por exemplo por máxima verosimilhança, foi estabelecida por Smith (1986)*. Pode também introduzir-se uma tendência mais complicada, tal como foi feito no artigo atrás referido, por exemplo uma tendência linear com componente sinusoidal de período P, im-*

[2]Pirazzoli, P. (1982) Maree estreme a Venezia (periodo 1872-1981). *Acqua Aria* **10**, 1023–1039.

[3]Ferreira, A. (1997) Extreme Sea Level in Venice. In Reiss, R.-D. & Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Verlag, Basel, Part V—Case Studies in Extreme Value Analysis, Study 1, 233–240.

pondo:

$$\lambda_t = \alpha + \beta t/m + \eta \cos\left(\frac{2\pi t}{P} + \Phi\right); \qquad \delta_t = \delta, \quad 1 \le t \le m.$$

Um gráfico em papel de probabilidade Gumbel *sugere uma possível adequabilidade desse modelo. Apresentamos em seguida as estimativas obtidas para diferentes valores de* k (*com tendência linear*)*:*

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
$k = 1$	96.8(4.1)	28.7(7.0)	14.5(1.5)
$k = 5$	104.2(2.0)	23.4(2.7)	12.8(0.7)
$k = 10$	105.2(2.0)	24.6(2.0)	13.1(0.5)

Admitamos que queremos responder à questão: Para um dado ano t_0 *e para um dado* p*, pequeno, qual o nível* ν *que é excedido com probabilidade* p *nesse ano? Em termos deste modelo, temos*

$$\nu = \alpha + \beta t_0/m - \delta \log(-\log(1-p)),$$

que pode ser estimado substituindo os parâmetros desconhecidos pelas suas estimativas de máxima verosimilhança.

Utilizando $k = 5$*, e para* $t_0 = 60$ (*ano* 1990) *e* $t_0 = 70$ (2000)*, temos as estimativas*

$$\hat{\nu}(0.5, 60) = 136.4(2.5) \quad \hat{\nu}(0.1, 60) = 160.5(3.4)$$
$$\hat{\nu}(0.5, 70) = 141.0(2.9) \quad \hat{\nu}(0.1, 70) = 165.0(3.7)$$

Comparem-se estes valores com os correspondentes na Tabela 9.5*.*

Podíamos também ter admitido uma tendência quadrática

$$\lambda_t = \alpha + \beta t/m + \epsilon(t/m)^2; \qquad \delta_t = \delta, \quad 1 \le t \le m.$$

A comparação dos diferentes modelos podia ser feita, por exemplo, em termos de função de log-verosimilhança (*que nos interessa maximizar*)*. Neste caso:*

Tendência	*Linear*	*Quadr.*	$P = 18.62$	$P = 11$
Log-Ver.	-705.1	-705.0	-702.3	-697.5

$P = 18.62$ *é o valor sugerido por* Pirazzoli (1982)*, o ciclo de marés astronómico. O melhor modelo parece pois ser o último referido.*

9.3 Um novo caso de estudo: 'soa.txt'

Trata-se de um conjunto de dados de grandes indemnizações, em que se têm $N_u = 75789$ excedências acima do nível $u = 25000$ USD (Beirlant *et al.*, 2004; `http://lstat.kuleuven.be/Wiley/`).

Numa análise preliminar de dados, começámos por obter o histograma e a caixa-com-bigodes associados a este conjunto de dados, apresentados na Figura 9.20. O histograma e e a caixa-om-bigodes das log-indemnizações evidenciam uma acentuada assimetria à direita.

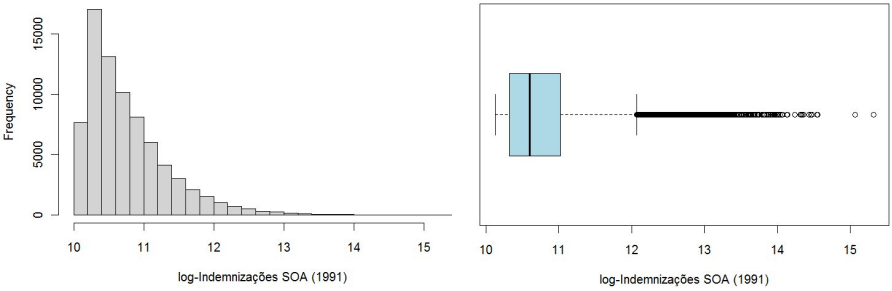


Figura 9.20: soa.txt: Histograma (*esquerda*) e caixa-com-bigodes (*direita*) das log-indemnizações.

Por outro lado, o comportamento crescente do ME-plot, apresentado na Figura 9.21, indica uma distribuição do montante de indemnização com cauda mais pesada do que a Exponencial. Além disso, a forma convexa do QQ-plot Exponencial associado aos excessos das indemnizações, $y_i = x_i - u$, para $u = 25000$ USD, também representado na Figura 9.21, evidencia uma distribuição do montante de indemnização com cauda mais pesada do que a Exponencial.

Consideremos um novo nível elevado, $u = 400000$ USD, a que está associado um **número de excedências** $N_u = 397$, tal como ilustrado na Figura 9.22, obtida com base no `script`:

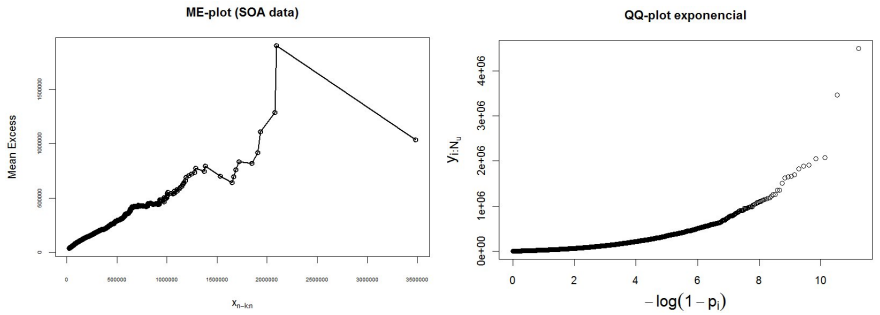


Figura 9.21: soa.txt: ME-plot (*esquerda*) e QQ-plot exponencial (*direita*) associado aos excessos das indemnizações dos dados.

```
soa<-read.table('soa.txt', header=FALSE,dec=".")
head(soa)
x<-soaV1; indice1<-c(1:75789)
u=400000; indice2 <- indice1[x>u] # n-k=75393, 1 inteiro tal que x[n-k]>400000
x2<-x[x>u]
plot(indice1,x,type="h",lwd=1,lty=1,xlab="index",ylab="Indemnizações")
abline(h=u,col="red")
lines(indice2,x2,type="p",col="red",pch=20)
text(100,400000, "u",col="red"); text(-2000,400000, "u",col="red")
text(50000,3000000, "Exced\^encias de u (POT: Peaks Over Threshold)",col="red")
```

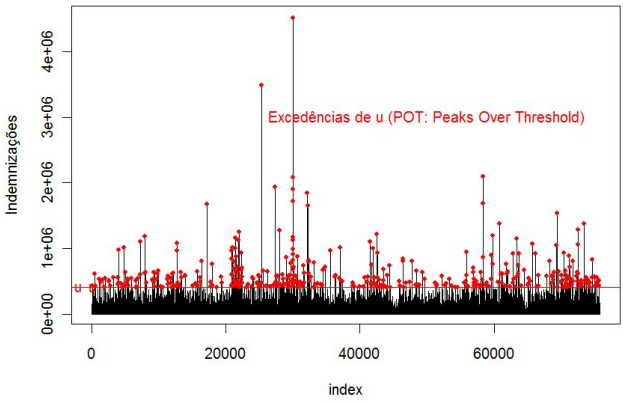


Figura 9.22: soa.txt: toda a base de dados e excedências de $u = 400000$ USD

Esquematizamos, em seguida, a forma usual de proceder mais detalhadamente:

- Obtenham-se as observações possíveis da v.a. dos *excessos* $(X-u)|X>u$, i.e., usem-se os dados acima desse nível u (excedências), transformando-os em excessos,

$$\{y_i = x_i - u\}_{i=1}^{Nu} = \{y_i = x_i - 400000\}_{i=1}^{397}.$$

- Ajuste-se o modelo GP, $H_\gamma(y;\sigma_u)$ $\sigma=\sigma_u$, à v.a. $(X-u)|X>u$

$$H_\gamma(y;\sigma) := 1 - \left(1+\gamma\frac{y}{\sigma}\right)_+^{-1/\gamma}$$

- Trace-se o QQ-plot associado a uma GP standard, $H_\gamma(y;1) = H_\gamma(y)$, i.e., marque-se a nuvem de pontos,

$$\left\{(H_\gamma^{\leftarrow}(p_i)\ ,\ y_{i:N_u}) : i=1,\ldots,N_u\right\},\ p_i := \frac{i}{N_u+1}.$$

- Proceda-se à **estimação preliminar** de γ: valor de $(-0.5, 0.5)$ que maximiza a correlação entre as observações $\{y_i\}_{i=1}^{Nu}$ e a recta ajustada ao QQ-plot. Obtivemos $\hat{\gamma} = 0.453717$.

- Face à **recta ajustada ao QQ-plot**, i.e., à recta com declive=a ajustada à nuvem de pontos

$$\left\{(H_\gamma^{\leftarrow}(p_i) = ((1-p_i)^{-\gamma}-1)/\gamma\ ,\ y_{i:N_u}) : i=1,\ldots,N_u\right\}$$

pelo *método dos mínimos-quadrados*, obtida pela minimização de

$$h(a) := \sum_{i=1}^{N_u} (y_{i:N_u} - a\,q_i)^2\ ,\ \text{com } q_i := ((1-p_i)^{-\gamma}-1)/\gamma,$$

temos

$$\hat{a} = \sum_{i=1}^{N_u} y_{i:n}q_i / \sum_{i=1}^{N_u} q_i^2\,.$$

A Figura 9.23 apresenta o QQ-plot associado ao modelo GP($\hat{\gamma}=0.453717$), i.e., a nuvem de pontos

$$\left\{(H_{\hat{\gamma}}^{\leftarrow}(p_i := i/(N_u+1)), y_{i:N_u}), i=1,\ldots,N_u\right\},$$

bem como o W-plot associado ao modelo GP($\hat{\gamma}$), ou seja a nuvem de pontos $(-\log(1-p_i), -\log(1-H_{\hat{\gamma}}(y_{i:N_u}, \hat{\sigma})))$ $\quad i = 1, \ldots, N_u$, que deverá estar perto da diagonal. Ajustando uma recta de mínimos quadrados, obtem-se para parâmetro de escala, $\hat{\sigma} = \hat{a} = 126788$, a que corresponde um coeficiente de correlação $r = 0.9932243$.

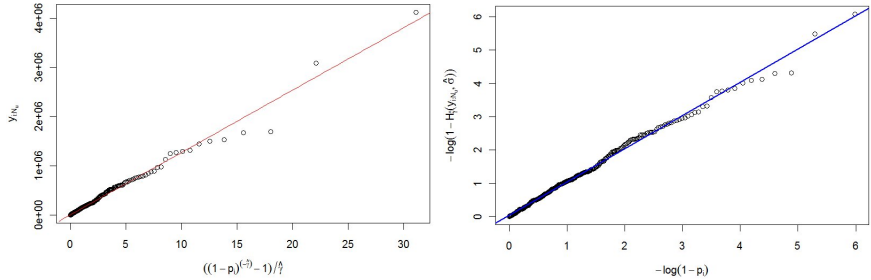


Figura 9.23: soa.txt: QQ-plot (*esquerda*) e W-plot (*direita*) associados ao modelo GP

Ao colocar-se a pergunta, '*qual a probabilidade de uma futura indemnização superior a* 25000 *USD vir a exceder o máximo disponível observado*', a resposta é a seguinte: Recordando que dada $X \frown F$, a distribuição condicional dos excessos acima de u é

$$F_u(y) = \mathbb{P}[X - u \leq y | X > u] = \frac{F(u+y) - F(u)}{1 - F(u)}, \qquad 0 \leq y \leq x^F - u,$$

então $F(u+y) = F_u(y)\{1 - F(u)\} + F(u)$ ou, equivalentemente,

$$\begin{aligned} 1 - F(u+y) &= 1 - [F_u(y)\{1 - F(u)\} + F(u)] \\ &= \{1 - F(u)\}[1 - F_u(y)], \end{aligned}$$

pelo que para o máximo dos excessos acima de 400000 USD observado,

$$y = x_{n:n} - u = y_{N_u:N_u} = 4518420 - 400000 = 4118420,$$

se tem

$$\begin{aligned} \mathbb{P}[X > x_{n:n}] &= 1 - F(x_{n:n}) \\ &= 1 - F(u + y_{N_u:N_u}) \\ &= \{1 - F(u)\}[1 - F_u(y_{N_u:N_u})]. \end{aligned}$$

Utilizando a aproximação ao modelo GP, $F_u(\cdot) \approx H_{\hat{\gamma}}(\cdot, \hat{\sigma})$, vem que

$$F_u(y_{N_u:N_u}) \simeq H_{\hat{\gamma}}(y_{N_u:N_u}, \hat{\sigma}) = H_{0.453717}(y_{397:397}, 126788) = 0.9976991.$$

Por outro lado, estimando $1-F(u)$ pela frequência relativa das indemnizações excedendo o nível u de entre as $n = 75789$ da amostra disponível (*por sua vez já acima dos* 25 000 *USD*),

$$\widehat{1-F(u)} \quad = \quad \frac{N_u}{n} = 397/75789 = 0.005238227,$$

obtem-se a resposta ao problema inicial

$$\mathbb{P}[X > x_{n:n}] \simeq \frac{N_u}{n}\left[1 - H_{\hat{\gamma}}(y_{N_u:N_u}, \hat{\sigma})\right] = 1.20528e-05.$$

A metodologia POT.

Passemos agora à aplicação da metodologia POT aos dados originais das $n = 75789$ grandes indenizações disponíveis *e acima de* 25000 *USD* (Beirlant et al., 2004).

Considere-se que na Figura 9.22 o nível u varia. A função de excesso médio, como função de u, pode ser obtida a partir das instruções:

```
library(ismev), mrl.plot
```

```
soa<-read.table('soa.txt', header=FALSE,dec="."); x<-soa$V1
library(ismev)
mrl.plot(x,conf = 0)     # library(ismev)
```

```
> abline(v=400000); x<-sort(x); n=length(x); k=200; x[n-k]; abline(v=x[n-k])
[1] 512458
```

Do gráfico da Figura 9.24 não se extrai nenhuma conclusão categórica acerca do nível u conveniente.

Vamos em seguida considerar o nível $u = 400000$ e posteriormente o nível $x_{n-k:n} = 512458$, $n = 75789$ e $k = 200$. Os 2 níveis estão assinalados pelas 2 rectas verticais na Figura 9.24 e foram igualmente abordados em Beirlant *et al.* (2004). Temos então $N_u = 397$ excessos acima de $u = 400000$ USD.

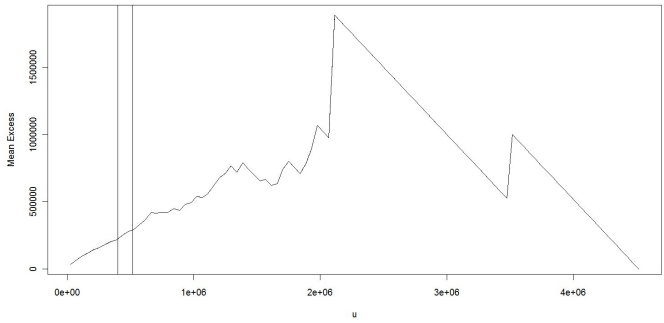


Figura 9.24: soa.txt: Função ME para os dados em estudo

Estimação ML,

`library(ismev)`

```
library(ismev)
u=400000
gpd.fit(x,threshold=u)   #  library(ismev)
```

Obtiveram-se a estimativas ML

$$\hat{\sigma} = 142558, \quad \hat{\gamma} = 0.3819 \quad e \quad L^*(\hat{\gamma}, \hat{\sigma}) = -5260\,.$$

`library(evir)`

```
library(evir)
u=400000
gpd(x,u, method = "ml")  #  library(evir)
```

Obtivémos as estimativas ML semelhantes às anteriores

$$\hat{\sigma} = 142147, \quad \hat{\gamma} = 0.3847 \quad e \quad L^*(\hat{\gamma}, \hat{\sigma}) = -5260\,.$$

Estimação PWM

`library(evir)`

```
gpd(x,u, method = "pwm") #  library(evir)
```

As estimativas PWM obtidas foram

$$\hat{\sigma} = 143410 \quad e \quad \hat{\gamma} = 0.3686\,.$$

Ainda para a **estimação ML**, usámos a

```
library(fitdistrplus)
```

```
library(fitdistrplus)
dGP<-function(x,g,a){ (1+g*x/a)^(-1/g-1)/a}
pGP<-function(x,g,a){1-(1+g*x/a)^(-1/g)}
qGP<-function(p,g,a){a*((1-p)^(-g)-1)/g}
a_GP= 126788  ; g_GP=0.453717    # (0.453717,126788)    from  QQ-Plot
fGP <- fitdist(y,"GP",start=list( g=g_GP,a=a_GP))   #  library(fitdistrplus)
print(fGP)
```

e fomos conduzidos às estimativas,

$$\hat{\sigma} = 142422 \quad e \quad \hat{\gamma} = 0.3829\,.$$

Construímos assim o **W-plot**, usando as estimativas ML, $\hat{\sigma} = 142147$ e $\hat{\gamma} = 0.3847$, representando graficamente a nuvem de pontos,

$$(-\log(1-p_i), -\log(1-H_{\hat{\gamma}}(y_{i:N_u}, \hat{\sigma}))) \quad i = 1, 2, \ldots, N_u.$$

Como se pode constatar na Figura 9.25, essa nuvem de pontos encontra-se perto da diagonal.

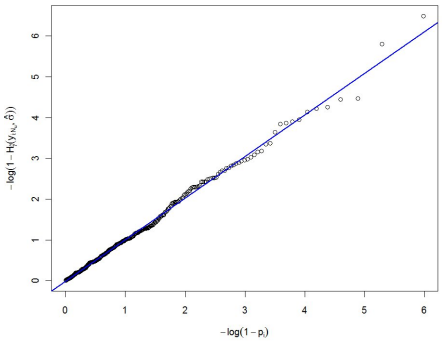


Figura 9.25: soa.txt: W-plot com estimativas ML

Relativamente à estimação do **EVI**, o gráfico da Figura 9.26, ilustra o comportamento das trajectórias das estimativas ML e PWM versus $k < 500$.

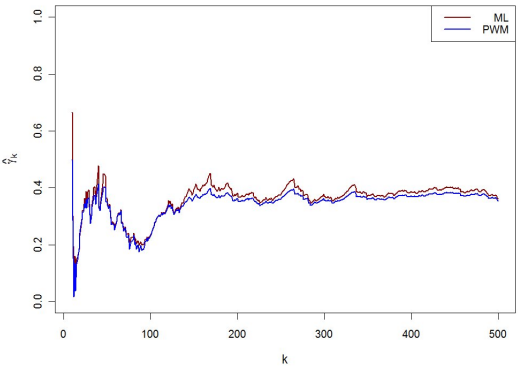


Figura 9.26: soa.txt: Estimativas ML e PWM versus k

Neste mesmo âmbito da metodologia POT, e como já vimos no Capítulo 7, um outro problema com interesse é o da estimação de *quantis extremais* e do *limite superior do suporte*.

O gráfico da Figura 9.27, ilustra o comportamento das trajectórias das estimativas de $U(100\,000)$ versus $k < 500$, obtidas através da abordagem POT.

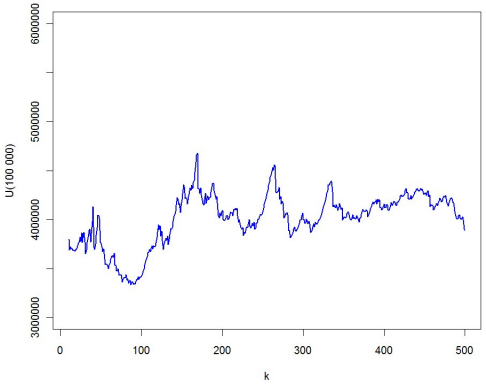


Figura 9.27: soa.txt: Estimativas de $U(100\,000)$ versus $k < 500$

Para estimação de $U(100\,000)$ foram incorporadas as estimativas de ML na expressão de $\hat{U}(1/p)$ para $p = 1/100\,000$, calculadas para $k < 500$. Para valores de k numa região sensivelmente em $(200, 500)$, parece haver alguma estabilidade à volta do valor $4\,000\,000$.

Observação 9.3.1. *Alternativamente, a estimativas* ML *de quantis elevados,* $\hat{U}(1/p)$*, podem ser obtidas directamente por reparametrização da log-verosimilhaça da* GP *em termos de* $U(1/p)$*, por exemplo, fazendo*

$$\sigma = \frac{\gamma\,(U(1/p) - u)}{\left(\frac{np}{N_u}\right)^{-\gamma} - 1}\,.$$

Estimação de EVI e de $U(100\,000)$.
Os 2 gráficos anteriores foram obtidos a partir do `script`:

```
############################################################
### Trajectórias: gama (ML) + Nível de Retorno U(100 000) ###
############################################################
library(evir)
qGP<-function(p,g,a){a*((1-p)^(-g)-1)/g}
soa<-read.table('soa.txt', header=FALSE,dec=".")
x<-soa$V1
n=length(x)
x <- sort(x)
gamma_ML <-c()
gamma_PWM <-c()
rlevel_ML <- c()
rlevel_PWM <- c()
for(k in seq(from = 10, to =500 )) { # ou ceiling(n/100)
   u=x[n-k]  ; Nu=k+1
 out <- gpd(x,u)                 #  library(evir)
   A <- out$par.ests
   gamma_ML[k-9] <- A[1]
   g_ML <- A[1] ; a_ML <- A[2]
   p=1/100000  ;z= 1-n*p/Nu
   rlevel_ML[k-9] <- u+qGP(z,g_ML,a_ML)
}
k <- seq(from = 10, to = 500)
plot(k,rlevel_ML,ylab="U(100 000)",ylim=c(3*10^6,6*10^6), xlab="k", cex.lab=1,lwd=2,type="l",col="blue
par(oma=c(0,2,0,0)) #Add space for main title
k <- seq(from = 10, to = 500)
plot(k,gamma_ML,ylab=bquote(hat(gamma)[k]),lwd=2,type="l",cex.lab=1.2,col="darkred",ylim=c(0,1) )
for(k in seq(from = 10, to =500 )) { # ou ceiling(n/100)
   u=x[n-k]  ; Nu=k+1
  out <- gpd(x,u, method = "pwm") #  library(evir)
   A <- out$par.ests
   gamma_PWM[k-9] <- A[1]
   g_PWM <- A[1] ; a_PWM <- A[2]
```

```
   p=1/100000  ;z= 1-n*p/Nu
   rlevel_PWM[k-9] <- u+qGP(z,g_PWM,a_PWM)
}
k <- seq(from = 10, to = 500)
lines(k,gamma_PWM,ylab=bquote(hat(gamma)[k]),  cex.lab=2,lwd=2,type="l",col="blue")
temp <- legend("topright", legend = c(" "," "),
              text.width = strwidth("      "),lwd=c(2,2),
               lty =c(1,1),col=c("darkred","blue"), xjust = 1, yjust = 1)
text(temp$rect$left + temp$rect$w, temp$text$y,
    c("ML","PWM"), pos=2)
```

Abordagem POT – comportamento das trajectórias das estimativas ML de γ e das bandas de confiança a 95%, versus $k < 500$:

Estimação ML de EVI vs k e IC a 95%.

Obtida muito simplesmente com o `script`

```
############## EVI & CI 95%  ###################
library(evir)
shape(x,models = 490,start = 15, end = 500,reverse = F)
```

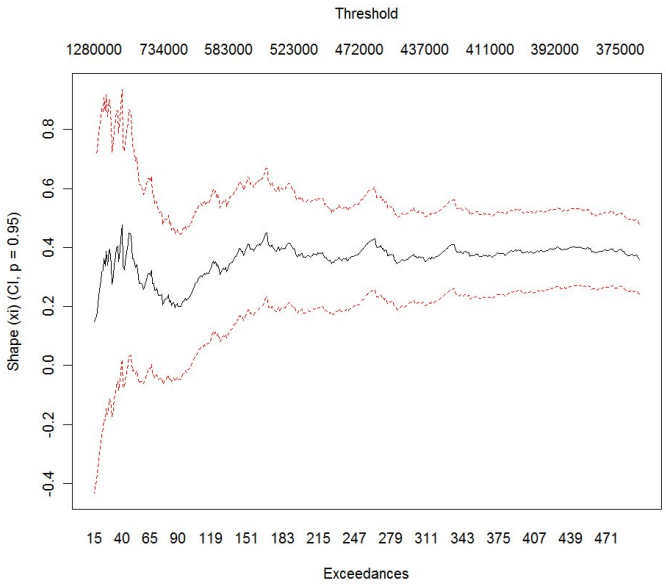


Figura 9.28: soa.txt: Estimação ML de EVI vs k e IC a 95%.

Em Beirlant et al. (2004) é considerado $k = 200$ excessos acima de $u = x_{n-k:n} = 512\,458$ USD para o estudo do

IC para γ baseado no *profile-likelihood*, com g.c. 95% ***aproximadamente***, apresentado na Figura 9.29, e obtido através do `script`:

```
########################################################################
##  optimização unidimensional  ###   L*(sigma, g=gi,fixed) optimize  ###
########################################################################
soa<-read.table('soa.txt', header=FALSE,dec=".")
x<-soa$V1 ;  x<-sort(x)
n<-length(x)          # n = 75789
k=200
index<-n-k            # n-k=75589
x[index]              # x[n-k]=x[n-k]=512458
                      #k=200   u~512458
length(x[x>x[index]]) # =k=200
u=x[index]
y<-x-u; length(y) ; y<-y[y>0]
Nu <- length(y)
Nu                    # =k=200
g.fix <- c(); log.L <-c(); sigma.hat <-c()
for(i in 1:5000) {
    gi=0.15+0.0001*i
    ## ----- (MLE using optimize) -------------
    logL <- function(sigma,g=gi) { #  L*(sigma;g=gi)
    return( -Nu*log(sigma)-(1/g+1)*sum(log(1+g*y/sigma)))
    }
    ##------(0.453717,126788 ) QQ-Plot --------
    out <- optimize(logL,lower=100,upper=2*10^5,maximum=T)
    g.fix [i]<- gi
    log.L[i] <- out$objective
    sigma.hat [i]<- out$maximum
    }
i <- c(1:5000)
i0 <- i[log.L==max(log.L)]
i0              # i0=2318
max(log.L)      # max(log.L)=-2701.813
g.fix [i0]      # g.fix [i0]=0.3818
log.L[i0]       # log.L[i0] =-2701.813
plot(g.fix,log.L,type="l", xlab=bquote(gamma))
abline(h=max(log.L))
q=qchisq(.95, df=1) # 1 degrees of freedom
abline(h=max(log.L)-0.5*q)
abline(v=g.fix [i0]); abline(v=0.217); abline(v=0.610)
text(0.45,-2703.5, "95%, optimize, L*_sigma",col="red")
```

Os resultados obtidos foram os seguintes:
Estimativa pontual para γ: $\hat{\gamma} = 0.3818$;
Estimativa intervalar para γ: $IC_\gamma(95\%) = (0.217, 0.610)$.

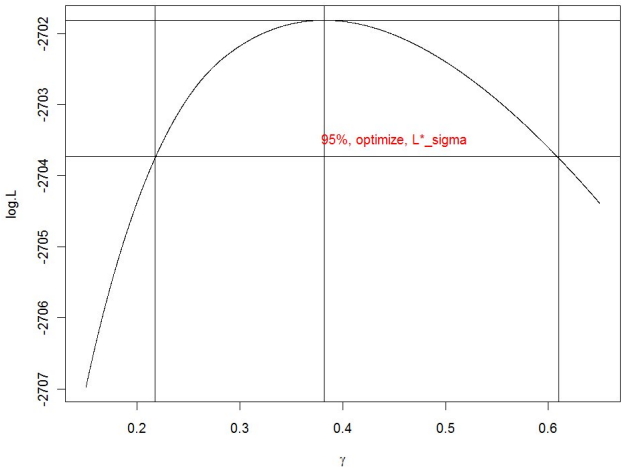


Figura 9.29: soa.txt: IC para γ baseado no *profile-likelihood* com g.c. 95% *aproximadamente*

O **IC para** $U(100\,000)$ **baseado no *profile-likelihood* com g.c.** 95% ***aproximadamente***, apresentado na Figura 9.30, foi obtido através do `script`:

```
##############################################################################
##    optimização bidimensional  ###   L*(gamma,sigma, rl=rli,fixed) optim  ##
##############################################################################
rl.fix <- c(); log.L <-c(); g.hat <- c(); sigma.hat <-c()
for(i in 1:750) {
    rli=2.5*10^6+10^4*i
    logL <- function(theta,y,u,n,Nu,p=1/100000,rl=rli) {# L*(gama,sigma)
        sigma <- theta[1]
        g <- theta[2]
        sigma <- (g*(rl-u))/(((n*p)/Nu)^(-g)-1)
        loglik <- -Nu*log(sigma)-(1/g+1)*sum(log(1+g*y/sigma))
        return(- loglik)
    }


    ##---- valores iniciais (0.453717,126788 ) QQ-Plot ----##
    optim(c(126788,0.453717), logL, y=y,u=u,n=n,Nu=Nu)
    out <- optim(c(126788,0.453717), logL, y=y,u=u,n=n,Nu=Nu)
    theta.hat <- out$par
    sigma.hat [i]<- theta.hat [1]
```

```
        g.hat [i]<- theta.hat [2]
    rl.fix [i]<- rli
    log.L[i] <- - out$value
    }
i <- c(1:1000)
i0 <- i[log.L==max(log.L)]
plot(rl.fix,log.L,type="l", xlab="U(100 000)")
abline(h=max(log.L))
q=qchisq(.95, df=1) # 1 degrees of freedom
abline(h=max(log.L)-0.5*q)
abline(v=rl.fix [i0]); abline(v=2.69*10^6);abline(v=8.25*10^6)
text(6*10^6,-2703.5, "95% , optim",col="red")
```

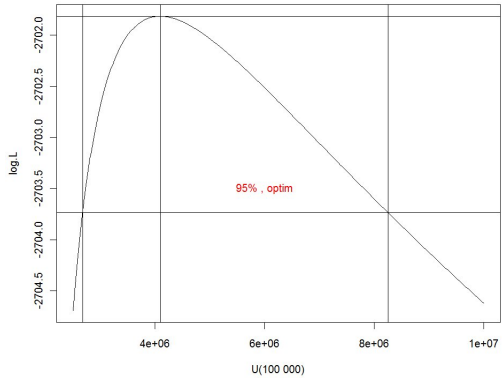


Figura 9.30: soa.txt: IC para $U(100\,000)$ baseado no *profile-likelihood* com g.c. 95% *aproximadamente*

Os resultados obtidos foram:
Estimativa pontual para $U(100\,000)$:

$$\hat{U}(100\,000) = 4\,100\,000\,USD.$$

Estimativa intervalar para $U(100,000)$:

$$IC_{U(100,000)}(95\%) = (2.69 \times 10^6\,USD,\ 8.25 \times 10^6\,USD).$$

Abordagem Semi-Paramétrica

Avancemos agora com uma abordagem semi-paramétrica dos dados `soa.txt`, das grandes indemnizações em 1991 acima de $u = 25000$ USD, com dimensão $n = 75789$ (Beirlant *et al.*, 2004), começando com a

Estimação do EVI – γ

Nas Figuras 9.31 e 9.32 apresentamos as estimativas de Hill, dos Momentos e de Pickands. Na Figura 9.33 apresentamos o gráfico global dessas estimativas.

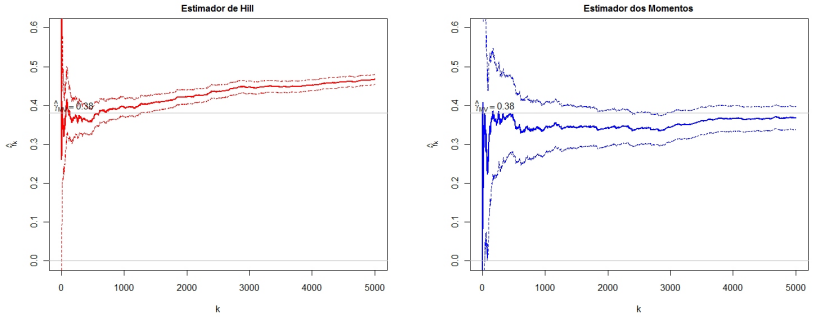


Figura 9.31: soa.txt: Estimador de Hill vs k (*esquerda*); Estimador de Momentos vs k (*direita*).

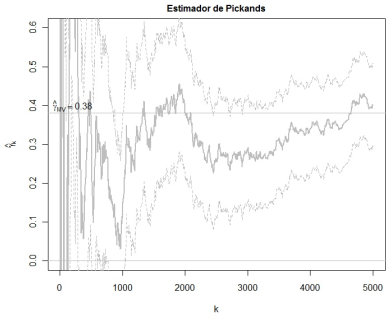


Figura 9.32: soa.txt: Estimador de Pickands vs k

Observação 9.3.2. *A maior variabilidade está claramente associada ao estimador de Pickands em comparação com os anteriores.*

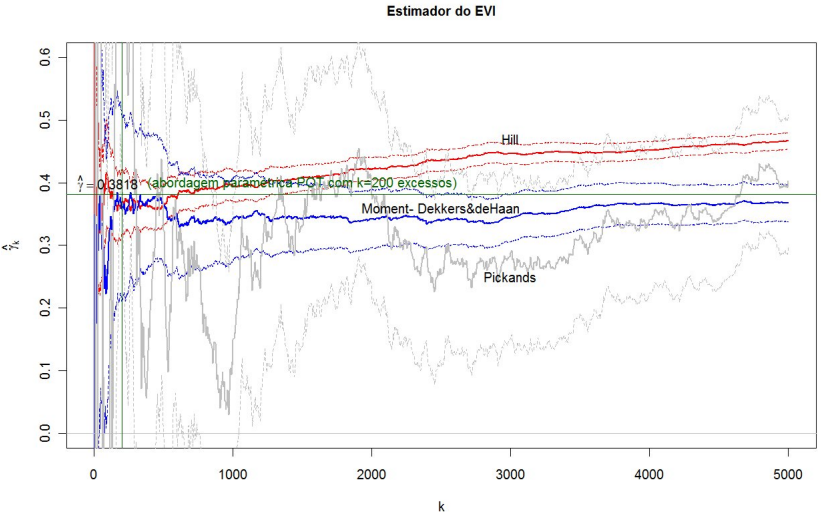


Figura 9.33: soa.txt: EVI – Abordagem semi-paramétrica – Estimador de Weissman-Hill, Weissman-Momentos (EVI positivo) e Weissman-Momentos (EVI real) vs k; abordagem paramétrica POT – a recta horizontal corresponde a ML $\hat{\gamma} = 0.3818$, para $k = 200$ – estudo comparativo.

O `script` que serviu de base a estas figuras é

```
##########################################
#      S E M I P A R A M E T R I C S     #
##########################################
soa<-read.table('soa.txt', header=FALSE,dec=".")
x<-soa$V1; length(x)   # 75789
x<-sort(x)
n<-length(x)          # n = 75789
##########################################
#            MOMENTO ordem r   (k o.s.)   #
##########################################
Mk_r <- function(x,r,k) { # x=vector
n=length(x)
x<-sort(x,decreasing = F)
mr <-mean((log(x[(n-k+1):n])-log(x[n-k]))^r)
return(mr)
}
```

```
##########################################
#            Plot_Hill até ordem k1      #
#     com bandas de confiança asympt     #
#          plots fora da function        #
##########################################
hill_est<-function(x,k1) {  #  k1 <= n-1
n=length(x)
x<-sort(x,decreasing = F)
kk<-numeric()
hk<-numeric()
for(k in 1:k1) {
  kk[k]<-k
  hk[k] = Mk_r(x,1,k)
}
k<-kk
Hill<-hk
out<-cbind(k,Hill)
return(out)
}
n1=5000
out <- hill_est(x,n1)
gamma_hat <- hill_est(x,n1)
g <-gamma_hat[,2]
lower <-   g - 1.96/sqrt(1:(n1))*g
upper <-   g + 1.96/sqrt(1:(n1))*g
plot(out,type = "l",ylim=c(0,0.6),col = "red",lwd=2,ylab=bquote(hat(gamma)[k]))
lines(1:(n1),lower,type = "l",lty=2 ,col = "red")
lines(1:(n1),upper,type = "l",lty=2 ,col = "red")
abline(h=0,col="grey")
##########################################
#          Plot_Moment até ordem k1      #
##########################################
Mom_est<-function(x,k1) {# k1 <= n-1; r integer
n=length(x)
x<-sort(x,decreasing = F)
kk<-numeric()
Momk<-numeric()
for(k in 1:k1) {
    kk[k]<-k
    M1=Mk_r(x,1,k);M2=Mk_r(x,2,k)
    Momk[k] = M1+1-0.5*(1-M1^2/M2)^(-1)
}
k<-kk
Mom<-Momk
out<-cbind(k,Mom)
#plot(out,type = "l", ylim=c(lower,upper),col = "blue")
return(out)
}
##########################################
n1=5000
n=length(x)
gamma_hat <- Mom_est(x,n1)
gamma_hat
```

```
g <-gamma_hat[,2]
##########################################
Var_Mom <- function(g) { # g vector # variância assintótica MOM
var=g
gmenos=g[g<0]
gmais=g[g>=0]
g1=gmenos;g2=gmais
var[g<0]=(1-g1)^2*(1-2*g1)*(4-8*(1-2*g1)/(1-3*g1)+(5-11*g1)*(1-2*g1)/((1-3*g1)*(1-4*g1)))
var[g>=0]=1+g2^2
return(var)
}
##########################################
lower <-   g - 1.96/sqrt(1:(n1))*sqrt(Var_Mom(g))
upper <-   g + 1.96/sqrt(1:(n1))*sqrt(Var_Mom(g))
lines(gamma_hat ,type = "l",ylim=c(0,0.6),lwd=2,col = "blue",ylab=bquote(hat(gamma)[k]))
lines(1:(n1),lower,type = "l",lty=2 ,col = "blue")
lines(1:(n1),upper,type = "l",lty=2 ,col = "blue")
abline(h=0,col="grey")
##########################################
#        Plot_Pickands até ordem k1       #
##########################################
Var_Pick <- function(g) {  # variância assintótica Pick
var=g^2*(2^(2*g+1)+1)/(((2^g-1)*log(2))^2)
return(var)
}
##########################################
Pick_est<-function(x,lower,upper) {
X<-x
n=length(X)
Xs=rev(sort(X))
n=5000
### ------   Estimador Pickands   ----- ###
gam=1/log(2)*log( (Xs[seq(1,length=trunc(n/4),by=1)]-Xs[seq(2,length=trunc(n/4),by=2)])/
 (Xs[seq(2,length=trunc(n/4),by=2)]-Xs[seq(4,length=trunc(n/4),by=4)]) )
### --  Bandas de Confiança  (assint)---###
g<-gam
gam_se=1.96/sqrt(seq(1,length=trunc(n/4),by=4))*sqrt( Var_Pick(g))
### ----      Bandas de Confiança  (assint) ----###
lines(seq(4,length=trunc(n/4),by=4),gam,col="grey",lwd=2,type="l",lty=1, ylim=c(y_L,y_U),xlab="k",ylab
lines(seq(4,length=trunc(n/4),by=4),pch=2,gam+gam_se,col="grey",lwd=1.5,lty=2)
lines(seq(4,length=trunc(n/4),by=4),pch=2,gam-gam_se,col="grey",lwd=1.5,lty=2)
abline(h=0,col="gray")
}
y_L=0;y_U=0.6
Pick_est(x,y_L,y_U)
text(3000,0.47,"Hill");text(2500,0.36,"Moment- Dekkers&deHaan");text(3000,0.25,"Pickands")
## A estimação paramétrica com um threshold elevado era 0.3818 (k=200 excessos)##
abline(h=0.3818,col="darkgreen")
abline(v=200,col="darkgreen");text(100,0.4,bquote(hat(gamma)==0.3818))
text(1500,0.4,"(abordagem paramétrica POT com k=200 excessos)",col="darkgreen")
title(main = list("Estimador do EVI", cex=1, col="black", font=2))
```

E finalmente na Figura 9.34 é feito um estudo comparativo das abordagens paramétrica e semi-paramétrica na estimação de um quantil extremal $U(100000)$.

Estimação de Quantil Extremal – $U(100000)$

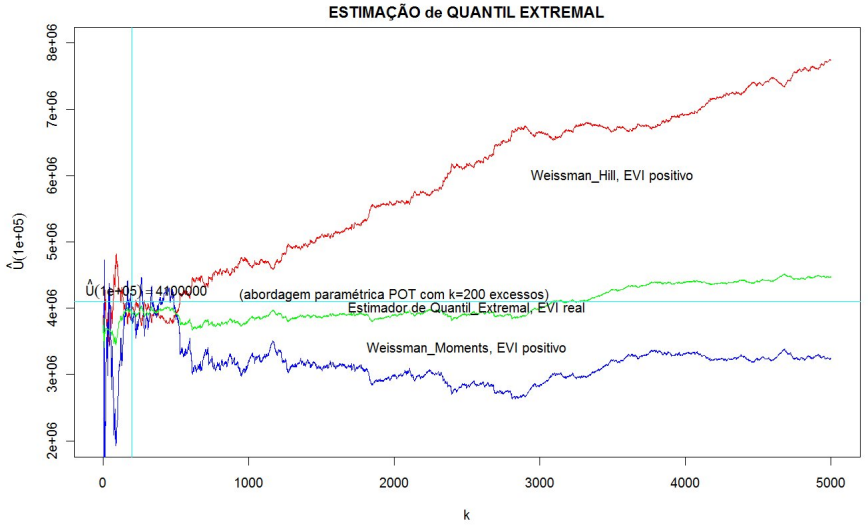


Figura 9.34: soa.txt: Quantil Extremal – Abordagem semi-paramétrica – Estimador de Hill, Momentos e Pickands vs k; abordagem paramétrica POT – a recta horizontal corresponde a ML estimativa $\hat{U}(100\,000) = 4\,100\,000\,USD.$, para $k = 200$ – estudo comparativo.

O `script` seguinte serviu de base para obter as Figuras 9.33 e 9.34.

```
##########################################
#     S E M I P A R A M E T R I C S      #
##########################################
soa<-read.table('soa.txt', header=FALSE,dec=".")
x<-soa$V1; x<-sort(x)
n<-length(x)          # n = 75789
##########################################
#   Plot_quantil_extremal até ordem k1   #
#        caso de gama positivo           #
```

```
##########################################
Quant_Ext_WeissHill<-function(x,p,k1) {# k1 <= n-1;
n=length(x)
x<-sort(x,decreasing = F)
kk<-numeric(); Qpk<-numeric()
for(k in 1:k1) {
    kk[k]<-k
    M1=Mk_r(x,1,k);
     Qpk[k]=x[n-k]*(k/(n*p))^ M1        # Hill
}
k<-kk ; Qp<-Qpk ; out<-cbind(k,Qp)
return(out)
}
n1=5000; n=length(x); p=1/100000; Quantil_hat <- Quant_Ext_WeissHill(x,p,n1)
par(mar=c(5,5,2,2))
plot(Quantil_hat,type="l",col="red",
    ylim=c(2*10^6,8*10^6),ylab=bquote(hat(U)(100000))) #Hill
text(3500,6*10^6,"Weissman_Hill, EVI positivo")



Quantile_Ext_WeissMom<-function(x,p,k1) {# k1 <= n-1;
n=length(x); x<-sort(x,decreasing = F)
kk<-numeric();  Momk<-numeric(); Qpk<-numeric()
for(k in 1:k1) {
    kk[k]<-k
    M1=Mk_r(x,1,k);M2=Mk_r(x,2,k)
    Momk[k] = M1+1-0.5*(1-M1^2/M2)^(-1)
    Qpk[k]=x[n-k]*(k/(n*p))^ Momk[k]      # Mom
}
k<-kk; Qp<-Qpk; out<-cbind(k,Qp)
return(out)
}
n1=5000; n=length(x); p=1/100000
Quantil_hat <- Quantile_Ext_WeissMom(x,p,n1)
lines(Quantil_hat ,type = "l",col = "blue")  # Mom
text(2500,3.4*10^6,"Weissman_Moments, EVI positivo")



##########################################
#       Quantil_extremal até ordem k1     #
#             caso de gama real           #
##########################################
Mk_r <- function(x,r,k) { # x=vector Momento_r
n=length(x); x<-sort(x,decreasing = F)
mr <-mean((log(x[(n-k+1):n])-log(x[n-k]))^r)
return(mr)
}
Quantil_Ext<-function(x,p,k1) {# k1 <= n-1;
n=length(x); x<-sort(x,decreasing = F)
kk<-numeric(); Momk<-numeric(); Qpk<-numeric()
for(k in 1:k1) {
    kk[k]<-k
```

```
    M1=Mk_r(x,1,k);M2=Mk_r(x,2,k)
    Momk[k] = M1+1-0.5*(1-M1^2/M2)^(-1)
    gam_=Momk[k]-M1
    ank=(1-gam_)*x[n-k]*M1
    Qpk[k]=x[n-k]+ank*((k/(n*p))^ Momk[k] -1 )/Momk[k]   # gama real
}
k<-kk; Qp<-Qpk; out<-cbind(k,Qp)
return(out)
}


##########################################
#   Plot_quantil_extremal até ordem k1    #
#               caso de gama real         #
##########################################
n1=5000
n=length(x)
p=1/100000
Quantil_hat <- Quantil_Ext(x,p,n1)
Quantil_hat
lines(Quantil_hat ,type = "l",col = "green")  # gama real
text(2500,4*10^6,"Estimador de Quantil_Extremal, EVI real")
title(main = "ESTIMAÇÃO de QUANTIL EXTREMAL")

## A estimação paramétrica com um threshold elevado era 4100000 (k=200 excessos)##
abline(h=4100000,col="cyan")
abline(v=200,col="cyan");text(300,4.3*10^6,bquote(hat(U)(100000)==4100000))
text(2000,4.2*10^6,"(abordagem paramétrica POT com k=200 excessos)")
```

Comentários Finais

- Nos gráficos anteriores verificamos exactamente que quando tomamos 200 excessos no estudo POT os resultados são similares aos obtidos nesta abordagem semi-paramétrica.
- É de realçar como comentário final que o grau de dificuldade de escolha de um nível elevado na abordagem POT ou das MMO é paralelo ao da escolha de k na abordagem semi-paramétrica.
- Devemos encarar as metodologias paramétricas e semi-paramétricas não como concorrentes, mas sim como complementares à inferência estatística em Valores Extremos.
- A escolha de um nível extremal não é trivial e, embora muitos tenham sido os contributos quer de índole teórica, quer utilizando técnicas heurísticas, constitui uma área onde muito haverá ainda a desenvolver.
- Muito mais haveria a dizer sobre o papel da EVT na Modelação de Acontecimentos Raros.

- O caso da não independência da amostra foi apenas aflorado, realçando nessa área o papel importante desempenhado por outro parâmetro designado por *índice extremal* θ que descreve e quantifica as características de agrupamentos de valores extremos (*clustering*) em séries temporais estacionárias ($\theta \approx$ (tamanho médio dos *clusters*)$^{-1}$)
- Nesta introdução ao estudo de Valores Extremos, focámos a nossa atenção no caso univariado, mas como facilmente se antevê a EVT quer no campo multivariado ou espacial tem igualmente relevância para a *Modelação de Acontecimentos Raros*.

Epílogo

Esperamos ter aguçado o vosso apetite por uma área relativamente recente em termos históricos, e com tantas áreas de aplicação quantas as que possamos conceber.

F I M

Bibliografia

[1] Aarssen, K. & de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* **4**:4, 259–281.

[2] Abramowitz, M. & Stegun, I.A. (1992). *Handbook of Mathematical Functions.* Dover, New York.

[3] Aragonés, J., Blanco, C. & Dowd, K. (2000). The Learning Curve: Extreme Value Theory for VaR (`http://www.fea.com/resources/pdf/a_evt_1.pdf`Part 1 & `http://www.fea.com/resources/pdf/a_evt_2.pdf`Part 2).

[4] Araújo Santos, P. (2011). *Excesses, Durations and Forecasting Value-at-Risk.* PhD Thesis, University Lisbon.

[5] Araújo Santos, P., Fraga Alves, M.I. & Gomes, M.I. (2006). Peaks over random threshold methododlogy for tail index and high quantile estimation. *Revstat* **4**:3, 227–247.

[6] Arnold, B.C. & Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics.* Springer-Verlag.

[7] Arnold, B., Balakhrishna, N. & Nagaraja, H. N. (1992; 2008). *A First Course in Order Statistics.* 1st Ed., Wiley; 2nd Ed., SIAM.

[8] Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580.

[9] Balkema, A.A. & de Haan, L. (1974). Residual life time at great age. *Annals of Probability* **2**, 792–804.

[10] Barlow R.E. & Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart & Winston.

[11] Barnett, V. (1975). Probability plotting methods and order statistics. *Applied Statistics* **24**, 95–108.

[12] Beirlant, J., Teugels, J.L. & Vynckier, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press.

[13] Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, England.

[14] Beirlant, J., Caeiro, F. & Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* **10**:1, 1–31.

[15] Bingham, N., Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation*. Cambridge Univ. Press, Cambridge.

[16] Blom, G. (1958). *Transformed Beta Variates*. Wiley.

[17] Box, G.E.P. & Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. Wiley.

[18] Bury, K.V. (1975). *Statistical Models in Applied Science*. Wiley.

[19] Castillo E., Hadi A.S., Balakrishnan N. & Sarabia, J.M. (2004). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley.

[20] Chernoff, H. & Lieberman, G.J. (1954). Use of normal probability paper. *J. Amer. Statist. Assoc.* **49**, 778–785.

[21] Chernoff, H. & Lieberman, G.J. (1956). The use of generalized probability paper for continuous distributions. *Ann. Math. Statist.* **27**, 806–818.

[22] Cole, R.H. (1951). Relations between moments of order statistics. *Ann. Math. Statist.* **22**, 308–310.

[23] Coles S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag.

[24] Csörgő, S. & Mason, D.M. (1989). Simple estimators of the endpoint of a distribution. In Hüsler, J., & Reiss, R.-D. (1989), *Extreme Value Theory, Proceedings Oberwolfach* 1987, 132–147, Springer-Verlag, Berlin, Heidelberg.

[25] David, H.A. (1981). *Order Statistics*. 2nd Ed., Wiley.

[26] David, H.A. & Nagaraja, H.N. (2003). *Order Statistics*. 3rd Ed., Wiley.

[27] Davison, A. (1984). Modeling excesses over high threshold with an application. In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, D. Reidel, 461–482.

[28] Davison, A.C. & Smith, R.L. (1990). Models for exceedances over high thresholds. *J. Royal Statist. Soc. B* **52**, 393–442.

[29] Dekkers, A.L.M. & de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics* **17**, 1795–1832.

[30] Dekkers, A.L.M., Einmahl, J.H.J. & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statististics* **17**, 1833–1855.

[31] Dijk, V. & de Haan, L. (1992). On the estimation of the exceedance probability of a high level. Order statistics and nonparametrics: theory and applications. In Sen, P.K., & Salama, I. A. (eds.), 79–92, Elsevier, Amsterdam.

[32] Drees, H., Ferreira, A. & de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**, 1179–1201.

[33] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* CMBS-NSF 9, SIAM.

[34] Embrechts, P., Klüpelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag.

[35] Falk, M., Hüsler, J. & Reiss, R.-D. (1994; 2005; 2010). *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser.

[36] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications.* Vol. 2, Wiley.

[37] Ferreira, A. (1997) Extreme Sea Level in Venice. In Reiss, R.-D. & Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Verlag, Basel, Part V—Case Studies in Extreme Value Analysis, Study 1, 233–240.

[38] Ferreira, A. (2002). Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Plannning and Inference* **104**, 83–102.

[39] Fisher, R.A. (1929). Tests of Significance in Harmonic Analysis. *Proc. Royal Statist. Soc. A* **125**:796, 54–59.

[40] Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings Cambridge Philosophical Society* **24**, 180–190.

[41] Fraga Alves, M.I., de Haan, L. & Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods of Statist.* **12**, 155–176.

[42] Fraga Alves, M.I., de Haan, L. & Lin, T. (2006). Third order extended regular variation. *Publications de l'Institut Mathématique* **80**:94, 109–120.

[43] Fraga Alves, M.I., Gomes, M.I., de Haan, L. & Neves, C. (2009). Mixed moment estimators and location invariant alternatives. *Extremes* **12**, 149–185.

[44] Fréchet, M. (1927). Sur le loi de probabilité de l'écart maximum. *Ann. Société Polonaise de Mathématique* **6**, 93–116.

[45] Galambos, J. (1987). *Asymptotic Theory of Extreme Order Statistics*. Krieger.

[46] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**:6, 423–453.

[47] Gomes, M.I. (1981). An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, D. Reidel, 389–410.

[48] Gomes, M.I. (1994). Penultimate behaviour of the extremes. In J. Galambos et al. (eds.), *Extreme Value Theory and Applications*, 403-418, Kluwer Academic Publishers.

[49] Gomes, M.I. & Oliveira, O. (2003). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events. *Comm. in Statist.—Simulation and Computation* **32**:4, 1005–1028.

[50] Gomes, M.I., Reiss, R.-D.& Thomas, M. (2007). Reduced-bias estimation. In Reiss, R.-D. & Thomas, M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd Ed., Chapter 6, 189–204, Birkhäuser Verlag, Basel-Boston-Berlin.

[51] Gomes, M.I., Canto e Castro, L., Fraga Alves, M.I. & Pestana, D.D. (2008). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. *Extremes* **11**:1, 3–34.

[52] Gomes, M.I., Figueiredo, F. & Barão, M.I. (2010). *Controlo Estatístico da Qualidade*. Edições INE.

[53] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research* **15**:5, 1049–1054.

[54] Gross, D., Shortle, J.F., Thompson, J.M. & Harris, C. (2008). *Fundamentals of Queueing Theory*. 4th Ed., Wiley.

[55] Gumbel E.J. (1958; 2004). *Statistics of Extremes*. Columbia University Press; Dover Publications Inc., New York.

[56] de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Thesis, University of Amsterdam / Mathematical Centre Tract 32.

[57] de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: an Introduction*. Springer Science+Business Media, LLC, New York.

[58] de Haan, L. & Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60–70.

[59] de Haan, L. & Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc.* **A61**, 381–395.

[60] Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10**, 556–568.

[61] Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Amer. Soc. Civil Engrs.* **77**, 1539–1659.

[62] Hazen, A. (1930). *Flood Flows. A Study of Frequencies and Magnitudes*. Wiley.

[63] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

[64] Hosking, J.R.M. (1985). Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Appl. Statist.* **34**, 301–310.

[65] Hosking, J.R.M. & Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.

[66] Hosking, J.R.M., Wallis, J.R. & Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261.

[67] Hüsler, J. & Peng, L. (2008). Review of testing issues in extremes: in honor of Professor Laurens de Haan. *Extremes* **11**:1, 99–111.

[68] Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Royal Meteorol. Soc.* **81**, 158–171.

[69] Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari* **4**, 83–91.

[70] Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions —Theory and Applications*. Imperial College Press, London.

[71] Leadbetter, M.R. (1973). On extreme values in stationary sequences. *Z. Wahrsch. und Verw. Gebiete* **28**, 289–303.

[72] Leadbetter M.R., Lindgren G. & Rootzen H. (1983). *Extremes and Related Properties of Random Sequences and Process*. Springer-Verlag.

[73] Loynes, R.M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36**:3, 993–999.

[74] Malmquist, S. (1950). On a property of order statistics from a rectangular distribution. *Skand. Aktuar.* **33**, 214–222.

[75] Mises, R. von (1936). La distribution de la plus grande de n valeurs. *Revue Math. Union Interbalcanique* **1**, 141-160. Reprinted in *Selected Papers of Richard von Mises*, Amer. Math. Soc. **2** (1954), 271–294.

[76] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*. Wiley.

[77] Montgomery, D.C. (1991). *Introduction to Statistical Quality Control*. 2nd Ed., Wiley.

[78] Moore, D.S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.* **39**, 263–265.

[79] Neves, C. & Fraga Alves, M.I. (2008). Testing extreme value conditions — an overview and recent approaches. *Revstat* **6**, 1, 83–100. Special issue on "Statistics of Extremes and Related Fields" edited by J. Beirlant, I. Fraga Alves & R. Leadbetter.

[80] O'Brien, G. (1974). Limit Theorems for the Maximum Term of a Stationary Process. *Ann. Probab.* **2**:3, 540–545.

[81] Pearson, E.S. & Hartley, H.O. (1970). *Biometrika Tables for Statisticians*. Cambridge Univ. Press.

[82] Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.

[83] Pirazzoli, P. (1982) Maree estreme a Venezia (periodo 1872-1981). *Acqua Aria* **10**, 1023–1039.

[84] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4**, 191–231.

[85] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer-Verlag.

[86] Reiss, R.-D. & Thomas, M. (1997; 2001; 2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. 1st Ed.; 2nd Ed.; 3rd. Ed., Birkhäuser Verlag, Basel-Boston-Berlin.

[87] Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag.

[88] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer-Verlag.

[89] Sarhan, A.E. & Greenberg, B.G. (1962). *Contributions to Order Statistics*. John Wiley & Sons.

[90] Schucany, R.W. (1972). Order statistics in simulation. *J. Statist. Comp. and Simul.* **1**, 281–286.

[91] Smith, R.L. (1986). Extreme value theory based on their largest annual events. *J. Hydrology* **86**, 27–43.

[92] Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–1207.

[93] Srikantan, K.S. (1962). Recurrence relations between the PDF's of order statistics and some applications. *Ann. Math. Statist.* **42**, 35–45.

[94] Steutel, F.W. (1967). Random division of an interval. *Statistica Neerlandica* **21**, 231–244.

[95] Stiegler, S.M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2**, 676–693.

[96] Tiago de Oliveira (1984) Bivariate models for extremes: statistical decision. In J. Tiago de Oliveira (ed.), *Statistical Extremes and Applications*. 131–153, D. Reidel.

[97] Tiago de Oliveira, J. (1997). *Statistical Analysis of Extremes*. Pendor.

[98] Tiago de Oliveira, J. & Gomes, M. I. (1984). Two test statistics for choice of univariate extreme models. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*. D. Reidel, Dordrecht, Holland, 651– 668.

[99] Watson, G.S. (1954). Extreme values in samples from m-dependent stationary stochastic processes. *Ann. Math. Statist.* **25**:4, 798–800.

[100] Weibull, W. (1939). *A Statistical Theory of Strength of Materials*. Ing. Vet. A.K, Handl., 151, Genelstabens Litografiska Anstals Forlg Stocklholm, Sweden.

[101] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Ass.* **73**, 812–815.

[102] Wilks, S.S. (1948). Order Statistics. *Bull. Amer. Math. Soc.* **54**:1, Part 1, 6–50.

[103] Wingo, D.R. (1972). Maximum likelihood estimation of the parameters of the Weibull distribution by modified quasilinearization. *IEEE Transactions on Reliability* **21**, 89–93.

[104] Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Analysis* **100**:4, 794–815.

[105] Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *J. Multivariate Analysis* **101**:4, 971–983.

Índice Remissivo