

MEMÓRIAS
DA
ACADEMIA DAS CIÊNCIAS
DE
LISBOA

CLASSE DE CIÊNCIAS

TOMO XLVI

**Estatística de extremos – Um
instrumento para predição de
tremores de terra?**

M. IVETTE GOMES, DINIS PESTANA



ACADEMIA DAS CIÊNCIAS
DE LISBOA

LISBOA • 2019

Estatística de extremos – Um instrumento para predição de tremores de terra?

M. Ivette Gomes¹, Dinis Pestana²

A *Estatística de Extremos* ajuda-nos a controlar acontecimentos potencialmente desastrosos, de enorme relevo para a sociedade e de grande impacto social. Os seus domínios de aplicação são muito variados. Mencionamos as áreas de *Bioestatística*, *Engenharia Estrutural*, *Finanças*, *Hidrologia*, *Meteorologia*, *Seguros* e também *Sismologia* (veja-se, entre outros livros, Reiss and Thomas, 2001, 2007; Beirlant *et al.*, 2004; Castillo *et al.*, 2005; Markovich, 2007; Gomes *et al.*, 2013). Embora seja possível encontrar alguns artigos de interesse histórico relacionados com acontecimentos extremos, o campo remonta a Gumbel, em artigos publicados a partir de 1935, e sumariados em Gumbel (1958; 2004). Gostaríamos ainda de realçar o nome de um Português pioneiro na área de extremos, José Tiago da Fonseca Oliveira, membro efectivo da Academia das Ciências de Lisboa desde 1985 até à sua morte prematura em 1992 (veja-se Gomes, 1993a; Tiago de Oliveira, J.C., ed., 1993, entre outros).

Gumbel desenvolveu procedimentos estatísticos essencialmente baseados no teorema de Gnedenko (Gnedenko, 1943), o chamado *teorema de tipos extremais* (ETT, do Inglês “*extremal types theorem*”), um dos resultados limite fundamentais em *Teoria de Valores Extremos* (EVT, do Inglês “*extreme value theory*”).

Em linhas muito gerais, o ETT permite identificar as distribuições de máximos com as chamadas leis de valores extremos (GEV, do Inglês “*general extreme value*”). Tratam-se das também chamadas leis max-estáveis (MS, do Inglês “*max-stable*”), definidas como leis para as quais é válida a equação funcional $MS^{(a_n x + \beta_n)} = MS(x)$, $n \geq 1$, para $a_n > 0$, $\beta_n \in \mathbb{P}$, onde \mathbb{P} denota o conjunto de números reais. Os modelos MS têm a forma funcional,

$$MS_{\xi}(x; \lambda, \delta) \circ GEV_{\xi}(x; \lambda, \delta) = \exp(-(1 + \xi(x - \lambda)/\delta)^{-1/\xi}), \quad 1 + \xi(x - \lambda)/\delta > 0,$$

onde λ , δ , ξ , possivelmente dependentes de co-variáveis adequadas, são parâmetros desconhecidos de localização, escala e forma, sendo ξ o chamado *índice de valores extremos* (EVI, do Inglês, “*extreme value*

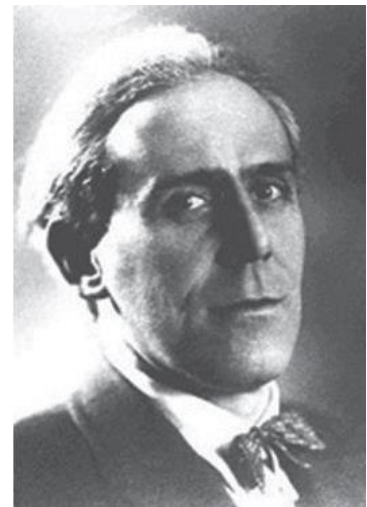


Figura 1
Emil Julius Gumbel (1891-1966) (baixo),
e José Tiago da Fonseca Oliveira
(1928-1992) (cima).

¹CEAUL and DEIO, FCUL, Universidade de Lisboa, Portugal. E-mail: ivette.gomes@fc.ul.pt

²CEAUL and DEIO, FCUL, Universidade de Lisboa, Portugal. E-mail: dinis.pestana@fc.ul.pt

index”). E na realidade, este modelo, contrariamente ao modelo normal, muito frequente em estatística clássica, adapta-se de forma bastante fidedigna, por exemplo, às magnitudes de tremores de terras, nas mais diversas regiões (veja-se Pisarenko & Sornette, 2003, Beirlant *et al.*, 2004, 2016, e Gomes *et al.*, 2013, entre outros).

O EVI, ξ , é o parâmetro fundamental em *Estatística de Extremos*. Se $\xi < 0$, a cauda direita é curta, com limite superior de suporte finito ($x_F < \infty$); se $\xi = 0$, a cauda é de tipo exponencial, e $x_F < \infty$ ou $x_F = \infty$; se $\xi > 0$, temos uma cauda pesada, de tipo Pareto, i.e. polinomial negativa, e $x_F = \infty$. Na **Figura 2** ilustramos o comportamento da cauda direita da densidade de valores extremos, $g_\xi(x) = dGEV_\xi(x)/dx$, comparativamente com a densidade Normal, $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$, com cauda direita muito leve, mesmo quando comparada com a do modelo Gumbel ($\xi = 0$, em GEV_0).

Inicialmente, no artigo de Gnedenko, surgiram 3 distribuições possíveis:

- Tipo I: $\mathcal{A}(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$ [Gumbel],
- Tipo II: $\Phi_a(x) = \exp(-x^a)$, $x > 0, a > 0$ [Fréchet],
- Tipo III: $\Psi_a(x) = \exp(-(-x)^a)$, $x < 0, a > 0$ [Max-Weibull],

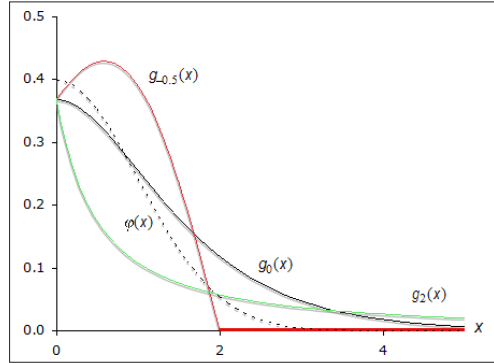


Figura 2
Caudas direitas de $g_\xi(x) = dGEV_\xi(x)/dx$ ($\xi = -0.5, 0, 2$) e da densidade Normal, $\varphi(x)$.

também frequentemente chamadas distribuições de valores extremos (ou EV, do Inglês, “*extreme values*”), associadas respectivamente com $\xi = 0$, $\xi = 1/a > 0$ e $\xi = -1/a < 0$, que podem obviamente ser unificadas na $GEV_\xi = MS_\xi$.

Mais geral do que a classe de modelos max-estáveis, podemos considerar a classe dos modelos max-semi-estáveis (MSS, do Inglês “*max-semi-stable*”), introduzida em Grienvich (1992a, 1992b), Pancheva (1992), e amplamente estudada em Canto e Castro *et al.* (2001) e em Temido & Canto e Castro (2003). A forma funcional das leis MSS é:

$$MSS_{\xi,\nu}(x) = \exp [-\nu \{ \ln(1+\xi x) / \xi \} (1+\xi x)^{-1/\xi}], \quad 1+\xi x > 0, \quad \xi \in \mathbb{R},$$

onde $\nu(\cdot)$ é uma função positiva, limitada e periódica, sendo $MS_\xi = MSS_{\xi,1}$.

Para excessos acima de um nível elevado, é sensato trabalhar com a distribuição generalizada de Pareto-MSS:

$$GP_{\xi,\nu}(x) = 1 + \ln MSS_{\xi,\nu}(x), \quad 1+\xi x > 0, \quad x \geq 0.$$

Os modelos MSS parecem ser interessantes para modelar algumas das variáveis relativas a tremores de terra, tal como sugerido em Sornette (1998). Temos no entanto dificuldades adicionais com a estimação dos parâmetros desconhecidos (veja-se Canto e Castro *et al.*, 2000, 2011, Canto e Castro & Dias, 2011), um ponto a favor dos modelos $GEV = MS$.

É perfeitamente natural perguntar qual o porquê da EVT. Reportando-nos unicamente à área de Extremos e Ambiente, damos em seguida alguns exemplos de grande relevância para a sociedade, e que envolvem ou podem envolver esta teoria. Dois destes exemplos, o segundo e o terceiro, foram retirados de Beirlant *et al.* (2004) e Gomes *et al.* (2013).

1. *O Terramoto de Lisboa, 1 de Novembro de 1755.* Um sismo de magnitude superior a 8.5, com epicentro a cerca de 240 quilómetros da capital portuguesa, criou um tsunami que, em cerca de 40 minutos, devastou a cidade, tendo provocado certamente mais de 10 mil mortos.

2. *As cheias no Mar do Norte, 1 de Fevereiro de 1953.* O nível das águas excedeu os 5.6 metros acima do nível do mar, destruiu as defesas marítimas, tendo inundado áreas na Holanda, Inglaterra, Bélgica, Dinamarca, França, e cerca de 2500 pessoas morreram.

3. *O furacão Katrina, 29 de Agosto de 2005.* A inundação provocada pelo Katrina deveu-se, sobretudo, a uma brecha de 60 metros num dique junto ao lago Pontchartrain, e provocou cerca de 2000 mortos.

4. *Terramoto no centro de Itália, 30 de Outubro de 2016.* Um poderoso tremor de terra abalou recentemente o centro da Itália, região que, apenas 4 dias antes, já havia sido castigada por uma série de tremores de terra. Contudo, o novo sismo causou danos, mas NÃO deixou mortos, inclusive em cidades que, em Agosto de 2016, foram destruídas por um tremor de terra que matou várias pessoas.

Reportamos, em seguida, de forma livre, parte de uma notícia do *New York Times*, Setembro 2005, intitulada '*New Orleans After Hurricane Katrina: An*



Figura 3
Imagem histórica da revista *Life*: Ilustração do terramoto de Lisboa de 1755.



Figura 4
Cheias no Mar do Norte.



Figura 5
Nova Orleães após o furacão Katrina.

Unnatural Disaster? Dizia o redator que teriam de construir um sistema de diques adequado, para o que necessitariam de engenheiros holandeses, capazes de desenhar essas estruturas... Tratar-se-ia de um plano que custaria biliões, mas seria sensato que se aprendesse a lição, de modo a NÃO se ter uma repetição do número de mortos numa catástrofe dentro dos próximos 20 anos.

Estes desastres deveriam realmente funcionar como um guia. E na realidade, como resultado das cheias do Mar do Norte às primeiras horas da manhã de 1 de Fevereiro de 1953, o governo holandês constituiu uma comissão (*'Delta Committee'*). E decretou que os diques deveriam ser construídos com uma altura tal que *'a probabilidade de uma inundação num determinado ano fosse de 1 em 10.000'*. No entanto, 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. Esse mesmo tipo de extrapolação é necessária relativamente a sismos que ocorrem em locais específicos, tal como o que ocorreu recentemente no centro de Itália. Mas na nossa opinião muito há ainda a fazer sobre este assunto...

Quanto à distribuição de tremores de terra no espaço, tempo e magnitude, acrescentamos ainda que durante todo o século passado, os sismologistas têm observado e localizado milhões de tremores de terra em todo o mundo. Com base nestas observações, que podemos dizer sobre a distribuição dos tremores de terra em espaço, tempo e magnitude? Existirão modelos estatísticos que descrevam de forma fiável algumas das medições relacionadas com sismos? E, por fim, podem esses modelos ser usados no futuro para fazer previsões (probabilísticas) ou mesmo predições de futuros sismos? Serão estas algumas das questões a que é preciso responder. Apresentamos, na **Figura 7**, um mapa global de sismos no período 1975-1979, colorido de acordo com a profundidade sísmica, cuja fonte é o "US Geological Service, NEIC (*National Earthquake Information Center*)".

Qualquer mapa sísmico, como o da **Figura 7** mostra-nos que os sismos ocorrem em grupos. Contudo, existem ocasionalmente sismos que ocorrem em lugares onde tal nunca aconteceu. Isso deve-se ao facto de só existirem registos desde há aproximadamente 100 anos, o que não é suficiente para obter a distribuição espacial das zonas sísmicas com baixa intensidade de sismos. Mas o maior desafio é não a identificação de regiões possivelmente sísmicas, mas sim prever qual a frequência e magnitude de sismos de relevo em determinada zona. Um estudo de dados deste tipo foi efectuado em Pisarenko &



Figura 6
Destruição em L'Aquila, cidade que tinha sido em 2009 castigada por um terramoto que matou mais de 300 pessoas.

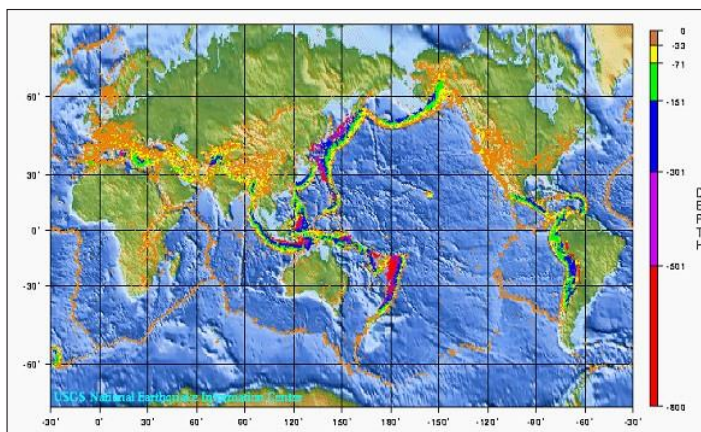


Figura 7
Mapa da profundidade de sismos entre 1975 e 1979.

Sornette (2003), onde são comparadas as caudas direitas das distribuições dos momentos sísmicos, em áreas de subducção e oceânicas. Ambas as caudas revelam um peso altamente positivo, i.e. $\xi > 0$, em GEV_{ξ} (veja-se também Beirlant *et al.*, 2004, 2016, e Gomes *et al.*, 2013).

Na realidade, a EVT consegue dar respostas fidedignas sobre a altura da referida barragem, e poderá também avançar com previsões sobre futuros sismos, entrando em linha de conta com aquilo a que chamamos período de retorno de um acontecimento extremo, que não é mais do que o intervalo de tempo médio entre ocorrências de um determinado acontecimento extremo, como o terramoto de Lisboa ou o furacão Katrina ou a cheia no Mar do Norte ou o recente terramoto no centro de Itália. Face às réplicas frequentes associadas a um tremor de terra, teremos ainda de dar atenção especial à estimação de um outro parâmetro de acontecimentos raros, o índice extremal, que traduz uma medida da dependência em grupos de valores elevados, e que pode frequentemente ser interpretado como o recíproco da dimensão média desses grupos de observações (veja-se Leadbetter *et al.*, 1983; Gomes, 1993b,c, 1995,a,b, 2015; Gomes *et al.*, 2008, 2015; Neves *et al.*, 2015, entre outros). Mas o controlo de tremores de terra é no nosso entender de extrema dificuldade, e requer um esforço multidisciplinar, que pensamos não ter sido totalmente conseguido até à data, particularmente quando tentamos abordar o carácter espacial e temporal do processo de tremores de terra. Gostaria no entanto de referir um trabalho de mestrado recente (Rosário, 2013), relacionado com dados também analisados em Beirlant *et al.* (2004, 2016).

As principais questões a ter em consideração são essencialmente as seguintes: Usualmente existem poucas observações na cauda da distribuição, e são requeridas estimativas muito para além do máximo observado. Necessitamos pois de recorrer a modelos para a cauda, usualmente baseados em resultados assintóticos. Será sensato usar esses modelos em todas as situações reais envolvendo acontecimentos raros? É preciso não esquecer, parafraseando George Box (1919-2013), genro de Sir Ronald Fisher, ‘... *all models are wrong but some models are useful*’ (Box & Draper, 1987, p. 424).

Não podemos deixar de referir três das frases célebres de Emil Gumbel, ‘*Il est impossible que l’improbable n’arrive jamais*’, ‘*Il y aura toujours une valeur qui dépassera toutes les autres*’ e ‘*It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis*’.

E a esta última frase, atrevemo-nos a acrescentar: ‘*Não só os rios, mas também os movimentos da crosta terrestre conhecem a teoria de valores extremos...*’ Detalhes sobre a *Estatística de Extremos* podem ser vistos nas recentes recensões críticas em Beirlant *et al.* (2012), Scarrott & MacDonald (2012) e Gomes & Guillou (2015). Em Beirlant *et al.* (2004) e Gomes *et al.* (2013), entre outros livros, são tratados diversos estudos de casos, num leque variado de áreas de aplicação de modelação de acontecimentos raros. Esperamos ter aguçado o vosso apetite por um tema relativamente recente em termos históricos, e com tantas áreas de aplicação quantas as que possamos conceber.

AGRADECIMENTOS

Investigação parcialmente financiada através de fundos nacionais, FCT – Fundação para a Ciência e a Tecnologia, projecto UID/MAT/00006/2013 (CEA/UL).

(Comunicação apresentada no debate
Terramoto 1755: o dia seguinte, no dia 2 de novembro de 2016)

REFERÊNCIAS

- [1] Beirlant, J., Goegebeur, Y., Segers, J. & J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley, England.
- [2] Beirlant, J., Caeiro, F. & M.I. Gomes (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* **10**:1, 1-31.
- [3] Beirlant, J., Fraga Alves, M.I. & M.I. Gomes (2016). Tail fitting for truncated and non-truncated Pareto-type distributions, *Extremes* **19**:3, 429-462.
- [4] Box, G.E.P. & N.R. Draper (1987). *Empirical Model-Building and Response Surfaces*. Wiley.
- [5] Canto e Castro, L. & S. Dias (2011). Generalized Pickands' estimators for the tail index parameter and max-semistability, *Extremes* **14**:4, 429-449.
- [6] Canto e Castro, L., Temido, G. & M.I. Gomes (2000). Inferência estatística em modelos max-semiestáveis. Em P. Oliveira & E. Athayde, eds., *Um Olhar sobre a Estatística*, 291-305, Edições S.P.E.
- [7] Canto e Castro, L., Haan, L. de & M.G. Temido (2001). Rarely observed maxima, *Th. Prob. Appl.* **45**, 658-662.
- [8] Canto e Castro, L., Dias, S. & M.G. Temido (2011). Looking for max-semistability: a new test for the extreme value condition, *Journal of Statistical Planning and Inference* **141**, 3005-3020.
- [9] Castillo, E., Hadi, A., Balakrishnan, N. & J.M. Sarabia (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, Hoboken, New Jersey.
- [10] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**:6, 423-453.
- [11] Gomes, M.I. (1993a). A obra científica de J. Tiago de Oliveira. In D. Pestana (ed.), *Estatística Robusta, Extremos e Mais Alguns Temas*, 241-248, Edições Salamandra.
- [12] Gomes, M.I. (1993b). Modelos extremos em esquemas de dependência. In D. Pestana (ed.), *Estatística Robusta, Extremos e Mais Alguns Temas*, pp. 209-220, Edições Salamandra.
- [13] Gomes, M.I. (1993c). On the estimation of parameters of rare events in environmental time series. In V. Barnett & K.F. Turkman (eds.), *Statistics for the Environment*, pp. 225-241, Wiley, New York.
- [14] Gomes, M.I. (1995a). Metodologias jackknife e bootstrap em Estatística de Extremos. *Actas da II Conferência Anual da S.P.E.*, 31-46.
- [15] Gomes, M.I. (1995b). The influence of the extremal index on the estimation of return periods of high levels. In Muirchearthaigh I.O. et al. (eds.), *Proceedings of the 61st International Meeting on Statistical Climatology*, pp. 299-302, University College, Galway.
- [16] Gomes, M.I. (2015). NAZARÉ and ARCH processes: extremal index estimation. In E. Gonçalves et al. (eds.), *Contributions in Statistics and Inference: Celebrating Nazaré Mendes Lopes' Birthday*. Textos de Matemática, N.º 47, pp. 1-12, Departamento de Matemática, Faculdade de Ciências e Tecnologia da Universidade de Coimbra.
- [17] Gomes, M.I. & A. Guillou (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review* **83**:2, 263-292.
- [18] Gomes, M.I., Hall, A. & C. Miranda (2008). Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. *J. Comput. Statist. and Data Analysis* **52**:4, 2022-2041.
- [19] Gomes, M.I., Fraga Alves, M.I. & C. Neves (2013). *Análise de Valores Extremos: uma Introdução*. Edições S.P.E. & I.N.E.
- [20] Gomes, M.I., Figueiredo, F., Martins, M.J. & M.M. Neves (2015). Resampling methodologies and reliable tail estimation. *South African Statistical Journal* **49**, 1-20.
- [21] Grienvich, I.V. (1992a). Max-semistable laws corresponding to linear and power normalizations, *Th. Probab. Appl.* **37**, 720-721.
- [22] Grienvich, I.V. (1992b). Domains of attraction of max-semistable laws under linear and power normalizations, *Th. Probab. Appl.* **38**, 640-650.
- [23] Gumbel, E.J. (1958; 2004). *Statistics of Extremes*. Columbia University Press, New York.
- [24] Leadbetter, M.R., Lindgren, G. & H. Rootzén (1983). *Extremal Theory for Stochastic Processes*. Springer-Verlag, New York.
- [25] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-tailed Data*, John Wiley & Sons, England.

- [26] Neves, M.M., Gomes, M.I., Figueiredo, F. & D. Prata-Gomes (2015). Modeling extreme events: sample fraction adaptive choice in parameter estimation. *Journal of Statistical Theory and Practice* **9**:1, 184-199.
- [27] Pancheva, E. (1992). Multivariate max-semistable distributions, *Th. Probab. and Appl.* **37**, 731-732.
- [28] Pisarenko, V.F. & D. Sornette (2003). Characterization of the frequency of extreme events by the generalized Pareto distribution, *Pure and Applied Geophysics* **160**, 2343-2364.
- [29] Reiss, R.-D. & M. Thomas (2001; 2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*, 2nd edition; 3rd edition, Birkhauser Verlag.
- [30] Rosário, P. (2013). *Valores Extremos em Sismologia—Caso Estudo*, Mestrado em Estatística e Investigação Operacional, DEIO, FCUL.
- [31] Scarrot, C. & A. MacDonald (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revstat* **10**:1, 33-60.
- [32] Sornette, D. (1998). Discrete scale invariance and complex dimensions, *Physics Reports* **297**, 239-270.
- [33] Temido, M.G. & L. Canto e Castro (2003). Max-semistable laws in extremes of stationary random sequences, *Theory Probab. Appl.* **47**:2, 365-374.
- [34] Tiago de Oliveira, J.C., ed. (1993). *J. Tiago de Oliveira: O Homem e a Obra*. Coleção Grandes Mestres, Edições Colibri.