Large Quantile Estimation for Distributions in the Domain of Attraction of a Max-Semistable Law

Luísa Canto e Castro[1]   Sandra Dias[2]

[1]Department of Statistics
University of Lisbon, Portugal

[2]Department of Mathematics
University of Trás-os-Montes e Alto Douro, Portugal

4th Conference on Extreme Value Analysis
Gothenburg, August 15-19, 2005

Introduction

- F is in the domain of attraction of a max-stable d.f. G, $F \in MS(G)$, if and only if there exist $a_n > 0$ and b_n such that

$$\lim_{n \to +\infty} F^n(a_n x + b_n) = G(x). \tag{1}$$

- F is in the domain of attraction of a max-semistable d.f. G, $F \in MSS(G)$, if and only if there exist $a_n > 0$ and b_n such that

$$\lim_{n \to +\infty} F^{k_n}(a_n x + b_n) = G(x), \ \ \forall\, x \in C_G \tag{2}$$

  where k_n verifies

$$\lim_{n \to +\infty} k_{n+1}/k_n = r \geq 1 \ \ (r < \infty). \tag{3}$$

Remark: Max-stable laws are a particular case of the max-semistable laws when $r = 1$.

Introduction (continuation)

Pioneers in max-semistable laws:

- Grinevich (1992, 1993)
- Pancheva (1992)

(Grinevich 1992) G is a max-semistable d.f. if and only if there exist $r > 1$, $a > 0$ and $b \in \mathbb{R}$ such that G is solution of the functional equation

$$G(x) = G^r(ax + b). \tag{4}$$

First characterization (Grinevich 1992)

Unifying standard expressions for max-semistable d.f.'s

- For $\gamma \neq 0$

  $$G_{\gamma,\nu}(x) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\nu(\log(1+\gamma x))\right\},\ x \in \mathbb{R},\ 1+\gamma x > 0$$

  $a \neq 1$ and $p = \log\ a = \gamma \log r$

- For $\gamma = 0$

  $$G_{0,\nu}(x) = \exp\{-e^{-x}\nu(x)\},\ x \in \mathbb{R}$$

  $a = 1$ and $p = b = \log\ r$

where ν is a positive, bounded and periodic function with period p.

Some graphical features of max-semistable laws

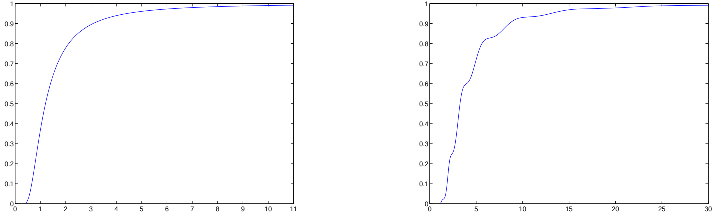


Figure: D.f.'s $G_{0.5}(x) = \exp(-x^{-2})$ and $G_{0.5,\nu}(x) = \exp(-x^{-2}(8 + \cos(4\pi \log x)))$

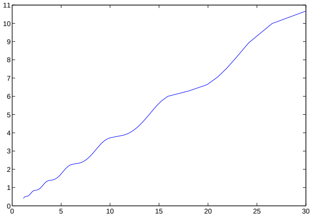


Figure: QQ-Plot of $G_{0.5,\nu}$ against $G_{0.5}$

Second characterization (Canto e Castro *et al.* 2000)

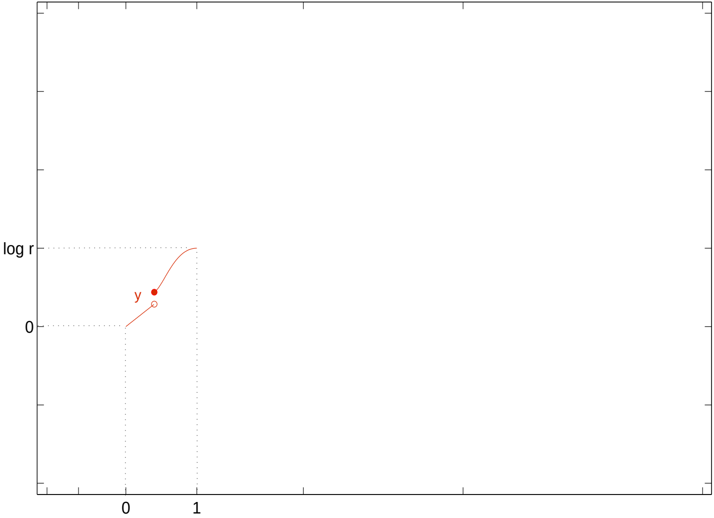
General to location and scale, for any d.f. G max-semistable we have:

$$-\log(-\log\ G(s_m + a^m x)) = m \log\ r + y(x),\ \forall\, x \in [0,1],\ m \in \mathbb{Z} \tag{5}$$

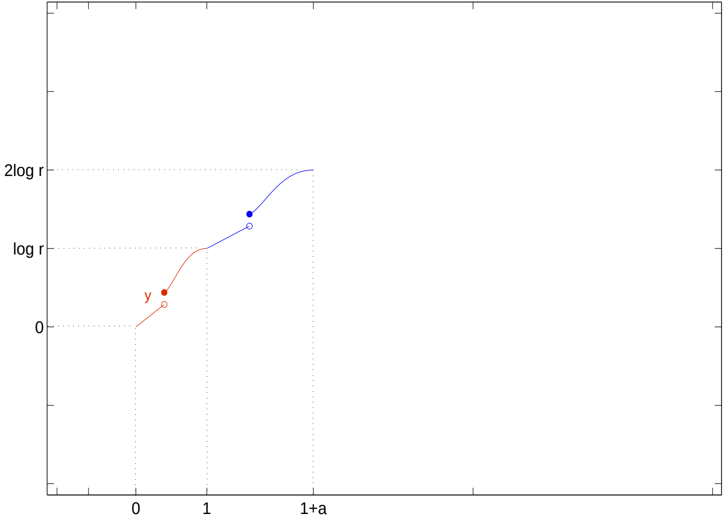
where

- $y : [0,1] \to [0, \log\ r]$ is non decreasing, right continuous and continuous at $x = 1$
- $s_m = m$ if $a = 1$
- $s_m = (a^m - 1)/(a-1)$ if $a \neq 1$ and $a > 0$

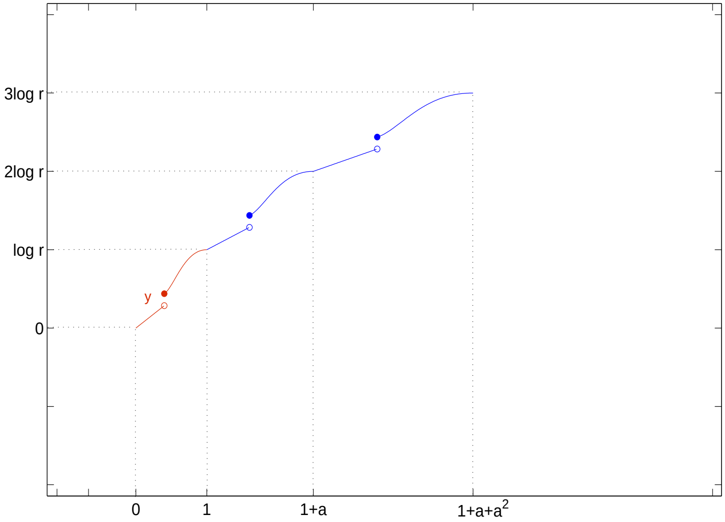
Graph of the function $-\log(-\log(G))$



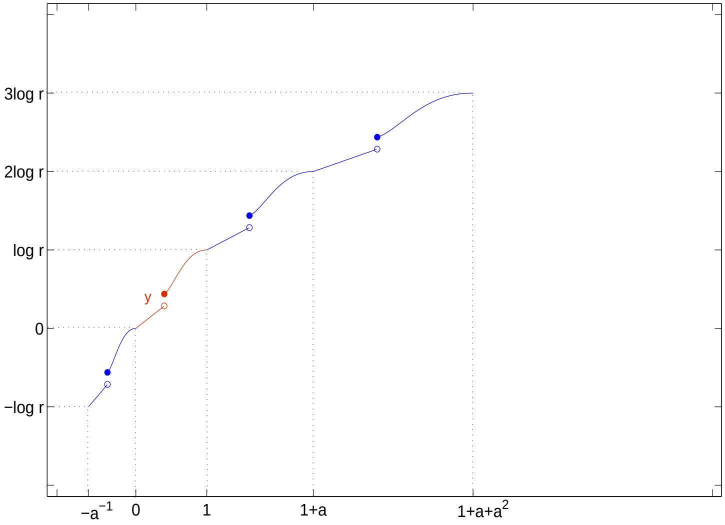
Graph of the function $-\log(-\log(G))$



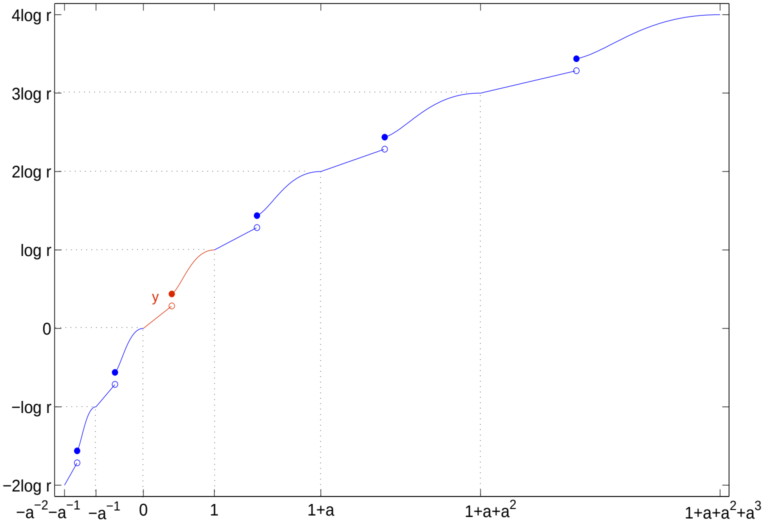
Graph of the function $-\log(-\log(G))$



Graph of the function $-\log(-\log(G))$



Graph of the function $-\log(-\log(G))$



Statistical inference in max-semistable models

Temido (2000) proposed that, in the estimation of the parameters r, p and γ, appropriated functions of the following sequence of statistics should be used

$$Z_s(m) := \frac{X_{(m/s^2)} - X_{(m/s)}}{X_{(m/s)} - X_{(m)}}$$

where

- $X_{(m)} := X_{N-[m]+1:N}$ are order statistics of a sample of size N from the random variable X
- $m := m_N$ is an intermediate sequence (that is, m_N is an integer sequence verifying $m_N \to +\infty$ and $m_N/N \to 0$)

Behaviour of the sequence of statistics Z_s

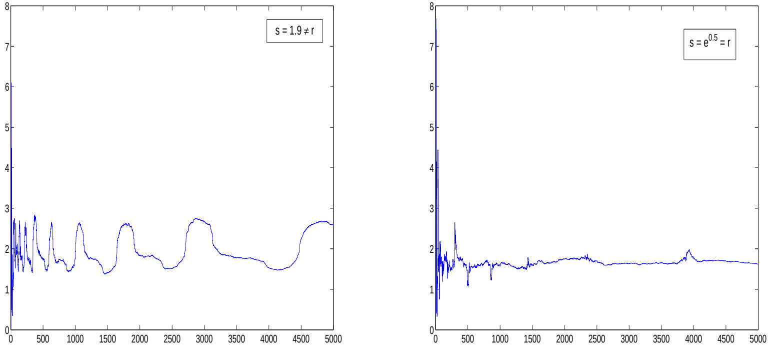


Figure: Sample trajectories of $Z_s(m)$ for $s = 1.9$ and $s = e^{0.5} \approx 1.65$

$$F(x) = 1 - x^{-1}(14 + \cos(4\pi \log x)), \quad r = e^{0.5}$$

Behaviour of the sequence of statistics Z_s (continuation)

Dias and Canto e Castro (2004) proved that

$$Z_s(m) \underset{n\to+\infty}{\overset{P}{\longrightarrow}} a^c \text{ if and only if } s = r^c,\ c \in \mathbb{N}.$$

Furthermore, if $s \neq r^c$ then $Z_s(m)$ as an oscillatory behaviour.

Using these result we have that if $s = r^c$, $c \in \mathbb{N}$, then

- $R_s(m) := \frac{Z_{s^2}(m)}{(Z_s(m))^2} \overset{P}{\longrightarrow} 1, \quad n \to +\infty$
- $\widehat{P}_s(m) := \log\left(Z_s(m)\right) \overset{P}{\longrightarrow} \log\ a^c = cp, \quad n \to +\infty,\ \gamma \neq 0$
- $\widehat{\gamma}_s(m) := \frac{\log(Z_s(m))}{\log\ s} \overset{P}{\longrightarrow} \gamma, \quad n \to +\infty$

Behaviour of the sequence of statistics Z_s (continuation)

Dias and Canto e Castro (2004) proved that

$$Z_s(m) \xrightarrow[n\to+\infty]{P} a^c \quad \text{if and only if } s = r^c,\ c \in \mathbb{N}.$$

Furthermore, if $s \neq r^c$ then $Z_s(m)$ as an oscillatory behaviour.

Using these result we have that if $s = r^c$, $c \in \mathbb{N}$, then

- $R_s(m) := \frac{Z_{s^2}(m)}{(Z_s(m))^2} \xrightarrow{P} 1, \quad n \to +\infty$
- $\widehat{P}_s(m) := \log\left(Z_s(m)\right) \xrightarrow{P} \log\ a^c = cp, \quad n \to +\infty,\ \gamma \neq 0$
- $\widehat{\gamma}_s(m) := \frac{\log(Z_s(m))}{\log\ s} \xrightarrow{P} \gamma, \quad n \to +\infty$

Proposed estimators for the parameters r, p and γ

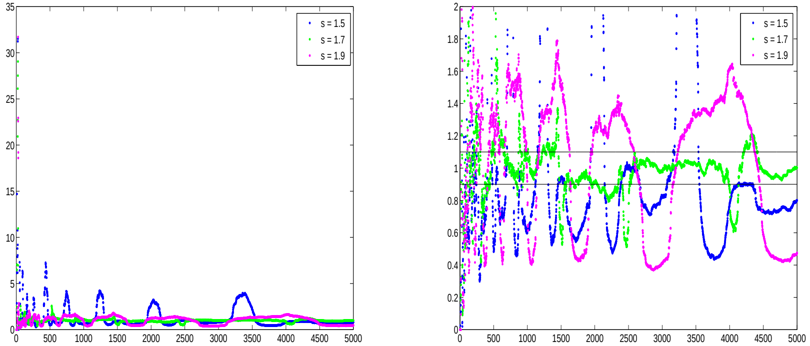


Figure: Left: Sample trajectories of $R_s(m)$ for $s = 1.5$, $s = 1.7$ and $s = 1.9$
Right: Magnified version

$$F(x) = 1 - x^{-1}(14 + \cos(4\pi \log x)), \quad r = e^{0.5} \approx 1.65$$

Proposed estimators for the parameters r, p and γ

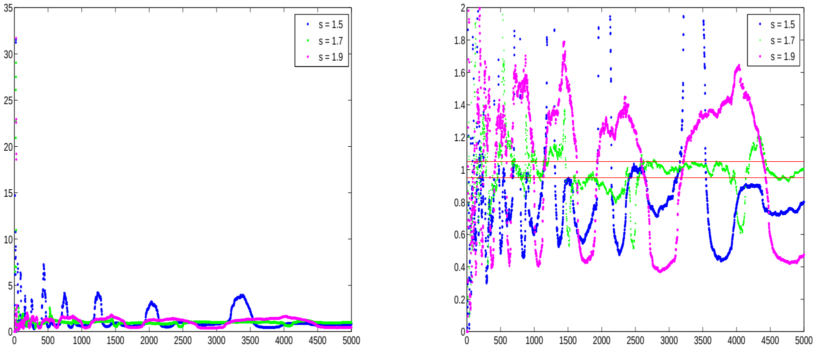


Figure: Left: Sample trajectories of $R_s(m)$ for $s = 1.5$, $s = 1.7$ and $s = 1.9$
Right: Magnified version

$$F(x) = 1 - x^{-1}(14 + \cos(4\pi \log x)), \quad r = e^{0.5} \approx 1.65$$

Proposed estimators for the parameters r, p and γ
(continuation)

- An estimate of r

$$\widehat{r} = \text{mode}\left\{ \underset{s=1.1,(0.1),3.0}{\arg\max}\ B_s(\epsilon), \epsilon = 0.01, (0.01), 0.1 \right\}$$

where

- $B_s(\epsilon) := \frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{\{m^{(i)} \in A_n : |R_s(m^{(i)})-1|<\epsilon\}}(m^{(i)})$ (percentage of time that the sequence of statistics R_s spends in a ϵ-neighborhood of 1)
- A_n is a set of suitable values of m $\quad A_n = \{m^{(1)}, m^{(2)}, ..., m^{(k)}\}$
- $k = \# A_n$

Proposed estimators for the parameters r, p and γ (continuation)

- An estimate of p
  - if $\gamma \neq 0$

$$\widehat{p} = \frac{1}{k} \sum_{m \in A_n} \widehat{P}_{\widehat{r}}(m) = \frac{1}{k} \sum_{m \in A_n} \log(Z_{\widehat{r}}(m))$$

  - if $\gamma = 0$

$$\widehat{p} = \log \widehat{r}$$

- An estimate of γ

$$\widehat{\gamma} = \widehat{p} / \log \widehat{r} \quad (\gamma \neq 0)$$

Simulation study

Selected d.f.'s

$$1 - F(x) = (1 - F_0(x))\theta(x)$$

where

- $F_0 \in MS(G)$ for some max-stable d.f. G. In particular
  - Generalized Pareto d.f.

$$F_0(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma} & x \in \mathbb{R},\ 1 + \gamma x > 0 \text{ and } \gamma \neq 0 \\ 1 - e^{-x} & x \in \mathbb{R} \text{ and } \gamma = 0 \end{cases}$$

  - Burr d.f.

$$F_0(x) = 1 - (1 + x^{-\rho/\gamma})^{-1/\rho},\ x \geq 0,\ \rho < 0 \text{ and } \gamma > 0$$

- θ positive and periodic function

Simulation study (continuation)

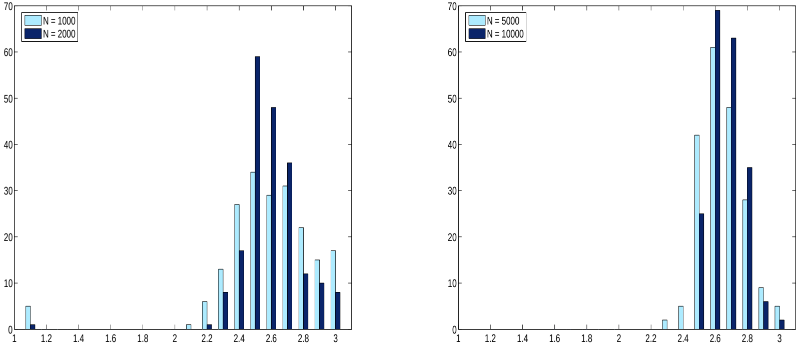


Figure: Empirical distribution of the estimates of r

$$F(x) = 1 - x^{-2}(8 + \cos(4\pi \log x)), \quad r = e \approx 2.71$$

Simulation study (continuation)

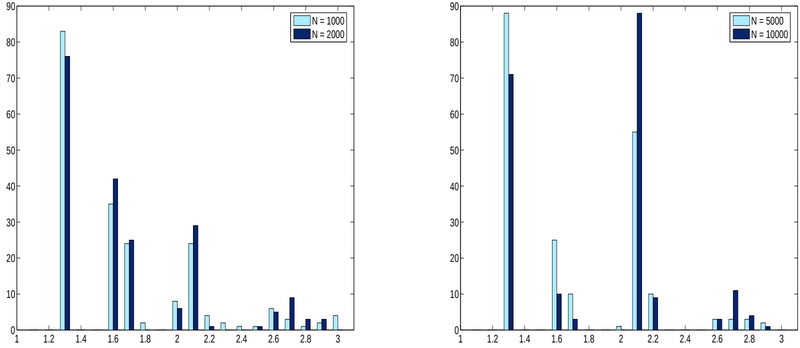


Figure: Empirical distribution of the estimates of r

$$F(x) = 1 - x^{-1}(27 + \cos(8\pi \log x)),$$

$$r = e^{0.25} \approx 1.28,\ r^2 \approx 1.65,\ r^3 \approx 2.12 \text{ and } r^4 \approx 2.72$$

Simulation study (continuation)

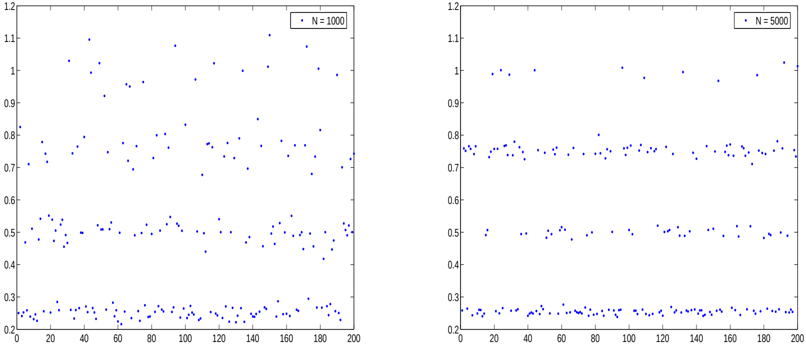


Figure: Estimates of the period p for each replica (sample sizes 1000 and 5000)

$$F(x) = 1 - x^{-1}(27 + \cos(8\pi \log x)), \quad p = 0.25$$

Results from the simulation study

- the root global mean square error, RGMSE

$$\text{RGMSE} = \sqrt{\frac{n_1}{n}\frac{\text{MSE}_1}{(\theta^{(1)})^2} + \frac{n_2}{n}\frac{\text{MSE}_2}{(\theta^{(2)})^2} + \ldots}$$

$$\text{MSE}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(\widehat{\theta}_i^{(j)} - \theta^{(j)}\right)^2$$

- the global bias, Gbias

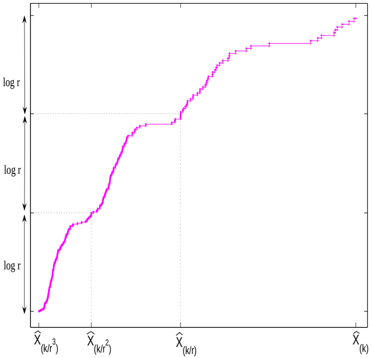
$$\text{Gbias} = \frac{n_1}{n}\frac{\text{bias}_1}{\theta^{(1)}} + \frac{n_2}{n}\frac{\text{bias}_2}{\theta^{(2)}} + \ldots \qquad \text{bias}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\widehat{\theta}_i^{(j)} - \theta^{(j)}$$

where $\widehat{\theta}_i^{(j)}$ are the estimates of θ which are nearer $\theta^{(j)}$ than from any other $\theta^{(k)}$, $k \neq j$ and n_j is the number of values $\widehat{\theta}_i^{(j)}$, $j = 1, 2, \ldots$

Remark: In the estimation of r: $\theta^{(j)} = r^j$. In the estimation of $p:\ \theta^{(j)} = jp$.

- Estimates of r
  - GRMSE
    - less than 0.13
    - decreases when n increases
    - slightly less in the Generalized Pareto models
    - increases when r increases
  - Gbias
    - less than 0.07 in terms of absolute value
    - does not have a monotonous behaviour with n
- Estimates of p
  - GRMSE
    - less than 0.18
    - decreases when n increases
    - slightly less in the Generalized Pareto models
    - increases when r increases
  - Gbias
    - less than 0.10 in terms of absolute value
    - does not have a monotonous behaviour with n
- Estimates of γ
  - RMSE/γ
    - less than 0.14
    - decreases when n increases
  - bias/γ
    - less than 0.09 in terms of absolute value
    - decreases when n increases
    - usually preserves the signal

Estimation of the function y



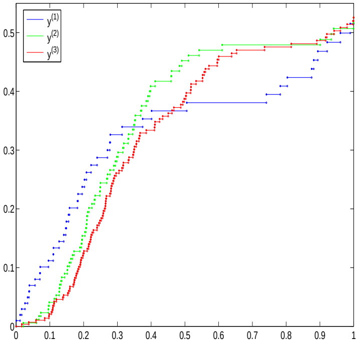
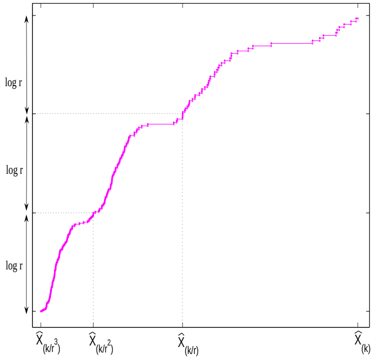


Figure: Left: Empirical function $-log(-log(\widehat{F}_n))$ Right: Empirical versions of y

$$F(x) = 1 - x^{-1}(14 + \cos(4\pi \log x))$$

Estimation of the function y



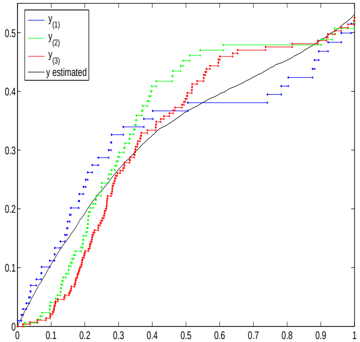


Figure: Left: Empirical function $-log(-log(\widehat{F}_n))$ Right: Empirical versions of y

$$F(x) = 1 - x^{-1}(14 + \cos(4\pi \log x))$$

Estimation of the function y (continuation)

$$\widehat{y}(x) = \sum_i \frac{n_i}{n_t}\, y^{(i)}(x), \quad x \in [0,1]$$

where

- $y^{(i)}(x) = \sum_j y_j^{(i)}\, \mathbb{1}_{[x_j^{(i)}, x_{j+1}^{(i)}[}(x)$ is the ith empirical version of y
- $x_j^{(i)}$ are the jump points of the ith empirical version of y
- $y_j^{(i)}$ are the correspondent points of $x_j^{(i)}$
- n_i is the number of points $x_j^{(i)}$ in the ith empirical version of y
- n_t is the total number of points $x_j^{(i)}$

Large quantile estimation

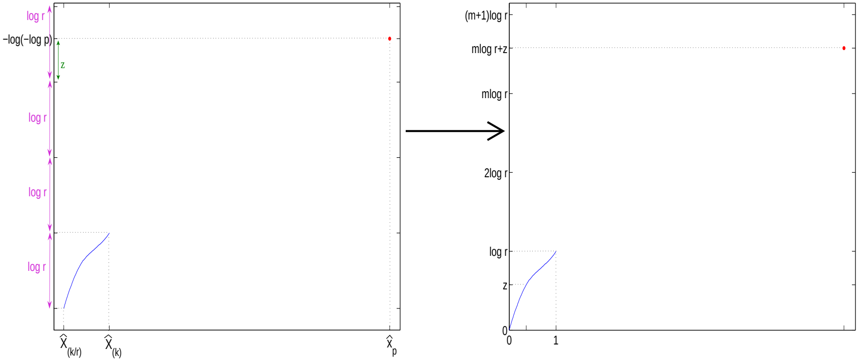


Figure: Empirical function $-log(-log(\widehat{G}))$

Large quantile estimation

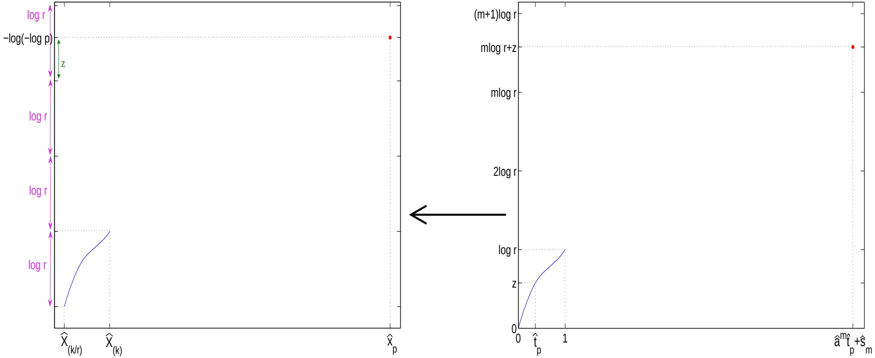


Figure: Empirical function $-log(-log(\widehat{G}))$

Large quantile estimation (continuation)

$$\widehat{x}_p = (\widehat{a}^m \widehat{t}_p + \widehat{s}_m)(\widehat{X}_{(k/r)} - \widehat{X}_{(k)}) + \widehat{X}_{(k)}$$

where

- $m \in \mathbb{N}$ such that

  $z = -\log(-\log p) - (-\log(-\log \widehat{G}(\widehat{X}_{(k/r)}))) - m \log r \in [0, 1]$

- $\widehat{t}_p = \widehat{y}^{-1}(z)$

Application to real data

Data: Major earthquake inter-arrival times registered in the period between January 1st, 1973 and March 31st, 2005. (A earthquake is a Major earthquake if its magnitude is greater or equal to 6.5.)

Sample Size: 1450 values

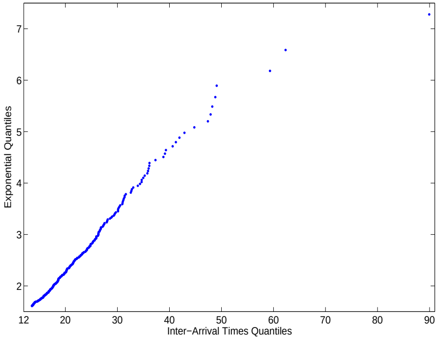


Figure: QQ-Plot of the inter-arrival times against the exponential (290 highest values)

Modeling with MSS laws

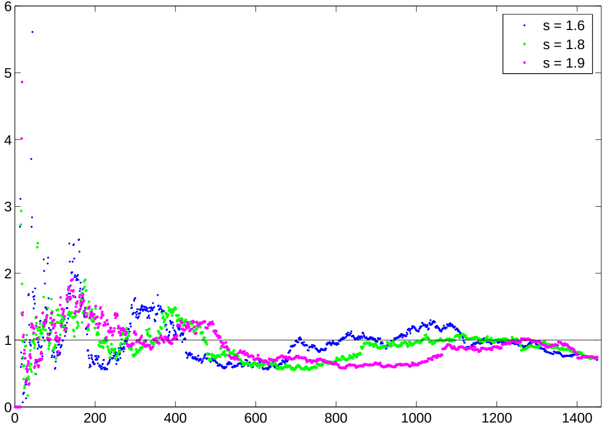


Figure: Sample trajectories of $R_s(m)$ for $s = 1.6$, $s = 1.8$ and $s = 1.9$

Estimation of the parameters

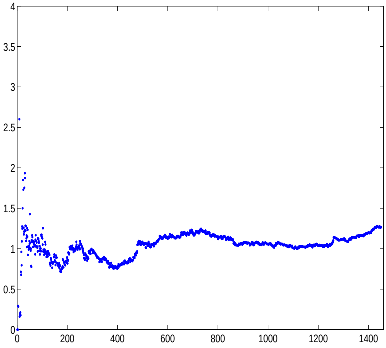


Figure: Sample trajectory of $Z_{\widehat{r}}(m)$, $\widehat{r} = 1.8$

Estimates for the parameters

Parameters	Estimates
r	1.8
a	1.000443
γ	0.000754

Estimated function y

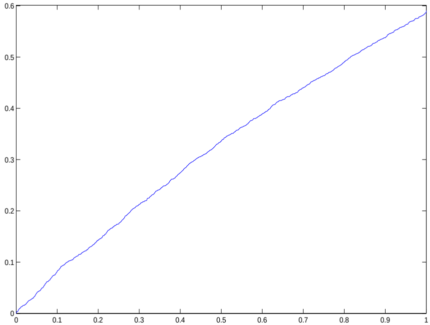


Figure: Estimated function y

Empirical functions $-\log(-\log \widehat{G})$ and $\widehat{G}$

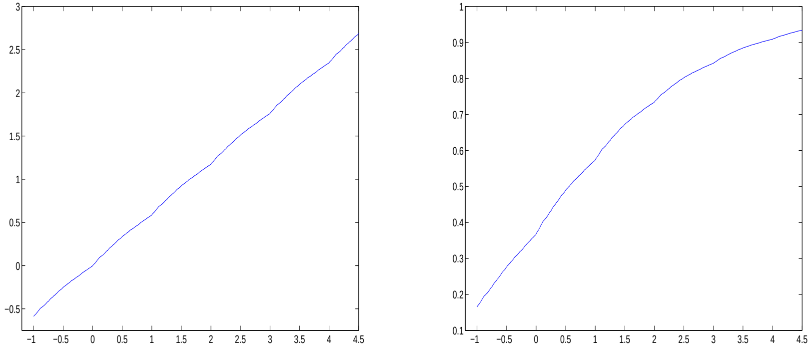


Figure: Left: Empirical function $-\log(-\log \widehat{G}))$ Right: Empirical function $\widehat{G}$

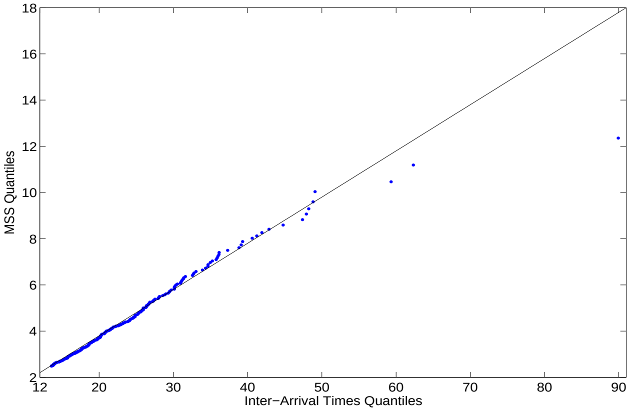


Figure: QQ-Plot of the inter-arrival times against the estimated max-semistable (290 highest values)

References

Canto e Castro, L., de Haan, L., and Temido, M. G. (2000).
Rarely observed sample maxima.
Theory of Probab. Appl., Vol. 45, p. 779-782.

Dias, S., e Canto e Castro, L. (2004).
Contribuições para a estimação em modelos max-semiestáveis.
In *Estatística com acaso e necessidade*, p. 177-187. Edições SPE, Lisboa.

Dias, S., e Canto e Castro, L. (2005).
Modelação de máximos em domínios de atracção max-semiestáveis.
In *Estatística Jubilar*, p. 223-234. Edições SPE, Lisboa.

Grinevich, I.V. (1992).
Max-semistable laws corresponding to linear and power normalizations.
Theory of Probab. Appl., Vol. 37, p. 720-721.

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

Pancheva, Z. (1992).
Multivariate max-semistable distributions.
Theory of Probab. Appl., Vol. 37, p. 731-732.

Temido, M.G. (2000)
Classes de Leis Limite em Teoria de Valores Extremos - Estabilidade e Semiestabilidade.
Ph.D. Thesis, University of Coimbra.