Estimation of the tail-index and extreme quantiles from a mixture of heavy-tailed distributions

Stéphane GIRARD[1] (Inria Grenoble Rhône-Alpes),
joint work with Emmanuel GOBET (CMAP, Ecole Polytechnique)

Stress Test
RISK Management and Financial Steering

May 2021

Risk
Univ. Grenoble Alpes

[1] Thanks to Michael Allouche and Jonathan Elmethni for their help with the numerical experiments.

Outline

1. Extreme-value theory for heavy-tailed distributions
2. Mixture of heavy-tailed distributions

1. Extreme-value theory for heavy-tailed distributions

A distribution is said to be **heavy-tailed** if the survival function is **regularly-varying** *i.e.*

$$\bar{F}(x) = x^{-1/\gamma}\ell(x),\ x > 0,$$

where

- $\gamma > 0$ is called the **tail-index** (or extreme-value index). The larger γ is, the heavier the tail: $\gamma > 1/2$ no variance, $\gamma > 1$ no expectation.
- ℓ is a **slowly-varying** function. It is positive and such that

$$\frac{\ell(xt)}{\ell(x)} \to 1 \text{ as } x \to \infty \text{ for all } t > 0.$$

Examples: Pareto (where ℓ is constant), Cauchy (where $\gamma = 1$), Student t_ν (where $\gamma = 1/\nu$), Burr(α, β), (where $\gamma = 1/(\alpha\beta)$), Fisher, ...

Goal: Estimation of **extreme quantiles**.

- Given a sample of iid observations $X_1, \dots, X_n$ from a survival function $\bar{F}$, estimate $q(p_n)$ such that $\bar{F}(q(p_n)) = p_n$ with $np_n \to 0$ as $n \to \infty$.
- Notation: $X_{1,n} \leq \cdots \leq X_{n,n}$ associated order statistics.
- Such extreme quantiles are asymptotically almost surely outside the sample:

$$
\begin{aligned}
\mathbb{P}(q(p_n) > X_{n,n}) &= \mathbb{P}^n(X_1 < q(p_n)) = F^n(q(p_n)) \\
&= (1-p_n)^n = \exp(n \log(1-p_n)) \\
&= \exp(-np_n(1+o(1))) \\
&\to 1
\end{aligned}
$$

  as $n \to \infty$.
- Need for dedicated statistical tools.

Weissman extrapolation device (Weissman, 1978).

- From the regular variation property:

$$\frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-1/\gamma}\frac{\ell(xt)}{\ell(x)} \to t^{-1/\gamma} \text{ as } x \to \infty \text{ for all } t > 0.$$

- Letting $x = q(\alpha_n)$ and $xt = q(p_n)$, an approximation can be derived for extreme quantiles:

$$q(p_n) \simeq q(\alpha_n)\left(\frac{p_n}{\alpha_n}\right)^{-\gamma},$$

  for all $\alpha_n \geq p_n$.

- The idea is then to choose $\alpha_n = k_n/n$ such that $k_n = n\alpha_n \to \infty$ and $q(\alpha_n)$ is then an **intermediate quantile**. It can be estimated by an order statistics $X_{n-k_n,n}$, leading to Weissman estimator:

$$\hat{q}(p_n) = X_{n-k_n,n}\left(\frac{np_n}{k_n}\right)^{-\hat{\gamma}}.$$

  Asymptotic normality: de Haan & Ferreira (2006), Theorem 4.4.6.

- Need for an estimator $\hat{\gamma}$ of the tail-index.

Hill estimator (Hill, 1975).

- Still using the regular variation property: For all $i = 0, \ldots, k_n - 1$,

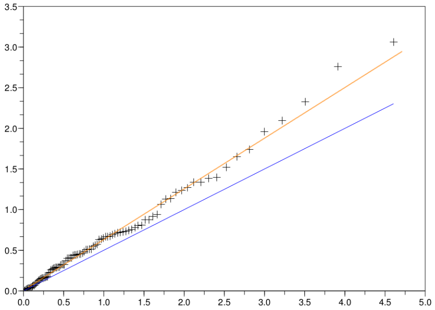
$$\log q(i/n) - \log q(k_n/n) \simeq \gamma \log(k_n/i).$$

- The intermediate quantiles are estimated by their empirical counterparts:

$$\log X_{n-i,n} - \log X_{n-k_n,n} \simeq \gamma \log(k_n/i),$$

- The quality of these approximations can be assessed graphically on a log quantile-quantile plot.

Simulation of a sample of size $n = 500$ from a Student t_2 distribution ($\gamma = 1/2$). Here, $k_n = 100$.



Horizontally: $\log(k_n/i)$. Vertically: $\log X_{n-i,n} - \log X_{n-k_n,n}$ for $i = 0, \ldots, k_n - 1$, $y = \gamma x$ and $y = \hat{\gamma} x$.

Hill estimator (Hill, 1975).

- Still using the regular variation property: For all $i = 0, \ldots, k_n - 1$,

$$\log q(i/n) - \log q(k_n/n) \simeq \gamma \log(k_n/i).$$

- The intermediate quantiles are estimated by their empirical counterparts:

$$\log X_{n-i,n} - \log X_{n-k_n,n} \simeq \gamma \log(k_n/i),$$

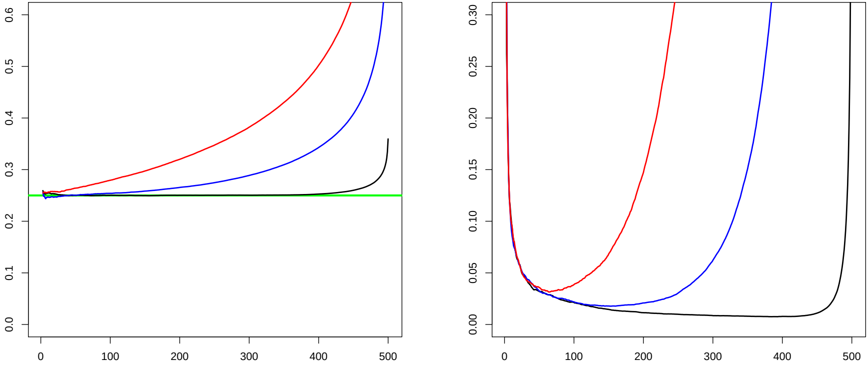
- The quality of these approximations can be assessed graphically on a log quantile-quantile plot.
- The estimator of the slope can be simplified as

$$\hat{\gamma}(k_n) := \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log X_{n-i,n} - \log X_{n-k_n,n}.$$

Illustration on simulated data.

3 Burr distributions, $n = 500$, computation on 1000 replications.

$(\alpha, \beta) = (40, 1/10)$, $(\alpha, \beta) = (8, 1/2)$, $(\alpha, \beta) = (4, 1)$.



Left: Mean Hill estimator as a function of k_n,

Right: MSE of the associated Weissman estimator as a function of k_n.

In all three cases: $\gamma = 1/(\alpha\beta) = 1/4$ but **very different behaviours**.

Second-order condition

(SO) There exist a function B verifying $B(x) \to 0$ as $x \to \infty$ with ultimately constant sign and $\rho \leq 0$ such that, for all $t > 1$,

$$\lim_{x\to\infty} \frac{1}{B(x)} \log\left(\frac{\ell(xt)}{\ell(x)}\right) = K_\rho(t) := \int_1^t u^{\rho-1} du.$$

It can be shown that, necessarily, $|B|$ is regularly varying with index ρ, referred to as the **second-order parameter**. The smaller ρ is, the faster the convergence $\ell(xt)/\ell(x) \to 1$ as $x \to \infty$.

Theorem 1 (Asymptotic normality: de Haan & Ferreira (2006), Theorem 3.2.5)

Suppose (SO) holds. If $k_n \to \infty$ with $k_n/n \to 0$ and $\sqrt{k_n} B(n/k_n) \to \lambda < \infty$, then

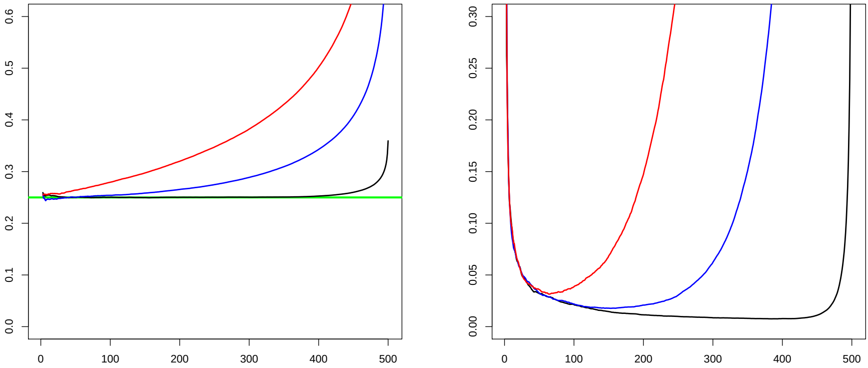
$$\sqrt{k_n}(\hat{\gamma}(k_n) - \gamma) \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2).$$

The asymptotic variance is γ^2/k_n (a decreasing function of k_n) while the asymptotic bias is $B(n/k_n)/(1-\rho)$ (an increasing function of k_n).

Illustration on simulated data.

3 Burr distributions, $n = 500$, computation on 1000 replications.

$(\alpha, \beta) = (40, 1/10)$, $\rho = -10$ $(\alpha, \beta) = (8, 1/2)$, $\rho = -2$, $(\alpha, \beta) = (4, 1)$, $\rho = -1$.

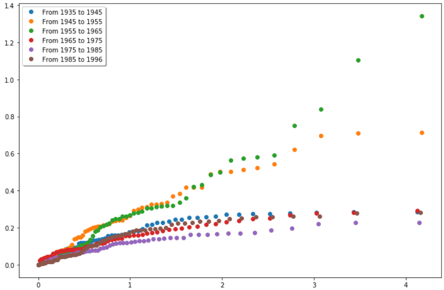


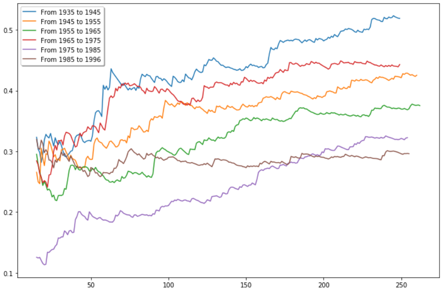
Left: Mean Hill estimator as a function of k_n,
Right: MSE of the associated Weissman estimator as a function of k_n.

In all three cases: $\gamma = 1/(\alpha\beta) = 1/4$, **behaviour mainly driven by** ρ.

Illustration on real data.

Log-returns of the Financial Times Stock Exchange 100 Index (FTSE100), 1935-1996. Share index of the 100 companies listed on the London Stock Exchange with the highest market capitalisation.





Left: log quantile-quantile plots on six different time periods ($k_n/n = 5\%$), Right: corresponding Hill estimators as functions of k_n.

$\Longrightarrow$ High heterogeneity in the tails.

2. Mixture of heavy-tailed distributions

Model. We assume:

- A heavy-tailed distribution for X conditionally on a tail-index $\gamma \in [\gamma_1, \gamma_2]$:

$$\bar{F}(x \mid \gamma) := \mathbb{P}(X \geq x \mid \gamma) = x^{-1/\gamma}\ell(x),\ x > 0,$$

  with $0 < \gamma_1 < \gamma_2 < +\infty$ and where ℓ is a slowly-varying function.

- A prior distribution on $\gamma \in [\gamma_1, \gamma_2]$:

$$\pi(\gamma) = (\gamma_2 - \gamma)^{\beta-1}\ell_\pi(1/(\gamma_2 - \gamma)),$$

  for some $\beta > 0$ where ℓ_π is a slowly-varying function.

Note that $\pi(\cdot)$ is regularly-varying in the neighbourhood of its upper endpoint γ_2.

- $\pi(\gamma) \to \infty$ as $\gamma \to \gamma_2$ when $\beta < 1$,
- $\pi(\gamma) \to 0$ as $\gamma \to \gamma_2$ when $\beta > 1$.
- The intermediate case $\beta = 1$ corresponds to a uniform-like behaviour.

Unconditional tails. The unconditional distribution of X is given by

$$\bar{F}(x) := \mathbb{P}(X \geq x) = \ell(x) \int_{\gamma_1}^{\gamma_2} x^{-1/\gamma} \pi(\gamma) d\gamma,$$

and its tail behaviour is provided in the next result:

Theorem 2

Under the above assumptions,

$$\bar{F}(x) \sim c_1(\gamma_1, \gamma_2, \beta)\, x^{-1/\gamma_2} (\log x)^{-\beta} \ell_\pi(\log x) \ell(x),$$

as $x \to +\infty$, *with* $c_1(\gamma_1, \gamma_2, \beta) = \gamma_2^{2\beta} (\gamma_2 - \gamma_1)^{-\beta}\, \Gamma(\beta)$.

- The unconditional distribution is also heavy-tailed, with tail-index γ_2.
- The associated slowly-varying function is given by

$$\tilde{\ell}(x) \sim c_1(\gamma_1, \gamma_2, \beta) (\log x)^{-\beta} \ell_\pi(\log x) \ell(x).$$

Asymptotic behaviour of Hill estimator on the mixture model.
It is easily seen that in

$$\tilde{\ell}(x) \sim c_1(\gamma_1, \gamma_2, \beta)(\log x)^{-\beta}\ell_\pi(\log x)\ell(x),$$

the orange part is the dominant component. Thus, $\tilde{\ell}(\cdot)$ also verifies (SO) with

$$\tilde{B}(x) = -\frac{\beta\gamma_2}{\log x},\ x > 0,$$

and second-order parameter $\tilde{\rho} = 0$, which is the worst scenario from the bias point of view. As a consequence of Theorem 1, Hill estimator is still asymptotically Gaussian in this mixture context:

Theorem 3

Suppose ℓ verifies (SO) with $\rho < 0$ and the mixture assumptions hold. If $k_n \to \infty$ with $k_n/n \to 0$ and $\sqrt{k_n}\tilde{B}(n/k_n) \to \lambda < \infty$, then

$$\sqrt{k_n}(\hat{\gamma}(k_n) - \gamma_2) \xrightarrow{d} \mathcal{N}(\lambda, \gamma_2^2).$$

There is a strong negative asymptotic bias:

$$\frac{\lambda}{\sqrt{k_n}} \sim \tilde{B}(n/k_n) = -\frac{\beta\gamma_2}{\log(n/k_n)},$$

meaning that

$$\mathbb{E}(\hat{\gamma}(k_n)) \simeq \gamma_2 \left(1 - \frac{\beta}{\log(n/k_n)}\right).$$

For instance $\beta = 1$ and $k_n/n = 10\%$ yield a relative error close to 43%.

A Jackknife estimator.

- Compute Hill estimator on two disjoint sub-samples with respective sizes $n_1 = \lfloor n^{\varepsilon} \rfloor$ and $n_2 = n - n_1$, for some $\varepsilon \in (0,1)$, to get $\hat{\gamma}_1(k_{n_1})$ and $\hat{\gamma}_2(k_{n_2})$.
- The Jackknife estimator is then defined as

$$\tilde{\gamma}_{\varepsilon}(k_{n_1}, k_{n_2}) = \frac{\hat{\gamma}_2(k_{n_2}) - \varepsilon\, \hat{\gamma}_1(k_{n_1})}{1 - \varepsilon}$$

(to make the above asymptotic bias vanish).

Asymptotic behaviour of Jacknife estimator on the mixture model.

Theorem 4

Suppose ℓ verifies (SO) with $\rho < 0$ and the mixture assumptions hold.
If $k_{n_j} \to \infty$ with $k_{n_j}/n_j \to 0$ and $\sqrt{k_{n_j}}\tilde{B}(n_j/k_{n_j}) \to \lambda_j < \infty$, $j \in \{1,2\}$, then

$$\sqrt{k_{n_1}}(\tilde{\gamma}_\varepsilon(k_{n_1}, k_{n_2}) - \gamma_2) \xrightarrow{d} \mathcal{N}\left(0,\ 2\left(\frac{\varepsilon}{1-\varepsilon}\right)^2 \gamma_2^2\right).$$

Estimator asymptotically unbiased.

Approximation of extreme quantiles. From Theorem 2, as $x \to +\infty$,

$$\frac{\bar{F}(xt)}{\bar{F}(x)} \sim t^{-1/\gamma_2}\left(\frac{\log(xt)}{\log(x)}\right)^{-\beta},$$

and thus the original approximation of extreme quantiles can also be refined:

$$q(p_n) \simeq q(\alpha_n)\left(\frac{p_n}{\alpha_n}\right)^{-\gamma_2}\left(\frac{\log p_n}{\log \alpha_n}\right)^{-\beta\gamma_2},$$

where $np_n \to 0$ while $n\alpha_n \to \infty$.

A refined Weissman estimator for extreme quantiles. Let $\alpha_n = k_n/n$ and

$$\tilde{q}(p_n) = X_{n-k_n,n} \left(\frac{p_n}{\alpha_n}\right)^{-\hat{\gamma}_2} \left(\frac{\log p_n}{\log \alpha_n}\right)^{-\hat{\beta}\hat{\gamma}_2},$$

where $\hat{\gamma}_2$ and $\hat{\beta}$ are appropriate estimators of γ_2 and β respectively.

Theorem 5

Suppose ℓ verifies (SO) with $\rho < 0$ and the mixture assumptions hold. Let $k_n = \lfloor (\ln n)^a \rfloor$, $\alpha_n = k_n/n$ and $p_n = 1/(n(\ln n)^b)$ with $a \geq 2$ and $b > 0$. Suppose we are given two estimators $\hat{\beta}$ and $\hat{\gamma}_2$ such that $\hat{\beta} \xrightarrow{\mathbb{P}} \beta$ and $\sqrt{k_n}(\hat{\gamma}_2 - \gamma_2) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, for some $v > 0$. Then,

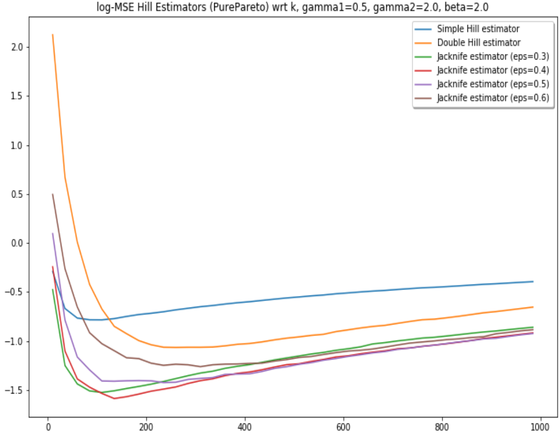
$$\frac{\sqrt{k_n}}{\log \log n} \left(\frac{\tilde{q}(p_n)}{q(p_n)} - 1\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2(a+b)^2),$$

as $n \to \infty$.

- $\tilde{q}(p_n)$ inherits its asymptotic normality from $\hat{\gamma}_2$.
- A consistent estimator of β has been proposed, although not presented here.

Illustration on simulated data. Estimation of the tail-index.

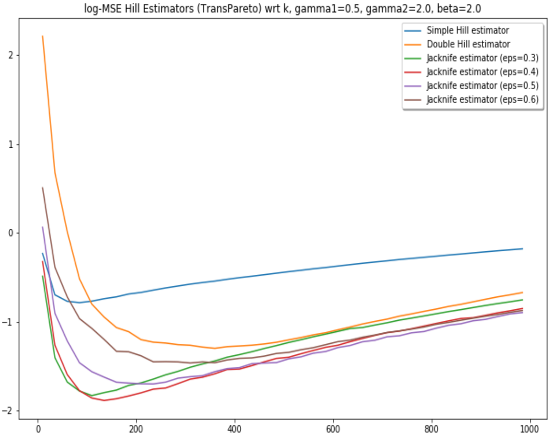
Pareto distribution ($B(x) = 0$) $\gamma_1 = 1/2$, $\gamma_2 = 2$, $\beta = 2$ and $n = 5,000$.



Log-MSE (computed on 500 replications) as a function of k_n.

Illustration on simulated data. Estimation of the tail-index.

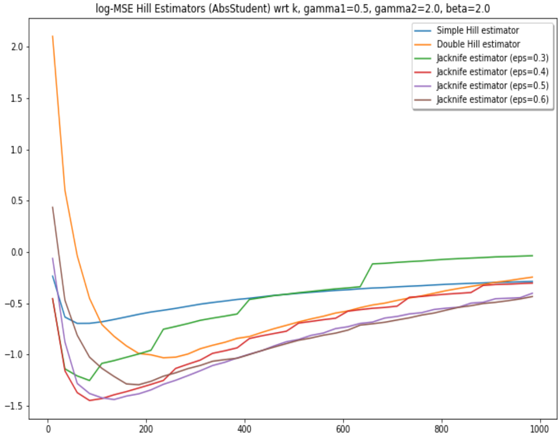
Translated Pareto distribution ($B(x) < 0$) $\gamma_1 = 1/2$, $\gamma_2 = 2$, $\beta = 2$ and $n = 5,000$.



Log-MSE (computed on 500 replications) as a function of k_n.

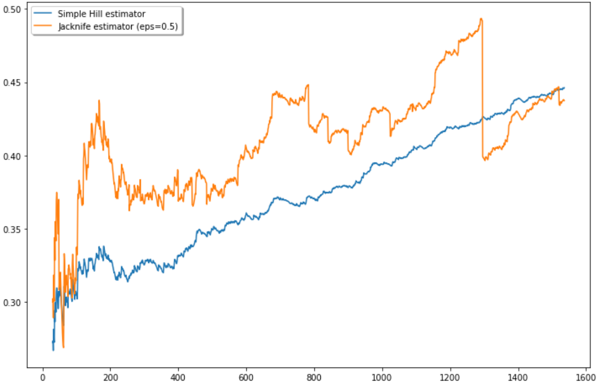
Illustration on simulated data. Estimation of the tail-index.

Student distribution ($B(x) > 0$) $\gamma_1 = 1/2$, $\gamma_2 = 2$, $\beta = 2$ and $n = 5,000$.



Log-MSE (computed on 500 replications) as a function of k_n.

Illustration on real data. Comparison of Hill and Jackknife tail-index estimators on the FTSE100 dataset, whole period 1935-1996.



Jackknife: nice stability when $k_n \in \{200, \ldots, 600\}$ pointing towards $\gamma \simeq 0.37$.
Hill: no stability and under-estimation suspected.