Estimation of a new parameter discriminating between Weibull tail-distributions and heavy-tailed distributions

INRIA

Jonathan El Methni [(1)], Laurent Gardes [(1)], Stéphane Girard [(1)] & Armelle Guillou [(2)]

Laboratoire Jean Kuntzmann

[(1)] Team Mistis, INRIA Rhône-Alpes & LJK, http://mistis.inrialpes.fr – [(2)] Université de Strasbourg & CNRS.
Contact: jonathan.elmethni@inrialpes.fr http://mistis.inrialpes.fr/people/elmethni

I. Statistical Framework

- Let $X_1, \dots, X_n$ be a sample of independent and identically distributed random variables driven from X with cumulative distribution function F, and let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics associated to this sample.
- We want to estimate the extreme quantile x_{p_n} of order p_n associated to the random variable $X \in \mathbb{R}$ defined by :

$$x_{p_n} = \overline{F}^{\leftarrow}(p_n) = \inf\{x, \bar{F}(x) \leq p_n\},$$

with $p_n \to 0$ when $n \to \infty$. The function $\overline{F}^{\leftarrow}$ is the generalized inverse of the non-increasing function $\overline{F} = 1 - F$.

- In [2], a family of distributions is introduced, it encompasses the whole Fréchet maximum domain of attraction as well as Weibull tail-distributions. These distributions depend on two parameters $\tau \in [0,1]$ and $\theta > 0$.

II. Model: Gumbel/Fréchet

Let us consider the family of survival distribution functions defined as
$\overline{F}(x) = \exp(-K_\tau^{\leftarrow}(\log H(x)))$ for $x \geq x_* > 0$ where

- $K_\tau(y) = \int_1^y u^{\tau-1} du$ where $\tau \in [0,1]$,
- H an increasing function such that $H^{\leftarrow} = x^{-\theta}\ell(x)$ where $\theta > 0$ and $\ell(x)$ is a slowly varying function.

Proposition

- $\tau = 0 \iff F$ is a Weibull-tail distribution function with Weibull tail-coefficient θ.
- $\tau \in [0,1)$ and H is twice differentiable $\implies F$ belongs to the Gumbel maximum domain of attraction.
- $\tau = 1 \iff F$ is in the Fréchet maximum domain of attraction with tail-index θ.

III. Estimators depending on τ

Denoting by (k_n) an intermediate sequence of integers, the following estimator of θ is considered in [2]:

$$\widehat{\theta}_{n,\tau}(k_n) = \frac{H_n(k_n)}{\mu_\tau(\log(n/k_n))}$$

where $H_n(k_n)$ is the Hill estimator [3],

$$H_n(k_n) = \frac{1}{k_n - 1}\sum_{i=1}^{k_n-1} \log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n}),$$

and with, for all $t > 0$, $\mu_\tau(t) = \int_0^\infty (K_\tau(x+t) - K_\tau(t))\, e^{-x} dx$.
An estimator of the extreme quantile x_{p_n} is proposed in [2]:

$$\widehat{x}_{p_n,\widehat{\theta}_{n,\tau}(k_n)} = X_{n-k_n+1,n} \exp\left(\widehat{\theta}_{n,\tau}(k_n)\left(K_\tau(\log(1/p_n)) - K_\tau(\log(n/k_n))\right)\right).$$

IV. Main Goal

The asymptotic distributions of $\widehat{\theta}_{n,\tau}(k_n)$ and $\widehat{x}_{p_n,\widehat{\theta}_{n,\tau}(k_n)}$ have been established in [2] under a second-order condition on ℓ:

There exist $\rho < 0$, a function b satisfying $b(x) \to 0$ and $|b|$ asymptotically decreasing such that uniformly locally on $\lambda > 0$

$$\log\left(\frac{\ell(\lambda x)}{\ell(x)}\right) \sim b(x) K_\rho(\lambda), \text{ when } x \to \infty.$$

► The main goal of this work is to propose an estimator for τ independent from θ.

This parameter controls the behavior of the tail-distribution: the larger the value of τ, the heavier is the tail.

V. Estimator of τ

Let us consider for $t > t'$

$$\psi(x;t,t') : \mathbb{R} \to (-\infty, \exp(t-t')) \text{ such that } \psi(x;t,t') = \frac{\mu_x(t)}{\mu_x(t')}$$

Denoting by (k_n) and (k'_n) two intermediate sequences of integers such that $k'_n > k_n$, the following estimator of τ is considered :

$$\widehat{\tau}_n = \begin{cases} \psi^{-1}\left(\frac{H_n(k_n)}{H_n(k'_n)}; \log(n/k_n), \log(n/k'_n)\right) & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} < \frac{k'_n}{k_n} \\ u & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} \geq \frac{k'_n}{k_n} \end{cases}$$

where $u \in [0,1]$.

- $\widehat{\tau}_n$ exists because $\psi(.;\log(n/k_n), \log(n/k'_n)) : \mathbb{R} \to (-\infty, k'_n/k_n)$ is a bijection.

VI. Asymptotic distribution

Replacing τ by $\widehat{\tau}_n$ we obtain :

- an estimator of θ: $\widehat{\theta}_{n,\widehat{\tau}_n}(k_n) = \frac{H_n(k_n)}{\mu_{\widehat{\tau}_n}(\log(n/k_n))}$
- an estimator of x_{p_n}:

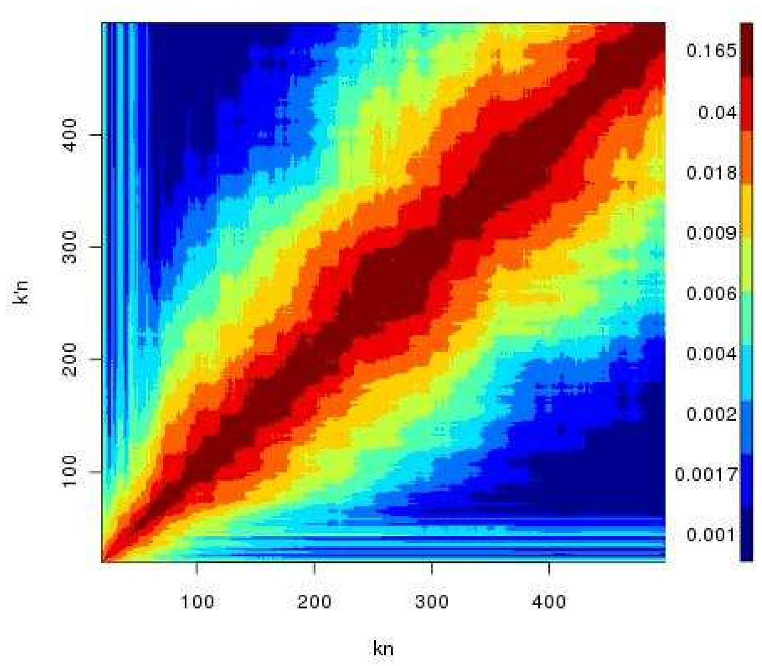
$$\widehat{x}_{p_n,\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)} = X_{n-k_n+1,n} \exp\left(\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)\left(K_{\widehat{\tau}_n}(\log(1/p_n)) - K_{\widehat{\tau}_n}(\log(n/k_n))\right)\right).$$

Under some assumptions on (k_n) and (k'_n) we establish the asymptotic normality of $\widehat{\tau}_n$, $\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)$ and $\widehat{x}_{p_n,\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)}$. In particular:

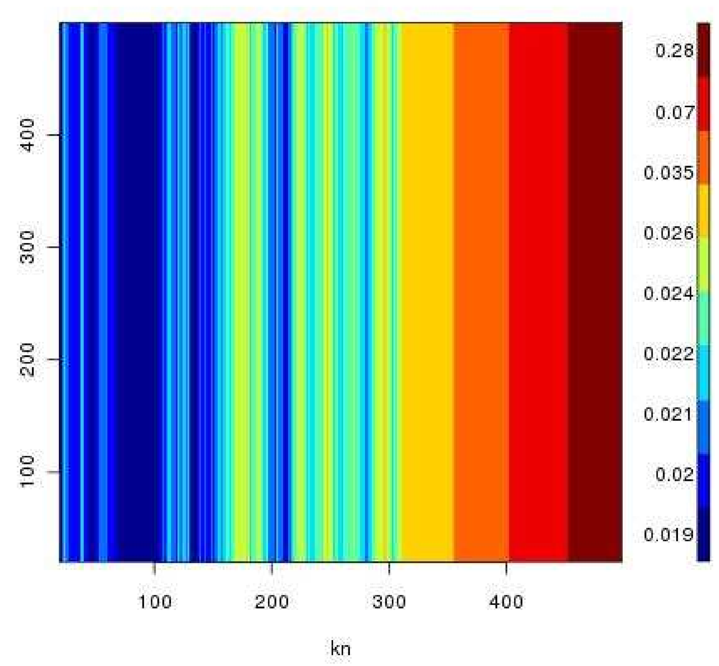
$$\frac{\sqrt{k_n}\left(\log_2(n/k_n) - \log_2(n/k'_n)\right)}{\int_{\log(n/k_n)}^{\log(1/p_n)} \log(u) u^{\tau-1} du}\left(\frac{\widehat{x}_{p_n,\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)}}{x_{p_n}} - 1\right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

VII. Numerical experiments on simulated data

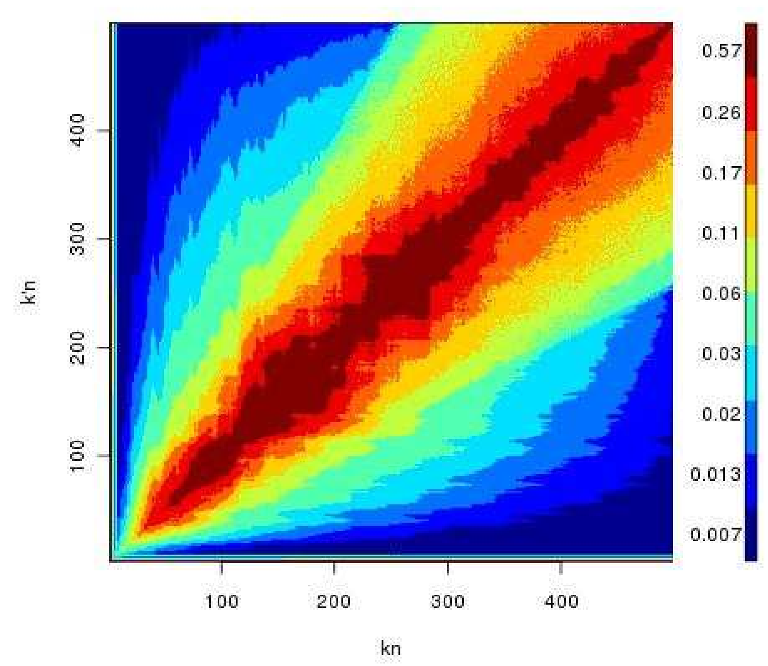
Gamma distribution $\mathcal{D}(Gumbel)$ **Pareto distribution** $\mathcal{D}(Fréchet)$



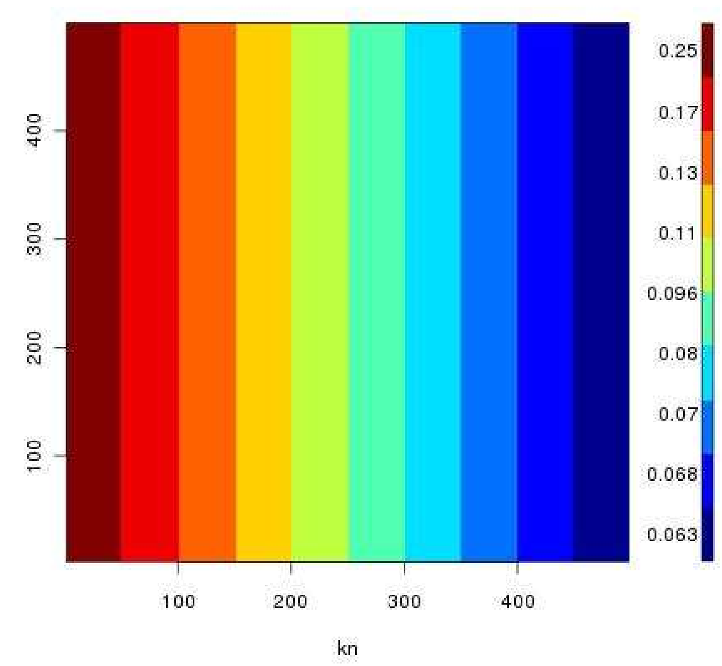
$\mathcal{MSE}$ $\widehat{x}_{p_n,\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)}$



$\mathcal{MSE}$ Dekkers et al.



$\mathcal{MSE}$ $\widehat{x}_{p_n,\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)}$



$\mathcal{MSE}$ Dekkers et al.

- The estimator $\widehat{x}_{p_n,\widehat{\theta}_{n,\widehat{\tau}_n}(k_n)}$ is computed on $N = 100$ samples of size $n = 500$ for $k_n = 2, \dots, 499$ and $k'_n = k_n, \dots, 500$ where $p_n = 10^{-3}$.
- The associated deciles of the empirical Mean-Squared Error $\mathcal{MSE}$ are plotted
- Comparison with an estimator of A. L. M. Dekkers, J.H.J. Einmahl & L. de Haan [1].

Bibliography

[1] Dekkers, A. L. M., Einmahl, J. H. J., De Haan, L., (1989). A Moment Estimator for the Index of an Extreme-Value Distribution, *The Annals of Statistics*, **17**, 1833–1855.

[2] Gardes, L., Girard, S., Guillou, A., (2011). Weibull tail-distributions revisited: a new look at some tail estimators, *Journal of Statistical Planning and Inference*, **141**, 429–444.

[3] Hill, B.M., (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**, 1163–1174.