Study of the relative extrapolation error associated with Weissman estimator for extreme quantiles

Clément ALBERT

Anne Dutfoy (EDF), Stéphane Girard (INRIA)

Journées de Statistique
June 2019

Outline

1. Extreme quantile estimation : Weissman estimator

2. Extrapolation error associated with Weissman estimator

3. Comparison to the extrapolation error associated with the Exponential Tail estimator

Outline

1. Extreme quantile estimation : Weissman estimator
2. Extrapolation error associated with Weissman estimator
3. Comparison to the extrapolation error associated with the Exponential Tail estimator

Extreme quantile estimation: Principle

- Let $X_1, \ldots, X_n$ be a sample of n random variables *i.i.d.* whose distribution function F (with associated survival function $\overline{F}$) is unknown.

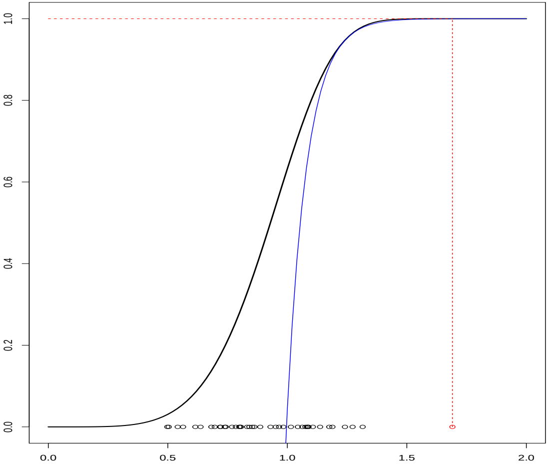
- Goal : Estimate $q(p_n)$ an extreme quantile of order $1 - p_n$, with $p_n \to 0$ when $n \to \infty$:

$$q(p_n) = F^{\leftarrow}(1 - p_n),$$

$F^{\leftarrow}$ generalized inverse of F.

- Difficulty : If p_n is small enough ($np_n \underset{n\to\infty}{\to} 0$) then

$$\mathrm{P}(q(p_n) > \max(X_1, \ldots X_n)) \underset{n\to\infty}{\longrightarrow} 1.$$



- Extrapolation beyond the range of the sample is required.

Weissman estimator

Weissman estimator [Weissman, 1978] is a semi-parametric estimator dedicated to distributions belonging to the maximum domain of attraction of Fréchet (MDA(Fréchet)). It is based on the assumption that the function q is regularly varying with index $-\gamma$ (denoted $q \in RV_{-\gamma}$), $\gamma > 0$ i.e that

$$q(x) = x^{-\gamma}\ell(x^{-1}), \tag{1}$$

with ℓ a slowly varying function ($\ell \in RV_0$) satisfaying, for all $t > 0$:

$$\lim_{x \to \infty} \frac{\ell(tx)}{\ell(x)} = 1.$$

Define two sequences $\alpha_n \to 0$, $p_n \to 0$ such that $p_n < \alpha_n$. Based on (1), for all $\gamma > 0$:

$$q(p_n) = p_n^{-\gamma}\ell(p_n^{-1}), \tag{2}$$

$$q(\alpha_n) = \alpha_n^{-\gamma}\ell(\alpha_n^{-1}). \tag{3}$$

Dividing (3) by (2) leads to :

$$q(p_n) = q(\alpha_n)\left(\frac{\alpha_n}{p_n}\right)^{\gamma} \frac{\ell(p_n^{-1})}{\ell(\alpha_n^{-1})}.$$

Weissman estimator

Assuming that $\limsup \log(1/p_n)/\log(1/\alpha_n) < \infty$ finally leads to the following quantile approximation :

$$\tilde{q}_{\mathrm{W}}(p_n; \alpha_n) = q(\alpha_n) \left(\frac{\alpha_n}{p_n} \right)^{\gamma_n},$$

which is called Weissman approximation in the sequel. Weissman estimator [Weissman, 1978] is then obtained by replacing the intermediate quantile $q(\alpha_n)$ and the tail index γ_n by appropriate estimators:

$$\hat{q}_{\mathrm{W}}(p_n; \alpha_n) = \hat{q}(\alpha_n) \left(\frac{\alpha_n}{p_n} \right)^{\hat{\gamma}_n}.$$

The most common choices are $\hat{q}(\alpha_n) = X_{n-k_n+1,n}$, $k_n = \lfloor n\alpha_n \rfloor$ and Hill estimator [Hill, 1975]:

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=1}^{k_n} \log X_{n-i+1,n} - \log X_{n-k_n+1,n}.$$

Extrapolation is performed from the quantile of order α_n times a correction which is proportional to the ratio between the two orders α_n and p_n.

Outline

Extrapolation error: Definition

The error $q(p_n) - \hat{q}_{\mathrm{W}}(p_n; \alpha_n)$ between the true quantile and Weissman estimator can be expanded as a sum of two terms:

$$q(p_n) - \hat{q}_{\mathrm{W}}(p_n; \alpha_n) = (\tilde{q}_{\mathrm{W}}(p_n; \alpha_n) - \hat{q}_{\mathrm{W}}(p_n; \alpha_n)) + (q(p_n) - \tilde{q}_{\mathrm{W}}(p_n; \alpha_n)),$$

the first one being a random estimation error

$$\tilde{q}_{\mathrm{W}}(p_n; \alpha_n) - \hat{q}_{\mathrm{W}}(p_n; \alpha_n)$$

and the second one being a deterministic extrapolation error

$$q(p_n) - \tilde{q}_{\mathrm{W}}(p_n; \alpha_n).$$

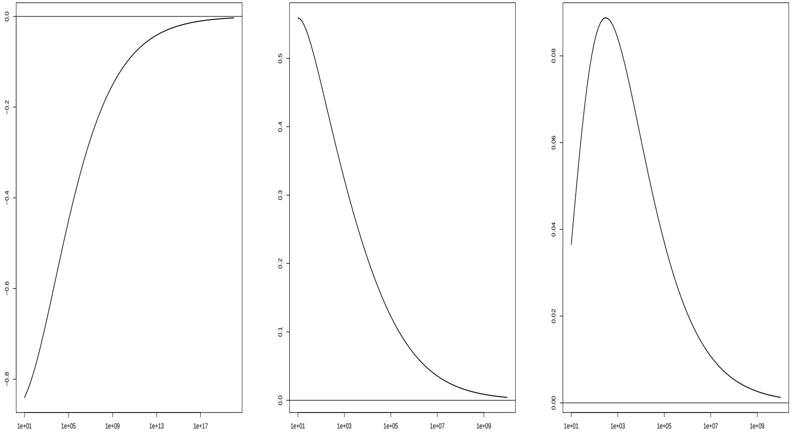
In this talk, we focus on the asymptotic behavior of the extrapolation error. Our goal is to quantify to which extent the extrapolation can be performed in a consistent way. More specifically, we provide necessary and sufficient conditions on the pair (p_n, α_n) such that the relative extrapolation error

$$\varepsilon_{\mathrm{W}}(p_n; \alpha_n) := (q(p_n) - \tilde{q}_{\mathrm{W}}(p_n; \alpha_n))/q(p_n)$$

tends to zero as $n \to \infty$.

Extrapolation error: Illustration

Figure: Vertically: Extrapolation error $\varepsilon_W(p_n; \alpha_n)$ with $p_n = n^{-5/4}$ and $\alpha_n = n^{-3/4}$. Horizontally: $\log n$. From left to right : Burr, Fréchet, Pareto distributions.



- Does the relative extrapolation error tend to zero whenever $F \in$MDA(Fréchet) ?
- What if the model is misspecified, for example if $F \in$MDA(Gumbel) ?

Extrapolation error: Model

Define $H(\cdot) = -\log \overline{F}(\cdot)$ the cumulative hazard rate function and $K(x) = x(\log\log H^{-1})'(x)$, $x > 0$. In the following, we suppose that $K \in RV_0$.

Proposition

Suppose F is increasing, twice differentiable and K' is ultimately monotone.

(i) If $\exp H \in RV_{1/\gamma}$, $\gamma > 0$, then $K \in RV_0$.

(ii) If $H \in RV_\beta$, $\beta > 0$, then $K \in RV_0$.

(iii) If $H(\exp(\cdot)) \in RV_\beta$, $\beta > 0$ then $K \in RV_0$.

(i) corresponds to Pareto-tail distributions (i.e heavy tails such as Burr, Cauchy, Fréchet, Pareto, Student...).

(ii) corresponds to Weibull-tail distributions (Exponential, Gamma, Weibull, Gaussian...).

(iii) corresponds to log-Weibull tail distributions (Lognormal).

Extrapolation error: Main results

Define $\delta_n := 1 - \log(1/\alpha_n)/\log(1/p_n)$ a proximity measure between α_n and p_n.

Theorem (Necessary and sufficient conditions on (α_n, p_n) for $\varepsilon_W(p_n;\alpha_n) \to 0$)

Let $0 < p_n \leq \alpha_n < 1$ such that $\limsup \delta(n) < 1$ or equivalently $\limsup \log(1/p_n)/\log(1/\alpha_n) < \infty$. Then, under the previous assumptions :

(i) Suppose $F \in$ MDA(Fréchet) with tail index $\gamma > 0$ and F satisfy a second order condition assumption.

- *If $\delta(n) \to \delta_\infty \in [0,1)$ then $\varepsilon_W(p_n;\alpha_n) \to 0$ as $n \to \infty$.*

(ii) Weibull tail-distributions. Suppose $H \in RV_\beta$, $\beta > 0$.

- *Then, $\varepsilon_W(p_n;\alpha_n) \to 0$ if and only if $\delta(n) \to 0$ as $n \to \infty$.*

(iii) Log-Weibull tail-distributions. Suppose $H(\exp \cdot) \in RV_\beta$, $\beta > 0$ and $\beta \neq 1$.

- *Then, $\varepsilon_W(p_n;\alpha_n) \to 0$ if and only if $\delta^2(n) \log q(p_n) \to 0$ as $n \to \infty$.*

Extrapolation error: Main results

Theorem (First order approximation of $\varepsilon_W(p_n;\alpha_n)$)

Under the same assumptions than previous Theorem :

(i) *Suppose $F \in$ MDA(Fréchet) with tail index $\gamma > 0$ and F satisfy a second order condition assumption.*

- *If* $\delta(n) \to \delta_\infty \in [0,1)$ *then* $\varepsilon_W(p_n;\alpha_n) \sim -\frac{1}{1-\delta_\infty}\delta(n)\log(1/\alpha_n)\eta(1/\alpha_n)$.

(ii) *Weibull tail-distributions. Suppose $H \in RV_\beta$, $\beta > 0$.*

- *If* $\delta(n) \to 0$ *then* $\varepsilon_W(p_n;\alpha_n) \sim -\frac{1}{2\beta}\delta^2(n)$.
- *If* $\delta(n) \to \delta_\infty \in (0,1)$ *then* $\varepsilon_W(p_n;\alpha_n) \to 1 - \exp\left(\frac{1}{\beta}\frac{\delta_\infty^2}{1-\delta_\infty}\right)$.

(iii) *Log-Weibull tail-distributions. Suppose $H(\exp\cdot) \in RV_\beta$, $\beta > 0$ and $\beta \neq 1$.*

- *If* $\delta^2(n)\log q(p_n) \to 0$ *then* $\varepsilon_W(p_n;\alpha_n) \sim \frac{1-\beta}{2\beta^2}\delta^2(n)\log q(p_n)$.
- *If* $\delta^2(n)\log q(p_n) \to a \in (0,\infty)$ *then* $\varepsilon_W(p_n;\alpha_n) \to 1 - \exp\left(-\frac{1-\beta}{2\beta^2}a\right)$.

Extrapolation error: Illustration

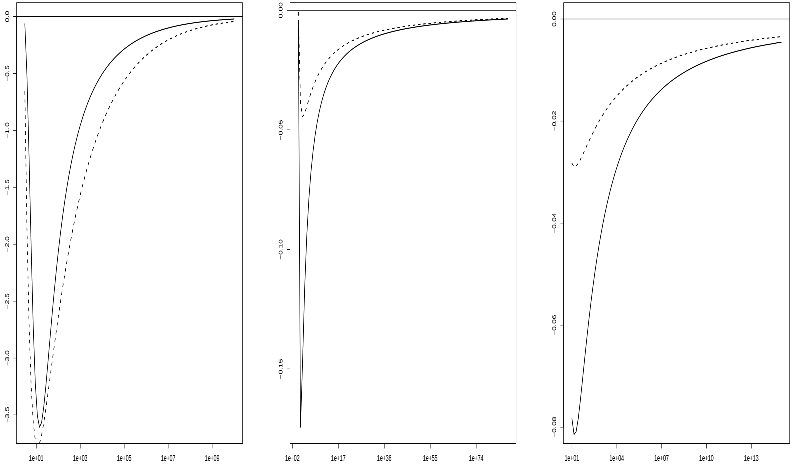


Figure: Vertically: Extrapolation error $\varepsilon_{\mathrm{W}}(p_n;\alpha_n)$ (solid line) and its first order approximation (dashed line) with $p_n = 1/(n \log n)$ and $\alpha_n = (\log n)/n$. Horizontally: $\log n$. From left to right : Burr, Lognormale, Weibull distributions.

Extrapolation error: Rates of convergence

Distribution	$\varepsilon_{\mathrm{W}}(p_n; \alpha_n)$
Gamma($a > 0$)	$-2\dfrac{(\log\log n)^2}{(\log n)^2}$
Weibull($\beta \neq 1$)	$-\dfrac{2}{\beta}\dfrac{(\log\log n)^2}{(\log n)^2}$
Gaussian	$-\dfrac{(\log\log n)^2}{(\log n)^2}$
Log-Weibull($\beta > 1$)	$\dfrac{2(1-\beta)}{\beta^2}\dfrac{(\log\log n)^2}{(\log n)^{2-1/\beta}}$
Lognormal	$-\dfrac{\sqrt{2\sigma^2}}{2}\dfrac{(\log\log n)^2}{(\log n)^{3/2}}$

Table: First order approximations of $\varepsilon_{\mathrm{W}}(p_n; \alpha_n)$ with $p_n = 1/(n \log n)$ and $\alpha_n = (\log n)/n$ for some usual distributions.

Outline

Exponential Tail approximation

Pickands theorem [Pickands, 1975] states that the distribution of excesses over a given threshold can be approximated by a Generalized Pareto Distribution (GPD). Based on this result, $q(p_n)$ can be approximated by

$$q(p_n) \approx q(\alpha_n) + \frac{\sigma_n}{\gamma_n}\left[\left(\frac{\alpha_n}{p_n}\right)^{\gamma_n} - 1\right], \tag{4}$$

where σ_n and γ_n are respectively the scale and shape parameters of the GPD distribution.

The Exponential Tail (ET) approximation [Breiman et al, 1990] corresponds to the case where $F \in$MDA(Gumbel). In such a situation, $\gamma_n = 0$ and the GPD distribution reduces to an Exponential distribution with scale parameter σ_n :

$$\tilde{q}_{\mathrm{ET}}(p_n) = q(\alpha_n) + \sigma_n \log(\alpha_n / p_n).$$

Extrapolation error: Exponential Tail

The Exponential Tail relative extrapolation error is given by :

$$\varepsilon_{\mathrm{ET}}(p_n;\alpha_n) := \frac{q(p_n) - \tilde{q}_{\mathrm{ET}}(p_n)}{q(p_n)} = \frac{q(p_n) - q(\alpha_n) - \sigma_n \log(\alpha_n/p_n)}{q(p_n)}.$$

Theorem (Necessary and sufficient conditions on (α_n, p_n) for $\varepsilon_{\mathrm{ET}}(p_n;\alpha_n) \to 0$)

Under some regularity assumptions :

(i) *Suppose F is a Weibull-tail distribution with index $\beta = 1$ (Exponential, Gamma, Logistique, Gumbel)*

- *Then $\varepsilon_{ET}(p_n;\alpha_n) \to 0$ as $n \to \infty$.*

(ii) *Suppose F is a Weibull-tail distribution with index $\beta \neq 1$ (Gaussian, Weibull)*

- *Then $\varepsilon_{ET}(p_n;\alpha_n) \to 0$ if and only if $\delta(n) \to 0$ as $n \to \infty$.*

(iii) *Suppose F is a log-Weibull tail distribution (Lognormal, Log-Weibull)*

- *Then $\varepsilon_{ET}(p_n;\alpha_n) \to 0$ if and only if $\delta(n)h(\log(1/p_n)) \to 0$ as $n \to \infty$, h a specific function, $h(\log(1/p_n)) \to \infty$.*

Extrapolation error: Rates of convergence

Distribution	$\varepsilon_{\text{ET}}(p_n;\alpha_n)$	$\varepsilon_{\text{W}}(p_n;\alpha_n)$
Gamma$(a>0)$	$2(1-a)\dfrac{(\log\log n)^2}{(\log n)^3}$	$-2\dfrac{(\log\log n)^2}{(\log n)^2}$
Weibull$(\beta\neq 1)$	$\dfrac{2(1-\beta)}{\beta^2}\dfrac{(\log\log n)^2}{(\log n)^2}$	$-\dfrac{2}{\beta}\dfrac{(\log\log n)^2}{(\log n)^2}$
Gaussian	$-\dfrac{1}{2}\dfrac{(\log\log n)^2}{(\log n)^2}$	$-\dfrac{(\log\log n)^2}{(\log n)^2}$
Log-Weibull$(\beta>1)$	$\dfrac{2}{\beta^2}\dfrac{(\log\log n)^2}{(\log n)^{2-2/\beta}}$	$\dfrac{2(1-\beta)}{\beta^2}\dfrac{(\log\log n)^2}{(\log n)^{2-1/\beta}}$
Lognormal	$\sigma^2\dfrac{(\log\log n)^2}{\log n}$	$-\dfrac{\sqrt{2\sigma^2}}{2}\dfrac{(\log\log n)^2}{(\log n)^{3/2}}$

Table: First order approximations of $\varepsilon_{\text{ET}}(p_n;\alpha_n)$ and $\varepsilon_{\text{W}}(p_n;\alpha_n)$ with $p_n = 1/(n\log n)$ and $\alpha_n = (\log n)/n$.

Extrapolation error: Illustration

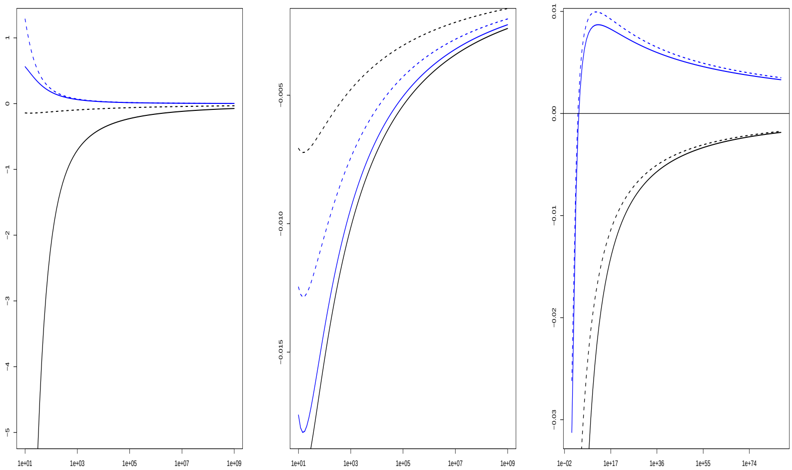


Figure: Vertically: $\varepsilon_{\mathrm{W}}(p_n;\alpha_n)$ (black solid line), $\varepsilon_{\mathrm{ET}}(p_n;\alpha_n)$ (blue solid line) and their corresponding first order approximation (dashed line) with $p_n = 1/(n\log n)$ and $\alpha_n = (\log n)/n$. Horizontally: $\log n$. From left to right : Gamma, Weibull, Log-Weibull.

Conclusions

- Extrapolation abilities of Weissman approximation are good for distributions belonging to MDA(Fréchet) and relatively poor for some parts of MDA(Gumbel) (Weibull-tail, log-Weibull tail), which is logical since it has not been designed for this purpose.

- In comparison, extrapolation abilities of the ET method are even poorer on the same parts of MDA(Gumbel), even if it is dedicated to MDA(Gumbel) precisely, doing good only for distributions near the exponential one.

- Estimators dedicated to Weibull-tail and log-Weibull tail distributions [Gardes et al, 2013] should thus be preferred to ET estimator when someone has some insights about the shape of a distribution $F \in$MDA(Gumbel).

Main references

[1] **Weissman, I. (1978).** Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364), 812-815.

[2] **Breiman, L., Stone, C. J., & Kooperberg, C. (1990).** Robust confidence bounds for extreme upper quantiles. *Journal of Statistical Computation and Simulation*, 37(3-4), 127-149.

[3] **Albert, C., Dutfoy, A., & Girard, S. (2018)**. *Asymptotic behavior of the extrapolation error associated with the estimation of extreme quantiles*, submitted.

[4] **Gardes, L., & Girard, S. (2013)**. Estimation de quantiles extrêmes pour les lois à queue de type Weibull: une synthèse bibliographique. *Journal de la Société Française de Statistique*, 154(2), 98-118.

[5] **Albert, C., Dutfoy, A., Gardes, L., & Girard, S. (2019)**. An extreme quantile estimator for the log-generalized Weibull-tail model. *Econometrics and Statistics*, to appear.