A unified statistical model for Pareto and Weibull tail- distributions

Stéphane Girard

INRIA Rhône-Alpes, Team Mistis
http://mistis.inrialpes.fr/people/girard/

June, 2009

Joint work with Laurent Gardes (Mistis, INRIA Rhône-Alpes)
and Armelle Guillou (Université de Strasbourg)

1 Motivation: Weibull tail- distributions

2 φ- tail distributions

3 Further work

Weibull tail- distributions

Definition. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F such that

$$1 - F(x) = \exp(-H(x)), \quad H^{\leftarrow}(t) = \inf\{x,\ H(x) \geq t\} = t^{\theta}\ell(t),$$

where

- $\theta > 0$ is the Weibull tail- index,
- ℓ is a slowly varying function *i.e.*

$$\ell(\lambda x)/\ell(x) \to 1 \text{ as } x \to \infty \text{ for all } \lambda > 0.$$

The inverse failure rate function $H^{\leftarrow}$ is said to be regularly varying at infinity with index θ and this property is denoted by $H^{\leftarrow} \in \mathcal{R}_{\theta}$.

Weibull tail- distributions

Remarks.

- Weibull tail- distributions are included in the Gumbel Maximum Domain of Attraction.
- **Second order condition:** There exist $\rho \leq 0$ and $b(x) \to 0$ such that uniformly locally on $\lambda \geq 1$

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

with $K_\rho(\lambda) = \int_1^\lambda u^{\rho-1}du$.

Weibull tail- distributions

Examples.

	θ	$\ell(x)$	$b(x)$	ρ
Gaussian(μ, σ^2)	$1/2$	$2^{1/2}\sigma - \frac{\sigma}{2^{3/2}} \frac{\log x}{x} + O\left(\frac{1}{x}\right)$	$\frac{1}{4} \frac{\log x}{x}$	-1
Gamma(α, β)	1	$\frac{1}{\beta} + \frac{\alpha - 1}{\beta} \frac{\log x}{x} + O\left(\frac{1}{x}\right)$	$(1-\alpha)\frac{\log x}{x}$	-1
Weibull(α, λ)	$1/\alpha$	λ	0	$-\infty$

Weibull tail- distributions

Estimation of the Weibull tail- index.

1. [Beirlant *et al*, 1996] proposed an estimator based on the Hill statistics:

$$\hat{\theta}_n = \frac{\frac{1}{k_n}\sum_{i=1}^{k_n}\left(\log X_{n-i+1,n} - \log X_{n-k_n+1,n}\right)}{\frac{1}{k_n}\sum_{i=1}^{k_n}\left(\log\log(n/i) - \log\log(n/k_n)\right)},$$

where (k_n) is an intermediate sequence, *i.e.* a sequence such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

Weibull tail- distributions

Estimation of the Weibull tail- index.

2. [Gardes & Girard, 2006] proposed different normalizing sequences:

$$\hat{\theta}_n = \frac{1}{T_n}\frac{1}{k_n}\sum_{i=1}^{k_n}(\log X_{n-i+1,n} - \log X_{n-k_n+1,n})$$

where (T_n) is such that $T_n \log(n/k_n) \to 1$ as $n \to \infty$.

3. Introduction of weights [Gardes & Girard, 2008]:

$$\hat{\theta}_n = \frac{\sum_{i=1}^{k_n} \alpha_{i,n}(\log X_{n-i+1,n} - \log X_{n-k_n+1,n})}{\sum_{i=1}^{k_n} \alpha_{i,n}(\log\log(n/i) - \log\log(n/k_n))},$$

Weibull tail- distributions

Estimation of the Weibull tail- index.

4. Other propositions: Mean residual life function [Beirlant *et al*, 1995], Bias correction [Diebolt *et al*, 2008], Mean excess function [Dierckx *et al*, 2009], ...

Remarks

- Most of the proposed estimators are based on linear combinations of log-spacings between upper order statistics.
- How to explain this similarity with the Hill estimator (dedicated to the Fréchet Maximum Domain of Attraction)?

1 Motivation: Weibull tail- distributions

2 φ- tail distributions

3 Further work

φ- tail distributions

Definition. The cumulative distribution function is such that

$$1 - F(x) = \exp(-\varphi^{\leftarrow}(\log H(x))), \text{ for } x \geq x_*,$$

with

- $H^{\leftarrow} \in \mathcal{R}_\theta$,
- $\varphi(x) \to \infty$ as $x \to \infty$,
- φ is continuously differentiable,
- $\varphi' \in \mathcal{R}_\tau$, with $\tau \in [-1, 0]$,
- there exists $M > 0$ such that $0 \leq \varphi'(.) \leq M$.

Examples.

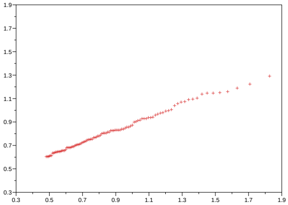
- $\varphi(t) = t$: Pareto type distributions (Fréchet Maximum Domain of Attraction), $\tau = 0$,
- $\varphi(t) = \log t$: Weibull tail- distributions, $\tau = -1$.

φ- tail distributions

The log-spacings between two quantiles x_u and x_v of $1 - F$

$$\begin{aligned}\log x_u - \log x_v &= \theta\left(\varphi(\log 1/u) - \varphi(\log 1/v)\right) \\ &+ \log\left(\frac{\ell(\exp\varphi(\log 1/u))}{\ell(\exp\varphi(\log 1/v))}\right) \\ &\simeq \theta\left(\varphi(\log 1/u) - \varphi(\log 1/v)\right)\end{aligned}$$

are approximately proportional to θ (if the orders u and v of the quantiles are small enough).



Pairs $(\varphi(\log n/i), \log(X_{n-i+1,n}))$ for $i = 1, \ldots, 100$ from a Gaussian sample with size $n = 500$.

φ- tail distributions

Inference. The following estimator of θ is introduced:

$$\widehat{\theta}_n = \frac{1}{\mu(\log(n/k_n))} \frac{1}{k_n} \sum_{i=1}^{k_n} (\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})),$$

where $\mu(t) = \int_0^\infty (\varphi(x+t) - \varphi(t))\, \mathrm{e}^{-x} dx.$

- Pareto type distributions: $\mu(t) = 1 \longrightarrow$ Hill estimator
- Weibull tail- distributions:
  i) $\mu(t) = \int_0^1 \log\left(1 - \frac{\log y}{t}\right) dy \longrightarrow$ [Gardes & Girard, 2006]
  ii) Approximation by a Riemann sum $\longrightarrow$ [Beirlant *et al*, 1996]
  iii) $\mu(t) \sim \varphi'(t) = 1/t \longrightarrow$ [Gardes & Girard, 2006]

φ- tail distributions

Theorem 1: Asymptotic normality of $\widehat{\theta}_n$

Let (k_n) be an intermediate sequence such that $k_n^{1/2} b(\exp\varphi(\log(n/k_n))) \to 0$. Then,

$$k_n^{1/2}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

φ- tail distributions

Estimation of extreme quantiles. Recall that an extreme quantile x_{p_n} of order p_n is defined by the equation $1 - F(x_{p_n}) = p_n$ with $0 < p_n < 1/n$. An estimator of x_{p_n} can be deduced from $\widehat{\theta}_n$ by:

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

- Pareto type distributions: $\varphi(t) = t \longrightarrow$ Weissman estimator
- Weibull tail- distributions: $\varphi(t) = \log t \rightarrow$ [Gardes & Girard, 2005]

φ- tail distributions

Theorem 2: Asymptotic distribution of $\widehat{x}_{p_n}$

Suppose the assumptions of Theorem 1 hold with $\rho < 0$. If, moreover,

$$k_n^{1/2} \frac{b(\exp \varphi(\log(n/k_n)))}{\log(n/k_n)\varphi'(\log(n/k_n))} \to 0$$

and there exists $c > 1$ such that

$$\frac{\log(1/p_n)}{\log(n/k_n)} \to c$$

then,

$$\frac{k_n^{1/2}}{\log(n/k_n)\varphi'(\log(n/k_n))} \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \theta^2 K_{\tau+1}^2(c)).$$

1 Motivation: Weibull tail- distributions

2 φ- tail distributions

3 Further work

Further work

The φ- tail model could help to discriminate between Pareto type and Weibull tail- distributions.

- Either by building a test based on the Hill statistics

$$\widehat{H}_n := \frac{1}{k_n} \sum_{i=1}^{k_n} (\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n}))$$

since $H_n \xrightarrow{P} \theta$ for Pareto type distributions and $H_n \xrightarrow{P} 0$ for Weibull tail- distributions.

- Or by estimating τ ($\tau = 0$ for Pareto type distributions and $\tau = -1$ for Weibull tail- distributions).

References

- Beirlant, J., Broniatowski, M., Teugels, J. & Vynckier, P. (1995) The mean residual life function at great age: Applications to tail estimation, *JSPI*, **45**, 21–48.
- Beirlant, J., Teugels, J. & Vynckier, P. (1996) *Practical analysis of extreme values*, Leuven university press.
- Diebolt, J., Gardes, L., Girard, S. & Guillou, A. (2008) Bias-reduced estimators of the Weibull tail-coefficient, *Test*, **17**, 311–331.
- Dierckx, G., Beirlant, J., De Waal, D. & Guillou, A. (2009) A new estimation method for Weibull-type tails based on the mean excess function, *JSPI*, **139**, 1905–1920.
- Gardes, L. & Girard, S. (2005) Estimating extreme quantiles of Weibull tail distributions, *Comm. Stat.*, **34**, 1065–1080.
- ______ (2006) Comparison of Weibull tail-coefficient estimators, *REVSTAT*, **4**, 163–188.
- ______ (2008) Estimation of the Weibull tail-coefficient with linear combination of upper order statistics, *JSPI*, **138**, 1416–1427.