Estimation of the functional Weibull-tail coefficient

Stéphane Girard

Inria Grenoble Rhône-Alpes & LJK, France
http://mistis.inrialpes.fr/people/girard/

June 2017

joint work with Laurent Gardes, Université de Strasbourg.
1 Weibull-tail distributions

2 Estimation of extreme conditional quantiles

3 Estimation of the functional Weibull-tail coefficient

4 Illustration on simulations
Let \((X, Y) \in E \times \mathbb{R}\) be a random pair where \(E\) is an arbitrary space associated with a semi-metric (or pseudometric) \(d\), see [3], Definition 3.2.

The conditional survival function of \(Y\) given \(X = x \in E\) is denoted by \(\bar{F}(y|x) := \mathbb{P}(Y > y|X = x)\) and is supposed to be continuous and strictly decreasing with respect to \(y\).

The associated conditional cumulative hazard function is defined by \(H(y|x) := -\log \bar{F}(y|x)\).

The conditional quantile is given by \(q(\alpha|x) := \bar{F}^{-1}(\alpha|x) = H^{-1}(\log(1/\alpha)|x)\), for all \(\alpha \in (0, 1)\).
(A.1) \( H(\cdot|x) \) is supposed to be regularly varying with index \( 1/\theta(x) \), i.e.

\[
\lim_{y \to \infty} \frac{H(ty|x)}{H(y|x)} = t^{1/\theta(x)},
\]

for all \( t > 0 \). In this situation, \( \theta(.) \) is a positive function of the covariate \( x \in E \) referred to as the functional Weibull tail-coefficient.

From [1], \( H^{-1}(\cdot|x) \) is regularly varying with index \( \theta(x) \). Thus, there exists a slowly-varying function \( \ell(\cdot|x) \) such that

\[
q(e^{-y}|x) = H^{-1}(y|x) = y^{\theta(x)} \ell(y|x).
\]

Recall that the slowly-varying function \( \ell(.)|x \) is such that

\[
\lim_{y \to \infty} \frac{\ell(ty|x)}{\ell(y|x)} = 1,
\]

(1)

for all \( t > 0 \).
Additional assumptions

(A.2) $\ell(.|x)$ is a normalised slowly-varying function.

In such a case, the Karamata representation (see [1]) of the slowly-varying function can be written as

$$
\ell(y|x) = c(x) \exp \left\{ \int_1^y \frac{\varepsilon(u|x)}{u} du \right\},
$$

where $c(x) > 0$ and $\varepsilon(u|x) \to 0$ as $u \to \infty$.

The function $\varepsilon(.|x)$ plays an important role in extreme-value theory since it drives the speed of convergence in (1) and more generally the bias of extreme-value estimators. Therefore, it may be of interest to specify how it converges to 0:

(A.3) $|\varepsilon(.|x)|$ is regularly varying with index $\rho(x) \leq 0$.

$\rho(x)$ is called the conditional second-order parameter.
Examples of (unconditional) Weibull-tail distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\theta$</th>
<th>$\ell(y)$</th>
<th>$\varepsilon(y)$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian $\mathcal{N}(\mu, \sigma^2)$</td>
<td>$1/2$</td>
<td>$\sqrt{2} \sigma - \frac{\sigma}{2 \sqrt{2}} \frac{\log y}{y} + O(1/y)$</td>
<td>$\frac{1}{4} \frac{\log y}{y}$</td>
<td>$-1$</td>
</tr>
<tr>
<td>Gamma $\Gamma(\alpha \neq 1, \lambda)$</td>
<td>$1$</td>
<td>$\frac{1}{\beta} + \frac{\alpha - 1}{\beta} \frac{\log y}{y} + O(1/y)$</td>
<td>$(1 - \alpha) \frac{\log y}{y}$</td>
<td>$-1$</td>
</tr>
<tr>
<td>Weibull $\mathcal{W}(\alpha, \lambda)$</td>
<td>$1/\alpha$</td>
<td></td>
<td></td>
<td>$-\infty$</td>
</tr>
</tbody>
</table>
Two goals

Starting from iid copies \((X_i, Y_i), \ i = 1, \ldots, n,\) of \((X, Y),\)

- Estimate the extreme conditional quantiles defined as

  \[
P(Y > q(\alpha_n, x)|X = x) = \alpha_n,\]

  when \(\alpha_n \to 0\) as \(n \to \infty.\)

- Estimate the functional Weibull-tail coefficient \(\theta(x).\)
1. Weibull-tail distributions

2. Estimation of extreme conditional quantiles

3. Estimation of the functional Weibull-tail coefficient

4. Illustration on simulations
First, $\hat{F}(y|x)$ is estimated by a kernel method. For all $(x, y) \in E \times \mathbb{R}$, let

$$\hat{F}_n(y|x) = \sum_{i=1}^{n} K(d(x, X_i)/h_n) \mathbb{I}\{Y_i > y\} / \sum_{i=1}^{n} K(d(x, X_i)/h_n),$$

where

- $h_n$ is a nonrandom sequence (called *bandwidth*) such that $h_n \to 0$ as $n \to \infty$,

- $K$ is assumed to have a support included in $[0, 1]$ such that $C_1 \leq K(t) \leq C_2$ for all $t \in [0, 1]$ and $0 < C_1 < C_2 < \infty$.

It is assumed without loss of generality that $K$ integrates to one. $K$ is called a *type I kernel*, see [3], Definition 4.1.

Second, $q(\alpha|x)$ is estimated via the generalized inverse of $\hat{F}_n(\cdot|x)$:

$$\hat{q}_n(\alpha|x) = \hat{F}_n^{-1}(\alpha|x) = \inf\{y, \hat{F}_n(y|x) \leq \alpha\},$$

for all $\alpha \in (0, 1)$. 

Notations:

- $B(x, h_n)$ the ball of center $x$ and radius $h_n$,
- $\varphi_x(h_n) := \mathbb{P}(X \in B(x, h_n))$ the small ball probability of $X$,
- $\mu^{(\tau)}_x(h_n) := \mathbb{E}\{K^\tau(d(x, X)/h_n)\}$ the $\tau$-th moment,
- $\Lambda_n(x) = (n\alpha_n(\mu^{(1)}_x(h_n))^2/\mu^{(2)}_x(h_n))^{-1/2}$.

It is easily shown that for all $\tau > 0$

$$0 < C_1^\tau \varphi_x(h_n) \leq \mu^{(\tau)}_x(h_n) \leq C_2^\tau \varphi_x(h_n),$$

and thus $\Lambda_n(x)$ is of order $(n\alpha_n\varphi_x(h_n))^{-1/2}$.
Since the considered estimator involves a smoothing in the $x$ direction, it is necessary to assess the regularity of the conditional survival function with respect to $x$. To this end, the oscillations are controlled by

$$\Delta \bar{F}(x, \alpha, \zeta, h) := \sup_{(x', \beta) \in B(x, h) \times [\alpha, \zeta]} \left| \frac{\bar{F}(q(\beta|x)|x')}{\bar{F}(q(\beta|x)|x)} - 1 \right|,$$

and

$$\Delta \bar{F}(x, \alpha, \zeta, h) := \sup_{(x', \beta) \in B(x, h) \times [\alpha, \zeta]} \left| \frac{\bar{F}(q(\beta|x)|x')}{\beta} - 1 \right|,$$

where $(\alpha, \zeta) \in (0, 1)^2$. 

11/ 29
Asymptotic normality

Theorem 1
Suppose (A.1), (A.2) hold.

- Let $0 < \tau_J < \cdots < \tau_1 \leq 1$ where $J$ is a positive integer.
- $x \in E$ such that $\phi_x(h_n) > 0$ where $h_n \to 0$ as $n \to \infty$.

If $\alpha_n \to 0$ and there exists $\eta > 0$ such that $n\phi_x(h_n)\alpha_n \to \infty$,

$$n\phi_x(h_n)\alpha_n(\Delta \bar{F})^2(x, (1 - \eta)\tau_J\alpha_n, (1 + \eta)\alpha_n, h_n) \to 0,$$

then, the random vector

$$\left\{ \log(1/\alpha_n)\Lambda_n^{-1}(x) \left( \frac{\hat{q}_n(\tau_j\alpha_n|x)}{q(\tau_j\alpha_n|x)} - 1 \right) \right\}_{j=1,\ldots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\theta^2(x)\Sigma$ where $\Sigma_{j,j'} = \tau_{j\wedge j'}^{-1}$ for $(j, j') \in \{1, \ldots, J\}^2$. 
Conditions on the sequences

\[ n \varphi_x(h_n) \alpha_n \to \infty: \text{Necessary and sufficient condition for the almost sure presence of at least one point in the region } B(x, h_n) \times [q(\alpha_n|x), +\infty) \text{ of } E \times \mathbb{R}. \]

\[ n \varphi_x(h_n) \alpha_n (\Delta \bar{F})^2(x, (1 - \eta)\tau_f \alpha_n, (1 + \eta)\alpha_n, h_n) \to 0: \text{The biais induced by the smoothing is negligible compared to the standard-deviation.} \]
1. Weibull-tail distributions

2. Estimation of extreme conditional quantiles

3. Estimation of the functional Weibull-tail coefficient

4. Illustration on simulations
We propose a family of estimators of $\theta(x)$ based on some properties of the log-spacings of the conditional quantiles. Recall that

$$q(e^{-y|x}) = H^{-1}(y|x) = y^\theta(x)\ell(y|x).$$

Let $\alpha \in (0, 1)$ small enough and $\tau \in (0, 1)$,

$$\log q(\tau\alpha|x) - \log q(\alpha|x)$$

$$= \log H^{-1}(-\log(\tau\alpha)|x) - \log H^{-1}(-\log(\alpha)|x)$$

$$= \theta(x)(\log_2(\tau\alpha) - \log_2(\alpha)) + \log \left( \frac{\ell(-\log(\tau\alpha)|x)}{\ell(-\log(\alpha)|x)} \right)$$

$$\approx \theta(x)(\log_2(\tau\alpha) - \log_2(\alpha))$$

$$\approx \theta(x) \frac{\log(1/\tau)}{\log(1/\alpha)},$$

where $\log_2(\cdot) := \log \log(1/\cdot)$,
Hence, for a decreasing sequence $0 < \tau_J < \cdots < \tau_1 \leq 1$, where $J$ is a positive integer, and for all (twice differentiable) function $\phi : \mathbb{R}^J \to \mathbb{R}$ satisfying the shift and location invariance condition

$$
\phi(\eta z) = \eta \phi(z), \\
\phi(\eta u + z) = \phi(z),
$$

for all $\eta \in \mathbb{R} \setminus \{0\}$, $z \in \mathbb{R}^J$ and where $u = (1, \ldots, 1)^t \in \mathbb{R}^J$, one has:

$$
\theta(x) \approx \log(1/\alpha) \frac{\phi(\log q(\tau_1 \alpha | x), \ldots, \log q(\tau_J \alpha | x))}{\phi(\log(1/\tau_1), \ldots, \log(1/\tau_J))}.
$$

Thus, the estimation of $\theta(x)$ relies on the estimation of conditional quantiles $q(\cdot | x)$:

$$
\hat{\theta}_n(x) = \log(1/\alpha_n) \frac{\phi(\log \hat{q}_n(\tau_1 \alpha_n | x), \ldots, \log \hat{q}_n(\tau_J \alpha_n | x))}{\phi(\log(1/\tau_1), \ldots, \log(1/\tau_J))},
$$

with $\alpha_n \to 0$ as $n \to \infty$. 
Suppose (A.1)–(A.3) hold. Let $x \in E$ such that $\varphi_x(h_n) > 0$ where $h_n \to 0$ as $n \to \infty$. If $\alpha_n \to 0$,

$$\sqrt{n \varphi_x(h_n) \alpha_n \varepsilon (\log(1/\alpha_n)|x)} \to \lambda \in \mathbb{R}$$

and there exists $\eta > 0$ such that $n \varphi_x(h_n) \alpha_n \to \infty$ and

$$\sqrt{n \varphi_x(h_n) \alpha_n \{\Delta \bar{F}(x, (1 - \eta) \tau_j \alpha_n, (1 + \eta) \alpha_n, h_n) \vee 1/\log(1/\alpha_n)\}} \to 0,$$

then,

$$\Lambda_n^{-1}(x)(\hat{\theta}_n(x) - \theta(x)) \xrightarrow{d} \mathcal{N}(\mu_\phi, \theta^2(x) V_\phi)$$

where $\mu_\phi = \lambda \nu^t \nabla \log \phi(\nu)$, $V_\phi = (\nabla \log \phi(\nu))^t \Sigma (\nabla \log \phi(\nu))$ and $\nu = (\log(1/\tau_1), \ldots, \log(1/\tau_J))^t$ do not depend on $(X, Y)$ distribution.
Corollary

Suppose (A.1)–(A.3) hold. Let $x \in E$ such that $\varphi_x(h_n) > 0$ and $y \varepsilon(y|x) \to \infty$ as $y \to \infty$. Assume there exist $L_\theta$, $L_c$ et $L_\varepsilon$ such that

\[
\left| \frac{1}{\theta(x)} - \frac{1}{\theta(x')} \right| \leq L_\theta d(x, x') ,
\]

\[
|\log c(x) - \log c(x')| \leq L_c d(x, x') ,
\]

\[
\sup_{u \in [1, \bar{y}_n(x)]} |\varepsilon(u|x) - \varepsilon(u|x')| \leq L_\varepsilon d(x, x') ,
\]

where $\bar{y}_n(x) := \sup\{ H(q(\alpha_n|x)|x'), x' \in B(x, h_n) \}$. Suppose

\[
\varphi_x^{-1}(1/y)(\log y)^{1+\xi-\rho(x)} \to 0 
\] (2)

for some $\xi > 0$ as $y \to \infty$. Then, letting $\lambda > 0$,

\[
\alpha_n = n^{-1+\xi} \text{ and } h_n = \varphi_x^{-1} \left( \lambda(1 - \xi)^2 \rho(x) n^{-\xi} (\varepsilon(\log n|x))^{-2} \right) ,
\]

Theorem 2 yields $\Lambda_n^{-1}(x)(\hat{\theta}_n(x) - \theta(x)) \xrightarrow{d} \mathcal{N}(\mu_\phi, \theta^2(x)V_\phi)$.

The key assumption (2) holds in the finite dimensional setting or for fractal-type and some exponential-type processes, see [3], Chapter 13.
Let us focus on the functions $\phi^{(p)}(z) = \left( \sum_{j=2}^{J} \beta_j (z_j - z_1)^p \right)^{1/p}$, where $z = (z_1, \ldots, z_J)^t \in \mathbb{R}^J$, $p \in \mathbb{N} \setminus \{0\}$ and for all $j \in \{2, \ldots, J\}$, $\beta_j \in \mathbb{R}$.

The corresponding estimator of $\theta$ writes:

$$\hat{\theta}^{(p)}(x) = \log(1/\alpha_n) \left( \frac{\sum_{j=2}^{J} \beta_j \left[ \log \hat{q}_n(\tau_j \alpha_n | x) - \log \hat{q}_n(\tau_1 \alpha_n | x) \right]^p}{\sum_{j=2}^{J} \beta_j [\log(\tau_1 / \tau_j)]^p} \right)^{1/p}.$$

As a consequence of Theorem 2, the associated asymptotic mean and variance of $\hat{\theta}^{(p)}(x)$ are given for an arbitrary vector $\beta$ by $\mu = \lambda$ and

$$V^{(p)} = \frac{(\eta^{(p)})^t A \Sigma A^t \eta^{(p)}}{(\eta^{(p)})^t A \Sigma \Sigma A^t \eta^{(p)}}.$$

where $A$ is a given matrix and $\eta^{(p)} = (\beta_j (v_j - v_1), j = 2, \ldots, J)^t$.

- The asymptotic bias $\mu$ does not depend neither on $p$ and nor on the weights $\{\beta_j, j = 2, \ldots, J\}$.
- It is possible to minimize $V^{(p)}$ with respect to $\eta^{(p)}$. 
Proposition

The asymptotic variance of $\hat{\theta}_n^{(p)}(x)$ is minimal for $\eta^{(p)}$ proportional to

$$\eta_{\text{opt}} = (A\Sigma A^t)^{-1}A\nu$$

and is given by

$$V_{\text{opt}} = \frac{1}{(A\nu)^t (A\Sigma A^t)^{-1}A\nu},$$

and is independent of $p$.

Moreover, for a fixed value of $J$, it is possible to minimize numerically the optimal variance $V_{\text{opt}}$ with respect to parameters $0 < \tau_J < \cdots < \tau_1 \leq 1$. The resulting values of $V_{\text{opt}}$ are displayed in the table below:

<table>
<thead>
<tr>
<th>$J$</th>
<th>$V_{\text{opt}}$</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$\tau_3$</th>
<th>$\tau_4$</th>
<th>$\tau_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.5441</td>
<td>1.0000</td>
<td>0.2032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.2191</td>
<td>1.0000</td>
<td>0.3615</td>
<td>0.0735</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.1223</td>
<td>1.0000</td>
<td>0.4703</td>
<td>0.1702</td>
<td>0.0346</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.0789</td>
<td>1.0000</td>
<td>0.5486</td>
<td>0.2585</td>
<td>0.0936</td>
<td>0.0190</td>
</tr>
</tbody>
</table>
1. Weibull-tail distributions

2. Estimation of extreme conditional quantiles

3. Estimation of the functional Weibull-tail coefficient

4. Illustration on simulations


- $E$ is a subset of $L^2([0, 1])$ made of trigonometric functions $\psi_z : [0, 1] \to [0, 1]$, $\psi_z(t) = \cos(2\pi z t)$ with different periods indexed by $z \in [1/10, 1/2]$.

- Two semi-metrics are considered:

  $$d_1(\psi_z, \psi_{z'}) = ||\psi_z||_2^2 - ||\psi_{z'}||_2^2,$$
  $$d_2(\psi_z, \psi_{z'}) = ||\psi_z - \psi_{z'}||_2,$$

for all $(z, z') \in [1/10, 1/2]^2$, where

$$||\psi_z||_2^2 = \int_0^1 \psi_z^2(t) dt = \frac{1}{2} \left( 1 + \frac{\sin(4\pi z)}{4\pi z} \right).$$

The semi-metric $d_2$ is built on the classical $L_2$ norm while $d_1$ measures some spacing between the periods of the trigonometric functions.
Simulated data

$N = 100$ copies of a sample of size $n = 1000$ from a random pair $(X, Y)$ defined as follows:

- The covariate $X$ is chosen randomly on $E$ by considering $X = \psi Z$ where $Z$ is a uniform random variable on $[1/10, 1/2]$.

- For a fixed function $x \in E$, the generalized inverse of the conditional hazard function $H(.|x)$ is given by the following Hall’s model:

\[
H^{-1}(y|x) = y^{\theta(x)} \left(1 - \gamma y^{\rho(x)}\right), \quad y \geq 0,
\]

with

\[
\theta(x) = \left(\frac{18}{5}||x||_2^2 + \frac{9}{50}\right)^{-1} - \frac{5}{18},
\]

\[
\rho(x) = \frac{50}{(60||x||_2^2 + 3)} - \frac{5}{2},
\]

\[
\gamma = \frac{1}{10}.
\]
Four simulated random functions $X(.)$
The previous estimator with optimal weights is used. Here, we limit ourselves to $J = 5$ and $p \in \{1, 3\}$.

A modified bi-quadratic kernel is adopted (type I kernel):

$$K(u) = \frac{10}{9} \left( \frac{3}{2} (1 - u^2)^2 + \frac{1}{10} \right) \mathbb{I}\{|u| \leq 1\}.$$

$h_n$ and $\alpha_n$ are selected simultaneously thanks to a data-driven procedure. For a fixed $x$, let $\{Z_1(x, h_n), \ldots, Z_m(x, h_n)\}$ be the $m_n$ random values $Y_i$ for which $X_i \in B(x, h_n)$. The idea [7] is to select the sequences $h_n$ and $\alpha_n$ such that the rescaled log-spacings

$$i \log(m_n/i)(\log Z_{m_n-i+1,m_n}(x, h_n) - \log Z_{m_n-i,m_n}(x, h_n)),$$

$i = 1, \ldots, \lfloor m_n \alpha_n \rfloor$, are approximately $\text{Exp}(\theta(x))$ distributed. The “optimal” sequences are obtained by minimizing a Kolmogorov-Smirnov distance.

Comparison with the non-conditional estimator proposed in [2]:

$$\hat{\theta}^{\text{NCE}}_n = \frac{\sum_{i=1}^{k_n} (\log Y_{n-i+1,n} - \log Y_{n-k_n+1,n})}{\sum_{i=1}^{k_n} (\log_2(n/i) - \log_2(n/k_n))}.$$
Influence of the exponent $p$

Let $e_{i,\ell,p}$ be the relative error obtained on the $i$th replication using the semi-metric $d_\ell$ and the estimator $\hat{\theta}(p)$.

Left: histogram of $e_{\bullet,1,3} - e_{\bullet,1,1}$ (semi-metric $d_1$), right: histogram of $e_{\bullet,2,3} - e_{\bullet,2,1}$ (semi-metric $d_2$).

Both histograms are nearly centered, small influence of $p$. 
Recall that $e_{i,\ell,p}$ is the relative error obtained on the $i$th replication using the semi-metric $d_\ell$ and the estimator $\hat{\theta}(p)$.

Left: histogram of $e_{\bullet,2,1} - e_{\bullet,1,1}$ ($p = 1$), right: histogram of $e_{\bullet,2,3} - e_{\bullet,1,3}$ ($p = 3$). Both histograms are skewed to the right, the semi-metric $d_1$ yields better result than $d_2$. 
Comparison with the non-conditional estimator

Recall that $e_{i,\ell,p}$ is the relative error obtained on the $i$th replication using the semi-metric $d_\ell$ and the estimator $\hat{\theta}(p)$. We moreover denote by $e_i$ the relative error obtained on the $i$th replication using the non-conditional estimator $\hat{\theta}_n^{NCE}$.

Left: histogram of $e_\bullet - e_{\bullet,1,1}$ ($p = 1$), right: histogram of $e_\bullet - e_{\bullet,1,3}$ ($p = 3$).
Both histograms are skewed to the right, the conditional estimator yields better results than the unconditional one.
References


