On the Convergence of the ET Method for Extreme Upper Quantile Estimation

Stéphane Girard, Jean Diebolt

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the convergence of the ET method for extreme upper quantile estimation

Stéphane Girard, Jean Diebolt

N° 3389

Mars 1998

THÈME 4

Rapport de recherche

ISSN 0249-6399

On the convergence of the ET method for extreme upper quantile estimation

Stéphane Girard, Jean Diebolt

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet is2

Rapport de recherche n° 3389 — Mars 1998 — 16 pages

Résumé : We examine the consistency of the Exponential Tail (ET) nonparametric method for estimating extreme quantiles of an unknown distribution. We show that, in general, the consistency of ET imposes strong limitations on the rate of convergence to 0 of the estimated quantile order.

Mots-clé : Exponential Tail, Extreme quantiles, Nonparametric.

(Abstract: pto)

This work is supported by the Direction des Etudes et Recherches (DER) of Electricité de France (EDF). We thank V. Durbec and B. Villain for helpful discussions.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Convergence de la méthode ET pour l'estimation de quantiles extrêmes

Abstract: Nous étudions la consistance de la méthode non-paramétrique ET (Exponential Tail) pour estimer les quantiles extrêmes de distributions inconnues. Nous montrons que, en général, la consistance de ET impose de fortes limitations sur les taux de convergence vers zéro des ordres des quantiles estimés.

Key-words: Queues exponentielles de distributions, Quantiles extrêmes, Non-paramétrique.

1 Introduction

We examine the consistency of a basic nonparametric method for estimating extreme quantiles of an unkown distribution. Given a n-sample, an extreme quantile is a $(1-\alpha_n)$-th quantile q_{α_n} of F, larger than the maximal observation, i.e. $\alpha_n \leq n^{-1}$. We say that an estimate sequence $\hat{q}_{\alpha_n}$ of q_{α_n} is consistent if the relative error $\varepsilon_n = (\hat{q}_{\alpha_n} - q_{\alpha_n})/q_{\alpha_n}$ converges to 0 in distribution.

Let $x_1, \dots, x_n$ be observations of i.i.d. random variables from a distribution F in the Gumbel attraction domain, and denote $x_{(1)} \leq \ldots \leq x_{(n)}$ the corresponding ordered sample. The Exponential Tail (ET) method (Breiman, Stone and Kooperberg 1990) computes an estimate $\hat{q}_{\alpha_n}$ of the $(1-\alpha_n)$-th quantile q_{α_n}, $0 < \alpha_n \leq n^{-1}$, of F. ET relies on an exponential approximation $\exp(-y/\sigma_n)$ of the tail distribution $1 - F_{u_n}(y)$ of $Y = X - u_n$ conditional on $X \geq u_n$, $X \sim F$, where $u_n = (1-F)^{-1}(m_n/n)$. In practice, $\hat{u}_n = x_{(n-m_n)}$, and we estimate σ_n by the empirical mean of the m_n excesses $y_{(j)} = x_{(j)} - \hat{u}_n$, $1 \leq j \leq m_n$: $\hat{\sigma}_n = m_n^{-1} \sum_{j=1}^{m_n} y_{(j)}$ with $m_n \to +\infty$ and $\lim_{n\to\infty} m_n/n = 0$. The ET estimate is then $\hat{q}_{\alpha_n} = \hat{u}_n - \hat{\sigma}_n \ln(\alpha_n/c_n)$ where $c_n = m_n/n$.

We investigate conditions on F and the rates of convergence to 0 of α_n and c_n under which $\varepsilon_n \xrightarrow{d} 0$. We focus on the most usual distributions in the Gumbel attraction domain. For simplicity, we only consider standardized forms of these distributions. To obtain reasonably general results, we write them as $1 - F(x) = \rho(x)\exp(-H(x))$ with ρ and $H \in \mathcal{C}^2(I\!R)$, $\lim_{x\to+\infty} \rho(x) = 1$, and $H(x)$ ultimately increasing to $+\infty$. For the considered standardized usual distributions, ρ and H take the following forms:

Weibull distribution: $H(x) = x^\beta, \rho(x) = 1$,
Normal distribution: $H(x) = x^2/2 + \ln x + (2\pi)^{1/2}, \rho(x) = 1 - x^{-2} + \mathcal{O}(x^{-4})$,
Lognormal distribution: $H(x) = (\ln x)^2/2 + \ln\ln x + \ln(2\pi)^{1/2}, \rho(x) = 1 + \ln^{-2} x + \mathcal{O}(\ln^{-4} x)$,
Gamma distribution: $H(x) = x + (1-\mu)\ln x + \ln\Gamma(\mu), \rho(x) = 1 + (\mu-1)x^{-1} + \mathcal{O}(x^{-2})$.

Furthermore, we limit ourselves to sequences $\alpha_n = n^{-(p+\eta_n)}$ and $c_n = n^{-(p'+\eta'_n)}$ where $\lim_{n\to\infty} \eta_n = \lim_{n\to\infty} \eta'_n = 0$ and $0 < p' \leq 1 \leq p$. Roughly speaking, our results show that, in general, $\varepsilon_n \xrightarrow{d} 0$ implies that $p = p' = 1$, and the converse is true in most cases. This sets a strong limitation on the order of the quantiles which can be consistently estimated through ET.

2 Assumptions and preliminary results.

We will need the following assumptions throughout the paper:

(A1) $\lim_{x\to+\infty} H''(x)/H'(x)^2 = 0.$
(A2) $\rho'(x) = \mathcal{O}(x^{-1})$ as $x \to +\infty$.
(A3) There exists a constant C such that $|H(t(1+\ell)) - H(t) - \ell t H'(t)| \le C\ell^2 t H'(t)$ for all t large enough and $|\ell| \le \ell_0$ for some $\ell_0 > 0$.

These assumptions hold for the distributions considered in Section 1. We will also need the following properties $(\mathbf{P}g)$ and $(\mathbf{Q}g)$.

$(\mathbf{P}g)$ $0 < \lim_{n\to\infty} \inf A_n/B_n \le \lim_{n\to\infty} \sup A_n/B_n < +\infty$
$\Rightarrow 0 < \lim_{n\to\infty} \inf g(A_n)/g(B_n) \le \lim_{n\to\infty} \sup g(A_n)/g(B_n) < +\infty.$
$(\mathbf{Q}g)$ $\lim_{n\to\infty} A_n/B_n = 1 \Rightarrow \lim_{n\to\infty} g(A_n)/g(B_n) = 1.$

The properties $(\mathbf{P}H^{-1})$, $(\mathbf{P}H)$, $(\mathbf{P}H')$, $(\mathbf{P}H'')$, $(\mathbf{Q}H^{-1})$, $(\mathbf{Q}H)$, $(\mathbf{Q}H')$ and $(\mathbf{Q}H'')$ are true for the Normal, Weibull, Exponential and Gamma distributions. For the Lognormal distribution, $(\mathbf{P}H^{-1})$ and $(\mathbf{Q}H^{-1})$ are not true.
The proofs of the following results are straightforward.

Lemma 2.1
Under **(A1)**, *there exists an increasing function* $A \in \mathcal{C}^3(\mathbb{R})$ *such that* $H(x) = A(x)\ln x$.

Lemma 2.2 *Under* **(A1)**, *we have that:*

(i) $xH'(x) \to +\infty$ *as* $x \to +\infty$,
(ii) $\lim_{x\to+\infty} \exp(-H(x))/H'(x) = 0.$

Proposition 2.1
Under **(A1)** *and* **(A2)**, $(1-F(x))^{-1} \int_x^{+\infty} (1-F(y))dy = 1/H'(x)(1+o(1))$ *as* $x \to +\infty$.

A result of Pickands (1975), shows that the exponential approximation to $1 - F_{u_n}(y)$ is valid when F is in the Gumbel attraction domain for extreme values (e.g. Galambos, 1987).

Proposition 2.2 *Under* **(A1)**–**(A3)**,

(i) the probability distributions are in the Gumbel attraction domain,
(ii) for all constant $D > 0$, $\lim_{u\to+\infty} \sup_{0\le y\le D/H'(u)} |1 - F_u(y) - \exp(-yH'(u))| = 0.$

3 Main results.

We split the relative error into $\varepsilon_n = \varepsilon_{est_n} + \varepsilon_{app_n} + \varepsilon_{est_n}\varepsilon_{app_n} \simeq \varepsilon_{est_n} + \varepsilon_{app_n}$, where $\varepsilon_{app_n} = (\tilde{q}_{\alpha_n} - q_{\alpha_n})/q_{\alpha_n}$ is the relative approximation error of q_{α_n} by $\tilde{q}_{\alpha_n} = u_n + \sigma_n \ln(\alpha_n/c_n)$ due to the exponential tail approximation, and $\varepsilon_{est_n} = (\hat{q}_{\alpha_n} - \tilde{q}_{\alpha_n})/\tilde{q}_{\alpha_n}$ is the relative estimation error due to the estimation of the parameters u and σ.

3.1 Approximation error

Theorem 1

Assume that $0 \le \alpha_n \le n^{-1}$, $c_n = m_n/n$ with $\{m_n\}$ an increasing integer-valued sequence such that $1 \le m_n \le n$, $m_n \to +\infty$ and $\lim_{n\to\infty} c_n = 0$, $-\ln \alpha_n = (p+\eta_n)\ln n$ and $-\ln c_n = (p' + \eta'_n)\ln n$ for some sequences $\{\eta_n\}$ and $\{\eta'_n\}$ such that $\lim_{n\to\infty} \eta_n = \lim_{n\to\infty} \eta'_n = 0$.

(i) Convergence to 0 of ε_{app_n}.

- *Normal, Lognormal and Weibull distributions ($\beta \neq 1$):*

$$\lim_{n\to\infty} \varepsilon_{app_n} = 0 \Rightarrow p = p' = 1. \tag{1}$$

The converse is true for the Normal and Weibull distributions, and under the condition $\lim_{n\to\infty} (\eta_n - \eta'_n)\ln^{1/2} n = 0$ for the Lognormal distribution.

- *Gamma distribution:*

$$\lim_{n\to\infty} \varepsilon_{app_n} = 0 \text{ for all } 0 < p' \le 1 \le p. \tag{2}$$

(ii) If in addition $\eta_n = q \ln\ln n/\ln n - \ln k/\ln n$ and $\eta'_n = q' \ln\ln n/\ln n - \ln k'/\ln n$ for some $q \ge 0$, $q' \le 0$, $k > 0$ and $k' > 0$, then the rates of convergence of ε_{app_n} are asymptotic to

Distribution	*1st-order asympt. to ε_{app_n}*	*Condition*
Weibull	$\dfrac{\beta-1}{2\beta^2}(q-q')^2\left(\dfrac{\ln\ln n}{\ln n}\right)^2$	$p = p' = 1$
Normal	$\dfrac{1}{8}(q-q')^2\left(\dfrac{\ln\ln n}{\ln n}\right)^2$	$p = p' = 1$
Lognormal	$\dfrac{1}{4}(q-q')^2\dfrac{(\ln\ln n)^2}{\ln n}$	$p = p' = 1$
Gamma	$\dfrac{(\mu-1)\ln(p/p')}{p\ln n}$	$p \neq p'$

This result highlights the asymptotic limitations of the ET method. We conjecture that the methods based on the generalized extreme-value or Pareto distributions suffer from similar limitations.
We give here the outline of the proof. Proofs of Lemmas can be found in appendices.

Sketch of the proof.

Step 1. We first assume that $\rho(x) = 1$ for all x large enough. We denote $\bar{q}_{\alpha_n} = H^{-1}(-\ln \alpha_n)$, $\bar{u}_n = H^{-1}(-\ln c_n)$, $\bar{\tilde{q}}_{\alpha_n} = \bar{u}_n + H'(\bar{u}_n)^{-1} \ln(\alpha_n / c_n)$ (the corresponding approximation), and $\varepsilon^{(2)}_{app_n} = (\bar{\tilde{q}}_{\alpha_n} - \bar{q}_{\alpha_n})/q_{\alpha_n}$. We define $K(x) = H^2(x) H''(x)/(x {H'}^3(x))$.

Lemma 3.1

1. *Suppose* $(\mathbf{P}H^{-1})$ *is true.*
   (i) If $0 < \lim_{x\to+\infty} \inf K(x) \le \lim_{x\to+\infty} \sup K(x) < +\infty$ *then*
$$\lim_{n\to\infty} \varepsilon^{(2)}_{app_n} = 0 \Rightarrow p = p' = 1. \tag{3}$$
   (ii) If $\lim_{x\to+\infty} K(x) = 0$ *then*
$$\lim_{n\to\infty} \varepsilon^{(2)}_{app_n} = 0 \textit{ for all } p \textit{ and } p' \textit{ such that } 0 < p' \le 1 \le p. \tag{4}$$
2. *If* F *is the Lognormal distribution, then (3) is still true.*
3. *If* $(\mathbf{P}H^{-1})$, $(\mathbf{Q}H^{-1})$ *and* $(\mathbf{Q}K)$ *are true then*
$$\textit{if } p = p' = 1, \varepsilon^{(2)}_{app_n} \sim \frac{1}{2}(\eta_n - \eta'_n)^2 K\left(H^{-1}(\ln n)\right) \textit{ as } n \to +\infty. \tag{5}$$
4. *If* F *is the Lognormal distribution,* $p = p' = 1$ *and* $\lim_{n\to\infty} (\eta_n - \eta'_n) \ln^{1/2} n = 0$ *then*
$$\varepsilon^{(2)}_{app_n} \sim \frac{1}{4}(\eta_n - \eta'_n)^2 \ln n \to 0 \textit{ as } n \to +\infty. \tag{6}$$

Step 2. We now examine the contributions to the approximation error related to $\rho(x)$. We split ε_{app_n} into $\varepsilon_{app_n} = \varepsilon^{(1)}_{app_n} + \varepsilon^{(2)}_{app_n} + \varepsilon^{(3)}_{app_n}$, where $\varepsilon^{(1)}_{app_n} = (\tilde{q}_{\alpha_n} - \bar{\tilde{q}}_{\alpha_n})/q_{\alpha_n}$, $\varepsilon^{(3)}_{app_n} = (\bar{q}_{\alpha_n} - q_{\alpha_n})/q_{\alpha_n}$, and $\varepsilon^{(2)}_{app_n}$ has been introduced in Step 1.

Lemma 3.2 *Suppose that* **(A1)**–**(A3)** *hold, and either* F *satisfies* $(\mathbf{P}H')$ *and* $(\mathbf{P}H^{-1})$ *or* F *is Lognormal. Let* $y_n \to +\infty$ *as* $n \to +\infty$ *and define* $x_n = (H + \ln\rho)^{-1}(y_n)$, $\bar{x}_n = H^{-1}(y_n)$ *and* $d(x_n) = (x_n - \bar{x}_n)/x_n$. *Then* $d(x_n) = \mathcal{O}(\ln \rho(x_n)/(x_n H'(x_n)))$ *and* $\lim_{n\to\infty} d(x_n) = 0$.

Lemma 3.3 *Assume* **(A1)**–**(A3)** *and* **(P**H**)**.

(i) We have, for some v_n *between* u_n *and* $\bar{u}_n$*:*

$$\varepsilon^{(1)}_{app_n} = \mathcal{O}\left(d(u_n)\left[1 \;-\; H'^{-2}(v_n)H''(v_n)\ln\left(\alpha_n/c_n\right)\right]\right). \tag{7}$$

(ii) We have

$$\varepsilon^{(3)}_{app_n} = d(q_{\alpha_n}) \to 0 \ \textit{as}\ n \to +\infty. \tag{8}$$

(iii) If in addition either $\lim_{x\to+\infty} |H''(x)|H(x)H'^{-2}(x) < +\infty$ *or* F *is the Lognormal distribution then* $\lim_{n\to+\infty} \varepsilon^{(1)}_{app_n} = 0$.

Step 3. Next, we investigate the important special case where $0 < \alpha_n = kn^{-p}\ln^{-q} n < n^{-1}$ ($k > 0$, $p \geq 1$, $q \geq 0$), and $1 \leq m_n = k'n^{1-p'}\ln^{-q'} n \leq n$, ($k' > 0$, $0 \leq p' \leq 1$, $q' \leq 0$). We need to compare the orders of magnitude of $\varepsilon^{(1)}_{app_n}$, $\varepsilon^{(2)}_{app_n}$ and $\varepsilon^{(3)}_{app_n}$.

Lemma 3.4 *Assume* **(A1)**–**(A3)**.
Define $L(x) = H'^2(x)H''^{-1}(x)\ln^{-2} H(x)$ *and* $M(x) = H'^2(x)H''^{-1}(x)H^{-2}(x)$.

1. *If* **(P**H'**)**, **(P**H^{-1}**)** *are true and* $\lim_{x\to+\infty} |H''(x)|H(x)H'^{-2}(x) < +\infty$, *then*

   (i) if $p = p' = 1$ *then, with* $x_n = H^{-1}(r_n \ln n)$ *for some* $r_n \in [1 + \eta'_n, 1 + \eta_n]$,

   $$\varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \mathcal{O}(\ln\rho(u_n)L(x_n)) \ \textit{and}\ \varepsilon^{(3)}_{app_n}/\varepsilon^{(2)}_{app_n} = \mathcal{O}(\ln\rho(q_{\alpha_n})L(x_n)); \tag{9}$$

   (ii) if $p \neq p'$ *then, with* $x_n = H^{-1}(r_n \ln n)$ *for some* $r_n \in [p' + \eta'_n, p + \eta_n]$,

   $$\varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \mathcal{O}(\ln\rho(u_n)M(x_n)) \ \textit{and}\ \varepsilon^{(3)}_{app_n}/\varepsilon^{(2)}_{app_n} = \mathcal{O}(\ln\rho(q_{\alpha_n})M(x_n)). \tag{10}$$

2. *If* F *is Lognormal, then if* $p = p' = 1$, *we have* $\lim_{n\to\infty} \varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \lim_{n\to\infty} \varepsilon^{(3)}_{app_n}/\varepsilon^{(2)}_{app_n} = 0$.

Step 4. End of the proof of Theorem 1.

(i) According to Lemma 3.3, the convergence of the ET method reduces to the convergence of $\varepsilon^{(2)}_{app_n}$ to 0.

   – Normal and Weibull ($\beta \neq 1$) distributions: their functions H verify condition 1.(i) of Lemma 3.1 and $K(x)$ converges to a finite limit as $x \to +\infty$, hence (3). The converse follows from (5) in Lemma 3.1.

   – Lognormal distribution: see Lemma 3.1-2. and 3.1-4.

   – Gamma distribution: see Lemma 3.1-1.(ii) since $K(x) = \mathcal{O}(x^{-1})$.

(ii) We first establish that the order of $\varepsilon^{(2)}_{app_n}$ is larger than the order of $\varepsilon^{(1)}_{app_n}$ and $\varepsilon^{(3)}_{app_n}$.

- Weibull distribution: $\varepsilon^{(1)}_{app_n} = \varepsilon^{(3)}_{app_n} = 0$.
- Normal distribution: $(\mathbf{P}H^{-1})$, $(\mathbf{P}H)$ and $(\mathbf{P}H')$ are true, $\ln \rho(x) = \mathcal{O}(x^{-2})$ and $L(x) = \mathcal{O}(x^2 \ln^{-2} x)$. Therefore (9) in Lemma 3.4-1.(i) implies that $\lim_{n\to\infty} \varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \lim_{n\to\infty} \varepsilon^{(3)}_{app_n}/\varepsilon^{(2)}_{app_n} = 0$.
- Gamma distribution: the proof is similar with $\ln \rho(x) = \mathcal{O}(x^{-1})$, $M(x) = \mathcal{O}(1)$ and (10).
- Lognormal distribution. See Lemma 3.4-2.

Finally, since $\varepsilon_{app_n} = \varepsilon^{(2)}_{app_n}(1 + o(1))$ in all cases, it suffices to find first-order asymptotics to $\varepsilon^{(2)}_{app_n}$ as $n \to +\infty$. For the Normal and Weibull cases, see (5) in Lemma 3.1-3. For the Lognormal case, see (6) in Lemma 3.1-4. The Gamma case requires a direct computation.

3.2 Estimation error

Up to the first order, $\varepsilon_{est_n} = \varepsilon^u_{est_n} + \varepsilon^\sigma_{est_n}$ with

$$\varepsilon^u_{est_n} = \frac{\partial \tilde{q}_{\alpha_n}}{\partial u}\Delta u = \frac{\Delta u}{u_n} \text{ and } \varepsilon^\sigma_{est_n} = \frac{\partial \tilde{q}_{\alpha_n}}{\partial \sigma}\Delta \sigma = \frac{\ln(\alpha_n/c_n)}{u_n}\Delta\sigma, \tag{11}$$

where $\Delta u = \hat{u}_n - u_n$ and $\Delta\sigma = \hat{\sigma}_n - \sigma_n$. Recall that the parameters u and σ are estimated by $\hat{u}_n = x_{(n-m_n)}$ and $\hat{\sigma}_n = m_n^{-1} \sum_{i=1}^{m_n} \left(x_{(n-m_n+i)} - x_{(n-m_n)}\right)$.

Theorem 2 *Assume that* **(A1)**–**(A3)** *and* $(\mathbf{P}H'')$ *are true. Then*

$$\varepsilon^u_{est_n} = \frac{1}{\sqrt{m_n}} \left.\frac{1}{xH'(x)}\right|_{x=H^{-1}(-\ln c_n)} \xi_n \tag{12}$$

$$\varepsilon^\sigma_{est_n} = \frac{\ln(\alpha_n/c_n)}{\sqrt{m_n}} \left.\frac{1}{xH'(x)}\right|_{x=H^{-1}(-\ln c_n)} \xi'_n \tag{13}$$

where ξ_n *(resp.* ξ'_n*)* $\xrightarrow{d} \xi \sim \mathcal{N}(0,1)$ *(resp.* ξ'*).*

Since $xH'(x) \to +\infty$ as $x \to +\infty$, the variance of $\varepsilon^u_{est_n}$ converges to 0. If $\ln\alpha_n = -(p + \eta_n)\ln n$ and $\ln c_n = -(p' + \eta'_n)\ln n$, then the variance of $\varepsilon^\sigma_{est_n}$ converges to 0 whenever either $p' < 1$ or $p = p' = 1$ with $\eta_n = q \ln\ln n/\ln n + \ln k/\ln n$ and $\eta'_n = q' \ln\ln n/\ln n + \ln k'/\ln n$. Moreover, in such cases $\varepsilon^u_{est_n} = o(\varepsilon^\sigma_{est_n})$. The asymptotic forms of (12) and (13) for the Weibull, Normal, Lognormal and Gamma distributions are given in Table 1.

Distribution	1st-order asympt. to $\varepsilon^u_{est_n}$	1st-order asympt. to $\varepsilon^\sigma_{est_n}$	Condition
Weibull	$\dfrac{(\ln n)^{q'/2-1}}{\beta\sqrt{k'}}\xi_n$	$\dfrac{(q-q')(\ln\ln n)(\ln n)^{q'/2-1}}{\beta\sqrt{k'}}\xi'_n$	$p=p'=1$
Normal	$\dfrac{(\ln n)^{q'/2-1}}{2\sqrt{k'}}\xi_n$	$\dfrac{(q-q')(\ln\ln n)(\ln n)^{q'/2-1}}{2\sqrt{k'}}\xi'_n$	$p=p'=1$
Lognormal	$\dfrac{(\ln n)^{q'/2-1/2}}{\sqrt{2k'}}\xi_n$	$\dfrac{(q-q')(\ln\ln n)(\ln n)^{q'/2-1/2}}{\sqrt{2k'}}\xi'_n$	$p=p'=1$
Gamma	$\dfrac{n^{p'/2-1/2}(\ln n)^{q'/2-1}}{p'\sqrt{k'}}\xi_n$	$\dfrac{(p-p')n^{p'/2-1/2}(\ln n)^{q'/2}}{p'\sqrt{k'}}\xi'_n$	$p\neq p'$

TAB. 1 – *Asymptotic forms of (12) and (13)*

Here (13) is proved assuming that the excess random variables $Y_i^{(n)} = X_{(n-m_n+i)} - u_n$, $1 \leq i \leq m_n$, are i.i.d. $Exp(\sigma_n)$. However, this is only an approximation. Taking into account the possible effects of this approximation on the limiting distribution of ξ'_n is beyond the scope of the present paper. Contiguity theory (Le Cam and Yang, 1990) leads us to conjecture that additional limitations on the growth of m_n (e.g. $p = p' = 1$ for Gamma and limitations on $|q'|$ in all cases) can cancel those effects. This is the subject of our current research.

References

Breiman, L., Stone and C.J., Kooperberg, C. (1990), Robust confident bounds for extreme upper quantiles, *J. Statist. Comput. Simul.*, **37**, 127–149.

Galambos, J. (1987), *The asymptotic theory of extreme order statistics*, (R.E. Krieger publishing company).

Le Cam, L. and Yang, G.L. (1990), *Asymptotics in Statistics. Some Basic Concepts*, (Springer series in Statistics, Springer-Verlag, New York).

Pickands, J. (1975), Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, 119–131.

Reiss, R.-D. (1989), *Approximate distributions of order statistics. With applications to nonparametrics statistics*, (Springer series in Statistics, Springer-Verlag, New York).

Appendices

Proof of Lemma 3.1

Writing the Taylor expansion of $\varepsilon^{(2)}_{app_n}$, first-order terms vanish and we get

$$\begin{aligned}
\varepsilon^{(2)}_{app_n} &= \frac{\ln^2(\alpha_n/c_n)}{2H^{-1}(-\ln\alpha_n)}\left.\frac{H''}{H'^3}\right|_{H^{-1}(-\ln\tau_n)}, \quad -\ln\tau_n \in [-\ln c_n, -\ln\alpha_n],\\
&= \frac{(p-p'+\eta_n-\eta'_n)^2\ln^2 n}{2H^{-1}((p+\eta_n)\ln n)}\left.\frac{H''}{H'^3}\right|_{H^{-1}(r_n\ln n)}, \quad r_n \in [p'+\eta'_n, p+\eta_n],\\
&= \frac{1}{2}\left(\frac{p-p'+\eta_n-\eta'_n}{r_n}\right)^2 \frac{H^{-1}(r_n\ln n)}{H^{-1}((p+\eta_n)\ln n)}K(H^{-1}(r_n\ln n)), r_n \in [p'+\eta'_n, p+\eta_n].
\end{aligned}$$

1. Using $(\mathbf{P}H^{-1})$ with $A_n = r_n\ln n$ and $B_n = (p+\eta_n)\ln n$, we obtain that:

   $$a(p-p'+\eta_n-\eta'_n)^2 K\left(H^{-1}(r_n\ln n)\right) \le \varepsilon^{(2)}_{app_n} \le b(p-p'+\eta_n-\eta'_n)^2 K\left(H^{-1}(r_n\ln n)\right), \tag{14}$$

   for some $0 < a \le b < +\infty$.

   (i) If $0 < \lim_{x\to+\infty}\inf K(x) \le \lim_{x\to+\infty}\sup K(x) < +\infty$, then $a'(p-p'+\eta_n-\eta'_n)^2 \le \varepsilon^{(2)}_{app_n} \le b'(p-p'+\eta_n-\eta'_n)^2$ for some $0 < a' \le b' < +\infty$. Therefore, in this case $(\varepsilon^{(2)}_{app_n} \to 0$ as $n \to 0) \Rightarrow (p = p' = 1)$.

   (ii) If $\lim_{x\to+\infty} K(x) = 0$ then $\lim_{n\to\infty}\varepsilon^{(2)}_{app_n} = 0$ $\forall(p,p')$ such that $0 < p' \le 1 \le p$.

2. If F is Lognormal, we make use of the Taylor formula with integral remainder. We obtain:

   $$\varepsilon^{(2)}_{app_n} = \frac{1}{2}\ln n\int_{p'+\eta'_n}^{p+\eta_n}\frac{p+\eta_n-v}{v}\exp\left(\sqrt{2\ln n}(\sqrt{v}-\sqrt{p+\eta_n})\right)dv. \tag{15}$$

   A saddlepoint type method can then be used to show that $\varepsilon^{(2)}_{app_n}$ is bounded below if $p \ne p'$.

3. Assuming $(\mathbf{P}H^{-1})$, $(\mathbf{Q}H^{-1})$ and $(\mathbf{Q}K)$ and $p = p' = 1$, (14) rewrites

   $$\varepsilon^{(2)}_{app_n} \sim \frac{1}{2}\left(\eta_n-\eta'_n\right)^2 K\left(H^{-1}(\ln n)\right) \text{ as } n \to +\infty.$$

4. Applying a saddlepoint method as in 2 to (15) gives the result. □

Proof of Lemma 3.2

– Case when $(\mathbf{P}H')$ and $(\mathbf{P}H^{-1})$ hold.

$$\begin{aligned}\bar{x}_n &= H^{-1}[(H+\ln\rho)(x_n)]\\ &= H^{-1}[H(x_n)] + \left(H^{-1}\right)'[H(x_n)+\theta_n\ln\rho(x_n)]\ln\rho(x_n), \theta_n\in[0,1].\end{aligned}$$

Consequently,

$$x_n - \bar{x}_n = \frac{\ln\rho(x_n)}{H'\left[H^{-1}\left(H(x_n)+\theta_n\ln\rho(x_n)\right)\right]},$$

and it follows that

$$d(x_n) = \frac{\ln\rho(x_n)}{x_n H'\left[H^{-1}\left(\Theta_n H(x_n)\right)\right]}.$$

with $\Theta_n = 1+\theta_n\ln\rho(x_n)/H(x_n)$.
Since $\lim\limits_{n\to+\infty}\Theta_n = 1$, there exist $\varepsilon>0, N>0$ such that $\Theta_n\in[1-\varepsilon,1+\varepsilon]$ for $n\geq N$.
Applying $(\mathbf{P}H')$ and $(\mathbf{P}H^{-1})$ to $A_n = H(x_n)$ and $B_n=\Theta_n H(x_n)$ if follows that $d(x_n) = \mathcal{O}\left(\ln\rho(x_n)/(x_n H'(x_n))\right)$, and with Lemma 2.2 (i) we obtain $\lim\limits_{n\to+\infty} d(x_n)=0$.

– Lognormal case.
An asymptotic expansion of H^{-1} yields $d(x_n) = \ln\rho(x_n)(1+o(1))/\ln x_n$, and the conclusion follows as above.

□

Proof of Lemma 3.3

(i) We have

$$\varepsilon^{(1)}_{app_n} = \left(\frac{u_n-\bar{u}_n}{q_{\alpha_n}}\right)\left(1+\left(\frac{1}{H'(u_n)}-\frac{1}{H'(\bar{u}_n)}\right)\ln(\alpha_n/c_n)\right).$$

There exists v_n between u_n et $\bar{u}_n$ such that

$$\begin{aligned}\varepsilon^{(1)}_{app_n} &= \left(\frac{u_n-\bar{u}_n}{q_{\alpha_n}}\right)\left(1-\frac{H''(v_n)}{H'^2(v_n)}\ln(\alpha_n/c_n)\right)\\ &= d(u_n)\frac{u_n}{q_{\alpha_n}}\left(1-\frac{H''(v_n)}{H'^2(v_n)}\ln(\alpha_n/c_n)\right).\end{aligned}$$

Then $u_n\leq q_{\alpha_n}$ implies

$$\varepsilon^{(1)}_{app_n} = \mathcal{O}\left(d(u_n)\left(1-\frac{H''(v_n)}{H'^2(v_n)}\ln(\alpha_n/c_n)\right)\right). \qquad (16)$$

(ii) By definition, $\varepsilon^{(3)}_{app_n} = d(q_{\alpha_n})$. Lemma 3.2 gives the result.

(iii) Let us write

$$\begin{aligned}\ln(\alpha_n/c_n) &= (p-p'+\eta_n-\eta'_n)\ln n\\ &= \frac{p-p'+\eta_n-\eta'_n}{p'+\eta'_n}\;\frac{H(H^{-1}((p'+\eta'_n)\ln n))}{H(v_n)}\;H(v_n)\\ &= \frac{p-p'+\eta_n-\eta'_n}{p'+\eta'_n}\;\frac{H(\bar{u}_n)}{H(v_n)}\;H(v_n).\end{aligned}$$

Applying $(\mathbf{P}H)$ with $A_n=\bar{u}_n$ and $B_n=v_n$, we get $\ln(\alpha_n/c_n)=\mathcal{O}(H(v_n))$. Thus

$$\frac{H''(v_n)}{H'^2(v_n)}\ln(\alpha_n/c_n)=\mathcal{O}\left(\frac{H''(v_n)H(v_n)}{H'^2(v_n)}\right). \tag{17}$$

Two cases appear:

- If $\lim\limits_{x\to+\infty}|H''(x)|H(x)H'^{-2}(x)<+\infty$, then $\varepsilon^{(1)}_{app_n}=\mathcal{O}(d(u_n))$ and the result follows from Lemma 3.2.
- Lognormal distribution: $\lim\limits_{x\to+\infty}|H''(x)|H(x)H'^{-2}(x)=+\infty$.

  Then, using (17), (16) rewrites:

  $$\varepsilon^{(1)}_{app_n}=\mathcal{O}\left(d(u_n)\frac{H''(v_n)H(v_n)}{H'^2(v_n)}\right),$$

  where $d(u_n)$ is given by Lemma 3.2. It results that

  $$\varepsilon^{(1)}_{app_n}=\mathcal{O}\left(\frac{\ln\rho(u_n)}{u_nH'(u_n)}\frac{H''(v_n)H(v_n)}{H'^2(v_n)}\right).$$

  After simplifications, we get

  $$\varepsilon^{(1)}_{app_n}=\mathcal{O}(\ln v_n/\ln^3 u_n).$$

  Lemma 3.2 shows that $\bar{u}_n=u_n(1+o(1))$. Therefore, since $v_n\in[u_n,\bar{u}_n]$, we have $\ln v_n\sim\ln u_n$, and

  $$\varepsilon^{(1)}_{app_n}=\mathcal{O}(\ln^{-2}u_n). \tag{18}$$

  Hence, $\lim\limits_{n\to+\infty}\varepsilon^{(1)}_{app_n}=0$.

□

Proof of Lemma 3.4

1. According to (14), $\varepsilon^{(2)}_{app_n}$ is bounded as follows:

$$a(p-p'+\eta_n-\eta'_n)^2 K\left(H^{-1}(r_n \ln n)\right) \le \varepsilon^{(2)}_{app_n} \le b(p-p'+\eta_n-\eta'_n)^2 K\left(H^{-1}(r_n \ln n)\right),$$

with $0 < a \le b < +\infty$ and $r_n \in [p'+\eta'_n, p+\eta_n]$. Besides, the proof of Lemma 3.3 shows that in this case $\varepsilon^{(1)}_{app_n} = \mathcal{O}(d(u_n))$ and $\varepsilon^{(3)}_{app_n} = d(q_{\alpha_n})$. Applying $(\mathbf{P}H')$ and $(\mathbf{P}H^{-1})$ with, in a first time $A_n = r_n$, $B_n = p'+\eta'_n$ and in a second time $A_n = r_n$, $B_n = p+\eta_n$ it follows that

$$\varepsilon^{(1)}_{app_n} = \mathcal{O}\left(\frac{\ln \rho(u_n)}{x_n H'(x_n)}\right) \text{ and } \varepsilon^{(3)}_{app_n} = \mathcal{O}\left(\frac{\ln \rho(q_{\alpha_n})}{x_n H'(x_n)}\right), \tag{19}$$

with $x_n = H^{-1}(r_n \ln n)$. Consequently, the proof is similar for $\varepsilon^{(1)}_{app_n}$ and $\varepsilon^{(3)}_{app_n}$. Let us consider the $\varepsilon^{(1)}_{app_n}$ case.

(i) If $p = p' = 1$, then

$$a(q-q')^2\left(\frac{\ln\ln n}{\ln n}\right)^2 K(x_n) \le \varepsilon^{(2)}_{app_n} \le b(q-q')^2\left(\frac{\ln\ln n}{\ln n}\right)^2 K(x_n), \tag{20}$$

with $r_n \in [1+\eta'_n, 1+\eta_n]$ and $x_n = H^{-1}(r_n \ln n)$.
Remark that $\ln n = H(x_n)/r_n$ and subsitute it in (20). For n large enough, there exist $0 < a' \le b' < +\infty$ such that

$$\frac{a'}{(q-q')^2}\,\frac{H^2(x_n)}{K(x_n)\ln H^2(x_n)} \le \frac{1}{\varepsilon^{(2)}_{app_n}} \le \frac{b'}{(q-q')^2}\,\frac{H^2(x_n)}{K(x_n)\ln H^2(x_n)}. \tag{21}$$

Finally, (19) and (21) provide $\varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \mathcal{O}(\ln \rho(u_n) L(x_n))$.

(ii) If $p \ne p'$, then (21) takes a different form:

$$\frac{a'}{(p-p')^2}\,\frac{1}{K(x_n)H^2(x_n)} \le \frac{1}{\varepsilon^{(2)}_{app_n}} \le \frac{b'}{(p-p')^2}\,\frac{1}{K(x_n)H^2(x_n)}.$$

As previously, we conclude with (19): $\varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \mathcal{O}(\ln \rho(u_n) M(x_n))$.

2. If F is Lognormal and $p = p' = 1$, then (6) shows that

$$\varepsilon^{(2)}_{app_n} \sim \frac{(q-q')^2}{4}\,\frac{(\ln\ln n)^2}{\ln n}.$$

Now, (18) can be rewritten as

$$\varepsilon^{(1)}_{app_n} = \mathcal{O}\left(\frac{1}{\ln n}\right),$$

and Lemma 3.2 provides

$$\varepsilon^{(3)}_{app_n} = \mathcal{O}\left(\frac{1}{\ln^{3/2} n}\right).$$

It follows that $\lim\limits_{n\to\infty} \varepsilon^{(1)}_{app_n}/\varepsilon^{(2)}_{app_n} = \lim\limits_{n\to\infty} \varepsilon^{(3)}_{app_n}/\varepsilon^{(2)}_{app_n} = 0.$

□

Proof of Theorem 2 (12)

We first prove that

$$\hat{u}_n - u_n = \frac{1}{f(u_n)} \frac{m_n^{1/2}}{n} \xi_n, \tag{22}$$

where $f = F'$. By Rényi's representation (Reiss, 1989),

$$\left(U_{(1)}, \ldots, U_{(n)}\right) \stackrel{d}{=} \left(\frac{T_1}{T_{n+1}}, \ldots, \frac{T_n}{T_{n+1}}\right),$$

where the U_i's are i.i.d. uniform over $[0,1]$, $U_{(1)} \le \ldots \le U_{(n)}$ is the corresponding ordered sample, and $T_k = \sum_{j=1}^{k} e_j$, $1 \le k \le n+1$, the e_j's being i.i.d. $Exp(1)$. Using the quantile transformation, $X_{(n-i+1)} = (1-F)^{-1}(U_{(i)})$, $1 \le i \le n$. Therefore,

$$\begin{aligned}
\hat{u}_n - u_n &= X_{(n-m_n)} - u_n \\
&\stackrel{d}{=} (1-F)^{-1}\left(U_{(m_n+1)}\right) - (1-F)^{-1}\left(\frac{m_n}{n}\right) \\
&\stackrel{d}{=} -\frac{1}{f(w_n)}\left(U_{(m_n+1)} - \frac{m_n}{n}\right),
\end{aligned}$$

where w_n is close to $(1-F)^{-1}(m_n/n) = u_n$. Then,

$$\begin{aligned}
U_{(m_n+1)} - \frac{m_n}{n} &= \frac{1}{n}\sum_{j=1}^{m_n}(e_j - 1) + \frac{e_{m_n+1}}{n} \\
&= \frac{m_n^{1/2}}{n}(\xi_n + o_{\mathbf{P}}(1)),
\end{aligned}$$

where

$$\xi_n = m_n^{-1/2}\sum_{j=1}^{m_n}(e_j - 1) \xrightarrow{d} \xi \sim \mathcal{N}(0,1),$$

by the Central Limit Theorem. Now,

$$f(u_n) = H'(u_n)\frac{m_n}{n}\left(1 \quad - \quad \frac{\rho'(u_n)}{\rho(u_n)H'(u_n)}\right).$$

Since $\lim_{x\to+\infty} \rho(x) = 1$ and by **(A2)**, $\rho'(u_n) = \mathcal{O}(1/u_n)$, it follows that

$$\frac{\rho'(u_n)}{\rho(u_n)H'(u_n)} = \mathcal{O}\left(\frac{1}{u_n H'(u_n)}\right) \to 0 \text{ as } n \to +\infty,$$

in view of Lemma 2.2 (i). Hence,

$$f(u_n) \sim \frac{m_n}{n} H'(u_n), \tag{23}$$

and

$$\hat{u}_n - u_n = \frac{1}{\sqrt{m_n} H'(u_n)} \xi_n.$$

□

Proof of Theorem 2 (13)

In this proof, we make the following approximation. The excess random variable's $\hat{Y}_i^{(n)} = X_{(n-m_n+i)} - \hat{u}_n$, $1 \le i \le m_n$ are considered as i.i.d. following an $Exp(1/H'(\hat{u}_n))$ distribution. The random variables $Z_i^{(n)} = H'(\hat{u}_n) Y_i^{(n)}$ are then i.i.d. $Exp(1)$. Therefore,

$$\begin{aligned} \hat{\sigma}_n &= \frac{1}{m_n} \sum_{i=1}^{m_n} Y_i^{(n)} \\ &= \frac{1}{H'(\hat{u}_n)} \left(1 + \frac{1}{m_n} \sum_{i=1}^{m_n} (Z_i^{(n)} - 1)\right) \\ &= \frac{1}{H'(\hat{u}_n)} + \frac{1}{H'(\hat{u}_n)\sqrt{m_n}} \xi_n', \end{aligned} \tag{24}$$

where

$$\xi_n' = m_n^{-1/2} \sum_{i=1}^{m_n} (Z_i^{(n)} - 1) \xrightarrow{d} \xi \sim \mathcal{N}(0,1).$$

Besides,

$$\begin{aligned} \frac{1}{H'(\hat{u}_n)} - \frac{1}{H'(u_n)} &= -\frac{H'(\hat{u}_n) - H'(u_n)}{H'(\hat{u}_n)H'(u_n)} \\ &= \frac{(\hat{u}_n - u_n)H''(z_n)}{H'(\hat{u}_n)H'(u_n)}, \end{aligned}$$

where z_n is between u_n and $\hat{u}_n$. By **($\mathbf{P}H''$)**, $H''(z_n) = \mathcal{O}_{\mathbf{P}}(H''(u_n))$. Then, using (22), we obtain

$$\frac{1}{H'(\hat{u}_n)} - \frac{1}{H'(u_n)} = \mathcal{O}_{\mathbf{P}}\left(\frac{H''(u_n)}{H'(\hat{u}_n)H'(u_n)} \frac{1}{f(u_n)} \frac{\sqrt{m_n}}{n}\right)$$

$$
\begin{aligned}
&= \mathcal{O}_{\mathbf{P}}\left(\frac{H''(u_n)}{H'^2(u_n)}\frac{1}{\sqrt{m_n}}\frac{1}{H'(\hat{u}_n)}\right) \text{ by (23)} \\
&= o_{\mathbf{P}}\left(\frac{1}{\sqrt{m_n}H'(\hat{u}_n)}\right) \text{ by } \textbf{(A1)}.
\end{aligned}
$$

Finally, replacing in (24), we obtain that

$$
\begin{aligned}
\hat{\sigma}_n &= \frac{1}{H'(u_n)}\left(1+\frac{\xi'_n + o_{\mathbf{P}}(1)}{\sqrt{m_n}}\right) \\
&= \sigma_n + \frac{1}{H'(u_n)}\frac{\xi'_n + o_{\mathbf{P}}(1)}{\sqrt{m_n}},
\end{aligned}
$$

and the conclusion follows. □

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399