Concentration results in extreme value theory

joint work with Stéphane Boucheron (LPMA - Univ. Paris Diderot)

Maud Thomas

Chalmers University of Technology
Gothenburg, Sweden

Workshop "Extremes - Copulas - Acturial Science"

CIRM - Marseille February 22-26, 2016

CHALMERS

Order statistics

- $X_{(1)} \geqslant \ldots \geqslant X_{(n)}$ order statistics of an i.i.d. sample $X_1, \ldots, X_n \sim F$.
- $Y_{(1)} \geqslant \ldots \geqslant Y_{(n)}$ order statistics of an **exponential** sample.

Rényi representation (Rényi 1953)

$$\left(Y_{(n)}, \ldots, Y_{(i)}, \ldots, Y_{(1)}\right) \stackrel{d}{=} \left(\frac{E_n}{n}, \ldots, \sum_{k=i}^{n} \frac{E_k}{k}, \ldots, \sum_{k=1}^{n} \frac{E_k}{k}\right)$$

where $E_1, \ldots, E_n$ are **independent** exponentially distributed random variables.

- $\mathbb{E}\left[Y_{(k)}\right] = \sum_{i=k}^{n} 1/i \sim \ln(n/k)$ and $\mathrm{Var}[Y_{(k)}] = \sum_{i=k}^{n} 1/i^2 \sim 1/k$

Quantile and tail quantile functions

$F^{\leftarrow}(p) = \inf\left\{x \colon F(x) \geqslant p\right\}, p \in (0,1)$ and $U(t) = F^{\leftarrow}(1 - 1/t), t \in (1, \infty)$

Representation for order statistics

$$\left(X_{(1)}, \ldots, X_{(n)}\right) \stackrel{d}{=} \left((U \circ \exp)(Y_{(1)}) \ldots (U \circ \exp)(Y_{(n)})\right) \quad .$$

Concentration of measure phenomenon

Concentration of measure phenomenon

Any function of many independent random variables that does not depend too much on any of them is concentrated around its mean value.

- Markov inequality: $X > 0$, $\mathbb{P}\{X > t\} \leqslant \mathbb{E}X/t$.
- Chebyschev inequality: $\mathbb{P}\{|X - \mathbb{E}X| > t\} \leqslant \operatorname{Var} X/t^2$.
- Gaussian vectors: $X \sim \mathcal{N}(0,1)$ and $Z = g(X)$
  - Poincaré inequality: $\operatorname{Var}[Z] \leqslant \mathbb{E}\left[||\nabla g||^2\right]$

Entropy

X positive v.a.

$$\operatorname{Ent}[X] = \mathbb{E}[X \ln X] - \mathbb{E}X \ln \mathbb{E}X$$

- Gross log-Sobolev inequality: $\operatorname{Ent}\left[Z^2\right] \leqslant 2\mathbb{E}||\nabla g||^2$
- Cirelson's inequality: if $||\nabla g|| \leqslant L$

$$\mathbb{P}\{Z \geqslant \mathbb{E}Z + t\} \leqslant \exp\left(-t^2/(2L^2)\right)$$

Gaussian order statistics

- $g(X_1, \ldots, X_n) = X_{(k)}$
  the kth order statistic is a smooth function of independent variables.
- $||\nabla g|| = 1$
- If $X_i \sim \mathcal{N}(0, 1)$,
  - Poincaré inequality $\Rightarrow$ $\text{Var}\left[X_{(k)}\right] \leqslant 1$
  - EVT $\Rightarrow$ $\text{Var}[X_{(1)}] = O(1/\ln n)$
  - Classical statistics $\Rightarrow$ $\text{Var}[X_{(n/2)}] = O(1/n)$

We do not understand (clearly) in which way the order statistics are smooth functions of the sample.

Variance bounds for order statistics

Poincaré inequality (Talagrand, 1991; Bobkov & Ledoux, 1997)

Let g be a differentiable function on $\mathbb{R}^n$, and $Z = g(E_1, \ldots, E_n)$ where $E_1, \ldots, E_n \sim_{\text{i.i.d.}} \mathcal{E}\text{xp}(1)$. Then

$$\text{Var}[Z] \leqslant 4\mathbb{E}\left[||\nabla g||^2\right] \; .$$

Application to order statistics

$$\text{Var}[X_{(k)}] \leqslant 4 \sum_{i=k}^{n} \frac{1}{i^2} \mathbb{E}\left[\frac{1}{h(X_{(k)})^2}\right] \leqslant \frac{4}{k}\left(1 + \frac{1}{k}\right) \mathbb{E}\left[\frac{1}{h(X_{(k)})^2}\right]$$

- **Hazard rate** h: if F has a positive density f, $h = f/(1-F)$
- Thanks to Efron-Stein's inequality (Efron & Stein, 1981 ; Steele, 1986), the factor 4 can be improved by a factor 2 if h is non-drecreasing (Boucheron & T., 2012)
- Tight variance bounds for Gaussian order statistics (Boucheron & T., (2012)).

Bernstein-type inequality for order statistics

Bernstein-type inequality (Bobkov & Ledoux, 1997)

Assume $\max_i |\partial_i g| < \infty$ and let v be the supremum of $||\nabla g||^2$. Then, for all $0 < \delta < 1/2$, with probability $\geqslant 1 - 2\delta$,

$$|Z - \mathbb{E}Z| \leqslant \sqrt{8v \ln(1/\delta)} + \max_i |\partial_i g| \ln(1/\delta) \ .$$

- $4v$ is the variance factor
- $\max_i |\partial_i g|$ is the scale factor.

Application to order statistics

If h is non-decreasing,, with probability $\geqslant 1 - 2\delta$,

$$|X_{(k)} - \mathbb{E}X_{(k)}| \leqslant \sqrt{\frac{8}{k}\left(1 + \frac{1}{k}\right) \mathbb{E}\left[\frac{1}{h(X_{(k)})^2}\right] \ln(1/\delta)} + \frac{\ln(1/\delta)}{k \inf_x h(x)} \ .$$

- Variance factor = variance bound
- Can be also obtained from the modified log-Sobolev inequality (Massart, 2000 ; Wu, 2000)

Model

- Fréchet domain: U is γ (>0)-regularly varying, *i.e.*

$$U(t) = c(t)t^{\gamma} \exp\left(\int_1^t \frac{\eta(s)}{s}\mathrm{d}s\right)$$

with $\lim_{t\to\infty} c(t) = c$ and $\lim_{t\to\infty} \eta(t) = 0$.

- Model: von Mises condition

$$U(t) = ct^{\gamma} \exp\left(\int_1^t \frac{\eta(s)}{s}\mathrm{d}s\right)$$

with c a constant and $\lim_{t\to\infty} \eta(t) = 0$.

- Hill estimator (1975), $\gamma > 0$

$$\widehat{\gamma}(k) = \frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{(i)}}{X_{(k+1)}} = \frac{1}{k}\sum_{i=1}^{k} i \ln \frac{X_{(i)}}{X_{(i+1)}} \quad .$$

Representations for Hill estimator

Under von Mises condition, Hill estimators $(\widehat{\gamma}(k))_{2 \leqslant k \leqslant n}$ represented as smooth functions of exponential variables.

Proposition (Boucheron & T., 2015)

- $$\widehat{\gamma}(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^{k} \int_0^{E_i} \left(\gamma + \eta(\mathrm{e}^{u + Y_{(k+1)}})\right) \mathrm{d}u \ .$$
- $$\widehat{\gamma}(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^{k} \int_0^{E_i} \left(\gamma + \eta(\mathrm{e}^{\frac{u}{i} + Y_{(i+1)}})\right) \mathrm{d}u \ .$$

with

- $E_1, \ldots, E_k$ independent exponentially distributed random variables
- $Y_{(i+1)}$ and $Y_{(k+1)}$ the $(i+1)$th and the $(k+1)$th largest order statistics of an exponential sample.

Hill Plot

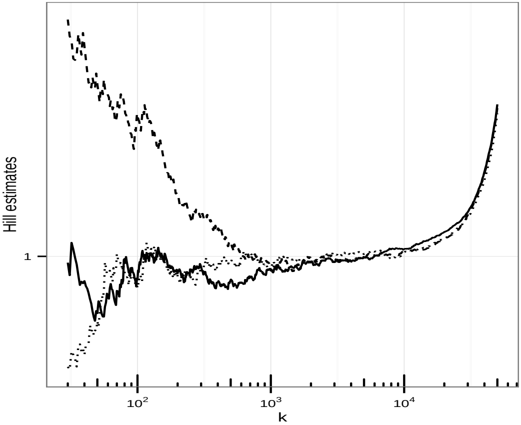


Figure: $\widehat{\gamma}(k)$ as a function of k computed on a sample of size 10^7 from Cauchy distribution ($\gamma = 1$)

Why an adaptive choice of k?

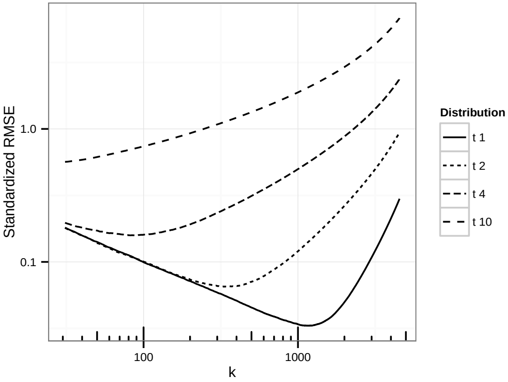


Figure: Estimated standardised RMSE as a function of k for samples of size 10^4 from Student's distributions with different degrees of freedom $\nu = 1, 2, 4, 10$.

Bias-variance dilemma

Problem

How to choose the number k of order statistics in the computation of $\widehat{\gamma}(k)$?

Risk of $\widehat{\gamma}(k)$

$$\mathbb{E}\left[(\gamma - \widehat{\gamma}(k))^2\right] = \mathrm{Var}\left[\widehat{\gamma}(k)\right] + \left(\gamma - \mathbb{E}\left[\widehat{\gamma}(k)\right]\right)^2 .$$

- Bias

$$\mathbb{E}\widehat{\gamma}(k) - \gamma = \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-v} \eta\left(\mathrm{e}^{Y_{(k+1)}}\mathrm{e}^{v}\right)\mathrm{d}v\right] = \mathbb{E}\left[\int_1^\infty \frac{\eta(\mathrm{e}^{Y_{(k+1)}}v)}{v^2}\mathrm{d}v\right] .$$

Optimal choice of k depends on η (unknown).

- Conditional bias given $\exp\left(Y_{(k+1)}\right) = t$

$$b(t) = t\int_t^\infty \frac{\eta(v)}{v^2}\mathrm{d}v$$

is a smooth function of $Y_{(k+1)}$.

Variance bounds for $\widehat{\gamma}(k)$

- $\sqrt{k_n}\left(\widehat{\gamma}(k_n) - \mathbb{E}\widehat{\gamma}(k_n)\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\gamma^2)$, if (k_n) intermediate.
- $\mathrm{Var}\left[\widehat{\gamma}(k_n)\right] \sim \gamma^2/k_n$.
- Define the non-increasing function $\overline{\eta}(t) = \sup_{s \geqslant t} |\eta(s)|$.

Proposition (Boucheron & T. (2015))

If F belongs to the Fréchet domain and satisfies the von Mises condition then

$$-\frac{2\gamma}{k}\mathbb{E}\left[\overline{\eta}\left(\mathrm{e}^{Y_{(k+1)}}\right)\right] \leqslant \mathrm{Var}[\widehat{\gamma}(k)] - \frac{\gamma^2}{k} \leqslant \frac{2\gamma}{k}\mathbb{E}\left[\overline{\eta}\left(\mathrm{e}^{Y_{(k+1)}}\right)\right] + \frac{5}{k}\mathbb{E}\left[\overline{\eta}\left(\mathrm{e}^{Y_{(k+1)}}\right)^2\right].$$

Proposition (Boucheron & T. (2015))

Assume η is regularly varying with index $\rho \leqslant 0$, then for all intermediate sequence $(k_n)_n$,

$$\lim_{n\to\infty} \frac{k_n \mathrm{Var}(\widehat{\gamma}(k_n)) - \gamma^2}{\eta(n/k_n)} = \frac{2\gamma}{(1-\rho)^2}.$$

Bernstein-type inequalities for $\widehat{\gamma}(k)$

- $\sqrt{\ln \log_2 n} \leqslant \ell \leqslant J \leqslant n$.
- $0 < \delta < 1/2$.
- $T = \exp\left(Y_{(J+1)}\right)$.

Proposition (Boucheron & T., 2015)

For $\ell \leqslant i \leqslant k \leqslant J$, with probability $\geqslant 1-\delta$,

$$|\widehat{\gamma}(i) - \mathbb{E}\left[\widehat{\gamma}(i) \mid T\right]| \leqslant \frac{(\gamma + 3\overline{\eta}(T))}{\sqrt{i}} \left(\sqrt{8s} + \frac{s}{\sqrt{i}}\right)$$

Proposition (Boucheron & T., 2015)

If $J\overline{\eta}(T) \leqslant 2\gamma^2 \ln \ln n$, the, for $\ell \leqslant k \leqslant J$, with probability $\geqslant 1-\delta$,

$$\begin{aligned}
\max_{\ell \leqslant i \leqslant n} \sqrt{i}\, |\widehat{\gamma}(i) - \mathbb{E}\left[\widehat{\gamma}(i) \mid T\right]| \quad & \leqslant \quad \gamma \left(1 + 3\sqrt{(2\ln\ln n)/J}\right) \left(c_1 \sqrt{2 \ln \log_2 n} + c_1'\right) \\
& + \left(1 + 3\sqrt{(2\ln\ln n)/J}\right) \sqrt{8\gamma^2 \ln(2/\delta)} \\
& + \gamma \left(1 + 3\sqrt{(2\ln\ln n)/J}\right) \frac{\ln(2/\delta)}{\sqrt{\ell}}
\end{aligned}$$

Adaptive estimation

- Hall and Welsh (1985): under the following assumption

$$\overline{F}(x) = Cx^{-1/\gamma}\left(1 + Dx^{\rho/\gamma} + o(x^{\rho/\gamma})\right) ,$$

 the optimal choice of k is equivalent to

$$K(C, D, \rho)n^{2|\rho|/(1+2|\rho|)}$$

 where C and D are constants > 0 and $\rho < 0$ **unknown** second order parameter.
- Drees et Kaufmann (1998)
- Grama and Spokoiny (2008)
- Carpentier and Kim (2014)

Minimax lower bound

For any $\widehat{\gamma}$, there exists a distribution P such that P belongs the Fréchet domain and $\overline{\eta}(t) \leqslant \gamma t^{\rho}$, then, with large probability,

$$|\widehat{\gamma} - \gamma| \geqslant C_\rho \gamma \left(\frac{v \ln \ln n}{n}\right)^{|\rho|/(1+2|\rho|)} .$$

Adaptive version of $\widehat{\gamma}_H$

- $\ell_n = \lceil c_2 \ln n \rceil$
- $k^\delta = k + \sqrt{2k \ln(1/\delta)} + 2 \ln(1/\delta)$
- $r_n = \sqrt{2 \ln \ln n}$
- $\widetilde{k}_n(r_n) = \max\left\{k \in \{\ell_n, \ldots, n\} : \sqrt{k}\overline{\eta}(n/k^\delta) \leqslant \gamma r_n\right\}$
- $z_{\delta,n} = c_1 \sqrt{2 \ln \log_2 n} + c_1' + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{\ell_n}}$
- $r_n(\delta) = 10\left(r_n + \left(1 + 3r_n/\sqrt{k_n}\right) z_{\delta,n}\right)$

Pivotal index

$$\widetilde{k}_n(r_n) = \max\left\{k \in \{\ell_n, \ldots, n\} : \sqrt{k}\overline{\eta}(n/k^\delta) \leqslant \gamma r_n\right\}$$

Selected index (Boucheron & T., 2015)

$$\widehat{k}_n = \max\left\{k : \ell_n \leqslant k \leqslant n \text{ et } \forall i \in \{\ell_n, \ldots, k\}, |\widehat{\gamma}(i) - \widehat{\gamma}(k)| \leqslant \frac{r_n(\delta)\widehat{\gamma}(i)}{\sqrt{i}}\right\}$$

Illustration

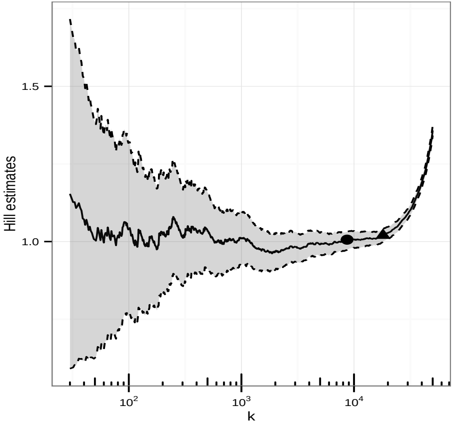


Figure: Cauchy distribution($\gamma = 1$). Hill estimators as functions of k and adaptive estimator for a sample of size 10^5 and $\delta = 1/20$.

Analysis of $\widehat{\gamma}(\widehat{k}_n)$

Theorem (Boucheron & T., 2015)

- With probability $\geqslant 1 - 3\delta$,

$$\left|\gamma - \widehat{\gamma}(k_n)\right| \leqslant \left|\gamma - \widehat{\gamma}(\widetilde{k}_n)\right| \left(1 + \frac{r_n(\delta)}{\sqrt{\widetilde{k}_n}}\right) + \frac{r_n(\delta)\gamma}{\sqrt{\widetilde{k}_n}}$$

- With probability $\geqslant 1 - 4\delta$,

$$|\gamma - \widehat{\gamma}(k_n)| \leqslant \frac{2r_n(\delta)\gamma}{\sqrt{\widetilde{k}_n}} (1 + \alpha(\delta, n))$$

where $\alpha(\delta, n) \to 0$ as $n \to \infty$.

Analysis of $\widehat{\gamma}(\widehat{k}_n)$

Corollary (Boucheron & T., 2015)

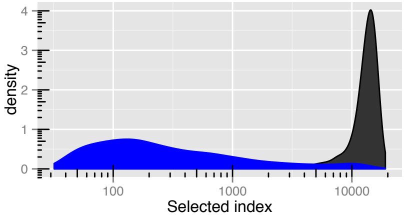
If there exist $C > 0$ and $\rho < 0$ such that for all n, k

$$|\gamma - \mathbb{E}\widehat{\gamma}(k)| \leqslant C\left(\frac{n}{k}\right)^{\rho}$$

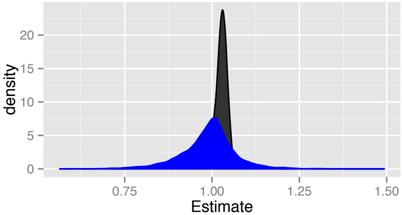
then, with probability $\geqslant 1 - 4\delta$,

$$\left|\gamma - \widehat{\gamma}(\widehat{k}_n)\right| \leqslant \kappa_{\delta,\rho}\left(\frac{\gamma^2 \ln((2/\delta)\ln n)}{n}\right)^{|\rho|/(1+2|\rho|)} (1 + \alpha(\delta, n))$$

A calibration problem for Cauchy distribution ($\gamma = 1$)

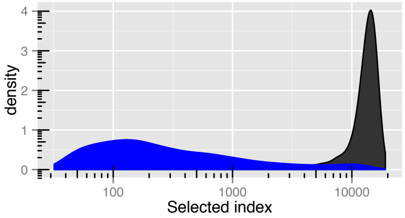


Distribution of the selected index with $r_n = \sqrt{3 \ln \ln(n)}$ (black area) and with $r_n = \sqrt{\ln(\ln(n))}$ (blue aera).

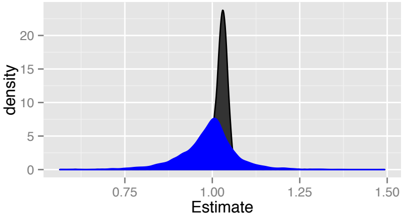


Distribution of $\widehat{\gamma}(\widehat{k}_n)$ with $r_n = \sqrt{3 \ln \ln(n)}$ (black area) and with $r_n = \sqrt{\ln(\ln(n))}$ blue aera).

A calibration problem for Cauchy distribution $(\gamma = 1)$



Distribution of the selected index with $r_n = \sqrt{3\ln\ln(n)}$ (black area) and with $r_n = \sqrt{\ln(\ln(n))}$ (blue aera).



Distribution of $\widehat{\gamma}(\widehat{k}_n)$ with $r_n = \sqrt{3\ln\ln(n)}$ (black area) and with $r_n = \sqrt{\ln(\ln(n))}$ blue aera).

Thank you for your attention!