Concentration inequalities for order statistics

Using the entropy method and Rényi's representation

S. Boucheron[1] M. Thomas[1]

[1]LPMA Université Paris-Diderot

Probability and Harmonic Analysis, Angers, September 2012

Background : order statistics

- Sample : $X_1, \ldots, X_n \sim_{\text{i.i.d.}} F$
- Order statistics

$X_{1,n} \geq \ldots \geq X_{n,n}$ non-increasing rearrengement of $X_1, \ldots, X_n$.
If n clear from context, $X_{1,n}, \ldots, X_{n,n}$ denoted by $X_{(1)}, \ldots, X_{(n)}$.

- $X_{(1)}$: sample maximum
- $X_{(n/2)}$: sample median ...
- Extreme value theory and classical statistics
  - *Asymptotic* distributions
  - Convergence of moments
  - ....
- Goal: derive simple, non-asymptotic variance/tail bounds for order statistics

Background : concentration

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.

A new (*non-asymptotic*) look at independence

- Example: Gaussian concentration

  (Bonami,Beckner, Nelson, Gross, Borell, Ehrhard, Bobkov, Ledoux, ...)

  - $X = (X_1, \ldots, X_n)$ a standard Gaussian vector
  - Poincaré's inequality: $\mathrm{Var}\, f(X) \leq \mathbb{E}\|\nabla f\|^2$
  - Gross logarithmic Sobolev inequality: $\mathrm{Ent}(f(X)^2) \leq 2\mathbb{E}\|\nabla f\|^2$
  - Cirelson's inequality: $\mathbb{P}\{f(X) \geq \mathbb{E}f(X) + t\} \leq \exp(-t^2/(2L^2))$ if $\|\nabla f\| \leq L$
- Product spaces: Talagrand's inequalities
- Order statistics are not (usually) sums of independent random variables

Off-the shelf concentration inequalities and order statistics

- $f(X_1, \ldots, X_n) = \max(X_1, \ldots, X_n)$:
  a simple function of many independent random variables that does not depend too much on any of them.
- Scenario : X_i are standard Gaussian
  - Almost surely, $\|\nabla f\| = 1$.
  - Poincaré's inequality $\Rightarrow$ $\mathrm{Var}(f(X_1, \ldots, X_n)) \leq 1$
  - Extreme Value Theory asserts : $\mathrm{Var}(\max(X_1, \ldots, X_n)) = O(1/\log n)$

We do not understand (clearly)

in which way the maximum is a smooth function of the sample.

Central, intermediate and extreme order statistics

$X_1, \ldots, X_n \sim_{\text{i.i.d.}} F$

Order statistics

$X_{1,n} \geq \ldots \geq X_{n,n}$ non-increasing rearrengement of $X_1, \ldots, X_n$.
If n clear from context, $X_{1,n}, \ldots, X_{n,n}$ denoted by $X_{(1)}, \ldots, X_{(n)}$.

$(X_{k,n})$ is a sequence of

extreme order statistics,	if k fixed, $n \to \infty$;
central order statistics,	if $k/n \to p \in (0,1)$ while, $n \to \infty$;
intermediate order statistics,	if $k/n \to 0$, $k \to \infty$.

Different asymptotics

Central and intermediate order statistics (often):	Gaussian
Extreme order statistics (sometimes):	Generalized Extreme Value

Variance bounds, order statistics and spacings

A connection

The variance (and more generally the higher moments) of the k^{th} order statistics can be upper-bounded by moments of the k^{th} *spacing* $X_{(k)} - X_{(k+1)}$.

Lemma (Jackknife bounds)

$$\mathrm{Var}[X_{(k)}] \leq k\mathbb{E}\left[\left(X_{(k)} - X_{(k+1)}\right)^2\right] .$$

Convention

$$\Delta_k = X_{(k)} - X_{(k+1)}$$

Proof (i)

Theorem (Efron-Stein inequalities, 1981)

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be measurable, and let $Z = f(X_1, \ldots, X_n)$.
Let $Z_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ where $f_i\colon \mathbb{R}^{n-1} \to \mathbb{R}$ is an arbitrary measurable function. Suppose Z is square-integrable.
Then

$$\mathrm{Var}[Z] \leq \mathbb{E}\left[\sum_{i=1}^{n} (Z - Z_i)^2\right] .$$

Efron-Stein inequalities provide a key ingredient in the derivation of Poincaré's inequality.

$\sum_{i=1}^{n} (Z - Z_i)^2$ is a jackknife estimate of variance.

Proof (ii)

- $Z = X_{(k)}$
- Z_i as the rank k statistic from subsample $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$:

$$Z_i = \begin{cases} Z_i = X_{(k+1)} & \text{if } X_i \geq X_{(k)} \\ Z_i = Z & \text{otherwise.} \end{cases}$$

- Jackknife estimate of variance of $X_{(k)}$:

$$\sum_{i=1}^{n} (Z - Z_i)^2 = \sum_{i: X_i \geq X_{(k)}} (X_{(k)} - X_{(k+1)})^2 = k\Delta_k^2$$

Asymptotic assessment for extreme order statistics

Definition (Quantile function)

$$F^{\leftarrow}(p) = \inf \{x : F(x) \geq p\}$$

Definition ($\mathrm{MDA}(\gamma), \gamma \in \mathbb{R}$)

$F \in \mathrm{MDA}(\gamma)$ if the exists a function $a : \mathbb{R}_+ \to \mathbb{R}_+$, such that

$$\mathbb{P}\left\{ \frac{\max(X_1, \ldots, X_n) - F^{\leftarrow}(1 - 1/n)}{a(n)} \leq x \right\} \to \exp\left(-(1 + \gamma x)^{-1/\gamma} \right)$$

according to the sign of extreme value index γ
$$\begin{cases} > 0 & \text{Frechet domain} \\ = 0 & \text{Gumbel domain} \\ < 0 & \text{Weibull domain} \end{cases}$$

Asymptotic assessment for extreme order statistics (ii)

If $F \in \mathrm{MDA}(\gamma)$ with $\gamma < 1/2$,

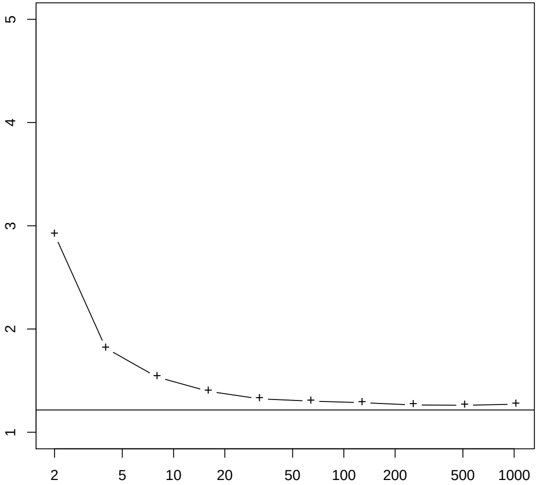
the ratio between the jackknife estimate and the variance converges toward a limit that depends on k and γ, for $k = 1$:

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\left(X_{(1)} - X_{(2)}\right)^2\right]}{\mathrm{Var}[X_{(1)}]} = \frac{\frac{2\Gamma(2(1-\gamma))}{(1-\gamma)(1-2\gamma)}}{\frac{\Gamma(1-2\gamma)-\Gamma(1-\gamma)^2}{\gamma^2}}$$

In the Guembel domain ($\gamma = 0$),

for $k = 1$, the limit is $12/\pi^2 \approx 1.2159$.

Graphical assessment



- Ratio between the Efron-Stein estimate and the variance of the maximum of n independent Gaussian random variables.
- $n = 2^p$ for $p = 1, \ldots, 10$.
- Empirical quantities evaluated over 5×10^6 replicates.
- The asymptote is the line $y = 12/\pi^2$.

Rényi's representation

The order statistics of an exponential sample ...

are partial sums of independent exponentially distributed random variables.

If $F(x) = 1 - \exp(-x)$ for $x > 0$, letting $X_{n+1,n} = 0$,

$$X_{k,n} = \sum_{i=k}^{n} (X_{i,n} - X_{i+1,n})$$

where the spacings $\Delta_i = (X_{i,n} - X_{i+1,n})_{i=1,\ldots,n}$ form an independent family of random variables and $i \times (X_{i,n} - X_{i+1,n}) \sim F$

Quantile transformation

Definition (Quantile function (bis))

$$F^{\leftarrow}(p) = \inf\{x \colon F(x) \geq p\}, p \in (0,1) \qquad U(t) = F^{\leftarrow}(1-1/t), t \in (1,\infty)$$

Representation for order statistics

If $Y_{(1)}, \ldots, Y_{(n)}$ are the order statistics of an exponential sample, then

$$\left(F^{\leftarrow}\left(1 - \exp(-Y_{(i)})\right)\right)_{i=1,\ldots,n}$$

is distributed as the order statistics of a sample drawn according to F.

Representation for order statistics

If $Y_{(1)}, \ldots, Y_{(n)}$ are the order statistics of an exponential sample, then

$$U(e^{Y_{(1)}}) \geq U(e^{Y_{(2)}}) \geq \ldots \geq U(e^{Y_{(n)}})$$

is distributed as the order statistics of a sample drawn according to F.

Hazard rate, spacings and order statistics

Definition (Hazard rate)

The hazard rate of a differentiable distribution function F is $F'/\overline{F} = F'/(1-F)$.

Lemma

The distribution function F has non-decreasing hazard rate, iff $U \circ \exp$ is concave.

Lemma

If the distribution function F has non-decreasing hazard rate, then $X_{(k+1)}$ and $\Delta_k = X_{(k)} - X_{(k+1)}$ are negatively associated.

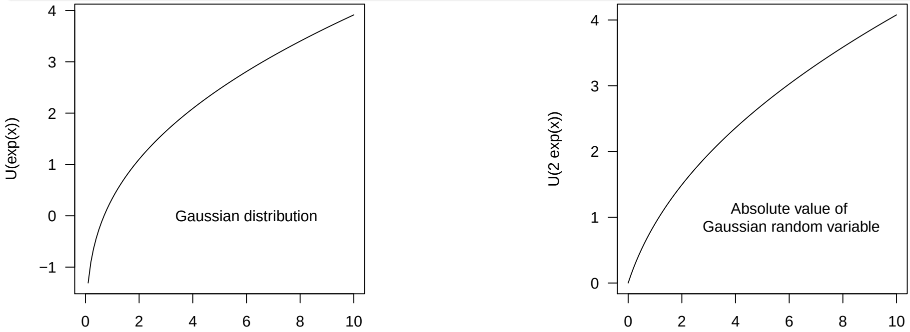
Negative association

For increasing functions f, g

$$\mathbb{E}\left[f(X_{(k+1)}g(\Delta_k)\right] \leq \mathbb{E}\left[f(X_{(k+1)}\right]\mathbb{E}\left[g(\Delta_k)\right]$$

Gaussian hazard rate

$U(t) = \Phi^{\leftarrow}(1 - 1/t)$ for $t > 1$.



Gaussian and absolute value of Gaussian

have non-decreasing hazard rate. The absolute value of a Gaussian random variable has both non-decreasing hazard rate and bounded inverse hazard rate.

Taking advantage of increasing hazard rate

Lemma

If F has non-decreasing hazard rate h, then for $1 \leq k \leq n/2$,

$$\mathrm{Var}\left[X_{(k)}\right] \leq \mathbb{E} V_k \leq \frac{2}{k}\mathbb{E}\left[\left(\frac{1}{h(X_{(k+1)})}\right)^2\right] ,$$

Lemma

Let $n \geq 3$, let $X_{(1)} \geq \ldots \geq X_{(n)}$ be the order statistics of absolute values of a standard Gaussian sample,

$$\text{For } 1 \leq k \leq n/2, \quad \mathrm{Var}[X_{(k)}] \leq \frac{1}{k \log 2} \frac{8}{\log \frac{2n}{k} - \log(1 + \frac{4}{k} \log\log \frac{2n}{k})} .$$

Goal

Context

If F has increasing hazard rate (more concentrated than exponential), extreme and intermediate order statistics have exponentiel moments.

Target

Derive

- Establishing Exponential Efron-Stein inequalities
- Bernstein-like deviation inequalities statistics.

for order statitics

Modified logarithmic Sobolev inequalities

Theorem

(Modified logarithmic Sobolev inequality. L. Wu, P. Massart, 2000)
Let $\tau(x) = e^x - x - 1$.
Then for any $\lambda \in \mathbb{R}$,

$$\begin{aligned}
\mathrm{Ent}\left[e^{\lambda Z}\right] &= \mathbb{E}\left[e^{\lambda Z}\log e^{\lambda Z}\right] - \mathbb{E}\left[e^{\lambda Z}\right]\log \mathbb{E}\left[e^{\lambda Z}\right] \\
&= \lambda \mathbb{E}\left[Z e^{\lambda Z}\right] - \mathbb{E}\left[e^{\lambda Z}\right]\log \mathbb{E}\left[e^{\lambda Z}\right] \\
&\leq \mathbb{E}\left[\sum_{i=1}^{n} e^{\lambda Z}\tau\left(-\lambda(Z - Z_i)\right)\right]
\end{aligned}$$

Remark

Logarithmic-Sobolev inequalities and Efron-Stein inequalities are derived in a similar way, proofs rely on variational representations of variance and entropy.

Application to order statistics

Notation

$\psi(x) = e^x \tau(-x) = 1 + (x-1)e^x$

Lemma

For all $\lambda \in \mathbb{R}$,

$$\begin{aligned} \text{Ent}\left[e^{\lambda X_{(k)}}\right] &\leq k\mathbb{E}\left[e^{\lambda X_{(k+1)}}\psi(\lambda(X_{(k)} - X_{(k+1)}))\right] \\ &= k\mathbb{E}\left[e^{\lambda X_{(k+1)}}\psi(\lambda\Delta_k)\right] \end{aligned}$$

Proof parallels the variance bounds derived from Efron-Stein inequalities.

Bernstein bounds, sub-Gamma distributions

Sub-gamma on the right tail with variance factor v and scale parameter c

$$\log \mathbb{E} e^{\lambda(X-\mathbb{E}X)} \leq \frac{\lambda^2 v}{2(1-c\lambda)} \text{ for every } \lambda \quad \text{such that} \quad 0 < \lambda < 1/c \,.$$

Bernstein's inequality

for $t > 0$, $\mathbb{P}\left\{X \geq \mathbb{E}X + \sqrt{2vt} + ct\right\} \leq \exp\left(-t\right).$

Exponential Efron-Stein inequality for order statistics

$V_k = k\Delta_k^2$: the Efron-Stein estimate of the variance of $X_{(k)}$.

Theorem

If F has non-decreasing hazard rate h, then for $\lambda \geq 0$, and $1 \leq k \leq n/2$,

$$\begin{aligned} \log \mathbb{E} e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})} &\leq \lambda \frac{k}{2} \mathbb{E}\left[\Delta_k \left(e^{\lambda \Delta_k} - 1\right)\right] \\ &= \lambda \frac{k}{2} \mathbb{E}\left[\sqrt{\frac{V_k}{k}} \left(e^{\lambda \sqrt{V_k/k}} - 1\right)\right] . \end{aligned}$$

Assessment

- Does not follow from exponential Efron-Stein inequality from B., Lugosi and Massart (Ann. Probab. 2003).

$$\log \mathbb{E} e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})} \leq \frac{\lambda\theta}{1-\lambda\theta} \log \mathbb{E} e^{\lambda V_k/\theta} \text{ for } \theta > 0, 0 \leq \lambda \leq 1/\theta$$

  as V_k may not have exponential moments!
- Sharp (up to constants) for exponential samples.
- Works both for central, intermediate and extreme order statistics.

Proof (i)

- $\psi(x) = x(e^x - 1)$ is non-decreasing over $\mathbb{R}_+$,
- $X_{(k+1)}$ and Δ_k are negatively associated:

$$
\begin{aligned}
\mathrm{Ent}\left[e^{\lambda X_{(k)}}\right] &\leq k\mathbb{E}\left[e^{\lambda X_{(k+1)}}\psi(\lambda\Delta_k)\right] \\
&\leq k\mathbb{E}\left[e^{\lambda X_{(k+1)}}\right] \times \mathbb{E}\left[\psi(\lambda\Delta_k)\right] \\
&\leq k\mathbb{E}\left[e^{\lambda X_{(k)}}\right] \times \mathbb{E}\left[\psi(\lambda\Delta_k)\right] .
\end{aligned}
$$

Multiplying both sides by $\exp(-\lambda\mathbb{E}X_{(k)})$, leads to

$$
\mathrm{Ent}\left[e^{\lambda(X_{(k)}-\mathbb{E}X_{(k)})}\right] \leq k\mathbb{E}\left[e^{\lambda(X_{(k)}-\mathbb{E}X_{(k)})}\right] \times \mathbb{E}\left[\psi(\lambda\Delta_k)\right] .
$$

Proof (ii) Herbst's argument

Let $G(\lambda) = \mathbb{E}e^{\lambda \Delta_k}$. Obviously, $G(0) = 1$, and as $\Delta_k \geq 0$, G and its derivatives are increasing on $[0, \infty)$,

$$\mathbb{E}\left[\psi(\lambda \Delta_k)\right] = 1 - G(\lambda) + \lambda G'(\lambda) = \int_0^\lambda s G''(s) \mathrm{d}s \leq G''(\lambda) \frac{\lambda^2}{2} .$$

Hence, for $\lambda \geq 0$,

$$\frac{\mathrm{Ent}\left[e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}\right]}{\lambda^2 \mathbb{E}\left[e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}\right]} = \frac{\mathrm{d}\frac{1}{\lambda} \log \mathbb{E}e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}}{\mathrm{d}\lambda} \leq \frac{k}{2} \frac{\mathrm{d}G'}{\mathrm{d}\lambda} .$$

Proof (iii) solving the differential inequality

Integrating both sides, using the fact that

$$\lim_{\lambda \to 0} \frac{1}{\lambda} \log \mathbb{E} e^{\lambda (X_{(k)} - \mathbb{E} X_{(k)})} = 0,$$

leads to

$$\begin{aligned} \frac{1}{\lambda} \log \mathbb{E} e^{\lambda (X_{(k)} - \mathbb{E} X_{(k)})} &\leq \frac{k}{2} (G'(\lambda) - G'(0)) \\ &= \frac{k}{2} \mathbb{E} \left[\Delta_k \left(e^{\lambda \Delta_k} - 1 \right) \right] . \end{aligned}$$

Maxima of Gaussians

Lemma

For n such that the solution v_n of equation

$$16/x + \log(1 + 2/x + 4\log(4/x)) = \log(2n)$$

is smaller than 1,
for all $0 \leq \lambda < \frac{1}{\sqrt{v_n}}$,

$$\log \mathbb{E} e^{\lambda(X_{(1)} - \mathbb{E}X_{(1)})} \leq \frac{v_n \lambda^2}{2(1 - \sqrt{v_n}\lambda)} \ .$$

For all $t > 0$,

$$\mathbb{P}\left\{X_{(1)} - \mathbb{E}X_{(1)} > \sqrt{v_n}(t + \sqrt{2t})\right\} \leq e^{-t} \ .$$

Median of Gaussians

...

The same approach works for extreme, intermediate and central order statistics

Lemma

Let $v_n = 8/(n \log 2)$.
For all $0 \leq \lambda < n/(2\sqrt{v_n})$,

$$\log \mathbb{E} e^{\lambda(X_{(n/2)} - \mathbb{E}X_{(n/2)})} \leq \frac{v_n \lambda^2}{2(1 - 2\lambda\sqrt{v_n/n})} \ .$$

For all $t > 0$,

$$\mathbb{P}\left\{X_{(n/2)} - \mathbb{E}X_{(n/2)} > \sqrt{2v_n t} + 2\sqrt{v_n/n}t\right\} \leq e^{-t} \ .$$

Assessment

Rényi's representation : order statistics are functions of sums of independent random variables (spacings of exponential samples).

If the function is concave, concavity may be used twice.

What about plugging tail bounds for order statistics of exponential samples ?

Ad hoc arguments

What can be obtained from Rényi's representation and exponential inequalities for sums of Gamma-distributed random variables ?

Lemma

Let $X_{(1)}$ be the maximum of the absolute values of n independent standard Gaussian random variables, and let $\widetilde{U}(s) = \Phi^{\leftarrow}(1 - 1/(2s))$ for $s \geq 1$. For $t > 0$,

$$\mathbb{P}\left\{X_{(1)} - \mathbb{E}X_{(1)} \geq t/(3\widetilde{U}(n)) + \sqrt{t}/\widetilde{U}(n) + \delta_n\right\} \leq \exp\left(-t\right) ,$$

where $\delta_n > 0$ and $\lim_n(\widetilde{U}(n))^3\delta_n = \frac{\pi^2}{12}$.

References

- S. B. and M. Thomas. Concentration inequalities for order statistics. http://arxiv.org/abs/1207.7209
- P. Massart. *Concentration inequalities and model selection* Springer. 2006. Lecture Notes in Mathematics 1896.
- L. de Haan and A. Ferreira. *Extreme value theory*. Springer. 2006