Concentration inequalities for order statistics

Using the entropy method and Rényi's representation

Maud Thomas[1]

in collaboration with Stéphane Boucheron[1]

[1]LPMA Université Paris-Diderot

PEPER Workshop

Aussois, December 15 - 19, 2013

Background: order statistics

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

Order statistics

$X_{(1)} \geqslant \ldots \geqslant X_{(n)}$ non-increasing rearrangement of $X_1, \ldots, X_n$.

- $X_{(1)}$: sample maximum.
- $X_{(n/2)}$: sample median.
- Classical statistic theory and Extreme Value Theory provide:
  - **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.

- Bound for the variance (*e.g* Efron-Stein inequality).
- Markov's inequality: let Y be non-negative random variable. Then,

$$\mathbb{P}\{Y \geqslant t\} \leqslant \frac{\mathbb{E}Y}{t} \ .$$

- Chebytchev's inequality:

$$\mathbb{P}\{|Z - \mathbb{E}Z| \geqslant t\} \leqslant \frac{\mathrm{Var}[Z]}{t^2} \ .$$

- More generally, let ϕ be a non-decreasing and positive function:

$$\mathbb{P}\{Z \geqslant t\} \leqslant \mathbb{P}\{\phi(Z) \geqslant \phi(t)\} \leqslant \frac{\mathbb{E}\phi(Z)}{\phi(t)} \ .$$

Concentration inequalities for Gaussian random variables

- X a standard Gaussian vector.
- Poincaré's inequality:

$$\mathrm{Var}[f(X)] \leqslant \mathbb{E}\|\nabla f\|^2 \ .$$

- Gross logarithmic Sobolev inequality:

$$\mathrm{Ent}[f(X)^2] \leqslant 2\mathbb{E}\|\nabla f\|^2 \ .$$

- Cirelson's inequality:

$$\mathbb{P}\{f(X) \geqslant \mathbb{E}f(X) + t\} \leqslant \exp(-t^2/(2L^2)) \quad \text{if } \|\nabla f\| \leqslant L \ .$$

Gaussian case: concentration inequalities for order statistics

- X_i are standard Gaussian.
- $f(X_1, \ldots, X_n) = X_{(k)}$:
  the rank k order statistic of a sample is a simple function of n independent random variables.
- Almost surely, $\|\nabla f\| = 1$.
- Poincaré's inequality $\Rightarrow$ $\mathrm{Var}[f(X_1, \ldots, X_n)] \leqslant 1$.
- Extreme Value Theory asserts: $\mathrm{Var}[X_{(1)}] = O(1/\log n)$.
- Classical statistic theory implies: $\mathrm{Var}[X_{(n/2)}] = O(1/n)$.

We do not understand (clearly)

in which way order statistics are a smooth function of the sample.

Variance bounds, order statistics and spacings

Proposition (Boucheron, T. (2012))

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

Without any assumption such as:

- F belongs to the max-domain of attraction of an extreme value distribution G.
- $(X_{(k)})$ 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 $k/n \to 0$, $k \to \infty$.

Proof

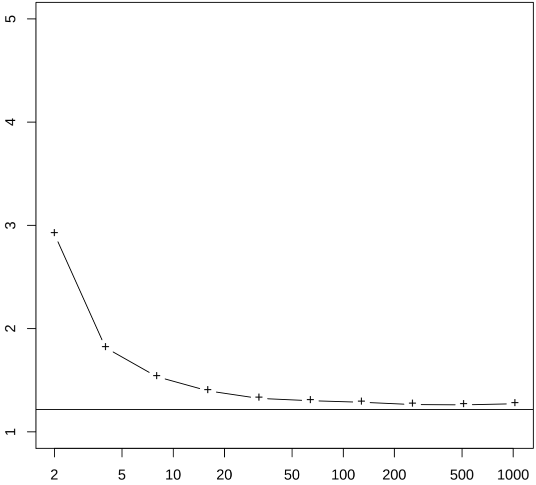
Efron-Stein inequality (Efron, Stein (1981))

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:

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

$\sum_{i=1}^{n} (Z - Z_i)^2$ is Efron-Stein estimate of the variance.
Here, the Efron-Stein estimate of the variance: $V_k = k{\Delta_k}^2$.

Graphical assessment



- Ratio between the Efron-Stein estimate and the variance of the maximum of n independent Gaussian random variables.
- $n = 2^p$ for $p = 1, \ldots, 10$.
- 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.

Rényi's representation (Rényi (1953))

Let $Y_{(1)} \geqslant Y_{(2)} \geqslant \ldots \geqslant Y_{(n)}$ be the order statistics of an independent sample of the standard exponential distribution, then

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

where $E_1, \ldots, E_n$ are i.i.d standard exponential random variables.

Quantile function

Definition (Quantile function)

$$F^{\leftarrow}(p) = \inf \{x \colon F(x) \geqslant p\}, p \in (0,1) \quad .$$

Notation

$$U(t) = F^{\leftarrow}(1 - 1/t), t \in (1, \infty) \, .$$

Representation for order statistics

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

$$(U \circ \exp)(Y_{(1)}) \geqslant \ldots \geqslant (U \circ \exp)(Y_{(n)})$$

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

Hazard rate, spacings and order statistics

Definition (Hazard rate)

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

Lemma

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

Link with the von Mises condition ($\gamma > 0$)

$$\lim_{x \to \infty} x h(x) = \frac{1}{\gamma} \ .$$

Variance bound for order statistics when the hazard rate is non-decreasing

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

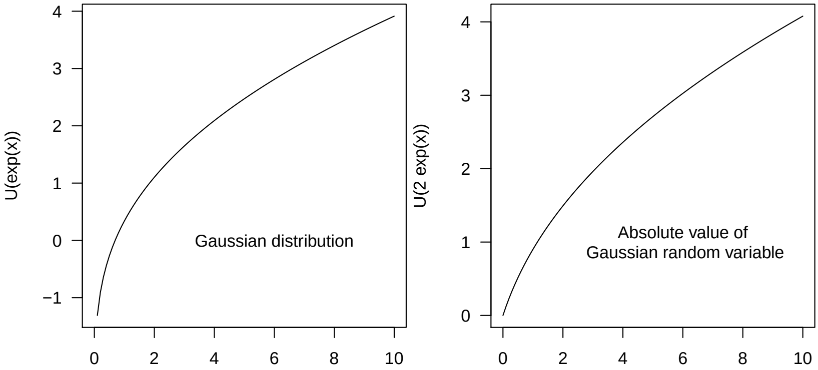
Theorem (Boucheron, T. (2012))

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

$$\operatorname{Var}\left[X_{(k)}\right] \leqslant \mathbb{E} V_k \leqslant \frac{2}{k}\mathbb{E}\left[\left(\frac{1}{h(X_{(k+1)})}\right)^2\right].$$

Gaussian hazard rate

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



Variance of absolute values of Gaussian random variables

Proposition (Boucheron, T. (2012))

Let $n \geqslant 3$, let $X_{(k)}$ be the rank k order statistic of absolute values of n standard independent Gaussian random variables,

$$\operatorname{Var}[X_{(k)}] \leqslant \frac{1}{k \log 2} \frac{8}{\log\left(\frac{2n}{k}\right) - \log\left(1 + \frac{4}{k} \log\log\left(\frac{2n}{k}\right)\right)} \quad .$$

- For the maximum ($k = 1$), the bound becomes:

$$\frac{1}{\log 2} \frac{8}{\log 2n - \log(1 + 4 \log\log 2n)}$$

- For the median ($k = n/2$),

$$\frac{2}{n \log 2} \frac{8}{\log 4 - \log(1 + \frac{8}{n} \log\log 4)}$$

Entropy bound, order statistics and spacings

Proposition (Boucheron, T. (2012))

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

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

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

Remark

Same comments as for the variance bound.

Sketch of proof

Modified logarithmic Sobolev inequality (Wu(2000); Massart(2000))

Let $\tau(x) = e^x - x - 1$.
Then for any $\lambda \in \mathbb{R}$,

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

Goal

- Context: if F has non-decreasing hazard rate (more concentrated than exponential), then extreme and intermediate order statistics have exponentiel moments.

Definition (Exponential Efron-Stein inequality)

Let $Z = f(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are independent random variables and V its Efron-Stein estimate of the variance of Z.
Z satisfies an exponential Efron-Stein inequality if for all θ, $\lambda > 0$ such that $\lambda\theta < 1$ and $\mathbb{E}\left[e^{\lambda V/\theta}\right] < \infty$:

$$\log \mathbb{E}\left[e^{\lambda(Z-\mathbb{E}Z)}\right] \leqslant \frac{\lambda\theta}{1-\lambda\theta} \log \mathbb{E}\left[e^{\lambda V/\theta}\right] \quad .$$

Problem

For an exponential sample, $\mathbb{E}\left[e^{\lambda V/\theta}\right] = \infty$.
↳ Find another decoupling inequality.
↳ Negative Association.

Decoupling inequality: negative association

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 non-decreasing functions f, g

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

Remember:

$$\mathrm{Ent}\left[e^{\lambda Z}\right] \leqslant k\mathbb{E}\left[e^{\lambda X_{(k+1)}}\psi(\lambda\Delta_k)\right] .$$

Exponential Efron-Stein inequality for order statistics

Theorem (Boucheron, T. (2012))

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

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

Bernstein bounds, sub-Gamma distributions

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

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

Bernstein's inequality

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

Bernstein inequality for the maximum of absolute values of Gaussian random variables

Theorem (Boucheron, T. (2012))

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 $0 \leqslant \lambda < \frac{1}{\sqrt{v_n}}$,

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

Remarks and perspectives

- Condition "non-decreasing hasard rate" should not be necessary for central order statistics.
- To obtain pleasant bound need another condition on the hazard rate.

- Concentration inequalities for tail index estimators (*e.g* Hill's estimator, moment estimators).
- Adaptive estimation
- Model selection.

Thank you for your attention !

- S. Boucheron. and M. T. Concentration inequalities for order statistics. Electronic Communications in Probability (2012). http://ecp.ejpecp.org/article/view/2210
- S. Boucheron, G. Lugosi and P. Massart **Concentration inequalities** Oxford University Press (2012).
- P. Massart. **Concentration inequalities and model selection** Springer (2006). Lecture Notes in Mathematics 1896.