Electron. Commun. Probab. **17** (2012), no. 51, 1–12.
DOI: 10.1214/ECP.v17-2210
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Concentration inequalities for order statistics

Stéphane Boucheron* Maud Thomas†

Abstract

This note describes non-asymptotic variance and tail bounds for order statistics of samples of independent identically distributed random variables. When the sampling distribution belongs to a maximum domain of attraction, these bounds are checked to be asymptotically tight. When the sampling distribution has a non-decreasing hazard rate, we derive an exponential Efron-Stein inequality for order statistics, that is an inequality connecting the logarithmic moment generating function of order statistics with exponential moments of Efron-Stein (jackknife) estimates of variance. This connection is used to derive variance and tail bounds for order statistics of Gaussian samples that are not within the scope of the Gaussian concentration inequality. Proofs are elementary and combine Rényi's representation of order statistics with the entropy approach to concentration of measure popularized by M. Ledoux.

Keywords: order statistics;concentration inequalities;entropy method;Efron-Stein inequalities; Rényi's representation.
AMS MSC 2010: 60E15;60G70;62G30;62G32.
Submitted to ECP on August 1, 2012, final version accepted on October 30, 2012.
Supersedes arXiv:1207.7209.

1 Introduction

The purpose of this note is to develop non-asymptotic variance and tail bounds for order statistics. In the sequel, $X_1, \ldots, X_n$ are independent random variables, distributed according to a certain probability distribution F, and $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$ denote the corresponding order statistics (the non-increasing rearrangement of $X_1, \ldots, X_n$). In Extreme Value Theory (EVT), the Fisher-Tippett-Gnedenko Theorem characterizes the asymptotic behavior of the maximum $X_{(1)}$ [3] while asymptotics of the median $X_{(\lfloor n/2 \rfloor)}$ and other central order statistics are well documented [12]. Although the distribution function of order statistics is explicitly known, simple variance or tail bounds for order statistics do not seem to be well documented when sample size is kept fixed.

The search for variance and tail bounds for order statistics is driven by the desire to understand some aspects of the concentration of measure phenomenon [5, 8]. Concentration of measure theory tells us that a function of many independent random variables that does not depend too much on any of them is almost constant. The best known results in that field are the Poincaré and Gross logarithmic Sobolev inequalities and the Tsirelson-Ibragimov-Sudakov tail bounds for functions of Gaussian vectors. If $X_1, \ldots, X_n$ are independent standard Gaussian random variables, and $f\colon \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz, then $Z = f(X_1, \ldots, X_n)$ satisfies $\mathrm{Var}(Z) \leq L^2$, $\log \mathbb{E}[\exp(\lambda(Z - \mathbb{E}Z))] \leq \lambda^2 L^2/2$

*LPMA UMR 7599 CNRS Université Paris-Diderot, France. E-mail: boucheron@univ-paris-diderot.fr
†LPMA UMR 7599 CNRS Université Paris-Diderot, France. E-mail: maud.thomas@univ-paris-diderot.fr

and $\mathbb{P}\{Z - \mathbb{E}Z \geq t\} \leq \exp(-t^2/(2L^2))$. If we apply these bounds to $Z = X_{(k)}$, the Lipschitz constant is (almost surely) $L = 1$, so Poincaré inequality allows us to establish $\mathrm{Var}(X_{(k)}) \leq 1$. This upper bound is far from being satisfactory: it is well-known that $\mathrm{Var}(X_{(1)}) = O\big(1/\log n\big)$ and $\mathrm{Var}(X_{(\lfloor n/2 \rfloor)}) = O\big(1/n\big)$ [3, 12]. Naive use of off-the-shelf concentration bounds does not work when handling order statistics. This situation is not uncommon: the analysis of the largest eigenvalue of random matrices from the Gaussian Unitary Ensemble (GUE) [6] provides a setting where the derivation of sharp concentration inequalities requires an ingenious combination of concentration inequalities and special representations.

Even so, our purpose is to show that the tools and methods of concentration of measure theory are relevant to the analysis of order statistics. To address this, the Efron-Stein inequalities are our main tools. They assert that, on average, the jackknife estimate(s) of the variance of functions of independent random variables are upper bounds. Extensions allow us to derive exponential bounds (see Theorem 2.1). We refer to [9, 10] and references therein for an account of the early interplay between jackknife estimates, order statistics, EVT and statistical inference. When properly combined with Rényi's representation for order statistics (see Theorem 2.5), the so-called entropy method [5] allows us to recover sharp variance and tail bounds. Proofs are elementary and parallel the approach followed by Ledoux [6] in a much more sophisticated setting. Ledoux builds on the determinantal structure of the joint density of the eigenvalues of random matrices from the GUE to upper bound tail bounds by sums of Gaussian integrals that can be handled by concentration arguments. In the sequel, we build on Rényi's representation of order statistics: $X_{(1)}, \ldots, X_{(n)}$ can be represented as the monotone image of the order statistics of a sample of the standard exponential distribution which turn out to be distributed as partial sums of independent random variables.

In Section 2, we derive simple relations between the variance or the entropy of order statistics $X_{(k)}$ and moments of spacings $\Delta_k = X_{(k)} - X_{(k+1)}$. When the sampling distribution has a non-decreasing hazard rate (a condition that is satisfied by Gaussian, exponential, Gumbel, logistic distributions, ...) we are able to build on these connections, obtaining Theorem 2.9 that may be considered as an exponential Efron-Stein inequality for order statistics. In Section 3, using the framework of EVT, these connections are checked to be asymptotically tight.

In Section 4, using explicit bounds on the Gaussian hazard rate, we derive Bernstein-like inequalities for the maximum and the median of a sample of independent standard Gaussian random variables with a correct variance and scale factors (Proposition 4.6). We provide non-asymptotic variance bounds for order statistics of Gaussian samples with the right order of magnitude in Propositions 4.2, and 4.4.

2 Order statistics and spacings

Efron-Stein inequalities [4] allow us to derive upper bounds on the variance of functions of independent random variables. The next version can be found in [1, p. 221]

Theorem 2.1. *(Efron-Stein inequalities.) 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*

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

The quantity $\sum_{i=1}^{n}(Z - Z_i)^2$ is called a jackknife estimate of variance. Efron-Stein inequalities form a special case of a more general collection of inequalities that encompasses the so-called modified logarithmic Sobolev inequalities [5, 7] or [8, p. 157]. Henceforth, the entropy of a non-negative random variable X is defined by $\mathrm{Ent}[X] = \mathbb{E}[X\log X] - \mathbb{E}X\log\mathbb{E}X$.

Theorem 2.2. *(Modified logarithmic Sobolev inequality.) Let $\tau(x) = e^x - x - 1$. Then for any $\lambda \in \mathbb{R}$,*

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

Theorems 2.1 and 2.2 provide a transparent connection between moments of order statistics and moments of spacings.

Henceforth, let $\psi\colon \mathbb{R}\to\mathbb{R}_+$ be defined by $\psi(x) = e^x\tau(-x) = 1 + (x-1)e^x$.

Proposition 2.3. *(Order statistics and spacings.) For all $1 \leq k \leq n/2$,*

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

and for all $\lambda \in \mathbb{R}$,

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

For all $n/2 < k \leq n$,

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

and for all $\lambda \in \mathbb{R}$,

$$\mathrm{Ent}\left[e^{\lambda X_{(k)}}\right] \leq (n-k+1)\mathbb{E}\left[e^{\lambda X_{(k)}}\tau(\lambda(X_{(k-1)} - X_{(k)}))\right] \ .$$

Proof of Proposition 2.3. Let $Z = X_{(k)}$ and for $k \leq n/2$ define Z_i as the rank k statistic from subsample $X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n$, that is, $Z_i = X_{(k+1)}$ if $X_i \geq X_{(k)}$ and $Z_i = Z$ otherwise. Apply Theorem 2.1.

For $k > n/2$, define Z_i as the rank $k-1$ statistic from $X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n$, that is $Z_i = X_{(k-1)}$ if $X_i \leq X_{(k)}$ and $Z_i = Z$ otherwise. Apply Theorem 2.1 again.

For $k \leq n/2$, define Z and Z_i as before, apply Theorem 2.2:

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

The proof of the last statement proceeds by the same argument. □

In the sequel, we focus on the setting $1 \leq k \leq n/2$: the setting $k > n/2$ can be treated in a similar way.

Proposition 2.3 can be fruitfully complemented by Rényi's representation of order statistics [3, and references therein].

In the sequel, if f is a monotone function from (a,b) (where a and b may be infinite) to (c,d), its generalized inverse $f^{\leftarrow} : (c,d) \to (a,b)$ is defined by $f^{\leftarrow}(y) = \inf\{x : a < x < b, f(x) \geq y\}$ [3, for properties of this transformation].

Definition 2.4. *The U- transform of a distribution function F is defined as a non-decreasing function on $(1,\infty)$ by $U = (1/(1-F))^{\leftarrow}$, $U(t) = \inf\{x \ : \ F(x) \geq 1 - 1/t\} = F^{\leftarrow}(1-1/t)$.*

Rényi's representation asserts that the order statistics of a sample of independent standard exponentially distributed random variables are distributed as partials sums of independent rescaled exponentially distributed random variables.

Theorem 2.5. *(Rényi's representation) Let $X_{(1)} \geq \ldots \geq X_{(n)}$ be the order statistics of a sample from distribution F, let $U = (1/(1-F))^{\leftarrow}$, let $Y_{(1)} \geq Y_{(2)} \geq \ldots \geq Y_{(n)}$ 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 independent and identically distributed standard exponential random variables, and $(X_{(n)}, \ldots, X_{(1)}) \sim \left(U \circ \exp(Y_{(n)}), \ldots, U \circ \exp(Y_{(1)})\right)$.

We may readily test the tightness of Proposition 2.3. By Theorem 2.5, $Y_{(k)} = \frac{E_n}{n} + \ldots + \frac{E_k}{k}$ and $\operatorname{Var}[Y_{(k)}] = \sum_{i=k}^{n} \frac{1}{i^2}$. Hence, for any sequence $(k_n)_n$ with $\lim_n k_n = \infty$, and $\limsup k_n/n < 1$, $\lim_{n\to\infty} k_n \operatorname{Var}[Y_{(k_n)}] = 1$, while by Proposition 2.3, $\operatorname{Var}[Y_{(k)}] \leq k\mathbb{E}\left[\left(E_k/k\right)^2\right] = \frac{2}{k}$.

The next condition makes combining Proposition 2.3 and Theorem 2.5 easy.

Definition 2.6. *(Hazard rate.) The hazard rate of an absolutely continuous probability distribution with distribution function F is: $h = f/\overline{F}$ where f and $\overline{F} = 1 - F$ are respectively the density and the survival function associated with F.*

From elementary calculus, letting $U = (1/(1-F))^{\leftarrow}$, we get $(U \circ \exp)' = 1/h(U \circ \exp)$.

Proposition 2.7. *Let F be an absolutely continuous distribution function with hazard rate h, let $U = (1/(1-F))^{\leftarrow}$. Then, h is non-decreasing if and only if $U \circ \exp$ is concave.*

Observe that if the hazard rate h is non-decreasing, then for all $t > 0$ and $x > 0$, $U(\exp(t+x)) - U(\exp(t)) \leq x/h(U(\exp(t)))$. Moreover, assuming that the hazard rate is non-decreasing warrants negative association between spacings and related order statistics.

Proposition 2.8 (Negative association)**.** *If F has non-decreasing hazard rate, then the k^{th} spacing $\Delta_k = X_{(k)} - X_{(k+1)}$ and $X_{(k+1)}$ are negatively associated: for any pair of non-decreasing functions g_1 and g_2,*

$$\mathbb{E}[g_1(X_{(k+1)})g_2(\Delta_k)] \leq \mathbb{E}[g_1(X_{(k+1)})]\mathbb{E}[g_2(\Delta_k)] \ .$$

Proof of Proposition 2.8. Let $Y_{(n)}, \ldots, Y_{(1)}$ be the order statistics of a standard exponential sample. Let $E_k = k(Y_{(k)} - Y_{(k+1)})$ be the rescaled k^{th} spacing of the exponential sample. By Theorem 2.5, E_k is standard exponentially distributed and independent of $Y_{(k+1)}$. Let g_1 and g_2 be two non-decreasing functions.

$$\begin{aligned}
\mathbb{E}[g_1(X_{(k+1)})g_2(\Delta_k)] &= \mathbb{E}[g_1(U(e^{Y_{(k+1)}}))g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}}))] \\
&= \mathbb{E}\left[\mathbb{E}\left[g_1(U(e^{Y_{(k+1)}}))g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}\right]\right] \\
&= \mathbb{E}\left[g_1(U(e^{Y_{(k+1)}}))\mathbb{E}\left[g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}\right]\right] \ .
\end{aligned}$$

The function $g_1 \circ U \circ \exp$ is non-decreasing. Almost surely, as the conditional distribution of kE_k with respect to $Y_{(k+1)}$ is the exponential distribution with scale parameter $\frac{1}{k}$,

$$\mathbb{E}\left[g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}\right] = \int_0^{\infty} e^{-x} g_2(U(e^{\frac{x}{k}+Y_{(k+1)}}) - U(e^{Y_{(k+1)}}))\mathrm{d}x \ .$$

As F has a non-decreasing hazard rate, $U(\exp(x/k+y))-U(\exp(y)) = \int_0^{x/k}(U\circ\exp)'(y+z)\mathrm{d}z$ is non-increasing with respect to y.
Hence, $\mathbb{E}\left[g_2(U(e^{E_k/k+Y_{(k+1)}})-U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}\right]$ is a non-increasing function of $Y_{(k+1)}$.
Hence, by Chebyshev's association inequality,

$$\begin{aligned}
&\mathbb{E}[g_1(X_{(k+1)})g_2(\Delta_k)] \\
&\quad\le \mathbb{E}\left[g_1(U(e^{Y_{(k+1)}}))\right]\mathbb{E}\left[\mathbb{E}\left[g_2(U(e^{E_k/k+Y_{(k+1)}})-U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}\right]\right] \\
&\quad= \mathbb{E}\left[g_1(X_{(k+1)})\right]\mathbb{E}\left[g_2(\Delta_k)\right] \ .
\end{aligned}$$

□

Negative association between order statistics and spacings allows us to establish our main result.

Theorem 2.9 (Exponential Efron-Stein inequality). *Let $X_1,\ldots,X_n$ be independently distributed according to F, let $X_{(1)}\ge\ldots\ge X_{(n)}$ be the order statistics and let $\Delta_k = X_{(k)}-X_{(k+1)}$ be the k^{th} spacing. Let $V_k = k\Delta_k^2$ denote the Efron-Stein estimate of the variance of $X_{(k)}$ (for $k=1,\ldots,n/2$).*

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

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

For $\lambda\ge 0$, and $1\le k\le n/2$,

$$\log\mathbb{E}e^{\lambda(X_{(k)}-\mathbb{E}X_{(k)})} \le \lambda\frac{k}{2}\mathbb{E}\left[\Delta_k\left(e^{\lambda\Delta_k}-1\right)\right] = \lambda\frac{k}{2}\mathbb{E}\left[\sqrt{\frac{V_k}{k}}\left(e^{\lambda\sqrt{V_k/k}}-1\right)\right] . \tag{2.1}$$

Inequality (2.1) may be considered as an exponential Efron-Stein inequality for order-statistics: it connects the logarithmic moment generating function of the k^{th} order statistic with the exponential moments of the square root of the Efron-Stein estimate of variance $k\Delta_k^2$. This connection provides correct bounds for the standard exponential distribution whereas the exponential Efron-Stein inequality described in [2] does not. This comes from the fact that negative association between spacing and order statistics dispenses us from the general decoupling argument used in [2]. It is then possible to carry out the Herbst's argument in an effortless way.

Proof of Theorem 2.9. Let $Y_{(k)},Y_{(k+1)}$ denote the k^{th} and $k+1^{\text{th}}$ order statistics of a standard exponential sample of size n. By Proposition 2.3, using Rényi's representation (Theorem 2.5), and Proposition 2.7, for $k\le n/2$,

$$\operatorname{Var}[X_{(k)}] \le k\mathbb{E}\left[\left(U\left(e^{Y_{(k+1)}}e^{Y_{(k)}-Y_{(k+1)}}\right)-U\left(e^{Y_{(k+1)}}\right)\right)^2\right] \le \frac{2}{k}\mathbb{E}\left[\left(\frac{1}{h(X_{(k+1)})}\right)^2\right] ,$$

as by Theorem 2.5, $Y_{(k)}-Y_{(k+1)}$ is independent of $Y_{(k+1)}$ and exponentially distributed with scale parameter $1/k$.

By Propositions 2.3 and 2.7, as ψ, defined by $\psi(x)=1+(x-1)e^x$ is non-decreasing

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

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

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

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 $[0, \infty)$,

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

Hence, for $\lambda \geq 0$,

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

Integrating both sides, using the fact that $\lim_{\lambda\to 0}\frac{1}{\lambda}\log\mathbb{E}e^{\lambda(X_{(k)}-\mathbb{E}X_{(k)})} = 0$,

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

□

3 Asymptotic assessment

Assessing the quality of the variance bounds from Proposition 2.3 in full generality is not easy. However, Extreme Value Theory (EVT) provides us with a framework where the Efron-Stein estimates of variance for maxima are asymptotically of the right order of magnitude.

Definition 3.1 (Maximum domain of attraction)**.** *The distribution function F belongs to a maximum domain of attraction with tail index $\gamma \in \mathbb{R}$ ($F \in \mathsf{MDA}(\gamma)$), if and only if there exists a non-negative auxiliary function a on $[1, \infty)$ such that for $x \in [0, \infty)$ (if $\gamma > 0$), $x \in [0, -1/\gamma)$ (if $\gamma < 0$), $x \in \mathbb{R}$ (if $\gamma = 0$)*

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

If $\gamma = 0$, $(1 + \gamma x)^{-1/\gamma}$ should read as $\exp(-x)$.

If $F \in \mathsf{MDA}(\gamma)$ and has a finite variance ($\gamma < 1/2$), the variance of $(\max(X_1, \ldots, X_n) - F^{\leftarrow}(1 - 1/n))/a(n)$ converges to the variance of the limiting extreme value distribution [3].

Membership in a maximum domain of attraction is characterized by the *extended regular variation* property of $U = (1/(1 - F))^{\leftarrow}$: $F \in \mathsf{MDA}(\gamma)$ with auxiliary function a iff for all $x > 0$

$$\lim_{t\to\infty}\frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} ,$$

where the right-hand-side should read as $\log x$ when $\gamma = 0$ [3].

Using Theorem 2.1.1 and Theorem 5.3.1 from [3], and performing simple calculus, we readily obtain that the asymptotic ratio between the Efron-Stein upper bound and the variance of $X_{(1)}$ converges toward a limit that depends only on γ (for $\gamma = 0$ this limit is $12/\pi^2 \approx 1.21$).

Proposition 3.2. *Assume $X_{(1)} \geq \ldots \geq X_{(n)}$ are the order statistics of an independent sample distributed according to F, where $F \in \mathsf{MDA}(\gamma), \gamma < 1/2$, with auxiliary function a. Then*

$$\lim_n \frac{\mathbb{E}\left[\left((X_{(1)} - X_{(2)})\right)^2\right]}{a(n)^2} = \frac{2\Gamma(2(1-\gamma))}{(1-\gamma)(1-2\gamma)} \quad \text{while} \quad \lim_n \frac{\mathrm{Var}\left(X_{(1)}\right)}{a(n)^2} = \frac{1}{\gamma^2}\left(\Gamma(1 - 2\gamma) - \Gamma(1 - \gamma)^2\right) .$$

For $\gamma = 0$, the last expression should read as $\pi^2/6$.

When the tail index γ is negative, the asymptotic ratio degrades as $\gamma \to -\infty$, it scales as -4γ.

4 Order statistics of Gaussian samples

We now turn to the Gaussian setting and establish Bernstein inequalities for order statistics of absolute values of independent standard Gaussian random variables.

A real-valued random variable X is said to be *sub-gamma on the right tail with variance factor v and scale parameter c* if

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

Such a random variable satisfies a so-called Bernstein inequality: for $t>0$, $\mathbb{P}\left\{X \geq \mathbb{E}X + \sqrt{2vt} + ct\right\} \leq \exp(-t)$. A real-valued random variable X is said to be *sub-gamma on the left tail with variance factor v and scale parameter c*, if $-X$ is sub-gamma on the right tail with variance factor v and scale parameter c. A Gamma random variable with shape parameter p and scale parameter c (expectation pc and variance pc^2) is sub-gamma on the right tail with variance factor pc^2 and scale factor c while it is sub-gamma on the left-tail with variance factor pc^2 and scale factor 0. The Gumbel distribution (with distribution function $\exp(-\exp(-x))$ is sub-gamma on the right-tail with variance factor $\pi^2/6$ and scale factor 1, and sub-gamma on the left-tail with scale factor 0 (note that this statement is not sharp, see Lemma 4.3 below).

Order statistics of Gaussian samples provide an interesting playground for assessing Theorem 2.9. Let Φ and ϕ denote respectively the standard Gaussian distribution function and density. Throughout this section, let $\widetilde{U}\colon]1,\infty) \to [0,\infty)$ be defined by $\widetilde{U}(t) = \Phi^{\leftarrow}(1-1/(2t))$, $\widetilde{U}(t)$ is the $1-1/t$ quantile of the distribution of the absolute value of a standard Gaussian random variable, or the $1-1/(2t)$ quantile of the standard Gaussian distribution.

Proposition 4.1. *Absolute values of Gaussian random variables have a non-decreasing hazard rate :*
i) $\widetilde{U}\circ\exp$ is concave;
ii) For $y>0$, $\phi(\widetilde{U}(\exp(y)))/\overline{\Phi}(\widetilde{U}(\exp(y))) \geq \sqrt{\kappa_1(y+\log 2)}$ where $\kappa_1 \geq 1/2$.
iii) For $t\geq 3$,

$$\sqrt{2\log(2t) - \log\log(2t) - \log(4\pi)} \leq \widetilde{U}(t) \leq \sqrt{2\log(2t)-\log\log(2t)-\log\pi}\ .$$

Proof of Proposition 4.1. i) As $(\widetilde{U}\circ\exp)'(t) = \overline{\Phi}(\widetilde{U}(e^t))/\phi(\widetilde{U}(e^t))$ it suffices to check that the standard Gaussian distribution has a non-decreasing hazard rate on $[0,\infty)$. Let $h = \phi/\overline{\Phi}$, by elementary calculus, for $x>0$, $h'(x) = \big(\phi(x) - x\overline{\Phi}(x)\big)\,\phi(x)/\overline{\Phi}^2(x) \geq 0$ where the last inequality is a well known fact.

ii) For $\kappa_1 = 1/2$, for $p\in(0,1/2]$, the fact that $p\sqrt{\kappa_1\log 1/p} \leq \phi\circ\Phi^{\leftarrow}(p)$ follows from $\phi(x) - x\overline{\Phi}(x) \geq 0$ for $x>0$. Hence,

$$\frac{\overline{\Phi}(\Phi^{\leftarrow}(1-e^{-y}/2))}{\phi(\Phi^{\leftarrow}(1-e^{-y}/2))} = \frac{e^{-y}/2}{\phi(\Phi^{\leftarrow}(e^{-y}/2))} \leq \frac{1}{\sqrt{\kappa_1(\log 2 + y)}}\ .$$

iii) The first inequality can be deduced from $\phi\circ\Phi^{\leftarrow}(p) \leq p\sqrt{2\log 1/p}$, for $p\in(0,1/2)$ [11], the second from $p\sqrt{\kappa_1\log 1/p} \leq \phi\circ\Phi^{\leftarrow}(p)$. □

The next proposition shows that when used in a proper way, Efron-Stein inequalities may provide seamless bounds on extreme, intermediate and central order statistics of Gaussian samples.

Proposition 4.2. *Let $n\geq 3$, let $X_{(1)} \geq \ldots \geq X_{(n)}$ be the order statistics of absolute values of a standard Gaussian sample,*

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

By [3, Theorem 5.3.1], $\lim_n 2\log n \operatorname{Var}[X_{(1)}] = \pi^2/6$, while the above described upper bound on $\operatorname{Var}[X_{(1)}]$ is equivalent to $(8/\log 2)/\log n$. If $\lim_n k_n = \infty$ and $\lim_n k_n/n = 0$, by Smirnov's lemma [3], $\lim_n k(\widetilde{U}(n/k))^2 \operatorname{Var}[X_{(k)}] = 1$. For the asymptotically normal median of absolute values, $\lim_n (4\phi(\widetilde{U}(2))^2 n) \operatorname{Var}[X_{(n/2)}] = 1$ [12]. Again, the bound in Proposition 4.2 has the correct order of magnitude.

Lemma 4.3. *Let $Y_{(k)}$ be the k^{th} order statistics of a sample of n independent standard exponential random variables, let $\log 2 < z < \log(n/k)$, then*

$$\mathbb{P}\left\{Y_{(k+1)} \le \log(n/k) - z\right\} \le \exp\left(-\tfrac{k(e^z-1)}{4}\right) .$$

Proof of Lemma 4.3.

$$\mathbb{P}\left\{Y_{(k+1)} \le \log(n/k) - z\right\} = \sum_{j=0}^{k} \binom{n}{j} \left(1 - \frac{ke^z}{n}\right)^{n-j} \left(\frac{ke^z}{n}\right)^{j} \le \exp\left(-\frac{k(e^z-1)^2}{2e^z}\right)$$

since the right-hand-side of the first line is the probability that a binomial random variable with parameters n and $\frac{ke^z}{n}$ is less than k, which is sub-gamma on the left-tail with variance factor less than ke^z and scale factor 0. $\square$

Proof of Proposition 4.2. By Propositions 2.9 and 4.1, letting $\kappa_1 = 1/2$

$$\begin{aligned}
\operatorname{Var}\left(X_{(k)}\right) &\le \frac{2}{k}\mathbb{E}\left[\frac{2}{\log 2 + Y_{(k+1)}}\right] \\
&\le \frac{1}{\log 2}\frac{4}{k}\mathbb{P}\Big\{Y_{(k+1)} \le \log(n/k) - z\Big\} + \frac{4}{k}\frac{1}{\log\frac{n}{k} - z + \log 2} \\
&\le \frac{4}{k\log 2}\frac{1}{\log\frac{2n}{k}} + \frac{4}{k}\frac{1}{\log\frac{2n}{k} - \log(1+\frac{4}{k}\log\log\frac{2n}{k})} ,
\end{aligned}$$

where we used Lemma 4.3 with $z = \log(1 + \frac{4}{k}\log\log\frac{2n}{k})$. $\square$

Our next goal is to establish that the order statistics of absolute values of independent standard Gaussian random variables are sub-gamma on the right-tail with variance factor close to the Efron-Stein estimates of variance derived in Proposition 4.1 and scale factor not larger than the square root of the Efron-Stein estimate of variance.

Before describing the consequences of Theorem 2.9, it is interesting to look at what can be obtained from Rényi's representation and exponential inequalities for sums of Gamma distributed random variables.

Proposition 4.4. *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 $s \ge 1$. For $t > 0$,*

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

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

This inequality looks like what we are looking for: $\widetilde{U}(n)(X_{(1)} - \mathbb{E}X_{(1)})$ converges in distribution, but also in quadratic mean, or even according to the Orlicz norm defined by $x \mapsto \exp(|x|) - 1$, toward a centered Gumbel distribution. As the Gumbel distribution is sub-gamma on the right tail with variance factor $\pi^2/6$ and scale factor 1, we expect $X_{(1)}$ to satisfy a Bernstein inequality with variance factor of order $1/\widetilde{U}(n)^2$ and scale factor $1/\widetilde{U}(n)$. Up to the shift δ_n, this is the content of the proposition. Note that the shift is asymptotically negligible with respect to typical fluctuations. The next proposition shows that Theorem 2.9 captures the correct order of growth for the right-tail of Gaussian maxima even though the constants are not sharp enough to make it competitive with Proposition 4.4

Proposition 4.5. *For n such that the solution v_n of the equation $16/x + \log(1 + 2/x + 4\log(4/x)) = \log(2n)$ is smaller than 1, for all $0 \le \lambda < \frac{1}{\sqrt{v_n}}$,*

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

For all $t > 0$,

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

Proof of Proposition 4.5. By Proposition 2.9,

$$\log \mathbb{E} e^{\lambda(X_{(1)} - \mathbb{E}X_{(1)})} \le \frac{\lambda}{2} \mathbb{E}\left[\Delta\left(e^{\lambda\Delta} - 1\right)\right]$$

where $\Delta = X_{(1)} - X_{(2)} \sim U(2e^{Y_{(2)}+E_1}) - U(2e^{Y_{(2)}})$, with E_1 a standard exponentially distributed random variable and independent of $Y_{(2)}$ which is distributed like the 2^{nd} largest order statistics of a standard exponential sample.
On the one hand, the conditional expectation

$$\mathbb{E}\left[\left(U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}})\right)\left(e^{\lambda(U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}}))} - 1\right)|Y_{(2)}\right]$$

is a non-increasing function of $Y_{(2)}$. The maximum is achieved for $Y_{(2)} = 0$, and is equal to :

$$2\int_0^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} x(e^{\lambda x} - 1)\mathrm{d}x \le 2\lambda e^{\frac{\lambda^2}{2}} \ .$$

On the other hand, by Proposition 4.1,

$$U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}}) \le \frac{\sqrt{2}E_1}{\sqrt{(\log 2 + Y_{(2)})}} \ .$$

For $0 \le \mu < 1/2$,

$$\int_0^\infty \mu x(e^{\mu x} - 1)e^{-x}\mathrm{d}x = \frac{\mu^2(2-\mu)}{(1-\mu)^2} \le \frac{2\mu^2}{1 - 2\mu} \ .$$

Hence,

$$\lambda\mathbb{E}\left[\left(U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}})\right)\left(e^{\lambda(U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}}))} - 1\right) \mid Y_{(2)}\right] \le \frac{4\lambda^2}{\log 2 + Y_{(2)}} \frac{1}{1 - \frac{2\sqrt{2}\lambda}{\sqrt{\log 2 + Y_{(2)}}}} \ .$$

Letting $\tau = \log n - \log(1 + 2\lambda^2 + 4\log(4/v_n))$,

$$\log \mathbb{E}\left[e^{\lambda(X_{(1)} - \mathbb{E}X_{(1)})}\right] \le \underbrace{\lambda^2 e^{\lambda^2/2}\mathbb{P}\left\{Y_{(2)} \le \tau\right\}}_{:=\mathrm{i}} + \underbrace{\frac{4\lambda^2}{\log 2 + \tau}\frac{1}{1 - \frac{2\sqrt{2}\lambda}{\sqrt{\log 2 + \tau}}}}_{:=\mathrm{ii}} \ .$$

By Lemma 4.3, (i) $\le \frac{v_n\lambda^2}{4}$.
As $\lambda \le 1/\sqrt{v_n}$ and by assumption on v_n, $\log 2 + \tau \ge 16/v_n$ and (ii) $\le \frac{v_n\lambda^2}{4(1-\sqrt{v_n}\lambda)}$. □

We may also use Theorem 2.9 to provide a Bernstein inequality for the median of absolute values of a standard Gaussian sample. We assume $n/2$ is an integer.

Proposition 4.6. *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 $t > 0$*,*

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

Proof of Proposition 4.6. By Proposition 2.9,

$$\log \mathbb{E} e^{\lambda(X_{(n/2)} - \mathbb{E}X_{(n/2)})} \leq \frac{n}{4}\lambda \mathbb{E}\left[\Delta_{n/2}\left(e^{\lambda \Delta_{n/2}} - 1\right)\right]$$

where $\Delta_{n/2} = X_{(n/2)} - X_{(n/2+1)} \sim U\left(2e^{E_{n/2}/(n/2) + Y_{(n/2+1)}}\right) - U\left(e^{Y_{(n/2+1)}}\right)$ where $E_{n/2}$ is standard exponentially distributed and independent of $Y_{(n/2+1)}$.
By Proposition 4.1,

$$\Delta_{n/2} \leq \frac{\sqrt{2}\lambda E_{n/2}}{(n/2)\sqrt{\log 2 + Y_{(n/2+1)}}} \leq \frac{\sqrt{2}E_{n/2}}{(n/2)\sqrt{\log 2}} = \sqrt{\frac{v_n}{n}} E_{n/2} \ .$$

Reasoning as in the proof of Proposition 4.5,

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

□

As the hazard rate $\phi(x)/\overline{\Phi}(x)$ of the Gaussian distribution tends to 0 as x tends to $-\infty$, the preceding approach does not work when dealing with order statistics of Gaussian samples. Nevertheless, Proposition 4.2 paves the way to simple bounds on the variance of maxima of Gaussian samples.

Proposition 4.7. *Let* $X_1, \ldots, X_n$ *be* n *independent and identically distributed standard Gaussian random variables, let* $X_{(1)} \geq \ldots \geq X_{(n)}$ *be the order statistics.*
For all $n \geq 11$*,*

$$\mathrm{Var}[X_{(1)}] \quad \leq \quad \frac{8/\log 2}{\log(n/2) - \log(1 + 4\log\log(n/2))} + 2^{-n} + \exp(-\tfrac{n}{8}) + \frac{4\pi}{n} \ .$$

Proof of Proposition 4.7. We may generate n independent standard Gaussian random variables in two steps: first generate n independent random signs ($\epsilon_1, \ldots, \epsilon_n$: $\mathbb{P}\{\epsilon_i = 1\} = 1 - \mathbb{P}\{\epsilon_i = -1\} = 1/2$), then generate absolute values of independent standard Gaussian random variables $(V_1, \ldots, V_n)$, the resulting sample $(X_1, \ldots, X_n)$ is obtained as $X_i = \epsilon_i V_i$. Let N be the number of positive random signs.

$$\mathrm{Var}(X_{(1)}) = \underbrace{\mathbb{E}\left[\mathrm{Var}\left(X_{(1)} \mid \sigma(N)\right)\right]}_{\text{i}} + \underbrace{\mathrm{Var}\left(\mathbb{E}\left[X_{(1)} \mid \sigma(N)\right]\right)}_{\text{ii}} \ .$$

Conditionally on $N = m$, if $m \geq 1$, $X_{(1)}$ is distributed as the maximum of a sample of m independent absolute values of standard Gaussian random variables. If $m = 0$, $X_{(1)}$ is negative, its conditional variance is equal to the variance of the minimum of a sample of size n. Hence, letting $V^m_{(k)}$ denote the k^{th} order statistic of a sample of m independent absolute values of standard Gaussian random variables. Let w_m denote the upper bound on the variance of $V^m_{(1)}$ from Proposition 4.2. Note that w_m is non-increasing, and that, by Poincaré's inequality, $\mathrm{Var}(V^n_{(k)}) \leq 1$.

(ii) $\ll$ (i), as

$$\begin{aligned}
\text{(i)} &= \sum_{m=1}^{n} \binom{n}{m} 2^{-n} \operatorname{Var}\left(V_{(1)}^{m}\right) + \binom{n}{0} 2^{-n} \operatorname{Var}\left(V_{(n)}^{n}\right) \\
&= \sum_{m=1}^{n} \binom{n}{m} 2^{-n} w_m + 2^{-n} \\
&\leq \sum_{m=1}^{n/4} \binom{n}{m} 2^{-n} + w_{n/4} + 2^{-n} \\
&\leq \exp(-n/8) + w_{n/4} + 2^{-n} \ .
\end{aligned}$$

Let H_N be the random harmonic number $H_N = \sum_{i=1}^{N} 1/i$,

$$\begin{aligned}
\text{(ii)} &= \mathbb{E}_{N,N'}\left[\Big(\mathbb{E}[X_{(1)} \mid N] - \mathbb{E}[X_{(1)} \mid N']\Big)_+^2\right] \\
&= \mathbb{E}_{N,N'}\left[\Big(\mathbb{E}\left[\widetilde{U}\Big(\exp\left(\textstyle\sum_{i=1}^{N} \frac{E_i}{i}\right)\Big) - \widetilde{U}\Big(\exp\left(\textstyle\sum_{i=1}^{N'} \frac{E_i}{i}\right)\Big)\right]\Big)_+^2\right] \\
&\leq \mathbb{E}_{N,N'}\left[\Big(1/h\big(\widetilde{U}\big(\exp\big(\textstyle\sum_{i=1}^{N'} \frac{E_i}{i}\big)\big)\big)(H_N - H_{N'})\Big)_+^2\right] \\
&\leq \frac{\pi}{2}\mathbb{E}_{N,N'}\left[(H_N - H_{N'})_+^2\right] \\
&= \frac{\pi}{2}\operatorname{Var}(H_N)\,.
\end{aligned}$$

Now, as $N = \sum_{i=1}^{n}(1+\epsilon_i)/2$, letting $Z = H_N$ and $Z_i = \sum_{j=1}^{N-(1+\epsilon_i)/2} 1/j$, by Efron-Stein inequality, $\operatorname{Var} Z \leq \mathbb{E}[0 \wedge 1/N]$. Finally, using Hoeffding's inequality [1] in a crude way leads to $\mathbb{E}[0 \wedge 1/N] \leq \exp(-n/8) + 4/n \leq 8/n$. We may conclude by ii $\leq (4\pi)/n$. □

Remark 4.8. *Trading simplicity for tightness, sharper bounds on ii could be derived and show that ii* $= O(1/(n \log n))$.

References

[1] S. Boucheron, O. Bousquet, and G. Lugosi, *Concentration inequalities*, Machine Learning Summer School 2003, Lecture Notes in Artificial Intelligence, vol. 3176, Springer-Verlag, 2004, 169–207.

[2] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities using the entropy method*, The Annals of Probability **31** (2003), 1583–1614. MR-1989444

[3] L. de Haan and A. Ferreira, *Extreme value theory*, Springer-Verlag, 2006. MR-2234156

[4] B. Efron and C. Stein, *The jackknife estimate of variance*, The Annals of Statistics **9** (1981), 586–596. MR-0615434.

[5] M. Ledoux, *The concentration of measure phenomenon*, AMS, 2001. MR-1849347

[6] ______, *A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices*, Séminaire de probabilités XXXVII, Lecture Notes in Mathematics, vol. 1832, Springer-Verlag, 2003, 360–369. MR-2053053

[7] P. Massart, *About the constants in Talagrand's concentration inequality*, The Annals of Probability **28** (2000), 863–885.MR-1782276

[8] ______, *Concentration inequalities and model selection. Ecole d'eté de Probabilité de Saint-Flour* xxxiv, Lecture Notes in Mathematics, vol. 1896, Springer-Verlag, 2007. MR-2319879

[9] R.G. Miller, *The jackknife : a review*, Biometrika **61** (1974), 1–15. MR-0391366

[10] J. Shao and C. F. J. Wu, *A general theory for jackknife variance estimation*, The Annals of Statistics **17** (1989), 1176–1197. MR-1015145

[11] J.P. Tillich and G. Zémor, *Discrete Isoperimetric Inequalities and the Probability of a Decoding Error*, Combinatorics, Probability and Computing **9** (2001), 465–479. MR-1810152

[12] A. van der Vaart, *Asymptotic statistics*, Cambridge University Press, 1998. MR-1652247

Acknowledgments. The authors gratefully thank Elisabeth Gassiat, Gábor Lugosi, Pascal Massart, Marie Théret and an anonymous reviewer for stimulating remarks and discussions.

A Proof of Proposition 4.4

Let $Y_{(1)}$ denote the maximum of a sample of absolute values of independent standard exponentially distributed random variables, so that $X_{(1)} \sim \widetilde{U}(e^{Y_{(1)}})$.

Thanks to the concavity of $\widetilde{U} \circ \exp$, and the fact that $\overline{\Phi}(x)/\phi(x) \leq 1/x$ for $x > 0$, $\widetilde{U}(\exp(Y_{(1)})) - \widetilde{U}(\exp(H_n)) \leq (Y_{(1)} - H_n)/(\widetilde{U}(\exp(H_n))$. Now, $Y_{(1)}$ satisfies a Bernstein inequality with variance factor not larger than $\mathrm{Var}(Y_{(1)}) \leq \pi^2/6 \leq 2$ and scale factor not larger than 1, so

$$\mathbb{P}\{Y_{(1)} - H_n \geq t\widetilde{U}(\exp(H_n))\} \leq \exp\Big(- \frac{t^2\widetilde{U}(\exp(H_n))^2}{2(2 + t\widetilde{U}(\exp(H_n))/3)}\Big).$$

Agreeing on $\delta_n = \widetilde{U}(e^{H_n}) - \mathbb{E}\widetilde{U}(e^{Y_{(1)}})$, using the monotonicity of $x^2/(2+x/3)$ and $\log n \leq H_n$, we obtain the first part of the proposition.

The second-order extended regular variation condition satisfied by the standard Gaussian distribution [3, Exercise 2.9, p. 61] allows us to assert $\widetilde{U}(n)^3\big(\widetilde{U}(e^x n) - \widetilde{U}(n) - \frac{x}{\widetilde{U}(n)}\big) \to -\frac{x^2}{2} - x$. This suggests that

$$\widetilde{U}(\exp(H_n))^3\left(\mathbb{E}\widetilde{U}(e^{Y_{(1)}-H_n}e^{H_n}) - \widetilde{U}(e^{H_n}) - \frac{\mathbb{E}Y_{(1)} - H_n}{\widetilde{U}(\exp(H_n))}\right) \to -\frac{\pi^2}{12}\,,$$

or that the order of magnitude of $\widetilde{U}\big(e^{H_n}\big) - \mathbb{E}X_{(1)}$ is $O(1/\widetilde{U}(n)^3)$ which is small with respect to $1/\widetilde{U}(n)$.

Use Theorem B.3.10 from [3], to ensure that for all $\delta, \epsilon > 0$, for n, x such that $\min(n, n\exp(x)) \geq t_0(\epsilon, \delta)$, $\left|\widetilde{U}(n)^3\left(\widetilde{U}(e^x n) - \widetilde{U}(n) - \frac{x}{\widetilde{U}(n)}\right) + \frac{x^2}{2} + x\right| \leq \epsilon\exp(\delta|x|)$. The probability that $Y_{(1)} \leq H_n/2$ is less than $\exp(-H_n/3)$. When for $Y_{(1)} - H_n \leq 0$, the supremum of $\left|\widetilde{U}(\exp(H_n))^3\left(\widetilde{U}(e^{Y_{(1)}-H_n}e^{H_n}) - \widetilde{U}(e^{H_n}) - \frac{Y_{(1)}-H_n}{\widetilde{U}(\exp(H_n))}\right) + \frac{(Y_{(1)}-H_n)^2}{2} + Y_{(1)} - H_n\right|$, is achieved when $Y_{(1)} - H_n = -H_n$, it is less than $4(\log n)^2$. The dominated convergence theorem allows us to conclude.