Improved Convergence Rates of Normal Extremes

Yijun Zhu and Han Xiao

Abstract It is well know that the convergence of the normal extremes to the limiting Gumbel distribution is extremely slow, at the rate of $(\log n)^{-1}$. We show that after a monotone transform, the convergence rate of the squared normal extremes can be improved to $(\log n)^{-3}$. Simulations confirm that the convergence is much faster than existing results uniformly, especially when the sample is of moderate sizes around hundreds or thousands. More importantly, it is observed that the convergence rate at the upper tail is substantially improved, which has direct implications for hypothesis tests based on maximum type test statistics.

1 Introduction

Let $X_1, X_2, \dots,$ be a sequence of independent standard normal random variables, and let $M_n := \max\{X_1, X_2, \dots, X_n\}$ be the maximum of the first n of them. According to the extreme value theory (see Leadbetter et al., 1983, for an overview), after proper centering and rescaling, the limiting distribution of M_n is the extreme value distribution of type I, or the so called Gumbel distribution, with the distribution function $G_1(x) = \exp(-e^{-x})$. In fact, if we define

$$\alpha_n = (2\log n)^{-1/2}$$
$$\beta_n = \sqrt{2\log n} - \frac{\log(\log n) + \log(4\pi)}{2\sqrt{2\log n}}$$

then $\alpha_n^{-1}(M_n - \beta_n)$ converges to G_1 in distribution, i.e.

Yijun Zhu
Rutgers University, Piscataway, NJ, USA, e-mail: yijunzhu@stat.rutgers.edu

Han Xiao
Rutgers University, Piscataway, NJ, USA, e-mail: hxiao@stat.rutgers.edu

$$\lim_{n\to\infty} P\left[\alpha_n^{-1}(M_n - \beta_n) \le x\right] = \lim_{n\to\infty} \Phi^n(\alpha_n x + \beta_n) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}, \tag{1}$$

where $\Phi(\cdot)$ is the distribution function of $N(0, 1)$.

The rate of convergence in (1) is extremely slow. The fact was noted by Fisher and Tippett (1928), and studied more precisely by Hall (1979), who proved that the convergence rate in (1) is no better than $(\log\log n)^2/\log n$. Hall (1979) also found that if β_{n1} is the solution of the equation

$$2\pi\beta_{n1}^2 \exp(\beta_{n1}^2) = n^2, \tag{2}$$

and $\alpha_{n1} = \beta_{n1}^{-1}$, then

$$\frac{C_1}{\log n} < \sup_{-\infty<x<\infty} \left|P[\alpha_{n1}^{-1}(M_n - \beta_{n1}) \le x] - G_1(x)\right| < \frac{C_2}{\log n}, \tag{3}$$

where C_1 and C_2 are absolute constants. In other words, the convergence rate can be improved to $(\log n)^{-1}$ by choosing a better centering constant β_{n1}. In the same paper, it was further proved that the rate cannot be better than $(\log n)^{-1}$ by choosing a different sequence of normalizing constants.

It is equivalent and sometimes more convenient to study the limiting behavior of M_n through its squared version M_n^2. There are counterparts of (1) and (3) for M_n^2. More importantly, Hall (1980) found that with a suitably chosen constants a_n and b_n, the normalized sequence $a_n^{-1}(M_n^2 - b_n)$ converges to $G_1(x)$ with the rate $(\log n)^{-2}$. A detailed overview of the progression regarding the convergence rates of normal extremes will be provided in Section 2.3 via the squared version M_n^2.

While the aforementioned results are all on the uniform convergence rates, the convergence to G_1 in the upper tail is of particular interests when performing hypothesis tests using maximum type statistics. For example, the stepdown procedure of Romano and Wolf (2005) for multiple testing requires the knowledge about the upper quantiles of the maximum test statistic. Cai et al. (2014) used the maximum coordinate-wise difference of two transformed sample mean vectors to test the equality of two high dimensional means.

In Figure 1 we plot the empirical distributions of M_n^2 with difference choices of normalizing sequences. The black line is the theoretical cumulative distribution function (CDF) G_1, the dashed, red and green lines (labeled by b_{n1}, b_{n2} and b_{n3} *resp.*) are empirical CDF corresponding to convergence rates in (1), (3) and $(\log n)^{-2}$ respectively. Figure 2 zooms in on the upper tails. Despite the fact that the red line is associated with a faster convergence rate than that of the dashed one, Figure 2 shows that it is consistently farther from the theoretical CDF in the upper tail, even when the sample size is as large as 10^5. This needs not contradicts the theories on the uniform convergence rates, because we see in Figure 1 that the dashed line deviates apparently from the black one in the lower tail. However, tests based on the statistic in (3) will be quite off, and have no advantage over the statistic in (1). On the other hand, the green line, corresponding to the rate $(\log n)^{-2}$, shows the potential to outperform the dashed one, when the sample size is sufficiently large, as shown in the bottom right panel of Figure 2. The issue is that the green line is below the theoretical CDF, indicating that the corresponding asymptotic test is not conservative.

Our major finding is that the convergence rate can be further improved to $(\log n)^{-3}$ by applying a monotone transform to M_n^2. Let $b_n := \frac{1}{2}[\Phi^{-1}(1 - 1/n)]^2$. Define Y_n through the following transform of M_n^2:

$$Y_n := \left[1 - \left(1 + \frac{M_n^2 - 2b_n}{8b_n^2}\right)^{-1}\right]\left(4b_n^2 + 2b_n - 2\right),$$

The results in Section 2 imply the following rate of convergence

$$\sup_{-\infty<x<\infty} |P(Y_n \le x) - G_1(x)| < \frac{C_3}{(\log n)^3}.$$

The blue lines in Figure 1 give empirical CDF of Y_n, which are almost identical with G_1 even when the sample size is as small as 200. When zoomed into the upper tail in Figure 2, the faster convergence of Y_n is more clearly seen. Furthermore, if Y_n is used as the test statistic for the asymptotic test, it is not only more accurate, but also always conservative, since the blue curve sits above the black one (for G_1) in the upper tail.

The rest of this article is organized as follows. We present and prove the pointwise and uniform convergence rates of Y_n in Section 2.1 and Section 2.2 respectively. In Section 2.3 we demonstrate how the faster convergence rate is achieved by comparing with existing results. Similar convergence rates regarding the k-th maxima are presented in Section 2.4. Numerical analysis and an application on testing the covariance structure are given in Section 3. Additional figures, tables, and some technical results are relegated in the Appendix.

We conclude this section by a brief review of the literature on the convergence rates of normal extremes. Cohen (1982b) showed that the penultimate approximation can achieve the $(\log n)^{-2}$ rate, and considered the extension to other types of extreme value distributions in Cohen (1982a). Daniels (1982) proposed another nonlinear transformation which leads to faster convergence. Rootzén (1983) investigated the convergence rates of the extremes from a stationary Gaussian process. Hall (1991) found that the extreme of a continuous time Gaussian process also has a logarithmic convergence rate. For convergence rates of extremes from a non-Gaussian sequence, we refer to Hall and Wellner (1979), Smith (1982), Leadbetter et al. (1983), de Haan and Resnick (1996), Peng et al. (2010) and references therein.

2 Main results

We will first consider the pointwise convergence rates in Section 2.1, and then illustrate how the faster rates are achieved by modifying the normalizing constants and applying a transform of M_n^2 in Section 2.3. The uniform convergence rates are given in Section 2.2. In Section 2.4 we present the corresponding results for the k-th maxima. We make the convention that $C, C_1, C_2, \ldots$ are generic absolute constants, whose values may vary from place to place.

2.1 Pointwise convergence rates

Let b_n be the solution of the equation $1 - \Phi(\sqrt{2b_n}) = 1/n$. Recall that Y_n is defined as:

$$Y_n := \left[1 - \left(1 + \frac{M_n^2 - 2b_n}{8b_n^2}\right)^{-1}\right] \left(4b_n^2 + 2b_n - 2\right). \tag{4}$$

According to the definition, $\sqrt{2b_n}$ is the $(1 - 1/n)$-th quantile of the standard normal distribution. Since $M_n^2 \geq 0$ and $b_5 \approx .35$, the transform given in (4) is strictly monotone when $n \geq 5$, which we shall assume in the sequel.

Using the Newton-Raphson approximation (see Appendix 5.1 for detailed derivations), it can be shown that

$$b_n = \log n - \tfrac{1}{2} \log \log n - \tfrac{1}{2} \log 4\pi + O(\log \log n / \log n).$$

We first prove the pointwise convergence rate of Y_n to G_1. It is convenient to express the result through b_n, which is of the order $\log n$.

Theorem 1 *For each fixed $-\infty < x < \infty$,*

$$P(Y_n \leq x) - G_1(x) = G_1(x)e^{-x} \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4}).$$

Proof Define the function $g_n(x)$ as the inverse transform of (4)

$$g_n(x) = \left[\left(1 - \frac{x}{4b_n^2 + 2b_n - 2}\right)^{-1} - 1\right] \cdot 8b_n^2 + 2b_n \tag{5}$$

Since (4) is a monotone transform, the event $[Y_n \leq x]$ is equivalent to $[M_n^2 \leq g_n(x)]$. It can be shown that

$$g_n(x) = 2b_n + 2x - \frac{x}{b_n} + \frac{x^2 + 3x}{2b_n^2} - \frac{2x^2 + 5x}{4b_n^3} + O(b_n^{-4}). \tag{6}$$

When n is large enough, $g_n(x) > 0$, and we let $x_n = [g_n(x)]^{1/2}$. Note that

$$P(M_n \leq x_n) > P(M_n^2 \leq x_n^2) = P(M_n \leq x_n) - P(M_n < -x_n) > P(M_n \leq x_n) - 2^{-n}. \tag{7}$$

According to Lemma 2.4.1 in Leadbetter et al. (1983), for any $0 \leq z \leq n$,

$$0 \leq e^{-z} - \left(1 - \frac{z}{n}\right)^n \leq \frac{z^2 e^{-z}}{2} \cdot \frac{1}{n-1}. \tag{8}$$

Let $\tau_n(x) = n\,[1 - \Phi(x_n)]$, it follows that

$$P(M_n \leq x_n) = [1 - (1 - \Phi(x_n))]^n = \exp[-\tau_n(x)] + O(n^{-1}). \tag{9}$$

To evaluate $\tau_n(x)$, we make use the following series expansion of the normal tail probability (Abramowitz and Stegun, 1964): for any $z > 0$, and any positive integer m,

$$1 - \Phi(z) = \frac{\phi(z)}{z}\left\{1 - \frac{1}{z^2} + \frac{1\cdot 3}{z^4} + \cdots + \frac{(-1)^m 1\cdot 3 \ldots (2n-1)}{z^{2m}} + R_m\right\},$$

where

$$R_m = (-1)^{m+1}(2m+1)!! \int_z^\infty \frac{\phi(t)}{t^{2m+2}}\,dt,$$

which is less in absolute value than the first neglected term. In particular, when $m = 3$, it holds that for any $z > 0$,

$$\left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - \frac{15}{z^7}\right)\phi(z) < 1 - \Phi(z) < \left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - \frac{15}{z^7} + \frac{105}{z^9}\right)\phi(z) \tag{10}$$

According to the definition of $\tau_n(x)$ and (10), we first do the Taylor expansion (up to the order b_n^{-4}) for

$$\begin{aligned}\phi(x_n) &= \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-b_n - x + \frac{x}{2b_n} - \frac{x^2+3x}{4b_n^2} + \frac{2x^2+5x}{8b_n^3}\right) \\ &= \frac{e^{-x}e^{-b_n}}{\sqrt{2\pi}} \cdot \left(1 + \frac{x}{2b_n} - \frac{x^2+6x}{8b_n^2} - \frac{5x^3+6x^2-30x}{48b_n^3} + O(b_n^{-4})\right),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{x_n} &= \left(2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3}\right)^{-1/2} \\ &= \frac{1}{\sqrt{2b_n}}\left(1 - \frac{x}{2b_n} + \frac{3x^2+2x}{8b_n^2} - \frac{5x^3+8x^2+6x}{16b_n^3} + O(b_n^{-4})\right).\end{aligned}$$

Combining the two preceding equations and rearranging the terms, we have

$$\frac{\phi(x_n)}{x_n} = \frac{e^{-x}e^{-b_n}}{\sqrt{4\pi b_n}} \cdot \left(1 - \frac{x}{2b_n^2} - \frac{4x^3+3x^2-6x}{24b_n^3} + O(b_n^{-4})\right).$$

According to (10), we also calculate

$$
\begin{aligned}
1-\frac{1}{x_n^2}+\frac{3}{x_n^4}-\frac{15}{x_n^6} &= 1-\frac{1}{2b_n}+\frac{2x+3}{4b_n^2}-\frac{4x^2+14x+15}{8b_n^3}+O(b_n^{-4}) \\
&= \left(1-\frac{1}{2b_n}+\frac{3}{4b_n^2}-\frac{15}{8b_n^3}\right)\cdot\left(1+\frac{x}{2b_n^2}-\frac{x^2+3x}{2b_n^3}+O(b_n^{-4})\right)
\end{aligned}
$$

Recall b_n is the solution of the equation $1-\Phi(\sqrt{2b_n}) = 1/n$. According to the approximation to normal probability function in (10), we have

$$
\frac{ne^{-b_n}}{\sqrt{4\pi b_n}} = \left(1-\frac{1}{2b_n}+\frac{3}{4b_n^2}-\frac{15}{8b_n^3}+O(b_n^{-4})\right)^{-1}. \tag{11}
$$

Therefore,

$$
\begin{aligned}
&\left(1-\frac{1}{x_n^2}+\frac{3}{x_n^4}-\frac{15}{x_n^6}\right)\frac{n\phi(x_n)}{x_n} \\
&\quad = e^{-x}\cdot\left(1-\frac{x}{2b_n^2}-\frac{4x^3+3x^2-6x}{24b_n^3}+O(b_n^{-4})\right) \\
&\qquad \cdot\left(1-\frac{1}{2b_n}+\frac{3}{4b_n^2}-\frac{15}{8b_n^3}+O(b_n^{-4})\right)^{-1} \\
&\qquad \cdot\left(1-\frac{1}{2b_n}+\frac{3}{4b_n^2}-\frac{15}{8b_n^3}\right)\cdot\left(1+\frac{x}{2b_n^2}-\frac{x^2+3x}{2b_n^3}+O(b_n^{-4})\right) \\
&\quad = e^{-x}\left(1-\frac{4x^3+15x^2+30x}{24b_n^3}+O(b_n^{-4})\right)
\end{aligned}
$$

Since $n\phi(x_n)/x_n^9 = O(b_n^{-4})$, we have by (10)

$$
\tau_n(x) = e^{-x}\left(1-\frac{4x^3+15x^2+30x}{24b_n^3}\right)+O(b_n^{-4}).
$$

According to (9), it follows that

$$
\begin{aligned}
P(Y_n \le x) - G_1(x) &= \exp(-\tau_n(x)) + O(n^{-1}) - G_1(x) \\
&= G_1(x)e^{-x}\cdot\frac{4x^3+15x^2+30x}{24b_n^3}+O(b_n^{-4}).
\end{aligned}
$$

The proof is complete. □

Using (10) and Newton-Raphson method, we have the following expansions for b_n

$$
b_n = \log n - \frac{\Delta}{2} + \frac{\Delta-2}{4\log n} + \frac{\Delta^2-6\Delta+14}{16(\log n)^2} + O\left(\frac{(\log\log n)^3}{(2\log n)^3}\right), \tag{12}
$$

where

$$
\Delta = \log\log n + \log 4\pi.
$$

Therefore, Theorem 1 implies that Y_n converges to G_1 with the rate $(\log n)^{-3}$. The detailed derivation of (12) is given in the Appendix.

2.2 Uniform convergence rate

In this section we establish the uniform convergence rate.

Theorem 2 *There exists an absolute constant* c_1*, such that*

$$\sup_{-\infty<x<\infty} |P(Y_n \le x) - G_1(x)| < \frac{c_1}{(\log n)^3}.$$

We prove Theorem 2 using two lemmas. Recall that $g_n(x)$, defined in (5), is the inverse transform of (4).

Lemma 1 *Let* $\{c_n\}$ *be an increasing sequence of positive integers such that* $c_n^4/b_n \to 0$*, then*

$$g_n(x) = 2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3} + \frac{d_{1n}(x)}{b_n^3},$$

where $\lim_{n\to\infty} \sup_{-c_n \le x \le c_n} |d_{1n}(x)| = 0$.

Proof According to (5), for $-c_n \le x \le c_n$, we can obtain the following expansion:

$$\begin{aligned} g_n(x) &= 2b_n + 8b_n^2 \cdot \left[\left(1 - \frac{x}{4b_n^2 + 2b_n + 2}\right)^{-1} - 1\right] \\ &= 2b_n + 2x\gamma_n + \frac{x^2\gamma_n^2}{2b_n^2} + \frac{x^3\gamma_n^3}{8b_n^4} \cdot \left(1 - \frac{x\gamma_n}{4b_n^2}\right)^{-1}, \end{aligned} \tag{13}$$

where

$$\gamma_n = \left(1 + \frac{1}{2b_n} - \frac{1}{2b_n^2}\right)^{-1}.$$

When $n \ge 13$, $b_n > 1$, by series expansion of γ_n, we have

$$\begin{aligned} \gamma_n &= 1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{5}{8b_n^3} + \frac{e_{1n}}{b_n^4} \\ \gamma_n^2 &= 1 - \frac{1}{b_n} + \frac{e_{2n}}{b_n^2} \\ \gamma_n^3 \left(1 - \frac{x\gamma_n}{4b_n^2}\right)^{-1} &= 1 + e_{3n}. \end{aligned}$$

The following bounds can be easily verified: $|e_{1n}| \le 1$, $|e_{2n}| \le 2$ and $|e_{3n}| \le 1$. Then by simplifying (13) we have

$$g_n(x) = 2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3} + \frac{16xe_{1n} + 4x^2e_{2n} + x^3(1+e_{3n})}{8b_n^4}.$$

The proof is completed by noting that

$$\sup_{-c_n \le x \le c_n} \left| \frac{16xe_{1n} + 4x^2e_{2n} + x^3(1+e_{3n})}{8b_n} \right| \le \frac{8c_n + 4c_n^2 + c_n^3}{4b_n} \to 0$$

under the condition $c_n^4/b_n \to 0$.

Lemma 2 *Let $\{c_n\}$ be the same sequence as used in Lemma 1, then*

$$\tau_n(x) = e^{-x}\left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + \frac{d_{2n}(x)}{b_n^3}\right),$$

where $\lim_{n\to\infty} \sup_{-c_n \le x \le c_n} |d_{2n}(x)| = 0$ *for all* $-c_n \le x \le c_n$.

Proof Recall that $x_n := [g_n(x)]^{1/2}$. Using the normal tail probability bound in (10), we have

$$\left|\tau_n(x) - n\phi(x_n)\left(\frac{1}{x_n} - \frac{1}{x_n^3} + \frac{3}{x_n^5} - \frac{15}{x_n^7}\right)\right| \le \frac{105n\phi(x_n)}{x_n^9}. \tag{14}$$

Write

$$n\phi(x_n)\left(\frac{1}{x_n} - \frac{1}{x_n^3} + \frac{3}{x_n^5} - \frac{15}{x_n^7}\right) = \left(\frac{x_n}{\sqrt{2b_n}}\right)^{-1} \cdot \frac{n\phi(x_n)}{\sqrt{2b_n}} \cdot \left(1 - \frac{1}{x_n^2} + \frac{3}{x_n^4} - \frac{15}{x_n^6}\right). \tag{15}$$

Let

$$x_{1n} := \frac{x}{b_n} - \frac{x}{2b_n^2} + \frac{x^2+3x}{4b_n^3} - \frac{2x^2+5x}{8b_n^4} + \frac{d_{1n}(x)}{2b_n^4},$$

where $d_{1n}(x)$ is defined in Lemma 1. For the first term on the right hand side of (15), by Lemma 1,

$$\left(\frac{x_n}{\sqrt{2b_n}}\right)^{-1} = (1+x_{1n})^{-1/2} = 1 - \frac{x_{1n}}{2} + \frac{3x_{1n}}{8} - \frac{5x_{1n}^3}{16} + R_{1n}(x_{1n}). \tag{16}$$

Under the condition $c_n^4/b_n \to 0$, it holds that $\sup_{-c_n \le x \le c_n} |x_{1n}| \le 5c_n/b_n$, and thus

$$\sup_{-c_n \le x \le c_n} |R_{1n}(x)| = \frac{o(1)}{b_n^3}.$$

The terms on the right hand side of (16) except for $R_{1n}(x_{1n})$ can be expanded as

$$\left(\frac{x_n}{\sqrt{2b_n}}\right)^{-1} - R_{1n}(x_{1n}) = 1 - \frac{x}{2b_n} + \frac{3x^2+2x}{8b_n^2} - \frac{5x^3+8x^2+6x}{16b_n^3} + \frac{d_{3n}(x)}{b_n^3}.$$

Note that for each fractional term in x_{1n}, the power of x is no greater than that of b_n, and the same claim holds for the series $d_{3n}(x)/b_n^3$. Furthermore, the first term (of the smallest power of x) in the expansion of $d_{3n}(x)$ is x^3/b_n, which goes to 0 uniformly over $-c_n \le x \le c_n$. Therefore, we conclude

$$\lim_{n\to\infty} \sup_{-c_n \le x \le c_n} |d_{3n}(x)| = 0$$

The other two terms in (15) can be treated similarly:

$$\frac{n\phi(x_n)}{\sqrt{2b_n}} = \frac{ne^{-x}e^{-b_n}}{\sqrt{4\pi b_n}} \cdot \left(1 + \frac{x}{2b_n} - \frac{x^2+6x}{8b_n^2} - \frac{5x^3+6x^2-30x}{48b_n^3} + \frac{d_{4n}(x)}{b_n^3} + R_{2n}(x)\right),$$

$$1 - \frac{1}{x_n^2} + \frac{3}{x_n^4} - \frac{15}{x_n^6} = \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{15}{8b_n^3}\right) \cdot \left(1 + \frac{x}{2b_n^2} - \frac{x^2+3x}{2b_n^3} + \frac{d_{5n}(x)}{b_n^3} + R_{3n}(x)\right),$$

where

$$\sup_{-c_n \le x \le c_n} |d_{4n}(x)| \to 0 \quad \text{and} \quad |R_{2n}(x)| = \frac{o(1)}{b_n^3},$$
$$\sup_{-c_n \le x \le c_n} |d_{5n}(x)| \to 0 \quad \text{and} \quad |R_{3n}(x)| = \frac{o(1)}{b_n^3}.$$

Combining all the preceding bounds together with (11), we have

$$n\phi(x_n)\left(\frac{1}{x_n} - \frac{1}{x_n^3} + \frac{3}{x_n^5} - \frac{15}{x_n^7}\right) = e^{-x}\left(1 - \frac{4x^3+15x^2+30x}{24b_n^3} + \frac{d_{6n}(x)}{b_n^3}\right).$$

Using similar arguments as those for d_{3n}, we can verify that

$$\lim_{n\to\infty} \sup_{-c_n \le x \le c_n} |d_{6n}(x)| = 0.$$

It is easy to show that $\sup_{-c_n \le x \le c_n} n\phi(x_n)/x_n^9 = o(b_n^{-3})$. So the proof is complete in view of (14).

We are now ready to prove Theorem 2.

Proof (Proof of Theorem 2) Let c_1 be a generic absolute constant which may vary from place to place. We consider three scenarios: $x < -c_n$, $-c_n \le x \le c_n$ and $x > c_n$, with $c_n = 4\log b_n$. Obviously, this choice of c_n satisfies the condition $c_n^4/b_n \to 0$.

We begin with the situation $-c_n \le x \le c_n$. By (7), it holds that

$$\left|P(Y_n \le x) - \left(1 - \frac{\tau_n(x)}{n}\right)^n\right| \le 2^{-n}.$$

By (8) and Lemma 2, we have

$$|P(Y_n \le x) - G_1(x)| \le 2G_1(x)e^{-x}\left(\frac{|4x^3+15x^2+30x|}{24b_n^3} + \frac{|d_{2n}(x)|}{b_n^3}\right) + \frac{1}{2^n} + \frac{1}{n},$$

when n is large enough. Since $\sup_{-c_n \le x \le c_n} |d_{2n}(x)| \to 0$, it suffices to show that

$$\sup_{-c_n \le x \le c_n} \left|G_1(x)e^{-x}(4x^3+15x^2+30x)\right| < \infty.$$

Numerical evaluations show that

$$\sup_{-\infty < x < \infty} \left|G_1(x)e^{-x}(4x^3+15x^2+30x)\right| < 20$$

Therefore, we have

$$\sup_{-c_n < x < c_n} |P(Y_n \le x) - G_1(x)| < \frac{c_1}{(\log n)^3}.$$

Now we consider the second scenario $x > c_n$. We will show that both $G_1(x)$ and $P(Y_n \le x)$ are close to 1, and their differences from 1 are of the order $1/(\log n)^3$. Since $x > c_n = 4\log b_n$,

$$G_1(x) = \exp(-e^{-x}) > \exp(-b_n^4) \ge 1 - 1/b_n^4. \tag{17}$$

On the other hand, recall the definition of $g(\cdot)$ in (5)

$$\begin{aligned} 1 - P(Y_n \le x) \le P(Y_n \ge 4\log b_n) &= P\left[M_n^2 \ge g(4\log b_n)\right] \\ &\le P\left(M_n^2 \ge 2b_n + 4\log b_n \cdot \frac{8b_n^2}{4b_n^2 + 2b_n - 2}\right) \end{aligned}$$

Note that $8b_n^2/(4b_n^2 + 2b_n - 2) > 1.5$ for $n \ge 33$. Let $y_n^2 = 2b_n + 6\log b_n$, then

$$P(M_n^2 \ge y_n^2) \le P(M_n \ge y_n) + 1/2^n.$$

Let $\tau_n = n[1 - \Phi(y_n)]$. Using the normal tail probability bounds (10), we have

$$\begin{aligned} \tau_n &\le \frac{n}{\sqrt{2\pi}}\,(2b_n + 6\log b_n)^{-1/2} \cdot \exp(-b_n - 3\log b_n) \\ &= \frac{ne^{-b_n}}{\sqrt{3\pi b_n}}\left(1 + \frac{3\log b_n}{b_n}\right)^{-1/2} \cdot \exp(-3\log b_n) \end{aligned}$$

Recall $1 - \Phi(\sqrt{2b_n}) = 1/n$, so that by (10)

$$\frac{ne^{-b_n}}{\sqrt{4\pi b_n}}\left(1 - \frac{1}{2b_n}\right) < 1.$$

When $n \ge 33$, we have

$$\left(1 + \frac{3 \log b_n}{b_n}\right)^{-1/2} \cdot \left(1 - \frac{1}{2b_n}\right)^{-1} < 1,$$

and it follows that

$$\tau_n < \exp(-3 \log b_n) = 1/b_n^3.$$

Using (8), we deduce that when n is large enough

$$P(M_n \geq y_n) = 1 - (\Phi(y_n))^n = 1 - \left(1 - \frac{\tau_n}{n}\right)^n \leq 1 - e^{-\tau_n} + \frac{1}{n-1} \leq \tau_n + \frac{1}{n-1}.$$

Therefore, we conclude

$$1 - P(Y_n \leq x) < \frac{1}{b_n^3} + \frac{1}{n-1} + \frac{1}{2^n} < \frac{c_1}{(\log n)^3},$$

for some absolute constant c_1. The preceding inequality, together with (17), completes the proof for $x > c_n$.

Finally we consider $x < -c_n$ by showing that both $G_1(x)$ and $P(Y_n \leq x)$ converge to 0 faster than $1/(\log n)^3$. Using the definition of b_n, we have when $n \geq 33$, and $x < -c_n = -4 \log b_n$,

$$G_1(x) = \exp(-e^{-x}) < \exp(-b_n^4) < 1/b_n^4.$$

On the other hand, when $x \leq -4 \log b_n$,

$$P(Y_n \leq x) \leq P[M_n^2 \leq g(-4 \log b_n)] \leq P\left(M_n^2 \leq 2b_n - 4 \log b_n \cdot \frac{8b_n^2}{4b_n^2 + 2b_n - 2}\right).$$

Again since $8b_n^2/(4b_n^2 + 2b_n - 2) > 1.5$ when $n \geq 33$, if we let ${y_n'}^2 = 2b_n - 6 \log b_n$, then

$$P(Y_n \leq x) \leq P(M_n \leq y_n).$$

Let $\tau_n' = n[1 - \Phi(y_n')]$, we have by (10)

$$\begin{aligned}
\exp(-\tau_n') &< \exp\left\{-\frac{ne^{-b_n}}{\sqrt{4\pi b_n}}\left(1 - \frac{3 \log b_n}{b_n}\right)^{-1/2}\right. \\
&\qquad \left.\cdot \left(1 - \frac{1}{(2b_n - 6 \log b_n)^2}\right) \cdot \exp(3 \log b_n)\right\} \\
&< \exp\{-\exp(3 \log b_n)\} \\
&< 1/b_n^3,
\end{aligned}$$

when n is large enough. We conclude by (8)

$$P(Y_n \leq x) < \frac{1}{b_n^3} + \frac{1}{n} < \frac{c_1}{(\log n)^3},$$

which completes the proof. □

2.3 Comparisons of different convergence rates

The best uniform convergence rate that can be obtained for M_n^2, if only centering and rescaling is allowed, is $(\log n)^{-2}$. We will give a summary of the progression in the literature. We also explain why the transformed M_n^2 can have a faster convergence rate $(\log n)^{-3}$.

In order for M_n^2 to have the limiting distribution G_1, the simplest option is to choose

$$b_{n1} = \log n - \log(\log n)/2 - \log(4\pi)/2;$$

then as a counterpart of (1), it holds that $\frac{1}{2}(M_n^2 - 2b_{n1}) \Rightarrow G_1$, where we use $\Rightarrow$ to denote the convergence in distribution. Using similar arguments as given in Hall (1979), it can be shown that the convergence rate is $(\log\log n)^2/\log n$. Similarly as (2), if b_{n1} is the solution of the equation

$$4\pi b_{n2} \exp(2b_{n2}) = n^2,$$

and M_n^2 is centered by b_{n2}, then the rate of convergence is analogous to (3)

$$\frac{C_1}{\log n} < \sup_{-\infty<x<\infty} \left| P\left[\tfrac{1}{2}(M_n^2 - 2b_{n2}) \leq x\right] - G_1(x) \right| < \frac{C_2}{\log n}. \tag{18}$$

Again (18) can be established following the proof in Hall (1979).

We note that $\sqrt{2b_{n1}}$ is an approximation of the $(1-1/n)$-th quantile of standard normal distribution obtained by using the following approximation of the tail probability:

$$1 - \Phi\left(\sqrt{2b_{n1}}\right) \approx \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\log n}} \cdot \exp(-b_{n1}) = \frac{1}{n};$$

and b_{n2} is obtained by the following approximation of $1 - \Phi(\sqrt{2b_{n2}})$:

$$1 - \Phi\left(\sqrt{2b_{n2}}\right) \approx \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2b_{n2}}} \cdot \exp(-b_{n2}) = \frac{1}{n}.$$

If we choose b_{n3} through a more precise approximation of $1 - \Phi(\sqrt{2b_{n3}})$:

$$1 - \Phi\left(\sqrt{2b_{n3}}\right) \approx \frac{1}{\sqrt{4\pi b_{n3}}} \left(1 - \frac{1}{2b_{n3}}\right) \exp\left(-b_{n3}\right) = \frac{1}{n},$$

and set $a_{n3} = 2 - 1/b_{n3}$, then $a_{n3}^{-1}(M_n^2 - 2b_{n3}) \Rightarrow G_1$ with the convergence rate

$$\frac{C_1}{(\log n)^2} < \sup_{-\infty<x<\infty} \left| P\left[a_{n3}^{-1}(M_n^2 - 2b_{n3}) \le x\right] - G_1(x) \right| < \frac{C_2}{(\log n)^2}. \tag{19}$$

The way we represent the preceding result is slightly different from the original one given by Hall (1980). The choices of a_{n3} and b_{n3} differ from those in Hall (1980) by smaller order terms, which do not affect the convergence rates. We choose the current formulation in order to have a better comparison with our main result.

To achieve a better rate of convergence, we first choose b_n precisely through $1 - \Phi(\sqrt{2b_n}) = 1/n$. Second, observe that the events in (18) and (19) can be written as

$$M_n^2 \le 2b_{n2} + 2x$$
$$M_n^2 \le 2b_{n3} + 2x - x/b_{n3}$$

respectively. According to (6), the event $[Y_n \le x]$ implies that

$$M_n^2 \le 2b_n + 2x - \frac{x}{b_n} + \frac{x^2 + 3x}{2b_n^2} + O\left(b_n^{-3}\right). \tag{20}$$

We see that a term of order $O(b_n^{-2})$ is needed on the right hand side to achieve the convergence rate $(\log n)^{-3}$ in Theorem 1. In fact, it is this expansion which motivates the proposed nonlinear transform Y_n.

2.4 k-th maxima

In this section we present pointwise and uniform convergence rates for the k-th maxima $M_{n,k}$, defined as the k-th largest among the first n variables $\{X_1, X_2, \ldots, X_n\}$. These results follow from almost the same arguments as those for the maxima, so we state them without proofs.

Theorem 3 *Let b_n be the solution of the equation $1 - \Phi(\sqrt{2b_n}) = 1/n$. For an given positive integer k, define*

$$Y_{n,k} := \left[1 - \left(1 + \frac{M_{n,k}^2 - 2b_n}{8b_n^2}\right)^{-1}\right]\left(4b_n^2 + 2b_n - 2\right).$$

(i) For each fixed $-\infty < x < \infty$, it holds that

$$P(Y_{n,k} \le x) - G_k(x) = G_1(x)\frac{e^{-kx}}{(k-1)!} \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4}),$$

where $G_k(x) := G_1(x)\sum_{j=0}^{k-1} e^{-jx}/j!$.

(ii) There exists a constant $c_2 > 0$, such that

$$\sup_{-\infty<x<\infty} |P(Y_{n,k} \le x) - G_k(x)| < \frac{c_2}{(\log n)^3}.$$

3 Applications and numerical comparisons

3.1 Numerical comparisons

In this section, we numerically compare the convergence rates of different versions of the normalized M_n^2, introduced in Section 2.3. Specifically, we compare with $G_1(x)$, the CDF of $Y_{n1} := \frac{1}{2}(M_n^2 - 2b_{n1})$, $Y_{n2} := \frac{1}{2}(M_n^2 - 2b_{n2})$, $Y_{n3} := (2 - 1/b_{n3})^{-1}(M_n^2 - 2b_{n3})$, and Y_n, labeled by b_{n1}, b_{n2}, b_{n3} and b_n respectively in Figure 1. The vertical lines mark 90%, 95% and 99% quantiles of the Gumbel distribution. We see that the distribution of Y_n (blue curve) is uniformly closer to $G_1(x)$, no matter what the sample size is. Figure 2 zooms into the upper tail for a clearer visualization. An interesting finding is that the faster theoretical convergence rates of Y_{n2} and Y_{n3} over Y_{n1}, are not reflected through the plots for Y_{n2} even when the sample size is as large as 10^5. The distribution of Y_{n3} starts to be closer to $G_1(x)$ in the upper tail when $n = 10^5$. We remark that the inferior performances of Y_{n2} and Y_{n3} need not necessarily contradicts the theoretical convergence rates: from Figure 1 it is seen that the convergence of Y_{n1} is much slower in the left tail. On the other hand, in Figure 2 it is more clearly seen that Y_n always has a faster convergence rate, compared with the rest. Further more, the CDF of Y_n lies above $G_1(x)$, indicating that if a hypothesis test is based on the maximum type statistic, then it is guaranteed to be conservative by using Y_n. This is in contrast to Y_{n3}, which is always below $G_1(x)$. Similar patterns are observed for the second maxima in Figure 3. Two additional figures for the 3rd and 4th maxima are given in the Appendix.

Let c_α be the $(1 - \alpha)$-th quantile of $G_1(x)$. We find the smallest sample size n such that $P(Y_n > c_\alpha)$ reaches ±10% of α. The results are summarized in Table 1 for all of Y_{ni}, $i = 1, 2, 3$ and Y_n. Overall Y_n needs much smaller sample sizes. Such sizes do not exist for Y_{n2} when $n \le 10^6$, so we choose not to report them.

Table 1: Smallest sample size to reach ±10% of the nominal level.

α	Y_{n1}	Y_{n2}	Y_{n3}	Y_n
10%	92	-	1230	293
5%	995	-	3639	686
1%	359965	-	38208	4126

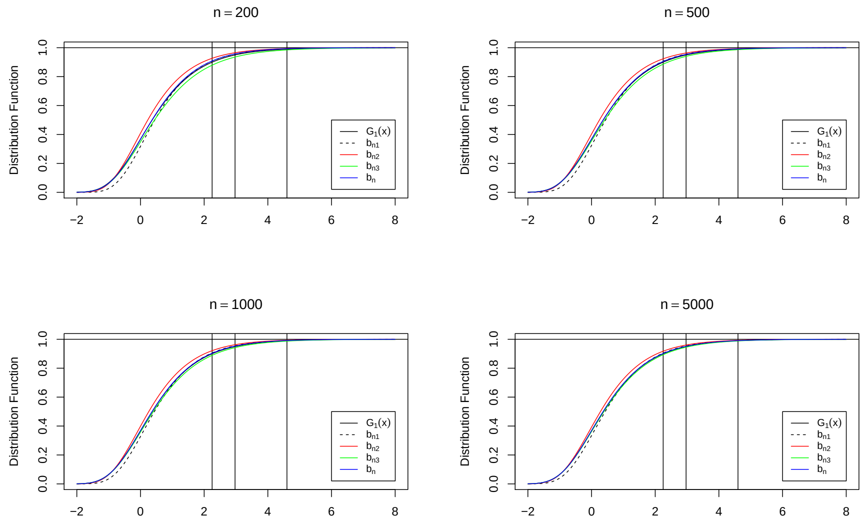

Fig. 1: Comparison of the CDFs. The black line is the true CDF of the Gumbel distribution. The dashed, red and green (labeled by b_{n1}, b_{n2} and b_{n3}) curves are the empirical CDFs, corresponding to the convergence rates $(\log\log n)^2/\log n$, $(\log n)^{-1}$ and $(\log n)^{-2}$ respectively. The blue line depicts the empirical CDF of the proposed Y_n, of convergence rate $(\log n)^{-3}$.

3.2 An example

In this section, we consider an example on testing the covariance structure. Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ is a sequence of independent and identically distributed p-dimensional random vectors. Let $R = \{\rho_{ij}\}_{1\le i,j\le p}$ be the correlation matrix of $\boldsymbol{x}_1$. Consider the hypothesis testing problem:

$$H_0 : R = I_p \quad \text{vs} \quad H_1 : R \neq I_p.$$

Jiang et al. (2004) proposed to use the maximum absolute sample correlation $L_N = \max_{1\le i<j\le p} |\hat{\rho}_{ij}|$ as the test statistic, and proved that $\frac{1}{2}(NL_N^2 - 2b_{n1})$ converges in distribution to G_1, where $n = p(p-1)$. We consider the test statistics T_{Ni}, $i = 1, 2, 3$ and T_N, which are defined in the same way as Y_{ni} and Y_n in Section 3.1, but replacing M_n^2 therein by NL_N^2. The p-values are calculated by comparing the test statistics with the Gumbel distribution G_1. By treating the sample correlations $N\hat{\rho}_{ij}$ as iid standard normal random variables, we obtain another approximation of the p-value, given by $1 - [\Phi(NL_N^2)]^n$. The test done this way is named as T_0.

For the asymptotic tests considered here, two approximations are involved: (i) Gaussian approximation of $\hat{\rho}_{ij}$, and (ii) approximation of the maximum by the Gumbel distribution. It has been understood that Gaussian approximation usually

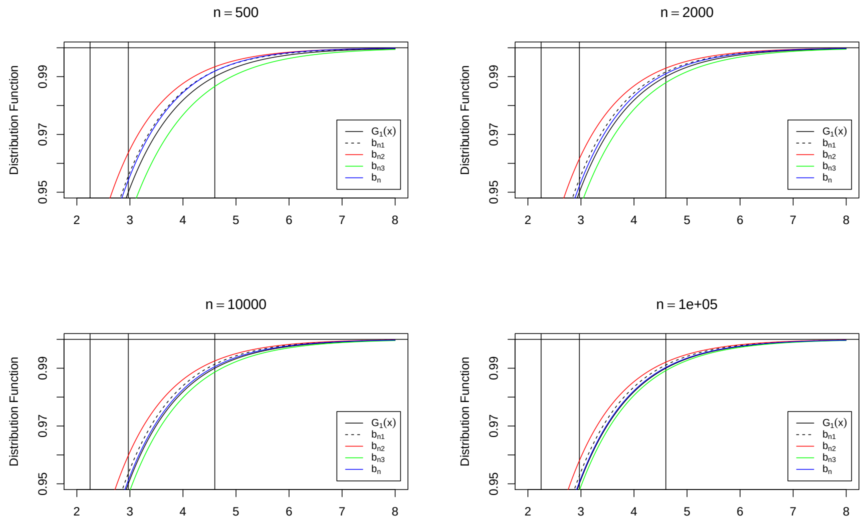

Fig. 2: Comparison of the CDFs in the upper tail. The black line is the true CDF of the Gumbel distribution. The vertical lines mark 90%, 95% and 99% quantiles of the Gumbel distribution. The dashed, red and green (labeled by b_{n1}, b_{n2} and b_{n3}) curves are the empirical CDFs, corresponding to the convergence rates $(\log\log n)^2/\log n$, $(\log n)^{-1}$ and $(\log n)^{-2}$ respectively. The blue line depicts the empirical CDF of the proposed Y_n, of convergence rate $(\log n)^{-3}$.

has a much higher convergence rate, especially in view of the recent development on the topic (see for example Chernozhukov et al., 2013, and a series of follow-up works). Therefore, the bottleneck is the convergence rate of the maximum to the Gumbel distribution. We report the empirical rejection probabilities based on 5000 repetitions in Table 2 and Table 3, where $\boldsymbol{x}_i \sim N(\mathbf{0}, I_p)$, and $\boldsymbol{x}_i$ has iid t_7 entries, respectively. We see that the empirical sizes of T_N, T_{N1} and T_0 are in general close to the nominal ones, and their performances are stable across different sample sizes and dimensions. The results are also consistent with our findings in Section 3.1. More extensive simulations, covering more sample sizes and dimensions, continue to support the observations above. These results are omitted for the sake of space.

4 Conclusion

We propose a monotone transform of the squared normal extreme, and prove that its pointwise and uniform convergence rates are both of the order $(\log n)^{-3}$, which improves the existing results in the literature. The theoretical improvements are also demonstrated and supported numerically.

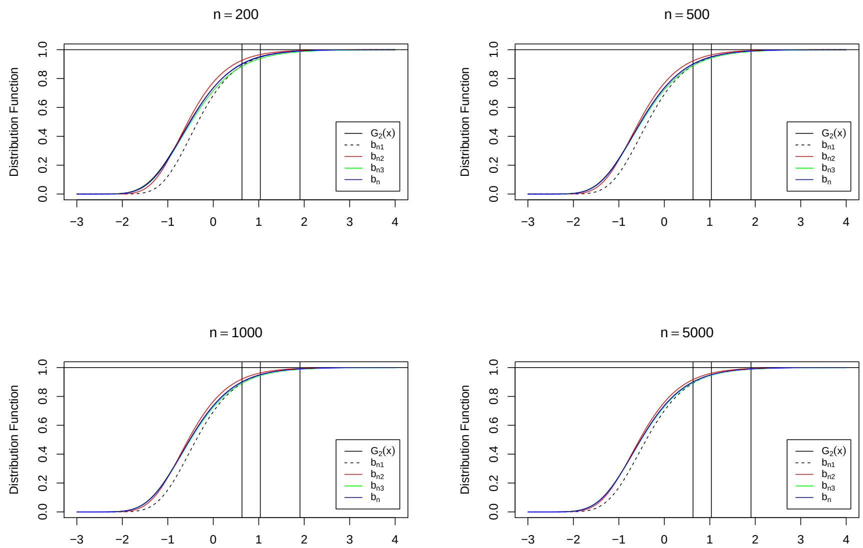

Fig. 3: Comparison of CDFs for second maxima.

Acknowledgment. The research is supported in part by National Science Foundation grant DMS-2027855 and DMS-2052949. The authors thank the Editors and three anonymous referees for their insightful comments which have led to significant improvement of the manuscript.

Table 2: The empirical rejection probabilities (%) when $\boldsymbol{x}_i$ is $\mathbb{N}(0, I_p)$.

p	Test	$n = 256$			$n = 512$			$n = 1024$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
32	T_0	8.96	4.42	0.72	9.64	4.86	0.94	10.74	5.60	1.28
	T_{N1}	8.62	4.02	0.62	8.94	4.48	0.78	10.28	5.02	1.10
	T_{N2}	7.10	3.36	0.48	7.72	3.80	0.54	8.48	4.18	0.80
	T_{N3}	10.06	5.12	1.02	10.64	5.46	1.16	11.92	6.42	1.62
	T_N	8.94	4.28	0.66	9.46	4.72	0.88	10.64	5.36	1.20
64	T_0	7.68	3.78	0.80	9.94	5.34	0.80	9.42	4.74	1.00
	T_{N1}	7.48	3.30	0.66	9.46	4.72	0.70	9.02	4.44	0.84
	T_{N2}	6.26	2.76	0.62	8.42	3.96	0.66	7.88	3.84	0.70
	T_{N3}	8.44	4.10	0.96	10.60	5.88	0.90	10.06	5.12	1.10
	T_N	7.68	3.76	0.72	9.88	5.24	0.80	9.36	4.70	0.96
128	T_0	7.60	3.32	0.62	9.30	4.86	0.80	9.86	4.82	0.98
	T_{N1}	7.34	3.12	0.60	8.90	4.52	0.72	9.56	4.58	0.82
	T_{N2}	6.26	2.72	0.60	7.86	3.82	0.66	8.20	4.00	0.68
	T_{N3}	8.16	3.82	0.66	9.78	5.14	0.90	10.14	5.30	1.14
	T_N	7.60	3.32	0.62	9.30	4.78	0.78	9.86	4.76	0.92
256	T_0	6.44	2.94	0.34	8.64	3.96	0.62	8.54	4.22	0.74
	T_{N1}	6.08	2.70	0.28	8.46	3.78	0.58	8.38	3.98	0.64
	T_{N2}	5.40	2.38	0.24	7.42	3.28	0.52	7.48	3.64	0.42
	T_{N3}	6.80	3.10	0.42	8.92	4.26	0.72	9.16	4.42	0.80
	T_N	6.44	2.92	0.34	8.66	3.94	0.60	8.54	4.20	0.70

Table 3: The empirical rejection probabilities (%) when $\boldsymbol{x}_i$ has iid t_7 entries.

p	Test	$n = 256$			$n = 512$			$n = 1024$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
32	T_0	9.84	4.82	1.02	9.18	4.70	1.24	10.22	5.28	1.12
	T_{N1}	9.28	4.36	0.76	8.82	4.34	1.06	9.58	4.66	0.92
	T_{N2}	7.84	3.74	0.66	7.28	3.72	0.92	8.26	3.90	0.68
	T_{N3}	10.74	5.70	1.36	10.04	5.32	1.40	11.38	6.12	1.20
	T_N	9.70	4.66	0.84	9.14	4.44	1.16	10.02	5.02	0.98
64	T_0	9.28	4.28	0.94	10.18	5.02	0.82	9.02	4.78	1.00
	T_{N1}	8.96	3.88	0.70	9.80	4.50	0.68	8.46	4.54	0.78
	T_{N2}	7.70	3.44	0.52	8.42	3.94	0.56	7.44	3.82	0.70
	T_{N3}	9.72	4.74	1.04	10.76	5.44	0.98	9.82	5.28	1.14
	T_N	9.24	4.22	0.84	10.18	4.94	0.78	9.00	4.74	0.88
128	T_0	9.14	4.64	0.82	9.74	4.90	1.30	9.96	4.76	0.94
	T_{N1}	8.90	4.32	0.76	9.32	4.60	1.14	9.58	4.44	0.84
	T_{N2}	7.82	3.90	0.56	8.10	3.76	0.84	8.26	3.84	0.70
	T_{N3}	9.64	4.84	0.90	10.20	5.16	1.50	10.32	5.10	1.08
	T_N	9.14	4.58	0.80	9.74	4.80	1.20	9.96	4.66	0.90
256	T_0	9.08	4.32	0.94	10.20	4.98	0.98	9.98	5.36	1.30
	T_{N1}	8.80	4.02	0.88	9.96	4.74	0.84	9.68	5.02	1.18
	T_{N2}	7.92	3.50	0.86	8.82	4.18	0.80	8.84	4.52	1.10
	T_{N3}	9.36	4.72	1.04	10.58	5.30	1.04	10.48	5.62	1.38
	T_N	9.08	4.30	0.94	10.20	4.98	0.94	9.98	5.30	1.20

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349–372.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Cohen, J. P. (1982a). Convergence rates for the ultimate and pentultimate approximations in extreme-value theory. *Advances in Applied Probability*, 14(4):833–854.

Cohen, J. P. (1982b). The penultimate form of approximation to normal extremes. *Adv. in Appl. Probab.*, 14(2):324–339.

Daniels, H. (1982). A transformation for normal extremes. *Journal of Applied Probability*, 19(A):201–206.

de Haan, L. and Resnick, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Probab.*, 24(1):97–124.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press.

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.*, 16(2):433–439.

Hall, P. (1980). Estimating probabilities for normal extremes. *Adv. in Appl. Probab.*, 12(2):491–500.

Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields*, 89(4):447–455.

Hall, W. J. and Wellner, J. A. (1979). The rate of convergence in law of the maximum of an exponential sample. *Statist. Neerlandica*, 33(3):151–154.

Jiang, T. et al. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865–880.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin.

Peng, Z., Saralees, N., and Lin, F. (2010). Convergence rate of extremes for the general error distribution. *J. Appl. Probab.*, 47(3):668–679.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Rootzén, H. (1983). The rate of convergence of extremes of stationary normal sequences. *Adv. in Appl. Probab.*, 15(1):54–80.

Smith, R. L. (1982). Uniform rates of convergence in extreme-value theory. *Adv. in Appl. Probab.*, 14(3):600–622.

5 Appendix

5.1 Expansion of b_n

Recall b_n is the solution of the equation $1 - \Phi(\sqrt{2b_n}) = 1/n$. We use the following approximation to the normal density:

$$1 - \Phi(z) = \left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - \frac{15}{z^7}\right)\phi(z).$$

Then $\sqrt{2b_n}$ is the solution of the following equation:

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \frac{15}{x^7}\right) = 1/n. \tag{21}$$

Our goal is to use three consecutive applications of the Newton-Rhapson approximation method to obtain the solution of (21) and then calculate b_n accordingly. Let

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \frac{15}{x^7}\right).$$

then the derivative of $f(x)$ is:

$$f'(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left(-1 + \frac{105}{x^8}\right).$$

We start from

$$x_0 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}},$$

where

$$\Delta = \log\log n + \log 4\pi.$$

By Newton-Rhapson approximation method,

$$f(x_0) + f'(x_0)(x_1 - x_0) = 1/n.$$

Then we can obtain:

$$x_1 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}} - \frac{\Delta^2 - 4\Delta + 8}{8\left(2\log n\right)^{3/2}}.$$

Repeat this procedure for two more times, we have

$$x_2 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}} - \frac{\Delta^2 - 4\Delta + 8}{8\left(2\log n\right)^{3/2}} - \frac{\Delta^3 - 8\Delta^2 + 32\Delta - 56}{16\left(2\log n\right)^{5/2}}.$$

$$x_3 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}} - \frac{\Delta^2 - 4\Delta + 8}{8\,(2\log n)^{3/2}} - \frac{\Delta^3 - 8\Delta^2 + 32\Delta - 56}{16\,(2\log n)^{5/2}}$$
$$- \frac{15\Delta^4 - 184\Delta^3 + 1152\Delta^2 - 4128\Delta + 7040}{384\,(2\log n)^{7/2}}.$$

Then by $b_n = x_3^2/2$, it can be easily calculated:

$$b_n = \log n - \frac{\Delta}{2} + \frac{\Delta - 2}{4\log n} + \frac{\Delta^2 - 6\Delta + 14}{16(\log n)^2} + O\left(\frac{(\log\log n)^3}{(2\log n)^3}\right).$$

5.2 Additional Figures

In this section we provide comparisons of the CDFs of the third and fourth maxima.

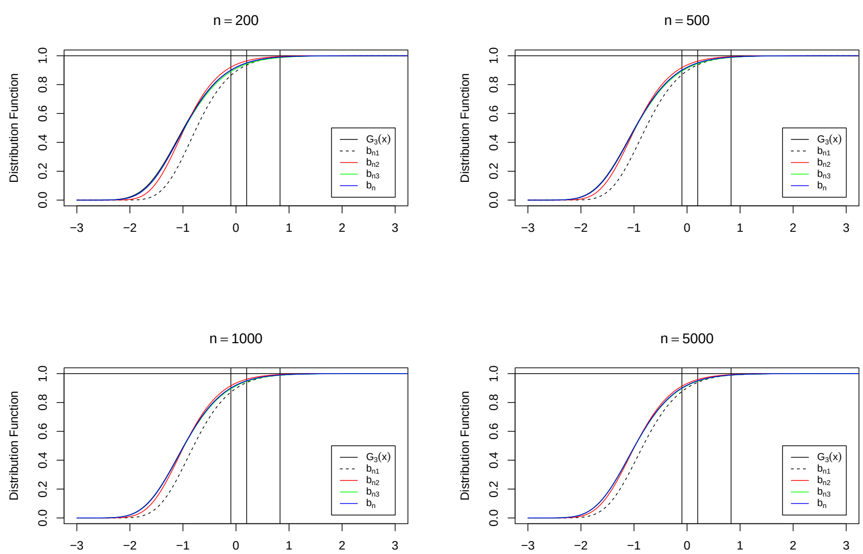

Fig. 4: Comparison of the CDFs for third maxima.

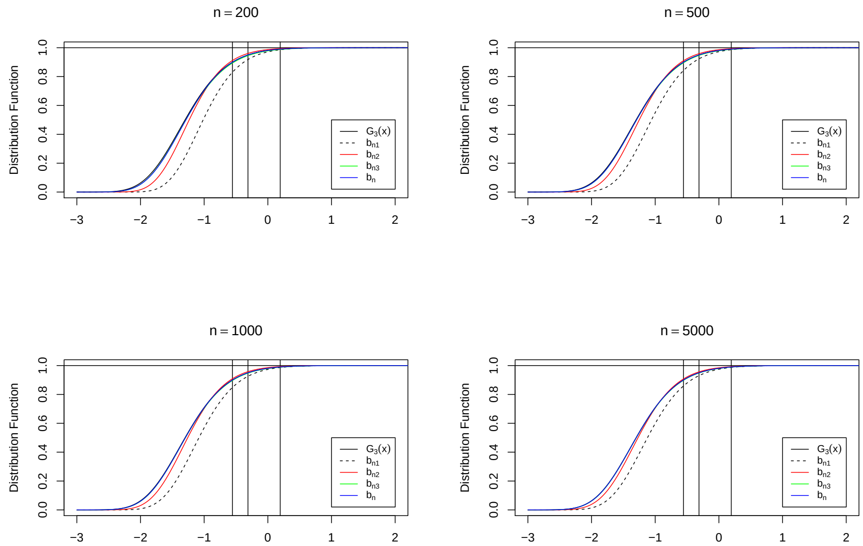

Fig. 5: Comparison of the CDFs for fourth maxima.