Electronic Journal of Statistics
Vol. 8 (2014) 3126–3140
ISSN: 1935-7524
DOI: 10.1214/14-EJS974

Bounding the maximum of dependent random variables

J. A. Hartigan

Yale University
e-mail: john.hartigan@yale.edu

Abstract: Let M_n be the maximum of n unit Gaussian variables $X_1, \dots,$ X_n with correlation matrix having minimum eigenvalue λ_n. Then

$$M_n \geq \lambda_n \sqrt{2 \log n} + o_p(1).$$

As an application, the log likelihood ratio statistic testing for the presence of two components in a 1-dimensional exponential family mixture, with one component known, is shown to be at least $\frac{1}{2} \log \log n (1 + o_p(n))$ under the null hypothesis that there is only one component.

MSC 2010 subject classifications: 60E15.
Keywords and phrases: Maxima of gaussian processes, likelihood ratio test, exponential family mixtures.

Received May 2014.

1. Introduction

The asymptotic behaviour of the maximum M_n of n independent and identically distributed random variables, with continuous distribution function F, is well known following Gumbel [8], namely that

$$-\log n - \log (1 - F(M_n)) \to G \text{ in distribution,}$$

where the Gumbel variable $G = -\log(-\log U)$ for U uniform.

When the variables are not independent, but identically distributed, upper bounds for the tail probabilities of M_n are available through

$$\Pr\{M_n \geq A\} \leq n \left(1 - F(A)\right),$$

but lower bounds are rare.

Berman [2] shows that for stationary Gaussian processes with correlations satisfying

$$\lim_{|i-j| \to \infty} \rho(X_i, X_j) \log |i - j| = 0,$$

the maximum behaves in distribution asymptotically like the maximum of i.i.d variables. Under the slightly weaker condition $\lim_{|i-j| \to \infty} \rho(X_i, X_j) = 0$,

$$\lim_{n \to \infty} M_n / (2 \mathrm{var}(X_1) \log n)^{\frac{1}{2}} = 1.$$

Darling and Erdos [5] state: For $X_1, \dots, X_n$ independent and identically distributed with expectation zero, variance 1, finite third absolute third moment, and standardized partial sums $Z_i = \sum_{1 \le j \le i} X_j/\sqrt{i}$,

$$\lim_{n\to\infty} \Pr\{a(n) \max_{1\le i\le n} Z_i < t + b(n)\} = \exp(-e^{-t}).$$

$$\text{where } a(n) = (2\log\log n)^{1/2},$$

$$b(n) = 2\log\log n + \frac{1}{2}\log\log\log n - \frac{1}{2}\log(4\pi).$$

Thus $\max Z_i$ behaves asymptotically like the maximum of $\log n$ independent unit normals. Essentially, only the sums Z_i with $i = o(n)$ determine the maximum asymptotically, and there are about $\log n$ nearly independent sums among those Z_i. Yao and Davis [13] apply this theorem to find the asymptotic distribution for the likelihood ratio statistic testing for a change in means of a sequence of n independent normal variables with variance 1. In fact, the test statistic differs negligibly from half the squared maximum of standardized sums. Gombay and Horvath [6] show that the Darling-Erdos asymptotics hold for general one parameter change point problems. Kim and Siegmund [10] study the log likelihood ratio with the change point constrained to lie in the interval $[\alpha n, 1 - \alpha n]$ for $0 < \alpha < \frac{1}{2}$, showing that the test statistic, under the null hypothesis, is distributed asymptotically as half the squared maximum on an internal segment of an Ornstein-Uhlenbeck process.

Our principal result, Theorem 3.4, provides lower bounds for the tail probabilities of the maximum of dependent Gaussian variables. We use the theorem to show the log likelihood ratio statistic testing for the presence of two components in a 1-dimensional exponential family mixture, with one component known, is at least $\frac{1}{2}\log\log n(1 - o_p(n))$ under the null hypothesis of there being only one component.

2. General bounds

We will use EX to denote the expectation of the random variable X, and $\{S\}$ to denote the function that is 1 when S is true, and 0 when S is false.

Theorem 2.1. *Let M_n denote the maximum of n random variables $X_1, \dots, X_n$ each with continuous distribution function F. Then, for each n, there exists an exponential variable W_n with*

$$-\log n - \log(1 - F(M_n)) \le W_n.$$

Proof. Let F_n denote the distribution function of M_n. Note that $U_n = F_n(M_n)$ is uniformly distributed, and define $W_n = \log(1 - U_n)$, which is exponentially distributed. Then

$$\begin{aligned} \{M_n > A\} &\le \sum_{i=1}^{n} \{X_i > A\}, \\ 1 - F_n(A) &\le n(1 - F(A)), \\ -\log n - \log(1 - F(M_n)) &\le W_n, \end{aligned}$$

since $1 - F_n(M_n) \sim \exp(-W_n)$. □

Theorem 2.2. *Let M_n denote the maximum of n independent random variables $X_i, 1 \leq i \leq n$, each with continuous distribution function F. Then the function*

$$G(M_n) = -\log(-n \log F(M_n)),$$

has a Gumbel distribution and,

$$G \leq -\log\left(n(1 - F(M_n))\right) \leq G + \exp(-G)/n.$$

Proof. Since $F(M_n) = F(\max_{1 \leq i \leq n} X_i) = \max_{1 \leq i \leq n} F(X_i)$ is the maximum of n independent uniforms, $F(M_n) \sim U^{\frac{1}{n}}$, so $G = -\log(-n \log F(M_n))$ is Gumbel. Then

$$\begin{aligned}
1 - F(M_n) &= 1 - e^{-\frac{1}{n}\exp(-G)}, \\
\frac{1}{n}\exp(-G)/(1 + \frac{1}{n}\exp(-G)) &\leq 1 - e^{-\frac{1}{n}\exp(-G)} \leq \frac{1}{n}\exp(-G), \\
G \leq -\log\left(n(1 - F(M_n))\right) &\leq G + \log(1 + \frac{1}{n}\exp(-G)), \\
G \leq -\log\left(n(1 - F(M_n))\right) &\leq G + \frac{1}{n}\exp(-G).
\end{aligned}$$

□

It follows that the limiting distribution of $-\log n - \log(1 - F(M_n))$ is the Gumbel distribution. Note that E and G are very close in their tail distributions, so there is not much difference between the upper bounds in the independent and dependent cases.

3. Gaussian bounds

In the Gaussian case, we invert the standard tail bounds for $1-\Phi(x)$ for large x so that we can accurately determine the asymptotic distribution of the maximum.

Theorem 3.1. *Let $V = -2\log(1 - \Phi(x)) - \log(2\pi)$.*

$$\begin{aligned}
&\textit{For } x \geq 2, \quad V - \log V \leq x^2, \\
&\textit{For } x \geq 1, \quad x^2 \leq V - \log V + \log V/V.
\end{aligned}$$

Proof. The standard bounds from Abramowitz and Stegun [1, p. 932]: For $x \geq 1$, with $y = x^2$,

$$\begin{aligned}
\phi(x)\left(\frac{1}{x} - \frac{1}{x^3}\right) \leq \quad & 1 - \Phi(x) \quad \leq \phi(x)\left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5}\right), \\
y + \log y - 2\log\left(1 - \frac{1}{y} + \frac{3}{y^2}\right) \leq \quad & V \quad \leq y + \log y - 2\log\left(1 - \frac{1}{y}\right).
\end{aligned} \tag{3.1}$$

We demonstrate the specified bounds by explicit calculation for moderate x, and by using the standard bounds for large x. For the lower bound, let $V - y - \log y = 2\log\frac{dV}{dy} = \Delta$ where, from (3.1), $\Delta \leq -2\log(1 - \frac{1}{y})$.

$$\begin{aligned} y - V + \log V &= \log(V/y) - \Delta \\ &\geq \log(1 + \log y/y) + 2\log\left(1 - \frac{1}{y}\right) \\ &\geq 0 \text{ for } y \geq 11. \end{aligned}$$

The lower bound is thus established for $y \geq 11$. The lower bound is exhibited in explicit calculation for $4 < y < 11$, so that the lower bound holds for $y \geq 4$ which is $x \geq 2$.

For the upper bound, using $V = y + \log y + \Delta$ and $\Delta \geq \frac{1}{y} - \frac{3}{y^2}$,

$$\begin{aligned} y - V + \log V - \log V/V &= \log(V/y) - \log V/V - \Delta \\ &= \log(1 + (\log y + \Delta)/y) - \log V/V - \Delta \\ &\leq \log y/y - \log V/V - (1 - 1/y)\Delta \\ &\leq (V - y)\log y/y^2 - (1 - 1/y)\Delta \\ &\leq (\log y + \Delta)\log y/y^2 - (1 - 1/y)\Delta \\ &\leq (\log y/y)^2 - (1 - 1/y - \log y/y^2)\left(\frac{1}{y} - \frac{3}{y^2}\right) \\ &\leq 0 \text{ for } y > 5 \end{aligned}$$

The upper bound is thus established for $y \geq 5$. The upper bound is exhibited in explicit calculation for $1 < y < 5$, so that the upper bound holds for $x \geq 1$. □

Theorem 3.2. *Let M_n be the maximum of n independent unit Gaussians. Let $N = \log(n^2/2\pi)$. For each n, there exists a Gumbel variable G, a monotone function of M_n, such that, for $M_n \geq 2$,*

$$\begin{aligned} (N + 2G) - \log(2G + N) &\leq M_n^2 \leq V - \log V + \log V/V, \\ \textit{where } V &= N + 2G + 2\exp(-G)/n. \end{aligned}$$

Proof. Substitute the Gaussian probability bounds from Theorem 3.1 into the probability bounds for $\log(1 - \Phi(M_n))$ given in Theorem 2.2. □

We see from Theorem 3.2, that as $n \to \infty$, $M_n^2 - N - \log N \to 2G$ in distribution. Note that the bounds apply only to tail probabilities $\Pr\{M_n \geq A\}$ for $A > 2$.

Theorem 3.3. *Let M_n denote the maximum of n unit Gaussian random variables. Let $N = \log(n^2/2\pi)$. Then there is an exponential variable W with*

$$M_n^2 \leq \max\left(1, N + 2W - \log(N + 2W) + \log(N + 2W)/(N + 2W)\right).$$

Proof. From Theorem 3.1, with $V = -\log(2\pi) - 2\log(1 - \Phi(M_n))$, we have, for $M_n \geq 1$, $M_n^2 \leq V - \log V + \log V/V$. From Theorem 2.1, there exists a function $W(M_n)$ distributed exponentially, with $V \leq N + 2W$. Thus,

$$M_n^2 \leq \max\left(1, N + 2W - \log(N + 2W) + \log(N + 2W)/(N + 2W)\right).$$

□

Asymptotically, $M_n^2 - N - \log N \leq 2W$ as $n \to \infty$.

Theorem 3.4. *Let M_n be the maximum of n zero-mean Gaussian variables $X_1, \ldots, X_n$. Define*

$$\begin{aligned} E_i &= E(X_i|X_1, \ldots, X_{i-1}), \\ R_i &= X_i - E_i, \\ \sigma_i^2 &= ER_i^2, \sigma^2 = \min_{1 \le i \le n} \sigma_i^2, \\ \tau_i^2 &= EE_i^2 = EX_i^2 - ER_i^2, \tau^2 = \max_{1 \le i \le n} \tau_i^2, \\ N &= \log(n^2/2\pi), \\ L_\alpha &= -2\log(-\log\alpha). \end{aligned}$$

Then, for $N + L_\alpha \ge 6$,

$$\Pr\{M_n \ge \sigma\left(N + L_\alpha - \log(N + L_\alpha)^{\frac{1}{2}} + \tau\Phi^{-1}(\alpha)\right)\} \ge 1 - 2\alpha.$$

The term $\sigma\sqrt{2\log n}$ dominates the lower bound for n large, so for each $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\{M_n \ge \sigma\sqrt{2\log n}(1-\epsilon)\} = 1.$$

Proof. We first show that for each real A and non-positive B,

$$\Pr\{M_n \ge A + B\} \ge \Pr\{\max_i R_i \ge A\} \min_i \Pr\{E_i \ge B\}.$$

Construct n disjoint events

$$H_i = \{R_i \ge A\} \prod_{j>i} \{R_j < A\}.$$

Note that $\sum_i H_i = \{\max_i R_i \ge A\}$. Since the terms $\{R_i \ge A\}$ are independent of all variables $X_j, j < i$, the individual terms in each H_i are l independent, and H_i is independent of $\{E_i \ge B\}$. Also, $\{M_n \ge A+B\} \ge \sum_i H_i\{E_i \ge B\}$, since at most one of the events $H_i\{E_i \ge B\}$ occurs, and if any occurs $M_n \ge A+B$. Then

$$\begin{aligned} \Pr\{M_n \ge A + B\} &\ge \sum_i EH_i \Pr\{E_i \ge B\} \\ &\ge \sum_i EH_i \min_j \Pr\{E_j \ge B\} \\ &\ge \Pr\{\max_i R_i \ge A\} \min_i \Pr\{E_i \ge B\}. \end{aligned}$$

For $A > 0$, $\{\max_i R_i \ge A\} \ge \{\sigma \max_i(R_i/\sigma_i) \ge A\}$.

Let $R_{(n)} = \sigma \max_i(R_i/\sigma_i)$. From Theorem 3.2, for $A/\sigma \ge 2$,

$$\Pr\{R_{(n)} \ge A\} \ge \Pr\{(N + 2G - \log(N + 2G))^{\frac{1}{2}} \ge A/\sigma\}.$$

Since $\Pr\{2G \ge L_\alpha\} = 1 - \alpha$, for $N + L_\alpha \ge 6$ (which implies $A/\sigma \ge 2$),

$$\Pr\{R_{(n)} \ge \sigma(N + L_\alpha - \log(N + L_\alpha))^{\frac{1}{2}}\} \ge 1 - \alpha.$$

Also $\min_i \Pr\{E_i \ge B\} = 1 - \Phi(B/\tau)$, so

$$\min_i \Pr\{E_i \ge \tau\Phi^{-1}(\alpha)\} = 1 - \alpha.$$

Combining the two bounds gives, for $N + L_\alpha \geq 6$,

$$\Pr\{M_n \geq \sigma \left(N + L_\alpha - \log(N + L_\alpha)^{\frac{1}{2}} + \tau\Phi^{-1}(\alpha) \right)\} \geq 1 - 2\alpha.$$

as asserted.

For each $\epsilon > 0$, $\sqrt{2\log n}(1-\epsilon) < N + L_\alpha - \log(N + L_\alpha)^{\frac{1}{2}} + \tau\Phi^{-1}(\alpha)$ for n large, so that, for each $\alpha > 0$,

$$\Pr\{M_n \geq \sigma\sqrt{2\log n}(1-\epsilon)\} \geq 1 - 2\alpha.$$

as required. □

4. Diagonally dominated correlation matrices

If the random variables $X_1, \ldots, X_n$ have correlation matrix C, then

$$1 - \tau^2 = \sigma^2 \quad \geq \quad \min_i \frac{1}{(C^{-1})_{ii}} \geq \min_x \frac{x'Cx}{x'x} = \lambda.$$

If a sequence of corrrelation matrices C for n gaussians has minimum eigenvalue exceeding λ, the lower bound of Theorem 3.4 holds with λ substituted for σ^2 and $1 - \lambda$ is substituted for τ^2:

$$\lim_{n\to\infty} \Pr\{M_n \geq \lambda\sqrt{2\log n}(1-\epsilon)\} = 1. \tag{4.1}$$

The principal contributor to the bound is the minimum eigenvalue. A correlation matrix C is *strictly diagonally dominated* if

$$1 > \max_i \left(\sum_{j|j\neq i} |C_{ij}| \right).$$

For such a matrix, Varah [12] shows that the minimum eigenvalue λ satisfies

$$\lambda \geq 1 - \max_i \left(\sum_{j|j\neq i} |C_{ij}| \right).$$

In practice then, to find a lower bound for the maximum, we look for a subset of the variables such that each is unpredictable from the rest. One way to find that selection is to make sure the absolute correlations in each row have a small sum.

5. Asymptotic lower bound for the log likelihood ratio in exponential family mixtures

In Hartigan [9] it was shown that the maximum log likelihood ratio for testing the mixture $(1-p)\phi(x) + p\phi(x-\theta)$ against the null hypotheses $\phi(x)$ diverged in probability to ∞ under the null hypothesis, and it was conjectured that the maximum would be $O_p(\log\log n)$. Bickel and Chernoff [3] used the Komlos, Major and Tusnady [11] "Hungarian construction" to show that the maximum log likelihood is indeed $\frac{1}{2}\log\log n(1+o_p(1))$. Chernoff and Lander [4], studying mixtures of binomials with k trials, show that the maximum log likelihood ratio is asymptotically finite, and conjecture that it would $O(\sqrt{\log k})$ for k large. See also Ghosh and Sen [7].

Theorem 5.1. *Let $f(x,\theta)=\exp(x\theta-c(\theta)),\theta\in T$, be a one-dimensional exponential family of densities with respect to a base measure μ, with the real interval T consisting of those parameter values θ for which $\int\exp(x\theta)d\mu(x)$ is finite. The Jeffreys density is $j(\theta)=\sqrt{c''(\theta)}$.*

Suppose that, for some $A>0$,

1. *The upper endpoint of T is ∞.*
2. *$\int_A^\infty j(\theta)d\theta=\infty$.*
3. *$\sup_{A<\theta<\infty}|j'(\theta)|/j(\theta)<\infty$.*

Then, for each $\theta_0\in T$, for n observations $X_1,\dots,X_n$ from $f(x,\theta_0)$, the log likelihood ratio test statistic for the mixture

$$((1-p)f(x,\theta_0)+pf(x,\theta),0\le p\le 1,\theta\in T)\ \textit{versus}\ f(x,\theta_0),$$

$$LL_n=\sup_{p,\theta}\sum_{1\le i\le n}\log\left(1+p\left(\frac{f(X_i,\theta)}{f(X_i,\theta_0)}-1\right)\right)$$

satisfies $LL_n\ge\frac{1}{2}\log\log n(1+o_p(n))$ as $n\to\infty$.

Proof. It will be convenient, without loss of generality, to assume

$$\theta_0=0,A=0,c(0)=1,c''(0)=1,\sup_{0<\theta<\infty}|j'(\theta)|/j(\theta)=1.$$

Define $J(\theta)=\int_0^\theta j(v)dv$.

In order to facilitate the uniform approximation of a likelihood process by a gaussian process, let U be uniformly distributed, and let X be a function on $[0,1]$, such that $X(U)$ has the density $f(x,\theta_0)$ with respect to μ. A random sample $U_1,\dots U_2$ from the uniform generates the sample $X(U_1),\dots,X(U_n)$ from the member of the exponential family at $\theta=0$.

Define, for $0\le\theta<\infty$,

$$\begin{aligned}
Y(U,\theta) &= \exp\left(X(U)\theta-c(\theta)\right)-1,\\
L_n(p,\theta) &= \sum_{1\le i\le n}\log\left(1+pY(U_i,\theta)\right),\\
L_n(\theta) &= \sup_{0\le p\le 1}L_n(p,\theta),\\
LL_n &= \sup_{0\le\theta<\infty}L_n(\theta),\\
l_n(\theta) &= \sum_i(Y(U_i,\theta)/\sqrt{\sum_i Y(U_i,\theta)^2},
\end{aligned}$$

For T_n a subset of $\{0 \le \theta < \infty\}$, define uniform orders of magnitude by

$$A(n,\theta) = o_p^{T_n}(1) \quad : \quad \max_{\theta \in T_n} |A(n,\theta)| = o_p(1) \text{ as } n \to \infty,$$
$$A(n,\theta) = O_p^{T_n}(1) \quad : \quad \max_{\theta \in T_n} |A(n,\theta)| = O_p(1) \text{ as } n \to \infty.$$

Outline of the proof:

Step 1: Identify a subset T_n of size $O(\sqrt{\log n}/(\log\log n)^2)$ such that the correlations between the neighboring values are less than $1/\log n$.

Step 2: For some gaussian process $g_n(\theta)$ with unit variances and the same correlations as $Y(\theta)$, $g_n(\theta) = l_n(\theta) + o_p^{T_n}(1)$.

Step 3: The maximum log likelihood ratio at θ, $L_n(\theta) \ge \frac{1}{2} l_n^+(\theta)^2(1 + o_p^{T_n}(1))$.

Step 4: Combine approximations to show $LL_n \ge \frac{1}{2} \log\log n(1 + o_p(1))$.

Step 1: Identify a subset T_n of size $O(\sqrt{\log n}/(\log\log n)^2)$ such that the correlations between the neighboring values are less than $1/\log n$.

The covariance of $Y(u)$ and $Y(v)$ is $\mathrm{cov}(u,v) = \exp(c(u+v) - c(u) - c(v)) - 1$. It may be verified, by integrating by parts, that

$$c(u+v) - c(u) - c(v) = \int \min(w, u, u+v-w)^+ c''(w)dw.$$

Since c is strictly convex and $\min(w, u, u+v-w)^+$ is non-decreasing in u and v for all w, and is increasing for some values of w, it follows that $c(u,v)$ is increasing in u and v. Also, since $c'(u+w) - c'(w) = \int_0^u c''(w+x)dx$ is non-decreasing in u, for $0 \le u \le v \le w$,

$$\frac{\partial}{\partial w}(\log \mathrm{cov}(u,w) - \log \mathrm{cov}(v,w)) = \frac{c'(u+w) - c'(w)}{1 + 1/\mathrm{cov}(u,w)} - \frac{c'(v+w) - c'(w)}{1 + 1/\mathrm{cov}(v,w)} \le 0.$$

Thus $\mathrm{cov}(u,w)/\mathrm{cov}(v,w)$ is decreasing in w, so the covariances are

$$submultiplicative : \text{for } 0 < u \le v \le w, \mathrm{cov}(u,w)\mathrm{cov}(v,v) \le \mathrm{cov}(u,v)\mathrm{cov}(v,w).$$

For $u \le v$, $\mathrm{cov}(u,u) \le \mathrm{cov}(u,v) \le cov(v,v)$, so the correlation $\rho(u,v)$ satisfies

$$\begin{aligned}\log(\rho(u,v)) + \frac{1}{2}\log(1 - 1/\mathrm{cov}(u,u)) &\le c(u+v) - \frac{1}{2}c(2u) - \frac{1}{2}c(2v) \\ &= -\frac{1}{2}\int (|v-u| - |w-(v+u)|)^+ c''(w)dw \\ &\le -\frac{1}{8}(J(2v) - J(2u))^2.\end{aligned}$$

The last assertion follows from condition (3): $|j'(\theta)| \le j(\theta)$. The smallest value of $I_2 = \int(|v-u| - |w-(v+u)|)^+ c''(w)dw$ for a given value of $I_1 = \int_{2u}^{2v} j(w)dw$ occurs when c'' is concentrated as much as possible near the endpoints of the interval $(2u, 2v)$, which is when $|j'(\theta)| = j(\theta)$ in the interval.

In this case,

$$\frac{I_2}{I_1^2} = \frac{e^{2(v-u)} - 2(v-u) - 1}{4(e^{(v-u)} - 1)^2},$$

which has a minimum value of $\frac{1}{4}$ at $v - u = \infty$. Thus for any choice of $j(\theta)$ satisfying condition (3),

$$\int (|v-u| - |w-(v+u)|)^+ c''(w) dw \geq \frac{1}{4}(J(2v) - J(2u))^2.$$

Choose θ_n such that $J(2\theta_n) = \sqrt{2\log n}/\log\log n$ to ensure that $l_n(\theta)$ is asymptotically gaussian for $\theta < \theta_n$ (to be shown later).

Choose a an initial sequence of parameter values $T_n^* = \{\theta^1, \theta^2, \ldots, \theta^{k_n}\}$ in the interval $0 < \theta < \theta_n$ such that the correlations are less than $\frac{16}{17}$ between all neighbouring pairs $Y(\theta^i), Y(\theta^{i+1})$. First choose θ^1 large enough so that $\mathrm{cov}(\theta^1, \theta^1) > \frac{1}{8}$, (possible since $\mathrm{cov}(u,u) \uparrow \infty$ as $u \to \infty$). Next, for $j = 1, 2, 3, , k_n$, choose θ^{j+1} such that $J(2\theta^{j+1}) - J(2\theta^j) = 1$. Then

$$\begin{aligned} \log \rho(\theta^j, \theta^{j+1}) &\leq \frac{1}{2}\log(1 - 1/\mathrm{cov}(\theta^j, \theta^{j+1})) - \frac{1}{8} \leq -\frac{1}{16} \\ \rho(\theta^j, \theta^{j+1}) &\leq \frac{16}{17}. \end{aligned}$$

We can find k_n^* such values, with k_n^* the integer part of $\sqrt{2\log n}/\log\log n - J(2\theta_1)$. Since the covariances of the $Y(\theta)$ are submultiplicative, so also are the correlations, and so $\rho(\theta^i, \theta^j) \leq (\frac{16}{17})^{-|i-j|}$.

Choose m_n to be the integer part of $1 + 17\log\log n$, so that $\rho(\theta^i, \theta^{i+m_n}) \leq 1/\log n$. We now choose the final set of parameter values $T_n = \{\theta^1, \theta^{1+m_n}, \theta^{1+2m_n}, \ldots\}$ of length $k_n \geq \sqrt{\log n}/(\log\log n)^2$, in which the neighbouring correlations are less than $1/\log n$, the sum of off-diagonal terms in each row of the correlation matrix is bounded by $2/(\log n - 1)$, and so the smallest eigenvalue of the diagonally dominated correlation matrix exceeds $(1 - 3/(\log n - 1))$ from Varah [12].

Step 2: For some gaussian process $g_n(\theta)$ with unit variances and the same correlations as $Y(\theta)$, $g_n(\theta) = l_n(\theta) + o_p^{T_n}(1)$.

It is straightforward to show a uniformly close gaussian approximation to $l_n(\theta)$ for any finite number of θ values, but far more difficult to show a uniform approximation when the number of θ values increases with n.

Let $F_n(u)$ be the empirical distribution of the sample $U_1, \ldots, U_n$ from a uniform distribution. Let B be a Brownian bridge on $[0, 1]$, a gaussian process having zero means and covariances $E(B_s B_t) = \min(s,t) - st, 0 \leq s \leq t \leq 1$, the same as the process $\sqrt{n}(F_n(u) - u)$. From Komlos, Major and Tusnady [11], there exists such a brownian bridge B_n, for each n, such that

$$\sup_{0<u<1} |\sqrt{n}\,(F_n(u) - u) - B_n(u)| = O_p(n^{-\frac{1}{2}}\log n).$$

The variable $Z(U, \theta) = \exp\left(X(U)\theta - \frac{1}{2}c(2\theta)\right)$, a linear transform of $Y(U, \theta)$, handles more easily because $EZ^2(U, \theta) = 1$.

Define, following Bickel and Chernoff [3],

$$\begin{aligned}
S_n(\theta) &= n^{-\frac{1}{2}}\sum_i (Z(U_i,\theta) - EZ(U,\theta)) \\
&= \sqrt{n}\int_0^1 Z(u,\theta)(dF_n(u) - du), \\
G_n(\theta) &= \int_0^1 Z(u,\theta)dB_n(u), \\
S_n^\alpha(\theta) &= \sqrt{n}\int_\alpha^1 Z(u,\theta)(dF_n(u) - du), \\
G_n^\alpha(\theta) &= \int_\alpha^1 Z(u,\theta)dB_n(u).
\end{aligned}$$

The integral $G_n(\theta)$ exists with probability 1. Since $\sqrt{n}(F_n(u) - u), 0 < u < 1$, and $B_n(u), 0 < u < 1$, have the same means and covariances, $G_n(\theta)$ is a gaussian process with the same variances and covariances as $Z(U,\theta)$ and $S_n(\theta)$. We will show that $G_n(\theta)$ differs negligibly from $S_n(\theta)$ on T_n. Then we show that $g_n(\theta) = G_n(\theta)/\sqrt{\mathrm{var} G_n(\theta)}$ differs negligibly from $l_n(\theta)$ on T_n. We do this in three steps:

Step 2.1: Bounding the upper tails of the integrals.
Step 2.2: Demonstrating the two truncated integrals are close.
Step 2.3: Showing the empirical variance is close to the theoretical variance.
Step 2.4: Combine previous approximations to construct the gaussian process $g_n(\theta)$ to uniformly approximate $l_n(\theta)$ over $\theta \in T_n$.

Step 2.1: Bounding the upper tails of the integrals.

We bound the tail integrals in the region $\{u > \alpha\} = \{Z(u,\theta) > n^{\frac{1}{2}}(\log n)^{-2}\}$. Note that $S_n^\alpha(\theta)$ and $G_n^\alpha(\theta)$ have the same second moment. Then, using $c(\theta) = \int_0^\theta vc''(v)dv$,

$$\begin{aligned}
E(G_n^\alpha(\theta))^2 &= \int_\alpha^1 Z^2(u,\theta)du - \left(\int_\alpha^1 Z(u,\theta)du\right)^2 \\
&\le \int_\alpha^1 Z^2(u,\theta)du \\
&\le \int_\alpha^1 Z^{2+2\epsilon}(u,\theta)(n(\log n)^{-4})^{-\epsilon}du \text{ (Lyaponov bound)} \\
&\le \exp\left(c(2\theta+2\epsilon) - (1+\epsilon)c(2\theta) - \epsilon(\log n - 4\log\log n\right) \\
&\le \exp\left(\epsilon\left(\int_0^{2(\theta+\epsilon)} vc''(v)dv - \log n + 4\log\log n\right)\right) \\
&\le \exp\left(\epsilon\left(\frac{1}{2}J^2(2\theta+2\epsilon)\log(J(2\theta+2\epsilon)+1) - \log n + 4\log\log n\right)\right)
\end{aligned}$$

The final inequality follows from condition (2): $|j'(\theta)| < j(\theta)$. The largest value of $\int_0^\theta vc''(v)dv$ for given $J(\theta) = \int_0^\theta \sqrt{c''(v)}dv$ occurs for $c''(v) = \exp(2v)$, which increases maximally, and the inequality

$$\int_0^\theta c''(v)dv \leq \frac{1}{2}J^2(\theta)\log(J(\theta)+1)$$

then holds.

We have chosen θ_n previously, and now choose ϵ_n:

$$\begin{aligned} \theta_n &: \quad J(2\theta_n) = \sqrt{2\log n}/\log\log n, \\ \epsilon_n &: \quad J(2\theta_n + 2\epsilon_n) = J(2\theta_n) + 1. \end{aligned}$$

Using condition (2), the smallest value of ϵ occurs when $j(\theta)$ increases as fast as possible, so that

$$\begin{aligned} j(\theta) &= \exp(\theta), \\ J(2\theta_n) &= \exp(2\theta_n) - 1, \\ 1 &= J(2\theta_n + 2\epsilon) - J(2\theta_n) = \exp(2\epsilon_n) - 1, \end{aligned}$$

from which we conclude $\epsilon_n \geq (4 + 2J(2\theta_n))^{-1}$.

Now dropping $\log\log n$ terms for large n we obtain, for $0 < \theta < \theta_n$,

$$E(G_n^\alpha(\theta))^2 \leq \int_\alpha^1 Z(u,\theta)du \leq \exp\Big(-\frac{1}{2}\sqrt{\log n}\Big) \text{ for } n \text{ large enough.}$$

Let T_n contain $O(\sqrt{\log n})$ values in $(0, \theta_n)$, so

$$\begin{aligned} \Pr\{\max_j G_n^\alpha(\theta^j) > \eta\} &\leq \sum_j \Pr\{G_n^\alpha(\theta^j) > \eta\} \\ &\leq \eta^{-2}\sum_j EG_n^\alpha(\theta^j)^2 \\ &\leq \eta^{-2}\sqrt{\log n}\exp\Big(-\frac{1}{2}\sqrt{\log n}\Big) \text{ for } n \text{ large enough} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus $G_n^\alpha(\theta) = o_p^{T_n}(1)$ and, by the same argument, $S_n^\alpha(\theta) = o_p^{T_n}(1)$, since $G_n^\alpha(\theta)$ and $S_n^\alpha(\theta)$ have the same second moments.

Step 2.2: Bounding the difference between the truncated integrals.

We next bound the difference $\Delta_n(\theta) = S_n(\theta) - S_n^\alpha(\theta) - G_n(\theta) + G_n^\alpha(\theta)$ integrating by parts, and noting that $Z(u,\theta)$ increases in u:

$$\begin{aligned} \text{Setting } \delta_n(u) &= \sqrt{n}(dF_n(u) - du) - dB_n(u), \\ \Delta_n(\theta) &= \int_0^\alpha Z(u,\theta)d\Delta_n(u) \\ &= Z(u,\alpha)\delta_n(\alpha) - \int_0^\alpha \Delta_n(u)dZ(u,\theta) \end{aligned}$$

$$\begin{aligned}
|\Delta_n(\theta)| &\leq Z(\alpha,\theta)|\delta_n(\alpha)| + \sup_{0<u<1} |\delta_n(u)| \int_0^\alpha dZ(u,\theta) \\
&\leq 2Z(\alpha,\theta) \sup_{0<u<1} |\Delta_n(u)| \\
&= O_p^{T_n}\left(\sqrt{n}(\log n)^{-2}\right) O_p^{T_n}\left(n^{-\frac{1}{2}}\log n\right) \\
&= O_p^{T_n}((\log n)^{-1}).
\end{aligned}$$

Combining the tail bounds and trunk difference bounds

$$|S_n(\theta^j) - G_n(\theta^j)| = o_p^{T_n}(1).$$

Step 2.3: The first two sample moments are uniformly close to their expectations.

For the first moment, $\mathrm{var}(\frac{1}{n}\sum Z(U_i,\theta)) \leq \frac{1}{n}EZ^2(U,\theta) = 1/n$ so, using chebyshev, $\frac{1}{n}\sum Z(U_i,\theta) = EZ(U,\theta) + o_p^{T_n}(1)$.

For the second moment, from step 2.1,

$$E\int_\alpha^1 Z^2(u,\theta)dF_n = \int_\alpha^1 Z^2(u,\theta)du \leq \exp\left(-\frac{1}{2}\sqrt{\log n}\right).$$

Also

$$\int_\alpha^1 du = \Pr\{Z^2(U,\theta) > \exp(\nu_n)\} \leq EZ^2(U,\theta)\exp(-2\nu_n) = (\log n)^2/n.$$

so $E|\int_\alpha^1 (Z^2(u,\theta)-1)dF_n(u)| \leq 2\exp(-\frac{1}{2}\sqrt{\log n})$.

For the truncated integrals, integrating by parts, and using the Dvoretzky-Kiefer-Wolfowitz inequality: $E\sqrt{n}\sup_u |F_n(u)-u| \leq 2$,

$$\int_0^\alpha Z^2(u,\theta)(dF_n - du) = Z^2(\alpha,\theta)(F_n(\alpha)-\alpha) - \int_0^\alpha (F_n(u)-u)dZ^2(\alpha,\theta)$$

$$\begin{aligned}
E\left|\int_0^\alpha Z^2(u,\theta)(dF_n - du)\right| &< 4n^{-\frac{1}{2}}Z^2(\alpha,\theta) \\
&< 4\exp\left(\frac{1}{2}\log n - 2\log_2 n - \frac{1}{2}\log n\right) = 4(\log n)^{-2}.
\end{aligned}$$

Since $\sum_i (Z^2(U_i,\theta) - 1)/n = \int_0^1 Z^2(u,\theta)(dF_n - du)$, summing over the $O(\sqrt{\log n})$ terms in T_n gives $\sum_{\theta^j \in T_n} E|\sum_i (Z^2(U_i,\theta^j)^2 - 1)/n| \leq 12(\log n)^{-1}$, which implies $\frac{1}{n}\sum_i (Z^2(U_i,\theta)) = 1 + o_p^{T_n}(1)$.

Step 2.4: Gaussian process approximation of $l_n(\theta)$.

Since $Z(U,\theta)$ is a linear function of $Y(U,\theta)$,

$$l_n(\theta) = \sum_i (Y(U_i,\theta) - EY(U,\theta)) \Big/ \sqrt{\sum_i (Y(U_i,\theta) - EY(U,\theta))^2}$$

$$= \sum_i \left(Z(U_i,\theta) - EZ(U,\theta)\right) / \sqrt{\sum_i \left(Z(U_i,\theta) - EZ(U,\theta)\right)^2}.$$

We have shown that

$$\begin{aligned} S_n(\theta) &= G_n(\theta) + o_p^{T_n}(1), \\ n^{-\frac{1}{2}} S_n(\theta) &= o_p^{T_n}(1), \\ \frac{1}{n}\sum_i Z^2(U_i,\theta) &= 1 + o_p^{T_n}(1). \end{aligned}$$

Since $\mathrm{var}Z(U,\theta) = 1 - \exp(2c(\theta) - c(2\theta))$ is increasing in θ and therefore exceeds $\mathrm{var}Z(U,\theta^1)$ for $\theta \in T_n$,

$$g_n(\theta) = G_n(\theta)/\sqrt{\mathrm{var}Z(U,\theta)} = S_n(\theta)/\sqrt{\mathrm{var}Z(U,\theta)} + o_p^{T_n}(1).$$

Since $\frac{1}{n}\sum_i \left(Z(U_i,\theta) - EZ(U,\theta)\right)^2 = \mathrm{var}Z(U,\theta) + o_p^{T_n}(1)$,

$$l_n(\theta) = g_n(\theta) + o_p^{T_n}(1),$$

which demonstrates step 2.

Step 3: The maximum log likelihood at θ, $L_n(\theta) \geq \frac{1}{2} l_n^+(\theta)^2(1+o_p^{T_n}(1))$.

We consider first the maximization of $L_n(p,\theta)$ over p with θ fixed. The log likelihood ratio is a concave function of p with maximum achieved at $p = 0$ when the derivative at $p = 0$ is non-positive, that is, when $\sum Y(U_i,\theta) \leq 0$.

Let $s_n^{[j]}(\theta) = \sum_i (Y^j(U_i),\theta), j = 1,2$. By a Taylor expansion of $L_n(p,\theta)$ about $p = 0$, for all p,

$$L_n(\theta) \geq p s_n^{[1]} - \frac{1}{2} p^2 s_n^{[2]}/(1-p),$$

and selecting $\hat{p} = (s_n^{[1]}(\theta))^+ / s_n^{[2]}(\theta)$,

$$\max_p L_n(p,\theta) \geq \frac{1-2\hat{p}}{2-2\hat{p}} \hat{l}_n^+(\theta)^2.$$

From step 2, noting that the maximum of $\sqrt{\log n}$ dependent unit gaussians is bounded by $\sqrt{\log n \log n}(1 + o_p(1))$, and that $\mathrm{var}Y(U,\theta)$ is increasing in θ,

$$\begin{aligned} n^{-1} s_n^2(\theta)/\mathrm{var}Y(U,\theta) &= 1 + o_p^{T_n}(1), \\ l_n(\theta) &= g_n(\theta) + o_p^{T_n}(1). \\ \hat{p} &= \left(g_n^+(\theta) + o_p^{T_n}(1)\right)(1 + o_p^{T_n}(1))/\sqrt{n \mathrm{var}Y(U,\theta)} \\ &\leq n^{-\frac{1}{2}} \left(\sqrt{\log\log n} + o_p^{T_n}(1)\right)(1 + o_p^{T_n}(1))/\sqrt{\mathrm{var}Y(U,\theta^1)} \\ &= o_p^{T_n}(1). \\ \max_p L_n(p,\theta) &\geq \frac{1}{2} l_n^+(\theta)^2 (1 + o_p^{T_n}(1)) \end{aligned}$$

Step 4: Combine approximations to show $LL_n \geq \frac{1}{2}\log\log n(1+o_p(1))$.

In step 1 we identified a set of parameter values T_n of size $k_n \geq \sqrt{\log n}/(\log\log n)^2$, in which the minimum eigenvalue of the correlation matrix exceeds $1 - 3/\log n$.

The approximating Gaussian process $g_n(\theta), \theta \in T_n$ has the same correlations as the $Y(\theta), \theta \in T_n$ and so, from equation (4.1), satisfies, for each $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{\max_{\theta\in T_n} g_n(\theta) \geq (1 - 3/\log n)\sqrt{2\log\left(\sqrt{\log n}/(\log\log n)^2\right)}(1-\epsilon)\right\} = 1,$$

which simplifies to $\max_{\theta\in T_n} g_n(\theta) = \sqrt{\log\log n}(1 + o_p(1))$.

From steps 2 and 4,

$$L_n(\theta) \geq \frac{1}{2} l_n^+(\theta)^2(1 + o_p^{T_n}(1)) = \frac{1}{2}\left(g_n^+(\theta) + o_p^{T_n}(1)\right)^2(1 + o_p^{T_n}(1)).$$

Thus, as proposed, $LL_n \geq \frac{1}{2}\log\log n(1 + o_p(1))$. □

6. Discussion

Theorem 5.1 applies for the gaussian location, the gamma scale, and the Poisson exponential families. The conditions do not apply for the binomial or the negative binomial. For the binomial having k trials, the Jeffreys integral is $\pi\sqrt{k}$, and from step 1 of the proof, this means there are order $\sqrt{k}/\log k$ parameter values for which the minimum eigenvalue of their correlation matrix exceeds $1 - 1/k$. Now letting $k \to \infty$, and letting n approach ∞ fast enough for each k so that the error in the gaussian process approximation becomes uniformly negligible,

$$LL_n(k) \geq \left(\frac{1}{2}\log k\right)(1 + o_p(1)),$$

as conjectured by Chernoff and Lander [4].

To apply these methods more widely, I would suspect that for a similar class of multiparameter exponential families of dimension d, the limiting rate for the maximum log likelihood would exceed $\frac{1}{2}d\log\log n$, on the grounds that a d-dimensional grid could be laid out, each direction containing about $\sqrt{\log n}$ points, and with all $(\sqrt{\log n})^d$ parameter settings having nearly independent likelihood sums. Certainly the Hungarian construction should be available for general mixture problems; and in principle, seeking out sets of nearly uncorrelated likelihood sums might be possible; however, that project works in the exponential family case only because of the submultiplicative propery.

A referee asked about similar bounds for maxima of non-independent variables in the non-Gaussian case. Following the present path would seek subsets of variables each of which is "nearly independent" of the remaining variables in the subset. "Nearly independent" generalised from the gaussian case would mean that the tail probabilities of each variable conditioned on all other variables are greater than the tail probabilities of the original variable rescaled. This will be a formidable task if the covariance matrix does not control everything, as it does in the Gaussian case.

References

[1] Abramowitz, M. and Stegun, I. A., eds. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York. MR0208797

[2] Berman, S. I.(1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.* **35** 502–516. MR0161365

[3] Bickel, P. J. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift* (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. Prakasa Rao, eds.) 83–96. Wiley, New York.

[4] Chernoff, H. and Lander, E. (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *J. Statist. Plann. Inference* **43** 19–40. MR1314126

[5] Darling, D. and Erdos, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155. MR0074712

[6] Gombay, E. and Horvath, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and Their Applications* **50** 161–171. MR1262337

[7] Ghosh, J. K. and Sen, P. K. (1985). On the asymptotic performance of the log likelihood ratio statistics for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer.* **2** 789–806. MR0822065

[8] Gumbel, E. J. (1941). The return period of flood flows. *Ann. Math. Statist.* **12** 163–190. MR0004457

[9] Hartigan, J. A. (1972). A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer.* **2** 807–810. MR0822066

[10] Kim, H. J. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76** 409–423. MR1040636

[11] Komlos, J., Major, P. and Tusnady, G. (1976). An approximation of partial sums of independent RV's and the sample DF. I. *Wahrsch verw. Gebiete/Probability Theory and Related Fields* **32** 111–131. MR0375412

[12] Varah, J. M. (1975). A lower bound for the smallest singular value of a matrix. *Linear Algebra Appl.* **11** 3–5. MR0371929

[13] Yao, Y.-C. and Davis, R.A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhya Ser. A* **48** 339–353. MR0905446