MAXIMA OF THE CELLS OF AN EQUIPROBABLE MULTINOMIAL

ARUP BOSE, AMITES DASGUPTA, AND KRISHANU MAULIK

Indian Statistical Institute

ABSTRACT. Consider a sequence of multinomial random vectors with increasing number of equiprobable cells. We show that if number of trials increase fast enough, the sequence of maxima of the cells after a suitable centering and scaling converges to the Gumbel distribution.

Keywords: Triangular array, limit distribution, normal approximation, large deviation.

Subject Classification: Primary 60G70, 60F05; Secondary 60F10.

September 27, 2006

1. Introduction and main result

Let $(Y_{1n}, \ldots, Y_{m_n n})$ be a triangular sequence of random variables. Let $M_n = \max\{Y_{1n}, \ldots, Y_{m_n n}\}$. The question of convergence in distribution of M_n with linear normalisation has been addressed under a variety of conditions.

The classical case is when there is one sequence of iid random variables $\{Y_i\}$ and $M_n = \max\{Y_1, \ldots, Y_n\}$. In this case, necessary and sufficient conditions for the convergence are known. See for example, de Haan (1970), Fisher and Tippett (1928), Gnedenko (1943). In particular, it follows from these results that if $\{Y_i\}$ are i.i.d. Poisson or i.i.d. binomial with fixed parameters, then M_n cannot converge to any non degenerate distribution under any linear normalisation (cf. Leadbetter et al., 1983, pp 24–27). On the other hand (cf. Leadbetter et al., 1983, Theorem 1.5.3), if Y_i are i.i.d. standard normal variables then

$$\lim_{n\to\infty} P[M_n \le \alpha_n x + \beta_n] = \exp(-e^{-x}),$$

where

$$\alpha_n = \frac{1}{\sqrt{2\log n}} \tag{1.1}$$

and

$$\beta_n = \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}}. \tag{1.2}$$

General triangular schemes under various suitable conditions have been considered by several authors. The classical large deviation results due to Cramér (cf. Petrov, 1975, pg 218) play an important role in the proofs of these results.

Consider, for example, the case where $Y_{m_n n} = \left(\sum_{1\le j\le m_n} U_j - m_n\mu\right)/(\sigma m_n^{1/2})$ and U_j are i.i.d. with mean μ and standard deviation σ. Assuming that U_j has a finite moment generating function in an open interval containing the origin and $\log n = o\left(m_n^{(R+1)/(R+3)}\right)$ for some integer $R \ge 0$, Anderson et al. (1997) showed that

$$\lim_{n\to\infty} P[M_n \le \alpha_n x + \beta_n^{(R)}] = \exp(-e^{-x})$$

for α_n as in (1.1) and some suitable sequences $\beta_n^{(R)}$.

They also consider the following case. Suppose $m_n = n$ and for each n, $Y_{m_n n}$, are independent Poisson with mean λ_n such that for some integer $R \ge 0$, $\log n = o(\lambda_n^{(R+1)/(R+3)})$. Then again

$$\lim_{n\to\infty} P[M_n \le \lambda_n + \lambda_n^{1/2}(\beta_n^{(R)} + \alpha_n x)] = \exp(-e^{-x}),$$

where α_n and $\beta_n^{(R)}$ are as before. In particular, in the above results, if $R = 0$ then we can choose α_n as in (1.1) and $\beta_n^{(0)} = \beta_n$, given by (1.2).

In this paper we consider the following dependent situation. Suppose $\boldsymbol{Y}_n = (Y_{1n}, \cdots, Y_{nn})$ follow multinomial $(m_n; 1/n, \ldots 1/n)$ distribution and define $M_n = \max_{1\le i\le n} Y_{in}$ to be the maximum of the n cell variables. If m_n tends to infinity fast enough, then the sequence M_n after a suitable linear normalization, converges to the Gumbel distribution. We summarize this result in the following theorem:

Theorem 1.1. *Suppose $\boldsymbol{Y}_n$ is distributed as multinomial $(m_n; \frac{1}{n}, \ldots, \frac{1}{n})$ and $M_n = \max_{1\le i\le n} Y_{in}$. If*

$$\lim_{n\to\infty} \frac{\log n}{m_n/n} = 0 \tag{1.3}$$

holds, then, for $x \in \mathbb{R}$,

$$P\left[\frac{M_n - (m_n/n) - \beta_n\sqrt{m_n/n}}{\alpha_n\sqrt{m_n/n}} \le x\right] \to \exp(-e^{-x}), \tag{1.4}$$

where α_n *is as in* (1.1) *and* β_n *is the unique solution of*

$$\log z + \frac{1}{2}z^2 + \frac{1}{2}\log(2\pi) + z^2\sum_{i=1}^{\infty}\frac{(-1)^i}{(i+1)(i+2)}\left(\frac{z}{\sqrt{m_n/n}}\right)^i = \log n \tag{1.5}$$

in the region $\beta_n \sim \sqrt{2\log n}$.

2. Proofs

We first give an outline of the proof. Fix x, a real number. Denote

$$y_n = x\alpha_n\sqrt{m_n/n} + \beta_n\sqrt{m_n/n} + (m_n/n), \tag{2.1}$$

and

$$x_n = \frac{y_n - m_n/n}{\sqrt{m_n/n}} = \alpha_n x + \beta_n \sim \sqrt{2\log n}, \tag{2.2}$$

using (1.2).

Then for any fixed l, for sufficiently large n, using inclusion-exclusion principle and the identical distribution of the marginals from the multinomial distribution, we have,

$$\begin{aligned}1 - &\sum_{k=1}^{2l-1}(-1)^{k+1}\frac{n(n-1)\cdots(n-k+1)}{k!}P(\cap_{i=1}^k\{Y_{in} > y_n\})\\ \le &P(\cap_{i=1}^n\{Y_{in} \le y_n\})\\ \le &1 - \sum_{k=1}^{2l}(-1)^{k+1}\frac{n(n-1)\cdots(n-k+1)}{k!}P(\cap_{i=1}^k\{Y_{in} > y_n\}).\end{aligned} \tag{2.3}$$

For each fixed k, we are going to show that

$$n^k P(\cap_{i=1}^k\{Y_{in} > y_n\}) \to e^{-kx}. \tag{2.4}$$

where y_n and x are related as in (2.1).

Combining (2.3) and (2.4), we get for each fixed l,

$$\begin{aligned}1 - \sum_{k=1}^{2l-1}(-1)^{k+1}\frac{e^{-kx}}{k!} &\le \liminf_{n\to\infty} P(\cap_{i=1}^n\{Y_{in} \le y_n\})\\ &\le \limsup_{n\to\infty} P(\cap_{i=1}^n\{Y_{in} \le y_n\}) \le 1 - \sum_{k=1}^{2l}(-1)^{k+1}\frac{e^{-kx}}{k!},\end{aligned}$$

which gives the desired result (1.4) since l is arbitrary.

Towards establishing (2.4), let $(Z_0, Z_1, \ldots, Z_k)$ has multinomial $(1; \frac{n-k}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ distribution. Denote by F_n the distribution of $(Z_1 - \frac{1}{n}, \cdots, Z_k - \frac{1}{n})$. Note that F_n has mean vector $\mathbf{0}$ and its covariance matrix is given by $((a_{ij}))$, $a_{ii} = 1/n - 1/n^2$, $a_{ij} = -1/n^2, i \ne j$. Let $\boldsymbol{U}_n^{(i)} = (U_{1n}^{(i)}, \ldots, U_{kn}^{(i)})$, $1 \le i \le m_n$, be i.i.d. F_n.

Define $\boldsymbol{X}_n = (X_{1n}, \cdots, X_{kn}) = \sum_{i=1}^{m_n} \boldsymbol{U}_n^{(i)}$. Using these notations (2.4) becomes

$$\begin{aligned} P_{n,k} \equiv P_n =& P[X_{1n} > x_n\sqrt{m_n/n}, \ldots, X_{kn} > x_n\sqrt{m_n/n}] \\ =& \int_{x_n\sqrt{m_n/n}}^{\infty} \cdots \int_{x_n\sqrt{m_n/n}}^{\infty} dF_n^{\star m_n}(u_1, \ldots, u_k) \sim n^{-k}e^{-kx}. \end{aligned} \tag{2.5}$$

As a first approximation, we shall show existence of $v_n \equiv v_{n,k}(x) \sim \sqrt{2\log n} \sim x_n$, such that, for each fixed k and x,

$$n^k\bar{P}_n := n^k P[X_{1n} > v_n\sqrt{m_n/n}, \ldots, X_{kn} > v_n\sqrt{m_n/n}] \to e^{-kx} \tag{2.6}$$

holds.

To show existence of v_n, we first simplify (2.6) assuming the existence of $v_n \sim \sqrt{2\log n}$, see (2.35). We apply Esscher transform or exponential tilting on the distribution of $\boldsymbol{X}_n$, and then approximate it by a k-variate normal distribution with i.i.d. components having marginal mean and variance same as that of the tilted distribution.

Let $\Psi_n(t_1, \ldots, t_k)$ be the cumulant generating function of F_n:

$$\Psi_n(t_1, \ldots, t_k) = -\frac{t_1 + \cdots + t_k}{n} + \log\Big(1 + \frac{e^{t_1} + \cdots + e^{t_k} - k}{n}\Big). \tag{2.7}$$

Let s_n be the unique solution of

$$m_n\partial_1\Psi_n(s, \ldots, s) = v_n\sqrt{m_n/n}. \tag{2.8}$$

The following lemma on the rate of growth of $u_n = e^{s_n} - 1$ will be useful later. Here and for almost all of the discussion that follows, the specific form of v_n is not important, but we shall always crucially use the fact that

$$v_n \sim \sqrt{2\log n}. \tag{2.9}$$

Lemma 2.1. *If v_n satisfies (2.9) and if m_n satisfies (1.3) given by $\log n = o(m_n/n)$, we have*

$$u_n = \frac{v_n}{\sqrt{m_n/n}}\left(1 + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\frac{v_n}{\sqrt{m_n/n}}\right)\right). \tag{2.10}$$

Proof. Note that the first partial of Ψ_n is

$$\partial_1\Psi_n(t_1, \ldots, t_k) = -\frac{1}{n} + \frac{e^{t_1}}{e^{t_1} + \cdots + e^{t_k} + n - k}.$$

Hence, using (2.8), we have

$$\frac{v_n}{\sqrt{m_n/n}} = \frac{(n-k)(e^{s_n} - 1)}{n + k(e^{s_n} - 1)} = \frac{(n-k)u_n}{n + ku_n}. \tag{2.11}$$

Solving, we get,

$$u_n = \left(1 - \frac{k}{n}\right)^{-1}\left(1 - \frac{k}{n-k}\frac{v_n}{\sqrt{m_n/n}}\right)^{-1}\frac{v_n}{\sqrt{m_n/n}}$$

and, the result follows using $\frac{v_n}{\sqrt{m_n/n}} \sim \sqrt{\frac{2\log n}{m_n/n}} \to 0$, from (1.3). □

Next we define the exponential tilting for the multivariate case as

$$dV_n(w_1,\dots,w_k) = e^{-\Psi_n(s_n,\dots,s_n)} e^{s_n(w_1+\cdots+w_k)} dF_n(w_1,\dots,w_k). \tag{2.12}$$

Then, the m_n-th convolution power of V_n is given by

$$dV_n^{\star m_n}(u_1,\dots,u_k) = e^{-m_n\Psi_n(s_n,\cdots,s_k)} e^{s_n(w_1+\cdots+w_k)} dF_n^{\star m_n}(w_1,\dots,w_k),$$

in terms of which $\bar{P}_n$ in (2.6) becomes

$$\bar{P}_n = e^{m_n\Psi_n(s_n,\dots,s_n)} \int_{v_n\sqrt{m_n/n}}^{\infty} \cdots \int_{v_n\sqrt{m_n/n}}^{\infty} e^{-s_n(w_1+\cdots+w_k)} dV_n^{\star m_n}(w_1,\dots,w_k). \tag{2.13}$$

V_n has mean vector $\mu_n \mathbf{1}_k$ and covariance matrix $\Sigma_n = a_n I_k - b_n J_k$, where $\mathbf{1}_k$ is the k-vector with all coordinates 1, I_k is the $k \times k$ identity matrix, J_k is the $k \times k$ matrix with all entries 1 and μ_n, a_n and b_n are given as follows:

$$\begin{aligned}
\mu_n &= \partial_1 \Psi_n(s_n,\dots,s_n) = -\frac{1}{n} + \frac{e^{s_n}}{n + k(e^{s_n}-1)} \\
&= \frac{(n-k)(e^{s_n}-1)}{n(n+k(e^{s_n}-1))} = \frac{(n-k)u_n}{n(n+ku_n)} \sim \frac{u_n}{n} \qquad (2.14) \\
b_n &= -\partial_1\partial_2 \Psi_n(s_n,\dots,s_n) = \frac{e^{2s_n}}{(n+k(e^{s_n}-1))^2} \\
&= \left(\frac{1+u_n}{n+ku_n}\right)^2 \sim \frac{1}{n^2} \qquad (2.15) \\
\tau_n^2 := a_n - b_n &= \partial_1^2 \Psi_n(s_n,\dots,s_n) = \frac{e^{s_n}(n-k+(k-1)e^{s_n})}{(n+k(e^{s_n}-1))^2} \\
&= \frac{(1+u_n)(n-k+(k-1)(1+u_n))}{(n+ku_n)^2} \sim \frac{1}{n}, \qquad (2.16)
\end{aligned}$$

where the asymptotics hold by Lemma 2.1, as $v_n \sim \sqrt{2\log n} = o(\sqrt{m_n/n})$, using (1.3). Then using (2.8) and (2.14), we have from (2.13),

$$\bar{P}_n = e^{m_n\Psi_n(s_n,\dots,s_n)} \int_{m_n\mu_n}^{\infty} \cdots \int_{m_n\mu_n}^{\infty} e^{-s_n(u_1+\cdots+u_k)} dV_n^{\star m_n}(u_1,\dots,u_k). \tag{2.17}$$

Now we replace V_n by a k-variate normal with mean vector $\mu_n \mathbf{1}_k$ and covariance matrix $\tau_n^2 I_k$ (i.e., independent coordinates). The result of this change of distribution leads to the approximation (for $\bar{P}_n$), given by

$$\begin{aligned}
A_{s_n} =& e^{m_n\Psi_n(s_n,\dots,s_n))} \left[\int_{m_n\mu_n}^{\infty} e^{-s_n y} \phi\left(\frac{y - m_n\mu_n}{\tau_n\sqrt{m_n}}\right) \frac{dy}{\tau_n\sqrt{m_n}}\right]^k \qquad (2.18) \\
=& e^{m_n\Psi_n(s_n,\dots,s_n))} \left[\int_0^{\infty} \phi(z) e^{-s_n(m_n\mu_n + z\tau_n\sqrt{m_n})} dz\right]^k \\
=& e^{m_n(\gamma_n - ks_n\mu_n)} \rho^k(s_n\tau_n\sqrt{m_n}), \qquad (2.19)
\end{aligned}$$

where $\gamma_n = \Psi_n(s_n,\dots,s_n)$ and $\rho(t) = \int_0^\infty e^{-zt}\phi(z)dz = e^{\frac{t^2}{2}}(1-\Phi(t))$ and ϕ and Φ are the univariate standard normal density and distribution functions respectively. We shall prove the following.

Proposition 2.1. *If v_n satisfies (2.9), namely, $v_n \sim \sqrt{2\log n}$ and m_n satisfies (1.3) given by*

$$\lim_{n\to\infty} \frac{\log n}{m_n/n} = 0,$$

then

$$A_{s_n} \sim \frac{1}{(z_n\sqrt{2\pi})^k} \exp\left(-\frac{k}{2}z_n^2 - kz_n^2\sum_{i=1}^{\infty}\frac{(-1)^i}{(i+1)(i+2)}\left(\frac{z_n}{\sqrt{\frac{m_n}{n}}}\right)^i\right), \qquad (2.20)$$

where $z_n = u_n\sqrt{m_n/n}$.

Proof. We first treat the exponent in the first factor of the expression (2.19) for A_{s_n}. Since $\gamma_n = -\frac{k}{n}s_n + \log(1+\frac{k}{n}(e^{s_n}-1)) = -\frac{k}{n}\log(1+u_n) + \log(1+\frac{k}{n}u_n)$ using (2.7), it follows from expression (2.14) for μ_n,

$$\begin{aligned}
&m_n(\gamma_n - ks_n\mu_n)\\
=&\frac{m_n}{n}n\log\left(1+\frac{k}{n}u_n\right) - \frac{m_n}{n}\frac{k(1+u_n)\log(1+u_n)}{1+\frac{k}{n}u_n}\\
=&\frac{m_n}{n}\left(1+\frac{k}{n}u_n\right)^{-1}\left[(n+ku_n)\log\left(1+\frac{k}{n}u_n\right) - (k+ku_n)\log(1+u_n)\right]\\
=&k\frac{m_n}{n}\left(1+\frac{k}{n}u_n\right)^{-1}\sum_{r=2}^{\infty}\frac{(-1)^{r-1}}{r(r-1)}\left[1-\left(\frac{k}{n}\right)^{r-1}\right]u_n^r\\
=&-\frac{k}{2}\frac{m_nu_n^2}{n} + \frac{k^2}{2n}\frac{m_nu_n^2}{n}\\
&\quad - k\frac{m_nu_n^2}{n}\sum_{i=1}^{\infty}(-1)^i\sum_{r=0}^{i}\frac{1}{(r+1)(r+2)}\left(\frac{k}{n}\right)^{i-r}\left[1-\left(\frac{k}{n}\right)^{r+1}\right]u_n^i\\
=&-\frac{k}{2}\frac{m_nu_n^2}{n} - k\frac{m_nu_n^2}{n}\sum_{i=1}^{\infty}\frac{(-1)^i}{(i+1)(i+2)}u_n^i + E_n^{(0)} + E_n^{(1)},
\end{aligned} \qquad (2.21)$$

where, using (2.10),

$$E_n^{(0)} = \frac{k^2}{2n}\frac{m_nu_n^2}{n} \sim k^2\frac{\log n}{n} \to 0 \qquad (2.22)$$

and

$$\begin{aligned}
E_n^{(1)} = \frac{m_nu_n^2}{n}\sum_{i=1}^{\infty}(-1)^iu_n^i\Bigg[&\sum_{r=0}^{i-1}\frac{1}{(r+1)(r+2)}\left(\frac{k}{n}\right)^{i-r}\left\{1-\left(\frac{k}{n}\right)^{r+1}\right\}\\
&-\frac{1}{(i+1)(i+2)}\left(\frac{k}{n}\right)^{i+1}\Bigg]
\end{aligned}$$

is bounded by S_1+S_2, where, using the fact, from (2.10) and (1.3) that $m_nu_n^2/n \sim v_n^2 \sim 2\log n$,

$$\begin{aligned}
S_1 \le&\frac{m_nu_n^2}{n}\sum_{i=0}^{\infty}\left(\frac{k}{n}u_n\right)^{i+1}\sum_{r=0}^{i}\frac{1}{(r+1)(r+2)}\left(\frac{k}{n}\right)^{-r}\left\{1-\left(\frac{k}{n}\right)^{r+1}\right\}\\
=&\frac{m_nu_n^2}{n}\sum_{r=0}^{\infty}\frac{1}{(r+1)(r+2)}\frac{k}{n}\left\{1-\left(\frac{k}{n}\right)^{r+1}\right\}u_n^{r+1}\left(1-\frac{k}{n}u_n\right)^{-1}
\end{aligned}$$

$$\sim 2k\frac{\log n}{n}u_n\sum_{r=0}^{\infty}\frac{1}{(r+1)(r+2)}\left\{1-\left(\frac{k}{n}\right)^{r+1}\right\}u_n^r\to 0,$$

since the sum is finite, and

$$S_2\le\frac{m_nu_n^2}{n}\sum_{i=1}^{\infty}\left(\frac{k}{n}u_n\right)^i\sim 2\log n\frac{k}{n}u_n\to 0.$$

Hence, we have

$$E_n^{(1)}\to 0. \tag{2.23}$$

Thus, using (2.21)–(2.23), we have

$$m_n(\gamma_n-ks_n\mu_n)=-\frac{k}{2}\frac{m_nu_n^2}{n}-k\frac{m_nu_n^2}{n}\sum_{i=1}^{\infty}\frac{(-1)^i}{(i+1)(i+2)}u_n^i+o(1)$$

and hence, using $z_n=u_n\sqrt{m_n/n}$, we have,

$$e^{m_n(\gamma_n-ks_n\mu_n)}\sim\exp\left(-\frac{k}{2}z_n^2-kz_n^2\sum_{i=1}^{\infty}\frac{(-1)^i}{(i+1)(i+2)}\left(\frac{z_n}{\sqrt{m_n/n}}\right)^i\right). \tag{2.24}$$

To complete the proof of the proposition, we consider the second factor in (2.19). Using asymptotic expression (2.16) for τ_n, the fact $u_n=e^{s_n}-1\sim s_n$, we have

$$\tau_ns_n\sqrt{m_n}\sim u_n\sqrt{m_n/n}=z_n.$$

Also, we know that $\rho(t)=e^{\frac{t^2}{2}}(1-\Phi(t))\sim\frac{1}{t\sqrt{2\pi}}$ as $t\to\infty$ (cf. Feller, 1968, Lemma 2, Chapter VII). So, the proof is completed using

$$\rho^k(s_n\tau_n\sqrt{m_n})\sim\frac{1}{(z_n\sqrt{2\pi})^k} \tag{2.25}$$

and (2.24) in (2.19). □

It turns out that A_{s_n} is a good approximation for $\bar{P}_n$. Let $\Phi_{\boldsymbol{\mu},A}$ denote the k-variate normal distribution function with mean vector $\boldsymbol{\mu}$ and covariance matrix A. Then using (2.17) and (2.18), we easily see that

$$\frac{\bar{P}_n-A_{s_n}}{e^{m_n\gamma_n}}=\int_{m_n\mu_n}^{\infty}\cdots\int_{m_n\mu_n}^{\infty}e^{-s_n(u_1+\cdots+u_k)}d(V_n^{\star m_n}-\Phi^{\star m_n}_{\mu_n\mathbf{1}_k,\tau_n^2I_k})(u_1,\ldots,u_k).$$

Denote the distribution function of the signed measure $V_n^{\star m_n}-\Phi^{\star m_n}_{\mu\mathbf{1}_k,\tau_n^2I_k}$ by H_n. Then, using Theorem 3.1 in Appendix and (2.19), we have

$$|\bar{P}_n-A_{s_n}|\le 2^k\|H_n\|_\infty e^{m_n(\gamma_n-ks_n\mu_n)}=2^kA_{s_n}\rho^{-k}(s_n\tau_n\sqrt{m_n})\|H_n\|_\infty, \tag{2.26}$$

using (2.18), where $\|H_n\|_\infty$ is the sup norm. Hence, using (2.25) and the fact that $z_n=\sqrt{m_n/n}u_n\sim v_n$, using (2.10), we have

$$\frac{\bar{P}_n}{A_{s_n}}=1+O\left(v_n^k\|H_n\|_\infty\right). \tag{2.27}$$

Finally we study $\|H_n\|_\infty$. We write H_n as sum of two signed measures by introducing the normal distribution with covariance matrix, Σ_n, same as that of V_n:

$$H_n=\left(V_n^{\star m_n}-\Phi^{\star m_n}_{\mu_n\mathbf{1}_k,\Sigma_n}\right)+\left(\Phi^{\star m_n}_{\mu_n\mathbf{1}_k,\Sigma_n}-\Phi^{\star m_n}_{\mu_n\mathbf{1}_k,\tau_n^2I_k}\right). \tag{2.28}$$

We estimate the first part by Berry-Esseen theorem and handle the second part directly, which comes next.

Lemma 2.2. *Recall V_n is the exponential tilting defined in* (2.12) *with mean vector $\mu_n \mathbf{1}_k$, covariance matrix Σ_n and marginal variance τ_n^2. Then*

$$\|\Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \tau_n^2 I_k}\|_\infty = O(1/n)$$

and if, further, v_n satisfies (2.9), *then*

$$\lim_{n\to\infty} v_n^k \|\Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \tau_n^2 I_k}\|_\infty = 0. \tag{2.29}$$

Proof. Observe that

$$\|\Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \tau_n^2 I_k}\|_\infty = \|\Phi_{\mathbf{0}, \tau_n^{-2}\Sigma_n} - \Phi_{\mathbf{0}, I_k}\|_\infty,$$

which is estimated easily using Slepian's inequality (see, for example, Leadbetter et al., 1983, Theorem 4.2.1). Observe that $\tau_n^{-2}\Sigma_n = \frac{a_n}{a_n - b_n} I_k - \frac{b_n}{a_n - b_n} J_k$, using (2.15) and (2.16). Hence, from Slepian's inequality and asymptotic behavior of a_n and b_n in (2.15) and (2.16), we have

$$\begin{aligned}\|\Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \tau_n^2 I_k}\|_\infty &= \|\Phi_{\mathbf{0}, \tau_n^{-2}\Sigma_n} - \Phi_{\mathbf{0}, I_k}\|_\infty \\ &\le \frac{1}{2\pi} \frac{k(k-1)}{2} \frac{b_n}{\sqrt{a_n(a_n - 2b_n)}} \sim O(1/n).\end{aligned}$$

If v_n satisfies (2.9) and hence, $v_n^k = o(n)$, then (2.29) follows. □

Next we study the first term of (2.28).

Lemma 2.3. *Recall V_n is the exponential tilting defined in* (2.12) *with mean vector $\mu_n \mathbf{1}_k$, covariance matrix Σ_n. Assume v_n satisfies* (2.9), *and hence by* (1.3), $v_n = o(\sqrt{m_n/n})$. *Then*

$$\lim_{n\to\infty} v_n^k \|V_n^{\star m_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n}\|_\infty = 0. \tag{2.30}$$

Proof. Suppose $\boldsymbol{\xi}_j$ are i.i.d. V_n with mean $\mu_n \mathbf{1}_k$, covariance Σ_n. Then

$$\begin{aligned}&V_n^{\star m_n}(u_1, \ldots, u_k) - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n}(u_1, \ldots, u_k)) \\ &\quad = P\left[\frac{1}{\sqrt{m_n}} \sum_{j=1}^{m_n} (\boldsymbol{\xi}_j - \mu_n \mathbf{1}_k) \le \frac{\boldsymbol{u} - m_n \mu_n \mathbf{1}_k}{\sqrt{m_n}}\right] - \Phi_{\mathbf{0}, \Sigma_n}\left(\frac{\boldsymbol{u} - m_n \mu_n \mathbf{1}_k}{\sqrt{m_n}}\right),\end{aligned}$$

and hence, by multivariate Berry-Esseen theorem, (see, for example, Bhattacharya and Ranga Rao, 1976, Corollary 17.2, pg. 165)

$$\begin{aligned}\|V_n^{\star m_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n}\|_\infty &= \sup_{\boldsymbol{u}} \left| P\left[\frac{1}{\sqrt{m_n}} \sum_{j=1}^{m_n} (\boldsymbol{\xi}_j - \mu_n \mathbf{1}_k) \le \boldsymbol{u}\right] - \Phi_{0, \Sigma_n}(\boldsymbol{u})\right| \\ &\le \frac{C_3}{\sqrt{m_n}} \frac{\kappa_n}{\lambda_n^{3/2}},\end{aligned} \tag{2.31}$$

where $\kappa_n = E\|\boldsymbol{\xi}_1 - \mu_n \mathbf{1}_k\|_2^{3/2}$, (the norm being Euclidean one),

$$\lambda_n = a_n - k b_n \sim \frac{1}{n}, \tag{2.32}$$

by (2.15) and (2.16), is the smallest eigenvalue of $\Sigma_n = a_n I - b_n J$, and C_3 is a universal constant. So, to complete the proof we need to estimate κ_n. Using the definition of V_n, (2.12), we have,

$$\kappa_n = e^{-m_n\gamma_n} \int \cdots \int e^{s_n(u_1+\cdots+u_k)} \left(\sum_{j=1}^k (u_j-\mu_n)^2\right)^{3/2} dF_n(u_1,\ldots,u_k).$$

Recall that F_n is the distribution of the last k coordinates of the centered multinomial $(1;(n-k)/n, 1/n, \ldots, 1/n)$ distribution, which puts mass $1/n$ at each of the k vectors which have all coordinates $-1/n$ except the ith one being $(n-1)/n$, for $i = 1, \ldots, k$, and $(n-k)/n$ at $(-1/n, \ldots, -1/n)$. Thus,

$$e^{m_n\gamma_n}\kappa_n = \frac{n-k}{n} e^{\frac{ks_n}{n}} k^{\frac{3}{2}} \left(\frac{1}{n}+\mu_n\right)^3 + \frac{k}{n} e^{\frac{(n-k)s_n}{n}} \left[(k-1)\left(\frac{1}{n}+\mu_n\right)^2 + \left(1-\frac{1}{n}-\mu_n\right)^2\right]^{\frac{3}{2}}.$$

Since, from (2.14) we have $\mu_n \sim \frac{u_n}{n}$, and by (2.10), we have $s_n = \log(1+u_n) \sim u_n \to 0$,

$$\kappa_n \sim e^{-m_n\gamma_n} \frac{k}{n}.$$

Thus, using (2.31) and (2.32), we have,

$$v_n^k \|V_n^{\star m_n} - \Phi^{\star m_n}_{\mu_n \mathbf{1}_k, \Sigma_n}\|_\infty \leq kC_3 \frac{v_n^k}{\sqrt{\frac{m_n}{n}}} e^{-m_n\gamma_n}. \tag{2.33}$$

Finally, from (2.7), we get, for fixed k,

$$\begin{aligned} m_n\gamma_n &= m_n\Psi(s_n,\ldots,s_n) = -\frac{km_n s_n}{n} + m_n \log\left[1 + \frac{k(e^{s_n}-1)}{n}\right] \\ &= \frac{m_n}{n}[-k\log(1+u_n) + n\log(1+\frac{k}{n}u_n)] \sim \frac{k}{2}\frac{m_n}{n}u_n^2 \sim \frac{k}{2}v_n^2 \to \infty \end{aligned}$$

using (2.10). Hence, the result follows from (2.33). □

Combining Lemmas 2.2 and 2.3, under the assumption that $v_n \sim \sqrt{2\log n}$, we get,

$$\lim_{n\to\infty} v_n^k \|H_n\|_\infty = 0.$$

Thus, under the assumption $v_n \sim \sqrt{2\log n}$, we get from (2.20) and (2.27),

$$\bar{P}_n \sim \exp\left(-k\left\{\log z_n + \frac{1}{2}\log(2\pi) + \frac{1}{2}z_n^2 + z_n^2 \sum_{i=1}^\infty \frac{(-1)^i}{(i+1)(i+2)} \left(\frac{z_n}{\sqrt{m_n/n}}\right)^i\right\}\right). \tag{2.34}$$

Modifying Lemmas 1 and 2 of Anderson et al. (1997), we can find $z_n = \bar{\alpha}_n x + \beta_n$, such that

$$\log z_n + \frac{1}{2}\log(2\pi) + \frac{1}{2}z_n^2 + z_n^2 \sum_{i=1}^\infty \frac{(-1)^i}{(i+1)(i+2)} \left(\frac{z_n}{\sqrt{m_n/n}}\right)^i - \log n \to x$$

will hold. Note that the referred lemmas require a polynomial instead of a power series in the defining equation. However, the proofs work verbatim in our case due to the specific form of the coefficients. Also using (16) and (17) of the same reference, we have $\bar{\alpha}_n \sim (2\log n)^{-\frac{1}{2}}$ and β_n is the unique solution of (1.5) satisfying

$\beta_n \sim (2\log n)^{\frac{1}{2}}$. Observe that, $\bar{\alpha}_n$ and β_n, and hence, z_n will be free from k. Then using (2.11), we obtain v_n, which satisfies (2.6). Also, by (2.10),

$$v_n \sim u_n \sqrt{m_n/n} = z_n \sim \sqrt{2\log n}, \tag{2.35}$$

as required. However, the only problem that remains is v_n would be dependent on k, as is evident from (2.11). The convergence in (2.6)

$$n^k P\left[\frac{\min_{1\le i\le k} X_{in}}{\sqrt{m_n/n}} > v_n = \left(\frac{k/(n-k)}{\sqrt{m_n/n}} + \frac{n}{n-k}\frac{1}{\bar{\alpha}_n x + \beta_n}\right)^{-1}\right] \to e^{-kx} \tag{2.36}$$

is locally uniform in x, since the left hand side is monotone non-increasing in x and the right hand side is continuous in x (cf. Resnick, 1987, pg. 1).

Now, take $\alpha_n = \sqrt{2\log n}$ as in (1.1) and, further, define ξ_n through

$$\frac{1}{\alpha_n x + \beta_n} = \frac{k/(n-k)}{\sqrt{m_n/n}} + \frac{n}{n-k}\frac{1}{\bar{\alpha}_n \xi_n + \beta_n}. \tag{2.37}$$

Solving (2.37), we get

$$\xi_n = \frac{1}{\bar{\alpha}_n}\left[\frac{(1-\frac{k}{n})^{-1}}{\frac{1}{\alpha_n x+\beta_n} - \frac{k}{n-k}\frac{1}{\sqrt{m_n/n}}} - \beta_n\right] = \frac{\alpha_n}{\bar{\alpha}_n} x\gamma_n + \frac{\beta_n}{\alpha_n}[\zeta_n - 1], \tag{2.38}$$

where

$$\zeta_n = \frac{(1-\frac{k}{n})^{-1}}{1 - \frac{k}{n-k}\frac{\alpha_n x+\beta_n}{\sqrt{m_n/n}}} = 1 + O\left(\frac{1}{n}\sqrt{\frac{\log n}{m_n/n}}\right).$$

Since $\bar{\alpha}_n \sim \alpha_n = (2\log n)^{-1}$ and $\beta_n/\alpha_n \sim 2\log n$, we have, from (1.3) and (2.38), $\xi_n \to x$, using (1.3). Hence, using (2.37) and by local uniform convergence in (2.36), we have

$$\begin{aligned} n^k P_n &= n^k P[X_{1n} > (\alpha_n x + \beta_n)\sqrt{m_n/n}, \ldots, X_{1n} > (\alpha_n x + \beta_n)\sqrt{m_n/n}] \\ &= n^k P\left[\frac{\min\limits_{1\le i\le k} X_{in}}{\sqrt{m_n/n}} > \alpha_n x + \beta_n = \left(\frac{k/(n-k)}{\sqrt{m_n/n}} + \frac{n}{n-k}\frac{1}{\bar{\alpha}_n \xi_n + \beta_n}\right)^{-1}\right] \\ &\to e^{-kx}, \end{aligned}$$

as required in (2.5).

The above analysis also provides a similar result for the maximum of i.i.d. Binomial random variables.

Corollary 2.1. *Let $\{Y_{in} : 1 \le i \le n\}$ be a triangular array of independent Binomial random variables, with Y_{1n} having Binomial $(m_n; 1/n)$ distribution. If we have*

$$\lim_{n\to\infty} \frac{\log n}{m_n/n} = 0,$$

then we have

$$P\left[\frac{M_n - (m_n/n) - \beta_n\sqrt{m_n/n}}{\alpha_n\sqrt{m_n/n}} \le x\right] \to \exp(-e^{-x}),$$

where $\sigma_n^2 = 1/n - 1/n^2$ and α_n and β_n are chosen as in Theorem 1.1.

Proof. We have, from (2.4), with $k = 1$,

$$-\log\left(P[Y_{1n} \le y_n]\right)^n \sim nP[Y_{1n} > y_n] \to e^{-x},$$

where y_n is as in (2.1). The result follows immediately. □

3. Appendix

Finally, we prove the result on integration by parts, which was used in approximating the error between $\bar{P}_n$ and A_{s_n}, see (2.26). Let H be the distribution function of a finite signed measure on $\mathbb{R}^k$. For any subset I of $\{1, \ldots, k\}$ and $a \in \mathbb{R}$, define,

$$y_i^I = \begin{cases} a, & i \in I \\ y_i, & i \notin I \end{cases}, \quad \text{for } 1 \le i \le k,$$

$$H^I(a; y_1, \ldots, y_k) = H(y_1^I, \ldots, y_k^I)$$

and

$$H_{\boldsymbol{y}}^I(y_i; i \in I) = H(y_1, \ldots, y_k)$$

considered as a function in coordinates indexed by I only.

Theorem 3.1. *For $1 \le l \le k$ and $I \subseteq \{1, \ldots, l\}$, we have,*

$$\int_a^\infty \cdots \int_a^\infty e^{-s(y_1+\cdots+y_k)} dH_{y_1,\ldots,y_k}^{\{1,\ldots,l\}}(y_1, \ldots, y_k) = \sum_{I \subset \{1,\ldots,l\}} (-1)^{|I|} s^l e^{-s(y_1+\cdots+y_k)} H^I(a; y_1, \ldots, y_k) dy_1 \cdots dy_k. \quad (3.1)$$

The bound (2.26) then follows immediately by considering $l = k$.

Proof. We prove (3.1) by induction on l. For $l = 1$, (3.1) is the usual integration by parts formula. Assume (3.1) for l. Then

$$\begin{aligned}
&\int_a^\infty \cdots \int_a^\infty e^{-s(y_1+\cdots+y_{l+1})} H_{y_1,\ldots,y_k}^{\{1,\ldots,l+1\}}(dy_1, \ldots, dy_{l+1}) \\
&= \sum_{I \subset \{1,\ldots,l\}} (-1)^{|I|} \int_a^\infty \cdots \int_a^\infty s^l e^{-s(y_1+\cdots+y_l)} \int_a^\infty e^{-sy_{l+1}} H_{y_1^I,\ldots,y_k^I}^{\{l+1\}}(dy_{l+1}) dy_1 \cdots dy_l \\
&= \sum_{I \subset \{1,\ldots,l\}} (-1)^{|I|} \int_a^\infty \cdots \int_a^\infty s^l e^{-s(y_1+\cdots+y_l)} \Big[e^{-sa} H^{I \cup \{l+1\}}(a; y_1, \ldots, y_k) \\
&\qquad + \int_a^\infty s e^{-sy_{l+1}} H^I(a; y_1, \ldots, y_k) dy_{l+1} \Big] dy_1 \cdots dy_l \\
&= \sum_{I \subset \{1,\ldots,l\}} \int_a^\infty \cdots \int_a^\infty e^{-s(y_1+\cdots+y_{l+1})} s^{l+1} \times \\
&\qquad \Big[(-1)^{|I|+1} \mu_{I \cup \{l+1\}}(a; y_1, \ldots, y_k) + (-1)^{|I|} \mu_I(a; y_1, \ldots, y_k) \Big] dy_1 \cdots dy_{l+1}
\end{aligned}$$

where we use the induction hypothesis for the first step and the usual integration by parts for the second step, and the final step is the required sum, since any subset of $\{1, \ldots, l+1\}$ either contains $l+1$ or does not and the remainder is a subset of $\{1, \ldots, l\}$. This completes the inductive step and the proof of the theorem. □

References

C. W. Anderson, S. G. Coles, and J. Hüsler. Maxima of Poisson-like variables and related triangular arrays. *Ann. Appl. Probab.*, 7(4):953–971, 1997. ISSN 1050-5164.

R. N. Bhattacharya and R. Ranga Rao. *Normal approximation and asymptotic expansions.* John Wiley & Sons, New York-London-Sydney, 1976. Wiley Series in Probability and Mathematical Statistics.

L. de Haan. *On regular variation and its application to the weak convergence of sample extremes*, volume 32 of *Mathematical Centre Tracts.* Mathematisch Centrum, Amsterdam, 1970.

W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest and smallest members of a sample. *Proc. Camb. Philos. Soc.*, 24: 180–190, 1928.

B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453, 1943. ISSN 0003-486X.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, 1983. ISBN 0-387-90731-9.

V. V. Petrov. *Sums of independent random variables.* Springer-Verlag, New York, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

S. I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York, 1987. ISBN 0-387-96481-9.

Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

E-mail address: abose@isical.ac.in

Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

E-mail address: amites@isical.ac.in

Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

E-mail address: krishanu@isical.ac.in