Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 78 (2008) 135–143
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

On the expectation of the maximum of IID geometric random variables

Bennett Eisenberg

Department of Mathematics #14, Lehigh University, Bethlehem, PA 18015, USA

Received 6 December 2006; received in revised form 11 April 2007; accepted 23 May 2007
Available online 8 June 2007

Abstract

A study of the expected value of the maximum of independent, identically distributed (IID) geometric random variables is presented based on the Fourier analysis of the distribution of the fractional part of the maximum of corresponding IID exponential random variables.

Keywords: Geometric random variables; Exponential random variables; Maximum; Asymptotic expectation; Fractional part; Fourier series; Bioinformatics

Introduction

The probability problem studied in this paper is motivated by a statistical problem in bioinformatics. In bioinformatics one often compares long sequences of nucleotides (denoted A, C, G, T) in DNA for similarities or sometimes observes a single string for long runs of a given nucleotide or pattern of nucleotides (see Ewens and Grant, 2005, sec. 5.4 and sec. 6.3.). In the simplest case of comparing two strands of DNA, one might line up the two DNA strands and count the length of matches. Under the null hypothesis of randomness, this might be modeled as a modified geometric random variable (the number of matches until the first non-match) with parameter $p = .75$, the probability of a non-match. One might also try to match triples of nucleotides called codons, which play a major role in molecular biology. Here we might use $p = 1 - .25^3 = .98$. These values for p are approximations since the nucleotides do not have equal probabilities in reality. In testing for statistical significance in these procedures, the test statistic is often the maximum of an extremely large number of independent, identically distributed (IID) modified geometric random variables. In doing such tests approximations are used for the distribution of the test statistic under the null hypothesis. Ewens and Grant state on p. 138, "unless very accurate approximations are used for the mean and variance…, serious errors in P-value approximations can arise…". In this paper we analyze the standard approximation used for the mean of the maximum of IID geometric random variables.

It is well known and easily shown that $E(M_n)$, the expected value of the maximum of n IID exponential random variables with mean $1/\lambda$ is $(1/\lambda)\sum_{k=1}^{n}(1/k)$. This formula is useful for large n since $\sum_{k=1}^{n}(1/k)$ is

E-mail address: BE01@Lehigh.edu

0167-7152/$ - see front matter
doi:10.1016/j.spl.2007.05.011

asymptotic to $(\log n + \gamma)$, where $\gamma = .577\ldots$ is Euler's constant. There is no such simple expression for $E(M_n^*)$, the expected value of the maximum of n IID geometric random variables with mean $1/p$. There is the infinite series expression from the tail probabilities of the maximum of the random variables and even a finite sum expression, but these expressions are not so useful.

It is also easily seen that

$$\frac{1}{\lambda}\sum_{k=1}^{n}\frac{1}{k} \leqslant E(M_n^*) < 1 + \frac{1}{\lambda}\sum_{k=1}^{n}\frac{1}{k},$$

where $q = (1-p) = \mathrm{e}^{-\lambda}$. This is better, but not good enough. The gap of 1 between the upper and lower bounds is significant for moderate values of λ and n. A finer analysis is needed.

Careful asymptotic approximations are given for $E(M_n^*)$ in Szpankowski and Rego (1990). In our notation the Szpankowski and Rego result for the first moment is as follows:

$$E(M_n^*) = \sum_{k=1}^{n}\frac{1}{\lambda k} + \frac{1}{2} - \frac{1}{\lambda}\sum_{m\neq 0}\Gamma\left(\frac{2\pi i m}{\lambda}\right)\mathrm{e}^{-2\pi i m \log n/\lambda} + \mathrm{O}(n^{-1}). \tag{1}$$

Since $|\Gamma(2\pi i m/\lambda)|$ is very small for $\lambda < 2$ and all $m \neq 0$, this result has the interpretation that for $\lambda < 2$ and large n that $E(M_n^* - \sum_{k=1}^{n}\frac{1}{\lambda k})$, is close to $1/2$. This is the interpretation used in Jeske and Blessinger (2004) and applied to bioinformatics by Ewens and Grant (2005).

This interpretation is true, but can be misleading. First, there is no convergence to $1/2$ since the gamma function terms do not go to zero. They are just very small. Also the error term $\mathrm{O}(n^{-1})$ means that here is an unknown constant C such that $\mathrm{O}(n^{-1}) < C/n$. Without knowing the value of C, one does not know what the bound really is. In this case this problem is compounded by the fact that for reasonable values of n and λ, the value $1/n$ is much greater than the value of the gamma function terms. Hence the $\mathrm{O}(n^{-1})$ term can easily dominate the infinite sum in determining how close $E(M_n^* - \sum_{k=1}^{n}\frac{1}{\lambda k})$ is to $1/2$.

In this paper we use simple Fourier analysis to show that $E(M_n^*) - \sum_{k=1}^{n}\frac{1}{\lambda k}$ is very close to $1/2$ not only for moderate values of λ, but also relatively small values of n and that this difference is logarithmically summable to $1/2$ for all values of λ. Moderate λ may be interpreted as $\lambda < 2$. This corresponds to $p < .865$, which is almost always the case. We see, however, that the codon example is an exception to this. A key component of this work is the analysis of the distribution of the fractional part of M_n.

1. A survey of formulas for $E(M_n)$ and $E(M_n^*)$

Let $X_1, X_2, \ldots, X_n$ be IID exponential random variables with $P(X \leqslant x) = 1 - \mathrm{e}^{-\lambda x}$ for $x > 0$ and let $M_n = \max(X_1, \ldots, X_n)$. We then have $P(M_n \leqslant x) = (1 - \mathrm{e}^{-\lambda x})^n$. It follows that

$$\begin{aligned} E(M_n) &= \int_0^\infty P(M_n > x)\,\mathrm{d}x = \int_0^\infty 1 - (1-\mathrm{e}^{-\lambda x})^n\,\mathrm{d}x \\ &= \int_0^1 \frac{1-u^n}{\lambda(1-u)}\,\mathrm{d}u = \int_0^1 \sum_{k=0}^{n-1}\frac{u^k}{\lambda}\,\mathrm{d}u = \sum_{k=1}^{n}\frac{1}{\lambda k}. \end{aligned}$$

This also follows by decomposing M_n as

$$\begin{aligned} M_n &= X_{(1)} + (X_{(2)} - X_{(1)}) + \cdots + (X_{(n)} - X_{(n-1)}) \\ &= Y_1 + Y_2 + \cdots + Y_n, \end{aligned} \tag{2}$$

where $X_{(i)}$ is the ith order statistic of $X_1, \ldots, X_n$. It follows from the lack of memory property of exponential random variables and the fact that the minimum of exponential random variables is exponential that $Y_1, \ldots, Y_n$ are independent exponential random variables with parameters $n\lambda, (n-1)/\lambda, \ldots, 1/\lambda$, respectively. This implies

$$E(M_n) = \sum_{k=1}^{n}\frac{1}{\lambda k} \quad \text{and} \quad \mathrm{Var}(M_n) = \sum_{k=1}^{n}\frac{1}{\lambda^2 k^2}. \tag{3}$$

Things are not so simple for discrete geometric random variables. Let $X_1^*, \ldots, X_n^*$ be IID geometric random variables with $P(X_i^* = k) = q^{k-1}p$, for $k = 1, \ldots$. We then have $P(X_i^* \leqslant k) = 1 - q^k$. Now let $M_n^* = \max(X_1^*, \ldots, X_n^*)$. Then $P(M_n^* \leqslant k) = (1 - q^k)^n$, where k is a positive integer.

In analogy with the exponential case, we have

$$E(M_n^*) = \sum_{k=0}^{\infty} P(M_n > k) = \sum_{k=0}^{\infty} (1 - (1 - q^k)^n).$$

There is no closed form expression for this sum and we cannot decompose M_n^* into the sum of independent geometric random variables analogous to the decomposition of M_n^* since the order statistics $X_{(1)}^*, \ldots, X_{(n)}^*$ are not necessarily distinct. However, letting $q = \mathrm{e}^{-\lambda}$, we can say

$$\int_0^{\infty} 1 - (1 - \mathrm{e}^{-\lambda x})^n \, \mathrm{d}x < \sum_{k=0}^{\infty} (1 - (1 - \mathrm{e}^{-\lambda k})^n) < 1 + \int_0^{\infty} 1 - (1 - \mathrm{e}^{-\lambda x})^n \, \mathrm{d}x.$$

It follows that

$$\frac{1}{\lambda} \sum_{k=1}^{n} \frac{1}{k} < E(M_n^*) < 1 + \frac{1}{\lambda} \sum_{k=1}^{n} \frac{1}{k}. \tag{4}$$

Let K_n equal the number of $X_1^*, \ldots, X_n^*$ equal to 1. Szpankowski and Rego (1990) use complicated analysis to solve the recursive relation

$$\begin{aligned} E(M_n^*) &= \sum_{k=0}^{n} P(K_n = k) E(M_n^* | K_n = k) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (1 + E(M_{n-k}^*)) \end{aligned}$$

to find exact and asymptotic expressions for $E(M_n^*)$. We can easily derive some of their results using the tail probability generating of M_n^*. We have

$$\begin{aligned} Q_n(s) &= \sum_{k=0}^{\infty} s^k P(M_n^* > k) = \sum_{k=0}^{\infty} s^k (1 - (1 - q^k)^n) \\ &= \sum_{k=0}^{\infty} s^k \sum_{j=1}^{n} \binom{n}{j} (-1)^{j+1} q^{jk} = \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} (1 - sq^j)^{-1}. \end{aligned}$$

Letting $s = 1$, we find $E(M_n^*) = Q_n(1) = \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} (1 - q^j)^{-1}$. This agrees with the formula (2.6) given in Szpankowski and Rego (1990). Higher moments can be found using derivatives of $Q_n(s)$. However, even these finite sums are not very useful in analyzing the behavior of the moments as $n \to \infty$.

2. The connection between $E(M_n)$ and $E(M_n^*)$

If x is any variable or number, $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x and $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of x. We also say that $\{x\} = x \bmod 1$. If X is exponential with parameter λ, then for k a non-negative integer and $0 \leqslant x < 1$, we have

$$\begin{aligned} P(\lfloor X \rfloor = k, \{X\} \leqslant x) &= P(k \leqslant X < k + x) \\ &= \mathrm{e}^{-\lambda k} - \mathrm{e}^{-\lambda(k+x)} = \mathrm{e}^{-\lambda k}(1 - \mathrm{e}^{-\lambda x}). \end{aligned}$$

It follows that $\lfloor X \rfloor$ and $\{X\}$ are independent with $P(\lfloor X \rfloor = k) = \mathrm{e}^{-\lambda k}(1 - \mathrm{e}^{-\lambda})$ and $P(\{X\} \leqslant x) = (1 - \mathrm{e}^{-\lambda x})(1 - \mathrm{e}^{-\lambda})^{-1}$ for $k = 0, 1, 2, \ldots$ and $0 \leqslant x < 1$. We thus have $\lfloor X \rfloor = X^* - 1$, where X^* is geometric with $q = \mathrm{e}^{-\lambda}$.

Also

$$M_n = \max(X_1, \ldots, X_n) = \max(\lfloor X_1 \rfloor, \ldots, \lfloor X_n \rfloor) + \{M_n\}$$
$$= \max(X_1^* - 1, \ldots, X_n^* - 1) + \{M_n\} = M_n^* - 1 + \{M_n\},$$

where the X_i^* are IID geometric random variables.

Thus

$$E(M_n^*) = E(M_n) + 1 - E(\{M_n\}) = \sum_{k=1}^{n} \frac{1}{\lambda k} + 1 - E(\{M_n\}). \tag{5}$$

This gives a probabilistic proof of an improved version of (4). It follows from (5) that accurate estimates of $E(M_n^*)$ can be found by analyzing $E(\{M_n\})$, the expected value of the fractional part of the maximum of n IID exponential random variables.

3. The asymptotic distribution of $\{M_n\}$

The asymptotic distribution of the fractional part of M_n received some attention in the 1990s. Jagers (1990) solved a problem of Steutel of proving that the limiting distribution did not exist. Jagers proved this using fairly complicated analysis. He remarked that the proof is much easier for very large values of the parameter λ. In this section we give a simple proof of this result using characteristic functions and sequences. We then use these same characteristic sequences to study the asymptotic expectation.

The proof is based on the well-known and easily proved result that $M_n - \log n/\lambda$ converges in distribution to a random variable W with cumulative distribution function $\exp(-e^{-\lambda w})$ and characteristic function $\Gamma(1 - it/\lambda)$ (e.g. Ewens and Grant, 2005, p. 108). The decomposition in (2) implies that M_n has characteristic function $\phi_n(t) = \prod_{k=1}^{n}(1 - \frac{it}{\lambda k})^{-1}$ and $M_n - \log n/\lambda$ has characteristic function $\phi_n(t)\exp(-it\log n/\lambda)$. It is worth noting here that for geometric random variables not only does $M_n^* - \log n/\lambda$ not converge in distribution, but there do not exist sequences a_n and b_n such that $(M_n^* - a_n)/b_n$ converges in distribution to a non-degenerate random variable (see Arnold et al., 1992, p. 217).

Theorem 1. *The limiting distribution of* $\{M_n\}$ *does not exist.*

Proof. Let $\phi_n(t)$ be the characteristic function of M_n. It follows that

$$\phi_n(2\pi m) = E(e^{2\pi i m(\lfloor M_n \rfloor + \{M_n\})}) = E(e^{2\pi i m\{M_n\}}).$$

That is, $\phi_n(2\pi m)$ is the characteristic sequence of $\{M_n\}$, where m is an arbitrary integer. Since $0 \leqslant \{M_n\} < 1$, it determines its distribution. Furthermore, $\{M_n\}$ converges in distribution if and only if $\phi_n(2\pi m)$ converges for every m.

We know that $M_n - \log n/\lambda$ converges to W with characteristic function $\Gamma(1 - it/\lambda)$, which is never 0. Thus for each m, $\phi_n(2\pi m)e^{-2\pi m i \log n/\lambda}$ converges to a non-zero value. If $\phi_n(2\pi m)$ converged, it would have to converge to a non-zero value as well since $|e^{2\pi m i \log n/\lambda}| = 1$. This would imply that $e^{2\pi m i \log n/\lambda}$ converges, but it clearly does not. Therefore $\{M_n\}$ does not converge in distribution. □

A direct analysis of $\phi_n(2\pi m)$ would also show that it does not converge for $m \neq 0$. Jaegers notes that

$$\lim_{\lambda\to\infty} \mathrm{Var}(M_n) \leqslant \lim_{\lambda\to\infty} \sum_{k=1}^{\infty} \frac{1}{\lambda^2 k^2} = 0,$$

while $E(M_n) = \sum_{k=1}^{n} 1/(\lambda k)$. It follows that as $\lambda \to \infty$ for large n the distribution of M_n is concentrated about $\sum_{k=1}^{n} 1/(\lambda k) \bmod 1$. As n increases, these values are dense in $[0, 1]$, so $\{M_n\}$ cannot possibly converge in distribution. It follows that for large enough λ, $E(\{M_n^*\})$ will vary almost all the way from 0 to 1 as $n \to \infty$. Combined with (5), we see that for large λ, $E(M_n^*) - \sum_{k=1}^{n} 1/(\lambda k)$ will not stay close to 1/2 as $n \to \infty$.

We now consider the asymptotic behavior of $E(\{M_n\})$. To do that, we find a useful formula for the density of $\{M_n\}$. Note that $\phi_n(2\pi m) = \prod_{k=1}^{n}(1 - \frac{2\pi m i}{\lambda k})^{-1}$ and thus $|\phi_n(2\pi m i)|^2 = \prod_{k=1}^{n}(1 + \frac{4\pi^2 m^2}{\lambda^2 k^2})^{-1}$, which decreases

as n increases and as λ decreases. Moreover,

$$\lim_{n\to\infty}\prod_{k=1}^{n}\left(1+\frac{4\pi^2m^2}{\lambda^2k^2}\right)^{-1}=\left|\Gamma\left(1-\frac{2\pi mi}{\lambda}\right)\right|^2=\frac{2\pi^2m}{\lambda\sinh(2\pi^2m/\lambda)}.$$

The last equality follows from standard identities (see Silverman, 1967, pp. 310–320).

Theorem 2. *The density of* $\{M_n\}$ *is given by*

$$g_n(x)=1+\sum_{m\neq0}\prod_{k=1}^{n}\left(1+\frac{2\pi im}{\lambda k}\right)^{-1}e^{2\pi imx}\quad \textit{for } 0\leqslant x\leqslant 1.$$

Proof. This follows by noting that $\phi_n(-2\pi m)=\int_0^1 g_n(x)e^{-2\pi imx}\,dx$ is the mth Fourier coefficient of $g_n(x)$. The expression for $g_n(x)$ is merely its Fourier series expansion. To check the convergence we note that $\sum|\phi_n(2\pi m)|^2\leqslant\sum|\phi_1(2\pi m)|^2=\sum(1+(4\pi^2m^2)/\lambda^2)^{-1}<\infty$. □

An interesting question is how close is $g_n(x)$ to a uniform distribution for large n. Although we can write $g_n(x)=\sum_{t=0}^{\infty}f_n(x+t)$, where $f_n(x)=n\lambda e^{-\lambda x}(1-e^{-\lambda x})^n$ is the density of $M_n(x)$, this expression is not very useful for asymptotic analysis. For the asymptotic analysis, we invoke dominated convergence with $|\phi_n(2\pi m)|^2\leqslant|\phi_1(2\pi m)|^2$, to show

$$\lim_{n\to\infty}\sum_{m\neq0}|\phi_n(2\pi m)|^2=\sum_{m\neq0}\frac{2\pi^2m}{\lambda\sinh(2\pi^2m/\lambda)}.\tag{6}$$

This gives the corollary.

Corollary 1. *If* $g_n(x)$ *is the density of* $\{M_n\}$, *then*

$$\int_0^1(1-g_n(x))^2\,dx=2\sum_{m=1}^{\infty}\prod_{k=1}^{n}\left(1+\frac{4\pi^2m^2}{\lambda^2k^2}\right)^{-1}.$$

and

$$\lim_{n\to\infty}\int_0^1|1-g_n(x)|^2\,dx=\sum_{m\neq0}\frac{2\pi^2m}{\lambda\sinh(2\pi^2m/\lambda)}.$$

Now, $\lim_{\lambda\to\infty}2\pi^2m/\lambda\sinh(2\pi^2m/\lambda)=1$, so the sum in the right side of (6) can be large for large values of λ. However, for reasonable values of λ, say $\lambda<2$, things are different.

Let $c=2\pi^2/\lambda$ in (4). Then

$$\lim_{n\to\infty}\sum_{m\neq0}|\phi_n(2\pi m)|^2=\sum_{m\neq0}\frac{cm}{\sinh(cm)}=\sum_{m=1}^{\infty}\frac{4cme^{-cm}}{(1-e^{-2cm})}.$$

It easily follows that

$$4ce^{-c}(1-e^{-c})^{-2}<\sum_{m=1}^{\infty}\frac{4cme^{-cm}}{1-e^{-2cm}}<4ce^{-c}(1-e^{-2c})^{-1}(1-e^{-c})^{-2}$$

and the bounds are relatively tight for say, $c>2$ or $\lambda<10$ since the ratio of the lower bound to the upper bound is $(1-e^{-2c})$. In particular, we can see from the upper bound that for $\lambda=2$ or $c=\pi^2$, that $\lim_{n\to\infty}\sum_{m\neq0}|\phi_n(2\pi m)|^2<.0021$, which is very small. Using Corollary 1, we find the results summarized in Table 1.

We note that $\int_0^1|g_n(x)-1|^2\,dx$ is always a decreasing function of n and an increasing function of λ. The most interesting cases in applications are where q is large and p is very small, so the table shows that in this case, even for $n\geqslant10$ and $q\geqslant.368$ that $\int_0^1|g_n(x)-1|^2\,dx<.000074$. On the other hand, we see that for $q<2\times10^{-9}$, no matter how large the sample size n is, we have $\int_0^1|g_n(x)-1|^2\,dx>4$. This confirms the earlier observation that for large λ, the distribution of $\{M_n\}$ is not close to uniform for large n.

Table 1
$\int_0^1 (1-g_n(x))^2\,dx$

λ	$q=e^{-\lambda}$	$n=10$	$n\to\infty$
1	.368	.0000074	.000000211
2	.135	.00258	.00204
4	.018	.185	.144
20	2×10^{-9}	4.26	4.00

We can use these formulas to estimate $E(\{M_n\})$. We have

$$\left|E(\{M_n\})-\frac{1}{2}\right|^2=\left|\int_0^1\left(x-\frac{1}{2}\right)(g_n(x)-1)\,dx\right|^2$$
$$<\int_0^1\left(x-\frac{1}{2}\right)^2dx\int_0^1(g_n(x)-1)^2\,dx=\frac{1}{24}\int_0^1(g_n(x)-1)^2\,dx.$$

It then follows, for example, from the table that for $\lambda\leqslant 2(q\geqslant .135)$ and $n\geqslant 10$ that $|E(\{M_n\})-1/2|<.011$ and for $\lambda\leqslant 1(q\geqslant .368)$ and $n\geqslant 10$ that $|E(\{M_n\})-1/2|<.0006$. Thus from (3), we have that for $q\geqslant .368$ and $n\geqslant 10$,

$$\left|E(M_n^*)-\left(\sum_{k=1}^{n}\frac{1}{\lambda k}+1/2\right)\right|<.0006.$$

Brands et al. (1994) derive bounds on $|\int_0^t g_n(x)\,dx-t|$ by different methods. In terms of the results above we can say

$$\left|\int_0^t g_n(x)\,dx-t\right|=\left|\int_0^1 1_{[0,t]}(g_n(x)-1)\,dx\right|<\sqrt{t}\left(\int_0^1(g_n(x)-1)^2\,dx\right)^{1/2}.$$

We can also use the Fourier series to get an exact expression for $E(\{M_n^*\})$. We have that $\int_0^1 xe^{-2\pi imx}\,dx=(-2\pi mi)^{-1}$ for $m\neq 0$ and $\int_0^1 x\,dx=1/2$. Thus

$$E(\{M_n^*\})=\int_0^1 xg_n(x)\,dx=\frac{1}{2}+\sum_{m\neq 0}\frac{1}{2\pi mi}\phi_n(-2\pi m).$$

Substituting this into (5) gives us the following corollary.

Corollary 2.

$$E(M_n^*)=\sum_{k=1}^{\infty}\frac{1}{\lambda k}+\frac{1}{2}-\sum_{m\neq 0}\frac{1}{2\pi mi}\prod_{k=1}^{n}\left(1+\frac{2\pi mi}{k\lambda}\right)^{-1}.$$

The main interest of this formula is that it gives an exact version of the approximation (1). To see the connection, note that

$$\lim_{n\to\infty}\exp\left(\frac{2\pi im\log n}{\lambda}\right)\prod_{k=1}^{n}\left(1+\frac{2\pi im}{\lambda k}\right)^{-1}=\Gamma\left(1+\frac{2\pi im}{\lambda}\right).$$

Moreover,

$$\Gamma\left(1+\frac{2\pi im}{\lambda}\right)=\frac{2\pi im}{\lambda}\Gamma\left(\frac{2\pi im}{\lambda}\right).$$

So

$$\frac{1}{2\pi mi}\prod_{k=1}^{n}\left(1+\frac{2\pi mi}{k\lambda}\right)^{-1}\approx\frac{1}{\lambda}\Gamma\left(\frac{2\pi im}{\lambda}\right)\exp\left(-\frac{2\pi im\log n}{\lambda}\right).$$

4. The limit of the logarithmic means of $\{M_n^*\}$

In this section we show that although the sequence $\{M_n^*\}$ does not converge in distribution,

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}P(\{M_k^*\}\leqslant x)/k=x.$$

Such convergence is called *logarithmic summability*. It follows that $E(\{M_n^*\})$ and $E(M_n^*)-\sum_{k=1}^{n}1/(\lambda k)$ are each logarithmically summable to $1/2$ and this is true even for large values of λ, where the distributions of $\{M_n\}$ are concentrated around values all the way from 0 to 1 as $n\to\infty$. Although we have not proved that $E(\{M_n^*\})$ does not converge for small λ, it seems highly unlikely given the formulas we have derived.

Let K_n equal the number of $\lfloor X_i\rfloor$ equal to $\max(\lfloor X_1\rfloor,\ldots,\lfloor X_n\rfloor)$. Then, using the fact that $\lfloor X_i\rfloor$ and $\{X_i\}$ are independent, it is easily seen that $M_n=\max(X_1,\ldots,X_n)=\lfloor M_n\rfloor+\{M_n\}$, where $\lfloor M_n\rfloor=\max(\lfloor X_1\rfloor,\ldots,\lfloor X_n\rfloor)$ and $\{M_n\}=W_{K_n}$, where W_{K_n} is the maximum of K_n IID random variables equal in distribution to $\{X_i\}$. For example, if $X_1=4.5$, $X_2=1.8$, and $X_3=4.2$, then $M_3=4.5$, $\lfloor M_3\rfloor=\max(\lfloor X_1\rfloor,\lfloor X_2\rfloor,\lfloor X_3\rfloor)=\max(4,1,4)=4$, $K_3=2$, and $\{M_3\}=W_2=\max(\{X_1\},\{X_3\})=\max(.5,.2)=.5$.

We therefore have

$$P(\{M_n\}\leqslant x)=P(W_{K_n}\leqslant x)=\sum_{k=1}^{\infty}P(K_n=k)F^k(x). \tag{7}$$

This result is also used in Brands et al. (1994). It is convenient to allow the sum to go to infinity even though $P(K_n=k)=0$ for $k>n$.

We recall that $\lfloor X_i\rfloor+1$ is geometric with $q=\mathrm{e}^{-\lambda}$. It then follows from (7) and Theorem 1 that for IID geometric random variables, the number of variables taking the maximum value does not converge in distribution. (If they did, $\{M_n\}$ would have to converge in distribution and it does not.) Eisenberg et al. (1993) used a much more complicated argument to prove this surprising result. However, it is shown in Baryshnikov et al. (1995) and reproved by a different method in Olofsson (1999) that

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{n=1}^{N}\frac{P(K_n=k)}{n}=\frac{(1-q)^k}{k|\log q|}\quad\text{for } k=1,2,\ldots. \tag{8}$$

This leads to the following theorem.

Theorem 3.

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{n=1}^{N}\frac{P(\{M_n\}\leqslant x)}{n}=x\quad \textit{for } 0\leqslant x\leqslant 1.$$

Proof. From (7) we have

$$\begin{aligned}\frac{1}{\log N}\sum_{n=1}^{N}\frac{P(\{M_n\}\leqslant x)}{n}&=\frac{1}{\log N}\sum_{n=1}^{N}\sum_{k=1}^{\infty}\frac{P(K_n=k)}{n}F^k(x)\\&=\sum_{k=1}^{\infty}\left(\frac{1}{\log N}\sum_{n=1}^{N}\frac{P(K_n=k)}{n}\right)F^k(x).\end{aligned}$$

Thus using (8) and dominated convergence from the summability of $F^k(x)$, we have

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{n=1}^{N}\frac{P(\{M_n\}\leqslant x)}{n}=\sum_{k=1}^{\infty}\frac{(1-\mathrm{e}^{-\lambda})^k}{k|\log(\mathrm{e}^{-\lambda})|}\left[\frac{(1-\mathrm{e}^{-\lambda x})}{(1-\mathrm{e}^{-\lambda})}\right]^k.$$

The right side immediately simplifies to

$$\sum_{k=1}^{\infty}\frac{(1-\mathrm{e}^{-\lambda x})^k}{k\lambda}=\frac{-\log(\mathrm{e}^{-\lambda x})}{\lambda}=x. \qquad \square$$

Corollary 3.

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{n=1}^{N}\frac{E(\{M_n\})}{n}=1/2.$$

Proof.

$$\left|\frac{1}{\log N}\sum_{n=1}^{N}\frac{1-P(\{M_n\}\leqslant x)}{n}\right|<3$$

for all $N>1$. Therefore from Theorem 3 and dominated convergence

$$\begin{aligned}\frac{1}{\log N}\sum_{n=1}^{N}\frac{E(\{M_n\})}{n}&=\frac{1}{\log N}\sum_{n=1}^{N}\frac{\int_0^1(1-P(\{M_n\}\leqslant x)\,\mathrm{d}x}{n}\\&=\int_0^1\frac{1}{\log N}\sum_{n=1}^{N}\frac{(1-P(\{M_n\}\leqslant x)}{n}\,\mathrm{d}x\to\int_0^1 1-x\,\mathrm{d}x=1/2. \qquad \square\end{aligned}$$

Corollary 3 plus (5) gives us the following corollary.

Corollary 4.

$$\lim_{N\to\infty}\frac{1}{\log N}\sum\nolimits_{n=1}^{N}\frac{E(M_n^*)-\sum_{k=1}^{n}\frac{1}{\lambda k}}{n}=1/2.$$

5. A puzzling phenomenon

We next consider the sequence of *new* record values. Call the nth new record value R_n. Then R_{n+1} is the value of the first exponential random variable after the time of the nth new record taking a greater value than R_n. It follows by the lack of memory property of exponential random variables that R_{n+1} is equal in distribution to R_n+X, where X is an exponential random variable independent of R_n. Thus R_n is equal in distribution to $X_1+\cdots+X_n$, where the X_i's are IID exponential random variables with parameter λ (e.g. Arnold et al., 1992, p. 243). R_n therefore has an Erlang(n,λ) distribution and characteristic function $\psi(t)=(1-it/\lambda)^{-n}$. $\{R_n\}$ has characteristic sequence $\psi_n(2\pi m)$ which converges to 0 for $m\neq 0$. It follows that $\{R_n\}$ converges in distribution to a uniform random variable on $[0,1]$.

This is a bit surprising after what we have seen for $\{M_n\}$. Let N_n denote the number of new records in the first n random variables. It then follows that $\{M_n\}=\{R_{N_n}\}$. That is, the maximum observation up to time n is the value of the most recent new record. Now it is known for large n that N_n is of the order $\log n$ (see Arnold

et al., 1992, p. 247) so at first glance one might think that $\{M_n\}$ has approximately the same asymptotic distribution as $\{R_{[\log n]}\}$, namely the uniform distribution. Evidently this is not the case. With some further analysis one can see where this reasoning goes wrong.

We have $P(R_{N_n} \leqslant x | N_n = m) = P(M_n \leqslant x | N_n = m)$. But clearly M_n and N_n are independent. The number of new records in n observations depends on the order of the observations and not on their absolute magnitudes. So, for example, $P(M_2 \leqslant x | N_2 = 1) = P(X_1 \leqslant x | X_1 > X_2)$ and $P(M_2 \leqslant x | N_2 = 2) = P(X_2 \leqslant x | X_2 > X_1)$. These two probabilities are the same by symmetry, so we see that M_2 is independent of N_2. In this way we conclude that

$$\begin{aligned} P(M_n \leqslant x) &= \sum_{m=1}^{n} P(N_n = m) P(R_{N_n} \leqslant x | N_n = m) \\ &= \sum_{m=1}^{n} P(N_n = m) P(R_{N_n} \leqslant x) = P(R_{N_n} \leqslant x). \end{aligned}$$

Thus the relation $\{M_n\} = \{R_{N_n}\}$ appears to be of no help in relating the asymptotic distribution of $\{M_n\}$ to that of $\{R_n\}$. Still there should be some connection since both represent sequences of maximum values. A description of the relationship would be of interest.

References

Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics. Wiley, New York.

Baryshnikov, Y., Eisenberg, B., Stengle, G., 1995. A necessary and sufficient condition for the existence of the probability of a tie for first place. Statist. Probab. Lett. 23, 203–209.

Brands, J.J.A.M., Steutel, F.W., Wilms, R.J.G., 1994. On the number of maxima in a discrete sample. Statist. Probab. Lett. 20, 209–217.

Eisenberg, B., Stengle, G., Strang, G., 1993. The asymptotic probability of a tie for first place. Ann. Appl. Probab. 3, 731–745.

Ewens, W.J., Grant, G.R., 2005. Statistical Methods in Bioinformatics. An Introduction, second ed. Springer, New York.

Jagers, A.A., 1990. Solution of problem 247 of F.W. Steutel. Statist. Neerlandica 44, 180.

Jeske, D.R., Blessinger, T., 2004. Tunable approximations for the mean and variance of the maximum of heterogeneous geometrically distributed random variables. Amer. Statist. 58, 322–327.

Olofsson, P., 1999. A Poisson approximation with applications to the number of maxima in a discrete sample. Statist. Probab. Lett. 44, 23–27.

Silverman, R.A., 1967. Introductory Complex Analysis. Prentice-Hall, Englewood Cliffs, NJ.

Szpankowski, W., Rego, V., 1990. Yet another application of binomial recurrence: order statistics. Computing 43, 401–410.