THE NUMBER OF ELEMENTS CLOSE TO NEAR-RECORDS IN GEOMETRIC SAMPLES

GUY LOUCHARD AND HELMUT PRODINGER

Abstract. The statistics of interest here are related to an independent sequence of geometrically distributed random variables. We look at the $(n-k)$th order statistic $(=(k+1)$st winner) and study the number of elements that fall exactly a away from the value of this "person," where a is a fixed integer. This is motivated by a recent paper by Balakrishnan and Stepanov who considered a continous analogue of the problem.

1. Introduction

Let the random variable X be geometrically distributed, i.e., $\mathbb{P}\{X = k\} = pq^{k-1}$, with $q = 1 - p$. Also, assume that n independent copies are given. Finally, let K count the number of winners, i.e. the number of random variables with highest value.

The probability distribution of K, the expectation $\mathbb{E}K$ of K and the asymptotic behaviour of $\mathbb{P}\{K = 1\}$ (probability of a single winner) for $n \to \infty$ was the subject of a series of papers some 10 years ago. Without trying to be complete here, we cite some of them: [5, 6, 2, 14, 3, 9, 4].

Assume that the sample is given by $X_1, \ldots, X_n$. In the theory of order statistics, they are sorted and being denoted by $X_{1,n} \le \cdots \le X_{n,n}$. Thus, $X_{n,n}$ is the winner, $X_{n-1,n}$ the second winner, etc.

Pakes and Steutel [13] considered a continuous analogue: Let $M_n = X_{n,n}$ be the maximum and $a > 0$ be a fixed parameter. The interest was in the random variable

$$K(n, a) = \#\{j \mid X_j \in (M_n - a, M_n]\}$$

i.e., in the number of elements "close to the maximum."

Balakrishnan and Stepanov [1] went one step further: They are interested in

$$K_-(n, k, a) = \#\{j \mid X_j \in (X_{n-k,n} - a, X_{n-k,n}]\},$$

i.e., in the number of elements "close to the $(k + 1)$st winner."

If $k > 0$, then the notion of

$$K_+(n, k, a) = \#\{j \mid X_j \in (X_{n-k,n}, X_{n-k,n} + a]\}$$

also makes sense.

In this paper, we come back to the original discrete (geometric) set-up, and study

$$\#\{j \mid X_j = X_{n-k,n} + a\}.$$

The integer parameter a might be positive or negative or zero. For $k = a = 0$, we are back at the original question of "the number of winners."

We will always think about an urn model, with urns labelled $1, 2, \ldots$, where we throw n balls, the probability of each ball falling into urn k being given by pq^{k-1}. Of course, we can neglect the order in which the n items arrive.

Date: March 30, 2006.

In particular, we will prove the following theorems.

Theorem 1. *Set $Q = 1/q$, $L = \ln Q$, C any positive constant and $\chi_v = \frac{2\pi \mathbf{i} v}{L}$. The average number of elements that are a below the level of the $(k+1)$st winner, is, for $n \to \infty$, given by*

$$\frac{p^2 Q^{1+a}(k+1)}{L} + \delta_{below}(\log_Q n) + \mathcal{O}(n^{-C}),$$

where the periodic function $\delta_{below}(\theta)$ is given by the following Fourier series:

$$\delta_{below}(\theta) = \frac{p^2 Q^{1+a}}{Lk!} \sum_{v \neq 0} \frac{\Gamma(\chi_v + k + 2)}{\chi_v + 1} e^{-2\pi \mathbf{i} v \theta}.$$

For $k = 0$, this formula appeared already in [9].

Theorem 2. *The average number of elements that are a above the level of the $(k+1)$st winner, is, for $n \to \infty$, and $k > 0$, given by*

$$\frac{p^2 Q^{1-a} k}{L} + \delta_{above}(\log_Q n) + \mathcal{O}(n^{-C}),$$

where the periodic function $\delta_{below}(\theta)$ is given by the following Fourier series:

$$\delta_{above}(\theta) = \frac{p^2 Q^{1-a}}{L(k-1)!} \sum_{v \neq 0} \frac{\Gamma(\chi_v + k + 1)}{\chi_v + 1} e^{-2\pi \mathbf{i} v \theta}.$$

Theorem 3. *The average number of elements that are at the level of the $(k+1)$st winner, is, for $n \to \infty$, given by*

$$\frac{pQ(pk+1)}{L} + \delta_{at}(\log_Q n) + \mathcal{O}(n^{-C}),$$

where the periodic function $\delta_{at}(\theta)$ is given by the following Fourier series:

$$\delta_{at}(\theta) = \frac{pQ}{Lk!} \sum_{v \neq 0} \frac{\Gamma(\chi_v + k + 1)}{\chi_v + 1} (pk + 1 + \chi_v) e^{-2\pi \mathbf{i} v \theta}.$$

Although the computations are not in the paper, we also got the variances. We state the main term, ignoring a small fluctuation and an error term of the form $\mathcal{O}(n^{-C})$.

Theorem 4. *With the same assumptions as in the previous theorems, the variance is asymptotic to:*
In the "below" instance:

$$\frac{1}{L} p^3 (1+q) Q^{2a+2} \binom{k+2}{2} + \frac{1}{L} p^2 Q^{1+a}(k+1) - \Big(\frac{1}{L} p^2 Q^{1+a}(k+1)\Big)^2.$$

In the "above" instance:

$$\frac{1}{L} p^3 (1+q) Q^{2-2a} \binom{k}{2} + \frac{1}{L} Q^{1-a} k - \Big(\frac{1}{L} Q^{1-a} k\Big)^2.$$

In the "at" instance:

$$\frac{p^2 Q^2}{L} \left[\binom{k+2}{2} - \binom{k}{2} q^2 \right] + \frac{pQ(pk+1)}{L} - \Big(\frac{pQ(pk+1)}{L}\Big)^2.$$

It is even possible to compute the sth factorial moment explicitly (leading term only, apart from small fluctuation). Here are the results:

Theorem 5. *With the same assumptions as in the previous theorems, the sth factorial moment is asymptotic to:*
In the "below" instance:

$$\frac{p^s Q^{s(1+a)}(k+1)^{\overline{s}}}{sL}(1-q^s).$$

In the "above" instance:

$$\frac{p^s Q^{s(1-a)}k^{\underline{s}}}{sL}(1-q^s).$$

In the "at" instance:

$$\frac{p^s Q^s}{sL}\Big[(k+1)^{\overline{s}}-q^s k^{\underline{s}}\Big].$$

We get also results on generating functions and probability distributions that the reader can find in the following sections.

The distributions do not converge in the weak sense, they do however converge along subsequences n_m for which the fractional part of $\log n_m$ is constant.

We would also like to emphasize that *integer compositions* are closely related to the instance $p = q = \frac{1}{2}$; the probability that digit 1 occurs is $\frac{1}{2}$, that digit 2 occurs is $\frac{1}{4}$, etc. The difference is that the sum of the digits must be n, whereas normally we are interested in n balls (or digits in this case). However, the differences are minor, and we refer to [8] and the references therein.

2. Notations

Looking at the problem as an urn model, we use the following notations:

$$\begin{aligned}
\Pi(\lambda,u) &:= e^{-\lambda}\lambda^u/u!,\ \text{(Poisson distribution)},\\
n^* &:= n\frac{p}{q},\\
Q &:= 1/q,\\
\log &:= \log_Q,\\
\eta &:= m-\log n^*,\\
L &:= \ln Q,\\
\{x\} &:=\ \text{fractional part of } x,\\
m &:=\ \text{the urn containing the } (k+1)\text{th winner},\\
\mathbf{I} &:=\ \text{number of balls in urn } m,\\
\mathbf{J} &:=\ \text{number of balls in urn } m-a \text{ or urn } m+a,\\
\mathbf{L} &:=\ \text{number of balls to the right of urn } m,\\
M &:= \mathbb{E}(\mathbf{J}),\\
\partial z\big|_{z=1} &:=\ \text{the operator that differentiats a function and then replaces } z \text{ by } 1.
\end{aligned}$$

3. Below the $(k+1)$st winner

3.1. **Probability Generating function of J.** In the urn model, we have asymptotic independence of urns behaviour: this is proved in Louchard, Prodinger and Ward [12], with analytic Poissonization-dePoissonization techniques. Also, replacing Binomials by Poisson and using a uniform integrability condition (by computing rates of convergence) are justified in [12] and in Louchard and Prodinger [11, 10].

Assume that the $(k+1)$st winner is in urn m, that there are l balls in urns above m, i balls in urn m and j balls in urn $m-a$. This leads to

$$\mathbb{E}(z^{\mathbf{J}}) \sim F_1(z,\eta) = \sum_{m=1}^{\infty} \exp\left(e^{-L(\eta-a)}(z-1)\right) \sum_{l=0}^{k} \Pi\left(\frac{q}{p}e^{-L\eta}, l\right) \sum_{i=k+1-l}^{\infty} \Pi\left(e^{-L\eta}, i\right)$$
$$= \sum_{m=1}^{\infty} \exp\left(e^{-L(\eta-a)}(z-1)\right) \sum_{l=0}^{k} \Pi\left(\frac{q}{p}e^{-L\eta}, l\right) \left[1 - \sum_{i=0}^{k-l} \Pi\left(e^{-L\eta}, i\right)\right].$$

This is a harmonic sum. So we set $y = Q^{-\eta} = e^{-L\eta} = n^*/Q^m$, and we define

$$g_1(z,y) = \exp\left(ye^{La}(z-1)\right) \sum_{l=0}^{k} \Pi\left(\frac{q}{p}y, l\right) \left[1 - \sum_{i=0}^{k-l} \Pi(y,i)\right].$$

But

$$\sum_{l=0}^{k} \Pi\left(\alpha y, l\right) \sum_{i=0}^{k-l} \Pi(\beta y, i) = \sum_{v=0}^{k} \Pi\left((\alpha+\beta)y, v\right), \qquad \alpha, \beta > 0.$$

So we have

$$g_1(z,y) = \exp\left(ye^{La}(z-1)\right) \sum_{l=0}^{k} \left[\Pi\left(\frac{q}{p}y, l\right) - \Pi\left(\frac{1}{p}y, l\right)\right].$$

Now we take the Mellin transform of $g_1(z,y)$ (for a good reference on Mellin transforms, see Flajolet et al. [7] or Szpankowski [15]). This gives

$$g_1^*(z,s) = \int_0^{\infty} y^{s-1} g_1(z,y)dy = \sum_{l=0}^{k} \left[T_{1,1}(z,s,l) - T_{1,4}(z,s,l)\right],$$

with fundamental strip $\langle -1, 0\rangle$, and

$$T_{1,1}(z,s,l) = \frac{(p/q)^s \left(\frac{q-pze^{La}+pe^{La}}{q}\right)^{-s-l} \Gamma(s+l)}{\Gamma(l+1)},$$
$$T_{1,4}(z,s,l) = \frac{p^s \left(1-p(z-1)e^{La}\right)^{-s-l} \Gamma(s+l)}{\Gamma(l+1)}.$$

This leads to

$$g_1^*(z,0) = \sum_{l=0}^{k} \left[T_{1,1}(z,0,l) - T_{1,4}(z,0,l)\right],$$

where

$$T_{1,1}(z,0,l) = \frac{\left(\frac{q-pze^{La}+pe^{La}}{q}\right)^{-l}}{l},$$
$$T_{1,4}(z,0,l) = \frac{\left(1-p(z-1)e^{La}\right)^{-l}}{l},$$

and, of course, the term in the sum for $l=0$ is to be interpreted in the limiting sense.

For further use, we also compute

$$T_{1,2}(z,s,i,l) = \int_0^{\infty} y^{s-1} \exp\left(ye^{La}(z-1)\right) \Pi\left(\frac{q}{p}y, l\right) \Pi(y,i)$$

$$= \frac{p^{i+s}\left(1-pze^{La}+pe^{La}\right)^{-i-s-l}q^l\Gamma(i+s+l)}{\Gamma(l+1)\Gamma(i+1)},$$

$$T_{1,2}(z,0,i,l) = \frac{p^i(1-pze^{La}+pe^{La})^{-i-l}q^l\Gamma(i+l)}{\Gamma(l+1)\Gamma(i+1)},$$

$$T_{1,2}(1,0,i,l) = \frac{p^iq^l\Gamma(i+l)}{\Gamma(i+1)\Gamma(l+1)}.$$

We now deal with the Mellin transform related to the Probability generating function; this is given by

$$\chi(z,s) = g_1^*(z,s)\frac{Q^s}{1-Q^s},$$

and to obtain the asymptotic form of the Probability generating function, we invert this Mellin transform (again, for details, see [7]). Two interesting particular cases of the inverted Mellin transform, for $s=0$ (wich represents the dominant term), are

$z=1$, which must lead to 1,

$\partial_z|_{z=1}$, which leads to the mean $M_1 := \mathbb{E}(\mathbf{J})$.

For the case $z=1$, we note that, for $s=0$, $l=0$, we must avoid the singularity at $y=0$. So we must compute separately

$$\int_0^\infty y^{-1}\left[\Pi\left(\frac{q}{p}y,0\right)-\Pi\left(\frac{1}{p}y,0\right)\right]dy = L,$$

$$\int_0^\infty y^{-1}\Pi\left(\alpha y,0\right)dy = \frac{1}{l}, \qquad l>0, \alpha>0.$$

This leads to

$$\frac{1}{L}\,L = 1,$$

as expected.

The periodic component is given by $w_1(\{\log n^*\}) = w_1(\log n^*)$ where, with $\chi_v := 2\pi\mathbf{i}v/L$,

$$w_1(\theta) = \frac{1}{L}\sum_{v\neq 0}\sum_{l=0}^{k}\left[T_{1,1}(1,\chi_v,l)-T_{1,4}(1,\chi_v,l)\right]e^{-2\pi\mathbf{i}v\theta}.$$

This must give identically 0. The second summation leads to (we just write χ without index):

$$\sum_{l=0}^{k}\left[(p/q)^\chi\frac{\Gamma(\chi+l)}{l!}-p^\chi\frac{\Gamma(\chi+l)}{l!}\right].$$

Note that $q^\chi = 1$, which evaluates the sum to 0, as it should.

For the case $\partial_z|_{z=1}$, we obtain, for the asymptotic dominant part of the mean M_1, the expression

$$\tilde{M_1} = \frac{1}{L}\sum_{l=0}^{k}\left[V_{1,1}(l)-V_{1,4}(l)\right],$$

with

$$V_{1,1}(l) = \partial_z T_{1,1}(z,0,l)|_{z=1} = \frac{pe^{La}}{q},$$

$$V_{1,4}(l) = \partial_z T_{1,4}(z,0,l)|_{z=1} = pe^{La},$$

and

$$\tilde{M_1} = \frac{pQ^{a+1}(k+1)}{L} - \frac{pQ^a}{L}(k+1)$$
$$= \frac{p^2Q^{a+1}(k+1)}{L}.$$

To obtain the periodic component of the mean, we compute

$$S_{1,1}(s,l) = \partial_z T_{1,1}(z,s,l)|_{z=1} = \frac{(p/q)^s e^{La} p\Gamma(s+l+1)}{q\Gamma(l+1)},$$
$$S_{1,4}(s,l) = \partial_z T_{1,4}(z,s,l)|_{z=1} = \frac{p^s e^{La} p\Gamma(s+l+1)}{\Gamma(l+1)},$$

and $M_1 = \tilde{M_1} + w_2(\{\log n^*\}) + \mathcal{O}(n^{-C})$, for any positive constant C, where the periodic component $w_2(\theta)$ is given by

$$w_2(\theta) = \frac{1}{L}\sum_{v\neq 0}\sum_{l=0}^{k}\left[S_{1,1}(\chi_v,l) - S_{1,4}(\chi_v,l)\right]e^{-2\pi \mathbf{i} v\theta}.$$

Indeed, the fundamental strip is now $\langle -1, \infty\rangle$.

All periodic functions we will obtain do have mean 0.

Again, we can simplify that:

$$\sum_{l=0}^{k}\left[S_{1,1}(\chi,l) - S_{1,4}(\chi,l)\right] = \sum_{l=0}^{k}\left[\frac{p^\chi Q^a p\Gamma(\chi+l+1)}{ql!} - \frac{p^\chi Q^a p\Gamma(\chi+l+1)}{l!}\right]$$
$$= \frac{p^\chi Q^a p\Gamma(\chi+1)}{q}\sum_{l=0}^{k}\binom{\chi+l}{\chi} - p^\chi Q^a p\Gamma(\chi+1)\sum_{l=0}^{k}\binom{\chi+l}{\chi}$$
$$= \frac{p^\chi Q^a p\Gamma(\chi+1)}{q}\binom{\chi+k+1}{\chi+1} - p^\chi Q^a p\Gamma(\chi+1)\binom{\chi+k+1}{\chi+1}$$
$$= \frac{p^\chi Q^a p\Gamma(\chi+k+2)}{q(\chi+1)k!} - \frac{p^\chi Q^a p\Gamma(\chi+k+2)}{(\chi+1)k!}$$
$$= \frac{p^{2+\chi}Q^{a+1}\Gamma(\chi+k+2)}{(\chi+1)k!}.$$

Hence

$$w_2(\theta) = \frac{1}{L}\sum_{v\neq 0}\frac{p^{2+\chi}Q^{a+1}\Gamma(\chi_v+k+2)}{(\chi_v+1)k!}e^{-2\pi \mathbf{i} v\theta}.$$

Note that $w_2(\theta)$ is a periodic function of θ. We have made some experiments for the case $k=3$, $a=1$, $p=2/3$. This is given in Figure 1. We obtain $\tilde{M_1} = 14.563382762\ldots$. For $\theta = 0$, we have $\tilde{M_1} + w_2(0) = 14.80212765\ldots$. Oscillations are of order .25.

Of course all moments can be derived by the same method.

3.2. **Probability distribution of J.** As in the previous section, we derive

$$\mathbb{P}(\mathbf{J}=j) \sim F_1(j,\eta) = \sum_{m=1}^{\infty}\Pi\left(e^{-L(\eta-a)},j\right)\sum_{l=0}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\left[1-\sum_{i=0}^{k-l}\Pi\left(e^{-L\eta},i\right)\right]$$

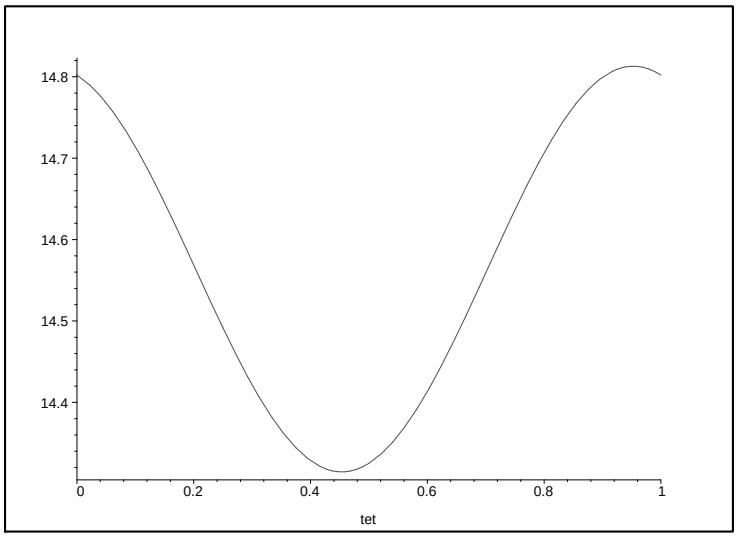

FIGURE 1. $\tilde{M}_1 + w_2(\theta)$, $\theta = 0..1$

$$= \sum_{m=1}^{\infty} \Pi\left(e^{-L(\eta-a)}, j\right) \sum_{l=0}^{k} \left[\Pi\left(\frac{q}{p}e^{-L\eta}, l\right) - \Pi\left(\frac{1}{p}e^{-L\eta}, l\right)\right].$$

Set $\eta = m - \lfloor \log n^* \rfloor - \{\log n^*\}$ and $u := m - \lfloor \log n^* \rfloor$. We obtain another expression: $\Psi_1(j, \{\log n^*\})$ for the probability distribution (it is easy to check that the boundary values are negligible, see [10]), where

$$\Psi_1(j,\theta) = \sum_{u=-\infty}^{\infty} \Pi\left(e^{-L(u-\theta-a)}, j\right) \sum_{l=0}^{k} \left[\Pi\left(\frac{q}{p}e^{-L(u-\theta)}, l\right) - \Pi\left(\frac{1}{p}e^{-L(u-\theta)}, l\right)\right].$$

This is a periodic function of θ.

In our example, Figure 2 gives $\Psi_1(j, 0)$ for $j = 1..50$. This curious behaviour is explained later on. Figure 3 gives $\Psi_1(j, 0)$ for $j = 50..120$.

Figures 4, 5, 6 give $\Psi_1(j, \theta), \theta = 0..1$ for $j = 0, 8, 17$. For $j = 8$, oscillations are of order .025.

We observe that these periodic functions have shifted phases.

The numerical value of the mean, for $\theta = 0$, based on $j = 0..150$ is given by $14.80212765\ldots$ This is identical to the value obtained in Section 3.1

3.3. **Mellin transform related to the probability distribution of J.** We first set

$$g_1(j, y) = \Pi\left(ye^{La}, j\right) \sum_{l=0}^{k} \left[\Pi\left(\frac{q}{p}y, l\right) - \Pi\left(\frac{1}{p}y, l\right)\right].$$

This leads to

$$g_1^*(j, s) = \int_0^{\infty} y^{s-1} g_1(j, y) dy.$$

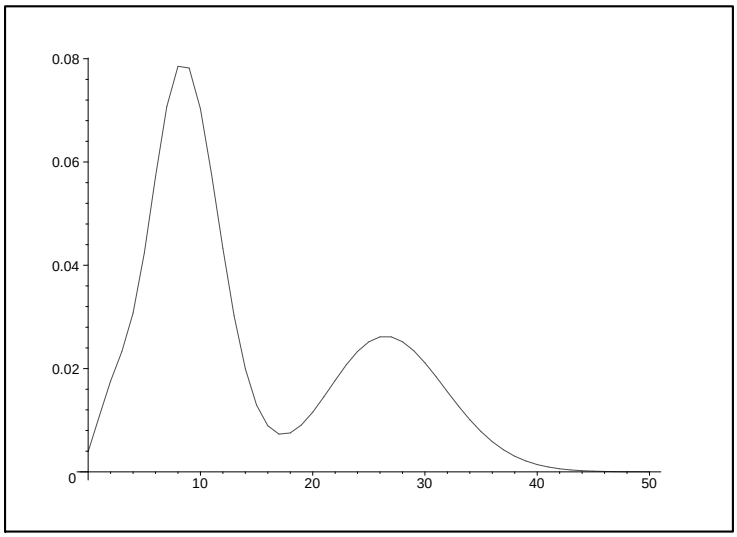

FIGURE 2. $\Psi_1(j,0)$ for $j = 0..50$

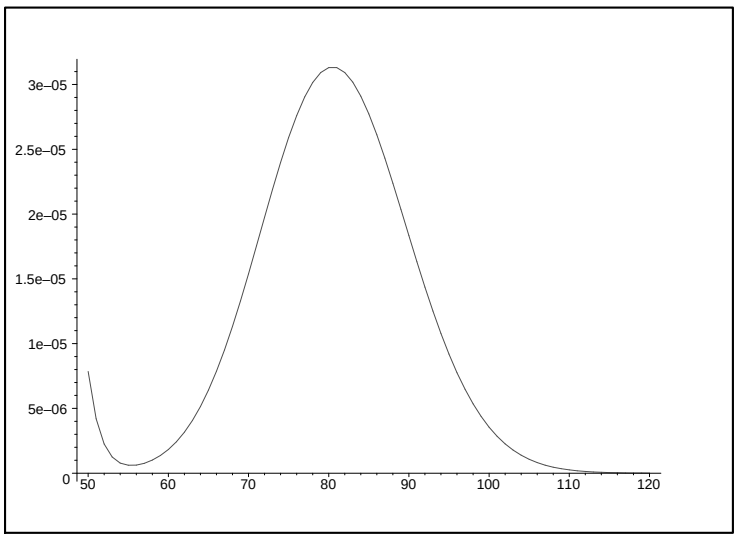

FIGURE 3. $\Psi_1(j,0)$ for $j = 50..120$

We note again that, for $l = 0$, we must avoid the singularity at $y = 0$. So we must consider separately

$$R_{1,1}(j,s) = \int_0^\infty y^{s-1}\Pi\left(ye^{La}, j\right)\left[\Pi\left(\frac{q}{p}y, 0\right) - \Pi\left(\frac{1}{p}y, 0\right)\right]dy,$$
$$R_{1,3}(j,s,l) = \int_0^\infty y^{s-1}\Pi\left(ye^{La}, j\right)\Pi\left(\frac{q}{p}y, l\right)dy,$$
$$R_{1,4}(j,s,l) = \int_0^\infty y^{s-1}\Pi\left(ye^{La}, j\right)\Pi\left(\frac{1}{p}y, l\right)dy.$$

The fundamental strip is now $\langle -1, \infty\rangle$.

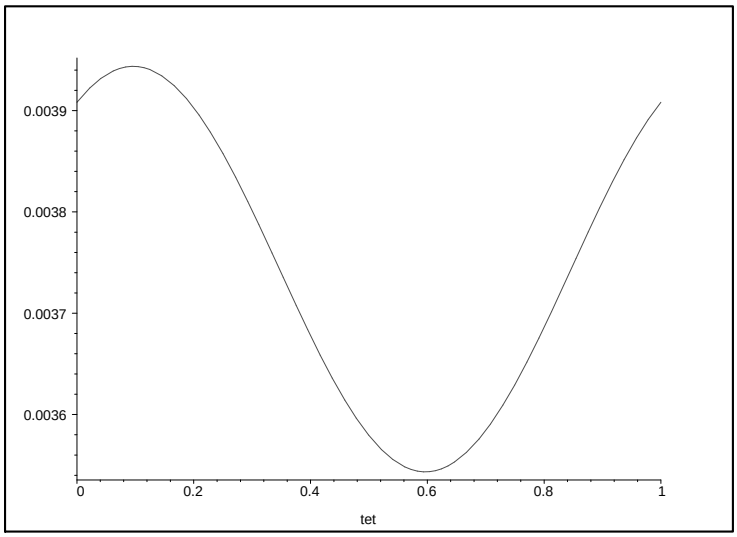

FIGURE 4. $\Psi_1(0,\theta)$, $\theta = 0..1$

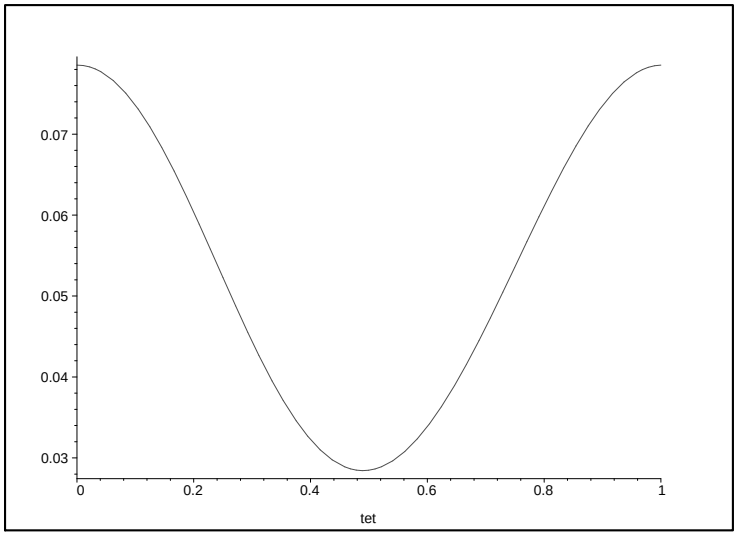

FIGURE 5. $\Psi_1(8,\theta)$, $\theta = 0..1$

Actually, we only need the following expressions

$$R_{1,1}(0,0) = \ln\left(\frac{1+pe^{La}}{q+pe^{La}}\right),$$

$$R_{1,3}(j,s,l) = \frac{e^{La(-s-l)}\left(\frac{p+qe^{-La}}{p}\right)^{-s-j-l}\left(\frac{q}{p}\right)^{l}\Gamma(s+j+l)}{\Gamma(l+1)\Gamma(j+1)},$$

$$R_{1,4}(j,s,l) = \frac{e^{La(-s-l)}\left(\frac{p+e^{-La}}{p}\right)^{-s-j-l}\left(\frac{1}{p}\right)^{l}\Gamma(s+j+l)}{\Gamma(l+1)\Gamma(j+1)},$$

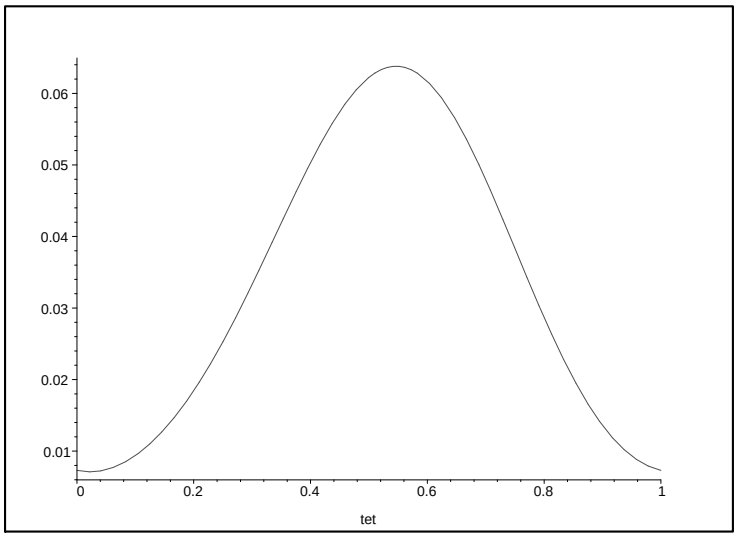

FIGURE 6. $\Psi_1(17,\theta)$, $\theta = 0..1$

$$R_{1,3}(j,0,l) = \frac{e^{-Lal}\left(\frac{p+qe^{-La}}{p}\right)^{-j-l}\left(\frac{q}{p}\right)^{l}\Gamma(j+l)}{\Gamma(l+1)\Gamma(j+1)},$$
$$R_{1,4}(j,0,l) = \frac{e^{-Lal}\left(\frac{p+e^{-La}}{p}\right)^{-j-l}\left(\frac{1}{p}\right)^{l}\Gamma(j+l)}{\Gamma(l+1)\Gamma(j+1)}.$$

We obtain

$$\mathbb{P}(\mathbf{J}=j) = \tilde{P}(j) + w_3(\{\log n^*\}) + \mathcal{O}(n^{-C}),$$

with

$$\tilde{P}(0) = \frac{1}{L}\left(R_{1,1}(0,0) + \sum_{l=1}^{k}\left[R_{1,3}(0,0,l) - R_{1,4}(0,0,l)\right]\right).$$

For $j > 0$, we derive

$$\tilde{P}(j) = \frac{1}{L}\left(\sum_{l=0}^{k}\left[R_{1,3}(j,0,l) - R_{1,4}(j,0,l)\right]\right).$$

Again, we must have $\sum_{j=0}^{\infty}\tilde{P}(j) = 1$. We will confirm this now by a direct computation:

$$R_{1,1}(0,0) + \sum_{l=1}^{k}\left[R_{1,3}(0,0,l) - R_{1,4}(0,0,l)\right] + \sum_{j\geq 1}\sum_{l=0}^{k}\left[R_{1,3}(j,0,l) - R_{1,4}(j,0,l)\right]$$
$$= R_{1,1}(0,0) + \sum_{j\geq 1}\left[R_{1,3}(j,0,0) - R_{1,4}(j,0,0)\right] + \sum_{l=1}^{k}\sum_{j\geq 0}\left[R_{1,3}(j,0,l) - R_{1,4}(j,0,l)\right]$$

$$= \ln\left(\frac{1+pQ^a}{q+pQ^a}\right) + \sum_{j\geq 1}\left[\frac{\left(\frac{p+Q^{-a-1}}{p}\right)^{-j}}{j} - \frac{\left(\frac{p+Q^{-a}}{p}\right)^{-j}}{j}\right]$$
$$+ \sum_{l=1}^{k}\frac{1}{l}Q^{-al}p^{-l}\sum_{j\geq 0}\binom{j+l-1}{j}\left[\left(\frac{p+Q^{-a-1}}{p}\right)^{-j-l}q^l - \left(\frac{p+Q^{-a}}{p}\right)^{-j-l}\right]$$
$$= \ln\left(\frac{1+pQ^a}{q+pQ^a}\right) + \ln\left(\frac{1+pQ^{a+1}}{1+pQ^a}\right) + \sum_{l=1}^{k}\frac{1}{l}Q^{-al}p^{-l}\left[p^lQ^{al} - p^lQ^{al}\right]$$
$$= \ln\left(\frac{1+pQ^{a+1}}{q+pQ^a}\right) = L,$$

so the formula is confirmed.

In our example, this gives $\tilde{P}(0) = .003743724888\ldots, \tilde{P}(8) = .05322292897\ldots$. A plot of $\tilde{P}(j)$ for $j = 0..50$ is given in Figure 7. This distribution is quite smooth.

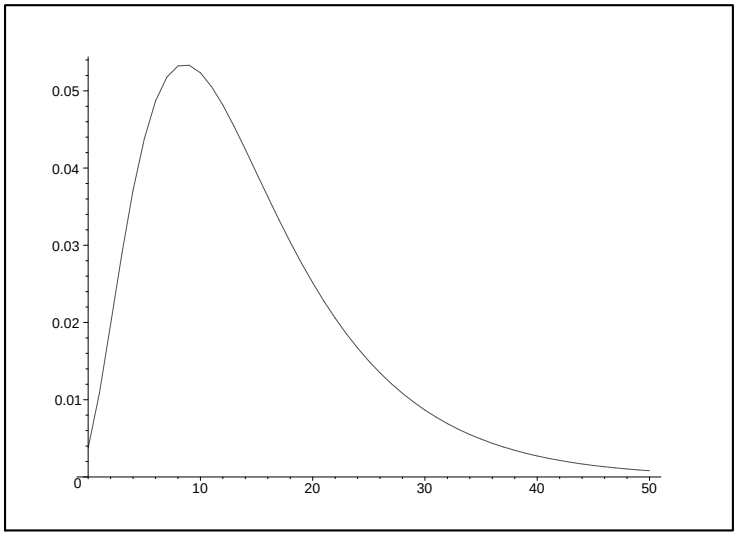

FIGURE 7. $\tilde{P}(j)$ for $j = 0..50$

The periodic component of $\mathbb{P}(\mathbf{J} = j)$ is asymptotically given by $w_3(j, \{\log n^*\})$, with

$$w_3(j,\theta) = \frac{1}{L}\sum_{v\neq 0}\sum_{l=0}^{k}\left[R_{1,3}(j,\chi_v,l) - R_{1,4}(j,\chi_v,l)\right]e^{-2\pi \mathbf{i} v\theta}.$$

As these periodic corrections have shifted phases, this explains now the behaviour of $\Psi_1(j,0)$ shown in Figures 2 and 3.

We must have $\sum_{j=0}^{\infty} w_3(j,\theta) = 0$, which we confirm now:

Since

$$R_{1,3}(j,\chi,l) = Q^{-a(l+\chi)}(1+\frac{q}{p}Q^{-a})^{-\chi-j-l}\Big(\frac{q}{p}\Big)^l\frac{\Gamma(\chi+j+l)}{l!j!},$$

we have

$$\sum_{j\geq 0}R_{1,3}(j,\chi,l) = p^{\chi}\frac{\Gamma(\chi+l)}{l!},$$

and, since we obtain the same expression from $R_{1,4}$, this leads indeed to 0.

Our example gives, in Figures 8 and 9 a comparison between $\tilde{P}(j) + w_3(j,\theta)$ and $\Psi_1(j,\theta)$, for $j = 0$ resp. $j = 8$.

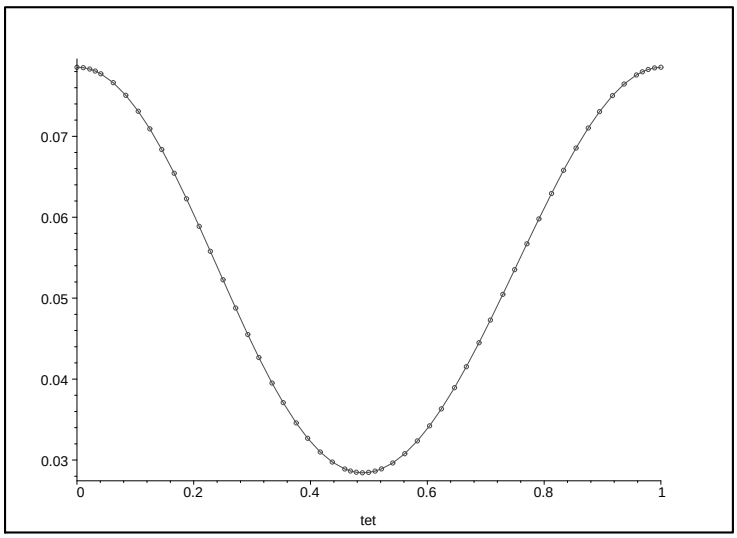

FIGURE 8. $\Psi_1(8,\theta)$ (line), $\tilde{P}(8) + w_3(8,\theta)$ (circles), $\theta = 0..1$

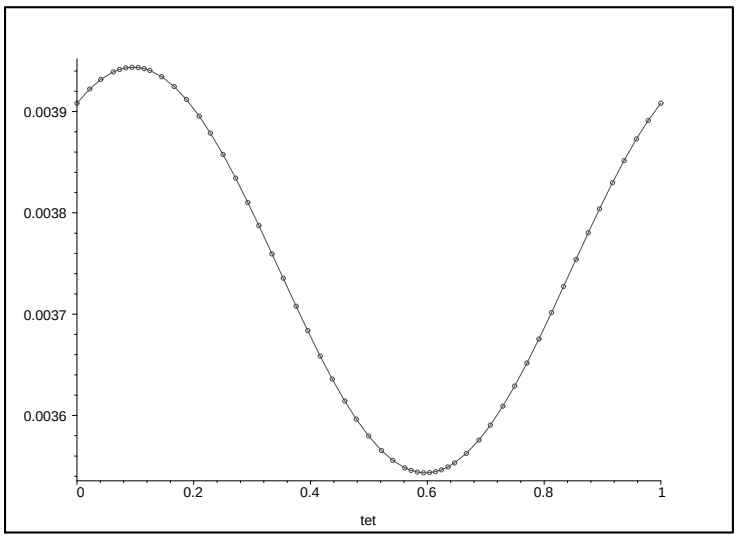

FIGURE 9. $\Psi_1(0,\theta)$ (line), $\tilde{P}(0) + w_3(0,\theta)$ (circles), $\theta = 0..1$

4. Above the $(k+1)$st winner

4.1. **Probability distribution of J.** In this case, we didn't get a compact expression for the generating function of **J**, so we deal directly with the probability distribution. Assume

that the $(k+1)$st winner is in urn m, that there are l balls in urns above m, i balls in urn m and j balls in urn $m+a$. This leads, with $0 \le j \le k$, to

$$\mathbb{P}(\mathbf{J}=j) \sim F_2(j,\eta) = \sum_{m=1}^{\infty} \Pi\left(e^{-L(\eta+a)}, j\right) \sum_{l=j}^{k} \Pi\left(\frac{q}{p}e^{-L\eta} - e^{-L(\eta+a)}, l-j\right) \left[1 - \sum_{i=0}^{k-l} \Pi\left(e^{-L\eta}, i\right)\right]$$

$$= \sum_{m=1}^{\infty} \Pi\left(e^{-L(\eta+a)}, j\right) \sum_{v=0}^{k-j} \left[\Pi\left(e^{-L\eta}\left(\frac{q}{p} - e^{-La}\right), v\right) - \Pi\left(e^{-L\eta}\left(\frac{1}{p} - e^{-La}\right), v\right)\right].$$

As in Section 3.2, this leads to another expression: $\Psi_2(j, \{\log n^*\})$ where

$$\Psi_2(j,\theta) = \sum_{u=-\infty}^{\infty} \Pi\left(e^{-L(u-\theta+a)}, j\right) \sum_{v=0}^{k-j} \left[\Pi\left(e^{-L(u-\theta)}\left(\frac{q}{p} - e^{-La}\right), v\right) - \Pi\left(e^{-L(u-\theta)}\left(\frac{1}{p} - e^{-La}\right), v\right)\right].$$

This is a periodic function of θ. In our example, this gives $\Psi_2(0,0) = 0.2388165797\ldots, \Psi_2(1,0) = 0.3929358425\ldots, \Psi_2(2,0) = 0.2803374485\ldots, \Psi_2(3,0) = 0.0879101291\ldots$. This is represented in Figure 10. The corresponding mean $\sum_{j=0}^{3} j\Psi_2(j,0)$ is given by $1.217341127\ldots$

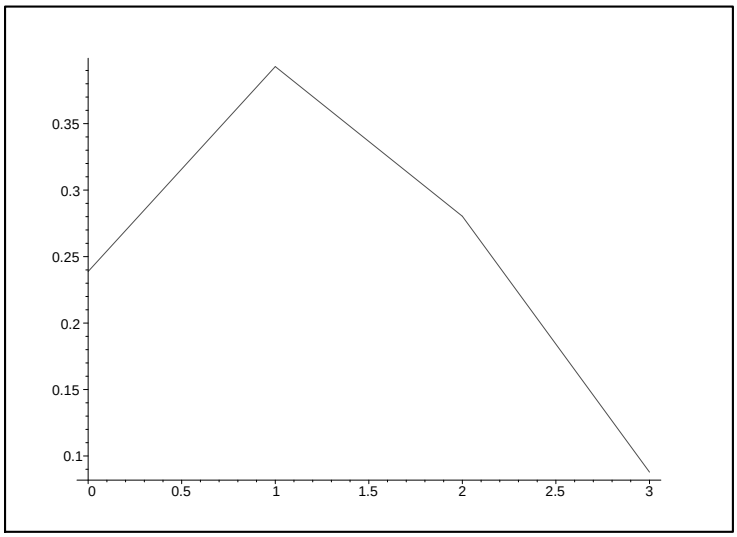

FIGURE 10. $\Psi_2(j,0)$, $j = 0..k$

The functions $\Psi_2(0,\theta)$ and

$$\Omega(\theta) := \sum_{j=1}^{k} j\Psi_2(j,\theta) \tag{1}$$

are given in Figures 11 and 12. The oscillations are of order .003 and .001 respectively.

4.2. **Mellin transform related to the probability distribution of J.** We first set

$$g_2(j,y) = \Pi\left(ye^{-La}, j\right) \sum_{v=0}^{k-j} \left[\Pi\left(y\left(\frac{q}{p} - e^{-La}\right), v\right) - \Pi\left(y\left(\frac{1}{p} - e^{-La}\right), v\right)\right].$$

This leads to

$$g_2^*(j,s) = \int_0^{\infty} y^{s-1} g_2(j,y)dy.$$

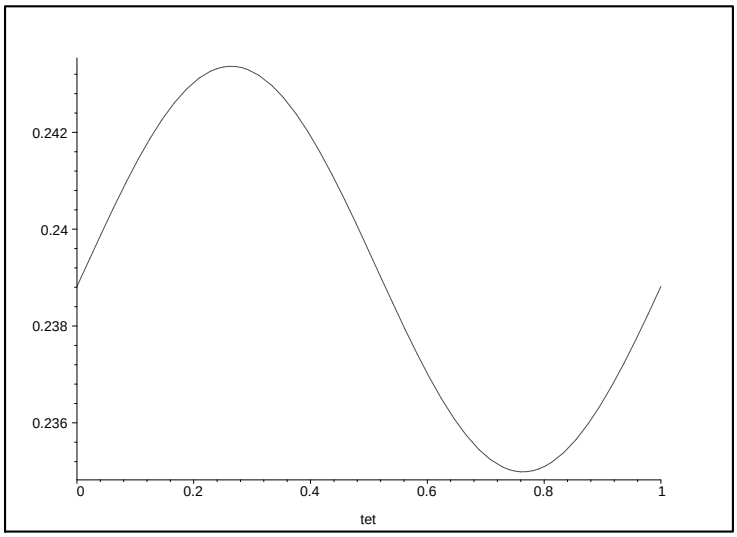

FIGURE 11. $\Psi_2(0,\theta)$, $\theta = 0..1$

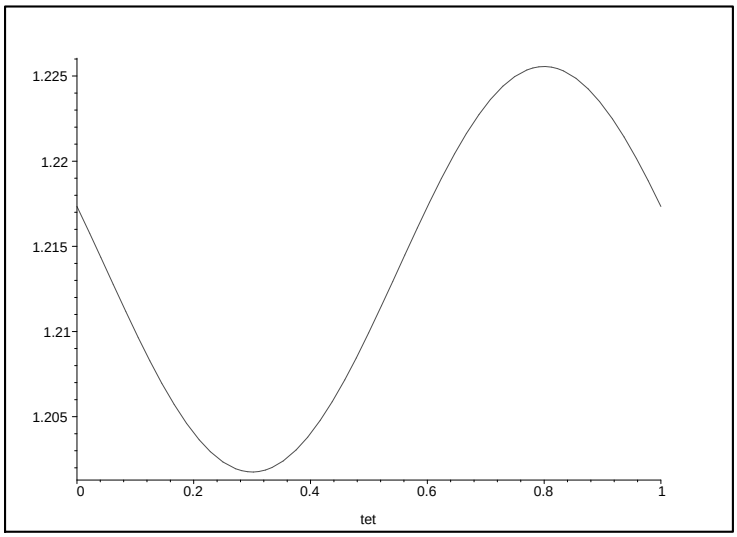

FIGURE 12. $\Omega(\theta) := \sum_{j=0}^{k} j\Psi_2(j,\theta)$, $\theta = 0..1$

Again, for $j = 0$, $v = 0$, we must we must avoid the singularity at $y = 0$. So we must consider separately

$$R_{2,1}(s) = \int_0^\infty y^{s-1}\Pi\left(ye^{-La},0\right)\left[\Pi\left(\frac{q}{p}y - ye^{-La},0\right) - \Pi\left(\frac{1}{p}y - ye^{-La},0\right)\right]dy,$$

$$R_{2,3}(j,s,v) = \int_0^\infty y^{s-1}\Pi\left(ye^{-La},j\right)\Pi\left(\frac{q}{p}y - ye^{-La},v\right)dy,$$

$$R_{2,4}(j,s,v) = \int_0^\infty y^{s-1}\Pi\left(ye^{-La},j\right)\Pi\left(\frac{1}{p}y - ye^{-La},v\right)dy.$$

The fundamental strip is $\langle -1, \infty\rangle$.

Actually, we only need the following expressions

$$\begin{aligned}
R_{2,1}(0) &= L,\\
R_{2,3}(j,s,v) &= \frac{e^{-Laj}\left(\frac{q-pe^{-La}}{p}\right)^{v}\left(\frac{q}{p}\right)^{-s-j-v}\Gamma(s+j+v)}{\Gamma(v+1)\Gamma(j+1)},\\
R_{2,4}(j,s,v) &= \frac{e^{-Laj}\left(\frac{1-pe^{-La}}{p}\right)^{v}\left(\frac{1}{p}\right)^{-s-j-v}\Gamma(s+j+v)}{\Gamma(v+1)\Gamma(j+1)},\\
R_{2,3}(j,0,l) &= \frac{e^{-Laj}\left(\frac{q-pe^{-La}}{p}\right)^{v}\left(\frac{q}{p}\right)^{-j-v}\Gamma(j+v)}{\Gamma(v+1)\Gamma(j+1)},\\
R_{2,4}(j,0,l) &= \frac{e^{-Laj}\left(\frac{1-pe^{-La}}{p}\right)^{v}\left(\frac{1}{p}\right)^{-j-v}\Gamma(j+v)}{\Gamma(v+1)\Gamma(j+1)},
\end{aligned}$$

We obtain

$$\mathbb{P}(\mathbf{J}=j)=\tilde{P}(j)+w_4(\{\log n^*\})+\mathcal{O}(n^{-C}),$$

with

$$\tilde{P}(0)=\frac{1}{L}\left(R_{2,1}(0)+\sum_{v=1}^{k}\left[R_{2,3}(0,0,v)-R_{2,4}(0,0,v)\right]\right).$$

For $j>0$, we derive

$$\tilde{P}(j)=\frac{1}{L}\left(\sum_{v=0}^{k-j}\left[R_{2,3}(j,0,v)-R_{2,4}(j,0,v)\right]\right).$$

As usual, again, we must have $\sum_{j=0}^{k}\tilde{P}(j)=1$.

A direct confirmation of this is very similar to a previous computation and therefore omitted.

We have

$$\tilde{M}_2=\sum_{j=1}^{k}j\tilde{P}(j).$$

Using the available information we got for $\tilde{P}(j)$, this can be simplified. Let us start with the constant term:

$$\sum_{j=1}^{k}j\sum_{v=0}^{k-j}\left[R_{2,3}(j,0,v)-R_{2,4}(j,0,v)\right]=\frac{p}{q}Q^{-a}k-pQ^{-a}k,$$

Consequently, the constant contribution to the average $\tilde{M}_2$ is given by

$$\frac{1}{L}(pQ^{1-a}k-pQ^{-a}k)=\frac{p^2Q^{1-a}k}{L}.$$

In our example, this gives $\tilde{P}(0) = .2391781864\ldots$, $\tilde{P}(1) = .3945618595\ldots$, $\tilde{P}(2) = .2706894194\ldots$, $\tilde{P}(3) = .0865705345\ldots$. This leads to $\tilde{M}_2 = 1.213652302\ldots$. The periodic component of $\mathbb{P}(\mathbf{J} = j)$ is asymptotically given by $w_4(\{\log n^*\})$, with

$$w_4(j,\theta) = \frac{1}{L}\sum_{v\neq 0}\sum_{u=0}^{k-j}\left[R_{2,3}(j,\chi_v,u) - R_{2,4}(j,\chi_v,u)\right]e^{-2\pi \mathbf{i} v\theta}.$$

We must have $\sum_{j=0}^{k} w_4(j,\theta) = 0$. Here is the computation:

$$\sum_{j=0}^{k}\sum_{u=0}^{k-j}\left[R_{2,3}(j,\chi,u) - R_{2,4}(j,\chi,u)\right]$$

$$= \sum_{j=0}^{k}\sum_{u=0}^{k-j}\left[\frac{Q^{-aj}\left(\frac{q-pQ^{-a}}{p}\right)^u\left(\frac{q}{p}\right)^{-\chi-u-j}\Gamma(\chi+u+j)}{u!j!} - \frac{Q^{-aj}\left(\frac{1-pQ^{-a}}{p}\right)^u\left(\frac{1}{p}\right)^{-\chi-u-j}\Gamma(\chi+u+j)}{u!j!}\right].$$

Now

$$\sum_{j=0}^{k}\sum_{u=0}^{k-j}\frac{Q^{-aj}\left(\frac{q-pQ^{-a}}{p}\right)^u\left(\frac{q}{p}\right)^{-\chi-u-j}\Gamma(\chi+u+j)}{u!j!} = p^{\chi}\sum_{l=0}^{k}\left(\frac{p}{q}\right)^l\frac{\Gamma(\chi+l)}{l!}\sum_{j=0}^{l}Q^{-aj}\left(\frac{q-pQ^{-a}}{p}\right)^{l-j}\binom{l}{j}$$

$$= p^{\chi}\sum_{l=0}^{k}\frac{\Gamma(\chi+l)}{l!} = p^{\chi}\frac{\Gamma(\chi+k+1)}{k!\chi}.$$

Similarly,

$$\sum_{j=0}^{k}\sum_{u=0}^{k-j}\frac{Q^{-aj}\left(\frac{1-pQ^{-a}}{p}\right)^u\left(\frac{1}{p}\right)^{-\chi-u-j}\Gamma(\chi+u+j)}{u!j!} = p^{\chi}\sum_{l=0}^{k}p^l\frac{\Gamma(\chi+l)}{l!}\sum_{j=0}^{l}Q^{-aj}\left(\frac{1-pQ^{-a}}{p}\right)^{l-j}\binom{l}{j}$$

$$= p^{\chi}\sum_{l=0}^{k}\frac{\Gamma(\chi+l)}{l!} = p^{\chi}\frac{\Gamma(\chi+k+1)}{k!\chi},$$

and since these two expression coincide, the formula is proved.

A comparison of $\tilde{P}(0) + w_4(0,\theta)$ and $\Psi_2(0,\theta)$, is given in Figure 13.

To analyze the periodic component of the mean we start from

$$G(s) = \sum_{j=1}^{k} j\sum_{v=0}^{k-j}\left[R_{2,3}(j,s,v) - R_{2,4}(j,s,v)\right] = \sum_{v=0}^{k}\left[A_1(s,v) - A_2(s,v)\right],$$

where

$$A_1(s,v) = \sum_{j=0}^{k-v} jR_{2,3}(j,s,v),$$

$$A_2(s,v) = \sum_{j=0}^{k-v} jR_{2,4}(j,s,v).$$

This leads to

$$w_5(\theta) = \frac{1}{L}\sum_{v\neq 0} G(\chi_v)e^{-2\pi \mathbf{i} v\theta}.$$

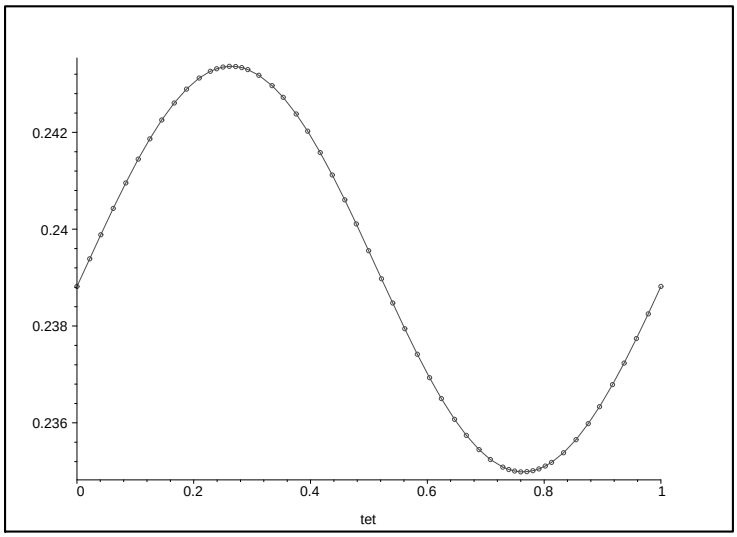

FIGURE 13. $\Psi_2(0,\theta)$ (line), $\tilde{P}(0)+w_4(0,\theta)$ (circles), $\theta = 0..1$

Figure 14 gives a comparison between $\tilde{M}_2 + w_5(\theta)$ and $\Omega(\theta)$ as given by (1).

In very much the same style the Fourier coefficients $G(\chi_v)$ of the periodic component of the average $\tilde{M}_2$ are computed as

$$G(\chi_v) = \sum_{j=1}^{k} j \sum_{u=0}^{k-j} [R_{2,3}(j,\chi_v,u) - R_{2,4}(j,\chi_v,u)] = p^{2+\chi_v} Q^{1-a} \frac{\Gamma(\chi_v + k + 1)}{(k-1)!(\chi_k + 1)}.$$

Note carefully that k cannot be 0 here, as there are no element *larger* than the maximum!

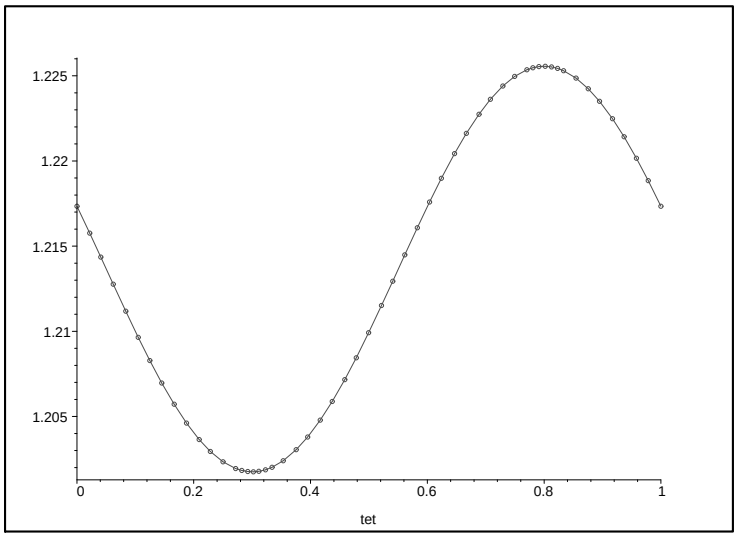

FIGURE 14. $\Omega(\theta)$ (line), $\tilde{M}_2 + w_5(\theta)$ (circles), $\theta = 0..1$

We also have $\tilde{M}_2 + w_5(0) = 1.217341127\ldots$ which is identical to the value given in Section 4.1.

5. At the level of the $(k+1)$st winner

We now analyze $\mathbb{P}(\mathbf{I}=i)$. We observe that $i\geq 1$, $i\geq k+1-l$ or $l\geq k+1-i$. So we have

$$\mathbb{P}(\mathbf{I}=i))\sim F_3(i,\eta)=\sum_{m=1}^{\infty}\sum_{l=k+1-i}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\Pi\left(e^{-L\eta},i\right),$$

with the convention $k+1-i=0$ if $i>k+1$. We derive another form

$$\Psi_3(i,\theta)=\sum_{u=-\infty}^{\infty}\sum_{l=k+1-i}^{k}\Pi\left(\frac{q}{p}e^{-L(u-\theta)},l\right)\Pi\left(e^{-L(u-\theta)},i\right).$$

We also have, with the notations of Section 3,

$$g_3(i,y)=\sum_{l=k+1-i}^{k}\Pi\left(\frac{q}{p}y,l\right)\Pi(y,i),$$

and

$$g_3^*(i,s)=\sum_{l=k+1-i}^{k}T_{1,2}(1,s,i,l),$$
$$g_3^*(i,0)=\sum_{l=k+1-i}^{k}T_{1,2}(1,0,i,l),$$

with fundamental strip $\langle -1,\infty\rangle$.

This gives

$$\mathbb{P}(\mathbf{I}=i)\sim\tilde{P}(i)+w_6(i,\{\log n^*\}),$$

with

$$\tilde{P}(i)=\frac{1}{L}g_3^*(i,0),$$
$$w_6(i,\theta)=\frac{1}{L}\sum_{v\neq 0}\sum_{l=k+1-i}^{k}T_{1,2}(1,\chi_v,i,l)e^{-2\pi\mathbf{i}v\theta}.$$

As before, we can explicitly compute that $\sum_{j\geq 0}\tilde{P}(j)=1$, but we don't include that, as it is similar to previous computations.

In our example, Figure 15 gives $\tilde{P}(i)$, $i=1..20$.

The generating function is given by

$$\begin{aligned}\mathbb{E}(z^{\mathbf{I}})\sim F_3(z,\eta)&=\sum_{m=1}^{\infty}\sum_{i=1}^{\infty}z^i\sum_{l=k+1-i}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\Pi\left(e^{-L\eta},i\right)\\&=\sum_{m=1}^{\infty}\sum_{l=0}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\sum_{i=k+1-l}^{\infty}z^i\Pi\left(e^{-L\eta},i\right)\\&=\sum_{m=1}^{\infty}\sum_{l=0}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\left[\exp\left(e^{-L\eta}(z-1)\right)-\sum_{i=0}^{k-l}z^i\Pi\left(e^{-L\eta},i\right)\right].\end{aligned}$$

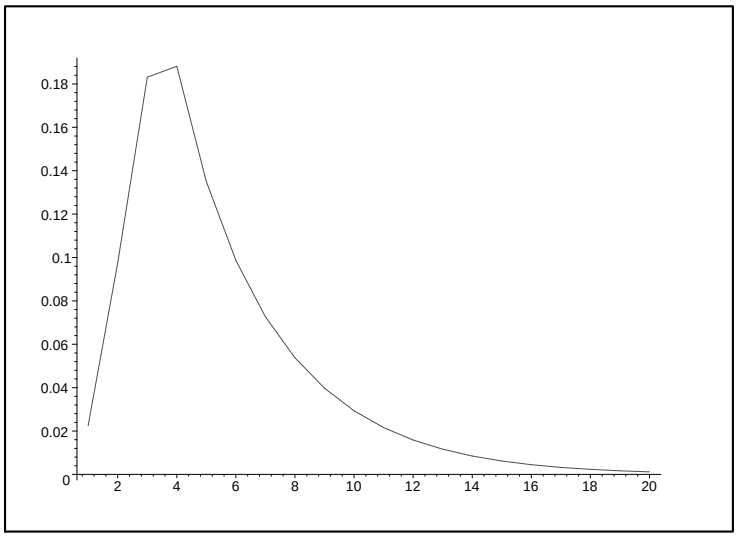

FIGURE 15. $\tilde{P}(i)$, $i = 1..20$

This doesn't seem to be managable. So we are modest and confine ourselves to the mean:

$$\begin{aligned} M_3 &\sim \sum_{m=1}^{\infty}\sum_{l=0}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\sum_{i=k+1-l}^{\infty} i\Pi\left(e^{-L\eta},i\right) \\ &= \sum_{m=1}^{\infty}\sum_{l=0}^{k}\Pi\left(\frac{q}{p}e^{-L\eta},l\right)\left[e^{-L\eta}-\sum_{i=1}^{k-l} i\Pi\left(e^{-L\eta},i\right)\right]. \end{aligned}$$

This leads to

$$\Psi_4(\theta) = \sum_{u=-\infty}^{\infty}\sum_{l=0}^{k}\Pi\left(\frac{q}{p}e^{-L(u-\theta)},l\right)\left[e^{-L(u-\theta)}-\sum_{i=1}^{k-l} i\Pi\left(e^{-L(u-\theta)},i\right)\right].$$

We also obtain

$$g_4(y) = \sum_{l=0}^{k}\Pi\left(\frac{q}{p}y,l\right)\left[y-\sum_{i=1}^{k-l} i\Pi(y,i)\right],$$

and

$$\begin{aligned} g_4^*(s) &= \sum_{l=0}^{k}\left[T_{1,1}(1,s+1,l)-\sum_{i=1}^{k-l} iT_{1,2}(1,s,i,l)\right], \\ g_4^*(0) &= \sum_{l=0}^{k}\left[T_{1,1}(1,1,l)-\sum_{i=1}^{k-l} iT_{1,2}(1,0,i,l)\right] = \sum_{l=0}^{k}\left[\frac{p}{q}-\sum_{i=1}^{k-l} iT_{1,2}(1,0,i,l)\right]. \end{aligned}$$

So

$$M_3 \sim \tilde{M}_3 + w_7(\{\log n^*\}),$$

with

$$\tilde{M}_3 = \frac{1}{L}g_4^*(0),$$

$$w_7(\theta)=\frac{1}{L}\sum_{v\neq 0}\sum_{l=0}^{k}\left[T_{1,1}(1,\chi_v+1,l)-\sum_{i=1}^{k-l}iT_{1,2}(1,\chi_v,i,l)\right]e^{-2\pi\mathbf{i}v\theta}.$$

We compute

$$\sum_{l=0}^{k}\left[\frac{p}{q}-\sum_{i=1}^{k-l}iT_{1,2}(1,0,i,l)\right]=\frac{p}{q}(k+1)-\sum_{l=0}^{k}\sum_{i=1}^{k-l}p^iq^l\frac{(i+l-1)!}{(i-1)!l!}=\frac{p}{q}(pk+1),$$

with the same technique as several times before. Likewise,

$$w_7(\theta)=\frac{1}{L}\sum_{v\neq 0}\frac{p^{\chi_v+1}}{q}\frac{\Gamma(\chi_v+k+1)}{k!(\chi_v+1)}(pk+1+\chi_v)e^{-2\pi\mathbf{i}v\theta}.$$

In our example, we derive $\tilde{M}_3=5.461435359\ldots$.

6. Conclusion

The approach of this paper, as already exemplified in our earlier papers [11, 10, 12] has the advantage of "relatively" simple computations, since we move to approximations at a very early stage. Of course, the computations are still heavy enough, but with "Rice's method" (as advocated for in [9]), they would be more cumbersome.

The techniques we used in this paper can lead to the joint distribution of our three random variables. With

$$\begin{aligned}\mathbf{I}&:=\text{ number of balls in urn } m,\\ \mathbf{J}&:=\text{ number of balls in urn } m+a,\\ \mathbf{U}&:=\text{ number of balls in urn } m-b,\end{aligned}$$

we have

$$\begin{aligned}&\mathbb{P}(\mathbf{J}=j,\mathbf{I}=i,\mathbf{U}=u)\\ &\sim\sum_{m=1}^{\infty}\sum_{l=\max(j,k+1-i)}^{k}\Pi\left(e^{-L(\eta+a)},j\right)\Pi\left(\frac{q}{p}e^{-L\eta}-e^{-L(\eta+a)},l-j\right)\Pi\left(e^{-L\eta},i\right)\Pi\left(e^{-L(\eta-b)},u\right).\end{aligned}$$

References

[1] N. Balakrishnan and A. Stepanov. A note on the number of observations near an order statistic. *Journal of Statistical Planning and Inference*, 134:1–14, 2005.

[2] Y. Baryshnikov, B. Eisenberg, and G. Stengle. A necessary and sufficient condition for the existence of the limiting probability of a tie for first place. *Statistics and Probability Letters*, 23:203–209, 1995.

[3] J.J.A.M. Brands, F.W Steutel, and R.J.G. Wilms. On the maximum and its uniqueness for geometric random samples. *Statistics and Probability Letters*, 20:209–217, 1994.

[4] F.T. Bruss and R. Grübel. On the multiplicity of the maximum in a discrete random sample. *Annals of Applied Probability*, 13(4):1252–1263, 2003.

[5] F.T. Bruss and C.A. O'Cinneide. On the number of maxima in a discrete sample. *Journal of Applied Probability*, 27:598–610, 1990.

[6] B. Eisenberg, G. Stengle, and G. Strang. The asymptotic probability of a tie for first place. *Statistics and Probability Letters*, 3:731–745, 1993.

[7] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[8] P. Hitczenko and G. Louchard. Distinctness of compositions of an integer: a probabilistic analysis. *Random Structures and Algorithms*, 19(3,4):407–437, 2001.

[9] P. Kirschenhofer and H. Prodinger. The number of winners a in discrete geometrically distributed sample. *Annals of Applied Probability*, 6:687–694, 1996.

[10] G. Louchard and H. Prodinger. The number of gaps in sequences of geometrically distributed random variables. 2005. http://www.ulb.ac.be/di/mcs/louchard/gaps3.ps.
[11] G. Louchard and H. Prodinger. Asymptotics of the moments of extreme-value related distribution functions. *Algorithmica*, 2006. To appear. Long version: http://www.ulb.ac.be/di/mcs/louchard/moml.ps.
[12] G. Louchard, H. Prodinger, and M.D. Ward. The number of distinct values of some multiplicity in sequences of geometrically distributed random variables. *Discrete Mathematics and Theoretical Computer Science*, AD:231–256, 2005. 2005 International Conference on Analysis of Algorithms.
[13] A. Pakes and F.W. Steutel. On the number of records near the maximum. *Australian Journal of Statistics*, 39:179–192, 1997.
[14] Y. Qi and R.J.G. Wilms. The limit behavior of maxima modulo one and the number of maxima. *Statistics and Probability Letters*, 34:75–84, 1997.
[15] W. Szpankowski. *Average case Analysis of Algorithms on Sequences*. Wiley, 2001.

Guy Louchard, Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium.
E-mail address: louchard@ulb.ac.be

Helmut Prodinger, Mathematics Department, Stellenbosch University, 7602 Stellenbosch, South Africa.
E-mail address: hproding@sun.ac.za