Injecting unique minima into random sets and applications to "Inverse Auctions"

F.Thomas Bruss[*] Guy Louchard [†] Mark Daniel Ward[‡]

February 5, 2007

Abstract

Consider N balls that are distributed among V urns according to some distribution G. We do not see the outcome and now have to place one ball into one urn with the goal of maximizing the probability that it will be the left-most urn containing a *single* ball. How should we proceed?

This is the urn-model translation of an interesting problem posed by an internet-auction offered by a German real-estate company. In the real problem only V is known (upper-price limit), whereas neither G (the way in which participants choose their offer) nor N (number of offers) is known. We would like to make an offer in such a way to maximize the probability that it turns out to be the minimum of the random set of single offers. We face a two-sided problem. On the one side we would like to choose a model which is convincing in terms of the expected behaviour of participants. On the other side, we want to solve an optimization problem; that is, the model should also be tractable and allow for asymptotic expansions, leading to a computable algorithm. Our attack is based on arguing that G should be (essentially) geometric and that some information on $\mathbb{E}(N)$ (expectation of N) and $\mathbb{V}(N)$ (variance of N) can be obtained in practice. Under certain conditions on possible dependencies of G and N, we can give answers. Poissonization (namely, changing the number N ofballs from a fixed quantity into a random quantity with Poisson distribution and mean N) and dePoissonization (i.e. reconciling with the original model) play here an important role to make the answers explicit.

Keywords: Urn model, unique minimum, Poisson approximation, asymptotic independence, asymptotic expansions, dominant terms, game theory, internet- auctions.

1 Introduction

Let $X_1, X_2, \dots$ be a sequence of discrete-valued nonnegative integers, and let N be a nonnegative integer-valued random variable (RV) with distribution function F. Given $N = n$, we suppose that $X_1, \dots, X_n$ are independent, identically distributed random variables (iid RV) with discrete distribution function G_n on the positive integers. Our objective is to fix an additional value $X = X_{n+1}$ (say) which has maximum probability of being the smallest among those values $X_1, X_2, \dots, X_{n+1}$, which are unique. Hence if

[*]Université Libre de Bruxelles, Département de Mathématique , CP 210 Boulevard du Triomphe, B-1050 Bruxelles, Belgium, `tbruss@ulb.ac.be`

[†]Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, `louchard@ulb.ac.be`

[‡]University of Pennsylvania, Department of Mathematics, 209 South 33rd Street, Philadelphia, PA 19104–6395, USA. `ward2@math.upenn.edu`

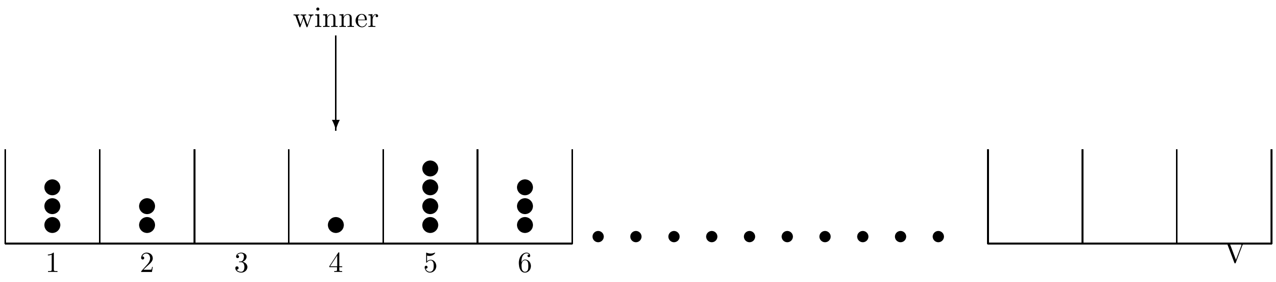


Figure 1: The number of balls in urn K denotes the number of submitted offers of K cents. V is the upper offer limit (in cents). The offer of 4 cents is here the winner because it is the smallest *single* offer. Urns close to V are naturally very likely to be empty.

$$A_n = \left\{ X_k : \sum_{j=1}^{n} [\![X_j = X_k]\!] = 1 \right\},$$

then our goal is to find a value X in the support of G_n such that

$$X = \arg \max_{\{X \in \text{ supp } G_n\}} \mathbb{P}\left(X < \min A_n, \sum_{j=1}^{n} [\![X_j = X]\!] = 0 \right),$$

and to find an algorithm to compute efficiently an asymptotic value of X.

Motivation

There are concrete applications behind this problem. The first we mention is an internet-auction which one may call “inverse auctions”, and which attracts a great deal of interest (see e.g. traumhause.de on the internet; traumhaus means dreamhouse). A house is put on auction (with photos and an upper price estimation V). Offers can be made in form of any nonnegative Euro/cents amount $\leq V$. Thus, for instance, €0 or €1.47 are admissible offers but €1.471 or $V + 1$ are not. In the version we have seen on the net, V was €$350,000.00$ and only the true buyer had to pay. After a bidding period of several months, the bidder with the *smallest unique offer* gets the deal. The description of the rules become somewhat less transparent if all offers are bid at least twice. This is why we concentrate on the true objective to make the smallest single offer which we want to activate with the highest possible probability.

Similar auctions or lotteries are now advertized in many places and seem to attract a lot of attention. The Karlsberg brewery for instance gives for each purchase of a case of their new “Urmild” beer the right to make an offer for a sports car, and the smallest single offer (if any) wins the car. Here the offers must be made in integer values of Euros. This shows that the advertizers have an adequate intuition about the right relationship between the number of “urns and balls” to make the problem intellectually interesting.

The sports car lottery example of Karlsberg brewery is likely to be an excellent advertizing campaign; moreover, it may directly increase sales since each purchase gives the right to make a new offer. Hence the motivation is here very clear. In the house auction example, the value of the prize is much higher and so one may wonder how such an auction can possibly be rewarding for the party who offers it. This party is a real estate company, and advertising aspects may again play a central role. Moreover, offers can be made by pay-phone which seems to be used frequently in particular since several offers can be made by the same person. Each participant is allowed to make more than one offer by calling or in writing, but not more than *one* per day.

Other versions of such auctions seem to us even more interesting. Unlike in the house auction where only the actual buyer of the house has to pay the money he or she offered we now assume that any offer which is made must be paid immediately with some entrance fee (with no return if it is not successful). This is like buying a ticket in a lottery and the offers can be expected to be smaller than in the first version, but now ticket prizes vary, and the buyer has full control of what ticket he buys! We see a good chance of this type of auction to be profitable or even very profitable for the one who offers it, and it may be only a question of time until it appears on the internet. This is why we try to be before our time and look also at this version and refer to it as version 2.

Novelty of the problem

We are not aware of any closely related problems in the literature. The so-called "unique maximum problem for i.i.d. random variables" which has attracted a great deal of interest (see e.g. Bruss and Grübel [2] for references) may sound somewhat similar but is a true maximum problem. Unlike our problems studied here, it has no strategic component to inject an extreme value in a random set.

The paper is organized as follows. Section 2 proposes an urn model, where each urn corresponds to one offer. We are led to a geometric distribution (with parameter p) for the distribution of offers. Two cases are analyzed: one where the participants want to have a certain win probability $P_1(v) = 1 - \eta$, say, the other one where we consider the event that urn K is empty, and all urns before K *do not contain* exactly one ball. The probability of this event is denoted by $P_2(K)$ and we want to maximize it. Section 3 proves the unicity of the maximum. Section 4 assumes that the participant has to pay immediately his offer in order to make it active. Section 5 is devoted to the case where p depends on the number n of participants. Section 6 considers another type of dependence of p on n. Section 7 concludes the paper.

2 Proposing a model

The difficulty of the essence of the problem (in both versions) is the fact that only the upper price limit V is known. To get explicit answers we need the distribution of the number of offers and also of the amount of offers, without which the optimization problem is not well defined. Hence we must first discuss aspects of the model. We confirm here our interest to the house auction. The reasoning would be similar for the sports car auction or offers of this type.

Modelling aspects

We first note that, in order to make the real-word problem meaningful, the distribution of offers G and of the number of offers F cannot be chosen independently. A trivial counterexample is $G(1) = \mathbb{P}(\text{ offer } \le 1) = 1$, in which case more than 1 participant would not try to make offers. Hence it is probable that nobody would in the end make offers. A similar reasoning shows that we should allow for no gaps, that is, the support of G should be $[1, V]$, that is, G should be stricly increasing to 1 on $[1, V]$. Indeed, suppose that the hypothesis of certain gaps was justified. Then there would be a smallest urn number for which it were justified. But then, individually, everybody had an interest to place there an offer which contradicts the hypothesis.

Now first some thoughts on N. By observing previous auctions, it should be possible to estimate $\mathbb{E}(N)$. One feels that modelling N by a binomial RV —and hence in the limit by a normal RV— should be the first choice. Indeed, if we know that approximately n people can read the auction on the internet, we may estimate the success parameter p of the binomial distribution by $p \simeq \mathbb{E}(N)/n$. However we do not want to be too restrictive and therefore ask only for certain moments conditions for N to hold.

Now back to modeling offer sizes. We think that somebody who offers €100.17, say, would not be less willing to increase (upon advice) his offer to €100.89 , say as somebody else would be to increase from €1.17 to €1.89, say. Hence we believe that the memoryless-property should generally hold true except, possibly, near the upper price limit V. These three conditions support assuming G to be (truncated) geometric, and this will be our first method of attack.

We will model the possible offers by urns, and the number of offers by the total number of balls in those V urns.

We first look at the case where the success parameter of the geometric distribution GEOM($p(n)$) is fixed.

The case of fixed p

Let N denote the number of offers with mean m, variance σ^2, $\sigma^2 = o(m^2)$. Distribution of offers: geometric, pq^{i-1}. More generally we assume that the tail is of the form Cq^i. We will also assume that m is large: all our asymptotics are computed as $m \to \infty$.

Set

$$K := \text{ index of the urn where we want to throw our ball (i.e. place our offer),}$$

and let

$$\begin{aligned}
N^* &:= Np/q, \\
m^* &:= mp/q, \\
\sigma^{*2} &:= \sigma^2 p^2/q^2, \\
Q &:= 1/q, \\
L &:= \ln(Q) = -\ln(q), \\
\log &:= \log_Q,
\end{aligned}$$

and finally

$$N^* = m^* + U = m^* \left(1 + \frac{U}{m^*}\right).$$

Of course, $\mathbb{E}(U) = 0, \mathbb{V}(U) = \sigma^{*2}$ by definition.

We now formulate the moments conditions described beforehand and assume that the central moments $\mu_k := \mathbb{E}(U^k)$ are such that $\mu_k = o(m^{k-2})\sigma^2$. Also, for simplicity, we suppose that U is distributed on integers, with distribution $\rho(u)$. Set $\pi := \frac{p}{q}q^K$. The probability

$P_1(K|N=n)$ that *no offer* coincides with K is given by $(1-\pi)^n$, conditioned on n offers. Set

$$n^* = np/q = m^* + u = m^*(1+\frac{u}{m^*}) = m^*\left(1+\varepsilon\right),$$

with

$$\varepsilon := u/m^*,$$

and also

$$K = \log m^* + v.$$

This gives

$$m^* q^K = e^{-Lv} = E, \text{ say},$$

and

$$n\pi = E(1+\varepsilon).$$

Also, the absolute probability of no offer coinciding with K is

$$P_1(K) = P_1(v) = \sum_u \rho(u)(1-\pi)^n. \tag{1}$$

We now use the by Poisson approximation to estimate $P_1(K)$. (See, for instance Barbour et al. [1]).

We define

$$\begin{aligned} P_0(v) := (1-\pi)^n &\sim \exp\left[-E - E\varepsilon - \frac{pE^2}{2qm^*} + \mathcal{O}\left(\frac{1}{m^{*2}}\right) + \mathcal{O}\left(\frac{\varepsilon}{m^*}\right)\right] \\ &\sim e^{-E}\left[1 - \varepsilon E - \frac{pE^2}{2qm^*} + \frac{\varepsilon^2 E^2}{2} + \mathcal{O}\left(\frac{1}{m^{*2}}\right) + \mathcal{O}\left(\frac{\varepsilon}{m^*}\right)\right], \end{aligned} \tag{2}$$

where $\mathcal{O}()$ are functions of E. We obtain of course the Gumbel distribution as the dominant term. So

$$\begin{aligned} P_1(v) &\sim \sum_u \rho(u) \exp\left[-e^{-Lv}\right]\left[1 - e^{-Lv}\frac{u}{m^*} + \frac{e^{-2Lv}}{2}\left[\left(\frac{u}{m^*}\right)^2 - \frac{p}{qm^*}\right] + \dots\right] \\ &= \exp\left[-e^{-Lv}\right] + 0 + \exp\left[-e^{-Lv}\right]\frac{e^{-(2Lv)}}{2}\left[\frac{\sigma^{*2}}{m^{*2}} - \frac{p}{qm^*}\right] + \dots. \end{aligned} \tag{3}$$

This is the starting point for Version 0. We give one specific example.

2.1 Version 0

We assume in this Version 0 that participants want to have a certain win probability $P_1(v) = 1-\eta$, say. Assume $\sigma^2 = O(m)$ and let first

$$\sigma^{*2} = \beta m^*.$$

Set

$$v = \tilde{v} + \gamma_0/m^* + \dots,$$

with $\exp(-e^{-L\tilde{v}}) = 1-\eta$, or $e^{-L\tilde{v}} = -\ln(1-\eta)$. Then we obtain after straightforward computation

$$\tilde{v} = -\ln[-\ln(1-\eta)]/L,$$

and

$$\gamma_0 = \frac{\ln(1-\eta)}{2L}\left[\beta - \frac{p}{q}\right]$$

.

Now we must find v_n such that

$$P_1(v_n) = 1-\eta.$$

We only give one example. The choice $q = 1/2, \beta = 1$ (hence $\gamma_0 = 0$), $\eta = 0.2$ and a Gaussian distribution for N, with mean $m = 1000$, gives, from (1), a numerical value $v_n = 2.18\ldots$, and $\tilde{v} = 2.16\ldots$, $\tilde{K} = \lfloor \log m^* + \tilde{v} \rfloor = 12$.

As a comparison, what is the probability that a uniformly placed offer on $\{1, 2, \ldots, V\}$ would be minimal among the unique offers?

Optimal strategies for uniform hypothesis

Recall that V (upper price limit) is known. We want to compute our optimal offer (that is, the one which maximizes the probability to win) under the hypothesis that the number of offers N is known ($N = n$) and that these offers are uniformly distributed on $\{1, 2, \ldots, V\}$. Our offer is the offer number $n+1$ and we can choose it as we want. Note that if different people make the same given offer, then these events must be seen as different events. Hence we have to use the classical urn model of distributing distinguishable balls in (distinguishable) urns. The probability that an offer K which we make is single equals $\left(\frac{V-1}{V}\right)^n$ (independently of K);

$$\begin{aligned}
&\mathbb{P}(\text{ no urn before } K \text{ is singly occupied})\\
&= 1 - \mathbb{P}(\exists \text{ a single-occupied urn before } K)\\
&= 1 - \mathbb{P}(S_1 \cup S_2 \cup \ldots \cup S_{K-1}) \text{ where } S_j := \{j\text{th urn is singly-occupied}\}\\
&= 1 - \sum_1^{K-1} \mathbb{P}(S_j) + \sum_{j_1<j_2} \mathbb{P}(S_{j_1} \cap S_{j_2}) - \ldots + (-1)^{K-1}\mathbb{P}(\cap_1^{K-1} S_j)\\
&= \sum_0^{K-1} (-1)^i \binom{K-1}{i} \frac{n!}{(n-i)!} \left(\frac{1}{V-1}\right)^i \left(\frac{V-1-i}{V-1}\right)^{n-i}.
\end{aligned} \tag{4}$$

So it is evident that we have to offer $K = 1$: the probability of all preceding occupied urns to be ocupied by at least 2 balls is decreasing in the number of preceding urns. (We can also verify this argument by showing the (4) is decreasing in K).

This comparison helps us to understand the bidder's behaviour trying to optimize their offer. They want to make small offers to increase the probability of their offer to be the smallest single offer. However, surmizing that all bidders would reason similarly, the conclusion is to avoid values which are too small. Thus, as n increases, the concentration around 1 must decrease.

2.2 Version 1

Consider the event that urn K is empty, and all urns before K *do not contain* exactly one ball. The probability of this event is denoted by $P_2(K)$. We will use asymptotic independence of urns, as far as fixed numbers of balls are concerned, as proved in Louchard, Prodinger and Ward, [5]. We must carefully study the effect of the dispersion of U around its mean 0. We have, with

$$\pi(i) := \frac{p}{q}q^{K-i}, \pi = \pi(0),$$

$$P_2(v) \sim \sum_u \rho(u)(1-\pi)^n \left\{ \prod_{i=1}^{\infty} \left[1 - n\pi(i)(1-\pi(i))^{n-1}\right] \right\}, \tag{5}$$

The product should go to K, but, as proved in [5], the error is exponentially negligible.

The numerical optimal value of (5), $q = 1/2, m = 1000$ is given by $v_n = 0.55\ldots$, with $P_2(v_n) = 0.263\ldots$. A plot of $P_2(v)$, using (5), is given in Figure 2.

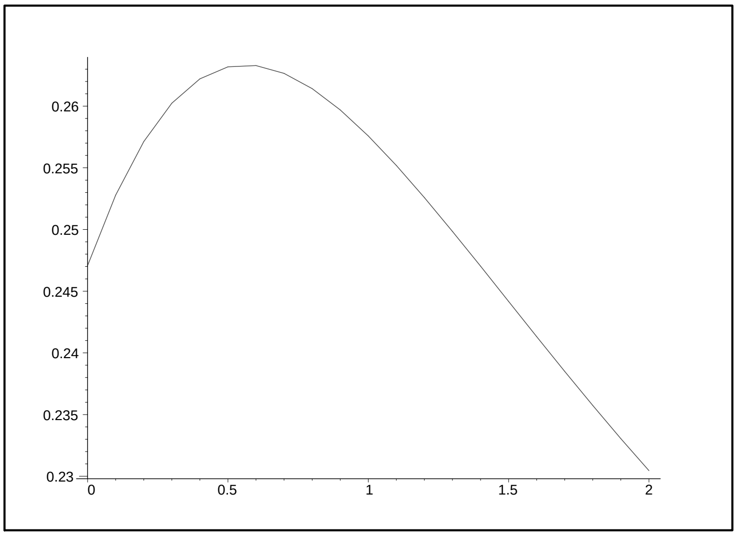


Figure 2: $P_2(v)$. This graph presents the probability that the Kth urn is empty and all non-empty urns with number smaller than K contain more than one ball. Here K is scaled to $K = \log m^* + v$.

Now (we drop the $i-$dependence, to ease the notation)

$$\begin{aligned} n\pi &= E(1+\varepsilon), \\ (1-\pi)^{n-1} &\sim \exp\left[-E - E\varepsilon - \frac{pE^2}{2qm^*} + \frac{pE}{qm^*} + \mathcal{O}\left(\frac{1}{m^{*2}}\right) + \mathcal{O}\left(\frac{\varepsilon}{m^*}\right)\right] \\ &\sim e^{-E}\left[1 - \varepsilon E - \frac{pE(-2+E)}{2qm^*} + \frac{\varepsilon^2E^2}{2} + \mathcal{O}\left(\frac{1}{m^{*2}}\right) + \mathcal{O}\left(\frac{\varepsilon}{m^*}\right)\right]. \end{aligned}$$

Note that the $1/m^*$ term is different from the one in (2). This leads to

$$P_2(v) \sim \exp\left[-e^{-Lv}\right] \sum_u \rho(u) \left[1 - e^{-Lv}\varepsilon + \frac{e^{-2Lv}}{2}\varepsilon^2 - \frac{pe^{-2Lv}}{2qm^*} + \ldots\right] \times$$

$$\times \left\{ \prod_{i=1}^{\infty} \left(1 - \exp\left[-e^{-L(v-i)} \right] \left[1 - e^{-L(v-i)}\varepsilon + \frac{e^{-2L(v-i)}}{2}\varepsilon^2 - \frac{pe^{-L(v-i)}}{2qm^*} \left[-2 + e^{-L(v-i)} \right] + \dots \right] \right. \right.$$

$$\left. \left. e^{-L(v-i)}[1+\varepsilon] \right) \right\}. \tag{6}$$

This shows that the dispersion of K around $\log m^*$ is $\mathcal{O}(1)$. Set

$$\begin{aligned}
\alpha_1(i,v) &:= -\exp\left[-e^{-L(v-i)} \right], \\
\alpha_2(i,v) &:= -e^{-L(v-i)}, \\
\alpha_3(i,v) &:= \frac{e^{-2L(v-i)}}{2}, \\
\alpha_4(i,v) &:= e^{-L(v-i)}, \\
\alpha_5 &= 1, \\
\alpha_6 &= 0, \text{ but this will be } \neq 0 \text{ in Sec. 5}, \\
\alpha_7(v) &:= -e^{-Lv}, \\
\alpha_8(v) &:= \frac{e^{-2Lv}}{2}, \\
\alpha_9(i,v) &:= -\frac{pe^{-2L(v-i)}}{2q}[-2 + e^{-L(v-i)}], \\
\alpha_{10}(v) &:= -\frac{pe^{-2Lv}}{2q}, \\
\alpha_{11} &= 0, \text{ Again, this will be } \neq 0 \text{ in Sec. 5.}
\end{aligned}$$

The bracketed term within $P_2(v)$ (see (6)) becomes

$$\left\{ \prod_{i=1}^{\infty} \left(1 + \alpha_1(i,v) \left[1 + \alpha_2(i,v)\varepsilon + \alpha_3(i,v)\varepsilon^2 + \frac{\alpha_9(i,v)}{m^*} + \dots \right] \alpha_4(i,v) \left[1 + \alpha_5\varepsilon + \alpha_6\varepsilon^2 + \frac{\alpha_{11}}{m^*} + \dots \right] \right) \right\},$$

This can be written as

$$\left\{ \prod_{i=1}^{\infty} \left(A_0(i,v) + A_1(i,v)\varepsilon + A_2(i,v)\varepsilon^2 + \frac{A_3(i,v)}{m^*} + \dots \right) \right\},$$

with

$$\begin{aligned}
A_0(i,v) &= 1 + \alpha_1(i,v)\alpha_4(i,v), \\
A_1(i,v) &= \alpha_1(i,v)\alpha_2(i,v)\alpha_4(i,v) + \alpha_1(i,v)\alpha_4(i,v)\alpha_5, \\
A_2(i,v) &= \alpha_1(i,v)\alpha_4(i,v)\alpha_6 + \alpha_1(i,v)\alpha_2(i,v)\alpha_4(i,v)\alpha_5 + \alpha_1(i,v)\alpha_3(i,v)\alpha_4(i,v), \\
A_3(i,v) &= \alpha_1(i,v)\alpha_4(i,v)[\alpha_9(i,v) + \alpha_{11}].
\end{aligned}$$

Set

$$B_k(i,v) := A_k(i,v)/A_0(i,v),$$

and

$$D_0(v) := \prod_{i=1}^{\infty} A_0(i,v).$$

The bracketed term becomes

$$\begin{aligned}&\left\{D_0(v)\prod_{i=1}^{\infty}\left[1+B_1(i,v)\varepsilon+B_2(i,v)\varepsilon^2+\frac{B_3(i,v)}{m^*}+\dots\right]\right\}\\&=\left\{D_0(v)\left[1+S_1\varepsilon+\left[S_2+S_1^2/2-S_{1,2}/2\right]\varepsilon^2+\frac{S_3}{m^*}+\dots\right]\right\},\end{aligned}$$

with

$$\begin{aligned}S_1(v)&=\sum_1^{\infty}B_1(i,v),\\S_2(v)&=\sum_1^{\infty}B_2(i,v),\\S_{1,2}(v)&=\sum_1^{\infty}B_1^2(i,v),\\S_3(v)&=\sum_1^{\infty}B_3(i,v).\end{aligned}$$

All these sums are easily shown to converge. So we get finally

$$P_2(v) \sim F_0(v) + \frac{F_1(v)}{m^*},$$

with

$$F_0(v) = \exp\left[-e^{-Lv}\right] D_0(v), \tag{7}$$

$$F_1(v) = \exp\left[-e^{-Lv}\right] D_0(v)\left\{\beta\left[S_2+S_1^2/2-S_{1,2}/2+S_1(v)\alpha_7(v)+\alpha_8(v)\right]+\left[\alpha_{10}(v)+S_3(v)\right]\right\}. \tag{8}$$

To obtain $\max_v P_2(v)$, we first compute $\tilde{v}$ as the solution of $F_0'(\tilde{v}) = 0$. With our usual choice of parameters (see Section 2.1), a plot of $F_0(v)$ is given in Figure 3. This leads to $\tilde{v} = 0.5613032851\dots, F_0(\tilde{v}) = 0.2642452648\dots$, $\tilde{K} = \lfloor \log m^* + \tilde{v} \rfloor = 10$.

A comparison between $P_2(v)$ and $F_0(v)$ is shown in Figure 4, where it seems that $F_0(v)$ dominates $P_2(v)$ on $[0,2]$.

Then we set

$$\bar{v} = \tilde{v} + \frac{\gamma_1}{m^*} + \dots,$$

and

$$P_2'(\bar{v}) = 0.$$

This leads to

$$F_0'(\bar{v}) + \frac{F_1'(\bar{v})}{m^*} = 0,$$

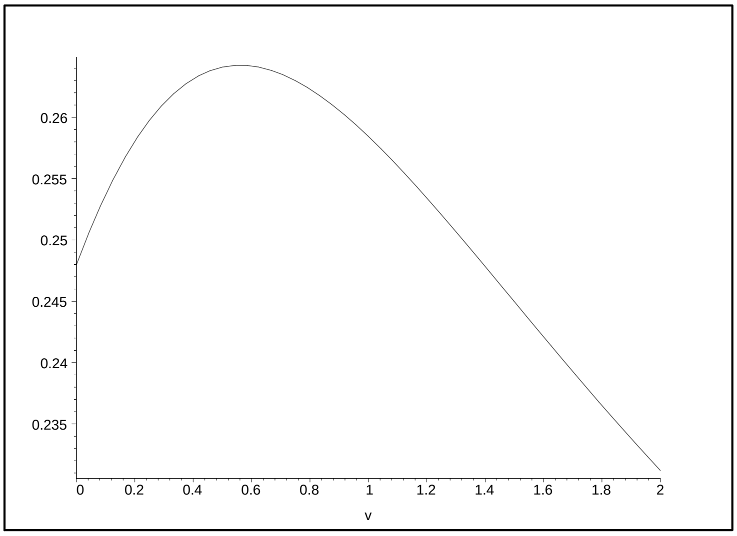


Figure 3: $F_0(v)$

or

$$\gamma_1 = -\frac{F_1'(\tilde{v})}{F_0''(\tilde{v})},$$

$$P_2(\bar{v}) \sim F_0(\bar{v}) + \frac{F_1(\bar{v})}{m^*} \sim F_0(\tilde{v}) + \frac{F_1(\tilde{v})}{m^*}.$$

To summarize, the algorithm works as follows.

Algorithm 1

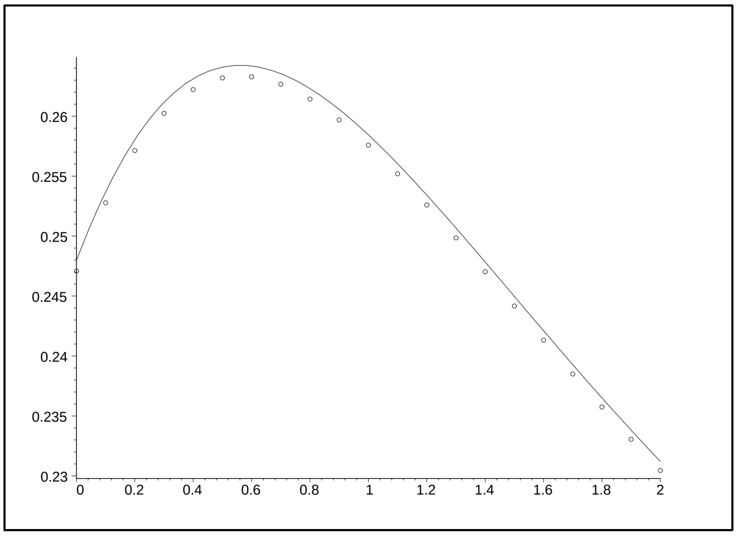
Input: p, m, β
Output: second order optimal value for $K : \bar{K}$
Solve $F_0'(\tilde{v}) = 0$;
Compute $\gamma_1 = -\frac{F_1'(\tilde{v})}{F_0''(\tilde{v})}$;
Compute $\bar{K} = \lfloor \log m^* + \tilde{v} + \frac{\gamma_1}{m^*} \rfloor$;
End

With our choice of parameters, we compute $\gamma_1 = 1.07903\ldots$. As K must be an integer, we see that the correction in our example is practically negligible .

3 Uniqueness of the Maximum

Please note that we cannot assure so far, that the the candidate for this maximum is unique. Let us work with the simple case $L = 1$, so

$$F_0(v) = \exp[-e^{-v}] \prod_{i=1}^{\infty} \left[1 - \exp\left[-e^{-(v-i)}\right] e^{-(v-i)}\right].$$



circle : $P_2(v)$
line : $F_0(v)$

Figure 4: A comparison between $P_2(v)$ and $F_0(v)$

We will deal with the logarithm, ie

$$LF_0(v) = -e^{-v} + \sum_{i=1}^{\infty} \ln\left[1 - \exp\left[-e^{-(v-i)}\right] e^{-(v-i)}\right], \tag{9}$$

and observe the following

i) First we observe numerically that $LF_0(v)$ possesses a maximum at $\tilde{v} = 0.7983134948\ldots$ with $LF_0(\tilde{v}) = -1.024695735\ldots$. A plot of $LF_0(v), v \in [-0.2..4]$ is given in Figure 5. We have used a $\sum_1^{30}$ summation.

ii) Next, we see that an excellent approximation for LF_0 for $v \in [0..4]$ is given by $i = 1..5$. Indeed,

$$\left|\sum_{6}^{\infty} \ln\left[1 - \exp\left[-e^{-(v-i)}\right] e^{-(v-i)}\right]\right| \leq \left|\sum_{6}^{\infty} \ln\left[1 - \exp\left[-e^{-(4-i)}\right] e^{-(4-i)}\right]\right| = 0.0045767767\ldots$$

This justifies using $\sum_1^{30}$ summation in(9) for numerical computations.

iii) Also, for $v < 0$, we have

$$LF_0(v) \leq LF_0(0) = -1.202264688\ldots$$

iv) For $v \geq 4$, we see that we can practically limit the sum to

$$\sum_{\max\{\lfloor v \rfloor - 13, 1\}}^{\lfloor v \rfloor + 2} \ln\left[1 - \exp\left[-e^{-(v-i)}\right] e^{-(v-i)}\right],$$

which entails an asymptotic periodicity.

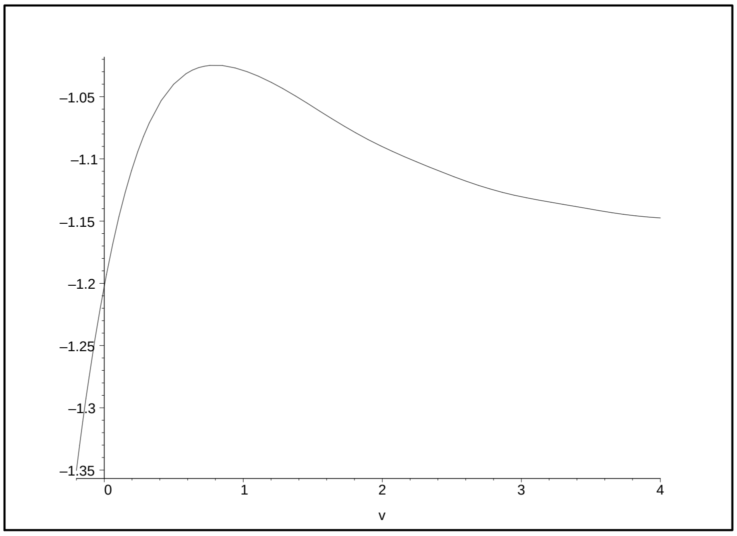


Figure 5: $LF_0(v), v \in [-0.2..4]$

Indeed, for $\lfloor v \rfloor \leq 13$, the lower bound of the summation remains 1. Next set $j = \lfloor v \rfloor - i$. For $\lfloor v \rfloor > 13$,

$$\left| \sum_{j=14}^{\lfloor v \rfloor -1} \ln \left[1 - \exp\left[-e^{-j}\right] e^{-j}\right] \right| \leq \left| \sum_{14}^{\infty} \ln \left[1 - \exp\left[-e^{-j}\right] e^{-j}\right] \right| = 1.31 \ldots 10^{-4},$$

and, for any v,

$$\sum_{j=-\infty}^{-3} \ln \left[1 - \exp\left[-e^{-j}\right] e^{-j}\right] = -3.80 \ldots 10^{-6}.$$

Again, this justifies using $\sum_1^{30}$ summation . A plot of $LF_0(v), v \in [4..12]$ is given in Figure 6, showing the asymptotic periodicity for large v. We could analyze the periodicity in detail with Mellin transforms (see, for instance Flajolet et al. [3], or Szpankowski [6]) but we will not pursue this matter further on here.

v) We conclude that the maximum (unique or not) occurs for some $v \in [0..1]$.
vi) A similar analysis of $LF_0''(v), v \in [0,1]$ shows that it is strictly < 0 in this range, proving the unicity. We have uniform convergence of the righthand side of (9) on $[0,1]$, so we can differentiate term by term. A plot of $LF_0''(v), v \in [0..1]$ is given in Figure 7.
vii) Finally, the effect of $F_1(v)/m^*$, for large m^*, does not destroy the maximum unicity.

4 Version 2

Here the participant has to pay immediately his offer in order to make it active. Thus the expected gain by offering K is $P_2(v)V - K$, so that the objective is to find

$$\max_v [P_2(v)V - K] = \max_v [P_2(v)V - v - \log m^*],$$

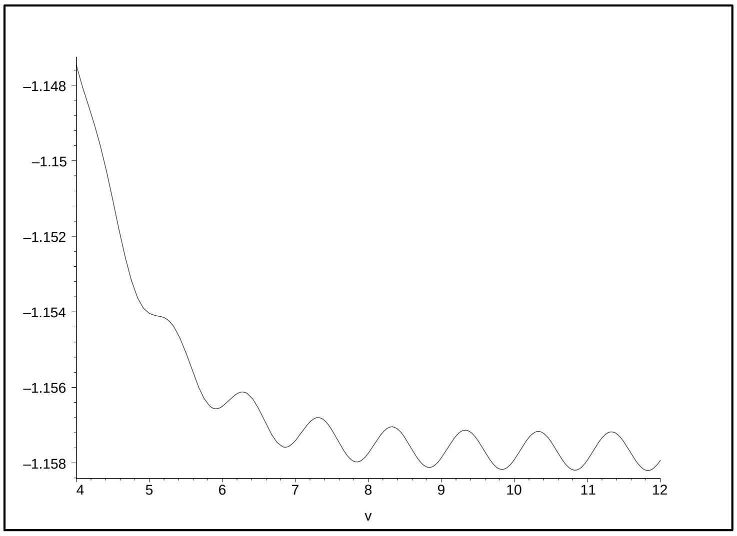


Figure 6: $LF_0(v), v \in [4..12]$

with v and m^* as before. Set

$$v = \tilde{v} + \gamma_2/V + \dots.$$

This leads, with the dominant term (7), to

$$\gamma_2 = \frac{1}{F_0''(\tilde{v})} = -12.71\dots$$

Again, as K must be an integer, we see that the correction in our example is practically negligible with our choice of parameters.

5 Version 3. p,q depend on n

We now let p and q depend on n, where our problem stays as before. We could reason as follows. The mean offer is given by $K = q/p$. The participant wants to specify the probability that the offer $K = q/p$ is unique. We consider the case where this probability is given be some real, $1/2$ say. This means that q must depend on n. Set $q = e^{-\varepsilon}$. This leads to

$$\pi = \frac{p}{q} q^K,$$
$$K \sim 1/\varepsilon,$$
$$1/2 \sim \left(1 - \varepsilon e^{-K\varepsilon}\right)^n,$$

or

$$1/2 \sim \exp\left[-n\varepsilon e^{-1}\right].$$

Hence

$$\varepsilon \sim \frac{\ln(2)}{e^{-1}n}.$$

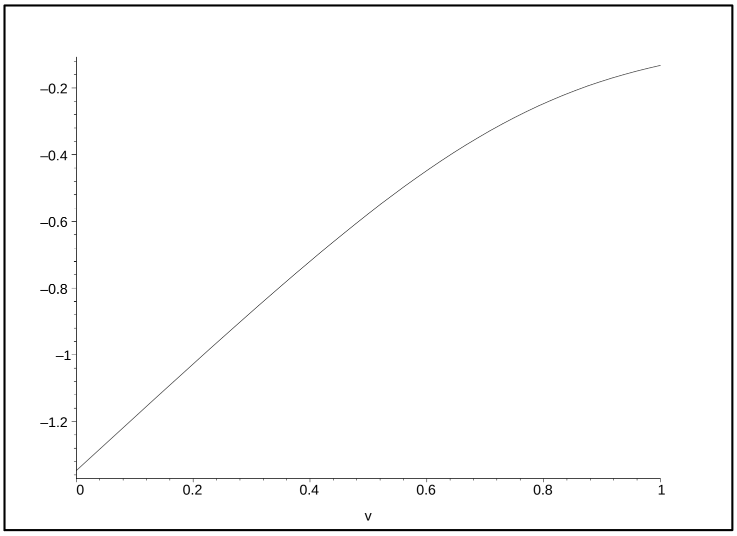


Figure 7: $LF_0''(v), v \in [0..1]$

So we choose $q(n) = e^{-C/n}$, for some constant C, $p(n) = 1 - q(n)$. Now we set $N = m + U, \mathbb{V}(U) = \sigma^2$. The probability $P_1(K)$ that urn K is empty is given by

$$\sum_u \rho(u)(1-\pi)^n.$$

Set $n = m + u = m\left(1 + \frac{u}{m}\right)$. This gives then

$$\pi = \frac{p(n)}{q(n)} e^{-Cv/(1+\varepsilon)},$$

if we set $K = mv$ (the scale is of course different from Sec.2.2). and $\varepsilon = \frac{u}{m}$. Set

$$P_0(v) := (1-\pi)^n.$$

We have

$$\begin{aligned} P_0(v) &\sim \exp\left[-Ce^{-Cv}\left[1 + Cv\varepsilon + (-Cv + \frac{(Cv)^2}{2})\varepsilon^2 + \frac{C[1+e^{-Cv}]}{2m} + \dots\right]\right] \\ &\sim \exp\left[-Ce^{-Cv}\right]\left[1 - C^2 v\varepsilon e^{-Cv} + \left[C^2 v e^{-Cv} - \frac{1}{2}C^3 v^2 e^{-Cv} + \frac{1}{2}C^4 v^2 e^{-2Cv}\right]\varepsilon^2\right. \\ &\left. -\frac{C^2 e^{-Cv}\left[1+e^{-Cv}\right]}{2m} + \dots\right]. \end{aligned}$$

Set

$$\begin{aligned} \alpha_7(v) &:= -C^2 v e^{-Cv}, \\ \alpha_8(v) &:= C^2 v e^{-Cv} - \frac{1}{2}C^3 v^2 e^{-Cv} + \frac{1}{2}C^4 v^2 e^{-2Cv}, \\ \alpha_{10}(v) &:= -\frac{C^2 e^{-Cv}\left[1+e^{-Cv}\right]}{2}. \end{aligned}$$

5.1 Asymptotic dependence of the urns

We first consider the model in which the balls are independent, and each ball has a GEOM($p(n)$) distribution, with $p(n) = 1 - e^{-C/n}$. In other words, the probability that a ball lands in the ith urn is exactly $p(n)q(n)^{i-1}$ (usually abbreviated as $\pi(i)$), where $q(n) = e^{-C/n}$. If k balls are placed into the urns, then let $X_i(k, n) = 1$ if exactly one ball lands in the ith urn, and $X_i(k, n) = 0$ otherwise. In other words,

$$X_i(k, n) := [\![\text{ value } i \text{ appears among the } k \text{ GEOM } p(n) \text{ RVs exactly once }]\!].$$

Define $X(k, n) = \sum_{i=1}^{\infty} X_i(k, n)$. So $X(k, n)$ denotes the number of urns that each contain exactly one ball (when starting with k balls, and each ball has GEOM($p(n)$) distribution).

In order to obtain the asymptotics of $X(n, n)$, we consider a similar model, where each ball has a GEOM($p(n)$) distribution, but there is a random number of balls placed into the urns, which is Poisson distributed with mean τ. Let $N_\tau = \text{Poisson}(\tau)$ denote the random number of balls placed into the urns. Then the urns are all independent, and the ith urn contains a Poisson number of balls with mean $\tau\pi(i)$. In analogy to our notation above, we define

$$\widetilde{X}_i(\tau, n) := [\![\text{ value } i \text{ appears among the } N_\tau \text{ GEOM } (p(n)) \text{ RVs exactly once }]\!].$$

Then define $\widetilde{X}(\tau, n) = \sum_{i=1}^{\infty} \widetilde{X}_i(\tau, n)$. So $\widetilde{X}(\tau, n)$ denotes the number of urns that each contain exactly one ball (when starting with N_τ balls, and each ball has GEOM($p(n)$) distribution).

We first consider $\widetilde{M}_n(\tau) := \mathbb{E}(\widetilde{X}(\tau, n))$. We compute

$$\widetilde{M}_n(\tau) = \sum_{k=0}^{\infty} P(N_\tau = k)\mathbb{E}(\widetilde{X}(\tau, n) \mid N_\tau = k) = \sum_{k=0}^{\infty} \frac{e^{-\tau}\tau^k}{k!}\mathbb{E}(X(k, n))$$

and also

$$\widetilde{M}_n(\tau) = \mathbb{E}\left(\sum \widetilde{X}_i(\tau, n)\right) = \sum_{i=1}^{\infty} \mathbb{E}(\widetilde{X}_i(\tau, n)) = \sum_{i=1}^{\infty} \tau\pi(i)e^{-\tau\pi(i)}.$$

In summary,

$$\sum_{k=0}^{\infty} \frac{e^{-\tau}\tau^k}{k!}\mathbb{E}(X(k, n)) = \widetilde{M}_n(\tau) = \sum_{i=1}^{\infty} \tau\pi(i)e^{-\tau\pi(i)}.$$

Now we consider $\widetilde{U}_n(\tau) := \mathbb{E}(\widetilde{X}^2(\tau, n))$. Similar reasoning yields

$$\sum_{k=0}^{\infty} \frac{e^{-\tau}\tau^k}{k!}\mathbb{E}(X^2(k, n)) = \widetilde{U}_n(\tau) = \sum_{i=1}^{\infty} \tau\pi(i)e^{-\tau\pi(i)}\left(1 + \sum_{j\neq i} \tau\pi(j)e^{-\tau\pi(j)}\right).$$

5.2 Diagonal DePoissonization

In this section, we compute $\mathbb{E}(X(n, n))$ and $\mathbb{E}(X^2(n, n))$ by first randomizing the number of balls using Poissonization. Of course, we must precisely compare the randomized model to the original model.

5.2.1 Diagonal dePoissonization Theorem

We first recall the Diagonal dePoissonization Theorem from Jacquet and Szpankowski,

Diagonal DePoissonization Lemma. (Jacquet and Szpankowski, [4])
Let $\widetilde{F}_n(\tau)$ be a sequence of Poisson transforms of $f_{k,n}$ which is assumed to be a sequence of entire functions of τ. Consider a polynomial cone $\mathcal{C} := \{\tau = x + iy \ : \ |y| \leq x\}$. Let the following two conditions hold for some $A > 0$, B, and $\alpha > 0$, β, and γ:
(I) For $\tau \in \mathcal{C}$ and $|\tau| \leq 2n$,

$$|\widetilde{F}_n(\tau)| \leq Bn^{\beta}\,,$$

(O) For $\tau \notin \mathcal{C}$ and $|\tau| = n$,

$$|\widetilde{F}_n(\tau)e^{\tau}| \leq n^{\gamma}\exp(n - An^{\alpha})\,.$$

Then, for large n,

$$f_{n,n} = \widetilde{F}_n(n) + O(n^{\beta-1})$$

and more generally, for every nonnegative integer m,

$$f_{n,n} = \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij} n^i \widetilde{F}_n^{\langle j\rangle}(n) + O(n^{\beta-m-1})$$

where $\widetilde{F}_n^{\langle j\rangle}(n)$ denotes the jth derivative of $\widetilde{F}_n(\tau)$ at $\tau = n$, and where the b_{ij} are defined by $B_j(x) = \sum_i b_{ij}x^i$ and $B_j(x) = [y^j](e^{-xy}(1+y)^x)$. (The relation of the coefficients b_{ij} to Poisson-Charlier polynomials and the Laguerre polynomials is also described briefly in [4].)

5.2.2 Comparison of Expectations

First we check (I) and (O) for $\widetilde{F}_n(\tau) = \widetilde{M}_n(\tau)$ and $f_{k,n} = \mathbb{E}(X(k,n))$. For $\tau \in \mathcal{C}$,

$$|\widetilde{M}_n(\tau)| = \left|\sum_{i=1}^{\infty}\tau\pi(i)e^{-\tau\pi(i)}\right| \leq \sum_{i=1}^{\infty}|\tau|\pi(i)e^{-\Re(\tau)\pi(i)} \leq |\tau|\sum_{i=1}^{\infty}\pi(i) = |\tau| \leq 2n.$$

For $\tau \notin \mathcal{C}$ with $|\tau| = n$,

$$|\widetilde{M}_n(\tau)e^{\tau}| = \left|e^{\tau}\sum_{i=1}^{\infty}\tau\pi(i)e^{-\tau\pi(i)}\right| \leq |e^{\tau}|\sum_{i=1}^{\infty}|\tau|\pi(i)e^{-\Re(\tau)\pi(i)} \sim ne^{n/\sqrt{2}}.$$

Thus (I) and (O) are both satisfied (with $\beta = 1$) for $\widetilde{F}_n(\tau) = \widetilde{M}_n(\tau)$. So, by the Diagonal dePoissonization Lemma we conclude that, for every nonnegative integer m,

$$\mathbb{E}(X(n,n)) = \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij}n^i\, \partial_{\tau}^{j}\mathbb{E}(\widetilde{X}(\tau,n))\Big|_{\tau=n} + O(n^{-m}).$$

In particular, when $m = 1$,

$$\mathbb{E}(X(n,n)) = \mathbb{E}(\widetilde{X}(n,n)) - \frac{1}{2}n\widetilde{M}_n''(n) + \frac{1}{3}n\widetilde{M}_n'''(n) + O(n^{-1}). \tag{10}$$

We note that $\widetilde{M}_n^{\langle j \rangle}(n) = \sum_{i=1}^{\infty}(-\pi(i))^j e^{-n\pi(i)}(n\pi(i) - j) = O(n^{-j+1})$. In particular, the $\frac{1}{3}n\widetilde{M}_n'''(n)$ term above is $O(n^{-1})$. Thus, (10) simplifies to

$$\mathbb{E}(X(n,n)) = \mathbb{E}(\widetilde{X}(n,n)) - \frac{1}{2}n\widetilde{M}_n''(n) + O(n^{-1}). \tag{11}$$

We use Euler-Maclaurin summation to compute

$$\mathbb{E}(\widetilde{X}(n,n)) = \sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)} = \int_1^{\infty} n\pi(i)e^{-n\pi(i)}\,di - \frac{1}{2}n\pi(i)e^{-n\pi(i)}\Big|_{i=1}^{\infty} + O(n^{-1}). \tag{12}$$

The integral evaluates to $\frac{1-e^{-C}}{C}n - \frac{1}{2}Ce^{-C} + O(n^{-1})$. Also, the Euler-Maclaurin correction term is $-\frac{1}{2}n\pi(i)e^{-n\pi(i)}\big|_{i=1}^{\infty} = \frac{1}{2}Ce^{-C} + O(n^{-1})$. So the mean in the Poisson model is

$$\mathbb{E}(\widetilde{X}(n,n)) = \frac{1-e^{-C}}{C}n + O(n^{-1}). \tag{13}$$

Next, we compute the correction term between $\mathbb{E}(X(n,n))$ and $\mathbb{E}(\widetilde{X}(n,n))$. We have

$$-\frac{1}{2}n\widetilde{M}_n''(n) = -\frac{1}{2}n\int_1^{\infty}(-\pi(i))^2 e^{-n\pi(i)}(n\pi(i)-2)\,di + O(n^{-1}) = \frac{1}{2}Ce^{-C} + O(n^{-1}). \tag{14}$$

By (11), we simply add (13) and (14) to obtain the expectation in the dependent model:

$$\mathbb{E}(X(n,n)) = \frac{1-e^{-C}}{C}n + \frac{1}{2}Ce^{-C} + O(n^{-1}). \tag{15}$$

In particular, $\mathbb{E}(X(n,n))$ and $\mathbb{E}(\widetilde{X}(n,n))$ both have leading terms $\frac{1-e^{-C}}{C}n$. On the other hand, $\mathbb{E}(X(n,n))$ also has $\frac{1}{2}Ce^{-C}$ as a constant term, but $\mathbb{E}(\widetilde{X}(n,n))$ *does not have* a $\Theta(1)$ term.

5.2.3 Comparison of Variances

Now we check conditions (I) and (O) for $\widetilde{F}_n(\tau) = \widetilde{U}_n(\tau)$ and $f_{k,n} = \mathbb{E}(X^2(k,n))$. We see that (by the same type of reasoning that we used above for $\widetilde{M}_n(\tau)$) for $\tau \in \mathcal{C}$, $|\widetilde{U}_n(\tau)| \le 4n^2 + 2n$, and for $\tau \notin \mathcal{C}$, $|\widetilde{U}_n(\tau)e^{\tau}| \sim e^{n/\sqrt{2}}n^2$. Since (I) and (O) are both satisfied (with $\beta = 2$) for $\widetilde{F}_n(\tau) = \widetilde{U}_n(\tau)$, then by the Diagonal dePoissonization Lemma we conclude that, for every nonnegative integer m,

$$\mathbb{E}(X^2(n,n)) = \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij}n^i\,\partial_\tau^j \mathbb{E}(\widetilde{X}^2(\tau,n))\Big|_{\tau=n} + O(n^{1-m}).$$

In particular, when $m = 1$,

$$\mathbb{E}(X^2(n,n)) = \mathbb{E}(\widetilde{X}^2(n,n)) - \frac{1}{2}n\widetilde{U}_n''(n) + \frac{1}{3}n\widetilde{U}_n'''(n) + O(1). \tag{16}$$

For ease of notation, we define $R_n(\tau) := \sum_{i=1}^{\infty}(\tau\pi(i))^2 e^{-2\tau\pi(i)}$. We note $\int_1^{\infty}(n\pi(i))^2 e^{-2n\pi(i)}\,di = \frac{1-e^{-2C}-2Ce^{-2C}}{4C}n - \frac{1}{2}C^2e^{-2C} + O(n^{-1})$. Additionally, the Euler-Maclaurin correction term is $-\frac{1}{2}\,(\tau\pi(i))^2e^{-2\tau\pi(i)}\big|_{i=1}^{\infty} = \frac{1}{2}C^2e^{-2C}$. Thus

$$R_n(n) = \left(\frac{1-e^{-2C}-2Ce^{-2C}}{4C}\right)n + O(n^{-1}). \tag{17}$$

Also $R''_n(n) = O(n^{-1})$ and $R'''_n(n) = O(n^{-2})$.

We compute

$$\mathbb{E}(\widetilde{X}^2(n,n)) = \sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)} \left(1 + \sum_{j\neq i} n\pi(j)e^{-n\pi(j)}\right) = \widetilde{M}_n(n)^2 + \widetilde{M}_n(n) - R_n(n). \quad (18)$$

From (13), it follows that the second moment in the Poisson model is

$$\mathbb{E}(\widetilde{X}^2(n,n)) = \left(\frac{1-e^{-C}}{C}\right)^2 n^2 + \left(\frac{2Ce^{-2C} + 3 - 4e^{-C} + e^{-2C}}{4C}\right) n + O(1). \quad (19)$$

We also compute

$$\begin{aligned} -\frac{1}{2}n\widetilde{U}''_n(n) &= -n\widetilde{M}'_n(n)^2 - n\widetilde{M}_n(n)\widetilde{M}''_n(n) - \frac{1}{2}n\widetilde{M}''_n(n) + \frac{1}{2}nR''_n(n) \\ \frac{1}{3}n\widetilde{U}'''_n(n) &= 2n\widetilde{M}'_n(n)\widetilde{M}''_n(n) + \frac{2}{3}n\widetilde{M}_n(n)\widetilde{M}'''_n(n) + \frac{1}{3}n\widetilde{M}'''_n(n) - \frac{1}{3}nR'''_n(n) \end{aligned} \quad (20)$$

We again use Euler-Maclaurin summation:

$$\begin{aligned} \widetilde{M}'_n(n) &= -e^{-C} + O(n^{-1}) \\ \widetilde{M}''_n(n) &= -Ce^{-C}n^{-1} + O(n^{-2}) \\ \widetilde{M}'''_n(n) &= -C^2e^{-C}n^{-2} + O(n^{-3}) \end{aligned} \quad (21)$$

By (16), we simply add (19) and both parts of (20) to obtain the second moment in the dependent model:

$$\mathbb{E}(X^2(n,n)) = \left(\frac{1-e^{-C}}{C}\right)^2 n^2 + \left(\frac{-6Ce^{-2C} + 3 - 4e^{-C} + e^{-2C} + 4Ce^{-C}}{4C}\right) n + O(1). \quad (22)$$

In particular, $\mathbb{E}(X^2(n,n))$ and $\mathbb{E}(\widetilde{X}^2(n,n))$ both have leading terms $\left(\frac{1-e^{-C}}{C}\right)^2 n^2$, but the linear (i.e., $\Theta(n)$) terms are different in the dependent versus Poisson models.

It follows from (15) and (22) that the variance in the original (dependent) model is

$$\mathbb{V}(n,n) = \left(\frac{-2Ce^{-2C} + 3 - 4e^{-C} + e^{-2C}}{4C}\right) n + O(1). \quad (23)$$

5.3 A direct approach to the correction term (*without* Poissonization)

In this section, we want to verify our computations. So we compute $X(n,n)$ directly, this time *without* using Poissonization. (We simply perform a brute-force computation!) We use $\pi(i) := p(n)q(n)^{i-1}$ throughout the discussion below. First, we make a direct calculation of the mean. We compute

$$\mathbb{E}(X(n,n)) = \sum_{i=1}^{\infty} n\pi(i)(1-\pi(i))^{n-1} \sim \sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)} \left(1 + \frac{n\pi(i) - (n\pi(i))^2/2}{n}\right)$$

By using Euler-Maclaurin summation, we have $\sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)} = \frac{1-e^{-C}}{C}n + O(n^{-1})$. Similarly, $\sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)} \frac{n\pi(i)-(n\pi(i))^2/2}{n} \sim e^{-C}C/2$.

Thus

$$\mathbb{E}(X(n,n)) = \frac{1-e^{-C}}{C}n + \frac{1}{2}Ce^{-C} + O(n^{-1}),$$

which completely agrees with (15).

Now we turn our attention to the variance. We have

$$\mathbb{V}(X(n,n)) = \sum_{i=1}^{\infty}\sum_{j\neq i}\mathbb{E}(X_i(n,n)X_j(n,n)) + \mathbb{E}(X(n,n)) - \mathbb{E}(X(n,n))^2. \tag{24}$$

We just computed $\mathbb{E}(X(n,n))$, so we focus on computing the double sum. We have

$$\sum_{i=1}^{\infty}\sum_{j\neq i}\mathbb{E}(X_i(n,n)X_j(n,n)) = \sum_{i=1}^{\infty}\sum_{j\neq i}n(n-1)\pi(i)\pi(j)[1-\pi(i)-\pi(j)]^{n-2}, \tag{25}$$

which is asymptotically

$$\begin{aligned}&\sim \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}n(n-1)\pi(i)\pi(j)e^{-n\pi(i)-n\pi(j)}\\ &\quad\times\left(1+\frac{2n\pi(i)+2n\pi(j)-n\pi(i)n\pi(j)-\frac{1}{2}(n\pi(i))^2-\frac{1}{2}(n\pi(j))^2}{n}\right)\\ &\quad-\sum_{i=1}^{\infty}n(n-1)\pi(i)^2e^{-2n\pi(i)}\left(1+\frac{4n\pi(i)-2(n\pi(i))^2}{n}\right).\end{aligned} \tag{26}$$

We use Euler-Maclaurin summation to analyze the right-hand side of (26). The double sum on the right-hand side of (26) simplifies to

$$\left(\frac{1-e^{-C}}{C}\right)^2 n^2 + (e^C-2)e^{-2C}n + O(1). \tag{27}$$

We also use Euler-Maclaurin summation to compute

$$\sum_{i=1}^{\infty}n(n-1)\pi(i)^2e^{-2n\pi(i)} = \left(\frac{1-e^{-2C}-2Ce^{-2C}}{4C}\right)n + O(1) \tag{28}$$

and

$$\sum_{i=1}^{\infty}n(n-1)\pi(i)^2e^{-2n\pi(i)}\left(\frac{4n\pi(i)-2(n\pi(i))^2}{n}\right) \sim \left(\frac{-2C^2-2C-1+4C^3+e^{2C}}{4C}\right)e^{-2C} = O(1). \tag{29}$$

Finally, equation (24) tells us exactly how to easily combine the equations (25)–(29). As a result, we obtain

$$\mathbb{V}(n,n) = \left(\frac{-2Ce^{-2C}+3-4e^{-C}+e^{-2C}}{4C}\right)n + O(1),$$

which completely agrees with (23).

5.4 Diagonal Exponential dePoissonization

In this section, we analyze the asymptotic dependence of the urns, by comparing the distributions of $X(n,n)$ and $\widetilde{X}(n,n)$. To do this, we first estimate $\mathbb{E}(z^{X(n,n)})$. We emphasize that we are unable to use the dePoissonization theorem from the previous sections, because $\mathbb{E}(z^{\widetilde{X}(\tau,n)})$ has exponential growth in terms of n. Due to this fast growth, we need to use a stronger version of dePoissonization.

For some reasons that will be clear later on (see Section 6), we will analyze the more general case $q = e^{-1/n^{\kappa}}, 0 < \kappa < 1$.

5.4.1 Diagonal Exponential dePoissonization Theorem

We adapt the Diagonal Exponential DePoissonization Theorem from Jacquet and Szpankowski.

Diagonal Exponential DePoissonization Theorem. (Jacquet and Szpankowski, [4]) Let $\widetilde{F}_n(\tau)$ be a sequence of Poisson transforms of $f_{k,n}$ which is assumed to be a sequence of entire functions of τ.

Consider a polynomial cone $\mathcal{C}(D) := \{\tau = x + iy \ : \ |y| \le Dx\}$, with $0 < D < 1$. Consider $\log \widetilde{F}_n(\tau)$ that exists in the polynomial cone $\mathcal{C}(D)$. Let the following conditions hold for some $A > 0$, $B > 0$, $\frac{1}{2} \le \beta < \frac{2}{3}$, and $\alpha > \beta$:

(I) For $\tau \in \mathcal{C}(D)$ and $(1-D)n \le |\tau| \le (1+D)n$,

$$|\log \widetilde{F}_n(\tau)| \le Bn^{\beta}\,;$$

(O) For $\tau \notin \mathcal{C}(D)$ and $|\tau| = n$,

$$|\widetilde{F}_n(\tau)e^{\tau}| \le \exp(n - An^{\alpha})\,.$$

Then, for all $\epsilon > 0$, we have

$$f_{n,n} = \widetilde{F}_n(n)\exp\left[-\frac{n}{2}(L_n'(n))^2\right](1 + O(n^{3\beta-2+\epsilon})),$$

where $L_n(\tau) = \log \widetilde{F}_n(\tau)$ and $L_n'(\tau) = \widetilde{F}_n'(\tau)/\widetilde{F}_n(\tau)$.

5.4.2 Comparison of the Distributions

Now we check (I) and (O) for $\widetilde{F}_n(\tau) = \mathbb{E}(z^{\widetilde{X}(\tau,n)})$ and $f_{k,n} = \mathbb{E}(z^{X(k,n)})$. We first note that

$$\sum_{k=0}^{\infty} \frac{e^{-\tau}\tau^k}{k!}\mathbb{E}(z^{X(k,n)}) = \widetilde{F}_n(\tau,z) = \prod_{i=1}^{\infty}\left[1 + (z-1)\tau\pi(i)e^{-\tau\pi(i)}\right]. \tag{30}$$

We fix $0 < \xi < 1$, and afterwards we consider only z such that $|1-z| < \xi$. In all of the calculations below, we are careful to make sure that the claims hold *uniformly* in terms of z with $|1-z| < \xi$. Also, for technical reasons that will be clear below, we choose $D > 0$ such that $\frac{1}{\sqrt{1+D^2}} + \xi < 1$.

First, we consider $\tau \in \mathcal{C}(D)$, and we check condition (I). All bounds and asymptotic estimates below hold *uniformly* for all $\tau \in \mathcal{D}(D)$. We compute

$$
\begin{aligned}
|\log \widetilde{F}_n(\tau, z)| &= \left|\log \prod_{i=1}^{\infty}\left[1 + (z-1)\tau\pi(i)e^{-\tau\pi(i)}\right]\right| \\
&\leq \pi/2 + \sum_{i=1}^{\infty} \log\left[1 + \xi|\tau|\pi(i)e^{-\Re(\tau)\pi(i)}\right] \\
&\leq \pi/2 + \xi\sum_{i=1}^{\infty} |\tau|\pi(i)e^{-\Re(\tau)\pi(i)} \qquad (31)
\end{aligned}
$$

Notice $\sum_{i=1}^{\infty} |\tau|\pi(i)e^{-\Re(\tau)\pi(i)}$ grows asymptotically no faster than

$$\sqrt{1+D^2}\, n^{\kappa}(1 - \exp(-\Re(\tau)(1 - e^{-1/n^{\kappa}}))).$$

We note that $(1-D)n \leq |\tau| \leq \Re(\tau)\sqrt{1+D^2}$, so $\frac{1-D}{\sqrt{1+D^2}}n \leq \Re(\tau)$. Thus, $|\log \widetilde{F}_n(\tau, z)|$ is asymptotically (ignoring lower-order terms) at most

$$n^{\kappa}\left(1 - \exp\left(-\frac{1-D}{\sqrt{1+D^2}} \cdot n^{1-\kappa}\right)\right).$$

So condition (I) holds for each β with $\kappa \leq \beta$. Since we also need $\frac{1}{2} \leq \beta < \frac{2}{3}$, we are specifically restricted to $\kappa < \frac{2}{3}$. We write $\beta = \max\{\frac{1}{2}, \kappa\}$.

Now we consider $\tau \notin \mathcal{C}(D)$ with $|\tau| = n$, and we check condition (O). We compute

$$
\begin{aligned}
|\widetilde{F}_n(\tau, z)e^{\tau}| &= \left|e^{\tau}\prod_{i=1}^{\infty}\left[1 + (z-1)\tau\pi(i)e^{-\tau\pi(i)}\right]\right| \\
&\leq \exp\left(\Re(\tau) + \xi n\sum_{i=1}^{\infty} \pi(i)e^{-\Re(\tau)\pi(i)}\right) \\
&\sim \exp\left(\Re(\tau) + \xi n\right) \qquad (32)
\end{aligned}
$$

We recall that $0 < \xi < 1$; also, $\Re(\tau) < \frac{1}{\sqrt{1+D^2}}n$ for $\tau \notin \mathcal{C}(D)$. Thus $\Re(\tau) + \xi n < \frac{1}{\sqrt{1+D^2}}n + \xi n$. Since $\frac{1}{\sqrt{1+D^2}} + \xi < 1$, then condition (O) is satisfied.

So conditions (I) and (O) in the Diagonal Exponential DePoissonization Theorem have been verified.

Now we explicitly compute the correction factor $\exp\left[-\frac{n}{2}\left(L'_{n,z}(n)\right)^2\right]$. We compute

$$
\begin{aligned}
L'_{n,z}(n) &= \sum_{j=1}^{\infty}(z-1)\pi(j)e^{-n\pi(j)}\left[1 - n\pi(j)\right]\left[1 + (z-1)n\pi(j)e^{-n\pi(j)}\right]^{-1} \\
&\sim -\frac{1}{n}\left(1 + n^{\kappa+1} - n^{\kappa}\ln\left(n(z-1)\exp(ne^{-1/n^{\kappa}})(e^{1/n^{\kappa}} - 1) + \exp(1/n^{\kappa} + n)\right)\right) \\
&\sim -\frac{1}{n}\left(1 + n^{\kappa+1} - n^{\kappa}\left(1/n^{\kappa} + n + \frac{n(z-1)\exp(ne^{-1/n^{\kappa}})(e^{1/n^{\kappa}} - 1)}{\exp(1/n^{\kappa} + n)}\right)\right) \\
&= n^{\kappa}(z-1)\exp(ne^{-1/n^{\kappa}} - 1/n^{\kappa} - n)(e^{1/n^{\kappa}} - 1)
\end{aligned}
$$

$$\sim (z-1)\exp(ne^{-1/n^\kappa} - 1/n^\kappa - n) \tag{33}$$

We note that $ne^{-1/n^\kappa} - 1/n^\kappa - n \sim -n^{1-\kappa} - 1/n^\kappa$, and thus

$$\exp\left(-\frac{n}{2}(L'_{n,z}(n))^2\right) \sim \exp\left(-\frac{n}{2}(z-1)^2 e^{-2n^{1-\kappa}}\right).$$

For $\kappa < \frac{2}{3}$ and $\beta = \max\{\frac{1}{2}, \kappa\}$, we conclude that

$$\mathbb{E}(z^{X(n,n)}) = \mathbb{E}(z^{\widetilde{X}(n,n)}) \exp\left[-\frac{n}{2}(L'_{n,z}(n))^2\right](1 + O(n^{3\beta-2+\epsilon})).$$

We emphasize that all of our bounds and estimates are *uniformly* valid for all z in the disk $|1-z| \le \xi$, and ξ does *not* depent on n; thus, for $\kappa < \frac{2}{3}$ and $\beta = \max\{\frac{1}{2}, \kappa\}$, we conclude that

$$\mathbb{E}(z^{X(n,n)}) = \mathbb{E}(z^{\widetilde{X}(n,n)}) \exp\left[-\frac{n}{2}(L'_{n,z}(n))^2\right](1 + O(n^{3\beta-2+\epsilon})),$$

which proves the asymptotic independence of the urns in the case $\kappa < 2/3$.

5.5 Looking for an optimum

We are now ready to tackle the optimization problem for $P_2(v), q(n) = e^{-C/n}$.

The covariance between two urns is asymptotically 0. Nevertheless, we have a correction to the Poisson variance of order n. However, as we will see, Version 3 is not interesting, so, we do not want to be completely precise and, in first approximation, we will assume asymptotic independence of urns. Proceeding now as in Sec. 2.2, we deal with the bracketed term in $P_2(v)$. We have $K - i = mv - i = m(v - i/m)$. After some algebra, this leads to

$$\begin{aligned}&\left\{\prod_{i=1}^{K}\left(1 - \exp\left[-Ce^{-C(v-i/m)}\right]\left[1 + \alpha_7(v-i/m)\varepsilon + \alpha_8(v-i/m)\varepsilon^2 + \frac{\alpha_{12}(v-i/m)}{m}\right] \times\right.\right.\\ &\left.\left.\times Ce^{-C(v-i/m)}\left[1 + C(v-i/m)\varepsilon + \left[-C(v-i/m) + \frac{C^2(v-i/m)^2}{2}\right]\varepsilon^2 + \frac{C}{2m}\right]\right)\right\},\end{aligned} \tag{34}$$

with

$$\alpha_{12}(v) = -\frac{Ce^{-Cv}\left[C + Ce^{-Cv} - 2\right]}{2}.$$

Note again that α_{12} is different from α_9 in (6). Note also that the product is now limited to K: the urns up to $\mathcal{O}(m)$ do have a similar (Poisson) behaviour.

We put

$$\begin{aligned}
\alpha_1(i,v) &:= -\exp\left[-Ce^{-C(v-i/m)}\right],\\
\alpha_2(i,v) &:= \alpha_7(v-i/m),\\
\alpha_3(i,v) &:= \alpha_8(v-i/m),\\
\alpha_4(i,v) &:= Ce^{-C(v-i/m)},\\
\alpha_5(i,v) &:= C(v-i/m),\\
\alpha_6(i,v) &:= -C(v-i/m) + \frac{C^2(v-i/m)^2}{2},
\end{aligned}$$

$$
\begin{aligned}
\alpha_9(i,v) &:= \alpha_{12}(v-i/m),\\
\alpha_{11}(i,v) &:= \frac{C}{2}.
\end{aligned}
$$

This translates into

$$
\begin{aligned}
P_2(v) \sim \exp\left[-Ce^{-Cv}\right] \sum_u \rho(u) \left[1+\alpha_7(v)\varepsilon+\alpha_8(v)\varepsilon^2+\frac{\alpha_{10}(v)}{m}\right] \times \\
\times \left\{ \prod_{i=1}^{K} \left(1+\alpha_1(i,v)\left[1+\alpha_2(i,v)\varepsilon+\alpha_3(i,v)\varepsilon^2+\frac{\alpha_9(i,v)}{m}\right]\alpha_4(i,v)\left[1+\alpha_5(i,v)\varepsilon+\alpha_6(i,v)\varepsilon^2+\frac{\alpha_{11}}{m}\right]\right)\right\} \\
\sim F_0(v)+\frac{F_1(v)}{m},
\end{aligned}
$$

where F_0, F_1 are given by (7),(8), with the new expressions for $\alpha_{\cdot}$ and the sums in $S_{\cdot}$ go to K.

We can now proceed as in Section 2.2. We will limit ourselves to the dominant term, i.e.

$$
\begin{aligned}
P_2(v) &\sim F_0(v),\\
F_0(v) &= \exp\left[-Ce^{-Cv}\right] D_0(v),\\
D_0(v) &:= \prod_{i=1}^{K} A_0(i,v),
\end{aligned}
$$

where

$$
A_0(i,v) = 1+\alpha_1(i,v)\alpha_4(i,v) = 1-\exp[-Ce^{-C(v-i/m)}]Ce^{-C(v-i/m)}.
$$

By Euler-Maclaurin, we conlude, setting $(v-i/m)=\eta$, that

$$
P_2(v) \sim \exp\left[-Ce^{-Cv}\right]\exp\left[m\int_0^v \ln\left[1-\exp\left[-Ce^{-C\eta}\right]Ce^{-C\eta}\right]d\eta\right].
$$

After a detailed analysis, this shows that for sufficiently large m, namely, $m>m^*$, with $m^* = -C^2/\ln\left[1-Ce^{-C}\right]$, the probability $P_2(v)$ is stricly decreasing, and hence no optimum is attainable.

Choosing $q(n)=e^{-C/n}$ is definitively too strong.

6 Version 4. p,q depend on n

If we try $q(n)=e^{-C\delta(n)}$, we can remove the absurdity related to $q(n)=e^{-C/n}$ by carefully choosing $\delta(n)$. For simplicity, let $C=1$. We change the scale into

$$
K = \varphi_0(m)+\varphi_1(m)v,
$$

and we want (we deal with the mean of N, i.e. m)

$$
m\pi = m\frac{p(m)}{q(m)}q^K = e^{-v}.
$$

Limiting our interest first to the dominant term yields

$$m\pi \sim m\delta(m)e^{-\delta(m)[\varphi_0(m)+\varphi_1(m)v]}. \tag{35}$$

Choose here

$$\varphi_1(m) = \frac{1}{\delta(m)},$$

and

$$m\delta(m)e^{-\delta(m)\varphi_0(m)} = 1.$$

For instance, if

$$\delta(m) = \frac{1}{m^\kappa}, \kappa < 1,$$

then we have

$$m^{1-\kappa}e^{-\varphi_0(m)/m^\kappa} = 1,$$

or equivalently

$$\varphi_0(m) = (1-\kappa)m^\kappa \ln(m),$$

and

$$\begin{aligned}\varphi_1(m) &= m^\kappa \ll \varphi_0(m),\\ m^\kappa &\ll m\end{aligned}$$

Eq. (35) leads now to

$$e^{-v},$$

as expected.

6.1 Poissonization

As above, we first consider the model in which the balls are independent, and each ball has a GEOM($p(n)$) distribution, with $p(n) = 1 - e^{-\delta(n)}$; here, $\delta(n) = 1/n^\kappa$ and $\kappa < 1$. So the probability that a ball lands in the ith urn is exactly $\pi(i) := p(n)q(n)^{i-1}$, where $q(n) = e^{-\delta(n)}$. If k balls are placed into the urns, then we define

$$X_i(k,n) := [\![\text{ value } i \text{ appears among the } k \text{ GEOM } p(n) \text{ RVs exactly once }]\!],$$

and $X(k,n) = \sum_{i=1}^{\infty} X_i(k,n)$. So $X(k,n)$ denotes the number of urns that each contain exactly one ball (when starting with k balls).

As before, in order to obtain the asymptotics of $X(n,n)$, we consider a model with a Poisson number N_τ of balls placed into the urns (where N_τ has mean τ). We define

$$\widetilde{X}_i(\tau,n) := [\![\text{ value } i \text{ appears among the } N_\tau \text{ GEOM } (p(n)) \text{ RVs exactly once }]\!].$$

Then define $\widetilde{X}(\tau,n) = \sum_{i=1}^{\infty} \widetilde{X}_i(\tau,n)$. So $\widetilde{X}(\tau,n)$ denotes the number of urns that each contain exactly one ball (when starting with N_τ balls).

As before, we consider $\widetilde{M}_n(\tau) := \mathbb{E}(\widetilde{X}(\tau,n))$ and $\widetilde{U}_n(\tau) := \mathbb{E}(\widetilde{X}^2(\tau,n))$. We have

$$\sum_{k=0}^{\infty} \frac{e^{-\tau}\tau^k}{k!}\mathbb{E}(X(k,n)) = \widetilde{M}_n(\tau) = \sum_{i=1}^{\infty} \tau\pi(i)e^{-\tau\pi(i)}$$

and
$$\sum_{k=0}^{\infty} \frac{e^{-\tau}\tau^k}{k!}\mathbb{E}(X^2(k,n)) = \widetilde{U}_n(\tau) = \sum_{i=1}^{\infty} \tau\pi(i)e^{-\tau\pi(i)}\Big(1 + \sum_{j\neq i} \tau\pi(j)e^{-\tau\pi(j)}\Big).$$

6.2 Diagonal DePoissonization

We compute $\mathbb{E}(X(n,n))$ and $\mathbb{E}(X^2(n,n))$ in a manner similar to the one used in the previous section. We note that both of these functions satisfy the requirements of Jacquet and Szpankowski's Diagonal dePoissonization Theorem stated earlier, with β values "1" and "2" (see the discussion in the previous section; the details do not change here).

Since $\mathbb{E}(X(n,n))$ satisfies conditions (I) and (O) (with $\beta = 1$) for $\widetilde{F}_n(\tau) = \widetilde{M}_n(\tau)$, then by the Diagonal dePoissonization Lemma we conclude that, for every nonnegative integer m,

$$\mathbb{E}(X(n,n)) = \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij}n^i\, \partial_\tau^j \mathbb{E}(\widetilde{X}(\tau,n))\Big|_{\tau=n} + O(n^{-m}).$$

In particular, when $m = 1$,

$$\mathbb{E}(X(n,n)) = \mathbb{E}(\widetilde{X}(n,n)) - \frac{1}{2}n\widetilde{M}''_n(n) + \frac{1}{3}n\widetilde{M}'''_n(n) + O(n^{-1}). \tag{36}$$

We note that $\widetilde{M}_n^{\langle j\rangle}(n) = \sum_{i=1}^{\infty}(-\pi(i))^j e^{-n\pi(i)}(n\pi(i) - j)$. (In fact, each of the correction terms in (36) decay exponentially in terms of n. In particular, the $\frac{1}{3}n\widetilde{M}'''_n(n)$ term above is not only $O(n^{-1})$ but in fact exponentially small in terms of n.) Thus (36) simplifies to

$$\mathbb{E}(X(n,n)) = \mathbb{E}(\widetilde{X}(n,n)) - \frac{1}{2}n\widetilde{M}''_n(n) + \cdots. \tag{37}$$

We use Euler-Maclaurin summation to compute

$$\mathbb{E}(\widetilde{X}(n,n)) = \sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)} = \int_1^{\infty} n\pi(i)e^{-n\pi(i)}\,di - \frac{1}{2}n\pi(i)e^{-n\pi(i)}\Big|_{i=1}^{\infty} + \cdots. \tag{38}$$

The integral evaluates to

$$n^\kappa(1 - \exp(n(e^{-1/n^\kappa} - 1))) = n^\kappa - n^\kappa\exp(-n^{1-\kappa}) + \cdots;$$

also, the lower-order terms, and all correction terms, decay exponentially in terms of n.

Thus, the expectation in the Poisson model is

$$\mathbb{E}(\widetilde{X}(n,n)) \sim n^\kappa. \tag{39}$$

Next, we use Euler-Maclaurin summation to compute the correction between $\mathbb{E}(X(n,n))$ and $\mathbb{E}(\widetilde{X}(n,n))$, namely

$$\begin{aligned}-\frac{1}{2}n\widetilde{M}''_n(n) &\sim -\frac{1}{2}n\int_1^{\infty}(-\pi(i))^2e^{-n\pi(i)}(n\pi(i) - 2)\,di\\ &= \frac{1}{2}n^{1+\kappa}\exp(n(e^{-1/n^\kappa} - 1))(1 - 2e^{-1/n^\kappa} + e^{-2/n^\kappa})\end{aligned}$$

$$= \Theta(n^{1-\kappa}\exp(-n^{1-\kappa})), \tag{40}$$

and thus decays exponentially in terms of n.

By (37), we simply add (39) and (40) to obtain the expectation in the dependent model:

$$\mathbb{E}(X(n,n)) \sim n^{\kappa}.$$

In summary, $\mathbb{E}(X(n,n))$ and $\mathbb{E}(\widetilde{X}(n,n))$ are each asymptotically n^{κ}, and the difference between $\mathbb{E}(X(n,n))$ and $\mathbb{E}(\widetilde{X}(n,n))$ is at most $O(n^{-1})$.

6.2.1 Comparison of Variances

Now we check conditions (I) and (O) for $\widetilde{F}_n(\tau) = \widetilde{U}_n(\tau)$ and $f_{k,n} = \mathbb{E}(X^2(k,n))$. We see that (by the same type of reasoning that we used above for $\widetilde{M}_n(\tau)$) for $\tau \in \mathcal{C}$, $|\widetilde{U}_n(\tau)| \le 4n^2 + 2n$, and for $\tau \notin \mathcal{C}$, $|\widetilde{U}_n(\tau)e^{\tau}| \le e^{n/\sqrt{2}}cn^2$ for some $c > 0$. Since (I) and (O) are both satisfied (with $\beta = 2$) for $\widetilde{F}_n(\tau) = \widetilde{U}_n(\tau)$, then by the Diagonal dePoissonization Lemma we conclude that, for every nonnegative integer m,

$$\mathbb{E}(X^2(n,n)) = \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij} n^i \, \partial_\tau^j \mathbb{E}(\widetilde{X}^2(\tau,n))\Big|_{\tau=n} + O(n^{1-m}).$$

In particular, when $m = 2$,

$$\mathbb{E}(X^2(n,n)) = \mathbb{E}(\widetilde{X}^2(n,n)) - \frac{1}{2}n\widetilde{U}_n''(n) + \frac{1}{3}n\widetilde{U}_n'''(n) + \frac{1}{8}n^2\widetilde{U}_n''''(n) + O(n^{-1}). \tag{41}$$

For ease of notation, we also define $R_n(\tau) := \sum_{i=1}^{\infty}(\tau\pi(i))^2 e^{-2\tau\pi(i)}$. We note

$$R_n(n) \sim -\frac{1}{4}n^{\kappa}\big(\exp(2n(e^{-1/n^{\kappa}} - 1))(2n(1 - e^{-1/n^{\kappa}}) + 1) - 1\big) \sim \frac{1}{4}n^{\kappa}$$

Also we note that $R_n''(n)$, $R_n'''(n)$, and $R_n'''(n)$ are each $\Theta\big(n^{1+\kappa}\exp(-2n^{1-\kappa})\big)$.

We compute

$$\mathbb{E}(\widetilde{X}^2(n,n)) = \sum_{i=1}^{\infty} n\pi(i)e^{-n\pi(i)}\left(1 + \sum_{j\neq i} n\pi(j)e^{-n\pi(j)}\right) = \widetilde{M}_n(n)^2 + \widetilde{M}_n(n) - R_n(n). \tag{42}$$

From (39), it follows that the second moment in the Poisson model is

$$\mathbb{E}(\widetilde{X}^2(n,n)) \sim n^{2\kappa} + \frac{3}{4}n^{\kappa}, \tag{43}$$

and the error terms have exponential decay in terms of n.

We also compute

$$-\frac{1}{2}n\widetilde{U}_n''(n) = -n\widetilde{M}_n'(n)^2 - n\widetilde{M}_n(n)\widetilde{M}_n''(n) - \frac{1}{2}n\widetilde{M}_n''(n) + \frac{1}{2}nR_n''(n)$$
$$\frac{1}{3}n\widetilde{U}_n'''(n) = 2n\widetilde{M}_n'(n)\widetilde{M}_n''(n) + \frac{2}{3}n\widetilde{M}_n(n)\widetilde{M}_n'''(n) + \frac{1}{3}n\widetilde{M}_n'''(n) - \frac{1}{3}nR_n'''(n)$$

$$\frac{1}{8}n^2\widetilde{U}_n''''(n) = \frac{3}{4}n^2\widetilde{M}_n''(n)^2 + n^2\widetilde{M}_n'(n)\widetilde{M}_n'''(n) + \frac{1}{4}n^2\widetilde{M}_n(n)\widetilde{M}_n''''(n) + \frac{1}{8}n^2\widetilde{M}_n'''' - \frac{1}{8}n^2R_n''''(n) \tag{44}$$

We recall, from the Euler-Maclaurin summation in (40), that

$$\widetilde{M}_n''(n) = -n^\kappa \exp(n(e^{-1/n^\kappa}-1))(1-2e^{-1/n^\kappa}+e^{-2/n^\kappa}) = \Theta(n^{-\kappa}\exp(-n^{1-\kappa})).$$

Similarly, we use Euler-Maclaurin summation to compute the following:

$$\begin{aligned}
\widetilde{M}_n'(n) &\sim n^\kappa \exp(n(e^{-1/n^\kappa}-1))(1-e^{-1/n^\kappa}) = \Theta(\exp(-n^{1-\kappa}))\\
\widetilde{M}_n'''(n) &\sim n^\kappa \exp(n(e^{-1/n^\kappa}-1))(1-3e^{-1/n^\kappa}+3e^{-2/n^\kappa}-e^{-3/n^\kappa}) = \Theta(n^{-2\kappa}\exp(-n^{1-\kappa}))\\
\widetilde{M}_n''''(n) &\sim n^\kappa \exp(n(e^{-1/n^\kappa}-1))(1-4e^{-1/n^\kappa}+6e^{-2/n^\kappa}-4e^{-3/n^\kappa}+e^{-4/n^\kappa}) = \Theta(n^{-3\kappa}\exp(-n^{1-\kappa}))
\end{aligned} \tag{45}$$

By (41), we simply add (43) and all three parts of (44) to obtain the second moment in the dependent model:

$$\mathbb{E}(X^2(n,n)) = n^{2\kappa} + \frac{3}{4}n^\kappa + O(n^{-1}).$$

In particular, $\mathbb{E}(X^2(n,n))$ and $\mathbb{E}(\widetilde{X}^2(n,n))$ each are asymptotically $n^{2\kappa}+\frac{3}{4}n^\kappa$. The difference between $\mathbb{E}(X^2(n,n))$ and $\mathbb{E}(\widetilde{X}^2(n,n))$ is at most $O(n^{-1})$.

6.3 A direct approach to the correction term

Now the correction term is more interesting: all contributions are asymptotically 0 (we omit the details). Everything fits with previous computations. However, here, we have no constraints on $\kappa < 1$. So we conjecture that far all $0 < \kappa < 1$, we have asymptotic independence of urns.

The dominant term of mean and variance are given by the Poisson approximation, i.e;

$$\mathbb{E}(X(n,n)) \sim n^\kappa, \qquad \mathbb{V}(X(n,n)) \sim n^\kappa.$$

6.4 Looking for an optimum for $P_2(v), q(n) = e^{-1/n^\kappa}, 0 < \kappa < 2/3$

Again, we proceed as in Section 2.2. Limiting ourselves to the dominant term, we have

$$\begin{aligned}
P_2(v) &\sim F_0(v),\\
F_0(v) &= \exp\left[-e^{-v}\right] D_0(v),\\
D_0(v) &:= \prod_{i=1}^{\infty} A_0(i,v),
\end{aligned}$$

where

$$A_0(i,v) = 1 + \alpha_1(i,v)\alpha_4(i,v) = 1 - \exp[-e^{-(v-i/m^\kappa)}]e^{-(v-i/m^\kappa)}.$$

Numerical experiments show that we must take $v << 0$. This entails

$$e^{-(v-i/m^\kappa)} >> 1,$$

and hence

$$\exp\left[-e^{-(v-i/m^\kappa)}\right] << 1,$$

and also

$$\ln(A_0(i,v)) \sim -\exp\left[-e^{-(v-i/m^\kappa)}\right] e^{-(v-i/m^\kappa)}.$$

Using Euler-Maclaurin, we obtain, setting $(v - i/m^\kappa) = \eta$,

$$P_2(v) \sim \exp\left[-e^{-v}\right] \exp\left[-m^\kappa \int_{-\infty}^{v} \exp\left[-e^{-\eta}\right] e^{-\eta} d\eta\right]. \tag{46}$$

By taking logarithms, maximizing $P_2(v)$ leads to

$$e^{-\tilde{v}} - m^\kappa \exp[-e^{-\tilde{v}}]e^{-\tilde{v}} = 0,$$

or

$$\exp\left[-e^{-\tilde{v}}\right] = 1/m^\kappa,$$

that is

$$e^{-\tilde{v}} = \ln(m^\kappa),$$

and so

$$\tilde{v} = -\ln[\ln(m^\kappa)] << 0 \text{ as expected} \tag{47}$$

We want now $\tilde{P}_2 = P_2(\tilde{v})$. Set $\eta = \tilde{v} - \tau$. From (46) we obtain

$$-m^\kappa \int_{-\infty}^{\tilde{v}} \exp\left[-e^{-\eta}\right] e^{-\eta} d\eta = -m^\kappa \int_{\tau=0}^{\infty} \exp[-e^{-\tilde{v}+\tau}]e^{-\tilde{v}+\tau} d\tau.$$

Set now

$$\begin{aligned} G_0 &= -\tilde{v} = \ln[\ln(m^\kappa)], \\ G_1 &= e^{G_0} = \ln(m^\kappa), \\ e^{-G_1} &= 1/m^\kappa, \\ e^\tau &= u. \end{aligned}$$

This gives then

$$-m^\kappa \int_0^\infty \exp\left[-G_1 e^\tau\right] G_1 e^\tau d\tau = -m^\kappa \int_{u=1}^{\infty} G_1 e^{-G_1 u} du,$$

and hence, after a straightforward computation

$$-m^\kappa e^{-G_1} = -1.$$

So, finally

$$\tilde{P}_2 \sim \frac{1}{m^\kappa} e^{-1}. \tag{48}$$

Remark: We have tried to understand whether the factor $1/e$ has a simple explanation, but we are not able to find an analogy of the problem to a problem of sequential selection.

Also

$$|\tilde{v}| \ll \varphi_1(m),$$
$$\tilde{K} = \varphi_0(m) + \varphi_1(m)\tilde{v} = (1-\kappa)m^\kappa \ln(m) - m^\kappa \ln[\ln(m^\kappa] = m^\kappa[(1-\kappa)\ln(m) - \ln(\ln(m)) - \ln(\kappa)] \gg 1.$$

A plot of $P_2(v)$, using (5), with $\kappa = 1/2, m = 1000$ is given in Figure 8. This leads to $v_n = -1.24\ldots, P_2(v_n) = 0.0117\ldots$. We also obtain $\tilde{v} = -1.24\ldots, \tilde{P}_2 \sim 0.01164\ldots, \tilde{K} = 70$.

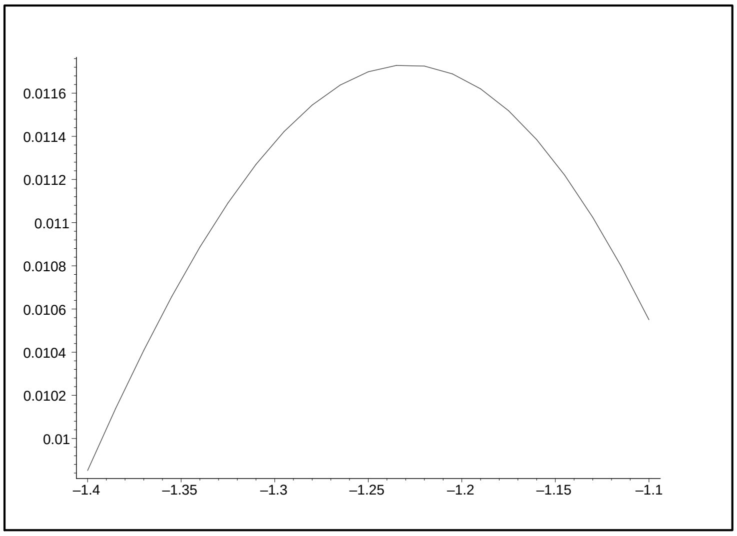


Figure 8: $P_2(v)$ for the case $q(n) = e^{-1/n^\kappa}$

A correction to the dominant term can be computed. We derive, with $\varepsilon = u/m$,

$$\begin{aligned} P_0(v) \sim \exp[-e^{-v}] \Bigg[1 - \frac{e^{-v}}{2m^\kappa} - e^{-v}\varepsilon[1-\kappa+\kappa v] \\ + e^{-v}\frac{\varepsilon^2}{2}\left[3\kappa^2 v - \kappa v - \kappa^2 - \kappa^2 v^2 + \kappa + e^{-v} + 2e^{-v}\kappa v - 2e^{-v}\kappa + e^{-v}\kappa^2 v^2 - 2e^{-v}\kappa^2 v + e^{-v}\kappa^2\right] \\ + \mathcal{O}\left(\frac{1}{m^{2\kappa}}\right) + \mathcal{O}\left(\frac{1}{m}\right) + \mathcal{O}\left(\frac{\varepsilon}{m}\right)\Bigg]. \end{aligned}$$

We see here that *the dominant term is related to the* $\frac{1}{m^\kappa}$ *term.* This is *independent from* β. Similarly, we obtain (up to the required precision)

$$(1-\pi)^{n-1} \sim (1-\pi)^n,$$

and

$$n\pi \sim e^{-v}[1 + \frac{1}{m^{2\kappa}} + \varepsilon[1-\kappa+\kappa v] - \kappa\frac{\varepsilon^2}{2}[-1+\kappa v^2 + v - 3\kappa v + \kappa]].$$

So, finally, the dominant term of bracketed in $P_2(v)$ becomes

$$\left\{ \prod_{i=1}^{\infty} \left(1 - \exp\left[-e^{-(v-i/m^\kappa)}\right] \left[1 - \frac{e^{-(v-i/m^\kappa)}}{2m^\kappa}\right] e^{-(v-i/m^\kappa)} \left[1 + \frac{1}{2m^\kappa}\right] \right) \right\}. \tag{49}$$

and

$$P_2(v) \sim \exp\left[-e^{-v}\right]\left[1-\frac{e^{-v}}{2m^\kappa}\right]\left\{\prod_{i=1}^{\infty}\left(1-\exp\left[-e^{-(v-i/m^\kappa)}\right]\left[1-\frac{e^{-(v-i/m^\kappa)}}{2m^\kappa}\right]e^{-(v-i/m^\kappa)}\left[1+\frac{1}{2m^\kappa}\right]\right)\right\}$$

We must now compute the solution $\bar{v}$ of $\partial_v \ln(P_2(v)) = 0$. Proceeding as above, this leads to

$$1+\frac{1}{2m^\kappa} - m^\kappa \exp\left[-e^{-\bar{v}}\right]\left[1-\frac{e^{-\bar{v}}}{2m^\kappa}\right]\left[1+\frac{1}{2m^\kappa}\right] = 0. \qquad (50)$$

Indeed, taking logarithms in (49) only induces an extra term of order $\exp[-2e^{-\tilde{v}}]e^{-2\tilde{v}} = \mathcal{O}\left(\frac{\ln(m^\kappa)^2}{m^{2\kappa}}\right)$ which does not affect (50).

Set therefore

$$e^{-\bar{v}} = \ln(m^\kappa) + \eta.$$

This gives the dominant equation

$$1+\frac{1}{2m^\kappa} - e^{-\eta}\left[1-\frac{\ln(m^\kappa)+\eta}{2m^\kappa}\right]\left[1+\frac{1}{2m^\kappa}\right] = 0,$$

which is asymptotically equivalent to

$$1+\frac{1}{2m^\kappa} - (1-\eta)\left[1-\frac{\ln(m^\kappa)}{2m^\kappa}\right]\left[1+\frac{1}{2m^\kappa}\right] = 0.$$

Therefore

$$\eta \sim -\frac{\ln(m^\kappa)}{2m^\kappa},$$

i.e.

$$e^{-\bar{v}} \sim \ln(m^\kappa)\left[1-\frac{1}{2m^\kappa}\right]$$

or again

$$\bar{v} \sim -\ln[\ln(m^\kappa)] + \frac{1}{2m^\kappa} = \tilde{v} + \frac{1}{2m^\kappa}$$

and

$$\bar{K} \sim \tilde{K} + 1/2.$$

As K must be an integer, we see that the correction is practically negligible with our choice of parameters. To summarize, the algorithm works as follows.

Algorithm 2

Input: C, κ, m
Output: second order optimal value for $K : \bar{K}$
Compute $\tilde{v} = -\ln[\ln(m^\kappa)]$;
Compute $\tilde{K} = \varphi_0(m) + \varphi_1(m)\tilde{v}$;
Compute $\bar{K} = \lfloor \tilde{K} + 1/2 \rfloor$;
End

---

7 Conclusion

The problem studied in this paper may be seen as a game where each player of an unknown number of players can choose infinitely many actions in the sense that he/she can choose a distribution in an (uncountable) set of distributions according to which to place his/her offer. There is no other constraint than that the support of his/her choice cannot exceed $[1..V]$. Hence the only approach to such a problem we see is to assume that individual strategic behaviour (to maximize the probability of placing the minimum single offer) leads to a common distribution for all players. We gave several good reasons why a (truncated) geometric distribution should model the situation more suitably than other choices, although, clearly, our arguments depend more on exclusion of unreasonable distributions than on actual preferences. Our next step was to assume some knowledge about $\mathbb{E}(N)$ and the variance of N, because, as we argued, with no information on N whatsoever, we still would have an ill-posed problem.

With these assumptions, the problem is sufficiently well defined to allow the search of an optimum. Finding in practice the optimum is seemingly only possible via asymptotic expansions and algorithms with which we can give, as we have seen, explicit answers. We conclude, there is no good rule of thumb for the optimal choice, that is there is no easy answer without calculation. However, in some cases, the effort is clearly rewarding. The optimum is frequently distinguishably better than some random choice in regions we may think of as being reasonable.

The general difficulty of the problem has, as we should add saying, one game-specific advantage. Since these results would probably be perceived as "too mathematical" by the large majority of participants, we do not expect any serious danger of distortion by publishing them.

8 Acknowledgments.

The problem was suggested to us by G.M. Ziegler. We gratefully acknowledge insightful suggestions of P.Jacquet and W.Szpankowski.

References

[1] A.D. Barbour, L. Holst, and S. Janson. *Poisson Approximation.* Oxford University Press, 1992.

[2] 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.

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

[4] P. Jacquet and W. Szpankowski. Analytic depoissonization and its applications. *Theoretical Computer Science*, 201(1–2):1–62, 1998.

[5] 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.

[6] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences.* Wiley, New york, 2001.