

RT-MAE 2001-13

**THE NUMBER OF EMPTY CELLS IN AN  
ALLOCATION SCHEME GENERATED BY  
A ZERO-INFLATED DISTRIBUTION:  
EXACT RESULTS AND POISSON CONVERGENCE**

by

**Ljuben Mutafchiev  
and  
Nikolai Kolev**

**Palavras-Chave:** Random allocations; Zero-inflated discrete distributions; Poisson convergence.

**Classificação AMS:** 60C05, 62E15, 60E05, 60F05.  
(AMS Classification)

# The Number of Empty Cells in an Allocation Scheme Generated by a Zero-Inflated Distribution: Exact Results and Poisson Convergence

Ljuben Mutafchiev<sup>1</sup> and Nikolai Kolev<sup>2</sup>

<sup>1</sup>Institute of Mathematics and Informatics  
of the Bulgarian Academy of Sciences,

<sup>1</sup>American University in Bulgaria

and <sup>2</sup>Department of Statistics - IME, São Paulo University  
C.P. 66281, 05315-970 São Paulo, Brazil

## Abstract

In the classical allocation scheme, where  $n$  labeled balls are placed into  $N$  labeled cells, the joint distribution of the counts of the cells contents is determined by a joint conditional distribution of  $N$  independent and Poisson distributed random variables having common positive parameter  $\lambda > 0$ . In this paper we explore the idea of adjusting the probability  $e^{-\lambda}$  of zero outcomes in the initial Poisson distribution formula by making it greater relative to the underlying non-zero probabilities. Thus, dealing with this type of distribution function, called zero-inflated Poisson, we derive a closed-form expression for the exact distribution of the number  $\mu_0(n, N)$  of empty cells, and afterwards, we obtain certain limiting distributions. It turns out that, as  $n, N \rightarrow \infty$ ,  $\mu_0(n, N)$  follows again the Poisson limit law, however, in certain cases the scaling parameters differ from those obtained by von Mises (1939) and Békéssy (1963) in the classical occupancy scheme.

**Key words and Phrases:** *Random allocations; Zero-inflated discrete distributions; Poisson convergence.*

**AMS (MOS) subject classification:** 60C05, 62E15, 60E05, 60F05

## 1 Introduction

Many classical occupancy problems concern models of placing  $n$  balls labeled by  $1, 2, \dots, n$  into  $N$  cells labeled by  $1, 2, \dots, N$ . It is then rather natural to assume that each of the  $N^n$  possible allocations has probability  $N^{-n}$ . Thus, each conceivable numerical characteristic of a random allocation becomes a random variable with a specified probability distribution. Various exact and asymptotic distributional results on this subject are included in classical textbooks and

monographs (see e.g. Feller (1967), Kolchin et al. (1978), Sachkov (1978)). If in this model, we let  $\eta_j$  to denote the content (the number of balls) of the  $j$ th cell,  $j = 1, 2, \dots, N$ , then the vector  $(\eta_1, \dots, \eta_N)$  has a multinomial distribution, i.e.

$$P(\eta_1 = k_1, \eta_2 = k_2, \dots, \eta_N = k_N) = \frac{n!}{k_1!k_2! \dots k_N!N^n}, \quad (1)$$

where the integers  $k_j$  satisfy

$$k_1 + k_2 + \dots + k_N = n, \quad k_j \geq 0, \quad j = 1, 2, \dots, N. \quad (2)$$

Consider now  $N$  independent random variables  $\xi_j$ ,  $j = 1, 2, \dots, N$ , having all the Poisson distribution with an arbitrary parameter  $\lambda > 0$ . It is then easy to verify using (1) that

$$P(\eta_1 = k_1, \dots, \eta_N = k_N) = P(\xi_1 = k_1, \dots, \xi_N = k_N | \xi_1 + \dots + \xi_N = n), \quad (3)$$

(see Kolchin (1986), Chap. 1 or Kolchin et al. (1978), Chap. 2) for  $k_j$ 's satisfying (2). Relation (3) can be also regarded as a definition of a scheme of allocating  $n$  labeled balls into  $N$  labeled cells. Assuming that  $\xi_j$ ,  $j = 1, 2, \dots, N$ , are independent and identically distributed random variables with

$$P(\xi_1 = k) = p_k, \quad k = 0, 1, \dots, \quad \sum_{k=0}^{\infty} p_k = 1, \quad (4)$$

one obtains the so-called generalized scheme of allocation defined by (3) (see again Kolchin (1986), Chap. 1). If  $p_k = \lambda^k e^{-\lambda} (k!)^{-1}$ ,  $k = 0, 1, \dots$ , we get the classical scheme (1).

In the present note we consider the generalized scheme of allocation that addresses the problem that the count data display a higher fraction of zeros, or non-occurrences. More precisely, we introduce a common alteration making the probability of zero (in the Poisson distribution formula) greater relative to the underlying non-zero probabilities. For properties, generalizations and certain statistical and economical applications of zero-inflated distributions we refer the reader to Johnson et al. (1992), Chap. 8; Winkelmann (2000), Chap. 3 and 4; Kolev et al. (2000). In what follows next we assume that (4) and (3) are valid with a distribution

$$\begin{aligned} p_0 &= \rho + (1 - \rho)e^{-\lambda}, \\ p_k &= (1 - \rho)\lambda^k e^{-\lambda} (k!)^{-1}, \quad k = 1, 2, \dots, \quad \lambda > 0, \quad 0 < \rho < 1. \end{aligned} \quad (5)$$

Our aim in this note is to derive the exact distribution of the number  $\mu_0(n, N)$  of empty cells in the generalized allocation model defined by (3)-(5) as well as to obtain some asymptotic results related to the Poisson convergence of  $\mu_0(n, N)$  as  $n, N \rightarrow \infty$ . Note that if  $\rho = 0$  we arrive at the classical occupancy scheme in which we adopt the notation  $m_0(n, N)$  for the number of empty cells.

Our paper is organized as follows. In the next Section 2 we summarize our results. Section 3 contains the proof of the formula for the exact distribution of  $\mu_0(n, N)$  together with a key lemma that implies our asymptotic results. Finally, in Section 4 we outline some directions for further research on this topic.

We conclude this section with two classical Poisson convergence results for  $m_0(n, N)$ , the number of empty cells in the classical occupancy scheme (1). We shall explore them in our proofs to show the differences that appear between classical and zero-inflated allocations. We start with an old theorem due to von Mises (1939). It can be seen in certain famous textbooks and monographs (see e.g. Feller (1967), Chap. 4; Kolchin *et al.* (1978) Chap. 1; Sachkov (1978), Chap. 3).

**Theorem 1.** *If  $n, N \rightarrow \infty$  so that*

$$n = N \log N + cN + o(N), \quad -\infty < c < \infty, \quad (6)$$

*then*

$$P[m_0(n, N) = k] = \frac{e^{-kc}}{k!} \exp(-e^{-c}) + o(1), \quad k = 0, 1, \dots \quad (7)$$

Relation (6) above shows that the number  $n$  of balls in Theorem 1 is larger relative to the number  $N$  of cells. It also defines the so-called “right” domain of allocations, Kolchin *et al.* (1978), Chap. 1.

The next limit theorem belongs to Békéssy (1963) and concerns the “left” domain of allocation.

**Theorem 2.** *If  $n, N \rightarrow \infty$  so that*

$$n^2(2N)^{-1} \rightarrow d < \infty, \quad (8)$$

*then*

$$P[m_0(n, N) - (N - n) = k] = \frac{d^k e^{-d}}{k!} + o(1), \quad k = 0, 1, \dots \quad (9)$$

Another proof of this result can be also found in Kolchin *et al.* (1978), Chap. 1.

## 2 Summary of Results

First, we introduce some conventions and notations that we need later. Throughout the paper we assume that our additional parameters  $\rho$  and  $\lambda$  in the zero-inflated allocation scheme satisfy the dependence

$$\rho = \rho(n, N), \quad \lambda = \lambda(n, N).$$

For the sake of convenience, we write

$$Q = Q(\rho, \lambda) = (1 - \rho)e^{-\lambda}. \quad (10)$$

We also let

$$S_{n, N} = \sum_{k=0}^N \binom{N}{k} Q^k \rho^{N-k} k^n. \quad (11)$$

Finally, we let  $\Delta$  to denote the finite difference operator, see e.g. Riordan (1958), Chap. 1.

**Theorem 3.** Under the above notations the following relation is true:

$$P[\mu_0(n, N) = k] = \frac{(\rho + Q)^k Q^{N-k}}{S_{n,N}} \binom{N}{k} \Delta^{N-k} 0^n, \quad k = 0, 1, \dots, N. \quad (12)$$

**Note.** The corresponding result in the classical scheme of allocation states that

$$P[m_0(n, N) = k] = \binom{N}{k} N^{-n} \Delta^{N-k} 0^n, \quad k = 0, 1, \dots, N, \quad (13)$$

see Riordan (1958), Chap. 5 or Sachkov (1978), Chap. 3. If we set  $\rho = 0$ , one can easily check that

$$P[\mu_0(n, N) = k] = P[m_0(n, N) = k], \quad k = 0, 1, \dots, N.$$

Now we proceed to asymptotic results in the "right" domain. We establish a weak convergence to the Poisson distribution of the type:

$$P[\mu_0(n, N) = k] = \frac{[(1+q)e^{-c}]^k}{k!} \exp\{-(1+q)e^{-c}\} + o(1), \quad k = 0, 1, \dots, \quad (14)$$

where  $q \geq 0$  is determined in the following theorem.

**Theorem 4.** Suppose that  $n, N \rightarrow \infty$  so that (6) holds.

(i) If  $N\rho \rightarrow 0$  and  $\rho e^\lambda \rightarrow \alpha_1 \in [0, \infty)$ , then (14) is valid with  $q = \alpha_1$ .

(ii) If  $N\rho \rightarrow b \in (0, \infty)$  and  $N e^{-\lambda} \rightarrow \alpha_2 \in (0, \infty)$ , then (14) is valid with  $q = \frac{b}{\alpha_2}$ .

(iii) If  $N\rho \rightarrow b \in (0, \infty)$  and  $N e^{-\lambda} \rightarrow \infty$ , then (14) is valid with  $q = \alpha_3 = 0$ .

(iv) If  $N\rho \rightarrow \infty$  and  $\frac{\rho e^\lambda}{1-\rho} \rightarrow \alpha_4 \in [0, \infty)$ , then (14) is valid with  $q = \alpha_4$ .

In the "left" domain determined by (8) we find a sufficient condition for  $\mu_0(n, N)$  and  $m_0(n, N)$  to be asymptotically identically distributed.

**Theorem 5.** If  $n, N \rightarrow \infty$  in a way that they satisfy (8), and, if  $n\rho e^\lambda = o(1)$ , then the limiting distribution of  $\mu_0(n, N)$  coincides with that of  $m_0(n, N)$  determined by (9).

## 3 Proofs

### 3.1 Proof of Theorem 3

First, using (4) and (5), we shall derive

$$P(\xi_1 = k_1, \dots, \xi_N = k_N | \xi_1 + \dots + \xi_N = n) = \frac{P(\xi_1 = k_1, \dots, \xi_N = k_N, \xi_1 + \dots + \xi_N = n)}{P(\xi_1 + \dots + \xi_N = n)}, \quad (15)$$

for  $k_j, j = 1, \dots, N$  satisfying (2). To find  $P(\xi_1 + \dots + \xi_N = n)$  we can use generating functions. Since  $E(x^{\xi_j}) = \rho + (1-\rho)e^{\lambda(x-1)}$ ,  $|x| < 1$ , by (5) one can write

$$E(x^{\xi_1 + \dots + \xi_N}) = [\rho + (1-\rho)e^{\lambda(x-1)}]^N, \quad |x| < 1.$$

Adopting the notation  $x^n[F(x)]$  for the coefficient of  $x^n$  in the power series  $F(x)$ , we obviously have

$$\begin{aligned}
P(\xi_1 + \dots + \xi_N = n) &= x^n \left[ (\rho + (1 - \rho)e^{\lambda(x-1)})^N \right] \\
&= x^n \left[ \sum_{k=0}^N \binom{N}{k} (1 - \rho)^k e^{\lambda k(x-1)} \rho^{N-k} \right] \\
&= \sum_{k=0}^N \binom{N}{k} (1 - \rho)^k e^{-\lambda k} \rho^{N-k} x^n \left[ e^{\lambda k x} \right] \\
&= \frac{\lambda^n}{n!} \sum_{k=0}^N \binom{N}{k} (1 - \rho)^k e^{-\lambda k} \rho^{N-k} k^n,
\end{aligned}$$

i.e.

$$P(\xi_1 + \dots + \xi_N = n) = \frac{\lambda^n}{n!} S_{n,N}, \quad (16)$$

with  $S_{n,N}$  determined by (11) and (10).

Further, for a given  $N$ -tuple of integers  $k_1, \dots, k_N$  let  $k$  denote the number of those  $k_j$ 's that are equal to 0. Then, using (5), we get

$$P(\xi_1 = k_1, \dots, \xi_N = k_N, \xi_1 + \dots + \xi_N = n) = \frac{[\rho + (1 - \rho)e^{-\lambda}]^k [(1 - \rho)e^{-\lambda}]^{N-k} \lambda^n}{k_1! \dots k_N!}. \quad (17)$$

Replacing first (16) and (17) into (15), and then, going back to definition (3), we find that

$$P(\eta_1 = k_1, \dots, \eta_N = k_N) = \frac{n! [\rho + (1 - \rho)e^{-\lambda}]^k [(1 - \rho)e^{-\lambda}]^{N-k} \lambda^n}{k_1! \dots k_N! S_{n,N}} \quad (18)$$

for any vector  $(k_1, \dots, k_N)$  whose coordinates satisfy (2) such that  $k$  of them equal 0. The relation (18) may be also regarded as an inflated parameter generalization of the multinomial distribution with coinciding probabilities of outcomes.

To find  $P[\mu_0(n, N) = k]$  we have to sum the probabilities from (18) over all vectors  $(k_1, \dots, k_N)$  with  $k$  zero-coordinates satisfying (2). Since the places of the empty cells can be chosen in  $\binom{N}{k}$  ways, we can write

$$P[\mu_0(n, N) = k] = \binom{N}{k} \frac{[\rho + (1 - \rho)e^{-\lambda}]^k [(1 - \rho)e^{-\lambda}]^{N-k}}{S_{n,N}} \sum \frac{n!}{k_1! \dots k_{N-k}!},$$

where the last summation is over all positive integers  $k_j \geq 1$ ,  $1 \leq j \leq N - k$  such that  $k_1 + \dots + k_{N-k} = n$ .

Now, using a relation between the Stirling numbers of second kind and the finite difference operator  $\Delta$  (see e.g. Riordan (1958), Chap.5) by notation (10), we get expression (12) of the theorem.

### 3.2 Proofs of the Limit Theorems

Our first task here is to find the asymptotic form of  $S_{n,N}$  given by (11). The result may be stated as:

**Lemma.** *If  $n, N \rightarrow \infty$  so that*

$$\frac{\rho\sqrt{n}}{Q}e^{-\frac{n}{N}} \rightarrow 0 \quad \text{and} \quad \frac{\rho\sqrt{N}}{Q}e^{-\frac{N}{N}} \rightarrow 0, \quad (19)$$

then

$$S_{n,N} \sim N^n Q^N \exp \left\{ \frac{\rho N}{Q} e^{-\frac{N}{N}} \right\}. \quad (20)$$

**Proof.** For the sake of convenience, denote the general term of  $S_{n,N}$  by  $a_k$ , i.e.

$$a_k = \binom{N}{k} Q^k \rho^{N-k} k^n, \quad k = 0, 1, \dots, N.$$

It is straightforward that

$$\frac{a_k}{a_N} = \binom{N}{k} \left( \frac{\rho}{Q} \right)^{N-k} \left( \frac{k}{N} \right)^n, \quad k = 0, 1, \dots, N. \quad (21)$$

We apply the obvious identity

$$\frac{a_k}{a_N} = \prod_{j=0}^{N-k-1} \frac{a_{k+j}}{a_{k+j+1}},$$

and find that

$$\frac{a_k}{a_N} = \left( \frac{N\rho}{Q} \right)^{N-k} \frac{1}{(N-k)!} \prod_{j=0}^{N-k-1} \left( \frac{k+j}{k+j+1} \right)^{n-1} \frac{k+j}{N}. \quad (22)$$

Now we break the sum into two parts in the following way:

$$S_{n,N} = a_N(S_1 + S_2), \quad (23)$$

where

$$S_1 = \sum_{0 \leq k \leq N - w(N)} \frac{a_k}{a_N} \quad \text{and} \quad S_2 = \sum_{N - w(N) < k \leq N} \frac{a_k}{a_N}, \quad (24)$$

where  $w(N) \rightarrow \infty$  arbitrarily slowly, so that

$$w(N) = o(N). \quad (25)$$

We will subsequently restrict  $w(N)$  by (31) and (33).

Consider first  $S_2$  with a general term given by (22). For  $N - w(N) < k$  and  $0 \leq j \leq N - k - 1$ , we get

$$1 - \frac{w(N)}{N} < \frac{k+j}{N} \leq 1 - \frac{1}{N},$$

which means that

$$\frac{k+j}{N} = 1 + O\left(\frac{w(N)}{N}\right). \quad (26)$$

Then we have

$$\begin{aligned} \left(\frac{k+j}{k+j+1}\right)^{n-1} &= \exp\left\{(n-1)\log\left(1 - \frac{1}{k+j+1}\right)\right\} \\ &= \exp\left\{-\frac{n-1}{k+j+1} - R(n, k, j)\right\} \\ &= \exp\left\{-\frac{n(1-\frac{1}{n})}{\frac{N(k+j+1)}{N}} - R(n, k, j)\right\}, \end{aligned}$$

i.e.

$$\left(\frac{k+j}{k+j+1}\right)^{n-1} = \exp\left\{-a + O\left(\frac{aw(N)}{N}\right) - R(n, k, j)\right\}, \quad (27)$$

where for sake of brevity, we let

$$a = a(n, N) = \frac{n}{N}, \quad (28)$$

and

$$R(n, k, j) = (n-1) \left[ \frac{1}{2(k+j+1)^2} + \frac{1}{3(k+j+1)^3} + \dots \right].$$

From (25) it follows that

$$R(n, k, j) = O\left(\frac{n}{N^2}\right) = O\left(\frac{a}{N}\right) \quad (29)$$

uniformly for  $N - w(N) + 1 < k + j + 1$ . Relations (27) and (29) imply now that

$$\left(\frac{k+j}{k+j+1}\right)^{n-1} = e^{-a} \left[ 1 + O\left(\frac{aw(N)}{N}\right) \right] \quad (30)$$

uniformly for all  $n$  and  $N - w(N) < k + j \leq N - 1$ .

In what follows later we need a restriction on  $w(N)$  stronger than that in (25). So we assume further that

$$w(N) = o\left(\frac{N}{\sqrt{n}}\right) \quad \text{and} \quad w(N) = o(\sqrt{N}) \quad (31)$$

as  $n, N \rightarrow \infty$ .

Inserting into the product (22) relations (26), (28), (30), (31) and using the fact that  $N - k < w(N)$  in  $S_2$  (see (24)), we find that

$$\prod_{j=0}^{N-k-1} \left(\frac{k+j}{k+j+1}\right)^{n-1} \frac{k+j}{N}$$

$$\begin{aligned}
&= e^{-a(N-k)} \left[ 1 + O\left(\frac{aw(N)}{N}\right) \right]^{N-k} \left[ 1 + O\left(\frac{w(N)}{N}\right) \right]^{N-k} \\
&= e^{-a(N-k)} \exp \left\{ (N-k) \left[ \log \left( 1 + O\left(\frac{aw(N)}{N}\right) \right) + \log \left( 1 + O\left(\frac{w(N)}{N}\right) \right) \right] \right\} \\
&= e^{-a(N-k)} \exp \left\{ O\left(w(N)\right) \left[ \log \left( 1 + O\left(\frac{aw(N)}{N}\right) \right) + \log \left( 1 + O\left(\frac{w(N)}{N}\right) \right) \right] \right\} \\
&= e^{-a(N-k)} \exp \left\{ O\left(w(N)\right) \left[ O\left(\frac{aw(N)}{N}\right) + O\left(\frac{w(N)}{N}\right) \right] \right\} \\
&= e^{-a(N-k)} \exp \left\{ O\left(\frac{aw^2(N)}{N}\right) + O\left(\frac{w^2(N)}{N}\right) \right\} \\
&= e^{-a(N-k)} \left[ 1 + O\left(\frac{aw^2(N)}{N}\right) + O\left(\frac{w^2(N)}{N}\right) \right] = e^{-a(N-k)} [1 + o(1)].
\end{aligned}$$

The last relation is justified by (31).

Thus, for  $S_2$  from (24) we get

$$\begin{aligned}
S_2 &= [1 + o(1)] \sum_{N-w(N) < k \leq N} \frac{\left(\frac{N\rho}{Q}\right)^{N-k}}{(N-k)!} e^{-a(N-k)} \\
&= [1 + o(1)] \sum_{0 \leq l < w(N)} \frac{\left(\frac{N\rho}{Q}\right)^l}{l!} e^{-al}.
\end{aligned}$$

If  $\frac{N\rho}{Q}e^{-a} = O(1)$ , then, since  $w(N) \rightarrow \infty$ , one can immediately conclude that

$$S_2 \rightarrow \exp\left(\frac{N\rho}{Q}e^{-a}\right), \quad (32)$$

as  $n, N \rightarrow \infty$ .

When  $\frac{N\rho}{Q}e^{-a} \rightarrow \infty$ , we choose  $w(N)$  of the form:

$$w(N) = \frac{N\rho}{Q}e^{-a} + \gamma(n, N) \sqrt{\frac{N\rho}{Q}e^{-a}}, \quad (33)$$

where  $\gamma(n, N)$  is a function tending to  $\infty$  arbitrarily slowly so that  $\frac{N\rho}{Q}e^{-a}$  determines the basic growth of  $w(N)$ . Conditions (19) imply that this choice of  $w(N)$  agrees with requirements (31). Now it remains to apply the normal approximation of the Poisson distribution with parameter  $M = \frac{N\rho}{Q}e^{-a}$  (see Feller (1967), Chap. 2):

$$\sum_{l < M + \gamma(n, N)\sqrt{M}} \frac{M^l e^{-M}}{l!} = \Phi(\gamma(n, N)) + o(1), \quad (34)$$

where  $\Phi(\cdot)$  is the standard normal distribution function. Since  $\gamma(n, N) \rightarrow \infty$  as  $n, N \rightarrow \infty$ , we conclude that with  $M = \frac{N\rho}{Q}e^{-a}$  and  $y = \gamma(n, N)$  to show that

$$S_2 \sim \exp\left(\frac{N\rho}{Q}e^{-a}\right). \quad (35)$$

We now turn our attention to the easier task to find an estimate for  $S_1$ . Substituting  $\frac{2k}{a_N}$  by expression (21) and changing the summation we obtain

$$S_1 = \sum_{w(N) \leq l \leq N} \binom{N}{l} \left(\frac{\rho}{Q}\right)^l \left(\frac{N-l}{N}\right)^n,$$

Since

$$\left(\frac{N-l}{N}\right)^n = \left(1 - \frac{l}{N}\right)^n = \exp\left\{n \log\left(1 - \frac{l}{N}\right)\right\} < \exp\left\{-\frac{nl}{N}\right\} = e^{-al}$$

and

$$\binom{N}{l} < \frac{N^l}{l!},$$

we get

$$S_1 < \sum_{w(N) \leq l \leq N} \left(\frac{N\rho}{Q}e^{-a}\right)^l \frac{1}{l!}.$$

The last sum can be easily estimated using again the normal approximation of the Poisson distribution (34) with  $w(N)$  given by (33) in the case  $M = \frac{N\rho}{Q}e^{-a} \rightarrow \infty$ . We have

$$S_1 e^{-M} < \sum_{w(N) \leq l \leq N} \frac{M^l e^{-M}}{l!} = 1 - \Phi(\gamma(n, N)) + o(1) = o(1),$$

and therefore

$$S_1 = o\left(\exp\left(\frac{N\rho}{Q}e^{-a}\right)\right). \quad (36)$$

It is clear that this estimate is obvious when  $\frac{N\rho}{Q}e^{-a} = O(1)$ .

To complete the proof it remains to recall that  $a_N = N^n Q^N$  and to combine (35), (36) and (23).

**Proof of Theorem 4.** We first recall the formula (13) for the exact distribution of the number  $m_0(n, N)$  of empty cells in the classical occupancy problem. We next replace  $n$  by (6) in relation (20) stated in the Lemma. Thus we get

$$S_{n,N} \sim N^n Q^N \exp\left\{\varepsilon^{-c} \frac{\rho}{Q} [1 + o(1)]\right\}, \quad (37)$$

as  $n, N \rightarrow \infty$ . Combining (13) and (37) with the exact distribution (12) obtained in Theorem 3, we conclude that

$$P[\mu_0(n, N) = k] \sim \left(1 + \frac{\rho}{Q}\right)^k \exp\left\{-\varepsilon^{-c} \frac{\rho}{Q} [1 + o(1)]\right\} P[m_0(n, N) = k], \quad k = 0, 1, \dots \quad (38)$$

One can easily check that

$$\frac{\rho}{Q} \rightarrow \begin{cases} \alpha_1 & \text{in case (i);} \\ \frac{b}{\alpha_2} & \text{in case (ii);} \\ 0 & \text{in case (iii);} \\ \alpha_4 & \text{in case (iv).} \end{cases}$$

Inserting these limits and the limiting distribution (7) of Theorem 1 into (38), we obtain the results of Theorem 4.

**Proof of Theorem 5.** One can still use here the reasoning that we applied to establish (38) to show that

$$P[\mu_0(n, N) = N - n + k] \sim \left(1 + \frac{\rho}{Q}\right)^{N-n+k} \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} P[m_0(n, N) = N - n + k], \quad (39)$$

$k = 0, 1, \dots$  This implies that  $\mu_0(n, N)$  and  $m_0(n, N)$  are asymptotically identically distributed if

$$\left(1 + \frac{\rho}{Q}\right)^{N-n+k} \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} \rightarrow 1$$

as  $n, N \rightarrow \infty$ . So, by (8) and (28) we find that

$$a = \sqrt{\frac{2d}{N}} + o\left(\frac{1}{\sqrt{N}}\right),$$

and thus

$$e^{-a} = 1 + O\left(\frac{1}{\sqrt{N}}\right). \quad (40)$$

Furthermore, the assumption  $n\rho e^\lambda = o(1)$  implies that

$$\frac{\rho}{Q} \sim \rho e^\lambda. \quad (41)$$

With the aid of (40) and (41) one can obtain

$$\begin{aligned} & \left(1 + \frac{\rho}{Q}\right)^{N-n+k} \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} \\ &= \exp\left\{(N - n + k) \log\left(1 + \frac{\rho}{Q}\right) - \frac{N\rho e^{-a}}{Q}\right\} \\ &= \exp\left\{\frac{(N - n + k)\rho}{Q} - \frac{(N - n + k)\rho^2}{2Q^2} - \frac{N\rho}{Q} \left[1 + O\left(\frac{1}{\sqrt{N}}\right)\right] + O\left(\frac{N\rho^3}{Q^3}\right)\right\} \\ &= \exp\left\{\frac{N\rho}{Q} + O\left(\frac{n\rho}{Q}\right) + O\left(\frac{N\rho^2}{Q^2}\right) - \frac{N\rho}{Q} + O\left(\frac{\sqrt{N}\rho}{Q}\right) + O\left(\frac{N\rho^3}{Q^3}\right)\right\} \\ &= \exp\left\{O(n\rho e^\lambda) + O(\sqrt{N}\rho e^\lambda)\right\} = \exp\left\{O(n\rho e^\lambda)\right\} = 1 + O(n\rho e^\lambda) = 1 + o(1). \end{aligned}$$

Inserting this into (39) we complete the proof.

## 4 Future Perspectives

In the present note we studied the asymptotic behavior of the number  $\mu_0(n, N)$  of empty cells in the zero-inflated four-parameter model when  $n$  and  $N$  satisfy either (6) or (8), (the “right” and “left” domains of the parameters  $n$  and  $N$ ). We conjecture a weak convergence to the Gaussian distribution and the validity of the corresponding local limit theorems as long as  $(n, N)$  belongs to the “central” domain, i.e. when  $\frac{n}{N}$  is bounded, and when  $\frac{n}{N} - \log N \rightarrow -\infty$ . Moreover, we consider the result of Theorem 5 as incomplete and plan to study the changes in the limiting distribution (9) under weaker assumptions on  $\rho$  and  $\lambda$ . Finally, we also note that the study of similar problems for random mappings of an  $n$  element set into itself and random rooted trees on  $n$  labeled vertices can be transferred into allocation problems in the “central” domain, see Kolchin *et al.* (1978).

## 5 Acknowledgments

The preparation of this paper was assisted by FAPESP Project No 2000/13505-3.

## References

- [1] Békéssy, A. (1963). On classical occupancy problems I. *Magyar Tud. Acad. Mat. Kutató Int. Közl.* **8**, 59-71.
- [2] Feller, W. (1967). *An Introduction to Probability Theory and Its Applications*, Vol. I. 2nd ed., Wiley, New York.
- [3] Johnson, N., Kotz, S. and Kemp, A. (1992). *Univariate Discrete Distributions*. 2nd. ed., Wiley, New York.
- [4] Kolchin, V.F. (1986). *Random Mappings*. Optimization Software Inc., New York.
- [5] Kolchin, V.F., Sevast'yanov, B.A. and Chistyakov, V.P. (1978). *Random Allocations*. Wiley, New York.
- [6] Kolev, N., Minkova, L. and Neytchev, P. (2000). Inflated-parameter family of generalized power series distributions and their application in analysis of overdispersed data. *ARCH Research Clearing House* **2**, 295-320.
- [7] von Mises, R. (1939). Über Abtaltung und Besetzungswahrscheinlichkeiten. *Revue de la Faculté des Sciences de l'Université d'Istanbul* **4**, 145-163.
- [8] Sachkov, V.N. (1978). *Probabilistic Methods in Combinatorial Analysis* (in Russian). Nauka, Moscow.
- [9] Riordan, J. (1958). *An Introduction to Combinatorial Analysis*. Wiley, New York.
- [10] Winkelmann, R. (2000). *Econometric Analysis of Count Data*, 3rd. ed. Springer, Berlin.

## ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2001-01 - KOTTAS, A., BRANCO, M.D., GELFAND, A.E., A Nonparametric bayesian modeling approach for cytogenetic dosimetry. 2001. 19p. (RT-MAE-2001-01)

2001-02 - AOKI, R., BOLFARINE, H., SINGER, J.M., Null Intercept measurement error regression models. 2001. 18p. (RT-MAE-2001-02)

2001-03 - AOKI, R., BOLFARINE, H., SINGER, J.M., Asymptotic Efficiency of null intercept measurement error regression models. 2001. 11p. (RT-MAE-2001-03).

2001-04 - ALMEIDA, S.S., LIMA, C.R.O.P., SANDOVAL, M.C., Linear Calibration in functional models without the normality assumption. 2001. 18p. (RT-MAE-2001-04).

2001-05 - GARCIA-ALFARO, K.E., BOLFARINE, H. Comparative calibration with subgroups. 2001. 18p. (RT-MAE-2001-05)

2001-06 - DUNLOP, F.M., FERRARI, P.A., FONTES, L.R.G. A dynamic one-dimensional interface interacting with a wall. 2001. 21p. (RT-MAE-2001-06).

2001-07 - FONTES, L.R., ISOPI, M., NEWMAN, C.M., STEIN, D.L. 1D Aging. 2001. 10p. (RT-MAE-2001-07).

2001-08 - BARROS, S.R.M., FERRARI, P.A., GARCIA, N.L., MARTÍNEZ, S. Asymptotic behavior of a stationary silo with absorbing walls. 2001. 20p. (RT-MAE-2001-08).

2001-09 - GONZALEZ-LOPEZ, V.A., TANAKA, N.Y. Characterization of copula and its relationships with  $TP_2$  ( $RR_2$ ) association. 2001. 24p. (RT-MAE-2001-09).

2001-10 - ORTEGA, E.M.M., BOLFARINE, H., PAULA, G.A. Influence Diagnostics in Generalized Log-Gamma Regression Models. 2001. 24p. (RT-MAE-2001-10).

2001-11 - M.D, BRANCO, BOLFARINE, H., IGLESIAS, P., ARELLANO-VALLE, R.B. Bayesian and classical solutions for binomial cytogenetic dosimetry problem. 2001. 16p. (RT-MAE-2001-11).

2001-12 - GONÇALEZ-LOPES, V.A., TANAKA, N.I. Dependence structures and a posteriori distributions. 2001. 41p. (RT-MAE-2001-12).

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística  
IME-USP  
Caixa Postal 66.281  
05315-970 - São Paulo, Brasil