

## A ZERO-INFLATED OCCUPANCY DISTRIBUTION: EXACT RESULTS AND POISSON CONVERGENCE

NIKOLAI KOLEV and LJUBEN MUTAFCHIEV

Received 2 September 2002

We introduce the generalized zero-inflated allocation scheme of placing  $n$  labeled balls into  $N$  labeled cells. We study the asymptotic behavior of the number of empty cells when  $(n, N)$  belongs to the “right” and “left” domain of attraction. An application to the estimation of characteristics of agreement among a set of raters which independently classify subjects into one of two categories is also indicated. The case when a large number of raters acts following the zero-inflated binomial law with small probability for positive diagnosis is treated using the zero-inflated Poisson approximation.

2000 Mathematics Subject Classification: 60C05, 60F05, 62H10.

**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 [2], Kolchin et al. [4]). If in this model, we let  $\eta_j$  denote 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, that is,

$$P(\eta_1 = k_1, \eta_2 = k_2, \dots, \eta_N = k_N) = \frac{n!}{k_1!k_2! \dots k_N!N^n}, \quad (1.1)$$

where the integers  $k_j$  satisfy

$$k_1 + k_2 + \dots + k_N = n, \quad k_j \geq 0, \quad j = 1, 2, \dots, N. \quad (1.2)$$

Consider now  $N$  independent and identically distributed random variables  $\xi_j$  with

$$P(\xi_1 = k) = p_k, \quad k = 0, 1, \dots, \quad \sum_{k=0}^{\infty} p_k = 1, \quad (1.3)$$

such that

$$P(\eta_1 = k_1, \dots, \eta_N = k_N) = P(\xi_1 = k_1, \dots, \xi_N = k_N \mid \xi_1 + \dots + \xi_N = n). \quad (1.4)$$

For  $k_j$ 's satisfying (1.2), one obtains the so-called generalized scheme of allocation defined by (1.4) (see [4, Chapter 2]). In the Poisson case, that is, if  $p_k = \lambda^k e^{-\lambda} (k!)^{-1}$ ,  $k = 0, 1, \dots$ , we get the classical scheme (1.1).

In the present paper, we consider the generalized scheme of allocation that addresses the problem that the count data displays a higher fraction of zeros or nonoccurrences. More precisely, we introduce a common alteration making the probability of zero (in the Poisson distribution formula) greater relative to the underlying nonzero probabilities. For properties, generalizations, and certain statistical and economical applications of zero-inflated distributions, we refer the reader to Johnson et al. [3, Chapter 8] and Winkelmann [8, Chapters 3 and 4]. In what follows, we assume that (1.3) and (1.4) are valid with a distribution

$$\begin{aligned} p_0 &= \rho + (1 - \rho)e^{-\lambda}, \\ p_k &= (1 - \rho)\lambda^k e^{-\lambda} (k!)^{-1}, \quad k = 1, 2, \dots, \lambda > 0, 0 < \rho < 1. \end{aligned} \quad (1.5)$$

Our aim in this paper is to derive the exact distribution of the number  $\mu_0(n, N)$  of empty cells in the generalized allocation model defined by (1.3), (1.4), and (1.5) 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. Considering the exact distribution of  $\mu_0(n, N)$ , we also obtain a zero-inflated generalization of the multinomial distribution (1.1) (see Section 3.1, formula (3.6)).

Our study is also motivated by an application of these occupancy distributions to statistical problems of estimating measures of agreement among a set of raters which independently classify subjects into one of two categories. Following the model description of Verducci et al. [6, Section 3.3], we assume that the population of subjects consists of two subpopulations: those that possess the characteristic in question and those that do not. A set of  $R$  raters acting independently is also given. The subjects which do not possess the characteristic are never misdiagnosed, but any subject possessing the characteristic has fixed probability  $\pi$  of receiving a positive diagnosis. If the prevalence of the characteristic in the population is  $1 - \rho$ , then the distribution of the number  $\xi$  of positive responses is a zero-inflated binomial one, that is,

$$\begin{aligned} P(\xi = 0) &= \rho + (1 - \rho)(1 - \pi)^R, \\ P(\xi = k) &= (1 - \rho) \binom{R}{k} \pi^k (1 - \pi)^{R-k}, \quad k = 1, \dots, R. \end{aligned} \quad (1.6)$$

If the number of raters  $R$  is large and  $\pi = \lambda/R$ ,  $\lambda > 0$ , one obtains the Poisson approximation of this distribution, given by (1.5). Let  $\xi_1, \dots, \xi_N$  be a random

sample from model (1.5) and let  $\mu_0(N) = \sum_{i=1}^N I_{(\xi_i=0)}$ , where  $I_{(\xi=0)}$  is the indicator of the event  $\{\xi = 0\}$ . Furthermore, we restrict the probability space, where the observed number of perfect negative agreements  $\mu_0(N)$  is defined, by equating the total number  $\xi_1 + \dots + \xi_N$  of positive responses to a new parameter  $n$ . Thus, we get the new random variable  $\mu_0(n, N)$  that counts the number of negative agreements in a sample of size  $N$ , given that the total number  $n$  of positive responses is preliminarily known.

We include two classical Poisson convergence results for  $m_0(n, N)$ , the number of empty cells in the classical occupancy scheme (1.1). We will 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 [7]. It can be also found in certain famous textbooks and monographs (see, e.g., [2, Chapter 4] and [4, Chapter 1]).

**THEOREM 1.1.** *If  $n, N \rightarrow \infty$  so that*

$$n = N \log N + cN + o(N), \quad -\infty < c < \infty, \tag{1.7}$$

then

$$P[m_0(n, N) = k] = \frac{e^{-kc}}{k!} \exp(-e^{-c}) + o(1), \quad k = 0, 1, \dots \tag{1.8}$$

Relation (1.7) shows that the number  $n$  of balls is larger relative to the number  $N$  of cells. It also defines the so-called “right” domain of allocations, see [4, Chapter 1].

The next theorem belongs to Békéssy [1] and concerns the “left” domain of allocation.

**THEOREM 1.2.** *If  $n, N \rightarrow \infty$  so that*

$$n^2 (2N)^{-1} \rightarrow d \in (0, \infty), \tag{1.9}$$

then

$$P[m_0(n, N) - (N - n) = k] = \frac{d^k e^{-d}}{k!} + o(1), \quad k = 0, 1, \dots \tag{1.10}$$

In 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.

**2. Summary of results.** First, we introduce some conventions and notations that will be needed later. Throughout the paper, we assume that our additional

parameters  $\rho$  and  $\lambda$  in the zero-inflated allocation scheme satisfy the dependencies  $\rho = \rho(n, N)$  and  $\lambda = \lambda(n, N)$ . For the sake of convenience, we write

$$Q = Q(\rho, \lambda) = (1 - \rho)e^{-\lambda}. \tag{2.1}$$

We also let

$$S_{n,N} = \sum_{k=0}^N \binom{N}{k} Q^k \rho^{N-k} k^n = \sum_{k=0}^N a_k. \tag{2.2}$$

Next, we let  $\Delta$  denote the finite difference operator (see, e.g., Riordan [5, Chapter 1]) as follows. For any function  $f(x)$  defined at points  $x, x + 1, \dots, x + r$ , where  $r$  is a natural number, let  $\Delta f(x) = f(x + 1) - f(x)$  and let  $\Delta^r f(x) = \Delta[\Delta^{r-1} f(x)]$  if  $r \geq 2$ . Finally, we introduce the notation  $\Delta^r 0^s$ ,  $s \geq 0$ , for the combinatorial numbers that will be used in the next statement by setting first  $f(x) = x^s$  in the finite difference operator  $\Delta^r$ , and then, by substituting  $x = 0$  in it, that is,  $\Delta^r 0^s = \Delta^r x^s|_{x=0}$ .

**THEOREM 2.1.** *Under the above notations, the following relation is true:*

$$P[\mu_0(n, N) = k] = \frac{(\rho + Q)^k Q^{N-k}}{S_{n,N}} \binom{N}{k} \Delta^{N-k} 0^n, \quad k = 0, 1, \dots, N. \tag{2.3}$$

**NOTE 2.2.** The corresponding result in the classical scheme of allocation states that

$$P[m_0(n, N) = k] = \binom{N}{k} N^{-n} \Delta^{N-k} 0^n, \quad k = 0, 1, \dots, N, \tag{2.4}$$

see [5, Chapter 5]. If we set  $\rho = 0$ , one can easily check that

$$P[\mu_0(n, N) = k] = P[m_0(n, N) = k], \quad k = 0, 1, \dots, N. \tag{2.5}$$

Now, we proceed to asymptotic results in the “right” domain. We apply a weak convergence to the Poisson distribution of type

$$P[\mu_0(n, N) = k] = \frac{[(1 + q)e^{-c}]^k}{k!} \exp\{- (1 + q)e^{-c}\} + o(1), \quad k = 0, 1, \dots, \tag{2.6}$$

where the parameter  $q \geq 0$  is specified in the next theorem.

**THEOREM 2.3.** *Suppose that  $n, N \rightarrow \infty$  so that (1.7) holds.*

- (i) *If  $N\rho \rightarrow 0$  and  $\rho e^\lambda \rightarrow \alpha_1 \in [0, \infty)$ , then (2.6) is valid with  $q = \alpha_1$ .*
- (ii) *If  $N\rho \rightarrow b \in (0, \infty)$  and  $Ne^{-\lambda} \rightarrow \alpha_2 \in (0, \infty)$ , then (2.6) is valid with  $q = b/\alpha_2$ .*

- (iii) If  $N\rho \rightarrow b \in (0, \infty)$  and  $Ne^{-\lambda} \rightarrow \infty$ , then (2.6) is valid with  $q = \alpha_3 = 0$ .
- (iv) If  $N\rho \rightarrow \infty$  and  $\rho e^\lambda / (1 - \rho) \rightarrow \alpha_4 \in [0, \infty)$ , then (2.6) is valid with  $q = \alpha_4$ .

In the “left” domain defined by (1.9), we find a sufficient condition for  $\mu_0(n, N)$  and  $m_0(n, N)$  to be asymptotically identically distributed.

**THEOREM 2.4.** *If  $n, N \rightarrow \infty$  in a way that they satisfy (1.9) 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 (1.10).*

### 3. Proofs

**3.1. Proof of Theorem 2.1.** First, using (1.5), we will derive

$$\begin{aligned}
 P(\xi_1 = k_1, \dots, \xi_N = k_N \mid \xi_1 + \dots + \xi_N = n) \\
 = \frac{P(\xi_1 = k_1, \dots, \xi_N = k_N, \xi_1 + \dots + \xi_N = n)}{P(\xi_1 + \dots + \xi_N = n)}
 \end{aligned}
 \tag{3.1}$$

for  $k_j, j = 1, \dots, N$ , satisfying (1.2). To find  $P(\xi_1 + \dots + \xi_N = n)$ , we can use generating functions. Since  $E(x^{\xi_1}) = \rho + (1 - \rho)e^{\lambda(x-1)}$ , by (1.5), one can write

$$E(x^{\xi_1 + \dots + \xi_N}) = [\rho + (1 - \rho)e^{\lambda(x-1)}]^N.
 \tag{3.2}$$

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

$$\begin{aligned}
 P(\xi_1 + \dots + \xi_N = n) &= [x^n][\rho + (1 - \rho)e^{\lambda(x-1)}]^N \\
 &= \sum_{k=0}^N \binom{N}{k} (1 - \rho)^k e^{-\lambda k} \rho^{N-k} [x^n]e^{\lambda k x} \\
 &= \frac{\lambda^n}{n!} \sum_{k=0}^N \binom{N}{k} (1 - \rho)^k e^{-\lambda k} \rho^{N-k} k^n,
 \end{aligned}
 \tag{3.3}$$

that is, with  $S_{n,N}$  determined by (2.2) and (2.1), we obtain

$$P(\xi_1 + \dots + \xi_N = n) = \frac{\lambda^n}{n!} S_{n,N}.
 \tag{3.4}$$

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 (1.5), we get

$$\begin{aligned}
 P(\xi_1 = k_1, \dots, \xi_N = k_N, \xi_1 + \dots + \xi_N = n) \\
 = \frac{[\rho + (1 - \rho)e^{-\lambda}]^k [(1 - \rho)e^{-\lambda}]^{N-k} \lambda^n}{k_1! \dots k_N!}.
 \end{aligned}
 \tag{3.5}$$

Inserting first (3.4) and (3.5) into (3.1), and then going back to definition (1.4), we find that

$$P(\eta_1 = k_1, \dots, \eta_N = k_N) = \frac{n! [\rho + (1-\rho)e^{-\lambda}]^k [(1-\rho)e^{-\lambda}]^{N-k}}{k_1! \cdots k_N! S_{n,N}} \quad (3.6)$$

for any vector  $(k_1, \dots, k_N)$  whose coordinates satisfy (1.2) such that  $k$  of them equals 0.

To find  $P[\mu_0(n, N) = k]$ , we have to sum the probabilities of (3.6) over all vectors  $(k_1, \dots, k_N)$  with  $k$  zero coordinates satisfying (1.2). Since the places of the empty cells can be chosen in  $\binom{N}{k}$  ways, we can write

$$\begin{aligned} P[\mu_0(n, N) = k] \\ = \binom{N}{k} \frac{[\rho + (1-\rho)e^{-\lambda}]^k [(1-\rho)e^{-\lambda}]^{N-k}}{S_{n,N}} \sum \frac{n!}{k_1! \cdots k_{N-k}!}, \end{aligned} \quad (3.7)$$

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

Now, using a relation between the Stirling numbers of the second kind and the finite difference operator  $\Delta$  (see, e.g., [5, Chapter 5]) by notation (2.1), we get expression (2.3) of the theorem.

**NOTE 3.1.** Relation (3.6) may be also regarded as an *inflated parameter generalization of the multinomial distribution* with coinciding probabilities of outcomes. It may have applications in such diverse areas as start-up demonstration tests, formulation of connection acceptance control in an asynchronous transfer model telecommunications network, clinical trials, and so forth.

**3.2. Proofs of the limit theorems.** Our first task here is to find the asymptotic form of  $S_{n,N}$  given by (2.2). The result may be stated as in the following lemma.

**LEMMA 3.2.** *If  $n, N \rightarrow \infty$  so that*

$$\frac{\rho\sqrt{n}}{Q} e^{-n/N} \rightarrow 0, \quad \frac{\rho\sqrt{N}}{Q} e^{-n/N} \rightarrow 0, \quad (3.8)$$

then

$$S_{n,N} \sim N^n Q^N \exp\left\{\frac{\rho N}{Q} e^{-n/N}\right\}. \quad (3.9)$$

**PROOF.** For the general term  $a_k$  of  $S_{n,N}$  in (2.2), it is straightforward that

$$\frac{a_k}{a_N} = \binom{N}{k} \left(\frac{\rho}{Q}\right)^{N-k} \left(\frac{k}{N}\right)^n, \quad k = 0, 1, \dots, N. \quad (3.10)$$

We apply the obvious identity

$$\frac{a_k}{a_N} = \prod_{j=0}^{N-k-1} \frac{a_{k+j}}{a_{k+j+1}} \tag{3.11}$$

and find that

$$\frac{a_k}{a_N} = \left(\frac{N\rho}{Q}\right)^{N-k} \frac{1}{(N-k)!} \prod_{j=0}^{N-k-1} \left(\frac{k+j}{k+j+1}\right)^{n-1} \frac{k+j}{N}. \tag{3.12}$$

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

$$S_{n,N} = a_N(S_1 + S_2), \tag{3.13}$$

where

$$S_1 = \sum_{0 \leq k \leq N-w(N)} \frac{a_k}{a_N}, \quad S_2 = \sum_{N-w(N) < k \leq N} \frac{a_k}{a_N}, \tag{3.14}$$

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

$$w(N) = o(N). \tag{3.15}$$

We will subsequently restrict  $w(N)$  by (3.21) and (3.25).

Consider first  $S_2$  with a general term given by (3.12). For  $N - w(N) < k$  and  $0 \leq j \leq N - k - 1$ , we get  $1 - w(N)/N < (k + j)/N \leq 1 - 1/N$ , which means that

$$\frac{k+j}{N} = 1 + O\left(\frac{w(N)}{N}\right). \tag{3.16}$$

Then we have

$$\begin{aligned} \left(\frac{k+j}{k+j+1}\right)^{n-1} &= \exp\left\{(n-1)\log\left(1 - \frac{1}{k+j+1}\right)\right\} \\ &= \exp\left\{-\frac{n(1-1/n)}{N(k+j+1)/N} - R(n,k,j)\right\}, \end{aligned} \tag{3.17}$$

that is,

$$\left(\frac{k+j}{k+j+1}\right)^{n-1} = \exp\left\{-a + O\left(\frac{aw(N)+1}{N}\right) - R(n,k,j)\right\}, \tag{3.18}$$

where, for the sake of brevity,  $R(n,k,j) = (n-1)[2^{-1}(k+j+1)^{-2} + 3^{-1}(k+j+1)^{-3} + \dots]$  and  $a = a(n,N) = n/N$ .

From (3.15), it follows that

$$R(n, k, j) = O\left(\frac{n}{N^2}\right) = O\left(\frac{a}{N}\right) \tag{3.19}$$

uniformly for  $N - w(N) + 1 < k + j + 1$ . Relations (3.18) and (3.19) imply now that

$$\left(\frac{k+j}{k+j+1}\right)^{n-1} = \exp\left\{-a + O\left(\frac{aw(N)}{N}\right)\right\} \tag{3.20}$$

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

In what follows, we need a restriction on  $w(N)$  stronger than that in (3.15). So, for  $n, N \rightarrow \infty$ , we assume further that

$$w(N) = o\left(\frac{N}{\sqrt{n}}\right), \quad w(N) = o(\sqrt{N}). \tag{3.21}$$

Inserting relations (3.16), (3.20), and (3.21) into the product in (3.12) and using the fact that  $N - k < w(N)$  in  $S_2$  (see (3.14)), we find that

$$\begin{aligned} & \prod_{j=0}^{N-k-1} \left(\frac{k+j}{k+j+1}\right)^{n-1} \frac{k+j}{N} \\ &= \exp\left\{(N-k)\left[-a + O\left(\frac{aw(N)}{N}\right)\right]\right\} \left[1 + O\left(\frac{w(N)}{N}\right)\right]^{N-k} \\ &= e^{-a(N-k)} \exp\left\{(N-k)\left[O\left(\frac{aw(N)}{N}\right) + \log\left(1 + O\left(\frac{w(N)}{N}\right)\right)\right]\right\} \tag{3.22} \\ &= e^{-a(N-k)} \exp\left\{O\left(w(N)\left[O\left(\frac{aw(N)}{N}\right) + \log\left(1 + O\left(\frac{w(N)}{N}\right)\right)\right]\right)\right\} \\ &= e^{-a(N-k)} \left[1 + O\left(\frac{aw^2(N)}{N}\right) + O\left(\frac{w^2(N)}{N}\right)\right] \\ &= e^{-a(N-k)} [1 + o(1)]. \end{aligned}$$

The last relation is justified by (3.21). Thus, for  $S_2$  from (3.14), we get

$$\begin{aligned} S_2 &= [1 + o(1)] \sum_{N-w(N) < k \leq N} \frac{(N\rho/Q)^{N-k}}{(N-k)!} e^{-a(N-k)} \\ &= [1 + o(1)] \sum_{0 \leq l < w(N)} \frac{(N\rho/Q)^l}{l!} e^{-al}. \end{aligned} \tag{3.23}$$

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

$$S_2 \sim \exp\left(\frac{N\rho}{Q} e^{-a}\right) \tag{3.24}$$

as  $n, N \rightarrow \infty$ .

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

$$w(N) = \frac{N\rho}{Q} e^{-a + \gamma(n,N)} \sqrt{\frac{N\rho}{Q} e^{-a}}, \tag{3.25}$$

where  $\gamma(n,N)$  is a function tending to  $\infty$  arbitrarily slowly so that  $(N\rho/Q)e^{-a}$  determines its basic growth. Conditions (3.8) imply that this choice of  $w(N)$  agrees with requirements (3.21). Now, it remains to apply the normal approximation of the Poisson distribution with parameter  $M = (N\rho/Q)e^{-a}$  (see [2, Chapter 2]):

$$\sum_{l < M + \gamma(n,N)\sqrt{M}} \frac{M^l e^{-M}}{l!} = \Phi(\gamma(n,N)) + o(1), \tag{3.26}$$

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

$$S_2 \sim \exp\left(\frac{N\rho}{Q} e^{-a}\right). \tag{3.27}$$

Finally, if  $\liminf (N\rho/Q)e^{-a} = M < \infty$  and  $\limsup (N\rho/Q)e^{-a} = \infty$ , then any sequence presenting  $(N\rho/Q)e^{-a}$  and converging to a point of accumulation  $M_0 \in [M, \infty)$  and any  $w(N) \rightarrow \infty$  satisfying requirements (3.21) yield the convergence of  $S_2$  to  $e^{M_0}$ . This shows that the asymptotic equivalence (3.27) is again valid in this particular case.

We now turn our attention to the easier task to find an estimate for  $S_1$ . Substituting  $a_k/a_N$  by expression (3.10) and changing the summation index, we obtain

$$S_1 = \sum_{w(N) \leq l \leq N} \binom{N}{l} \left(\frac{\rho}{Q}\right)^l \left(\frac{N-l}{N}\right)^n. \tag{3.28}$$

Since

$$\left(\frac{N-l}{N}\right)^n = \left(1 - \frac{l}{N}\right)^n = \exp\left\{n \log\left(1 - \frac{l}{N}\right)\right\} < \exp\left\{-\frac{nl}{N}\right\} = e^{-al} \tag{3.29}$$

and  $\binom{N}{l} < N^l/l!$ , we get

$$S_1 < \sum_{w(N) \leq l \leq N} \left(\frac{N\rho}{Q} e^{-a}\right)^l \frac{1}{l!}. \tag{3.30}$$

The last sum can be easily estimated using again the normal approximation of the Poisson distribution (3.26) with  $w(N)$  given by (3.25) in the case  $M = (N\rho/Q)e^{-a} \rightarrow \infty$ . We have

$$S_1 e^{-M} < \sum_{w(N) \leq l \leq N} \frac{M^l e^{-M}}{l!} = 1 - \Phi(y(n, N)) + o(1) = o(1), \tag{3.31}$$

and therefore

$$S_1 = o\left(\exp\left(\frac{N\rho}{Q}e^{-a}\right)\right). \tag{3.32}$$

This estimate is obvious either when  $(N\rho/Q)e^{-a} = O(1)$  or when

$$\liminf\left(\frac{N\rho}{Q}\right)e^{-a} < \infty, \quad \limsup N\rho Q e^{-a} = \infty, \tag{3.33}$$

and  $(N\rho/Q)e^{-a}$  follows a sequence that converges to a point of accumulation  $M_0 < \infty$ . To complete the proof, it remains to recall that  $a_N = N^n Q^N$  and to combine (3.27), (3.32), and (3.13). □

**PROOF OF THEOREM 2.3.** We first recall formula (2.4) for the exact distribution of the number  $m_0(n, N)$  of empty cells in the classical occupancy problem. Then, note that both convergences in (3.8) are valid under the assumptions of Theorem 2.3. So, replacing (1.7) into relation (3.9), we get

$$S_{n,N} \sim N^n Q^N \exp\left\{e^{-c} \frac{\rho}{Q} [1 + o(1)]\right\} \tag{3.34}$$

as  $n, N \rightarrow \infty$ . Combining (2.4) and (3.34) with the exact distribution (2.3) obtained in Theorem 2.1, we conclude that

$$\begin{aligned} &P[\mu_0(n, N) = k] \\ &\sim \left(1 + \frac{\rho}{Q}\right)^k \exp\left\{-e^{-c} \frac{\rho}{Q} [1 + o(1)]\right\} P[m_0(n, N) = k], \quad k = 0, 1, \dots \end{aligned} \tag{3.35}$$

One can easily check that

$$\frac{\rho}{Q} \rightarrow \begin{cases} \alpha_1 & \text{in case (i),} \\ \frac{b}{\alpha_2} & \text{in case (ii),} \\ 0 & \text{in case (iii),} \\ \alpha_4 & \text{in case (iv).} \end{cases} \tag{3.36}$$

Inserting these limits and the limiting distribution (1.8) into (3.35), we obtain the results of Theorem 2.3. □

**PROOF OF THEOREM 2.4.** One can still use here the reasoning we applied to establish (3.35) to show that

$$\begin{aligned}
 &P[\mu_0(n, N) = N - n + k] \\
 &\sim \left(1 + \frac{\rho}{Q}\right)^{N-n+k} \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} P[m_0(n, N) = N - n + k], \tag{3.37}
 \end{aligned}$$

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

$$\left(1 + \frac{\rho}{Q}\right)^{N-n+k} \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} \rightarrow 1 \tag{3.38}$$

as  $n, N \rightarrow \infty$ . So, by (1.9) and  $a = n/N$ , we find that  $a = \sqrt{2d/N} + o(1/\sqrt{N})$ , and thus

$$e^{-a} = 1 + O\left(\frac{1}{\sqrt{N}}\right). \tag{3.39}$$

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

$$\frac{\rho}{Q} \sim \rho e^\lambda. \tag{3.40}$$

With the aid of (3.39) and (3.40), one can obtain

$$\begin{aligned}
 &\left(1 + \frac{\rho}{Q}\right)^{N-n+k} \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} \\
 &= \exp\left\{(N - n + k) \log\left(1 + \frac{\rho}{Q}\right) - \frac{N\rho e^{-a}}{Q}\right\} \\
 &= \exp\left\{\frac{N\rho}{Q} + O\left(\frac{n\rho}{Q}\right) + O\left(\frac{N\rho^2}{Q^2}\right) - \frac{N\rho}{Q} + O\left(\frac{\sqrt{N}\rho}{Q}\right) + O\left(\frac{N\rho^3}{Q^3}\right)\right\} \tag{3.41} \\
 &= \exp\{O(n\rho e^\lambda) + O(\sqrt{N}\rho e^\lambda)\} \\
 &= \exp\{O(n\rho e^\lambda)\} = 1 + O(n\rho e^\lambda) = 1 + o(1).
 \end{aligned}$$

Inserting this into (3.37), we complete the proof. □

**NOTE 3.3.** Under the second assumption in (3.8), one can study in a similar way the behavior of  $\mu_0(n, N)$  in the “central” domain, where  $a = n/N$  is bounded away from  $-\infty$  or  $\infty$ . In this case, for

$$k = Ne^{-a} + u\sqrt{N}, \quad |u| \leq C < \infty, \tag{3.42}$$

it is easy to verify that

$$\begin{aligned} & \left(1 + \frac{\rho}{Q}\right)^k \exp\left\{-e^{-a} \frac{N\rho}{Q}\right\} \\ &= \exp\left\{\frac{u\rho\sqrt{N}}{Q} - \frac{N\rho^2 e^{-a}}{2Q^2} O\left(\frac{\sqrt{N}\rho^2}{Q^2}\right) + O\left(\frac{N\rho^3}{Q^3}\right)\right\} = 1 + o(1). \end{aligned} \quad (3.43)$$

Hence,  $P[\mu_0(n, N) = k] \sim P[m_0(n, N) = k]$  uniformly for all  $k$  satisfying (3.42). This implies a local limit theorem for  $\mu_0(n, N)$  of the same form as stated in [4, Chapter 2, Theorem 2.2].

**4. Conclusions.** In the present paper, we studied the “right” and “left” domains of the parameters  $n$  and  $N$ . We conjecture a weak convergence of  $\mu_0(n, N)$  to the Gaussian distribution and the validity of the corresponding local limit theorems as long as  $(n, N)$  belongs to the “central” domain, that is, when  $n/N$  is bounded. We also plan to study the changes in the limiting distribution (1.10) under weaker assumptions on  $\rho$  and  $\lambda$ .

**ACKNOWLEDGMENTS.** We thank the referees for helpful comments and suggestions. The preparation of this paper was assisted by FAPESP Project no. 2000/13505-3 during the visit of the second author to the Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, Brazil.

#### REFERENCES

- [1] A. Békéssy, *On classical occupancy problems. I*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **8** (1963), 59–71.
- [2] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. 1*, 2nd ed., Izdat. Mir, Moscow, 1967 (Russian).
- [3] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, 2nd ed., Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, 1992.
- [4] V. F. Kolchin, B. A. Sevast'yanov, and V. P. Chistyakov, *Random Allocations*, John Wiley & Sons, New York, 1978.
- [5] J. Riordan, *An Introduction to Combinatorial Analysis*, Wiley Publications in Mathematical Statistics, John Wiley & Sons, New York, 1958.
- [6] J. S. Verducci, M. E. Mack, and M. H. DeGroot, *Estimating multiple rater agreement for a rare diagnosis*, J. Multivariate Anal. **27** (1988), no. 2, 512–535.
- [7] R. von Mises, *Über Aufteilungs- und Besetzungswahrscheinlichkeiten*, Rev. Fac. Sci. Univ. Istanbul, N.S. **4** (1939), 145–163 (German).
- [8] R. Winkelmann, *Econometric Analysis of Count Data*, 3rd ed., Springer-Verlag, Berlin, 2000.

Nikolai Kolev: Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, CP 66281, 05311-970 São Paulo, Brazil  
E-mail address: [nkolev@ime.usp.br](mailto:nkolev@ime.usp.br)

Ljuben Mutafchiev: American University in Bulgaria, 2700 Blagoevgrad, Bulgaria  
E-mail address: [ljuben@aubg.bg](mailto:ljuben@aubg.bg)