Z. Wahrscheinlichkeitstheorie verw. Geb. 21, 109–120 (1972)

Asymptotic Normality in a Generalized Occupancy Problem

LARS HOLST

1. Introduction and Notation

Let us suppose that n balls are distributed among N cells so that each ball may fall into the $k:th$ cell with probability p_k, $p_1+\cdots+p_N=1$, independently of what happens to the other balls. To every cell a real number is associated, the *occupancy value*, a_k for the $k:th$ cell. We call $\{(p_k, a_k), k=1, \ldots, N\}$ a *cell situation.* The *occupancy sum*, Z_n, is the sum of the occupancy values for the cells which contain at most r balls, r being a fixed number.

If $r=0$, $a_1=\cdots=a_N=1$, and $p_1=\cdots=p_N=1/N$, then Z_n is the number of empty cells in the classical occupancy problem [1, 4]. Therefore we call the problem of determining the distribution of Z_n, in the general situation, the *generalized occupancy problem.*

A number of papers deal with limiting distributions of Z_n in various special cases indicated by names such as: occupancy problems [6, 8, 9], the coupon collector's problem [7], the empty cell test [2, 5], only to mention a few of the names and papers.

The above situation has also applications in sampling theory. Suppose that we sample with varying probabilities and with replacement n times from a finite population and that we use the Horwitz-Thompson estimator as an estimator of the population total. The limiting behaviour of this estimator can be deduced from that of Z_n.

The main aim of this paper is to show that, when n and N increase, under general conditions, Z_n is asymptotically normally distributed. Rosén [7] treats the case $r=0$ under the same conditions.

In order to give a precise formulation of the asymptotic behaviour of Z_n when n and N increase, we consider a sequence

$$\{(p_{kv}, a_{kv}), k=1, \ldots, N_v\}, \qquad v=1, 2, \ldots,$$

of cell situations. Let d be an arbitrary natural number and consider simultaneously the occupancy sums after the distribution of n_{1v}, $n_{1v}+n_{2v}$, $\ldots$, $n_{1v}+n_{2v}+\cdots+n_{dv}$ balls. Call these occupancy sums

$$Z_{n_{1v}}, Z_{n_{1v}+n_{2v}}, \ldots, Z_{n_{1v}+\cdots+n_{dv}}$$

and set

$$V_{N_v}=(Z_{n_{1v}}, Z_{n_{1v}+n_{2v}}, \ldots, Z_{n_{1v}+\cdots+n_{dv}}), \qquad v=1, 2, \ldots.$$

In the following we assume that

$$N_\nu \to \infty, \quad \text{when } \nu \to \infty,$$

$$0 < \liminf_{\nu\to\infty} n_{s\nu}/N_\nu \leqq \limsup_{\nu\to\infty} n_{s\nu}/N_\nu < \infty, \quad s=1, \ldots, d,$$

and

$$N_\nu p_{k\nu} \leqq C < \infty, \quad k=1, \ldots, N_\nu, \ \nu=1, 2, \ldots, \text{for some real number } C.$$

To facilitate the writing the index ν will be suppressed.

In Section 2 we compute the characteristic function of V_N. Some auxiliary results concerning moments of V_N are obtained in Section 3. Asymptotic normality of V_N is proved in Section 4.

2. The Characteristic Function

Consider a cell situation and denote by ξ_{ks} the number of balls placed in the $k{:}th$ cell after the distribution of n_s balls. Obviously $(\xi_{1s}, \ldots, \xi_{Ns})$ is multinomial $(n_s, p_1, \ldots, p_N)$, i.e.

$$P(\xi_{1s}=y_1, \ldots, \xi_{Ns}=y_N) = \frac{n_s!}{y_1! \ldots y_N!} \cdot p_1^{y_1} \ldots p_N^{y_N}, \qquad \sum_{k=1}^{N} y_k = n_s.$$

Let $\varepsilon_{ks}=1$, if the $k{:}th$ cell contains at most r balls after the distribution of $n_1+\cdots+n_s$ balls, i.e., if $\xi_{k1}+\cdots+\xi_{ks} \leqq r$, $\varepsilon_{ks}=0$ otherwise, $k=1, \ldots, N$, $s=1, \ldots, d$. Introduce the generating function

$$A_N(z_1, \ldots, z_d) = \sum_{n_1, \ldots, n_d=0}^{\infty} E_{n_1, \ldots, n_d}\left(\prod_{k,s} x_{ks}^{\varepsilon_{ks}}\right) \cdot \prod_{s=1}^{d} (N z_s)^{n_s} \cdot e^{-N z_s}/n_s!$$

where $z_1, \ldots, z_d, x_{11}, \ldots, x_{Nd}$ are complex numbers, and $E_{n_1, \ldots, n_d}$ denotes expectation for fixed values of $n_1, \ldots, n_d$.

Lemma 2.1.

$$A_N(z_1, \ldots, z_d) = A_N(z_1, \ldots, z_d; x_{11}, \ldots, x_{Nd}; p_1, \ldots, p_N)$$

$$= \prod_{k=1}^{N} \left[1 + \sum_{s=1}^{d} P\left(N p_k (z_1 + \cdots + z_s)\right) \cdot x_{k1} \ldots x_{k,s-1} \cdot (x_{ks} - 1)\right]$$

where

$$P(z) = \sum_{s=0}^{r} z^s \cdot e^{-z}/s!.$$

Proof. Let $z_1, \ldots, z_d$ be real positive numbers and let $\eta_{11}, \ldots, \eta_{Nd}$ be independent random variables, where η_{ks} is Poisson $(N p_k z_s)$, i.e.

$$P(\eta_{ks}=y) = (N p_k z_s)^y \cdot e^{-N p_k z_s}/y!, \qquad y=0, 1, 2, \ldots.$$

It is well-known that the conditional distribution of $(\eta_{1s}, \ldots, \eta_{Ns})$ given $\eta_{1s}+\cdots+\eta_{Ns}=n_s$ is multinomial $(n_s, p_1, \ldots, p_N)$. Using this observation it is easily seen that

$$E_{n_1, \ldots, n_d}\left(\prod_{k,s} x_{ks}^{\varepsilon_{ks}}\right) = E\left(\prod_{k,s} x_{ks}^{\varepsilon_{ks}} \,\middle|\, \sum_{k=1}^{N} \eta_{ks} = n_s, \ s=1, \ldots, d\right)$$

where $\varepsilon'_{ks}=1$, if $\eta_{k1}+\cdots+\eta_{ks}\leqq r$, $\varepsilon'_{ks}=0$ otherwise, $k=1,\ldots,N$, $s=1,\ldots,d$. Hence we have

$$\begin{aligned}A_N(z_1,\ldots,z_d)&=\sum_{n_1,\ldots,n_d=0}^{\infty}E\left(\prod_{k,s}x_{ks}^{\varepsilon'_{ks}}\,\middle|\,\sum_{k=1}^{N}\eta_{ks}=n_s,\ s=1,\ldots,d\right)\cdot\prod_{s=1}^{d}P\left(\sum_{k=1}^{N}\eta_{ks}=n_s\right)\\&=E\left(\prod_{k,s}x_{ks}^{\varepsilon'_{ks}}\right).\end{aligned}$$

As $\eta_{11},\ldots,\eta_{Nd}$ are independent we find

$$A_N(z_1,\ldots,z_d)=E\left(\prod_{k,s}x_{ks}^{\varepsilon'_{ks}}\right)=\prod_{k=1}^{N}E\left(\prod_{s=1}^{d}x_{ks}^{\varepsilon'_{ks}}\right).$$

By direct computation we get

$$\begin{aligned}E\left(\prod_{s=1}^{d}x_{ks}^{\varepsilon'_{ks}}\right)&=\left(1-P(\eta_{k1}\leqq r)\right)\cdot 1+\left(P(\eta_{k1}\leqq r)-P(\eta_{k1}+\eta_{k2}\leqq r)\right)\cdot x_{k1}\\&\quad+\cdots+P(\eta_{k1}+\cdots+\eta_{kd}\leqq r)\cdot x_{k1}\ldots x_{kd}\\&=1+\sum_{s=1}^{d}P\left(N\,p_k(z_1+\cdots+z_s)\right)\cdot x_{k1}\ldots x_{k,s-1}\cdot(x_{ks}-1)\end{aligned}$$

with $P(z)$ defined as stated in the lemma.

Hence the lemma is proved for real positive $z_1,\ldots,z_d$. By analytical continuation it follows that the lemma is valid for all complex $z_1,\ldots,z_d$. Q.E.D.

Lemma 2.2. *If $Z_{n_1}, Z_{n_1+n_2},\ldots,Z_{n_1+\cdots+n_d}$ are occupancy sums for a cell situation $\{(p_k,a_k),k=1,\ldots,N\}$, then the characteristic function of $V_N=(Z_{n_1},\ldots,Z_{n_1+\cdots+n_d})$ can be written*

$$\begin{aligned}&E\left(\exp\left(i(t_1Z_{n_1}+\cdots+t_dZ_{n_1+\cdots+n_d})\right)\right)\\&\quad=\frac{n_1!\ldots n_d!}{N^{n_1+\cdots+n_d}}\cdot(2\pi i)^{-d}\oint\cdots\oint\exp\left(N(z_1+\cdots+z_d)\right)\cdot\frac{A_N(z_1,\ldots,z_d)}{z_1^{n_1+1}\ldots z_d^{n_d+1}}\,dz_1\ldots dz_d\end{aligned}$$

integrating along the circles:

$$\begin{aligned}|z_s|&=R_s>0,\qquad s=1,\ldots,d,\\x_{ks}&=\exp(i\,a_k\,t_s),\qquad k=1,\ldots,N,\ s=1,\ldots,d,\end{aligned}$$

in the expression for $A_N(z_1,\ldots,z_d)$.

Proof. With $x_{ks}=\exp(i\,a_k\,t_s)$ it follows that

$$\begin{aligned}&E\left(\exp\left(i(t_1Z_{n_1}+\cdots+t_dZ_{n_1+\cdots+n_d})\right)\right)\\&\quad=E\left(\exp\left(i\,t_1(a_1\varepsilon_{11}+\cdots+a_N\varepsilon_{N1})+\cdots+i\,t_d(a_1\varepsilon_{1d}+\cdots+a_N\varepsilon_{Nd})\right)\right)\\&\quad=E_{n_1,\ldots,n_d}\left(\prod_{k,s}x_{ks}^{\varepsilon'_{ks}}\right).\end{aligned}$$

From the definition of A_N and by Cauchy's integral formula the lemma follows. Q.E.D.

3. Some Auxiliary Results

In this section we compute asymptotic expressions for the expectation and covariance of V_N and prove some inequalities which we shall need in Section 4. We shall use the following lemma by Fabius [3] concerning the Poisson approximation of the binomial distribution.

Lemma 3.1. *If ξ is binomial (n, p), X is Poisson (np), and A is an arbitrary set of real numbers, then for any n and p*

$$|P(\xi \in A) - P(X \in A)| \leqq n p^2.$$

Let us call the general conditions introduced in Section 1:

A) $0 < \liminf n_s/N \leqq \limsup n_s/N < \infty$, $s = 1, \ldots, d$,

B) $N p_k \leqq C < \infty$, for all N and k, and some real number C.

Let $X_{11}, \ldots, X_{Nd}$ be independent random variables, where X_{ks} is Poisson $(n_s p_k)$, $k = 1, \ldots, N$, $s = 1, \ldots, d$. Put

$$P(k, s) = P(X_{k1} + \cdots + X_{ks} \leqq r),$$

$$Q(k, s) = P(X_{k1} + \cdots + X_{ks} = r),$$

$$\mu_s = \sum_{k=1}^{N} a_k \cdot P(k, s),$$

$$\sigma_{uv} = \begin{cases} \sum_{k=1}^{N} a_k^2 \cdot P(k, u) \cdot (1 - P(k, v)) - (n_1 + \cdots + n_v) \cdot \sum_{k=1}^{N} p_k a_k Q(k, u) \cdot \sum_{k=1}^{N} p_k a_k Q(k, v) \\ \qquad \text{for } 1 \leqq v \leqq u \leqq d, \\ \sigma_{vu} \qquad \text{for } 1 \leqq u < v \leqq d. \end{cases}$$

Lemma 3.2. *If $Z_{n_1}, Z_{n_1+n_2}, \ldots, Z_{n_1+\cdots+n_d}$ are occupancy sums for a cell situation $\{(p_k, a_k), k = 1, \ldots, N\}$, then*

$$E(Z_{n_1+\cdots+n_s}) = \mu_s + O\left(\left(\sum_1^N a_k^2/N\right)^{\frac{1}{2}}\right), \qquad s = 1, \ldots, d,$$

$$\operatorname{Cov}(Z_{n_1+\cdots+n_u}, Z_{n_1+\cdots+n_v}) = \sigma_{uv} + O\left(\sum_1^N a_k^2/N\right), \qquad u, v = 1, \ldots, d.$$

Proof. With the ε's and the ξ's defined as in Section 2 we find

$$E(Z_{n_1+\cdots+n_s}) = E\left(\sum_{k=1}^{N} a_k \varepsilon_{ks}\right) = \sum_{k=1}^{N} a_k P(\xi_{k1} + \cdots + \xi_{ks} \leqq r).$$

As $\xi_{k1} + \cdots + \xi_{ks}$ is binomial $(n_1 + \cdots + n_s, p_k)$, we get by Lemma 3.1, Cauchy's inequality, and conditions A) and B), that

$$\begin{aligned} |E(Z_{n_1+\cdots+n_s}) - \mu_s| &\leqq \sum_{k=1}^{N} |a_k| \cdot |P(\xi_{k1} + \cdots + \xi_{ks} \leqq r) - P(X_{k1} + \cdots + X_{ks} \leqq r)| \\ &\leqq \sum_{k=1}^{N} |a_k| \cdot (n_1 + \cdots + n_s) \cdot p_k^2 \leqq K \cdot \left(\sum_1^N a_k^2/N\right)^{\frac{1}{2}}. \end{aligned}$$

This proves the first statement.

We now consider the variances. Obviously it is sufficient to treat the case $d=1$. For convenience we suppress the index 1. We have

$$\mathrm{Var}(Z_n)=\mathrm{Var}\left(\sum_1^N a_k\,\varepsilon_k\right)=\sum_1^N a_k^2\cdot\mathrm{Var}(\varepsilon_k)+\sum_{j\neq k}a_j\,a_k\,\mathrm{Cov}(\varepsilon_j,\varepsilon_k)=S_1+S_2.$$

From Lemma 3.1, and conditions A) and B) we obtain

$$\mathrm{Var}(\varepsilon_k)=P(\xi_k\leqq r)-\big(P(\xi_k\leqq r)\big)^2=P(X_k\leqq r)-\big(P(X_k\leqq r)\big)^2+O(1/N)$$

where $|O(1/N)|\leqq K/N$, with K independent of j. Hence

$$S_1=\sum_1^N a_k^2\cdot P(X_k\leqq r)\cdot P(X_k>r)+O\left(\sum_1^N a_k^2/N\right).$$

To find S_2 we first note that since $(\xi_1,\ldots,\xi_N)$ is multinomial $(n,p_1,\ldots,p_N)$ we have

$$\begin{aligned}\mathrm{Cov}(\varepsilon_j,\varepsilon_k)&=P(\xi_j\leqq r,\xi_k\leqq r)-P(\xi_j\leqq r)\cdot P(\xi_k\leqq r)\\&=\sum_{x,y=0}^{r}\binom{n}{x}\binom{n}{y}p_j^x\,p_k^y(1-p_j)^{n-x}(1-p_k)^{n-y}\\&\quad\cdot\left[-1+\frac{(n-x)!\,(n-y)!}{n!\,(n-x-y)!}\cdot\frac{(1-p_j-p_k)^{n-x-y}}{(1-p_j)^{n-x}(1-p_k)^{n-y}}\right].\end{aligned}$$

Using Stirling's formula

$$n!=\sqrt{2\pi n}\cdot n^n\cdot\exp\big(-n+1/12n+O(n^{-3})\big)$$

taking logarithms, expanding into a Taylor series, and using conditions A) and B), we see that

$$\frac{(n-x)!\,(n-y)!}{n!\,(n-x-y)!}=\exp\big(-x\,y/n+O(n^{-2})\big)$$

and, further, that

$$\frac{(1-p_j-p_k)^{n-x-y}}{(1-p_j)^{n-x}(1-p_k)^{n-y}}=\exp\big(-n\,p_j\,p_k+y\,p_j+x\,p_k+O(n^{-2})\big)$$

where $|O(n^{-2})|\leqq K/n^2$, with K independent of j and k. Introducing these expressions into the formula for the covariances, expanding the exponential function, using Lemma 3.1, and conditions A) and B), we get

$$\begin{aligned}\mathrm{Cov}(\varepsilon_j,\varepsilon_k)&=\sum_{x,y=0}^{r}P(\xi_j=x)\,P(\xi_k=y)\big[-x\,y/n-n\,p_j\,p_k+y\,p_j+x\,p_k+O(n^{-2})\big]\\&=-n\,p_j\,p_k\,P(X_j=r)\,P(X_k=r)+O(n^{-2})\end{aligned}$$

where $|O(n^{-2})| \leqq K/n^2$, and K is independent of j and k. Hence we find

$$S_2 = -n \cdot \sum_{j \neq k} p_j p_k a_j a_k P(X_j = r) P(X_k = r) + O\Big(\sum_{j \neq k} |a_j a_k|/n^2\Big).$$

By conditions A) and B),

$$n \cdot \sum_1^N p_k^2 a_k^2 \big(P(X_k = r)\big)^2 \leqq K \cdot \sum_1^N a_k^2/N,$$

and by Cauchy's inequality,

$$\Big|O\Big(\sum_{j \neq k} |a_j a_k|/n^2\Big)\Big| \leqq \Big|O\Big(\sum_{j,k} |a_j| \cdot |a_k|/n^2\Big)\Big| \leqq K \cdot \sum_1^N a_k^2/N.$$

Hence we find

$$S_2 = -n \left(\sum_1^N p_k a_k P(X_k = r)\right)^2 + O\left(\sum_1^N a_k^2/N\right).$$

Finally we see

$$\mathrm{Var}(Z_n) = S_1 + S_2 = \sum_1^N a_k^2 P(X_k \leqq r) P(X_k > r) - n \left(\sum_1^N p_k a_k P(X_k = r)\right)^2 + O\left(\sum_1^N a_k^2/N\right)$$

which proves the second statement when $u = v$.

It remains to consider the covariances. It is sufficient to treat the case $d = 2$.

$$\mathrm{Cov}(Z_{n_1}, Z_{n_1+n_2}) = \sum_1^N a_k^2 \mathrm{Cov}(\varepsilon_{k1}, \varepsilon_{k2}) + \sum_{j \neq k} a_j a_k \mathrm{Cov}(\varepsilon_{j1}, \varepsilon_{k2}).$$

As in the derivation of the variances we find

$$\begin{aligned} \mathrm{Cov}(\varepsilon_{k1}, \varepsilon_{k2}) &= P(\xi_{k1} + \xi_{k2} \leqq r) - P(\xi_{k1} \leqq r) \cdot P(\xi_{k1} + \xi_{k2} \leqq r) \\ &= P(X_{k1} + X_{k2} \leqq r) \cdot P(X_{k1} > r) + O(1/N) \end{aligned}$$

and

$$\mathrm{Cov}(\varepsilon_{j1}, \varepsilon_{k2}) = P(\xi_{j1} \leqq r, \xi_{k1} + \xi_{k2} \leqq r) - P(\xi_{j1} \leqq r) \cdot P(\xi_{k1} + \xi_{k2} \leqq r).$$

As ξ_{k1} and ξ_{k2} are independent the last expression can be written

$$\sum_{\substack{x \leqq r \\ y+z \leqq r}} P(\xi_{k2} = z) \cdot \big(P(\xi_{j1} = x, \xi_{k1} = y) - P(\xi_{j1} = x) \cdot P(\xi_{k1} = y)\big)$$

and, in the same way as we obtained $\mathrm{Cov}(\varepsilon_j, \varepsilon_k)$ above, we get

$$\sum_{\substack{x \leqq r \\ y+z \leqq r}} P(\xi_{k2} = z) P(\xi_{j1} = x) P(\xi_{k1} = y) \cdot [-x y/n_1 - n_1 p_j p_k + y p_j + x p_k + O(n_1^{-2})].$$

By Lemma 3.1 we realize that this is equal to

$$-n_1 p_j p_k P(X_{j1} = r) P(X_{k1} + X_{k2} = r) + O(n_1^{-2}).$$

As in the derivation of the variance it follows that

$$\mathrm{Cov}(Z_{n_1}, Z_{n_1+n_2}) = \sigma_{12} + O\left(\sum_1^N a_k^2/N\right). \quad \text{Q.E.D.}$$

Lemma 3.3. *There exist real and positive numbers K_1 and K_2 such that*

$$K_1\left(\left(\sum_1^N a_k^2/N\right)\Big/\left(\sum_1^N p_k a_k^2\right)\right)\leqq\sum_1^N a_k^2/\sigma_{uu}\leqq K_2\left(\left(\sum_1^N a_k^2/N\right)\Big/\left(\sum_1^N p_k a_k^2\right)\right)^{2r+2}.$$

Proof. It is sufficient to consider $u=1$. We write $\sigma^2=\sigma_{11}$ and suppress the index 1. We find

$$\sigma^2=\sum a_k^2\,P(X_k\leqq r)\,P(X_k>r)-n\left(\sum p_k a_k\,P(X_k=r)\right)^2\leqq\sum a_k^2\,P(X_k\leqq r)\,P(X_k>r)$$

and after some reflection, using conditions A) and B),

$$\sigma^2\leqq K'\sum a_k^2(n p_k)^{r+1}\leqq K''\cdot N\sum p_k a_k^2.$$

This proves the first part of the lemma.

Further, using Cauchy's inequality, we find

$$\sigma^2\geqq\sum a_k^2\cdot\left[P(X_k\leqq r)\,P(X_k>r)-n p_k\left(P(X_k=r)\right)^2\right]\geqq K'\sum a_k^2(n p_k)^{2r+2}$$

where the latter part is proved below. By Hölder's inequality we get

$$\left(\sum a_k^2\right)^{(2r+1)/(2r+2)}\cdot\left(\sum a_k p_k^{2r+2}\right)^{1/(2r+2)}\geqq\sum a_k^2 p_k,$$

and from this it follows that

$$\sigma^2\geqq K'\cdot n^{2r+2}\cdot\left(\sum p_k a_k^2\right)^{2r+2}/\left(\sum a_k^2\right)^{2r+1}.$$

This proves the second inequality in the statement.

It remains to be proved that, if X is Poisson (m) and $0\leqq m\leqq C<\infty$, then for some $K'>0$ we have

$$f_r(m)=P(X\leqq r)\,P(X>r)-m\left(P(X=r)\right)^2\geqq K'\cdot m^{2r+2}.$$

It is easily seen that the inequality holds for m sufficiently small or great. Therefore it is sufficient to prove that $f_r(m)>0$ for $m>0$. Now

$$e^m f_0(m)=e^m-1-m>0\qquad\text{for } m>0.$$

Taking derivatives we find

$$f_r'(m)-f_{r-1}(m)=e^{-m}\cdot m^{2r-1}\cdot\left((r-1)!\right)^{-2}(m/r-1)^2>0\qquad\text{for } m>0.$$

Hence we have proved that $f_{r-1}(m)>0$, implying $f_r'(m)>0$. Since $f_r(0)=0$, it follows by induction that $f_r(m)>0$. Q.E.D.

From the definition of σ_{uu}, and Lemma 3.3, we obtain

Lemma 3.4. 1) $\liminf\left(\sum_1^N a_k^2/\sigma_{uu}\right)\geqq 1$, *and* 2) $\limsup\left(\sum_1^N a_k^2/\sigma_{uu}\right)<\infty$ *if and only if* $\limsup\left(\left(\sum_1^N a_k^2/N\right)\Big/\left(\sum_1^N p_k a_k^2\right)\right)<\infty$.

4. A Limit Theorem

Theorem. *Let* $Z_{n_1}, Z_{n_1+n_2}, \ldots, Z_{n_1+\cdots+n_d}$ *be occupancy sums for a cell situation* $\{(p_k, a_k), k=1, \ldots, N\}$. *If*

1. $0<\liminf_{N\to\infty} n_s/N \leqq \limsup_{N\to\infty} n_s/N<\infty$, $s=1, \ldots, d$,
2. $N p_k \leqq C<\infty$, *for some* C *and all* N *and* k,
3. $\limsup_{N\to\infty}\left(\left(\sum_1^N a_k^2/N\right)\Big/\left(\sum_1^N p_k a_k^2\right)\right)<\infty$,
4. $\lim_{N\to\infty}\left(\max_{1\leqq k\leqq N}|a_k|\Big/\left(\sum_1^N a_k^2\right)^{\frac{1}{2}}\right)=0$,
5. $\lim_{N\to\infty}[\mathrm{Corr}(Z_{n_1+\cdots+n_u}, Z_{n_1+\cdots+n_v})]=[\rho_{uv}^{\infty}; u, v=1, \ldots, d]=\rho^{\infty}$,

then

$$\left((Z_{n_1}-E(Z_{n_1}))/\sqrt{\mathrm{Var}(Z_{n_1})}, \ldots, (Z_{n_1+\cdots+n_d}-E(Z_{n_1+\cdots+n_d}))/\sqrt{\mathrm{Var}(Z_{n_1+\cdots+n_d})}\right)$$

is asymptotically normal $(0, \rho^{\infty})$ *as* $N\to\infty$.

Proof. From conditions 1, 2, and 3, and Lemmas 3.2 and 3.4, it follows that

$$\mathrm{Var}(Z_{n_1+\cdots+n_s})=\sigma_{ss}\cdot(1+O(1/N)),$$
$$\mathrm{Corr}(Z_{n_1+\cdots+n_u}, Z_{n_1+\cdots+n_v})=\sigma_{uv}/\sqrt{\sigma_{uu}\sigma_{vv}}+O(1/N)=\rho_{uv}+O(1/N),$$
$$(E(Z_{n_1+\cdots+n_s})-\mu_s)/\sqrt{\mathrm{Var}(Z_{n_1+\cdots+n_s})}=O(N^{-\frac{1}{2}}),$$

and that conditions 3, 4, and 5 are equivalent to

3. $\limsup_{N\to\infty}\left(\sum_1^N a_k^2/\sigma_{uu}\right)<\infty$,
4. $\lim_{N\to\infty}\left(\max_{1\leqq k\leqq N}|a_k|/\sqrt{\sigma_{uu}}\right)=0$,
5. $\lim_{N\to\infty}\rho_{uv}=\rho_{uv}^{\infty}$, $u, v=1, \ldots, d$.

Hence it is sufficient to prove asymptotic normality for

$$U_N=\left((Z_{n_1}-\mu_1)/\sqrt{\sigma_{11}}, \ldots, (Z_{n_1+\cdots+n_d}-\mu_d)/\sqrt{\sigma_{dd}}\right).$$

In the following we write σ_s instead of $\sqrt{\sigma_{ss}}$.

By Lemma 2.2 the characteristic function φ_N of U_N can be written

$$\varphi_N(t_1, \ldots, t_d)=\exp\left(-i\sum_{s=1}^{d} t_s\mu_s/\sigma_s\right)\cdot n_1!\ldots n_d!\cdot N^{-(n_1+\cdots+n_d)}$$
$$\cdot(2\pi i)^{-d}\cdot\oint\cdots\oint\frac{\exp(N(z_1+\cdots+z_d))\cdot A_N(z_1, \ldots, z_d)}{z_1^{n_1+1}\ldots z_d^{n_d+1}}\,dz_1\ldots dz_d$$

where the integrals are taken along the circles $|z_s|=n_s/N$, $s=1,\ldots,d$, and where

$$A_N(z_1,\ldots,z_d)$$
$$=\prod_{k=1}^{N}\left[1+\sum_{s=1}^{d}P(Np_k(z_1+\cdots+z_s))\exp\left(i\,a_k\sum_{v=1}^{s-1}(t_v/\sigma_v)\right)(\exp(i\,a_k\,t_s/\sigma_s)-1)\right]$$

with

$$P(z)=\sum_{s=0}^{r}z^s\cdot e^{-z}/s!.$$

We will show that, when $N\to\infty$,

$$\varphi_N(t_1,\ldots,t_d)\to\exp\left(-\frac{1}{2}\sum_{u,v=1}^{d}t_u\,t_v\,\rho_{uv}^{\infty}\right).$$

By applying Stirling's formula for $n_s!$, $s=1,\ldots,d$, and changing to polar coordinates, we obtain

$$\varphi_N(t_1,\ldots,t_d)=\exp\left(-i\sum_{s=1}^{d}t_s\,\mu_s/\sigma_s\right)\cdot e^{o(1)}\cdot\left(\prod_{s=1}^{d}(n_s/2\pi)\right)^{\frac{1}{2}}$$
$$\cdot\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\exp\left(\sum_{s=1}^{d}n_s(e^{i\theta_s}-1-i\,\theta_s)\right)\cdot A_N(n_1\,e^{i\theta_1}/N,\ldots,n_d\,e^{i\theta_d}/N)\,d\theta_1\ldots d\theta_d.$$

Now we break up the integration region into two parts:

$$S=\{\theta_s\colon|\theta_s|\leqq\delta,\ s=1,\ldots,d\}$$
$$\bar{S}=\{\theta_s\colon|\theta_s|\leqq\pi,\ s=1,\ldots,d\}-S.$$

First we prove

Lemma 4.1. *For all fixed values of* $t_1,\ldots,t_d$, *and* $\delta>0$, *the integral*

$$\left(\prod_{s=1}^{d}n_s\right)^{\frac{1}{2}}\int\cdots\int_{\bar{S}}\exp\left(\sum_{s=1}^{d}n_s(e^{i\theta_s}-1-i\,\theta_s)\right)A_N(n_1\,e^{i\theta_1}/N,\ldots,n_d\,e^{i\theta_d}/N)\,d\theta_1\ldots d\theta_d$$

converges to 0 *when* $N\to\infty$.

Proof. From the conditions 1, 2, and 3 it follows that

$$|A_N(\ldots)|\leqq\prod_{k=1}^{N}(1+K_1\,|a_k|/\sigma_1)\leqq\exp\left(K_2\sum_{1}^{N}|a_k|/\sigma_1\right)\leqq\exp(K_3\sqrt{N}).$$

For $0<\delta\leqq|\theta_s|\leqq\pi$,

$$|\exp(n_s(e^{i\theta_s}-1-i\,\theta_s))|\leqq\exp(n_s(\cos\theta_s-1))\leqq\exp(-2n_s\sin^2\delta/2)\leqq\exp(-K_4\,N).$$

Hence for some $K_5>0$ and $K_6>0$ the integral can be majorized by $K_5\cdot N^{d/2}\cdot\exp(-K_6\cdot N)\to 0$, when $N\to\infty$. Q.E.D.

In the region S we expand the logarithm of the integrand in powers of θ_s and t_s/σ_s. We find, with $P(k, s)$ and $Q(k, s)$ defined as in Section 3, that

$$\begin{aligned}
&\sum_{s=1}^{d} n_s(e^{i\theta_s}-1-i\theta_s)+\log\left(A_N(n_1 e^{i\theta_1}/N, \ldots, n_d e^{i\theta_d}/N)\right)\\
&\quad=-\frac{1}{2}\sum_{s=1}^{d} n_s\theta_s^2+i\sum_{s=1}^{d}\left(\sum_{k=1}^{N} a_k P(k,s)\right)t_s/\sigma_s\\
&\qquad+\sum_{s=1}^{d}(n_1\theta_1+\cdots+n_s\theta_s)\cdot(t_s/\sigma_s)\cdot\sum_{k=1}^{N} p_k a_k Q(k,s)\\
&\qquad-\sum_{s=1}^{d}\sum_{v=1}^{s-1}(t_s t_v/\sigma_s\sigma_v)\cdot\sum_{k=1}^{N} P(k,s)\,a_k^2-\frac{1}{2}\sum_{s=1}^{d}(t_s^2/\sigma_s^2)\cdot\sum_{k=1}^{N} P(k,s)\,a_k^2\\
&\qquad+\frac{1}{2}\sum_{k=1}^{N} a_k^2\cdot\left(\sum_{s=1}^{d}(t_s/\sigma_s)\cdot P(k,s)\right)^2+O\left(\sum_{1}^{N}|a_k|^3/\sigma_1^3\right)\\
&\qquad+O\left(\sum_{s=1}^{d}|\theta_s|\cdot\sum_{1}^{N} a_k^2/\sigma_1^2\right)+O\left(\sum_{s=1}^{d}\theta_s^2\cdot\sum_{1}^{N}|a_k|/\sigma_1\right)+O\left(\sum_{1}^{d} n_s|\theta_s|^3\right).
\end{aligned}$$

From this result and the conditions of the theorem we find that the logarithm of the integrand can be written in the form

$$\begin{aligned}
&i\sum_{s=1}^{d}\mu_s t_s/\sigma_s-\frac{1}{2}\sum_{s=1}^{d} n_s(\theta_s-m_s)^2-\frac{1}{2}\sum_{s=1}^{d} t_s^2-\sum_{s=1}^{d}\sum_{v=1}^{s-1} t_s t_v\rho_{sv}\\
&\qquad+o(1)+O\left(\sum_{1}^{d}|\theta_s|\right)+O\left(\sum_{1}^{d}\theta_s^2\cdot\sqrt{N}\right)+O\left(\sum_{1}^{d} n_s|\theta_s|^3\right)
\end{aligned}$$

where

$$m_s=\sum_{v=s}^{d}(t_v/\sigma_v)\sum_{k=1}^{N} p_k a_k Q(k,v)=O(N^{-\frac{1}{2}}).$$

Using the coordinate transformation $\varphi_s=\sqrt{n_s}\cdot\theta_s$, $s=1, \ldots, d$, we get

$$\begin{aligned}
\varphi_N(t_1, \ldots, t_d)&=\exp\left(-\frac{1}{2}\sum_{u,v=1}^{d} t_u t_v\rho_{uv}\right)\cdot(2\pi)^{-d/2}\\
&\quad\cdot\int_{S_1}\cdots\int\exp\left(-\frac{1}{2}\sum_{1}^{d}(\varphi_s-O(1))^2+O(\ldots)\right)d\varphi_1\ldots d\varphi_d,
\end{aligned}$$

where

$$S_1=\{\varphi_s: |\varphi_s|\leqq\sqrt{n_s}\cdot\delta, s=1, \ldots, d\}.$$

If we could neglect the error terms, then the integral would converge to $(2\pi)^{d/2}$ when $N\to\infty$. Therefore we first only consider integration over the region

$$S_1'=\{\varphi_s: |\varphi_s|\leqq n_s^{\frac{1}{7}}, s=1, \ldots, d\}$$

where $O(\ldots)$ converges to 0 when $N\to\infty$. Hence we have proved

Lemma 4.2. *For all fixed values of* $t_1, \ldots, t_d$, *and*

$$S' = \{\theta_s : |\theta_s| \leqq n_s^{\frac{1}{7}-\frac{1}{2}}, s = 1, \ldots, d\},$$

we have when $N \to \infty$

$$\exp\left(-i\sum_{s=1}^{d} \mu_s t_s/\sigma_s\right) \cdot \left(\prod_{s=1}^{d} (n_s/2\pi)\right)^{\frac{1}{2}}$$

$$\cdot \int_{S'} \cdots \int \exp\left(\sum_{s=1}^{d} n_s(e^{i\theta_s} - 1 - i\theta_s)\right) \cdot A_N(n_1 e^{i\theta_1}/N, \ldots, n_d e^{i\theta_d}/N)\, d\theta_1 \ldots d\theta_d$$

$$\to \exp\left(-\frac{1}{2}\sum_{u,v=1}^{d} t_u t_v \rho_{uv}^{\infty}\right).$$

It now remains to prove

Lemma 4.3. *For all fixed values of* $t_1, \ldots, t_d$, *and for a sufficiently small value of* $\delta > 0$, *the expression*

$$\left(\prod_{s=1}^{d} n_s\right)^{\frac{1}{2}} \int_{S-S'} \cdots \int \exp\left(\sum_{s=1}^{d} n_s(e^{i\theta_s} - 1 - i\theta_s)\right) \cdot A_N(n_1 e^{i\theta_1}/N, \ldots, n_d e^{i\theta_d}/N)\, d\theta_1 \ldots d\theta_d$$

converges to 0 *when* $N \to \infty$.

Proof. For sufficiently small $\delta > 0$ we have with $c_\delta > 0$

$$\cos\theta_s - 1 = -2 \cdot \sin^2 \theta_s/2 \leqq -2c_\delta^2(\theta_s/2)^2 = -K_0\theta_s^2.$$

Hence the integral can be majorized by

$$\left(\prod_{s=1}^{d} n_s\right)^{\frac{1}{2}} \int_{S-S'} \cdots \int \exp\left(-K_0 \sum_{1}^{d} n_s \theta_s^2\right) \cdot |A_N(\ldots)|\, d\theta_1 \ldots d\theta_d.$$

Now we use the expansion for $\log A_N(\ldots)$, and we obtain

$$|A_N(\ldots)| \leqq \exp\left(K_1 \sum_{1}^{d} \sqrt{n_s}\,|\theta_s| + K_2\right).$$

Changing coordinates to $\varphi_s = \sqrt{n_s}\,\theta_s$, $s = 1, \ldots, d$, we get the estimate

$$\int_{S_1 - S_1'} \cdots \int \exp\left(-K_0 \cdot \sum_{1}^{d} \varphi_s^2 + K_1 \cdot \sum_{1}^{d} |\varphi_s| + K_2\right) d\varphi_1 \ldots d\varphi_d \to 0,$$

when $N \to \infty$. Q.E.D.

Combining Lemmas 4.1, 4.2, and 4.3, we conclude that when $N \to \infty$

$$\varphi_N(t_1, \ldots, t_d) \to \exp\left(-\frac{1}{2}\sum_{u,v=1}^{d} t_u t_v \rho_{uv}^{\infty}\right).$$

By the continuity theorem the assertion follows.

Remark 1. As $|\mathrm{Corr}(Z_{n_1+\cdots+n_u}, Z_{n_1+\cdots+n_v})| \leqq 1$ it follows that it is always possible to select a subsequence $\{N'\}$ of $\{N\}$ so that Condition 5 is fulfilled. Hence Conditions 1, 2, 3, and 4 imply that the only possible limiting distribution is the normal.

Remark 2. In Rosén [7] it is shown that, in the case $r=0$, Conditions 1, 2, and 3 imply that the limiting correlation matrices are non-singular. In this case it is easily seen that our theorem can be stated as Rosén's Theorem 1.

Acknowledgement. I am grateful to Professor Gunnar Blom for valuable advice and encouragement.

References

1. Barton, D. E., David, F. N.: Combinatorial chance. London: Griffin 1962.
2. Chistyakov, V. P.: On the calculation of the power of the test of empty boxes. Theor. Probab. Appl. **9**, 648–653 (1964).
3. Fabius, J.: De poisson benadering voor der binomiale verdelning. Statistica Neederlandica **21**, 287–289 (1967).
4. Feller, W.: An introduction to probability theory and its applications, vol. 1, 2nd ed. New York-London: John Wiley & Sons 1957.
5. Okamoto, M.: On a non-parametric test. Osaka math. J. **4**, 77–85 (1952).
6. Rényi, A.: Three new proofs and a generalization of a theorem of Irving Weiss. Publ. math. Inst. Hungar. Acad. Sci., Ser. B **7**, 203–214 (1962).
7. Rosén, B.: Asymptotic normality in a coupon collector's problem. Z. Wahrscheinlichkeitstheorie verw. Geb. **13**, 256–279 (1969).
8. Sevastyanov, B. A.: Convergence of the distribution of the number of empty boxes to gaussian and poisson processes in the classical ball problem. Theor. Probab. Appl. **12**, 126–134 (1967).
9. Weiss, I.: Limiting distributions in some occupancy problems. Ann. math. Statistics **29**, 878–884 (1958).

Lars Holst
Department of Mathematics
Uppsala University
Sysslomansgatan 8
S-752 23 Uppsala
Sweden

(Received August 17, 1970)