The Annals of Mathematical Statistics
1971, Vol. 42, No. 5, 1671–1680

LIMIT THEOREMS FOR SOME OCCUPANCY AND SEQUENTIAL OCCUPANCY PROBLEMS

BY LARS HOLST
Uppsala University

Consider a situation in which balls are falling into N cells with arbitrary probabilities. A limiting distribution for the number of occupied cells after n falls is obtained, when n and $N \to \infty$, so that $n^2/N \to \infty$ and $n/N \to 0$. This result completes some theorems given by Chistyakov (1964), (1967). Limiting distributions of the number of falls to achieve a_N+1 occupied cells are obtained when $\limsup a_N/N < 1$. These theorems generalize theorems given by Baum and Billingsley (1965), and David and Barton (1962), when the balls fall into cells with the same probability for every cell.

1. Introduction. Suppose that we throw balls into N cells, so that each ball may fall into the kth cell with probability p_k, $p_1+\cdots+p_N = 1$, independently of what happens to other balls. Let Z_n be the number of occupied cells after n throws, and let T_N be the throw on which, for the first time, a_N+1 cells are occupied, $0 \leqq a_N < N$.

The classical occupancy problem deals with the distribution of Z_n when $p_1 = \cdots = p_N = 1/N$, see e.g. David and Barton (1962) and Feller (1968). For this case a complete characterization of the limiting behavior of Z_n, under different assumptions on how n and N tend to infinity, was given by Rényi (1962). In the general situation, when the p's are allowed to be different, Chistyakov (1964), (1967) has obtained limiting distributions of Z_n, when n and N tend to infinity, so that $\log (n/N)$ or n^2/N are bounded. In Section 3 we will give Chistyakov's theorems and prove a limit theorem when n and $N \to \infty$ so that $n^2/N \to \infty$ and $n/N \to 0$. Further aspects on occupancy problems are considered by Rosén (1969) and by Holst (1971).

The classical sequential occupancy problem (or coupon collector's problem) deals with the distribution of T_N when $p_1 = \cdots = p_N = 1/N$, see e.g. David and Barton (1962) and Feller (1968). In this case David and Barton derived the limiting distributions of T_N, when $a_N/N \to \alpha$ for $0 < \alpha < 1$, or when $a_N+1 = N$. Baum and Billingsley (1965) obtained limiting distributions for this classical situation for every asymptotic behavior of a_N when $N \to \infty$. The methods used in these papers cannot be used when the p's are different. In Section 4 we will give limiting distributions of T_N in the above mentioned situation when $\limsup (a_N/N) < 1$, with the aid of the theorems of Section 3. Similar problems are considered by Rosén (1970) and by Holst (1971).

2. Notation and assumptions. In order to give a precise probabilistic formulation of the asymptotic behavior of Z_n or T_N, we introduce a sequence of probability vectors $(p_{1\nu}, p_{2\nu}, \cdots, p_{N_\nu \nu})$, $p_{1\nu}+\cdots+p_{N_\nu \nu} = 1$, two sequences of integers n_ν and a_{N_ν},

Received September 29, 1970.

and the corresponding sequences of random variables Z_{n_ν} and T_{N_ν}, $\nu = 1, 2, 3, \cdots$. We suppose that

$$N_\nu \to \infty \text{ when } \nu \to \infty,$$
$$N_\nu p_{k\nu} \leqq C < \infty, \text{ for all } k \text{ and } \nu, \text{ and some real number } C,$$
$$n_\nu \text{ and } a_{N_\nu} \to \infty \text{ when } \nu \to \infty.$$

To facilitate the notation we suppress the index ν in the following. Hence we speak about the asymptotic behavior of Z_n or T_N (when n or $N \to \infty$) instead of that of $\{Z_{n_\nu}\}_{\nu=1}^{\infty}$ or $\{T_{N_\nu}\}_{\nu=1}^{\infty}$ (when $\nu \to \infty$).

3. Limit theorems for some occupancy problems.

THEOREM 1. *If $n^2 \sum_1^N p_k^2/2 \to m < \infty$ (implying that n^2/N is bounded), then $n - Z_n$ is asymptotically Poisson (m), when $n \to \infty$.*

THEOREM 2. *If $n^2/N \to \infty$ and $n/N \to 0$, then $N - Z_n$ is asymptotically normal $(\sum_1^N \exp(-np_k), (n^2 \sum_1^N p_k^2/2)^{\frac{1}{2}})$, when $n \to \infty$.*

THEOREM 3. *If $\log(n/N)$ is bounded, then $N - Z_n$ is asymptotically normal $(\sum_1^N \exp(-np_k), \sigma_n)$, when $n \to \infty$, where*

$$\sigma_n^2 = \sum_1^N \exp(-np_k) \cdot (1 - \exp(-np_k)) - n \cdot (\sum_1^N p_k \exp(-np_k))^2.$$

Chistyakov (1964), (1967) proved Theorems 1 and 3 using methods similar to those initiated by Rényi (1962). In proving Theorem 2 we will use the same methods.

PROOF OF THEOREM 2. Depending on how n and $N \to \infty$ we discuss two cases separately.

(i) *The case $N^3/n^4 \to 0$.* In Chistyakov (1964) it is shown that the characteristic function of $N - Z_n$ can be written

$$E(\exp(it(N - Z_n))) = \frac{n!}{2\pi i N^n} \oint_{|z| = n/N} \frac{e^{Nz} \prod_1^N (1 + (e^{it} - 1) \cdot e^{-Np_k z})}{z^{n+1}} \, dz.$$

From this expression, using Stirling's formula for $n!$ and changing to polar coordinates, it follows that

$$(1) \quad E(\exp(it(N - Z_n - \mu)/\sigma)) = e^{-it\mu/\sigma} \cdot e^{o(1)} \cdot (n/2\pi)^{\frac{1}{2}} \cdot \int_{-\pi}^{\pi} \exp(n(e^{i\theta} - 1 - i\theta)) \prod_1^N (1 + (e^{it/\sigma} - 1) \cdot \exp(-np_k e^{i\theta})) \, d\theta$$

where

$$\mu = \mu_n = \sum_1^N \exp(-np_k)$$

and

$$\sigma^2 = \sigma_n^2 = \sum_1^N \exp(-np_k) \cdot (1 - \exp(-np_k)) - n \cdot (\sum_1^N p_k \exp(-np_k))^2.$$

Lemmas 3.1 and 3.2 below show that (1) converges to $\exp(-t^2/2)$. By the continuity theorem for characteristic functions it follows that $(N-Z_n-\mu_n)/\sigma_n$ is asymptotically normal (0, 1). As $Np_k \leqq C$ and $n/N \to 0$ it is easily proved that

$$\sigma_n^2 = n^2 \textstyle\sum_1^N p_k^2(1+o(1))/2$$

and that $\sigma_n^2 \to \infty$ if and only if $n^2/N \to \infty$. Hence

$$(N-Z_n-\mu_n)/(n^2 \textstyle\sum_1^N p_k^2/2)^{\frac{1}{2}}$$

is asymptotically normal (0, 1), which proves the theorem when $N^3/n^4 \to 0$.

(ii) *The case* $n^5/N^4 \to 0$. In Chistyakov (1967) it is shown that the characteristic function of $n-Z_n$ can be written

$$E(\exp(it(n-Z_n))) = \frac{n!}{2\pi i N^n}\oint_{|z|=n/N} \frac{\prod_1^N(1+(\exp(Np_k z e^{it})-1)\cdot e^{-it})}{z^{n+1}}\,dz.$$

In the same way as in case (i) we obtain

$$(2) \quad E(\exp(it(N-Z_n-\mu)/s)) = \exp(-it\textstyle\sum_1^N(e^{-np_k}-1+np_k)/s)\cdot e^{o(1)}\cdot(n/2\pi)^{\frac{1}{2}}$$
$$\cdot \textstyle\int_{-\pi}^{\pi}\exp(n(-1-i\theta))\prod_1^N(1+(\exp(np_k e^{i(\theta+t/s)})-1)\cdot e^{-it/s})\,d\theta$$

where

$$s^2 = s_n^2 = n^2\textstyle\sum_1^N p_k^2/2.$$

Lemmas 3.3, 3.4, and 3.5 show that (2) converges to $\exp(-t^2/2)$. Hence $(N-Z_n-\mu_n)/s_n$ is asymptotically normal (0, 1) when $n^5/N^4 \to 0$.

Combining (i) and (ii) the assertion follows. The lemmas remain.

Let t belong to a fixed bounded interval, and let $d > 0$ be a fixed sufficiently small real number.

LEMMA 3.1. *If* $N^2/n^3 \to 0$, *then*

$$(n/2\pi)^{\frac{1}{2}}\textstyle\int_{\pi \geqq |\theta| \geqq d}\exp(n(e^{i\theta}-1-i\theta))\cdot\prod_1^N(1+(e^{it/\sigma}-1)\exp(-np_k e^{i\theta}))\,d\theta \to 0$$

when $n \to \infty$.

PROOF. Expanding into a Taylor series we obtain

$$\begin{aligned}\log \textstyle\prod(1+(e^{it/\sigma}-1)\cdot\exp(-np_k e^{i\theta})) &= \textstyle\sum((e^{it/\sigma}-1)\exp(-np_k e^{i\theta})+O(1/\sigma^2)) \\ &= Nit/\sigma + O(N/\sigma^2)+O(n/\sigma).\end{aligned}$$

Now $N/\sigma^2 \leqq K\cdot N/(n^2/N)$ and $\cos\theta - 1 \leqq -2\sin^2 d/2$ for $d \leqq |\theta| \leqq \pi$. Thus the absolute value of the integral can be estimated from above by

$$K_1\cdot n^{\frac{1}{2}}\cdot\exp(-n(2\sin^2 d/2+K_2N^2/n^3+K_3/\sigma)).$$

From the conditions it follows that the estimate converges to zero. □

LEMMA 3.2. *If* $N^3/n^4 \to 0$, *then*

$$\exp(-it\mu/\sigma)\cdot(n/2\pi)^{\frac{1}{2}}\int_{|\theta|\leqq d}\exp(n(e^{i\theta}-1-i\theta)) \cdot\prod_1^N(1+(e^{it/\sigma}-1)\exp(-np_k e^{i\theta}))\,d\theta$$

converges to $\exp(-t^2/2)$ *when* $n\to\infty$.

PROOF. Expanding into a Taylor series we find

$$\begin{aligned}\log\prod(1+(e^{it/\sigma}-1)\cdot\exp(-np_k e^{i\theta})) &= (it/\sigma)\cdot\sum\exp(-np_k e^{i\theta})\\ &+((it/\sigma)^2/2!)\cdot\sum[\exp(-np_k e^{i\theta})-\exp(-2np_k e^{i\theta})]\\ &+((it/\sigma)^3/3!)\cdot[-n\,e^{i\theta}+O(n^2/N)]\\ &+O(n/\sigma^4)+O(N/\sigma^5)=(it/\sigma)\cdot\sum\exp(-np_k)\\ &+(t\theta/\sigma)\cdot\sum np_k\exp(-np_k)+O(n\theta^2/\sigma)\\ &+((it/\sigma)^2/2!)\sum[\exp(-np_k)-\exp(-2np_k)]+O(n|\theta|/\sigma^2)+o(1).\end{aligned}$$

The last equality follows from

$$n/\sigma^3\leqq K\cdot N^3/n^4\to 0,$$
$$N/\sigma^5\leqq K\cdot(N^7/n^{10})^{\frac{1}{2}}\to 0,$$

and

$$\exp(-np_k e^{i\theta})=\exp(-np_k)\cdot(1-i\theta np_k+O(np_k\theta^2)).$$

Introduce $f(n)=N^3/n^4\to 0$ and break up the integration interval into $|\theta|\leqq g(n)=f(n)^{1/7}\cdot n^{1/6}/n^{\frac{1}{2}}$ and $g(n)\leqq|\theta|\leqq d$. We find from the expression above that

$$\exp(-it\mu/\sigma)\cdot(n/2\pi)^{\frac{1}{2}}\int_{|\theta|\leqq g(n)}\exp(n(e^{i\theta}-1-i\theta))\cdot\prod_1^N(\cdots)\,d\theta = e^{-t^2/2}\cdot(n/2\pi)^{\frac{1}{2}}\cdot e^{o(1)}\cdot\int_{|\theta|\leqq g(n)}\exp(h(\theta))\,d\theta$$

where

$$h(\theta)=-n(\theta-t\sum p_k e^{-np_k}/\sigma)^2/2+\theta^2 O(n/\sigma)+|\theta|O(n/\sigma^2)+|\theta|^3 O(n).$$

Changing coordinates to $v=n^{\frac{1}{2}}\cdot\theta$ we see that every error term converges to zero and

$$(n^{\frac{1}{2}}\sum p_k e^{-np_k}/\sigma)/(f(n)^{1/7}\cdot n^{1/6})\leqq K\cdot f(n)^{1/6}/f(n)^{1/7}\to 0.$$

As

$$n^{\frac{1}{2}}\cdot g(n)=(N/n)^{3/7}\cdot n^{1/42}\to\infty$$

it follows that the integral converges to $\exp(-t^2/2)$.

It remains to be proved that the integral over $g(n) \leqq |\theta| \leqq d$ can be neglected. From the expansion above and as $\cos\theta - 1 \leqq -K_d\theta^2$ for $|\theta| \leqq d$, with $K_d > 0$ for d sufficiently small, we get the following estimate

$$(n/2\pi)^{\frac{1}{2}} \int_{g(n) \leqq |\theta| \leqq d} \exp(-n \cdot K_d\theta^2 + K_1|\theta|n/\sigma + K_2)\, d\theta$$
$$\leqq K_3 \int_{n^{1/6} \cdot f(n)^{1/7} \leqq |v| \leqq d \cdot n^{1/2}} \exp(-K_d v^2 + v^2 o(1))\, dv \to 0,$$

when $n \to \infty$. □

We note that the condition $N^3/n^4 \to 0$ probably cannot be weakened, because t^3 disappears due to this condition.

LEMMA 3.3.

$$(n/2\pi)^{\frac{1}{2}} \int_{\pi \geqq |\theta| \geqq d} \exp(-n - ni\theta) \cdot \prod_1^N [1 + (\exp(np_k e^{i(\theta + t/s)}) - 1) \cdot e^{-it/s}]\, d\theta$$

converges to zero when $n \to \infty$.

PROOF. Expanding into a Taylor series we find

$$\log \prod [1 + (\exp(np_k e^{i(\theta + t/s)}) - 1) \cdot e^{-it/s}]$$
$$= n\, e^{i\theta} + e^{2i\theta}(e^{it/s} - 1)s^2 - it \sum (np_k)^3/6s + |\theta| \cdot O(n^3/N^2 s)$$
$$+ O(n^3/N^2 s^2) + O(n^4/N^3 s) + O(n^5/N^4).$$

From this we see that the integral can be written

$$(n/2\pi)^{\frac{1}{2}} \int_{\pi \geqq |\theta| \geqq d} \exp[n(e^{i\theta} - 1 - i\theta) + e^{2i\theta}(e^{it/s} - 1)s^2$$
$$+ O(\sum (np_k)^3/s) + O(\sum (np_k)^4)]\, d\theta.$$

The absolute value can be estimated by

$$(n/2\pi)^{\frac{1}{2}} \int_{\pi \geqq |\theta| \geqq d} e^{n(\cos\theta - 1)} \cdot \exp[O(s) + O(n^3/N^2 s) + O(n^4/N^3)]\, d\theta$$
$$\leqq (n/2\pi)^{\frac{1}{2}} \cdot 2\pi \cdot \exp[-n(2\sin^2 d/2 + O(s/n) + O(n^2/N^2 s) + O(n^3/N^3))] \to 0$$

when $n \to \infty$. □

LEMMA 3.4. *If $n^5/N^4 \to 0$, then*

$$\exp(-it(\sum_1^N (\exp(-np_k) - 1 + np_k)/s)) \cdot (n/2\pi)^{\frac{1}{2}}$$
$$\cdot \int_{|\theta| \leqq n^{1/7}/n^{1/2}} \exp(-n - ni\theta) \cdot \prod_1^N [1 + (\exp(np_k e^{i(\theta + t/s)}) - 1) \cdot e^{-it/s}]\, d\theta$$

converges to $\exp(-t^2/2)$ *when $n \to \infty$.*

PROOF. Using the expansion used in the proof of the preceding lemma we find that the integral can be written

$$\exp(-t^2/2) \cdot (n/2\pi)^{\frac{1}{2}} \int_{|\theta| \leqq n^{1/7}/n^{1/2}} \exp(h(\theta))\, d\theta$$

where

$$h(\theta) = -n(\theta+2ts/n)^2/2+2t^2s^2/n+\theta^2O(s)+|\theta|O(1)+O(1/s)+|\theta|O(n^3/N^2s)$$
$$+o(1)+|\theta|^3O(n).$$

As in Lemma 3.2 we see that the integral converges to $\exp(-t^2/2)$, when $n \to \infty$. □

LEMMA 3.5. *If $n^5/N^4 \to \infty$ then*

$$(n/2\pi)^{\frac{1}{2}} \int_{n^{1/7}/n^{1/2} \leqq |\theta| \leqq d} \exp(-n-ni\theta) \cdot \prod_1^N [1+(\exp(np_k e^{i(\theta+t/s)})-1)e^{-it/s}]\, d\theta$$

converges to zero when $n \to \infty$.

PROOF. From the expansion in Lemma 3.3 it follows that the integral can be estimated by

$$(n/2\pi)^{\frac{1}{2}} \int_{n^{1/7}/n^{1/2} \leqq |\theta| \leqq d} \exp(n(\cos\theta-1))$$
$$\cdot |\exp(e^{2i\theta}(e^{it/s}-1)s^2)| \exp(|\theta|O(n^3/N^2s)+o(1))\, d\theta.$$

As

$$e^{2i\theta}(e^{it/s}-1)s^2 = its+|\theta|O(s)+O(1)$$

and $\cos\theta-1 \leqq -K_d\theta^2$ for $|\theta| \leqq d$, with $K_d > 0$ for d sufficiently small, we get the estimate

$$(n/2\pi)^{\frac{1}{2}} \int_{n^{1/7}/n^{1/2} \leqq |\theta| \leqq d} \exp(-nK_d\theta^2+|\theta|O(s)+|\theta|O(n^3/N^2s)+O(1))\, d\theta.$$

Changing coordinates to $v = n^{\frac{1}{2}} \cdot \theta$ we see that the integral converges to zero when $n \to \infty$. □

4. Limit theorems for some sequential occupancy problems.

THEOREM 4. *If $a_N^2 \cdot \sum_1^N p_k^2/2 \to m < \infty$, then $T_N - a_N - 1$ is asymptotically Poisson (m), when $N \to \infty$.*

PROOF. From the definitions of T_N and Z_n it follows that

$$T_N-a_N-1 > x \Leftrightarrow T_N > a_N+1+x \Leftrightarrow Z_{a_N+1+x} < a_N+1$$

and hence

$$P(T_N-a_N-1 > x) = P(x < a_N+1+x-Z_{a_N+1+x}).$$

Since $a_N^2 \cdot \sum_1^N p_k^2/2 \to m$ and $a_N \to \infty$ we have that, for every fixed x, $(a_N+1+x)^2 \cdot \sum_1^N p_k^2/2 \to m$.
Using Theorem 1, for every fixed integer x, we have when $N \to \infty$

$$P(x < a_N+1+x-Z_{a_N+1+x}) \to P(x < Y),$$

where Y is Poisson (m). □

REMARK. By Cauchy's inequality and the assumption $Np_k \leqq C$ it is easily seen that

$$\limsup a_N{}^2/N < \infty \Leftrightarrow \limsup a_N{}^2 \cdot \sum_1^N p_k{}^2 < \infty.$$

Hence, if lim sup $a_N{}^2/N < \infty$, then the only possible limiting distribution is the Poisson distribution.

THEOREM 5. *If* $a_N{}^2/N \to \infty$ *and* $a_N/N \to 0$, *then* T_N *is asymptotically normal* (m_N, σ_N), *when* $N \to \infty$, *where* m_N *is the unique solution of the equation in* t.

$$\sum_1^N \exp(-tp_k)/N = 1 - a_N/N \tag{3}$$

and

$$\sigma_N{}^2 = m_N{}^2 \cdot \sum_1^N p_k{}^2/2.$$

REMARK. The existence of a unique solution of (3), for N sufficiently large, is obvious since the left side of (3) is strictly decreasing and equal to one for $t = 0$, and the right side of (3) converges to one (from below) when $N \to \infty$.

Before proving the theorem we state two lemmas.

LEMMA 4.1. *If* $a_N/N \to 0$, *then* $m_N/a_N \to 1$ *and* $\sum_1^N p_k \exp(-m_N p_k) \to 1$.

PROOF. Introduce the distribution function defined by

$$H_N(x) = \text{(the number of } p_k \text{ satisfying } Np_k \leqq x)/N.$$

As $Np_k \leqq C < \infty$ we have

$$\int_0^C dH_N(x) = 1,$$

$$\int_0^C x\, dH_N(x) = \sum_1^N p_k = 1.$$

$$\int_0^C \exp(-x \cdot m_N/N)\, dH_N(x) = \sum_1^N \exp(-m_N p_k)/N = 1 - a_N/N.$$

First we will show that $m_N/N \to 0$ when $N \to \infty$. Let us assume the contrary, i.e. that there exists a subsequence (N') so that $m_{N'}/N' \to d$, $0 < d \leqq \infty$. By Helly's theorems we can select from (N') a subsequence (N'') so that

$$H_{N''}(x) \to H(x), \qquad \text{when } N \to \infty$$

$$\int_0^C dH(x) = 1,$$

$$\int_0^C x\, dH(x) = 1.$$

But when $N \to \infty$

$$\int_0^C \exp(-x m_{N''}/N'')\, dH_{N''}(x) = 1 - a_{N''}/N'' \to \int_0^C \exp(-xd)\, dH(x) = 1.$$

As $\int_0^C x\, dH(x) = 1$ the total probability mass is not in zero, which contradicts $d > 0$. Hence $m_N/N \to 0$ when $N \to \infty$. Now we can expand into a Taylor series

$$a_N/N = \int_0^C (1 - \exp(-x \cdot m_N/N))\, dH_N(x)$$
$$= (m_N/N) \cdot \left(\int_0^C x\, dH_N(x) + O(m_N/N)\right) = (m_N/N) \cdot (1 + o(1)).$$

Hence $m_N/a_N = 1 + o(1)$.

In the same way we find for every convergent subsequence

$$\sum_1^N p_k \exp(-m_N p_k) = \int_0^C x \cdot \exp(-x \cdot m_N/N)\, dH_N(x) \to \int_0^C x\, dH(x) = 1.$$

As we from every subsequence are able to select a convergent subsequence, the second assertion is proved. □

LEMMA 4.2. *For every fixed real number* x *we have*

$$(N - a_N - \sum_1^N \exp(-[x\sigma_N + m_N]p_k))/([x\sigma_N + m_N]^2 \cdot \sum_1^N p_k^2/2)^{\frac{1}{2}} \to x.$$

PROOF. As $Np_k \leqq C$ we have by Lemma 4.1 when $N \to \infty$,

$$\sigma_N = m_N(\sum p_k^2/2)^{\frac{1}{2}} = (a_N^2 \cdot \sum_1^N p_k^2/2)^{\frac{1}{2}} \cdot (1 + o(1)) \to \infty,$$

and hence

$$[x\sigma_N + m_N]^2 \cdot \sum p_k^2/2 = \sigma_N^2 \cdot (1 + o(1)).$$

Expanding in series we find

$$\sum_1^N \exp(-(x\sigma_N + m_N)p_k) = \sum_1^N \exp(-m_N p_k) - x\sigma_N \cdot \sum_1^N p_k \exp(-m_N p_k) \cdot (1 + o(1)).$$

By Lemma 4.1 this expression can be written

$$\sum_1^N \exp(-m_N p_k) - x\sigma_N \cdot (1 + o(1)).$$

From the definition of m_N and the results above we have

$$(N - a_N - \sum_1^N \exp(-(x\sigma_N + m_N)p_k))/([x\sigma_N + m_N]^2 \sum p_k^2/2)^{\frac{1}{2}} = x \cdot (1 + o(1)).$$

The same result is obtained if we change $x\sigma_N$ to $x\sigma_N - 1$. Since

$$x\sigma_N + m_N - 1 < [x\sigma_N + m_N] \leqq x\sigma_N + m_N$$

the lemma is proved.

PROOF OF THEOREM 5. In the same way as in Theorem 4 we find

$$\begin{aligned} P(T_N > x\sigma_N + m_N) &= P(Z_{[x\sigma_N + m_N]} < a_N + 1) \\ &= 1 - P(N - Z_{[x\sigma_N + m_N]} < N - a_N) \\ &= 1 - P(Y_N < (N - a_N - \textstyle\sum_1^N \exp(-[x\sigma_N + m_N]p_k))/([x\sigma_N + m_N]^2 \sum p_k^2/2)^{\frac{1}{2}}) \end{aligned}$$

where according to Theorem 2

$$Y_N = (N - Z_{[x\sigma_N + m_N]} - \sum_1^N \exp(-[x\sigma_N + m_N]p_k))/([x\sigma_N + m_N]^2 \sum p_k^2/2)^{\frac{1}{2}}$$

is asymptotically normal (0, 1).
By Lemma 4.2, for every fixed x,

$$P(Y_N < \cdots) = P(Y_N < x \cdot (1 + o(1)))$$

which converges to $\Phi(x)$ when $N \to \infty$. Hence when $N \to \infty$

$$P(T_N > x\sigma_N + m_N) \to 1 - \Phi(x). \quad \square$$

Finally we consider the case $0 < \liminf a_N/N \leqq \limsup a_N/N < 1$. Let us as in Lemma 4.1 introduce the distribution function

$$H_N(x) = \text{(the number of } p_k \text{ satisfying } Np_k \leqq x)/N,$$

then

$$\sum_1^N \exp(-np_k) = N \cdot \int_0^C \exp(-xn/N)\, dH_N(x)$$

and for σ_n^2 defined as in Theorem 3

$$(4) \quad \sigma_n^2 = N \cdot (\int_0^C \exp(-xn/N) \cdot (1 - \exp(-xn/N))\, dH_N(x)) \\ - n \cdot (\int_0^C x \cdot \exp(-xn/N)\, dH_N(x))^2.$$

As in the preceding theorem we consider the equation

$$(5) \qquad \int_0^C \exp(-tx/N)\, dH_N(x) = 1 - a_N/N.$$

The following condition is sufficient to give the equation a unique solution:

$$1 - H_N(0) > a_N/N$$

i.e. the number of $p_k \neq 0$ is at least $a_N + 1$.

THEOREM 6. *If* $\liminf a_N/N > 0$ *and* $\liminf (1 - H_N(0) - a_N/N) > 0$, *then* T_N *is asymptotically normal* $(m_N, \sigma_{m_N}/(\sum_1^N p_k \exp(-m_N p_k)))$, *when* $N \to \infty$, *where* m_N *is the solution of* (5).

PROOF. To facilitate the notation set $\sigma_N = \sigma_{m_N}$. As in Theorem 5 we get

$$P(T_N \leqq x\sigma_N + m_N) = P(Y_N < (N - a_N - \sum_1^N \exp(-[x\sigma_N + m_N]p_k))/\sigma_{[x\sigma_N + m_N]})$$

where according to Theorem 3

$$Y_N = (N - Z_{[x\sigma_N + m_N]} - \sum_1^N \exp(-[x\sigma_N + m_N]p_k))/\sigma_{[x\sigma_N + m_N]}$$

is asymptotically normal (0, 1). From the conditions of the theorem and the definition of m_N we find in the same way as in Lemma 4.1 that

$$0 < \liminf m_N/N \leqq \limsup m_N/N < \infty.$$

By this observation and by (4) it follows that, for every fixed x,

$$\sigma_{m_N}/\sigma_{[x\sigma_N + m_N]} = 1 + o(1).$$

As in the proof of Theorem 5 we find

$$(N - a_N - \sum_1^N \exp(-(x\sigma_N + m_N)p_k))/\sigma_N = x \cdot \sum_1^N p_k \exp(-m_N p_k) \cdot (1 + o(1)).$$

We obtain the same result if $x\sigma_N$ is replaced by $x\sigma_N - 1$. Hence

$$P(T_N \leqq x\sigma_N + m_N) = P(Y_N < x \cdot \sum_1^N p_k \exp(-m_N p_k) \cdot (1 + o(1))).$$

Since Y_N is asymptotically normal (0, 1) the theorem is proved.

REMARK. The case $p_1 = \cdots = p_N = 1/N$ corresponds to the distribution function $H_N(x) = 1$ for $x \geqq 1$ and $= 0$ otherwise. In this situation we find $m_N = -N \cdot \log(1 - a_N/N)$ and that, the expression for the variance in Theorem 5 can be written

$$N \cdot (\log(1 - a_N/N))^2/2 \sim a_N{}^2/2N$$

and that of Theorem 6

$$N \cdot (a_N/N + (1 - a_N/N) \cdot \log(1 - a_N/N))/(1 - a_N/N).$$

From these observations we see that our theorem in this special case is the same as the corresponding theorems of Baum and Billingsley (1965) and David and Barton (1962).

REFERENCES

BAUM, L. E. and BILLINGSLEY, P. (1965). Asymptotic distributions for the coupon collector's problem. *Ann. Math. Statist.* **36** 1835–1839.

CHISTYAKOV, V. P. (1964). On the calculation of the power of the test of empty boxes. *Theor. Probability Appl.* **9** 648–653.

CHISTYAKOV, V. P. (1967). Discrete limit distributions in the problem of balls falling in cells with arbitrary probabilities. *Math. Notes Acad. Sc. USSR* **1** 6–11.

DAVID, F. N. and BARTON, D. E. (1962). *Combinatorial Chance.* Griffin, London.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications* **1**, 3rd ed. Wiley, New York.

HOLST, L. (1971). A note on finding the size of a finite population. *Biometrika* **58** 228–229.

HOLST, L. (1971). Asymptotic normality in a generalized occupancy problem. To appear in *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.*

RÉNYI, A. (1962). Three new proofs and a generalization of a theorem of Irving Weiss. *Publ. Math. Inst. Hungar. Acad. Sci.* **7** 203–214.

ROSÉN, B. (1969). Asymptotic normality in a coupon collector's problem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **13** 256–279.

ROSÉN, B. (1970). On the coupon collector's waiting time. *Ann. Math. Statist.* **41** 1952–1969.