The Annals of Mathematical Statistics
1972, Vol. 43, No. 5, 1698-1701

A NOTE ON THE CLASSICAL OCCUPANCY PROBLEM[1]

BY C. J. PARK

University of Wisconsin

Assume that n balls are randomly distributed into N equiprobable cells. The ball is presumed to have probability p, $0 < p < 1$ of staying in the cell and $(1-p)$ of falling through. Let S_0 denote the number of empty cells. In this note we establish the asymptotic normality of S_0 as n and N tend to infinity so that $np/N \to c > 0$, $np/N^{\frac{5}{8}} \to \infty$ and $n/N \to 0$, or $3np/N - \log N \to -\infty$ and $n/N \to \infty$. We accomplish this by estimating the factorial cumulants of S_0.

1. Introduction and summary. Assume that n balls are randomly distributed into N cells with equal probabilities, i.e. each ball has probability $1/N$ of falling into ith cell, $i = 1, 2, \cdots, N$. The ball is presumed to have probability p, $0 < p < 1$ of staying in the cell and $(1-p)$ of falling through. Let S_0 denote the number of empty cells. In this note we will show that the asymptotic distribution of S_0 is normal as n and N tend to infinity with one of the following conditions being satisfied:

(i) $np/N \to c$, $0 < c < \infty$,
(ii) $n/N \to 0$ and $np/N^{\frac{5}{8}} \to \infty$,
(iii) $n/N \to \infty$ and $3np/N - \log N \to -\infty$.

We establish the asymptotic normality of S_0 by estimating the factorial cumulants of S_0 and utilizing the similar method given by Harris and Park [4]. For the special case when $p = 1$, the asymptotic distribution of S_0 has been extensively studied (see for example [5], [6], [8] and [9]). Harkness [3] gives numerous examples of situations for which the distribution of S_0 can be applied (see also the references therein).

2. Asymptotic normality of S_0. The probability distribution of S_0 is well known (see for example [3]) and given by

$$P[S_0 = x; n, N, p] = \binom{N}{x} \sum_{k=0}^{N-x} (-1)^k \binom{n-x}{k} \left(1 - \frac{(k+x)}{N} p\right)^n, \qquad x = 0, 1, \cdots, N.$$

The mth factorial moment of S_0 is given by

$$\mu_{[m]} = N^{(m)} \left(1 - \frac{mp}{N}\right)^n, \qquad m = 0, 1, \cdots, \tag{1}$$

where $N^{(m)} = N(N-1) \cdots (N-m+1)$. Consequently the factorial moment

Received June 22, 1970; revised March 9, 1972.

[1] This research was primarily done while the author was at the University of Nebraska. Sponsored by the United States Army under Contract No. DA-31-124-ARO-D-462.

generating function can be written as,

$$\varphi_{n,N}(t) = \textstyle\sum_{m=0}^{\infty} \frac{\mu[m]}{m!} t^m = \sum_{m=0}^{N} \binom{N}{m} \left(1 - \frac{mp}{N}\right)^n t^m . \tag{2}$$

Let $K_{n,N}(t)$ be the corresponding factorial cumulant generating function, then

$$K_{n,N}(t) = \log \varphi_{n,N}(t) = \textstyle\sum_{m=1}^{\infty} k_{[m]} \frac{t^m}{m!} , \tag{3}$$

where $k_{[m]} = k_{[m]}(n, N)$ is the mth factorial cumulant of S_0. The factorial cumulants are related to the cumulants in the same way as the factorial moments are related to the moments, that is,

$$k_m = \textstyle\sum_{j=1}^{m} \alpha_{j,m} k_{[j]} , \tag{4}$$

where $\alpha_{j,m}$ are the Stirling numbers of the second kind. To establish the asymptotic normality of S_0, we will show that for $m > 2$

$$k_m k_2^{-m/2} \rightarrow 0$$

as n and N tend to infinity. Now we introduce the following theorem.

THEOREM 1. *The* mth *cumulant of* S_0,

$$k_m = O(N) \quad as \quad N \rightarrow \infty , \qquad for \ m = 1, 2, \cdots .$$

PROOF. Let

$$P(t) = (1 + t)^N = \textstyle\sum_{\nu=0}^{N} \binom{N}{\nu} t^\nu ,$$

a polynomial of degree N with every root -1. Then let

$$P_1(t) = P(t) - p \frac{t}{N} P'(t)$$
$$= \textstyle\sum_{\nu=0}^{N} \binom{N}{\nu} \left(1 - p \frac{\nu}{N}\right) t^\nu .$$

For $\nu \geqq 1$, define

$$P_{\nu+1}(t) = P_\nu(t) - \left(p \frac{t}{N}\right) P_\nu'(t) ;$$

then we readily see that

$$P_n(t) = \textstyle\sum_{\nu=0}^{N} \binom{N}{\nu} \left(1 - p \frac{\nu}{N}\right)^n t^\nu = \varphi_{n,N}(t)$$

where $\varphi_{n,N}(t)$ is defined in (2). Now define

$$Q_n(t) = \frac{N}{p} P_{n+1}(t) = \frac{N}{p} P_n(t) - tP_n'(t) .$$

Then it can be verified (cf. Lemma 1 and Lemma 2 in [4]) that for every $n \geqq 1$ $Q_n(t)$ has N real roots and all of its roots $\leqq -1$ because $P_n(t)$ is a polynomial of degree N and has N real roots $\leqq -1$. Hence, $N^{-1}\log P_n(t) = N^{-1}\log \varphi_{n,N}(t) = N^{-1}K_{n,N}(t)$ is analytic in $|t| < 1$. Thus for $|t| < 1$,

$$\operatorname{Re}(N^{-1}\log P_n(t)) = N^{-1}\log |P_n(t)|$$
$$\leqq N^{-1}\log \textstyle\sum_{\nu=0}^{N}\binom{N}{\nu}|t|^{\nu} = \log(1+|t|) \leqq \log 2 .$$

We can now apply a well-known theorem of Carathéodory (see [1], [2] and [7]), that is, if $f(z) = \sum_{j=1}^{\infty}\alpha_j z^j$, $|z| < 1$ and $\operatorname{Re}[f(z)] \leqq 1$ for $|z| < 1$, then $|\alpha_j| < 2$ for all j. Thus, since

$$K_{n,N}(t) = \textstyle\sum_{m=1}^{\infty} k_{[m]}t^m/m! ,$$

we have

$$|k_{[m]}| \leqq Nm! \log 4;$$

thus the theorem follows from (4).

Now from (1), we have

$$E(S_0) = \mu(S_0) = N\left(1 - \frac{p}{N}\right)^n ,$$

$$\operatorname{Var}(S_0) = \sigma^2(S_0)$$
$$= N^2\left(\left(1 - \frac{2p}{N}\right)^n - \left(1 - \frac{p}{N}\right)^{2n}\right) + N\left(\left(1 - \frac{p}{N}\right)^n - \left(1 - \frac{2p}{N}\right)^n\right).$$

We now establish the limiting distribution of

$$S_0{}^* = (S_0 - \mu(S_0))/\sigma(S_0) .$$

THEOREM 2. *If one of the conditions* (i)—(iii) *in Section* 1 *is satisfied, the limiting distribution of* $S_0{}^*$, *as n and N tend to infinity, is the standard normal distribution.*

PROOF. To establish the theorem it suffices to show that $k_2{}^{-m/2} \to 0$ for $m > 2$. From Theorem 1, this is equivalent to showing that $Nk_2{}^{-\frac{3}{2}} \to 0$. Let $n/N = \alpha(n, N)$ and since $\alpha(n, N) = o(N)$, we have

$$k_2 = \sigma^2(S_0) = Ne^{-\alpha p}(1 - e^{-\alpha p} - \alpha p e^{-\alpha p}) + O(\psi(\alpha))$$

where $\psi(\alpha) = \max(\alpha, \alpha^2)$. Thus, the conclusion holds for $\alpha \to 0$ as n and N tend to infinity with $np/N^{\frac{2}{3}} \to \infty$, and for $\alpha \to \infty$ as n and N tend to infinity with $3np/N - \log N \to -\infty$. The conclusion clearly holds if α has a positive limit as n and N tend to infinity.

REMARK. The probability distribution of S_0 can be written as

$$P[S_0 = x;\ n, N, p] = \textstyle\sum_{t=0}^{n} P(S_0 = x;\ t, N, 1)\binom{n}{t}p^t(1-p)^{n-t}$$

where $P(S_0 = x;\ t, N, 1)$ denotes the probability distribution of the number of empty cells when t balls are randomly distributed into N equi-probable cells and $p = 1$.

The limiting distribution of the number of cells occupied by i balls, $i \neq 0$, is under investigation and we hope to report the result in the future.

3. Acknowledgment. The author is indebted to the referee for his useful comments.

REFERENCES

[1] CARATHÉODORY, C. (1907). Über den Variabilitätbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.* **64** 95–115.

[2] CARATHÉODORY, C. (1911). Über den Variabilitätsbereich der Fourierchen Konstanten von positiven harmonischen Funktionen. *Rend. Circ. Mat. Palermo* **32** 193–217.

[3] HARKNESS, W. L. (1969). The classical occupancy problem revisited. Technical Report No. 11, Dept. of Statistics, The Penn. State Univ.

[4] HARRIS, B. and PARK, C. J. (1968). A note on the asymptotic normality of the distribution of the number of empty cells in occupancy problems. Math. Research Center Summary Report No. 955, Univ. of Wisconsin. To appear in *Ann. Inst. Statist. Math.*

[5] OKAMOTO, M. (1952). On a non-parametric test. *Osaka J. Math.* **4** 77–85.

[6] RÉNYI, A. (1962). Three new proofs and a generalization of a theorem of Irving Weiss. *Magyar Tud. Akad. Mat. Fiz. Oszt. Kozl. A* **7** 203–214.

[7] RIESZ, F. (1911). Sur certaines systemes singuliers d'equations integrales. *Ann. Sci. École Norm. Sup.* **28** 33–62.

[8] SEVAST'YANOV, B. A. and CHISTYAKOV, V. P. (1964). Asymptotic normality in the classical ball problem. *Theor. Probability Appl.* **9** 198–211.

[9] WEISS, I. (1958). Limiting distributions in some occupancy problems. *Ann. Math. Statist.* **29** 878–884.

MATH RESEARCH CENTER
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706