A NOTE ON THE CLASSICAL OCCUPANCY PROBLEM

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 \( S_n \) denote the number of empty cells. In this note we establish the asymptotic normality of \( S_n \) as \( n \) and \( N \) tend to infinity so that \( np/N \to c > 0 \), \( np/N^2 \to \infty \) 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 \( S_n \).

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 \( i \)th cell, \( i = 1, 2, \ldots, N \). The ball is presumed to have probability \( p \), \( 0 < p < 1 \) of staying in the cell and \( (1 - p) \) of falling through. Let \( S_n \) denote the number of empty cells. In this note we will show that the asymptotic distribution of \( S_n \) 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 \( np/N^2 \to \infty \),
   (iii) \( n/N \to \infty \) and \( 3np/N - \log N \to -\infty \).

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

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

\[
P[S_n = x; \ n, N, p] = \binom{n}{x} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left( \frac{k + x}{N} \right)^n,
\]

where \( x = 0, 1, \ldots, N \).

The \( m \)th factorial moment of \( S_n \) is given by

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

where \( N^{(m)} = N(N - 1) \cdots (N - m + 1) \). Consequently the factorial moment
generating function can be written as,

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

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

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

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

\[ k_m = \sum_{j=1}^{m} \alpha_{j,m} k_{(j)}, \]

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

\[ k_m k_{-m/2} \rightarrow 0 \]

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

**Theorem 1.** The \( m \)th cumulant of \( S_n \),

\[ k_m = O(N) \quad \text{as} \quad N \rightarrow \infty, \quad \text{for} \quad m = 1, 2, \ldots. \]

**Proof.** Let

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

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

\[ P_\nu(t) = P(t) - p \frac{t}{N} P'(t) \]

\[ = \sum_{\nu=0}^{N} \binom{N}{\nu} \left( 1 - p \frac{\nu}{N} \right) t^\nu. \]

For \( \nu \geq 1 \), define

\[ P_{\nu+1}(t) = P_\nu(t) - \left( p \frac{t}{N} \right) P'_\nu(t); \]

then we readily see that

\[ P_\nu(t) = \sum_{\nu=0}^{N} \binom{N}{\nu} \left( 1 - p \frac{\nu}{N} \right)^\nu = \varphi_{n,N}(t) \]

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

\[ Q_\nu(t) = \frac{N}{p} P_{\nu+1}(t) = \frac{N}{p} P_\nu(t) - tP'_\nu(t). \]

Then it can be verified (cf. Lemma 1 and Lemma 2 in [4]) that for every \( n \geq 1 \)

\( Q_\nu(t) \) has \( N \) real roots and all of its roots \( \leq -1 \) because \( P_\nu(t) \) is a polynomial
of degree $N$ and has $N$ real roots $\leq -1$. 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 $|t| < 1$,

$$\text{Re} (N^{-1} \log P_n(t)) = N^{-1} \log |P_n(t)| \leq N^{-1} \log \sum_{j=0}^{N} (j)!/|t|^j = \log (1 + |t|) \leq \log 2.$$ 

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 $\text{Re} \{f(z)\} \leq 1$ for $|z| < 1$, then $|\alpha_j| < 2$ for all $j$. Thus, since

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

we have

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

thus the theorem follows from (4).

Now from (1), we have

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

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

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 $k_{\frac{m}{2}} \to 0$ for $m > 2$. From Theorem 1, this is equivalent to showing that $Nk_{\frac{2}{N}} \to 0$. Let $n/N = \alpha(n, N)$ and since $\alpha(n, N) = o(N)$, we have

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

where $\psi(\alpha) = \max(\alpha, \alpha^2)$. Thus, the conclusion holds for $\alpha \to 0$ as $n$ and $N$ tend to infinity with $np/N^4 \to \infty$, and for $\alpha \to \infty$ as $n$ and $N$ tend to infinity with $3np/N - \log N \to -\infty$. The conclusion clearly holds if $\alpha$ has a positive limit as $n$ and $N$ tend to infinity.

**Remark.** The probability distribution of $S_0$ can be written as

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

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 $\rho = 1$.

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


Math Research Center
University of Wisconsin
Madison, Wisconsin 53706