J. Korean Math. Soc.
Vol. 14, No. 2, 1978.

A NOTE ON THE ASYMPTOTIC NORMALITY OF THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS UNDER BOSE-EINSTEIN STATISTICS

BY C. J. PARK

1. Introduction and summary.

Assume that m indistinguishable balls are randomly distributed into n distinct cells such that each of the distinct emplacement is equally likely with the probability

$$\frac{1}{\binom{m+n-1}{m}}.$$

We shall refer to this assumption as Bose-Einstein statistics. Let $X_i=1$ if the ith cell is empty and $X_i=0$ otherwise. Let $S_0=\sum_{i=1}^{n} X_i$, thus S_0 is the number of empty cells expressed as sum of dependent Bernoulli random variables. The distribution of S_0 under Bose-Einstein statistics has been studied by others; see, for example, [2], [6], and [7]. Employing the technique used by Harris or Park [3] and Park [5], we will show that, as $m, n\to\infty$ so that $m/n^{5/6}\to\infty$ or $n/m^{3/4}\to\infty$, the limiting distribution of $(S_0-\mu(S_0))/\sigma(S_0)$ is the standard normal distribution.

We also give a representation of S_0 as a sum of independent Bernoulli variables.

2. Asymptotic normality of S_0.

The probability distribution of S_0 is well known and it can be written

$$P[S_0=s_0]=\frac{\binom{n}{s_0}\binom{m-1}{n-s_0-1}}{\binom{m+n-1}{m}}.$$

Also see Neuts [4], p. 150.

Received by the editors Oct. 20, 1977.

The kth factorial moment of S_0 is given by, see Wilks [8], p. 445,

(1) $$E(S_0^{(k)})=\mu_{[k]}=n^{(k)}\binom{m+n-k-1}{m}\Big/\binom{m+n-1}{m},\quad k=0,1,2,\cdots,$$

where $n^{(k)}=n(n-1)\cdots(n-k+1)$..
Consequently the factorial moment generating function of S_0 can be written

(2) $$\phi_{m,n}(t)=\sum_{k=0}^{\infty}\mu_{[k]}\frac{t^k}{k!}=\sum_{k=0}^{n-1}\frac{\binom{n}{k}\binom{m+n-k-1}{m}t^k}{\binom{m+n-1}{m}}$$

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

(3) $$K_{m,n}(t)=\ln\phi_{m,n}(t)=\sum_{k=1}^{\infty}K_{[r]}\frac{t^r}{r!},$$

where $K_{[r]}=K_{[r]}(m,n)$ is the rth 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

(4) $$K_r=\sum_{j=1}^{r}\alpha_{j,r}K_{[j]},\quad \text{where}$$

$\alpha_{j,r}$ are the Stirling numbers of the second kind. To establish the asymptotic normality of S_0, we will show that for $r>2$, $K_rK_2^{-r/2}\to 0$ as m and n tend to infinity. Now we introduce the following theorem.

THEOREM 1. *The rth cumulant of* S_0

$$K_r=0(n)\ \textit{as}\ n\to\infty\ \textit{for}\ r=1,2,\cdots.$$

Proof. Let $P(t)=(1+t)^n=\sum_{\nu=0}^{n}\binom{n}{\nu}t^\nu$,

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

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

Thus $P_1(t)$ is a monic polynomial of degree $(n-1)$ with all roots -1. For $\nu\geq 1$, define

$P_{\nu+1}(t)=P_\nu(t)-\left(\frac{t}{n+\nu}\right)P_\nu'(t)$; then we readily see that

$P_m(t)=\phi_{m,n}(t)$ where $\phi_{m,n}(t)$ is defined in (2). Now define

$$Q_m(t) = (n+m)P_{m+1}(t)$$
$$= (n+m)P_m(t) - tP_m'(t).$$

Then it can be verified (c.f. Lemma 1 and Lemma 2 in [3]) that for every $m \geq 1$, $Q_m(t)$ has $(n-1)$ real roots and all of its roots ≤ -1 because $P_m(t)$ is a polynomial of degree $(n-1)$ and has $(n-1)$ real roots ≤ -1. Hence,

$$n^{-1}\ln P_m(t) = n^{-1}\ln\phi_{m,n}(t) = n^{-1}K_{m,n}(t)$$

is analytic in $|t|<1$.
Thus for $|t|<1$,

$$\mathrm{Re}(n^{-1}\ln P_m(t)) = n^{-1}\ln|P_m(t)|$$
$$\leq n^{-1}\ln\textstyle\sum_{\nu=0}^{n}\binom{n}{\nu}|t|^{\nu} = \ln(1+|t|) \leq \ln 2.$$

We can now apply a well-known theorem of Caratheordory (see [1] and references in [3]). That is, if $f(z)=\sum_{j=1}^{\infty}\alpha_j z^j$; $|z|<1$ and $\mathrm{Re}[f(z)]\leq 1$ for $|z|<1$, then $|\alpha_j|<2$ for all j. Thus, since

$$K_{m,n}(t) = \sum_{r=1}^{\infty} K_{[r]}\frac{t^r}{r!}, \text{ we have}$$
$$|K_{[r]}| \leq nr!\ln 4.$$

Hence the theorem follows from (4).

Corollary: *$S_0=\sum_{j=1}^{n-1}Y_j$, where Y_j are independent Bernoulli random variables, with $P[Y_j=1]=-\gamma_j$, and $\frac{1}{\gamma_j}$ is the jth root of $\phi_{m,n}(t)$, $j=1,2,\cdots, n-1$.*

Proof. Since the $(n-1)$ roots of $\phi_{m,n}(t)$ are real and ≤ -1, we can write

$$\phi_{m,n}(t) = \prod_{j=1}^{n-1}(1-\gamma_j t), \quad -1\leq\gamma_j<0.$$

The factorial moment generating function of a Bernoulli random variable is $E(1+t)^Y = 1+pt$, where $P(Y=1)=p=1-P[Y=0]$. By setting $p=-\gamma_j$, the corollary follows.

From (1), we have

$$K_1 = E(S_0) = \mu(S_0) = n\frac{\binom{m+n-2}{m}}{\binom{m+n-1}{m}} = n\frac{(n-1)}{(m+n-1)}$$

(5) $K_2=\mathrm{Var}(S_0)=\sigma^2(S_0)=n\left(\frac{n-1}{m+n-1}\right)\left(\frac{m}{m+n-1}\right)\left(\frac{m}{m+n-2}\right)$

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) *is satisfied, the limiting distribution of* S_0^* *is the standard normal distribution as* m *and* $n\to\infty$,

(i) $\lim_{n\to\infty}\frac{m}{n}=\rho>0$,

(ii) $\lim_{n\to\infty}\frac{m}{n}=0$ *and* $\frac{m}{n^{5/6}}\to\infty$, *or*

(iii) $\frac{m}{n}\to\infty$ *and* $\frac{n}{m^{3/4}}\to\infty$.

Proof: To establish the theorem it suffices to show that $K_r/K_2^{r/2}\to 0$ for $r>2$. From Theorem 1, this is equivalent to showing that $nK_2^{-3/2}\to 0$.

Let $m/n=\rho(n)$ and since $\rho(n)/(1+\rho(n))=0(n)$, from (5) we have

$$K_2=n\left[\frac{\rho^2}{(1+\rho)^3}\right]+0\left(\frac{\rho}{1+\rho}\right)$$

Thus, the conclusion holds for $\rho\to 0$ as $n\to\infty$ whenever $m/n^{5/6}\to\infty$ and for $\rho\to\infty$ as m and n tend to infinity with $n/m^{3/4}\to\infty$. The conclusion clearly holds if ρ has a positive limit as m and n tend to infinity.

REMARK: (1) With a slight modification of notation, one can establish a conclusion similar to Theorem 2 for the hypergeometric distribution.

(2) Let $W_{n,j}$ denote the number of balls required to get n-j cells occupied for the first time. Then the asymptotic normality of $W_{n,j}$ can be deduced from the following identity,

$$P[W_{n,j}\le m]=P[S_0\le n-j].$$

EXAMPLE: Let $m=3$, $n=3$, then the factorial moment generating function of S_0 is

$$\phi_{3,3}(t)=1+\frac{12}{10}t+\frac{3}{10}t^2.$$

The roots of $\phi_{3,3}(t)=0$ are

$$r_1=-2-2\sqrt{\frac{1}{6}}\text{ and }r_2=-2+2\sqrt{\frac{1}{6}}.$$

From the Corollary, we have

$$-\gamma_1=\frac{1}{2+2\sqrt{\frac{1}{6}}} \qquad -\gamma_2=\frac{1}{2-2\sqrt{\frac{1}{6}}}, \text{ and}$$

$$P[S_0=2]=\left(\frac{1}{2+2\sqrt{\frac{1}{6}}}\right)\left(\frac{1}{2-2\sqrt{\frac{1}{6}}}\right)=\frac{3}{10}$$

$$P[S_0=1]=\left(\frac{1}{2+2\sqrt{\frac{1}{6}}}\right)\left(1-\frac{1}{2-2\sqrt{\frac{1}{6}}}\right)+\left(\frac{1}{2-2\sqrt{\frac{1}{6}}}\right)\left(1-\frac{1}{2+2\sqrt{\frac{1}{6}}}\right)$$

$$=\frac{6}{10}$$

$$P[S_0=0]=\left(1-\frac{1}{2+2\sqrt{\frac{1}{6}}}\right)\left(1-\frac{1}{2-2\sqrt{\frac{1}{6}}}\right)=\frac{1}{10}.$$

3. Acknowledgement.

The author wishes to thank Dr. Macky for his comments.

References

[1] Caratheodory, C. (1911). *Über den Variabilitätsbereich der Fourierchen Konstanten von positive harmoischen Funktionen*. Rend. Circ. Mat. Palermo **32**, 193-217.

[2] Csorgo, M. and Guttman, I. (1964). *On the consistency of the two-sample empty cell test*. Canad. Math. Bull., **7**, 57-63.

[3] Harris, B. and Park, C. J. (1971). *A note on the asymptotic normality of the distribution of the number of empty cells in occupancy problems*. Ann. Inst. Statist. Math. **23**, 507-513.

[4] Neuts, M. F. (1973). *Probability*, Allyn and Bacon, Inc., Boston.

[5] Park, C. J. (1972). *A note on the classical occupancy problem*. Ann. Math. Statist. **43**, 1698-1701.

[6] Park, C. J. (1973). *The distribution of frequency counts of the geometric distribution*, Sankhya, **35**, 106-111.

[7] Wilks, S. S. (1961). *A combinatorial test for the problem of two samples from continuous distributions*. Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., 707-717.

[8] ______ (1962). *Mathematical Statistics*, John Wiley and Sons, Inc., New York.

Seoul National University and San Diego State University