A NOTE ON THE ASYMPTOTIC NORMALITY OF THE DISTRIBUTION OF THE NUMBER OF EMPTY CELLS IN OCCUPANCY PROBLEMS*

B. HARRIS AND C. J. PARK

(Received April 30, 1969)

Introduction and summary

In this note we present two new and interesting proofs of the asymptotic normality of the distribution of the number of empty cells in occupancy problems. More specifically, we suppose that we have a random sample of n observations from a multinomial distribution on N equiprobable cells. Then, letting s be the number of empty cells, we will show that as n, $N\to\infty$ so that $n/N^{5/6}\to\infty$ and $n/N-1/3\log N\to-\infty$, then the distribution of $V=(s-E(s))/\sigma_s$ has the standard normal distribution. We accomplish this by estimating the factorial cumulants of V. Since the cumulants are linear combinations of the factorial cumulants (with fixed coefficients), the factorial cumulants can easily be exploited for this purpose. In particular, in order to show that V has asymptotically the standard normal distribution, it suffices to show that all cumulants beyond the second tend to zero as n, $N\to\infty$. In F. N. David and D. E. Barton [1], the factorial cumulants were exploited to show that V is asymptotically normal when $n/N\to c$, a constant. Their method of estimating the factorial cumulants is somewhat different than that employed here. Using the closely related but substantially more complicated method of moments, I. Weiss [9] and M. Okamoto [5] established the asymptotic normality of V under the hypothesis $n/N\to c$, some constant. The asymptotic normality of V under the hypothesis $n/N\to c$ has been established by Sevast'yanov and Chistyakov [8] using saddle point methods. In fact, Sevast'yanov and Chistyakov examined a multivariate extension of this problem. A. Rényi [6] obtained the most general result dealing with the asymptotic normality of V. Employing characteristic functions, he established that V has an asymptotically standard normal distribution whenever $n/N^{1/2}\to\infty$ and $n/N-\log N\to-\infty$. Thus, Rényi's results are in fact more general than those presented herein. Despite this, we still felt that it was worthwhile to

* Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No.: DA-31-124-ARO-D-462.

present these arguments, since they are of distinct methodological interest and far more elementary than those of Rényi. In addition, despite the elementary character, they still lead to more general conclusions than all but Rényi's result. Some of the ideas in this manuscript are consequences of conversations and correspondence with Professor N. G. de Bruijn of Eindhoven, Netherlands. The second proof of Theorem 1 is a slight extension of an unpublished note of de Bruijn's [4].

2. The asymptotic normality of the distribution of the number of empty cells

The probability distribution of s is well-known and given by

$$P_{n,N}^{(s)}=\frac{N!}{s!N^n}\alpha_{N-s,n}\ , \qquad s=0, 1, \cdots, n\ ,$$

where $\alpha_{N-s,n}$ are the Stirling numbers of the second kind defined by

$$x^k=\sum_{j=1}^{k}\alpha_{j,k}x^{(j)}\ .$$

The factorial moments of s are given by

$$\mu_{[m]}=N^{(m)}\left(1-\frac{m}{N}\right)^n, \qquad m=0, 1, 2, \cdots , \tag{1}$$

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

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

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

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

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

$$\kappa_m=\sum_{j=1}^{m}\alpha_{j,m}\kappa_{[j]}\ , \tag{4}$$

where $\alpha_{j,m}$ are the Stirling numbers of the second kind. For $m\geqq 2$, the mth cumulant of V is $\kappa_m/\kappa_2^{m/2}$; thus, we need only show that $\kappa_m/\kappa_2^{m/2}\to 0$ for $m>2$ to establish the asymptotic normality of V. As a preliminary step, we now produce two proofs on the remarkable fact that $\kappa_m = O(N)$, $m=1, 2, \cdots$ as $N \to \infty$ with no conditions on n whatever.

THEOREM 1. *The mth cumulant of s, $\kappa_m = O(N)$, $N \to \infty$, for $m = 1, 2, \cdots$.*

FIRST PROOF. We proceed by first establishing two auxiliary lemmas.

LEMMA 1. *If $P(x)$ is a polynomial of degree $p>0$, then, for any $M \geqq p$ and $0<\theta<1$, $Q(x)=(M/\theta)P(x)-xP'(x)$ has at least as many real zeros as $P(x)$. If $P(x)$ has only real roots, then $Q(x)$ has only real roots.*

PROOF. Write

$$Q(x)=\begin{cases} -x^{M/\theta+1}\dfrac{d}{dx}(x^{-M/\theta}P(x)) , & x>0 \\ (M/\theta)P(x) , & x=0 \\ (-x)^{M/\theta+1}\dfrac{d}{dx}((-x)^{-M/\theta}P(x)) , & x<0 . \end{cases}$$

$Q(x)$ is a polynomial of degree p. Since $M/\theta>p$, as $x \to \pm\infty$, $(\pm x)^{-M/\theta}P(x) \to 0$. Thus for any $a>0$, the intervals (a, ∞), $(-\infty, -a)$ have at least as many zeros of $d/dx((\pm x)^{-M/\theta}P(x))$ as they have of $P(x)$. Consequently, $Q(x)$ has at least as many real zeros as $P(x)$.

LEMMA 2. *If $P(x)$ is a polynomial of degree $p>0$ with real roots $x_1 \leqq x_2 \leqq \cdots \leqq x_p < 0$, then the roots of $Q(x)$ are negative and do not exceed x_p.*

PROOF. For $P(x) \neq 0$, every zero of $Q(x)$ is a solution of

$$\frac{P'(x)}{P(x)}=\frac{d}{dx}\log P(x)=\frac{M}{\theta x} . \tag{5}$$

For $x>0$, we can assume $P(x)>0$ with no loss of generality. Then,

$$\frac{d}{dx}\log P(x)=\sum_{i=1}^{p}\frac{1}{(x-x_i)} \leqq \sum_{i=1}^{p}\frac{1}{x-x_p}=\frac{p}{x-x_p}<\frac{M}{x}<\frac{M}{\theta x} .$$

Thus there can be no positive roots of $Q(x)$. Trivially, zero is not a root of $Q(x)$. Hence all real roots are negative. For $x<0$, $P'(x)/P(x)$ has a simple pole at every zero of $P(x)$ (including multiple zeros). The conclusion is now immediate.

We now proceed to prove the theorem. Let

$$P(t)=(1+t)^N=\sum_{\nu=0}^{N}\binom{N}{\nu}t^\nu , \tag{6}$$

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

$$P_1(t)=P(t)-\frac{t}{N}P'(t)
=\sum_{\nu=0}^{N}\binom{N}{\nu}t^\nu-\sum_{\nu=1}^{N}\binom{N-1}{\nu-1}t^\nu
=\sum_{\nu=0}^{N}\binom{N}{\nu}\left(1-\frac{\nu}{N}\right)t^\nu
=(1+t)^{N-1}. \tag{7}$$

Thus, $P_1(t)$ is a monic polynomial of degree $N-1$ with all roots -1. We now proceed inductively. For $n\geqq 1$, define $P_{n+1}(t)=P_n(t)-(t/N)P_n'(t)$. Carrying out the induction and comparing with (2), we readily see that

$$P_n(t)=\sum_{\nu=0}^{N}\binom{N}{\nu}\left(1-\frac{\nu}{N}\right)^n t^\nu=\varphi_{n,N}(t). \tag{8}$$

Now let $M=N-1$, $\theta=(N-1)/N$. Then, define $Q_n(t)=NP_{n+1}(t)=NP_n(t)-tP_n'(t)$. $\varphi_1(t)$ satisfies the hypotheses of Lemma 1 and therefore has $N-1$ real roots. Thus $P_2(t)$ has $N-1$ real roots. Clearly, each $P_n(t)$ is of degree $N-1$. By induction, $Q_n(t)$ satisfies the hypotheses of Lemma 1 for each n and has $N-1$ real roots. From Lemma 2, each $\varphi_n(t)$ has all of its roots $\leqq -1$ and consequently for $n\geqq 1$, $P_n(t)$ has all roots $\leqq -1$. Hence, $N^{-1}\log P_n(t)=N^{-1}\log\varphi_{n,N}(t)=K_{n,N}(t)$ is analytic in $|t|<1$. Thus, for $|t|<1$,

$$Re\left(\frac{1}{N}\log P_n(t)\right)=\frac{1}{N}\log|P_n(t)|\leqq\frac{1}{N}\sum_{\nu=0}^{N}\binom{N}{\nu}|t|^2
=\log(1+|t|)\leqq\log 2. \tag{9}$$

We can now apply a well-known theorem of Carathéodory (see [2], [3] and [7]), that is, if $f(z)=\sum_{j=1}^{\infty}\alpha_j z^j$, $|z|<1$ and $Re[f(z)]\leqq 1$ for $|z|<1$, then $|\alpha_j|\leqq 2$ for all j. Thus, since

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

we have

$$|\kappa_{[\nu]}|\leqq N\nu!\log 4,$$

the conclusion now follows from (4).

Remark. $P_n(-1)$ has an interesting combinatorial interpretation. It is easily seen that $P_n(-1)$ is the probability that no cell is empty. Thus for $n<N$, $P_n(-1)=0$, $P_N(-1)=N!/N^N$.

SECOND PROOF. We employ the following introductory lemma.

LEMMA 3. *If $P(t)$ is a polynomial of degree p, with real roots in $[-1, 0]$, then $(t(d/dt))^n P(t)$ has p real roots in $[-1, 0]$ for $n=1, 2, \cdots$.*

PROOF. Clearly $tP'(t)$ is of degree p. The roots of $P'(t)$ always fall in the same interval as those of $P(t)$; the factor t introduces a root at zero. The conclusion the follows for $(t(d/dt))^n P(t)$.

The proof of the theorem follows. Let $P(t)=(1+t)^N=\sum_{\nu=0}^{N}\binom{N}{\nu}t^\nu$. Then let

$$P_n(t)=\left(t\frac{d}{dt}\right)^n P(t)=\sum_{\nu=0}^{N}\binom{N}{\nu}\nu^n t^\nu . \tag{10}$$

$P(t)$ has all zeros at -1, $P_n(t)$ has a simple root at zero for $n\geqq 1$. The remaining $N-1$ roots lie in $[-1, 0)$. Then, write

$$P_n(t)=N^n t\prod_{j=1}^{N-1}(t-\gamma_j)\ , \qquad -1\leqq\gamma_j<0\ . \tag{11}$$

Thus, from (2) and (10),

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

Hence,

$$\varphi_{n,N}(t)=t^{N-1}\prod_{j=1}^{N-1}(t^{-1}-\gamma_j)=\prod_{j=1}^{N-1}(1-\gamma_j t)\ . \tag{12}$$

Thus for $|t|<1$, $|\gamma_j t|<1$, $j=1, 2, \cdots, N-1$ and

$$\begin{aligned} K_{n,N}(t)&=\sum_{j=1}^{N-1}\log(1-\gamma_j t)\\ &=\sum_{j=1}^{N-1}\left(-\gamma_j t-\frac{\gamma_j^2 t^2}{2}-\frac{\gamma_j^3 t^3}{3}-\cdots\right)\end{aligned}$$

and from (3) we see that

$$\frac{\kappa_{[\nu]}}{\nu!}=-\sum_{j=1}^{N-1}\frac{\gamma_j^\nu}{\nu}\ .$$

Hence,

$$(\nu!)^{-1}|\kappa_{[\nu]}|\leqq\frac{1}{\nu}\sum_{j=1}^{N-1}|\gamma_j^\nu|\leqq\frac{N-1}{\nu}\ , \tag{14}$$

establishing the theorem.

COROLLARY. $s=\sum_{j=1}^{N-1}Y_j$, *where Y_j are independent Bernoulli random* *variables, with* $P\{Y_j=1\}=-\gamma_j$, $j=1,2,\cdots,N-1$.

PROOF. From (12),

$$\varphi_{n,N}(t)=\prod_{j=1}^{N-1}(1-\gamma_j t)$$

and $-1\leqq\gamma_j<0$. Now the factorial moment generating function for a Bernoulli random variable Y is

$$E[(1+t)^Y]=(1-p)+p(1+t)=1+pt\ ,$$

where $P\{Y=1\}=p$, $P\{Y=0\}=1-p$. The conclusion follows on setting $p=-\gamma_j$.

We now establish the asymptotic normality of $V=(s-E(s))/\sigma_s$.

THEOREM 2. *V is asymptotically distributed by the standard normal distribution whenever as $N\to\infty$,*

1. $\lim_{N\to\infty}\dfrac{n}{N}=c>0$,

2. $\lim_{N\to\infty}\dfrac{n}{N}=0$ *and* $\dfrac{n}{N^{5/6}}\to\infty$

or

3. $\dfrac{n}{N}\to\infty$, $\dfrac{3n}{N}-\log N\to-\infty$.

PROOF. In order to show that V has asymptotically ($N, n\to\infty$) the standard normal distribution, we need to show that $\kappa_m/\kappa_2^{m/2}\to 0$ for m>2. From Theorem 1, this is equivalent to showing that $N/\kappa_2^{m/2}\to 0$ for $m>2$ and this reduces to showing that $N/\kappa_2^{3/2}\to 0$. Moreover, elementary calculations show that

$$\kappa_2=N^2\left[\left(1-\frac{2}{N}\right)^n-\left(1-\frac{1}{N}\right)^{2n}\right]-N\left[\left(1-\frac{2}{N}\right)^n-\left(1-\frac{1}{N}\right)^n\right]. \tag{15}$$

Let $n/N=\alpha(N)$ and since $\alpha^k(N)=o(N)$ for every positive integer k, we have

$$\begin{aligned}\kappa_2&=-N\alpha e^{-2\alpha}+O(\alpha^2)+Ne^{-\alpha}(1-e^{-\alpha})+O(\alpha)\\&=Ne^{-\alpha}(1-e^{-\alpha}-\alpha e^{-\alpha})+O(\psi(\alpha))\ ,\end{aligned} \tag{16}$$

where $\psi(\alpha)=\max(\alpha,\alpha^2)$. Thus, the conclusion holds for $\alpha\to 0$ as $N\to\infty$, whenever $n/N^{5/6}\to\infty$ and for $\alpha\to\infty$ as $N\to\infty$ provided $3n/N-\log N\to-\infty$. The conclusion is obvious if α has a positive limit as $N\to\infty$.

THE UNIVERSITY OF WISCONSIN

REFERENCES

[1] David, F. N. and Barton, D. E. (1962). *Combinatorial Chance*, Charles Griffin and Company, Ltd., London.

[2] Carathéodory, C. (1970). Über den Variabilitätbereich der Koeffizieten von Potenzreihen, die gegebene Werte nicht annehmen, *Math. Ann.*, **64**, 95-115.

[3] Carathéodory, C. (1911). Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen, *Rend. Circ. Mat. Palermo*, **32**, 193-217.

[4] Bruijn, N. G. (1966). An asymptotic problem, *Notitie nr.* **30**, *Technical University, Eindhoven, Netherlands*, (mimeographed).

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

[6] Rényi, A. (1960). Three new proofs and a generalization of a theorem of Irving Weiss, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* A. **7**, 203-214.

[7] Riesz, F. (1911). Sur certaines systèmes singuliers d'équations intégrales, *Ann. Sci. Ecole Norm. Sup.*, **28**, 33-62.

[8] Sevast'yanov, B. A. and Chistyakov, V. P. (1964). Asymptotic normality in the classical ball problem, *Theory of probability and its applications*, **9**, 198-211.

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