LIMITING DISTRIBUTIONS IN SOME OCCUPANCY PROBLEMS[1, 2]

By Irving Weiss

Bell Telephone Laboratories, North Andover, Mass.

Section I

Introduction. The classical occupancy problem is concerned with the random distribution of a specified number of objects (r) in a given number of cells (N). No restriction is placed on the number of objects in any cell other than that the total number of objects equals r. The problem of finding exactly m cells empty for the case with r and N finite, and with all arrangements of r objects having equal probability can be expressed in closed form [1]. However, for large N, use of this formula for computation becomes exceedingly tedious. Several authors, [2] and [3] have stated without proof that under suitable restrictions on N, r the limiting distribution of the number of unoccupied cells as N, r approach infinity is normal.

By imposing the restriction $\alpha = r/N$, $\alpha > 0$, it will be shown that in the above occupancy problem the asymptotic distribution of the number of unoccupied cells is normal.

A modification of the above occupancy problem is the following: q objects are randomly distributed among N cells such that no more than one object is in any cell. The procedure is repeated w times. For example, with $w = k$, the maximum number of objects in any cell is k, one for each of k trials. It can be shown that by restricting $qw = \alpha N$, $\alpha > 0$, the normal asymptotic result given above holds. Also, by imposing the restriction $qw = N \log N/\lambda$ the number of unoccupied cells has asymptotically a Poisson distribution. This is an extension of the same results listed by Feller [1] for the classical occupancy problem. Proofs for the modified occupancy problem have been given by the author [7] and will not be given in this paper.

2. Outline of proof. In showing asymptotic normality our method will employ moments. We show that the moments converge to the moments of the normal distribution. From this it follows (by a theorem in Uspensky [4]) that the distribution of our random variable converges uniformly to the normal distribution.

3. Main results. With $\alpha = r/N$, $\alpha > 0$, we define a random variable X_j as follows:

$$X_j = 1 \quad \text{if cell } j \text{ is unoccupied after } r \text{ tosses.}$$
$$= 0 \quad \text{otherwise.}$$

Received January 9, 1956; revised December 26, 1957.

[1] This paper is part of the author's Ph.D. dissertation at Stanford University.

[2] Sponsored by the Office of Naval Research.

Assuming all N events are equally likely and that the r trials are independent of each other:

$$E(X_{i_1} \cdot X_{i_2}, \cdots, X_{i_s}) \doteq \left(1 - \frac{s}{N}\right)^r.$$

Let X equal the number of unoccupied cells

$$X = \sum_{i=1}^{N} X_i$$

$$E(X) = N\left(1 - \frac{1}{N}\right)^r = N\left(1 - \frac{1}{N}\right)^{\alpha N}$$

$$\lim_{N\to\infty} \frac{E(X)}{N} = e^{-\alpha}$$

As N becomes infinite, $E(X)$ becomes infinite but $E(X)/N$ approaches a finite limit.

We will prove that the random variable X has an asymptotically normal distribution by showing that

$$\lim_{N\to\infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = 1 \cdot 3 \cdots (k - 1) \quad \text{for } k \text{ even}$$

$$= 0 \quad \text{for } k \text{ odd}$$

The general kth moment, μ_k, is

$$\mu_k = E(X - E(X))^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} E(X^{k-r})(E(X))^r.$$

As shown in Theorem 1 of Section II, by using Stirling numbers of the first and second kind, μ_k can be expressed as follows:

$$\mu_k = \sum_{r=0}^{k} \sum_{p=1}^{k-r} \sum_{j=1}^{p} (-1)^r \binom{k}{r} N^{j+r} S_p^j \mathfrak{S}_{k-r}^p \left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r}.$$

It can be shown (see [5]) that

$$\left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r} = \exp\left[-\alpha(p + r)\right] \exp\left[-\sum_{t=1}^{\infty} \frac{\alpha(r + p^{t+1})}{(t + 1)N^t}\right].$$

Now

$$\exp\left[-\sum_{t=1}^{\infty} \frac{\alpha(r + p^{t+1})}{(t + 1)N^t}\right] = \sum_{n=0}^{\infty} \sum_{\{m_i : \sum_{i=1}^{n} i m_i = n\}} \frac{a_1^{m_1} \cdots a_n^{m_n}}{m_1! \cdots m_n!} \frac{1}{N^n}$$

$$= \sum_{n=0}^{\infty} K_n(r, s) \frac{1}{N^n}$$

where

$$a_i = \frac{-(r + p^{i+1})}{i.+ 1}$$

Substituting above and noting that $\mathcal{S}_p^0 = \mathcal{S}_{k-r}^0 = 0$ we have

$$\mu_k = \sum_{r=0}^{k} \sum_{s=r}^{k} \sum_{v=r}^{s} (-1)^r \binom{k}{r} S_{s-r}^{v-r} \mathcal{S}_{k-r}^{s-r} N^v e^{-\alpha s} \sum_{n=0}^{\infty} K_n(r, s) \frac{1}{N^n}$$

where

$$p + r = s$$

$$j + r = v$$

Collecting like powers of N

$$\mu_k = \sum_{v=k}^{-\infty} N^v \left[\sum_{s=v}^{k} e^{-\alpha s} b_{s,v,0} + \sum_{s=v+1}^{k} e^{-\alpha s} b_{s,v+1,1} + \cdots + e^{-\alpha k} b_{k,k,k-v} \right]$$

(1) $$= \sum_{v=k}^{-\infty} N^v \left[\sum_{s=v}^{k} e^{-\alpha s} \left(\sum_{n=0}^{s-v} b_{s,v+n,n} \right) \right]$$

$$= \sum_{v=k}^{[k/2]} N^v \left[\sum_{s=v}^{k} e^{-\alpha s} a_s(v) \right] + R(r, N)$$

where

$$b_{s,v+n,n} = \sum_{r=0}^{k} (-1)^r \binom{k}{r} S_{s-r}^{v+n-r} \mathcal{S}_{k-r}^{s-r} K_n(r, s)$$

$$= \sum_{r=0}^{k} (-1)^r \binom{k}{r} f(k - r)$$

$$= \Delta^k f(0)$$

and

$$[k/2] = k/2 \qquad \text{for } k \text{ even}$$

$$= \frac{k-1}{2} \qquad \text{for } k \text{ odd}$$

As shown in [6] $b_{s,v+n,n}$ is the kth difference of $f(0)$. By Lemma 1,

$$b_{s,v+n,n} \equiv 0 \qquad \text{for} \qquad v > k/2$$

$$= ck! \qquad \text{for} \qquad v = k/2$$

where c is the product of the coefficients of the highest degree terms in r of S_{s-r}^{v+n-r}, $\mathcal{S}_{k-r}^{s-r}$ and $K_n(r, s)$.

For a given k, $R(r, N)$ is a bounded function of r and N. This is an immediate consequence of the analyticity of μ_k. From (1), the highest power of N in

$R(r, N)$ is $N^{[k/2]-1}$. Therefore

$$R(r, N) = O(N^{[k/2]-1}).$$

Incorporating these results in (1) for k even

$$\mu_k = N^{k/2} \sum_{s=k/2}^{k} e^{-\alpha s} a_s(k) + O(N^{k/2-1})$$

where

$$a_s(k) = \sum_{n=0}^{s-k/2} b_{s,k/2+n,n} = \sum_{n=0}^{s-k/2} ck!$$

Using Lemma 1, it follows that

$$a_s(k) = D_{k/2,0}(-1)^h(\alpha + 1)^h \binom{k/2}{h}$$

where

$$D_{k/2,0} = 1 \cdot 3 \cdots (k - 1)$$

$$h = s - k/2$$

Substituting above

$$\mu_k = N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} D_{k/2,0} + O(N^{k/2-1})$$

Noting that $\sigma^2 = \mu_2$, forming the ratio

$$\frac{\mu_k}{(\sigma^2)^{k/2}} = \frac{D_{k/2,0} N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} + O(N^{k/2-1})}{N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} + O(N^{k/2-1})}$$

dividing numerator and denominator by $N^{k/2}$ and then letting $N \rightarrow \infty$,

$$\lim_{N \to \infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = D_{k/2,0} \quad \text{for} \quad k = 2, 4, \cdots.$$

For k an odd positive integer,

$$\lim_{N \to \infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = 0.$$

This follows from the fact that $b_{s,v+n,n} = 0$ for $v \geqq k/2$ as v being a positive integer cannot equal $k/2$. Therefore,

$$\mu_k = O(N^{(k-1)/2})$$

while

$$(\sigma^2)^{k/2} = O(N^{k/2}).$$

Section II

THEOREM 1.

$$\mu_k = \sum_{r=0}^{k} \sum_{p=1}^{k-r} \sum_{j=1}^{p} (-1)^r \binom{k}{r} N^{j+r} S_p^j \mathfrak{S}_{k-r}^p \left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r}$$

where S_p^j and $\mathfrak{S}_{k-r}^p$ are Stirling numbers of the first and second kind respectively.

PROOF.

$$\mu_k = E(X - E(X))^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} E(X^{k-r})[E(X)]^r. \tag{2}$$

By the multinomial expansion with

$$\begin{aligned} \lambda(s_i) &= 1 \quad \text{for} \quad s_i > 0 \\ &= 0 \quad \text{for} \quad s_i = 0 \end{aligned}$$

and

$$\sum_{i=1}^{N} \lambda(s_i) = p \qquad 1 \leqq p \leqq s$$

$$X_i^{s_i} = X_i^{\lambda(s_i)} \qquad \text{as } X_i = 0 \text{ or } 1.$$

we have

$$X^s = \sum_{p=1}^{s} \sum_{\left\{\substack{s_i : \sum_{i=1}^N \lambda(s_i) = p \\ \sum_{i=1}^N s_i = s}\right\}} \frac{s!}{s_1! \cdots s_N!} \binom{N}{p} X_1^{\lambda(s_1)} \cdots X_N^{\lambda(s_N)}$$

From Jordan [5]

$$\mathfrak{S}_s^p = \frac{s!}{p!} \sum_{\left\{\substack{s_i : s_i > 0 \\ \sum_{i=1}^p s_i = s}\right\}} \frac{1}{s_1! \cdots s_p!}$$

Substituting above to eliminate the second summation and taking expectations,

$$\begin{aligned} E(X^s) &= \sum_{p=1}^{s} p! \binom{N}{p} \mathfrak{S}_s^p E(X_1^{\lambda(s_1)} \cdots X_N^{\lambda(s_N)}) \\ &= \sum_{p=1}^{s} (N)_p \left(1 - \frac{p}{N}\right)^{\alpha N} \mathfrak{S}_s^p \end{aligned}$$

From Jordan [5]

$$(N)_p = \sum_{j=1}^{p} S_p^j N^j$$

Substituting above in (2) with

$$[E(X)]^r = N^r\left(1 - \frac{1}{N}\right)^{\alpha N r}$$

yields the desired result.

THEOREM 2. *The degree of $K_n(r, s)$ defined in equation (1), considered as a polynomial in r, is obtained from the term of the summation in which $m_1 = n$ and $m_2 = m_3 = \cdots = m_n = 0$.*

PROOF. The highest power of r in a_i is $i + 1$. For a given n we have to determine $m_1, \cdots, m_n$ which will maximize the highest power of r subject to the restriction that

$$\sum_{i=1}^{n} i m_i = n.$$

Maximizing the power of r is equivalent to maximizing

$$2m_1 + 3m_2 + \cdots + (n + 1)m_n = n + \sum_{i=1}^{n} m_i$$

Maximizing $\sum_{i=1}^{n} m_i$ subject to the above restraint yields

$$\sum_{i=1}^{n} m_i = n - \sum_{i=2}^{n} (i - 1)m_i$$

The maximum is attained when $m_i = 0$; $i = 2, \cdots, n$. Therefore, the power of r is maximized when $m_1 = n$. From the definition of a_1, it is readily seen that the degree of r in $K_n(r, s)$ is $2n$ and that the coefficient of this highest degree term is $(-\alpha)^n/2^n n!$.

LEMMA 1. *Let $b_{s,v+n,n}$ be defined as above. Then*

$$b_{s,v+n,n} \equiv 0 \quad \textit{for} \quad v > k/2$$
$$= ck! \quad \textit{for} \quad v = k/2$$

where

$$c = \frac{C_{(s-k/2-n),0}}{[2(s - k/2 - n)]!} \cdot \frac{D_{(k-s),0}}{[2(k - s)]!} \cdot \frac{(-\alpha)^n}{n!2^n}.$$

PROOF. From Jordan [5], S_n^{n-m} and $\mathfrak{S}_n^{n-m}$ are polynomials in n of degree $2m$, i.e.:

$$S_n^{n-m} = C_{m,0}\frac{(n)_{2m}}{(2m)!} + \text{terms in } n \text{ of degrees less than } 2m$$

$$\mathfrak{S}_n^{n-m} = D_{m,0}\frac{(n)_{2m}}{(2m)!} + \text{terms in } n \text{ of degrees less than } 2m$$

where

$$C_{m,0} = (-1)^m D_{m,0}$$

As the product of a finite number of polynomials is also a polynomial,

$$S_{s-r}^{v+n-r} S_{k-r}^{s-r} K_n(r, s)$$

is also a polynomial. Its degree in r for fixed v, s, n is

$$2(s - v - n) + 2(k - s) + 2n = 2(k - v)$$

It follows from elementary properties from the calculus of finite differences that

$$\begin{aligned} b_{s,v+n,n} &\equiv 0 \qquad \text{for} \quad v > k/2 \\ &= ck! \qquad \text{for} \quad v = k/2 \end{aligned}$$

where c is the coefficient of r^k in the product polynomial. That c is the product of the above three factors is apparent from the polynomial expansion of Stirling numbers and from Theorem 2.

Acknowledgment. I am deeply indebted to Dr. Joseph F. Daly whose helpful guidance made this work possible.

REFERENCES

[1] WILLIAM, FELLER, "An introduction to probability theory and its applications," John Wiley & Sons, New York, 1950.
[2] F. N. DAVID, "Two combinatorial tests of whether a sample has come from a given population," *Biometrika*, 37, 1950, pp. 97–110.
[3] V. W. SCHAFER, "Das Mutungsproblem der Besetzungs-Verteilung," *Mitteilungsblatt Math. Statist.* 5, (1954), 1–38.
[4] J. V. USPENSKY, "Introduction to mathematical probability," McGraw-Hill Book Co., New York, 1937.
[5] CHARLES JORDAN, "Calculus of Finite Differences," Chelsea Publishing Co., New York, 1950.
[6] TOMLINSON FORT, "Finite differences and difference equations in the real domain," Oxford, 1948.
[7] IRVING WEISS, "Limiting distributions in some occupancy problems," Ph.D. dissertation, Stanford University, Stanford, California.