TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 32, Núm. 2, 1981, pp. 131 a 134

NOTAS

A NOTE ON THE POLYA-EGGENBERGER DISTRIBUTIONS WITH s = 1 AND SOME APPLICATIONS

William C. Guenther
Department of Statistics
University of Wyoming
Laramie, Wyoming 82071

Although cumulative probabilities for the Pólya-Eggenberger distributions can be easily found on modern computing facilities, the special cases $s = 1$ can be handled with hypergeometric tables. Applications are mentioned.

Key Words
Pólya-Eggenberger distribution
Inverse Pólya-Eggenberger distribution
Hypergeometric distribution
Inverse hypergeometric distribution

Johnson and Kotz (1969, Chapter 9, Section 4, pp. 229-232) give a short discussion of Pólya-Eggenberger distributions, together with references. Similar material has been published by Patil and Joshi (1968, pp. 30-37). On their page 230 the former authors state: "Initially it is supposed that there are a white balls and b black balls in an urn. One ball is drawn at random, and then replaced, together with s balls of the same color. If this procedure is repeated n times and U represents the total number of times a white ball is drawn, then the distribution of U is a Pólya-Eggenberger distribution with parameters n, a, b, s." The distribution of U is given by

$$Pr\,(U=u)=\binom{n}{u}\frac{a\,[a+s]\cdots[a+(u-1)s]\,b\,[b+s]\cdots[b+(n-u-1)s]}{[a+b]\,[a+b+s]\cdots[a+b+(n-1)s]}\,, \quad (1)$$

$$u=0, 1, 2, ..., n.$$

With modern desk and hand-held calculators it is a simple matter to calculate $Pr\ (U = u)$ and $Pr\ (U \leqslant u)$. However, for the case $s = 1$, which has occurred numerous times in the literature, these probabilities can be read from the hypergeometric table of Lieberman and Owen (1961).

If $s = 1$, (1) can be rewritten as

$$Pr\,(U = u) = \frac{\binom{a+u-1}{a-1}\binom{b+n-u-1}{b-1}}{\binom{a+b+n-1}{a+b-1}}, u = 0, 1, 2, \ldots, n. \quad (2)$$

If we let $U + a = X$ we have

$$Pr\,(X = x) = \frac{\binom{x-1}{a-1}\binom{a+b+n-1-x}{b-1}}{\binom{a+b+n-1}{a+b-1}}, x = a, a+1, \ldots, n+a. \quad (3)$$

Both (2) and (3) are forms the inverse hypergeometric probability function with parameters $N = a + b + n - 1,\ k = a + b - 1,\ c = a$. Hence

$$Pr\,(X \leqslant x) = 1 - P(a + b + n - 1, x, a + b - 1, a - 1) \quad (4)$$

and

$$Pr\,(U \leqslant u) = 1 - P(a + b + n - 1, u + a, a + b - 1, a - 1) \quad (5)$$

where we have used the Lieberman and Owen (1961) notation

$$P\,(N, n, k, r) = \sum_{x=d}^{r} p\,(N, n, k, x), \quad (6)$$

$$p\,(N, n, k, x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, d \leqslant x \leqslant g \quad (7)$$

where $d = \max\,[0, n - (N - k)]$, $g = \min\,[k, n]$.

For more details and numerous applications of the inverse hypergeometric, see a review type paper by the author (Guenther (1975)).

If balls are drawn from the urn until k white balls have been selected, then the probability that $k + v$ drawings are required is given by Johnson and Kotz (1969, p. 232) as

$$Pr(V=v) = \frac{a[a+s]\cdots[a+(k-1)s]b[b+s]\cdots[b+(v-1)s]}{[a+b][a+b+s]\cdots[a+b+(k+v-1)s]} \binom{k+v-1}{v}, \tag{8}$$

$$v = 0, 1, 2, \ldots$$

For the case $s = 1$ this can be rewritten

$$Pr(V=v) = \frac{a}{a+k} \frac{\binom{a+b-1}{a}\binom{k+v-1}{k-1}}{\binom{a+b+k+v-1}{a+k}}, \quad v=0,1,2,\ldots \tag{9}$$

If we let $V+k=Y$ this can be rewritten as

$$Pr(Y=y) = \frac{a}{a+k} \frac{\binom{a+b-1}{a}\binom{y-1}{k-1}}{\binom{a+b+y-1}{a+k}}, \quad y=k, k+1, \ldots \tag{10}$$

The distribution whose density is given by (8) is called the inverse Pólya-Eggenberger distribution by Johnson and Kotz. They state that the distribution is related to the binomial. To obtain the relationship between cumulative sums, the same argument relating negative binomial sums to binomial sums or inverse hypergeometric sums to hypergeometric sums can be used. (See, i.e., Guenther (1975), p. 131). Hence we get

$$\begin{aligned} Pr(k \leqslant Y \leqslant y) &= Pr(k \leqslant U \leqslant y) \\ &= 1 - Pr(U \leqslant k-1) \end{aligned} \tag{11}$$

where y has replaced n in the notation of (2), (3), (4), (5). Using (5) on (11) gives

$$Pr(k \leqslant Y \leqslant y) = P(a+b+y-1,\ k+a-1,\ a+b-1,\ a-1). \tag{12}$$

Also

$$Pr(0 \leqslant V \leqslant v) = Pr(k \leqslant Y \leqslant v+k). \tag{13}$$

Apparently statistical applications of the inverse Pólya-Eggenberger distributions are scarce. One did occur recently in a paper by Wolfe (1977). In deriving the distribution of a statistic N_m for his two-sample median test, he found that

$$Pr\,(N_m = n) = \frac{m+1}{2n-2r+m+1} \; \frac{\binom{m}{(m+1)/2}\binom{n-1}{r-1}}{\binom{m+n}{(m+2r-1)/2}}, n = r, r+1, \ldots$$

which is equivalent to

$$Pr\,(N_m = n) = \frac{(m+1)/2}{r+(m+1)/2} \; \frac{\binom{m}{(m+1)/2}\binom{n-1}{r-1}}{\binom{m+n}{r+(m+1)/2}}, n = r, r+1, \ldots \quad (14)$$

Identifying $a = (m+1)/2,\ k = r,\ a+b-1 = m,\ y = n$ we have that

$$Pr\,(N_m \leqslant n) = P\left(m+n,\ r+\frac{m-1}{2},\ m,\ \frac{m-1}{2}\right), \quad (15)$$

Hence the hypergeometric table and (15) can be used to produce Wolfe's Table 1.

REFERENCIAS

GUENTHER, WILLIAM C.: (1975). The inverse hypergeometric-a useful model. *Statistica Neerlandica*, 29, 129-144.

JOHNSON, N. L. and KOTZ, S.: (1969). *Distributions in Statistics: Discrete Distributions.* Houghton Mifflin Company, Boston.

LIEBERMAN, G. J. and OWEN, D. B.: (1961). *Tables of the Hypergeometric Probability Distribution.* Stanford University Press, Stanford, Calif.

PATIL, G. P. and JOSHI, S. W.: (1968). *A Dictionary and Bibliography of Discrete Distributions.* Hafner Publishing Company, New York.

WOLFE, DOUGLAS A.: (1977). Two-stage two-sample median test. *Technometrics,* 19, 495-501.