The Annals of Statistics
1981, Vol. 9, No. 5, 1123–1126

A REPRESENTATION FOR MULTINOMIAL CUMULATIVE DISTRIBUTION FUNCTIONS

BY BRUCE LEVIN

Columbia University

A re-expression of the usual representation of the multinomial distribution as the conditional distribution of independent Poisson random variables given fixed sum provides a convenient new way to compute multinomial cumulative distribution functions.

1. A representation of the multinomial cumulative distribution function. The following theorem offers a convenient way to compute multinomial cdf's, although it appears not to have been explicitly stated in the literature on multinomial distributions.

THEOREM. *Let* $(n_1, \cdots, n_t)$ *have a t-category multinomial distribution with sample size N and parameters* $(p_1, \cdots, p_t)$. *Let* $(a_1, \cdots, a_t)$ *be non-negative integers, and define*

$$p_N = p_N(t, p_1, \cdots, p_t, a_1, \cdots, a_t) = P(n_1 \leqq a_1, \cdots, n_t \leqq a_t).$$

Then for any real number $s > 0$,

$$p_N = \frac{N!}{s^N e^{-s}} \left\{ \prod_{i=1}^{t} P(X_i \leqq a_i) \right\} P(W = N), \tag{1}$$

where $X_i \sim$ indep $\mathbb{P}(sp_i)$ = *independent Poisson r.v.'s with mean* sp_i *and W is a sum of independent truncated Poisson r.v.'s, namely* $W = \sum_1^t Y_i$ *where* $Y_i \sim T\mathbb{P}_{a_i}(sp_i)$ = *truncated Poisson*(sp_i) *with range* $0, 1, \cdots, a_i$.

The theorem may be proved by applying Bayes' Theorem to the usual representation of multinomial frequencies as independent Poisson frequencies conditional on their sum being fixed. Let A_i denote the event $[X_i \leqq a_i]$ where $X_i \sim$ indep $\mathbb{P}(sp_i)$. Then the multinomial cdf is

$$P(A_1 \cdots A_t \mid \textstyle\sum_1^t X_i = N) = \dfrac{P(A_1 \cdots A_t)}{P(\sum_1^t X_i = N)} P(\textstyle\sum_1^t X_i = N \mid A_1 \cdots A_t).$$

The result follows by noting that $\sum X_i \sim \mathbb{P}(s)$ and that the conditional distribution of X_i given A_i is $T\mathbb{P}_{a_i}(sp_i)$. An alternative proof follows directly from the generating function for p_N which is well-known to be

$$\sum_{N=0}^{\infty} p_N \frac{u^N}{N!} = \prod_{i=1}^{t} \{1 + p_i u + \cdots + (p_i u)^{a_i}/a_i!\}, \tag{2}$$

(see, e.g. Good, 1957, or Barton and David, 1959). We write $u = sx$ for any positive s, and multiply and divide the i-th factor on the right of (2) by $\sum_{j=0}^{a_i} (sp_i)^j/j!$ to find

Received December 19, 1980; revised February, 1981.

AMS 1970 *subject classifications.* Primary 62E15, 62G30, 62H10; secondary 05A15, 41A60.

Key words and phrases. Multinomial cdf, maximum cell distribution, generating functions, representations.

$$\textstyle\sum_{N=0}^{\infty} p_N \frac{s^N x^N}{N!} = e^s \prod_{\iota=1}^{t} \{\sum_{j=0}^{a_\iota} e^{-sp_\iota}(sp_\iota)^j/j!\} \prod_{\iota=1}^{t} \left[\sum_{j=0}^{a_\iota} \frac{\{(sp_\iota)^j/j!\}x^j}{\{\sum_{j=0}^{a_\iota} (sp_\iota)^j/j!\}}\right]$$

$$= e^s \prod_{\iota=1}^{t} P(X_\iota \leqq a_\iota) \prod_{\iota=1}^{t} Ex^{Y_\iota} = e^s \prod_{\iota=1}^{t} P(X_i \leqq a_\iota)\, Ex^W,$$

where $X_\iota \sim$ indep $\mathbb{P}(sp_\iota)$ and $Y_\iota \sim$ indep $T\mathbb{P}_{a_\iota}(sp_\iota)$. Comparing coefficients of x^N concludes the proof. We are indebted to C.L. Mallows for the succinct first proof.

As a corollary we obtain a representation for the cdf of the maximum multinomial cell frequency by taking $a_1 = \cdots = a_t = m$, so that (1) gives $P(\max_\iota \mathrm{n}_\iota \leqq \mathrm{m})$. The theorem shows that for small t the multinomial cdf is as easy to compute (exactly) as the convolution of t truncated Poisson r.v.'s. Of greater practical importance is the fact that for large t the Central Limit Theorem offers an approximation to the last term in (1) with two virtues: (a) the approximation is quick to compute with hand-held calculators and does not require special tables; and (b) the approximation works well even for cases in which the Bonferroni-Mallows bounds (see Mallows, 1968)

$$1 - \textstyle\sum_{\iota=1}^{t} P(n_\iota > a_\iota) \leqq p_N \leqq \prod_{\iota=1}^{t} P(n_\iota \leqq a_\iota), \tag{3}$$

are wide, e.g. when each term $P(n_\iota > a_\iota)$ is of order $1/t$. We illustrate the approximation in Section 2 through two examples that have appeared in the literature.

In general, the use of a first-order normal approximation to $P(W = N)$ in (1) is not guaranteed to produce an estimate between the Bonferroni or Mallows bounds. In our second example below we use an Edgeworth expansion through terms of order $O(t^{-1})$ to improve accuracy. While Edgeworth corrections are not guaranteed to work either, numerical experience suggests they are quite adequate for three- or four-place accuracy. To state the approximations we need the first four moments of a truncated Poisson r.v., say $Y \sim T\mathbb{P}_m(\lambda)$. These are furnished through the following lemma, which we state without proof, concerning the factorial moments of Y,

$$\mu_{(r)} = E\{Y(Y-1)\cdots(Y-r+1)\}, \qquad \mu_{(0)} = 1.$$

LEMMA. *Let $Y \sim T\mathbb{P}_m(\lambda)$. Then*

$$\mu = \mu_{(1)} = EY = \lambda\left(1 - \frac{\lambda^m/m!}{\sum_{k=0}^{m} \lambda^k/k!}\right) = \lambda\left\{1 - \frac{P(X=m)}{P(X \leqq m)}\right\} \quad \textit{where} \quad X \sim \mathbb{P}(\lambda)$$

$$\sigma^2 = \operatorname{Var} Y = \mu - (m - \mu)(\lambda - \mu)$$

and in general,

$$\mu_{(r+1)} = \lambda\mu_{(r)} - m^{(r)}(\lambda - \mu) \quad \textit{for} \quad r = 1, 2, \cdots$$

where

$$m^{(r)} = m(m-1)\cdots(m-r+1).$$

The third and fourth central moments are obtained from the usual formulae

$$\mu_2 = \mu_{(2)} + (\mu - \mu^2),$$

$$\mu_3 = \mu_{(3)} + \mu_{(2)}(3 - 3\mu) + (\mu - 3\mu^2 + 2\mu^3),$$

$$\mu_4 = \mu_{(4)} + \mu_{(3)}(6 - 4\mu) + \mu_{(2)}(7 - 12\mu + 6\mu^2) + (\mu - 4\mu^2 + 6\mu^3 - 3\mu^4).$$

The Edgeworth approximation we use is

$$P(W = N) \doteq f\left(\frac{N - \sum_1^t \mu_\iota}{\sqrt{\sum_1^t \sigma_\iota^2}}\right)\frac{1}{\sqrt{\sum_1^t \sigma_\iota^2}}, \tag{4}$$

where $\mu_\iota = EY_\iota$, $\sigma_\iota^2 = \operatorname{Var} Y_\iota$ and where

$$f(x) = \left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right)$$

$$\cdot\left\{1 + \frac{\gamma_1}{6}(x^3 - 3x) + \frac{\gamma_2}{24}(x^4 - 6x^2 + 3) + \frac{\gamma_1^2}{72}(x^6 - 15x^4 + 45x^2 - 15)\right\},$$

$$\gamma_1 = \text{coefficient of skewness} = \frac{1}{\sqrt{t}} \frac{\left(\frac{1}{t} \sum_1^t \mu_{3,i}\right)}{\left(\frac{1}{t} \sum_1^t \sigma_i^2\right)^{3/2}},$$

and

$$\gamma_2 = \text{coefficient of excess} = \frac{1}{t} \frac{\frac{1}{t} (\sum_1^t \mu_{4,i} - 3\sigma_i^4)}{\left(\frac{1}{t} \sum_1^t \sigma_i^2\right)^2}.$$

The first-order normal approximation is (4) with $f(x)$ replaced by its first factor.

2. Examples.

Example 1. Mallows (1968) discusses an example of the maximum cell frequency in the multinomial distribution with $N = 500$, $t = 50$, $p_1 = \cdots = p_{50} = .02$ and $a_1 = \cdots = a_{50} = m = 19$. The parameter s in (1) is a 'tuning' parameter which may be chosen for convenience and stable computation. With N large, a reasonable choice is $s = N$ so that the first factor in (1) becomes $\sqrt{2\pi N}$, ignoring the error in Stirling's approximation for N! While the choice of s that optimizes accuracy of the normal approximation is an open question, we have found $s = N$ to be generally satisfactory, and it is also natural in that Np_i are the expected cell frequencies. For the final factor in (1) we apply the first-order normal approximation to $P(W = N)$ with $Y_i \sim T\mathbb{P}_{19}(10)$ to find $P(W = N) \doteq .018120$. Thus

$$P(n_1 \leqq 19, \cdots, n_{50} \leqq 19) \doteq \sqrt{2\pi \cdot 500}\ (.99655)^{50}(.018120) = .8545,$$

which lies between Mallows' bounds of $.8437 \leqq p_N \leqq .8551$.

Example 2. Barton and David (1959) discuss the distribution of max n_i for $N = t = 12$, $p_1 = \cdots = p_{12} = 1/12$ and $a_1 = \cdots = a_{12} = m$ for $m = 1, 2, \cdots$. We consider the case $m = 3$. Proceeding as we did in Example 1, we choose $s = 12$ and find that the first-order normal approximation yields

$$P(n_1 \leqq 3, \cdots n_{12} \leqq 3) \doteq \frac{12!}{12^{12}e^{-12}} (.981012)^{12}(.12441) = .8643.$$

The Bonferroni-Mallows bounds (3) yield $.8340 \leqq p_{12} \leqq .8461$, while the exact value is .8371 to four decimals. Thus the first-order normal approximation to $P(W = N)$ has overestimated the correct value by 3.2%. The Edgeworth approximation (4) gives $P(W = N) \doteq .12044$ yielding $p_{12} \doteq .8367$, reducing the relative error to 0.05%.

Note that for the case $m = 2$ the first-order normal approximation yields $P(n_1 \leqq 2, \cdots, n_{12} \leqq 2) \doteq .3210$, (the exact value is .3127) with a relative error of 2.7%, while the Bonferroni-Mallows bounds are rather wide, $.1359 \leqq P(n_1 \leqq 2, \cdots, n_{12} \leqq 2) \leqq .4079$.

3. Discussion. While the generating function (2) for p_N is well-known, the representation (1) and its consequent normal approximation appear to be not widely recognized. Kozelka (1956) gave a different normal approximation, based on the asymptotic normality of the binomial terms in the first Bonferroni inequality. Because his normal approximation is to a bound rather than an exact term, we expect our application to be superior. Good (1957) and Barton and David (1959) give different asymptotic expansions for large t based on the generating function (2), although our application of the Edgeworth expansion to $P(W = N)$ is more straightforward, is applicable to any $(p_1, \cdots p_t)$ and $(a_1, \cdots, a_t)$, and appears to offer excellent accuracy. Good (1957) and Riordan (1958) give recursive relations in the equiprobable case which are awkward. For exact computation, the convolution of truncated Poisson distributions is easy to program and offers an attractive alternative to enumerative methods such as Freeman's (1979), whose program covers only the equiprobable case. We also remark that our method of computing $\{p_N : N = 1, 2, \cdots\}$ for fixed t, $(p_1, \cdots, p_t)$ and $(a_1, \cdots, a_t)$ is especially well-suited for calculating the operating characteristics of certain sequential methods for multinomials, e.g. the inverse sampling procedure of Cacoullos and Sobel (1966) for selecting the most probable multinomial outcome. Perhaps most important is the observation that the Bonferroni-Mallows bounds, which are so easy to calculate, are occasionally too wide to be useful, so that an equally simple point approximation is handy to have in order to 'fill the gap'.

REFERENCES

Barton, D. E. and David, F. N. (1959) Combinatorial Extreme Value Distributions. *Mathematika* **6** 63–76.

Cacoullos, T. and Sobel, M. (1966) An Inverse Sampling Procedure for Selecting the Most Probable Event in a Multinomial Distribution. *Proc. Int. Symp. on Multivariate Analysis.* Ed. P. R. Krishnaiah 423–444. Academic, New York.

Freeman, P. R. (1979) Exact Distribution of the Largest Multinomial Frequency. *Appl. Statist.* **28**(3) 333–336.

Good, I. J. (1957) Saddle-point Methods for the Multinomial Distribution. *Ann. Math. Statist.* **28** 861–881.

Kozelka, Robert M. (1956) Approximate Upper Percentage Points for Extreme Values in Multinomial Sampling. *Ann. Math. Statist.* **27** 507–512.

Mallows, C. L. (1968) An Inequality Involving Multinomial Probabilities. *Biometrika* **55** 422–424.

Riordan, John (1958) *An Introduction to Combinatorial Analysis.* 104. Wiley, New York.

School of Public Health
Division of Biostatistics
Columbia University
600 West 168th Street
New York, New York 10032