ON MULTIVARIATE MODIFIED POLYA AND INVERSE POLYA DISTRIBUTIONS AND THEIR PROPERTIES

K. G. JANARDAN AND G. P. PATIL

(Received Oct. 13, 1969; revised June 21, 1973)

1. Introduction and summary

In connection with the negative multinomial distribution, Sibuya, Yoshimura, and Shimizu [5] discuss a slight modification of the Eggenberger Polya Urn Scheme and obtain a multivariate distribution which tends to the negative multinomial under certain limiting conditions. In this paper we treat this distribution to be called "multivariate modified Polya (MMP) distribution" and discuss some of its interesting structural properties, which we hope are at least of pedagogical interest. If the random vector (rv) $(x_1, x_2, \cdots, x_s)$ has s-variate modified Polya distribution, then it is shown that the marginal distribution of the subvector $(x_1, x_2, \cdots, x_r)$, and the conditional distribution of $(x_{r+1}, x_{r+2}, \cdots, x_s)$ given $(x_1, x_2, \cdots, x_r)$ do not belong to the family of MMP distributions unlike in the case of multivariate Polya and inverse Polya families (see [2], [3]), these are shown to be products of Gaussian hypergeometric function (see [1]) and a r-variate modified Polya distribution. We also introduce "multivariate inverse Polya (MMIP) distribution," and give similar properties for MMIP distribution.

2. Genesis of the MMP and MMIP distributions

(a) Suppose an urn contains $N_1, N_2, \cdots, N_s$ $\left(\sum_{i=1}^{s} N_i = M\right)$ balls of s-different colors as well as N_0 white balls. We draw a ball and if it is colored, it is replaced with additional $c_1, c_2, \cdots, c_s$ balls of s-colors and if it is white, it is replaced with additional $d=\sum_{i=1}^{s} c_i$ white balls. This is repeated n times. If x_i denotes the number of balls of ith color, and x_0 the number of white balls such that $\sum_{i=0}^{s} x_i = n$, then the rv $\boldsymbol{x}=(x_1, x_2, \cdots, x_s)$ has been shown (see [5]) to have a probability distribution whose probability function (pf) is given by

$$\text{mmp}\,(\boldsymbol{x}; d, n, N) = \binom{n}{\boldsymbol{x}}\left(\frac{N_0^{(x_0,d)}M^{(x,d)}}{N^{(n,d)}}\right)\prod_{i=1}^{s}\left(\frac{N_i}{M}\right)^{x_i} \tag{2.1}$$

where $x_i = 0, 1, 2, \cdots, n$ for $i = 1, 2, \cdots, s$ so that $x = \sum_{i=1}^{s} x_i$, $N = N_0 + M$; $a^{(k,d)} = a(a+d)(a+2d)\cdots(a+(k-1)d)$ and $\binom{n}{\boldsymbol{x}}$ denotes the multinomial coefficient.

We note that when $s=1$ the pf (2.1) reduces to the univariate Polya distribution (see [4]). This is also evident from the urn scheme. For, in the case when $s=1$ there are only "white" and "black" balls in the urn and the scheme of sampling would reduce to Eggenberger-Polya sampling scheme for univariable case.

(b) In this paragraph, we give a different chance mechanism which generates the MMP distribution. Put $\boldsymbol{x} = (x_1, x_2, \cdots, x_s)$ and $x = \sum x_i$. If the conditional distribution of $\boldsymbol{x}$ given x is singular multinomial (SM) distribution (see [4]) with parameters x and $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_s)$ and the distribution of x is univariate Polya distribution with parameters n, ν and p_0, then the distribution of $(\boldsymbol{x}, x)$ is MMP distribution as in (2.1).

Remark. It may be noted, as the referee pointed out, that (1) marginal distribution of the component x of $(\boldsymbol{x}, x)$ follows a univariate Polya distribution and that (2) the sum $x_1 + x_2 + \cdots + x_s + x = 2(x_1 + x_2 + \cdots + x_s)$ takes even numbers with probability one. For,

$$\begin{aligned} \mathrm{P}\,(\boldsymbol{x}, x) &= \mathrm{P}\,(\boldsymbol{x}|x)\,\mathrm{P}\,(x) \\ &= \left(x! \Big/ \prod_{i=1}^{s} x_i!\right) \prod_{i=1}^{s} \phi_i^{x_i} \cdot \binom{-q_0/\nu}{x}\binom{-p_0/\nu}{n-x} \Big/ \binom{-1/\nu}{n} \end{aligned} \tag{2.2}$$

where $q_0 = 1 - p_0$. The pf (2.2) can be easily shown to be equivalent to the pf (2.1), by taking $\phi_i = p_i/(1-p_0)$, $p_i = N_i/N$ and $\nu = d/N$.

(c) Now in the above urn sampling scheme of (a) suppose we draw balls until we observe exactly k white balls following the same procedure. Let x_i denote the number of balls of ith color drawn in the process of getting k white balls, then it is easy to show that the pf of the *rv* $\boldsymbol{x}$ is

$$\text{mmip}\,(\boldsymbol{x}; d, k, N) = \binom{k+x-1}{\boldsymbol{x}}\left(\frac{N_0^{(k,d)}M^{(x,d)}}{N^{(k+x,d)}}\right)\prod_{i=1}^{s}\left(\frac{N_i}{M}\right)^{x_i} \tag{2.3}$$

where $x_i = 0, 1, 2, \cdots, \infty$ for $i = 1, 2, \cdots, s$; $x = \sum_{i=1}^{s} x_i$, $M = \sum_{i=1}^{s} N_i$ and $N = N_0 + M$.

We note when $s=1$ the pf (2.3) reduces to the pf of univariate inverse Polya distribution (see [4]).

(d) Here we give a different chance mechanism which generates the MMIP distribution. Again, as in (b), put $\boldsymbol{x} = (x_1, x_2, \cdots, x_s)$ and

$x=\sum x_i$. If the conditional distribution of $\boldsymbol{x}$ given x is SM distribution and the distribution of x is univariate inverse Polya distribution with parameters k, ν and p_0, then the distribution of $(\boldsymbol{x}, x)$ is MMIP as in (2.3). A similar remark as in (b) applies here also. For,

$$(2.4)\quad \mathrm{P}(\boldsymbol{x}, x)=\mathrm{P}(\boldsymbol{x}|x)\,\mathrm{P}(x)$$
$$=\left(x!\Big/\prod_{i=1}^{s} x_i!\right)\prod_{i=1}^{s}\phi_i^{x_i}\cdot\left(\frac{k}{k+x}\right)\binom{-p_0/\nu}{k}\binom{-q_0/\nu}{x}\Big/\binom{-1/\nu}{k+x}$$

where $q_0=1-p_0$. The pf (2.4) can be easily seen to be equivalent to the pf (2.3) after little algebra with $\phi_i=p_i/(1-p_0)$, $p_i=N_i/N$ and $\nu=d/N$.

3. Moments of the distributions

The probability generating functions (pgfs) of MMP and MMIP distributions are respectively,

$$(3.1)\qquad G_{\boldsymbol{x}}(\boldsymbol{t})=C_1\,{}_2F_1(-n;\,q_0\nu^{-1},\,-n+1;\,Z)$$

and

$$(3.2)\qquad G_{\boldsymbol{x}}(\boldsymbol{t})=C_2\,{}_2F_1(k;\,q_0\nu^{-1},\,\nu^{-1}+k;\,Z)$$

where

$$C_1=\frac{(-p_0/\nu)^{(n)}}{(-1/\nu)^{(n)}},\qquad C_2=\frac{(-p_0/\nu)^{(k)}}{(-1/\nu)^{(k)}}$$

$$Z=\phi_1t_1+\phi_2t_2+\cdots+\phi_st_s;\qquad a^{(b)}=a(a-1)\cdots(a-b+1)$$

and ϕ_i, p_0 and ν are as defined earlier. For the definition of ${}_2F_1$ see [1].

Remark. It may be noted that the pgfs of MMP and MMIP are equivalent to the pgfs of univariate Polya (with parameters n, ν and p_0), and univariate inverse Polya (with parameters k, ν and p_0), evaluated at $Z=\phi_1t_1+\phi_2t_2+\cdots+\phi_st_s$.

The factorial moment generating functions of MMP and MMIP, obtained by substituting $t_i=1+\alpha_i$ $(i=1, 2, \cdots, s)$ in (3.1) and (3.2) are clearly,

$$(3.3)\qquad M_{\boldsymbol{x}}(\boldsymbol{\alpha})=C_1\,{}_2F_1(-n;\,q_0\nu^{-1},\,-p_0\nu^{-1}-n+1;\,W)$$

and

$$(3.4)\qquad M_{\boldsymbol{x}}(\boldsymbol{\alpha})=C_2\,{}_2F_1(k;\,q_0\nu^{-1},\,\nu^{-1}+k;\,W)$$

where C_1 and C_2 are as given above and $W=1+\phi_1\alpha_1+\phi_2\alpha_2+\cdots+\phi_s\alpha_s$.

Hence the factorial moments $\mu'[r_1, r_2, \cdots, r_s]$ of MMP and MMIP are respectively given by,

$$(n^{(r)}M^{(r,d)})\prod_{i=1}^{s}\left(\frac{N_i}{M}\right)^{r_i} \quad \text{and} \quad \left(\frac{(k)_r M^{(r,d)}}{(N_0-rd)^{(r,d)}}\right)\prod_{i=1}^{s}\left(\frac{N_i}{M}\right)^{r_i}$$

where $r=r_1+r_2+\cdots+r_s$ and $(k)_r=k(k+1)(k+2)\cdots(k+r-1)$.

4. Structural properties of MMP and MMIP

The MMP and MMIP distributions possess several remarkable and interesting properties which are not similar to the properties of multivariate Polya and inverse Polya distributions. The more important of these properties refer to any group $r<s$ variables out of the s considered. These properties, which can be proved using the pgf techniques or otherwise, are stated as theorems without proofs in this section.

4.1. *Properties of MMP*

If the *rv* $\boldsymbol{x}$ has MMP as in (2.1), then

THEOREM 1. *The sum of the components of $\boldsymbol{x}$ follows univariate Polya distribution with parameters n, d and M.*

THEOREM 2. *The conditional distribution of $\boldsymbol{x}$ given $\sum_{i=1}^{s} x_i=m$ follows SM distribution with parameters m and $(\phi_1, \phi_2, \cdots, \phi_s)$ where $\phi_i=p_i/(1-p_0)$ and $p_i=N_i/N$.*

THEOREM 3. *The rv $(Z_1, Z_2, \cdots, Z_r)$ also follows MMP with parameters n, d and $\boldsymbol{M}$, where $Z_j=\Sigma_j x_i$, $M_j=\Sigma_j N_i$ with Σ_j standing for the summation taken over $i \in G_j$ for $j=1, 2, \cdots, s$ such that $\bigcup_{j=1}^{r} G_j=\{1, 2, \cdots, s\}$ and $G_j \cap G_{j1}=\phi$* for $j\neq j^1=1, 2, \cdots, r$.

THEOREM 4. *The conditional distribution of the rv $\boldsymbol{x}$ given the rv $(Z_1, Z_2, \cdots, Z_r)$ is the product of r independent SM distributions of the type $SM(\{p_{ij}\}, Z_j)$ where $P_{ij}=N_i/M_j$ for $i \in G_j$ $(j=1, 2, \cdots, r)$ and Z_j and G_j are as defined in the above theorem.*

THEOREM 5. *The marginal distribution of $\boldsymbol{x}_1=(x_1, x_2, \cdots, x_r)$ where $1\leqq r\leqq s$ is available as*

$$\binom{n}{\boldsymbol{x}_1}\left(\frac{N_0^{(n-y,d)}M^{(y,d)}}{N^{(n,d)}}\right)\prod_{i=1}^{s}\left(\frac{N_i}{M}\right)^{x_i}$$
$$\cdot {}_2F_1\left(-n+y;\frac{M+yd}{d},\frac{-N_0}{d}+n-y-1;\sum_{i=r+1}^{s}\frac{N_i}{M}\right)$$

where $y=\sum_{i=1}^{r} x_i$ *and* ${}_2F_1$ *is Gaussian hypergeometric function* (*see* [1]).

THEOREM 6. *The conditional distribution of* $\boldsymbol{x}_2$ *given* $\boldsymbol{x}_1$, *where* $\boldsymbol{x}_2=(x_{r+1}, x_{r+2}, \cdots, x_s)$ *is given by*

$$\binom{n-y}{\boldsymbol{x}_2}\left(\frac{N_0^{(n-x,d)}M^{(x,d)}}{N_0^{(n-y,d)}M^{(y,d)}{}_2F_1(\cdots)}\right)\prod_{i=r+1}^{s}\left(\frac{N_i}{M}\right)^{x_i}$$

where ${}_2F_1(\cdots)$ *is as in Theorem* 5.

4.2. *Properties of MMIP*

Let the *rv* $\boldsymbol{x}$ have MMIP as in (2.3), then

THEOREM 7. *The sum* $\sum_{i=1}^{s} x_i$ *has the univariate inverse Polya distribution with parameters* k, d *and* M.

THEOREM 8. *The conditional distribution of* $\boldsymbol{x}$ *given* $\sum_{i=1}^{s} x_i=m$ *follows SM distribution with parameters* m *and* $\phi_1, \phi_2, \cdots, \phi_s$.

THEOREM 9. *The distribution of the rv* $(Z_1, Z_2, \cdots, Z_r)$ *is also MMIP with parameters* k, d *and* $\boldsymbol{M}$, *where* Z_j, M_j $(j=1, 2, \cdots, r)$ *are as defined in Theorem* 4.

THEOREM 10. *The conditional distribution of* $\boldsymbol{x}$ *given* $(Z_1, Z_2, \cdots, Z_r)$ *is the product of* r *independent SM distributions of the type*, $SM(\{P_{ij}\}, Z_j)$ *where* $P_{ij}=N_i/M_j$ *for* $i \in G_j$.

THEOREM 11. *The marginal distribution of* $\boldsymbol{x}_1=(x_1, x_2, \cdots, x_r)$ *is available as*

$$\binom{k+y-1}{x_1, x_2, \cdots, x_r}\left(\frac{N_0^{(k,d)}M^{(y,d)}}{N^{(k+y,d)}}\right)\prod_{i=1}^{r}\left(\frac{N_i}{M}\right)^{x_i}$$
$$\cdot {}_2F_1\left(k+y;\frac{M+yd}{d}, \frac{N}{d}+k+y;\sum_{r+1}^{s}\frac{N_i}{M}\right)$$

where $y=\sum_{i=1}^{r} x_i$.

THEOREM 12. *The conditional distribution of* $\boldsymbol{x}_2$ *given* $\boldsymbol{x}_1$ *is given by*

$$\binom{k^*+z-1}{x_{r+1}, \cdots, x_s}\left(\frac{(M+yd)^{(Z,d)}}{(N+k^*d)^{(Z,d)}{}_2F_1(\cdots)}\right)\prod_{r+1}^{s}\left(\frac{N_i}{M}\right)^{x_i}$$

where $k^*=k+y$, $\boldsymbol{x}_2=(x_{r+1}, \cdots, x_s)$ *and* ${}_2F_1(\cdots)$ *is as in Theorem* 11.

Acknowledgements

The authors wish to thank the referee for his suggestions which greatly improved this paper. In particular, the referee's comments led to results (b) and (d) of Section 2.

SANGAMON STATE UNIVERSITY
PENNSYLVANIA STATE UNIVERSITY

REFERENCES

[1] Erdelyi, A. (1953). *Higher Transcendental Functions* 1, McGraw-Hill, New York.

[2] Janardan, K. G. and Patil, G. P. (1970). On the multivariate Polya distribution: a model of contagion for data with multiple counts, *Random Counts in Scientific Work*, **3**, 143-161, Ed. G. P. Patil.

[3] Janardan, K. G. and Patil, G. P. (1972). The multivariate inverse Polya distribution: a model of contagion for data with multiple counts in inverse sampling, *Studi di probalita, statistica e Ricerca Operative in Onore di G. Pompilj—Oderisi, Gubbie*, 1-15.

[4] Patil, G. P. and Joshi, S. W. (1968). *A dictionary and bibliography of discrete distributions*, 286, Oliver and Boyd, Edinburgh.

[5] Sibuya, Masaaki, Yoshimura, Isao and Shimizu, Ryoichi (1964). Negative multinomial distribution, *Ann. Inst. Statist. Math.*, **16**, 409-426.