APPLICATIONES MATHEMATICAE
25,4 (1999), pp. 457–461

I. KOPOCIŃSKA (Wrocław)

BIVARIATE NEGATIVE BINOMIAL DISTRIBUTION OF THE MARSHALL–OLKIN TYPE

Abstract. The bivariate negative binomial distribution is introduced using the Marshall–Olkin type bivariate geometrical distribution. It is used to the estimation of the distribution of the number of accidents in standard data.

1. Introduction. The bivariate negative binomial (BVNB) distribution was used in insurance theory to the description of the number of accidents in transportation. The Bates and Neyman [1] version of BVNB is based on Poisson independent random variables with mixed parameter. The Edwards and Gurland [2] version of BVNB is based on a compound correlated bivariate Poisson distribution function. In both cases the mixing is a univariate gamma random variable. Kopociński [3] introduced BVNB distributions using Poisson independent random variables mixed by a Marshall–Olkin bivariate exponential distribution [4]. In this paper we introduce BVNB distributions using bivariate geometrical (BVG) distributions of Marshall–Olkin type and their convolutions.

2. Bivariate geometrical distribution. Let U, V, W be independent random variables geometrically distributed with parameters p_1, p_2, p_3, respectively. Recall that a distribution function $\{p_n\}$ is *geometrical* with parameter p if $p_n = (1-p)p^n$, $n = 0, 1, \ldots$ We say that X, Y have a *BVG distribution* if $X = \min(U, W)$, $Y = \min(V, W)$. Let

$$p_{m,n} = P(X = m,\ Y = n),$$
$$P_{m,n} = P(X \ge m,\ Y \ge n) = \sum_{i=m}^{\infty}\sum_{j=n}^{\infty} p_{i,j},$$

where $m, n = 0, 1, \ldots$

1991 *Mathematics Subject Classification*: Primary 60E05; Secondary 60E10.
Key words and phrases: bivariate negative geometrical, negative binomial distribution.

It is easy to see that

$$p_{m,m} = P(U \ge m,\ V \ge m,\ W = m) + P(U = m,\ V = m,\ W \ge m+1),$$

$$p_{m,n} = \begin{cases} P(U \ge m,\ V = n,\ W = m) \\ \qquad\qquad + P(U = m,\ V = n,\ W \ge m+1) & \text{for } m > n, \\ P(U = m,\ V \ge n,\ W = n) \\ \qquad\qquad + P(U = m,\ V = n,\ W \ge n+1) & \text{for } m < n, \end{cases}$$

and also

$$P_{m,n} = p_1^m p_2^n p_3^{\max(m,n)}.$$

THEOREM 1. *The generating function* (*gf*) *of* $\{P_{m,n}\}$ *is*

$$\Phi(u,v) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P_{m,n} u^m v^n = \frac{1}{1-\gamma uv}\left[1 + \frac{\alpha u}{1-\alpha u} + \frac{\beta v}{1-\beta v}\right], \tag{1}$$

and the gf of $\{p_{m,n}\}$ *is*

$$\phi(u,v) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p_{m,n} u^m v^n = \frac{1}{1-\gamma uv}\left[A + \frac{B\alpha u}{1-\alpha u} + \frac{C\beta v}{1-\beta v}\right], \tag{2}$$

where $\alpha = p_1p_3$, $\beta = p_2p_3$, $\gamma = p_1p_2p_3$, $A = 1-\alpha-\beta+\gamma$, $B = (1-p_2)(1-\alpha)$, $C = (1-p_1)(1-\beta)$.

The proof is omitted.

COROLLARY 1. *The gfs of the boundary distribution functions are*

$$\phi(u,1) = \frac{1-\alpha}{1-\alpha u}, \qquad \phi(1,v) = \frac{1-\beta}{1-\beta v},$$

i.e. they are geometrical with parameters α *and* β, *respectively.*

COROLLARY 2. *The zero cell probability of* X, Y *is*

$$P(X=0,\ Y=0) = 1-\alpha-\beta+\gamma. \tag{3}$$

COROLLARY 3. *The covariance of* X, Y *is positive*:

$$\mathrm{Cov}(X,Y) = \frac{\gamma(1-p_3)}{(1-\gamma)(1-\alpha)(1-\beta)}. \tag{4}$$

For the proof of (4) we have

$$\Phi(1,1) = \frac{1}{1-\gamma}\left[1 + \frac{\alpha}{1-\alpha} + \frac{\beta}{1-\beta}\right].$$

Because

$$\begin{aligned}\Phi(1,1) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P_{m,n} &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} (m+1)(n+1)p_{m,n} \\ &= \mathrm{E}(XY) + \mathrm{E}(X) + \mathrm{E}(Y) + 1,\end{aligned}$$

we have

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)\\ &= \Phi(1,1)-\frac{\alpha}{1-\alpha}-\frac{\beta}{1-\beta}-\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}-1.\end{aligned}$$

Simple calculations lead to (4). ■

3. BVNB distribution functions. We say that random variables X, Y have a *BVNB distribution* with parameters p_1, p_2, p_3, r if their gf is

$$\phi_r(u,v)=\phi^r(u,v), \qquad r>0. \tag{5}$$

PROPOSITION 1. *The boundary distribution functions are BVNB distributions with parameters* α, r *and* β, r, *respectively.*

PROPOSITION 2. *For the distribution function* (5) *we have*

$$P(X=0,\ Y=0)=A^r, \qquad \mathrm{Cov}(X,Y)=\frac{r\gamma(1-p_3)}{(1-\gamma)(1-\alpha)(1-\beta)}.$$

THEOREM 2. *The distribution function* (5) *is the convolution of the bivariate sequences* $\{a_{m,n}\}$ *and* $\{b_{m,n}\}$:

$$P(X=m,\ Y=n)=p^{(r)}_{m,n}=\sum_{k=0}^{m}\sum_{l=0}^{n}a_{m-k,n-l}b_{k,l},$$

where

$$\begin{aligned} a_{m,n} &= \begin{cases}\binom{r+m-1}{m}\gamma^m, & m=n=0,1,\ldots,\\ 0, & m\neq n.\end{cases}\\ b_{k,l} &= A^r(\alpha)^k(\beta)^l\\ &\quad\times\sum_{i=0}^{k}\sum_{j=0}^{l}\binom{r}{i}\binom{r-i}{j}\binom{k-1}{k-i}\binom{l-1}{l-j}(B/A)^i(C/A)^j.\end{aligned} \tag{6}$$

P r o o f. The gf of $\{a_{m,n}\}$ is $(1-\gamma uv)^{-r}$. We have

$$\begin{aligned}&\left(1+\frac{(B/A)\alpha u}{1-\alpha u}+\frac{(C/A)\beta v}{1-\beta v}\right)^r\\ &\qquad=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{r}{i}\binom{r-i}{j}\left(\frac{(B/A)\alpha u}{1-\alpha u}\right)^i\left(\frac{(C/A)\beta v}{1-\beta v}\right)^j\end{aligned}$$

$$= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{r}{i}\binom{r-i}{j}((B/A)\alpha u)^i((C/A)\beta v)^j$$

$$\times \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\binom{i+m-1}{m}(\alpha u)^m\binom{j+n-1}{n}(\beta v)^n$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(\alpha u)^k(\beta v)^l\sum_{i=0}^{k}\sum_{j=0}^{l}\binom{r}{i}\binom{r-i}{j}\binom{k-1}{k-i}\binom{l-1}{l-j}(B/A)^i(C/A)^j$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}c_{k,l}u^kv^l.$$

From the above and (3) we have $b_{k,l} = A^r c_{k,l}$, which proves (6). ■

4. Application. The Edwards–Gurland BVNB distribution was used in [2], [5] to the estimation of the distribution of the number of accidents sustained by London omnibuses in two consecutive time intervals. We use these data to the illustration of the distribution (5). Similarly to [2] we assume $r = 5$ and we estimate the other parameters of the distribution by the minimum chi-square method and obtain $p_1 = 0.519$, $p_2 = 0.502$, $p_3 = 0.473$. The details of the parameter estimation are omitted. For the estimation of parameters of the boundary distributions, correlation or the Zero-Zero Cell Frequency Method may be used. The observed and expected numbers of accidents for 166 London omnibuses are presented in Table 1.

TABLE 1. Observed and expected numbers of motor vehicle accidents among 166 London omnibus drivers during two time intervals (data from [2])

N. of a.	Number of accidents, time interval 1											
t. i. 2	0		1		2		3		≥ 4		Total	
0	15	17.87	15	12.15	4	6.19	2	2.70	1	1.68	37	40.59
1	17	12.87	18	18.01	9	10.58	3	5.02	5	3.35	52	49.82
2	4	6.87	16	11.16	12	9.71	6	5.16	5	3.80	43	36.69
3	2	3.13	6	5.54	5	5.43	2	3.76	4	3.15	19	21.02
≥ 4	2	2.07	4	3.95	5	4.29	0	3.41	4	4.14	15	17.87
Total	40	42.80	59	50.81	35	36.20	13	20.05	19	16.12	166	165.99

References

[1] G. E. Bates and J. Neyman, *Contribution to the theory of accident proneness, I. An optimistic model of the correlation between light and severe accidents*, Univ. of California Publ. Statist. 1 (1952), 215–254.

[2] C. B. Edwards and J. Gurland, *A class of distributions applicable to accidents*, J. Amer. Statist. Assoc. 56 (1961), 503–517.

[3] B. Kopociński, *Bivariate negative binomial distribution based on a bivariate exponential distribution function*, to appear.

[4] A. W. Marshall and I. Olkin, *A multivariate exponential distribution*, J. Amer. Statist. Assoc. 62 (1967), 30–44.

[5] K. Subrahmaniam and K. Subrahmaniam, *On the estimation of the parameters in the bivariate negative binomial distribution*, J. Roy. Statist. Assoc. 35 (1973), 131–146.

Ilona Kopocińska
Mathematical Institute
University of Wrocław
Pl. Grunwaldzki 2/4
50-384 Wrocław, Poland
E-mail: ibk@math.uni.wroc.pl

Received on 21.4.1998