ON A BIVRIATE COMPOUND POISSON MODEL

Arun K. Sinha

Department of Statistics, Patna University,
Patna 800005, INDIA
[arunkrsinha@yahoo.com]

The Model

We presume that the births in human populations follow the Poisson distribution with λ as a fixed birthrate. The proportions of male and female offspring are denoted by p and q respectively. As the two types of births occur independently we have

$$P(X=x|\lambda)=\frac{e^{-\lambda p}(\lambda p)^x}{\Gamma(x+1)},\ x=0,1,2,\ldots$$

and

$$P(Y=y|\lambda)=\frac{e^{-\lambda q}(\lambda q)^y}{\Gamma(y+1)},\ y=0,1,2,\ldots,p+q=1.$$

The joint probability mass function of the births is given by

$$P(X=x, Y=y \mid \lambda) = \frac{e^{-\lambda}\lambda^{x+y}p^{x}q^{y}}{\Gamma(x+1)\,\Gamma(y+1)}. \tag{1}$$

It is, further, assumed that λ follows the Pearson's type III distribution as given below

$$f(\lambda) = \frac{(k/\mu)^{k}\,\lambda^{k-1}e^{-k\lambda/\mu}}{\Gamma(k)}. \tag{2}$$

Here μ and k denote the average number of births per mother and a constant respectively.

Case (I). BIVARIATE COMPOUND POISSON DISTRIBUTION

These assumptions lead us to the probability mass function of the bivariate compound Poisson distribution (also known as a bivariate negative binomial distribution), which is mentioned below:

$$P(X=x, Y=y) = \frac{\Gamma(k+x+y)\,[k/(k+\mu)]^{k}\,[\mu p/(k+\mu)]^{x}\,[\mu q/(k+\mu)]^{y}}{\Gamma(k)\,\Gamma(x+1)\,\Gamma(y+1)}$$

where $x, y = 0, 1, 2, \ldots.$ (3)

This appears in a reparameterized form of the bivariate compound Poisson distribution referred to by Arbous and Kerrich (1951).

The marginal distributions of (3) are obtained as

$$P(X=x)=\frac{\Gamma(k+x)}{\Gamma(k)\,\Gamma(x+1)}\left(\frac{k}{k+\mu p}\right)^{k}\left(\frac{\mu p}{k+\mu p}\right)^{x},\quad x=0,1,2,\ldots$$

and

$$P(Y=y)=\frac{\Gamma(k+y)}{\Gamma(k)\,\Gamma(y+1)}\left(\frac{k}{k+\mu q}\right)^{k}\left(\frac{\mu q}{k+\mu q}\right)^{x},\quad y=0,1,2,\ldots$$

The conditional distribution of Y given X is given below:

$$P(Y=y|X=x)=\frac{\Gamma(k+x+y)}{\Gamma(1+y)\,\Gamma(k+x)}\left(\frac{\mu q}{k+\mu}\right)^{y}\left(1-\frac{\mu q}{k+\mu}\right)^{k+x}\qquad (4)$$

This yields that

$$E(Y|X=x)=\left(\frac{\mu q}{k+\mu p}\right)(k+x)\qquad (5)$$

and

$$\operatorname{var}(Y|X=x)=\left[\frac{\mu q\,(k+x)}{(k+\mu p)^2}\right](k+\mu) \qquad (6)$$

Similarly, using $\mathrm{P}(\mathrm{X}=\mathrm{x}|\mathrm{Y}=\mathrm{y})$ we derive

$$E(X|Y=y)=\left(\frac{\mu p}{k+\mu q}\right)(k+y) \qquad (7)$$

and

$$\operatorname{var}(X|Y=y)=\left[\frac{\mu p(k+y)}{(k+\mu q)^2}\right](k+\mu). \qquad (8)$$

The Operating Characteristic for "Birth Proneness"

This may be defined as the probability of a mother with a known number of male offspring (x) to have α or more female offspring (Y). This is given below:

$$P(Y \geq \alpha | X = x) = 1 - \frac{1}{\Gamma(x+k)}\left(1 - \frac{\mu q}{k+\mu}\right)^{k+x} \sum_{y=0}^{\alpha-1} \frac{\Gamma(k+x+y)}{\Gamma(1+y)}\left(\frac{\mu q}{k+\mu}\right)^{y}. \quad (9)$$

Likewise, we define $\mathrm{P}(\mathrm{X} \geq \beta | \mathrm{Y} = \mathrm{y})$ where β denotes the given number of male offspring.

The formula for the product moment correlation coefficient in terms of the parameters of the model (3) is given by

$$\rho_{XY} = \frac{(pq)^{1/2}\mu}{(k+\mu p)^{1/2}(k+\mu q)^{1/2}}. \quad (10)$$

This is given by Rao, Mazumdar, Waller and Li (1973).

(1) Applications of the Model

To examine the adequacy of model (3) we consider the bivariate distribution of births occurred in France, which is referred to by James (1975).

The univariate negative binomial distribution (NBD) is fitted to the distribution of births of male and female offspring taken together. The adequacy of the model examined by the chi-square test indicates that the model provides a good fit to the observed distribution.

The maximum likelihood method used for the estimation of the parameters yields the following estimates:

$\hat{\mu}=1.3916$, $\hat{k}=1.18764$, $\hat{p}=0.5164$ and $\hat{q}=0.48356$.

Note that the proportion of male is significantly different from 0.5 at 5% level of significance.

The equation (13) provides the value of the product moment correlation coefficient as 0.36919.

It is included in the 95% confidence limits for the Pearson's correlation coefficient (0.36063), which are computed as 0.33289 (lower) and 0.38774 (upper).

The model (3) describes the bivariate distribution fairly well as judged by the chi-square test. The observed and the expected frequencies are given in Table 1.

Further, we obtain the conditional expected number of male and female offspring on the basis of (6) and (8), which are given below:

$$E\left(X \mid Y = y\right) = 0.45877 + 0.38628\, y$$

and

$$E\left(Y \mid X = x\right) = 0.41925 + 0.35301\, x$$

The results so obtained are given in Tables 2 and 3. The values compare reasonably well with the observed ones.

Next, we find P(Y ≥ α|X = x) and P(X ≥ β|Y = y), which are given below.

$$P(X \geq \beta | Y = y) = 1 - \frac{(0.72133)^{1.18764+y}}{\Gamma(1.18764+y)} \sum_{x=0}^{\beta-1} (0.27865)^{x} \frac{\Gamma(x+y+1.18764)}{\Gamma(x+1)}$$

and

$$P(Y \geq \alpha | X = x) = 1 - \frac{(0.73909)^{1.18764+x}}{\Gamma(1.18764+x)} \sum_{y=0}^{\alpha-1} (0.26091)^{y} \frac{\Gamma(x+y+1.18764)}{\Gamma(y+1)}.$$

Tables 4 and 5 present the probabilities, the observed and the expected number of offspring.

The resemblance between the two is very good.

Table 1. The observed and the expected (inside parentheses) number of mothers having x sons and y daughters

X / Y	**0**	**1**	**2**	**3**	**4**	**5**	**6+**	**Total**
0	1499 (1539.81)	552 (509.58)	163 (155.32)	45 (45.99)	13 (13.41)	4 (3.88)	2 (1.56)	2278 (2269.55)
1	506 (477.14)	290 (290.85)	114 (129.22)	43 (50.24)	16 (18.16)	7 (6.26)	3 (3.09)	979 (974.96)
2	142 (136.17)	107 (120.94)	60 (70.56)	30 (34.01)	15 (14.66)	7 (5.87)	3 (3.48)	364 (385.69)
3	37 (37.75)	38 (44.05)	29 (31.83)	19 (18.30)	11 (9.16)	6 (4.18)	3 (2.95)	143 (148.22)
4	10 (10.31)	14 (14.90)	13 (12.85)	11 (8.58)	8 (4.89)	5 (2.51)	3 (2.11)	64 (56.15)
5	3 (2.79)	6 (4.81)	6 (4.82)	6 (3.66)	5 (2.34)	3 (1.33)	0 (1.30)	29 (21.05)
6+	1 (1.01)	2 (2.15)	3 (2.59)	3 (2.33)	2 (1.74)	0 (1.15)	0 (1.41)	11 (12.38)
Total	2198 (2204.98)	1009 (987.28)	388 (407.19)	157 (163.11)	70 (64.36)	32 (25.18)	14 (14.49)	3868 (3868.00)

Table 2: The observed ($\bar{x}$) and expected ($E(X|Y=y)$) number of sons to a mother having Y daughters

y	$\bar{x}$	$E(X\|Y=y)$
0	0.482	0.459
1	0.780	0.845
2	1.181	1.231
3	1.713	1.618
4	2.312	2.004
5	2.448	2.390
6	2.273	2.776

Table 3: The observed ($\bar{y}$) and the expected ($E(Y|X=x)$), number of daughters to a mother having X sons

x	$\bar{y}$	$E(Y\|X=x)$
0	0.438	0.419
1	0.710	0.772
2	1.085	1.125
3	1.605	1.478
4	2.114	1.831
5	2.312	2.184
6	2.143	2.537

Table 4: The observed and the expected number of mothers to have β (β = 1, 2, ... , 6) or more sons when they have Y (y = 0, 1, 2, ..., 6) daughters

Y	0	1	2	3	4	5	6
$P(X \geq 1 \mid Y)$	0.32156	0.51062	0.64699	0.74537	0.81632	0.86751	0.90443
Exp. No.	732.50	499.89	235.50	106.59	52.54	25.16	9.95
Obs. No.	779	473	222	106	54	26	10
$P(X \geq 2 \mid Y)$	0.09703	0.21230	0.33344	0.44824	0.55082	0.63907	0.71302
Exp. No.	221.04	207.84	112.37	64.10	33.25	18.53	7.84
Obs. No.	227	183	115	68	40	20	8
$P(X \geq 3 \mid Y)$	0.02860	0.07981	0.15050	0.23348	0.32192	0.41031	0.49467
Exp. No.	65.15	78.13	54.78	33.39	20.60	11.90	5.44
Obs. No.	64	69	55	39	27	14	5
$P(X \geq 4 \mid Y)$	0.00834	0.02827	0.06235	0.11006	0.16911	0.23634	0.30834
Exp. No.	19.00	27.68	22.70	15.74	10.82	6.85	3.39
Obs. No.	19	26	25	20	16	8	2
$P(X \geq 5 \mid Y)$	0.00243	0.00942	0.02436	0.04826	0.08195	0.12499	0.17610
Exp. No.	5.53	9.22	8.87	6.90	5.24	3.62	1.94
Obs. No.	6	10	10	9	8	3	0
$P(X \geq 6 \mid Y)$	0.00072	0.00323	0.00914	0.02006	0.03732	0.06177	0.09365
Exp. No.	1.64	3.16	3.32	2.87	2.39	1.79	0.84
Obs. No.	2	3	3	3	3	0	0

Table 5: The observed and the expected number of mothers to have α (α = 1, 2, ..., 6) or more daughters when they have X (x = 0, 1, 2, ..., 6) son

X	0	1	2	3	4	5	6
$P(Y \geq 1\|X)$ Exp. No. Obs. No.	0.30167 663.07 699	0.48387 448.23 457	0.61853 239.99 225	0.71806 112.74 112	0.79162 55.41 57	0.84599 27.07 28	0.88617 12.41 12
$P(Y \geq 2\|X)$ Exp. No. Obs. No.	0.08528 187.45 193	0.18928 190.28 167	0.30127 116.89 111	0.41002 64.37 69	0.50958 35.67 41	0.59736 19.11 21	0.67271 9.42 9
$P(Y \geq 3\|X)$ Exp. No. Obs. No.	0.02353 51.71 51	0.06677 67.37 60	0.12796 49.65 51	0.20155 31.64 39	0.28192 19.73 26	0.36422 11.66 14	0.44471 6.22 6
$P(Y \geq 4\|X)$ Exp. No. Obs. No.	0.00641 14.08 14	0.02216 22.36 22	0.04976 19.31 22	0.08936 14.03 20	0.13960 9.77 15	0.19821 6.34 8	0.26252 3.67 3
$P(Y \geq 5\|X)$ Exp. No. Obs. No.	0.00173 3.80 4	0.00706 7.12 8	0.01820 7.06 9	0.03677 5.77 9	0.06360 4.45 7	0.09872 3.16 3	0.14146 1.98 0
$P(Y \geq 6\|X)$ Exp. No. Obs. No.	0.00080 1.76 1	0.00218 2.21 2	0.00636 2.47 3	0.01430 2.24 3	0.02716 1.90 2	0.04583 1.47 0	0.07078 0.99 0

See Sinha, Rai and Kumar (2003) for more details.

Case (II). THE ZERO-TRUNCATED BIVARIATE COMPOUND POISSON DISTRIBUTION

In many situations the frequency of the zeroth cell of a bivariate distribution is either not available or difficult to record. For such cases we need to have the zero-truncated bivariate compound Poisson distribution with unequal proportion of male and female births. It is defined below:

$$P(X=x,Y=y)=\frac{\Gamma(k+x+y)\left[k/(k+\mu)\right]^{k}\left[\mu p/(k+\mu)\right]^{x}\left[\mu q/(k+\mu)\right]^{y}}{(1-P_{00})\,\Gamma(k)\,\Gamma(x+1)\,\Gamma(y+1)} \tag{11}$$

where $P_{00}=\left(\frac{k}{k+\mu}\right)^{k}$ and $x+y=1, 2, 3, \ldots$ and $x, y = 0, 1, 2, \ldots$

The marginal probability mass function of X is obtained as follows:

$$P(X=0)=\left(\frac{1}{1-P_{00}}\right)\left[\left(\frac{k}{k+\mu\, p}\right)^{k}-P_{00}\right] \qquad (12)$$

and

$$P(X=x)=\frac{\Gamma(k+x)}{(1-P_{00})\Gamma(k)\Gamma(x+1)}\left(\frac{k}{k+\mu p}\right)^{k}\left(\frac{\mu p}{k+\mu p}\right)^{x}$$

(13)

where x = 1, 2, ….

The conditional probability mass function of Y given X is derived as follows

$$P(Y=y|X=0)=\frac{\left[\frac{\Gamma(k+y)}{\Gamma(1+y)\Gamma(k)}\right]\left(\frac{k}{k+\mu}\right)^{k}\left(\frac{\mu\, q}{k+\mu}\right)^{y}}{\left[\left(\frac{k}{k+\mu\, p}\right)^{k}-P_{00}\right]}$$

(14)

where y = 1, 2, 3, … and

$$P(Y=y|X=x)=\frac{\Gamma(k+x+y)}{\Gamma(1+y)\Gamma(k+x)}\left(\frac{\mu q}{k+\mu}\right)^{y}\left(1-\frac{\mu q}{k+\mu}\right)^{k+x} \tag{15}$$

where x = 1, 2, 3, … and y = 0, 1, 2, 3, ….

These expressions help obtain the conditional mean and variance of *Y*, which are given below:

$$E(Y|X=0)=\frac{\mu q\left(\frac{k}{k+\mu p}\right)^{k+1}}{\left[\left(\frac{k}{k+\mu p}\right)^{k}-P_{00}\right]} \tag{16}$$

$$E(Y|X=x)=\left(\frac{\mu q}{k+\mu p}\right)(k+x), \quad x = 1, 2, 3, \ldots \tag{17}$$

$$\operatorname{var}(Y|X=0)=\frac{k\mu q\{k(\mu q+1)+\mu\}}{(k+\mu p)^2\left\{1-\left(\frac{k+\mu p}{k+\mu}\right)^k\right\}}\left[1-\frac{k\mu q}{\{k(\mu q+1)+\mu\}\left\{1-\left(\frac{k+\mu p}{k+\mu}\right)^k\right\}}\right]$$

(18)

and for x = 1, 2, 3, …,

$$\operatorname{var}(Y|X=x)=\left[\frac{\mu q(k+x)}{(k+\mu p)^2}\right](k+\mu) \tag{19}$$

The Operating Characteristic for "Birth Proneness"

These are obtained by the Prob(Y ≥ α|X=0) and Prob(Y ≥ α|X=x) when x = 1, 2, ….. These are given below:

$$P(Y\geq\alpha|x=0)=\frac{1-\left(\frac{k}{k+\mu}\right)^k\frac{1}{\Gamma(k)}\sum_{y=1}^{\alpha-1}\frac{\Gamma(k+y)}{\Gamma(1+y)}\left(\frac{\mu q}{k+\mu}\right)^y}{\left[\left(\frac{k}{k+\mu\ p}\right)^k-P_{00}\right]}$$

(20)

For x = 1, 2, 3, . . . we have

$$P(Y \geq \alpha | X = x) = 1 - \frac{1}{\Gamma(x+k)}\left(1 - \frac{\mu q}{k+\mu}\right)^{k+x} \sum_{y=1}^{\alpha-1} \frac{\Gamma(k+x+y)}{\Gamma(1+y)}\left(\frac{\mu q}{k+\mu}\right)^{y} \tag{21}$$

The formula to obtain the correlation coefficient between X and Y in terms of the parameters when Z = X + Y follows the zero-truncated negative binomial distribution is given by

$$\rho_{XY} = \frac{(pq)^{1/2}\mu[A(k+1)-k]}{[A(k+\mu p)-k\mu p(1-A)]^{1/2}[A(k+\mu q)-k\mu q(1-A)]^{1/2}} \tag{22}$$

where $A = 1 - \left(\frac{k}{k+\mu}\right)^{k}$

See Hamdan (1975) for this result.

(2) Applications of the Model

The American births of offspring referred to by Rao, Mazumdar, Waller and Li (1973) is used to examine the adequacy of the model.

The estimates of the parameters are obtained by fitting the zero-truncated negative binomial distribution to the observed distribution of the total births.

The distribution of Z = X+Y is given by

$$P(Z=z)=\frac{\Gamma(k+z)}{\Gamma(k)\,\Gamma(z+1)}\left(\frac{k}{k+\mu}\right)^{k}\left(\frac{\mu}{k+\mu}\right)^{z},\quad z=0,1,2,\ldots$$

Its zero-truncated form appears as follows:

$$P(Z=z) = \frac{\Gamma(k+z)}{\left[1-(k/(k+\mu))^k\right]\Gamma(k)\Gamma(z+1)}\left(\frac{k}{k+\mu}\right)^k\left(\frac{\mu}{k+\mu}\right)^z, \quad z=1,2,\ldots$$

For fitting the model the maximum likelihood estimates of the parameters are obtained as:

$$\hat{\mu} = 3.65678 \, and \, \hat{k} = 2.64649.$$

The model provides a good fit to the observed distribution.

For the observed bivariate distribution we obtain the proportion of male offspring, i.e., an estimate of p, as 0.521 and the proportion of the female offspring, i.e., an estimate of q, as 0.479.

The proportion of male offspring is found significantly different from 0.5 for both one tailed and two tailed tests.

This, in turn, reveals that the two proportions of births cannot be considered equal.

On putting these estimates we obtain the following bivariate model:

$$P(X = x, Y = y) = \frac{(0.11184)\,\Gamma(2.64649 + x + y)(0.30225)^{x}(0.27788)^{y}}{\Gamma(2.64649)\,\Gamma(x+1)\Gamma(y+1)}.$$

Table 6. The observed and expected number of American families with X male offspring and Y female offspring

X |Y	0	1	2	3	4	5	6	7+	Total
0	-	183 (154.5)	90 (78.3)	44 (33.7)	13 (13.2)	3 (4.9)	1 (1.7)	2 (0.9)	336 (287.2)
1	211 (168.1)	206 (170.3)	101 (110.0)	61 (57.5)	17 (26.6)	6 (11.3)	6 (4.5)	7 (2.7)	615 (551.0)
2	96 (92.6)	105 (119.6)	94 (93.8)	46 (57.8)	23 (30.7)	13 (14.7)	2 (6.6)	3 (4.7)	382 (420.5)
3	43 (43.4)	60 (68.0)	67 (62.8)	37 (44.5)	18 (26.7)	11 (14.3)	9 (7.1)	6 (5.7)	251 (272.5)
4	12 (18.5)	37 (34.2)	30 (36.3)	30 (29.1)	22 (19.5)	9 (11.5)	6 (6.2)	4 (5.9)	152 (161.2)
5	6 (7.4)	19 (15.8)	18 (19.0)	10 (17.0)	12 (12.5)	6 (8.1)	4 (4.8)	1 (5.1)	76 (89.7)
6	3 (2.9)	8 (6.9)	4 (9.2)	9 (9.1)	4 (7.4)	3 (5.2)	4 (3.3)	0 (3.9)	35 (47.9)
7+	3 (1.8)	8 (4.9)	5 (7.3)	4 (8.3)	5 (7.1)	3 (8.2)	3 (4.5)	1 (6.9)	32 (49.0)
Total	374 (334.7)	626 (574.2)	409 (416.7)	241 (257.0)	114 (143.7)	54 (78.2)	37 (38.7)	34 (35.8)	1879 (1879.0)

The equation (22) gives the correlation coefficient as 0.327, but the value of the product moment correlation coefficient is 0.24173, which is significantly different from zero at 1% level of significance.

Its 99% confidence limits almost include the population correlation coefficient based on (22). This illustrates the suitability of the theory with the observed data.

Further, we get from (16) and (17):

$$E(Y|X=0) = 1.763425$$

and for x = 1,2, . . .

$$E(Y|X=x) = 0.38482(2.64649 + x).$$

Similarly, we obtain

$$E\left(X \mid Y = 0\right) = 1.866$$

and for y = 1, 2, 3, …,

$$E\left(X \mid Y = y\right) = 0.43318193\left(2.64649 + y\right).$$

Table 7. The expected number, $\mathrm{E}\left(Y \mid X\right)$ and $\mathrm{E}\left(X \mid Y\right)$ and the observed average number of sons ($\bar{x}$) and daughters ($\bar{y}$)

X or Y	$\mathrm{E}\left(Y \mid X\right)$	$\bar{y}$	$\mathrm{E}\left(X \mid Y\right)$	$\bar{x}$
0	1.76	1.73	1.89	1.75
1	1.40	1.27	1.58	1.52
2	1.79	1.63	2.01	1.87
3	2.17	2.11	2.45	2.14
4	2.56	2.62	2.88	2.86
5	2.94	2.57	3.31	3.18
6	3.33	2.80	3.75	3.62
7	3.71	3.26	4.18	2.37
8	4.10	3.50	4.61	4.00

Operating Characteristics for Birth Proneness

Equations (20) and (21) provide

$$P(Y \geq \alpha | x = 0) = 1 - \frac{0.73152}{\Gamma(2.64649)} \sum_{y=1}^{\alpha-1} \frac{\Gamma(2.64649 + y)}{\Gamma(1+y)} (0.27788)^{y}$$

and

$$P(Y \geq \alpha | X = x) = 1 - \frac{(0.72212)^{2.64649+x}}{\Gamma(2.64649 + x)} \sum_{y=1}^{\alpha-1} \frac{\Gamma(2.64649 + x + y)}{\Gamma(1+y)} (0.27788)^{y} .$$

Similarly, we can obtain

$$P(X \geq \beta | y = 0) \text{ and } \mathrm{P}(\mathrm{X} \geq \beta | \mathrm{Y} = \mathrm{y})$$

for y =1, 2, 3, ...

For more details see Sinha and Mishra (2004).

Case (III). THE SYMMETRICAL BIVARIATE COMPOUND

POISSON DISTRIBUTION (THE SYMMETRICAL BIVARIATE NEGATIVE BINOMIAL DISTRIBUTION)

In many situations the two proportions may not be significantly different and so one could safely consider that $p = q = 0.5$. In this case the bivariate compound Poisson model may be referred to as the symmetrical bivariate compound Poisson distribution.

It is more commonly called as the symmetrical bivariate negative binomial distribution (SBNBD).

Its probability mass function is given by

$$P(X=x,Y=y)=\frac{\Gamma(k+x+y)\,[k/(k+2m)]^{k}\,[m/(k+2m)]^{x+y}}{\Gamma(k)\Gamma(x+1)\Gamma(y+1)} \tag{23}$$

Here,

$$E(Y|X=x)=\left(\frac{m}{k+m}\right)(k+x) \tag{24}$$

and

$$\operatorname{var}(Y|X=x)=\left[\frac{m\,(k+2m)}{(k+m)^{2}}\right](k+x) \tag{25}$$

Similarly, using $P(X=x|Y=y)$ we derive

$$E(X|Y=y)=\left(\frac{m}{k+m}\right)(k+y)$$

(26)

and

$$\operatorname{var}(X|Y=y)=\left[\frac{m(k+2m)}{(k+m)^2}\right](k+y).$$

(27)

We obtain

$$P(Y\geq\alpha|X=x)=1-\frac{1}{\Gamma(x+k)}\left(1-\frac{m}{k+2m}\right)^{k+x}\sum_{y=0}^{\alpha-1}\frac{\Gamma(k+x+y)}{\Gamma(1+y)}\left(\frac{m}{k+2m}\right)^{y}.$$

(28)

Similarly, one could get an expression for $P(X\geq\beta|Y=y)$.

Here,

$$\rho_{XY}=\frac{m}{k+m}. \quad (29)$$

In fact, $\rho_{XY}=\dfrac{\operatorname{var}(N)-E(N)}{\operatorname{var}(N)+E(N)}$, when N = X+Y.

This is given by Rao, Mazumdar, Waller and Li (1973), and Arbous and Sichel (1954). Interestingly, Newbold (1927) has also referred to the formula.

All the above expressions follow directly from the case (I), i.e., the bivariate compound Poisson distribution by putting $p = q = 0.5$ and $\mu=2m$.
See Sinha (1985).

(3) Applications of the Model

Arbous and Sichel (1954) investigated the model to describe the phenomenon of "absence – proneness" in a group of individuals. Its applications in the field of doctor-patient consultations in general practice, one day industrial absence and accidents are studied by Froggatt (1970).

Sinha (1985) considers the model to describe the bivariate distributions with male and female offspring for Indian and American families. For the Indian families the product moment correlation turns out to be

$$r = 0.31183.$$

The SBNBD provides a good fit to the observed distributions. It is given in Table 10. The parameters are estimated by estimating the mean μ (i.e., $2m$) and k

of the univariate negative binomial distribution for the family size by the maximum likelihood method. The univariate binomial distribution provides a good fit to the distribution of the family size. Thus we obtain

$$m = 2.18766,\ \hat{k} = 3.8158,\ \rho_{XY} = 0.36440.$$

The 95 percent confidence limits for the product moment correlation coefficient (0.31183) are 0.24762 (lower) and 0.37330 (upper). This includes the value of ρ_{XY}. This illustrates the suitability of the theory with the observed data.

Next, we get $E(Y|X = x) = (1.39048 + 0.3644x)$.

Table 8: The actual ($\bar{y}$) and predicted $E(Y|X = x)$

X	$\bar{y}$	$E(Y\|X = x)$
0	1.38750	1.39048
1	1.91623	1.75488
2	2.35849	2.11928
3	2.62931	2.48368
4	3.09639	2.84808
5	2.81081	3.21248
6	3.05882	3.57688

Also, we get

$$P(y \geq \alpha | x) = 1 - (0.73292)^{x+3.81581} \frac{1}{\Gamma(x+3.81581)} \sum_{y=0}^{\alpha-1} \frac{\Gamma(x+y+3.81581)}{\Gamma(y+1)} (0.26708)^{y}$$

.

Table 9: The expected and observed number

X	$P(y \geq 1 \mid x)$	Expected number	Actual number
0	0.69445	111.112	110
1	0.77606	148.227	151
2	0.83587	132.903	136
3	0.87970	102.045	107
4	0.91183	75.682	80
5	0.93538	34.609	35
6	0.95264	32.390	31
7	0.96529	6.757	7
8	0.97456	2.924	3
9	0.98135	2.944	3
10	0.98633	0.986	1

Table10: The observed and expected frequencies of the Indian data (x: daughters, and y: sons).

y / x	0	1	2	3	4	5	6	7	8	9	10	Total
0	50 (43.04)	40 (43.87)	23 (28.21)	9 (14.61)	3 (6.65)	2 (2.78)	3 (1.09)	(0.41)	(0.15)	(0.05)	(0.03)	130 (140.89)

1	54 (43.87)	52 (56.42)	34 (43.82)	19 (26.59)	11 (13.88)	6 (6.53)	5 (2.86)	1 (1.18)	(0.47)	(0.18)	(0.10)	182 (195.90)
2	29 (28.21)	47 (43.82)	43 (39.88)	38 (27.75)	20 (16.34)	12 (8.57)	9 (4.12)	2 (1.86)	(0.80)	1 (0.33)	(0.20)	201 (171.88)
3	15 (14.61)	21 (26.59)	21 (27.75)	18 (21.78)	20 (14.28)	6 (8.25)	4 (4.34)	(2.12)	(0.98)	(0.43)	(0.30)	105 (121.43)
4	6 (6.65)	11 (13.88)	18 (16.34)	16 (14.28)	13 (10.31)	5 (6.51)	4 (3.71)	3 (1.96)	2 (0.97)	1 (0.45)	(0.35)	79 (75.41)
5	2 (2.78)	12 (6.53)	9 (8.57)	10 (8.25)	9 (6.51)	3 (4.46)	3 (2.74)	(1.55)	1 (0.82)	1 (0.41)	(0.34)	50 (42.96)
6	1 (1.09)	4 (2.86)	5 (4.12)	3 (4.34)	4 (3.71)	1 (2.74)	4 (1.81)	1 (1.09)	(0.61)	(0.32)	1 (0.30)	24 (22.99)
7	2 (0.41)	3 (1.18)	(1.86)	1 (2.12)	2 (1.96)	1 (1.55)	2 (1.09)	(0.70)	(0.42)	(0.23)	(0.22)	15 (11.74)
8	(0.15)	1 (0.47)	1 (0.80)	1 (0.98)	(0.97)	1 (0.82)	(0.61)	(0.42)	(0.26)	(0.15)	(0.17)	4 (5.80)
9	(0.05)	(0.18)	1 (0.33)	1 (0.43)	(0.45)	(0.41)	(0.32)	(0.23)	(0.15)	(0.10)	(0.11)	2 (2.76)
10+	1 (0.03)	(0.10)	(0.20)	(0.30)	1 (0.35)	(0.34)	(0.30)	(0.22)	(0.17)	(0.11)	(0.12)	2 (2.24)
Total	160 (140.89)	191 (195.90)	159 (171.88)	116 (121.43)	83 (75.41)	37 (42.96)	34 (22.99)	7 (11.74)	3 (5.80)	3 (2.76)	1 (2.24)	794 (794.00)

Case (IV). THE ZERO-TRUNCATED SYMMETRICAL BIVARIATE NEGATIVE BINOMIAL DISTRIBUTION

The probability mass function of the model is given by

$$P(X=x,Y=y)=\frac{\Gamma(k+x+y)\,[k/(k+2m)]^{k}\,[m/(k+2m)]^{x+y}}{(1-P_{00})\Gamma(k)\Gamma(x+1)\Gamma(y+1)}. \tag{30}$$

Here x+y = 1, 2, . . . but x,y = 0, 1, 2, . . . and

$$P_{00}=\left(\frac{k}{k+2m}\right)^{k}.$$

The marginal probability mass function of X is obtained as follows:

$$P(X=0)=\left(\frac{1}{1-P_{00}}\right)\left[\left(\frac{k}{k+m}\right)^{k}-P_{00}\right] \tag{31}$$

and

$$P(X=x)=\frac{\Gamma(k+x)}{(1-P_{00})\Gamma(k)\Gamma(x+1)}\left(\frac{k}{k+m}\right)^{k}\left(\frac{m}{k+m}\right)^{x} \qquad (32)$$

where x = 1, 2, …

The conditional probability mass function of *Y* given *X* is derived as follows

$$P(Y=y|X=0)=\frac{\left[\frac{\Gamma(k+y)}{\Gamma(1+y)\Gamma(k)}\right]\left(\frac{k}{k+2m}\right)^{k}\left(\frac{m}{k+2m}\right)^{y}}{\left[\left(\frac{k}{k+m}\right)^{k}-P_{00}\right]} \qquad (33)$$

where y = 1, 2, 3, … and

$$P(Y=y|X=x)=\frac{\Gamma(k+x+y)}{\Gamma(1+y)\Gamma(k+x)}\left(\frac{m}{k+2m}\right)^{y}\left(1-\frac{m}{k+2m}\right)^{k+x} \qquad (34)$$

where x = 1, 2, 3, … and y = 0, 1, 2, 3, …

These expressions help obtain the conditional mean and variance of *Y*, which are given below:

$$E\left(Y \mid X = 0\right) = \frac{m\left(\frac{k}{k+m}\right)^{k+1}}{\left[\left(\frac{k}{k+m}\right)^{k} - P_{00}\right]} \qquad (35)$$

$$E\left(Y \mid X = x\right) = \left(\frac{m}{k+m}\right)\left(k+x\right), \qquad x = 1, 2, 3, \ldots \qquad (36)$$

$$\operatorname{var}\left(Y \mid X = 0\right) = \frac{km\{k(m+1)+2m\}}{(k+m)^2\left\{1-\left(\frac{k+m}{k+2m}\right)^{k}\right\}}\left[1-\frac{km}{\{k(m+1)+2m\}\left\{1-\left(\frac{k+m}{k+2m}\right)^{k}\right\}}\right] \qquad (37)$$

and for x = 1, 2, 3, …,

$$\operatorname{var}(Y|X = x) = \left[\frac{m(k + x)}{(k + m)^2}\right](k + 2m) \tag{38}$$

The Operating Characteristic for "Birth Proneness"

$$P(Y \geq \alpha|x = 0) = 1 - \frac{\left(\frac{k}{k + 2m}\right)^k \frac{1}{\Gamma(k)}}{\left[\left(\frac{k}{k + m}\right)^k - P_{00}\right]} \sum_{y=1}^{\alpha-1} \frac{\Gamma(k + y)}{\Gamma(1 + y)} \left(\frac{m}{k + 2m}\right)^y \tag{39}$$

For x = 1, 2, 3, . . . we have

$$P(Y \geq \alpha|X = x) = 1 - \frac{1}{\Gamma(x + k)} \left(1 - \frac{m}{k + 2m}\right)^{k+x} \sum_{y=0}^{\alpha-1} \frac{\Gamma(k + x + y)}{\Gamma(1 + y)} \left(\frac{m}{k + 2m}\right)^y \tag{40}$$

The formula to obtain the correlation coefficient between X and Y in terms of the parameters when Z = X + Y follows the zero-truncated negative binomial distribution is given by

$$\rho_{XY} = \frac{m\left[A(k+1)-k\right]}{\left[A(k+m)-km(1-A)\right]} \quad . \qquad (41)$$

where $A = 1 - \left(\dfrac{k}{k+2m}\right)^k$

This formula may also be obtained from the formula derived by Hamdan (1975) after putting $\theta = \dfrac{k}{k+2m}$ in his formula for the correlation coefficient.

Note that the model and its characteristics could be directly obtained by putting p = q = 0.5 and μ=2m in case (II).

The model is investigated by Sinha and Kumar (2001).

(4) Applications of the Model

For describing the application of the model we use the data set (Indian Data) referred to by Sinha (1985). The zero-truncated NBD model is fitted to the observed distribution of children (Z=X+Y) using the maximum likelihood estimates of the parameters involved. These estimates are obtained as $\hat{k}=4.6129$ and $\hat{\theta}=0.50730$.

The model provides a fairly good fit to the observed distribution. Also, we obtain

$2\hat{m} = 4.66667$, and $\hat{\rho}_{XY} = 0.29035$.

Using the estimates of *m* and *k* the zero-cell truncated SBNBD is fitted. Table 11 shows the observed as well as the expected frequency. It is a fairly good fit.

The product moment correlation coefficient is obtained as 0.23890 using Karl Pearson's method. As it is significantly different from zero at 5% level of significance,

Its 95 percent confidence limits are obtained as 0.16994 (lower) and 0.30553 (upper), which include the estimated value of the population correlation coefficient (ρ) obtained on the basis of the parameters of the model.

Further, we use the regression equations involving the parameters to obtain the expected number of daughters (Y) and sons (X) a mother can have who has already a given number of sons and daughters respectively. For this, we use the following equations:

$$E(Y|X=0) = 2.10224 = E(X|Y=0)$$
$$E(Y|x) = 1.54954 + 0.33591\,x;$$
$$E(X|y) = 1.54954 + 0.33591\,y$$

where x = 1, 2, …, and y = 1, 2, ….

The observed and the expected values are given in Table 12.

Table 11: Observed and expected number (under parenthesis) of Indian mothers with X daughters and Y sons

X / Y	0	1	2	3	4	5	6	7+	Total
0	-	40 (35.8)	23 (25.2)	9 (14.0)	3 (6.7)	2 (2.9)	3 (1.2)	- (0.7)	80 (86.5)
1	54 (35.8)	52 (50.5)	34 (42.0)	19 (26.8)	11 (14.5)	6 (7.0)	5 (3.1)	1 (2.1)	182 (181.8)
2	29 (25.2)	47 (42.0)	43 (42.0)	38 (29.0)	20 (17.5)	12 (9.3)	9 (4.6)	3 (3.6)	201 (171.4)
3	15 (14.0)	21 (26.8)	21 (29.0)	18 (23.4)	20 (15.6)	6 (9.1)	4 (4.8)	- (4.2)	105 (126.9)
4	6 (6.7)	11 (14.5)	18 (17.5)	16 (15.6)	13 (11.4)	5 (7.2)	4 (4.1)	6 (4.1)	79 (81.1)
5	2 (2.9)	12 (7.0)	9 (9.3)	10 (9.1)	9 (7.2)	3 (4.9)	3 (3.0)	2 (3.3)	50 (46.7)
6	1 (1.2)	4 (3.1)	5 (4.6)	3 (4.8)	4 (4.1)	1 (3.0)	4 (2.0)	2 (2.4)	24 (25.2)
7+	3 (0.7)	4 (2.1)	6 (3.6)	3 (4.2)	3 (4.1)	2 (3.3)	2 (2.4)	- (4.0)	23 (24.4)
Total	110 (86.5)	191 (181.8)	159 (171.4)	116 (126.9)	83 (81.1)	37 (46.7)	34 (25.2)	14 (24.4)	744 (744.0)

The resemblance between the observed and the expected values appears reasonably good.

Table 12: Acutal ($\bar{x}$ and $\bar{y}$) and expecte number of [E(X|y) or E(Y|x)] of daughters (X) and sons (Y) per mother having x daughters and y sons.

X or Y	$\bar{y}$	E(X|y) or E(Y|x)	$\bar{x}$
0	2.01	2.10	1.91
1	1.92	1.88	1.58
2	2.36	2.22	2.31
3	2.63	2.56	2.39
4	3.08	2.89	3.06
5	2.81	3.23	2.92
6	3.06	3.56	3.54
7	3.29	3.90	2.60
8	4.33	4.24	2.75

For the birth proneness we have

$$P(Y \geq \alpha | X=0) = 1 - \frac{0.35668}{\Gamma(4.61294)} \sum_{y=1}^{\alpha-1} \frac{\Gamma(y+k)}{\Gamma(y+1)} (0.25145)^y$$

$$P(X \geq \beta | Y=0) = 1 - \frac{0.35668}{\Gamma(4.61294)} \sum_{x=1}^{\beta-1} \frac{\Gamma(x+k)}{\Gamma(x+1)} (0.25145)^x$$

$$P(Y \geq \alpha | x) = 1 - \frac{(0.2629)(0.74855)^x}{\Gamma(4.61294 + x)} \sum_{y=0}^{\alpha-1} \frac{\Gamma(x+y+4.61294)}{\Gamma(y+1)} (0.25145)^y$$

and

$$P(X \geq \beta | y) = 1 - \frac{(0.2629)(0.74855)^y}{\Gamma(4.61294 + y)} \sum_{x=0}^{\beta-1} \frac{\Gamma(x+y+4.61294)}{\Gamma(x+1)} (0.25145)^x$$

where x = 1, 2, … and y = 1, 2, …

Here β(1, 2, …) denotes the number of daughters. The expected values along with the observed values are given in Tables 13 and 14.

Table 13: Observed and expected number of mothers having α (1, 2, ... 7) or more sons given x number of daughters

Frequency	X (Daughters)								
	0	1	2	3	4	5	6	7	8
Obs. ($Y \geq 1$ | x) Exp.	110 110.0	151 153.4	136 135.6	107 103.2	80 76.1	35 34.7	31 32.4	7 6.7	3 2.9
Obs. ($Y \geq 2$ | x) Exp.	56 64.5	99 100.3	102 96.6	88 78.7	69 61.3	29 29.2	26 28.2	6 6.0	3 2.6
Obs. ($Y \geq 3$ | x) Exp.	27 32.4	52 56.2	59 59.3	50 52.2	49 43.4	17 21.8	17 22.1	4 4.9	3 2.2
Obs. ($Y \geq 4$ | x) Exp.	12 14.6	31 28.1	38 32.4	32 30.8	29 27.4	11 14.6	13 15.6	4 3.6	3 1.8
Obs. ($Y \geq 5$ | x) Exp.	6 6.0	20 12.9	20 16.2	16 16.6	16 15.8	6 8.9	9 10.1	1 2.4	1 1.2
Obs. ($Y \geq 6$ | x) Exp.	4 2.4	8 5.5	11 7.5	6 8.2	7 8.4	3 5.0	6 6.0	1 1.5	0 0.8
Obs. ($Y \geq 7$ | x) Exp.	3 0.9	4 2.2	6 3.3	3 3.8	3 4.2	2 2.6	2 3.3	0 0.9	0 0.5

Table 14: The observed and expected number of mothers having β (1, 2, 7) or more daughters given y number of sons

Frequency	Y (Sons)								
	0	1	2	3	4	5	6	7	8
Obs. $(X \geq 1 \mid y)$	80	128	172	90	73	48	23	13	4
Exp.	80.0	146.2	171.4	93.4	72.5	46.9	22.9	14.4	3.9
Obs. $(X \geq 2 \mid y)$	40	76	125	69	62	36	19	10	3
Exp.	46.9	95.6	122.2	71.3	58.4	39.4	19.9	12.9	3.5
Obs. $(X \geq 3 \mid y)$	17	42	82	48	44	27	14	6	2
Exp.	23.5	53.6	75.0	47.2	41.3	29.5	15.6	10.5	3.0
Obs. $(X \geq 4 \mid y)$	8	23	44	30	28	17	11	5	1
Exp.	10.6	26.7	41.0	27.9	26.1	19.8	11.0	7.8	2.4
Obs. $(X \geq 5 \mid y)$	5	12	24	10	15	8	7	3	1
Exp.	4.4	12.2	20.4	15.0	15.0	12.1	7.1	5.3	1.6
Obs. $(X \geq 6 \mid y)$	3	6	12	4	10	5	6	2	0
Exp.	1.7	5.2	9.5	7.5	8.0	6.8	4.2	3.3	1.1
Obs. $(X \geq 7 \mid y)$	0	1	3	0	6	2	2	0	0
Exp.	0.6	2.1	4.1	3.5	4.0	3.6	2.3	1.9	0.7

The models provide a good fit to the observed distributions of sibship sizes and are helpful in the investigations of various characteristics of the observed phenomena.

These models may be used by policy and decision makers as well as research workers for the study of other bivariate situations too.

Weighted Distribution and Its Impact

While studying "the effect of method of ascertainment upon the estimation of frequencies" Fisher (1934) initiated and explained the need for an adjustment in specification depending on the way data are ascertained. But Professor C. R. Rao formalized the concept and developed the theory, which is called weighted distributions as a method of adjustment applicable to many situations. See Rao (1997), and Patil and Rao (1977) for more details.

Some the situations where this could arise are given below.

(1) Truncation
(2) Partial destruction of observations
(3) Sampling with unequal chances of observations.

The situations that generate weighted distributions involve "non-response" . In fact, these situations refer to instances where the recorded observations cannot be considered as a random sample from the original distribution.

Let X be a non-negative observable random variable with its natural probability function $f(x,\theta)$ where $\theta \in \Omega$, the parameter space. Suppose x of X under $f(x,\theta)$ enters the investigation record with probability proportional to $\omega(x,\beta)$. Here the recording function is a non-negative function with $\omega(x,\beta)$ parameter β, which represents the recording mechanism. The recording function is also called weight function.

Note that the recorded x is not an observation on X but on the rv X^{ω} with the following pf:

$$f^{\omega}(x;\theta,\beta)=\frac{\omega(x,\beta)\,f(x,\theta)}{\omega}$$

Here $\omega = E[\omega(x,\beta)]$ and it is known as the normalizing factor. The rv X^{ω} is the weighted version of X and its distribution is referred to as the weighted distribution with the weight function ω. In case $\omega(x,\beta)=x$, $X^* = X^\omega$ is said to be the size-biased form of X. Then the distribution of X is called to the size-biased distribution with the following probability function:

$$f^*(x;\theta)=\frac{x\,f(x,\theta)}{\mu}$$

where $\mu = E(X)$.

Note that f^* denotes the size-biased form of f. The resulting sighting mechanism is referred to the size-biased sampling. For example, the size-biased negative binomial distribution is a special case of the size-biased distribution when X follows the negative binomial distribution.

Bivariate Weighted Distributions

Suppose (X,Y) is a pair of non-negative rvs with a joint pdf $f(x,y)$ and $\omega(x,y)$ is a non-negative weight function. We assume here that $\mathrm{E}[\omega(\mathrm{X,Y})]$ exists for the joint pdf of (X,Y).

Thus the weighted form of $f(x,y)$ is defined below:

$$f^{\omega}(x,y)=\frac{\omega(x,y)f(x,y)}{E[\omega(X,Y)]}.$$

Suppose $\omega(x,y)=x^{\alpha}$. This leads to

$$f^{\omega}(x,y)=\frac{x^{\alpha}f(x,y)}{E[X^{\alpha}]}.$$

On putting $\alpha = 1$ we can find out the size-biased form of the bivariate distributions. Mahfoud (1978) mentions a number of useful weights functions that include $\omega(x, y)$ as 1, x, y, x+y, x*y, x(x-1)+y(y-1), $x^2 + y^2$, max (x,y) etc. One could obtain a number of weighted distributions corresponding to the bivariate compound Poisson models to describe various bivariate data sets.

REFERENCES

[1] Arbous, A.G. and Kerrich, J.E. (1951). Accident statistics and the concept of accident proneness. *Biometrics*, 7, 340-432.

[2] Arbous, A.G. and Sichel, H.S. (1954). New technique for the analysis of absenteeism data. *Biometrika*, 41, 77-90.

[3] Froggatt, P. (1970). Applications of discrete distribution theory to the study of noncommunicable events in medical epidemiology. In *Random Counts in Scientific Work: Biomed. and Soc. Sc.*, 2, (Ed. G. P. Patil), 15-40.

[4] Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Ann. Eugen.*, 6, 13-25.

[5] Hamdan, M. A. (1975). Correlation between the numbers of two types of children when the family size distribution is zero-truncated negative binomial. *Biometrics*, 31, 765-766.

[6] James, W.H. (1957). Sex ratio and the sex composition of the existing sibs. *Annals of Human Genetics,* 38, Part 3, 371-378.

[7] Mahfoud, M. (1978). A study of weighted distributions and size-biased sampling. A Ph. D. thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Department of Statistics, Pennsylvania State University, University Park, Pa 16802, USA.

[8] Newbold, E. M. (1927). Practical applications of repeated events, particularly to industrial accidents. *J. Roy. Statist. Soc.,* 90,487-535.

[9] Patil, G. P. and Rao, C. R (1977). The weighted distributions: a survey of their applications. In *Applications of Statistics* (Ed. P. K. Krishnaiah), 383-405. North-Holland Publishing Company, USA.
[10] Rao, B.R., Mazumdar, S., Waller, J.H. and Li, C.C. (1973). Correlation between the number of two types of children in a family. *Biometrics*, 29, 271-279.
[11] Rao, C. R. (1997). *Statistics and Truth: Putting Chance to Work*, Second Edition. World Scientific, Singapore.
[12] Sinha, Arun K., and Mishra, R.S. (2004). The zero-truncated bivariate compound Poisson distribution and its applications. Manuscript.

[13] Sinha, Arun K., Rai, M. and Kumar, R. (2003). A bivariate compound Poisson model and its applications. Manuscript.

[14] Sinha, Arun K. and Kumar, R. (2001). The zero-truncated symmetrical bivariate negative binomial distribution. *American Journal of Mathematical and Management Scieces,* 21, Nos.1&2, 57-68.

[15] Sinha, Arun K. and Kumar, D. (1989). An application of the symmetrical bivariate negative binomial distribution. In *Population Transition in India,* (Eds. S. N. Singh et al.), *Vol. 1,* 435 – 444. B. R. Publication Corporation, Delhi.

[16] Sinha, Arun K. (1985). On the symmetrical bivariate negative binomial distribution. In *Statistical Theory and Data Analysis* (Ed. K. Matusita), Elsevier Science Publishers, B.V., North Holland, New York, 649-657.

THANK YOU VERY MUCH.