INFLATED AND MODIFIED BIVARIATE DISCRETE DISTRIBUTIONS

by
G. S. Lingappiah *

Summary

A bivariate distribution is considered in general, where (i,j)-th count is misreported as (ℓ,m)-th count, which leads to "modification" of size λ $(0 \leqslant \lambda \leqslant 1)$. In this case, variance and covariance are expressed in the form of inverted parabolas. In order to obtain the asymptotic distribution of the maximum likelihood estimates, terms of the variance-covariance matrix are provided for the bivariate negative binomial case. Next, a bivariate "inflated" distribution is considered where all the terms except (0,0) are inflated by the quantity λ $(0 \leqslant \lambda \leqslant 1)$. In this case, correlation is being analysed and also the Bayesian estimates of the parameters are obtained for the bivariate negative binomial.

Key Words: Inflation; Modification; Bivariate distribution; Bayesian estimates.

1. Introduction

Inflated and modified distributions have received much attention recently. Singh [14] and Singh [16] consider inflated binomial and Poisson distributions respectively. Sobic and Szynal [17] give the distribution of the sum of independent variables from the inflated binomial distribution. Lauchenbruch [4] considers the inflated negative binomial distribution. Lingappaiah [5, 6] deals with the inflated double binomial and generalised Poison distributions and discusses the variance in each case in detail with

*Department of Mathematics, Sir George William Campus, Concordia University, Montreal, Canada.

graphical displays and the estimation of the parameters is also discussed. Lingappaiah and Patel [9] deal with the inflated distribution of the general type. In this paper, on the above lines, bivariate distribution is considered in general in which each term except (0,0) is inflated by a quantity λ $(0 \leqslant \lambda \leqslant 1)$. Correlation in this inflated case is expressed in terms of λ and the correlation in the simple (non-inflated) case.

Further the Bayesian estimates of the parameters in the case of bivariate negative binomial are estimated. Next, another situation is considered by many of the works in the reference and this deals with that of "modification" where some count or counts are misreported. Cohen [1, 2, 3] deals with such a situation in binomial and Poisson cases. Parikh and Shah [11] consider the same problem as related to power series distribution. Lingappaiah [7, 8] generalises the above situations to a case where one set of counts are misreported as another set of counts or where a set of counts are misreported as a single count and many other similar possibilities. In this note, bivariate modified distribution is considered where a single count (i,j) is misreported as the count (ℓ,m). The variance and covariances are expressed as parabolic functions. The elements of the variance-covariance matrix are given to evaluate the asymptotic distributions of the estimates of the parameters in the case of bivariate negative binomial distribution.

2. Bivariate Modified distribution

Suppose count (i,j) is misreported as the count (l,m), then the modified distribution can be written as

$$P(x,y) = \begin{cases} P_{ij} + \lambda P_{\ell m} & \text{if } x = i,\ y = j \\ (1-\lambda)P_{\ell m} & \text{if } x = \ell,\ y = m \\ P_{rt} & \text{if } \begin{matrix} r \neq i, \ell \\ t \neq j, m \end{matrix} \end{cases} \tag{1}$$

where $0 \leqslant \lambda \leqslant 1$ and $P_{ij} = P(x = i, y = j)$ and similarly $P_{\ell m}$ and P_{rt}.

From (1), we have

$$\mu_x = m_x + a\lambda P_{\ell m}$$

$$\mu_x = m_x + a\lambda P_{\ell m}$$

$$\mu_y = m_y + b\lambda P_{\ell m} \tag{2a}$$

and

$$\mu'_{2x} = m_{2y} + (\lambda P_{\ell m})\ (i^2 - \ell^2)$$

$$\mu'_{2y} = m_{2y} + (\lambda P_{\ell m})\ (j^2 - m^2) \tag{2b}$$

where m_x, μ_x are E(x) for the simple (non-modified) and modified distributions respectively. Similarly m_y and μ_y. μ'_{2x} and m_{2m} are E(x^2) for the modified and the simple cases. μ'_{2y}, m_{2y} are the same quantities for y and a = $(i - \ell)$, b = $(j - m)$. If $\lambda = 0$ in (1), we have the simple distribution.

2(a): Variances and Covariances

Now from (2*a*) and (2*b*) it follows that

$$\sigma^2_{0(x)} = \sigma^2_x + (\lambda a P_{\ell m})\ (i + \ell) - 2\lambda a m_x P_{\ell m} - \lambda^2 a^2 P^2_{\ell m} \tag{3a}$$

where $\sigma^2_{0(x)}$, σ^2_x are the variances for the modified and simple cases respectively and if $z = \sigma^2_{0(x)} - \sigma^2_x$, we can write (3*a*) in the form of an inverted parabola as

$$\left[Z - \frac{(i + \ell)-2m_x}{4}\right]^2 = - a^2 P^2_{\ell m} \left[\lambda - \frac{(i + \ell)-2m_x}{2aP_{\ell m}}\right]^2 . \tag{3b}$$

This parabola is truncated at right at $Z = (aP_{\ell m})\ [(i + \ell) - 2m_x - aP_{\ell m}]$ which is the value of Z at $\lambda = 1$. Sometimes (3*b*) gives most of the parabola in the region $0 \leqslant \lambda \leqslant 1$ and sometimes a very small part of parabola in this region depending on a and $P_{\ell m}$ and the location of the vortex of the parabola. Now from (1) again, we have

$$\mu_{xy} = m_{xy} + \lambda ijP_{\ell m} - \lambda \ell m P_{\ell m} \qquad (4)$$

where μ_{xy}, m_{xy} are $E(xy)$ for the modified and simple distributions respectively. From (2a) and (4), we get

$$\sigma_{0(xy)} = \sigma_{xy} + (\lambda P_{\ell m}) \; [(ij - \ell m) - (bm_x + am_y) - \lambda abP_{\ell m}] \, (4a)$$

with $\sigma_{0(xy)}$, σ_{xy}, the covariances for he modified and simple cases. Again (4a) can be expressed as an inverted parabola in λ as

$$U - \left[\frac{(ij - \ell m) - (am_y + bm_x)^{\;2}}{4ab}\right]$$
$$= - abP_{\ell m}^2 \left[\lambda - \frac{(ij - \ell m) - (am_y + bm_x)}{2abP_{\ell m}}\right]^2 . \qquad (5)$$

with $U = \sigma_{0(xy)} - \sigma_y$. This parabola is again truncated in its right arm at $U = (P_{\ell m}) \; [(ij - \ell m) - (am_y + bm_x) - ab]$ which is the value of U when $\lambda = 1$. Again, as before, we get most of the parabola in the region $0 \leqslant \lambda \leqslant 1$ or only a part of it in this region depending upon the values of a, b, $P_{\ell m}$ and the location of the vortex of this parabola.

2(b): Asymptotic distributions: Now from (1) we have the likelihood function as

$$L = (P_{ij} + \lambda P_{\ell m})^{n_{ij}} [(1 - \lambda)P_{\ell m}]^{n_{\ell m}} \prod_{t \neq i, \ell,\; t \neq j,m} (P_{rt})^{n_{rt}} \qquad (6)$$

from which we get

$$D_\lambda = \frac{\partial}{\partial \lambda} \log L = \frac{n_{ij} P_{\ell m}}{P_{ij} + \lambda P_{\ell m}} - \frac{n_{\ell m}}{1 - \lambda} \qquad (7)$$

which gives with $D_\lambda = 0$,

$$\hat{\lambda} = \frac{n_{ij}P_{\ell m} - n_{\ell m}P_{ij}}{P_{\ell m}(n_{ij} + n_{\ell m})} \quad . \tag{7a}$$

Now consider the bivariate negative binomial distribution given by

$$f(x, y) = \frac{\Gamma(k + x + y)}{\Gamma(k)x!y!} \theta_1^x \; \theta_2^y \; (1 - \theta_1 - \theta_2)^k \tag{8}$$

$0 < \theta_i < 1, i = 1, 2,; 0 < K < \infty, \theta_1 + \theta_2 < 1, x, y = 0, 1, 2, \ldots.$

From (6) and (8), we get

$$D_{\theta_1} = \frac{\partial}{\partial \theta_1} \log L = \frac{n_{ij} \frac{\partial}{\partial \theta_1}[P_{ij} + \lambda P_{\ell m}]}{P_{ij} + \lambda P_{\ell m}} + n_{\ell m} \frac{\frac{\partial}{\partial \theta_1}(P_{\ell m})}{P_{\ell m}}$$

$$+ \Sigma \; \Sigma \, n_{rt} \; \frac{\left(\frac{\partial}{\partial \theta_1}\right) P_{rt}}{P_{rt}} \tag{9}$$

where $\Sigma\,\Sigma$ is $\underset{r \neq i, l,}{\Sigma} \; \underset{t \neq j, m}{\Sigma}$. Now for the case (8), we have

$$\frac{\partial}{\partial \theta_1} P_{ij} = \left(\frac{i}{\theta_1} - \frac{k}{A}\right) P_{ij}. \tag{9a}$$

Using (9a) in (9), we get

$$D_{\theta_1} = \frac{n_{ij}}{\theta_1} \left[\frac{iP_{ij} + \lambda \ell P_{\ell m}}{P_{ij} + \lambda P_{\ell m}}\right] + n_{\ell m} \frac{\ell}{\theta_1} - \frac{nk}{A} + \Sigma \; \Sigma \; \frac{(rn_{rt})}{\theta_1} \tag{10}$$

Further $D_{\theta_1} = 0$ implies

$$n_{ij} \left[\frac{\frac{\partial}{\partial \theta_1}(P_{ij} + \lambda P_{\ell m})}{P_{ij} + P_{\ell m}}\right] = \frac{k}{A}(n - n_{ij}) - \frac{1}{\theta_1}[\ell n_{\ell m} + \Sigma\;\Sigma\; rn_{rt}]. \tag{11}$$

From (9a) and (11) we get

$$\frac{n_{ij}}{\theta_1}\left[\frac{iP_{ij} + \lambda \ell P_{\ell m}}{P_{ij} + \lambda P_{\ell m}}\right] = \frac{nk}{A} - \frac{1}{\theta_1}\left[\ell n_{\ell m} + \Sigma\ \Sigma\ r n_{rt}\right] \tag{12}$$

Using (7) in $D_\lambda^2 = \frac{\partial^2}{\partial\lambda^2}\log L$, we get

$$D_\lambda^2 = \frac{n_{\ell m}}{n_{ij}}\left(\frac{n_{\ell m} + n_{ij\cdot}}{(1-\lambda)^2}\right) \tag{13}$$

Using (7), (11) and (12), we get

$$D_{\lambda\theta_1}^2 = \left(\frac{\partial^2}{\partial\lambda\partial\theta_1}\log L\right) = \left[\frac{1}{(1-\lambda)}\right]\left[\frac{1}{\theta_1}\ell n_{\ell m} + \frac{n_{\ell m}}{n_{ij}}(\ell n_{\ell m} + \Sigma\Sigma\, r n_{rt}) - \frac{k}{A}\frac{n\, n_{\ell m}}{n_{ij}}\right]. \tag{14}$$

Similarly, we have using (10), (11) and (12)

$$D_{\theta_1}^2 = \frac{\partial^2}{\partial\theta_1^2}\log L = -\frac{nk}{A}\left(\frac{1}{A} + \frac{1}{\theta_1}\right) + \frac{n_{ij}}{\theta_1^2}\left[\frac{i^2P_{ij} + \lambda\ell^2 P_{\ell m}}{P_{ij} + \lambda P_{\ell m}}\right] - \frac{1}{n_{ij}}\left[\frac{nk}{A} - \frac{1}{\theta_1}(\ell n_{\ell m} + \Sigma\ \Sigma\ r n_{rt})\right]^2 \tag{15}$$

$D_{\lambda\theta_2}^2$, $D_{\theta_2}^2$ are very similar in forms to those of $D_{\kappa\theta_1}$ and $D_{\theta_1}^2$. From these quantities, elements of the variance-covariance matrix can be found and the asymptotic distributions of $\hat{\lambda}$, $\hat{\theta}_1$ and $\hat{\theta}_2$ can be evaluated.

3. Bivariate inflated distribution

Now consider the situation where there is more concentration of probability at (0, 0) while all other terms are inflated by a quantity λ $(0 \leqslant \lambda \leqslant 1)$. In such a case, density function can be written as

$$\tilde{P}(x,y) = \begin{cases} 1-\lambda+\lambda P_{00}, & \text{if } x=0, y=0 \\ \lambda P_{xy}, & \text{if } x\neq 0,\ y\neq 0 \end{cases} \tag{16}$$

as before $P_{ij} = P(x=i,\ y=j)$. From (16), we get

$$\mu_x = \lambda m_x,\ \sigma^2_{0(x)} = \lambda\sigma_x^2 + \lambda(1-\lambda)m_x^2 \tag{17}$$

where μ_x, m_x are $E(x)$ for the inflated and the simple (non-inflated) distributions respectively and similarly $\sigma^2_{0(x)}$ and σ_x^2 are the variances for the inflated and the simple cases. μ_y, m_y, $\sigma^2_{0(y)}$ and σ_y^2 are for the variable y. For the case of bivariate negative binomial given by (8), we have for $\theta_1 = .4$, $\theta_2 = .2$, $k = 2$, $(m_x/\sigma_x)^2 = k\theta_1/(1-\theta_2) = 1$ and

$$(y-1) = (\lambda-1)^2 \text{ where } y = \sigma^2_{0(x)}/\sigma_x^2 \tag{18a}$$

which is an inverted parabola. Similarly for $\theta_1 = .5$, $\theta_2 = .4$,, $k = 2$, we get

$$[y-(16/15)] = -(5/3)\ [\lambda-(4/5)]^2. \tag{18b}$$

Expression (18b) is also an inverted parabola and we have more of the parabola (18b) in the region $(0 \leqslant \lambda \leqslant 1)$ than that of (18a). This is shown in the table below giving the values of y in (18a) and (18b)

Table I

λ	Values of y from (18a) and (18b)	
	$\theta_1 = .4,\ \theta_2 = .2,\ k = 2$	$\theta_1 = .5,\ \theta_2 = .4,\ k = 2$
0	0	0
.1	.19	.2500
.2	.36	.4667
.3	.51	.6500
.4	.64	.8000
.5	.75	.9167
.6	84	1.0000
.7	.91	1.0500
.8	.96	1.0667
.9	.99	1.0500
1.0	1.00	1.0000

Again from (16), we get $\mu_{xy} = \lambda m_{xy}$ and

$$\sigma_{0(xy)} = \lambda\sigma_{xy} + \lambda(1-\lambda)\, m_x m_y \tag{19}$$

where $\sigma_{0(xy)}$, σ_{xy} are the covariances for the inflated and the simple cases and μ_{xy},, m_{xy} are $E(xy)$for the inflated and the simple distributions respectively. From (17) and (19), we get

$$r_0 = \frac{r + (1-\lambda)\, ab}{[1 + (1-\lambda)a^2]^{1/2}\, [1 + (1-\lambda)b^2]^{1/2}} \tag{20}$$

where r_0, r are the correlations for the inflated and the simple cases respectively and $a = m_x/\sigma_x$ and $b = m_y/\sigma_y$. Expression (20) can be written as

$$A(1-\lambda)^2 + B(1-\lambda) + C = 0 \tag{21}$$

with $A = a^2b^2(1-r_0^2)$

$$B = 2rab - r_0^2(a^2 + b^2) \tag{22}$$

$$C = r^2 - r_0^2$$

and $\Delta = B^2 - 4AC$ can be again expressed as a parabola in r as

$$\Delta = A_1 r^2 + B_1 r + C_1 \tag{23}$$

where $A_1 = 4a^2b^2r_0^2$

$$B_1 = 4r_0^2 ab(a^2+b^2)$$

$$C_1 = r_0^4(a^2-b^2)^2 + 4r_0^2a^2b^2 \tag{24}$$

if $r_0 =$ then $\Delta = [2abr - (a^2 + b^2)]^2$ (25a)

and if $r_0 \neq 1$, $a = b$, then

$$\Delta = 4a^4\,(1-r)^2 \tag{25b}$$

3a: Bayesian estimates

From (16), we get with a sample size n, the likelihood function as

$$L = (1-\lambda+\lambda P_{00})^{n_{00}} \prod_{i\neq 0} \prod_{j\neq 0} (\lambda P_{ij})^{n_{ij}} \tag{26}$$

with $\sum_i \sum_j n_{ij} = n$ and n_{00} is at (0,0). Now consider the bivariate negative binomial given by (8), for which $P_{00} = A_0^k = (1-\theta_1-\theta_2)^k$ Taking the priors for θ_1, θ_2 and λ as

$$f(\theta_1, \theta_2, \lambda) = g(\theta_1, \theta_2)\, h(\lambda) = [T]\lambda^{u-1}(1-\lambda)^{v-1}\theta_1^{a_1-1}\theta_2^{a_2-1}A_0^{a_3-1} \tag{27}$$

where $T = \Gamma(a)/\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)B(u, v)$ with $a = a_1 + a_2 + a_3$. Then from (26) and (27) we can write

$$L(x,y;\theta_1,\theta_2,\lambda)f(\theta_1,\theta_2,\lambda) = \sum_{t=0}^{n_{00}} (1-\lambda)^{c-1}\lambda^{d-1}(Q)^{n-n_{00}}$$

$$\binom{n_{00}}{t} \cdot \theta_1^{e-1}\,\theta_2^{f-1}\,A_0^{q-1}[T] \tag{28}$$

where

$$c = n_{00} + v - t, \qquad e = \sum_i i n_{i.} + a_1, \qquad q = k(t+n-n_{00}) + a_3$$

$$d = n - n_{00} + u + t, \qquad f = \sum_j j n_{.j} + a_2,$$

$$Q = \left[\prod_{i\neq 0} \prod_{j\neq 0} [\Gamma(k+i+j)/\Gamma(k)i!\,j!\,]^{n_{ij}}\right].$$

From (28), we get the Bayesian estimate of λ as

$$E(\hat{\lambda}) = \sum_{t=0}^{n_{00}} B(d+1, c) \binom{n_{00}}{t} A_0^{kt} \Big/ \sum_{t=0}^{n_{00}} B(d, c) A_0^{kt} \binom{n_{00}}{t} \quad (30)$$

and similarly the estimate of θ_1 as

$$E(\hat{\theta}_1) = \frac{\sum_{t=0}^{n_{00}} [\lambda/(1-\lambda)]^t \binom{n_{00}}{t} (e)\Gamma(q)/\Gamma(e+f+q+1)}{\sum_{t=0}^{n_{00}} [\lambda/(1-\lambda)]^t \binom{n_{00}}{t} \Gamma(r)/\Gamma(e+f+q)} \quad (31)$$

Suppose, we have $n = 4$, $n_{00} = 2$, $u = v = a_1 = a_2 = a_3 = 2$, $k = 2$, $\theta_1 = .2$, $\theta_2 = .4$, then we get from (30)

$$E(\hat{\lambda}) = .54938 \quad (32)$$

and suppose $n_{01} = 1$, $n_{10} = 1$, $n = 4$, $n_{00} = 1$, and for the value of $E(\hat{\lambda})$ in (32) we get from (31), $E(\hat{\theta}_1) = .23214$.

4. *References*

[1] Cohen, A.C. (1960) Estimating the parameters of a modified Poisson distribution. *Journal of American Statistical Association*, Vol. 55, pp. 139-143.

[2] Cohen, A.C. (1960) Misclassified data from a binomial population. *Technometrics*, Vol. 2, pp. 109-113.

[3] Cohen, A.C. (1960) Estimation in the Poisson distribution when the sample values (c + 1) are sometimes erroneously as c. *Annals of Institute of Statistical Mathematics*, Vol. 9, pp. 181-193.

[4] Lachenbruch, P.A. (1975) Estimation of parameters of the Poisson with excess zeroes and negative binomial with excess zeroes distribution. *Biometrische Zeitshrift*, Vol. 17, pp. 339-344.

[5] Lingappaiah, G.S. (1977) On some inflated generalised discrete distributions. *Communications in Statistics*, Vol. A6(3), pp. 231-141.

[6] Lingappaiah, G.S. (1977) On the inflated discrete distributions. *Statistica,* Vol. 37, pp. 379-386.

[7] Lingappaiah, G.S. (1977) On some discrete distributions with varying probabilities. *Egyptian Statistical Journal,* Vol. 21, pp. 1-15.

[8] Lingappaiah, G.S. (1978) Further investigations into the discrete distributions with jumps in probabilities. *Philippine Statiscian,* Vol. 27, pp. 26-35.

[9] Lingappaiah, G.S. and Patel, I.D. (1979) On the modified and inflated discrete distributions of general type. *Gujarat Statistical Review,* Vol. 6, No. 2, pp. 50-60.

[10] Pandey, K.N. (1965) On generalized inflated Poisson distribution. *Journal of Scientific Research, Benares Hindu University,* Vol. 15, pp. 157-162.

[11] Parikh, N.T. and Shah, S.M. (1969) Misclassification in power series distribution in which the value one is sometimes reported as zero. *Journal of Indian Statistical Association,* Vol. 7, pp. 11-19.

[12] Patel, I.D. and Shah, S.M. (1969) On generalized inflated power series distribution with its application to Poisson distribution. *Journal of Indian Statistical Association,* Vol. 7, pp. 20-25.

[13] Rychlik, Zdzislaw. and Szynal Dominik (1977) Inflated truncated negative binomial acceptance sampling plan. *Applikace Matematiky,* Vol. 22, No. 3, pp. 157-164.

[14] Singh, M.P. (1965) Inflated binomial distribution. *Journal of Scientific Research, Benares Hindu University,* Vol. 16, pp. 87-90.

[15] Singh, M.P. (1966) A note on generalised binomial distribution. *Sankhya,* Series A, Vol. 28, pp. 99.

[16] Singh, S.N. (1963) A note on inflated Poisson distribution. *Journal of Indian Statistical Association,* Vol. 1, pp. 140-144.

[17] Sobic, Lusza and Szynal, Dominik (1974) Some properties of the inflated binomial distribution. *Canadian Mathematical Bulletin,* Vol. 17, pp. 611-614.

[18] Varahamurthy, Krishnan (1967) A modified Poisson distribution. *Portugalai Mathematica,* Vol. 26, pp. 319-328.

[19] Williford, W.O. and Bingham, S.F. (1979) Bayesian estimation of the parameters in two modified Poisson distributions. *Communications in Statistics,* Vol. A8(13), pp. 1315-1326.