CHAPTER 8

THE DISPLACED NEGATIVE BINOMIAL DISTRIBUTION

8.1 Introduction

Most of the biological data are well fitted by the negative binomial distribution ([3], [4]). But there are some instances [28], where the negative binomial fits are worse. Staff [51] has defined the displaced Poisson distribution which can fit well in some cases to biological data where the negative binomial fits are worse. In this chapter, the displaced negative binomial distribution has been defined on the same lines as Staff [51]. The displaced negative binomial distribution is generated by the number of events in the negative binomial distribution in excess of a threshold value r when it is assumed that at least r events do occur.

The displaced negative binomial distribution can also be obtained as under. Let Y be a random variate having a truncated negative binomial distribution given by

$$(8.1.1)\quad P(Y=j) = \left(\Gamma(m+j) / \Gamma(j+1)\right)\theta^j / \sum_{j=r}^{\infty} \left(\Gamma(m+j) / \Gamma(j+1)\right)\theta^j,$$

$$(m > 0,\ 0 < \theta < 1;\quad j=r,\ r+1,\ r+2, \ldots).$$

Let $X = Y - r$. Then the distribution of X is given by

$$(8.1.2)\quad P(X=j) = \left(\Gamma(m+r+j)/\Gamma(r+j+1)\right)\theta^j / \sum_{j=0}^{\infty} \left(\Gamma(m+r+j)/\Gamma(r+j+1)\right)\theta^j,$$

$$(m > 0,\quad 0 < \theta < 1;\quad j=0,\ 1,2,\ldots).$$

We, therefore, define the displaced negative binomial distribution (in which r need not be a positive integer), as

$$(8.1.3)\quad P(X=j) = \left(\Gamma(m+r+j)/\Gamma(r+j+1)\right)\theta^j / \sum_{j=0}^{\infty} \left(\Gamma(m+r+j)/\Gamma(r+j+1)\right)\theta^j,$$

$$(m > 0,\ 0 < \theta < 1,\quad r > -1;\ j = 0,1,2,\ldots\infty).$$

It is suggested that the displaced negative binomial distribution may be found useful in fitting to biological data where the negative binomial fits are not good. Here, three methods of estimating the parameters are discussed: (i) Maximum likelihood method, (ii) Method A - using zero-cell frequency and the first three sample moments and (iii) Method B - using the first four sample moments. The asymptotic variance-covariances of the estimators obtained by the above three methods are derived and it is found that Method A is reasonably efficient while Method B is inefficient. The maximum likelihood estimators are complicated and laborious and so Method A is recommended.

Notation. (i) Bar below the letter indicates a vector, (ii) Dash over a vector or a matrix indicates its transpose.

8.2 The maximum likelihood estimation

The probability law for the displaced negative binomial (8.1.3) can be written as

$$(8.2.1) \qquad P(X = j) = p_j = L_j / \sum_{j=0}^{\infty} L_j, \qquad (j = 0, 1, 2, \ldots, \infty)$$

where $L_j = (\Gamma(m + r + j) / \Gamma(r + j + 1))\theta^j$ and $m > 0$, $r > -1$ and $0 < \theta < 1$.

Consider a random sample of size n from the population (8.2.1) and let n_j be the observed frequency in the sample corresponding to $X = j$. The likelihood function L is given by

(8.2.2) $$L = \prod_{j=0}^{\infty} p_j^{n_j} .$$

Let $R_j = \psi(m+r+j)$, $S_j = \psi(r+j+1)$, $T_j = R_j - S_j$, where $\psi(x) = \frac{\partial \log \Gamma x}{\partial x}$ is the digamma function tabulated by Eleanor Pairman [16]. Let $\bar{x} = \Sigma j n_j/n$, $m_R = \Sigma n_j R_j/n$, $m_S = \Sigma n_j S_j/n$, wherein the summation is taken over all sample values, $m_T = m_R - m_S$; $\mu_R = \sum_{j=0}^{\infty} p_j R_j$, $\mu_S = \sum_{j=0}^{\infty} p_j S_j$, $\mu_T = \mu_R - \mu_S$ and $\mu_1' = \sum_{j=0}^{\infty} j p_j$. Then we find that

(8.2.3) $$\frac{\partial p_j}{\partial \theta} = \theta^{-1}(j - \mu_1')p_j ,$$

(8.2.4) $$\frac{\partial p_j}{\partial m} = p_j(R_j - \mu_R) ,$$

(8.2.5) $$\frac{\partial p_j}{\partial r} = p_j(T_j - \mu_T) .$$

Taking logarithm of L and differentiating respectively w.r.t. θ, m and r, we obtain

(8.2.6) $$\frac{\partial \log L}{\partial \theta} = n\theta^{-1}(\bar{x} - \mu_1') ,$$

(8.2.7) $$\frac{\partial \log L}{\partial m} = n(m_R - \mu_R) ,$$

$$(8.2.8) \qquad \frac{\partial \log L}{\partial r} = n(m_T - \mu_T).$$

From (8.2.6), (8.2.7) and (8.2.8), we get the maximum likelihood equations for estimating θ, m and r as

$$(8.2.9) \qquad \begin{aligned} \bar{x} &= \mu_1' , \\ m_R &= \mu_R , \\ m_T &= \mu_T . \end{aligned}$$

Further, the information matrix M of the maximum likelihood estimators can be found to be as

$$(8.2.10) \quad M = n \begin{bmatrix} \theta^{-2}V(j) & , \theta^{-1}Cov(j, R_j), & \theta^{-1}Cov(j, T_j) \\ \theta^{-1}Cov(j, R_j), & V(R_j), & Cov(R_j, T_j) \\ \theta^{-1}Cov(j, T_j), & Cov(R_j, T_j) & , V(T_j) \end{bmatrix}.$$

Then M^{-1} will be the asymptotic variance-covariance matrix of the maximum likelihood estimators of θ, m, r.

The method of scoring (Rao [35], p.169) can be applied to solve the maximum likelihood equations (8.2.9). Let M_0 be the information matrix at the initial solution

(θ_0, m_0, r_0). Let $\underset{\sim}{\alpha}_0 = (\theta^{-1}\bar{x},\ m_R,\ m_T)_0'$ = column vector at the initial solution for the sample and $\underset{\sim}{\beta}_0 = (\theta^{-1}\mu_1',\ \mu_R,\ \mu_T)_0'$ = column vector at the initial solution for the population. Then the increments $\Delta\theta$, Δm and Δr in the initial values are calculated from

$$\begin{bmatrix} \Delta\theta \\ \Delta m \\ \Delta r \end{bmatrix} = n\, M_0^{-1}\ (\underset{\sim}{\alpha}_0 - \underset{\sim}{\beta}_0), \tag{8.2.11}$$

where M_0^{-1} is the reciprocal of the matrix M_0. The process is repeated till the increments in θ, m and r are negligible.

8.3 Method A- using zero-cell frequency and the first three sample moments

We now give simplified method for estimating the parameters θ, m and r, which makes use of the zero-cell frequency and the first three sample moments. We call this method as Method A in short. The probabilities of the displaced negative binomial distribution satisfy the following recurrence relation.

$$p_j = \theta(m + r + j - 1)p_j/(r + j), \tag{8.3.1}$$

$$(j = 1, 2, \ldots, \infty).$$

Using (8.3.1), we get the following recurrence relation for the moments μ_n' of the displaced negative binomial distribution.

(8.3.2) $$\mu_1' = h(1 - p_0) + p_0 u ,$$

(8.3.3) $$\mu_{n+1}' = h\,\mu_n' + u \sum_{j=0}^{n-1} \binom{n}{j} \mu_j' + p \sum_{j=1}^{n} \binom{n}{j-1} \mu_j' ,$$

$$(n = 1, 2, \ldots).$$

where $h = pm - r$, $u = p(m + r)$ and $p = \theta/(1 - \theta)$.

From (8.3.2) and (8.3.3), we get

(8.3.4) $$\mu_1' = h(1 - p_0) + up_0 ,$$

(8.3.5) $$\mu_2' = h\,\mu_1' + u + p\,\mu_1' ,$$

(8.3.6) $$\mu_3' = h\,\mu_2' + u(1 + 2\,\mu_1') + p(\mu_1' + 2\,\mu_2').$$

We write (8.3.4), (8.3.5) and (8.3.6) in the matrix notation as

(8.3.7) $$\underline{P} = A\,\underline{e} ,$$

where $\underline{P} = (\mu_1', \mu_2', \mu_3')'$, $\underline{e} = (h, u, p)'$ and

$$A = \begin{bmatrix} (1 - p_0), & p_0 & , & 0 \\ \mu_1' & , & 1 & , & \mu_1' \\ \mu_2' & , & 2\mu_1' + 1, & 2\mu_2' + \mu_1' \end{bmatrix}.$$

From (8.3.7), we get

$$\underline{e} = A^{-1}\underline{p} , \tag{8.3.8}$$

where A^{-1} = inverse of the matrix $A = (|A|)^{-1} [a_{ij}]$, $i, j = 1, 2, 3$; and the elements a_{ij} are given by

$$a_{11} = 2(\mu_2' - \mu_1'^2),$$

$$a_{12} = - p_0(2\mu_2' + \mu_1'),$$

$$a_{13} = p_0 \mu_1',$$

$$a_{21} = - \mu_1'(\mu_2' + \mu_1'),$$

$$a_{22} = (1 - p_0)(2\mu_2' + \mu_1'),$$

$$a_{23} = - (1 - p_0) \mu_1',$$

$$a_{31} = \mu_1'(2\mu_1' + 1) - \mu_2',$$

$$a_{32} = p_0 \mu_2' - (1 - p_0)(2\mu_1' + 1),$$

$$a_{33} = 1 - p_0(\mu_1' + 1),$$

and $|A| = 2(\mu_2' - \mu_1'^2) - p_0(\mu_2'\mu_1' + 2\mu_2' - \mu_1'^2)$.

Let $m_r' = \Sigma\, j^r n_j/n$ denote the rth order raw moment for the sample. Then equation (8.3.8) suggests that we can estimate the parameters h, u and p, if we substitute m_r' for μ_r' $(r = 1, 2, 3)$ and n_0/n for p_0 in the r.h.s. expression of (8.3.8). Thus if $\hat{\underline{e}} = (\hat{h}, \hat{u}, \hat{p})'$ denotes the column vector of the estimators $\hat{h}$, $\hat{u}$, $\hat{p}$, then $\hat{\underline{e}}$ is given by

(8.3.9) $$\hat{\underline{e}} = (\hat{A})^{-1}\,\hat{\underline{P}},$$

where $\hat{A}$ and $\hat{\underline{P}}$ denote the values of A and $\underline{P}$ obtained by replacing the population moments μ_r' by the sample moments m_r' $(r = 1, 2, 3)$ and p_0 by n_0/n. From the estimators $\hat{h}$, $\hat{u}$, $\hat{p}$, we find the estimators of θ, m and r as

(8.3.10) $$\hat{\theta} = \hat{p}/(1 + \hat{p}),$$

(8.3.11) $$\hat{m} = (\hat{u} + \hat{p}\hat{h})/\,\hat{p}(1 + \hat{p}),$$

(8.3.12) $$\hat{r} = (\hat{u} - \hat{h})\,/\,(1 + \hat{p}).$$

We now derive the asymptotic variance-covariance matrix of the estimators $\hat{\theta}$, $\hat{m}$ and $\hat{r}$. Let $\overset{*}{X}$ denote the differential of X. Then, noting that $(XY)^* = X^*Y + XY^*$, we have from (8.3.9),

$$\hat{\underline{e}}^* = A^{-1}(\hat{\underline{P}}^* - \hat{A}^*\underline{e}). \tag{8.3.13}$$

Hence, by the δ-method (Kendall and Stuart [23], §10.6), the variance-covariance matrix V of $\hat{\underline{e}}$, to the order n^{-1}, is found to be

$$V = E(\underline{e}^* . \underline{e}^{*'}) = A^{-1}B(A')^{-1}/n, \tag{8.3.14}$$

where $B = nE(\underline{P}^* - A^*\underline{e})(\underline{P}^* - A^*\underline{e})' = [b_{ij}]$. The elements b_{ij}, $(i, j = 1, 2, 3)$ are given by

$$b_{11} = \mu_2' - \mu h + \mu_1'(u - h),$$

$$b_{12} = b_{21} = \mu_3' - (h + p)\mu_2' - uh,$$

$$b_{13} = b_{31} = D - uh,$$

$$b_{22} = C - u^2,$$

$$b_{23} = b_{32} = E - (h + p)D - u^2,$$

$$b_{33} = F - (h + 2p)E - (2u + p)D - u^2,$$

where

$$C = \mu_4' - 2(h + p)\,\mu_3' + (h + p)^2\,\mu_2' ,$$

$$D = \mu_4' - (h + 2p)\,\mu_3' - (2u + p)\,\mu_2' ,$$

$$E = \mu_5' - (h + 2p)\,\mu_4' - (2u + p)\,\mu_3' ,$$

$$F = \mu_6' - (h + 2p)\,\mu_5' - (2u + p)\,\mu_4' .$$

The variance-covariance matrix W, to the order n^{-1}, of $(\hat{\theta}, \hat{m}, \hat{r})'$ can be obtained by using (8.3.10), (8.3.11), (8.3.12) and the δ-method (Kendall and Stuart [23], §10.6) and is given by

$$W = T' V T, \tag{8.3.15}$$

where

$$T = \begin{bmatrix} (1-\theta), & -(1-\theta), & 0 \\ (1-\theta)^2/\theta, & (1-\theta), & 0 \\ -(1-\theta)(m+r-r\theta)/\theta, & -r(1-\theta), & (1-\theta)^2 \end{bmatrix}.$$

The joint asymptotic efficiency E_A, of $(\hat{\theta}, \hat{m}, \hat{r})$ relative to the maximum likelihood estimators of θ, m, r is given by

$$E_A = 1 \; / \; |M| . |W| \;, \tag{8.3.16}$$

where $|X|$ denotes the determinant of the matrix X and M is the information matrix given by (8.2.10).

8.4 Method B- using first four sample moments

We now derive the estimators of θ, m, r by the method of moments, which makes use of the first four sample moments. We call this method as Method B in short.

From (8.3.3), by taking $n = 1$, 2 and 3, we get

$$\mu_2' = h\,\mu_1' + u + p\,\mu_1' \;, \tag{8.4.1}$$

$$\mu_3' = h\,\mu_2' + u(2\,\mu_1' + 1) + p(2\,\mu_2' + \mu_1') \;, \tag{8.4.2}$$

$$\mu_4' = h\,\mu_3' + u(3\,\mu_2' + 3\,\mu_1' + 1) + p(3\,\mu_3' + 3\,\mu_2' + \mu_1') \;, \tag{8.4.3}$$

where h, u and p have the same meanings as in (8.3.2) and (8.3.3). Writing equations (8.4.1), (8.4.2) and (8.4.3) in the matrix notation, we get

$$\underline{Q} = R\,\underline{e} \;, \tag{8.4.4}$$

where $\underline{Q} = (\mu_2', \mu_3', \mu_4')'$ and $\underline{e} = (h, u, p)'$ and

$$R = \begin{bmatrix} u_1' , & 1 & , u_1' \\ u_2' , & (2 u_1' + 1) & , (2 u_2' + u_1') \\ u_3' , & (3 u_2' + 3 u_1' + 1), & (3 u_3' + 3 u_2' + u_1') \end{bmatrix} .$$

From (8.4.4), we get

$$\underline{e} = R^{-1} \underline{Q} , \qquad (8.4.5)$$

where R^{-1} = inverse of the matrix $R = (|R|)^{-1} [r_{ij}]$. The elements r_{ij}, $(i, j = 1, 2, 3)$ are given by

$$r_{11} = \mu_1'(6 \mu_3' - 3 \mu_2' - \mu_1') + 3 \mu_3' - 6 \mu_2'^2 + \mu_2' ,$$

$$r_{12} = 3 \mu_1'(\mu_2' + \mu_1') - 3(\mu_3' + \mu_2') ,$$

$$r_{13} = 2(\mu_2' - \mu_1'^2) ,$$

$$r_{21} = \mu_3' \mu_1' - \mu_2'(\mu_3' + 3 \mu_2' + \mu_1'),$$

$$r_{22} = \mu_1'(2 \mu_3' + 3 \mu_2' + \mu_1') ,$$

$$r_{23} = - \mu_1'(\mu_2' + \mu_1') ,$$

$$r_{31} = \mu_2'(3\,\mu_2' + 3\,\mu_1' + 1) - \mu_3'(2\,\mu_1' + 1),$$

$$r_{32} = \mu_3' - \mu_1'(3\,\mu_2' + 3\,\mu_1' + 1),$$

$$r_{33} = \mu_1'(2\,\mu_1' + 1) - \mu_2'\,,$$

and $|R| = \mu_1'(4\,\mu_3'\,\mu_1' + 3\,\mu_3' - 3\,\mu_2'^2 + \mu_2' - \mu_1'^2) - \mu_2'(\,\mu_3' + 3\,\mu_2')$.

Equation (8.4.5) suggests that we can estimate $\underline{e} = (h, u, p)'$ by substituting the sample moments m_r' for the population moments μ_r' in the r.h.s. expression of (8.4.5). Thus, the estimator $\tilde{\underline{e}}$ of $\underline{e}$ is given by

(8.4.6) $$\tilde{\underline{e}} = (\,\tilde{R}\,)^{-1}\,\tilde{\underline{Q}}\,,$$

wherein $\tilde{R}$ and $\tilde{\underline{Q}}$ denote the values of R and $\underline{Q}$ obtained by replacing the population moments μ_r' by the sample moments m_r' $(r = 1, 2, 3)$.

From the estimators $\tilde{h}$, $\tilde{u}$, and $\tilde{p}$, the estimators of θ, m and r are found to be

(8.4.7) $$\tilde{\theta} = \tilde{p}\,/\,(1 + \tilde{p}),$$

(8.4.8) $$\tilde{m} = (\,\tilde{u} + \tilde{p}\,\tilde{h}\,)/\,\tilde{p}\,(1 + \tilde{p}),$$

(8.4.9) $$\tilde{r} = (\tilde{u} - \tilde{h}) / (1 + \tilde{p}).$$

Following the method given in Section 8.3 for deriving the asymptotic variance-covariance matrix of the estimators $\hat{h}$, $\hat{u}$ and $\hat{p}$, we obtain the asymptotic variance-covariance matrix U of $\underline{\tilde{e}}$ as

(8.4.10) $$U = R^{-1} S(R')^{-1} / n ,$$

where R^{-1} = inverse of the matrix R defined in (8.4.4) and $S = [s_{ij}] = nE(\underline{Q}^* - R^*\underline{e})(\underline{Q}^* - R^*\underline{e})'$. The elements s_{ij} are given by

$$s_{11} = C - u^2,$$

$$s_{12} = s_{21} = E - (h+p)D - u^2,$$

$$s_{13} = s_{31} = H - (h+p)G - u^2,$$

$$s_{22} = F - (h+2p)E - (2u+p)D - u^2,$$

$$s_{23} = s_{32} = I - (h+2p)H - (2u+p)G - u^2,$$

$$s_{33} = J - (h+3p)I - 3(u+p)H - (3u+p)G - u^2,$$

where C, D, E, F have the same meanings as in (8.3.14) and

$$G = \mu_5' - (h+3p)\,\mu_4' - 3(u+p)\,\mu_3' - (3u+p)\,\mu_2' ,$$

$$H = \mu_6' - (h+3p)\,\mu_5' - 3(u+p)\,\mu_4' - (3u+p)\,\mu_3' ,$$

$$I = \mu_7' - (h+3p)\,\mu_6' - 3(u+p)\,\mu_5' - (3u+p)\,\mu_4' ,$$

$$J = \mu_8' - (h+3p)\,\mu_7' - 3(u+p)\,\mu_6' - (3u+p)\,\mu_5' .$$

The variance-covariance matrix N of the estimators $\tilde{\theta}$, $\tilde{m}$, $\tilde{r}$, to the order n^{-1} can be obtained by using (8.4.7), (8.4.8) and (8.4.9) and the δ-method (Kendall and Stuart [23], §10.6) and is given by

(8.4.11) $$N = T' \, U \, T ,$$

where T and U have respectively the same meanings as in (8.3.15) and (8.4.10). The joint asymptotic efficiency E_B, of the estimators $\tilde{\theta}$, $\tilde{m}$, $\tilde{r}$ relative to the maximum likelihood estimators of θ, m, r is given by

(8.4.12) $$E_B = 1 \,/\, |M| . |N| ,$$

where $|M|$ = determinant of the information matrix M defined by (8.2.10).

8.5 Comparison of the joint asymptotic efficiencies of Method A and Method B

For sake of comparison, we evaluate the efficiencies E_A and E_B of Method A and Method B for selected values of θ, m, r. We take $m = 2$, $r = 1$, $p = 1$ and hence $\theta = 1/2$.

(i) Maximum likelihood method. Using the Tables of the digamma function (Eleanor Pairman [16]) and (8.2.10), we get,

$$M = n \begin{bmatrix} 14.22222, & -1.43412, & -0.444444 \\ -1.43412, & 0.153656, & -0.050504 \\ -0.444444, & -0.050534, & 0.017654 \end{bmatrix}$$

and $|M| = n^3 \times 0.0000179125$.

(ii) Method A. Using (8.3.2) and (8.3.3), we obtain the values of the moments as $\mu_1' = 5/3$, $\mu_2' = 19/3$, $\mu_3' = 101/3$, $\mu_4' = 691/3$, $\mu_5' = 5765/3$, $\mu_6' = 56659/3$, $\mu_7' = 640421/3$, $\mu_8' = 8178931/3$.

From (8.3.7), we obtain

$$A = (1/3) \begin{bmatrix} 2, & 1, & 0 \\ 5, & 3, & 5 \\ 19, & 13, & 43 \end{bmatrix}$$

and hence

$$A^{-1} = \begin{bmatrix} 24, & -16.125, & 1.875 \\ -45, & 32.25, & -3.75 \\ 3, & -2.625, & 0.375 \end{bmatrix} .$$

Also, using (8.3.14), we get $C = 121$, $D = 85$, $E = 995$, $F = 11509$ and

$$B = \begin{bmatrix} 20/3, & 18, & 82 \\ 18 \;\;, & 112, & 816 \\ 82 \;\;, & 816, & 7920 \end{bmatrix}$$

and hence

$$V = (1/n) \begin{bmatrix} 4911 \;\;, & -9751.5, & 1033.5 \\ -9751.5, & 19422, & -2056.5 \\ 1033.5, & -2056.5, & 240 \end{bmatrix} .$$

Using (8.3.15), we get

$$T = \begin{bmatrix} 0.5, & -0.5, & 0 \\ 0.5, & 0.5, & 0 \\ -2.5, & -0.5 & 0.25 \end{bmatrix}$$

and hence

$$W = (1/n) \begin{bmatrix} 5265 & 8046 & -277.875 \\ 8046, & 12564 & -416.25 \\ -277.875, & -416.25, & 15 \end{bmatrix}$$

and $|W| = 100419.75/n^3$.

Hence, the joint asymptotic efficiency E_A, of Method A relative to the maximum likelihood method is

$$E_A = 1 \ / \ |M|.|W| = 0.556 \quad (\text{i.e. } 55.6\,\%).$$

(iii) Method B. Using (8.4.4), we get

$$R = (1/3) \begin{bmatrix} 5 & 3 & 5 \\ 19, & 13, & 43 \\ 101, & 75, & 365 \end{bmatrix}$$

and hence

$$R^{-1} = \begin{bmatrix} 11.875, & -5.625, & 0.5 \\ -20.25, & 10.3125, & -0.9375 \\ 0.875, & -0.5625, & 0.0625 \end{bmatrix}.$$

Further, we have from (8.4.10), $G = 533$, $H = 8099$, $I = 112565$, $J = 1626563$, and

$$S = \begin{bmatrix} 112, & 816, & 7024 \\ 816, & 7920, & 84528 \\ 7024, & 84528, & 1073776 \end{bmatrix}$$

and hence

$$U = (1/n) \begin{bmatrix} 33759\ , & -64525.5\ , & 5107.5 \\ -64525.5, & 123405.75, & -9780.75 \\ 5107.5, & -9780.75, & 807.75 \end{bmatrix}.$$

Using (8.4.11), we get

$$N = (1/n) \begin{bmatrix} 23760, & 43200, & -1089 \\ 43200, & 79200, & -1962 \\ -1089, & -1962, & 50.4844 \end{bmatrix}$$

and $|N| = 1195560/n^3$.

Hence, the joint asymptotic efficiency E_B of Method B relative to the maximum likelihood method is

$$E_B = 1/|N|.|M| = 0.047, \text{ (i.e., 4.7 \%).}$$