Ann. Inst. Statist. Math.
Vol. 49, No. 1, 101–115 (1997)

SOME METHODS FOR ESTIMATION IN A NEGATIVE-BINOMIAL MODEL

EIJI NAKASHIMA

Department of Statistics, Radiation Effects Research Foundation,
5-2 Hijiyama Park, Minami-ku, Hiroshima 732, Japan

(Received September 7, 1995; revised March 25, 1996)

Abstract. To clarify the advantage of using the quasilikelihood method, lack of robustness of the maximum likelihood method was demonstrated for the negative-binomial model. Efficiency calculations of the method of moments and the pseudolikelihood method in the estimation of extra-Poisson parameters in a negative-binomial model were carried out. Especially when the overdispersion parameter is small, both methods are relatively highly efficient and the pseudolikelihood estimate is more efficient than the method of moments estimate. Two examples of the quasilikelihood analyses of count data with overdispersion are given. The bootstrap method also is applied to the data to illustrate the advantage of the method of moments or pseudolikelihood method in the estimation of the standard errors of the mean parameter estimates under the negative-binomial model.

Key words and phrases: Method of moments, negative-binomial, overdispersion, pseudolikelihood, quasilikelihood.

1. Introduction

Poisson models are widely used in the regression analysis of count data obtained experimentally (e.g., Fisher (1949)) or epidemiologically (e.g., Frome (1983)). It is common that putatively Poisson data will have overdispersion when the mean is large, since the coefficient of variation for the Poisson distribution with mean μ is $1/\sqrt{\mu}$. A natural model to consider is the negative-binomial, resulting from a compound model where μ has a gamma distribution (e.g., Margolin *et al.* (1981)). However, maximum likelihood methods in this model are inconvenient and may suffer from lack of robustness within the class of models having the same mean-variance relationship as the negative-binomial. There are two quasilikelihood methods that are more convenient and robust, quasilikelihood/method of moments (QL/M) (Breslow (1984)) and quasilikelihood/pseudolikelihood (QL/PL) method (Carroll and Ruppert (1982, 1988)). Here, we study the loss in efficiency under the negative-binomial model when using these in place of maximum likelihood

(ML). Furthermore, in the estimation of standard errors, the bootstrap method is compared to the method of moments and the pseudolikelihood method.

In Section 2, we introduce the negative-binomial maximum likelihood method and the alternative methods, the method of moments and the pseudolikelihood method. In Section 3, we compare the asymptotic relative efficiency of the two methods under a negative-binomial distribution; in Section 4, we give a robustness consideration of the QL/M and QL/PL methods and the asymptotic bias of the negative-binomial ML estimate of the extra-Poisson parameter when underlying distribution is a Poisson-log-normal distribution; in Section 5, we give two examples; and in Section 6, the bootstrap method is applied to one of the example data sets.

2. Estimation for a negative-binomial model

Let Y be the response variable with a Poisson distribution having the mean $\mu(x)$, $Y \sim \text{Poisson}(\mu(x))$. If Y has extra-Poisson variability, a negative-binomial model is often used since it has mathematical flexibility and tractability. If the distribution, given ν and x, is $\text{Poisson}(\nu\mu(x))$, when the distribution of ν is gamma with index $k = \alpha^{-1}$, $f(\nu) = k^k\nu^{k-1}e^{-k\nu}/\Gamma(k)$, the marginal distribution Y is negative-binomial. We denote this as $Y \sim \text{NB}(\mu(x), \alpha)$. Lawless (1987) discussed $\text{NB}(\mu, \alpha)$ in detail.

The negative-binomial likelihood can be written

$$(2.1) \qquad \Pr(Y = y \mid x) = \frac{\Gamma(y + \alpha^{-1})}{y!\Gamma(\alpha^{-1})}\left(\frac{\alpha\mu(x)}{1 + \alpha\mu(x)}\right)^y \left(\frac{1}{1 + \alpha\mu(x)}\right)^{\alpha^{-1}}; \quad y = 0, 1, 2, \ldots .$$

The moment-generating function can be written

$$(2.2) \qquad M(t) = E(e^{tY}) = \{1 + (1 - e^t)\alpha\mu(x)\}^{-\alpha^{-1}}.$$

The mean and variance can be given as

$$(2.3) \qquad \begin{cases} E(Y \mid x) = \mu(x) \\ \text{Var}(Y \mid x) = \mu(x) + \alpha\mu(x)^2. \end{cases}$$

Note, from (2.2), when α tends to zero, the distribution of Y becomes a Poisson distribution with mean μ. In this sense, the parameter α can be called the extra-Poisson parameter or overdispersion parameter of the negative-binomial model.

Let $Y_i \sim \text{NB}(\mu_i, \alpha)$, $i = 1, \ldots, n$ be independent with $\mu_i = \exp(x_i\beta)$, and x_i be the $p \times 1$ covariate vector. The likelihood score equation of β in the negative-binomial model is

$$(2.4) \qquad \frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mu_i(1 + \alpha\mu_i)} \frac{\partial \mu_i}{\partial \beta} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{1 + \alpha\mu_i} x_i = 0,$$

where l is the negative-binomial log-likelihood of (β, α). When α is known, the likelihood score equation of β is both a quasilikelihood equation and weighted least squares equation (McCullagh and Nelder (1989), Breslow (1984)). The likelihood score equation of α in the negative-binomial model is

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \sum_{j=0}^{y_i - 1} \frac{j}{1 + \alpha j} + \alpha^{-2} \log(1 + \alpha\mu_i) - \frac{(y_i + \alpha^{-1})\mu_i}{1 + \alpha\mu_i} \right\} = 0. \tag{2.5}$$

We can estimate the negative-binomial parameters by the Newton-Raphson method using the score equations (2.4) and (2.5) and observed Fisher information matrix (Lawless (1987)).

The Fisher information matrix $I(\beta, \alpha)$ has the entries

$$I_{11}(\beta, \alpha) = \sum_{i=1}^{n} \frac{\mu_i}{1 + \alpha\mu_i} x_i x_i^T, \tag{2.6}$$

$$I_{12}(\beta, \alpha) = I_{21}(\beta, \alpha)^T = 0, \tag{2.7}$$

and

$$I_{22}(\beta, \alpha) = \alpha^{-4} \sum_{i=1}^{n} \left\{ E\left(\sum_{j=0}^{y_i - 1} (\alpha^{-1} + j)^{-2} \right) - \frac{\alpha\mu_i}{\mu_i + \alpha^{-1}} \right\} = i(\beta, \alpha). \tag{2.8}$$

$I_{22}(\beta, \alpha)$ is most easily obtained by rewriting l in terms of β and $k = \alpha^{-1}$, calculating $\partial^2 l / \partial k^2$ and then noting that $E(-\partial^2 l/\partial\alpha^2) = \alpha^{-4} E(-\partial^2 l/\partial k^2)$ (Lawless (1987)). The i-th term of the expectation $I_{22}(\beta, \alpha)$ is equal to

$$\alpha^{-4} \left(\sum_{j=0}^{\infty} (\alpha^{-1} + j)^2 \Pr(Y_i \geq j + 1) - \frac{\alpha\mu_i}{\mu_i + \alpha^{-1}} \right).$$

Assuming $\alpha > 0$ and mild conditions on the x_i's to ensure that $n^{-1} I(\beta, \alpha)$ approaches a positive definite limit as n tends to infinity, $\sqrt{n}(\hat{\beta} - \beta, \hat{\alpha} - \alpha)$ tends to normal distribution under $\mathrm{NB}(\mu, \alpha)$, with mean 0 and variance-covariance matrix

$$nI(\hat{\beta}, \hat{\alpha})^{-1} = n \begin{pmatrix} I_{11}(\hat{\beta}, \hat{\alpha})^{-1} & 0 \\ 0^T & i(\hat{\beta}, \hat{\alpha})^{-1} \end{pmatrix}. \tag{2.9}$$

Note that parameter β and α are orthogonal, i.e., estimates of β and α are asymptotically independent, so the estimation of α has little effect on the estimation of β (Cox and Reid (1987)). We call the variance function, $\mathrm{Var}(Y) = \mu(1 + \alpha\mu)$, the negative-binomial variance function. There is another parameterization of the negative-binomial model, the McCullagh and Nelder (1989) type variance function, i.e., $\mathrm{Var}(Y) = (1 + \alpha)\mu(\beta)$, in which variance function the parameters α and β are not orthogonal (Manton *et al.* (1981)). Further, the score equation for β does not become a quasilikelihood score equation.

The likelihood score equation of α is a little complicated. There are two alternative methods to the maximum likelihood method in the estimation of α. Breslow (1984) suggested that, given estimate $\tilde{\mu}$, we estimate α by solving the moment equation

$$\sum_{i=1}^{n} \frac{(y_i - \tilde{\mu}_i)^2}{V(\tilde{\mu}_i, \alpha)} - (n-p) = \sum_{i=1}^{n} \frac{(y_i - \tilde{\mu}_i)^2}{\tilde{\mu}_i(1+\alpha\tilde{\mu}_i)} - (n-p) = 0, \tag{2.10}$$

where $V(\mu, \alpha) = \mu(1 + \alpha\mu)$. Carroll and Ruppert (1982) suggested another unbiased estimating equation, originally derived from the score equation of the approximate normal log-likelihood, i.e., the pseudolikelihood equation,

$$\sum_{i=1}^{n} \frac{(y_i - \tilde{\mu}_i)^2 - V(\tilde{\mu}_i, \alpha)}{2V(\tilde{\mu}_i, \alpha)^2} \frac{\partial V}{\partial \alpha}(\tilde{\mu}_i, \alpha) = \sum_{i=1}^{n} \frac{(y_i - \tilde{\mu}_i)^2 - \tilde{\mu}_i(1+\alpha\tilde{\mu}_i)}{2(1+\alpha\tilde{\mu}_i)^2} = 0. \tag{2.11}$$

This is an approximate quasilikelihood equation for $(Y-\mu)^2$ by using the relation $E\{(y-\mu)^2\} = V(\mu, \alpha)$ and $\mathrm{Var}\{(y-\mu)^2\} = 2V^2(\beta, \alpha)(1 + \gamma_2/2) \approx 2V^2(\beta, \alpha)$, where γ_2 is the kurtosis of Y. When p is large relative to n, we use an adjusted pseudolikelihood equation (Davidian and Carroll (1987)),

$$\begin{aligned} &\sum_{i=1}^{n} \frac{(y_i - \tilde{\mu}_i)^2 - h_i^* V(\tilde{\mu}_i, \alpha)}{2V(\tilde{\mu}_i, \alpha)^2} \frac{\partial V}{\partial \alpha}(\tilde{\mu}_i, \alpha) \\ &\quad = \sum_{i=1}^{n} \frac{(y_i - \tilde{\mu}_i)^2 - h_i^* \tilde{\mu}_i(1-\alpha\tilde{\mu}_i)}{2(1+\alpha\tilde{\mu}_i)^2} = 0, \end{aligned} \tag{2.12}$$

where $h_i^* = 1 - h_i$ and h_i is a i-th diagonal element of $Q(Q^TQ)^{-1}Q^T$, the leverage of the i-th data point, where Q is an $n \times p$ matrix with i-th row $(V_i^{-1/2})(\partial\mu_i/\partial\beta^T)$.

To obtain the estimates of β and α, we use the quasilikelihood equation $\partial l/\partial\beta = 0$ and the moments equation (2.10) or the pseudolikelihood equation (2.11). These estimates are obtained by solving $\partial l/\partial\beta = 0$ after which the equation (2.10) or (2.11) is solved. These procedures are iterated back and forth to obtain the final estimates. We call each of these procedures the QL/M method and the QL/PL method. Both estimates are consistent estimates. Asymptotics of the estimates obtained using the QL/M and QL/PL methods are given in Appendix.

3. Efficiency comparison of the method of moments and the pseudolikelihood method

The estimate $\tilde{\beta}$ obtained from the QL/M or QL/PL equations is asymptotically equivalent to the maximum likelihood estimator $\hat{\beta}$, and its asymptotic covariance matrix is consistently estimated to be $I_{11}(\tilde{\beta}, \tilde{\alpha})^{-1}$. There is, however, some loss of efficiency in the estimation of α using $\tilde{\alpha}_M$ and $\tilde{\alpha}_{PL}$, which are the solutions of the method of moments equation (2.10) and the pseudolikelihood equation (2.11), respectively, compared to the maximum likelihood estimate $\hat{\alpha}$. The

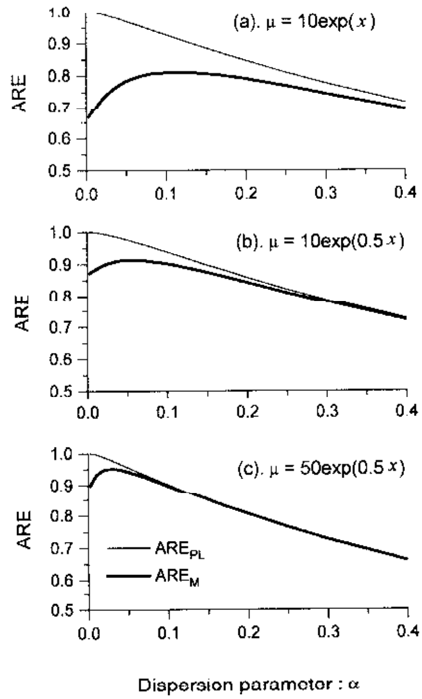

Fig. 1. Asymptotic relative efficiency for the method of moments (ARE_M) and the pseudolikelihood method (ARE_{PL}) in the case of (a). $\mu_i = 10 \cdot \exp(x_i)$, (b). $\mu_i = 10 \cdot \exp(0.5 \cdot x_i)$, and (c). $\mu_i = 50 \cdot \exp(0.5 \cdot x_i)$ for $x_i = -1/3$, 0 and 1/3. The thick line is the method of moments and the thin line is the pseudolikelihood method.

asymptotic relative efficiencies (ARE) of $\tilde{\alpha}_M$ to $\hat{\alpha}$ and $\tilde{\alpha}_{PL}$ to $\hat{\alpha}$ are given by

$$ARE_M = \frac{asvar(\sqrt{n}(\hat{\alpha} - \alpha))}{asvar(\sqrt{n}(\tilde{\alpha}_M - \alpha))} = \frac{i^*(\beta, \alpha)^{-1}}{b_{M,p+1}^{-2}(c_{M,p+1} - b_M^T I_1^*(\beta, \alpha)^{-1} b_M)}, \tag{3.1}$$

and

$$ARE_{PL} = \frac{asvar(\sqrt{n}(\hat{\alpha} - \alpha))}{asvar(\sqrt{n}(\tilde{\alpha}_{PL} - \alpha))} = \frac{i^*(\beta, \alpha)^{-1}}{b_{PL,p+1}^{-2}(c_{PL,p+1} - b_{PL}^T I_1^*(\beta, \alpha)^{-1} b_{PL})}, \tag{3.2}$$

where $i^*(\beta, \alpha) = \lim_{n\to\infty}(1/n)i(\beta, \alpha)$, and b_M, $b_{M,p+1}$, $c_{M,p+1}$, b_{PL}, $b_{PL,p+1}$ and $c_{PL,p+1}$ are given in Appendix. Detail of the calculation of the ARE are shown in Appendix.

Three cases are considered for ARE_M and ARE_{PL}, $p = 2$, $\mu_i = \exp(\beta_0 + \beta_1 x_i)$ and one-third of the x_i's each of -1, 0, 1: (a) $\exp(\beta_0) = 10$, $\beta_1 = 1$, (b) $\exp(\beta_0) = 10$, $\beta_1 = 0.5$, and (c) $\exp(\beta_0) = 50$, $\beta_1 = 0.5$. The results are shown in Fig. 1.

Similar comparisons have been made in Dean *et al.* (1989). If α is relatively large, e.g., 0.3, the efficiency loss of the QL/M and QL/PL methods are severe. The pseudolikelihood method is always preferred to the method of moments especially for a small value of α. Further, by calculation of ARE_{PL}/ARE_M, which is easily calculated, we found that the pseudolikelihood method is superior to the method of moments, especially in small β_0 and in large β_1. Note that if no covariate exists, i.e., iid case, then $ARE_M = ARE_{PL}$.

4. Robustness consideration

In general, the distribution with the mean-variance relationship (2.3) can be derived as follows: when ν is a positive random variable with mean 1 and variance α, assume that given ν, the distribution of Y is Poisson($\nu\mu$), then $E(Y) = \mu$ and $\text{Var}(Y) = E(\text{Var}(Y \mid \nu)) + \text{Var}(E(Y \mid \nu)) = E(\nu\mu) + \text{Var}(\nu\mu) = \mu + \alpha\mu^2$. Therefore, the marginal distribution of Y has a mean-variance relationship of (2.3). In some special cases, if ν is a gamma distribution then Y is distributed as a NB(μ, α), as stated in Section 2; if ν is an inverse-Gaussian, then Y is distributed as a Poisson-inverse Gaussian (P-IG) (Dean *et al.* (1989)). If ν is a log-normal, i.e., $\log(\nu)$ is distributed as a normal with mean $-\log(1+\alpha)/2$ and variance $\log(1+\alpha)$, then Y is distributed as a Poisson-log-normal (P-LN) (Hinde (1982)), where the P-LN distribution has a heavier tail than the P-IG distribution with relation (2.3). Even in these more general situations with relation (2.3), the estimate obtained using the method of moments, $\tilde{\alpha}_M$, and the estimate obtained using the pseudolikelihood method, $\tilde{\alpha}_{PL}$, are consistent estimates of α, and $\sqrt{n}(\tilde{\alpha}_M - \alpha)$ and $\sqrt{n}(\tilde{\alpha}_{PL} - \alpha)$ have asymptotic normal distributions. Further, by similar arguments as noted in Appendix, the pseudolikelihood estimate is more efficient than the estimate obtained using the method of moments.

We simulated the asymptotic bias of the solution of negative-binomial likelihood score equation for α, equation (2.5). Lawless (1987) showed that the negative-binomial ML estimate of α results in a negative asymptotic bias under P-IG distribution which has a heavier tail than the NB distribution. We assume $Y \sim P\text{-}LN$ with mean-variance relationship (2.3). Then, from White (1982), the solution of the score equation (2.5), $\hat{\alpha}$ tends to α_1^* which satisfies

$$E_{P\text{-}LN}\left(\frac{\partial l(\beta, \alpha_1^*)}{\partial \alpha_1^*}\right) = 0,$$

where $E_{P\text{-}LN}(\cdot)$ means the expectation under Poisson-log-normal with mean-variance relationship (2.3). Table 1 shows the α_1^*'s for various α's. The simulation suggests that the negative-binomial ML estimate of α results in a negative asymptotic bias under the P-LN distribution.

The above argument suggests the robustness of the method of moments and the pseudolikelihood method. In this sense, both methods can be called semi-parametric methods. The negative-binomial maximum likelihood method is not robust if the underlying distribution is not negative-binomial even if relation (2.3) holds, and the variance estimate of $\hat{\beta}$ would be biased. The method of moments and the pseudolikelihood method give a consistent variance estimate of $\hat{\beta}$ as long

Table 1. Asymptotic bias of the solution of the ML equation (2.5) under Poisson log normal (μ, α).

μ	α	α_1^*	e^{β_0} a)	α	α_1^*
5	0.5	0.422	10	0.2	0.185
5	0.1	0.0980	10	0.1	0.0969
20	0.1	0.0960	10	0.01	0.00999
20	0.01	0.00999	20	0.2	0.182
50	0.1	0.0949	20	0.1	0.0959
50	0.01	0.00998	20	0.01	0.00999

a) $\mu = \exp(\beta_0 + x)$ where $x = -1/3,\ 0,\ 1/3$.

as relation (2.3) holds. Furthermore, the pseudolikelihood variance estimate is more efficient than the variance estimate from the method of moments as shown in the previous section.

5. Examples

Two sets of data which are employed by Breslow (1984) will be examined to illustrate some points discussed above.

Example 1. (Ames salmonella assay data) Margolin *et al.* (1981) presented data from the Ames salmonella reverse mutagenicity assay. We will work with an approximation to Margolin *et al.*'s "single hit" model which is considered by Breslow (1984):

$$E(Y_i \mid x_i) = \mu_i = \exp\{\beta_0 + \beta_1 x_i + \beta_2 \log(x_i + 10)\}.$$

We did score tests for the variance functions; for a McCullagh and Nelder (1989) type variance function, i.e., $V = (1+\alpha)\mu$, the adjusted score test statistic for small sample data set, which is the test P_C' in Dean ((1992), p. 453), was 6.16 and, for a negative-binomial variance function, the adjusted score test statistic, which is the test P_B' in Dean ((1992), p. 453), was 6.24. Test of H: $\beta_2 = 0$ is of special interest, with $\beta_2 > 0$ representing the mutagenic effect. Fitting $Y_i \sim \mathrm{NB}(\mu_i, \alpha)$ using maximum likelihood, using QL/M and using QL/PL with adjustment (2.12), yields the estimates (standard errors)

$$\hat\alpha = 0.0488(0.0275), \quad \hat\beta_0 = 2.198(0.321),$$
$$\hat\beta_1 = -0.000980(0.000381), \quad \hat\beta_2 = 0.313(0.0868);$$
$$\tilde\alpha_M = 0.0718(0.0303), \quad \tilde\beta_{M0} = 2.203(0.359),$$
$$\tilde\beta_{M1} = -0.000974(0.000430), \quad \tilde\beta_{M2} = 0.311(0.0974);$$
$$\tilde\alpha_{PL} = 0.0533(0.0323), \quad \tilde\beta_{PL0} = 2.199(0.333),$$
$$\tilde\beta_{PL1} = -0.000979(0.000397), \quad \tilde\beta_{PL2} = 0.312(0.0902).$$

These standard errors (s.e.'s) are model-based s.e.'s. The Huber-White "sandwich" robust s.e.'s (e.g., White (1982), Liang and Zeger (1986)) for QL/PL analysis were

Table 2. Estimates and standard errors (s.e.) of parameter estimates for prostate cancer death data.

Parameters Estimated	Method of analysis			
	Poisson ML[a)]	NB ML[b)]	QL/M[c)]	QL/PL[d)]
Grand mean	−8.625	−8.642	−8.645	−8.642
Age group (y)				
50–54	(0.0)	(0.0)	(0.0)	(0.0)
55–59	0.819(0.029)	0.822(0.039)	0.823(0.045)	0.822(0.040)
60–64	1.553(0.028)	1.549(0.039)	1.549(0.045)	1.549(0.040)
65–69	2.133(0.028)	2.128(0.039)	2.128(0.046)	2.128(0.041)
70–74	2.687(0.028)	2.695(0.041)	2.696(0.048)	2.695(0.042)
75–79	3.134(0.029)	3.166(0.043)	3.172(0.050)	3.168(0.044)
80–84	3.455(0.031)	3.472(0.046)	3.474(0.053)	3.472(0.047)
Birth cohort				
1855–59	(0.0)	(0.0)	(0.0)	(0.0)
1860–64	0.365(0.109)	0.357(0.123)	0.355(0.133)	0.356(0.125)
1865–69	0.523(0.102)	0.520(0.116)	0.519(0.125)	0.520(0.118)
1870–74	0.773(0.100)	0.775(0.113)	0.774(0.123)	0.775(0.115)
1875–79	1.010(0.099)	1.012(0.112)	1.012(0.121)	1.012(0.114)
1880–84	1.151(0.098)	1.151(0.112)	1.151(0.121)	1.151(0.114)
1885–89	1.307(0.098)	1.301(0.112)	1.299(0.120)	1.301(0.113)
1890–94	1.510(0.099)	1.541(0.113)	1.546(0.122)	1.542(0.115)
1895–99	1.553(0.099)	1.572(0.114)	1.575(0.123)	1.573(0.116)
1900–04	1.592(0.099)	1.623(0.115)	1.628(0.125)	1.624(0.117)
1905–09	1.461(0.101)	1.464(0.117)	1.464(0.128)	1.464(0.120)
1910–14	1.369(0.104)	1.373(0.122)	1.372(0.133)	1.373(0.124)
1915–19	1.237(0.116)	1.253(0.137)	1.256(0.150)	1.254(0.140)
α	0.00	0.00209	0.00362	0.00239
s.e. of α	(—)	(0.000680)	(0.00108)	(0.000929)
DF	Goodness of fit chi-square			
30	127.69	43.78	30.00	40.06

[a)] Poisson maximum likelihood, [b)] negative-binomial maximum likelihood, [c)] quasilikelihood/method of moments, [d)] quasilikelihood/pseudolikelihood method.

0.318, 0.000428 and 0.0890 for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively, which are similar to the model based s.e.'s though the robust s.e.'s are not efficient (Firth (1992)). The estimates of β_0, β_1, and β_2 do not change drastically, however, the estimate of α in the pseudolikelihood method gives an intermediate value between the other two methods. For example, the values of the Wald ratio for β_2 of each of the maximum likelihood methods, the QL/M, and the QL/PL with small sample adjustment are

3.61, 3.19 and 3.46, in which the adjusted pseudolikelihood method also gives the intermediate value between the other two methods.

Example 2. (Prostate cancer death data) Holford (1983) presented data from prostate cancer deaths and a midperiod population for nonwhites in the US by age and calendar period. Epidemiological data with a large population are usually fit by a Poisson model. If there is overdispersion, the quasilikelihood method using the McCullagh and Nelder type variance function or the quasi-likelihood method using a negative-binomial variance function are usually used (Breslow (1984)). This is because, in general, epidemiological data are not well controlled, therefore a robust method is appropriate for the fit. First, we fit the full model, the *age* + *period* + *birth cohort* model, with the midperiod population as an offset. The adjusted score test (Dean (1992), p. 453) for overdispersion with the McCullagh and Nelder type variance function was 14.58 and that for the negative-binomial variance function was 14.17, by which a severe overdispersion was suggested. The period effect was not significant (Breslow (1984)). So, we fit the *age* + *birth cohort* model with a negative-binomial variance function using the maximum likelihood method, QL/M, and QL/PL methods with a small sample adjustment, equation (2.12). The estimated parameters of α (standard errors) are $\hat{\alpha} = 0.00209(0.000680)$, $\tilde{\alpha}_M = 0.00362(0.00108)$ and $\tilde{\alpha}_{PL} = 0.00239(0.000929)$. The pseudolikelihood method gives an intermediate value of the α estimate between the other two methods. Parameter estimates from the Poisson maximum likelihood method, the negative-binomial maximum likelihood method, the QL/M method, and the QL/PL method with small sample adjustment are given in Table 2. Parameter estimates from the four methods are consistent and similar, however, the standard errors of the Poisson fit are underestimated since the Poisson fit did not consider the overdispersion.

6. Comparison to the bootstrap method

Another method of analysis of the overdispersed count data is the jackknife or nonparametric bootstrap method (e.g., Efron and Tibshirani (1993)), which was suggested by a referee. Poisson regression parameter estimates using all data is used for the parameter estimates. The Huber-White "sandwich" variance of this estimate under this misspecified model (e.g., White (1982)) is

$$\left(\sum_{i=1}^{n} \mu_i x_i x_i^T\right)^{-1} \left(\sum_{i=1}^{n} \mathrm{Var}(Y_i) x_i x_i^T\right) \left(\sum_{i=1}^{n} \mu_i x_i x_i^T\right)^{-1}, \tag{6.1}$$

where x_i is the covariate vector and $\mu_i = E(Y_i)$, Poisson variance. For estimating the standard errors (s.e.'s) of the Poisson parameter estimates, i.e., the s.e.'s from the robust variance (6.1), the bootstrap s.e.'s are calculated from the 4,000 bootstrap Poisson parameter estimates by the 4,000 same sized resamplings from the original data set. Note that 4,000 is enough to produce stable bootstrap standard errors. This method is also a semi-parametric method that avoids the variance function estimation. We show the results applying this method to Example 1.

The result of Poisson regression parameter estimates (nonparametric bootstrap s.e.'s) are $\hat{\beta}_0 = 2.173(0.351)$, $\hat{\beta}_1 = -0.00101(0.000643)$, $\hat{\beta}_2 = 0.320(0.100)$ where the Wald test statistic for β_2 is 3.20. The bootstrap biases are -0.0345, -0.0000457 and 0.00761 for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively. Since the bootstrap biases are small relative to the parameter estimates, we used the estimates without bias adjustment. The jackknife method also can be considered but the jackknife s.e. is less efficient than the bootstrap s.e. in nonlinear problems (Efron and Tibshirani (1993), Chap. 11). The Huber-White "sandwich" robust estimate for s.e. (White (1982)) also can be considered but is not efficient (Firth (1992)) since this estimate is obtained by replacing $\mathrm{Var}(Y_i)$ by a single squared residual $(Y_i - \hat{\mu}_i)^2$ in (6.1). Thus, we used the bootstrap method.

The Poisson parameter estimates are consistent and do not differ greatly from the QL/PL method. Comparison of the bootstrap s.e.'s and the model-based s.e.'s from the QL/PL method show that the bootstrap method gives larger s.e. estimates. These results can be expected since the Poisson regression probably uses less appropriate weights than the QL/M and QL/PL methods under this extra-Poisson situation, i.e., $\mathrm{Var}(Y) = \mu(1 + \alpha\mu)$, hence would produce less efficient parameter estimates than the quasilikelihood methods. That is, asymptotically, the bootstrap s.e.'s for β are larger than the s.e.'s from the quasilikelihood methods, i.e., if underlying distribution is $\mathrm{NB}(\mu, \alpha)$, then, asymptotically, we have the inequality,

$$
\begin{aligned}
(6.2) \quad asvar(\sqrt{n}(\tilde{\beta}_B - \beta)) &= \lim_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} \mu_i x_i x_i^T\right)^{-1} \\
&\quad \cdot \left\{\frac{1}{n}\sum_{i=1}^{n} \mu_i(1 + \alpha\mu_i) x_i x_i^T\right\} \left(\frac{1}{n}\sum_{i=1}^{n} \mu_i x_i x_i^T\right)^{-1} \\
&\geq \lim_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mu_i}{1 + \alpha\mu_i} x_i x_i^T\right)^{-1} \\
&= asvar(\sqrt{n}(\tilde{\beta} - \beta)),
\end{aligned}
$$

where $\tilde{\beta}_B$ is the bootstrap estimate for β, i.e., Poisson estimate for β, $\tilde{\beta}$ is ML, QL/M or QL/PL estimate for β and the inequality is in the sense of positive semi-definite. The bootstrap method results in conservative Wald test statistics. Thus, the estimation of the correct variance function contributes to the increase of the efficiency of parameter estimates for β. Note if the variance function is a McCullagh and Nelder type variance function, i.e., $\mathrm{Var}(Y) = (1 + \alpha)\mu$, then the Poisson parameter estimate, $\tilde{\beta}_B$, is as efficient as the quasilikelihood estimate for β, since, in this case, the variance estimate (6.1) gives the same variance estimate as the quasilikelihood estimate.

7. Discussion

The coefficient of variation (CV) of the Poisson distribution is $1/\sqrt{\mu}$, and the CV of the distribution with relation (2.3) is $\sqrt{\alpha + 1/\mu}$. The CV of the Poisson distribution decreases as the mean increases. Although the effect of the non-Poisson source of variability, α, will preclude the CV being extremely small, this situation might occur when μ is large. When the value of α is about 0.3, the efficiencies of the QL/M and QL/PL methods are low, then the maximum likelihood method would be recommended though it is less robust. The three cases are considered in the calculation of asymptotic relative efficiency. Cases (a) and (b) correspond to experimental data, e.g., Margolin *et al.*'s data and case (c), which has a larger mean, corresponds to epidemiological data, e.g., Holford's cancer epidemiological data. The two examples show that when the value of α is small, the use of the pseudolikelihood method is advocated although the advantage of using the pseudolikelihood for data with large means is small.

The small sample size of Example 1 would result in some bias in the α and β estimates from ML, QL/M, and QL/PL. To evaluate this bias is, however, a future problem. Note the bias of the estimate α has little effect on the estimate of β since both parameters are orthogonal (Cox and Reid (1987)). The uncertainty of the true distribution of the example also weakens our reasoning in Section 6. For example, someone might say that the bootstrap estimates and the QL estimates for s.e.'s are similar. This is because the example data does not discriminate between the two variance functions, the McCullagh and Nelder type variance function and the NB type variance function, as shown by score test for overdispersion. This phenomenon is inevitable for small sample data. So if we want to avoid this weakness, we should generate a moderate to large sized example data set by NB(μ, α) distribution.

To extra-binomial data, the two types of variance function also can be applied, beta-binomial type variance function and McCullagh and Nelder type variance function (Liang and McCullagh (1993)). In the beta-binomial type variance function, the dispersion parameter also would be more efficiently estimated by the pseudolikelihood method than by the method of moments as seen in negative-binomial type variance function. Further, although the method of moments allows the estimation of only a single parameter, the pseudolikelihood method allows the estimation of more than one parameter. Liang and McCullagh (1993) investigated the appropriateness of these two types of variance function in five extra-binomial data examples. They concluded that, from the finite sample, they can not distinguish which variance function gives a better fit to the data. In the two examples of the negative-binomial applications in this paper, we also could not conclude which variance function fits better to the data between two variance functions, i.e., the McCullagh and Nelder type variance function and our negative-binomial variance function.

Regarding the adequacy of the negative-binomial assumption for a data, the following can be suggested. If there are some candidate distributions for the data, one can use the AIC criterion (Akaike (1973)). However, the basic assumption of the AIC is that the model includes the true model, which is generally unknown.

Since the actual data set has a finite number of data points, it would be difficult to specify the distribution of that data. Instead, the mean-variance relationship can be specified. Then we can apply the quasilikelihood method and can use the recently proposed AIC_c criterion (Hurvich and Tsai (1995)) for the selection of extended quasilikelihood models in small sample cases. Note the quasilikelihood method does not deteriorate the efficiency of the estimation (Firth (1987)). For estimating the variance function, the graphical method can be applied. Carroll and Ruppert (1988) suggested the use of the absolute, log-absolute, and cube root of the squared residuals. Lambert and Roeder (1995) proposed the convexity plot or C-plot to detect overdispersion, and the relative variance curve and the relative variance test to give an approximate mean-variance relation for the random-coefficient GLM. Hence, several techniques are applicable to our data sets. However, because of the small sample size of our data sets, the graphical methods would fail to suggest an appropriate variance function. Hence we only show the results of score tests for overdispersion (Dean (1992)), since our example data sets have been analyzed before using $\mathrm{NB}(\mu, \alpha)$.

Dean (1994) has done small sample simulation studies and has given seven methods of estimation for overdispersion parameter α in $\mathrm{NB}(\mu, \alpha)$ including the maximum likelihood method. Under condition of the mean-variance relation (2.3), the method of moments, the pseudolikelihood method, and the optimal quadratic estimating equation give consistent estimates for α, i.e., they are robust. Specifying the third and fourth moments would give an additional efficiency gain to the estimate from the optimal quadratic estimating equation approach, however, the specification of the higher order moments is an obvious disadvantage of this approach. Other method of estimation in the paper are not robust under relation (2.3). The extended quasilikelihood estimator, EQL, in the paper do not give a consistent estimate (Davidian and Carroll (1988)). NB and PIG estimators in the paper also would give inconsistent estimate under relation (2.3) as shown in Section 4. Dean (1994) suggested that the modified pseudolikelihood estimator, MPL, performs well, i.e., low bias and high efficiency, under a small α, even in small samples. However, the MPL estimator is not good nor consistent unless the approximation of the Anscombe residual is good. The approximation of the Anscombe residual is good for moderate to small values of $\alpha\mu$. So there is a possibility that for large values of μ the MPL estimator is no longer good even in cases of a small α.

Acknowledgements

The author thanks the referees for their constructive comments. This publication is based on research performed at Radiation Effect Research Foundation (RERF), Hiroshima and Nagasaki, Japan. RERF is a private nonprofit foundation funded equally by the Japanese Ministry of Health and Welfare and United States Department of Energy through the National Academy of Sciences.

Appendix. Asymptotic variance of the QL/M and QL/PL estimates

This appendix is largely after the Appendix A in Lawless (1987). We will explain here the derivations of the ARE formulae (3.1) and (3.2) for the method of moments and pseudolikelihood method, respectively. We assume that $Y_i \sim \mathrm{NB}(\mu_i, \alpha)$ with $\mu_i = \exp(x_i\beta)$, and we will consider the likelihood score equation for β, which is a quasilikelihood score equation, the method of moments equation for α, and the pseudolikelihood estimating equation for α

$$U_{QL}(\beta,\alpha) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{1+\alpha\mu_i} x_i = 0,$$

$$U_M(\beta,\alpha) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i(1+\alpha\mu_i)} - n = 0,$$

and

$$U_{PL}(\beta,\alpha) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2 - \mu_i(1+\alpha\mu_i)}{2(1+\alpha\mu_i)^2} = 0,$$

which, when p is fixed, are asymptotically equivalent to (2.4), (2.10) and (2.11), respectively, as n tends to infinity. Write that $\theta = (\beta, \alpha)$. Then from the results of Inagaki (1973), the estimator $\tilde{\theta} = (\tilde{\beta}, \tilde{\alpha})$ obtained solving the equations above, under conditions similar to those for which standard maximum likelihood asymptotics hold, is consistent and asymptotically normal with the covariance matrix

$$\text{(A.1)} \qquad asvar(\sqrt{n}(\tilde{\theta} - \theta)) = A(\theta)^{-1} B(\theta) (A(\theta)^{-1})^T,$$

where $A(\theta)$ and $B(\theta)$ are $(p+1) \times (p+1)$ matrices with respective entries

$$A(\theta) = \lim_{n\to\infty} \frac{1}{n} E\left(-\frac{\partial U}{\partial \theta^T}\right), \quad \text{and} \quad B(\theta) = \lim_{n\to\infty} \frac{1}{n} E(UU^T),$$

where $U^T = (U_{QL}^T, U_M)$ or $U^T = (U_{QL}^T, U_{PL})$. It follows after some algebra that $A(\theta)$ and $B(\theta)$ are the limits, respectively, of

$$A_n(\theta) = \begin{pmatrix} n^{-1} I_{11}(\beta,\alpha) & 0 \\ b^T & b_{p+1} \end{pmatrix}, \quad \text{and} \quad B_n(\theta) = \begin{pmatrix} n^{-1} I_{11}(\beta,\alpha) & c \\ c^T & c_{p+1} \end{pmatrix},$$

where $I_{11}(\beta,\alpha)$ is as in (2.9) and where, under $\mathrm{NB}(\mu,\alpha)$, we have $b = c$ for the method of moments and pseudolikelihood methods, respectively. Inverting $A(\theta)$ and using (A.1), we get

$$asvar(\sqrt{n}(\tilde{\beta} - \beta, \tilde{\alpha} - \alpha)) = \begin{pmatrix} I_1^*(\beta,\alpha)^{-1} & 0 \\ 0^T & \dfrac{1}{b_{p+1}^2}(c_{p+1} - b^T I_1^*(\beta,\alpha)^{-1} b) \end{pmatrix},$$

where $I_1^*(\beta,\alpha) = \lim_{n\to\infty}(1/n) I_{11}(\beta,\alpha)$.

Therefore, the estimates, $\tilde{\beta}$ and $\tilde{\alpha}_M$ or $\tilde{\alpha}_{PL}$, which are from the moments equation (2.10) and pseudolikelihood equation (2.11), respectively, are consistent and asymptotically independent and normally distributed, with

$$asvar(\sqrt{n}(\tilde{\beta}-\beta)) = I_1^*(\beta,\alpha)^{-1},$$
$$asvar(\sqrt{n}(\tilde{\alpha}_M-\alpha)) = b_{M,p+1}^{-2}(c_{M,p+1} - b_M^T I_1^*(\beta,\alpha)^{-1} b_M),$$

and

$$asvar(\sqrt{n}(\tilde{\alpha}_{PL}-\alpha)) = b_{PL,p+1}^{-2}(c_{PL,p+1} - b_{PL}^T I_1^*(\beta,\alpha)^{-1} b_{PL}),$$

where $I_1^*(\beta,\alpha) = \lim_{n\to\infty} n^{-1} I_{11}(\beta,\alpha)$; b_M and b_{PL} are $p\times 1$ covariate vector; $b_{M,p+1}$, $c_{M,p+1}$, $b_{PL,p+1}$, and $c_{PL,p+1}$ are scalars. More precisely, noting that when $Y_i \sim \mathrm{NB}(\mu_i,\alpha)$, we have $E(Y_i-\mu_i) = 0$, $E\{(Y_i-\mu_i)^2\} = \mu_i(1+\alpha\mu_i)$, $E\{(Y_i-\mu_i)^3\} = \mu_i(1+\alpha\mu_i)(1+2\alpha\mu_i)$, and $E\{(Y_i-\mu_i)^4\} = \mu_i(1+\alpha\mu_i)(1+3\mu_i + 6\alpha\mu_i + 3\alpha\mu_i^2 + 6\alpha^2\mu_i^2)$ from the moment-generating function (2.2) of a negative-binomial model, then we get

$$b_M = \lim_{n\to\infty}\frac{1}{n}E\left(-\frac{\partial U_M}{\partial\beta}\right) = \lim_{n\to\infty}\frac{1}{n}E(U_M U_{QL}) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\frac{1+2\alpha\mu_i}{1+\alpha\mu_i}\right)x_i,$$

$$b_{M,p+1} = \lim_{n\to\infty}\frac{1}{n}E\left(-\frac{\partial U_M}{\partial\alpha}\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\frac{\mu_i}{1+\alpha\mu_i},$$

$$c_{M,p+1} = \lim_{n\to\infty}\frac{1}{n}E(U_M U_M^T) = 2+6\alpha+\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\frac{1}{\mu_i(1+\alpha\mu_i)},$$

$$b_{PL} = \lim_{n\to\infty}\frac{1}{n}E\left(-\frac{\partial U_{PL}}{\partial\beta}\right) = \lim_{n\to\infty}\frac{1}{n}E(U_{PL}U_{QL})$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\frac{1+2\alpha\mu_i}{2(1+\alpha\mu_i)^2}\mu_i x_i,$$

$$b_{PL,p+1} = \lim_{n\to\infty}\frac{1}{n}E\left(-\frac{\partial U_{PL}}{\partial\alpha}\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\frac{\mu_i^2}{2(1+\alpha\mu_i)^2},$$

and

$$c_{PL,p+1} = \lim_{n\to\infty}\frac{1}{n}E(U_{PL}U_{PL}^T) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left\{\frac{\mu_i}{4(1+\alpha\mu_i)^3} + \frac{(1+3\alpha)\mu_i^2}{2(1+\alpha\mu_i)^2}\right\}.$$

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267–281, Akademia Kiado, Budapest.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models, *Appl. Statist.*, **33**, 38–44.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models., *Ann. Statist.*, **10**, 429–441.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Chapman and Hall, London.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation, *J. Amer. Statist. Assoc.*, **82**, 1033–1048.

Davidian, M. and Carroll, R. J. (1988). A note on extended quasi-likelihood, *J. Roy. Statist. Soc. Ser. B*, **50**, 74–82.

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models, *J. Amer. Statist. Assoc.*, **87**, 451–457.

Dean, C. B. (1994). Modified pseudo-likelihood estimator of the overdispersion parameter in Poisson mixture models, *J. Appl. Statist.*, **21**, 523–532.

Dean, C. B., Lawless, J. F. and Willmot, G. E. (1989). A mixed Poisson-inverse-Gaussian regression model, *Canad. J. Statist.*, **17**, 171–181.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London.

Firth, D. (1987). On the efficiency of quasi-likelihood estimation, *Biometrika*, **39**, 665–674.

Firth, D. (1992). Discussion of the paper: Multivariate regression analysis for categorical data (by Liang, K.-Y., Zeger, S. L. and Qaqish, B.), *J. Roy. Statist. Soc. Ser. B*, **54**, 3–40.

Fisher, R. A. (1949). A biological assay of tuberculin, *Biometrics*, **5**, 300–316.

Frome, E. L. (1983). The analysis of rates using Poisson regression models, *Biometrics*, **39**, 665–674.

Hinde, J. (1982). Compound Poisson regression model, *GLIM82: Proc. Internat. Conf. Generalized Linear Models* (ed. R. Gilchrist), 109–121, Springer, Berlin.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates, *Biometrics*, **39**, 311–324.

Hurvich, C. M. and Tsai, C.-H. (1995). Model selection for extended quasi-likelihood models in small samples, *Biometrics*, **51**, 1077–1084.

Inagaki, N. (1973). Asymptotic relation between the likelihood estimating function and the maximum likelihood estimator, *Ann. Inst. Statist. Math.*, **25**, 1–26.

Lambert, D. and Roeder, K. (1995). Overdispersion diagnostics for generalized linear models, *J. Amer. Statist. Assoc.*, **90**, 1225–1236.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression, *Canad. J. Statist.*, **15**, 209–225.

Liang, K.-Y. and McCullagh, P. (1993). Case studies in binary dispersion, *Biometrics*, **49**, 623–630.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

Manton, K. G., Woodbury, M. A. and Stallard, E. (1981). A variance components approach to categorical data models with heterogeneous populations: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties, *Biometrics*, **37**, 259–269.

Margolin, B. H., Kaplan, N. and Zeiger, E. (1981). Statistical analysis of the Ames salmonella/microsome test, *Proc. Nat. Acad. Sci. U.S.A.*, **76**, 3779–3783.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–25.