Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.4 : 2024
ISSN : **1906-9685**

ASYMPTOTIC COMPARISON OF METHOD OF MOMENT ESTIMATORS AND MAXIMUM LIKELIHOOD ESTIMATORS OF PARAMETERS IN ZERO-INFLATED NEGATIVE BINOMIAL DISTRIBUTION: USING EM ALGORITHM

Dr. Raveendra Naika T Assoc. Professor, Dept. of Statistics, Maharani's Science College for Women, Bengaluru, raviisec@yahoo.co.uk
Dr. Shivanna B K Assoc. Professor, Dept. of Statistics, Maharani's Science College for Women, Mysuru. shivustasts@gmail.com
Dr. Srinivasa A T Assoc. Professor, Dept. of Statistics, BLDE Association Smt. Bangaramma Sajjan Arts Commerce and Science College for Women, Vijayapura, srinivasaat1974@gmail.com

ABSTRACT
This paper discusses the estimation of parameters in the zero-inflated Negative Binomial (ZINB) model by the method of moments and maximum likelihood estimation. The method of moments estimators (MMEs) are analytically compared with the maximum likelihood estimators (MLEs).

Keywords: Zero-Inflated Negative Binomial Model; Maximum Likelihood and Moment Estimators; EM Algorithm; Two parameter exponential family; fisher information matrix; Asymptotic Relative Efficiency

1. Introduction
In the recent period, enormous research activity has been observed in generalizing of the standard discrete distribution. The important idea was to apply the comprehensive versions for modeling different kinds of dependent count data structure in various fields of traffic, insurance, health, engineering, medical, public, manufacturing, road safety, epidemiology, sociology, economic, agriculture, etc.,

The ZINB model is introduced in this section in the context of a practical situation. Maximum likelihood estimation of the parameters involved in the model is discussed in Section 2. The ZINB model is shown to be a member of the two-parameter exponential family in section 3 and hence the asymptotic normality of the MMEs is established in section 4. Further, in Section 5, the details of computing the Fisher information matrix corresponding to this model are shown. In Section 6 the MMEs and the MLEs are asymptotically compared.

A random variable X is said to have a zero-inflated Negative Binomial distribution, if its probability mass function (p.m.f) is given by

$$p(x;\theta,\varphi)=\begin{cases}\varphi+(1-\varphi)\theta^{r}, & x=0\\ (1-\varphi)\binom{x+r-1}{x}\theta^{r}(1-\theta)^{x}, & x=1,2,3,\ldots;\ 0<\varphi<1,\ 0\le\theta\le 1\end{cases} \quad (1.1)$$

$$p(x;\theta,\varphi)=\varphi\, p_1(0;\theta)+(1-\varphi)\,p_1(x;\theta)$$

Where, $p_1(0;\theta)=\begin{cases}1, & if\ \ x=0\\ 0, & if\ \ x\neq 0\end{cases}$ and $p_1(x;\theta)=\binom{x+r-1}{x}\theta^{r}(1-\theta)^{x}$, $x=0,1,2,\ldots,0\le\theta\le 1.$

Thus, the distribution of X is a convex combination of a distribution degenerate at zero and a Negative Binomial distribution and the parameter *"r"* is assumed to be known.

2. Maximum Likelihood Estimation

Let $X = (X_1, X_2, X_3, \cdots, X_n)$ be a random sample on X with the p.m.f. specified in (1.1). Then the likelihood function is given by

$$L(\theta,\varphi \mid \underline{x}) = \prod_{j=1}^{n} P(X_j = x_j)$$

$$= \prod_{j=1}^{n} \left\{\varphi + (1-\varphi)\theta^r\right\}^{1-a_j} \left\{(1-\varphi)\binom{x_j + r - 1}{x_j}\theta^r (1-\theta)^{x_j}\right\}^{a_j}$$

Where, $a_j = \begin{cases} 0, \text{ if } x_j = 0 \\ 1, \text{ if } x_j \geq 1 \end{cases}$.

Since the above likelihood function does not yield closed form expressions for the maximum likelihood estimators (MLEs) of θ and φ.

2.1 EM Algorithm

When the likelihood function has a complicated structure and maximizing it by numerical methods is difficult, a simple alternative procedure is the EM-algorithm developed by Dempster et al.

The adaptation of the EM algorithm is discussed for computing the ML estimates of *q* and *j* in *p(x,q,j)* specified in (1.1). The EM algorithm is an iterative procedure to compute the MLEs of the parameters involved in a model when the likelihood equations do not admit closed form solutions. There are E- and M-steps at each of the iteration. To implement the EM algorithm the likelihood has to be rewritten so as to accommodate missing data.

Let $Z_j = 0$ or 1 according as the j[th] observed leaf is unsuitable or suitable. If $X_j > 0$, then $Z_j = 1$. On the other hand, when $X_j = 0$, then $Z_j = 0$ or 1. Therefore $\{Z_j : X_j = 0\}$ becomes the set of missing observations. When $(X_1, X_2, \ldots, X_n)$ is augmented with $(Z_1, Z_2, \ldots, Z_n)$, $((X_1,Z_1), (X_2,Z_2), \ldots, (X_n,Z_n))$becomes the complete data set.

The likelihood function of the complete data is given by

$$L_c(\theta,\varphi \mid \underline{x},\underline{u}) = \prod_{j=1}^{n} \varphi^{1-u_j} \left\{(1-\varphi)\binom{x_j + r - 1}{x_j}\theta^r (1-\theta)^{x_j}\right\}^{u_j},$$

Where,

$$u_j = \begin{cases} 1, \text{ if } x_j > 0 \\ Z_j, \text{ if } x_j = 0 \end{cases}$$

In the E-step, the expectation of the likelihood function of the complete data is taken and $E(Z_j)$ is replaced by the conditional Expectation $E(Z_j|\theta_o, \varphi_o, X_j = 0)$, where θ_o and φ_o are respectively the initial estimates of θ and φ. In the M-step, $E[L_c(\theta,\varphi|\underline{x}, \underline{u})]$ is maximized with respect to θ and φ. If $\theta_1$and φ_1 are the values of θ and φ which maximize $E[L_c(\theta,\varphi|\underline{x}, \underline{u})]$, then the E-step is repeated using θ_1 and φ_1.

The computational details of these steps can be summarized as follows:

a) Choose the initial estimates $\theta_0 = \dfrac{r}{(r+\bar{X})}$ and $\varphi_0 = \dfrac{n_0}{n}$.

b) Compute, $w = \dfrac{(1-\varphi_0)\theta_0^{-r}}{\varphi_0 + (1-\varphi_0)\theta_0^{-r}}$.

c) Using the realization $(x_1, x_2, \ldots, x_n)$ of the observed sample, compute

$$\theta_1 = \frac{r(n_g + n_0 w)}{\sum_{j:x_j>0} x_j + r(n_g + n_0 w)} \quad \text{and} \quad \varphi_1 = \frac{n_0(1-w)}{n},$$

Where, n_g and n_o are respectively the number of observations greater than zero and equal to zero.

d) Repeat the steps b) and c) by fixing $\theta_o = \theta_1$ and $\varphi_o = \varphi_1$.

A reasonable initial estimate of φ is n_o/n and the ratio of sample mean to sample variance can be taken as an initial estimate of θ. If $\{\theta_n\}_{n=1}^{\infty}$ and $\{\varphi_n\}_{n=1}^{\infty}$ are respectively the sequence of iterates of the estimates of θ and φ and they converge, then their limits are the MLEs of θ and φ.

3. Two Parameter Exponential Family

In this section, along the lines of Kale (1998) we show that the zero-inflated power series model belongs to two parameter exponential family.

The probability mass function specified in (1.1) can be written as

$$p(x;\theta,\varphi) = \left\{\varphi + (1-\varphi)\theta^r\right\}^{t(x)} \left\{(1-\varphi)\binom{x+r-1}{x}\theta^r(1-\theta)^x\right\}^{1-t(x)} \tag{3.1}$$

Where, $t(x) = \begin{cases} 1, & if \ x = 0 \\ 0, & if \ x \geq 1 \end{cases}$.

Taking log on both sides to (3.1)

$$\log p(x;\theta,\varphi) = t(x)\log\left\{\varphi + (1-\varphi)\theta^r\right\} + (1-t(x))\log\left\{(1-\varphi)\binom{x+r-1}{x}\theta^r(1-\theta)^x\right\}$$

$$\log p(x;\theta,\varphi) = t(x)\left\{\log\left(\varphi + (1-\varphi)\theta^r\right) - \log(1-\varphi) - r\log\theta\right\} + \left\{\log(1-\varphi) + r\log\theta\right\}$$
$$+ x(1-t(x))\log(1-\theta) + (1-t(x))\log\binom{x+r-1}{x}$$

On simplification we get,

$$\log p(x;\theta,\varphi) = t(x)u_1(\theta,\varphi) + v(\theta,\varphi) + x(1-t(x))u_2(\theta,\varphi) + w(x)$$
$$p(x;\theta,\varphi) = w(x)v(\theta,\varphi)\exp\left[t(x)u_1(\theta,\varphi) + x(1-t(x)u_2(\theta,\varphi)\right]$$

Which is the general form of two parameter exponential family. Hence the zero-inflated negative binomial model belongs to two parameter exponential family and thus $\left(\sum_{i=1}^{n} t(X_i), \sum_{i=1}^{n} X_i(1-t(X_i))\right)'$ is minimal sufficient and complete for (θ,φ).

4. Method of Moment Estimators

In zero-inflated negative binomial model specified in (1.1), we get first and second moments

$$E(X) = (1-\varphi)r\left(\frac{1-\theta}{\theta}\right) \quad and \quad E(X^2) = (1-\varphi)r\left(\frac{1-\theta}{\theta}\right)\left(1 + \frac{(1+r)\theta}{(1-\theta)}\right)$$

When $\underline{X} = (X_1, X_2, X_3, \ldots, X_n)$ is a random sample on X, the moment estimators of θ and φ are given by the simultaneous equations

$$(1-\varphi)r\left(\frac{1-\theta}{\theta}\right) = M_{1n} \text{ and } (1-\varphi)r\left(\frac{1-\theta}{\theta}\right)\left(1 + \frac{(1+r)\theta}{(1-\theta)}\right) = M_{2n}$$

Note that

$$\frac{M_{2n}}{M_{1n}} = \frac{(1-\varphi)r\left(\frac{1-\theta}{\theta}\right)\left(1+\frac{(1+r)\theta}{(1-\theta)}\right)}{(1-\varphi)r\left(\frac{1-\theta}{\theta}\right)}. = \left(1+\frac{(1+r)\theta}{(1-\theta)}\right)$$

Hence, the moment estimator of θ is

$$\hat{\theta}_m = \frac{M_{2n}-M_{1n}}{rM_{1n}+M_{2n}} \quad .$$

Similarly, the moment estimator of φ is

$$\hat{\varphi}_m = 1-\frac{M_{2n}-M_{1n}}{r(r+1)}.$$

In both these estimators, we have the problem of inadmissibility when either $M_{2n} = 0$ or $M_{2n} - M_{1n} > rM_{1n} + M_{2n}$ *and* $M_{2n} - M_{1n} > r(r+1)$. Hence it is reasonable to redefine these estimators as follows:

$$\hat{\theta}_m = \begin{cases} \frac{M_{2n}-M_{1n}}{rM_{1n}+M_{2n}}, & if\ M_{1n} \neq M_{2n} \\ 0, & if\ M_{1n} = M_{2n} \end{cases}$$

and $$\hat{\varphi}_m = \begin{cases} 1-\frac{M_{2n}-M_{1n}}{r(r+1)}, & if\ M_{1n} \neq M_{2n} \\ 0, & if\ M_{1n} = M_{2n} \end{cases} .$$

Since the ZINB model belongs to two parameter exponential family and the MMEs are based on these minimal sufficient statistics for the parameters θ *and* φ,

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_m \\ \hat{\varphi}_m\end{pmatrix} \xrightarrow{L} Z' \sim N\left(\begin{pmatrix}\theta \\ \varphi\end{pmatrix}, \ \Sigma^{-1}\right), \ \ as\ n \to \infty,$$

5. Fisher information matrix

Taking log on both sides to (1.1)

$$\log p(x;\theta,\varphi) = \begin{cases} \log\left(\varphi + (1-\varphi)\theta^r\right), \ x=0 \\ \log(1-\varphi) + \log\begin{pmatrix} x+r-1 \\ x \end{pmatrix} + r\log\theta + x\log(1-\theta), \ \ x = 1,\ 2,\ 3, \ldots \end{cases}$$

We get the following partial derivatives:

$$\frac{\partial \log p(x;\theta,\varphi)}{\partial\theta} = \begin{cases} \frac{(1-\varphi)r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r}, & x=0 \\ \frac{r}{\theta} - \frac{x}{(1-\theta)}, & x = 1, 2, 3, \ldots. \end{cases}$$ and

$$\frac{\partial \log p(x;\theta,\varphi)}{\partial\varphi} = \begin{cases} \frac{(1-\theta^r)}{\varphi+(1-\varphi)\theta^r}, & x=0 \\ \frac{-1}{(1-\varphi)}, & x = 1, 2, 3, \ldots. \end{cases}$$

Using above expressions, we can verify that

$$E\left[\left(\frac{\partial \log p(x;\theta,\varphi)}{\partial\theta}\right)\right] = 0 \ \ and \ \ \ E\left[\left(\frac{\partial \log p(x;\theta,\varphi)}{\partial\varphi}\right)\right] = 0$$

Further, we get

$$I_{\varphi\varphi} = E\left[\left(\frac{\partial \log p(x;\theta,\varphi)}{\partial \varphi}\right)\right]^2$$

$$\boxed{I_{\varphi\varphi} = \frac{1-\theta^r}{(1-\varphi)\left(\varphi+(1-\varphi)\theta^r\right)}} \qquad (5.1)$$

$$I_{\varphi\theta} = I_{\theta\varphi} = -E\left(\frac{\partial^2 \log p(x;\theta,\varphi)}{\partial\varphi\,\partial\theta}\right)$$

$$I_{\varphi\theta} = \frac{\left(1-p_1(0;\theta)\right)(1-\varphi)\dfrac{\partial p_1(0;\theta)}{\partial\theta}}{p(0;\theta,\varphi)} - \frac{1}{(1-\varphi)}\sum_{x\neq 0}^{n}\frac{\partial \log p_1(x,\theta)}{\partial\theta}(1-\varphi)\,p_1(x,\theta)$$

$$I_{\varphi\theta} = \frac{\left(1-\theta^r\right)(1-\varphi)r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r} - \sum_{x=0}^{n}\frac{\partial}{\partial\theta}\left\{\log\binom{x+r-1}{x} + r\log\theta + x\log(1-\theta)\right\}p_1(x,\theta) + \frac{\partial\theta^r}{\partial\theta}\,p_1(0,\theta)$$

$$I_{\varphi\theta} = \frac{\left(1-\theta^r\right)(1-\varphi)r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r} - \frac{r}{\theta}\sum_{x=0}^{n}\binom{x+r-1}{x}\theta^r(1-\theta)^x + \frac{1}{(1-\theta)}\sum_{x=0}^{n}\left(x\binom{x+r-1}{x}\theta^r(1-\theta)^x\right) + \frac{r}{\theta}$$

$$I_{\varphi\theta} = \frac{\left(1-\theta^r\right)(1-\varphi)r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r} - 1 + \sum_{x=1}^{n}\left(\frac{(x+r-1)!}{(x-1)!(r)!}\theta^r(1-\theta)^{x-1}\right) + r\theta^{r-1}$$

$$\boxed{I_{\varphi\theta} = \frac{r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r}} \qquad (5.2)$$

$$I_{\theta\theta} = E\left[\left(\frac{\partial \log p(x;\theta,\varphi)}{\partial\theta}\right)\right]^2$$

$$I_{\theta\theta} = (1-\varphi)\left[I(\theta) - \frac{\varphi\left(\dfrac{\partial p_1(0;\theta)}{\partial\theta}\right)^2}{p(0;\theta,\varphi)\;\;p_1(0;\theta)}\right] \qquad (5.3)$$

Where,

$$I(\theta) = E\left(\frac{\partial \log p_1(x_j;\theta)}{\partial\theta}\right)^2 = -\sum_{x_j=0}^{\infty}\left(\frac{\partial \log p_1(x_j;\theta)}{\partial\theta}\right)^2\; p_1(x_j;\theta)$$

$$I(\theta) = -\sum_{x_j=0}^{\infty}\frac{\partial^2}{\partial\theta^2}\left\{\log\binom{x_j+r-1}{x_j} + r\log\theta + x_j\log(1-\theta)\right\}p_1(x_j;\theta)$$

$$I(\theta) = \frac{r(1-\theta)+\theta^2}{\theta^2(1-\theta)} \qquad (5.4)$$

$$I_{\theta\theta} = (1-\varphi)\left[\frac{r(1-\theta)+\theta^2}{\theta^2(1-\theta)} - \frac{\varphi\left(r\theta^{r-1}\right)^2}{(\varphi+(1-\varphi)\theta^r)\theta^r}\right]$$

$$I_{\theta\theta} = (1-\varphi)\left[\frac{\left(r(1-\theta)+\theta^2\right)\left((\varphi+(1-\varphi)\theta^r)\right) - (1-\theta)\varphi r^2\theta^r}{\theta^2(1-\theta)(\varphi+(1-\varphi)\theta^r)}\right] \qquad (5.5)$$

Therefore, the Fisher information matrix becomes

$$\Sigma = \begin{matrix} \varphi \\ \theta \end{matrix}\overset{\begin{matrix}\varphi & & & \theta\end{matrix}}{\begin{bmatrix} I_{\varphi\varphi} & I_{\varphi\theta} \\ I_{\theta\varphi} & I_{\theta\theta} \end{bmatrix}}$$

.

By using equation (5.1) to (5.5) we get the Fisher information matrix

$$\Sigma = \begin{bmatrix} \dfrac{1-\theta^r}{(1-\varphi)(\varphi+(1-\varphi)\theta^r)} & \dfrac{r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r} \\ \dfrac{r\theta^{r-1}}{\varphi+(1-\varphi)\theta^r} & \dfrac{(1-\varphi)\left\{(r(1-\theta)+\theta^2)(\varphi+(1-\varphi)\theta^r)-(1-\theta)\varphi r^2\theta^r\right\}}{\theta^2(1-\theta)(\varphi+(1-\varphi)\theta^r)} \end{bmatrix} \quad (5.6)$$

The inverse of the above Fisher information matrix (5.6) is the variance co-variance matrix of MLEs of the parameters.

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\left(Adj(\Sigma)\right)'$$

$$\Sigma^{-1} = \frac{1}{\left(I_{\theta\theta}I_{\varphi\varphi} - I_{\theta\varphi}\,I_{\varphi\theta}\right)}\begin{bmatrix} I_{\theta\theta} & -I_{\theta\varphi} \\ -I_{\varphi\theta} & I_{\varphi\varphi} \end{bmatrix} \quad (5.7)$$

Where,

$$\Sigma_{11}^{-1} = \frac{(1-\varphi)\{(r(1-\theta)+\theta^2)(\varphi+(1-\varphi)\theta^r)-\varphi(1-\theta)r^2\theta^r\}}{(1-\theta^r)(r(1-\theta)+\theta^2)-\frac{(1-\theta)r^2\theta^r(\varphi(1-\theta^r)+\theta^r)}{(\varphi+(1-\varphi)\theta^r)}}$$

$$\Sigma_{21}^{-1} = \Sigma_{12}^{-1} = \frac{-r(1-\theta)\theta^{r+1}}{(1-\theta^r)(r(1-\theta)+\theta^2)-\frac{(1-\theta)r^2\theta^r(\varphi(1-\theta^r)+\theta^r)}{(\varphi+(1-\varphi)\theta^r)}}$$

$$\Sigma_{22}^{-1} = \frac{\theta^2(1-\theta)}{(1-\varphi)(r(1-\theta)+\theta^2)-\frac{(1-\theta)r^2\theta^r(\varphi(1-\theta^r)+\theta^r)}{(\varphi+(1-\varphi)\theta^r)(1-\theta^r)}}$$

Asymptotic relative efficiency of MLEs over MMEs of the parameters has been presented in the next section.

6. Asymptotic Relative Efficiency

An empirical comparison of MLEs and MMEs of the parameters in a ZIP model is made by Nanjundan, Loganathan and Raveendra Naika [2009]. The relative efficiency (ARE) of MMEs with respect to MLEsof the parameters are compared analytically in case of ZIP model [see Nanjundan and Raveendra Naika [2012]. The MLEs of the parameters q and jin ZINB model do not yield closed form expressions but MMEs are obtained closed form expressions. According to section 3 the estimators of the parameters are asymptotically normally distribution. Hence the asymptotic relative efficiencies of the estimators are compared analytically. Since the zero inflated negative binomial distribution (1.1) belongs to two parameter exponential family, the MLEs of *q* and *j* are also asymptotically normal and

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_{mle}\\ \hat{\varphi}_{mle}\end{pmatrix}\xrightarrow{L} Z' \sim N\left(\begin{pmatrix}\theta\\ \varphi\end{pmatrix},\ \Sigma^{-1}\right),\quad as\ n\to\infty.$$

Hence the asymptotic relative efficiency of $\hat{\theta}_{mme}$ with respect to $\hat{\theta}_{mle}$ is

$$ARE(\hat{\theta}_{mle},\hat{\theta}_{mme}) = \frac{AV(\theta_{mme})}{AV(\theta_{mle})}$$

$$= 1.$$

Therefore, the MMEs and the MLEs of *q* are asymptotically equally efficient. The same is true in the case of j too.

7. Conclusion and Discussion

The MLEs of the parameters in the ZINB model have no closed form expressions and computing them even by the EM algorithm needs computer facility. Whereas the MMEs have simple closed form expressions and they can be computed easily without computer assistance. The MMEs and the MLEs are asymptotically equally efficient. Hence MMEs can be used instead of the MLEs when the sample size is sufficiently large.

References:

1. B. K. Kale, “Optimal Estimating Equations for discrete Data with Higher Frequencies at a Point,” Journal of the Indian Statistical Association, Vol. 36, 1998, pp. 125- 136.
2. Dempster A.P., N.M. Laird, and D.B. Rubin (1977). Maximum likelihood Estimation from incomplete data via the EM algorithm (with discussion), J. Roy. Statist. Soc. Ser. B,39,1 -38.
3. McLachlan G.J. and Krishnan T., (2008). The EM Algorithm and Extensions,2/e, John Wiley, New York.
4. McLachlan, G.J. and Krishnan, T. (1997). The EM Algorithm and Extensions. John Wiley and Sons, New York.
5. Nanjundan, G, A Loganathan and T Raveendra Naika, (2009) “An Empirical Comparison of Maximum Likelihood and Moment Estimators of Parameters in a Zero-Inflated Poisson Model,” Mapana Journal of Sciences, Vol. 8, No. 2, 2009, pp. 59-72.
6. Nanjundan, G and Raveendra Naika, T. (2012) “Asymptotic Comparison of Method of Moments Estimators and Maximum Likelihood Estimators of Parameters in Zero-Inflated Poisson Model”, Applied Mathematics, 2012, 3, 610-617
7. Nanjundan, G. (2006). An EM algorithmic approach to maximum likelihood estimation in a mixture model. Vignana Bharathi, Vol. 18, 7 – 13.
8. Nanjundan G and T. Raveendra Naika (2011). Maximum Likelihood Estimation of Parameters in a Perturbed Exponential Model in International Journal of Agricultural and Statistical Sciences, Vol.7, No.2, pp. 527-534, December 2011 issue.