International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 10, 2010, pp 37-43

A Fast Algorithm for Computing Maximum Likelihood Estimates of the Negative-Binomial Lindley Distribution

Naushad Mamode Khan
Department of Economics and Statistics
Faculty of Social studies and Humanities
University of Mauritius
Mauritius
E-mail: n.mamodekhan@uom.ac.mu

Maleika Heenaye-Mamode Khan
Department of Computer Science and Engineering
Faculty of Computer Science and Engineering
University of Mauritius
Mauritius
E-mail: m.mamodekhan@uom.ac.mu

[Received August 15, 2010; Revised September 19, 2010; Accepted September 30, 2010]

Abstract

Negative-binomial Lindley distribution is a two-parameter discrete model that has been recently introduced in statistical literature. It has been used mostly in the analysis of insurance data and has shown to provide suitable fits as compared with the Poisson and Negative -binomial distributions. The parameters of the Negative- binomial Lindley model were estimated using separate maximum likelihood equations via the Newton-Raphson iterative technique. However, this method of estimation does not require the construction of a joint Hessian matrix. Further to this, it becomes difficult to estimate the joint covariance matrix. To overcome this shortcoming, we propose a joint maximum likelihood estimation approach that is based on a diagonal Jacobian approximation of the joint Hessian matrix. We further compare this estimation methodology with the separate maximum likelihood approach and show that the joint maximum likelihood approach is computationally faster.

Keywords and Phrases: Negative-Lindley, Maximum-likelihood, Hessian matrix, Newton-Raphson, Jacobian.

AMS Classification: 62E$_{xx}$.

1 Introduction

Traditionally, Poisson and Negative binomial (NB) models are considered to be the most convenient models to represent count data. In the recent years, some new discrete distributions have been introduced such as the Com-Poisson model (Shmueli et al. 2005) and Negative- binomial Lindley (NL) distributions (Zamani and Ismail, 2010). In this paper, we focus on the Negative- binomial Lindley (NL) distribution. This is a two-parameter model and is formed by mixing the negative-binomial distribution and the Lindley distribution. Zamani and Ismail (2010) have applied it on two samples of insurance data and compared the fits with Poisson and negative binomial models. Their results show that NL is slightly better than NB and more efficient than Poisson since the counts are dispersed. To estimate the two parameters, the authors formulate separate maximum likelihood estimating equations (SML) and find the estimates iteratively using the Newton-Raphson technique. We note that, in their approach, they did not construct the joint hessian matrix. In fact, the hessian component of the likelihood function is very difficult to express and may yield huge computational troubles. In this paper, we propose an iterative algorithm based on a joint maximum likelihood principle (JML) that consists of approximating the joint hessian matrix by a diagonal Jacobian approximation following Waziri et al.(2010). We will also compare the computational time of this new algorithm with the approach of Zamani and Ismail (2010). The outline of the paper is as follows: In the next section, we review the Negative- binomial Lindley (NL) distribution and its statistical properties as demonstrated by Zamani and Ismail (2010). In section 3, we provide their estimating equations of the parameters under separate maximum likelihood functions. In section 4, we present a new version of the Newton iterative algorithm based on a diagonal Jacobian transformation of the hessian matrix. In section 5, we apply both algorithms on simulated samples of over- and under-dispersed data and compare the computational time of both algorithms. In section 6, we present the conclusions and recommendations.

2 Negative- binomial Lindley Model

Very often, mixed Poisson and mixed negative binomial provide better fits on count data than other discrete distribution. Zamani and Ismail (2010) introduce a new mixed distribution known as Negative-binomial Lindley distribution (NL) by mixing the distributions of the negative binomial (r, p) and Lindley (θ) where the parameterization of $p = \exp(-\lambda)$ is taken into account. They proved that the marginal distribution resulting out of this mixture can be expressed in the following probability mass function form

$$\Pr(x) = \frac{\theta^2}{\theta+1} \begin{pmatrix} r+x-1 \\ x \end{pmatrix} \sum_{j=0}^{k} \begin{pmatrix} k \\ j \end{pmatrix} (-1)^{j\frac{\theta+r+j+1}{(\theta+r+j)^2}}, x = 0, 1, 2, ... \tag{1}$$

where $E(X) = r\left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1\right]$ and
$Var(X) = (r + r^2)\frac{\theta^2(\theta-1)}{(\theta+1)(\theta-2)^2} - (r + 2r^2)\frac{\theta^3}{(\theta+1)(\theta-1)^2} + r^2 \text{-} r^2\left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1\right]^2$

3 Separate Maximum Likelihood Estimating Equations Approach (SML)

To estimate the parameters (θ, r), Zamani and Ismail (2010) used the maximum likelihood approach. The log-likelihood function is expressed as:

$$\log L(r,\theta) = \sum_{x=0}^{k} n_x \log\left[\frac{\theta^2}{\theta+1}\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\theta+r+j+1}{(\theta+r+j)^2}\right] \tag{2}$$

where n_x is the value at the x^{th} index and the partial derivatives are given as follows:

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial}{\partial \theta}\ln L(r,\theta) = n\left(\frac{2}{\theta} - \frac{1}{\theta+1}\right) + \sum_{x=0}^{k} n_x \left\{\frac{\sum_{j=0}^{x}(-1)^{j+1}\binom{x}{j}\frac{\theta+r+j+2}{(\theta+r+j)^3}}{\sum_{j=0}^{x}(-1)^{j}\binom{x}{j}\frac{\theta+r+j+1}{(\theta+r+j)^2}}\right\} = 0 \tag{3}$$

$$\frac{\partial \ell}{\partial r} = \frac{\partial}{\partial r}\ln L(r,\theta) = \frac{\partial}{\partial r}\sum_{x=0}^{k} n_x \log\binom{x+r-1}{x} + \sum_{x=0}^{k} n_x \left\{\frac{\sum_{j=0}^{x}(-1)^{j+1}\binom{x}{j}\frac{\theta+r+j+2}{(\theta+r+j)^3}}{\sum_{j=0}^{x}(-1)^{j}\binom{x}{j}\frac{\theta+r+j+1}{(\theta+r+j)^2}}\right\} = 0 \tag{4}$$

where $n = \sum_{x=0}^{k} n_x$

Following Klugman et al. (2008),

$$\frac{\partial}{\partial r}\sum_{x=0}^{k} n_x \log\binom{r+x-1}{x} = \sum_{x=0}^{k} n_x \sum_{m=0}^{x-1}\ln(r+m) \tag{5}$$

Zamani and Ismail (2010) re-write equation (3) as follows:

$$n\left(\frac{2}{\theta} - \frac{1}{\theta+1}\right) = \sum_{x=0}^{k} n_x \sum_{m=0}^{x-1}\ln(r+m) \tag{6}$$

The steps to estimate the two parameters can be then summarized as follows:

Step 1: The estimate of θ: $\widehat{\theta}$ is obtained by solving equation (6) using the quadratic formula.

Step 2: Following equation (5), they re-write equation (6) as follows:

$$H(\widehat{r}) = \frac{\partial}{\partial r}\ln L(r,\theta) = \sum_{x=0}^{k} n_x \sum_{m=0}^{x-1}\frac{1}{\widehat{r}+m} + \sum_{x=0}^{k} n_x \left\{ \frac{\sum_{j=0}^{x}(-1)^{j+1}\begin{pmatrix} x \\ j \end{pmatrix}\frac{\widehat{\theta}+r+j+2}{(\widehat{\theta}+r+j)^3}}{\sum_{j=0}^{x}(-1)^{j}\begin{pmatrix} x \\ j \end{pmatrix}\frac{\widehat{\theta}+r+j+1}{(\widehat{\theta}+r+j)^2}} \right\} = 0 \tag{7}$$

Then $\widehat{r}$ is obtained iteratively by using the Newton-Raphson technique

$$\widehat{r}_{k+1} = \widehat{r}_k - \dot{H}^{-1}(\widehat{r}_k)H(\widehat{r}_k) \tag{8}$$

where $\dot{H}(\widehat{r}_k)$ is the estimate of the gradient matrix at the k^{th} iterated value.

The algorithm works as follows: For an initial value of $\widehat{r}$, we calculate the estimate of $\widehat{\theta}$ using equation (6). Using this value of $\widehat{\theta}$, we solve iteratively equation (8) until convergence. Having obtained an update of $\widehat{r}$, we replace in equation (6) to update $\widehat{\theta}$, then solve equation (8) iteratively again. This cycle continues until both estimates converge based on the criteria $|| \widehat{r}_{k+1} - \widehat{r}_k || < 10^{-5}$.

4 Joint Maximum Likelihood Approach Based on the Diagonal Jacobian Transformation (JML)

In the previous approach, we note that there is no construction of a joint Hessian matrix to estimate the parameters. This is in fact quite difficult because the second partial derivatives of the parameters (θ, r) are complicated following equations (3) and (4). To overcome this problem, we propose to construct a diagonal Jacobian transformation matrix that approximates the Hessian matrix. That is, using the equations (3) and (4), we may write the Newton-Raphson iterative equation as follows:

$$\begin{bmatrix} \widehat{\theta}_{k+1} \\ \widehat{r}_{k+1} \end{bmatrix} = \begin{bmatrix} \widehat{\theta}_{k} \\ \widehat{r}_{k} \end{bmatrix} - \begin{bmatrix} \left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_k} & \left.\frac{\partial^2 \ell}{\partial \theta \partial r}\right|_{\theta=\theta_k, r=r_k} \\ \left.\frac{\partial^2 \ell}{\partial \theta \partial r}\right|_{\theta=\theta_k, r=r_k} & \left.\frac{\partial^2 \ell}{\partial r^2}\right|_{r=r_k} \end{bmatrix}^{-1} \begin{bmatrix} \left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_k} \\ \left.\frac{\partial \ell}{\partial r}\right|_{r=r_k} \end{bmatrix} \tag{9}$$

To obtain the diagonal approximation of the Hessian matrix, we follow the estimation method of Waziri et al. (2010), i.e, based on the classical Taylor series expansion, we may write

$$F(x_{k+1}) = F(x_k) + F'(x_k)(x_{k+1} - x_k) + o(||x_{k+1} - x_k||)^2 \tag{10}$$

By letting $F'(x_k) = D_k$ where $D_k$is the diagonal matrix, we obtain

$$D_{k+1}(x_{k+1} - x_k) = F(x_{k+1}) - F(x_k) \tag{11}$$

Then the i^{th} diagonal component of the matrix becomes

$$d^i_{k+1} = \frac{F_i(x_{k+1}) - F_i(x_k)}{x^i_{k+1} - x^i_k} \tag{12}$$

By applying equation (12) to equation (9) and by letting $F_1(\theta_k,\ r_k)= \left.\frac{\partial L}{\partial \theta}\right|_{\theta=\theta_k}$ and $F_2(\theta_k,\ r_k)= \left.\frac{\partial L}{\partial r}\right|_{r=r_k}$,

we may write

Step 1:

$$d^1_{k+1} = \frac{F_1(\theta_{k+1}, r_{k+1}) - F_1(\theta_k, r_k)}{\theta_{k+1} - \theta_k} \tag{13}$$

Step 2:

$$d^2_{k+1} = \frac{F_2(\theta_{k+1}, r_{k+1}) - F_2(\theta_k, r_k)}{r_{k+1} - r_k} \tag{14}$$

Step 3:

$$\begin{bmatrix} \widehat{\theta}_{k+1} \\ \widehat{r}_{k+1} \end{bmatrix} = \begin{bmatrix} \widehat{\theta}_{k} \\ \widehat{r}_{k} \end{bmatrix} - \begin{bmatrix} d^1_k & 0 \\ 0 & d^2_k \end{bmatrix}^{-1} \begin{bmatrix} \left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_k} \\ \left.\frac{\partial \ell}{\partial r}\right|_{r=r_k} \end{bmatrix} \tag{15}$$

The algorithm works as follows: For an initial value of $(\widehat{\theta}_0, \widehat{r}_0)^T$, we calculate the components $\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\theta_0}$ and $\left.\frac{\partial L}{\partial r}\right|_{r=r_0}$ assuming the diagonal matrix is the identity matrix. By replacing in equation (15), we obtain $(\widehat{\theta}_1, \widehat{r}_1)^T$. Then, we compute the diagonal elements d^1_1 and d^2_1using $(\widehat{\theta}_0, \widehat{r}_0)^T$ and $(\widehat{\theta}_1, \widehat{r}_1)^T$to obtain $(\widehat{\theta}_2, \widehat{r}_2)^T$. Then, we calculate d^1_2 and d^2_2 to obtain $(\widehat{\theta}_3, \widehat{r}_3)^T$using the iterative formula in equation (15). The whole process continues until convergence based on the criteria $||\ \widehat{r}_{k+1} - \widehat{r}_k\ || < 10^{-5}$.

5 Comparison of SML and JML

In this section, we perform a simulation study to compare the estimates of (θ, r) under SML and JML. Initially, we generate a set of over-dispersed counts using the *nbinom* function in R 1.6.1 with various parameters. Then, we apply (6), (8) and (15) to calculate the estimates of $(\widehat{\theta}, \widehat{r})$under SML and JML approaches respectively using Matlab 7. The results of the study are shown in the table below:

Table 1: Estimates of $(\widehat{\theta}, \widehat{r})$ under SML and JML based on Negative-binomial simulated data

Sample size	SML estimates	JML estimates	Computational time for SML	Computational time for JML
10	$(1.23, 3.44)$	$(1.21, 3.46)$	2 s	1.8 s
30	$(2.56, 1.11)$	$(2.53, 1.10)$	2 s	1.8s
50	$(0.23, 1.61)$	$(0.26, 1.61)$	3 s	2s
100	$(4.09, 2.11)$	$(4.09, 2.11)$	15 s	10s
500	$(3.13, 1.12)$	$(3.13, 1.12)$	35 s	20s
1000	$(5.55, 0.21)$	$(5.55, 0.21)$	57 s	40s
2000	$(0.93, 5.21)$	$(0.92, 5.20)$	60 s	40s
5000	$(2.53, 6.02)$	$(2.53, 6.02)$	80 s	56s
10,000	$(1.18, 2.41)$	$(1.18, 2.40)$	110 s	90s

In Table 1, we present the estimates under SML and JML approaches. To start the algorithms, we assume small initial values for $(\widehat{\theta}, \widehat{r})$. We note that under the SML approach, certain of the initial values lead to divergence problems whereas under the JML approach, such problems were not noted. As the sample size increases, the estimates under both approaches become almost equal and as for the computational time, SML is quite time-consuming compared with the JML approach. Next, we generate a set of under-dispersed Com-Poisson counts following the simulation process of Shmueli et al. (2005) and use SML and JML to estimate the parameters.

Table 2: Estimates of $(\widehat{\theta}, \widehat{r})$ under SML and JML based on Com-Poisson simulated data

Sample size	SML estimates	JML estimates	Computational time for SML	Computational time for JML
10	$(2.13, 0.92)$	$(2.25, 0.96)$	2 s	1.7 s
30	$(1.16, 0.76)$	$(1.23, 0.85)$	2 s	1.6s
50	$(4.03, 0.86)$	$(3.97, 0.91)$	2 s	1.8s
100	$(3.09, 0.56)$	$(3.09, 0.62)$	12 s	11s
500	$(1.13, 0.82)$	$(1.22, 0.86)$	30 s	23s
1000	$(3.55, 0.41)$	$(3.62, 0.45)$	60 s	45s
2000	$(1.43, 0.68)$	$(1.35, 0.72)$	62 s	45s
5000	$(2.51, 1.02)$	$(2.51, 0.98)$	80 s	52s
10,000	$(0.88, 0.79)$	$(0.91, 0.82)$	120 s	95s

In Table 2, we note practically the same pattern in the estimates and the computational times. That is, JML is lesser time consuming and as the cluster size increases, the difference in the estimated values become lesser. However, JML yields quite a few non-convergent simulations.

6 Conclusion

Based on the simulation results, we note that as the sample size increases, the estimates under SML and JML approaches become almost equal. As for the computational time, we note that SML takes significantly more time to run the algorithm than JML especially for large sample size. Thus, we may conclude that JML is a more suitable technique to estimate the parameters of the Negative-Binomial Lindley model.

References

[1] Klugman, S. A., Panjer, H. H., Willmot, G. E. (2008). Loss models: From data to decision, 3rd Edition, John Wiley and Sons, USA, pg.101-159

[2] Shmueli, G, Minka T, Borle J and Boatwright P. (2005). "A useful model for fitting discrete data: revival of the Com-Poisson distribution". Journal of the Royal statistical society (54):127-142.

[3] Waziri, M. Y., Leong, W. J., Hassan, M. A., Monsi, M. (2010). "A New Newton's method with diagonal Jacobian approximation for systems of Non-Linear equations", Journal of Mathematics and Statistics, 6(3): Pg. 246-252, ISSN 1549-3644, Science Publication

[4] Zamani, H., Ismail, N. (2010). "Negative Binomial-Lindley distribution and its application", Journal of Mathematics and Statistics, 6(1): Pg 4-9, ISSN 1549-3644, Science Publication.