

Estimating the Parameters of the Negative-Lindley Distribution using Broyden-Fletcher-Goldfarb-Shanno

Naushad Mamode Khan
Department of Economics and Statistics,
Faculty of Social Studies and Humanities,
University of Mauritius, Mauritius

Abstract: Problem statement: The Maximum Likelihood Estimation (MLE) technique is the most efficient statistical approach to estimate parameters in a cross-sectional model. Often, MLE gives rise to a set of non-linear systems of equations that need to be solved iteratively using the Newton-Raphson technique. However, in some situations such as in the Negative-Lindley distribution where it involves more than one unknown parameter, it becomes difficult to apply the Newton-Raphson approach to estimate the parameters jointly as the second derivatives of the score functions in the Hessian matrix are complicated. **Approach:** In this study, we propose an alternate iterative algorithm based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach that does not require the computation of the higher derivatives. **Conclusion:** To assess the performance of BFGS, we generate samples of over-dispersed count with various dispersion parameters and estimate the mean and dispersion parameters. **Results:** BFGS estimates the parameters of the Negative-Lindley model efficiently.

Key words: Maximum likelihood, Negative-Lindley, Hessian matrix, Newton-Raphson, Broyden-Fletcher-Goldfarb-Shanno (BFGS), Maximum Likelihood Estimation (MLE), dispersion parameters

INTRODUCTION

Traditionally, Poisson and the Negative Binomial (NB) models are regarded as the most suitable 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)(Khan and Khan, 2010) and Negative-Lindley (NL) distributions (Zamani and Ismail, 2010). In this study, we focus on the Negative-Lindley (NL) distribution. This is a two-parameter model and is developed through a mixture of 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 formulated separate maximum likelihood estimating equations and found 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 quite difficult to compute as the second

derivatives are difficult to express and may lead to numerical instability. In this study, we propose an alternate iterative algorithm based on the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach (Yuan, 1991). This algorithm does not require the computation of the second derivatives of the score function and can yield equally reliable estimates. The outline of the study is as follows: We review the Negative-Lindley distribution and the maximum likelihood approach as demonstrated by Zamani and Ismail (2010). We next introduce the BFGS method of estimation. We then perform a simulation study where we generate samples of over-dispersed counts and estimate the parameters using BFGS algorithm. Finally, we present the conclusions and recommendations.

MATERIALS AND METHODS

Zamani and Ismail (2010) showed that the marginal distribution of the Negative-Lindley distribution is expressed as:

$$\Pr(x) = \frac{\theta^2}{\theta + 1} \binom{r+x-1}{x} \sum_{j=0}^k \binom{k}{j} (-1)^j \frac{\theta+r+j+1}{(\theta+r+j)^2}, x = 0, 1, 2, \dots \quad (1)$$

Where:

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

and:

$$\begin{aligned} \text{Var}(X) &= (r+r^2) \frac{\theta^2(\theta-1)}{(\theta+1)(\theta-1)^2} \\ &- (r+2r^2) \frac{\theta^3}{(\theta+1)(\theta-1)^2} \\ &+ r^2 - r^2 \left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1 \right]^2 \end{aligned}$$

To estimate the parameters (θ, r) , they used the maximum likelihood approach. The log-likelihood function yields:

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

Where:

n_x = The value at the
 x^{th} = Index and the partial derivatives

$$\begin{aligned} \frac{\partial L}{\partial \theta} &= \frac{\partial}{\partial \theta} \log L(r, \theta) = n \left(\frac{2}{\theta} - \frac{1}{\theta+1} \right) \\ &+ \sum_{x=0}^k n_x \left\{ \frac{\sum_{j=0}^x (-1)^{j+1} \binom{x}{j} \frac{\theta+r+j+2}{(\theta+r+j)^3}}{\sum_{j=0}^x (-1)^j \binom{x}{j} \frac{\theta+r+j+1}{(\theta+r+j)^2}} \right\} = 0 \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\partial L}{\partial r} &= \frac{\partial}{\partial r} \log L(r, \theta) = \frac{\partial}{\partial r} \sum_{x=0}^k n_x \log \left(\frac{x+r-1}{x} \right) \\ &+ \sum_{x=0}^k n_x \left\{ \frac{\sum_{j=0}^x (-1)^{j+1} \binom{x}{j} \frac{\theta+r+j+2}{(\theta+r+j)^3}}{\sum_{j=0}^x (-1)^j \binom{x}{j} \frac{\theta+r+j+1}{(\theta+r+j)^2}} \right\} = 0 \end{aligned} \quad (4)$$

Where:

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

As it can be noted, the first derivatives of the log-likelihood function are quite complicated. Thus, the construction of the Hessian matrix becomes difficult. To eliminate this problem, Zamani and Ismail (2010) formulated separate estimating equations to estimate θ and r . Following Klugman *et al.* (2008):

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

Then Eq. 3 can be re-written as:

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

The estimate of $\hat{\theta}$ is obtained by solving Eq. 6 using the quadratic formula. In the same way, Eq. 4 can be written as:

$$\begin{aligned} H(\hat{r}) &= \frac{\partial}{\partial r} \log L(r, \theta) = \sum_{x=0}^k n_x \sum_{m=0}^{x-1} \frac{1}{\hat{r}+m} + \\ &\sum_{x=0}^k n_x \left\{ \frac{\sum_{j=0}^x (-1)^{j+1} \binom{x}{j} \frac{\hat{\theta}+r+j+2}{(\hat{\theta}+r+j)^3}}{\sum_{j=0}^x (-1)^j \binom{x}{j} \frac{\hat{\theta}+r+j+1}{(\hat{\theta}+r+j)^2}} \right\} = 0 \end{aligned} \quad (7)$$

Then \hat{r} is obtained iteratively by using the Newton-Raphson technique:

$$\hat{r}_{k+1} = \hat{r}_k - \hat{H}^{-1}(\hat{r}_k) H(\hat{r}_k) \quad (8)$$

where, $\hat{H}(\hat{r}_k)$ is the estimate of the Hessian matrix at the k^{th} iterated value. The algorithm works as follows: For an initial value of \hat{r} , we calculate the estimate of $\hat{\theta}$ using Eq. 6. Using this value of $\hat{\theta}$, we solve iteratively Eq. 8 until convergence. Having obtained an update of \hat{r} , we replace in Eq. 6 to update $\hat{\theta}$, then solve Eq. 8 iteratively again. This cycle continues until both estimates converge. The Newton-Raphson method described above has one major drawback. It does not provide an estimate of the joint Hessian matrix of the parameters θ and r . As a result, the variance of the estimates θ and r may be over- or under estimated. We note that Eq. 8 yields only an estimate of the Hessian matrix for the parameter r . To overcome this problem, we propose an iterative algorithm that is based on a quasi-Newton Broyden-Fletcher-Goldfarb-Shanno

(BFGS) approach. The BFGS algorithm can be summarized in the following steps:

- The quasi-Newton direction is given by:

$$\begin{pmatrix} \nabla\theta_{k-1} \\ \nabla\hat{r}_{k-1} \end{pmatrix} = -H_{k-1}^{-1} \begin{pmatrix} \frac{\partial L}{\partial \theta_{k-1}} \\ \frac{\partial L}{\partial \hat{r}_{k-1}} \end{pmatrix} \quad (9)$$

- The iterative equation yields:

$$\begin{bmatrix} \hat{\theta}_k \\ \hat{r}_k \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{k-1} \\ \hat{r}_{k-1} \end{bmatrix} + \begin{pmatrix} \nabla\theta_{k-1} \\ \nabla\hat{r}_{k-1} \end{pmatrix} \quad (10)$$

- Compute the Hessian matrix at the k^{th} iteration using:

$$H_k = H_{k-1} + \frac{yy^T}{y^T s} - \frac{H_{k-1} s s^T H_{k-1}}{s^T H_{k-1} s} \quad (11)$$

$$H_k^{-1} = \left(I - \frac{sy^T}{y^T s} \right) H_{k-1}^{-1} \left(I - \frac{ys^T}{y^T s} \right) + \frac{ss^T}{y^T s} \quad (12)$$

Where:

$$S = \begin{bmatrix} \hat{\theta}_k \\ \hat{r}_k \end{bmatrix} - \begin{bmatrix} \hat{\theta}_{k-1} \\ \hat{r}_{k-1} \end{bmatrix}$$

And:

$$y = \begin{pmatrix} \frac{\partial L}{\partial \theta_k} \\ \frac{\partial L}{\partial \hat{r}_k} \end{pmatrix} - \begin{pmatrix} \frac{\partial L}{\partial \theta_{k-1}} \\ \frac{\partial L}{\partial \hat{r}_{k-1}} \end{pmatrix} \begin{pmatrix} \frac{\partial L}{\partial \theta_k} \\ \frac{\partial L}{\partial \hat{r}_k} \end{pmatrix} \quad (13)$$

Initially, we set the parameters to zero and the hessian matrix at the $(k-1)^{\text{th}}$ iteration to be the identity matrix. Having obtained the first set of estimates, we follow the steps (9)-(13). The standard errors of $(\hat{\theta}, \hat{r})$ can be obtained through the diagonal entries of H_k^{-1} .

RESULTS

We perform a simulation study to estimate (θ, r) using the BFGS approach.

Table 1: Estimates of parameters under BFGS approach

Sample size	Estimates $(\hat{\theta}, \hat{r})$	Standard errors $(\hat{\theta}, \hat{r})$	Computational time (sec)
10	(1.23,3.44)	(0.213,0.112)	1.8
30	(2.56,1.11)	(0.131,0.107)	1.8
50	(0.23,1.161)	(0.146,0.110)	2.0
100	(4.09,2.11)	(0.132,0.096)	10.0
500	(3.13,1.12)	(0.142,0.110)	20.0
1000	(5.55,0.21)	(0.115,0.010)	40.0
2000	(0.93,5.21)	(0.009,0.008)	40.0
5000	(2.53,6.02)	(0.009,0.007)	56.0
10,000	(1.18,2.41)	(0.007,0.005)	90.0

Initially, we generate a set of over-dispersed counts using the `nbinom` function in R 1.6.1 with various parameters. Then, we apply (10) to calculate the estimates of $(\hat{\theta}, \hat{r})$. The programs were implemented in MATLAB. The results of the study are shown in the Table 1.

DISCUSSION

The simulation study was run for different sample size starting from 10-10,000. Each simulation was run 5000 times and the estimates obtained were averaged. To estimate the parameters, we start with very small values of $(\hat{\theta}, \hat{r})$. We note that as the cluster size increases, the standard errors decrease significantly. This justifies the consistency property of the estimators. Practically, we have reported very few non-convergence simulations even for large sample size. In comparison with the MLE approach, we note that for sample size ranging from 10-50, there were around 2455 non-convergent simulations while BFGS yields approximately 600 non-convergent simulations. As the sample size increases, the number of non-convergent simulations decreases in both methods but BFGS yields still fewer ones. The computational time is also very encouraging and it indicates that BFGS is quite fast.

CONCLUSION

Based on the simulation results, we conclude that BFGS is a suitable technique to estimate the parameters of the Negative-Lindley model. It yields reliable estimates and provides consistent results. Moreover, this method of estimation provides an estimate of the standard errors of both parameters. We therefore recommend that to estimate parameters of the Negative-Lindley model, the BFGS algorithm is a more convenient estimation approach than the traditionally MLE technique.

REFERENCES

- Klugman, S.A., H.H. Panjer and G.E. Willmot, 2008. Loss Models: From Data to Decision, 3rd Edn., John Wiley and Sons, USA, pp: 726.

- Khan, N.M. and M.H.M. Khan, 2010. Model for analyzing counts with over, equi- and under-dispersion in actuarial statistics. *J. Math. Stat.*, 6: 92-95. DOI: 10.3844/jmssp.2010.92.95
- Shmueli, G., T.P. Minka, Joseph B. Kadane, J. Borle and P. Boatwright, 2005. A useful distribution for fitting discrete data: Revival of the conway-maxwell-poisson distribution. *J. Royal Stat. Soc.*, 54: 127-142. <http://www.jstor.org/pss/3592603>
- Yuan, Y.X., 1991. A modified BFGS algorithm for unconstrained optimization. *IMA J. Numerical Anal.*, 11: 325-332. DOI: 10.1093/imanum/11.3.325
- Zamani, H. and N. Ismail, 2010. Negative binomial-lindley distribution and its application. *J. Math. Stat.*, 6: 4-9. ISSN: 1549-3644. DOI: 10.3844/jmssp.2010.4.9