Journal of Mathematics and Statistics 6 (2): 92-95, 2010
ISSN 1549-3644
© 2010 Science Publications

Model for Analyzing Counts with Over-,Equi-and Under-Dispersion in Actuarial Statistics

[1]Naushad Mamode Khan and [2]Maleika Heenaye-Mamode Khan
[1]Department of Mathematics, Faculty of Science,
[2]Department of Computer Science and Engineering,
University of Mauritius, Mauritius

Abstract: Problem statement: Actuarial science has grown much popularity in the recent years due to the growth of insurance companies. In practice, the data involved in actuarial science are mostly counts which may be over-,equi-or under-dispersed. Many probability distributions were proposed to model such data such as the mixed Poisson distributions. However, the estimation methodologies based on such mixed Poisson distributions may be complicated and may not yield consistent and efficient estimates. **Approach:** In this study, we consider a recently introduced model known as the two-parameter Com-Poisson distribution that is flexible in modeling both over-,equi-and under-dispersed data. **Results:** The estimation of parameters is carried out using a quasi-likelihood estimation technique based on a joint estimation approach and a marginal approach via Newton-Raphson iterative procedure. **Conclusion:** The Com-Poisson distribution is applied on two samples of insurance data and we compare the estimates with the estimates based on the Negative-Lindley distribution. Based on the results, it is shown that both Com-Poisson and Negative Lindley yield almost equally efficient estimates of the parameters with fitted values almost close to the actual values under both the joint and marginal quasi-likelihood approaches.

Key words: Actuarial, count data, insurance, Com-Poisson, Negative-Lindley, Quasi-likelihood, Joint estimation, Marginal estimation

INTRODUCTION

The modeling of count data is one of the most important issues in actuarial theory. Various probability distributions have been proposed to model these data but the fundamental question is which model yields the best fits. These distributions comprise of the Poisson distribution, the negative binomial distribution, the Generalized Poisson distribution and mixed distributions (Johnson *et al*., 1993). However, actuarial data may be over-, equior under-dispersed. Poisson and negative binomial distributions may not be suitable to model such data because of the restriction on the mean-variance ratio. The Generalized Poisson distribution may not also be suitable when the data is under-dispersed (Jahangeer *et al*., 2009). To overcome these problems, Shmueli *et al*. (2005) have recently re-introduced a discrete model known as Com-Poisson. This model has the ability to account for over-, equi- and under-dispersion irrespective of the type of dispersion of the count data. In this study, we provide an overview of the Com-Poisson model and discuss its statistical properties. We then apply a quasi-likelihood estimation equation to estimate the parameters of the model using an iterative procedure. Ultimately, we apply the Com-Poisson model to two insurance claim data collected by Klugman *et al*. (2008). Moreover, we compare the fits based on the Com-Poisson model with the fits based on the Negative-Lindley distribution (Zamani and Ismail, 2010).

MATERIALS AND METHODS

Com-Poisson model (CMP): Recently, Shmueli *et al*. (2005) proposed the Conway Maxwell Poisson (Com Poisson) distribution to model counts which may be equi-, over- and under-dispersed. Kadane *et al*. (2006) and Shmueli *et al*. (2005) studied the basic properties of this distribution and the fitting of this distribution to over -and under -dispersed cross sectional count data. To estimate the parameters, Shmueli *et al*. (2005) has used the Maximum-likelihood estimation technique while Lord *et al*. (2008) used Bayesian techniques. Following, Jowaheer and Mamode Khan (2009), we use a Joint Quasi-Likelihood technique (JGQL) to estimate the parameters of the model. The JGQL approach

Corresponding Author: Naushad Mamode Khan, Department of Mathematics, Faculty of Science, University of Mauritius, Mauritius

provides consistent and equally efficient estimates as the maximum likelihood approach. The Com-Poisson model is given by:

$$f(y_i) = \frac{\lambda_i^{y_t}}{(y_i!)^{\nu}} \frac{1}{Z(\lambda_i, \nu)} \tag{1}$$

where, y_i is the value of the insurance claim corresponding to the i^{th} accident. In Eq. 1, ν corresponds the dispersion index. More specifically, the values ν>1 correspond to equi-,over- and under-dispersion. Since Eq. 1 does not have closed form expressions, we use an asymptotic expression for $Z(\lambda_i,\nu)$ proposed by Shmueli *et al*. (2005) given by:

$$Z(\lambda_i, \nu) \simeq \frac{\exp\left(\nu \lambda_i^{\frac{1}{\nu}}\right)}{\lambda_i^{\frac{\nu-1}{2\nu}} (2\pi)^{\frac{\nu-1}{2}} \sqrt{\nu}} \tag{2}$$

Shmueli *et al*. (2005) derive the following moment expressions:

$$E(Y_i) = \theta_i = \lambda_i^{1/\nu} - \frac{\nu - 1}{2\nu} \tag{3}$$

and

$$\mathrm{Var}(Y_i) = -\frac{\lambda_i^{1/\nu}}{\nu} \tag{4}$$

To estimate the parameters λ and ν, we consider the Quasi-likelihood Equation (QLE) developed by Wedderburn (1974). We extend his approach and review the joint quasi-likelihood estimating equations and develop marginal quasi-likelihood estimating equations. The joint quasi-likelihood equation is given by:

$$\sum_{i=1}^{1} D_i^T V_i^{-1} (f_i - \mu_i) = 0 \tag{5}$$

Where:

$$f_i = \left(y_i, y_i^2\right)^T$$

$$\mu_i = E(f_i)$$

$$V_i = \mathrm{cov}(f_i)$$

$$D_i = \frac{\partial E(f_i)}{\partial(\lambda, \nu)}$$

The matrix components are as follows:

$$D_i = \begin{pmatrix} \partial\theta_i / \partial\lambda & \partial\theta_i / \partial\nu \\ \partial m_i / \partial\lambda & \partial m_i / \partial\nu \end{pmatrix}$$

Where:

$$\partial\theta_i / \partial\lambda = \frac{\lambda_i^{\frac{1}{\nu}}}{\nu} \tag{6}$$

$$\partial\theta_i / \partial\nu = \frac{1}{2}\frac{\nu - 1}{2\nu^2} - \frac{1}{2\nu} - \frac{\lambda_i^{\frac{1}{\nu}}}{\nu^2} \tag{7}$$

$$\partial m_i / \partial\lambda = \left(\frac{2\lambda_i^{\frac{1}{\nu}} + 2\nu\lambda_i^{\frac{2}{\nu}} - \nu\lambda_i^{\frac{1}{\nu}}}{\nu^2} \right) \tag{8}$$

$$\partial m_i / \partial\nu = \frac{1}{2\nu^3} \begin{bmatrix} 2\lambda_i^{\frac{1}{\nu}} \nu \ln(\lambda_i) + \nu - 1 \\ -4\lambda_i^{\frac{2}{\nu}} \ln(\lambda_i)\nu - 4\lambda_i^{\frac{1}{\nu}}\nu - 4\lambda_i^{\frac{1}{\nu}} \ln(\lambda_i) \end{bmatrix} \tag{9}$$

The covariance matrix of f_i is expressed as:

$$V_i = \begin{pmatrix} \mathrm{var}(Y_i) & \mathrm{cov}(Y_i, Y_i^2) \\ & \mathrm{var}(Y_i^2) \end{pmatrix}$$

The elements in V_i are derived iteratively from the moment generating function of y_{it} which is given by:

$$E[Y_i^{r+1}] = \lambda \frac{d}{d\lambda} E[Y^r] + E[Y]E[Y^r] \tag{10}$$

By deriving the moments for Y_i^2, Y_i^3 and Y_i^4, we obtain:

$$\mathrm{cov}(Y_i, Y_i^2) = E(Y_i^3) - E(Y_i)E(Y_i^2)$$
$$\frac{2\lambda_i^{\frac{1}{\nu}} + 2\nu\lambda_i^{\frac{2}{\nu}} - \nu\lambda_i^{\frac{1}{\nu}}}{\nu^2} \tag{11}$$

$$\mathrm{Var}(Y_i^2) = E(Y_i^4) - E(Y_i^2)^2$$
$$\frac{\lambda_i^{\frac{1}{\nu}}\nu^2 + 4\lambda_i^{\frac{3}{\nu}}\nu^2 + 10\lambda_i^{\frac{2}{\nu}}\nu - 4\lambda_i^{\frac{1}{\nu}}\nu + 4\lambda_i^{\frac{1}{\nu}} - 4\lambda_i^{\frac{2}{\nu}}\nu^2}{\nu^3} \tag{12}$$

The QL estimates of λ and ν are obtained by solving Eq. 7 iteratively until convergence using Newton-Raphson technique. At r^{th} iteration:

$$\begin{pmatrix}\hat{\lambda}_{r+1}\\ \hat{\nu}_{r+1}\end{pmatrix}=\begin{pmatrix}\hat{\lambda}_{r}\\ \hat{\nu}_{r}\end{pmatrix}+\left[\sum_{i=1}^{1}D_i^T V_i^{-1}D_i\right]_r^{-1}\left[\sum_{i=1}^{1}D_i^T V_i^{-1}(f_i-\mu_i)\right]_r \qquad (13)$$

Where:

$\hat{\lambda}_r$ = The value of $\hat{\lambda}$ at the rth iteration

$[.]_r$ = The value of the expression at the rth iteration

The estimators are consistent and under mild regularity conditions, for I→∞, it may be shown that $I^{\frac{1}{2}}((\hat{\lambda},\hat{\nu})-(\lambda,\nu))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I\left[\sum_{i=1}^{I}D_i^T V_i^{-1}D_i\right]^{-1}\left[\sum_{i=1}^{I}D_i^T V_i^{-1}(f_i-\mu_i)(f_i-\mu_i)^T V_i^{-1}D_i\right]$ $\left[\sum_{i=1}^{I}D_i^T V_i^{-1}D_i\right]^{-1}$.

The marginal quasi-likelihood equations under the Com-Poisson regression model is as follows: The first QLE is to estimate λ while the second QLE is to estimate the dispersion index ν. The QLE to estimate λ is given by:

$$\sum_{i=1}^{1}D_{i,\lambda}^T V_{i,\lambda}^{-1}(y_i-\theta_i)=0 \qquad (14)$$

where:

$$V_{i,\lambda}=\frac{\lambda_i^{1/\nu}}{\nu} \qquad (15)$$

and:

$$D_{i,\beta}=\frac{\partial\theta_i}{\partial\lambda}=\frac{\lambda_i^{\frac{1}{\nu}}}{\nu} \qquad (16)$$

The QLE to estimate ν is given by:

$$\sum_{i=1}^{1}D_{i,\alpha}^T V_{i,\alpha}^{-1}(y_i^2-\eta_i)=0 \qquad (17)$$

where:

$$\eta_i=E(Y_i^2)=\frac{\lambda_i^{1/\nu}}{\nu}+\left[\lambda_i^{1/\nu}-\frac{\nu-1}{2\nu}\right]^2$$
$$\sum_{i=1}^{1}D_{i,\alpha}^T V_{i,\alpha}^{-1}(y_i^2-\eta_i)=0 \qquad (18)$$

and:

$$D_{i,\nu}=\frac{1}{2\nu^3}\left[2\lambda_i^{\frac{1}{\nu}}\ln(\lambda_i)+\nu-1-4\lambda_i^{\frac{2}{\nu}}\ln(\lambda_i)\nu\right]$$
$$+\frac{1}{2\nu^3}\left[-4\lambda_i^{\frac{1}{\nu}}\nu-4\lambda_i^{\frac{1}{\nu}}\ln(\lambda_i)\right] \qquad (19)$$

$V_{i,\nu}$ is the variance of Y_i^2 and is calculated using:

$$V_{i,\nu}=E(Y_i^4)-E(Y_i^2)^2 \qquad (20)$$

where the moments are derived iteratively from the moment generating function. The Newton-Raphson technique is then applied to the two estimating equations. The iterative equations are given as follows: At the rth iteration:

$$\left(\hat{\lambda}_{r+1}\right)=\left(\hat{\lambda}_r\right)+\left[\sum_{i=1}^{1}D_{i,\lambda}^T V_{i,\lambda}^{-1}D_{i,\lambda}\right]_r^{-1}\left[\sum_{i=1}^{1}D_{i,\lambda}^T V_{i,\lambda}^{-1}(y_i-\theta_i)\right] \qquad (21)$$

$$\left(\hat{\nu}_{r+1}\right)=\left(\hat{\nu}_r\right)+\left[\sum_{i=1}^{1}D_{i,\nu}^T V_{i,\nu}^{-1}D_{i,\nu}\right]_r^{-1}\left[\sum_{i=1}^{1}D_{i,\nu}^T V_{i,\nu}^{-1}(y_i^2-\eta_i)\right] \qquad (22)$$

Where:

$\hat{\lambda}_r$ and $\hat{\nu}_r$ = The values of $\hat{\lambda}$ and $\hat{\nu}_r$ at the rth iteration

$[.]_r$ = The value of the expression at the rth iteration

The estimators are consistent and under mild regularity conditions, for I→∞, it may be shown that $I^{\frac{1}{2}}((\hat{\lambda})-(\lambda))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I\left[\sum_{i=1}^{I}D_{i,\lambda}^T V_{i,\lambda}^{-1}D_{i,\lambda}\right]^{-1}\left[\sum_{i=1}^{I}D_{i,\lambda}^T V_{i,\lambda}^{-1}(y_i-\theta_i)(y_i-\theta_i)^T V_{i,\lambda}^{-1}D_{i,\lambda}\right]$ $\left[\sum_{i=1}^{I}D_{i,\lambda}^T V_{i,\lambda}^{-1}D_{i,\lambda}\right]^{-1}$ and $I^{\frac{1}{2}}((\hat{\nu})-(\nu))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I\left[\sum_{i=1}^{I}D_{i,\nu}^T V_{i,\nu}^{-1}D_{i,\nu}\right]^{-1}\left[\sum_{i=1}^{I}D_{i,\nu}^T V_{i,\nu}^{-1}(y_i-\eta_i)(y_i-\eta_i)^T V_{i,\nu}^{-1}D_{i,\nu}\right]$ $\left[\sum_{i=1}^{I}D_{i,\nu}^T V_{i,\nu}^{-1}D_{i,\nu}\right]^{-1}$. The algorithm to estimate the parameters works as follows: For an initial estimate of λ and ν, we iterate Eq. 14 until convergence, then use the updated λ to update ν in Eq. 17. We then replace the updated λ and ν in Eq. 14 and iterate until convergence. Having obtained the new λ, we replace in Eq. 17 to obtain a new ν and the cycle continues until both values converge.

RESULTS AND DISCUSSION

The first set of insurance claim data is taken from Klugman *et al*. (2008), whereby it was collected by Dropkin in 1956-1958 and analyzed in a paper in Dropkin (1959). The methods JGQL and MGQL are implemented in MATLAB. The fitted values and estimates of λ and ν are provided in Table 1.

Table 1: Fits of the insurance claims under estimated value of $\lambda \mathrm{M\hat{G}QL} = 1.241$, $\nu \mathrm{M\hat{G}QL} = 0.521$ and $\lambda \mathrm{J\hat{G}QL} = 1.252$, $\nu \mathrm{J\hat{G}QL} = 0.527$

No. of accidents	No. of claims	MGQL	JGQL
0	81,714	81,719.2	81,719.5
1	11,306	11,295.1	11,295.7
2	1,618	1622.1	1623.2
3+	297	286.5	287.1

Table 2: Fits of the insurance claims under estimated value of $\lambda \mathrm{M\hat{G}QL} = 5.241$, $\nu \mathrm{M\hat{G}QL} = 0.3112$ and $\lambda \mathrm{J\hat{G}QL} = 5.233$, $\nu \mathrm{J\hat{G}QL} = 0.3125$

No. of accidents	No. of claims	MGQL	JGQL
0	7,840	7842.10	7843.00
1	1,317	1,322.20	1321.20
2	239	242.50	243.20
3	42	43.10	43.10
4	14	13.10	13.30
5	4	3.80	4.10
6	4	3.40	3.60
7	1	0.99	1.10
8+	0	0.20	0.21

The second set of data is taken from Klugman *et al*. (2008) and provides information on 9,461 automobile insurance policies. The methods JGQL and MGQL are implemented in MATLAB. The fitted values and estimates of ν and ν are provided in Table 2.

CONCLUSION

The Table 1 and 2 show the fitted values of the insurance claims. It is clear that under both MGQL and JGQL, CMP yields suitable fits and when compared with the analysis of Zamani and Ismail (2010), we note there is no huge difference in the value of the estimates. Thus, the Com-Poisson model is a very suitable model to analyze actuarial counts and the quasi-likelihood estimation technique is an efficient estimation procedure in terms of both computational and statistical performance.

REFERENCES

Dropkin, L.B., 1959. Some considerations on automobile rating systems utilizing individual driving records. PCAS, 46: 165. http://casualtyactuaries.com/pubs/proceed/proceed59/59165.pdf

Jahangeer, C., N. Mamode Khan and M. Heenaye Mamode Khan, 2009. Analyzing the factors influencing exclusive breastfeeding using the generalized Poisson regression model. Int. J. Math. Stat. Sci., 1: 107-110. http://www.waset.org/journals/ijmss/v1/v1-2-19.pdf

Johnson, N.L., S. Kotz and A.W. Kemp, 1993. Univariate Discrete Distributions. Wiley, New York, 2nd Edn., John Wiley and Sons, ISBN: 0-471-54897-9, pp: 163.

Jowaheer, V. and N. Mamode Khan, 2009. Estimating regression effects in Com-Poisson generalized linear model. Int. J. Math. Comput. Sci., 3: 169-174. http://www.waset.org/journals/ijcms/v3/v3-4-35.pdf

Kadane, J., G. Shmueli, G. Minka, T. Borle and P. Boatwright, 2006. Conjugate analysis of the Conway Maxwell Poisson distribution. Bayesian Anal., 1: 363-374.

Klugman, S.A., H.H. Panjer and G.E. Willmot, 2008. Loss Models: From Data to Decision. 3rd Edn., John Wiley and Sons, USA., ISBN: 10: 0470187816, pp: 101-159.

Lord, D., S. Guikema and S. Geedipally, 2008. Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. Accident Anal. Prev., 40: 1123-1134.

Shmueli, G., T. Minka, J. Borle and P. Boatwright, 2005. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. J. R. Stat. Soc. Ser. C., 54: 127-142. DOI: 10.1111/j.1467-9876.2005.00474.x

Wedderburn, R., 1974. Quasi-likelihood functions, generalized linear models and the Gauss Newton method. Biometrics, 61: 439-447. DOI:10.1093/biomet/61.3.439

Zamani, H. and N. Ismail, 2010. Negative binomial-Lindley distribution and its application. J. Math. Stat., 6: 4-9. http://www.scipub.org/fulltext/jms2/jms2614-9.pdf