PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **19** (2009), 207–216

PLISKA
STUDIA MATHEMATICA
BULGARICA

A NOTE ON BAYESIAN ESTIMATION FOR THE NEGATIVE-BINOMIAL MODEL

Y. L. Lio

The Negative Binomial model, which is generated by a simple mixture model, has been widely applied in the social, health and economic market prediction. The most commonly used methods were the maximum likelihood estimate (MLE) and the moment method estimate (MME). Bradlow et al. (2002) proposed a Bayesian inference with beta-prime and Pearson Type VI as priors for the negative binomial distribution. It is due to the complicated posterior densities of interest not amenable to closed-form integration. A polynomial type expansion for the gamma function had been used to derive approximations for posterior densities by Bradlow et al. (2002). In this note, different parameters of interest are used to re-parameterize the model. Beta and gamma priors are introduced for the parameters and a sampling procedure is proposed to evaluate the Bayes estimates of the parameters. Through the computer simulation, the Bayesian estimates for the parameters of interest are studied via mean squared error and variance. Finally, the proposed Bayesian estimate is applied to model two real data sets.

1. Introduction

The negative-binomial model is generated in the following manner. Consider a population in which the count data, X_i, for each individual member has a Poisson process with the rate parameter λ. Thus given, λ, X_i has the probability function as follows:

2000 *Mathematics Subject Classification:* 62F15.

Key words: Negative Binomial Model, Bayes Estimation, Prior Distribution, Posterior Distribution.

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots \tag{1.1}$$

Suppose that λ varies across the population and has a gamma with shape parameter $\gamma > 0$ and scale parameter $\alpha > 0$:

$$f(\lambda|\gamma, \alpha) = \frac{\alpha^\gamma \lambda^{\gamma-1} e^{-\alpha\lambda}}{\Gamma(\gamma)}, \quad \lambda > 0, \tag{1.2}$$

where $\Gamma(\gamma)$ is the gamma function. Since λ is not observable, the marginal distribution for X has the following simple mixture probability function:

$$Pr_{NB}(x|\gamma, \alpha) = \int f(\lambda|\gamma, \alpha) p(x|\lambda) d\lambda. \tag{1.3}$$

Therefore, the negative binomial model can be rewritten as,

$$Pr_{NB}(x|\gamma, \alpha) = \frac{\Gamma(\gamma + x)}{x!\Gamma(\gamma)} \alpha^\gamma (1 + \alpha)^{\gamma+x}, \quad x = 0, 1, 2, \ldots \tag{1.4}$$

The first application of a negative binomial model was presented by Greenwood and Yule (1920) to model accident statistics. Since then, the negative binomial model has been applied to model phenomena as diverse as the purchasing of consumer packaged goods (Ehrenberg, 1959), salesperson productivity (Carroll et al., 1986), library circulation (Burrell, 1990), and in the Biological sciences (Breslow 1984, Margolin et al. 1981). Letting $\mu = \dfrac{\gamma}{\alpha}$, the Negative-binomial can be re-parameterized as:

$$Pr_{NB}(x|\mu, \gamma) = \frac{\Gamma(\gamma + x)}{x!\Gamma(\gamma)} \left(1 + \frac{\mu}{\gamma}\right)^{-\gamma} \left(\frac{\mu}{\gamma + \mu}\right)^x, \quad x = 0, 1, 2, \ldots \tag{1.5}$$

Here, $\mu = E(X)$ and $var(X) = \mu + \dfrac{\mu^2}{\gamma}$. Anscombe (1950) observed that the maximum likelihood estimate (MLE) of γ does not have a distribution, since there is a finite probability of observing a data set from which the MLE of γ may not be calculated. (This occurs when the sample mean exceeds the sample variance.) Letting $\beta = 1/\gamma$, the Negative-binomial probability function $Pr_{NB} = (x|\mu, \gamma)$ can be re-parameterized as follows:

$$Pr_{NB}(x|\mu, \gamma) = \frac{\Gamma(\beta^{-1} + x)}{x!\Gamma(\beta^{-1})} (1 + \beta\mu)^{-1/\beta} \left(\frac{\beta\mu}{1 + \beta\mu}\right)^x, \quad x = 0, 1, 2, \ldots \tag{1.6}$$

Clark and Perry (1989) adapted the extended quasi-likelihood for the Negative-binomial distribution given by McCullagh and Nelder (1983), proposed the maximum quasi-likelihood estimate (MQLE) for β and found that many simulation cases had negative MQLEs and negative moment method estimates (MME). Piegorsch (1990) studied MLE for β and compared MLE with MQLEs of Clark and Perry (1989). Again, Piegorsch (1990) showed that many simulation cases had negative MLE values for β. Bradlow et al. (2002) introduced a beta-prime prior on α and a Pearson type VI prior on γ in the Negative-binomial (1.4). Approximating the ratio of two gamma functions using a polynomial expansion in the posterior, Bradlow et al. (2002) arrived at closed-form expressions for the moments of both the marginal posterior densities and the predictive distribution. However, the closed-form expressions involve an infinite series of complicated terms which cannot be programmed easily. Bradlow et al. (2002) used finite-term approximations to the infinite series for the simulation studies. Letting $p = \alpha/(\alpha + 1)$, equation (1.4) can be re-parameterized as:

$$Pr_{NB}(x|\gamma, p) = \frac{\Gamma(\gamma + x)}{x!\Gamma(\gamma)}(p)^{\gamma}(1 - p)^{x}, \quad x = 0, 1, 2, \ldots \tag{1.7}$$

where $0 < p < 1$ and $\gamma > 0$.

In this note, the Beta prior and the Gamma prior are introduced for p and γ, respectively. The Bayes estimations for the parameters will be studied. Section 2 describes the models in a Bayesian framework with input data. Section 3 presents and discusses the results of the simulation. Examples of fitting the model to two real data sets are presented in Section 4, and Section 5 contains concluding remarks.

2. Formulas for estimations and predictions

Let a sample of size N be selected from the model given in (1.7). Then the data will consist of the number of occurrences, k, for each of the N units. The data can be summarized as $n_k, k = 0, 1, 2, \ldots, r$, where n_k is the number of units with k occurrences, r is the largest possible occurrence for each sample and $N = \sum_{k=0}^{r} n_k$. Hence, the likelihood function is given as:

$$L(\gamma, p) = \prod_{k=0}^{r} Pr_{NB}(k|\gamma, p)^{n_k}. \tag{2.1}$$

Assume that γ and p are independent *a priori* and have a Gamma distribution with shape parameter δ_2, scale parameter δ_1 and a beta distribution with shape parameters α_1 and α_2, respectively. Then the joint prior distribution of γ and p is expressed as:

$$g(\gamma, p) \propto p^{\alpha_1 - 1}(1 - \alpha)^{\alpha_2 - 1}\gamma^{\delta_2 - 1}e^{-\delta_1\gamma}, \quad 0 < p < 1, \quad 0 < \gamma. \tag{2.2}$$

Combining the likelihood function (2.1) with the joint prior $g(\gamma, p)$ (2.2), the joint posterior distribution of γ and p, given data $n_k, k = 0, 1, 2, \ldots, r$, can be presented as follows:

$$(2.3) \quad Pr(\gamma, p|r, n_k) = \prod_{k=0}^{r} \left[\frac{\Gamma(\gamma + k)}{k!\Gamma(\gamma)} (p)^{\gamma} (1-p)^k \right]^{n_k} g(\gamma, p)/\Phi(r, n_k),$$
$$0 < p < 1, \quad 0 < \gamma,$$

where

$$(2.4) \quad \Phi(r, n_k) = \iint L(\gamma, p|r, n_k) g(\gamma, p) dp d\gamma.$$

$Pr(\gamma, p|r, n_k)$ can also be rewritten as:

$$Pr(\gamma, p|r, n_k) \propto \prod_{k=1}^{r} \left[(\gamma + k - 1)^{N - \sum_{i=0}^{k-1} n_i} (1-p)^{k n_k} \right] (p)^{N\gamma + \alpha_1 - 1} (1-p)^{\alpha_2 - 1}$$
$$(2.5) \quad \delta_1^{\delta_2} \gamma^{\delta_2 - 1} e^{-\delta_1 \gamma}, \quad 0 < p < 1, \quad 0 < \gamma.$$

The equation (2.5) can be represented as follows:

$$Pr(\gamma, p|r, n_k) \propto f_1(\gamma, p, n_k)(p)^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1} \delta_1^{\delta_2} \gamma^{\delta_2 - 1} e^{-\delta_1 \gamma}, 0 < p < 1, 0 < \gamma.$$

Let $K(r, n_k) = \iint f_1(\gamma, p, n_k)(p)^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1} \delta_1^{\delta_2} \gamma^{\delta_2 - 1} e^{-\delta_1 \gamma} dp d\gamma$. The marginal posterior distribution of γ could be obtained as $Pr(\gamma|r, n_k) = \int Pr(\gamma, p|r, n_k) dp$ and the marginal posterior distribution of p could be obtained as $Pr(p|r, n_k) = \int Pr(\gamma, p|r, n_k) d\gamma$. Since the marginal posterior distributions for γ and for p are not amenable to closed-form integration, the moments of the marginal posterior can be computed in the following ways:

$$E(\gamma^l) = \iint \gamma^l Pr(\gamma, p|r, n_k) dp d\gamma$$
$$(2.6) \quad \propto \iint \gamma^l f_1(\gamma, p, n_k)(p)^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1} \delta_1^{\delta_2} \gamma^{\delta_2 - 1} e^{-\delta_1 \gamma} dp d\gamma$$

and

$$E(p^l) = \iint p^l Pr(\gamma, p|r, n_k) dp d\gamma$$
$$(2.7) \quad \propto \iint p^l f_1(\gamma, p, n_k)(p)^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1} \delta_1^{\delta_2} \gamma^{\delta_2 - 1} e^{-\delta_1 \gamma} dp d\gamma.$$

When $l = 1$, (2.6) and (2.7) are the Bayesian estimates of γ and p, respectively (under quadratic loss).

Let Y_i be the count variable for individual i in a non-overlapping time period, which is equal to the length of the time period for the count variable, X_i. For the Bayesian prediction in the negative binomial model, mean, $E(Y_i|x)$, and variance,$Var(Y_i|x)$, of the predictive distribution are emphasized in this note. Given γ and p, the mean and variance of Y_i, conditional on x_i, are $E(Y_i|\gamma, x_i) = (\gamma + x_i)(1-p)$ and $Var(Y_i|\gamma, x_i) = (\gamma + x_i)((1-p) + (1-p)^2)$. Hence,

$$(2.8) \quad E(Y_i|x_i) = \iint (\gamma + x_i)(1-p)Pr(\gamma, p|r, n_k)dpd\gamma \propto \iint (\gamma + x_i)(1-p)f_1(\gamma, p, n_k)(p)^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1}\delta_1^{\delta_2} \gamma^{\delta_2 - 1}e^{-\delta_1\gamma}dpd\gamma$$

and

$$(2.9) \quad Var(Y_i|x_i) = \iint (\gamma + x_i)((1-p) + (1-p)^2)Pr(\gamma, p|r, n_k)dpd\gamma \propto \iint (\gamma + x_i)((1-p) + (1-p)^2)f_1(\gamma, p, n_k)(p)^{\alpha_1 - 1} (1-p)^{\alpha_2 - 1}\delta_1^{\delta_2}\gamma^{\delta_2 - 1}e^{-\delta_1\gamma}dpd\gamma.$$

It should be mentioned that the proportionality for (2.6), (2.7), (2.8) and (2.9) is the reciprocal of $K(r, n_k)$. Since the closed forms for all of the above moments of posteriors are not available, an importance simulation procedure which will be described in the Section 3 is applied to implement the Bayesian estimations.

3. Simulation studies

As the closed forms for the posterior lth moments of γ and p and the conditional mean and variance of Y_i are not available given x, an importance sampling method can be used to estimate those parameters. The importance sampling process for Bayesian estimations is described in the following steps.

1. Randomly generate observations, p_i, of size n from a Beta distribution with parameters α_1 and α_2, and randomly generate observations, γ_j, of size m from a Gamma distribution with scale parameter δ_1 and shape parameter δ_2.
2. Calculate $\hat{K}(r, n_k) = \sum\sum f_1(\gamma_j, p_i, n_k)$ as the estimate of $K(r, n_k)$ and $\rho_l(r, n_k) = \sum\sum \gamma_j^l f_1(\gamma_j, p_i, n_k)$ and $\theta_l(r, n_k) = \sum\sum p_i^l f_1(\gamma_j, p_i, n_k)$.

3. $E(\gamma^l)$ is estimated by $\rho_l(r, n_k)/\hat{K}(r, n_k)$ and $E(p^l)$ is estimated by $\theta_l(r, n_k)/\hat{K}(r, n_k)$.

Similarly, $E(Y_i|x)$ can be estimated by $\sum\sum(\gamma_j + x)(1-p)f_1(\gamma_j, p_i, n_k)/\hat{K}(r, n_k)$ and $Var(Y_i|x)$ can be estimated by $\sum\sum(\gamma_j + x)((1-p) + (1-p)^2)f_1(\gamma_j, p_i, n_k)/\hat{K}(r, n_k)$. The main purpose of this section is to estimate the parameters, γ and p, for each of the nine models by using a computing simulation process. These nine models represent quite distinct negative-binomial models which have γ selected from 1.00, 2.00 and 4.82 and p selected from 0.25, 0.50 and 0.75. Assume that the prior for p is non-informative prior which has $\alpha_1 = 1$ and $\alpha_2 = 1$ and the prior of γ is the Gamma distribution which has δ_1 selected from 0.5, 1.0 or 2.0 and δ_2 selected from 0.5, 1.0, 2.0 or 4.5. Given a Negative-binomial model mentioned in this section, 1000 samples of size 100 are generated. For each random sample of size 100, we have $r, n_k, k = 0, 1, 2, \ldots, r$ and $100 = \sum_{k=0}^{r} n_k$. A Bayes' estimate, $\hat{\gamma}$, for γ and a Bayes' estimate, $\hat{p}$, for p are calculated via Steps 1-3 of the simulation procedure using a sample of size 100 from a non-informative Beta distribution of p and a sample of size 1000 from one of the Gamma prior distributions. Then the mean squared error (MSE), the variance (VAR) and the mean absolute deviation (MAD) for the Bayes' estimator of γ and the mean squared error (MSE), the variance (VAR) and the mean absolute deviation (MAD) for the Bayes' estimator of p are calculated from the 1000 Bayes' estimates of γ and the 1000 Bayes' estimates of p, respectively. The simulation was conducted in the R language (R Development Core Team, 2006), which is a non-commercial, open source software package for statistical computing and graphics that was originally developed by Ihaka and Gentleman (1996). This can be obtained at no cost from `http://www.r-project.org`. Tables 1, 2 and 3 show the parts of simulation results. In general, the MSEs, Mads and VARs for the Bayes' estimates of p are consistently small and are less sensitive to the priors than the MSEs, MADs and VARs for the Bayes' estimates of γ. When misinformed prior are given for γ such as a Gamma prior of $\delta_1 = 1.5$ and $\delta_2 = 4.5$ in Table 1 or a Gamma prior of $\delta_1 = 1$ and $\delta_2 = 1$ in Table 3 the resulting Bayes' estimator of γ has largest MSEs, VARs and MADs among all the Bayes' estimators of γ.

4. Examples

In this section, two real data sets, shown in the most left two columns of Tables 4 and 5, are used to demonstrate the model fitting via Bayes' procedure. Table 4 presents the counts of red miles on apple leaves published in Table 1 of Bliss and Fisher (1953) and Table 5 shows the consumer purchase data from a continuous purchase diary panel reported by Paull (1978). Again, the non-informative Beta

Table 1: Comparison of Estimation with different priors using simulated data with $n = 100$

Estimator	$\gamma = 1.0$	$p = 0.75$	$\gamma = 1.0$	$p = 0.50$	$\gamma = 1.0$	$p = 0.25$
	Gamma Prior: $\delta_1 = 1$ $\delta_2 = 1$					
MSE	0.179557	0.004448	0.087331	0.003853	0.02607	0.001077
MAD	0.323705	0.052176	0.223079	0.050689	0.12336	0.025848
VAR	0.152805	0.004233	0.074866	0.003701	0.02460	0.001036
	Gamma Prior: $\delta_1 = 1.5$ $\delta_2 = 4.5$					
MSE	1.323391	0.008262	0.251243	0.007492	0.040073	0.001500
MAD	1.016934	0.081787	0.382288	0.070267	0.151633	0.030315
VAR	0.311923	0.002722	0.128836	0.004265	0.028215	0.001138
	Gamma Prior: $\delta_1 = 2$ $\delta_2 = 2$					
MSE	0.09494	0.003617	0.070360	0.003453	0.025040	0.001059
MAD	0.24297	0.046509	0.202435	0.047655	0.120301	0.025470
VAR	0.08758	0.003314	0.061450	0.003345	0.023756	0.001023
	Gamma Prior: $\delta_1 = 0.5$ $\delta_2 = 0.5$					
MSE	0.320178	0.0050550	0.109089	0.004256	0.026132	0.001082
MAD	0.416518	0.0561287	0.238814	0.052389	0.122382	0.025677
VAR	0.250235	0.0049468	0.092561	0.004049	0.024670	0.001042

Table 2: Comparison of Estimation with different priors using simulated data with $n = 100$

Estimator	$\gamma = 2.0$	$p = 0.75$	$\gamma = 2.0$	$p = 0.50$	$\gamma = 2.0$	$p = 0.25$
	Gamma Prior: $\delta_1 = 1$ $\delta_2 = 1$					
MSE	0.276458	0.004240	0.420968	0.004721	0.155142	0.001463
MAD	0.435936	0.050983	0.495805	0.055685	0.300584	0.029966
VAR	0.272459	0.003023	0.318147	0.003844	0.152815	0.001460
	Gamma Prior: $\delta_1 = 1.5$ $\delta_2 = 4.5$					
MSE	0.649105	0.002659	1.014109	0.009184	0.234313	0.002043
MAD	0.671170	0.041486	0.831086	0.082107	0.368583	0.035343
VAR	0.316033	0.002447	0.388099	0.003580	0.175435	0.001647
	Gamma Prior: $\delta_1 = 1.0$ $\delta_2 = 2$					
MSE	0.336196	0.004375	0.714790	0.006681	0.175175	0.001606
MAD	0.467596	0.050117	0.654885	0.067385	0.318833	0.030862
VAR	0.325430	0.003603	0.407898	0.004053	0.157175	0.001508
	Gamma Prior: $\delta_1 = 0.5$ $\delta_2 = 0.5$					
MSE	0.439792	0.007603	0.850066	0.007009	0.214090	0.001868
MAD	0.538471	0.067097	0.687934	0.067822	0.347927	0.033497
VAR	0.439387	0.005147	0.550115	0.004899	0.198489	0.001782

Table 3: Comparison of Estimation with different priors using simulated data with $n = 100$

Estimator	$\gamma = 4.82$	$p = 0.75$	$\gamma = 4.82$	$p = 0.50$	$\gamma = 4.82$	$p = 0.25$
	Gamma Prior: $\delta_1 = 1$ $\delta_2 = 1$					
MSE	3.033319	0.013333	0.790359	0.002682	0.645612	0.001087
MAD	1.625727	0.105445	0.730432	0.041957	0.649552	0.026600
VAR	0.395297	0.002221	0.667995	0.002198	0.584681	0.000947
	Gamma Prior: $\delta_1 = 1.5$ $\delta_2 = 4.5$					
MSE	1.622687	0.006589	0.598370	0.001863	0.560104	0.000942
MAD	1.122014	0.069939	0.625610	0.035148	0.597425	0.024507
VAR	0.421808	0.001764	0.591799	0.001825	0.550842	0.000909
	Gamma Prior: $\delta_1 = 1$ $\delta_2 = 2$					
MSE	2.193756	0.009467	0.796009	0.002343	0.651842	0.001071
MAD	1.318259	0.085680	0.718760	0.039652	0.643820	0.026129
VAR	0.503779	0.002187	0.790702	0.002284	0.640656	0.001031
	Gamma Prior: $\delta_1 = 0.5$ $\delta_2 = 0.5$					
MSE	1.966285	0.008827	1.454981	0.003169	0.847494	0.001264
MAD	1.195164	0.078841	0.939309	0.045719	0.710896	0.028211
VAR	0.907115	0.002997	1.344393	0.003093	0.843361	0.001264

prior for p is used with Gamma prior 1 ($\delta_1 = 1.0$ and $\delta_2 = 1.0$), Gamma prior 2 ($\delta_1 = 1.5$ and $\delta_2 = 4.5$) and Gamma prior 3 ($\delta_1 = 1.0$ and $\delta_2 = 2.0$), respectively. The Negative-binomial fittings are presented in the three right columns (NBD_1, NBD_2 and NBD_3, for Gamma priors 1, 2 and 3, respectively) of Tables 4 and 5. Using data from Table 4, the Bayes' estimates of (γ, p) are $(1.066, 0.4747)$ for using Gamma prior 1, $(1.368, 0.5371)$ for using Gamma prior 2 and $(1.137, 0.4886)$ for using Gamma prior 3. The data set of Table 5 produces the Bayes' estimates of (γ, p) of $(1.665, 0.6802)$ for Gamma prior 1, $(1.637, 0.6736)$ for Gamma prior 2 and $(1.616, 0.6731)$ for Gamma prior 3. The Chi-square test for the goodness of fit for these two data sets has the p-value reported in Tables 4 and 5. For each model fitting we see no evidence to reject the null hypothesis at the 0.05 significance level. Those results show that the impact from different gamma priors is not significantly different, since the sample sizes for both data sets are large.

5. Concluding Remark

Bayesian method for the Negative-binomial model provides a viable alternative to the maximum likelihood approach. Unlike the maximum likelihood estimates and the moment method estimates, the Bayes' estimate always produces values in the feasible regions of parameters. By using the proposed sampling procedure, the

Table 4: Frequency Distribution of Counts of Red Mites

X_i	Observed Frequency	Oberved Percentage	NBD_1 fitted Percentage	NBD_2 fitted Percentage	NBD_3 fitted Percentage
0	70	46.67	45.19	42.73	44.28
1	38	25.33	25.31	27.06	25.76
2	17	11.33	13.73	14.83	14.08
3	10	6.67	7.37	7.71	7.53
4	9	6.00	3.94	3.90	3.98
5	3	2.00	2.10	1.94	2.09
6	2	1.33	1.11	0.95	1.10
7	1	0.67	0.59	0.46	0.57
8	0	0.00	0.31	0.22	0.30
9	0	0.00	0.17	0.11	0.15
Goodness of fit test:		p-Value	0.4747	0.5371	0.4886

Table 5: Frequency Distribution of Consumer Purchase Data

X_i	Observed Frequency	Oberved Percentage	NBD_1 fitted Percentage	NBD_2 fitted Percentage	NBD_3 fitted Percentage
0	382	53.80	52.64	52.38	52.74
1	193	27.19	28.03	27.98	27.86
2	81	11.41	11.94	12.04	11.91
3	37	5.21	4.67	4.76	4.70
4	11	1.55	1.74	1.80	1.77
5	5	0.70	0.63	0.66	0.65
6	0	0.00	0.22	0.24	0.23
7	1	0.14	0.08	0.09	0.08
8	0	0.00	0.02	0.03	0.03
9	0	0.00	0.01	0.01	0.01
Goodness of fit test:		p-Value	0.6802	0.6736	0.6731

Bayesian approach for the Negative-binomial model can be gainfully implemented in real life applications.

Acknowledgments

The author would like to thank Dr. Dan Van Peursem, the editor and anonymous referee for their suggestions and comments, which significantly improved this manuscript.

REFERENCES

[1] F. J. Anscombe. Sampling Theory of the Negative Binomial And Logarithmic Series Distributions. *Biometrika* **36** (1950), 358–382.

[2] C. I. Bliss,R. A. Fisher. Fitting the negative binomial distribution to biological data. *Biometrics* **9** (1953), 176–200.

[3] E. T. Bradlow, B. G. S. Hardie, P. S. Fader. Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions. *Journal of Computational And Graphical Statistics* **11** (2002), 189–201.

[4] N. Breslow. Extra-Poisson Variation in Log-Linear Models. *Applied Statistics* **33** (1984), 38–44.

[5] Q. L. Burrell. Using the Gammer-Poisson Model to Predict Library Circulations. *Journal of the American Society for Information Science* **41** (1990), 164–170.

[6] V. P. Carroll, H. L. Lee, A. G. Rao. Implications of Salesforce Productivity Heterogeneity and Demotivation: A Navy Recruiter Case Study. *Management Science* **32** (1986), 1371–1388.

[7] S. J. Clark, J. N. Perry. Estimation of the Negative Binomial Parameter k by Maximum Quasi-Likelihood. *Biometrics* **45** (1989), 309–316.

[8] A. S. C. Ehrenberg. The Pattern of Consumer Purchases. *Applied Statistics* **8** (1959), 26–41.

[9] M. Greenwood, G. Udny Yule. An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeat Accidents. *Journal of the Royal Statistical Society* **83** (1920), 255–279.

[10] R. Ihaka, R. Gentleman. R:A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5** (1996), 299–314.

[11] B. H. Margolin, N. Kaplan, E. Zeiger. Statistical Analysis of the Ames Salmonella/microsome Test. *Proc. Nat. Acad. Sci. USA* **76** (1981), 3779–3783.

[12] P. McCullagh, J. A. Nelder. Generalized Linear Models. Chapman And Hall/Londa, 1983.

[13] W. W. Piegorsch. Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. *Biometrics* **46** (1990), 863–867.

[14] R Development Core Team. A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.

Y. L. Lio
Department of Mathematical Sciences
University of South Dakota, Vermillion, SD 57069,USA