Size-Biased Generalized Negative Binomial Distribution

Khurshid Ahmad Mir
Govt. Degree College Bemina, mirkhurshid_ahmad@yahoo.co.in

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm
Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
DOI: 10.22237/jmasm/1225512480
Available at: http://digitalcommons.wayne.edu/jmasm/vol7/iss2/9

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
Size-Biased Generalized Negative Binomial Distribution

Khurshid Ahmad Mir
Govt. Degree College Bemina,
Srinagar (J&K) India.

A size biased generalized negative binomial distribution (SBGNBD) is defined and a recurrence relationship for the moments of SBGNBD is established. The Bayes’ estimator for a parametric function of one parameter when two other parameters of a known size-biased generalized negative binomial distribution is derived. Prior information on one parameter is given by a beta distribution and the parameters in the prior distribution are assigned by computer using Monte Carlo and R-software.

Key words: Generalized negative binomial distribution, size-biased generalized negative binomial distribution, zero-truncated generalized negative binomial distribution; size biased negative binomial distribution, goodness of fit, Bayes’ estimation.

Introduction

Jain and Consul (1971) first defined generalized negative binomial distribution (GNBD), and it was subsequently obtained by Consul and Shenton (1972, 1975) as a particular family of the Lagrangian distribution. The parameter space of the distribution was further modified by Consul and Gupta (1995). The probability function of the GNBD is given by

\[ P(X = x) = \frac{m}{m + \beta x} \left( \frac{m + \beta x}{x} \right)^{m + \beta x - 1} \alpha^{x} (1-\alpha)^{m+\beta x - 1}, \]

where

\[ 0 < \alpha < 1, m > 0 \text{ and } |\alpha \beta| < 1. \]

The probability model (1.1) reduces to the binomial distribution when \( \beta = 0 \), and to the negative binomial distribution when \( \beta = 1 \). It also resembles the Poisson distribution at \( \beta = \frac{1}{2} \) because, for this value of \( \beta \), the mean and variance are approximately equal. Jain and Consul (1971) obtained the first four non-central moments by using a recurrence relation and Shoukri (1980) obtained a recurrence relation among the central moments. The model (1.1) has many important applications in various fields of study and is useful in queuing theory and branching processes. Famoye and Consul (1989) considered a stochastic model for the GNBD and gave some other interesting applications of this model. The moments about the origin of the model (1.1) are given as:

\[ \mu_1' = \frac{m \alpha}{(1-\alpha \beta)} \]

(1.2)

\[ \mu_2' = \frac{(m \alpha)^2}{(1-\alpha \beta)^2} + \frac{m \alpha (1-\alpha)}{(1-\alpha \beta)^3} \]

(1.3)

\[ \mu_3' = \frac{(m \alpha)^3}{(1-\alpha \beta)^3} + \frac{3(m \alpha)^2 (1-\alpha)}{(1-\alpha \beta)^4} + \frac{m \alpha (1-\alpha)}{(1-\alpha \beta)^5} [1-2\alpha + \alpha \beta (2-\alpha)] \]
Size-biased generalized negative binomial distribution (SBGNBD) taking the weights of the probabilities as the variate values, are defined in this study. The moments of size-biased GNBD are also obtained. As far as estimation the parameters of a size-biased generalized negative binomial distribution (SBGNBD) is concerned, no method seems to have evolved to date, thus a Bayes’ estimator of size-biased generalized negative binomial distribution is presented. A computer program in R-software has been developed to ease computations while estimating the parameters for data. A goodness of fit test is employed to test the program’s improvement over the Bayes’ estimator of the zero truncated generalized negative binomial distribution (ZTGNBD) and of the size biased negative binomial distribution (SBNBD).

The Truncated Generalized Negative Binomial Distribution

Jain and Consul’s (1997) generalized negative binomial distribution (1.1) can be truncated at \( x = 0 \). The probability function of the zero-truncated GNBD is given by:

\[
P_2(X = x) = \frac{m^m (m + \beta x)^x \alpha^x (1-\alpha)^{m + \beta x - x}}{[1-(1-\alpha)^m]}, \quad x = 1, 2, \ldots
\]  

(2.1)

where \( 0 < \alpha < 1, m > 0 \) and \(|\alpha\beta| \leq 1\).

Bansal and Ganji (1997) obtained the Bayes’ estimator of zero-truncated generalized negative binomial distribution using (1.1); they obtained an estimator of its parameters by using different estimation methods.

Methodology

A size-biased generalized negative binomial distribution (SBGNBD) - a particular case of the weighted generalized negative binomial - taking weights as the variate value is defined and moments of SBGNBD are obtained.
Using (1.1) and (1.2), results in the following:

\[ \sum_{x=0}^{\infty} x \cdot P_1(X = x) = \frac{m\alpha}{(1 - \alpha\beta)} \]

thus,

\[ \sum_{x=1}^{\infty} P_3(X = x) = 1 \]

represents a probability distribution. This gives the size-biased generalized negative binomial distribution (SBGNBD) as:

\[ P_3(X = x) = (1 - \alpha\beta) \left( \frac{m + \beta x - 1}{x - 1} \right) \alpha^{x-1} \left( 1 - \alpha \right)^{m + \beta x - x} \]

\( x = 1, 2, \ldots, \) where \( 0 < \alpha < 1, \) \( m > 0, \) \( |\alpha\beta| < 1 \)

(3.1)

Putting \( \beta = 0 \) and \( \beta = 1, \) results in size-biased binomial (SBB) and size-biased negative binomial (SBNB) distributions.

Moments of SBGNBD

The \( r \)th moment, \( \mu'_r(s), \) about origin of the size-biased GNBD (3.1) can be defined as:

\[ \mu'_r(s) = \sum_{x=1}^{\infty} x^r \cdot P_3(X = x); r = 1, 2, 3, \ldots \]

(3.2)

\( \mu'_0(s) = 1, \) and for \( r \geq 1, \) and

\[ \mu'_r(s) = \frac{1 - \alpha\beta}{m\alpha} \sum_{x=0}^{\infty} x^{r+1} P_1(X = x) \]

\[ \mu'_r(s) = \frac{1 - \alpha\beta}{m\alpha} \mu'_{r+1} \]

(3.3)

where \( \mu'_{r+1} \) is the \((r + 1)\)th moment about the origin of (1.1). The first three moments of (3.1) about the origin using relations from (1.2) to (1.5) in (3.2) can be obtained by:

\[ \mu'_1(s) = \frac{1 - \alpha\beta}{m\alpha} \mu'_2 \]

\[ \mu'_1(s) = \frac{m\alpha}{1 - \alpha\beta} + \frac{1 - \alpha}{(1 - \alpha\beta)^2} \]

(3.4)

which is the mean of (3.1). Similarly, for \( r = 2 \) in (3.2) using relation (1.4):

\[ \mu'_2(s) = \frac{(m\alpha)^2}{(1 - \alpha\beta)^2} + \frac{3m\alpha(1 - \alpha)}{(1 - \alpha\beta)^3} + \frac{(1 - \alpha)}{(1 - \alpha\beta)^4} \left[ 1 - 2\alpha + \alpha\beta(2 - \alpha) \right] \]

(3.5)

Using relation (1.5) for \( r = 3 \) in (3.2) results in:

\[ \mu'_3(s) = \frac{(m\alpha)^3}{(1 - \alpha\beta)^3} + \frac{6(m\alpha)^2(1 - \alpha)}{(1 - \alpha\beta)^4} + \frac{m\alpha(1 - \alpha)}{(1 - \alpha\beta)^5} \left[ 7 - 11\alpha - 4\alpha\beta(2 - \alpha) \right] \]

(3.6)

The variance \( \mu'_2(s) \) of (3.1) using (3.3) and (3.4) is obtained by:

\[ \mu'_2(s) = \frac{m\alpha(1 - \alpha)}{(1 - \alpha\beta)^3} + \frac{\alpha(1 - \alpha)}{(1 - \alpha\beta)^4} \left[ \beta(2 - \alpha) - 1 \right] \]

(3.7)

The higher moments of (3.1) about the origin can also be obtained similarly by using (3.2).

Bayes’ Estimation in Size-biased Generalized Negative Binomial Distribution

The likelihood function of SBGNBD (3.1) is:

\[ L(x | \alpha, \beta) = \frac{(1 - \alpha\beta)^n \prod_{i=1}^{n} \left( \frac{m + \beta x_i - 1}{x_i - 1} \right)^{x_i} (1 - \alpha)^{m + \beta \sum_{i=1}^{n} x_i} (1 - \alpha\beta)^{\sum_{i=1}^{n} x_i}}{\sum_{x=0}^{\infty} x^r \cdot P_3(X = x)} \]

448
\[ = K(1 - \alpha \beta)^n \alpha^{y-n} (1 - \alpha)^{mn+\beta y-y} \quad (4.1) \]

where

\[ y = \sum_{i=1}^{n} x_i \text{ and } K = \prod_{i=1}^{n} \left( m + \beta x_i - 1 \right). \]

Because \( 0 < \alpha < 1 \), it is assumed that prior information about \( \alpha \) came from the beta distribution. Thus,

\[ f(\alpha) = \frac{\alpha^{y-a-1}(1-\alpha)^{b-1}}{B(a,b)} ; 0 < \alpha < 1 , a>0, b>0. \]

(4.2)

Using Bayes’ Theorem, the posterior distribution of \( \alpha \) from (4.1) and (4.2) can be written as:

\[ p(\alpha | y) = \frac{(1 - \alpha \beta)^n \alpha^{y+a-n-1}(1 - \alpha)^{mn+\beta y-y+b-1}}{\int_0^1 (1 - \alpha \beta)^n \alpha^{y+a-n-1}(1 - \alpha)^{mn+\beta y-y+b-1} d\alpha} \cdot \]

(4.3)

Under square error loss function the Bayes’ estimator of parametric function \( \alpha^2 \) is the posterior mean given as

\[ \hat{\alpha}^2 = \int_0^1 \alpha^2 p(\alpha | y) \, d\alpha \]

\[ = \frac{\int_0^1 (1 - \alpha \beta)^n \alpha^{y+a-n+z}(1 - \alpha)^{mn+\beta y-y+b-1} \, d\alpha}{\int_0^1 (1 - \alpha \beta)^n \alpha^{y+a-n-1}(1 - \alpha)^{mn+\beta y-y+b-1} \, d\alpha}. \]

(4.4)

where

\[ \int_0^1 (1 - \alpha \beta)^n \alpha^{y+a-n+z}(1 - \alpha)^{mn+\beta y-y+b-1} \, d\alpha = \]

\[ \Gamma(y+a-n+z) \Gamma(\beta y+mn+b-y) \]

\[ \frac{\frac{1}{\beta} \left(-n, y+a-n+z, \beta y+mn+a+b-n+1, \beta \right)}{\Gamma(\beta y+mn+a+b-n+1)} \]

and

\[ \int_0^1 (1 - \alpha \beta)^n \alpha^{y+a-n-1}(1 - \alpha)^{mn+\beta y-y+b-1} \, d\alpha = \]

\[ \Gamma(y+a-n) \Gamma(\beta y+mn+b-y) \]

\[ \frac{\frac{1}{\beta} \left(-n, y+a-n, \beta y+mn+a+b-n, \beta \right)}{\Gamma(\beta y+mn+a+b-n+1)} \]

(4.5)

Using relations (4.5) and (4.6) in (3.4), the Bayes’ estimator of \( \alpha^2 \) becomes:

\[ \hat{\alpha}^2 = \]

\[ \int_0^1 \frac{(1 - \alpha \beta)^n \alpha^{y+a-n+z}(1 - \alpha)^{mn+\beta y-y+b-1}}{\Gamma(y+a-n+z) \Gamma(\beta y+mn+b-y)} \, d\alpha \]

\[ \frac{\frac{1}{\beta} \left(-n, y+a-n+z, \beta y+mn+a+b-n+1, \beta \right)}{\Gamma(\beta y+mn+a+b-n+1)} \]

(4.6)

Similarly, the Bayes’ estimator of the parametric function \( (1 - \alpha)^2 \) can also be obtained as:

\[ (1 - \alpha)^2 = \]

\[ \int_0^1 \frac{(1 - \alpha \beta)^n \alpha^{y+a-n+1}(1 - \alpha)^{mn+\beta y-y+b-1+z}}{\int_0^1 (1 - \alpha \beta)^n \alpha^{y+a-n-1}(1 - \alpha)^{mn+\beta y-y+b-1} \, d\alpha} \, d\alpha \]

(4.8)

where
Using the values from (4.9) and (4.6) in (4.8), the Bayes’ estimator of the parametric function $(1-\alpha)^\gamma$ can be obtained as

$$(1-\alpha)^\gamma = 
\frac{\Gamma(\beta y + mn + a + b - n) \Gamma(\beta y + mn + a - n - b - y + z)}{\Gamma(\beta y + mn + a + b - n) \Gamma(\beta y + mn + a - n - b - y + z)}
$$

The Bayes’ estimator for some parametric functions $\phi(\alpha)$ and for particular models of SBGNBD are shown in Tables 4.1 and 4.2.

Conclusion

A computer program in R-Software was developed to ease computations while estimating the parameters for data. The expected frequencies and Chi-square obtained are shown in tables 5.1, 5.2 and 5.3. Assuming that the parameter $\alpha$ is unknown and that it has a beta distribution with parameters $a$ and $b$, the Bayes’ relative frequencies are estimated by using the estimator of (2.1) and (3.1). Since no other information is provided about the values of $a$ and $b$, except that they are both positive and real, a range of values from 1 to 50 were considered for $a$ and $b$, and the values of (2.1) and (3.1) were computed. Three sets of simulated values were obtained with the help of R-software: one each for the parameter combination $(a=0.5, b=0.3, a=b=1)$, $(a=0.6, b=0.5, a=b=2)$ and $(a=0.6, b=0.7, a=b=3)$. We noted that the estimated Bayes’ frequencies were quite close to the simulated sample frequencies when $a$ and $b$ were equal and that the variation in the Bayes’ frequencies was very little as the equal values of $a$ and $b$ increased. The graph also reveals that the simulated frequencies and the estimated Bayes’ frequencies are very close to each other for almost all values of $X$.

References


Table 4.1: Bayes’ Estimators of SBGNBD

<table>
<thead>
<tr>
<th>Parametric Function $\phi(\alpha)$</th>
<th>Bayes’ Estimator of SBGNBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$(y+a-n)^{\frac{1}{2}} F_1[-n, y+a-n+1, b+mn + \beta y-n + a + 1, \beta]$</td>
</tr>
<tr>
<td>$1-\alpha$</td>
<td>$(\beta y + mn + a + b - n)^{\frac{1}{2}} F_1[-n, y+a-n, \beta y+mn+a+b-n+1, \beta]$</td>
</tr>
</tbody>
</table>

Table 4.2: Bayes’ $\hat{\alpha}$ Estimators

<table>
<thead>
<tr>
<th>$\beta$ Distribution</th>
<th>Bayes’ Estimator $\hat{\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 SBNBD</td>
<td>$\frac{y+a-n}{y+mn+a+b}$</td>
</tr>
<tr>
<td>0 SBBBD</td>
<td>$\frac{y+a-n}{mn+a+b-n}$</td>
</tr>
</tbody>
</table>
Table 5.1: Number of mothers ($f_x$) in Sri Lanka having at least one neonatal death according to number of neonatal deaths ($x$) Meegama (1980) ($a=b=2$, $m=5$, $\beta=0.3$)

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_x$</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BSBNBD</td>
</tr>
<tr>
<td>1</td>
<td>567</td>
<td>545.25</td>
</tr>
<tr>
<td>2</td>
<td>135</td>
<td>154.67</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>27.31</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>16.61</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2.16</td>
</tr>
<tr>
<td>Total</td>
<td>746</td>
<td>746</td>
</tr>
</tbody>
</table>

Estimates $\hat{\alpha}$

$\chi^2$ 3.7953 3.0477 2.738

Table 5.2: Number of workers ($f_x$) having at least one accident according to number of accidents ($x$) ($a=b=2$, $m=7$, $\beta=0.5$)

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_x$</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BSBNBD</td>
</tr>
<tr>
<td>1</td>
<td>2039</td>
<td>2033.32</td>
</tr>
<tr>
<td>2</td>
<td>312</td>
<td>325.33</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>29.28</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1.95</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.12</td>
</tr>
<tr>
<td>Total</td>
<td>2,390</td>
<td>2,390</td>
</tr>
</tbody>
</table>

Estimates $\hat{\alpha}$

$\chi^2$ 2.428 0.68 0.4077

Table 5.3: Number of households ($f_x$) having at least one migrant according to number of migrants ($x$) Singh and Yadav (1980) ($a=b=2$, $m=9$, $\beta=0.7$)

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_x$</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BSBNBD</td>
</tr>
<tr>
<td>1</td>
<td>375</td>
<td>370.87</td>
</tr>
<tr>
<td>2</td>
<td>143</td>
<td>156.29</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>48.42</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>11.44</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2.40</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.47</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>Total</td>
<td>590</td>
<td>590</td>
</tr>
</tbody>
</table>

Estimates $\hat{\alpha}$

$\chi^2$ 6.9458 4.06227 3.16908
Graph 1: Sample Relative Frequency and Bayes’ Relative Frequency for $a=b=2, m=5, \beta=0.3$

Graph 2: Sample Relative Frequency and Bayes’ Relative Frequency for $a=b=2, m=7, \beta=0.5$

Graph 3: Sample Relative Frequency and Bayes’ Relative Frequency for $a=b=2, m=9, \beta=0.7$