Journal of Statistical Research *ISSN 0256 - 422 X*
2008, Vol. 42, No. 1, pp. 117-126
Bangladesh

ON BAYESIAN ESTIMATION OF SIZE-BIASED GENERALIZED GEOMETRIC SERIES DISTRIBUTION AND ITS APPLICATIONS

Anwar Hassan

PG. Department of Statistics, University of Kashmir, Srinagar, India
Email: anwar.hassan2007@gmail.com

Khurshid Ahmad Mir

Department of Statistics, Amar Singh College, Srinagar, India
Email: khrshdmir@yahoo.com

SUMMARY

A size biased generalized geometric series distribution (SBGGSD) is defined and studied. Estimation of its parameters by the Bayes method is proposed. A goodness of fit is done in order to test its improvement over the zero truncated generalized geometric series distribution (ZTGGSD) and the size biased geometric series distribution (SBGGSD).

Keywords and phrases: Squared error loss function, Bayes estimator, Beta distribution, size-biased geometric series distribution, zero-truncated generalized geometric series distribution.

AMS Classification: 62E10, 62E15

1 Introduction

Mishra (1982) defined the generalized geometric series distribution (GGSD) by using the results of the lattice path analysis with the probability function as

$$P(X=x)=\frac{1}{1+\beta x}\binom{1+\beta x}{x}\alpha^{x}(1-\alpha)^{1+\beta x-x};0<a<1,|\alpha\beta|<1,x=0,1,2,\dots \quad (1.1)$$

It can be seen that at $\beta=1$, the model (1.1) reduces to simple geometric distribution and is a particular case of Jain and Consul's (1971) generalized negative binomial distribution (GNBD) in the same way as the geometric distribution is a particular case of the negative binomial distribution.

The first four moments about origin of the GGSD are given as

$$\begin{aligned}
\mu_1' &= \frac{\alpha}{1-\alpha\beta} \\
\mu_2' &= \frac{\alpha(1-\alpha)}{(1-\alpha\beta)^3} + \frac{\alpha^2}{(1-\alpha\beta)^2} \\
\mu_3' &= \frac{\alpha^3}{(1-\alpha\beta)^3} + \frac{3\alpha^2(1-\alpha)}{(1-\alpha\beta)^4} + \frac{\alpha(1-\alpha)}{(1-\alpha\beta)^5}[1-2\alpha+\alpha\beta(2-\alpha)] \\
\mu_4' &= \frac{\alpha^4}{(1-\alpha\beta)^4} + \frac{6\alpha^3(1-\alpha)}{(1-\alpha\beta)^5} + \frac{\alpha^2(1-\alpha)[7-11\alpha+4\alpha\beta(2-\alpha)]}{(1-\alpha\beta)^6} \\
&\quad + \frac{\alpha(1-\alpha)[1-60+6\alpha^2+2\alpha\beta(4-9\alpha+4\alpha^2)+\alpha^2\beta^2(6-6\alpha+\alpha^2)]}{(1-\alpha\beta)^7} \qquad (1.2)
\end{aligned}$$

The central moments are given as

$$\begin{aligned}
\mu_2 &= \frac{\alpha(1-\alpha)}{(1-\alpha\beta)^3} \\
\mu_3 &= \frac{\alpha(1-\alpha)}{(1-\alpha\beta)^5}[1-2\alpha+\alpha\beta(2-\alpha)] \\
\mu_4 &= \frac{3\alpha^2(1-\alpha)^2}{(1-\alpha\beta)^6} \\
&\quad + \frac{\alpha(1-\alpha)[1-60+6\alpha^2+2\alpha(4-9\alpha+4\alpha^2)+\alpha^2\beta^2(6-6\alpha+\alpha 2)]}{(1-\alpha\beta)^7} \qquad (1.3)
\end{aligned}$$

The various interesting properties and estimation of (1.1) have been discussed by Mishra (1979, 1982), Singh (1989), Mishra and Singh (1992), Hassan (1995) and Hassan et al. (2002, 2003, and 2007). They found this distribution to provide much closer fits to all those observed distributions where the geometric distribution and the various compound geometric distributions have been fitted earlier by many authors. A brief list of authors and their works can be seen in Johnson et al. (1992) and Consul and Famoye (2006).

In this paper, a size-biased geometric series distribution (SBGGSD) taking the weights of the probabilities as the variate values, has been defined. The moments of the size-biased GGSD have also been obtained. As far as estimation of parameters of the size-biased generalized geometric series distribution (SBGGSD) is concerned, no method seems to have been evolved so far. The estimation of the parameters of the GGSD is itself very difficult and only the method of moments give the estimate of the parameters easily but it also fails to provide estimates of the parameters in some cases. In this paper we have proposed a Bayes estimator of the size-biased generalized geometric series distribution. A goodness of fit test is done in order to test its improvement over the zero truncated generalized geometric series distribution (ZTGGSD) and the size-biased generalized geometric series distribution (SBGGSD).

2 Truncated Generalized Geometric Series Distribution

A discrete random variable is said to have a zero truncated generalized geometric series distribution (ZTGGSD) of its probability mass function is given by

$$P_1(X=x) = \frac{1}{1+\beta x}\binom{1+\beta x}{x}\alpha^{x-1}(1-\alpha)^{1+\beta x-x};\ 0<\alpha<1, |\alpha\beta|<1, x=1,2,\ldots. \quad (2.1)$$

The moments of (2.1) about the origin may be obtained by just dividing the corresponding moments of the GGSD (1.1) by α. We get

$$\begin{aligned}
\mu_1' &= \frac{1}{(1-\alpha\beta)} \\
\mu_2' &= \frac{\alpha}{(1-\alpha\beta)^2} + \frac{(1-\alpha)}{(1-\alpha\beta)^3} \\
\mu_3' &= \frac{3\alpha(1-\alpha)}{(1-\alpha\beta)^4} + \frac{\alpha^2}{(1-\alpha\beta)^3} + \frac{(1-\alpha)}{(1-\alpha\beta)^5}[1-2\alpha+\alpha\beta(2-\alpha)] \qquad (2.2)
\end{aligned}$$

The central moments of the zero truncated GGSD (2.1) may be obtained by multiplying the corresponding central moments of the GGSD (1.2) by β. We get

$$\mu_2 = \frac{\alpha\beta(1-\alpha)}{(1-\alpha\beta)^3} \quad \text{and} \quad \mu_3 = \frac{\alpha\beta(1-\alpha)}{(1-\alpha\beta)^5}[1-2\alpha+\alpha\beta(2-\alpha)] \qquad (2.3)$$

3 Size-biased Generalized Geometric Series Distribution

A size biased GGSD is obtained by taking the weight of (1.1) as X. We have from (1.1) and (1.2)

$$\sum_{x=0}^{\infty} xP(X=x) = \frac{\alpha}{1-\alpha\beta} \qquad \text{and}$$

$$\sum_{x=0}^{\infty} x\cdot\frac{1}{1+\beta x}\binom{1+\beta x}{x}\alpha^{x}(1-\alpha)^{1+\beta x-x} = \frac{\alpha}{1-\alpha\beta}.$$

Thus

$$\sum_{x=1}^{\infty}(1-\alpha)\binom{\beta x}{x-1}\alpha^{x-1}(1-\alpha)^{1+\beta x-x} = 1$$

which is implies that $\sum_{x=1}^{\infty} P_2[X=x]=1$ where $P_2[X=x]$ represents a probability function. This gives the size-biased generalized geometric series distribution (SBGGSD) as

$$P_2[X=x] = (1-\alpha\beta)\binom{\beta x}{x-1}\alpha^{x-1}(1-\alpha)^{1+\beta x-x};\ 0<\alpha<1,\ |\alpha\beta|<1,\ x=1,2,\ldots \quad (3.1)$$

When $\beta = 1$ and $\beta = 0$, the SBGGSD (3.1) reduces to the size-biased geometric series distribution and the size biased Bernoulli distributions (SBBD). The r^{th} moment of SBGGSD (3.1) about origin is defined as

$$\mu_r'(s) = \sum_{x=1}^{\infty} x^r \frac{1-\alpha\beta}{\alpha} \binom{\beta x}{x-1} \alpha^x 1-\alpha)^{1+\beta x - x}, \quad r = 1, 2, 3 \tag{3.2}$$

Obviously $\mu_0'(s) = 1$ and for $r \geq 1$

$$\begin{aligned} \mu_r'(s) &= \frac{1-\alpha\beta}{\alpha} \sum_{x=0}^{\infty} x^{r+1} \frac{1}{1+\beta x} \binom{1+\beta x}{x} \alpha^x (1-\alpha)^{1+\beta x - x} \\ &= \frac{1-\alpha\beta}{\alpha} \mu_{r+1}' \end{aligned} \tag{3.3}$$

where μ_{r+1}' is the $(r+1)$-th moments about the origin of the GGSD (1.1).

Using relations (1.3) to (1.5) for $r = 1, 2, 3$ in (3.3), the first three moments about origin of (3.1) we have

$$\mu_1'(s) = \frac{(1-\alpha^2\beta)}{(1-\alpha\beta)^2} \tag{3.4}$$

$$\mu_2'(s) = \frac{1}{(1-\alpha\beta)^4} \Big[\alpha^4\beta^2 + 2\alpha^3\beta - 6\alpha^2\beta + 2\alpha\beta + 1\Big]. \tag{3.5}$$

Also, the variance $\mu_2(s)$ of the SBGGSD (3.1) is

$$\mu_2(s) = \frac{1}{(1-\alpha\beta)^4}[2\alpha^3\beta - 4\alpha^2\beta + 2\alpha\beta] \tag{3.6}$$

$$\begin{aligned} \mu_3'(s) &= \frac{1}{(1-\alpha\beta)^6}[\alpha^3(1-\alpha\beta)^3 + 6\alpha^2(1-\alpha)(1-\alpha\beta)^2) \\ &\quad +\alpha(1-\alpha)[7 - 11\alpha - 4\alpha\alpha\beta(2-\alpha)(1-\alpha\beta) \\ &\quad [1 - 6\alpha + 6\alpha^2 + 2\alpha\beta(4 - 9\alpha + 4\alpha^2) + \alpha^2\beta^2(6 - 6\alpha + \alpha^2)]] \end{aligned} \tag{3.7}$$

The higher moments of the SBGGSD (3.1) about the origin can be obtained similarly using (3.3) if so desired.

4 Recurrence Relationship of Moments about Origin of Size Biased GGSD

A recurrence relationship of (3.1) can be obtained by differentiating (3.3) we have

$$\begin{aligned}\frac{\partial\mu_r'(s)}{\partial\alpha} &= \sum_{x=1}^{\infty} x^r\binom{\beta x}{x-1}\left[\frac{\partial}{\partial\alpha}\{\alpha^{x-1}(1-\alpha\beta)(1-\alpha)^{1+\beta x-x}\}\right] \\ &= \sum_{x=1}^{\infty} x^{r+1}\frac{(1-\alpha\beta)^2}{\alpha}\binom{\beta x}{x-1}\alpha^{x-1}(1-\alpha)^{\beta x-x} \\ &\quad -\frac{1}{\alpha}\sum_{x=1}^{\infty} x^r\binom{\beta x}{x-1}(1-\alpha\beta)\alpha^{x-1}(1-\alpha)^{\beta x-x} \\ &\quad -\beta\sum_{x=1}^{\infty} x^r\binom{\beta x}{x-1}\alpha^{x-1}(1-\alpha)^{1+\beta x-x} \\ &= \frac{(1-\alpha\beta)}{\alpha(1-\alpha)}\mu_{r+1}'(s)-\frac{1}{\alpha(1-\alpha)}\mu_r'(s)-\frac{\beta}{(1-\alpha\beta)}\mu_r'(s).\end{aligned}$$

Thus

$$\mu_{r+1}'(s)=\frac{\alpha(1-\alpha)}{(1-\alpha\beta)}\frac{\partial\mu_r'(s)}{\partial\lambda}+\frac{(1-\alpha^2\beta)}{(1-\alpha\beta)^2}\mu_r'(s). \tag{4.1}$$

We can also obtain the moments of the SBGGSD about the origin from (4.1).

5 Bayesian Estimation of Truncated Generalized Geometric Series Distribution

Let $X_1, X_2, \cdots, X_n$ be a random sample from distribution (2.1). The likelihood function is given by

$$\begin{aligned}L(\underline{x}\mid\alpha,\beta) &= \prod_{i=1}^{n}\left[\frac{1}{\beta x_i+1}\binom{betax_i+1}{x_i}\alpha^{i-1}(1-\alpha)^{\beta x_i-x_i+1}\right] \\ &= C\alpha^{\Sigma x_i-n}(1-\alpha)^{\beta\Sigma x_i-\Sigma x_i+n} \\ &= C\alpha^{n\overline{x}-n}(1-\alpha)^{\beta n\overline{x}-\overline{n}+n},\end{aligned} \tag{5.1}$$

where

$$C=\prod_{i=1}^{n}\frac{1}{\beta x_i+1}\binom{\beta x_i+1}{x_i}.$$

Since $0<\alpha<1$, the prior distribution on α can be summarized by a beta distribution $B(a,b)$ with probability density function

$$g(\alpha;a,b)=\frac{\alpha^{a-1}(1-\alpha)^{b-1}}{B(a,b)};\quad a>0, b>0, 0<\alpha<1. \tag{5.2}$$

When β is known the posterior distribution of α becomes

$$\begin{aligned}
\pi(\alpha/\underline{x}) &= \frac{\alpha^{n\overline{x}-n+a-1}(1-\alpha)^{\beta n\overline{x}-n\overline{x}+n+b-1}}{\int_0^1 \alpha^{n\overline{x}-n+a-1}(1-\alpha)^{\beta n\overline{x}-n\overline{x}+n+b-1}d\alpha} \\
&= \frac{\alpha^{n\overline{x}-n+a-1}(1-\alpha)^{\beta n\overline{x}-n\overline{x}+n+b-1}}{B(n\overline{x}-n+a,\beta n\overline{x}-n\overline{x}+n+b)} \qquad (5.3)
\end{aligned}$$

Using a squared error loss function, Bayes estimator of α is given as

$$\begin{aligned}
\hat{\alpha} &= \frac{\int_0^1 \alpha^{n\overline{x}-n+a}(1-\alpha)^{\beta n\overline{x}-n\overline{x}+n+b-1}d\alpha}{B(n\overline{x}-n+a,\beta n\overline{x}-n\overline{x}+n+b)} \\
&= \frac{B(n\overline{x}-n+a+1,\beta n\overline{x}-n\overline{x}+n+b)}{B(n\overline{x}-n+a,\beta n\overline{x}-n\overline{x}+n+b)}
\end{aligned}$$

which on simplification gives

$$\hat{\alpha} = \frac{n\overline{x}-n+a}{\beta n\overline{x}+a+b} \qquad (5.4)$$

6 Bayesian Estimator of Parameter of Size Biased Generalized Geometric Series Distribution

The likelihood function of the SBGGSD (3.1) can be given as

$$\begin{aligned}
L(\underline{x} \mid \alpha\beta) &= (1-\alpha\beta)^n \prod_{i=1}^{n} \binom{\beta x_i}{x_i-1} \alpha^{\sum_{i=1}^{n} x_i - n}(1-\alpha)^{\beta\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i + n} \\
L(y \mid \alpha,\beta) &= K(1-\alpha\beta)^n \alpha^{y-n}(1-\alpha)^{\beta y-y+n}, \qquad (6.1)
\end{aligned}$$

where $y = \sum_{i=1}^{n} x_i$ and $K = \prod_{i=1}^{n} \binom{\beta x_i}{x_i-1}$.

Since $0 < \alpha < 1$, we assume that prior information for α when β is known is from a beta distribution. Thus

$$f(\alpha) = \frac{\alpha^{a-1}(1-\alpha)^{b-1}}{B(a,b)}; \quad 0 < \alpha < 1, \quad a, b > 0. \qquad (6.2)$$

The posterior distribution from (6.1) and (6.2) can be written as

$$p(\alpha|y) = \frac{(1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1}}{\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1}d\alpha}. \qquad (6.3)$$

The Bayes estimator of the parametric function α^z is given as

$$\begin{aligned}
\alpha^{*z} &= \int_0^1 \alpha^z p(\alpha|y)d\alpha \\
&= \frac{\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n+z-1}(1-\alpha)^{\beta y-y+n+b-1}d\alpha}{\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1}d\alpha}, \qquad (6.4)
\end{aligned}$$

where

$$\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n+z-1}(1-\alpha)^{\beta y-y+n+b-1} d\alpha$$
$$= \frac{\Gamma(y+a-n+z)\Gamma(\beta y+b-y+n)}{\Gamma(\beta y+a+b+z)}$$
$$\times\ {}^2F_1[-n, y+a-n+z, \beta y+a+b+z, \beta] \qquad (6.5)$$

and

$$\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1} d\alpha$$
$$= \frac{\Gamma(y+a-n)\Gamma(\beta y+b-y+n)\ {}^2F_1[-n, y+a-n, \beta y+a+b, \beta]}{\Gamma(\beta y+a+b)} \qquad (6.6)$$

where ${}^2F_1$ refers to the generalized hyper geometric function with two arguments 2-numerator and 1-denominator.

Putting these values in (6.4), the Bayes estimator of α^z is obtained as

$$\alpha^{*z} = \frac{\Gamma(y+a-n+z)\Gamma(\beta y+a+b)^2F_1[-n, y+a-n+z, \beta y+a+b+z, \beta]}{\Gamma(y+a-n)\Gamma(\beta y+a+b+z)^2F_1[-n, y+a-n, \beta y+a+b, \beta]}. \qquad (6.7)$$

Similarly, the Bayes estimator of the parametric function $(1-\alpha)^z$ can be obtained as

$$(1-\alpha)^z = \frac{\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1+z} d\alpha}{\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1} d\alpha}, \qquad (6.8)$$

where

$$\int_0^1 (1-\alpha\beta)^n \alpha^{y+a-n-1}(1-\alpha)^{\beta y-y+n+b-1+z} d\alpha$$
$$= \frac{\Gamma(y+a-n)\Gamma(\beta y+b-y+n+z)^2F_1[-n, y+a-n, \beta y+a+b+z, \beta]}{\Gamma(\beta y+a+b+z)}. \qquad (6.9)$$

Using the values from (6.9) and (6.6) in (6.8), the Bayes estimator of the parametric Function $(1-\alpha)^z$ can be obtained as

$$(1-\alpha)^{*z} = \frac{{}^2F_1[-n, y+a-n, \beta y+a+b+z, \beta]}{{}^2F_1[-n, y+a-n, \beta y+a+b, \beta]}$$
$$\times \frac{{}^2F_1[-n, y+a-n, \beta y+a+b+z, \beta]}{{}^2F_1[-n, y+a-n, \beta y+a+b, \beta]} \qquad (6.10)$$

Similarly, the Bayes estimator of some parametric functions $\phi(\alpha)$ and of some particular models of the SBGGSD are listed in Tables 1.

Table 1: Bayes' estimators of some pparametric functions in a SBGGSD and some of its particular models

Parametric function α^z	Bayes estimate
α	$\frac{(y+a-n)^2F_1[-n,y+a-n+1,\beta y+a+b+1,\beta]}{(\beta y+a+b)^2F_1[-n,y+a-n,\beta y+a+b,\beta]}$
$(1-\alpha)$	$\frac{(\beta y+b-y+n)^2F_1[-n,y+a-n,\beta y+a+b+1,\beta]}{(\beta y+a+b)^2F_1[-n,y+a-n,\beta y+a+b,\beta]}$

β	Particular distribution	Bayes estimate
1	SBGSD	$\frac{y+a-n}{y+a+b+n}$
0	SBBD	$\frac{y+a-n}{a+b}$

7 Application

An attempt has been made to fit the SBGGSD (3.1) and the zero truncated GGSD (2.1) to some zero-truncated biological data of McGuire et al. and Student on counts of the number of European red mites on apple leaves and Haemayeytometerye yeast cell counts observed per square respectively. Estmates of parameter has been obtained by using the Bayesian method of estimation. Monte Carlo simulation technique and R-Software have been used to obtain best fit for different values of a, b and β. In each of the following Tables 2 and 3 the expected frequencies for $a = b = 2$ and $\beta = 0.3$ and the values of chi-square according to the SBGGSD and the ZTGGSD are also given. So that a quick comparison can be made.

Table 2: McGuire et al. and Student on counts of the number of European red mites on apple leaves, $a = b = 2$ and $\beta = 0.3$

No. of cells per square	Observed No. of squares	Expected frequency (Based on Bayes Estimators)	
		TGGSD	SBGGSD
1	128	126.36	127.54
2	37	39.00	37.84
3	18	19.26	18.34
4	3	1.06	1.83
5	1	1.32	1.45
Total	187	187.00	187.00
Mean	1.45991		
χ^2		0.129838	0.026985
d.f.		1	1
Estimates of α		0.814	0.875

Table 3: Haemayeytometerye yeast cell counts observed per square $a = b = 3$ and $\beta = 0.2$

No. of per Plant	Observed No. of Plants	Expected frequency (Based on Bayes Estimators)	
		TGGSD	SBGGSD
1	83	84.03	83.04
2	36	37.05	35.08
3	14	13.06	13.87
4	2	1.25	2.87
5	1	0.61	1.14
Total	136	136.00	136.00
Mean	1.54412		
χ^2		0.332355	0.080308
d.f.		1	1
Estimates of α		0.1807	0.1048

It may be seen that from Tables 2 and 3 that the SBGGSD provides a better fits than the ZTGGSD model at the same degree of freedom in all the cases.

Acknowledgements

The authors are thankful to the editors and referees for useful suggestions which improved the presentation.

References

[1] Consul, P. C. and Famye, F. (2006). *On the Lagrangian probability distributions*. Springer publications.

[2] Jain, G. C. and Consul, P. C. (1971). A generalized negative binomial distribution, *SIAM J. Appl. Math*, **21**(4); 501–513.

[3] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate discrete distributions*. Second Edition. Wiley.

[4] Hassan, A. (1995). *Problems of estimation in Lagrangian probability distributions*, Ph.D. Thesis, Patna University, Patna.

[5] Hassan, A., Mishra, A. and Jan, T. R. (2002). A quick method of estimating generalized geometric series distribution, *Studia Sci. Math Hungarica*, **39**, 291–295.

[6] Hassan, A., Raja, T. A., and Mishra, A. (2003). On estimation of generalized geometric series distribution, *Research J, SKUAST (K)*, **1**(5), 55–60.

[7] Hassan, A., Mir, K. A., and Ahmad, M. (2007). On Bayesian analysis of generalized geometric series distribution under different priors. *Pakistan Journal of Statistics*, **23**(3), 255–260.

[8] Mishra, A. (1979). *Generalization of some discrete distributions*, Ph.D. Thesis, Patna University, Patna.

[9] Mishra, A. (1982). A generalization of geometric series distribution, *J. Bihar Math. Soc*, **6**, 18–22.

[10] Singh, S. K. (1989). *Some contributions to quasi–binomial and discrete Lagrangian probability distributions*, Ph.D. thesis, Patna University, Patna.

[11] Mishra, A. and Singh, S. K. (1992). On a characteristic of the geometric series Distribution, *J. Bihar Math Soc*, **15**, 32–35.