11-1-2008

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

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

$$P_1(X = x) = \frac{m}{m + \beta x} \binom{m + \beta x}{x} \alpha^x (1-\alpha)^{m+\beta x - x};$$

$$x = 0, 1, 2.....$$

$$(1.1)$$

where

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

The probability model (1.1) reduces to the

Khurshid Ahmad Mir is a Senior Assistant Professor in the Department of Statistics. Email: him at mirkhurshid_ahmad@yahoo.co.in or khrshdmir@yahoo.com. Mailing Address: Khaiwan Narwara Srinagar, Kashmir (India) Postal Code: 190002.

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 β, 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:

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

$$(1.2)$$

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

$$(1.3)$$

$$\mu_3' = \frac{(m\alpha)^3}{(1-\alpha\beta)^3} + \frac{3(m\alpha)^2(1-\alpha)}{(1-\alpha\beta)^4}$$
$$+ \frac{m\alpha(1-\alpha)}{(1-\alpha\beta)^5}[1 - 2\alpha + \alpha\beta(2-\alpha)]$$

$$\mu'_4 = \frac{(m\alpha)^4}{(1-\alpha\beta)^4} + \frac{6(m\alpha)^3(1-\alpha)}{(1-\alpha\beta)^5}$$

$$+ \frac{(m\alpha)^2(1-\alpha)\left[7-11\alpha+4\alpha\beta(2-\alpha)\right]}{(1-\alpha\beta)^6}$$

$$+ \frac{m\alpha(1-\alpha)}{(1-\alpha\beta)^7} + \begin{bmatrix} 1-6\alpha+6\alpha^2 \\ +2\alpha\beta\left(4-9\alpha+4\alpha^2\right) \\ +\alpha^2\beta^2\left(6-6\alpha+\alpha^2\right) \end{bmatrix}$$

(1.4)

(1.5)

and variance

$$\mu_2 = \frac{m\alpha(1-\alpha)}{(1-\alpha\beta)^3}.$$

(1.6)

Jain and Consul (1971) discussed the method of moments of estimation, and Gupta (1972, 1975) and Hassan (1995) obtained maximum likelihood estimations. Jani (1977), Kumar and Consul (1980), and Consul and Famoye (1989) studied the minimum variance unbiased estimation of GNBD, while Islam and Consul (1986) examined its Bayesian method of estimation. Recently, Consul and Famoye (1980) and Famoye (1997) discussed these methods in brief with respect to the model (1.1). Estimation techniques in the case of GNBD are not simple, all involve computation and can become tedious and time intensive.

The weighted distributions arise when observations generated from a stochastic process are not given an equal chance of being recorded, but instead are recorded according to some weight function. When the weight function depends on the lengths of the units of interest, the resulting distribution is called length biased. More generally, when the sampling mechanism selects units with probability proportional to some measure of the unit size, the resulting distribution is called size-biased. Such distributions arise, for example, in life length studies (see Blumenthal, 1967; Consul, 1989; Gupta 1975, 1976, 1979, 1984; Gupta & Tripathi, 1987, 1992; Schaeffer, 1972).

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:

$$P_2\left(X=x\right) = \frac{\dfrac{m}{m+\beta x}\begin{pmatrix} m+\beta x \\ x \end{pmatrix}\alpha^x\left(1-\alpha\right)^{m+\beta x-x}}{\left[1-\left(1-\alpha\right)^m\right]},$$

$$x=1,2\ldots..$$

(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 (2.1). Famoye and Consul (1993) defined a truncated GNBD 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:

$$P_3(X = x)$$

$$= (1 - \alpha\beta)\binom{m + \beta x - 1}{x - 1} \alpha^{x-1} (1 - \alpha)^{m + \beta x - x} \quad;$$

$x = 1, 2,...,$ 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 rth moment, $\mu'_r(s)$, about origin of the size-biased GNBD (3.1) can be defined as:

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

$$(3.2)$$

$\mu'_0(s) = 1$, 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 μ'_{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:

$$L(\underline{x} \mid \alpha, \beta) =$$

$$(1 - \alpha\beta)^n \prod_{i=1}^{n}\binom{m + \beta x_i - 1}{x_i - 1} \alpha^{\sum_{i=}^{} x_i - n} (1 - \alpha)^{mn + \beta\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i}$$

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

where

$$y = \sum_{i=1}^{n} x_i \quad \text{and} \quad K = \prod_{i=1}^{n}\binom{m+\beta x_i - 1}{x_i - 1}.$$

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

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

$$(4.2)$$

Using Bayes' Theorem, the posterior distribution of α from (4.1) and (4.2) can be written as:

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

$$(4.3)$$

Under square error loss function the Bayes' estimator of parametric function α^z is the posterior mean given as

$$\hat{\alpha}^z = \int_0^1 \alpha^z \, p(\alpha \mid y) d\alpha$$

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

$$(4.4)$$

where

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

$$\Gamma(y+a-n+z)\,\Gamma(\beta y+mn+b-y)$$
$$\frac{{}^2F_1\left[-n, y+a-n+z, \beta y+mn+a+b-n+1, \beta\right]}{\Gamma(\beta y+mn+a+b-n+z)}$$

$$(4.5)$$

and

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

$$\Gamma(y+a-n)\,\Gamma(\beta y+mn+b-y)$$
$$\frac{{}^2F_1\left[-n, y+a-n, \beta y+mn+a+b-n, \beta\right]}{\Gamma(\beta y+mn+a+b-n)}$$

$$(4.6)$$

Using relations (4.5) and (4.6) in (3.4), the Bayes' estimator of α^z becomes:

$$\hat{\alpha}^z =$$

$$\Gamma(y+a-n+z)\ \Gamma(\beta y+mn+a+\text{b-n})$$

$$\frac{\dfrac{{}^2F_1\left[-n, y+a-n+z, \beta y+mn+a+b-n+1, \beta\right]}{\Gamma(y+a-n)\,\Gamma(\beta y+mn+a+b-n+z)}}{{}^2F_1\left[-n, y+a-n, \beta y+mn+a+b-n, \beta\right]}$$

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

$$(1-\alpha)^z =$$

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

$$(4.8)$$

where

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

$$\frac{\Gamma(y+a-n)\ \Gamma(\beta y+mn+b-y+z)\ {}_2F_1\left[-n,y+a-n,\beta y+mn+a+b-n+z,\beta\right]}{\Gamma(\beta y+mn+a+b-n)}.$$

$$(4.9)$$

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

$$(1-\alpha)^z =$$

$$\frac{\Gamma(\beta y+mn+b-y+z)\ \Gamma(\beta y+mn+a+b-n)\ {}_2F_1\left[-n,y+a-n,\beta y+mn+a+b-n+z,\beta\right]}{\Gamma(\beta y+mn+a+b-n+z)\ \Gamma(\beta y+mn+b-y)\ {}_2F_1\left[-n,y+a-n,\beta y+mn+a+b-n,\beta\right]}$$

$$(4.10)$$

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 α 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 (α=0.5, β=0.3, a=b=1), (α=0.6, β=0.5, a=b=2) and (α=0.6, β=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

Bansal, A. K., & Ganji, M. (1997). Bayes' estimation in a decapitated generalized negative binomial distribution and some of its applications. *Journal of Information and optimization* Sciences, *18*(1), 189-198.

Blumenthal, S. (1967). Proportional sampling in life length studies. *Technometrics, 9*, 205-218.

Consul, P. C., & Famoye, F. (1989). Minimum variance unbiased estimation for the Lagrange power series distributions. *Statistics, 20*, 401-415.

Consul, P. C. (1975). Some new characterization of discrete Lagrangian distributions. In *Statistical distributions in scientific work, characterizations and applications*, G. P. Patil, Kotz, S.,& Ord, J. K. *(Eds.)*, 279-290. Dordrecht: Reidel.

Consul, P. C., & Famoye, F. (1995). On the generalized negative binomial distribution. *Communication in Statistics, Theory and Methods, 24*(2), 495-472.

Consul, P. C. & Gupta, R. C. (1980). The generalized negative binomial distribution and its characterization by zero regression. *SIAM Journal of Applied Mathematics*, *39*(2), 231-237.

Consul, P. C., & Shenton, L. R. (1972). Use of Lagrange expansion for generating discrete generalized probability distribution. *SIAM Journal of Applied Mathematics, 23*(2), 239-248.

Consul, P. C., & Shenton, L. R. (1975). On the probabilistic structure and properties of discrete Lagrangian distributions. In *Statistical distributions in scientific work, characterizations and applications*, G. P. Patil, Kotz, S.,& Ord, J. K. *(Eds.)*, 41-48. Dordrecht: Reidel.

Famoye, F. (1997). Parameter estimation of generalized negative binomial distribution, *Communications in Statistics-Simulation and Computation, 26*, 269-279.

Famoye, F., & Consul, P. C. (1989). A stochastic urn model for the generalized negative binomial distribution. *Statistics*, *20*, 607-613.

Famoye, F., & Consul, P. C. (1993). The truncated generalized negative binomial distribution, *Journal of Applied Statistical Science*, *1*(2), 141-157.

Gupta, R. C. (1974). Modified power series distribution and some of its applications. *Sankhya: The Indian Journal of Statistics*, *36B*(3), 288-298.

Gupta, R. C. (1975). Some characterizations of discrete distributions. *Communication in Statistics, Theory and Methods*, *5*(1), 45-48.

Gupta, R. C. (1976). Some characterizations of distributions by properties of their forward and backward recurrence times in a renewal process. *Scandinavian Journal of Statistics*, *3*, 215-216.

Gupta, R. C. (1979). Waiting time paradox and size-biased sampling. *Communication in Statistics, Theory and* Methods, *8*, 601-607.

Gupta, R. C. (1984). Some characterization of renewal densities with emphasis in reliability. *Mathematische Operations-forschung und Statistics*, *15*, 571-579.

Gupta, R. C., & Tripathi, R. C. (1987). A comparison between the ordinary and the length biased modified power series distributions with applications. *Communication in Statistics, Theory and Methods*, *16*(4), 1195-1206.

Gupta, R. C., & Tripathi, R. C. (1992). Statistical inference based on the length-biased data for modified power series distributions. *Communication in Statistics, Theory and Methods,* 21(2), 519-537.

Hassan, A (1995). Problems of estimation in Lagrangian probability distribution. Ph.D Thesis, *Patna University, Patna.*Islam, M. N., & Consul, P. C. (1986). Bayesian estimation in generalized negative binomial distribution, *Biometrics*, *28*, 250-256.

Jain, G. C., & Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics*, *21*(4), 501-513.

Jani, P. N. (1977). Minimum variance unbiased estimate for some left-truncated modified power series distributions, *Sankhya, B3*(39), 258-278.

Kumar, A., & Consul, P. C. (1980). Minimum variance unbiased estimation for modified power series distribution. *Communication in Statistics, Theory and Methods,* *9*(12), 1261-1275.

Meegama, S. A. (1980). Socio-economic determinants of infant and child mortality in Sri Lanka. *An analysis of post war experience international statistical institute (World Fertility Survey) Netherlands*.

Scheaffer, R. L. (1972). Size biased sampling. *Technometrics, 14*, 635-644.

Shoukri, M. M. (1980). Estimation of generalized distributions. Unpublished Ph.D. thesis, *University of Calgary, Calgary, Canada*.
Singh, S. N., & Yadav, R. C. (1971). Trends in rural out-migration at household level. *Rural Demography*, *8*, 53-61.

Table 4.1: Bayes' Estimators of SBGNBD

Parametric Function $\phi(\alpha)$	Bayes' Estimator of SBGNBD
α	$\dfrac{(y+a-n)^2 \, F_1[-n, y+a-n+1, b+mn+\beta y-n+a+1, \beta]}{(\beta y+mn+a+b-n)^2 \, F_1[-n, y+a-n, b+mn+\beta y-n+a, \beta]}$
$(1-\alpha)$	$\dfrac{(\beta y+mn+b-y)^2 \, F_1[-n, y+a-n, \beta y+mn+a+b-n+1, \beta]}{(\beta y+mn+a+b-n)^2 \, F_1[-n, y+a-n, \beta y+mn+a+b-n, \beta]}$

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

β	Distribution	Bayes' Estimator $\hat{\alpha}$
1	SBNBD	$\dfrac{y+a-n}{y+mn+a+b}$
0	SBBD	$\dfrac{y+a-n}{mn+a+b-n}$

SIZE-BIASED GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION

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, β=0.3)

x	f$_x$	Expected Frequency		
		BSBNBD	BZTGNBD	BSBGNBD
1	567	545.25	549.22	547.45
2	135	154.67	153.03	150.47
3	28	27.31	29.65	29.41
4	11	16.61	12.69	15.65
5	5	2.16	1.41	3.02
Total	746	746	746	746
Estimates $\hat{\alpha}$		0.48	0.49	0.51
χ^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, β=0.5)

x	f$_x$	Expected Frequency		
		BSBNBD	ZTGNBD	BSBGNBD
1	2039	2033.32	2031.45	2033.45
2	312	325.33	322.78	320.15
3	35	29.28	32.98	33.26
4	3	1.95	2.56	2.89
5	1	0.12	0.23	0.25
Total	2,390	2,390	2,390	2,390
Estimates $\hat{\alpha}$		0.465	0.493	0.503
χ^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, β=0.7)

| x	f$_x$	Expected Frequency		
		BSBNBD	BZTGNBD	BSBGNBD
1	375	370.87	368.37	371.81
2	143	156.29	155.79	151.49
3	49	48.42	49.12	50.21
4	17	11.44	13.24	12.51
5	2	2.40	3.01	2.89
6	2	0.47	0.33	0.73
7	1	0.09	0.11	0.30
8	1	0.02	0.03	0.06
Total	590	590	590	590
Estimates $\hat{\alpha}$		0.475	0.489	0.493
χ^2		6.9458	4.06227	3.16908

Graph 1: Sample Relative Frequency and Bayes' Relative Frequency for a=b=2, m=5, β=0.3
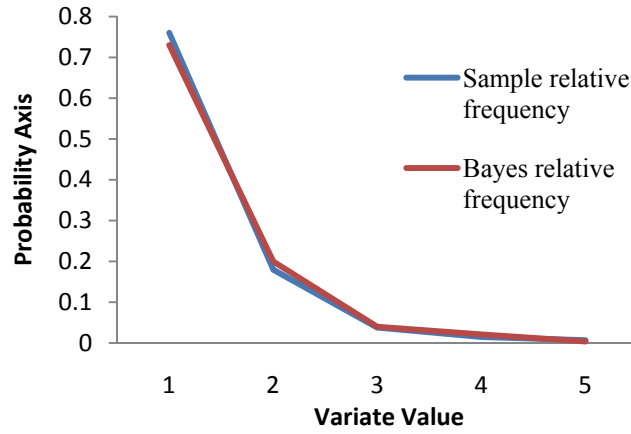
Graph 2: Sample Relative Frequency and Bayes' Relative Frequency for a=b=2, m=7, β=0.5
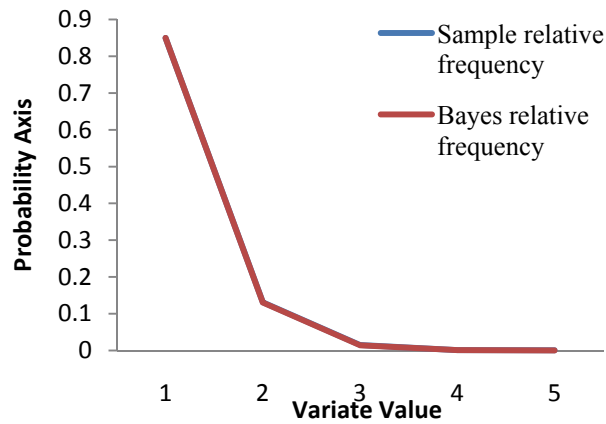
Graph 3: Sample Relative Frequency and Bayes' Relative Frequency for a=b=2, m=9, β=0.7