Bayesian Analysis of Generalized Geometric Series Distribution under Different Loss Functions

Chandan Kumer Podder
Department of Statistics
University of Chittagong
Chittagong-4331, Bangladesh
Email: podder_ck@yahoo.com

Abstract

In this paper, a Bayesian analysis of generalized geometric series distribution (GGSD) under different types of loss functions have been studied.

Keywords and Phrases: Generalized geometric series distribution, Parameter, Likelihood function, Conjugate prior distribution, Posterior distribution, Loss functions, Bayes' estimators.

AMS Classification: Primary 62J02; Secondary 62J20.

1   Introduction

The probability function of generalized geometric series distribution (GGSD) was given by Mishra [4] by using the results of the lattice path analysis as

$$
\begin{aligned}
P(X = x) &= \frac{1}{1 + \beta x}\binom{1 + \beta x}{x}\theta^x(1 - \theta)^{1+\beta x - x}; \quad x = 0, 1, 2, \ldots, \infty, \\
&\qquad\qquad 0 < \theta < 1,\ |\theta\beta| < 1,\ (1 + \beta x) \text{ is an integer}, \\
&= 0; \quad \text{otherwise.}
\end{aligned}
\tag{1}
$$

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

The various properties and estimation of (1) have been discussed by Mishra [4], Mishra and Singh [5]. However, Hassan et. al. [2] discussed the Bayesian analysis under non-informative and conjugate priors.

The purpose of this paper is to estimate the parameter of generalized geometric series distribution (GGSD) in a Bayesian approach under different loss functions.

2 Main Results

Let us consider the case of estimating the single parameter θ of the generalized geometric series distribution (GGSD) in the model (1), assuming β is known. The likelihood function of (1) is given by

$$
\begin{aligned}
l(\theta|x) &= \prod_{i=1}^{n} \left\{ \frac{1}{1+\beta x_i} \binom{1+\beta x_i}{x_i} \right\} \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n+\beta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i} \\
&= \kappa\, \theta^y (1-\theta)^{n+\beta y - y},
\end{aligned}
\tag{2}
$$

$$
\text{where} \quad \kappa = \prod_{i=1}^{n} \left\{ \frac{1}{1+\beta x_i} \binom{1+\beta x_i}{x_i} \right\} \quad \text{and} \quad y = \sum_{i=1}^{n} x_i.
$$

The maximum likelihood estimator of θ is $\frac{y}{n+\beta y}$, where y is defined above. It is noted that when β is known, the part of the likelihood function which is relevant to Bayesian inference on the unknown parameter θ is $\theta^y (1-\theta)^{n+\beta y - y}$.

A mathematically convenient prior density for the problem under consideration is conjugate prior given by

$$
\pi(\theta) \propto \theta^{p-1}(1-\theta)^{q-1}; \quad 0 < \theta < 1, \quad p > 0,\ q > 0,
\tag{3}
$$

which is simply a member of the beta family of distributions. The advantage of taking the prior distribution to be conjugate lies in the fact that the likelihood function $l(\theta|x)$, the prior density $\pi(\theta)$ and the posterior density $\pi(\theta|x)$ are all of the same functional form, thus ensuring mathematical tractability.

In this paper, estimation of the parameter θ, assuming β is known, considered in a Bayesian approach under the prior distribution (3) and the following loss functions :

  i. $L_1(\hat{\theta}, \theta) = c(\hat{\theta} - \theta)^2;\ c > 0.$

  ii. $L_2(\hat{\theta}, \theta) = c(\hat{\theta} - \theta)^2/\theta^2;\ c > 0.$

  iii. $L_3(\hat{\theta}, \theta) = c(\sqrt{\hat{\theta}} - \sqrt{\theta})^2;\ c > 0.$

  iv. $L_4(\hat{\theta}, \theta) = c(\sqrt{\hat{\theta}} - \sqrt{\theta})^2/\theta\ ;\ c > 0.$

  v. $L_5(\hat{\theta}, \theta) = \begin{cases} 0 & \text{if } |\hat{\theta} - \theta| < \delta \\ 1 & \text{otherwise} \end{cases}$
   where δ is a small known quantity.

vi. $L_6(\hat{\theta}, \theta) = \begin{cases} 0 & \text{if } \delta_1 < \hat{\theta} - \theta < \delta_2 \\ 1 & \text{if } \quad \hat{\theta} - \theta < \delta_1 \\ 1 & \text{if } \quad \hat{\theta} - \theta > \delta_2 \end{cases}$

where δ_1 and δ_2 are two small known quantities, and

vii. $L_7(\hat{\theta}, \theta) = \varpi \left[\left(\frac{\hat{\theta}}{\theta} \right)^\gamma - \gamma \, ln \left(\frac{\hat{\theta}}{\theta} \right) - 1 \right]; \; \gamma \neq 0, \; \varpi > 0.$

Combining (2) and (3), the posterior distribution of θ for the given sample $x = (x_1, x_2, \ldots, x_n)$ is

$$\begin{aligned} P(\theta|x) &\propto l(\theta|x)\pi(\theta) \\ \Rightarrow P(\theta|x) &\propto \theta^{y+p-1}(1-\theta)^{n+\beta y-y+q-1}. \end{aligned} \quad (4)$$

This implies that the posterior distribution of θ for the given sample $x = (x_1, x_2, \ldots, x_n)$ is

$$P(\theta|x) = \frac{1}{B(y+p, n+\beta y - y + q)} \theta^{y+p-1}(1-\theta)^{n+\beta y-y+q-1}; 0 \leq \theta \leq 1, \quad (5)$$

which follows that $\theta \sim Beta(y+p, \; n+\beta y - y + q)$.

The mean of the posterior distribution is $E(\theta|x) = \frac{y+p}{n+\beta y+p+q}$.

Different shapes and existence of mode of the posterior distribution $Beta(p', q')$ in (5), when $p' = y + p$ and $q' = n + \beta y - y + q$ are given below;

- **Case I :** For $p' = q' = 1$, the beta distribution simply becomes a uniform distribution between zero and one.
- **Case II :** For $p' = 1$ and $q' = 2$ or vice versa, we get triangular shaped distributions.
- **Case III :** For $p' = q' = 2$, we obtain a distribution of parabolic shape.
- **Case IV :** For $p' = q' \geq 3$, we obtain distributions of symmetrical and bell-shape.
- **Case V :** If p' and q' both are greater than one, the distribution has a unique mode at $\theta = \frac{(p'-1)}{(p'+q'-2)}$ and is zero at the end points.
- **Case VI :** If p' and/or q' is less than one $f(0) \to \infty$ and/or $f(1) \to \infty$ and the distribution is said to be $J-$ shaped.

In above, the mode of the posterior distribution in (5) does not exist in case of **I** and **VI**. In case of **II**, with vertex 0 (or 1) the mode will be attained at 0 (or 1). The mode is 0.5 in case of **III** and **IV**. In case of **V**, the mode is at $\theta = \frac{(p'-1)}{(p'+q'-2)}$ if $p' > 1$, $q' > 1$.

Hence, more generally, the mode of the posterior distribution in (5) is

$$M_0 = \frac{y + p - 1}{n + \beta y + p + q - 2},$$

if $(y + p)$ and $(n + \beta y - y + q)$ both are greater than or equal to one.

3 Bayes' Estimation

In this section, we provide the Bayes' estimators of the parameter of generalized geo-metric series distribution (GGSD) using above seven different loss functions.

The first four loss functions i, ii, iii and iv considered in section 2, are particular cases of the form

$$L(\hat{\theta}, \theta) = c\theta^a \left(\hat{\theta}^b - \theta^b \right)^2, \tag{6}$$

where c is a positive constant, a and b are known quantities.

For the loss function given by (6), the Bayes' estimator $\hat{\theta}$ is

$$\hat{\theta}^b = \frac{E_\theta(\theta^{a+b}|x)}{E_\theta(\theta^a|x)}$$

$$\Rightarrow \hat{\theta} = \left[\frac{E_\theta(\theta^{a+b}|x)}{E_\theta(\theta^a|x)} \right]^{\frac{1}{b}}, \tag{7}$$

where $E_\theta(\theta|x)$ is the posterior expectation with respect to parameter θ, if exists.

Therefore, using (7), the Bayes' estimator $\hat{\theta}$ under the loss function (6) is given by

$$\hat{\theta} = \left[\frac{\Gamma(y + p + a + b)\Gamma(n + \beta y + p + q + a)}{\Gamma(y + p + a)\Gamma(n + \beta y + p + q + a + b)} \right]^{\frac{1}{b}}, \text{ if exists.} \tag{8}$$

(a) Substituting $a = 0$ and $b = 1$, the loss function (6) becomes $L_1(\hat{\theta}, \theta)$ and the Bayes' estimator under the loss function L_1 using (8) is given by

$$\hat{\theta}_1 = \frac{y + p}{n + \beta y + p + q},$$

which is the mean of the posterior distribution.

(b) Substituting $a = -2$ and $b = 1$, the loss function (6) becomes $L_2(\hat{\theta}, \theta)$ and the Bayes' estimator under the loss function L_2 using (8) is given by

$$\hat{\theta}_2 = \frac{y + p - 2}{n + \beta y + p + q - 2}.$$

(c) Substituting $a = 0$ and $b = \frac{1}{2}$, the loss function (6) becomes $L_3(\hat{\theta}, \theta)$ and the Bayes' estimator under the loss function L_3 using (8) is given by

$$\hat{\theta}_3 = \left[\frac{\Gamma(y + p + \frac{1}{2})\Gamma(n + \beta y + p + q)}{\Gamma(y + p)\Gamma(n + \beta y + p + q + \frac{1}{2})} \right]^2.$$

(d) Substituting $a = -1$ and $b = \frac{1}{2}$, the loss function (6) becomes $L_4(\hat{\theta}, \theta)$ and the Bayes' estimator under the loss function L_4 using (8) is given by

$$\hat{\theta}_4 = \left[\frac{\Gamma(y + p - \frac{1}{2})\Gamma(n + \beta y + p + q - 1)}{\Gamma(y + p - 1)\Gamma(n + \beta y + p + q - \frac{1}{2})} \right]^2.$$

Remarks 3.1. If the expression (8) does not exist in general for the chosen super parameters and the specific observed data then the Bayes' estimation is not possible. In particular, for example, in case of $a = -2$, $b = 1$, if $y + p - 1$ is < 0, the Bayes' estimator does not exist.

(e) The Bayes' estimator for the zero-one type of loss function L_5 is the mode of the posterior distribution (5) if it is symmetrical. The posterior distribution (5) will be symmetrical when $p' = q' \geq 2$ (in case of **III** and **IV**) where $p' = y + p$, $q' = n + \beta y - y + q$ and the mode is $M_0 = 0.5$ in the symmetrical family of posterior distributions (5).

Hence, $$\hat{\theta}_5 = M_0 = 0.5.$$

Otherwise, in case of **II** and **V**, when both p' and q' are greater than or equal to one the Bayes' estimator is M_0, the mid-point of the interval I of length 2δ which maximizes the posterior distribution $P(\theta \varepsilon I | x)$ in (5).

(f) The Bayes' estimator for the special zero-one type of loss function L_6 is

$$\hat{\theta}_6 = M_0 + \frac{\delta_1 + \delta_2}{2}, \tag{9}$$

where M_0 is the mode of the posterior distribution, if it is symmetrical. The posterior distribution (5) will be symmetrical when $p' = q' \geq 2$ (in case of **III** and **IV**) where $p' = y + p$, $q' = n + \beta y - y + q$ and the mode is $M_0 = 0.5$ in the symmetrical family of posterior distributions (5). Here, δ_1 and δ_2 are two small known quantities.

Otherwise, in case of **II** and **V**, when both p' and q' are greater than or equal to one, M_0 will be the mid-point of the interval I of length 2δ which maximizes the posterior distribution $P(\theta \varepsilon I | x)$ in (5).

(g) The Bayes' estimator for the loss function L_7 is

$$\hat{\theta}_7 = \left[E_\theta(\theta^{-\gamma} | x) \right]^{-\frac{1}{\gamma}}. \tag{10}$$

Here,

$$E_\theta(\theta^{-\gamma}|x) = \int_0^1 \theta^{-\gamma} P(\theta|x)\mathrm{d}\theta = \frac{\Gamma(y+p-\gamma)\Gamma(n+\beta y+p+q)}{\Gamma(y+p)\Gamma(n+\beta y+p+q-\gamma)}. \tag{11}$$

From (10), using (11) gives,

$$\hat{\theta}_7 = \left[\frac{\Gamma(y+p-\gamma)\Gamma(n+\beta y+p+q)}{\Gamma(y+p)\Gamma(n+\beta y+p+q-\gamma)}\right]^{-\frac{1}{\gamma}}.$$

4   Discussion

It has been seen that as $\delta_1 = -\delta_2$, the Bayes' estimators $\hat{\theta}_6$ and $\hat{\theta}_5$ are identical. If we put $\gamma = -1$, the Bayes' estimators $\hat{\theta}_7$ and $\hat{\theta}_1$ are also identical.

It has been also found that as $\beta = 1$, the above estimators are the Bayesian estimators of the parameter of simple geometric distribution under the above loss functions L_1, L_2, L_3, L_4, L_5, L_6 and L_7 respectively.

Acknowledgement

References

[1] Jain, G. C. and Consul, P. C. (1971). A generalized negative binomial distribution, *Siam J. Appl. Math.* **21(4)**, 501-513.

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

[3] Lehmann, E. L. (1983). *Theory of Point Estimation*, John Wiley & Sons, Inc.

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

[5] Mishra, A. and Singh, S. K. (1992). On a characteristic of the geometric series distribution, *J. Bihar Math.* **15**, 32-35.

[6] Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss function, *Journal of American Statistical Association*, **81**, 446-451.