



ON ESTIMATION OF GEETA DISTRIBUTION AND ITS APPLICATION TO THE STUDY OF MICROBES

Khurshid Ahmad Mir*

Department of Statistics, Govt Degree College Bemina Srinagar, Kashmir, India

Received 02 November 2009; Accepted 10 April 2010
Available online 14 October 2010

Abstract: *In this paper, different estimation methods for the parameters of the Geeta distribution (GET) are discussed. The model is applied to study microorganisms based on difference-differential equations leading to Geeta distribution. R- Software has been used for making a comparison among the three different estimation methods.*

Keywords: *Geeta distribution, Bayes' estimator, R-Software.*

1. Introduction

Consul (1990a) defined the Geeta distribution over the set of all positive integers with the probability mass function as:

$$P[X = x] = \frac{1}{\beta x - 1} \binom{\beta x - 1}{x} \alpha^{x-1} (1 - \alpha)^{\beta x - x} ; x=1, 2, \dots \quad (1)$$

where $1 < \beta < \alpha^{-1}$ and $0 < \alpha < 1$.

The Geeta distribution has a maximum at $x=1$ and is L-shaped for all values of α and β . It may have a short tail or a long tail depending upon the values of α and β . Its mean and variance are given by:

$$\mu = (1 - \alpha)(1 - \alpha\beta)^{-1} \quad (2)$$

*Email: khurshdmir@yahoo.com

$$\mu'_2 = \frac{\alpha(1-\alpha)(\beta-1) + (1-\alpha)^2(1-\alpha\beta)}{(1-\alpha\beta)^3}$$

$$\sigma^2 = (\beta-1)\alpha(1-\alpha)(1-\alpha\beta)^{-3} = \mu(\mu-1)(\beta\mu-1)(\beta-1)^{-1}. \quad (3)$$

The family of Geeta probability models belong to the classes of Consul and Shenton's (1972) Lagrangian series distributions and Gupta's (1974) modified power series distributions (MPSD). Consul (1990b) also expressed the Geeta distribution as a location-parameter probability model with probability function as:

$$P_1[X = x] = \frac{1}{\beta^x - 1} \binom{\beta^x - 1}{x} \left[\frac{\mu - 1}{\beta\mu - 1} \right]^{x-1} \left[\frac{\mu(\beta - 1)}{\beta\mu - 1} \right]^{\beta^x - x}; \quad x=1, 2, 3, \dots \quad (4)$$

where $\mu \geq 1$ and $\beta > 1$.

Consul (1990c) showed that the Geeta distribution can be characterized by its variance $\sigma^2 = (\beta - 1)^{-1} \mu(\mu - 1)(\beta\mu - 1)$, when the probability is non-zero over all integral values of X .

2. Estimation Methods

In this section, we discuss the various estimation methods for Geeta distribution and verify their efficiencies.

2.1 Method of Moments

In the method of moments replacing the population mean and variance by the corresponding sample mean and variance, we have:

$$\bar{x} = \frac{1-\alpha}{1-\alpha\beta} \quad \text{and} \quad (5)$$

$$s^2 = \frac{(\beta-1)\alpha(1-\alpha)}{(1-\alpha\beta)^3} \quad (6)$$

On simplifying (5), we get:

$$\hat{\alpha} = \frac{\bar{x} - 1}{\beta\bar{x} - 1} \quad (7)$$

Using (7) in (6), we get the estimate for β as:

$$\hat{\beta} = \frac{s^2 - \bar{x}(\bar{x} - 1)}{s^2 - \bar{x}^2(\bar{x} - 1)} \tag{8}$$

Substituting the value of $\hat{\beta}$ in equation (7), we get the estimate for α .

2.2 Method of Maximum Likelihood Estimation

The likelihood function of GET (1) is given as:

$$L(x; \alpha, \beta) = \prod_{i=1}^n \binom{\beta x_i - 1}{x_i} \left(\frac{1}{\beta x_i - 1} \right) \alpha^{\sum_{i=1}^n x_i - n} (1 - \alpha)^{\beta \sum_{i=1}^n x_i - \sum_{i=1}^n x_i}$$

$$L(x; \alpha, \beta) = k \alpha^{y-n} (1 - \alpha)^{\beta y - y} \tag{9}$$

Where $y = \sum_{i=1}^n x_i$ and $k = \prod_{i=1}^n \binom{\beta x_i - 1}{x_i} \left(\frac{1}{\beta x_i - 1} \right)$

The log likelihood function is given as:

$$\text{Log } L = \sum_{i=1}^n \log \left(\frac{1}{\beta x_i - 1} \right) + \sum_{i=1}^n \log \binom{\beta x_i - 1}{x_i} + (n\bar{x} - n) \log \alpha + (\beta n\bar{x} - n\bar{x}) \log (1 - \alpha) \tag{10}$$

The likelihood equations are given as:

$$\frac{\partial \log L}{\partial \alpha} = \frac{(n\bar{x} - n)}{\alpha} - \left(\frac{\beta n\bar{x} - n\bar{x}}{1 - \alpha} \right) = 0 \tag{11}$$

$$\frac{\partial \log L}{\partial \beta} = \frac{-n\bar{x}}{\log(\beta x - 1)^2} + \log nx + n\bar{x} \log(1 - \alpha) = 0 \tag{12}$$

Solving above two likelihood equations, we get the maximum likelihood estimators of α and β .

2.3 Bayesian Estimation of the Parameter of Geeta Distribution

Since $0 < \alpha < 1$, therefore we assume that prior information about α when β is known is from beta distribution as:

$$f(\alpha) = \frac{\alpha^{a-1} (1 - \alpha)^{b-1}}{B(a, b)}; \quad 0 < \alpha < 1, \quad a > 0, \quad b > 0. \tag{13}$$

The posterior distribution from (9) and (13) can be written as:

$$\prod(\alpha/y) = \frac{\alpha^{y+a-n-1} (1-\alpha)^{\beta y-y+b-1}}{\int_0^1 \alpha^{y+a-n-1} (1-\alpha)^{\beta y-y+b-1} d\alpha}$$

The Bayes' estimator of α is given as:

$$\begin{aligned} \hat{\alpha} &= \int_0^1 \alpha \prod(\alpha/y) d\alpha \\ &= \frac{\int_0^1 \alpha^{y+a-n} (1-\alpha)^{\beta y-y+b-1} d\alpha}{\int_0^1 \alpha^{y+a-n-1} (1-\alpha)^{\beta y-y+b-1} d\alpha} \\ \hat{\alpha} &= \frac{y-n+a}{\beta y-n+a+b} \end{aligned} \tag{14}$$

3. Model Based On Difference-Differential Equations

In this section, we have obtained the Geeta distribution on the basis of difference-differential equations. For this consider a regenerative process initiated by a single microbe, bacteria, or cell and which may grow into any number. Let the probability of x cells in a location be $P_x(\alpha, \beta)$.

Theorem: If the mean μ for the distribution of the microbes is increased by changing α to $\alpha + \Delta\alpha$ in such a way that:

$$\begin{aligned} \frac{dP_x(\alpha, \beta)}{d\alpha} + \frac{x(\beta-1)}{1-\alpha} P_x(\alpha, \beta) \\ = \frac{(x-1)(\beta x-x)^{[x-1]} (1-\alpha)^{\beta-1}}{x(\beta x-x-\beta+1)^{[x-2]}} P_{x-1}(\alpha, \beta), \end{aligned} \tag{15}$$

for all integral values $x \geq 1$ with the initial conditions. $P_1(0, \beta) = 1$ and $P_0(0, \beta) = 0$ for $x \geq 2$, then the probability model $P_x(\alpha, \beta)$ is the Geeta distribution, where $a^{[x]} = a(a+1)\dots(a+x-1)$.

Proof: For $x=1$, equation (15) becomes:

$$\frac{dP_1(\alpha, \beta)}{d\alpha} + \frac{(\beta-1)}{1-\alpha} P_1(\alpha, \beta) = 0,$$

which is a simple differential equation with the solution

$$P_1(\alpha, \beta) = (1 - \alpha)^{\beta-1}. \tag{16}$$

For x=2, equation (15) with (16) gives:

$$\frac{dP_2(\alpha, \beta)}{d\alpha} + \frac{2(\beta-1)}{1-\alpha} P_2(\alpha, \beta) = (\beta-1)(1-\alpha)^{2\beta-2}.$$

The solution to the above equation is:

$$P_2(\alpha, \beta) = (2\beta - 2)^{[1]} \alpha(1-\alpha)^{2\beta-2} / 2!. \tag{17}$$

By putting x=3 in equation (15) and by using (17), we obtain another linear differential equation whose solution is:

$$P_3(\alpha, \beta) = (3\beta - 3)^{[2]} \alpha^2(1-\alpha)^{3\beta-3} / 3!. \tag{18}$$

By using the principle of mathematical induction, the solution for x=k is given by:

$$\begin{aligned} P_k(\alpha, \beta) &= (\beta k - k)^{[k-1]} \alpha^{k-1} (1-\alpha)^{k(\beta-1)} / k! \\ &= \frac{1}{\beta k - 1} \binom{\beta k - 1}{k} \alpha^{k-1} (1-\alpha)^{\beta k - k}, \end{aligned}$$

which is the Geeta distribution defined in (1).

4. Computer Simulation

It is very difficult to compare the theoretical performances of different estimators proposed in the previous section. Therefore, we perform extensive simulations to compare the performances of the different methods of estimation mainly with respect to their biases and mean squared errors (MSE's), for different sample sizes and of different parametric values. Regarding the choice of values of (a, b) in the Bayes' estimator (14), there was no information about their values except that they are real and positive numbers. Therefore 25 combinations of values of (a, b) were considered for a, b=1, 2,3,4,5 and those values of a, b were selected for which the Bayes' estimator $\hat{\alpha}$ has minimum variance. It was found that for a=b=5, the Bayes' estimator has minimum variance and χ^2 values between the simulated sample frequencies and the estimated Bayes' frequencies were the least. Data given in tables (1) and (2) are the zero-truncated data of P.Garman (1923) and Student (1907) on counts of the number of European red mites on apple leaves and yeast blood cell counts observed per square respectively.

Table 1. Counts of the number of European red mites on apple leaves

No. of mites per leaf	Leaves Observed	Expected frequency		
		Mom	MLE	Bayes'
1	38	36.76	37.15	37.91
2	17	15.21	16.95	16.99
3	10	10.25	9.55	9.78
4	9	9.75	9.34	8.92
5	3	3.15	2.95	2.98
6	2	1.85	1.91	1.99
7	1	1.34	1.56	0.99
≥8	0	1.69	0.59	0.44
Total	80	80	80	80
χ^2		0.832	0.198	0.031

Table 2. Yeast blood cell counts observed per square

No. of cells per square	Observed No. of squares	Expected frequency		
		Mom	MLE	Bayes'
1	128	126.38	126.94	127.42
2	37	36.45	36.81	36.71
3	18	17.25	17.42	18.02
4	3	2.37	2.83	2.95
5	1	1.14	1.19	1.02
≥ 6	0	3.41	1.81	0.88
Total	187	187	187	187
χ^2		0.224	0.077	0.038

5. Discussion and Conclusions

The paper discusses how a Geeta distribution can be used to study a regenerative process initiated by a single microbe, bacteria, or cell which may grow into any number. Also in tables 1 and 2, we made a comparison among different estimation methods of Geeta distribution and it was observed that the Bayes' estimator provides us the better fit against maximum likelihood or moment estimator.

Acknowledgement

The author is highly thankful to the editor and the two anonymous referees for their valuable suggestions.

References

- [1]. Consul, P.C. (1990a). Geeta distribution and its properties. *Communications in Statistics-Theory and Methods*, 19, 3051-3068.
- [2]. Consul, P.C. (1990b). Two stochastic models for the Geeta distribution. *Communications in Statistics-Theory and Methods*, 19, 3699-3706.
- [3]. Consul, P.C. (1990c). New class of location-parameter discrete probability distributions and their characterizations. *Communications in Statistics-Theory and Methods*, 19, 4653-4666.
- [4]. Consul, P.C., Shenton, L.R. (1972). Use of Lagrange expansion for generating discrete generalized probability distributions. *SIAM Journal of Applied Mathematics*, 23,239-248.
- [5]. Garman, P. (1923). The European red mites in Connecticut apple orchards. *Connecticut Agricultural Experimental Station Bulletin*, 252,103-125.
- [6]. Gupta, R. C. (1974). Modified power series distribution and some of its applications. *Sankhya-Series B*, 36, 288-298.
- [7]. Student (1907). On the error of counting with a haemocytometer. *Biometrika*, 5, 351-360.