

Parametric Interval Estimation of the Geeta Distribution

Betty C. Korir¹ & Josphat K. Kinyanjui²

¹Department of Mathematics & Computer Science, University of Eldoret, Kenya

²School of Pure and Applied Sciences, Karatina University, Kenya

Correspondence: Betty C. Korir, Department of Mathematics & Computer Science, University of Eldoret, Kenya.

Received: September 29, 2018 Accepted: October 30, 2018 Online Published: November 18, 2018

doi:10.5539/ijsp.v8n1p1

URL: <https://doi.org/10.5539/ijsp.v8n1p1>

Abstract

It is well known that the sample mean is the estimator of a population mean in mathematical statistics from a given population of interest as a point estimator which assume a single number that is obtained by taking a random sample of a specified size from the entire population, depending on whether the population mean and variance is known or unknown. In the interval estimation, the sample mean is accompanied with a plus or a minus margin of an error that is assumed that the estimator is contained within the range of values with certain degree of confidence. This paper investigated and obtained the interval estimators of the unknown constants of Geeta distribution model through the construction of confidence interval using; the pivotal quantity method, the shortest-length confidence interval, unbiased confidence interval estimators, Bayesian confidence interval estimators and statistical method. Geeta distribution is a new discrete random variable distribution defined over all the positive integers, with two unknown parameters. The properties and characteristics of the Geeta distribution model were discussed and reviewed that is, the existence of the mean, variance, moment generating function and that the sum of all probabilities is unity. These are common properties of any given probability density function.

Keywords: asymptotic variance, Bayesian confidence intervals, Geeta distribution and pivotal method

1. Introduction

Generally, in the interval estimation, we seek to construct the confidence interval which the population mean and population variance is contained within a range of values and has high confidence coefficient with the shortest-length of the interval (Harold J. Larson, 1934). The construction of these confidence intervals considers the factors such as when the population is known or unknown and similarly when the population variance is either known or unknown.

Let $X = (X_1, X_2, \dots, X_n)$ be a random sample from the normal distribution with mean μ and variance S^2 . Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ be the sample mean and sample variance for X , respectively (Traor é et al, 2018).

Let $Z(X_1, X_2, \dots, X_n; \mu)$ be a pivotal quantity where X_1, X_2, \dots, X_n is a random variable from the distribution of $f(X; \mu, \sigma^2)$.

Then

$$P_{\mu}[a \leq Z \leq b] = \phi(b) - \phi(a) = 1 - \alpha \quad (1)$$

for any constants a and b .

Then

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad \text{when } \sigma^2 \text{ is known and}$$

$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ with confidence coefficient of $(1-\alpha)$ 100%

But the pivotal quantity becomes $Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}, \frac{\alpha}{2}$, when σ^2 is unknown.

When Z is inverted, it yields a corresponding $(1-\alpha)$ confidence interval for mean, μ ;

$$\left[\bar{X} - b \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - a \frac{\sigma}{\sqrt{n}} \right]$$

Such that

$$P_{\mu} \left[\bar{X} - b \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - a \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha \tag{2}$$

Therefore confidence interval (1-α) 100% for the mean, μ is given by

$$\left[\bar{X} - b \frac{\sigma}{\sqrt{n}}; \bar{X} - a \frac{\sigma}{\sqrt{n}} \right] \tag{3}$$

Now [a;b] is chosen such that b-a is the shortest length of the interval for a given confidence coefficient (1-α).

The symmetric solution is such that;

$$1-\alpha = \phi(b) - \phi(a) = \phi(b) - \phi(-b) = 2\phi(b) - 1$$

Hence $b = \phi^{-1}(1-\alpha/2)$.

To find the shortest length $L(a,b) = b-a$ of the confidence interval for the mean at (1-α) is to minimize the length

$L(a,b) = b-a$ subject to

$$\int_a^b f(z) dz = 1 - \alpha \tag{4}$$

The solution of equation (4) yields $f(a) = f(b) \Rightarrow a = b$ or $a = -b$ but $a \neq b$, hence $a = -b$.

That means, the confidence interval (1-α) 100% for mean, μ s given by $\left[\bar{X} \pm b \frac{\sigma}{\sqrt{n}} \right]$;

such that $P_{\mu} \left[\bar{X} \pm b \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$ and its shortest-length interval is given by $L(a, b) = 2b \frac{\sigma}{\sqrt{n}}$.

The mathematical form (Hogg, R.V and Craig, A.T. ,1956) of probability distribution defined here, Geeta distribution contains the unknown constants θ and β, and are therefore estimated through construction of confidence intervals which forms the subject of discussion in this paper

1.1 The Geeta Distribution

Geeta distribution is a newly introduced distribution which has two unknown parameters and is of the form L-shaped model, it belongs to a family of Modified Power Series distribution(MPSD), the Langrangian series distributions and location parameter distribution. The Yule distribution and Pareto which belongs to the same family(MPSD) have a single parameter and therefore fails the test of handling large data sets when it comes to applications in modern technologies. Geeta distribution model is very versatile in meeting the needs of modern complex data sets and this attributed to the presence of the two unknown constants when compared to the distribution of the same class. The unknown constants can be estimated using the estimation techniques and it is believed that these constants contain a lot of information.

Geeta distribution is defined as a discrete random variable, X, over the set of all positive integers, with the probability mass function given by

$$\Pr(X = x) = \frac{1}{\beta x - 1} \binom{\beta x - 1}{x} \theta^{x-1} (1 - \theta)^{\beta x - x}; \quad x = 1, 2, 3, \dots \tag{5}$$

$= 0, \text{ otherwise}$

where

$$0 < \theta < 1; \quad 1 < \beta < \theta^{-1}$$

The upper limit on β has been imposed for the existence of the mean. When β → 1 the model degenerates to a single point at x=1.

The Geeta distribution has a maximum as x=1 and is L – shaped for all values of θ and β. It may have a short tail or a long tail and heavy tail depending upon the values of θ and β. Its mean μ and variance σ² are given by

$$\mu = (1 - \theta)(1 - \beta\theta)^{-1} \tag{6}$$

and

$$\sigma^2 = (\beta - 1)\theta(1 - \theta)(1 - \beta\theta)^{-3} \tag{7}$$

and from the formula (6)

$$\theta = \frac{(\mu - 1)}{(\beta\mu - 1)} \tag{8}$$

It can also be expressed as a location parameter probability distribution given below:

$$\Pr(X = x) = \frac{1}{\beta^{x-1}} \binom{\beta x - 1}{x} \left(\frac{\mu - 1}{\beta\mu - 1} \right)^{x-1} \left(\frac{\mu(\beta - 1)}{\beta\mu - 1} \right)^{\beta x - x}; \quad x = 1, 2, 3, \dots \tag{9}$$

= 0, otherwise

where μ is then mean and $\beta > 1$. Note that this form does not have an upper limit on β . (Consul, 1990) has shown that the Geeta distribution (9) can be characterized by its variance:

$$\sigma^2 = (\beta - 1)^{-1} \mu(\mu - 1)(\beta\mu - 1) \tag{10}$$

and the domain of X.

It is clear from the expression of σ^2 that $\frac{d\sigma^2}{d\mu} > 0$ and accordingly, σ^2 increases monotonically as μ increases in

value and that the smallest value of σ^2 is zero when $\mu = 1$, that is, when the model reduces to a single point $x=1$.

Also

$$\frac{d\sigma^2}{d\beta} = \frac{-\mu(\mu - 1)^2}{(\beta - 1)^2} < 0 \tag{11}$$

Thus, the variance σ^2 decreases monotonically as β increases and the smallest value of σ^2 , for the largest value of β becomes $\mu^2(\mu - 1)$. From this we conclude that when $\beta\mu - 1 \leq (\beta - 1)(\mu - 1)^{-1}$ the variance will be less than the mean μ and will have the range:

$$\mu^2(\mu - 1) < \sigma^2 \leq \mu \tag{12}$$

If $(\beta\mu - 1) > (\beta - 1)(\mu - 1)^{-1}$ the value of σ^2 will become larger than μ .

1.2 Bar – Diagram for Geeta Model

The successive probabilities for various values of x can be easily computed from the values:

$$\Pr(X = 1) = \left[\frac{(\beta - 1)\mu}{\beta\mu - 1} \right]^{\beta - 1}, \Pr(X = 2) = \frac{\mu - 1}{\mu} \left[\frac{(\beta - 1)\mu}{\beta\mu - 1} \right]^{2\beta - 1}, \dots, \tag{13}$$

and the recurrence formula

$$\Pr(X = K + 1) = \prod_{i=1}^k \left[1 + \frac{\beta}{k\beta - i} \right] \frac{\mu - 1}{\mu} \left[\frac{(\beta - 1)}{\beta\mu - 1} \right]^\beta \Pr(X = k), \quad \text{for } k = 2, 3, 4, \dots \tag{14}$$

The probabilities for the Geeta distribution (13) were computed for $\mu = 1.2$ to 5.2 varying by 0.2 and for values of β varying from 1.2 to 4.2 (with increments of 0.6) and bar-diagrams were drawn for all of them to see the variations. Twelve of these bar-diagrams are shown below for two typical values of $\mu = 1.2$ to 5.2 and for $\beta = 1.2, 1.8, 2.4, 3.0, 3.6, 4.2$ corresponding to each values of μ .

Bar-diagrams of Geeta Distribution for $\mu = 1.2$

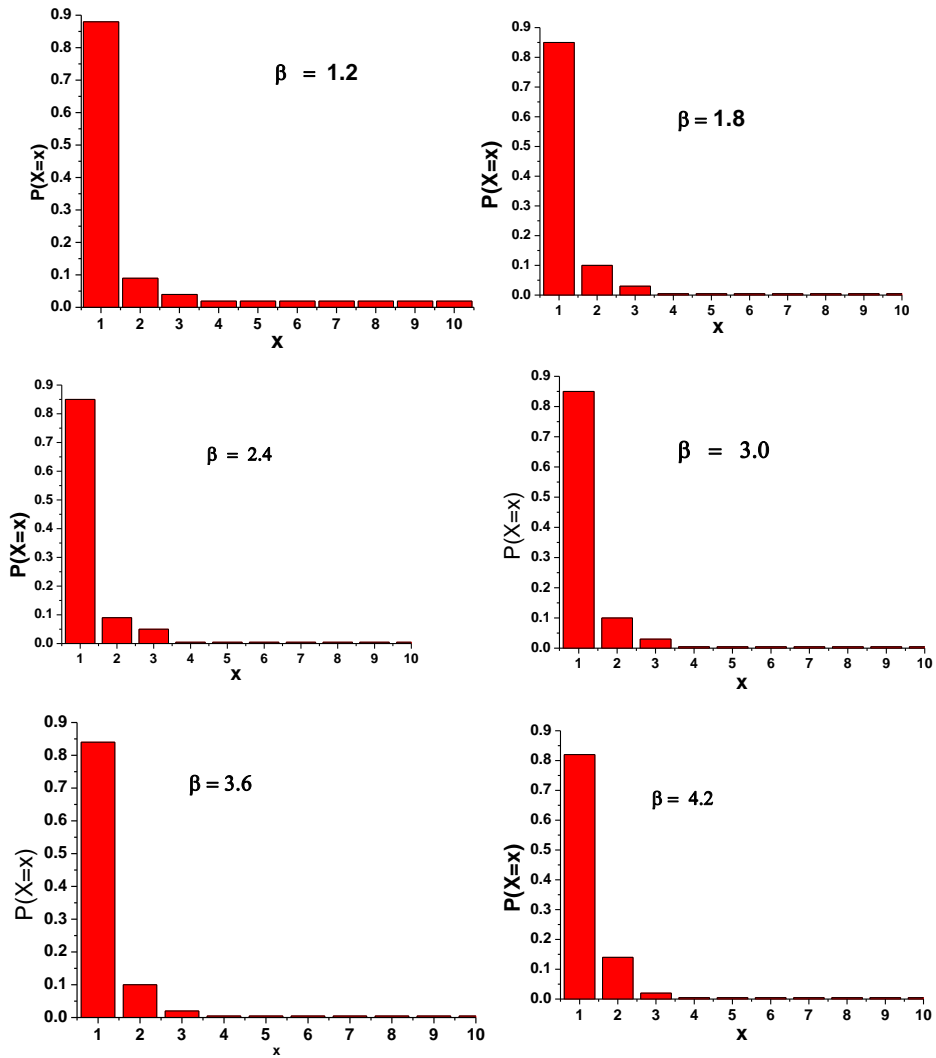


Figure 1. Probabilities for values of X at $\mu = 1.2$ for different values of β (Consul, 1990a)

Bar-diagrams of Geeta Distribution for $\mu = 5.2$

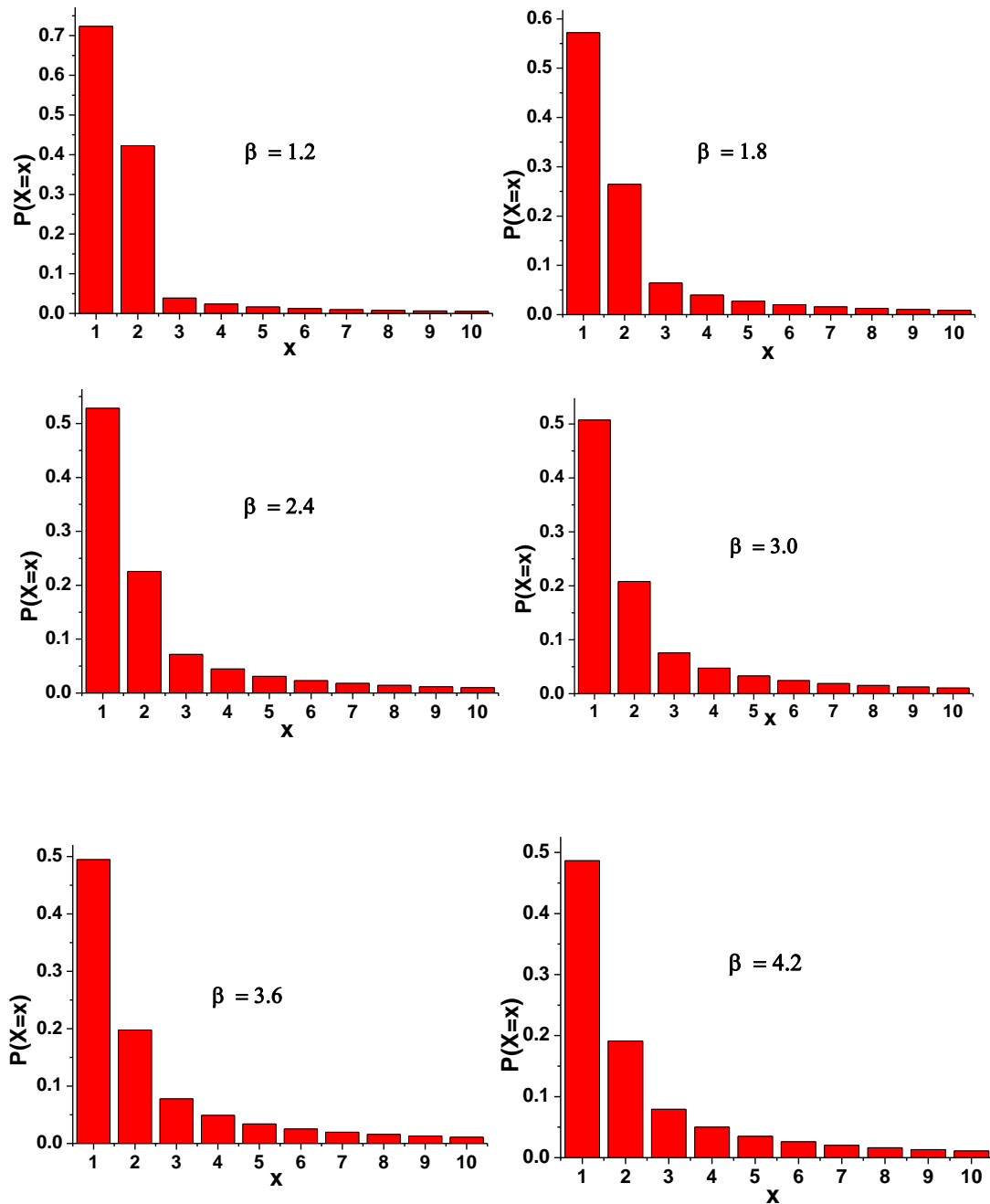


Figure 1. Probabilities for values of X at $\mu=5.2$ for different values of β (Consul, 1990a)

It is clear from these graphs that $\Pr(X = 1)$ reduces as μ increases and the probabilities for all other values of x increase but the model always remains L-shaped. Thus the tail becomes more and more heavy and longer with the increase in the value of μ . There is a similar effect when the value of μ is kept fixed and the value of β is slowly increased. The value of $\Pr(X = 1)$ decreases and the probabilities for other values of x increase as β increases. However, these changes for β are at a much slower pace than the changes for μ with the result that the Geeta probability model becomes more suitable and versatile than some other models for abundance data sets.

2. Methodology

The parametric interval estimation is twofold; finding interval estimators and then determining good, or optimum interval estimator. In the preliminaries stages the identification of a statistics is obtained by taking a random sample of

size, n from the entire population. A statistic can be expressed a function of sample observation and there are many statistics from same population that can be obtained. The statistics is chosen from either a sample mean or sample proportion and the confidence coefficient is obtained by determining its significant level such as, for 5 % significant level its confidence coefficient is 95 % or 0.95. The statistics can be biased or unbiased estimator and the properties of a good estimator informs the choice of a statistics. The construction of confidence intervals (CI) by choosing appropriate $Z_{\alpha/2}$ values which are obtained from the standard normal tables (Arnab et al, 2017, Warisa et al, 2017). It largely depends on whether the population mean is known or unknown and similarly whether population variance is known or unknown. There are other cases of constructing the CI for the difference of two means where the student's t distribution tests are used (Suparat et al, 2017). An approximate confidence interval for a population mean can be constructed for random variables that are not normally distributed in the population relying on the central theorem for sample sizes which are large

$$(n \geq 30)$$

This paper is divided into sections on which the following are discussed, shortest-length confidence interval, unbiased confidence intervals and Bayesian confidence interval.

2.1 Confidence Interval Estimation

Definition 1:

An interval estimate for a real-valued parameter θ based on a sample $\underline{X} \equiv (X_1, \dots, X_n)$ is a pair of functions $L(\underline{X})$ and $U(\underline{X})$ so that $L(\underline{X}) \leq U(\underline{X})$ for all \underline{X} , that is $[L(\underline{X}), U(\underline{X})]$ (Wei Zhu, 2017)

Theorem 2.1:

The confidence interval estimate of $\hat{\theta}$ is given by

$$\Pr \left[L_1 < \frac{\hat{\theta} - \theta}{\sigma} < L_2 \right] = 1 - \alpha$$

where L_1 and L_2 are lower and upper limits respectively

Proof:

It follows then that, from the central limit theorem $\frac{\hat{\theta} - \theta}{\sigma}$ is approximately normal with limit so that

$$\Pr[-Z < \mu < Z] = 1 - \alpha \quad (15)$$

where

$$1 - \alpha = \Phi(Z) - \Phi(-Z)$$

Therefore, $T = \frac{\hat{\theta} - \theta}{\sigma}$ is an approximate pivotal quantity. Thus

$$\Pr[L_1 < \mu < L_2] = 1 - \alpha \quad (16)$$

where α is the confidence coefficient and, L_1 and L_2 are lower and upper limits of the parameter θ and $(1-\alpha)$ 100% is confidence interval for θ (Mood, A.N. et al, 1963)

2.2 Pivotal Method

Definition2:

A pivotal quantity is a function of the sample and the parameter of interest. Furthermore, its distribution is entirely known.

Let X_1, X_2, \dots, X_n be a random sample for a normal population with mean, μ and variance, σ^2 . That is,

$$X_i \sim N(\mu, \sigma^2), i = 1, \dots, n, \text{ and } X_i \text{ is identically and independently distributed.}$$

The method of finding a confidence interval involves finding, Z , the pivotal quantity which is a function of the samples and the parameter to be estimated is as shown below in the examples (Arnab et al, 2017).

Example 1

When σ is known, the pivotal quantity given by;

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

The $(1-\alpha)$ 100 % CI for a given $\alpha=0.05$, then the 95 % CI for μ is given by

$$P_{\mu} \left[\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$$

then, $Z_{\alpha/2} = Z_{0.025} = 1.96$ from the standard normal tables

and hence

$$P_{\mu} \left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right] = 0.95$$

therefore the 95 % CI for mean, μ

$$\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right].$$

Example 2

When σ is unknown, the pivotal quantity is given by;

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

The $(1-\alpha)$ 100 % CI for a given $\alpha=0.05$, then the 95 % CI for μ is given by

$$P_{\mu} \left[\bar{X} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \right] = 1 - \alpha$$

Thus $(1-\alpha)$ 100 % CI for mean, μ is given by

$$\left[\bar{X} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}; \bar{X} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \right]$$

where $t_{n-1,\alpha/2}$ is the student's t distribution test with $(n-1)$ degrees of freedom and α is the significant level.

Definition 3:

Let x_1, x_2, \dots, x_n random sample X_1, X_2, \dots, X_n from density $f_x(\cdot; \theta)$.

Let $T_1 = t_1(x_1, x_2, \dots, x_n)$ and $T_2 = t_2(x_1, x_2, \dots, x_n)$ be two statistics satisfying $T_1 \leq T_2$ for which

$$P_{\theta}[T_1 < \mu < T_2] = 1 - \alpha$$

where $(1-\alpha)$ does not depend on θ , then the random interval (T_1, T_2) is called a $(1-\alpha)$ 100 % interval for $\tau(\theta)$ is called the confidence coefficient while T_1 and T_2 are called lower and upper coefficient limits respectively. The confidence interval estimators of the parameters θ and β for the distribution given in (10). For small sample size n , confidence interval estimation is not possible because we do not have the Geeta distribution tables to refer to, hence large sample distributions are considered where the normal table is going to be used. The large sample distribution for maximum likelihood estimators is used to derive the confidence intervals for the unknown parameter. If the maximum likelihood estimator $\hat{\theta}$ of a parameter θ is for large sample size n , it is approximately normally distributed with mean θ and variance σ^2 where

$$\sigma^2 = \frac{1}{n E \left[\frac{\partial}{\partial \theta} \ln P(x) \right]^2} \tag{17}$$

in which $P(x)$ is a discrete distribution.

2.3 Shortest – Length Confidence Interval

Large-sample confidence intervals based on maximum likelihood estimator will be shorter on the average than intervals determined by any other estimator.

If $\Pr[a < Z < b] = 0.95$

then to determine the length of the interval

$$L = \frac{b-a}{\sqrt{n}} \sigma$$

But from the solution of eq. (4), $a = -b$ thus

$$L = 2b \frac{\sigma}{\sqrt{n}}$$

2.4 Unbiased Confidence Interval

Let $S(x)$ be a family of $(1 - \alpha)$ level confidence intervals of finite length, that is $S(x) = [\underline{\theta}(x), \bar{\theta}(x)]$ and suppose $[\underline{\theta}(x), \bar{\theta}(x)]$ is finite where $\underline{\theta}(x)$ and $\bar{\theta}(x)$ are functions of x alone not θ with $\underline{\theta}(x)$ and $\bar{\theta}(x)$ as upper and lower bound respectively. $\underline{\theta}(x)$ may be $-\infty$ and $\bar{\theta}(x)$ may be $+\infty$ and $S(x)$ is a random interval $[\underline{\theta}(x), \bar{\theta}(x)]$.

2.5 Bayesian Confidence Interval

The Bayesian confidence interval estimate of θ is given by

$$\int_{t_1}^{t_2} K(\theta/X) d\theta = 1 - \alpha$$

where t_1 and t_2 are the upper and lower limits of Bayesian interval.

3. Results of the Interval Estimators

The following results were obtained.

Theorem 3.1.1

Using the pivotal quantity method interval, the confidence estimate interval of θ is given by

$$\left(\hat{\theta} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}} ; \hat{\theta} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}} \right)$$

where $\hat{\theta}$ is the point estimator of θ using the method of moments given as

$$\hat{\theta} = \frac{\bar{X} - 1}{\bar{X}\hat{\beta} - 1}$$

Proof:

From the distribution given equation in (5), we have

$$\log[f(x, \theta, \beta)] = -\log(\beta x - 1) + \log\left(\frac{\beta x - 1}{x}\right) + (x - 1)\log \theta + (\beta x - x)\log(1 - \theta) \tag{18}$$

Therefore differentiating partially w.r.t θ , we have

$$\frac{\partial}{\partial \theta} \log[f(x, \theta, \beta)] = \frac{1}{\theta(1-\theta)} [-(1-\theta) + x(1-\beta\theta)] \tag{19}$$

and

$$\left[\frac{\partial}{\partial \theta} \log(f(x, \theta, \beta)) \right]^2 = \frac{1}{\theta^2(1-\theta)^2} [-(1-\theta) + x(1-\beta\theta)]^2 \tag{20}$$

hence

$$\begin{aligned} E \left[\frac{\partial}{\partial \theta} \log(f(x, \theta, \beta)) \right]^2 &= \frac{E[(1-\theta)^2 - 2(1-\theta)x(1-\beta\theta) + x^2(1-\beta\theta)^2]}{\theta^2(1-\theta)^2} \\ &= \frac{[(1-\theta)^2 - 2(1-\theta)E(x)(1-\beta\theta) + E(x^2)(1-\beta\theta)^2]}{\theta^2(1-\theta)^2} \end{aligned} \tag{21}$$

$$E \left[\frac{\partial}{\partial \theta} \log(f(x, \theta, \beta)) \right]^2 = \frac{(\beta-1)}{\theta(1-\theta)(1-\beta\theta)} \tag{22}$$

therefore

$$\sigma_n(\theta) = \frac{1}{nE \left[\frac{\partial}{\partial \theta} \log(f(x, \theta, \beta)) \right]^2} = \frac{\theta(1-\theta)(1-\beta\theta)}{n(\beta-1)} \tag{23}$$

thus from (15) we have

$$\Pr \left[-Z_{1-\alpha/2} \leq \frac{\theta - \theta\sqrt{n}}{\sigma_n(\theta)} \leq Z_{\alpha/2} \right] = 1 - \alpha \tag{24}$$

which further simplifies to give

$$\Pr \left[\hat{\theta} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}} \leq \theta \leq \hat{\theta} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}} \right] = 1 - \alpha \tag{25}$$

Hence the confidence estimate interval of θ is given by

$$\left(\hat{\theta} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}} ; \hat{\theta} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}} \right) \tag{26}$$

Theorem 3.1.2:

Using the pivotal quantity method, the confidence estimate interval of β is given by

$$\left(\hat{\beta} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\sigma_n^2(\beta)} ; \hat{\beta} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\sigma_n^2(\beta)} \right)$$

Proof:

Assuming X_1, X_2, \dots, X_n random sample X_1, X_2, \dots, X_n from density $f(X, \theta, \beta)$.

Let $T = t(X_1, X_2, \dots, X_n)$ be a statistic, in this case $T = \sum X_i$ is given by

$$f(t; \theta, \beta) = \theta^{t-n} (1-\theta)^{\beta t - t} \prod_{i=1}^n \frac{1}{\beta x_i - 1} \binom{\beta x - 1}{x} \tag{27}$$

Let t_0 , denote an observed value of T ; that is $t_0 = t(X_1, X_2, \dots, X_n)$, then

$$f(t; \theta, \beta) = \frac{n}{\beta x_i - 1} \binom{\beta x - 1}{x} \theta^{x-n} (1-\theta)^{\beta x - x}, \quad x = 1, 2, 3, \dots \tag{28}$$

For a given observed value $t_0 = t(x_1, x_2, \dots, x_n)$, we need to find $V_1 = V_1(x_1, x_2, \dots, x_n)$ and $V_2 = V_2(x_1, x_2, \dots, x_n)$, such that

$$\Pr[V_1(x_1, x_2, \dots, x_n) \leq \theta \leq V_2(x_1, x_2, \dots, x_n)] = 1 - p_1 - p_2 \tag{29}$$

The following equations are used to solve for θ

$$P_1 = \sum_{t=0}^{t_0} f_T(t; \theta, \beta) \text{ and } P_2 = \sum_{t=t_0}^n f_T(t; \theta, \beta)$$

The solution for the above equations, gives

$$P_1 = \sum_{t=0}^{t_0} \frac{1}{\beta t - n} \binom{\beta t - 1}{t} \theta^{t-n} (1-\theta)^{\beta t - t} \text{ and } P_2 = \sum_{t=t_0}^n \frac{1}{\beta t - n} \binom{\beta t - 1}{t} \theta^{t-n} (1-\theta)^{\beta t - t}$$

where V_1 and V_2 are upper and lower limits.

Therefore $[V_1; V_2]$ is a $100(1 - p_1 - p_2)\%$ confidence interval for θ where

$$V_i = V_i(x_1, x_2, \dots, x_n) \text{ for } i = 1, 2.$$

Following the same procedure the interval for parameter β is obtained as

$$\left(\hat{\beta} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\sigma_n^2(\beta)}; \hat{\beta} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\sigma_n^2(\beta)} \right) \tag{30}$$

and statistically

$$\Pr[V_1(x_1, x_2, \dots, x_n) \leq \beta \leq V_2(x_1, x_2, \dots, x_n)] = 1 - p_1 - p_2$$

Theorem 3.1.3:

The shortest-length confidence interval for parameter θ is obtained as

$$\left(\hat{\theta} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sigma_n(\theta); \hat{\theta} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sigma_n(\theta) \right)$$

and hence the length $L = \frac{2 Z_{1-\alpha/2}}{\sqrt{n}} \sigma_n(\theta)$

Proof:

We follow this procedure:

The distance $(b - a)$ can be minimized for fixed area when $\Phi(a) = \Phi(b)$ given that

$$\Pr\left[a < \frac{\hat{\theta} - \theta}{\sigma/\sqrt{n}} < b\right] = 1 - \alpha \tag{31}$$

which can be re-written in this form

$$\Pr\left[\hat{\theta} - \frac{b}{\sqrt{n}}\sigma_n < \theta < \hat{\theta} - \frac{a}{\sqrt{n}}\sigma_n\right] = 1 - \alpha \tag{32}$$

Therefore, the length of this interval is

$$L = \frac{(b - a)}{\sqrt{n}}\sigma_n(\theta)$$

Minimizing L such that

$$\Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

we have

$$\frac{dL}{da} = \frac{\sigma}{\sqrt{n}} \left(\frac{db}{da} - 1 \right) \text{ and } \psi(b) \frac{db}{da} - \psi(a) = 0$$

which further simplifies to $\frac{db}{da} = \frac{\psi(a)}{\psi(b)}$

thus

$$\frac{dL}{da} = \frac{\sigma}{\sqrt{n}} \left(\frac{\psi(a)}{\psi(b)} - 1 \right)$$

The minimum occurs when $\Phi(b) - \Phi(a)$ that is when $a = b$ or $a = -b$, $a = b$ does not satisfy

$\int_a^b \phi(t) dt = 1 - \alpha$ therefore we choose $a = -b$. The shortest length confidence interval based on $T = \frac{\hat{\theta} - \theta}{\sigma/\sqrt{n}}$ is

given by

$$\left(\bar{X} - \frac{Z_{\alpha/2}}{\sqrt{n}}\sigma_n(\theta); \bar{X} + \frac{Z_{\alpha/2}}{\sqrt{n}}\sigma_n(\theta) \right) \tag{33}$$

and hence the length

$$L = 2 \frac{Z_{\alpha/2}}{\sqrt{n}}\sigma_n(\theta)$$

Following the same procedure as above, the shortest-length confidence interval for parameter β is obtained as

$$\left(\hat{\beta} - \frac{Z_{\alpha/2}}{\sqrt{n}} \sigma_n(\beta); \hat{\beta} + \frac{Z_{\alpha/2}}{\sqrt{n}} \sigma_n(\beta) \right) \tag{34}$$

and hence the length

$$L = 2 \frac{Z_{\alpha/2}}{\sqrt{n}} \sigma_n(\beta)$$

Theorem 3.1.4:

By use of Unbiased confidence interval estimation, the following results were obtained;

If $T_{\theta}(x) = \frac{\hat{\theta} - \theta}{\sigma_n(\theta)} = T$ is a pivotal quantity for Geeta distribution then we choose $\lambda_1(\alpha)$ and $\lambda_2(\alpha)$ such that

$$\Pr[\lambda_1(\alpha) < T_{\theta} < \lambda_2(\alpha)] = 1 - \alpha \tag{35}$$

which can be converted to

$$P_{\theta}[\underline{\theta}(x) < \hat{\theta} < \bar{\theta}(x)] = 1 - \alpha \tag{36}$$

In order for $[\underline{\theta}, \bar{\theta}]$ to be unbiased we must have

$$P(\underline{\theta}, \bar{\theta}) = P_{\theta}[\underline{\theta}(x) < \hat{\theta} < \bar{\theta}(x)] = 1 - \alpha \text{ if } \underline{\theta} = \bar{\theta} \tag{37}$$

$$P(\underline{\theta}, \bar{\theta}) = P_{\theta}[\underline{\theta}(x) < \hat{\theta} < \bar{\theta}(x)] < 1 - \alpha \text{ if } \underline{\theta} \neq \bar{\theta}$$

If $P(\underline{\theta}, \bar{\theta})$ depends only on a function γ of $\underline{\theta}$ and $\bar{\theta}$ which we may write as

$$\begin{aligned} P(\gamma) &= 1 - \alpha \text{ if } \underline{\theta} = \bar{\theta} \\ P(\gamma) &< 1 - \alpha \text{ if } \underline{\theta} \neq \bar{\theta} \end{aligned} \tag{38}$$

Then

$$P(\underline{\theta}, \bar{\theta}) = P_{\theta} \left(\lambda_1 < \frac{\hat{\theta} - \theta}{\sigma_n(\theta)} < \lambda_2 \right) = P_{\theta}(\hat{\theta} - \lambda_2 \sigma_n(\theta) < \bar{\theta} < \hat{\theta} + \lambda_1 \sigma_n(\theta)) = 1 - \alpha$$

Thus we need λ_1 and λ_2 such that

$$P(\gamma) = 1 - \alpha \text{ and } P(\gamma) < 1 - \alpha$$

Having obtained λ_1 and λ_2 which is $Z_{1-\frac{\alpha}{2}}$ and $Z_{\frac{\alpha}{2}}$ respectively

We get

$$P \left(\hat{\theta} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}} \sigma_n(\theta) < \bar{\theta} < \hat{\theta} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma_n(\theta) \right) = 1 - \alpha$$

The unbiased confidence interval for θ is therefore given by

$$\left(\hat{\theta} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}} \sigma_n(\theta), \hat{\theta} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma_n(\theta) \right) \tag{39}$$

where

$$\hat{\theta} = \frac{\bar{X} - 1}{\bar{X}\beta - 1}$$

and

$$\sigma_n(\theta) = \frac{\theta(1-\theta)(1-\beta\theta)}{(\beta-1)}$$

The pivotal quantity for the parameter β is given by

$$T_\beta = \frac{\hat{\beta}-\beta}{\sigma_n(\beta)} \tag{40}$$

Therefore

$$P(\underline{\beta}, \bar{\beta}) = P_\beta\left(\lambda_1 < \frac{\hat{\beta}-\beta}{\sigma\sqrt{n}(\beta)} < \lambda_2\right) = P_\beta(\hat{\beta} - \lambda_2\sigma_n(\beta) < \bar{\beta} < \hat{\beta} + \lambda_1\sigma_n(\beta)) = 1 - \alpha$$

Then λ_1 and λ_2 are chosen such that

$$P_\beta(\hat{\beta} - \lambda_2\sigma_n(\beta) < \bar{\beta} < \hat{\beta} + \lambda_1\sigma_n(\beta)) = 1 - \alpha$$

Given that $\lambda_1 = Z_{1-\frac{\alpha}{2}}$ and $\lambda_2 = Z_{\frac{\alpha}{2}}$ then,

$$P_\beta\left(\hat{\beta} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}\sigma_n(\beta) < \bar{\beta} < \hat{\beta} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sigma_n(\beta)\right) = 1 - \alpha$$

and thus the unbiased confidence interval for β is

$$\left(\hat{\beta} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}\sigma_n(\beta), \hat{\beta} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sigma_n(\beta)\right) \tag{41}$$

Theorem 3.1.5:

The Bayesian confidence interval estimate of θ is given by (Saxena, H.C. and Surendran, P.U., 1967)

$$\int_{t_1}^{t_2} K\left(\frac{\theta}{x}\right) d\theta = 1 - \alpha$$

where t_1 and t_2 are the upper and lower limits of Bayesian interval

Proof:

Given that

$$g(x/\theta, \beta) = \frac{1}{\beta x - 1} \binom{\beta x - 1}{x} \theta^{x-1} (1 - \theta)^{\beta x - x}, \quad x = 1, 2, 3, \dots$$

then

$$\begin{aligned} g(\underline{x}/\theta, \beta) &= \prod_{i=1}^n \frac{1}{\beta x_i - 1} \binom{\beta x_i - 1}{x_i} \theta^{x_i-1} (1 - \theta)^{\beta x_i - x_i}, \quad x_i = 1, 2, 3, \dots \\ &= \theta^{\sum x_i - n} (1 - \theta)^{\beta \sum x_i - \sum x_i} \prod_{i=1}^n \binom{\beta x_i - 1}{x_i} \frac{1}{\beta x_i - 1} \end{aligned} \tag{42}$$

The joint p .d .f of X and θ is given by

$$K(\theta/x) = h(\theta)g(\underline{x}/\theta, \beta) = \theta^{x+\sum x_i - n} (1 - \theta)^{n-x+\beta \sum x_i - \sum x_i} \binom{n}{x} \prod_{i=1}^n \binom{\beta x_i - 1}{x_i} \frac{1}{\beta x_i - 1} \tag{43}$$

and the marginal p .d .f of X is given by

$$\begin{aligned} K_1(x) &= \int_0^1 h(\theta)g(\underline{x}/\theta, \beta) d\theta \\ &= \binom{n}{x} \prod_{i=1}^n \binom{\beta x_i - 1}{x_i} B(x + \sum x_i - n, n - x + \beta \sum x_i - \sum x_i + 1) \frac{1}{\beta x_i - 1} \end{aligned} \tag{3.25}$$

The posterior p.d.f is given by

$$K(\theta/x) = \frac{h(\theta)g(\underline{x}/\theta, \beta)}{k_1(x)} = \frac{\theta^{x+\sum x_i - n} (1 - \theta)^{n-x+\beta \sum x_i - \sum x_i}}{B(x+\sum x_i - n + 1, n - x + \beta \sum x_i - \sum x_i + 1)} \tag{44}$$

The Bayesian confidence interval estimate of θ is therefore given by

$$\int_{t_1}^{t_2} k(\theta/x) d\theta = 1 - \alpha$$

therefore

$$\int_{t_1}^{t_2} k\left(\frac{\theta}{x}\right) d\theta = \Phi\left(\frac{t_2 - \bar{x}}{\sigma}\right) - \Phi\left(\frac{t_1 - \bar{x}}{\sigma}\right) = 1 - \alpha \tag{45}$$

which implies that

$$\Phi\left(\frac{t_2 - \bar{x}}{\sigma}\right) - \Phi\left(\frac{t_1 - \bar{x}}{\sigma}\right) = 1 - \alpha$$

Therefore the distribution of $K(\theta/x)$ is beta, $B(\alpha, \beta)$ where

$a = x + \sum x_i - n + 1$ and $\beta = n - x + \beta t - t + 1$, with mean

$$E(x) = \frac{\alpha}{\alpha + \beta} = \frac{x + \sum x_i - n + 1}{\beta \sum x_i + 2}$$

and

$$E(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{(x + \sum x_i - n + 1)(n - x + \beta \sum x_i - \sum x_i + 1)}{(\beta \sum x_i + 2)^2(\beta \sum x_i + 3)}$$

thus

$$t_2 = \frac{x + \sum x_i - n + 1}{\beta \sum x_i + 2} + z \sqrt{\frac{(x + \sum x_i - n + 1)(n - x + \beta \sum x_i - \sum x_i + 1)}{(\beta \sum x_i + 2)^2(\beta \sum x_i + 3)}}$$

and

$$t_1 = \frac{x + \sum x_i - n + 1}{\beta \sum x_i + 2} - z \sqrt{\frac{(x + \sum x_i - n + 1)(n - x + \beta \sum x_i - \sum x_i + 1)}{(\beta \sum x_i + 2)^2(\beta \sum x_i + 3)}} \tag{46}$$

4. Discussion and Conclusion

In most branches of knowledge, experiments are a way of life and in probability and statistics, there are special types of experiments which are in practice the distribution models of either discrete or continuous distribution. In most of the scientific experiments, manufacturing industries, and the medicine world, this distribution model has been applied and it has eased the handling of abundance data sets. The Geeta distribution is more versatile than the other given distribution belonging to the same class, this is because of the presence of the two unknown parameters. These unknown parameters are the most vital part of the distribution, because the information derived is based on them.

This new model is versatile and flexible to fit observed count data sets and can be used efficiently to model different types of sets. In this study, Geeta distribution and its properties were reviewed and discussed, more so the unknown parameters were obtained, and it is a preferred distribution model to be used because of its relevance in modern research and application in genetics, sales and fashion industry, branching processes(queue) and bunching models.

Therefore, the properties and characteristics the Geeta distribution satisfy and meet the conditions of any probability distribution, that is sum of all probabilities for all values of a random variable x is unity. The mean, variance, moment generating function, probability generating function and the central theorem exist (Rohatgi, V.K., 1976). The confidence interval estimators of the parameters were obtained by use of large sample distribution, and it was proven that the optimum confidence intervals estimators occur for shortest –length intervals. Large-sample confidence intervals based on maximum likelihood estimator will be shorter on the average than intervals determined by any other estimator.

References

Consul, P. C. (1990). New class of location-parameter discrete probability distributions and their characterizations. *Communications in Statistics-Theory and Methods*, 19(12), 4653-4666. <https://doi.org/10.1080/03610929008830465>

Consul, P. C. (1990a). Geeta distribution and its Properties, *Communication in Statistics-Theory and Methods*, 19, 3051-3068[7.2.4]. <https://doi.org/10.1080/03610929008830364>

Harold, J. L. (1934). *Statistics: An introduction to Statistics*.Pg. 185-209.

- Hogg, R. V., & Craig, A. T. (1956). Introduction to mathematical statistics. Pg.341-357.
- Koley, A., & Kundu, D. (2017). Interval Estimation of the Unknown Exponential Parameter Based on Time Truncated Data. *arXiv preprint arXiv:1703.01051*.
- Mood, A. N., Gray, B. F. A., & Boes, D. C. (1963). Introduction to the theory of statistics. Pg.372-396.
- Rohatgi, V. K. (1976). An introduction to probability theory and mathematical statistics. Pg.499-530.
- Saxena, H. C., & Surendran, P. U. (1967). Statistical Inference. Pg.58-70.
- Suparat, N., & Sa-aat, N. (2017). Confidence intervals for the difference between normal means with known coefficients of variation. *Annals of Operations Research*, 256(2), 237–251. <https://doi.org/10.1007/s10479-016-2388-9>
- Thangjai, W., Niwitpong, S., & Niwitpong, S. A. (2017). Confidence Intervals for Mean and Difference between Means of Normal Distributions with Unknown Coefficients of Variation. *Mathematics*, 5(3), 39. <https://doi.org/10.3390/math5030039>
- Traoré B., Diabaté L., Touré B., & Fané A. (2018). The Shortest Confidence Interval for the Mean of a Normal Distribution, *International Journal of Statistics and Probability*, 7(2).
- Wei, Z. (2017). <http://www.ams.sunysb.edu/~zhu/ams571/ams571.pdf>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).