Banneheka, B.M.S.G., Ekanayake, G.E.M.U.P.D. Viyodaya Journal of Science, 2009. Vol 14. pp. 95-103

A NEW POINT ESTIMATOR FOR THE MEDIAN OF GAMMA DISTRIBUTION

B.M.S.G. Banneheka
Department of Statistics and Computer Science,
University of Sri Jayewardenepura,
Nugegoda, Sri Lanka
and
G.E.M.U.P.D Ekanayake
Department of Census and Statistics
Prices and wages division,
104A, Kitulwatta Road, Colombo 8, Sri Lanka.

Abstract

In this paper, we consider the problem of estimating the median of a gamma distribution. We introduce a new point estimator based on an approximation that we derive for the median of a gamma distribution. We compare the new estimator with two conventional estimators, namely the sample median and the maximum likelihood estimator (mle). Comparison is based on the amount of computations required to calculate the estimates and the root mean square errors of the estimators. The new estimator is shown to be 'optimum' with respect to these two criteria.

Keywords and phrases: gamma distribution, median, point estimate, maximum likelihood estimate, moment estimate.

1. Introduction

Estimation of population 'average' or 'central tendency' is a common inferential problem. Population mean and population median are the commonly used parameters to represent the population average. Most researchers consider mean to represent the average because the inference concerning the mean is easy. Sample mean is an unbiased estimator for the population mean. The central limit theorem can be used to derive sample confidence intervals and to test hypotheses when large samples are available.

However, when the underlying distribution is skewed, the population mean tends to be larger (when positively skewed) or smaller (when negatively skewed) than the typical population 'average'. For example, consider the monthly income of households in a fixed area. The monthly incomes of most of the households are small to moderately large. There may be few households with very large monthly incomes. Then the distribution of household incomes is positively skewed and the population mean can be significantly larger than the typical 'average' monthly household income. In such situations, the population median is better than the population mean to represent the population 'average'.

When the population median is selected to represent the population average, the next problem is how to make inference regarding the population median. The parametric approach is to select a suitable model for the distribution of the variable of interest and make inference regarding the median of the selected model distribution. The gamma distribution is often used as a model for positively skewed distributions. Literature related to inference concerning the mean of a gamma distribution can be found in Anita S. et.al. (2002) and references therein. However, we could not find any literature related to the inference concerning the median of a gamma distribution. In this paper we consider the problem of estimating the median of a gamma distribution. We intend to present a way to construct confidence intervals for the median of a gamma distribution, in another paper.

2. An Approximation for the Median of Gamma Distribution

If a random variable X has a gamma distribution with shape parameter α (>0) and scale parameter β (>0), it is denoted as $X \sim G(\alpha, \beta)$ (Anita S. et.al.,2002). Its density function is given by

$$f_X(x;\alpha,\beta) = \frac{e^{-x/\beta} x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}, \quad x > 0, \alpha > 0, \beta > 0. \qquad (1)$$

Using simple calculus, it is easy see that

$$\lim_{x \to 0} f_X(x;\alpha,\beta) = \begin{cases} \infty & \alpha < 1 \\ \frac{1}{\beta} & \alpha = 1 \\ 0 & \alpha > 1 \end{cases} \qquad (2)$$

Figure 1 shows the three different shapes arising from the above three cases.

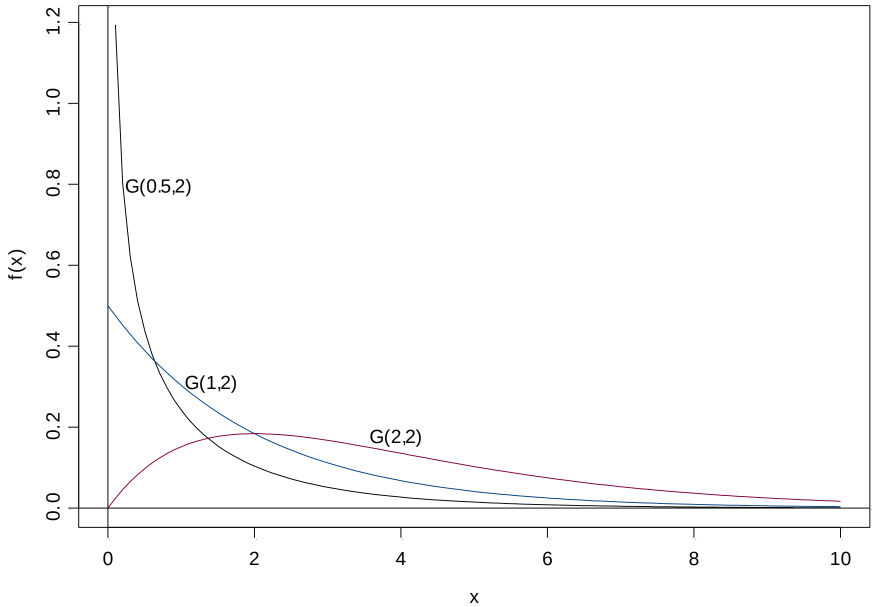

Figure 1: Densities of G(0.5,2), G(1,2) and G(2,2)

For the above distribution, mean (μ) = $\alpha\beta$, standard deviation (σ) = $\sqrt{\alpha}\beta$, and skewness = $\frac{2}{\sqrt{\alpha}}$ (Anita S. et.al.,2002). The skewness depends only on the shape parameter. As α increases, skewness decreases, and consequently the gamma distribution approaches a normal distribution when α is large (e.g., $\alpha \exists 10$) (Anita S. et.al., 2002).

Let ν be the median of the above gamma distribution. According to the definition, ν satisfies the equation

$$\int_0^{\nu} f_x(x;\alpha,\beta)dx = 0.5. \tag{3}$$

It is not possible to write ν in terms of α and β explicitly (http://en.wikipedia.org/wiki/Gamma_distribution). However, the value of ν for given values of α and β can be obtained using the 'INVCDF' function in the statistical package Minitab or 'qgamma' function in the statistical package R (http://www.r-project.org/),

Here we derive an approximation for ν using two interesting features that we observed of the ratio μ/(μ−ν). The first is that μ/(μ−ν) is free of β. In order to see this, suppose X~ G(α,β). Then, using the moment generating function technique (Mood A.M., et.al., 2001, pg. 189) it can be shown that X/β ~ G(α,1).

If ν is the median of X, then $\Pr(X<\nu)=0.5$. Hence, $\Pr(X/\beta<\nu/\beta)=0.5$. This implies that the median of $(X/\beta) = \nu/\beta$. In other words, $\nu = \beta*$ the median of $G(\alpha,1)$ distribution. Therefore, $\mu/(\mu-\nu) = \alpha\beta/(\alpha\beta - \beta*$median of a $G(\alpha,1)$ distribution). This implies

$$\mu/(\mu-\nu) = \alpha/(\alpha - \text{median of a } G(\alpha,1) \text{ distribution}). \qquad (4)$$

From (4), it is clear that $\mu/(\mu-\nu)$ is free of β and it is a function of α only. Figure 2 shows the relationship between $\mu/(\mu-\nu)$ and α.

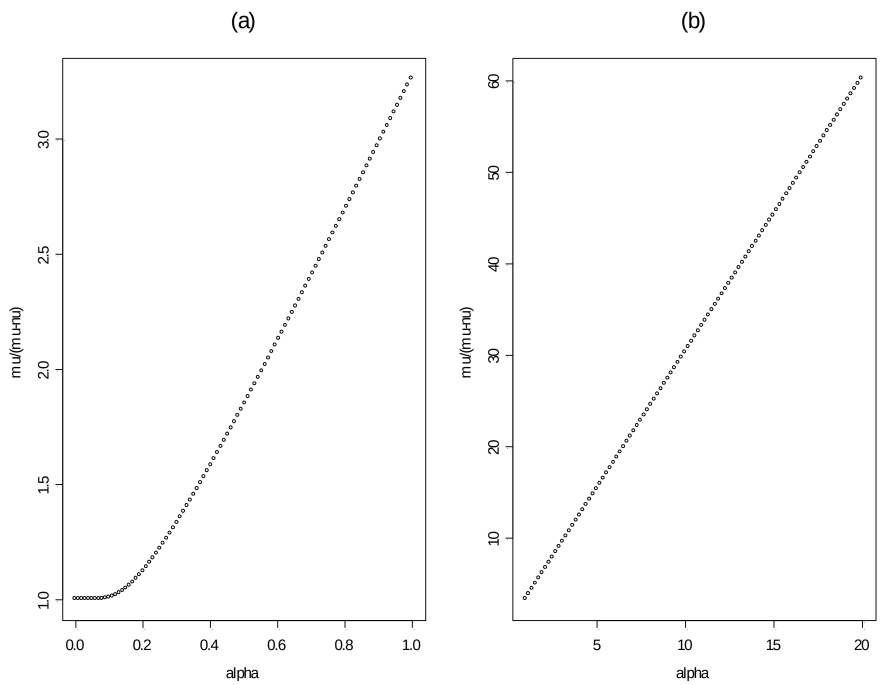

Figure 2: $\mu/(\mu-\nu)$ versus α

Figure 2 (a) is the plot of $\mu/(\mu-\nu)$ against α when $\alpha < 1$. Figure 2 (b) is the same when $\alpha \geq 1$. In order to produce these graphs, the medians of $G(\alpha,1)$ distributions for different values of α were obtained using the function 'qgamma' of the statistical package R. When $\alpha<1$, the relationship is non-linear. However, when $\alpha \geq 1$ $\mu/(\mu-\nu)$ is almost perfectly linear in α. This is the second interesting feature.

When $\alpha \geq 1$, the suitable values for the slope and intercept of the linear relationship can be obtained using the least square method. Based on 100 equally spaced α values between 1 and 20 and the corresponding $\mu/(\mu-\nu)$ values, the least square estimates for the slope and intercept are 0.2096 and 2.998 respectively. For simplicity, using 0.2 and 3 as the intercept and slope, we can write $\frac{\mu}{\mu-\nu} \approx 0.2 + 3\alpha$ or equivalently $\nu \approx \mu\frac{(3\alpha - 0.8)}{(3\alpha + 0.2)}$. We denote this approximation as

$$\nu_{BE} = \mu \frac{(3\alpha - 0.8)}{(3\alpha + 0.2)}, \quad \alpha \geq 1 \tag{5}$$

Table 1 shows the absolute error of the approximation ν_{BE} calculated as a percentage of the actual median ν ($\frac{|\nu - \nu_{BE}|}{\nu} * 100$).

Table 1: Absolute error of ν_{BE} as a percentage of actual median.

α	$\frac{\lvert \nu - \nu_{BE} \rvert}{\nu} * 100$ ν=actual median ν_{BE} = approximation for ν
1	0.8147
5	0.0031
10	0.0017
20	0.0005

These values show that our approximation (5) is very good when $\alpha \geq 1$. According to (2), the gamma distribution with α<1 is suitable only if the relative frequency of values near zero are very high. Such situations are rare in practice. Gamma distribution with $\alpha \geq 1$ fits in most practical situations. Therefore, our approximation is suitable for most practical applications.

3. Conventional Estimators for the Median of Gamma Distribution

Let ν be the median of gamma distribution with shape parameter α (>0) and scale parameter β (>0). The sample median and maximum likelihood estimator are two possible estimators for the median ν.

The sample median

The sample median of a sample of size *n* is calculated as follows:

$$\text{Sample median} = \begin{cases} \dfrac{(n+1)}{2} th\ ordered\ value & when\ n\ is\ odd \\ (\dfrac{n}{2} th\ ordered\ value + (\dfrac{n}{2}+1) st\ ordered\ value)/2 & when\ n\ is\ even \end{cases}$$

We shall denote this estimator by $\hat{\nu}_{sm}$.

The maximum likelihood estimator

Since it is not possible to write ν in terms of α and β explicitly, it is also not possible to obtain the maximum likelihood estimator of ν in a closed form. However, the maximum likelihood estimate of ν can be obtained using the invariance property of the maximum likelihood estimators (Mood A.M., et. al., 2001). This can be done by first deriving the maximum likelihood estimates $\hat{\alpha}_{mle}$ and $\hat{\beta}_{mle}$ of α and β respectively, and then finding $\hat{\nu}_{mle}$ that satisfies

$$\int_0^{\hat{\nu}_{mle}} f_X(x; \hat{\alpha}_{mle}, \hat{\beta}_{mle})dx = 0.5 . \qquad (6)$$

using the 'INVCDF' function in the statistical package Minitab or 'qgamma' function in the statistical package R. Anita S. et. al. (2002) have discussed the maximum likelihood estimation of α and β. For the convenience of the reader, we reproduce some of their results in this paper.

Let $x_1, x_2, ..., x_n$ be a random sample from a $G(\alpha, \beta)$ distribution. Then, maximum likelihood estimator $\hat{\beta}_{mle}$ of β is given by

$$\hat{\beta}_{mle} = \frac{\bar{x}}{\hat{\alpha}} . \qquad (7)$$

It is not possible to obtain $\hat{\alpha}_{mle}$ in a closed form. The authors have provided the following iterative procedure to obtain $\hat{\alpha}_{mle}$.

$$\hat{\alpha}_k = \hat{\alpha}_{k-1} - \frac{\log(\hat{\alpha}_{k-1}) - \Psi(\hat{\alpha}_{k-1}) - M}{1/\hat{\alpha}_{k-1} - \Psi'(\hat{\alpha}_{k-1})}, \ k=1,2,\ldots.. \qquad (8)$$

In equation (8),

$$M = \log(\bar{x}) - \frac{1}{n}\sum \log(x_i),$$

$$\Psi(\alpha) = \frac{d}{d\alpha}\left(\log \Gamma(\alpha)\right), \quad and$$

$$\Psi^{'}(\alpha) = \frac{d}{d\alpha}\left(\Psi(\alpha)\right).$$

$\Psi(\alpha)$ is the digamma function and $\Psi^{'}(\alpha)$ is the trigamma function. These functions are available in R statistical software. Authors have suggested several starting values for $\hat{\alpha}_0$ in the iterative procedure (8). We found that the moment estimator

$$\hat{\alpha}_{me} = \frac{(\bar{X})^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2} \tag{9}$$

of α (Wiens et. al.,2003) also works well as the initial value $\hat{\alpha}_0$.

As it can be seen from the above description, the derivation of the maximum likelihood estimate $\hat{\nu}_{mle}$ requires intensive computations. In the next section, we introduce a new estimator which requires fewer computations.

4. A New Estimator for the Median of Gamma Distribution

Based on our approximation (5), we propose the following new estimator for the median ν of a gamma distribution.

$$\hat{\nu}_{BE} = \frac{(3\hat{\alpha}_{me} - 0.8)}{(3\hat{\alpha}_{me} + 0.2)}\,\bar{x} \tag{10}$$

Here, $\hat{\alpha}_{me}$ is the moment estimate of α, given by (9)

5. Comparison of Estimators

Table 2 shows the root mean square errors of the three estimators $\hat{\nu}_{sm}$, $\hat{\nu}_{mle}$ and $\hat{\nu}_{BE}$ as a percentage of the actual median. We consider β =1 and three values for α. Results do not depend on the value of β. For each value of α, we consider four sample sizes (n). For each combination of α, and n, the root mean square errors were calculated based on 10000 Monte Carlo simulations.

Table 2: Root mean square errors of estimators as percentages of actual medians.

α	β	ν	n	$\frac{RMSE(\hat{\nu})}{\nu}*100$		
				$\hat{\nu}_{sm}$	$\hat{\nu}_{mle}$	$\hat{\nu}_{BE}$
1	1	0.69	5	68	54	57
			10	46	37	40
			20	32	25	29
			30	26	21	24
5	1	4.67	5	25	21	21
			10	17	15	15
			20	13	10	10
			30	10	8	8
10	1	9.67	5	17	14	14
			10	12	10	10
			20	9	7	7
			30	7	6	6

According to the values in Table 2,

- The sample median $\hat{\nu}_{sm}$ has the highest root mean square error.
- When $\alpha = 1$, the maximum likelihood estimator $\hat{\nu}_{mle}$ has the smallest root mean square error.
- When $\alpha > 1$, estimators $\hat{\nu}_{BE}$ and $\hat{\nu}_{mle}$ have the same root mean square error.

6. Conclusion

The sample median $\hat{\nu}_{sm}$ is the easiest estimate to calculate. Maximum likelihood estimator $\hat{\nu}_{mle}$ is the most difficult estimate to calculate. It requires intensive computations. Our estimator $\hat{\nu}_{BE} = \frac{(3\hat{\alpha}_{me} - 0.8)}{(3\hat{\alpha}_{me} + 0.2)}\bar{x}$ requires slightly more computations than that for the sample median and much less computations than that for the maximum likelihood estimate. Sample median has the highest root mean square error. Maximum likelihood estimator (mle) has the smallest root mean square error when α=1. The root mean square error of our estimator is slightly above that of the mle when α=1, but the same when α>1. Therefore, considering the required amount of computations and the root mean square error, our estimator can be considered as an 'optimum' estimator for the population median, when the gamma distribution with $\alpha \geq 1$ is a suitable model for the distribution of the variable of interest.

7. References

Anita Singh, Ashok K. Singh, and Ross J. Iaci. Estimation of the Exposure Point Concentration Term Using a Gamma Distribution, 2002. EPA Technology Support Center Issue, United States Environmental Protection Agency. Available at: http://www.hanford.gov/dqo/training/289cmb02.pdf

Mood, A.M., Graybill, F., Boes, D.C. Introduction to the theory of Statistics (2001). Tata McGraw Hill Publishing Company Limited, New Delhi.

Wiens, D.P., Cheng, J., Beaulieu, N.C. A class of method of moments estimators for the two-parameter gamma family. Pakistan Journal of Statistics, 2003. Vol 19(1). pp. 129-141.