ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 43(2023) 87-96

Local Information Based Parameter Estimation for Beta Distribution of First Kind

Ch.Yugandhar and V.V. Haragopal
[Received on June, 2022. Accepted on April, 2023]

ABSTRACT

The process of parameter estimation in order to Characterise a population using Method of moments and MLE is well known and popular. The purpose of this article is a little different in estimating the parameters for Beta distribution of first kind by the partial information available, and when the partial information is available, how should the parameter be estimated? If estimated, how far can these parameters be considered good enough when compared with the estimators obtained by using the full sample information. In this present study we explored the parameter estimation by "Local Frequency Ratio Method" to estimate the parameters and found that this method estimates the parameters effectively with less information as compared with standard estimation procedure.

1. Introduction

In the study and use of data science problems, it is always important to give the best possible description of the data and its parameter estimations by various methods being looked at. Recent research has shown how the statistical distributions can be used to model data in applied sciences especially in medical sciences. Statisticians often explore new statistical methodologies to suit the existing distributions and data sets in diverse domains. Statistical models/methods are very useful in describing and predicting a real phenomena. Many distributions have been widely used for data modeling in several domains during the last decades. Recent developments focus on defining new families that extend well known estimation procedures and at the same time provide a great flexibility in various estimations in practice. These procedures are quite helpful

: Ch.Yugandhar
Email: yugandhar0203@gmail.com

Extended authors information available after reference list of the article.

and better understanding in many fields of virus spreads, in particular Covid-19 etc.

By Hogg and Tanis (2001), estimation is defined as a process of assigning numerical values to the parameter that is to be estimated based on a sample observations following a specified distribution. The function of the sample value used for this purpose is a statistic and is considered as a specified function of the parameter or taken as the parameter value of the distribution. This statistic which is being used so far is called an estimator of the parameter and the particular value obtained from the data using this estimator is called an estimate. Estimators themselves are random variables having their own probability distribution.

Estimation of parameters in general are based on the complete information of sample under study. However, it is also possible and often necessary to construct estimators based on partial information available from samples i.e. by information obtained only on sample values which falls into two or few of their lines or bins in a frequency distribution ignoring the values falling into other regions in the frequency distribution. Estimators are based not on global but on local information from the sample.

This approach is of course not entirely new. Representatively, dealing with the problem of estimating the parameters in situations where sample observations are censored or truncated can obviously be claimed to belong to this category. But any detailed study of such estimation procedure and the properties of such estimators do not seem to have been reported so far. The present problem is an effort in this direction.

Our investigation aims at answering the following prominent questions.

1) Using only local information from different localities (locals) in the sample set, how good an estimator of the parameter can one hope to obtain?

2) How do these estimators compare with the usual, full global sample based estimators?

3) Particularly, when only partial data is considered how the local information estimator compares with the global estimators that takes into account the entire sample.

This paper is Organised as follows. In section 2 we explained the Beta distribution of first kind along with various methods of estimation procedures with corresponding illustrations viz, Method of moments and newly introduced frequency ratio methods.

Section 3 devotes to the explanation of frequency ratio method of estimation.Section4 illustrates the computational evidence for different sample sizes and different parameters.

2. Beta Distribution of First Kind

The Beta distribution of first kind is defined by the following pdf.

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, 0 < x < 1$$

Where a>0 and b>0 both are shape parameters.

A few well known properties are:

$$E(X) = \frac{a}{a+b}; \quad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Parameter Estimation

We are interested in estimating the parameters of the Beta distribution from which the sample comes. A few estimation methods are outlined below.

Method of Moments

Under this method, we equate the sample mean and variance with the distribution's theoretical expected value and variance. We obtain two equations in two unknowns:

$$\bar{x} = \frac{a}{a+b} \quad \text{and} \quad S^2 = \frac{ab}{(a+b)^2(a+b+1)}$$

Solving these equations yields the following estimators:

$$a = \bar{x}\left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right) \quad \text{and} \quad b = (1-\bar{x})\left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right)$$

For Example:

We generate 50 random samples, each of size 1000 from Beta distribution by taking (a=2, b=3) using MATLAB function. For each sample we estimate parameters a and b by using above procedure. The Mean, Standard Error, $\sqrt{\beta_1}$, β_2 of these 50 estimates were computed. The Estimated bias was calculated as the mean minus the true value of the parameter. The Mean Squared Error (MSE) was calculated as the bias squared plus the variance. The results are shown in the following table.

Table 1

	Method of Moments	
	a	**b**
Mean	2.0216	3.0248
SE	0.1081	0.1562
$\sqrt{\beta_1}$	0.2720	0.7451
β_2	2.1383	2.8472
Bias	0.0216	0.0248
MSE	0.0005	0.0006

3. Frequency Ratio Method of Estimation

Let $y_1, y_2 \ldots\ldots y_n$ be a random sample from a distribution. From this sample a frequency distribution is constructed with an appropriate bin width 'h'. The midpoint of these bins are denoted by x_i, i = 1,2…… k (number of bins). The corresponding frequencies are denoted by f_i , i = 1,2….k. Thus $\frac{f_i \times h}{n}$ is an estimate of the probability of y falling in the corresponding bin 'i' and is an estimate of the probability lying in the interval. Thus $f(x_i) \times h$ can be estimated by $\frac{f_i}{n}$, using the ratios of $f(x_i)$'s and equating them with corresponding observed frequency ratios gives a way of estimating the parameters similar to the moments method of estimation.

Let f_1, f_2 and f_3 are the frequency densities at the points x_1, x_2 and x_3 given by

$$f_1 = \frac{x_1^{a-1}(1-x_1)^{b-1}}{B(a,b)} \quad ; \quad f_2 = \frac{x_2^{a-1}(1-x_2)^{b-1}}{B(a,b)} \quad ; \; f_3 = \frac{x_3^{a-1}(1-x_3)^{b-1}}{B(a,b)}$$

The ratio of the frequencies f_1 and f_2 is

$$\frac{f_1}{f_2} = \frac{x_1^{a-1}(1-x_1)^{b-1}}{x_2^{a-1}(1-x_2)^{b-1}} = \left(\frac{x_1}{x_2}\right)^{a-1}\left(\frac{1-x_1}{1-x_2}\right)^{b-1}$$

Taking logarithms on both sides,

$$\log\left(\frac{f_1}{f_2}\right) = (a-1)\log\left(\frac{x_1}{x_2}\right) + (b-1)\log\left(\frac{1-x_1}{1-x_2}\right)$$

For notational convenience, let $lf_{12} = \log\left(\frac{f_1}{f_2}\right), lx_{12} = log\left(\frac{x_1}{x_2}\right), llx_{12} = log\left(\frac{1-x_1}{1-x_2}\right)$

$$lf_{12} = (a-1)lx_{12} + (b-1)llx_{12} \tag{3.1}$$

Similarly, the ratio of the frequencies f_2 and f_3 is

$$\frac{f_2}{f_3} = \frac{x_2^{a-1}(1-x_2)^{b-1}}{x_3^{a-1}(1-x_3)^{b-1}} = \left(\frac{x_2}{x_3}\right)^{a-1}\left(\frac{1-x_2}{1-x_3}\right)^{b-1}$$

Taking logarithms on both sides,

$$\log\left(\frac{f_2}{f_3}\right) = (a-1)\log\left(\frac{x_2}{x_3}\right) + (b-1)\log\left(\frac{1-x_2}{1-x_3}\right)$$

$$lf_{23} = (a-1)lx_{23} + (b-1)llx_{23} \tag{3.2}$$

Where, $lf_{23} = \log\left(\frac{f_2}{f_3}\right), lx_{23} = log\left(\frac{x_2}{x_3}\right), llx_{23} = log\left(\frac{1-x_2}{1-x_3}\right)$

Solving (1) and (2), we get estimates of a and b as

$$\hat{a} = \frac{1}{lx_{12}}\left[lf_{12} - \left(\frac{lf_{12}lx_{23} - lf_{23}lx_{12}}{llx_{12}lx_{23} - llx_{23}lx_{12}}\right)llx_{12}\right] + 1 \tag{3.3}$$

and

$$\hat{b} = \frac{lf_{12} lx_{23} - lf_{23} lx_{12}}{llx_{12} lx_{23} - llx_{23} lx_{12}} + 1 \quad (3.4)$$

Illustration:

For each of the sample generated above, we construct a frequency Distribution given below.

Table 2

x (Mid-value)	0.0484	0.1452	0.2419	0.3387	0.4355	0.5322	0.6290	0.7258	0.8225	0.9193
f	45	116	151	157	168	155	107	63	24	14

From the above table, f_1, f_2 and f_3 are 168,157 and 155 (first three maximum frequencies) and the corresponding midpoints are x_1, x_2 and x_3 .
Using these frequencies and mid values in the above formulae (3.3) and (3.4), we get estimates of a and b as $\hat{a} = 2.366$ *and* $\hat{b} = 3.17$
The above procedure is repeated for 50 samples and we get 50 estimates. The mean, Standard error , $\sqrt{\beta_1}$, β_2 of these 50 estimates were computed. The estimated bias was calculated as the mean minus the true value of the parameter. The Mean Squared Error (MSE) was calculated as the bias squared plus the variance.

Table 3

	Frequency ratio method	
	a	**b**
Mean	2.0388	3.0659
SE	1.1239	2.2443
$\sqrt{\beta_1}$	0.3788	0.5050
β_2	2.5355	2.7952
Bias	0.0388	0.0659
MSE	0.0028	0.0094

From the above tables, we notice that the actual values of (a, b) and the mean estimated values of (a, b) under the frequency ratio method and Method of moments are almost same. Therefore, it can be taken as a good estimator. Similar procedure is followed for different sample sizes and different values of (a, b) and the results are tabulated in the following tables.

4. Comparison of Method of Moments and Frequency Ratio Method for Different Sample Sizes and Different Parameters

Table 4 : Simulation statistics for beta (1,5).

(1,5)	ns=50				ns=100			
	Method of moments		**Frequency Ratio method**		**Method of moments**		**Frequency Ratio method**	
	a	b	a	b	a	b	a	b
Mean	0.9939	4.9999	0.9981	5.0960	1.0066	5.0418	1.0087	5.0742
SE	0.0547	0.2974	0.2353	2.4049	0.0452	0.2689	0.2457	2.4929
$\sqrt{\beta_1}$	0.1738	0.0788	0.0081	0.0049	0.0411	0.0877	0.109	0.0216
β_2	2.5259	2.3274	2.3978	2.2512	2.3523	2.4176	2.7163	2.9372
Bias	-0.006	0	-0.0019	0.0960	0.0066	0.0418	0.0087	0.0742
MSE	0	0	0.0001	0.0150	0	0.0018	0.0001	0.0117

Table 5 : Simulation statistics for beta (3, 1).

(3,1)	ns=50				ns=100			
	Method of moments		**Frequency Ratio method**		**Method of moments**		**Frequency Ratio method**	
	a	b	a	b	a	b	a	b
Mean	3.0151	1.0061	3.1538	1.0209	3.0145	1.0095	3.0364	1.0060
SE	0.1493	0.0451	1.6009	0.2233	0.1766	0.0524	1.6409	0.2286
$\sqrt{\beta_1}$	0.0027	-0.535	0.2121	0.2766	0.9717	0.7116	-0.290	-0.338
β_2	2.4480	3.508	2.9718	2.722	5.0290	3.7251	2.5931	2.6515
Bias	0.0151	0.0061	0.1538	0.0209	0.0145	0.0095	0.0364	0.0060
MSE	0.0003	0	0.0262	0.0005	0.0002	0.0001	0.0040	0.0001

Table 6 : Simulation statistics for beta (4,3).

(4,3)	ns=50				ns=100			
	Method of moments		**Frequency Ratio method**		**Method of moments**		**Frequency Ratio method**	
	a	b	a	b	a	b	a	b
Mean	4.0283	3.0102	3.999	2.9986	3.9836	2.9692	4.0362	3.0161
SE	0.1804	0.1381	2.6090	1.7411	0.1746	0.1306	2.1844	1.3516
$\sqrt{\beta_1}$	0.2512	0.3099	-0.0436	0.0270	0.9974	0.8995	0.3054	0.4703
β_2	2.8736	3.266	2.6150	2.7492	4.9974	4.7166	2.9499	2.8962
Bias	0.0283	0.0102	-0.0008	-0.0014	-0.016	-0.030	0.0362	0.0161
MSE	0.0008	0.0001	0.0068	0.0030	0.0003	0.0010	0.0061	0.0021

Table 7 : Simulation statistics for beta (3 ,2).

(3,2)	ns=50				ns=100			
	Method of moments		**Frequency Ratio method**		**Method of moments**		**Frequency Ratio method**	
	a	b	a	b	a	b	a	b
Mean	2.9663	1.9860	3.0435	2.0235	3.0122	1.9965	2.9463	1.9811
SE	0.1306	0.0923	2.3736	1.2595	0.1374	0.0949	2.6093	1.3951
$\sqrt{\beta_1}$	0.4621	0.0964	0.0938	0.0434	0.2088	0.0861	-0.399	-0.330
β_2	2.3359	3.612	2.4233	2.6296	2.1267	2.010	3.0701	3.1817
Bias	-0.033	-0.014	0.0435	0.0235	0.0122	-0.003	-0.053	-0.018
MSE	0.0012	0.0002	0.0075	0.0021	0.0002	0	0.0097	0.0023

Table 8 : Simulation statistics for beta (5, 1).

(5,1)	ns=50				ns=100			
	Method of moments		**Frequency Ratio method**		**Method of moments**		**Frequency Ratio method**	
	a	b	a	b	a	b	a	b
Mean	5.0836	1.0096	4.9818	1.0011	4.9791	0.9973	5.1635	1.0265
SE	0.0102	0.0018	0.0526	0.0074	0.0078	0.0014	0.0510	0.0070
$\sqrt{\beta_1}$	0.1310	0.0093	0.4945	0.2647	0.1757	0.1414	-0.149	-0.260
β_2	3.122	3.688	3.7453	3.4875	2.9967	2.7970	3.412	3.4515
Bias	0.0839	0.0096	-0.0182	0.0011	-0.021	-0.002	0.1635	0.0265
MSE	0.0071	0.0001	0.0031	0.0001	0.0005	0	0.0293	0.0008

5. Conclusions

For both Estimation methods, the statistical distributions are summarized by its mean, standard error, $\sqrt{\beta_1}$, β_2, Bias and Mean Square Error computed from the simulated data. Thus, from the empirical study of the type of distribution, the estimates computed using the various estimation procedures including the one based on full information is reported.

We observed from the above tables that the mean estimated values based on Method of moments with full data and the Local frequency ratio method based on partial information is nearly equal to the true value of the parameter. However, the standard errors of Local Frequency Ratio method are slightly more than that of the estimator based on full information sample. But in the particular case where out liars may affect the estimation procedure based on global information, this aspect is insignificant. Thus, when full information is available, the local information-based estimators are effectively as good as the corresponding Method of moments with full information. With sample sizes 50 and 100 itself the accuracy is being tallied, and if we consider a larger sample size say 10000 the results may be more accurate. This new approach of estimation can be applied in any simulation, medical, Big-data analytics approaches and any data science problems for a small or bigger sample sizes resulting an optimum time with accurate estimation.

References

Aris Spanos (1999): Probability Theory and Statistical Inference: *United Kingdom at the University Press, Cambridge*.

Ch.Yugandhar , V.V. Haragopal, S.N.N. Pandit (2011): Local Information Based Parameter Estimation for Exponential distribution, *ANU Journal of Physical Sciences*,**3(1&2)**.

Ch. Yugandhar and V.V. Haragopal (2014)**:** Estimating the parameters of Pareto Distribution through Local Frequency Ratio method, *Aligarh Journal of Statistics* **34**, 118-128.

Ch. Yugandhar and V.V. Haragopal (2015): Estimating the parameters of Shifted Exponential distribution through Local Frequency Ratio method, *International journal of Statistics and Systems (IJSS),* **10(1), 159-163. (ISSN 0973-2675).**

Hossan E., Abdulrahman A.T. and Gemeay A.M *et al*., (2022): A Noval extension of gumbel distribution: Statistical Inference with Covid-19 applications, *Alexandria engineering Journal*, **61(11), 8823- 8842**

Hogg and Tanis (2001): Probability and Statistical Inference *Sixth addition-Prentice-Hall publications, Newjersy*.

Manthil K.Thamar and Raoudha Zine (2021): Comparision of Five methods to estimate the parameters for the 3-parameter Lindley Distribution with application to Life data- *Computational and Mathematical methods in Medicine Journal, Volume* **2021**, *Article ID 2689000, 14 pages.*

Rao, C.R (1973): Linear Statistical Inference and its Applications-*John Wiley and sons, New York.*

Robert V. Hogg and Allent T. Craig (2002): Introduction to Mathematical Statistics-*Pearson Publications, Singapore*.

Rudra Pratap (2004): Getting Started with MATLAB-*Oxford University press, New York.*

Authors and Affiliations

Ch.Yugandhar[2] and V.V.Haragopal[1]

V.V.Haragopal
Email: haragopalvajjha@gmail.com

[1]Retd.Professor, Department of Statistics, Osmania University, Hyderabad -500 007 (Telangana), India
[2] Department of Statistics, St. Francis College for Women, Hyderabad – 500 016 (Telangana), India