World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Investigating the parameters of the beta distribution

Ali H *, Akanihu CN and Felix J

Department of Mathematics, University of Jos, Nigeria.

World Journal of Advanced Research and Reviews, 2023, 19(01), 815–830

Publication history: Received on 04 June 2023; revised on 12 July 2023; accepted on 15 July 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.19.1.1409

Abstract

The Beta distribution takes on many different shapes and may be described by two shape parameters, α and β, that can be difficult to estimate. Maximum likelihood and method of moments estimation are possible, though method of moments is much more straightforward and easier to compute by hand. In this paper, general information about the beta distribution were given. Usage areas of Beta distribution are also specified. The method of moment estimators and the system of linear equations of the maximum likelihood estimators are presented here. At the end of the investigation, it was discovered that the method of moment estimators is most preferable for small sample size (samples less than 50) whereas the Maximum Likelihood method is good when the sample size is large, this is because the estimates provided by this method are more consistent and asymptotically efficient; that is, they converge in probability to the parameter they are estimating and achieve the lower bound on variance.

Keywords: Estimator; Maximum Likelihood; Moments; Monograph; Mean Square Error

1. Introduction

The beta distribution is a continuous probability distribution that models random variables with values falling inside a finite interval. Analysts commonly use it to model the time to complete a task, the distribution of order statistics, and the prior distribution for binomial proportions in Bayesian analysis.

The standard beta distribution uses the interval [0,1]. This range is ideal for modeling probabilities, particularly for experiments with only two outcomes. However, other intervals are possible.

Consequently, numerous studies on the various generalized beta distribution forms were conducted. Chotikapanich, *et al.* (2007) studied a three-parameter Beta-2 model which is a generalized version of the normalized beta distribution; also, Ng *et al.* (2019) introduced the generalized beta model.

A statistician's chore, in distinct, often amounts to identifying an eminent probability distribution which sufficiently describes the variations found in experimental data. Ill-advisedly, no "best" scheme exists for executing this identification. Parameter estimation is a branch of statistics that involves using sample data to estimate the parameters of a probability distribution (be it discrete or continuous) this study considers Beta distribution.

Parameter estimation is la-di-da by the kind of forfeits placed on different kinds of errors in the estimate and even when the suitable penalties are agreed on, the "best" estimate may be grim to find. Hitherto for many families of distributions, good estimators exist for illustration, maximum likelihood estimators, method of moment estimators, least squares estimators among others.

* Corresponding author: Ali H https://orcid.org/0000-0001-5517-8226; Email: auhills1@gmail.com

The Beta distribution is a continuous probability distribution characterized by two shape parameters, α and β, and is used to model singularities that are constrained to be between 0 and 1, such as probabilities, proportions, and percentages. The Beta distribution is also used as the conjugate prior distribution for binomial probabilities in Bayesian statistics.

Gelman, *et al* (2004). With the extensive applicability of the Beta distribution, it is important to estimate, with approximately degree of precision, the parameters of the observed data. The study presents a simulation study to reconnoiter the efficacy of the two distinct parameter estimation methods for determining the parameters of the distribution.

If a random variable ***X*** has a Beta distribution with shape parameters α and β then, it has a probability density function (pdf) given as;

$$f_x(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 \leq x \leq 1, \alpha > 0, \beta > 0 \text{ ……………………………… (1)}$$

$$\text{Where } \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{1}{\mathrm{B}(\alpha,\beta)}$$

And $\mathrm{B}(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ is the integral definition of a beta function $\Gamma(\alpha)$ *and* $\Gamma(\beta)$ are gamma functions, α and β are the two positive shape parameters which control the shape of the beta distribution.

David and Edwards's dissertation on the history of statistics cites the first modern treatment of the beta distribution, in 1911, using the beta designation that has become standard, due to Corrado Gini, an Italian statistician, demographer, and sociologist, who developed the Gini coefficient.

Vijay (1999), detailed, in reliability safety analysis of civil engineering systems, “we encounter parameters which are generally bounded and skewed random quantities.” Exemplifying these parameters are factors of safety or safety indexes, variables representing strength of materials, intensity of loads, etc.

Oboni and Bourdeau C, (1985) abridged use of the beta distribution and investigated its sensitivity to the bound locations.

Harr C (1977) established the ability of the beta Cor Pearson type 1 distribution to approximate most of the geotechnical parameters.

Romesburg (1976) studied that formulation of the problem in terms of smallest order statistics would permit the use of the method of maximum likelihood estimation MLE to estimate the parameters of the beta distribution with little more effort than MOM. In multivariate cases, however, MOM would be the only practical method for parameter estimation.

Fielitz and Myers (1975) argued for the method of moments (MOM) to estimate the parameters of the beta distribution for ease of computation. And concluded that the method of moments is the easiest technique for estimating the parameters of the beta distribution.

Bayes, in a posthumous paper published in 1763 by Richard Price, obtained a beta distribution as the density of the probability of success in Bernoulli trials, but the paper does not analyze any of the moments of the beta distribution or discuss any of its properties.

Eugene, *et al* (2002), introduced the beta-normal distribution, based on a composition of the classical beta distribution and the normal distribution. Its importance is more than just generalize the normal distribution. The beta-normal distribution generalizes the normal distribution and has flexible shapes, giving it greater applicability. Since then, many authors generalized other distributions similar to the beta-normal distribution.

2. Material and methods

The beta-normal distribution is obtained as follows:

$$\text{Let, } G(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \ldots\ldots\ldots\ldots (2)$$

be the cumulative density function of normal distribution with parameters μ and σ, and

$g(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$ be the probability density function of the normal distribution.

Then, the density function of beta-normal distribution is given by

$$f(x;\alpha,\beta,\mu,\sigma) = \frac{\sigma^{-1}\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}\left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\beta-1}}{B(\alpha,\beta)} . \phi\left(\frac{x-\mu}{\sigma}\right) \ldots\ldots (3)$$

Where α>0, β>0, σ>0, $\mu \in \boldsymbol{R}$, $x \in \boldsymbol{R}$

The parameters α and β are the shape parameters characterizing the skewness, kurtosis and bimodality of the beta normal distribution. The parameters μ and σ have the same role as in normal distribution where, μ is a location parameter and σ is a scale parameter that stretches out or shrinks the distribution.

Cassela, *et al* (2002) attempted deriving the estimators of α and β using the maximum likelihood method and the method of moments estimates. The maximum likelihood is by far the most popular technique for deriving estimators. Recall that if $X_1, \ldots, X_n$ are i.i.d samples from a population with pdf $f(x|\theta_1, \ldots, \theta_k)$, the likelihood is defined by

$$L(\theta|x) = L(\theta_1, \ldots, \theta_k|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k$$

Intuitively, the MLE is a reasonable choice of estimator. The MLE is the parameter point for which the observed data is most likely. If the likelihood function is differentiable (in θ_i), possible candidates for the MLE are the values of $(\theta_1, \ldots, \theta_k$ that solve

$$\frac{\partial}{\partial\theta_i} L(\theta|x) = 0, i = 1, \ldots, k$$

The solutions to the equation above are only possible for the MLE since the first derivative being 0 is only a necessary condition for a maximum, not a sufficient condition. The method of moments is perhaps the oldest method of finding point estimators, dating back at least to Karl Pearson in the late 1800s. It has the virtue of being quite simple to use and almost always yields some sort of adequate estimate. In many cases, unfortunately, this method is yields estimators that may be improved upon. However, it is a good place to start when other methods prove difficult to be solved.

Let $X_1, \ldots, X_n$ be a sample from a population with pdf $f(x|\theta_1, \ldots, \theta_k)$. Method of moments estimators are found by equating the first *k* sample moments to the corresponding *k* population moments, and solving the resulting the system simultaneously.

Kumachov, et al (2006), studied the contoured curves of (γ_1, γ_2). The study showed that the curves exhibit turning points when a = b, making $\gamma_1 = 0$ and $\gamma_2 = \frac{3(2a+1)}{2b+3}$

When a=b=0, the turning point is at $\gamma_2 = 1$ and when a=b=1 the turning point is at $\gamma_2 = 1.8$. Karian also showed that the points of (γ_1, γ_2) are bounded by $1 + {\gamma_1}^2 < \gamma_2 < 3 + {\gamma_1}^2$.

Some approximations of γ_1 and γ_2 are;

$${\gamma_1}^2 \approx \frac{1}{a} + \frac{1}{b} - \frac{4}{a+b} \text{ for a and b large, } \ldots\ldots (4)$$

$$\gamma_1{}^2 \approx 4\left(\frac{1}{a}+\frac{1}{b}\right)-\frac{16}{a+b}, \text{......} (5)$$

$$\gamma_2 \approx 3+\frac{2}{a}+\frac{2}{b}-\frac{6}{a+b}. \text{ And ......} (6)$$

$$\gamma_2 = 3+6\left(\frac{1}{b}+\frac{1}{a}\right)-\frac{30}{a+b} \text{......} (7)$$

Kong, *et al* (2007), introduced the beta-gamma distribution, based on a composition of the classical beta distribution and the gamma distribution. The study derived some properties of the limit of the density function and of the hazard function. The study also presented an expression for the moments when the shape parameter α is an integer and made an application of the beta-gamma distribution.

The beta-gamma distribution is obtained as follows:

$$\text{Let } G(x) = \frac{\Gamma_{\frac{x}{\lambda}}(\rho)}{\Gamma(\rho)} \text{......} (8)$$

be the cdf of gamma distribution where;

$$\Gamma_x(\rho) = \int_0^x y^{\rho-1} e^{-y} dy \text{......} (9)$$

is the incomplete gamma function and

$$g(x) = \left(\frac{x}{\lambda}\right)^{\rho-1} e^{-\frac{x}{\lambda}} \text{......} (10)$$

is the pdf of the gamma function.

Then, the density function of the beta-gamma distribution is given by

$$f(x;\alpha,\beta,\rho,) = \frac{x^{\rho-1} e^{\frac{x}{\lambda}} \Gamma_{\frac{x}{\lambda}}(\rho)^{\alpha-1}\left[1-\frac{\Gamma_{\frac{x}{\lambda}}(\rho)}{\Gamma(\rho)}\right]^{\beta-1}}{B(\alpha,\beta)\Gamma(\rho)^{\alpha}\lambda^{\rho}} \text{......} (11)$$

Where a,,ρ,λ,x > 0, is the beta-gamma distribution introduced by Kong (2007).

Johnson, *et al* (1996) in their comprehensive and very informative monograph on statistical sciences credited Corrado Gini as "an early Bayesian" who dealt with the problem of eliciting the parameters of an initial Beta distribution, by singling out techniques which anticipated the advent of the empirical Bayes approach.

Ahmed (2011) provided a characterization of the beta distribution in terms of the failure rate functions. Namely, if X is a non-negative continuous random variable with cdf (F) and pdf (f) and mean m, then X has the beta distribution if and only if

$$E(X|X \geq t) = m + \frac{m}{\alpha} t(1-t)\lambda(t) \text{......} (12)$$

$$m = \frac{\alpha}{\alpha+\beta}$$

For the generalized beta distribution with parameter (a, b, c, d), these conditions become

$$E(X|X \geq t) = m + \frac{m}{bc+ab}(t-c)(d-t)\lambda(t) \text{......} (14)$$

$$m = \frac{bc+ab}{a+b}$$

For the power function distribution (the special case of the beta distribution for b=1), the condition reduces to

$$E(X|X \geq t) = m + \frac{m}{a}t(1-t)\lambda(t) \quad \text{.................................} (15)$$

$$m = \frac{a}{a+1}.$$

Chris Piech (2016) in his lecture on "Beta Distribution" noted that one can set ***X ∼Beta (a, b)*** as a prior to reflect how biased one thinks a coin is apriori to flipping it. This is a subjective judgement that represent ***a+b−2*** "imaginary" trials with ***a−1*** heads and ***b−1*** tails. If one then observes ***n + m*** real trials with ***n*** heads one can update one's belief. The new belief would be, ***X| (n heads in n+ m trials) ∼Beta (a + n, b + m)***. Using the prior ***Beta(1,1) =Uni(0,1)*** is the same as saying, one hasn't seen any "imaginary" trials, so apriori nothing is known about the coin. This form of thinking about probabilities is representative of the "Bayesian" field of thought where computer scientists explicitly represent probabilities as distributions (with prior beliefs). That school of thought is separate from the "Frequentist" school which tries to calculate probabilities as single numbers evaluated by the ratio of successes to experiments.

Mark Carpenter and Satya N. Mishra, (2018), conducted a study on the generalized beta distribution and the standard beta distribution which is one of the few well-studied distributions with [0,1] support. Oftentimes, the flexibility of the standard beta is desired as a model but the [0, 1] support represents an unreasonable restriction. Accordingly, the model is transformed by location-scale and/or ratio transformations to expand the support of the distribution and/or add flexibility. The several resulting classes of distributions are referred to in an umbrella fashion as the generalized beta distributions, which is the parent distribution of the standard beta. The generalized beta is equivalent to the Pearson Type I distribution. The extreme flexibility of this class makes it very useful in fitting distributions to data sets similar to the way generalized lambda distributions are used. The paper, discussed the properties and draw connections between the various forms of the generalized beta distribution. Estimators such as maximum likelihood, method of moments, and others are developed, compared, and applied.

The study adapted the simple random sampling technique for the choice of the samples of the random samples to be used in estimating the parameters of the Beta distribution.

2.1. Maximum Likelihood Estimators (MLE).

A well-known method of estimating parameters is the maximum likelihood approach and it is, by far, the most popular technique for deriving estimators. Maximum Likelihood Estimation (MLE) is a method of estimating the parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The logic of maximum likelihood is both intuitive and flexible, and as such the method has become a dominant means of statistical inference.

The likelihood function for an independent identically distributed random samples, say, $X_1, X_2, \ldots, X_n$ from a population with probability density function $f(x|\theta_1, \theta_2, \ldots, \theta_k)$ is defined as

$$L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x|\theta_1, \ldots, \theta_k) \quad \text{.................................} (16)$$

The maximum likelihood estimator (MLE) is the parameter value for which the observed sample is most likely. Possible MLEs are solutions to

$$\frac{\partial}{\partial \theta_i} L(\theta_i | x) = 0, i = 1, \ldots, k \text{.................................} (17)$$

It is easier to work with the log likelihood function, $logL(\theta|x)$, as derivatives of sums are more easily solved than derivatives of products. MLEs are desirable estimators because they are more consistent and asymptotically efficient; that is, they converge in probability to the parameter they are estimating and achieve the lower bound on variance, Berger (2002).

The likelihood function for the beta distribution is

$$L(\alpha, \beta) = \prod_{i=1}^{n} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \right] = \left[\Gamma(\alpha+\beta)^n \Gamma(\alpha)^{-n} \Gamma(\beta)^{-n} \sum_{i=1}^{n} {x_i}^{\alpha-1} \sum_{i=1}^{n} (1-x_i)^{\beta-1} \right]$$
$$\text{.................................} (18)$$

Taking the log gives

$$logL(\alpha,\beta) = nlog\Gamma(\alpha+\beta) - nlog\Gamma(\alpha) - nlog\Gamma(\beta) + (\alpha-1)\sum_{i=1}^{n} log\, x_i + (\beta-1)\sum_{i=1}^{n} log\,(1-x_i) \quad \text{.......... (19)}$$

To solve for the MLEs of α and β, we take the partial derivatives of the log likelihood function with respect to α and β and set the partial derivatives equal to zero and solve for $\hat{\alpha}$ and $\hat{\beta}$.

$$\frac{\partial}{\partial\alpha} logL(\alpha,\beta) = n\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n} log\, x_i = 0 \quad \text{.......... (20)}$$

$$\frac{\partial}{\partial\beta} logL(\alpha,\beta) = n\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - n\frac{\Gamma'(\beta)}{\Gamma(\beta)} + \sum_{i=1}^{n} log\,(1-x_i) = 0 \quad \text{.......... (21)}$$

But, $\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} = \psi(\alpha+\beta), \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \psi(\alpha)\ and\ \frac{\Gamma'(\beta)}{\Gamma(\beta)} = \psi(\beta)$ .... (22)

therefore equations (I) and (II) becomes

$$n\psi(\alpha+\beta) - n\psi(\alpha) + \sum_{i=1}^{n} log\, x_i = 0 \quad \text{.......... (23)}$$

$$n\psi(\alpha+\beta) - n\psi(\beta) + \sum_{i=1}^{n} log\,(1-x_i) = 0 \quad \text{.......... (24)}$$

There is no closed-form solution to this system of linear equations, so the values for $\hat{\alpha}$ and $\hat{\beta}$ iteratively using Newton-Raphson method in a package called EnvStats in R, a tangent method for root finding. In our case we will estimate

$\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ iteratively.

$$\hat{\theta}_{i+1} = \hat{\theta} - G^{-1}g \quad \text{.......... (25)}$$

g is the vector of normal equations for which we want

$$g = [g_1\ g_2] \quad \text{.......... (26)}$$

With $g_1 = \psi(\alpha+\beta) - \psi(\alpha) + \frac{\sum_{i=1}^{n} logx_i}{n}$ and .... (27)

$$g_2 = \psi(\alpha+\beta) - \psi(\beta) + \frac{\sum_{i=1}^{n} log(1-x_i)}{n} \quad \text{.......... (28)}$$

This implies that

$$g_1 = \psi(\alpha) - \psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} log\, x_i \quad \text{.......... (29)}$$

$$g_2 = \psi(\beta) - \psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} log\,(1-x_i) \quad \text{.......... (30)}$$

And G is a matrix of second order derivatives, that is,

$$G = \begin{bmatrix} \frac{\partial g_1}{\partial\alpha} & \frac{\partial g_1}{\partial\beta} \\ \frac{\partial g_2}{\partial\alpha} & \frac{\partial g_2}{\partial\beta} \end{bmatrix}$$

where,

$$\frac{\partial g_1}{\partial\alpha} = \psi'(\alpha) - \psi'(\alpha+\beta) \quad \text{.......... (31)}$$

$$\frac{\partial g_1}{\partial\beta} = -\psi'(\alpha+\beta) \quad \text{.......... (32)}$$

$$\frac{\partial g_2}{\partial \alpha} = -\psi'(\alpha + \beta) \quad \ldots\ldots (33)$$

$$\frac{\partial g_2}{\partial \beta} = \psi'(\beta) - \psi'(\alpha + \beta) \quad \ldots\ldots (34)$$

And $\psi'(.)$ are the trigamma functions defined as

$$\psi'(.) = \frac{\Gamma(.)''}{\Gamma(.)} - \frac{\Gamma'(.)^2}{\Gamma(.)^2} \quad \ldots\ldots (35)$$

The solution to these equations is very tedious and may seem difficult to compute so the software packages were used to estimate the MLEs, Owen (2008).

2.2. Method of Moment Estimators (MME)

The method of moments, is perhaps, the oldest method of finding point estimators, dating back to Karl Pearson in the 1800s. It has the virtue of being quite simple to use and always yields some sort of estimates. In many cases, unfortunately, this method yields estimates that may be improved upon. However, it is a good start when other methods prove intractable, Berger (2002). The k-th moment of a beta distributed random X is

$$\mu_x(k) = E(X^k) = \frac{B(\alpha+k,\beta)}{B(\alpha,\beta)} = \prod_{n=0}^{k-1} \frac{\alpha+n}{\alpha+\beta+n} \quad \ldots\ldots (36)$$

Proof;

By the definition of moments, we have

$$\mu_x(k) = E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x)dx \quad \ldots\ldots (37)$$

$$= \int_0^1 x^k \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \quad \ldots\ldots (38)$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 x^k\, x^{\alpha-1}(1-x)^{\beta-1}dx \quad \ldots\ldots (39)$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{(\alpha+k)-1}\,(1-x)^{\beta-1}dx \quad \ldots\ldots (40)$$

By the integral representation of the beta function, we have

$$= \frac{1}{B(\alpha,\beta)} * B(\alpha+k,\beta) \quad \ldots\ldots (41)$$

By the definition of the beta function, we have

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} * \frac{\Gamma(\alpha+k)\Gamma(\beta)}{\Gamma(\alpha+\beta+k)} \quad \ldots\ldots (42)$$

$$= \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)\Gamma(\beta)}} * \frac{\alpha\cdot(\alpha+1)\cdot\ldots\cdot(\alpha+k-1)\cancel{\Gamma(\alpha)\Gamma(\beta)}}{(\alpha+\beta)\cdot(\alpha+\beta+1)\cdot\ldots\cdot(\alpha+\beta+k-1)\cancel{\Gamma(\alpha+\beta)}} \quad \ldots\ldots (43)$$

$$= \frac{\alpha\cdot(\alpha+1)\cdot\ldots\cdot(\alpha+k-1)}{(\alpha+\beta)\cdot(\alpha+\beta+1)\cdot\ldots\cdot(\alpha+\beta+k-1)} \quad \ldots\ldots (44)$$

Therefore,

$$\mu_x(k) = \prod_{n=0}^{k-1} \frac{\alpha+n}{\alpha+\beta+n} \quad \ldots\ldots (45)$$

The first moment which is equal to the mean and second of the beta distribution are given as

$$\mu_x(1) = \frac{\alpha}{(\alpha+\beta)} \quad \ldots\ldots (46)$$

$$\mu_x(2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \quad \text{......} (48)$$

The sample mean and variance are;

$$\bar{X} = \frac{1}{n}\sum x_i \quad \text{......} (49)$$

$$S^2 = \frac{1}{n-1}\sum(x-\bar{x})^2 \quad \text{......} (50)$$

Now comparing equation (46) and (49)

$$\bar{X} = \frac{\alpha}{\alpha+\beta} \quad \text{......} (51)$$

$$\alpha = \bar{X}(\alpha+\beta)$$

$$\alpha = \alpha\bar{X} + \beta\bar{X}$$

$$\alpha - \alpha\bar{X} = \beta\bar{X}$$

$$\beta = (\frac{1}{\bar{X}} - 1)\alpha \quad \text{......} (52)$$

Comparing equation (49) and (50)

$$S^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad \text{......} (53)$$

$$\alpha\beta = S^2(\alpha+\beta)^2(\alpha+\beta+1) \quad \text{......} (54)$$

Substituting equation (52) into (54)

$$\alpha \times \alpha\left(\frac{1}{\bar{X}} - 1\right) = S^2(\alpha + \alpha\left(\frac{1}{\bar{X}} - 1\right))^2(\alpha + (\frac{1}{\bar{X}} - 1)\alpha + 1) \quad \text{......} (55)$$

$$\frac{\alpha^2}{\bar{X}} - \alpha^2 = S^2(\alpha + \frac{\alpha}{\bar{X}} - \alpha)^2(\alpha + \frac{\alpha}{\bar{X}} - \alpha + 1) \quad \text{......} (56)$$

$$\left(\frac{1-\bar{X}}{\bar{X}}\right)\alpha^2 = S^2\frac{\alpha^2}{\bar{X}^2}(\frac{\alpha+\bar{X}}{\bar{X}}) \quad \text{......} (57)$$

$$\left(\frac{1-\bar{X}}{\bar{X}}\right)\alpha^2 = S^2(\frac{\alpha^3+\bar{X}\alpha^2}{\bar{X}^3}) \text{ multiplying through by } \bar{X} \text{ gives} \quad \text{......} (58)$$

$$(1-\bar{X})\alpha^2 = S^2\frac{(\alpha^3+\bar{X}\alpha^2)}{\bar{X}^2} \quad \text{......} (59)$$

$$(1-\bar{X})\alpha^2 = \alpha^2(\frac{S^2\alpha + S^2\bar{X}}{\bar{X}^2}) \quad \text{......} (60)$$

Dividing through by α^2 and multiplying by $\bar{X}^2$ gives

$$(1-\bar{X})\bar{X}^2 = \alpha S^2 + S^2\bar{X} \quad \text{......} (61)$$

$$(1-\bar{X})\bar{X}^2 - S^2\bar{X} = \alpha S^2 \quad \text{......} (62)$$

Dividing through by S^2 gives

$$\frac{(1-\bar{X})\bar{X}^2 - S^2\bar{X}}{S^2} = \alpha \quad \text{......} (63)$$

Now,

$$\beta = \left(\frac{1-\bar{X}}{\bar{X}}\right) \times \bar{X}\left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right] \ldots\ldots (64)$$

Therefore, the Method of moments estimators for the beta distribution are

$$\hat{\alpha} = \bar{X}\left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right] \ldots\ldots (65)$$

$$\hat{\beta} = (1-\bar{X})\left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right] \ldots\ldots (66)$$

This paper estimated the following parameter values of the beta distribution as tabulated below

Table 1 Parameter values considered

Parameter					
A	0.5	1	1.5	1	2
B	0.5	1	1.5	2	2

The parameter values were chosen due to their special properties as described below:

- when α=β=0.5, the distribution is an Arcsine distribution with the following properties

$$f(x) = \frac{1}{x\sqrt{x(1-x)}}, 0 < x < 1 \ldots\ldots (67)$$

$$F(x) = \frac{2}{\pi}\arcsin(\sqrt{x}), \text{where } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \ldots\ldots (68)$$

- When α=β=1, the distribution approximates to the Uniform distribution with mean 0.5, variance 0.125, both skewness and kurtosis are zero.
- When α=1 and β=2, the distribution approximates to the triangular distribution with its pdf given as

$$f(x;1,2) = \frac{(1-x)}{B(1,2)} = 2 - 2x \ldots\ldots (69)$$

$$F(x) = 2x - x^2 \ldots\ldots (70)$$

- when α=β=2 the distribution takes on a Parabolic shaped.
- When α=β=1.5 the distribution is a Wigner Semicircle distribution.

In the equations above, $f(x)$is the probability density function and $F(x)$ is cumulative distribution function. We apply the MLE and MOM methods for determine the point estimates for the two-parameter beta distribution.

The two parameter estimation methods were applied to the simulated beta distributed random variables generated from R using the package called EnvStats with different parameter combinations, and therefore different shapes, to examine the performance of the two estimation methods. The two-point parameter estimators are said to adequately estimate the parameters of the beta distribution if they obtain estimates close to the actual values used to generate the data. The effect of sample size on the performance of each of the estimators was considered by simulating samples of size 50, 100, 500 and 1000. The mean squared error (MSE), bias and variance of each estimator were calculated for every combination of parameters and sample size used in the simulation study to determine the goodness of the estimates from the two methods.

For the simulation study, realizations from the beta distribution were obtained using the R command *rbeta(n,shape1,shape2),* where ***n*** is the desired sample size, shape1 is the desired $\boldsymbol{\alpha}$, and shape2 is the desired $\boldsymbol{\beta}$**.** The five (5) parameter combinations examined in this study were chosen to capture the range of profiles of the beta distribution.

These five (5) combinations may be found in Table 1 and Figure 1.

3. Results and discussion

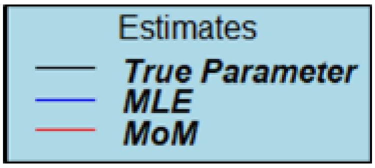

Figure 1 Parameter Estimates

Figure five (5) above is the legend to the Figures six 6 through eleven (11) which are graphs of the estimates of the parameters from the maximum likelihood and method of moment estimators.

Table 2 Parameter estimates for Beta (0.5,0.5)

Sample size (n)	α = 0.5		β = 0.5	
	MLE	MME	MLE	MME
50	0.5417264	0.5095458	0.4680386	0.6098021
100	0.3927370	0.3743463	0.4025905	0.3567893
500	0.5419481	0.5285822	0.5178041	0.4989743
1000	0.5250422	0.5260358	0.5268454	0.5200939

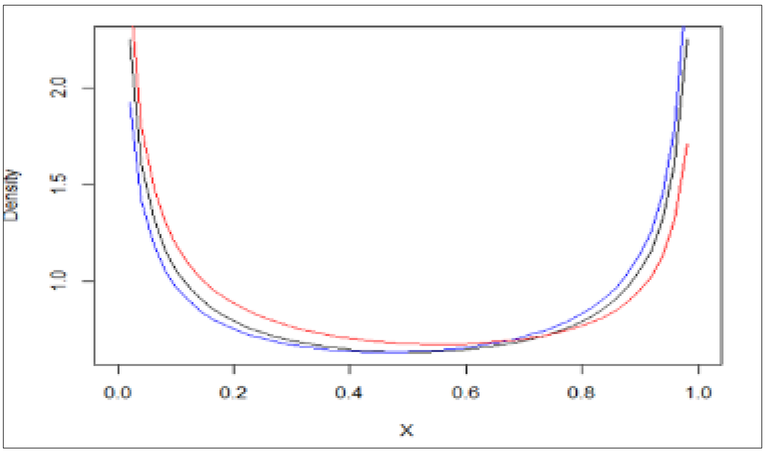

Figure 2 Graph of Estimates for Beta (0.5,0.5) from MLE and MoM methods

In Figure six (6), it evident that the maximum likelihood estimates give estimates that are close to the true parameters as shown in the plots, MLE (blue) is closer to the true parameter (black) than the MoM (red).

In Table 2, both methods appear to be consistent except for the case when the sample size is 100. The maximum likelihood method gives the most efficient estimate, this is shown in Table 3 as the variance of the estimates from the MLE appears to be relatively small as compared to the MME.

Table 3 Properties of the estimates for beta (0.5, 0.5)

(α=0.5,β =0.5)	**n=50**		**n=100**		**n=500**		**n=1000**	
	MLE	**MME**	**MLE**	**MME**	**MLE**	**MME**	**MLE**	**MME**
Mean	0.5364876	0.4330998	0.4723382	0.49782	0.490165	0.5096278	0.5149026	0.4942909
Variance	0.1237302	0.1138298	0.1554563	0.1291051	0.1188201	0.1259534	0.1203437	0.1226355
Standard Deviation	0.3517531	0.3641289	0.3923291	0.389989	0.3484024	0.3558872	0.3508694	0.3513045
Bias	0.0364876	-0.06690018	-0.0276618	-0.00218002	-0.00983503	0.009627838	0.01490263	-0.005709054
MSE	0.1250616	0.1183054	0.1562214	0.1291099	0.1189168	0.1260461	0.1205658	0.1226681

Table 4 Parameter estimates for the Beta (1,1)

Sample size (n)	**α = 1**		**β = 1**	
	MLE	**MME**	**MLE**	**MME**
50	0.8721207	0.7774687	0.9387141	0.8092706
100	1.000432	1.066623	1.813092	1.912086
500	0.9786284	0.9567314	1.0293154	1.0048733
1000	0.9886397	0.9750243	0.9415424	0.9386461

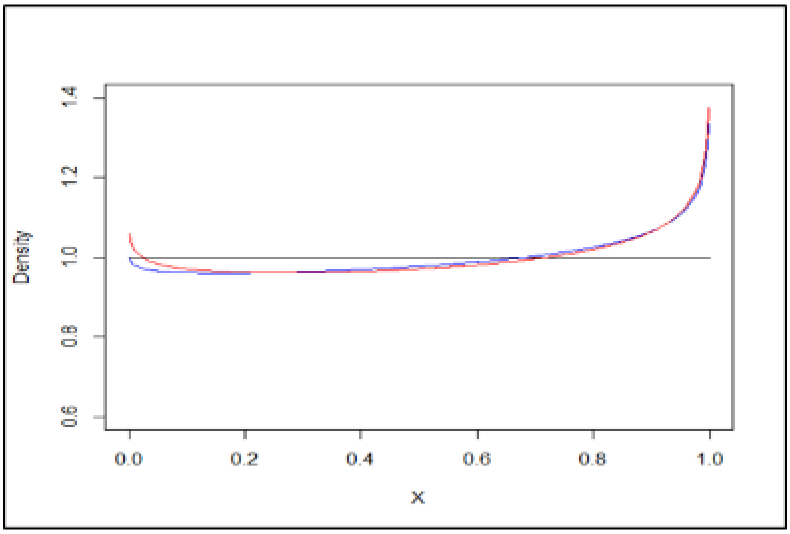

Figure 3 Graph of estimates for Beta (1,1) from both MLE and MoM

Both methods over-estimated the parameter α which makes the plots right skewed on the graph in Figure seven (7) above.

Table 5 Properties of the estimates for Beta (1,1)

(α=1,β=1)	n=50		n=100		n=500		n=1000	
	MLE	MME	MLE	MME	MLE	MME	MLE	MME
Mean	0.4651984	0.4532159	0.3265646	0.4102596	0.5022884	0.5033177	0.4976819	0.4952415
Variance	0.06286783	0.08156651	0.05839583	0.0615938	0.08206677	0.08498205	0.08540052	0.08867845
Standard Deviation	0.3387106	0.2924843	0.2371998	0.2328596	0.2967503	0.2915168	0.2922337	0.2977893
Bias	-0.5348016	-0.5467841	-0.6734354	-0.589740	-0.4977116	-0.4966823	-0.5023181	-0.5047585
MSE	0.3488806	0.3805394	0.5251244	0.4042544	0.3297837	0.3316754	0.337724	0.3434596

Both methods gave estimates which are greater than 1 for the parameter β when the sample size is 100. From table 4 the variance of the maximum likelihood estimators appears to be the smallest, there the maximum likelihood estimators are most efficient. Both methods are consistent as the sample size increases. Looking at the biases, the maximum likelihood methods tend to be unbiased as the difference between the expected values of the estimated parameters and the true parameters tends to zero as the sample size increases.

Table 6 Parameter estimates for Beta (1.5,1.5)

Sample size (n)	α = 1.5		β = 1.5	
	MLE	MME	MLE	MME
50	1.410892	1.365026	1.592117	1.563280
100	1.816022	1.821379	1.598315	1.569878
500	1.395002	1.362485	1.483906	1.449812
1000	1.504829	1.538588	1.464962	1.486546

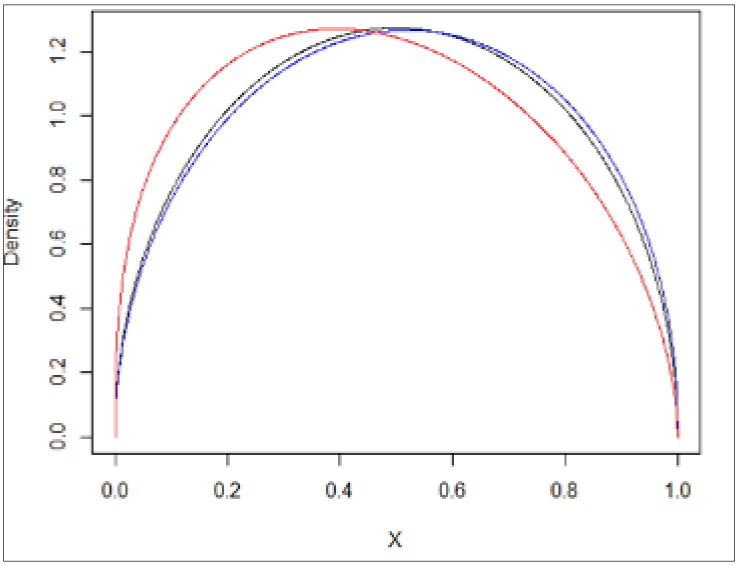

Figure 4 Graph of estimates for Beta (1.5,1.5) from both MLE and MoM

From Figure eight (8) above is the graph for the beta (1.5,1.5) which is also called the Wigner semi-circle distribution, the method of moments estimators plots is left skewed which indicates that the estimate for the parameter β is greater than the true parameter of β which is 1.5.

Table 7 Properties of the estimates for Beta (1.5,1.5)

(α=1.5,β=1.5)	**n=50**		**n=100**		**n=500**		**n=1000**	
	MLE	MME	MLE	MME	MLE	MME	MLE	MME
Mean	0.4966661	0.451472	0.5489327	0.4901267	0.0678343 6	0.4739501	0.5086016	0.4937091
Variance	0.0491644 7	0.0461503 8	0.0487570 5	0.056889 7	0.0652553 1	0.0678343 6	0.568225	0.0622354 5
Standard Deviation	0.2217306	0.2148264	0.22081	0.238515 6	0.2554512	0.2604503	0.753807	0.2494703
Bias	-1.003334	-1.048528	-0.9510673	-1.009873	-1.019851	-1.02605	-0.991398	-1.006291
MSE	1.055843	1.145561	0.9532861	1.076734	1.105352	1.120613	1.551096	1.074857

From table 7 above, the Maximum likelihood method gives the most efficient, most consistent and an unbiased estimate relative to the method of moment. The difference in the estimates given by the two methods do not vary significantly much.

Table 8 Parameter estimates for Beta (1,2)

Sample size (n)	**α = 1**		**β = 2**	
	MLE	**MME**	**MLE**	**MME**
50	1.104358	1.061187	2.365482	2.247744
100	0.9815375	0.9858293	1.9190667	1.9123410
500	1.023645	1.014421	2.047710	2.052318
1000	1.029208	1.016262	2.038727	2.008221

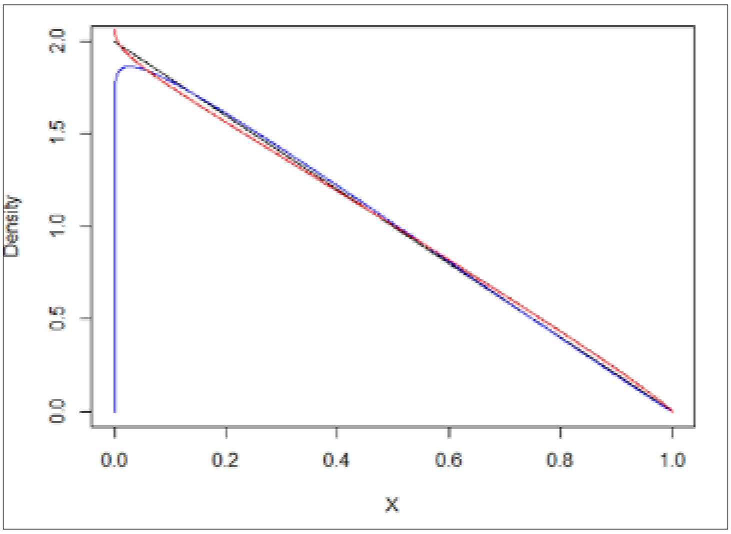

Figure 5 Graph of estimates for Beta (1,2) from both MLE and MoM

The beta (1,2) which is also called the triangular distribution is shown in Figure nine (9) above, both methods appear to be close to the true parameters

Table 9 Properties of the estimates for Beta (1,2)

(α=1,β=2)	n=50		n=100		n=500		n=1000	
	MLE	MME	MLE	MME	MLE	MME	MLE	MME
Mean	0.3182735	0.3207039	0.3383907	0.3401558	0.3332878	0.3307817	0.3354726	0.3360118
Variance	0.04854211	0.05055846	0.05739686	0.05757825	0.05457815	0.05443309	0.05480194	0.05543765
Standard Deviation	0.2203227	0.2248521	0.2395764	0.2399547	0.2336197	0.233309	0.2340981	0.235452
Bias	-0.6817265	-0.6792961	-0.6616093	-0.6598442	-1.166712	-1.169218	-1.164527	-1.163988
MSE	0.5132932	0.5120017	0.4951237	0.4929727	1.415796	1.421505	1.410926	1.410306

From table 8, overleaf both methods gave estimates almost exactly to the Beta (1,2) but from when both methods were tested for efficiency, the maximum likelihood is the most efficient with the smallest bias and mean square error.

Table 10 Parameter estimates for Beta (2,2)

Sample size (n)	α = 2		β = 2	
	MLE	MME	MLE	MME
50	2.484930	2.477630	2.890436	2.950747
100	1.943518	1.872087	2.136747	2.066923
500	1.937936	1.946307	1.909107	1.922622
1000	2.113234	2.145247	2.133996	2.166377

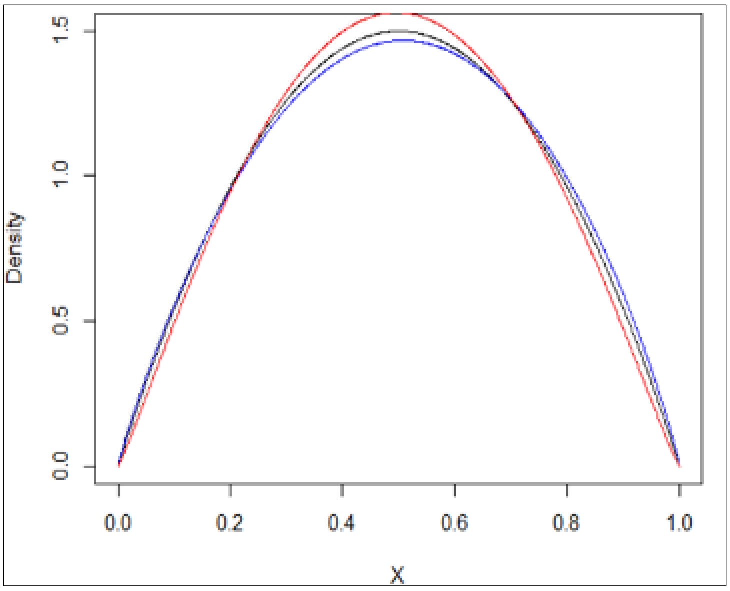

Figure 6 Graph of estimates for Beta (2,2) from both MLE and MoM

In the figure above, estimates from the maximum likelihood estimators appear to be close to the true parameters as compared to the method of moments.

Table 11 Properties of the estimates for Beta (2,2)

(α=2,β=2)	n=50		n=100		n=500		n=1000	
	MLE	**MME**	**MLE**	**MME**	**MLE**	**MME**	**MLE**	**MME**
Mean	0.4622811	0.4564219	0.4763215	0.4752684	0.5037469	0.5030609	0.4975558	0.4975496
Variance	0.03899028	0.03859465	0.0473993	0.05049359	0.05157494	0.05134407	0.04764305	0.04706545
Standard Deviation	0.1974596	0.1964552	0.2177138	0.2247078	0.2271012	0.2265923	0.2182729	0.2169457
Bias	-1.537719	-1.961405	-1.523678	-1.524732	-1.496253	-1.496939	-1.502444	-1.50245
MSE	2.40357	2.421228	2.368995	2.3753	2.292402	2.292171	2.304982	2.304423

The estimates estimated for the parabolic Beta (2,2) distribution by the two methods were consistent as the sample sizes increased. Both methods are consistent but ML estimates are efficient in all sample sizes as it has the least variance.

4. Conclusion

The method of moment technique is recommended for the estimations of parameters if the experimenter wishes to do so without a software, this is because of the simplicity of the method of moment estimators of the beta distribution. The maximum likelihood technique is very tedious and difficult to compute without the use of a software package. Both methods proved to be consistent. The method of moments and maximum likelihood over-estimated the parameters when the sample size is 100. The efficiency test showed the maximum likelihood method gives the most consistent, efficient and sufficient estimates as it does so by differentiating the logs of the likelihood function, by so doing the it minimizes the error thereby making it the best method for estimating the shape parameter of the beta distribution.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed

References

[1] Arjun K. Gupta and Saraless Nadarajah (2018), Handbook of Beta Distribution and its Applications, CRC Press, 600 Broken Sound Parkway, NW Suite 300, Boca Raton, 3,33-229.

[2] Ahmed, A.A. (2011) Financial Performance Evaluation of Some Selected Jordanian Commercial Banks. International Research Journal of Finance and Economics, 68, 50-63.

[3] Casella G, Robert C P and Wells M. T. (2002). Lecture Notes-Monograph Series 342–347

[4] Eugene, N., Lee, C., Famoye, F. (2002), Beta Normal distribution and its applications. Vol. 31, 496-513.

[5] Fielitz, B.D and Myers, B.L., (1975), Estimation of parameters in the beta distribution. Decision Sciences, Vol.6, No.1, 1-12.

[6] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) Bayesian Data Analysis. 2nd Edition, Chapman & Hall/CRC.

[7] Claire Elayne Bangerter Owen (2008), Parameter Estimation for the Beta Distribution, BYU ScholarsArchive, Provo, Utah, 1-3 and 8-12.

[8] Chotikapanich, D.; Rao, D.S.P.; Tang, K.K (2007). Estimating income inequality in China using grouped data and the generalized beta distribution. In Review of Income and Wealth; Wiley: Hoboken. Vol. 53, 127 – 147.

[9] Granadesikan, R.S. Pinikham & Laura P. Hughes (1997), Maximum Likelihood Estimation of the Parameters of the Beta Distribution from Smallest Order Statistics, Technometrics.Vol.9 No.4, 607-620.

[10] Harr, M.E (1977), Mechanics of Particle Media; A Probabilistic Approach. McGraw-Hill Book Company, New York.

[11] Johnson N L, Kotz S and Balakrishnan N 1996 Continuous univariate distributions, Computational Statistics & Data Analysis vol 2

[12] Kong.L., Carl L, and Sepanski, J.H. (2007), On the properties of Beta-Gamma Distribution, Journal of modern Applied Statistical method; Vol. 6 187-211.

[13] Kumachov, Muchadin and Vamling, Karina. 2006. ?rgativnostb v cërkesskich jabikach Circassian clause structure. Malmö: Malmö Univ., IMER. xxxi+210pp. (In Russian with an introduction and summary in English).

[14] Ng, D.W.W.; Koh, S.K.; Sim, S.Z.; Lee, M.C (2019). The study of properties in generalized beta distribution. IOP Conf. Series J. Phys.

[15] Vijay P. Singh (1999), Entropy-Based Parameter Estimation in Hydrology, Kluwer Academic Publishers, London, 275-281.

[16] Oboni, F and Boudreau, P.L, (1985), Simplified use of the beta distribution and sensitivity to Bounded locations. Structural safety, Vol.3 63-66.

[17] Chris Piech (2016). Beta Distribution, handout Number 22, 1-2.

[18] Romesburg, H.C., (1976). Estimation of Parameters in the beta distribution, comment Decision science, vol.7, 162.