Journal of Mathematical Problems, Equations and Statistics

E-ISSN: 2709-9407
P-ISSN: 2709-9393
JMPES 2022; 3(2): 78-84

www.mathematicaljournal.com
Received: 16-08-2022
Accepted: 01-10-2022

Khalid Talib Othman
Al-Iraqia University Economic & Administration College,
Baghdad, Iraq

Corresponding Author:
Khalid Talib Othman
Al-Iraqia University Economic & Administration College,
Baghdad, Iraq

Using the nonparametric methods to estimate parameters of The Beta distribution

Khalid Talib Othman

Abstract
As a probability function that permits modeling the natural behavior of random variables that may only take values within a specified interval of real numbers with finite length, the Beta distribution has gained significance in the field of applied statistics. Variables with ranges between 0 and 1 (percentages) are a specific instance that arises frequently in practice. The Beta distribution is typically used as a prior distribution when the sample data can be fitted to a binomial distribution; however, in the case of interest for this work, the sample data can be fitted to a Beta distribution, which would entail having a likelihood function that is different from the binomial likelihood function. Therefore, a methodology is proposed in this article to obtain prior distributions for the shape parameters of the Beta distribution. External information on the prior distribution's mean and variance was used to generate a joint distribution for the distribution's parameters. Finally, the marginal distributions of the joint distribution were used to derive the prior distributions for the shape parameters.

Keywords: Beta distribution, nonparametric methods, Particle swarm optimization

Introduction

Many research problems include random variables whose values fall inside bounded spaces, that is, intervals with both upper and lower bounds on the real number line. Various probability functions, such as the triangular, Kumaraswamy, and Beta distribution, as well as truncation on models developed for random variables with a path in the real line or on the subset of positive real numbers, can be found in the literature and used to probabilistically model the behavior of such variables. Most often, variables that take on values within intervals (c, d), where c and d are positive real numbers, are represented by the Beta distribution with shape parameters and. The conventional Beta probabilistic model, a special case of the extended Beta model, is assumed when modeling proportions, indicators generated from latent variable theory, and other random variables whose ranges of motion are defined in (0,1) space, as demonstrated in [7]. Parameters for this distribution can be estimated using tools like moments and maximum likelihood. When deriving the log-likelihood function with respect to the parameters, the equation is indeterminate, making it necessary to employ iterative approaches such the Newton Raphson algorithm for maximum likelihood estimate.

The Bayesian paradigm appears as a crucial alternative when the frequentist paradigm cannot be used to make the estimation or when it is required to have an additional source of information to generate the estimates. Estimates derived from posterior distributions are helpful because they consider the expertise of the people who provided the data.

focus of investigation (Check out) [1, 6, 16, 17]. Using a Bayesian methodology can be difficult when a conventional Beta distribution can be connected with the sample data to predict its behavior (Beta likelihood), because there are no a priori prior probability distributions that allow producing least friendly forms for the posterior distribution. Further, the two form parameters cannot be interpreted in a way that is adaptive to a particular real-world context; rather, their interpretations are only known from a mathematical (constants that allow the function to take different forms) or probabilistic (shape parameters of the distribution of) standpoint. Using an approximation to the distribution of the parameters, some writers (such as) [11] have performed Bayesian estimation in Beta Mixture Models (BMM) in order to find a closed form solution. figuring out the precise value using the rules of the approximation approach.

A product of Gamma distributions is used as a joint prior distribution for the parameters under the assumption of their independence. Using a hierarchical model, they obtain the estimates of the parameters by attaching the $n - \acute{e}sima$ Beta observation to a latent indicator variable, this generates a set of indicator vectors that are assumed to be independent, obtaining a mixture of conditioning proportions of the distribution of the indicator variable a which are assigned Dirichlet distributions. After that, we have a joint distribution, and the hyperparameters are produced repeatedly using the updating strategy of variational inference.

Parameters of the Beta distribution were estimated by [13] using maximum likelihood, the method of moments, and three other techniques, including the program review and evaluation technique (PERT), in which the mean and variance of the distribution are calculated using a modular value and with a procedure analogous to the method of moments, yielding results that are equal to the mean and variance of the Beta distribution. An alternative approach uses a variant of the two-tailed power distribution (TSP) to estimate the Beta distribution's mean and variance by first calculating the TSP's mean and variance and then equating them to the Beta distribution's mean and variance. The last method consists of obtaining an estimator based on the first and third quartiles of the Beta distribution, which is an empirical method that uses the first and third population quartiles of the distribution assuming that the real value of the parameter is known and then samples of the distribution whose parameter values lie in an interval built around the estimate obtained with the method of moments.

In addition [9], used the Metropolis-Hastings Tailored Random Blocks (TaRBMH) technique, which is based on the already-existing algorithm of Random Walk Metropolis-Hastings, to estimate the parameters of the Generalized Beta (BG) distribution from pooled data from a Bayesian perspective (RWMH).

This new approach eliminates the possibility of making a faulty decision about which blocks to use by encoding the parameters and then randomly grouping them into those blocks. The parameters in the remaining blocks are then changed one at a time using a Metropolis-Hastings step that accounts for the proposed density, until the block's succeeding density is adapted in position and curvature to the most recent value of the parameters.

The authors consider the Generalized Beta Likelihood using their parameters; I give each parameter a priori distribution, considering the constraints the authors impose. Each distribution has its own set of hyperparameters; the location parameter has a Normal prior distribution, the scale parameter and the two shape parameters each have Gamma prior distributions. The posterior distribution density function was created by multiplying the prior distributions by the BG likelihood, and the TaRBMH algorithm was used to get the estimates.

To approximate complex distributional forms with non-closed analytical structures is one of the weaknesses of the Bayesian approach to statistical inference. In mathematics and statistics, approximating integrals and distributions is a well-studied problem. Using the Newton-Cotes formulas, which include the Trapezoid rule and the Simpson rule among others, one can find an approximate value of the integral by evaluating the function at equidistant points; the greater the number of intervals into which the function is divided, the more precise the result will be. In order to determine which method provides the best approximation of an integral, this comparison was performed. The authors then carried out the respective procedures, taking into account all the theoretical constraints that the methods employ, and, upon obtaining the results of the approximations of the integrals of a function, they reached the following conclusion: algorithms based on the trapezoid scheme have proven to produce more accurate results in terms of approximating the value of the integral, using few iterations, and with small errors. However, various Markov Chain (MC) Monte Carlo (MCMC) methods were utilized in [4], such as the Metropolis-Hasting algorithms and Gibbs Sampling by blocks, both of which are common in Bayesian statistics. The authors did this because direct sampling of the C4 sand of the Misoa Formation in Lake Maracaibo is difficult, and they wanted to get their hands on some distribution estimates instead. Next, the authors used the collected values to create stratigraphic pseudo-columns that describe a deposit. Maps of the sand content were then made using data from both the actual and estimated column heights. Finally, after analyzing the data, they determined that the Metropolis-Hasting technique was able to discriminate between more than 70% of the silt in the research area to be sands along the created pseudo-sequences. When using the right block length, the Gibbs Block Sampling technique can also pick up on the presence of very thin layers of the other lithologies that have been spotted in the area. Estimating the parameters of the Beta distribution, for example, is a time-consuming procedure due to the high level of complexity needed in performing algebraic operations on the density function of the Beta distribution. In addition, the challenge is exacerbated from a Bayesian perspective by the fact that both parameters of the Beta distribution are shape parameters with no straightforward meaning from an applied perspective. Since the Beta distribution is typically regarded to be the best fit for describing the behavior of observed data, this article's primary goal is to devise a technique for estimating its parameters that yields forms for the a priori distributions.

The technique assumes that the mean and variance of the prior distribution have supported the process of extracting information outside the sample (which may be in the expertise of a specialist in the area, in publications, historical data, etc.). We also make the assumption that the sample of working data satisfies the property of interchangeability. The proposed methodology takes into account a bi parametric prior distribution during estimation by first considering the construction of a joint prior distribution for the two parameters and then carrying out a procedure to obtain the marginal distributions that would be that of each parameter individually. It was proposed to use the sample generation algorithm for Metropolis Hastings distributions in conjunction with the trapezoidal rule, a numerical approximation method for complex integrals, because the 4 form of the joint distribution for the two parameters of the prior Beta distribution is analytically intractable.

Methodology

The Kumaraswamy distribution is a member of a family of probability distributions for continuous random variables. It is defined on the interval [0, 1], and its shape parameters [a1, b1] correspond to the set of non-negative reals. The Kumaraswamy distribution is used to [8]. The probability density function (PDF) and the cumulative density function (CDF) for a distributed variable X that is studied by Kumaraswamy are as follows:

$$\begin{aligned} &\mathrm{f(x)} = \mathrm{a_1 b_1 x^{a_1-1}(1-x^{a_1})^{b_1-1}}\ 0 < x < 1; a_1 > 0; b_1 > 0 \\ &\mathrm{F(x)} = 1-(1-\mathrm{x}^{a_1})^{b_1} \end{aligned} \tag{1}$$

The moment function of the distribution can be obtained using the expression:

$$m_n = \frac{b_1\Gamma\left(1+\frac{n}{a_1}\right)\Gamma(b_1)}{\Gamma\left(1+b_1+\frac{n}{a_1}\right)} \quad (2)$$

So, the mathematical expectation and the variance are:

$$\mu^* = m_1 = \frac{b_1\Gamma\left(1+\frac{1}{a_1}\right)\Gamma(b_1)}{\Gamma\left(1+b_1+\frac{1}{a_1}\right)}$$
$$\sigma^{2*} = m_2 - m_1^2 \quad (3)$$

It is also possible to obtain the median using the expression:

$$m_e^* = \left(1-2^{-1/b_1}\right)^{1/a_1} \quad (4)$$

Metropolis–Hastings (MH) algorithm

For arbitrarily creating and sampling posterior distributions, a great number of ingenious methods have been developed. The Markov chain Monte Carlo (MCMC) simulation is a general method that is based on obtaining values of from approximate distributions and then correcting these values to approximate the target posterior distribution,
$f(\theta|x)$ [5]. This is done by obtaining values of from approximate distributions and then correcting these values. An example of a Markov chain Monte Carlo (MCMC) technique is the MH Algorithm, which is utilized to get a sequence of random samples from a probability distribution when direct sampling is difficult to do [5].

When paired with any arbitrary transition kernel q (proposed distribution), the acceptance probability, as it is defined here, can be used to build a reversible chain. This distribution is called a proposed distribution. The following idiom has been used most frequently to express the likelihood of acceptance:

$$\alpha(\theta,\phi) = min\left\{1, \frac{\pi(\phi)q(\phi,\theta)}{\pi(\theta)q(\theta,\phi)}\right\} = min\left\{1, \frac{\pi(\phi)w(\phi|\theta)h(\theta)}{\pi(\theta)w(\theta|\phi)h(\phi)}\right\} \quad (5)$$

The proposed transition kernel q is arbitrary and, therefore, it is a flexible tool for the construction of the algorithm. See [5] for more details.

Proposed method

The statistical model that is associated with the situation assumes that the set of observations obtained after sampling in order to obtain information about a population parameter presents a natural behavior that can be modeled with a standard Beta distribution. Assuming this to be the case, the model considers the function of density to be the one that appears in equation (6).

$$f_x(x \mid \alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, 0 \le x \le 1; \alpha > 0, \beta > 0 \quad (6)$$

Where

$$E(X) = \mu = \frac{\alpha}{\alpha+\beta}\ ;\ Var(X) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (7)$$

It is not possible to give a practical interpretation to the shape parameters of the distribution, which is a fact that makes it quite difficult to carry out an elicitation process, which is a necessary procedure when Bayesian inference is desired. Where and are the shape parameters of the distribution (See [12], for expand concepts of elicitation). However, the mean and variance of this distribution are straightforward to interpret from any contextual perspective, since 5 is an indicator of both the central tendency and dispersion of the data, respectively. This is the case since 5 is a central number. Because the mean and the variance are functions of and, it is possible to elicit information for and 2 as parameters of interest, assign prior distributions to them, and then derive the prior distributions for and. This is possible due to the fact that the mean and the variance are functions of and. In this setting, it is essential to have a solid understanding of a number of prerequisites in order to formulate a concept of which distributions will be appropriate and to assign them as prior distributions for and 2. The initial step in this process is to determine the actual number space in which the two parameters look for their values. Therefore, given that is more than or equal to 0 and y is greater than or equal to 0, it can be demonstrated that:

$$\lim_{\beta\to\infty}\mu = \lim_{\beta\to\infty}\frac{\alpha}{\alpha+\beta} = \lim_{\alpha\to 0}\mu = \lim_{\alpha\to 0}\frac{\alpha}{\alpha+\beta} = 0 \quad (8)$$

$$\lim_{\beta\to 0}\mu = \lim_{\beta\to 0}\frac{\alpha}{\alpha+\beta} = \lim_{\alpha\to\infty}\mu = \lim_{\alpha\to\infty}\frac{\alpha}{\alpha+\beta} = 1 \quad (9)$$

Considering the above limits, it is evident that $\mu \in (0,1)$. Proceeding analogously for $\sigma 2$, we have:

$$\lim_{\beta\to\infty}\sigma^2 = \lim_{\beta\to\infty}\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \lim_{\alpha\to\infty}\sigma^2 = \lim_{\alpha\to\infty}\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0 \quad (10)$$

$$\lim_{\beta \to 0} \sigma^2 = \lim_{\beta \to 0} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \lim_{\alpha \to 0} \sigma^2 = \lim_{\alpha \to 0} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0 \quad (11)$$

Analysis of the limits does not provide enough information if the goal is to establish a range of real numbers for 2 based on the result presented above. It is evident that the variance approaches zero when either of the shape parameters approaches zero or when both approach infinity. This is the case regardless of which of these two scenarios is being considered.

In order to better understand how σ^2 behaves in different locations, a graphical representation of the data was created (See Figure 1). It is possible to observe that the maximum value that the variance can take is very close to 0.25, which agrees with what was inferred in the calculation of its limits. This is the case since regardless of the value of, as the values of the parameter approach to zero or increment after two, σ^2 always approaches zero. This is the case because regardless of the value of, as the values of the parameter approach to zero or increment after two, σ^2 always approaches zero. Because of this, it is not unreasonable to believe that the variance of the Beta distribution is always contained inside the interval (0, 0.25).

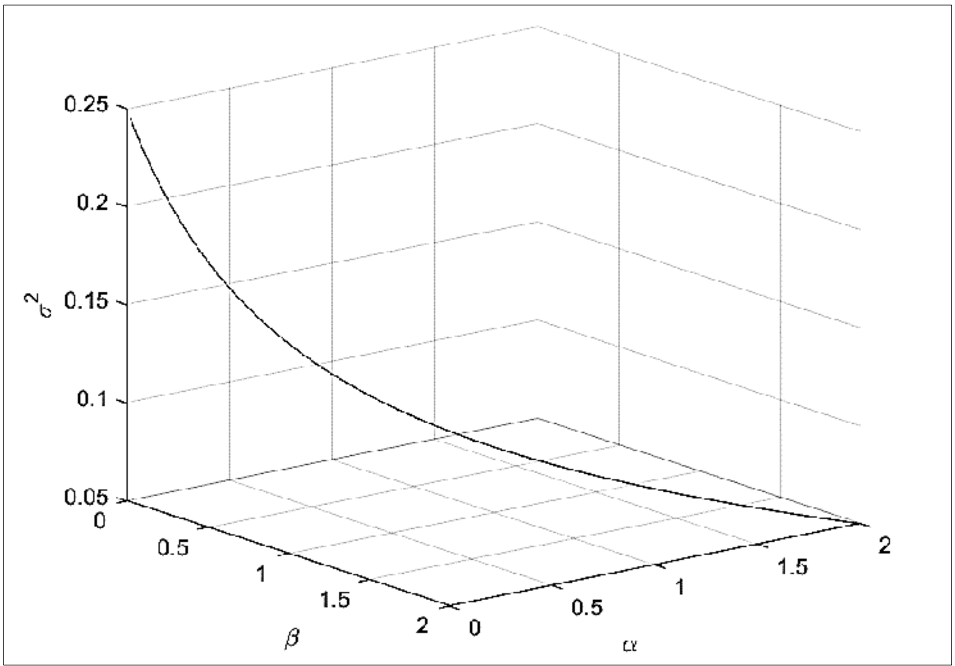

Fig 1: Behavior of $\sigma 2$ before different values of (α, β)

As μ and $\sigma 2$ are functions of the shape parameters, what is proposed is to obtain a joint prior distribution $f(\mu, \sigma 2)$, in such a way that, using the transformation method for distributions of random variables, it is possible to derive a joint prior distribution of the form $f(\alpha, \beta)$, and finally marginalize $f(\alpha)$ and $f(\beta)$.

In this context, it is necessary to know if the transformation is one to one. So far, it is known that positive values of α and β establish that $\mu \in (0,1)$ and $\sigma 2 \in (0, 0.25)$. it is then required to know if the latter result in positive values for the shape parameters when performing the inverse transformation:

$$\alpha = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right], \beta = (1-\mu)\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right] \quad (12)$$

Maximum Likelihood Estimators

Maximum likelihood is one of the most popular methods for estimating parameters. The likelihood function $f(x \mid \theta_1, \dots, \theta_k)$ as $L(\theta_1, \dots, \theta_k \mid x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta_1, \dots, \theta_k)$ is defined for an independent and identically distributed sample X1,.., Xn drawn from a population with pdf f(x). The MLE is the parameter value that best fits the data in the sample. Potential MLEs are the solutions to the equations I L(|X) = 0, I = 1,..., k. Maximums, as opposed to minimums, can be confirmed by ensuring the second derivative of the probability function is positive and hence greater than zero. Since derivatives of sums are more appealing than derivatives of products, the log likelihood function, log L(|X), is frequently the more convenient choice (Casella and Berger 2002). Consistent and asymptotically efficient, MLEs are highly sought after because they provide a lower bound on variance and converge in probability to the parameter being estimated.

The likelihood function for the beta distribution is

$$L(\alpha, \beta \mid \mathbf{X}) = \prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1} = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^n \prod_{i=1}^{n} (x_i)^{\alpha-1} \prod_{i=1}^{n} (1-x_i)^{\beta-1} \quad (13)$$

yielding the log likelihood

$$\log L(\alpha,\beta \mid \mathbf{X}) = n\log(\Gamma(\alpha+\beta)) - n\log(\Gamma(\alpha)) - n\log(\Gamma(\beta)) + (\alpha-1)\sum_{i=1}^{n}\log(x_i) + (\beta-1)\sum_{i=1}^{n}\log(1-x_i). \quad (14)$$

To solve for our MLEs of α and β we take the derivative of the log likelihood with respect to each parameter, we will solve for $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$ iteratively, using the Particle Swarm optimization where the objective function for algorithm equal to :

$$f = n\log(\Gamma(\alpha+\beta)) - n\log(\Gamma(\alpha)) - n\log(\Gamma(\beta)) + (\alpha-1)\sum_{i=1}^{n}\log(x_i) + (\beta-1)\sum_{i=1}^{n}\log(1-x_i). \quad 0 \le x \le 1; \alpha > 0, \beta > 0 \quad (15)$$

Particle Swarm Optimization

The swarm strategy for optimization was first proposed by Kennedy and Eberhart (Kennedy, 2011). Particle swarm optimization (PSO) is an example of a population-based search method that mimics the cooperative efforts of a flock of birds. In PSO, individuals are "flown" through a hyperdimensional search space using a method similar to that used in simulations. Particles' positions inside the search space are updated according to the social-psychological desire to replicate the accomplishments of others.

Two are the key aspects by which that PSO has become so popular:

1. 1.To begin, PSO's fundamental algorithm is clear and easy to implement since, unlike most evolutionary algorithms, it only uses a single operator to generate new solutions in its original form.
2. PSO has been shown to be useful in a wide range of applications due to its ability to generate high-quality output with little computational expense (Engelbrecht, 2006; Kennedy, 2006).

Following are definitions of various regularly used technical terms in order to build a common vocabulary:

Swarm: Population of the algorithm.

Particle: Swarm inhabitant or member. Every speck stands for a different approach to the situation at hand. Particles have fixed locations that are determined by the solutions they currently represent.

Pbest (personal best): Your best measured position for a particular particle. The most fruitful particle location, in other words (measured in terms of a scalar value analogous to the fitness adopted in evolutionary algorithms).

Gbest (global best): Position of the best particle of the entire Population.

Velocity (vector): This vector is what ultimately decides which way a particle needs to "fly" (move) in order to achieve a better position than it is in now.

Learning factor: A particle's propensity to cheer on its own achievements or those of its peers. C1 and C2 are the two cognitive components used for learning. The cognitive learning factor, or C1, shows an individual's propensity to be drawn to their own achievements. The social learning coefficient, denoted by C2, shows a particle's propensity to be drawn to its neighbors' achievements. Constants C1 and C2 have conventional definitions.

Inertia weight: The inertia weight, denoted by W, is used to modify how much a particle's past velocities affect its present velocity.

Each individual swarm particle stands in for an option for resolving the problem at hand. Particles move about to reflect their interactions with their surroundings and with other particles. To illustrate, let's write, $x_i(t)$ to represent particle i's location at time t. Adding the velocity $v_i(t)$ to the particle's present position yields the following expression:

$$x_i(t+1) = x_i(t) + v_i(t) \quad (16)$$

The velocity vector, which represents the socially transmitted data, is typically defined as follows:

$$v_i(t+1) = WV_i(t-1) + C_1 r_1(P_{best\ i} - x_i(t)) + C_2 r_2(g_{best\ i} - x_i(t)) \quad (2.2)$$

where and $r_1, r_2 \in [0;\ 1]$ are random values.

The success of a particle's connections often has a profound effect on the particle's own fortunes. The social structure of the swarm is defined by the topology of the neighborhoods; therefore neighbors are not always particles that are close to each other in parameter (choice variable) space (Kennedy, 2006).

Simulation Results

To evaluate the efficacy of five parameter estimation strategies, we generated simulated data from beta distributions with varying parameter combinations, yielding a wide range of distributional shapes. If these estimates of the beta distribution's parameters were near to the true values that were utilized to generate the data, then they provided a good approximation of those parameters. Our estimators were also subjected to simulated studies with several sample sizes, including 25, 50, 100, and 500, to examine how different sizes of samples impacted their accuracy. Each estimator's bias and MSE were determined for each parameter and sample size combination. In this paper, we detail our simulation setup and the findings we were able to draw from it.

Table 1: Parameter estimates for Beta distribution

	α=2				β=2			
n	MLE	MOM	LMOM	MLE-PSO	MLE	MOM	LMOM	MLE-PSO
25	2.183717	2.136287	1.939994	1.935166	1.914336	1.830342	1.851962	1.916218
50	2.018886	1.901713	1.878638	1.983083	2.10288	1.82158	2.027529	1.979428
100	1.85545	2.125714	1.900434	1.958526	2.101492	2.012319	1.987756	1.931122
500	1.859718	1.89741	2.046418	1.954972	1.952178	2.111667	1.804761	1.952853
	α=1				β=1			
n	MLE	MOM	LMOM	MLE-PSO	MLE	MOM	LMOM	MLE-PSO
25	1.040793	0.980217	1.130327	0.910665	1.147478	0.972566	0.854427	0.985303
50	0.905189	0.833529	1.015337	0.99619	0.833774	1.164259	1.147717	0.962206
100	1.061632	0.891591	1.198454	0.900463	0.959913	0.872739	1.031882	0.935095
500	1.075686	1.165335	0.83127	0.977491	0.903948	0.905521	1.019944	0.951325
	α=2				β=1			
n	MLE	MOM	LMOM	MLE-PSO	MLE	MOM	LMOM	MLE-PSO
25	1.830387	1.966907	1.995701	1.978025	0.852789	0.893912	0.867596	0.954701
50	1.895966	1.819862	1.935088	1.938974	1.17682	0.941263	1.059646	0.929632
100	1.849328	2.161086	2.160022	1.924169	1.182454	1.128478	1.092689	0.974469
500	1.873563	2.177915	1.947699	1.940391	1.030083	0.806161	1.059098	0.918896
	α=1				β=2			
n	MLE	MOM	LMOM	MLE-PSO	MLE	MOM	LMOM	MLE-PSO
25	0.873404	1.171754	0.92254	0.964432	2.175601	1.883097	1.877906	1.93111
50	0.947394	1.110285	1.003403	0.937861	2.150377	1.920499	1.890369	1.992338
100	1.050247	0.994717	1.004309	0.981158	2.020063	1.988369	1.868283	1.943021
500	1.112091	0.974343	1.127051	0.953283	2.04899	1.892195	1.891066	1.918482
	α=0.5				β=0.5			
n	MLE	MOM	LMOM	MLE-PSO	MLE	MOM	LMOM	MLE-PSO
25	0.691899	0.537958	0.346967	0.408552	0.592132	0.685235	0.549624	0.403774
50	0.475548	0.404885	0.41867	0.426248	0.495444	0.518722	0.571654	0.488517
100	0.344448	0.541137	0.427511	0.480101	0.53141	0.508454	0.458206	0.491329
500	0.403226	0.584486	0.469667	0.402922	0.394913	0.392638	0.446975	0.479618

The bias in the MLE-PSO estimations of and was the smallest across the board and decreased with increasing sample size. In a similar vein, more samples resulted in less error for the ML estimations of and. The bias of the MOM estimators marginally rose from sample size 25 to sample size 50, but then stabilized for all bigger sample sizes. Also, the LMOM estimators' biases were very stable across different data sets. The biases of LMOM estimators were consistently the worst. However, when the sample size grew, the MSE for both MLE-PSO and MOM estimations fell. MSE levels for MOM, MLE, and LMOM estimations were quite similar across different sample sizes. The MLE-PSO estimates for both and had the highest MSE across the board. Both the MLE and MOM estimators struggled to capture the U shape of the distribution, as shown by the density figure.

Conclusion

It might be challenging to obtain credible previous distributions for the Beta distribution's parameters. Both factors are form, making it hard to give them a practical interpretation, which is a major roadblock to success. This article develops and describes a method for determining the form of the prior distributions for y and the corresponding values for hyperparameters from a knowledge of the shape parameters and their reparameterizations, and σ^2.

Based on the data, we may infer that and σ^2 are intertwined, since the scale parameter is constrained to be less than the curve $\mu(1-\mu)$, in any case $\mu \in (0,1)$. The joint probability distribution for the shape parameters and can then be calculated using the transformation and Jacobian approach for mathematical functions, after first obtaining a joint prior distribution for the first two moments of the distribution. To derive the marginal distributions for each parameter, it is helpful to first know the joint distribution. The suggested method provides a closed analytic form for the distribution of the parameter, from which random numbers may be easily derived.

Although a closed analytic form for the parameter's distribution was not obtained, a numerical approximation of the density was achieved by employing an algorithm based on the trapezoidal rule in conjunction with the Metropolis Hastings method. Since the derived distribution begins with inferred information for the initial instant of the prior distribution, it can be categorized as an informative prior distribution.

References

1. Box G, Tiao G. Bayesian Inference in Statistical Analysis, Wiley, USA; c1992.
2. Canavos G. Probability and statistics: applications and methods, Limusa, Mexico; c1984.
3. Casella G, Berger R. Statistical Inference, Thomson Learning, USA; c2002.
4. Durán E, Soto M, Aldana M. Metropolis – Hastings and Gibbs by blocks: use of Monte Carlo algorithms via Markov chains in the modeling of stratigraphic columns, field lama, Lake Maracaibo, Magazine of the Faculty of Engineering U.C.V. 2014;29(4):41-54.

5. Gamerman D, Lopes H. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition, CRC Press, USA; c2006.
6. El-Alosey AR, Eledum H. Discrete Erlang-truncated exponential distribution. International Journal of Statistics and Applied Mathematics. 2021;6(1):230-236.
7. Johnson N, Kotz S, Balakrishnan N. Continuous univariate distributions, number v.2, Wiley & Sons, New York, USA; c1995.
8. Jones MC. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages, Journal Statistical Methodology. 2009;6(1):70-81.
9. Larson R, Hostetler R, Edwards B. Calculus with analytical geometry, McGraw Hill, Mexico; c2006.
10. MA Z, Leijon A. Bayesian Estimation of Beta Mixture Models with Variational Inference, Journal IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011;33(11):2160-2173.
11. O'hagan A, Garthwaite P, Kadane J. Statistical Methods for Eliciting Probability Distributions, Journal of the American Statistical Association. 2005;100(470):680-700.
12. Owen C. Parameter Estimation For The Beta Distribution, (Thesis Master of Science, Brigham Young University, Utah, USA), Available at https://scholarsarchive.byu.edu/etd/1614. consulted 15-9; c2018.
13. Raffo E, Mayta R, Perez V. Numerical integration algorithms, Department of Systems and Computer Engineering. UNMSM. 2007;10(2).
14. Tovar J. Eliciting Beta prior distributions for binomial sampling. Brazilian magazine of Biometrics. 2012;30(1):159-172.
15. Tovar J. Bayesian Inference and Health Research: a case of application in diagnosis Clinician, Medical Journal of Risaralda. 2015;21(1):9-16.
16. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian Data Analysis, Third Edition, Taylor & Francis; c2014.