



# Bayesian Estimation of the Shape Parameters of Mcdonald Generalized Beta-Binomial Distribution

Immaculate Kathomi Murithi, Justin Obwoye Okenye, Ali Salim Islam, Ronald Waliaula Wanyonyi

Department of Mathematics, Egerton University, Egerton, Nakuru, Kenya

Email: [immacleo23@gmail.com](mailto:immacleo23@gmail.com), [justinoke69@gmail.com](mailto:justinoke69@gmail.com), [asislam54@yahoo.com](mailto:asislam54@yahoo.com), [ronaldwaliaula@gmail.com](mailto:ronaldwaliaula@gmail.com)

**How to cite this paper:** Murithi, I.K., Okenye, J.O., Islam, A.S. and Wanyonyi, R.W. (2020) Bayesian Estimation of the Shape Parameters of Mcdonald Generalized Beta-Binomial Distribution. *Open Access Library Journal*, 7: e6651.

<https://doi.org/10.4236/oalib.1106651>

**Received:** July 23, 2020

**Accepted:** August 22, 2020

**Published:** August 25, 2020

Copyright © 2020 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The paper explores and establishes a unique Bayesian framework for estimating three shape parameters of the McDonald generalized beta-binomial distribution. The mixture distribution is used in modelling overdispersed binomial data. Foundations of the framework have been enriched by knowledge of Bayesian statistics and Markov Chain Monte Carlo methods. A Metropolis within Gibbs Monte Carlo method to sample from the unknown posterior form of the distribution was used. The shape parameters ( $\alpha$ ,  $\beta$  and  $\gamma$ ) were assigned flat gamma priors to ensure equal probabilities for all the values. McDonald generalized beta-binomial variables were simulated with fixed shape parameters set at  $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$  respectively and samples generated were used to estimate the parameters, to evaluate if the method recovers estimates close to the true parameter values. Standard errors were also computed for the simulated data and real data. Further, credible regions and highest probability density intervals (HPD) were computed and their corresponding lengths. To evaluate the marginal posterior samples for every shape parameter generated trace plots presented, their respective correlation plots were also presented and the histograms to show the distributions assumed by every parameter. Bayesian framework provides a direct and flexible method of computation for a mixture distribution whose complexity may pose challenges of integration when using the classical methods of estimation.

## Subject Areas

Statistics

## Keywords

Markov Chain Monte Carlo Methods, Bayesian Framework, Credible Regions, Highest Probability Density

## 1. Introduction

Data is an essential tool that is multidisciplinary in the world today because it informs on the decision making process of all the subjects that exist. Therefore, understanding the data as well as modelling this data to extract useful information is a key step to achieving a good decision. The confounding aspects of data also form an integral part in extracting information from a dataset, they are preliminary aspects that should be identified in a dataset. For instance presence of over dispersion [1]. With the changes that are happening in the world today like technology advancement and new industrial era, big volumes of data are generated on daily basis [2]. Therefore, models of higher dimension are becoming more useful in modelling such data. One characteristic of such data is the diversity that it presents in its features and therefore controlling such data requires flexible modelling techniques [3]. Bayesian framework for example, solves such problems through its unique nature of incorporating prior information about the parameters that is, allowing the parameters to vary as well and quantifying this variation using probability distributions [4]. For this reason, Bayesian modelling has been adopted in areas of artificial intelligence, bioinformatics, agriculture, and economics among many others [5]. Data in the form of proportions or count is not an exception to the emerging era of big data. It is often encountered in many scientific and social fields. One common feature that may be noticed from such data is over or under dispersion [6]. Over or under dispersion can be as result of variation in the success probability which in the case of a simple binomial model is treated as a constant but in reality, is usually not the case. Therefore, mixture distributions are developed to capture over dispersion where the mixing distribution is defined on the  $[0, 1]$  interval due to the property that the success probability should range on this interval. McDonald Generalized Beta-Binomial distribution (McGGB) has been proven to be superior in modelling overdispersed binomial data by [7]. The maximum likelihood estimation procedure was applied by [7] and [8] introduced the estimating equations for the McGGB distribution. Previously, estimation of the McGGB parameters using the Bayesian framework has not been addressed. Therefore, in this paper we develop a Bayesian framework through a Monte Carlo simulation technique specifically the Metropolis-Hasting step within Gibbs sampler to obtain the marginal posterior samples of the parameters of the McGGB, which will in turn help in Bayesian inference in particular estimation of the estimates of the parameters. The first part of this paper introduces the reader to the topic, the second part covers methodology which is subdivided into sections, the third part presents results and the fourth part gives a conclusion. In the paper different symbols have been used to identify different parameters, the symbols are defined as:

$\alpha$ ,  $\beta$  and  $\gamma$ : Shape parameters for McGGB distribution, which the paper wishes to estimate using the Bayesian framework.

$\theta$ : A notation that represents the parameter space, which contains the three parameters named ( $\alpha$ ,  $\beta$  and  $\gamma$ ).

$\pi_1(\alpha)$ ,  $\pi_2(\beta)$  and  $\pi_3(\gamma)$ : notation for prior distribution of each shape parameter.

$R_\varrho$ : A notation of the metropolis hastening ratio.

$\theta^{prop}$ : Joint proposal distribution used for the metropolis hastening algorithm.

$\left(\alpha_{\lfloor \psi M/2 \rfloor}, \alpha_{\lfloor (1-\psi/2)M \rfloor}\right)$ : Credible region showing upper and lower limits for the shape parameter  $\alpha$ .

$\left(\beta_{\lfloor \psi M/2 \rfloor}, \beta_{\lfloor (1-\psi/2)M \rfloor}\right)$ : Credible region showing upper and lower limits for the shape parameter  $\beta$ .

$\left(\gamma_{\lfloor \psi M/2 \rfloor}, \gamma_{\lfloor (1-\psi/2)M \rfloor}\right)$ : Credible region showing upper and lower limits for the shape parameter  $\gamma$ .

## 2. Methodology

In the methodology section, brief discussions of the distribution, the simulation algorithm and Bayesian method are discussed. This ensures that the reader is able to follow through each step of what forms the entire part of developing the Bayesian framework.

### 2.1. McDonald Generalized Beta Binomial Distribution

The McGGB is a mixture of distribution that is used to capture and model over-dispersion in binomial data. The mixture distribution is obtained by mixing the binomial distribution and the McDonald generalized beta distribution of the first kind [7]. A random variable  $X$  follows a MCGBB with parameters  $(n, \alpha, \beta, \gamma)$  if the probability mass function of  $X$  is expressed as:

$$f_{\text{McGGB}}(x | n, \alpha, \beta, \gamma) = \binom{n}{x} \frac{\gamma}{B(\alpha, \beta)} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} B(x + \alpha\gamma + \gamma i, n - x + 1)$$

For  $\alpha > 0, \beta > 0, \gamma > 0$ .

where,  $B(w, z) = \frac{\Gamma(w)\Gamma(z)}{\Gamma(w+z)}$  is the beta function.

An alternative form of the probability mass function of  $X$  is given as:

$$f_{\text{McGGB}}(y; n, \alpha, \beta, \gamma) = \binom{n}{y} \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{n-y} (-1)^j \binom{n-y}{j} B\left(\frac{y}{\gamma} + \alpha + \frac{j}{\gamma}, \beta\right)$$

For  $\alpha > 0, \beta > 0, \gamma > 0$ . Source [7].

Where the parameters  $\alpha, \beta$  and  $\gamma$  are the shape parameters of the distribution that can only assume positive values.

The properties of the distribution have been discussed in detail by [7], this paper primary attention is given to the discussion of the Bayesian framework to estimate the parameters of the distribution.

### 2.2. Simulation of McDonald Generalized Beta-Binomial Variables

Simulation of the McGGB variables has not been implemented using the following algorithm. The algorithm provides a direct approach to simulate directly from a

McGGBB distribution while being in control of the parameters. The variables obtained are McGGBB variables as opposed to the algorithm suggested by [7] whose variables were over dispersed random binomial variables:

*Step 1:* Setting fixed values of  $\alpha$ ,  $\beta$  and  $\gamma$ .

*Step 2:* Generate  $K$  random variables from Beta ( $\alpha$ ,  $\beta$ ) (i.e.  $U_i \sim \text{BETA}(\alpha, \beta)$ ) for  $i = 1, \dots, K$ .

*Step 3:* For each of the random variables that is  $P_i = U_i^{\frac{1}{\gamma}}$  for  $i = 1, \dots, K$ .

*Step 4:* For each  $P_i$  in step 3 generate binomial random variables that is:  $X_i \sim \text{Bin}(n, P_i)$ , then  $X_i \sim \text{McGGBB}(n, \alpha, \beta, \gamma)$  for  $i = 1, \dots, K$ .

### 2.3. Bayesian Method

Prior information is what makes Bayesian framework a unique method of parameter estimation as it provides a framework through which expert opinion that may have been obtained from previous studies is incorporated to the study [4]. Bayesian theory demands that the parameters of a distribution be treated random variables and thus the prior distribution is the main way in which such information (beliefs) is quantified [9]. The first step of Bayesian analysis is to define a probability distribution that is best suited to model a given dataset. The second step is to choose appropriate prior distributions for the parameters that characterizes the distribution selected. The choice of prior distributions is mostly guided by intuition and information that exists and is known about these parameters [10]. The methodology of choosing a prior distribution is however criticized by the classical statisticians since it is subjective as it is guided by intuitive knowledge which often lead to the use of informative priors. However, to circumvent this limitation, theory has highlighted the method of non-informative priors [11]. Non-informative prior distributions provide no information or provide an equal chance for all the possible parameters values in the parametric space before the data is observed [12]. The subject of obtaining the non-informative priors is the current in thing in Bayesian research, but a commonly used non-informative prior is the flat or diffuse priors. In this paper flat priors were used. In the Bayesian framework inferences are made based on the marginal posterior distributions of the parameters. However, a closed form of the joint posterior distribution may not be feasible thus making it more challenging to sample from such a distribution. Further, in cases of high dimensional parametric spaces, the integrations of such joint posterior distributions become intractable and complex and thus not easy to obtain the marginal posterior distributions [13]. In order to avoid the problem of intractability problem, the most commonly used methods are the Markov Chain Monte Carlo methods (MCMC) [14]. With these methods it is possible to obtain a samples from the marginal posterior distributions of the parameters from the joint without performing the integrations [15]. In this paper the Bayesian framework developed utilized the MCMC methods, specifically the use of a Metropolis-Hasting step within the Gibbs sampling.

In the Bayesian framework, the unknown parameters of the model were assumed to be random variables. Thus, there was a need to make appropriate assumptions about the distributions (prior distributions) of unknown parameters. The McGBB distribution has three shape parameters  $\underline{\theta} = (\alpha, \beta, \gamma)^T$  which we are interested in estimating, the study assumed the prior distribution of  $\alpha$ ,  $\beta$  and  $\gamma$  had jointly a flat prior, which were represented as:

$$\pi(\underline{\theta}) \propto \pi(\alpha, \beta, \gamma) \propto 1$$

The joint posterior distribution of  $\underline{\theta} = (\alpha, \beta, \gamma)$  was obtained by multiplying the conditional distribution  $f(\underline{y} | \underline{\theta})$  (essentially the likelihood function) with  $\pi(\underline{\theta})$  the joint prior distribution of  $\alpha$ ,  $\beta$  and  $\gamma$ . Let  $\underline{y} = y_1, y_2, \dots, y_N$  be a random sample of size  $N$  from a McGBB distribution with unknown parameter vector  $\underline{\theta} = (\alpha, \beta, \gamma)^T$ . The conditional distribution of  $f(\underline{y} | \underline{\theta})$  is obtained as:

$$\begin{aligned} f(\underline{y} | \underline{\theta}) &= f(y_1, y_2, \dots, y_N | \alpha, \beta, \gamma) = \prod_{k=0}^N f(y_k | \alpha, \beta, \gamma, n) \\ &= \prod_{k=1}^N \binom{n}{y_k} \times \frac{1}{B(\alpha, \beta)} \times \sum_{j=0}^{n-y_k} (-1)^j \binom{n-y_k}{j} \times B\left(\left(\frac{y_k}{\gamma} + \alpha + \frac{j}{\gamma}\right), \beta\right) \end{aligned}$$

The posterior distribution becomes:

$$\begin{aligned} \pi(\underline{\theta} | \underline{y}) &\propto f(\underline{y} | \underline{\theta}) \times \pi(\underline{\theta}) \\ &\propto \prod_{k=1}^N \binom{n}{y_k} \times \frac{1}{B(\alpha, \beta)} \times \sum_{j=0}^{n-y_k} (-1)^j \binom{n-y_k}{j} \times B\left(\left(\frac{y_k}{\gamma} + \alpha + \frac{j}{\gamma}\right), \beta\right) \times 1 \end{aligned}$$

It is evident that sampling from this joint posterior distribution is complicated, thus the study employed MCMC methods and in particular, the Metropolis Hasting step within the Gibbs sampling technique. In order to implement this algorithm, computation of the full conditional distributions for the parameters was obtained as follows:

$$\begin{aligned} \pi_\alpha(\alpha | \beta, \gamma, \underline{y}) &= \frac{\pi(\alpha, \beta, \gamma | \underline{y})}{\pi(\beta, \gamma | \underline{y})} = \frac{\pi(\alpha, \beta, \gamma | \underline{y})}{\int \pi(\alpha, \beta, \gamma | \underline{y}) d\alpha} \propto \pi(\alpha, \beta, \gamma | \underline{y}) \\ \pi_\beta(\beta | \alpha, \gamma, \underline{y}) &= \frac{\pi(\alpha, \beta, \gamma | \underline{y})}{\pi(\alpha, \gamma | \underline{y})} = \frac{\pi(\alpha, \beta, \gamma | \underline{y})}{\int \pi(\alpha, \beta, \gamma | \underline{y}) d\beta} \propto \pi(\alpha, \beta, \gamma | \underline{y}) \\ \pi_\gamma(\gamma | \alpha, \beta, \underline{y}) &= \frac{\pi(\alpha, \beta, \gamma | \underline{y})}{\pi(\alpha, \beta | \underline{y})} = \frac{\pi(\alpha, \beta, \gamma | \underline{y})}{\int \pi(\alpha, \beta, \gamma | \underline{y}) d\gamma} \propto \pi(\alpha, \beta, \gamma | \underline{y}) \end{aligned}$$

Then the metropolis within Gibbs sampling algorithm involved the following steps:

*Step 1.* Start with  $j = 1$  and the initial values of  $\{\underline{\theta}^{(1)} = (\alpha^{(1)}, \beta^{(1)}, \gamma^{(1)})\}$ .

*Step 2.* Using the proposal distributions of  $\underline{\theta}$ , where the proposal for the parameters was chosen as  $\alpha \sim \text{Normal}(\alpha^{(k-1)}, \sigma_\alpha^2)$ ,  $\beta \sim \text{Normal}(\beta^{(k-1)}, \sigma_\beta^2)$ ,  $\gamma \sim \text{Exponential}(\lambda)$  sample a candidate value for  $\underline{\theta}^{prop}$ .

*Step 3.* Generate  $U$  from a Uniform (0, 1) distribution (*i.e.*  $u \sim UNIF(0, 1)$ ).

*Step 4.* Calculate the Metropolis-Hasting (MH) ratio at the candidate value  $\underline{\theta}^{prop}$  and the previous value  $\underline{\theta}^{(k-1)}$ , using block updating.

$$R_{\underline{\theta}} = \frac{\pi_{\underline{\theta}}(\underline{\theta}^{prop} | \underline{y}) \times q_{\underline{\theta}}(\underline{\theta}^{(k-1)} | \underline{\theta}^{prop})}{\pi_{\underline{\theta}}(\underline{\theta}^{(k-1)} | \underline{y}) \times q_{\underline{\theta}}(\underline{\theta}^{prop} | \underline{\theta}^{(k-1)})}$$

*Step 5:* If  $u \leq \min(1, R_{\underline{\theta}})$ , accept the candidate point with probability  $\min(1, R_{\underline{\theta}})$ , i.e., set  $\underline{\theta}^{(k)} = \underline{\theta}^{prop}$ . Otherwise set  $\underline{\theta}^{(k)} = \underline{\theta}^{(k-1)}$ . Therefore for all  $j = 1, 2, \dots, M$  a sample of size  $M$  is obtained as the joint posterior distribution  $\{\underline{\theta}_j = (\alpha_j, \beta_j, \gamma_j), j = 1, 2, \dots, M\}$  sample.

Let  $B_0$  be the burn-in period for the markov chains for the parameters then under squared error loss function of the Bayesian estimates parameters were obtained as the mean of the samples generated using the algorithm above, i.e.,

$$\hat{\alpha} = E(\alpha | \underline{y}) = \frac{1}{M - B_0} \sum_{j=B_0+1}^M \alpha_j$$

$$\hat{\beta} = E(\beta | \underline{y}) = \frac{1}{M - B_0} \sum_{j=B_0+1}^M \beta_j$$

$$\hat{\gamma} = E(\gamma | \underline{y}) = \frac{1}{M - B_0} \sum_{j=B_0+1}^M \gamma_j$$

The  $100(1-\psi)\%$  Bayesian Credible Intervals and  $100(1-\psi)\%$  Highest Probability Density (HPD) intervals for  $\alpha$ ,  $\beta$  and  $\gamma$  were obtained using the algorithm proposed by [16]. The algorithm packaged in the package *CODA* in *R* language. Let  $\{(\alpha_{(j)}, \beta_{(j)}, \gamma_{(j)}), j = 1, 2, \dots, M\}$  be an ordered sample corresponding to the MCMC chain  $\{(\alpha_j, \beta_j, \gamma_j), j = 1, 2, \dots, M\}$  obtained using the algorithm above.

Then the approximate  $100(1-\psi)\%$  Bayesian Credible Intervals for  $\alpha$ ,  $\beta$  and  $\gamma$  were obtained as:  $(\alpha_{([\psi M/2])}, \alpha_{([(1-\psi/2)M] ]})$ ,  $(\beta_{([\psi M/2])}, \beta_{([(1-\psi/2)M] ]})$  and  $(\gamma_{([\psi M/2])}, \gamma_{([(1-\psi/2)M] ]})$  respectively.

## 2.4. Real Dataset

The framework was also applied to a real dataset and point estimates obtained. The dataset is secondary alcohol dataset which has also been used in literature by [7] and [8] for illustration. This dataset is documented in the *fitOD-BOD* package in *R* language. It consist of data collected in Netherlands for self-reported alcohol consumption frequencies from 399 randomly selected sample for two independent weeks. When number of days a respondent consumes alcohol out of 7 days is treated as a binomial random variable, traits of over-dispersion is portrayed from the variations of different individuals need to consume alcohol.

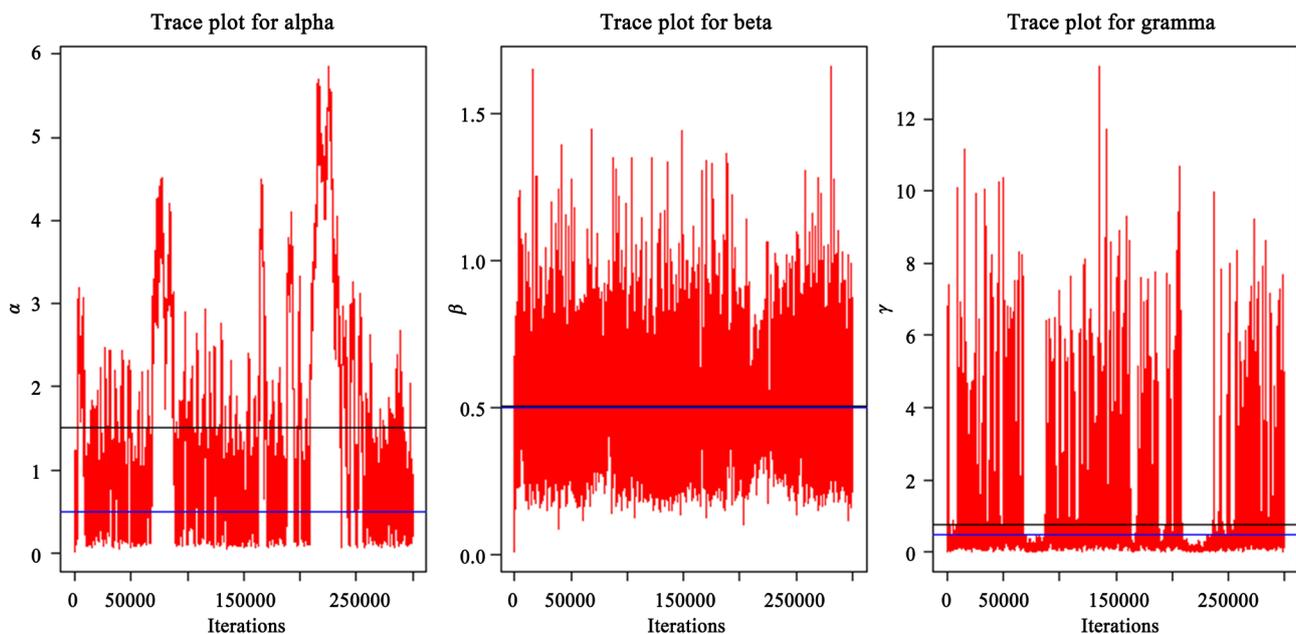
## 3. Results and Discussion

A small sample of size  $k = 25$  and a large sample of size  $k = 1000$  was simu-

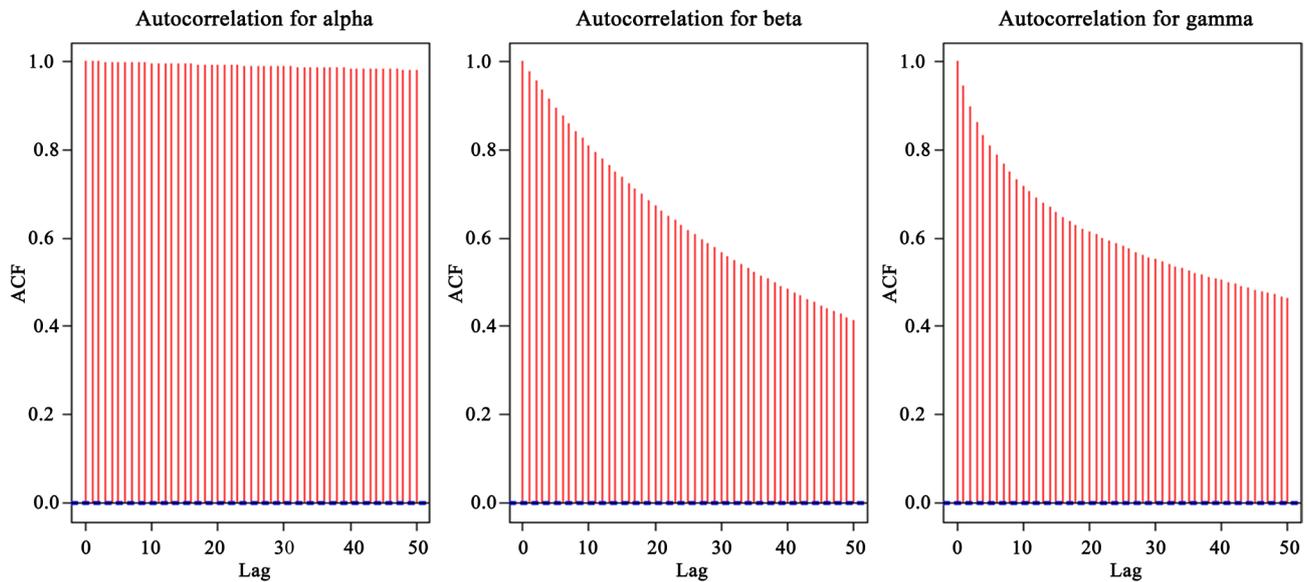
lated from the McGBB distribution while fixing the shape parameters at  $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$  respectively and applied to the Bayesian framework to get estimates of these parameters (essentially to recover the fixed parameters mentioned). The results obtained are presented using visuals of trace plots to show the behaviour of the samples generated, histogram plots to show the distribution or the unique shapes that the parameter estimates assume, autocorrelation plots, trace plots for means and over dispersion and the estimates presented in a table.

**Figure 1** shows the different chains generated from the posterior sampling when the sample size  $k = 25$  which was generated by setting the true values at  $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$  respectively. The starting points were set at  $(0.1, 0.1, 0.1)$  for the parameters  $(\alpha, \beta, \gamma)$ . The values of sigma for the proposal distributions were set at  $(\alpha, \beta, \gamma) = (0.1, 0.1, 0.5)$ . The first ten thousand values of three hundred thousand iterations were discarded as the burn-in-period. The blue horizontal line shows the true parameter value while the black horizontal line shows the parameter estimate that was given by the mean value for every chain. The horizontal lines on the plot for  $\alpha$  shows a bigger difference between the true parameter value and the estimate. However, for  $\beta$  and  $\gamma$  the horizontal lines are close to each other which is an indication that the estimates were relatively close to the true value. From the figure it can be seen that there was a higher serial correlation for  $\alpha$  and  $\gamma$  than it was for  $\beta$ . To see the autocorrelation behaviour, autocorrelation plots are presented in **Figure 2**.

From the plots, there was high correlation in  $\alpha$  values which is almost close to 1; confirming the behaviour of the trace plot for  $\alpha$ . The parameter  $\beta$  and  $\gamma$  show a decreasing correlation as more iterations are performed. Beta had the lowest correlation of all the parameters which is also evident from its trace plot.



**Figure 1.** Trace plots for the three shape parameters of the McGBB when  $k = 25$ .



**Figure 2.** Autocorrelation plots for  $\alpha$ ,  $\beta$  and  $\gamma$  when  $k = 25$ .

**Figure 3** shows the different histogram plots for the shape parameters. This shows the shape of the distribution that each of the parameter takes from the iterations performed. From the figure the histogram for  $\alpha$  and  $\gamma$  are skewed to the left with  $\gamma$  being the most skewed.  $\beta$  has a relatively normal curve.

**Figure 4** shows the trace plots for each mean computed at every parameter estimate obtained from the chains generated and a plot for the over-dispersion parameter. The blue horizontal line shows on average the mean and over-dispersion parameter across the chains. The mean was a value close to 0.7 and the over dispersion of a value close to 0.5. For values of over dispersion close to one indicate high over-dispersion while values of row close to zero indicate low over-dispersion. From the plot it can be seen that over-dispersion was present and it was relatively on the average.

**Figure 5** shows the different chains generated from the posterior sampling when a sample of size  $k = 1000$ , with true values of  $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$  respectively. The starting points were set at  $(0.1, 0.1, 0.1)$  for the parameters  $(\alpha, \beta, \gamma)$ . The values of sigma for the proposal distribution were set at  $(\alpha, \beta, \gamma) = (0.1, 0.01, 5)$  respectively. The black horizontal line shows the true parameter value while the blue horizontal line shows the parameter estimate that was given by the mean value for every chain. From the figure, the plots of  $\alpha$  and  $\gamma$  it can be seen that autocorrelation was high and low for  $\beta$  plot.

From **Figure 6**, correlation was relatively average for the parameter  $\beta$ , and high for  $\alpha$  and  $\gamma$ . This can be seen from the behaviour of the traceplots. Comparing the autocorrelation plots for when sample size is  $k = 25$ , it can be seen that autocorrelation has increased for sample size  $k = 1000$ .

From **Figure 7**, the shape of  $\alpha$  is skewed so is  $\gamma$ . However,  $\gamma$  is more skewed in shape than  $\alpha$ .  $\beta$  approaches a normal distribution shape. It can be noticed that with the increase in sample size the shapes are becoming flatter compared to when the sample size was  $k = 25$ .

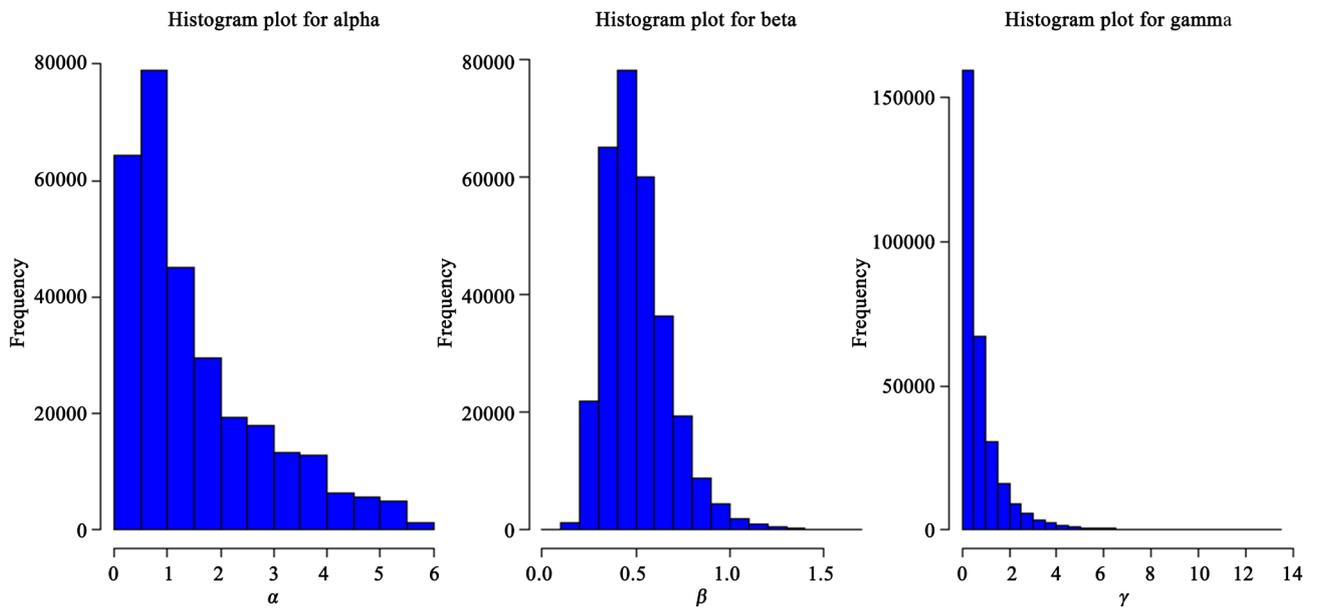


Figure 3. Histogram plots for the parameters when  $k = 25$ .

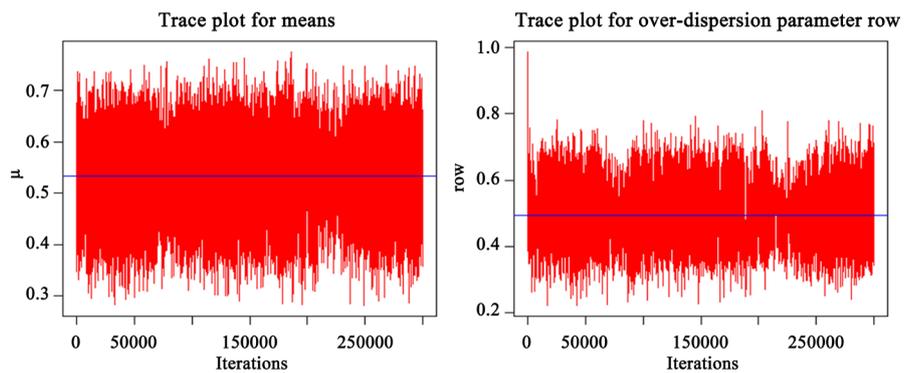


Figure 4. Trace plots for means and over dispersion parameter  $\rho$  computed from the estimates generated.

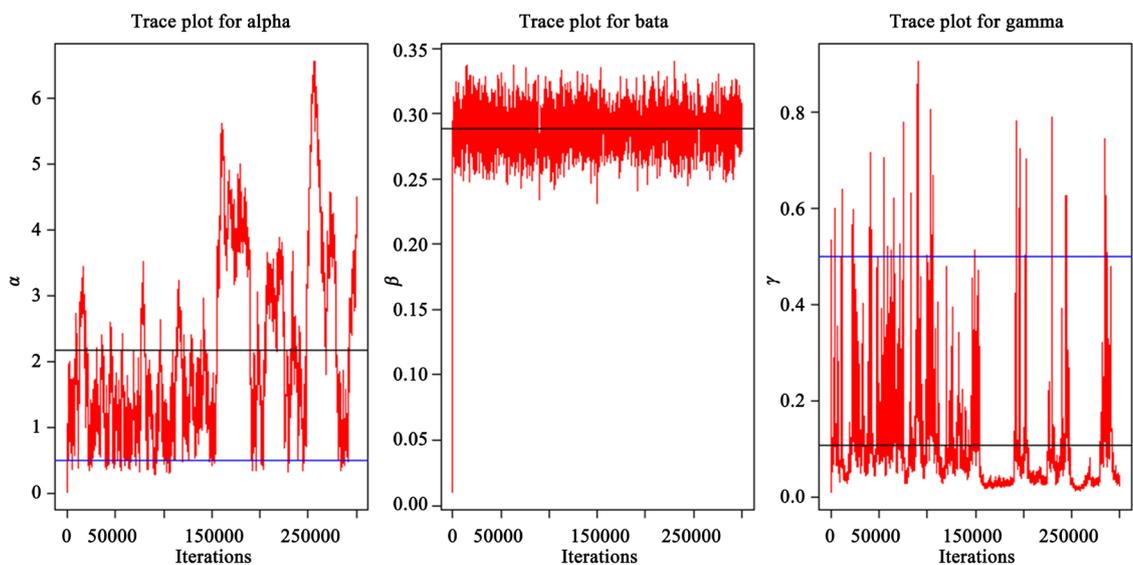


Figure 5. Trace plots when  $k = 1000$ .

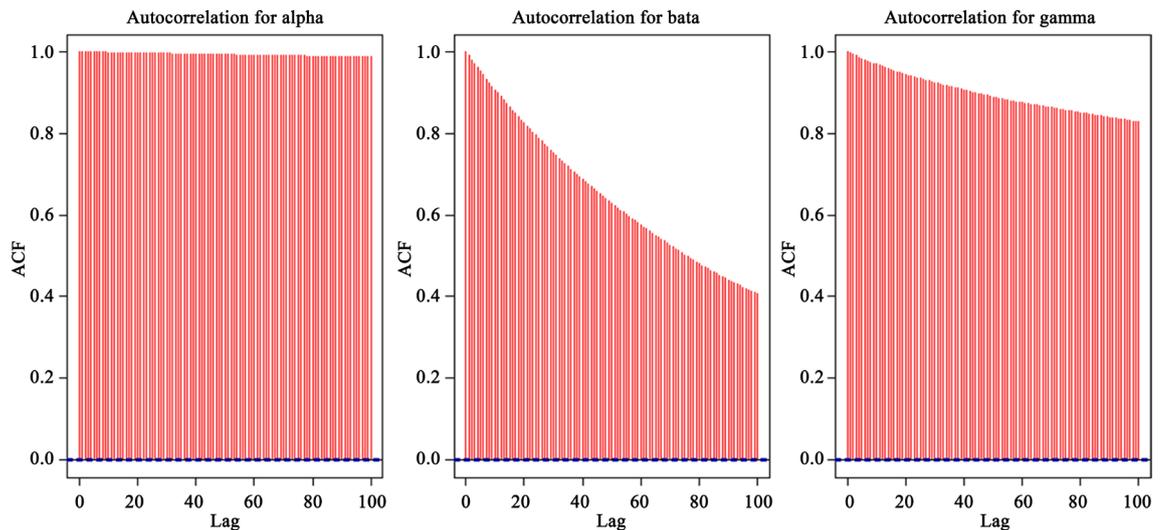


Figure 6. Correlation plots for  $\alpha$ ,  $\beta$  and  $\gamma$ .

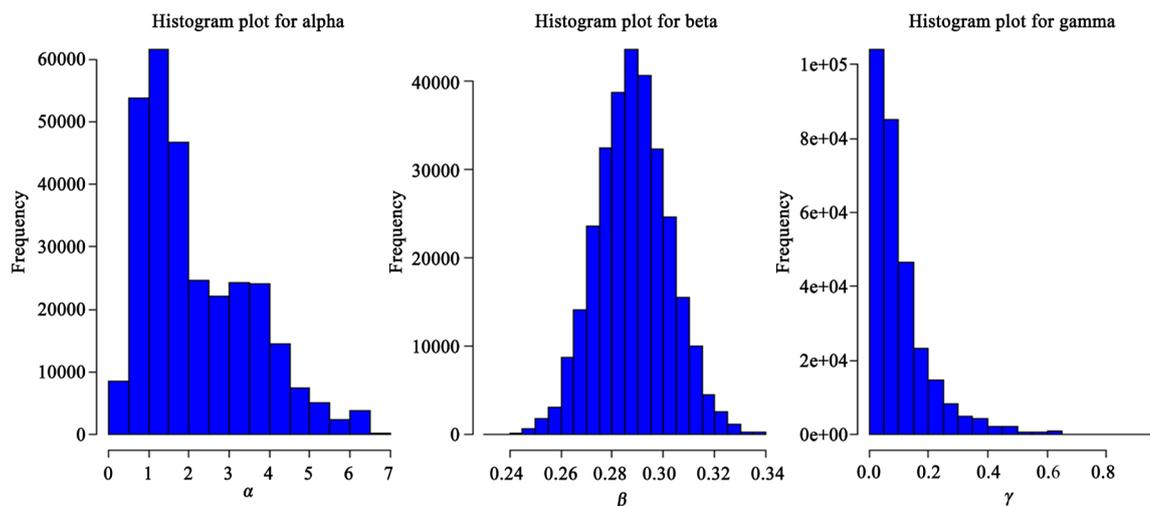


Figure 7. Histogram plots for  $\alpha$ ,  $\beta$  and  $\gamma$ .

Figure 8 shows trace plots for each mean computed at every parameter estimate obtained from the chains generated and a plot for the over-dispersion parameter. The blue horizontal line shows on average the mean and over-dispersion parameter across the chains. When the sample size is increased it can be noticed that there was an increase in over dispersion as well. On average for every combination of estimates computed the mean was 0.50. The average over dispersion was about 0.65 for every combination of the estimates computed across three hundred thousand iterations.

From Table 1 the estimates were obtained from the posterior mean with true values of the three parameters set at  $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$  respectively. The first one thousand values sampled were discarded as the burn-in-period. The acceptance rates of the sampled candidate values were relatively within the acceptable limits for small samples however, relatively declined for large samples. The acceptable limits of the acceptance rates should be between 13% - 50% [5].

Acceptance rates show the percentage of the values from the specified iterations which in this case was three hundred thousand that mimic the posterior distribution. The estimates for the parameter  $\beta$  and  $\gamma$  were close to the true parameter values. However, the parameter estimates of  $\alpha$  were not as close as  $\beta$  and  $\gamma$  to the true parameter value, but was not as far from the true value as much. The Bayesian framework produced estimates that were close to the true parameter values for small samples.

**Table 2** shows the standard errors for the estimates obtained. Standard errors shows how precise the estimates are. From the table the parameter  $\alpha$  had higher standard errors, which is also reflected in the trace plots figure where the variability is high. The parameter  $\beta$  had lower standard errors, which is evident from the low variation displayed in the trace plot figures while  $\gamma$  had relatively low standard errors.

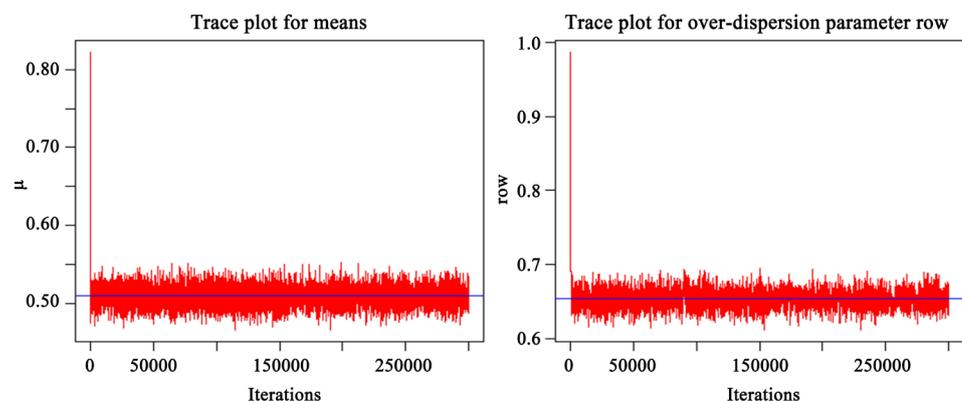
From **Table 3**, all the credible regions for all the parameters contained the true parameter value. Therefore, there is a 95% chance that the true value of each of the parameters  $(\alpha, \beta, \gamma)$  will lie within their respective credible regions computed.

**Table 4** shows the confidence lengths which were short especially for the parameter  $\beta$ . This further stamps the preciseness of the Bayesian framework in the estimation of the McGBB parameters.

**Table 5** shows the highest posterior densities for the parameters. The HPD is the shortest credible interval among all the intervals. Any point within the HPD has a higher density than any other point outside [17].

The article also uses a real dataset to apply the Bayesian framework. **Table 6** shows the estimates that were obtained when the Bayesian framework was applied for the alcohol dataset. Other methods of estimation by [7] and [8] have been applied for the same dataset. While the primary attention of this article was to come up with a Bayesian framework, readers can compare the estimates for the alcohol dataset with those obtained by [7] and [8].

**Table 7** shows the standard errors of the estimates when a real dataset was used. From the table the standard errors were high for  $\alpha$  which could have been contributed by the high variation portrayed in  $\alpha$ .



**Figure 8.** Trace plots for means and over dispersion parameter.

**Table 1.** Parameter estimates for  $\alpha$ ,  $\beta$  and  $\gamma$ .

Sample size ( $k$ )	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	Acceptance rate
25	1.507	0.502	0.792	15.7%
1000	2.178	0.288	0.1072	4.8%

**Table 2.** Standard errors of the estimates.

Sample size ( $k$ )	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	1.4014	0.1699	0.9246
1000	1.9704	0.01403	0.0801

**Table 3.** Bayesian credible intervals.

Sample size ( $k$ )	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	(0.1853, 4.8902)	(0.2532, 0.9051)	(0.0586, 3.5261)
1000	(0.02006, 0.9974)	(0.2712, 0.5567)	(0.0039, 1.67005)

**Table 4.** Lengths of the credible regions.

Sample size ( $k$ )	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	1.4769	0.5855	0.4783
1000	0.9773	0.2855	1.6697

**Table 5.** Bayesian HPD interval.

Sample size ( $k$ )	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	(0.0824, 4.2585)	(0.2134, 0.8355)	(0.0082, 2.6365)
1000	(0.0145, 0.8497)	(0.2712, 0.5566)	(0.001005, 1.3652)

**Table 6.** Application to a real dataset (Alcohol dataset).

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Alcohol week 1	0.1038	0.3031	21.8814
Alcohol week 2	0.0793	0.3619	32.2472

**Table 7.** Standard Errors for the Estimates of the Real data.

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Alcohol week 1	9.9839	0.00732	0.00832
Alcohol week 2	10.1964	0.00873	0.00425

## 4. Conclusion

This article explores a Bayesian framework for the mixture distribution known as McDonald generalized beta-binomial distribution. From the framework, the paper has been able to obtain the point estimates, credible regions and HPD intervals. It is evident that the point estimates obtained do not deviate from the true parameter which can also be seen from the standard errors obtained. The credible regions obtained are of short lengths and all include the true parameter value, which further emphasizes on the precision of the Bayesian framework. This paper emphasizes the importance of the Bayesian framework on modelling mixture models or distributions. The use of Bayesian framework for estimation is preferably good especially for a mixture distribution that is susceptible to problems of integration. The ability of the framework to overcome the challenge of intractable integrations especially in high-dimensional distributions by applying a random search procedure to obtain estimates makes it a preferable method. McGBB is one of the many beta-type generated distributions, this work can be extended to other distributions of the same class. Moreover, future studies can explore the Bayesian methodology by using informative priors as opposed to non-informative that were used in this study.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Stoffel, M.A., Nakagawa, S. and Schielzeth, H. (2017) Repeatability Estimation and Variance Decomposition by Generalized Linear Mixed-Effects Models. *Methods in Ecology and Evolution*, **8**, 1639-1644. <https://doi.org/10.1111/2041-210X.12797>
- [2] Coveney, P., Dougherty, E. and Highfield, R. (2016) Big Data Need Big Theory Too. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **374**, 1-11. <http://www.jstor.org/stable/26115982>  
<https://doi.org/10.1098/rsta.2016.0153>
- [3] Richterich, A. (2018) Big Data: Ethical Debates. In: *The Big Data Agenda: Data Ethics and Critical Data Studies*, University of Westminster Press, London, 33-52. <https://doi.org/10.2307/j.ctv5vddsw>
- [4] Petzschner, F.H., Glasauer, S. and Stephan, K.E. (2015) A Bayesian Perspective on Magnitude Estimation. *Trends in Cognitive Sciences*, **19**, 285-293. <https://doi.org/10.1016/j.tics.2015.03.002>
- [5] Entezari, R. (2018) Bayesian Computations via MCMC, with Applications to Big Data and Spatial Data. Doctoral Dissertation.
- [6] Zhang, C. (2019) Statistical Modeling of Count Data with Over-Dispersion or Zero-Inflation Problems.
- [7] Manoj, C., Wijekoon, P. and Yapa, R.D. (2013) The McDonald Generalized Beta-Binomial Distribution: A New Binomial Mixture Distribution and Simulation Based Comparison with Its Nested Distributions in Handling over Dispersion. *International Journal of Statistics and Probability*, **2**, 24. <https://doi.org/10.5539/ijsp.v2n2p24>

- 
- [8] Janiffer, N., Islam, A. and Luke, O. (2014) Estimating Equations for Estimation of McDonald Generalized Beta-Binomial Parameters. *Open Journal of Statistics*, **4**, 702-709. <https://doi.org/10.4236/ojs.2014.49065>
- [9] Dodwell, T.J., Ketelsen, C., Scheichl, R. and Teckentrup, A.L. (2015) A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. *SIAM/ASA Journal on Uncertainty Quantification*, **3**, 1075-1108. <https://doi.org/10.1137/130915005>
- [10] Lee, M.D. and Vanpaemel, W. (2018) Determining Informative Priors for Cognitive Models. *Psychonomic Bulletin & Review*, **25**, 114-127. <https://doi.org/10.3758/s13423-017-1238-3>
- [11] Sprenger, J. (2018) The Objectivity of Subjective Bayesianism. *European Journal for Philosophy of Science*, **8**, 539-558. <https://doi.org/10.1007/s13194-018-0200-1>
- [12] Lynch, S.M. (2007) Introduction to Applied Bayesian Statistics and Estimation for Social Scientists. Springer Science & Business Media, Berlin. <https://doi.org/10.1007/978-0-387-71265-9>
- [13] Li, J., Nott, D.J., Fan, Y. and Sisson, S.A. (2017) Extending Approximate Bayesian Computation Methods to High Dimensions via a Gaussian Copula Model. *Computational Statistics & Data Analysis*, **106**, 77-89. <https://doi.org/10.1016/j.csda.2016.07.005>
- [14] Martino, S. and Riebler, A. (2019) Integrated Nested Laplace Approximations (INLA). <https://doi.org/10.1002/9781118445112.stat08212>
- [15] Alquier, P., Friel, N., Everitt, R. and Boland, A. (2016) Noisy Monte Carlo: Convergence of Markov Chains with Approximate Transition Kernels. *Statistics and Computing*, **26**, 29-47. <https://doi.org/10.1007/s11222-014-9521-x>
- [16] Le, H., Pham, U., Nguyen, P. and Pham, T.B. (2018) Improvement on Monte Carlo Estimation of HPD Intervals. *Communications in Statistics-Simulation and Computation*. <https://doi.org/10.1080/03610918.2018.1513141>
- [17] Grzenda, W. (2015) The Advantages of Bayesian Methods over Classical Methods in the Context of Credible Intervals. *Information Systems in Management*, **4**, 53-63.