BAYESIAN ESTIMATION OF THE MCDONALD GENERALIZD BETA –BINOMIAL DISTRIBUTION PARAMETERS

IMMACULATE KATHOMI MURITHI

A Thesis Submitted to the Graduate School in Partial Fulfilment of the Requirements for the of Master of Science Degree in Statistics of Egerton University

EGERTON UNIVERSITY

JULY, 2021

DECLARATION AND RECOMMENDATION

Declaration

This thesis is my original work and has not been submitted for examination in any institution.

Signature: ____________________ Date: 24/06/2021

Immaculate Kathomi Murithi
SM123/12105/17

Recommendation

This thesis has been submitted with our approval as Egerton University supervisors.

Signature: ____________________ Date: 24/06/2021

Prof. Ali Salim Islam, PhD.
Department of Mathematics.
Egerton University.

Signature: ____________________ Date: 24/06/2021

Dr. Ronald Waliaula Wanyonyi, PhD.
Department of Mathematics.
Egerton University.

COPYRIGHT

DEDICATION

I dedicate this research work to my father Eston Eston Murithi, mother Elosy Murugi and my late grandmother Esther Kagendo who have always believed in me and prayed for me throughout this journey.

ACKNOWLEDGEMENTS

First, I would like to thank God for his blessings and good health during my research. I would also like to appreciate my supervisors Prof Ali Salim Islam and Dr. Ronald Waliaula Wanyonyi for their selfless efforts, positive criticism and guidance in throughout the period I was preparing the dissertation. I would also like to acknowledge Mr Justin Obwoge Okenye for being very resourceful and for his consistent advice throughout the time I was preparing the dissertation.

I would also like to acknowledge my father Eston Murithi who has been very instrumental through his financial support and his words of encouragement every time. My mother for praying for me every day and having my back through her guiding words. I would also like to thank my colleagues Paul Kimani, Rafael Mweleli, Purity Kinyua and Birech Benjamin for their support through scholarly comments that were vital in the preparation of this document. To the many people who are not mentioned but contributed to the success of this work, your support, care and understanding is acknowledged. I am deeply indebted to you all, thank you.

ABSTRACT

McDonald Generalized Beta-Binomial (McGBB) distribution is a mixture distribution that is used in modelling over-dispersed binomial outcome data. McGBB model has been shown to be more flexible and superior in handling over-dispersion compared to its nested distributions; Kumaraswamy Binomial (KB) distribution and Beta-Binomial (BB) distribution. Previous studies have estimated the parameters of the McGBB distribution using the maximum likelihood Estimation (MLE) method and estimating functions method, but the Bayesian estimation method has not been implemented to estimate the parameters of McGBB distribution. This study explored and established a Bayesian framework for estimating three shape parameters of the McGBB. Further the study compared the point and interval estimates computed from the Bayesian framework to the maximum likelihood estimates. In order to implement the Bayesian framework prior distributions were assigned to the parameters, in this study the parameters (α, β and γ) were assigned flat priors to ensure equal probabilities for all the values. The form of the joint posterior distribution obtained was intractable hence the study used the Markov Chain Monte Carlo methods (MCMC), specifically Metropolis-Hasting step within Gibbs sampler method to sample from the unknown posterior distribution and obtain samples of the marginal posterior densities of each parameter. McDonald generalized beta-binomial variables were simulated with fixed shape parameters set at $(\alpha, \beta, \gamma)=(0.5, 0.5, 0.5)$ following a new algorithm developed based on the mixing distributions that make up the distribution respectively. Small samples of sizes $k= 25$ and $k= 29$ and large samples of sizes $k= 500$ and $k=1000$ were simulated. The Bayesian estimates for the parameter α had higher standard errors across all sample sizes, which was also reflected by the high variation in the trace plots, β had very low standard errors across all the sample sizes and γ varied with low and high standard errors depending on each sample size. Further, the Bayesian credible regions had relatively short widths, which all included the true parameter value. The study also compared the point and credible regions of the Bayesian method to the point and confidence intervals of the MLE method. The study found out that the credible regions had shorter widths compared to the widths of the confidence intervals of the MLE. The standard errors and short confidence lengths obtained from applying the Bayesian framework shows the preciseness of this method in estimating the parameters of McGBB. This framework can be extended to distributions in the same class of beta-type generated distributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

McGBB	McDonald Generalized beta-binomial distribution
KB	Kumaraswamy binomial distribution
BB	Beta-Binomial distribution
MLE	Maximum Likelihood Estimation
PMF	Probability Mass Function
HPD	Highest Probability Density
MCMC	Markov Chain Monte Carlo

LIST OF SYMBOLS

α	Shape parameter for McGBB distribution
β	Shape parameter for McGBB distribution
γ	Shape parameter for McGBB distribution
ρ	Over-dispersion parameter
$\int$	Shows the integral sign for integration
θ	Parametric space
μ	Population mean
Σ	Sum of all the observation
Ψ	Significance level in the interval estimates

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Data is an essential tool in the world today because it informs on the decision making process of all disciplines. Therefore, understanding the data as well as modelling this data to extract useful information is a key step to achieving a good decision. However, 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 (Blazquez & Domenech 2018). Therefore, models of higher dimension are becoming more useful in modelling such data. One characteristic of such data is the diversity that it presents and therefore controlling such diversities requires flexible modelling techniques (Iqbal *et al.,* 2020). 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 also and quantifying this information using probability distributions (Bolstad & Curran, 2016). For this reason, Bayesian methods have been adopted in areas of artificial intelligence, bioinformatics, agriculture, and economics among many others (Entezari, 2018).

Data in the form of proportions is not an exception to the emerging era of big data. It is often encountered in many scientific fields. For instance proportions or count data is encountered in toxicology, ecological studies (Paul & Islam, 1995), epidemiology (Griffins, 1973), media exposure (Greene, 1970) and other similar fields that involve biological behaviour (Crowder, 1978). Traditionally, the simple binomial model is used to model such proportions or count. However, it fails to capture the extra variation that is exhibited by the data. This extra variation is called Extra-Binomial variation or the data is said to be over dispersed. In order to address over dispersion several parametric models have been proposed such as correlated binomial (Kupper & Haseman, 1978), the additive and multiplicative binomial models (Altham, 1974), BB (Lee & Sabavala, 1987; Pack, 1986; Paul, 1982; Paul & Islam, 1995; Skellam, 1948) and McGBB distribution introduced by McDonald and Xu (1995). The superiority of the McGBB in modelling proportions in presence of extra binomial variation has been shown by Manoj *et al.* (2013).

Over-dispersion can be as a 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. Among the binomial mixture distributions, a three parameter distribution McGBB was found to have a better fit and an improved goodness of fit compared to its nested two parameter distributions Beta- Binomial (BB) and the Kuraswamy-Binomial (KB) distributions. The additional parameter of the McGBB distribution makes it possible to accommodate wide range of shapes more to the already incorporated shapes by its nested distributions, the BB and KB (Manoj *et al.,* 2013). McGBB distribution portrays superiority to its nested distribution in capturing the over/under-dispersion which is common in binomial outcome data (Janiffer *et al.,* 2014).

Estimation of parameters in statistical models can be done using the Bayesian approach, method of moments, least squares method, estimating equations method and the maximum likelihood estimation method among other method (Cumming & Callin, 2016). However, the method of maximum likelihood estimation is extensively applied in most estimation procedures (Kakamu & Nishino, 2018). It is desirable for its asymptotic properties, thus it maybe preferred when the sample sizes are large but inappropriate when the sample sizes are small (Tarima & Flournoy, 2019). Moreover, the regularity conditions may not hold for all cases and consequently the ML estimates may not be defined (Lee & Sabavala, 1987). Unconstrained ML estimation could, however, lead to negative estimates. Practically, it is rarely clear how large n should be for these asymptotic results to be reasonable. Moreover, when parameters are bounded, asymptotic interval estimates could fall outside the valid range. Bayesian approach is a statistical method of estimation that uses probability theory to provide some degree of belief in the data in question. Bayesian methods that have been developed using the Bayes rule initially introduced by Thomas Bayes, have gained popularity in the field of statistics because they are able to address most of the limitations experienced using the classical methods (Fienberg, 2006). The Bayesian framework allows the use of prior information in the estimation of parameters thus enhancing the knowledge available about the parameters (Stokes *et al.,* 2014). Moreover, the Bayesian estimation allows for a stochastic search process that allows us to examine the parametric space in such a way that we minimize the probability of being locked up in a local maximum in cases where the posterior form is intractable. Bayesian method farther allows a researcher to choose estimates either as mean, median or mode from the form of marginal posterior. However in some cases the Bayesian estimates may approach the maximum likelihood estimate as the sample size increases.

Estimation of the McGBB parameters has been addressed in the past but not in the Bayesian framework. The maximum likelihood estimation procedure by Manoj *et al.* (2013) and Janiffer *et al*. (2014) introduced the estimating equations for the McGBB distribution. Therefore, in addition to the available methods of estimating parameters there is need to consider other estimation methods, so as to evaluate which methods are more precise and computationally efficient for McGBB. This study used the Bayesian framework in the estimation of McGBB parameters. The study also illustrated its application using simulated data as well as real data set. Further, a comparison study was done for the Bayesian method of estimation and maximum likelihood method.

1.2 Statement of the Problem

McDonald generalized beta-binomial (McGBB) distribution is a three-parameter model that has been used to model over-dispersion in binomial data. McGBB is preferred as a mixture model for binomial data due to its flexibility in capturing over and under dispersion over its nested BB and KB distributions. The estimation of the shape parameters of the McGBB has been established through the maximum likelihood method and the estimating functions method. However, the MLE and estimating function estimators may not be found in closed form and therefore require numerical methods of computation, which do not guarantee convergence. Further, the MLE and estimating functions methods of estimation possess desirable asymptotic properties that perform well for large samples but may perform poorly for small samples. MLE method is intensive in computation and involves sampling many times so that you can be able to study the asymptotic properties of the estimates. The estimating functions methods and the MLE methods do not allow for incorporation of prior knowledge. Therefore, there was a need to use the Bayesian approach for three reasons; first, it enables researchers to incorporate prior information, secondly it provides a framework of making statistical inference regarding the parameters, through creation of credible sets and finally it provides a framework where in case integrations are intractable that is cannot be solved numerically, one can apply a Monte Carlo approach.

1.3 Objectives

1.3.1 General Objective

To establish a Bayesian estimation framework for the of the McGBB parameters for the purpose of modelling proportions data that exhibit over dispersion.

1.3.2 Specific Objectives

i. To evaluate the performance of point estimates for the Bayesian framework and maximum likelihood methods of estimation for small and large samples using standard errors.

ii. To compare the interval estimates using confidence intervals for the MLE and credible intervals for the Bayesian in small and large samples.

1.4 Assumptions of the Study

i. The parameters of the McGBB α, β, γ are random variables.

ii. The proposal distribution for generating the Metropolis hasting ratio for each of the parameters is normal for α and β, and exponential for the parameter γ.

1.5 Significance of the Study

This study will give more insights to researchers in disciplines that encounter proportions data that may exhibit over dispersion. Specifically, such data can be modelled using the McGBB distribution whose parameters can be estimated using the Bayesian framework. Studies show that McGBB is flexible and has a better goodness of fit compared to its nested distributions; BB and KB in handling over dispersion that arises from proportions data. Proportions data is commonly experienced in toxicology studies, ecological studies, teratology studies, medical studies and economics demand pattern studies.

In toxicology studies, where researchers in this field seek to study effect of toxins on humans, animals and environment, proportions data is one example of data that arises from this discipline. For example in a study to determine the level of toxic substances present in foods due to application of pesticides. In this study a specified number of foods are sampled from different farms. Therefore, the data will be in form of proportions as the characteristics of interest will be; the proportion of foods that portray the presence of toxins over the total foods sampled from each farm. Since the farms are subjected to different conditions that support growth and presence of pests, it is highly likely that this data will have traits of over dispersion. Therefore, McGBB distribution can be used to model the data and the Bayesian framework used in estimation of its parameter. Therefore, the study will equip researchers in the named fields with understanding of the Bayesian estimation of the McGBB distribution parameters, a distribution that is widely applicable in their areas of study.

1.6 Definition of Significant Terms

Parameter	These are characteristics that define and describe a population.
Population	This is the totality of all individuals under study
Statistic	This is a characteristic of a sample, it describes the sample
Sample	This is a portion of the population or a section of the population which a researcher aims to investigate its characteristics.
Shape parameter	It characterizes the outlook of the population. Its distortion only affects the shape of the distribution.
Scale parameter	This is a parameter that stretches or squeezes the distribution in which it is contained.
Location parameter	This type of parameter defines where the distribution is located, its distortion will cause the distribution to move to a ne point.
Over dispersion	The term refers to extra variation that is observed from a model, that is the variation that has exceeded the normal variation of a model as is defined in its properties.
Estimation	Specified procedure that computes the value of some property of the population.
Prior distribution	A quantification of beliefs that exist from expert knowledge, intuition or literature concerning certain parameters of interest.
Posterior distribution	It is a probability distribution that summarizes the prior beliefs given the observed data.
Credible regions	An interval where an unobserved parameter value falls with a particular value of specified probability.

Highest probability density intervals	This is the shortest interval of all the credible I intervals and is the most optimal interval.
Confidence interval	refers to the probability that a population parameter will fall between a set of values for a certain proportion of times

CHAPTER TWO

LITERATURE REVIEW

2.1 McDonald Generalized Beta Binomial Distribution

The McGBB 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 (Manoj *et al.,* 2013). The distributions making up the mixture are discussed below, however not in detail some of the properties and reason for using the distributions are highlighted.

2.1.1 Binomial Distribution

Binomial distribution is used to model experiments with binary outcomes for a fixed known number of trials (n). The possible outcomes are non-overlapping and mutually exclusive and therefore this type of distribution is discrete (Johnson *et al.,* 2005). The success probability in theoretical sense is normally treated as a constant. However, in real life application, the success probability may not portray as a constant but a random variable, leading to an exaggeration of the variance of the observed binomial outcome data (Manoj *et al.,* 2013). The nominal binomial distribution fails to capture a trait referred to as over or under dispersion. Hence, there is a need to mix the binomial distribution by allowing the parameter p to vary and quantifying it using an appropriate mixing distribution (Jannifer *et al.,* 2014).

The probability mass function of a binomial random variable V with success probability p is expressed as;

$$f_v(v|\, n, p)=\begin{cases}\binom{n}{v}\ p^v\,(1-p)^{n-v} & v=1, 2, \ldots, n,\ 0 \le p \le 1,\\ 0 & \text{elsewhere.}\end{cases} \tag{1}$$

A random variable V that follows a binomial distribution has the following properties:

Moment generating function: $M_v\,(t) = (p\, e^t \,+\, (1-\, p)\,)^n$ (2)

Mean: $E(V) = np$ (3)

Variance: $Var\,(V) = E(V - E(V)\,)^2 = np\,(1-p)$ (4)

2.1.2 McDonald Generalized Beta Distribution of the First Kind

The McDonald generalized beta distribution of the first kind is a three parameter distribution that is constructed from the generalization of the classical beta of type one (Otieno, 2008). The probability density function of McDonald generalized beta distribution of the first kind (Manoj *et al.*, 2013) is known by the expression;

$$f_U\,(u|\ \alpha,\beta,\gamma) = \frac{\gamma}{B(\ \alpha,\beta)} u^{\alpha\gamma\,-\,1}(\,1 -\ u^{\gamma}\,)^{\beta-1},\quad 0\ \le\ u\ \le\ 1, \alpha\ >\ 0, \beta\ >\ 0, \gamma\ >\ 0 \qquad (5)$$

The beta distribution is an appropriate prior to binomial density because it allows for different variety of shapes and it is mapped on the interval [0, 1], on which the success probability *p* should lie (Ogal, 2012). Generalized forms of the beta distributions are more flexible in beta-type of distributions due to the additional parameter compared to the classical beta distributions (Alexander *et al.,* 2012). The distribution can also be obtained from the transformation highlighted below; if a random variable U is a beta distributed variable with parameters(α,β), then $U^{\frac{1}{\gamma}}$ follows a GB1 distribution with parameters $(\alpha,\ \beta,\ \gamma)$.

$$f(u|\alpha,\beta) = \frac{1}{B(\alpha,\,\beta)} u^{\alpha\,-\,1}(1\,-\,u)^{\beta\,-\,1}, 0\ \le\ u\ \le\ 1 \qquad (6)$$

Let $p = u^{1/\gamma} \Leftrightarrow p^{\gamma} = u$.

The Jacobian of transformation is;

$$\left|\frac{du}{dp}\right| = \gamma p^{\gamma\text{-}1}$$

$$\Rightarrow f(p|\,\alpha,\beta,\gamma) = \frac{1}{B(\alpha,\beta)} p^{\gamma(\alpha\,-\,1)}(1 -\ p^{\gamma})^{\beta\,-\,1}(\gamma p^{\gamma\,-\,1})$$

$$= \frac{\gamma}{B(\alpha,\ \beta)} p^{\gamma\alpha\,-\,1}(1 -\ p^{\gamma})^{\beta\,-\,1}. \qquad (7)$$

If a random variable U follows a GB1 (α, β, γ) distribution, then the k^{th} moment as illustrated by Manoj *et al.* (2013) is expressed as;

$$E(u^k) = \frac{B\left(\alpha+\beta,\ \frac{k}{\gamma}\right)}{B\left(\alpha,\ \frac{k}{\gamma}\right)} \qquad (8)$$

The generalized beta distribution of the first kind is supported on a bounded domain and is useful in the modelling of size phenomena, particularly the distribution of income.

It is preferred compared to beta type of the second kind and this has been illustrated in its superiority in modelling income data in the Unites States for the years 1970, 1975 and 1980.the According to Manoj *et al.* (2013), random variable X follows a McDonald generalized beta binomial distribution with parameters (n, α, β, γ) if the probability mass function of X is known by the expression:

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

for $\alpha > 0, \beta > 0, \gamma > 0$. (9)

Where;

$B(w, z) = \frac{\Gamma(w)\Gamma(z)}{\Gamma(w+z)}$ is the beta function represented as a gamma function. However, beta function can also be expressed as;

$B(w, z) = \int_0^1 m^{(w-1)}(1 - m)^{(z-1)} dm$, or as a factorial as $B(w, z) = \frac{(w-1)!(z-1)!}{(w+z-1)!}$.

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

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

for $\alpha > 0, \beta > 0, \gamma > 0$ (10)

If a random variable X follows a McGBB $(n, \alpha, \beta, \gamma)$ distribution, then

The kth moment: $E(X^k) = n\left[\frac{B\left(\alpha + \beta, \frac{k}{\gamma}\right)}{B\left(\alpha, \frac{k}{\gamma}\right)}\right]$.

The mean: $E(X) = n\frac{B\left(\alpha + \beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)} = n\pi$, where $\pi = \frac{B\left(\alpha + \beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}$

The variance: $var(X) = n\pi(1 - \pi)(1 + (n - 1)\rho)$

Where;

$$\rho = \frac{\left[\frac{B\left(\alpha + \beta, \frac{2}{\gamma}\right)}{B\left(\alpha, \frac{2}{\gamma}\right)}\right] - \left[\frac{B\left(\left(\alpha + \beta, \frac{1}{\gamma}\right)\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}\right]^2}{\left[\frac{B\left(\alpha + \beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}\right] - \left[\frac{B\left(\left(\alpha + \beta, \frac{1}{\gamma}\right)\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}\right]^2}$$

$$0 \leq \pi \leq 1 \text{ and } \frac{-1}{(n-1)} < \rho \leq 1$$

Note that if $\rho \longrightarrow 1$ we have severe over-dispersion.

2.2 Beta-Type Generated Distributions as Mixing Distributions

Generalized beta distributions are classified into three classes. These classes are based on those generated using the transformation approach, beta-type distributions generated using the cumulative distribution function (CDF) approach and distributions that are based on special functions (Aldeni & Famoye, 2017). The generalized beta form of distributions that were derived based on transformation have largely been contributed by McDonald (1984). McDonald and his associates contributed in the development of the generalized beta distributions as an income distribution and in unifying the various research activities in closely related fields. McDonald and Xu (1995) introduced a five parameter beta distribution which nest the generalized beta and gamma distributions and include more than thirty distributions as limiting or special cases. Others like McDonald and Richards (1987) presented a four parameter generalized beta distribution which includes as special cases three and two parameter beta, generalized gamma, Weibull, power function, Pareto, lognormal, half-normal, uniform and others. Further Libby and Novick (1982) proposed a three parameter generalized beta distribution for utility fitting. Beta function is a special function. Other special functions have been applied in literature to generalize the beta distribution. One of them is Gauss hypergeometric that was suggested by Armero and Bayarri (1994) in connection with marginal prior/posterior distribution. They obtained the generalized beta distribution by dividing the classical beta distribution by a given set of algebraic function. Further, Nadarajah and Kotz, (2004) also studied the beta distribution based on the Gauss hypergeometric function. Others like Gordy (1988a) introduced confluent hypergeometric distribution and applied it to auction theory. He (Gordy, 1988b) similarly introduced the compound confluent hypergeometric distribution which contains McDonald and Xu‘s generalized beta, Gauss hypergeometric and confluent hypergeometric as special cases. Gordy generalized the classical beta distribution using the confluent hypergeometric function in the way Armero and Bayarri (1994) used Gauss hypergeometric function to generalize the beta to the Gauss hypergeometric distribution. More to the Gauss hypergeometric distribution, Nadarajah and Kotz (2006) introduced $F1$ beta distribution based on Appell function of the first kind. The distribution was found to be very flexible and contains a number of the known generalization of the classical beta distribution as certain cases is the most flexible beta-type distributions, because it includes the most distributions. For example, it can be used as an income distribution and includes special or limiting cases. Comparisons of the hypothetical distributions have been examined in studies.

2.3 Methods of Parameter Estimation

There are many methods of parameter estimation that have been developed with time. These methods include; method of moments, maximum likelihood estimation method, method of least squares, estimating functions method and Bayesian estimation method (Marmarelis, 2012). Estimation of parameters in the McGBB distribution has been done using the maximum likelihood method and the estimating functions method.

2.3.1 Maximum Likelihood Estimation Method

MLE method is a popular method of estimation in statistics due to its desirable asymptotic properties and applicability when handling large sample size data. Other properties like the invariance property, sufficiency, consistency are associated with the MLE method especially for particular classes of distribution (Haan, 1977; Mood *et al.,* 1974). However for small sample sizes some of these properties may not hold. This is because the likelihood function may not be a smooth surface and the regularity conditions may not hold for all cases (Lee & Sabavala, 1987).

Let $\underline{\boldsymbol{y}} = y_1, y_2, \dots, y_N$ be a random sample of size N from a McGBB distribution with unknown parameter vector $\underline{\boldsymbol{\theta}} = (\alpha, \beta, \gamma)$ then the likelihood function was computed as;

$$L(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}}) = L(\alpha, \beta, \gamma | y_1, y_2, \dots, y_N) = \prod_{k=1}^{N} f(y_k | \alpha, \beta, \gamma, n)$$

$$= \prod_{k=1}^{N} \binom{n}{y_k} \times \frac{1}{B(\alpha, \beta)} \times \sum_{j=0}^{n-y_k} (-1)^j \binom{n-y_k}{j} \times B\left(\left(\frac{y_k}{\gamma} + \alpha + \frac{j}{\gamma}\right), \beta\right) \qquad (11)$$

Thus, the log-likelihood becomes,

$$l(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}}) = log_e(L(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}}))$$

$$= \sum_{k=1}^{N} log \binom{n}{y_k} - \sum_{k=1}^{N} log\,(B(\alpha, \beta)) + \sum_{k=1}^{N} log \left[\sum_{j=0}^{n-y_k} (-1)^j \binom{n-y_k}{j} \times B\left(\left(\frac{y_k}{\gamma} + \alpha + \frac{j}{\gamma}\right), \beta\right)\right] \qquad (12)$$

The Maximum Likelihood Estimates (MLE) $\hat{\underline{\boldsymbol{\theta}}}$ of $\underline{\boldsymbol{\theta}}$ were obtained by Manoj *et al.* (2013) by maximizing the likelihood function, $L(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}})$. However, it is often easier to maximize the log-likelihood $l(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}})$ that is;

$$\hat{\underline{\boldsymbol{\theta}}} = arg \max_{\theta \in \Theta} L(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}}) = arg \max_{\theta \in \Theta} l(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{y}}).$$

This was achieved by directly maximizing the above log-likelihood function with respect to $\underline{\boldsymbol{\theta}}$ (Using a suitable algorithm such as the "Simplex algorithm for minimization" (Nelder, 1965) or by solving a system of equation by equating the score vector function $U\left(\underline{\boldsymbol{\theta}}\right)$ to zero that is;

$$U\left(\underline{\boldsymbol{\theta}}\right) = \nabla(l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})) = \frac{\partial(l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\underline{\boldsymbol{\theta}}}$$

$$= \begin{pmatrix} \frac{\partial(l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\alpha} \\ \frac{\partial(l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\beta} \\ \frac{\partial(l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\gamma} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad (13)$$

Closed form solutions to the system of equations $U\left(\underline{\boldsymbol{\theta}}\right) = \underline{\mathbf{0}}$ were not feasible because of its non-linear in nature.

Further, the exact distribution of Maximum Likelihood Estimates $\underline{\widehat{\boldsymbol{\theta}}}$ for $\underline{\boldsymbol{\theta}}$ could not be obtained explicitly instead asymptotic confidence intervals were constructed. Under some regularity conditions, the MLEs $\underline{\widehat{\boldsymbol{\theta}}}$ are approximately multivariate normal with mean $\underline{\boldsymbol{\theta}}$ and variance matrix $\boldsymbol{I}^{-\mathbf{1}}(\underline{\widehat{\boldsymbol{\theta}}})$, where $\boldsymbol{I}(\underline{\widehat{\boldsymbol{\theta}}})$, is the observed Fishers information matrix and is defined as;

$$\boldsymbol{I}(\underline{\widehat{\boldsymbol{\theta}}}) = -\nabla\nabla^{T}(l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))\Big|_{\underline{\boldsymbol{\theta}}=\underline{\widehat{\boldsymbol{\theta}}}}$$

$$= -\left.\begin{bmatrix} \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\alpha^2} & \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\alpha\,\partial\beta} & \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\alpha\,\partial\gamma} \\ \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\beta\,\partial\alpha} & \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\beta^2} & \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\beta\,\partial\gamma} \\ \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\gamma\,\partial\alpha} & \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\gamma\,\partial\beta} & \frac{\partial^2 l(\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\gamma^2} \end{bmatrix}\right|_{(\alpha=\hat{\alpha},\ \beta=\hat{\beta},\ \gamma=\hat{\gamma})} \qquad (14)$$

The $100(1-\psi)\%$ normal approximation confidence interval of α, β and γ was obtained as $(\hat{\alpha} \pm Z_{\psi/2}\sqrt{var(\hat{\alpha})})$, $(\hat{\beta} \pm Z_{\psi/2}\sqrt{var(\hat{\beta})})$ and $(\hat{\gamma} \pm Z_{\psi/2}\sqrt{var(\hat{\gamma})})$ respectively, where the diagonal elements of $\boldsymbol{I}^{-\mathbf{1}}(\underline{\widehat{\boldsymbol{\theta}}})$ provide the asymptotic variances for the parameters α, β and γ respectively and $Z \sim N(0,1)$ **.**

Maximum Likelihood Estimation method was used by Manoj *et al.* (2013) to estimate the parameters of McGBB distribution in their quest to show that it (McGBB) is superior in handling over/under dispersion compared to its nested distributions based on this work they developed a "*fitODOB"* R package. However a simulation study to evaluate the performance of MLE for different sample sizes was not illustrated as their main focus was to show the superiority of the distribution in capturing over/under dispersion. This package was used to obtain the maximum likelihood estimates and the Hessian Matrix that was used to obtain the observed Fishers matrix using the *optim* function in *R* language.

The parameters of the BB distribution which is nested in the McGBB have been estimated using the MLE method by Griffiths (1973) and using the Bayesian framework by Lee and Sabavala (1987).) Unconstrained ML estimates will give negative estimates of the over dispersed parameter (Griffith, 1973). A comparison of the ML estimates and the Bayesian estimates of the BB distribution by Lee and Sabavala (1987) showed some inconsistency with the ML estimates.

MLE method was used by Manoj *et al.* (2013) to estimate the parameters of the mixture distribution McGBB. MLE approach is also be faced with the problem of the maximum points being trapped in a local maximum (Johansen *et al.,* 2008). Moreover, there are no explicit solutions for the ML estimates. Therefore, one has to apply numerical methods for example the simplex method (Nelder & Mead, 1965) as applied by Manoj *et al.* (2013) to solve for the non-linear equations.

2.3.2 Estimating Functions Method

The concept of "Quasi-likelihood" was introduced by Wedderburn (1974). In this method it is possible to carry out estimation when the parametric form of the likelihood is miss-specified or when only the first two moments are definable. In other words, the Quasi-likelihood estimation is a semi-parametric approach that only the knowledge of the mean function and the relationship between the mean function and the variance function (i.e., only the form of the first two moments are required) of the response variable.

Let $\underline{\boldsymbol{y}} = y_1, y_2, \ldots., y_N$ be a random sample of size N from a McGBB distribution with unknown parameter vector $\underline{\boldsymbol{\theta}} = (\alpha, \beta. \gamma)$ (i.e., $\sim McGBB\ (n, \alpha,\ \beta,\ \gamma)$). Further define a variable $Z_k = \frac{Y_k}{n}$, then

$$E(Z_k) = E\left(\frac{Y_k}{n}\right) = \frac{1}{n}E(Y_k) = \frac{B\left(\alpha+\beta,\frac{1}{\gamma}\right)}{B\left(\alpha,\frac{1}{\gamma}\right)} = \pi$$

$$Var(Z_i) = Var\left(\frac{Y_k}{n}\right) = \frac{1}{n^2}Var(Y_k) = \frac{\pi(1-\pi)(1+(n-1)\rho}{n}$$

Where, $\rho = \dfrac{\left[\frac{B\left(\alpha+\beta,\ \frac{2}{\gamma}\right)}{B\left(\alpha,\ \frac{2}{\gamma}\right)}\right] - \left[\frac{B\left((\alpha+\beta,\frac{1}{\gamma})\right)}{B\left(\alpha,\frac{1}{\gamma}\right)}\right]^2}{\left[\frac{B\left(\alpha+\beta,\frac{1}{\gamma}\right)}{B\left(\alpha,\frac{1}{\gamma}\right)}\right] - \left[\frac{B\left((\alpha+\beta,\frac{1}{\gamma})\right)}{B\left(\alpha,\frac{1}{\gamma}\right)}\right]^2}$ $\quad 0 \leq \pi \leq 1$ and $\frac{-1}{(n-1)} < \rho \leq 1$

Following Wedderburn (1974) and McCullagh and Nelder (1989), contribution of the k^{th} observation, z_k with the above mean and variance to the quasi-likelihood, or more appropriately the quasi log-likelihood, is given by;

$$Q_k(\pi, \rho\ |z_k) = \frac{n}{\{1+(n-1)\rho\}}\int_{z_k}^{\pi} \frac{(z_i - t)}{t(1-t)}dt$$

$$= \frac{n}{\{1+(n-1)\rho\}}\left[z_k log_e\left(\frac{\pi}{z_k}\right) + (1-z_k)log_e\left(\frac{1-\pi}{1-z_k}\right)\right]$$

$$\Rightarrow Q_k(\pi, \rho\ |y_k) = \frac{1}{\{1+(n-1)\rho\}}\left[y_k log_e\left(\frac{n\pi}{y_k}\right) + (n-y_k)log_e\left(\frac{n(1-\pi)}{n-y_k}\right)\right]$$

for $k = 1, \dots, N$

Since $y_1, y_2, \dots., y_N$ are independent then the log quasi-likelihood for the complete data is the sum of the individual contributions

$$Q(\pi, \rho\ |y) = \sum_{k=1}^{N} Q_k(\pi, \rho\ |y_k) = \frac{1}{\{1+(n-1)\rho\}}\sum_{k=1}^{N}\left[y_k log_e\left(\frac{n\pi}{y_k}\right) + (n-y_k)log_e\left(\frac{n(1-\pi)}{n-y_k}\right)\right]$$

The Quasi-Likelihood Estimates (QLE) $\underline{\widehat{\boldsymbol{\theta}}}_{\boldsymbol{QL}}$ of $\underline{\boldsymbol{\theta}}$ is obtained by maximizing the Quasi-likelihood function, $Q(\ \underline{\boldsymbol{\theta}}|\ \underline{\boldsymbol{y}})$ i.e.,

$$\underline{\widehat{\boldsymbol{\theta}}}_{\boldsymbol{QL}} = arg \max_{\theta \in \Theta} Q(\ \underline{\boldsymbol{\theta}}|\ \underline{\boldsymbol{y}})$$

Just like in the standard likelihood method this can be achieved either by directly maximizing the above Quasi-likelihood function with respect to $\underline{\boldsymbol{\theta}}$ (Using a suitable algorithm such as the “Simplex algorithm for minimization” by Nelder (1965) or by solving a system of equation by equating the Quasi-likelihood score vector function U_{QL} $(\underline{\boldsymbol{\theta}})$ to zero. i.e.,

$$U_{QL}\left(\underline{\boldsymbol{\theta}}\right) = \nabla(Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})) = \frac{\partial(Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\underline{\boldsymbol{\theta}}} = \begin{pmatrix} \frac{\partial(Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\alpha} \\ \frac{\partial(Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\beta} \\ \frac{\partial(Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))}{\partial\gamma} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Closed form solutions to the system of equations $U_{QL}\left(\underline{\boldsymbol{\theta}}\right) = \underline{\mathbf{0}}$ are not feasible because of its highly non-linear nature and will require iterative procedures such as the Newton-Raphson algorithm, "Simplex algorithm for minimization" by Nelder (1965), the fixed-point iterative method.

The exact distribution of Quasi-Likelihood Estimates $\underline{\widehat{\boldsymbol{\theta}}}_{\boldsymbol{QL}}$ for $\underline{\boldsymbol{\theta}}$ cannot be obtained explicitly instead asymptotic confidence intervals are constructed. Under some regularity conditions, the QLEs $\underline{\widehat{\boldsymbol{\theta}}}_{QL}$ are approximately multivariate normal with mean $\underline{\boldsymbol{\theta}}$ and variance matrix $\boldsymbol{I}^{-\mathbf{1}}(\underline{\widehat{\boldsymbol{\theta}}}_{QL})$, where $\boldsymbol{I}(\underline{\widehat{\boldsymbol{\theta}}}_{QL})$, is the observed Fishers information matrix and is defined as;

$$\boldsymbol{I}\left(\underline{\widehat{\boldsymbol{\theta}}}_{QL}\right) = -\nabla\nabla^{T}(Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}}))\Big|_{\underline{\boldsymbol{\theta}}=\underline{\widehat{\boldsymbol{\theta}}_{QL}}}$$

$$= -\left.\begin{bmatrix} \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\alpha^2} & \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\alpha\,\partial\beta} & \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\alpha\,\partial\gamma} \\ \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\beta\,\partial\alpha} & \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\beta^2} & \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\beta\,\partial\gamma} \\ \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\gamma\,\partial\alpha} & \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\gamma\,\partial\beta} & \frac{\partial^2 Q(\,\underline{\boldsymbol{\theta}}|\,\underline{\boldsymbol{y}})}{\partial\gamma^2} \end{bmatrix}\right|_{(\alpha=\widehat{\alpha}_{QL}\ \ \beta=\widehat{\beta}_{QL},\ \gamma=\widehat{\gamma}_{QL})}$$

The $100(1\ -\psi\ )\%$ normal approximation confidence interval of α, β and γ will be obtained as $(\hat{\alpha}_{QL} \pm Z_{\psi/2}\sqrt{var(\hat{\alpha}_{QL})})$, $(\hat{\beta}_{QL} \pm Z_{\psi/2}\sqrt{var(\hat{\beta}_{QL})})$ and $(\hat{\gamma}_{QL} \pm Z_{\psi/2}\sqrt{var(\hat{\gamma}_{QL})})$ respectively, where the diagonal elements of $\boldsymbol{I}^{-\mathbf{1}}(\underline{\widehat{\boldsymbol{\theta}}}_{QL})$ provide the asymptotic variances for the parameters α, β and γ respectively and $Z{\sim}N(0{,}1)$.

Estimating functions approach have been used by Jannifer *et al.* (2013) used to estimate the parameters of McGBB distribution in particular the Quasi-likelihood and Quadratic estimating equations. Further, Bichanga *et al.* (2015) used the quasi-likelihood and extended quasi-likelihood estimates in the development of the $C(\alpha)$ tests for testing homogeneity of proportions in the presence of over/under dispersion using the McGBB.

2.3.3 Bayesian Framework

The Bayesian approach to estimation incorporates three components. These components are the likelihood function, the observed data and the prior beliefs about the parameters. The approach gives an opportunity to use probability theory to represent subjective beliefs. The data is fixed while the parameters are treated as random variables.

If M and N are two events then the Bayes rule conferring to Aksoy and Guner (2015) is expressed by;

$$P(M/\mathrm{N}) = \frac{P(\mathrm{N}/\mathrm{M})\,p(M)}{P(N)} \tag{15}$$

Where $P(M)$ and $P(N)$is the independent probabilities of each event occurring, and $P(M/N)$ is the probability that event M occurs given that event N is true. Equation 11 forms the basis of Bayesian thinking. Bayesian thinking is applied in estimation of parameters in statistical and machine learning models (Kruschke & Lidell, 2018). In model case, the Bayesian equation is now expressed as;

$$P(\Theta/\ \mathrm{data}) = \frac{P(\mathrm{data}/\ \Theta)\,p(\Theta)}{P(data)}$$

Where $p(\Theta)$ is the prior distribution of the parameters in the parametric space of interest, $P(\Theta/\ \mathrm{data})$ is the posterior distribution and Θ is the parametric space.

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 (Petzschner *et al.,* 2015). 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 (Dodwell *et al.,* 2015). 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 (Lee & Vanpaemel, 2018). 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 (Sprenger, 2018). 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 (Lynch, 2007). 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. When there is substantial prior information for θ then the prior distribution dominates the posterior distribution that is $\pi(\theta|x) \sim \pi(\theta)$ (Stoner *et al.,* 2019). When prior information about θ is limited, the pragmatic approach is to choose a distribution which makes the Bayes updating from prior to posterior mathematically straightforward, and use what prior information is available to determine the parameters of this distribution (Beck & Taflanidis, 2013). For example Poisson random sample, Gamma prior distribution leads to Gamma posterior distribution, Normal random sample (known variance), and Normal prior distribution leads to Normal posterior distribution. In this study 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 (Li *et al.*, 2017). In order to avoid the problem of intractability problem, the most commonly used methods are the Markov Chain Monte Carlo methods (MCMC) (Martino & Riebler, 2019). With these methods it is possible to obtain samples of the marginal posterior distributions of the parameters from the joint without performing the integrations (Alquier *et al.,* 2016). If we have a statistical model $f\ (x|\theta)$ for data $x = (x1, x2, \ldots, xn)$, together with a prior distribution $\pi(\theta)$ for θ then $q\, J(\theta\hat{})\ (\theta - \theta\hat{})|x\ D \longrightarrow N(0,1)$ as $n \to \infty$, where θˆ is the likelihood mode and $J(\theta)$ is the observed information, this means that, with increasing amounts of data, the posterior distribution looks more and more like a normal distribution. The result also gives us a useful approximation to the posterior distribution for θ when n is large. Note that this limiting result is similar to one used in Frequentist statistics for the distribution of the maximum likelihood estimator where Fisher's information$I(\theta)$ is the expected value of the observed information, where the expectation is taken over the distribution of$X|\theta$ (Wang *et al.,* 2015). In this study the Bayesian framework developed utilized the MCMC methods, specifically the use of a Metropolis- Hasting step within the Gibbs sampling.

2.4 Frequentist Versus Bayesian Approaches to Parameter Estimation

Frequentists, also referred to as the classical statisticians have different approaches to estimation and inferencing compared to Bayesian statisticians. In Bayesian statistics the parameters are treated as random variables while the frequentist approach treats the parameters are fixed (Sjolander & Vansteelandt, 2019). In frequentists methods the data is generated from a repeated random process hence varied while in Bayesian approach the observed data does not vary unless new data set is observed and the posterior is updated (McNeish, 2016). Frequentists carry out point estimation alongside confidence intervals while Bayesian statisticians use probability distribution to describe the behaviour of the parameters with means and quantiles as the possible candidates for estimation alongside credible intervals for uncertainties (Feit *et al.,* 2017).

To illustrate this difference, let $y_1, \cdots, y_n \underline{iid}\, f(.|\theta)$ be a set of observed data from a specific distribution with parameter vector. The main aim is to make inference aboutθ. As a frequentist, one would consider θ to be _fixed and concentrate on its estimation, for example using the Maximum Likelihood Estimate (MLE) $\hat{\theta}$ which is calculated by maximizing the likelihood function:

$$\underset{\theta}{\operatorname{argmax}}\, L(\theta|y_1, \cdots, y_n) = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} f(yi|\theta)$$

2.5 Markov Chain Monte Carlo Techniques

Markov Chain Monte Carlo methods gained popularity after the discovery of computers. However, physicist and chemists initially used these methods for simulation purposes (Berg, 2017). Statisticians later adopted these methods in the late 1990s. MCMC methods have helped to shed light in the Bayesian computational problems since they enabled full Bayesian inference of all kinds (Bolstad & Curran, 2016). Monte Carlo methods are computational algorithms that are built on the criteria of random sampling. They are used to estimate exact numerical value of a problem like complex integration, which cannot be achieved using the numerical methods (Hammersley, 2013). The idea is to generate a Markov chain; which is a sequence of random elements where the conditional distribution of x_{n+1} given $x_1, x_2, \dots, x_n$ depends on x_n only. The algorithm is uniquely designed to the specifications of the posterior distribution, whereby it generates a sample that mimics the values that would ideally be drawn from the posterior distribution (Gamerman & Lopes, 2006). Poor mixing of a Markov chain Monte Carlo algorithm may pose the problem of convergence of the chain. Several methods have been suggested in literature to improve the convergence of MCMC chains. Such methods include; blocking, collapsing, re-parametrization and multiple try metropolis hasting among other methods. Gelman and Rubin (1992) proposed a general method for checking convergence of the MCMC chains. The method that is similar to the classical analysis of variance, is based on comparisons of variances between chains and within chains. Metropolis-hasting algorithm is the most used MCMC method die to its fast rate of convergence property. It requires a proposal distribution to help sample from an intractable posterior distribution. In the M-H algorithm, the first step is to set the initial values that is the starting point. The loops in the algorithms consist of three parts; the first part is to generate the proposal candidate from the proposal distribution. The second part of the loop defines the computation of the acceptance probability that is the M-H ratio and the last part of the loop defines the decision criteria of the acceptance or rejection of the candidate sampled based on the probability computed in the previous loop. Gibbs sampling is a special case of the M-H algorithm. It generates posterior samples by sweeping through each or the block of variables to sample from its conditional form of distribution, with the remaining variables fixed at current state (Gamerman & Lopes, 2006).

MCMC plays a big role in both art and scientific fields like physics, statistics, econometrics psychology and sociology among other fields of study. Some of the MCMC techniques that have been widely explored include the metropolis –hasting algorithm and the Gibbs sampling technique.

2.6 Markov Chain Properties

A discrete-time Markov Chain is defined by a sequence of random variables $x_1, x_2, x_3, \cdots$ that can take possible values in the state space X with an initial distribution defined for X_0 and transition probabilities defined as;

$$p(x, A) = p(x_{n+1} \in A | x_n = x), \forall\, A \subseteq X$$

The Markov property is a fundamental property of the Markov chain, it is stated as,

$p(x_{n+1} \in A | x_1, x_2, \cdots, x_n) = p(x_{n+1} \in A | x_n), \forall\, A \subseteq X$

Which means that the probability that a state moves to the next state is dependent only on the current state and not the previous state.

Definition 1. Consider a Markov chain $\{X_i\}$ on state space X with transition probability P(x,.). Let π (.) be a probability distribution defined on X. Then π is a stationary distribution for the Markov chain if for x, y ∈ X:

$Z\, \pi\, (dx) P(x, dy) \; = \; \pi(dy)\; x \; \in X$

Definition 2. A Markov chain is φ-irreducible, if there exists a non-zero σ-finite measure φ on X such that:

$\forall A : \; A \; \subseteq \; X\, with\, \varphi(A) > \; 0\, \forall x \; \in \; X$

$== \Rightarrow \; \exists\, n \; \in \; N : Pn(x, A) \; > \; 0$

Definition 3. A Markov chain is aperiodic, if there are no disjoint non-empty subsets $X1, \ldots, Xd \; \subseteq \; X\, for\, d \; \geq \; 2$, such that $P(x, Xi + 1) \; = \; 1$ for all $x \; \in \; Xi$ $(1 \; \leq \; i \; \leq \; d - 1)$ and $P(x, X1) \; = \; 1$ for all $x \; \in \; Xd$.

Now the Markov chain convergence theorem which specifies how MCMC algorithms work is stated as;

Theorem 1. Consider an aperiodic and φ-irreducible Markov chain defined on a state space X with stationary distribution π. Then for $\pi. a. e. x \; \in \; X$:

$$lim \, ||Pn(x, .) \; - \; \pi(.)|| \; = \; 0\, n \to \infty$$

That is $limn \to \infty\; Pn(x, A) \; = \; \pi(A)$ for all measurable $A \; \subseteq \; X$.

Proof. See Sean and Richard (2012) and Jeffrey (2006).

Assuming the goal is to sample from a complex π, MCMC methods are designed such that a Markov chain is generated with π stationary distribution. The *aperiodicity* of such Markov chains almost always hold (as there is usually positive probability of rejection in most MCMC methods), and *φ-irreducibility* is also straightforward to check and usually holds for MCMC algorithms

(Auger & Hansen, 2016). Hence, MCMC methods become reliable as they are built under the foundation of Markov chains.

2.7 Markov Chain Diagnostics

To ensure that the MCMC method is mixing well and hence converging, examining the choice of proposal distributions, as they can control the optimality in performance. For this purpose, *acceptance rates*, which are the proportion of *accepted* proposals to the total proposals, are useful tools to help determine such issues. The acceptance rates can specify how much the target distribution is being explored, that said, they should not be too low (close to 0), as this shows more rejection of proposals and hence a chain that is stuck and is not moving much. On the other hand, acceptance rates that are too high (close to 1) also show small movements of the chain, as only close-by states can be accepted highly. Hence, there needs to be an acceptance rate in between that can help the chain move more and explore the target distribution. It is proven that under mild conditions, the *optimal acceptance rates* of a *d*-dimensional Metropolis algorithm with a Gaussian proposal distribution is 0.234 as $d \rightarrow \infty$ (Gareth *et al.,* 1997). It is also shown that acceptance rates between 15% and 50% can still perform well (Gareth & Jeffrey, 2001).

2.8 Interval Estimation

In interval estimation given data x, we replace the point estimate $\hat{\theta}(x)$ for the parameter θ by a statistic that is subset$C\hat{}(x)$ of the parameter space. The classical and Bayesian approaches have different ways of choosing $\hat{}(x)$. Therefore, the two approaches have very different interpretations and ways of computations (Bodnar *et al.,* 2017). Interval estimation is an alternative to reporting a single number that is point estimate hence the goal is to report an entire interval of plausible values (Greenland *et al.,* 2016). We look at both approaches in detail that is the Bayesian and classical way of interval estimation.

2.8.1 Bayesian Credible Regions

A more useful summary of the posterior distribution is one which also reflects its variation. For example, $100(1 - \alpha)\%$ Bayesian confidence interval for θ is any region $C\alpha$ that satisfies$P\,r\,(\theta \in C\alpha|x) = 1 - \alpha$. If θ is a continuous quantity with posterior probability density function $\pi(\theta|x)$ then$C\alpha\;\pi(\theta|x)\;d\theta = 1 - \alpha$. The usual correction is made for discrete θ, that is, we take the largest region $C\alpha$ such that$P\,r\,(\theta\ \in\ C\alpha|x)\ \leq\ 1\ -\ \alpha$. Bayesian confidence intervals are sometimes called credible regions or plausible regions. Clearly these intervals are not unique, since there will be many intervals with the correct probability coverage for a given posterior distribution (Martin & Liu, 2013). A $100(1\ -\ \alpha)\%$ highest density interval (HDI) for θ is the region $C\alpha\ =\ \{\theta\ :\ \pi(\theta|x)\ \geq\ \gamma\}$ where γ is chosen so that$P\,r\,(\theta\ \in\ C\alpha|x)\ =\ 1\ -\ \alpha$. This region is sometimes called a most plausible Bayesian confidence interval. If the posterior distribution has many modes then it is possible that the HDI will be the union of several disjoint regions (Martin & Liu, 2015).

2.9 Sample Size Determination

The sample size (n) can be defined as the number of units in a group under study (Boody, 2016). Determining a sample size is of great importance when carrying out any type of study. It is often important step in planning a statistical study. This is because studies that are too small sizes of samples may lead to inaccurate results but studies that are too large sizes of samples may waste time and resources (Kim *et al.,* 2020). In selecting a sample size there are important factors that are considered on calculating the appropriate sample size. These factors, include: the type of data, level of precision, the confidence level, the degree of variability, effect size, type I error, Type II error, significance level, power of the test and other factors (Wolf *et al.,* 2013). An adequate sample size allows the researcher to report his results with sufficient degree of confidence and acceptable statistical power and helps ensure that the study will yield reliable information, interpretable results and minimizes research waste, but with inadequate sample size, making the interpretation of negative results difficult and yielding statistically in conclusive results. In addition to, sample size is important for economic reasons: an undersized study can be waste time and resources without to produce useful information, while an oversized study uses more resources than are necessary. Most of statistical studies are always better when the sample size is adequate relative to the purpose of the study (Russell, 2001). To avoid undersized and oversized studies, must be calculating of the adequate sample size, but before calculating the sample size of the study, we must answer to some the following questions: What is type the collective data of the study? , What is the purpose of the study? And how many samples will you calculate? To answer these questions this study further evaluated each factor that influences the determination of sample size

2.9.1 The Type of Collective Data of the Study

The statistical methods of sample size determination depend on the type of data being collected. If the data are continuous, our concern will be about studying the means, but if the data are qualitative (categorical data), our concern will be about studying the proportions (Ralph *et al.,* 2002).

2.9.2 The Level of Precision

The level of precision refers to the survey error or the sampling error. It is defined to be the difference between the parameter of population and the sample statistic that associated with this parameter. The level of precision is inversely proportional with the sample size. It means, the lower the precision level, the more sample size will be needed to reach the goal. This level is often expressed in percentage (such as ± 5% or ± 7%). Note that, while determining the sample size, the level should be as low as possible without increasing the sample size to more than your resources will allow (Chadha, 2006; Israel, 2009).

2.9.3 The Confidence Level.

The confidence level is necessary when the results will be presented using confidence intervals. This level indicates to the probability value that the sample contains the parameter being estimated. Common confidence levels used in most of studies are 90% or 95%. If 95% confidence level is selected, this means that, if you repeat the experiment 100 times, there are 95 out of 100 times will have the true value of population parameter within the range of precision pre-specified. The confidence level is directly proportional with the sample size. That is, the higher the confidence level predetermined, the larger the sample size will be needed to achieve the purpose of the study (Rahul, 2007).

2.9.4 The Significant Level

The significant level refer to the probability of spotting a statistically significant difference that is the result of chance in other words, this level defines the probability of obtaining an erroneously significant results. Also, this factor refers to the maximum P-value for which a difference is to be considered statistically significant. The significant level is inversely proportional with the sample size, as this level is decreased; the sample size is needed to detect the difference increased. The level of significant is most commonly set a 5% or 1% chance of inaccurately recording a significant effect (Jon *et al.,* 2003).

2.9.5 The Effect Size

The effect size (minimum expected difference) is the smallest measured difference between comparisons groups that the researcher would like the study to detect. It measures the distance between the null hypothesis and a specified value of the alternative hypothesis. The effect size has a reverse relationship with the sample size. It means, the smaller the minimum expected difference is predetermined; the larger sample size is needed to detect a statistical significance. Note that, estimation of the quantity of effect size is not always straight forward because the setting of this factor is subjective and based on experience with the problem of the study. There are different ways to calculate effect size depending on the evaluation design you use, general size is calculated by taking the difference between the two groups and dividing it by the standard deviation of one of the groups (Adcock, 1997; Eng, 2003).

2.9.6 The Degree of Variability

The degree of variability in the variables (attributes) being measured refers to the distribution of variables in the population. The more heterogeneous a statistical population (the more the degree of variability), the larger the sample size required to obtain level of precision. The more homogeneous a population, the smaller the sample size. This factor is considered a vital component of the sample size determination and is represented by the expected variance in the primary variables of interest in the study as listed (Cochran, 1977).

2.10 Summary of Previous Studies

A study of estimating over-dispersed corona virus data in China was done by Endo (2020). In this study, the over dispersed data was however modelled using the negative binomial model. The study compared the negative binomial model to the Poisson model using the Bayesian information criteria. The study found out that the negative binomial model provided better estimation than the Poisson model. However, this study was limited in the data that they used for comparison of the models.

A study to estimate the parameters of beta-binomial distribution was done by Martinez *et al.* (2015), however they used the *SAS* software to carry out the maximum like-likelihood estimation. Beta-binomial captures over- dispersion and is a distribution that is nested in McGBB distribution. In their study they noted that maximum likelihood method may pose challenges when carrying out estimation for a mixture distribution since explicit solutions may not exist. Therefore they recommended use of SAS software in carrying out the maximum likelihood estimation and noted that the software satisfactorily performs well in carrying out such estimation procedures.

The parameters of McGBB have been estimated in literature using the Estimating functions method and the maximum-likelihood method both for which are classical methods of estimation (Jannifer *et al.,* 2014; Manoj *et al.,* 2013). However, Jannifer *et al.* (2014) showed that the estimating functions method was more robust than the maximum likelihood method.

A comparison of maximum likelihood method to the Bayesian method of estimation in three parameter Weibull distribution was done by Teimouri *et al.* (2013).In their study, Teimouri *et al.* (2013) found that the estimators were completely different, however, the Bayesian approach provided the best solutions to the estimation since the maximum likelihood presented some inconsistency issues. Which is also in line with findings from Lee and Sabavala (1987).A Bayesian estimation study that used tailored randomized block Metropolis–Hastings by Kakamu and Nishino (2018) found out that the choice of algorithm to sample from the posterior also plays a key role in the Bayesian methods of estimation. They compared tailored randomized block Metropolis–Hastings that was proposed by Chib and Ramamurthy (2010) with Randomized Weighted Metropolis-Hasting algorithm proposed by Chotikapanich and Griffiths (2000). From the results they obtained, there was a confirmation that used tailored randomized block Metropolis–Hastings is more efficient than Randomized Weighted Metropolis-Hasting algorithm in terms of mixing. In addition, using empirical results of U.S. income data, they showed that the estimated parameters of the Bayesian approach differed from those of the MLE. Finally, they found out that the Bayesian approach could estimate the parameters of the GB distribution even if the number of groups are relatively small.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Bayesian Estimation Procedure

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 which the study was interested in estimating;

$$\underline{\boldsymbol{\theta}}=(\alpha, \beta, \gamma)^{T}.$$

The study assumed the prior distribution of α, β and γ had jointly a flat prior, which were represented as;

$$\pi(\underline{\boldsymbol{\theta}}) \propto \pi(\alpha,\beta,\gamma) \propto 1. \tag{16}$$

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

$$f\left(\underline{\boldsymbol{y}}\middle|\ \underline{\boldsymbol{\theta}}\right) = f(y_1,\ y_2, \dots, y_N|\ \alpha,\ \beta,\ \gamma) = \prod_{k=1}^{N} f(y_k\ |\alpha,\ \beta,\ \gamma,\ n)$$

$$= \prod_{k=1}^{N}\binom{n}{y_k}\times\frac{1}{B(\alpha,\ \beta)}\times\sum_{j\,=\,0}^{n\,-\,y_k}(-1)^j\binom{n-y_k}{j}\times B\left(\left(\frac{y_k}{\gamma}+\alpha+\frac{j}{\gamma}\right),\ \beta\right). \tag{17}$$

The posterior distribution was therefore obtained as;

$$\pi(\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{y}}) \propto f(\underline{\boldsymbol{y}}|\underline{\boldsymbol{\theta}})\times\pi(\underline{\boldsymbol{\theta}})$$

$$\propto \prod_{k=1}^{N}\binom{n}{y_k}\times\frac{1}{B(\alpha,\ \beta)}\times\sum_{j=0}^{n-y_k}(-1)^j\binom{n-y_k}{j}\times B\left(\left(\frac{y_k}{\gamma}+\alpha+\frac{j}{\gamma}\right),\ \beta\right)\times 1 \tag{18}$$

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:

$$\pi_\alpha\left(\alpha\middle|\beta,\gamma,\ \underline{\boldsymbol{y}}\right)=\frac{\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})}{\pi(\beta,\ \gamma\ |\underline{\boldsymbol{y}})}=\frac{\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})}{\int\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})\,d\alpha}\propto\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}}),\tag{19}$$

$$\pi_\beta\left(\beta\middle|\alpha,\gamma,\underline{\boldsymbol{y}}\right)=\frac{\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})}{\pi(\alpha,\ \gamma|\underline{\boldsymbol{y}})}=\frac{\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})}{\int\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})d\beta}\propto\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})\qquad\text{and}\tag{20}$$

$$\pi_\gamma\left(\gamma\middle|\alpha,\ \beta,\ \underline{\boldsymbol{y}}\right)=\frac{\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})}{\pi(\alpha,\ \beta|\underline{\boldsymbol{y}})}=\frac{\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})}{\int\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}})d\gamma}\propto\pi(\alpha,\ \beta,\ \gamma|\underline{\boldsymbol{y}}).\tag{21}$$

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

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

Step 2. Using the proposal distribution of $\alpha,\ q_\alpha\left(\alpha\ \middle|\alpha^{(j-1)}\right)$ sample a candidate value for $\alpha_{(prop)}^{(j)}$

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 $\alpha_{(prop)}^{(j)}$ and

the previous value $\alpha^{(j-1)}$.

$$R_\alpha=\frac{\pi_\alpha\left(\alpha_{(prop)}^{(j)}\middle|\beta^{(j-1)},\gamma^{(j-1)},\underline{\boldsymbol{y}}\right)\times q_\alpha\left(\alpha^{(j-1)}\ \middle|\alpha_{(prop)}^{(j)}\right)}{\pi_\alpha\left(\alpha^{(j-1)}\middle|\beta^{(j-1)},\gamma^{(j-1)},\underline{\boldsymbol{y}}\right)\times q_\alpha\left(\alpha_{(prop)}^{(j)}\ \middle|\alpha^{(j-1)}\right)}$$

Step 5: If$u\ \leq\ min(1,\ R_\alpha)$, accept the candidate point with probability $min(1,\ R_\alpha)$,

i.e., set $\alpha^{(j)}\ =\ \alpha_{(prop)}^{(j)}$. Otherwise set $\alpha^{(j)}\ =\ \alpha^{(j-1)}$

Step 6. Using the proposal distribution of $\beta,\ q_\beta\left(\beta\middle|\beta^{(j-1)}\right)$ sample a candidate value for $\beta_{(prop)}^{(j)}$

Step 7. Generate U from a Uniform (0, 1) distribution (i.e. $u{\sim}UNIF(0,1)$.

Step 8. Calculate the Metropolis-Hasting (MH) ratio at the candidate value $\beta_{(prop)}^{(j)}$ and

the previous value $\beta^{(j-1)}$.

$$R_\beta=\frac{\pi_\beta\left(\beta_{(prop)}^{(j)}\middle|\alpha^{(j)},\gamma^{(j-1)},\underline{\boldsymbol{y}}\right)\times q_\beta\left(\beta^{(j-1)}\ \middle|\beta_{(prop)}^{(j)}\right)}{\pi_\beta\left(\beta^{(j-1)}\middle|\alpha^{(j)},\gamma^{(j-1)},\underline{\boldsymbol{y}}\right)\times q_\beta\left(\beta_{(prop)}^{(j)}\ \middle|\beta^{(j-1)}\right)}$$

Step 9: If $u\ \leq\ min(1,\ R_\beta)$, accept the candidate point with probability $min(1,\ R_\beta)$,

i.e., set $\beta^{(j)}\ =\ \beta_{(prop)}^{(j)}$. Otherwise set $\beta^{(j)}\ =\ \beta^{(j-1)}$

Step 10. Using the proposal distribution of $\gamma,\ q_\gamma\left(\gamma\ \middle|\gamma^{(j-1)}\right)$ sample a candidate value

for $\gamma_{(prop)}^{(j)}$

Step 11. Generate U from a Uniform (0, 1) distribution (i.e. $u{\sim}UNIF(0,1)$.

Step 12. Calculate the Metropolis-Hasting (MH) ratio at the candidate value $\gamma_{(prop)}^{(j)}$ and the previous value $\gamma^{(j-1)}$.

$$R_\gamma = \frac{\pi_\gamma\left(\gamma_{(prop)}^{(j)} \middle| \alpha^{(j)}, \beta^{(j)}, \underline{\boldsymbol{y}}\right) \times q_\gamma\left(\gamma^{(j-1)} \middle| \gamma_{(prop)}^{(j)}\right)}{\pi_\gamma\left(\gamma^{(j-1)} \middle| \alpha^{(j)}, \beta^{(j)}, \underline{\boldsymbol{y}}\right) \times q_\gamma\left(\gamma_{(prop)}^{(j)} \middle| \gamma^{(j-1)}\right)}$$

Step 13: If $u \le min(1,\ R_\gamma)$, accept the candidate point with probability $min(1,\ R_\gamma)$, i.e., set $\gamma^{(j)} = \gamma_{(prop)}^{(j)}$. Otherwise set $\gamma^{(j)} = \gamma^{(j-1)}$

Step 14. Repeat the Steps 2-13 for all $j = 1, 2, \ldots, M$ and thus obtaining sample of size M from the joint posterior distribution $\{\underline{\boldsymbol{\theta}}_j = (\alpha_j, \beta_j, \gamma_j),\ \ j = 1,2,\ldots,M\}$ Therefore for all $j = 1,\ 2,\ \ldots,\ M$ a sample of size M is obtained as the joint posterior distribution $\{\underline{\boldsymbol{\theta}}_j = (\alpha_j,\ \beta_j,\ \gamma_j),\ \ j = 1, 2, \ldots\ ,\ M\}$ 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 Metropolis- Hasting step within Gibbs sampling algorithm above, i.e.,

$$\hat{\alpha} = E(\alpha|\underline{\boldsymbol{y}}) = \frac{1}{M-B_0}\sum_{j=B_0+1}^{M} \alpha_j\ , \tag{23}$$

$$\hat{\beta} = E(\beta|\underline{\boldsymbol{y}}) = \frac{1}{M-B_0}\sum_{j=B_0+1}^{M} \beta_j \qquad \text{and} \tag{24}$$

$$\hat{\gamma} = E(\gamma|\underline{\boldsymbol{y}}) = \frac{1}{M-B_0}\sum_{j=B_0+1}^{M} \gamma_j\ . \tag{25}$$

The $100(1-\psi)\%$ Bayesian Credible Intervals and $100(1-\psi)\%$ Highest Probability Density (HPD) intervals for α, β and γ were obtained using the algorithm proposed by Chen and Shao (1999). The algorithm packaged in the package *CODA* in *R* language. Let $\{(\alpha_{(j)}, \beta_{(j)}, \gamma_{(j)}),\ \ j = 1,2,\ldots,M\}$ be an ordered sample corresponding to the MCMC chain $\{(\alpha_j, \beta_j, \gamma_j),\ \ j = 1,2,\ldots,M\}$ obtained using the Metropolis- Hasting step within the Gibbs sampling algorithm above.

Then the approximate $100(1-\psi)\%$ Bayesian Credible Intervals for α, β and γ were obtained as;

$$\left(\alpha_{([\psi M/2])},\ \ \alpha_{([(1-\psi/2)M])}\right), \tag{26}$$

$$\left(\beta_{([\psi M/2])},\ \beta_{([(1-\psi/2)M])}\right) \quad \text{and} \tag{27}$$

$$\left(\gamma_{([\psi M/2])},\ \gamma_{([(1-\psi/2)M])}\right). \tag{28}$$

3.3 Graphical Presentation

The study used three types of graphical representation to report results obtained. These graphics include; trace plots, autocorrelation plots and histogram plots. Trace plots is a visual that was used to show the sampled values of the parameters from the marginal posterior densities over time, that is the sampled values against the number of iterations performed. Trace plot is used in MCMC to visually inspect the marginal posterior and convergence in distribution (Gamerman & Lopes, 2006). They are also used to identify visually potential problems (if any) affecting the quality of the sample generated. Autocorrelation plots were also used in this study to check for the randomness of the sampled values. Histogram plots were used to show the distribution of the shape parameters of the McGBB distribution.

3.4 Simulation Algorithm

Simulation of the McGBB variables was implemented using the following algorithm. The algorithm provided a direct approach to simulate directly from a McGBB distribution while being in control of the parameters. The variables obtained are McGBB variables as opposed to the algorithm suggested by Manoj *et al.* (2013) whose variables were over dispersed random binomial variables;

Step 1: Setting fixed values of α, β and γ.

Step 2: Generate *K* random variables from Beta (α, β) that is;

$U_i \sim BETA(\alpha, \beta)$ for $k = 1, \dots, K$

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

Step 4: For each P_i in step 3 generate binomial random variables that is:

$X_i \sim Bin\ (n, P_i)$, then $X_i \sim McGBB(n, \alpha, \beta, \gamma)$ for $i = 1, \dots, K$.

The parameters α, β and γ will be fixed values following from the assumption that they are real positive values.

The study simulated different samples of different sizes in particular; k=25 and k=29 to represent small samples and, k= 500 and k= 1000 to represent large samples, for fixed values of the parameters α, β and γ. This was done to ensure that the behavior of the parameter estimates and inferences was captured for small sample size and large sample size

Table 1: Simulated Dataset

Week	Frequencies							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
K = 25	7	1	1	3	2	4	3	4
K = 29	7	7	2	3	1	1	4	4
K = 500	195	59	35	40	25	33	35	78
K = 1000	399	90	91	70	73	63	74	140

Table 1 shows the simulated data set that was applied for this study. The McGBB variables were simulated by setting the true parameters values at $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$ respectively.

3.5 Real Data Description

In this study different sets of real data were applied in evaluating the performance of the Bayesian method and maximum likelihood methods of estimation. The alcohol dataset for instance has been used in literature by Jannifer *et al.* (2014) and Manoj *et al.* (2013) for illustration. These datasets are documented in the *fitODBOD* 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.

Table 2: Alcohol Dataset

Week	Frequencies							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Week 1	47	54	43	40	40	41	39	95
Week 2	42	47	54	40	49	40	43	83

Table 2 shows the frequencies for the number of times the respondents self-recorded their rate of consuming alcohol. From the table k = 399, the sample size is a large sample.

Table 3: Chromosome Association dataset

	Frequencies			
	0	**1**	**2**	**3**
Observations	32	103	122	80

In this dataset the secondary association of chromosomes was investigated. Data in this example refer to 337 observations hence a large sample data set. The number of chromosomes showing bivalent association was recorded for each trial done four three times. This data was used by Paul (1985) as well for illustration in their study.

Table 4: Course dataset

	Frequencies								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Observations	1	4	4	8	9	6	8	12	13

The data refer to the numbers of courses taken by a class of 65 students from the first year of the Department of Statistics of Athens University of Economics. The students enrolled in this class, attended 8 courses during the first year of their study. The total numbers of successful examinations that also included resits were recorded. This data was adopted by Karlis and Xekalaki (2008) for illustration in their study.

Table 5: Epidemic cold dataset

	Frequencies				
	0	**1**	**2**	**3**	**4**
Families	423	199	39	3	0
Father	53	31	4	1	0
Mother	75	25	4	1	0
School child	148	77	22	2	0

In this investigation, families of the same size, two parents and three children, living in different circumstances of domestic overcrowding were visited at fortnightly intervals. The date of onset and the clinical nature of upper respiratory infectious experienced by each member of the family were charted on a time scale marked off in days. Family epidemics of acute coryza-or common colds-were thus available for analysis. This data was adopted by Heasman and Reid (1961) for illustration.

3.6 Software

Analysis was prepared through the R programming software version 3.5.1. This is because R software is an open source software and best suited with tidy verse tools for handling statistical simulations and analysis. The package *fitODBOD* was a key package used in the analysis due to its specific tools in analysing and modelling McGBB data. Library *CODA* in *R* language was also used to check the convergence diagnostics of and implement the HPD intervals.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Bayesian Estimates

The main objective of this study was to develop a Bayesian framework by employing the MCMC methods, specifically Metropolis-Hasting step within Gibbs for the estimation of McGBB parameters. The point estimates, Bayesian credible intervals and HPD intervals were computed. Using the algorithm in section 3.6, small samples from of sizes $k = 25$ and $k = 29$ and large samples of sizes $k = 500$ and $k = 1000$ were simulated. Samples generated from the marginal posteriors of the parameters were visually presented as trace plots. The results obtained are presented in section 4.1.1 for the small samples and section 4.1.2 for the large samples graphically while parameter estimates for both large and small samples in section 4.1.3.

4.1.1 Bayesian Plots for Small Samples

For this study small samples of sizes $k = 25$ and $k = 29$ were employed, these were McGBB variables simulated by setting the true parameter values at $(\alpha, \beta, \gamma) = (0.5, 0.5, 0{,}5)$ respectively from the algorithm highlighted in section 3.6. The results obtained are explained and illustrated below;

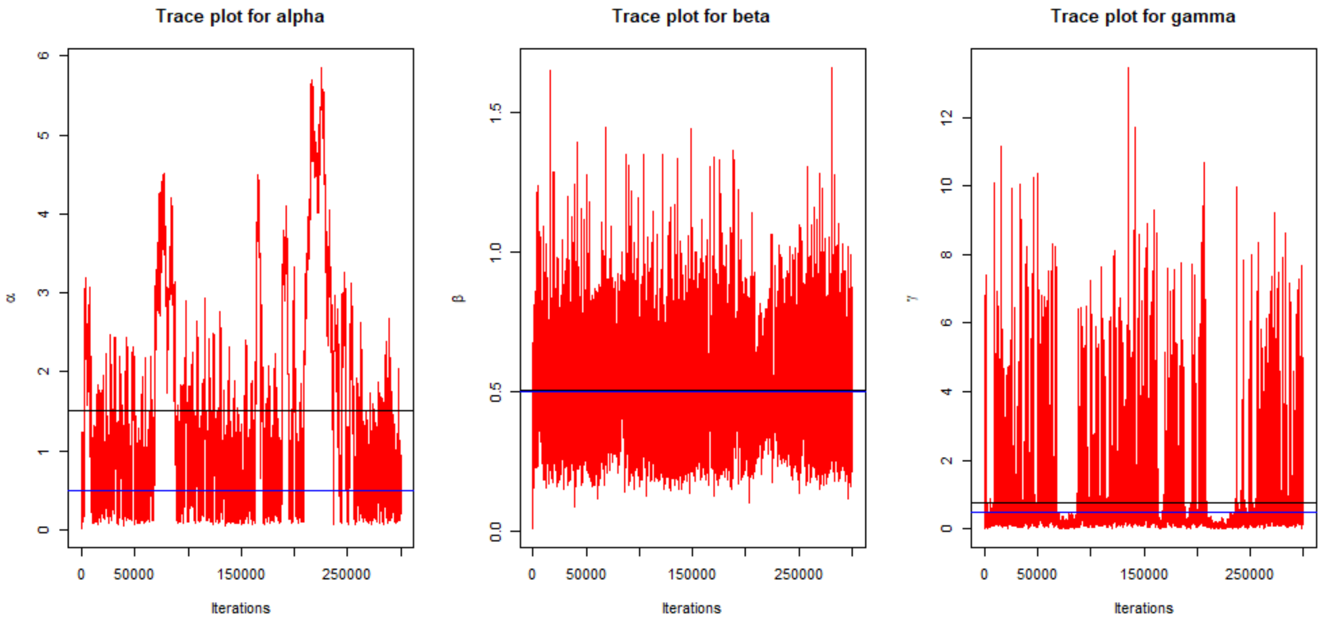


Figure 1: Trace plots for the parameters α, β and γ when the sample size k=25

Figure 1 shows the trace plots for each parameter, that is the samples generated for each marginal posterior of each parameter. The starting points were set at $(0.1, 0.1, 0.1)$ for the parameters (α, β, γ). The proposal distributions were chosen as normal for α and β, and exponential for γ. The values of sigma for the proposal distributions were set at $(\alpha, \beta, \gamma) = (0.1, 0.1, 0.5)$. Small sample sizes provide little information about the parameters (Button *et al.,* 2013). Therefore picking higher values of sigma ensures more information is available and allows for proper mixing of chains. The first ten thousand values of three hundred thousand values sampled were discarded as the burn-in-period. Burn-in-period is weeding out the first values of the chains before the chain stabilizes (Plummer *et al.,* 2006). The blue horizontal line shows the true parameter value while the black horizintal line shows the parameter estimate that was given by the mean of the sample generated while excluding the burn-in-period. From the figure, all the trace plots have converged. When trace plots follow a horizontal pattern or band , and the horizontal line fitted touches on the band then the trace plot has converged (Nylander *et al.,* 2008). Further, it can be seen that the variation in the parameter α β and γ were high. The horizontal lines on the plot for α shows a bigger difference between the true parameter value and the estimate. However, for β and γ the horizontal lines are close to each other which is an indication that the estimates were relatively close to the true value.

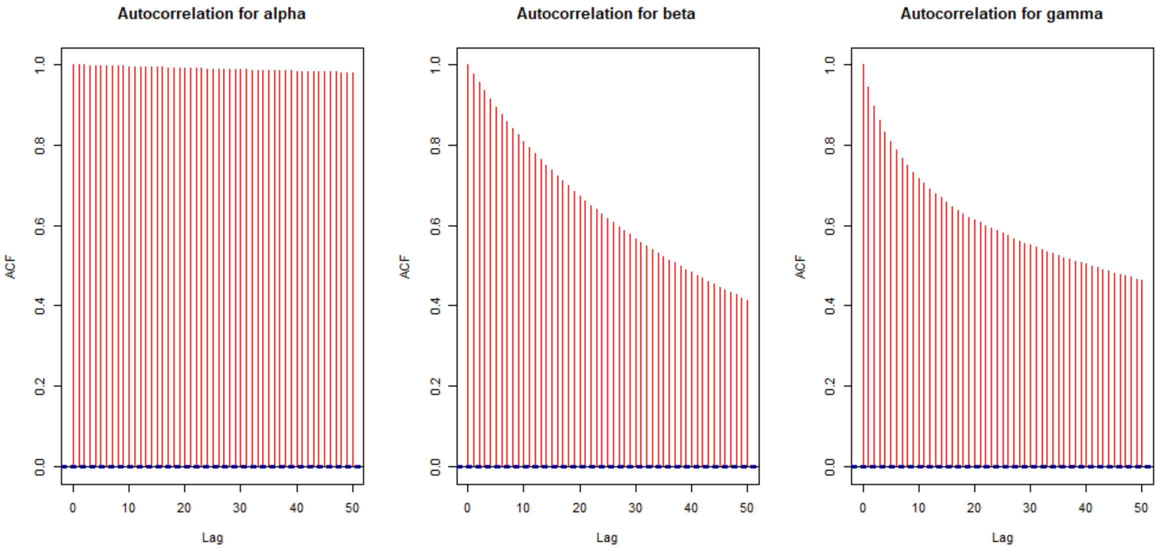


Figure 2: Autocorrelation plots for α, β and γ when k=25

Figure 2 shows the autocorrelation plots for the three parameters.The autocorrelation plots were used to evaluate the randomness of the sampled values. The figure shows strong autocorrelation for α that is almost one. There was average autocorrelation that is decreasing for β and γ with β having the lowest correlation. This shows that α was not as random in the search process of the sampled candidates while β and γ was very random in the search process of the sampled candidates that come from their marginal posterior densities.

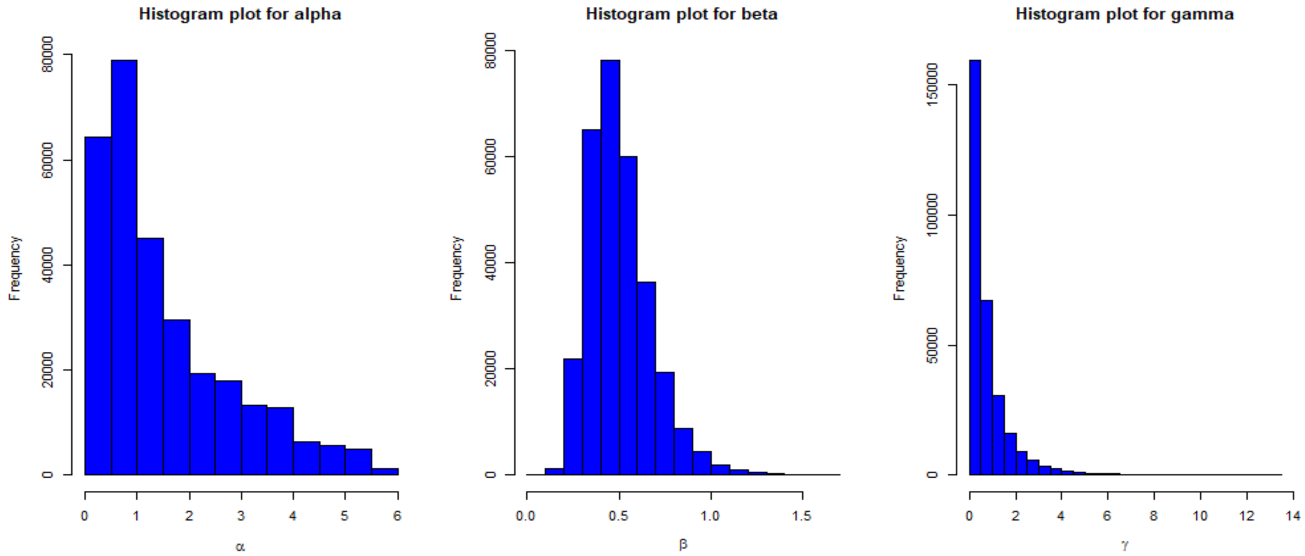


Figure 3: Histogram plots for α, β and γ when k=25

Figure 3 shows the histogram plots showing the distribution of the parameters α, β and γ. The shape parameters take a wide range of shapes. From the figure when $k = 25$, α and γ take a skewed shape of the distribution, however, γ is more skewed than α. β takes a relatively centred shaped however, it is not completely normal.

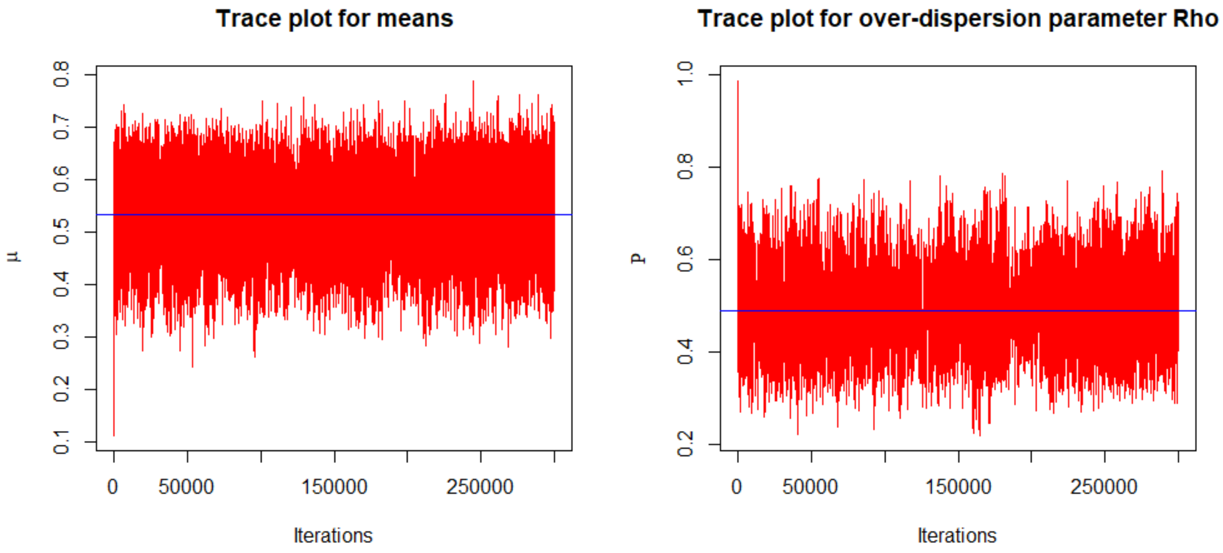


Figure 4: Trace plots for means and overdispersion parameter ρ

Figure 4 shows the trace plots of the different means and over-dispersion parameters computed for each combination of the parameter estimates for all the sampled values of the marginal posteriors. From the over-dispersion plot there was evidence of average over-dispersion in the data. The blue horizontal line in both plots shows on average the mean value and overdispersion value that was obtained across all the means and over-dispersions, that is the estimate of the mean was a value between 0.5 and 0.6, while the estimate of rho was a value between 0.4 and 0.6.

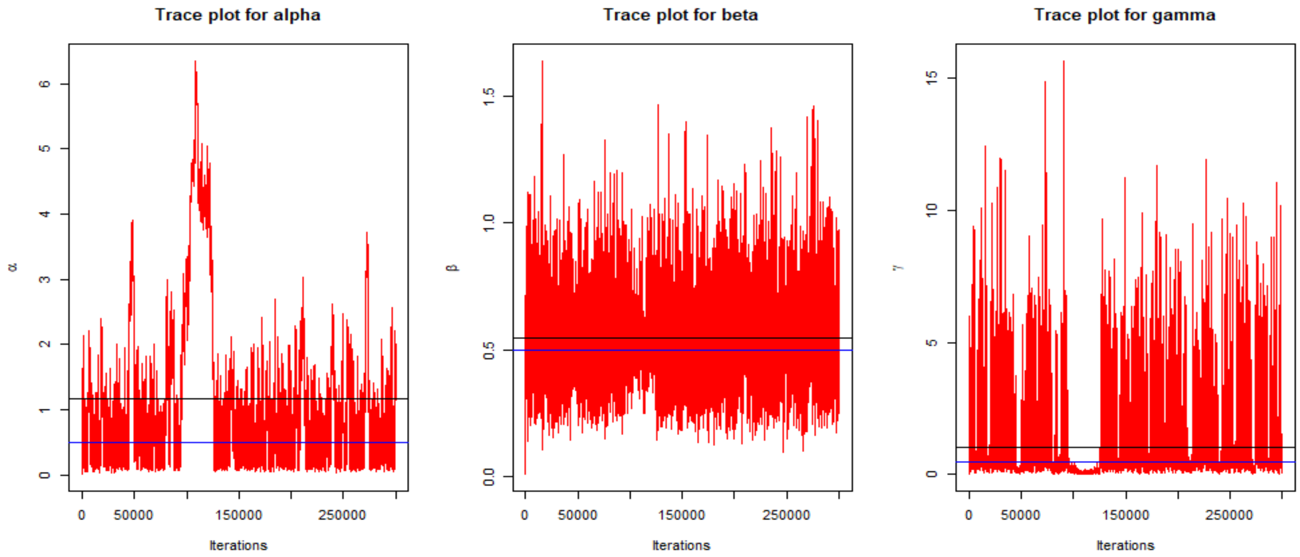


Figure 5: Trace plots for the parameters α, β and γ when the sample size k=29

Figure 5 shows the different trace plots generated for the parameters α, β and γ from the marginal posterior sampling when a sample size of size $k = 29$ was simulated from McGBB distribution 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 (α, β, γ). The values of sigma for the proposal distribution were set at $(\alpha, \beta, \gamma) = (0.5, 0.1, 8)$ respectively. The sigma values are scale parameter values for each of the proposal distribution used for this study. In this case, the choice was to use sigma values that are slightly higher since this would ensure more information is available given that the choice of priors was flat prior and the sample was a samll sample size. Hence these sigma values played a role in populating the information also. All the trace plots converged. However, for the trace plot for α and γ shows that at some point the chain got locked in a local maxima and local minima respectively, but was able toget out of the locked areas. However, from the figures,it can be noted that the parameter α and γ were slow in converging, while β was fast in converging. The horizontal lines are within the band of values sampled hence in line with literature from Gelman and Shirley

(2011). Further, it can be seen that the variation in the parameter α and γ was high, and slightly lower for the parameter β. 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. For β the estimates recovered were close to the true parameter values however for the parameter α and γ the estimate was not close to the true parameter value.

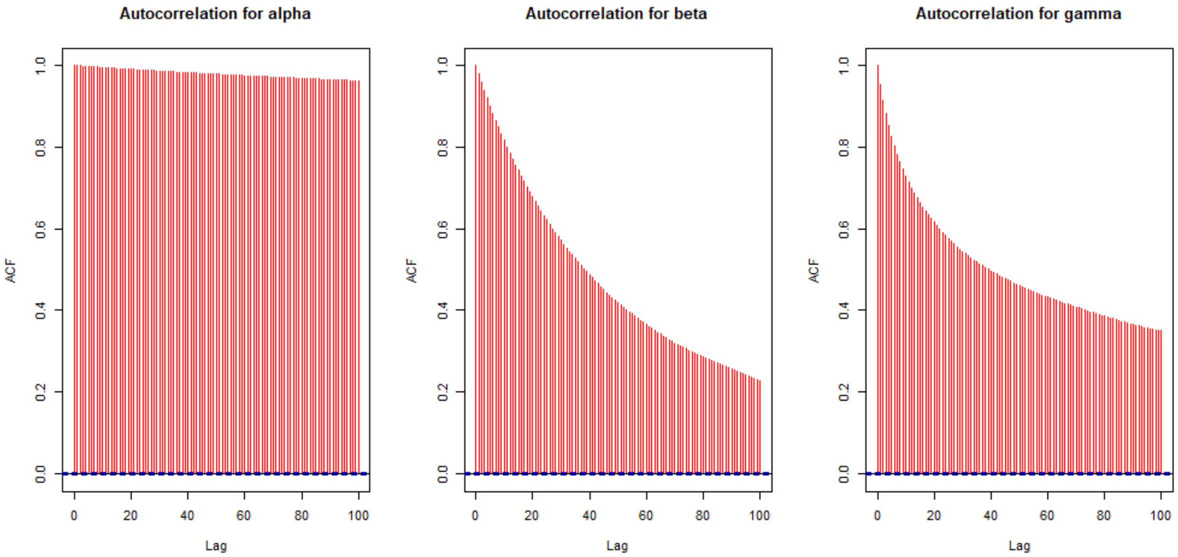


Figure 6: Autocorrelation plots for α, β and γ when k=29

Figure 6 shows there was high autocorrelation for α, low autocorrelation for the parameter β and it was average for the parameter γ. This shows that β was the most random in sampling the values from its marginal posterior density. The high autocorrelation in the trace plot for α is an indication of why the trace plot was slow in converging. The low and decreasing auto-correlation in the parameter β is evidence that the convergence was fast and there was good mixing of the chain. This also applies to the parameter γ. From the high autocorrelation in the parameter α, it shows that poor mixing of chains that could have been caused by value of the choice of scale parameter used to generate these samples.

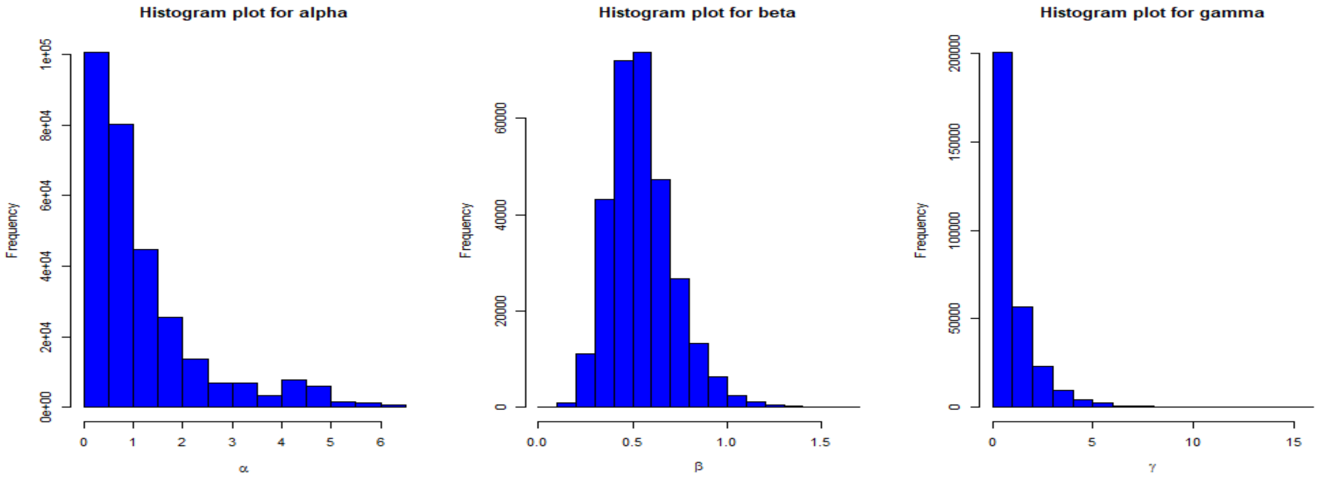


Figure 7: Histogram plots for α, β and γ when k=29

Figure 7 shows the distribution of α is skewed however flater compared to when the sample size k=25, the distribution of β is taking a normal direction while gamma is sharply skewed more to the right.

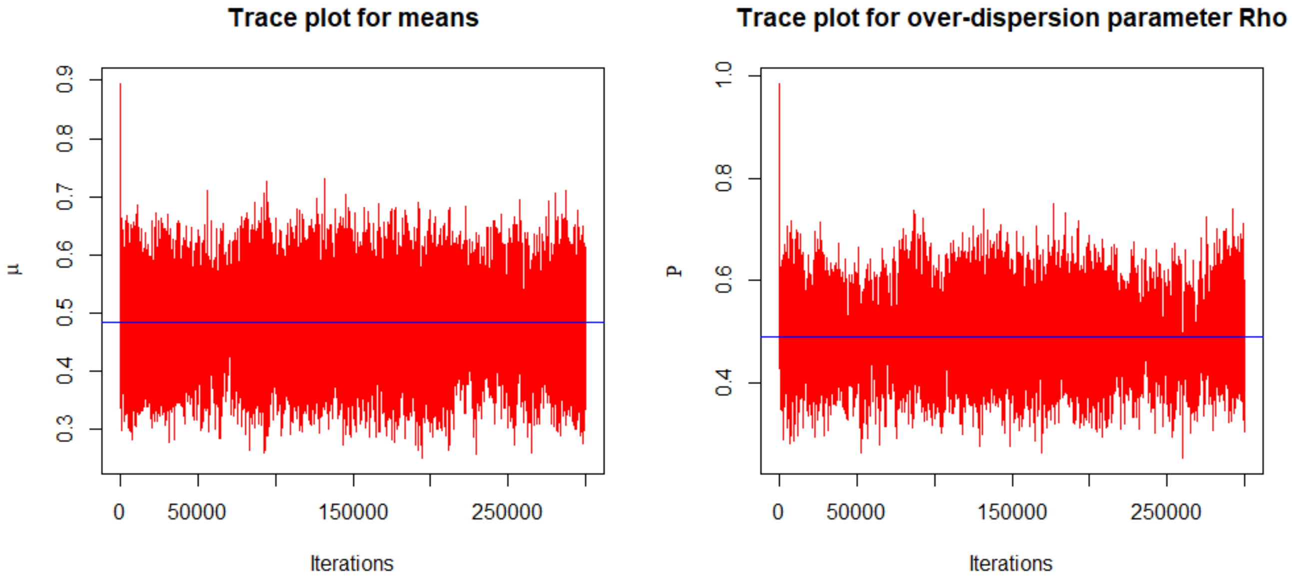


Figure 8: Trace plots for means and overdispersion parameter *ρ* computed from the estimates generated

Figure 8 shows the plots show each mean value and over-dispersion parameter computed at every estimate generated from the sampled values. From the figure the over-dispersion parameter was on average 0.5, while the mean was on average less than 0.5, as indicated by the blue horizontal line.

4.1.2 Bayesian Plots for Large Samples

An evaluation of the performance of the Bayesian method for large samples of sizes $k = 500$ and $k = 1000$ simulated from McGBB was done. The visual results in form of trace plots, correlation plots and histogram plots are presented below:

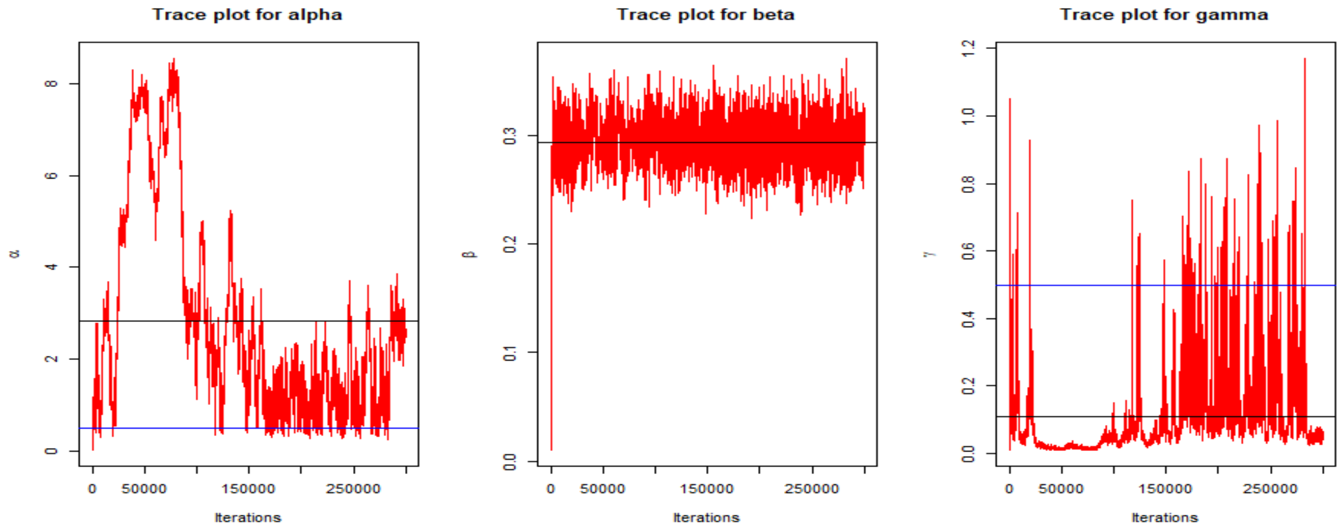

Trace plot for alpha
Trace plot for beta
Trace plot for gamma

Figure 9: Trace plot for α, β and γ when k=500

Figure 9 shows the different trace plots from the posterior sampling when $k = 500$ was simulated from McGBB distribution while setting 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 (α, β, γ). The values of sigma for the proposal distribution were set at $(\alpha, \beta, \gamma) = (0.1, 0.1, 0.5)$ respectively. Sigma values in this case were selected at lower values compared to the small sample size because, large samples provided more information already, pupulating the bayesian procedure with more information from the sigma values would bring about poor mixing of the chains. It can be seen that the burn-in-period is evident in the trace plot for α. From the figures all the plots have converged, however α and γ converge slowly compared to β. The variation is high in the plots for α and γ and low for β. The blue horizontal line shows true parameter value while the black horizontal line shows estimate of the parameters. From the figure there is a big difference in the blue and black horizontal lines for α and γ.

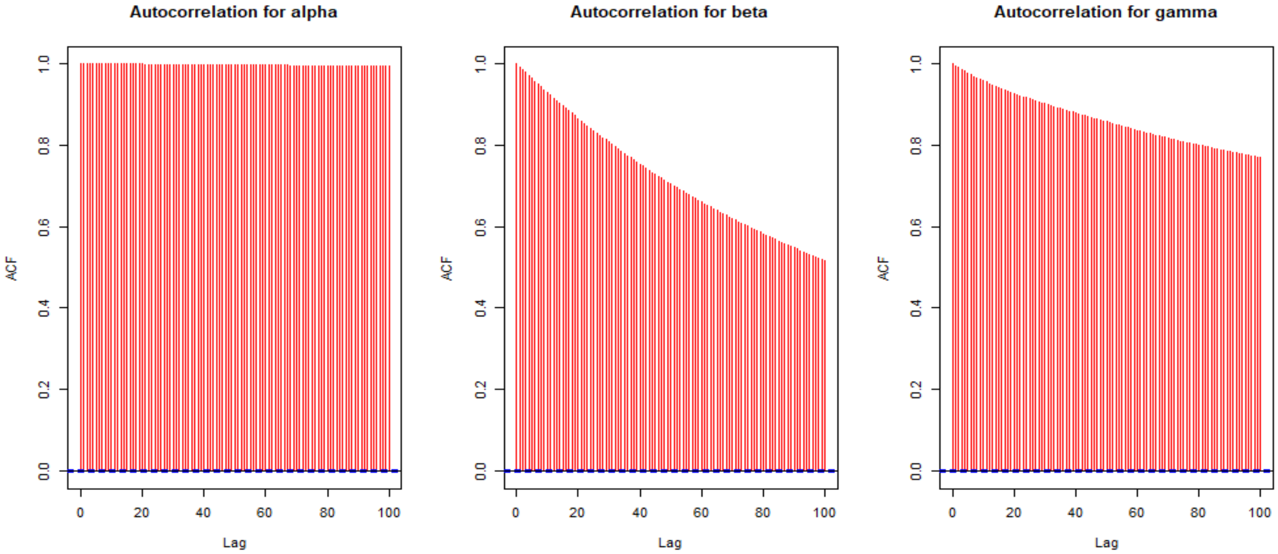


Figure 10: Autocorrelation plots for α, β and γ

Figure 10 shows the autocorrelation of all the parameters. From the figure autocorrelation was low for the parameter β, high for both α and γ. However, comparing with small samples the autocorrelation was relatively high for all the parameters as opposed to when small samples were used. This shows that β was random in sampling the values that follow its marginal posterior density compared to α and γ.

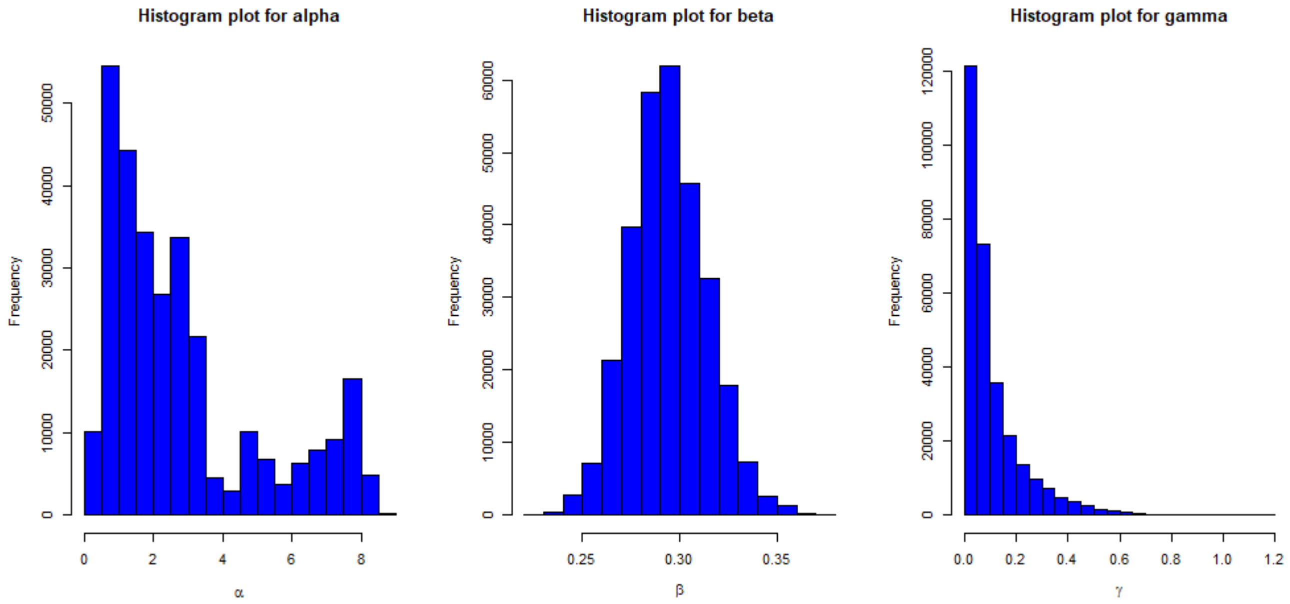


Figure 11: Histogram plots for α, β and γ

Figure 11 shows different distributions obtained for the three parameters. It can be noticed that β takes a normal distribution while γ a skewed distribution. The distribution space for α does not seem to be a smooth space however, it also takes a skewed distribution.

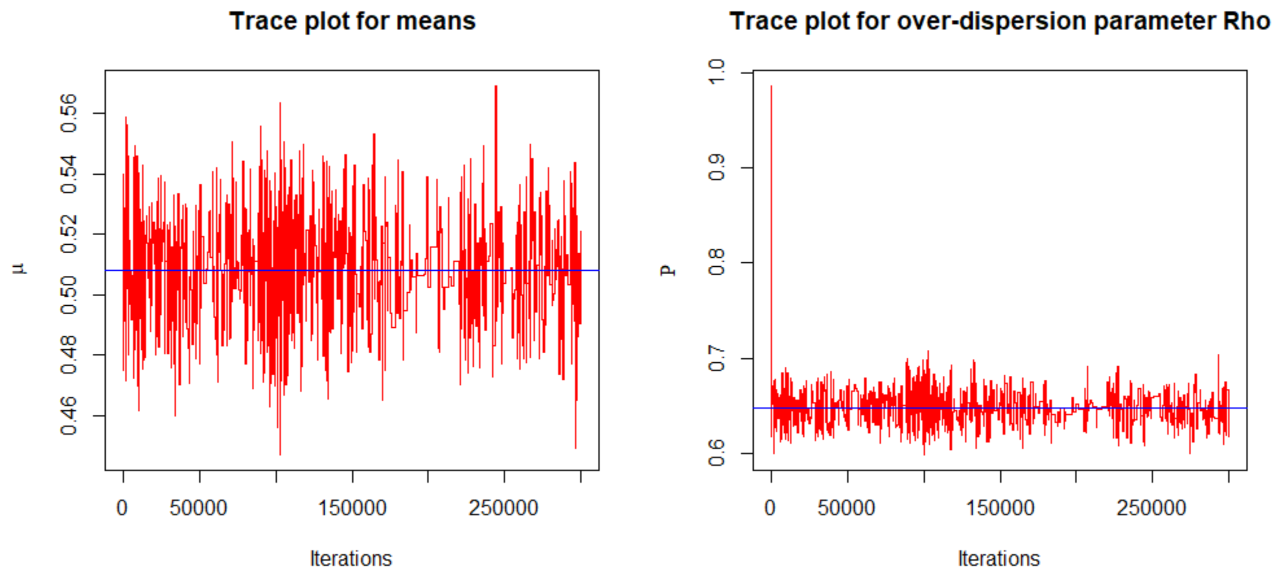
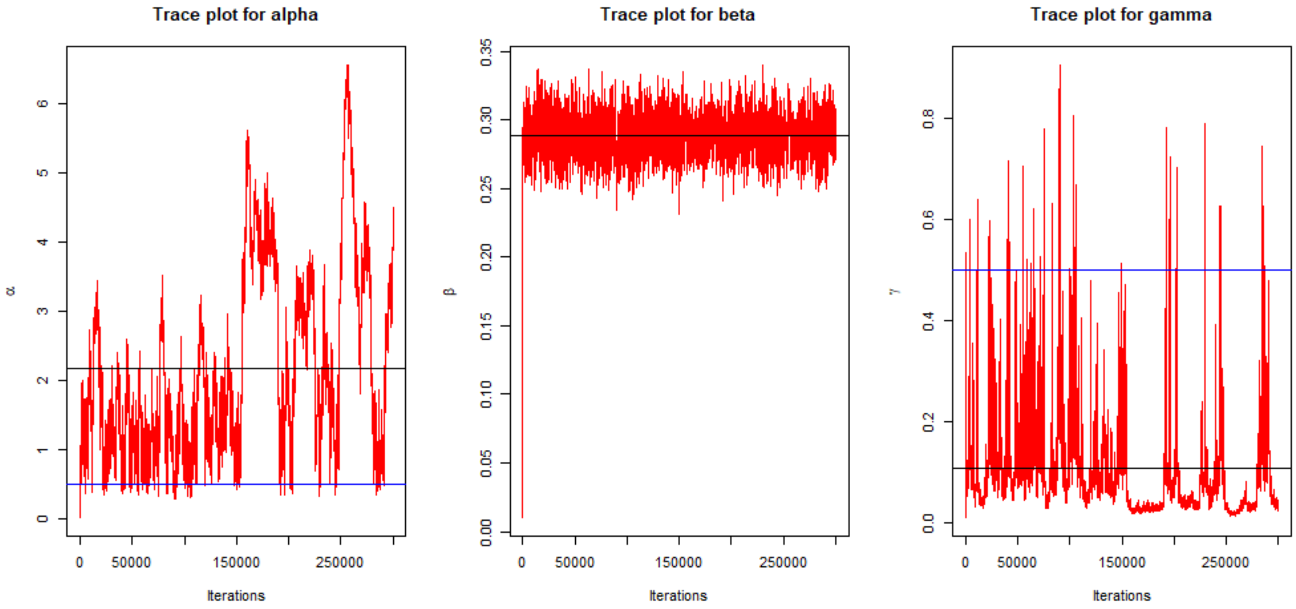


Figure 12: Trace plots for Mean and over-dispersion parameter ρ

Figure 12 shows over-dispersion and mean computed at every combination of estimate generated from the chains. From the figure it can be noticed that the on average the estimate for mean was a value between 0.5 and 0.6, while over-dispersion was a value between 0.4 and 0.6. Over-dispersion is higher compared to when the sample size is small, because of the increased sample size.


Trace plot for alpha

Trace plot for beta

Trace plot for gamma

Figure 13: Trace plot for α, β and γ when k=1000

Figure 13 shows the different trace plots 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 sigma values in this sample size were also chosen at smaller values, since this was a large sample data set, hence containing more information, therefore, the goal was to provide less information to enusre proper mixing of the chains (Robert &Casella, 2013). 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. The trace plots have converged however α and γ were slow in converging. From the figures its can be noted that there was a big difference observed from the blue and black horizontal lines in the plots for α and γ. Moreover the variance in the values sampled is also high for α and γ and low for β.

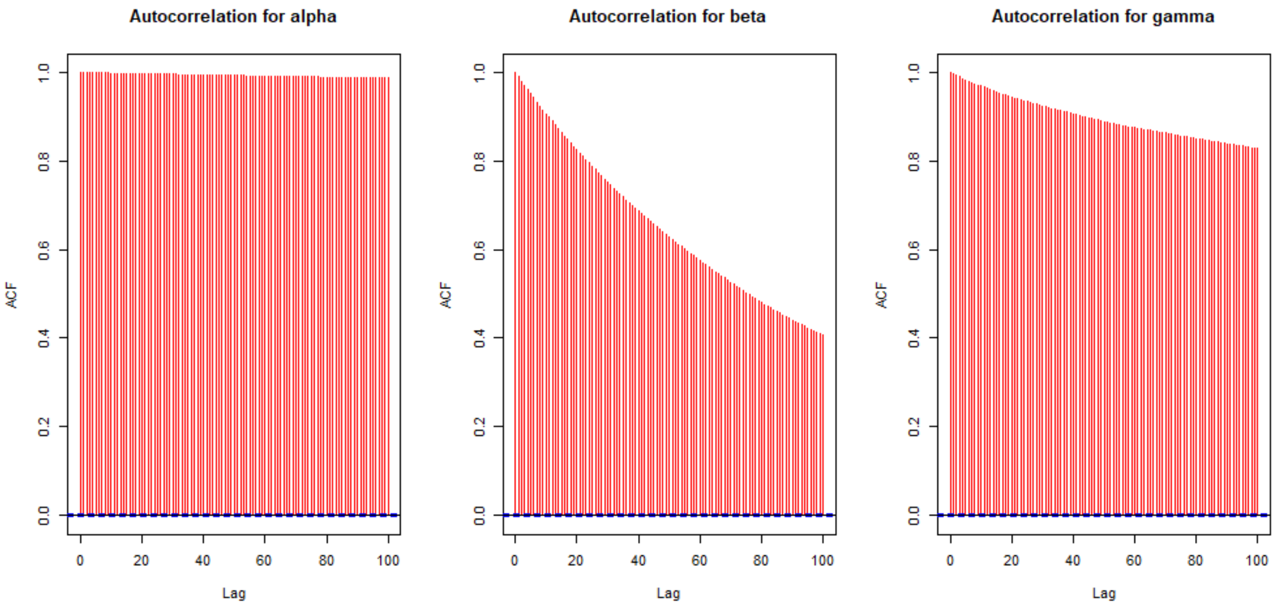


Figure 14: Autocorrelation plots for α, β and γ when k=1000

Figure 14 shows autocorrelation was relatively average for the parameter β, and high for α and γ. This can be seen from the behaviour of the traceplots. Comparing the autocorrelation plots for when the sample size is $k = 25$, it can be seen that autocorrelation has increased for sample size $k = 1000$. This shows that α and γ were not as random in the search process of the sampled candidates while β was very random in the search process of the sampled candidates that come from their marginal posterior densities.

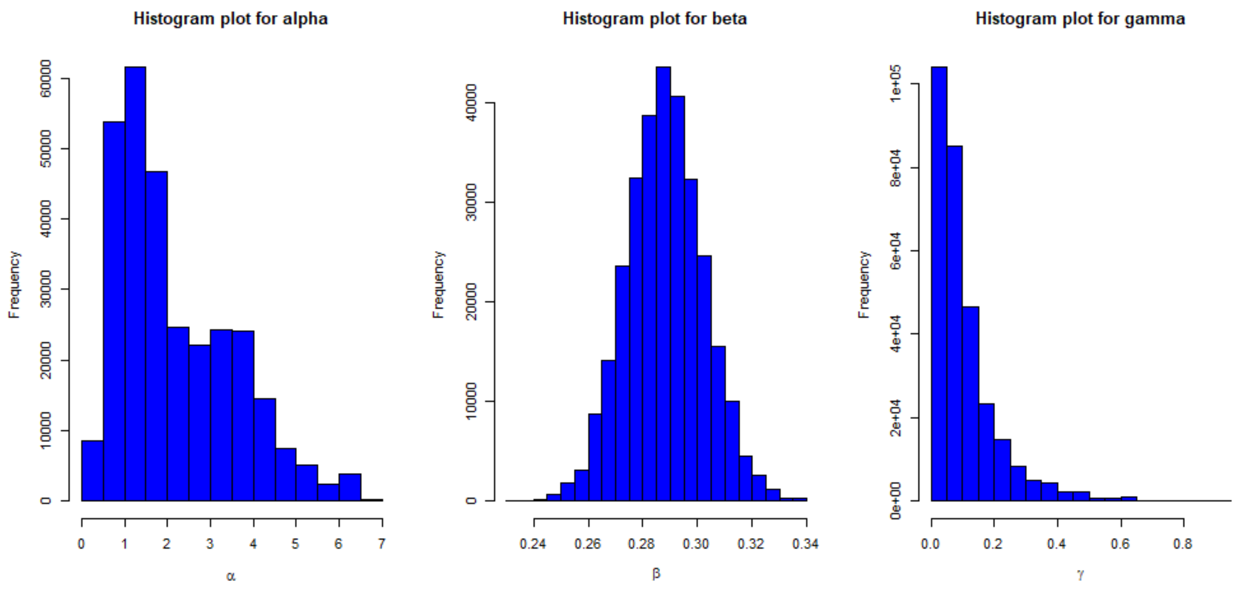


Figure 15: Histogram plots for α, β and γ when k=1000

Figure 15 shows the shape of α is skewed so is γ. However, γ is more skewed in shape than α. β 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.

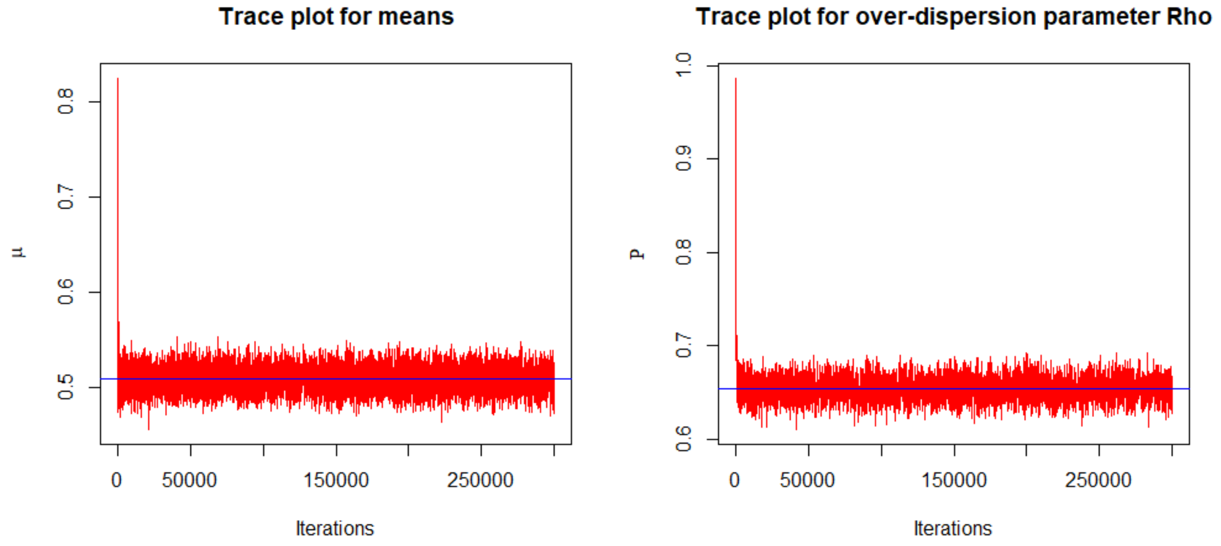


Figure 16: Trace plots for Mean and over-dispersion parameter ρ

Figure 16 shows the means and over-dispersion for every combonation of the estimates. Comparing the mean and over-dispersion plots for small samples and when the sample size is large it can be noticed that when the sample size is increased it can be noticed that there was an increase in over-dispersion also. 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. As the sample size increases so does the over-dispersion increase (Endo, 2020).

4.1.3 Bayesian Point and Interval Estimates

Point estimates were obtained as the mean of every individual marginal posterior sample for each parameter while discarding the first ten thousand values out of three hundred thousand values as the burn-in-period. The credible regions and HPD intervals were also computed.

Table 6: Bayesian Point Estimates

Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	**Acceptance rate%**
25	1.2812	0.4624	0.2159	24.507
29	2.5314	0.7139	0.2069	23.29
500	0.4696	0.4598	1.0384	2.8343
1000	0.2178	0.4109	0.3573	2.25

Table 6 shows the estimates were obtained from the marginal posterior samples generated. True values that we ought to recover for the three parameters were set at $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.5)$ respectively. The acceptance rates of the sampled candidate values were relatively within the accepatble limits for the small samples that is 24.507% for k = 25 and 23.29% for k = 29. The acceptable limits of the acceptance rates should be between 13- 50 % (Entezari, 2018). With large samples the acceptance rates was very low that is 2.8343% for k = 500 and 2.25% for k = 1000, however since the trace plots converged, the estimates used were still relevant. The acceptance rate shows the candidate values that were accepted out of the overall iterations performed. From the table it can be noted that for the parameter α the estimates improved as the sample size increase.This is is clear since in samll sample sizes the estimates were 1.2812 and 2.5314 when k = 25 and k = 29 respectively, while in large sample sizes k = 500 and k = 1000 the estimates were 0.4696 and 0.2178 respectively showing they were close t the true parameter value of 0.5 compared to the latter. The estimates for the parameter β were close to the true parameter value across all sample sizes. These estimates were ;0.4624, 0.7139, 0.4598 and 0.4109 for k = 25, 29, 50 and 1000 respectively.while the parameter γ had estimates close to the true parameter value compared to α. The estimates of γ were; 0.21, 0.2069, 1.0384 and 0.3573 for k = 25, 29, 500 and 1000 respectively.

Table 7: Bayesian 95 % credible regions

Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	(0.4167,1.8936)	(0.2233,0.8088)	(0.0529,0.5312)
29	(0.2176, 4.7035)	(0.3017, 1.2586)	(0.0515, 0.7397)
500	(0.0904,1.3581)	(0.3309,0.5667)	(0.1363,3.4954)
1000	(0.02006,0.9974)	(0.2712,0.5567)	(0.0039,1.67005

Table 7 shows the credible regions for all the sample sizes. From the table, 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 (α, β, γ) will lie within their respective credible regions computed.

Table 8: Bayesian HPD interval

Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	(0.4025,1.8979)	(0.1957,0.7598)	(0.0289,0.4681)
29	(0.1419,4.5364)	(0.2579, 1.1929)	(0.0222, 0.6212)
500	(0.0642,1.1707)	(0.3379,0.5718)	(0.0793,2.9573)
1000	(0.0145,0.8497)	(0.2712,0.5566)	(0.001005,1.3652)

Table 8 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. The HPD interval identify the most probable values. From the table each parameter had a unique HPD interval. The HPD interval is used as the interval estimate if it is unimodal since it consist of the most probable values (Maneerat *et al.,* 2020).

4.2 Maximum Likelihood Estimates

The study sought to examine the performance of maximum likelihood estimation method for small sample sizes and large sample size variables that follow a McGBB distribution. Point and interval estimates were obtained following the procedure stated in section 3.2. In order to examine the performance of the maximum likelihood method of estimation samples that were used in the Bayesian method were also used for the maximum likelihood estimation. To obtain the point estimates the function in library *fitODBOD* that highlights the maximum likelihood estimation procedure of the McGBB was applied. To compute the coverage probabilities of each parameter estimate one thousand iterations were used. The results for the point estimates obtained from the small samples simulated are presented in the table 6.

Table 9: Maximum Likelihood point Estimates

Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	4.122	0.521	0.0788
29	0.09758	0.4769	5.2125
500	0.1508	0.4069	1.9407
1000	1.2738	0.5184	0.1507

Table 9 shows parameter estimates for α and γ are far from the true parameter value that was set at 0.5. The point estimates for the parameter α are; 4.122, 0.09758 for k =25 and k = 29 that is small sample sizes and 0.1508, 1.2738 for k = 500 and k = 1000 in large sample sizes. For α estimates in the small sample sizes were far from the true parameter value 0.5 compared to the estimates in the large sample size. The estimates for the parameter γ were 0.0788, 5.2125, 1.9407, 0.1597 for k = 25, 29, 500 and 1000 respectively. For parameter γ the estimates in small samples were far from the true parameter value 0.5 compared to the estimates from the large samples. Parameter estimates for β are closer to the true parameter value which was also set at 0.5. This behaviour was also observed with the Bayesian estimates. To examine the behaviour of the estimates confidence intervals were computed using the hessian matrix generated as explained in section 3.2.

Table 10: Maximum Likelihood 95% confidence intervals

Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	(3.7466, 4.4778)	(0, 1.5418)	(0, 25.5613)
29	(0, 0.4741)	(0, 1.2339)	(0, 26.4381)
500	(0, 2.2327)	(0.4027, 0.6454)	(0, 2.3948)
1000	(0, 8.5882)	(0.4049, 0.6294)	(0, 5.8762)

Table 10 shows the confidence intervals for the parameter α in the case of small samples does not include the true parameter value that is when k = 25 and when k = 29. The intervals for the parameter γ are wider for small samples compared to those of large samples, however in both sets of sample sizes they all include the true parameter value unlike those of the parameter α. The intervals for beta are relatively short compared to the confidence intervals of the parameters α and γ and they all include the true parameter value. The parameters of McGBB α, β, and γ can only assume positive values, therefore negative side of the intervals especially with maximum likelihood method are truncated.

Table 11: Coverage probabilities

Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	0.5090	0.7510	0.5890
29	0.5200	0.7600	0.6120
500	0.7310	0.9480	0.7690
1000	0.7770	0.9470	0.7610

Table 11 shows the coverage probabilities computed. Coverage probabilities show the proportion of times that the intervals contained the true parameter value, over repeated sampling (Ledell *et al.,* 2015). Coverage probabilities are obtained after repeated sampling is done. In this study, different samples were generated however under same conditions. Repeated sampling was done one thousand times. The proportion of times the confidence interval included the true parameter value were recorded. From the table it can be observed that as the sample size increases, the chances of the true parameter value to be included in the confidence were increasing across all the parameters. However, the coverage probabilities of the parameter α and γ were low compared to the coverage probabilities for the parameter β. Low coverage probabilities mean the estimates are biased or anti-conservative standard error estimates or both (Link *et al.,* 2018). The coverage probabilities for the parameter α were the lowest in all the sample sizes compared to the other parameters β and γ. For the parameter α coverage probabilities were 0.5090, 0.5200, 0.7310 and 0.7770 for k = 25, 29, 500 and 1000 respectively. To interpret the coverage probabilities for instance when k = 25, it shows that the interval estimates contained the true parameter value 50.90% of the time out of 1000 times that repeated sampling was done. Parameter β had good coverage probabilities overall compared to α and γ. The coverage probabilities were 0.7510, 077600, 0.9480 and 0.9470 for k = 25, 29, 500 and 1000 respectively. For the parameter γ the coverage probabilities were 0.5890, 0.6120, 0.7690 and 0.7610 for k = 25, 29, 500 and 1000 respectively.

4.3 Comparison of Point Estimates Using Standard Errors

To evaluate the performance of the maximum likelihood method and Bayesian method in small samples and large samples, standard errors were used to evaluate the performance of point estimates and confidence length the performance of credible regions and confidence interval. The results obtained were compared sand presented in table 12 below.

Table 12: Standard Errors for Bayesian and Maximum likelihood point estimates

	Bayesian standard errors			**Maximum Likelihood**		
Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	1.4014	0.1699	0.9246	41.9503	0.1894	0.8495
29	9.1440	0.1719	0.0684	0.1900	0.3862	10.8294
500	0.5908	0.0222	0.3308	0.1108	0.0865	1.6809
1000	1.9704	0.01403	0.0801	2.7907	0.0316	0.3807

Table 12 shows the standard errors for the parameter estimates. Standard errors are used to evaluate the precision of the estimates (Imbens & Kolesar, 2016). From the table the parameter α were relatively higher compared to the parameters β and γ in the Bayesian method. The standard errors for the parameter β are low and decrease as the sample size increases in both the Bayesian and maximum likelihood method. The lowest standard error in the maximum likelihood method was 0.0316 that was obtained for the estimate when the sample size k = 1000, the largest sample size considered in this study, hence in line with literature from (Pace & Lesage, 2016). This shows that the parameter estimates for the parameter β were a true representation of true parameter value (King & Roberts, 2015). The standard errors for the parameter γ were also relatively low showing that the estimates were a good representation of the true parameter values in the Bayesian method. In the maximum likelihood method the standard error for α was very high for k = 25, and that of γ was also highest when k = 29, which are both small sample cases. The instability in the parameter α in the Bayesian method could have been caused by the choice of proposal distribution and the values of the scale parameters of the chosen proposal distributions. However, for the maximum likelihood method the behaviour of the parameter α was also unstable as is shown by the high standard errors obtained. This shows that the maximum likelihood method may not perform well in small samples and is line in literature from Li (2016).

Table 13: Length of the 95 % credible regions

	Bayesian			MLE		
Sample size (k)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
25	1.4769	0.5855	0.4783	0.7312	1.5418	25.5613
29	4.4859	0.9569	0.6882	0.4741	1.2339	26.4381
500	1.2677	0.2358	3.3591	2.2327	0.2427	2.3948
1000	0.9773	0.2855	1.6697	8.5882	0.2245	5.8762

Table 13 shows the length of each credible region across all the sample sizes and the length of the confidence intervals. From the table the lengths of the credible intervals were relatively short across all the parameters. However, the parameter α had slightly higher lengths of its credible interval for both sample sizes that is 1.4769, 4.4859, 1.2677 and 0.9773 for k = 25, 29, 500 and 1000 respectively compared to β and γ. The parameter β had credible lengths across all the sample sizes that is; 0.5855, 0.9569, 0.2358 and 0.2855 for k = 25, 29, 500 and 1000 respectively. The credible lengths for the parameter γ were 0.4783, 0.6882, 3.3591 and 1.6697 for k = 25, 29, 500 and 1000 respectively, showing that small sample sizes for γ had shorter credible intervals than large sample sizes. Small sample sizes credible intervals were slightly higher than larger sample size credible intervals for the parameter β and γ. Short lengths show a good precision in the estimates and also the method (Svec *et al.,* 2015). From the maximum likelihood confidence lengths the parameter γ had the widest for small samples and relatively short for large samples.

They were; 25.5613, 26.4381, 2.3948 and 5.8762 for k = 25, 29, 500 and 1000 respectively. However, the parameter γ had the widest intervals in the maximum likelihood method compared to the Bayesian method of estimation. The confidence length for β are short for all samples. The confidence lengths for the parameter α are relatively short too across all the sample sizes that is 1.5418, 1.2339, 0.2427 and 0.2245 for k = 25, 29, 500 and 1000 respectively. The confidence lengths for β were;1.5418, 1.2339, 0.2427 and 0.2245 for k = 25, 29, 500 and 1000 respectively Comparing the Bayesian credible regions lengths to the confidence interval lengths the Bayesian regions had shorter lengths across all the sample sizes (Sobhi & Soliman, 2016).

4.4 Real Data Set

Different sets of data from *fitODBOD* package were used in this study for illustration. The description of these datasets is highlighted in section 3.5. The alcohol data for example has been used by Manoj *et al.* (2013) and Janiffer *et al.* (2014) to find the estimates using the MLE and estimating functions methods. Therefore, this study used the Bayesian method as well to compute the estimates and compared this with the estimates from the other methods.

Table 14: Point Estimates for alcohol dataset

	Bayesian			**MLE**		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Alcohol week 1	0.1038	0.3031	21.8814	0.0381	0.1995	23.39
Alcohol week 2	0.0793	0.3619	32.2472	0.0451	0.3103	23.39

Table 14 shows the point estimates obtained for the parameters after applying the Bayesian framework and the maximum likelihood method for the alcohol dataset. The estimates from alcohol week 1 from the Bayesian method are 0.1038, 0.3031 and 21.8814 while the estimates of the maximum likelihood method for alcohol week 1 are 0.0381, 0.1995 and 23.39 for the parameters α, β and γ respectively. The estimates from alcohol week 2 from the Bayesian method are 0.0793, 0.3619 and 32.2472 while the estimates of the maximum likelihood method for alcohol week 2 are 0.0451, 0.3103 and 23.39 for the parameters α, β and γ respectively. The estimates from both methods of estimation do not vary a lot, however, in order to justify this standard errors were also computed.

Table 15: Standard errors for Alcohol dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Alcohol week 1	9.9839	0.00732	0.00832	0.00276	0.0404	0.000261
Alcohol week 2	10.1964	0.00873	0.00425	0.00322	0.0702	0.000639

Table 15 shows the standard errors for the parameter estimates of the Bayesian and MLE methods. The standard errors for the Bayesian method in alcohol week 1 are 9.9839, 0.00732 and 0.00832 while the maximum likelihood standard errors for alcohol week 1 were 0.00276, 0.0404 and 0.000261 for the parameters α, β and γ respectively. The standard errors for the Bayesian method in alcohol week 2 are 10.1964, 0.00873 and 0.00425 while the maximum likelihood standard errors for alcohol week 2 were 0.00322, 0.0702 and 0.000639 for the parameters α, β and γ respectively. From the table all the parameter estimates for MLE had low standard errors, showing that they were true representation of the true parameter values. The standard errors for the parameter β and γ were low in both methods showing that the estimates were true representation of the true parameter value. The standard error for β were low for Bayesian that the MLE but for the parameters α and γ the standard errors for MLE were low than the Bayesian. Standard error for the parameter γ was the lowest at alcohol week 2 that is 0.000639.

Table 16: Point Estimates for Chromosome dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Chromosome association	3.773	1.7606	0.9852	4.3391	4.4459	1.3231

From the table point estimates were obtained in both methods of estimation. The estimate for the parameters α was 3.733 in Bayesian and 4.3391 in the maximum likelihood method. These two did not differ much. The estimate for the parameter β was 1.7606 in Bayesian while in maximum likelihood method it was 4.4459, these two methods differed much in the estimate for this parameter. The estimate for the parameter γ was 0.9852 in the Bayesian method and 1.3231 in the maximum likelihood method, these two did not differ much also. To check for the precision of estimates their respective standard errors were also computed.

Table 17: Standard errors of chromosome dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Chromosome association	3.141	0.1575	0.9748	15.311	2.178	3.858

From the table the overall standard errors in the Bayesian method were lower than those of the maximum likelihood method. The parameter α had the highest standard error in the maximum likelihood method. Parameters β and γ had low standard errors in the Bayesian method of 0.1575 and 0.9748 respectively. For this dataset of chromosome association the Bayesian method outperformed the maximum likelihood method of estimation.

Table 18: Point estimates for Course dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Course observation	4.592	1.0976	1.0279	1.8152	0.9656	1.001

Table 18 shows the estimates after the Bayesian and maximum likelihood method were used to estimate the parameters. From the table the estimate for the parameter α was 4.592 in the Bayesian method and 1.8152 in the maximum likelihood method, this shows a big difference in the two methods. The estimates for the parameter β were 1.0976 for the Bayesian method and 0.9656 in the maximum likelihood method of estimation, these two estimates did not differ much. The estimates for the parameter γ were 1.0279 in the Bayesian method and 1.001 in the maximum likelihood method, these two did not also differ much. To further compare the estimates standard errors were obtained.

Table 19: Standard errors for Course dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Course observation	3.2401	0.3169	1.3727	4.128	0.2284	1.5670

From table 19 the standard errors for the parameter α were high in both methods of estimation that is 3.2401 in the Bayesian method and 4.128 in the maximum likelihood method. For the parameter β the standard errors were low in both methods of estimation. For the parameter γ the standard errors were also slightly higher compared to those of β that is 1.3727 in Bayesian method and 1.5670 in the maximum likelihood method. In this dataset, generally the Bayesian method outperformed the maximum likelihood method.

Table 20: Point estimates of the Epidemic cold dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Families	0.1194	0.1491	1.4891	8.3983	72.3319	0.5639
Father	0.6929	0.2159	0.2885	0.5659	0.9345	0.5689
Mother	0.4980	0.1309	0.4477	5.077	14.284	0.5203
School child	0.1338	0.1615	1.435	9.6220	5.3451	0.6617

From the table 20 the estimates showed a big difference in both methods of estimation. For instance in families the parameter β had very high estimates of 72.3319, which is a very big difference from the one obtained in the Bayesian method of 0.1491.For data observed from fathers the Bayesian estimates for the parameters α, β and γ were 0.6929,0.2159 and 0.2885 respectively, while the maximum likelihood estimates were 0.5659, 0.9345 and 0.5689 respectively. They two methods did not differ in the estimates for the case of fathers as much. In the data from mothers, the estimates in the Bayesian method varied from those in the maximum likelihood method a lot. The maximum likelihood estimates were higher than the Bayesian estimates. In the school child data, the estimates in the maximum likelihood method were also higher than those of the maximum likelihood method.

Table 21: Standard errors of the epidemic cold dataset

	Bayesian			MLE		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
Families	0.0137	0.0186	0.12980	5.8036	22.2732	0.0425
Father	0.05029	0.0386	0.0312	1.6041	0.7890	2.8686
Mother	0.6011	0.03284	0.4374	5.0763	10.5860	4.8326
School child	0.0116	0.0096	0.1696	1.180	5.9833	2.8678

In families, the standard errors in the Bayesian method were very low showing that the estimates obtained through the Bayesian method are a true representation of the true parameter values. The Maximum likelihood standard errors were high especially for the parameter β which is also reflected in the value of estimate that was gotten. From the standard errors in the father's row, the Bayesian standard errors were very low hence a true representation of the true parameter values, while those of the maximum likelihood method were slightly higher than those of the Bayesian method. Therefore in this case the Bayesian method performed better. Bayesian standard errors were lower than the maximum likelihood standard errors in all the three parameters for the case of mothers. The standard errors in the Bayesian method for the school child data were very low that is 0.0116, 0.0096, and 0.1696 respectively while those of the maximum likelihood method were slightly higher at 1.180, 5.9833 and 2.8678 respectively. Generally in this dataset, the Bayesian method of estimation outperformed the maximum likelihood method of estimation. This is evident by the low standard errors obtained in the Bayesian method compared to those obtained in the maximum likelihood method of estimation. Therefore, the Bayesian estimates were better and were a true representation of the true parameter values.

CHAPTER FIVE

SUMMARRY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

Estimation of the parameters of a distribution that is used to model a given dataset is a very key thing in statistical inference. This study focused on the Bayesian estimation of the shape parameters of a McGBB. The Bayesian method was compared to the maximum likelihood method. The Bayesian method applied Markov Chain Monte Carlo methods in particular the Metropolis-Hasting step within Gibbs sampler to sample from the unknown form of joint posterior. The trace plots graphs used to show the behaviour of the samples generated over a specified number of iterations showed convergence for all the sample sizes. However, in small sample sizes all the parameters were showing high variations, while for large samples the parameter β had low variations in its trace plots. As the sample size increased the over-dispersion in the dataset increases as well. The parameter β was able to randomly search the sample space as it can be noted from its decreasing autocorrelation. The choice of proposal distribution for the Metropolis-Hasting algorithm majorly affected the mixing of the chains especially since the distributions chosen had scale parameters. A wrong choice of proposal distribution will lead to biased results due to poor mixing of the chains. The proposal distribution were selected from a class of distributions that only assume positive values since the shape parameters of McGBB can only take positive values, for the case of nornmal distribution, the truncated form of normal distribution was used. The standard errors were obtained to evaluate the precision of the point estimates. Parameters β and γ had relatively low standard errors across all the sample sizes in the Bayesian framework using simulated data, however γ had higher standard errors across all sample sizes in the maximum likelihood method in the simulated data. When both methods of estimation that is MLE and Bayesian were applied to a real datasets, the parameter β had the lowest standard errors in the Bayesian framework. However, α and γ had lower standard errors in the MLE method. Different real data application were used and in most of the datasets applied for this study the Bayesian method provided better estimates compared tom the maximum likelihood method In the different sets of the data used in this study the Bayesian approach to estimation outperformed the maximum likelihood method as illustrated through the low standard errors obtained. As the sample size increased the standard errors for beta decreased for Bayesian and MLE. Confidence intervals and lengths were computed as well. In small samples it is notable that the Bayesian credible regions had the shortest lengths. The shapes of the parameter α and γ are skewed while β tends to take a normal shape. This may inform farther research in exploring more on the Bayesian framework of estimation by using informative priors. The class of distributions with shapes that are skewed and are positive may be used as priors for the parameter α and γ, while the truncated normal be used as prior for the parameter β.

5.2 Conclusions

This research work explored a Bayesian framework for the McGBB distribution. Through the implementation of the framework the study was able to obtain the point estimates, credible regions and HPD intervals of the McGBB distribution. The study also computed the point estimates, confidence intervals and coverage probabilities for the MLE method. Standard errors and confidence lengths were used to compare Bayesian framework to the MLE method.

i. For the point estimates Bayesian estimates had low standard errors for the case of simulated data. When applied to a real dataset, α and γ had low standard errors for the MLE method but β had lower standard errors for Bayesian framework.
ii. The credible regions had shorter widths compared to the confidence intervals.

5.3 Recommendations

This study explored one method of sampling from the posterior form of the distribution that is the metropolis hasting step within the Gibbs method while blocking. This can be extended further to explore sampling from the posterior using other the MCMC approaches for instance; Metropolis-Hating algorithm, Gibbs sampling algorithm and Hamilton Monte Carlo and therefore determining which approach converges fast. Other techniques of MCMC of sampling from the posterior that do not include the use of a proposal distribution can be explored to monitor the behaviour of the parameter α since its high deviations from the true parameter value could have been caused by the choice of proposal distribution used for this study. This study also used flat priors, studies can extend this to the case of using informative priors. The determination of the proposal distribution did not follow a specific procedure. The Bayesian framework developed can be extended to other beta-type generated distributions. The Bayesian method has only been compared with the MLE, there is need to compare this framework with the estimating functions method.

REFERENCES

Adcock, C.J. (1997). Sample size determination: A Review. *Statistician, 46,* 261– 283.

Alexander, C., Cordeiro, G., Ortega, E., & Sarabia, J. (2012). Generalized beta-generated distributions. *Computational Statistics and Data Analysis*, *56*(6), 1880-1897.

Aldeni, M., Lee, C., & Famoye, F. (2017). Families of distributions arising from the quantile of generalized lambda distribution. *Journal of Statistical Distributions and Applications*, *4*(1), 1-18.

Armero, C., & Bayarri, M. J. (1994). Prior assessments for prediction in queues. *Journal of the Royal Statistical Society, 43*(1)*,* 139-153.

Aksoy, H. K., & Guner, A. (2015). A Bayesian Approach to Demand Estimation. *Procedia Economics and Finance*, *26*, 777-784.

Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Applied Statistics 27,* 162-167.

Alquier, P., Friel, N., Everitt, R., & Boland, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Statistics and Computing*, *26*(2), 29-47.

Andreu, c., De Freitas, N., Doucet, A., & Jordan, M. I. (2003). An Introduction to MCMC for Machine Learning. *Journal of Machine Learning, 50*(2), 5-43.

Auger, A., & Hansen, N. (2016). Linear convergence of comparison-based step-size adaptive randomized search via stability of Markov chains. *SIAM Journal on Optimization, 26*(3), 1589-1624.

Berg, B. A. (2017). A brief history of the introduction of generalized ensembles to Markov chain Monte Carlo simulations. *The European Physical Journal Special Topics*, *226*(4), 551-565.

Bera, A. K., Bilias, Y., & Simlai, P. (2006). Estimating functions and equations: An essay on historical developments with applications to econometrics. *Palgrave Journal of Econometrics*, *1*(11), 427-476.

Beck, J. L., & Taflanidis, A. (2013). Prior and posterior robust stochastic predictions for dynamical systems using probability logic. *International Journal for Uncertainty Quantification*, *3*(4)*,* 56-68.

Bichanga, L, A., Islam, A.S., & Orawo L, (2015)**.** C (α) Tests for Testing Homogeneity of Proportions in Presence of McDonald Generalized Beta- Binomial Over-Dispersion. *International Journal of Science and Research, 4*(11)*,* 206-214.

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Journal of Technological Forecasting and Social Change*, *130*, 99-113.

Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian statistics*. John Wiley & Sons.

Bodnar, O., Link, A., Arendacká, B., Possolo, A., & Elster, C. (2017). Bayesian estimation in random effects meta-analysis using a non-informative prior. *Journal of Statistics in Medicine*, *36*(2), 378-399.

Boddy, C. R. (2016). Sample size for qualitative research. *Qualitative Market Research: An International Journal*.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Journal of Nature Reviews Neuroscience*, *14*(5), 365-376.

Chadha, V.K., (2006). Sample size determination In Health studies. *National Tuberculosis Institute Bulletin, 42*(5), 55 – 62.

Crowder, M. J. (1978). Beta-binomial anova for proportions. *Applied Statistics, 27,* 34-37.

Cappe, O. & Robert, C.P. (2000). Markov Chain Monte Carlo: 10 years and still running! *Journal of the American Statistical Association, 95,* 1282–86.

Chen, M., & Shao, Q. (1999). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics, 8* (1), 69-92.

Chib, S. & Ramamurthy S., (2010). Tailored randomized block MCMC methods with applications to DSGE models. *Journal of Econometrics, 155*(31), 19–38.

Chotikapanich, D. & Griffiths W.E., (2000). Posterior distributions for the Gini coefficient using grouped data. *Australian and New Zealand Journal of Statistics, 42,* 383–392.

Crowder, M. J. (1978). Beta-binomial Anova for proportions. *Journal of Applied Statistics*, *27,* 34-37.

Cumming, G., & Callin-jageman, R. (2016). *Introduction to the new statistics; Estimation, open science and beyond*. Routledge Publishing.

Cochran, W.G., (1977). *Sampling Techniques* (Vol. 2). John Wiley & Sons.

Dodwell, T. J., Ketelsen, C., Scheichl, R., & 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*(1), 1075-1108.

Entezari, R. (2018). *Bayesian Computations via MCMC, with applications to Big Data and Spatial Data* [Doctoral dissertation, University of Toronto].

Eng, J.M., (2003). Sample size estimation. How Many Individuals should be studied. *Journal of Radiology, 277*(17), 309 – 313.

Endo, A. (2020). Estimating the over dispersion in COVID-19 transmission using outbreak sizes outside China. *Well-come Open Research*, *5*(1), 32-41

Fienberg, S. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, *1*(1), 1-40.

Feit, E. M., Feinberg, F. M., & Lenk, P. J. (2017). Bayesian analysis. *Advanced Methods for Modeling Markets,* 493-554.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference* (Vol. 2). Boca Raton: Taylor & Francis.

Gareth O. R., Andrew G., & Walter R G. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, *7*(1), 110–120.

Gareth O. R., & Jeffrey S.R. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Journal of Statistical Science*, *16*(4), 351–367.

Griffiths, D. (1973). Maximum Likelihood Estimation for the Beta-Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of a Disease. *Biometrics, 29*(4), 637-648.

Gelman, A., & Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics*, *4*(1), 625-631.

Gordy, M.B., (1988a). Computationally convenient distributional assumptions for common-value auctions. *Computational Economics 12*(1), 61-78.

Gordy, M.B., (1988b). A generalization of generalized beta distributions. *Journal in Computational Economics 12*(2), 18-25.

Gelman, A., & Shirley, K. (2011). Inference from simulations and monitoring convergence. *Journal of Markov Chain Monte Carlo Algorithms*, *6*(4), 163-174.

Greene, J. D. (1970). Personal Media Probabilities. *Journal of Advertising Research*, *10*(1), 12-20.

Haan, C. T. (1977). *Statistical methods in hydrology*. The Iowa State University Press.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337-350.

Hammersley, J. (2013). *Monte Carlo Methods*. Springer Science and Business Media.

Heasman, M. A. & Reid, D. D. (1961). Theory and observation in family epidemics of the common cold. *Journal of Statistical methods in Medicine, 15*(1), 12-16.

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Journal of Review of Economics and Statistics*, *98*(4), 701-712.

Israel, G.D., (1992). *Determining sample size in Program Evaluation and organizational development.* University of Florida Press.

Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big Data analytics and Computational Intelligence for Cyber–Physical Systems: Recent trends and state of the art applications. *Journal of Future Generation Computer Systems*, *105*(4), 766-778.

Janiffer, N., Islam, A., & Luke, O. (2014). Estimating Equations for Estimation of McDonald Generalized Beta— Binomial Parameters. *Open Journal of Statistics*, *4*(9), 702-709.

Johansen, A.M., Doucet, A., & Davy, M. (2008). Practical methods for maximum likelihood estimation in latent variables models. *Journal of Statistics and Computing, 18*(1), *47- 57.*

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (Vol. 444). John Wiley & Sons.

Jon, K.M., Lars, E.M., & Katie, D.M., (2003). Considerations on sample size and power calculations in Randomized clinical trials. *The Journal of Arthroscopic and Related surgery, 19*(5), 997-999.

Jeffrey S. R. (2006). *A First Look at Rigorous Probability Theory*. World Scientific Publishing Company.

Kakamu, K., & Nishino, H. (2018). Bayesian Estimation of Beta-type Distribution Parameters Based on Grouped Data. *Journal of Computational Economics, 10*(7), 55-62.

Karlis, D. & Xekalaki, E., 2008. The Polygonal Distribution*. Journal of Advances in Mathematical and Statistical Modelling, 8*(3), 21-33.

Kim, D. G., Lee, S. H., Cho, B. K., Byeon, D. H., Lee, J., & Lee, W. H. (2020). Statistical analysis for determining optimal sample size for living modified organism (LMO) seed detection. *Journal of Crop Science and Biotechnology*, *23*(1), 1-7.

Korteweg, A.G (2011). Markov chain Monte Carlo methods in corporate finance. *Journal of Finance Computation Methods, 6*(3), 103-114.

Kupper, L. L., & Haseman, J. K. (1978). The Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments. *Biometrics*, *34*, 69-75.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178-206.

Lee, J., & Sabavala, D. (1987). Bayesian Estimation and Prediction for the Beta-Binomial. Model. *Journal of Business and Economic Statistics, 5*(3), 357-367.

Lee, M. D., & Vanpaemel, W. (2018). Determining informative priors for cognitive models. *Psychonomic Bulletin and Review*, *25*(1), 114-127.

LeDell, E., Petersen, M., & van der Laan, M. (2015). Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Journal of Statistics*, *9*(1), 1583-1590.

Li, J., Nott, D. J., Fan, Y., & Sisson, S. A. (2017). Extending approximate Bayesian computation methods to high dimensions via a Gaussian copula model. *Computational Statistics and Data Analysis*, *106*, 77-89.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936-949.

Link, R. M., Schuldt, B., Choat, B., Jansen, S., & Cobb, A. R. (2018). Maximum-likelihood estimation of xylem vessel length distributions. *Journal of Theoretical Biology*, *455*, 329-341.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer Science & Business Media.

Manoj, C., Wijekoon, P., & 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*(2), 24-31.

Maneerat, P., Niwitpong, S. A., & Niwitpong, S. (2020). A Bayesian approach to construct confidence intervals for comparing the rainfall dispersion in Thailand. *Journal of Bayesian Statistics, 8,* 502-508.

Martino, S., & Riebler, A. (2019). Integrated nested Laplace approximations (inla). *Journal of the Royal Statistical Society*, *71*(2), 319–392.

Marmarelis, V. Z. (Ed.). (2012). *Advanced methods of physiological system modelling* (Vol. 2). Springer Science & Business Media.

Martin, R., & Liu, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, *108*(501)*,* 301-313.

Martin, R., & Liu, C. (2015). Conditional inferential models: combining information for prior-free probabilistic inference. *Journal of the Royal Statistical Society in Statistical Methodology*, *71*(3), 195-217.

Martinez, E. Z., Achcar, J. A., & Aragon, D. C. (2015). Parameter estimation of the beta-binomial distribution: an application using the SAS software. *Portuguese Journal of Statistical Sciences, 37*(3), 12-19.

McCullagh, P. & Nelder, J. A. (1989). *Generalised Linear Model* (Vol. 1). Chapman & Hall.

McDonald, J.B. & Xu, Y.J. (1995). A Generalization of the Beta Distribution with Applications. *Journal of Econometrics*, *66*(6), 133-152.

McDonald, J.B., (1984). Some generalized functions for the size distribution of income. *Econometrica, 52*(5), 647-664.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modelling Journal*, *23*(5), 750-773.

Mood, A., Graybill, F., & Duane, C. D. Boes (1974). *Introduction to the Theory of Statistics* (Vol. 1). Taylor & Francis.

Nadarajah, S., & Gupta, A. K., (2004). The Beta Frèchet Distribution. *Far East Journal of Theoretical Statistics, 14* (44), 15–24.

Nadarajah, S., & Kotz, S., (2006). The Beta Exponential Distribution. *Reliability Engineering and System Safety,* 689-697.

Nylander, J. A., Wilgenbusch, J. C., Warren, D. L., & Swofford, D. L. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetic. *Bioinformatics*, *24*(4), 581-583.

Otieno, J. (2008). *From the classical beta distribution to generalized beta distributions,* [Doctoral dissertation, University of Nairobi].

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal,* 7(4), 308-313.

Otieno, J. (2008). *From the classical beta distribution to generalized beta distributions* [Doctoral dissertation, University of Nairobi].

Ogal, R. (2012). *Binomial mixtures with continuous mixing priors* [Doctoral Dissertation, University of Nairobi].

Paul, S. R. (1982). Analysis of proportions of affected fetuses in tautological experiments. *Biometrics, 38*(10), 361-370.

Paul, S.R., 1985. A three-parameter generalization of the binomial distribution. *Communications in Statistics Theory and Methods, 14*(6), 1497-1506.

Pack, S. E. (1986). Hypothesis testing for proportions with over-dispersion. *Biometrics, 42*(5), 967-972.

Pace, R. K., & LeSage, J. P. (2016). Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. *Journal of Spatial Econometrics: Qualitative and Limited Dependent Variables*, *14*(7), 33-45.

Paul, S. R., & Islam, A. S. (1995). Analysis of proportions in the presence of over-/under-dispersion. *Biometrics, 38*(13), 1400-1410.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285-293.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news Journal*, *6*(1), 7-11.

Ralph, B.D., Steve, H., & Rajasekhar, R., (2002). Sample size Determination. *ILAR Journal, 43*(3), 207 – 213.

Russel, V.L., (2001). Some Practical Guidelines for effective sample size determination. *The American Statistician, 55*(12), 187 – 193.

Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Sean P. M., & Richard L. T. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between the sets of trials. *Journal of the Royal Statistical Society, 10*(4), 257-261.

Sprenger, J. (2018). The objectivity of subjective Bayesianism. *European Journal for Philosophy of Science, 8*(3), 539-558.

Sobhi, M. M. A., & Soliman, A. A. (2016). Estimation for the exponentiated Weibull model with adaptive Type-II progressive censored schemes. *Applied Mathematical Modelling*, *40*(2), 1180-1192.

Stokes, M., Chen, F., & Gunes, F. (2014). An introduction to Bayesian analysis with SAS/STAT® software. *In Proceedings of the SAS Global Forum 2014 Conference, SAS Institute Inc Cary, 24*(9), 87-95.

Stoner, O., Economou, T., & da Silva, G. D. M. (2019). A hierarchical framework for correcting under-reporting in count data. *Journal of the American Statistical Association, 15*(9), 215-223.

Sjölander, A., & Vansteelandt, S. (2019). Frequentist versus Bayesian approaches to multiple testing. *European Journal of Epidemiology*, *34*(9), 809-821.

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W., & Kubista, M. (2015). How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*, *3*(1), 9-16.

Tarima, S., & Flournoy, N. (2019). Asymptotic properties of maximum likelihood estimators with sample size recalculation. *Journal of Statistical Methods, 60*(2), 23-44.

Teimouri, M., Hoseini, S. M., & Nadarajah, S. (2013). Comparison of estimation methods for the Weibull distribution. *Journal of Statistics Theory, 47**(1),*** 93-109.

Wedderburn, R.M. (1974). Quasi-Likelihood Functions, Generalized Linear Models and the Gauss Newton Method. *Biometrics*, *61*(5), 439-447.

Wang, Y. X., Fienberg, S., & Smola, A. (2015, June). Privacy for free: Posterior sampling and stochastic gradient Monte Carlo. *International Journal on Machine Learning*, *8*(5), 2493-2502.

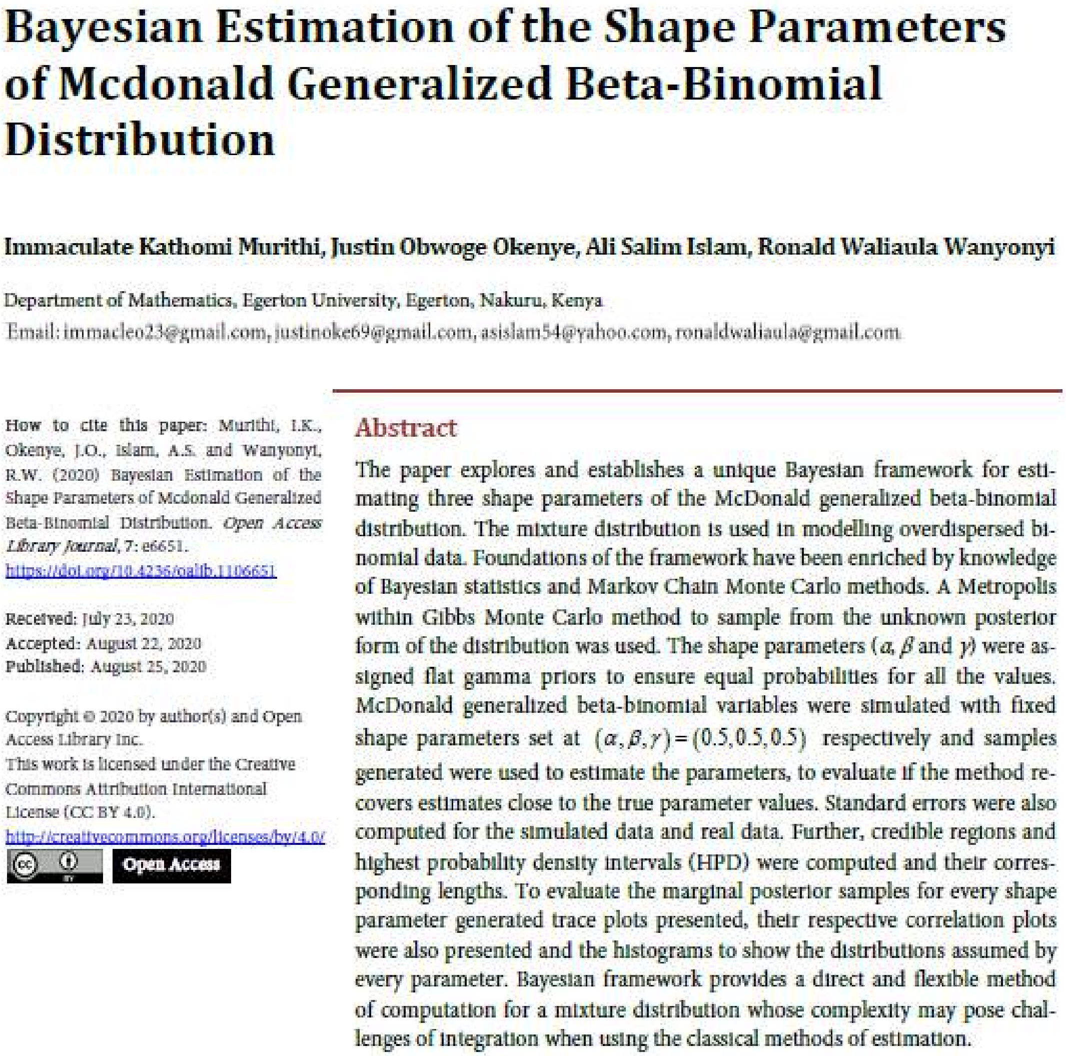
Withers, S. D. (2002). Quantitative methods: Bayesian inference, Bayesian thinking. *Progress in Human Geography*, *26*(4), 553-566.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913-934.

.

APPENDICES

Publication



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

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

Department of Mathematics, Egerton University, Egerton, Nakuru, Kenya
Email: immacleo23@gmail.com, justinoke69@gmail.com, asislam54@yahoo.com, ronaldwaliaula@gmail.com

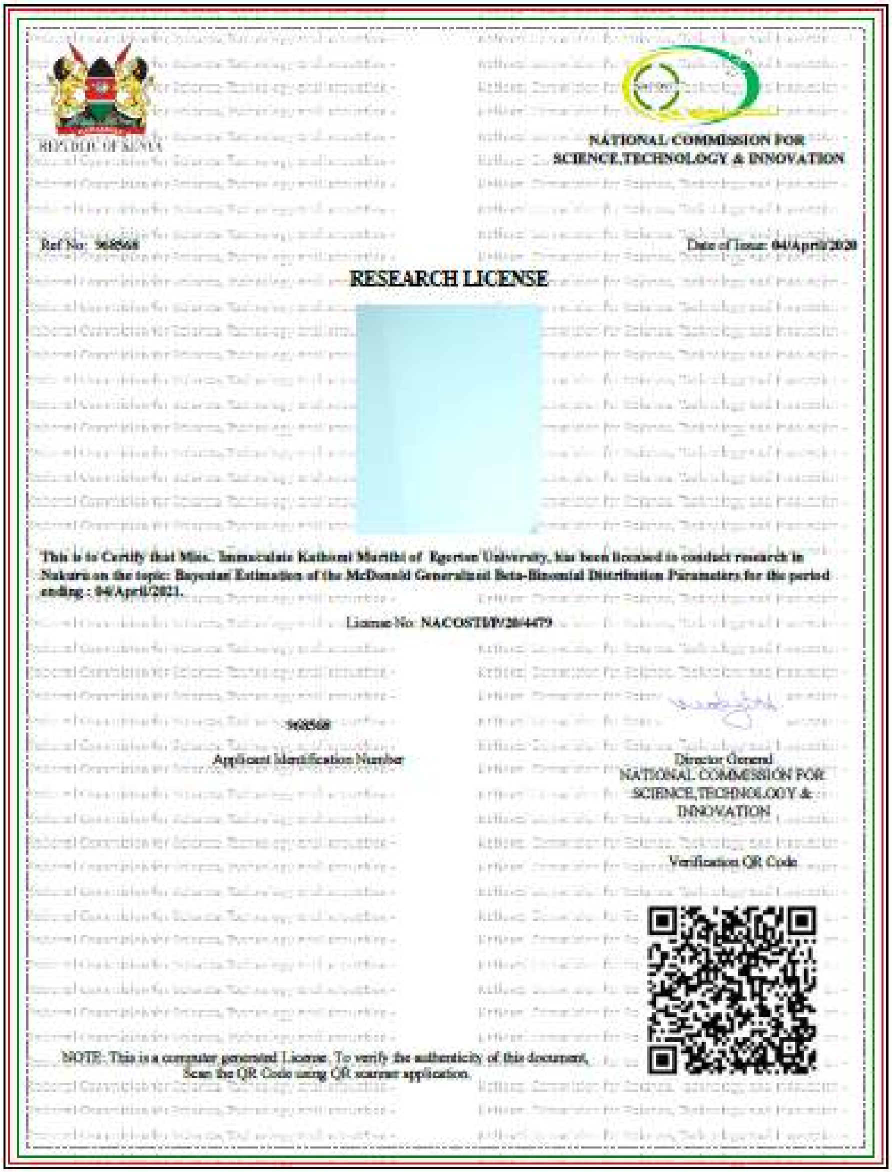
How to cite this paper: Murithi, I.K., Okenye, J.O., Islam, A.S. and Wanyonyi, R.W. (2020) Bayesian Estimation of the Shape Parameters of Mcdonald Generalized Beta-Binomial Distribution. *Open Access Library Journal*, 7: e6651.
https://doi.org/10.4236/oalib.1106651

Received: July 23, 2020
Accepted: August 22, 2020
Published: August 25, 2020

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 (α, β and γ) 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.

Research permit



REPUBLIC OF KENYA

NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 968568

Date of Issue: 04/April/2020

RESEARCH LICENSE

[redacted]

This is to Certify that Miss.. Immaculate Kathomi Murithi of Egerton University, has been licensed to conduct research in Nakuru on the topic: Bayesian Estimation of the McDonald Generalized Beta-Binomial Distribution Parameters for the period ending : 04/April/2021.

License No: NACOSTI/P/20/4479

968568
Applicant Identification Number

Director General
NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY &
INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document,
Scan the QR Code using QR scanner application.

Key R codes

Author: Immaculate Kathomi Murithi

5/18/2020: Report by R markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

simulating for simulating from the McGBB model given the parameters of the model (parameters are known).

```
McBBB=function(K,n,alpha,beta,gamma)
  K=number of trials
  n=number of possible outcomes
  X=rep(NA,K)
  B=(rbeta(K,alpha,beta))^(1/gamma)
  for (i in 1:K)
  {
    X[i]=rbinom(1,n,B[i])


  }
  return(X)
}
```

The likelihood function for McGBB model

You can also embed plots, for example:

```
library(truncnorm)

McGBBL<-function(x,a,b,c,fre,n){
  density=c()
  for( i in 0:n){
    j = 0:(n-i)
    term=sum(((-1)^j)*(choose(n-i,j))*(beta((a+((j+i)/c)),b)))
    vector.density=choose(n,i)*(1/beta(a,b))*term
    density[i+1]=vector.density
  }
  M=prod(density^fre)
```

```
  return(M)
}
```

The log likelihood function

```
McGBBLL1<-function(x,a,b,c,fre,n){
  MM=log(McGBBL(x,a,b,c,fre,n))
  return(MM)
}
```

```
McGBBLL<-function(a,b,c,fre,n){
  adopted and adjusted from Manoj et.al (2013)
  density=c()
  for( i in 0:n){
    j = 0:(n-i)
    term=sum(((-1)^j)*(choose(n-i,j))*(beta((a+((j+i)/c)),b)))
   # print(term)
    vector.density=choose(n,i)*(1/beta(a,b))*term
    density[i+1]=vector.density
  }
  M=sum(fre*log(density))
  return(M)
}
```

The logarithms of the full conditionals for theparameters of McGBB model updated as a block

```
L_Cond_abc=function(a,b,c,fre,n,a_a,b_a,a_b,b_b,a_c,b_c){
  MM=McGBBLL(a,b,c,fre,n)
  return(MM)
}
```

The bayesian code

```
BayesEst1B=function(x,n,fre,start,
                a_a,b_a,a_b,b_b,a_c,b_c,
                sig,K){
  The starting values
  K =number of interations
  (a_a,b_a,a_b,b_b,a_c,b_c)-hyperparameters
  sig= scale paramter for the proposal distribution
  x, n,fre-are as defined above
  Storage (the chains of the parameters)
  a=rep(NA,K)
  b=rep(NA,K)
  c=rep(NA,K)
 The starting values (initializing the chains)
  a[1]=start[1]
  b[1]=start[2]
  c[1]=start[3]
   check for acceptance rates
  c1=1
  for(k in 2:K){
    if (k%%10000==0){print(k)}

    a_prop=rnorm(1,a[k-1],sig[1]) #rexp(1,.5) #rtruncnorm(1,0,1,a[k-1],sig[1])#exp(rnorm(1,a[k-1],sig[1]))
    b_prop=rnorm(1,b[k-1],sig[2])#rtruncnorm(1,0,Inf,b[k-1],sig[2])#exp(rnorm(1,b[k-1],sig[2]))
    c_prop=rexp(1,sig[3]) #rnorm(1,c[k-1],sig[3])#rtruncnorm(1,0,Inf,c[k-1],sig[3])#exp(rnorm(1,c[k-1],sig[3]))

    u1=runif(1)
    log_R_abc1=L_Cond_abc(a_prop,b_prop,c_prop,fre,n,a_a,b_a,a_b,b_b,a_c,b_c)
    -log(dexp(c_prop,sig[3]))    #-log(dtruncnorm(a_prop,0,01,a[k-1],sig[1]))#-log(dtruncnorm(b_prop,0,Inf,b[k-1],sig[2]))-log(dtruncnorm(c_prop,0,Inf,c[k-1],sig[3]))
```

```
    log_R_abc2= L_Cond_abc(a[k-1],b[k-1],c[k-1],fre,n,a_a,b_a,a_b,b_b,a_c,b_c)
    -log(dexp(c[k-1],sig[3]))    #-log(dtruncnorm(a[k-1],0,1,a_prop,sig[1])) # -log(dtruncnorm(b[k-1],0,Inf,b_prop,sig[2]))-log(dtruncnorm(c[k-1],0,Inf,c_prop,sig[3]))

    log_R_abc=log_R_abc1-log_R_abc2
    if ((log(u1)<log_R_abc )&& !is.nan(log_R_abc) )
      {a[k]=a_prop;b[k]=b_prop;c[k]=c_prop
        c1=c1+1} else
        {a[k]=a[k-1];b[k]=b[k-1];c[k]=c[k-1]}
  }
  Pa=(c1/K)*100
 print(c1)

  return(list(a=a,b=b,c=c,Pa=Pa))
}
Frequecies

k=25

fre=c(7,1,1,3,2,4,3,4)
k=29
fre=c(7,7,2,3,1,1,4,4)
k=500
fre=c(195,59,35,40,25,33,35,78)
k=1000
fre=c(399,90,91,70,73,63,74,140)
alcohol week1
fre=c(47,54,43,40,40,41,39,95)
alcohol week2
fre=c(42,47,54,40,49,40,43,84)
start=c(.01,.01,.01)
BBB1=suppressWarnings(BayesEst1B(0:7,8,fre,start,10^-10,10^-10,
```

```
                        10^-10,10^-10,10^-10,10^-10,
                        c(.1,.01,10),300000))

plot(BBB1$a,type = 'l',col="red",xlab = "Iterations",
    ylab=expression(alpha),main = "Trace plot for alpha")
abline(h=c(0.5,mean(BBB1$a[-(1:1000)])),col=c("blue","black"))

plot(BBB1$b,type = 'l',col="red",xlab = "Iterations",
    ylab=expression(beta),main = "Trace plot for beta")
abline(h=c(0.5,mean(BBB1$b[-(1:1000)])),col=c("blue","black"))

plot(BBB1$c,type = 'l',col="red",xlab = "Iterations",
    ylab=expression(gamma),main = "Trace plot for gamma")
abline(h=c(0.5,mean(BBB1$c[-(1:1000)])),col=c("blue","black"))
```