



On Classical and Bayesian Reliability of Systems Using Bivariate Generalized Geometric Distribution

Nasir Abbas¹

Received: 31 August 2022 / Accepted: 2 May 2023 / Published online: 31 May 2023
© The Author(s) 2023

Abstract

The study of system safety and reliability has always been vital for the quality and manufacturing engineers of varying fields for which generally the continuous probability distributions are proposed. Bivariate and multivariate continuous distributions are the candidates while studying more than one characteristic of the system. In this article, an attempt is made to address this issue when the reliability systems generate bivariate and correlated count datasets. The bivariate generalized geometric distribution (BGGD) is believed to serve as a potential candidate to model such types of datasets. Bayesian approach of data analysis has the potential of accommodating the uncertainty associated with the model parameters of interest using uninformative and informative priors. A real life bivariate correlated dataset is analyzed in Bayesian framework and the results are compared with those produced by the classical approach. Posterior summaries including posterior means, highest density regions, and predicted expected frequencies of the bivariate data are evaluated. Different information criteria are evaluated to compare the inferential methods under study. The entire analysis is carried out using Markov chain Monte Carlo (MCMC) set-up of data augmentation implemented through WinBUGS.

Keywords Bayesian analysis · Bivariate generalized geometric distribution · Correlated count data · Maximum likelihood estimation · MCMC methods · Model selection criteria · Posterior estimates

Mathematical Subject Classification 62F15 · 62C10

Abbreviations

BGGD	Bivariate generalized geometric distribution
MCMC	Markov chain Monte Carlo
WinBUGS	Bayesian inference using Gibbs sampling
ML	Maximum likelihood

✉ Nasir Abbas
nabbascj@yahoo.com

¹ Government Graduate College, Jhang, Pakistan

AIC	Akaike information criterion
AICC	Corrected AIC
BIC	Bayes information criterion
HQC	Hannan-Quinn criterion

1 Introduction

Reliability and system engineers often encounter the difficulty of dealing with uncertainties present in the system where more than one study variables are of interest to them. Medical experts also face the similar situations when the life of patients goes at stake for the failure of vital organs like heart, brain, kidney, liver, lungs, and the likes. Choice of discrete or continuous and bivariate or multivariate distributions depends on the nature and number of the study variables. A vast literature exists on the construction of probability distributions. For this, one may refer to [1]. No hard and fast criteria could be established to construct probability distributions. More details on this issue may be found in [1–3].

If bivariate continuous distributions are to be used, we could choose from parametrical distributions to analyze bivariate lifetime data suggested in the literature, [4–8]. As our study corresponds to the reliability of system generating bivariate count data, so the most suitable distribution to model such types of datasets is believed to be the bivariate generalized geometric distribution (BGGD) proposed by [9]. Many bivariate distributions for continuous random variables are introduced in the literature to be used in data analysis, especially in applications of survival data in the presence of censored data and covariates (see, for example, [10–20]). The recent study includes [21]. Alternatively, it can be observed in the literature that it is not very common the use of bivariate distributions for survival data assuming discrete data. Some discrete bivariate distributions have been introduced in the literature as the bivariate geometric distribution of [4] or the bivariate geometric distribution of [22], but these discrete distributions are still not very popular in the analysis of bivariate lifetime data, especially in the presence of censored data and covariates (see also, [4, 23–28]).

Classical methods are frequently used in the analysis but they suffer from a certain drawbacks. The frequentists consider parameters to be unknown fixed quantities and they just rely on the current data and deprive the results of any prior information available about the parameters of interest. However, the Bayesians treat the parameters as random quantities and hence assign a probability distribution to the parameters. The Bayesian analysis is a modern inferential technique that endeavors to estimate the model parameters taking both the current data and prior information about the parameters into account. As a result, we get a posterior distribution that is believed to average the current data and the prior information. The posterior distribution thus derived is the achilles heel and the work-bench of the Bayesians to infer about the parameters based on numerical procedures and the entire estimation is then based on the very posterior distribution. A good review of the advantages of using Bayesian methods may be seen in [29]. The posterior distributions often have complex multidimensional functions that require the use of Markov chain Monte

Carlo (MCMC) methods to draw results, [30–34]. Its use is very popular for analyzing bivariate continuous or discrete random variables in presence of censored data and covariates (see for example, [29, 35–40]. Recently, [41] has considered weighted bivariate geometric distribution and [42] has used the q -Weibull distribution in classical as well as Bayesian frameworks. In recent years, the use of Markov Chain Monte Carlo (MCMC) methods has gained much popularity, [43, 44] and [45].

It has been established that the BGGD is a good choice to model and analyze reliability count data appearing in medicine, engineering. The probability mass function (pmf) of BGGD is given as

$$f(x, y; \alpha, \theta_1, \theta_2) = \frac{\alpha \theta_1^x \theta_2^y (1 - \theta_1)(1 - \theta_2)(1 - \alpha'^2 \theta_1^{2x+1} \theta_2^{2y+1})}{(1 - \alpha' \theta_1^{x+1} \theta_2^y) (1 - \alpha' \theta_1^{x+1} \theta_2^{y+1}) (1 - \alpha' \theta_1^x \theta_2^y) (1 - \alpha' \theta_1^x \theta_2^{y+1})}, \quad (1)$$

for $x = 0, 1, \dots, y = 0, 1, \dots, 0 < \theta_1 < 1, 0 < \theta_2 < 1, \alpha > 0, \alpha' = 1 - \alpha$. And the cumulative distribution function (CDF) is

$$F(x, y) = \frac{1 - \theta_1^x \theta_2^y}{1 - \alpha' \theta_1^x \theta_2^y}, \alpha > 0, \alpha' = 1 - \alpha.$$

Here α, θ_1 and θ_2 are unknowns parameters that control the behavior of the datasets emerging from the BGGD. Estimating the unknown parameters is ultimate goal of the inferential statistics.

Due to variety of applications of the BGGD, the efficient estimation of the PDF and the CDF of the BGGD is the purpose of the present study. In [9] have recently worked out the classical maximum likelihood estimators for the BGGD. Taking into account and to avail the aforesaid advantages, we estimate the parameters of the BGGD in Bayesian framework. We have used the MCMC methods to draw results and applied different model selection criteria to compare the methods under consideration. Such as ML, AIC, AICC, BIC is also known as Schwarz criterion), and HQC.

2 The Frequentist Approach of Statistical Analysis

In statistical terminology, the data generating pattern of any system or model depends on the system-specific characteristics, called parameters. So the data being generated from the model is believed to advocate the values of the parameters causing the system to generate the dataset. The uncertainty associated with the data values is defined in terms of frequencies of the data values emerging again and again from the system under study. The objective of the analysis is to infer the characteristics of the system or model from the relevant data collected randomly. It is considered as the default approach to be used in variety of areas of sciences. Commonly used frequentist methods of statistical inference include uniformly minimum variance unbiased estimation, maximum likelihood (ML) estimation, percentile estimation, least squares estimation, weighted least

squares estimation, etc. But we just report the most commonly used ML estimation method whose results will henceforth be compared with their Bayesian counterparts.

2.1 Maximum Likelihood (ML) Estimation

The likelihood function gives the probability of the situation that the model, system or distribution under study have witnessed to generate the observed sample. The frequentist method of maximum likelihood estimation professed by [46] calls for choosing those values of the parameters that maximize the probability of the very observed sample. We generally opt for algebraic maximization of the likelihood function to find the ML estimates, but we may also opt for evaluating the probabilities of the observed samples at all possible values of the parameters and to choose those parametric values as the estimates that maximize the evaluated probabilities of the observed samples.

Algebraically, the ML estimation may be proceeded as follows. Let us consider the random sample of size n from the bivariate correlated data (x_i, y_j) for $i = 1, 2, \dots, n_1$ and $j = 1, 2, \dots, n_2$ from the BGGD $f(x, y; \alpha, \theta_1, \theta_2)$ given in (11). The log likelihood function $l(x, y; \alpha, \theta_1, \theta_2)$ may be written as

$$\begin{aligned}
 l(x, y; \alpha, \theta_1, \theta_2) = & (n_1 + n_2) \log(\alpha) + n_1 \bar{x} \log(\theta_1) + n_2 \bar{y} \log(\theta_2) \\
 & + n_1 \log(1 - \theta_1) + n_2 \log(1 - \theta_2) \\
 & + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (1 - \alpha'^2 \theta_1^{2x+1} \theta_2^{2y+1}) - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (1 - \alpha' \theta_1^{x+1} \theta_2^y) \\
 & - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (1 - \alpha' \theta_1^{x+1} \theta_2^{y+1}) - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (1 - \alpha' \theta_1^x \theta_2^y) \\
 & - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (1 - \alpha' \theta_1^x \theta_2^{y+1})
 \end{aligned}$$

Equating to zero the first partial derivatives of the log-likelihood function $l(x, y; \alpha, \theta_1, \theta_2)$ with respect to the set of unknown parameters $\alpha, \theta_1, \theta_2$ yields the normal equations which may be solved simultaneously to get the required ML estimates. However, if the normal equations are too complex to be solved simultaneously, we have to proceed to numerical methods or by direct maximization of the log-likelihood function.

2.2 ML estimation–Algebraic approach

Let us consider the real dataset for the BGGD given in Appendix A₁ in Table 8 appearing in [47–49], where X represents the counts of surface and Y, the count of interior faults in 100 lenses. The summary of the data is presented in Table 1 along with its figurative representation made in Fig. 1.

Table 1 Summary of the observed data

\bar{x}	s_x^2	\bar{y}	s_y^2	r_{xy}
3.250	5.46750	2.930	6.14510	− 0.17468

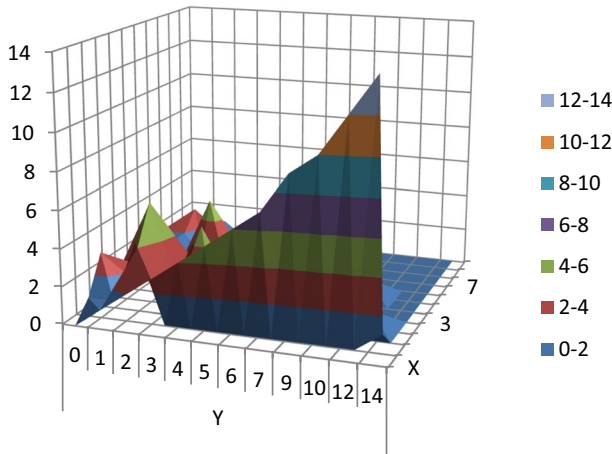


Fig. 1 Graphs of the BGGD at the observed data points

Table 2 ML estimates, standard error and 95% confidence intervals

Parameters	MLEs	Standard Errors	95% confidence intervals
α	2.288	0.198	(1.8999; 2.6761)
θ_1	0.676	0.024	(0.6290; 0.7230)
θ_2	0.652	0.025	(0.6030; 0.7010)

Obviously, the observed data is positively skewed and negatively correlated. As stated in Sect. 3.1, the normal equations we get using the observed dataset are complicated and hence the estimates are found using the numerical methods. Following [9], the ML estimates, standard error (SE) and 95% confidence intervals for the parameters are reported in Table 2.

2.3 ML Estimation–Graphical approach

As already discussed in Sect. 3.1, we may opt for direct maximization of the likelihood function to find the ML estimates. The ML estimation theory suggested by [46] calls for choosing those values of the parameters that maximize the probability of the observed sample, and these values are regarded as the ML estimates. This technique is used here to find the ML estimates by plotting the observed dataset of the BGGD at different parametric values. The resulting plots generated in R package and are displayed in Fig. 2. We observed that the highest probability is obtained at $\alpha = 2.288$, $\theta_1 = 0.676$, $\theta_2 = 0.652$ (the 1st one of the plots of Fig. 2), hence they may be regarded as the ML estimates.

The graph in first cell corresponds to that produced by using the ML estimates. The maximum likelihood is found to be $3.2619E-192$ at the ML estimates, i.e., $\alpha = 2.288$, $\theta_1 = 0.676$, $\theta_2 = 0.652$. It is also interesting to note that the value of highest ordinate,

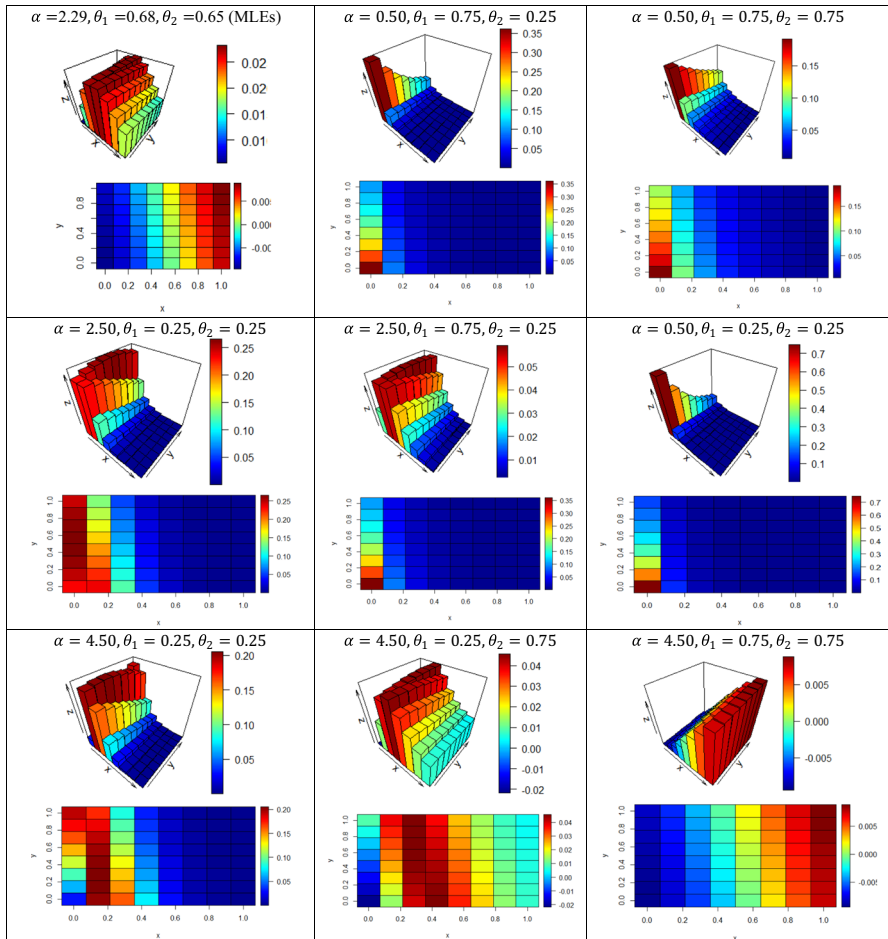


Fig. 2 BGGD at the observed data points and different parametric values and the parameters including the ML estimates, i.e., $\alpha = 2.288, \theta_1 = 0.676, \theta_2 = 0.652$ (First cell)

i.e., 0.02822 appears at the data pair $(x = 2, y = 1)$ at the ML estimates and the value of negative log-likelihood is found to be -432.957.

3 The Bayesian Approach of Statistical Analysis

A brief overview on this topic is already given in Sect. 1. Bayesian method combines prior information about the model parameters with dataset using Bayes rule yielding the posterior distribution. The Bayes rule is named after Thomas Bayes, whose work on this topic was published in 1763, 2 years after his death, [50]. To establish Bayesian inference set-up, we need a model or system in the form of a probability distribution controlled by a set of parameters to be estimated, the sample

dataset generated by the model or distribution, and a prior distribution based on the prior knowledge of experts regarding the parameters of interest. These elements are formally combined in to posterior distribution which is regarded as a key-distribution and work-bench for the subsequent analyses. The algorithm is explained in [42].

If $f(D|\theta)$ is the data distribution depending upon the vector of parameters θ , $p(\theta)$ is the joint prior distribution of vector of parameters θ , $L(D|\theta)$ is the likelihood function defining the joint probability of the observed sample data D conditional upon the parameter vector θ , then the posterior distribution of θ conditional upon the data D , denoted by $f(\theta|D)$, is given by

$$f(\theta|D) = \frac{p(\theta)L(D|\theta)}{\int_{\theta} p(\theta)L(D|\theta)d\theta},$$

The denominator is also termed as predictive distribution of data and is usually treated as the normalizing constant to make it a proper density. It may be omitted in evaluating the Bayes estimates but must be retained in comparing the models. The posterior distribution has the potential to balance the information provided by the data and prior distribution. It is of the supreme interest of the Bayesians but often has very complex and complicated nature and hence needs numerical methods to evaluate it.

3.1 The Prior Distributions

It has already been highlighted that main difference between frequentist and Bayesian approaches is to incorporate prior information regarding the model parameters into the analysis. The formal way of doing so it to quantify the initial knowledge of experts in the form of a prior distribution that can adequately fit to the nature of the parameter and the experts' opinion. The parameters of the prior distribution, known as hyperparameters, are elicited in the light of the subjective expert opinion. So, the prior is leading if selected and elicited carefully and adequately, otherwise it may be misleading.

3.1.1 Uninformative Priors

In the situations when there is lack of knowledge about the model parameters, we choose vague, defuse or flat priors. As the BGGD is based on the set of parameters $\alpha, \theta_1, \theta_2$, so we assume uninformative uniform priors for the parameters as follows:

$$\alpha \sim \text{uniform}(h_1, h_2), \theta_1 \sim \text{uniform}(h_3, h_4), \theta_2 \sim \text{uniform}(h_5, h_6),$$

where $\alpha > 0, 0 < \theta_1 < 1, 0 < \theta_2 < 1$, and $h_i > 0$, for $i = 1, 2, \dots, 6$ are the set of hyperparameters associated with the priors.

3.1.2 Informative Priors

When there is sufficient information available about the model parameters, we assign informative priors to the model parameters in such a way that they can adequately represent the knowledge available for the model parameters being examined. In the present situation, we assign an exponential distribution to α and independent beta distributions to θ_1 and θ_2 given as under

$$\left. \begin{aligned} p(\alpha) &= h_7 e^{-h_7 \alpha}, \alpha > 0, \\ p(\theta_1) &= \frac{\Gamma(h_8+h_9)}{\Gamma(h_8)\Gamma(h_9)} \theta_1^{h_8-1} (1-\theta_1)^{h_9-1}, 0 < \theta_1 < 1, \\ p(\theta_2) &= \frac{\Gamma(h_{10}+h_{11})}{\Gamma(h_{10})\Gamma(h_{11})} \theta_2^{h_{10}-1} (1-\theta_2)^{h_{11}-1}, 0 < \theta_2 < 1, \end{aligned} \right\}$$

Here again $h_i > 0$, for $i = 7, 8, \dots, 11$ are the set of hyperparameters associated with the priors that should be elicited in the light of expert opinion. It is to notice that the elicitation of hyperparameters is beyond the scope of this study, so we would opt for merely choosing the values of the hyperparameters to be used in the subsequent Bayesian analysis.

3.2 The Posterior Distribution

Being specific to estimation of the parameters of BGGD, let the vector of parameters of interest and the data are denoted by $\theta = (\alpha, \theta_1, \theta_2)$ and $D = (X, Y)$ respectively. The data distribution is denoted by $f(D|\theta)$ and prior distribution by $p(\theta)$. Then using the Bayes rule, the posterior distribution denoted by $f(\theta|D)$ may be written as

$$f\{\theta|D\} = \frac{p(\alpha)p(\theta_1)p(\theta_2)L\{D|\theta\}}{\int_{\alpha=1}^{\infty} \int_{\theta_1=0}^1 \int_{\theta_2=0}^1 p(\alpha)p(\theta_1)p(\theta_2)L\{D|\theta\} d\alpha d\theta_1 d\theta_2}$$

where, all the notations are already defined. The posterior distribution $f\{\theta|D\}$ may also be written as kernel density in proportional form as

$$f\{\theta|D\} \propto p(\alpha)p(\theta_1)p(\theta_2)L\{D|\theta\}.$$

The marginal posterior distributions $g(\alpha|D)$, $g(\theta_1|D)$ and $g(\theta_2|D)$ of the parameters α, θ_1 and θ_2 may be found by integrating out the nuisance parameters from the posterior distribution $f\{\alpha, \theta_1, \theta_2|D\}$ as follows:

$$g(\alpha|D) = \int_{\theta_1=0}^1 \int_{\theta_2=0}^1 f\{\theta|D\} d\theta_1 d\theta_2,$$

$$g(\theta_1|D) = \int_{\alpha=1}^{\infty} \int_{\theta_2=0}^1 f\{\theta|D\} d\alpha d\theta_2,$$

And

$$g(\theta_2|D) = \int_{\alpha=1}^{\infty} \int_{\theta_1=0}^1 f\{\theta|D\} d\alpha d\theta_1.$$

3.3 Bayes Estimates

To work out the Bayes estimates of the parameters $(\alpha, \theta_1, \theta_2)$, we need to specify some loss function. A variety of loss functions is used to derive the Bayes estimates. Under the well-known squared error loss function, the Bayes estimates $\hat{\alpha}$, $\hat{\theta}_1$ and $\hat{\theta}_2$ are the arithmetic means of their marginal posterior distributions, and are evaluated as

$$\begin{aligned}\hat{\alpha} &= E(\alpha|D) = \int_0^{\infty} \alpha \cdot g(\alpha|D) d\alpha, \\ \hat{\theta}_1 &= E(\theta_1|D) = \int_0^1 \theta_1 \cdot g(\theta_1|D) d\theta_1, \\ \hat{\theta}_2 &= E(\theta_2|D) = \int_0^1 \theta_2 \cdot g(\theta_2|D) d\theta_2.\end{aligned}$$

The marginal posterior distributions are generally of complicated and complex forms and hence need numerical methods to evaluate them. Markov Chains Monte Carlo (MCMC) is the most frequently used numerical method to be used in Bayesian inference. So we also proceed with MCMC with the WinBUGS package to find the posterior summaries of the parameters of interest.

3.4 The MCMC Method

The MCMC method selects random sample from the probability distribution according to random process termed as Markov Chain where every new step of the process depends on the current state and is completely independent of previous states. The MCMC methods can be implemented using any of the standard softwares like R, Python, etc., but the most specific software being used for the Bayesian analysis is Windows based Bayesian inference using Gibbs Sampling (WinBUGS). We have implemented WinBUGS using the following scheme.

- (i) Define model based on probability mass function (11) of BGGD and then click *Check Model* menu in WinBUGS software.
- (ii) *Load data* given in Appendix.
- (iii) Specify the *nodes* and run the codes for 10,000 times following a burn-in of 5000 iterations.

WinBUGS codes used to analyze the data are given in Appendix A2.

3.5 Bayesian Results Under Uniform Non-informative Priors

Here we have assumed uniform priors for all the set of parameters under study as defined in section 0, and the resulting Bayes estimates, standard errors, medians

Table 3 Summary of Bayes estimates under the uninformative priors

Uninformative uniform priors	Parameters	Means	Medians	SEs	MC error	95% Credible intervals
$\alpha \propto \text{uniform}(0, 50)$	α	25.020	25.010	14.4700	0.1030	(1.2330; 48.750)
$\theta_1 \propto \text{uniform}(0.25, 0.75)$	θ_1	0.4993	0.4977	0.1450	0.0010	(0.2623; 0.7376)
$\theta_2 \propto \text{uniform}(0.25, 0.75)$	θ_2	0.5008	0.4999	0.1446	0.0010	(0.2631; 0.7373)
$\alpha \propto \text{uniform}(0.25, 10)$	α	5.1290	5.1260	2.8210	0.0201	(0.4905; 9.7570)
$\theta_1 \propto \text{uniform}(0.1, 0.9)$	θ_1	0.4990	0.4964	0.2319	0.0016	(0.1197; 0.8802)
$\theta_2 \propto \text{uniform}(0.1, 0.9)$	θ_2	0.5013	0.4998	0.2313	0.0017	(0.1209; 0.8797)
$\alpha \propto \text{uniform}(1, 5)$	α	3.0020	3.0000	1.1570	0.0082	(1.0990; 4.9000)
$\theta_1 \propto \text{uniform}(0.1, 0.99)$	θ_1	0.5438	0.5410	0.2580	0.0018	(0.1220; 0.9680)
$\theta_2 \propto \text{uniform}(0.2, 0.95)$	θ_2	0.5762	0.5749	0.2168	0.0016	(0.2196; 0.9309)
$\alpha \propto \text{uniform}(0, 30)$	α	15.010	15.0000	8.6800	0.0618	(0.7400; 29.250)
$\theta_1 \propto \text{uniform}(0.25, 0.75)$	θ_1	0.4993	0.4977	0.1450	0.0010	(0.2623; 0.7376)
$\theta_2 \propto \text{uniform}(0.25, 0.75)$	θ_2	0.5008	0.4999	0.1446	0.0010	(0.2631; 0.7373)
$\alpha \propto \text{uniform}(0, 25)$	α	12.510	12.500	7.2340	0.0515	(0.6167; 24.380)
$\theta_1 \propto \text{uniform}(0.25, 0.75)$	θ_1	0.4993	0.4977	0.1450	0.0010	(0.2623; 0.7376)
$\theta_2 \propto \text{uniform}(0.25, 0.75)$	θ_2	0.5008	0.4999	0.1446	0.0010	(0.2631; 0.7373)
$\alpha \propto \text{uniform}(0, 40)$	α	20.020	20.000	11.5700	0.0824	(0.9867; 39.000)
$\theta_1 \propto \text{uniform}(0.25, 0.75)$	θ_1	0.4993	0.4977	0.1450	0.0010	(0.2623; 0.7376)
$\theta_2 \propto \text{uniform}(0.25, 0.75)$	θ_2	0.5008	0.4999	0.1446	0.0010	(0.2631; 0.7373)

and 95% highest density regions along with the values of the hyperparameters are presented in Table 3.

It is observed that the elicited hyperparameters have high impact on the Bayes estimates. The initial values have no effect no significant effects on the posterior estimates if the true convergence is achieved.

3.6 Convergence Diagnostics

Sequential plots are used in WinBUGS to assess difficulties in MCMC and realization of the model. In MCMC simulations, the values of the parameters of interest are sampled from the posterior distributions. So the estimates will be convergent if the posterior distributions are stationary in nature and the Markov chain will seem to be mixing well. To check convergence, different graphical representations of parametric behavior are used in MCMC implemented through WinBUGS.

3.6.1 History Time Series Plots

The time series history plots of parameters are presented in Fig. 3. Here, Markov chain seems to be mixing well enough and is being sampled from the stationary

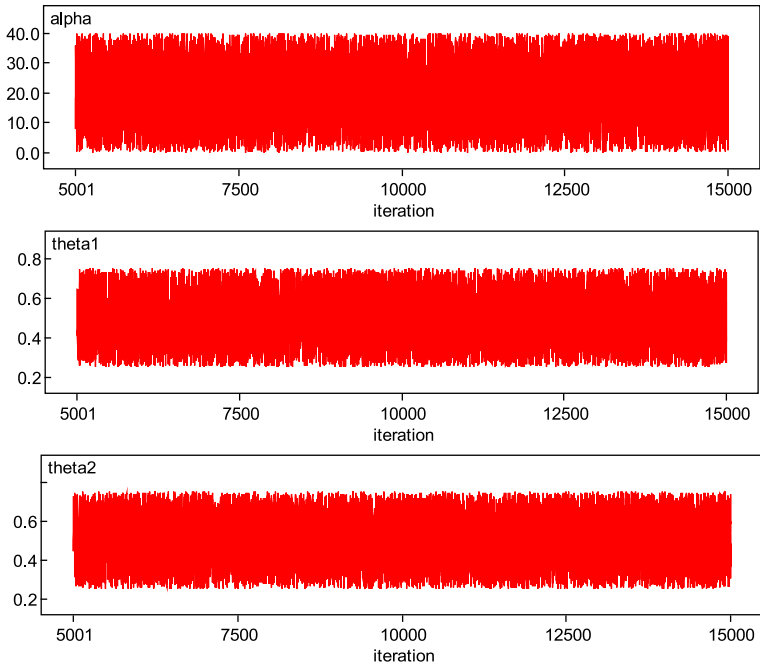


Fig. 3 Time series plots of the parameters $\alpha, \theta_1, \theta_2$

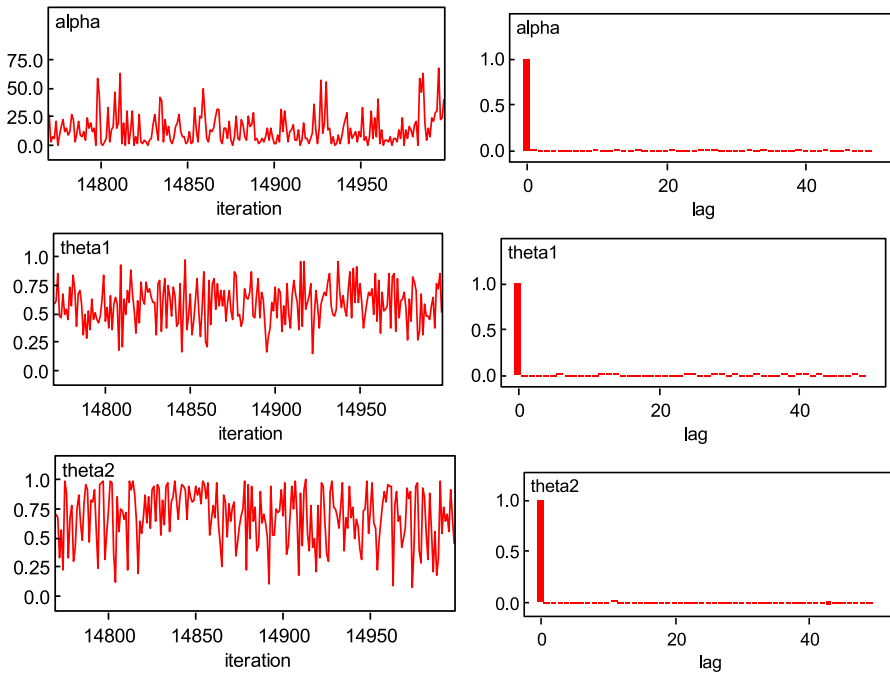


Fig. 4 Dynamic traces and autocorrelation function

distribution. The plots are in the form of horizontal bands with no long upward or downward trends. It indicates that the Markov chain has converged.

3.6.2 Dynamic Traces and Autocorrelation Function

The traces of the parameters and autocorrelation graphs of α , θ_1 and θ_2 are presented in Fig. 4. These graphs also confirm convergence.

3.7 Bayes Estimates Using Informative Exponential-Beta Priors

The ideal characteristic of Bayesian analysis is that it can accommodate the prior information shared by the field experts about the unknown parameters in the analysis. It is important to notice that the experts may not be the experts of statistics and hence cannot translate their expertise in statistical terms. So it is the sole responsibility of the statisticians to formally utilize the experts' prior information to elicit the values of the hyperparameters of the prior density which are to be subsequently used in the Bayesian analysis. Elicitation of hyperparameters is beyond the scope of our study. However, an exhaustive discussion on the elicitation of hyperparameters may be found in [51]. We have chosen the values of the hyperparameters with drastic changes and a summary of the Bayes estimates against all values are presented in Table 4.

Table 4 Summary of Bayes estimates under the informative priors

Exponential-Beta Priors	Parameters	Means	Medians	SEs	MC error	95% Credible intervals
$\alpha \propto \exp(0.15)$	α	6.6710	4.6880	6.6380	0.0448	(0.1756; 24.720)
$\theta_1 \propto \text{beta}(4, 3)$	θ_1	0.5720	0.5780	0.1749	0.0012	(0.2237; 0.8816)
$\theta_2 \propto \text{beta}(2, 1)$	θ_2	0.6657	0.7045	0.2360	0.0016	(0.1592; 0.9874)
$\alpha \propto \exp(0.20)$	α	5.0030	3.5160	4.9780	0.0336	(0.1317; 18.540)
$\theta_1 \propto \text{beta}(4, 3)$	θ_1	0.5720	0.5780	0.1749	0.0012	(0.2237; 0.8816)
$\theta_2 \propto \text{beta}(2, 1)$	θ_2	0.6657	0.7045	0.2360	0.0016	(0.1592; 0.9874)
$\alpha \propto \exp(0.10)$	α	10.0100	7.0320	9.9560	0.0671	(0.2634; 37.080)
$\theta_1 \propto \text{beta}(4, 3)$	θ_1	0.5720	0.5780	0.1749	0.0012	(0.2237; 0.8816)
$\theta_2 \propto \text{beta}(2, 1)$	θ_2	0.6657	0.7045	0.2360	0.0016	(0.1592; 0.9874)
$\alpha \propto \exp(0.07)$	α	14.3000	10.0500	14.2200	0.0959	(0.3762; 52.980)
$\theta_1 \propto \text{beta}(4, 3)$	θ_1	0.5720	0.5780	0.1749	0.0012	(0.2237; 0.8816)
$\theta_2 \propto \text{beta}(2, 1)$	θ_2	0.6657	0.7045	0.2360	0.0016	(0.1592; 0.9874)
$\alpha \propto \exp(0.05)$	α	20.0100	14.0600	19.9100	0.1342	(0.5267; 74.170)
$\theta_1 \propto \text{beta}(4, 3)$	θ_1	0.5720	0.5780	0.1749	0.0012	(0.2237; 0.8816)
$\theta_2 \propto \text{beta}(2, 1)$	θ_2	0.6657	0.7045	0.2360	0.0016	(0.1592; 0.9874)
$\alpha \propto \exp(0.02)$	α	50.2800	34.8200	50.0600	0.3880	(1.2530; 183.000)
$\theta_1 \propto \text{beta}(1, 5)$	θ_1	0.1666	0.1300	0.1396	0.0010	(0.0052; 0.5172)
$\theta_2 \propto \text{beta}(2, 10)$	θ_2	0.1671	0.1483	0.1040	0.0007	(0.0231; 0.4158)

Table 5 The predicted expected data of Counts of surface (X) and interior faults (Y) in 100 lenses

X	Y											
	0	1	2	3	4	5	6	7	9	10	12	14
0	0.37	0.73	1.56	2.81	3.56	3.21	2.26	1.36	0.39	0.20	0.05	0.01
1	0.73	1.57	2.81	3.56	3.21	2.26	1.36	0.75	0.20	0.10	0.03	0.01
2	1.57	2.82	3.56	3.21	2.25	1.36	0.75	0.39	0.10	0.05	0.01	0.00
3	2.82	3.56	3.20	2.25	1.35	0.74	0.39	0.20	0.05	0.03	0.01	0.00
4	2.82	3.56	3.20	2.25	1.35	0.74	0.39	0.20	0.05	0.03	0.01	0.00
5	3.56	3.20	2.25	1.35	0.74	0.39	0.20	0.10	0.03	0.01	0.00	0.00
6	2.24	1.35	0.74	0.39	0.20	0.10	0.05	0.03	0.01	0.00	0.00	0.00
7	1.34	0.74	0.39	0.20	0.10	0.05	0.03	0.01	0.00	0.00	0.00	0.00
8	0.74	0.39	0.20	0.10	0.05	0.03	0.01	0.01	0.00	0.00	0.00	0.00
11	0.10	0.05	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.05	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6 Summary of the predicted dataset based on Bayes estimates

\bar{x}	s_x^2	\bar{y}	s_y^2	r_{xy}
2.764	4.382704	2.6526	4.328313	-0.47433

It is observed that the elicited hyperparameters have high impact on the Bayes estimates. The initial values have no significant effect on the posterior estimates if the convergence in Markov chain is achieved. However, the change in the initial values causes a slight change in the parametric estimates.

3.8 Possible Predictive Inference

After finding out the Bayes estimates, it is necessary to evaluate them based on the predictive inference. As we have the data having 100 observations, so predicted sample of size 100 observations is generated based on the Bayes estimates obtained through MCMC based analysis. The predicted data along with their summaries are presented in Tables 5, 6.

Obviously, there exist some differences in the predicted estimates as compared to those of the original observed dataset. Definitely, these changes may be due to different Bayes estimates that are evaluated after accommodating the prior information about the model parameters via the hyperparameters.

4 Comparison of the Frequentist and Bayesian Approaches

An important aspect of this study is to compare the Bayes estimation method with the classical ML estimation method. We have accomplished this by using different model selection criteria presented as under.

4.1 Model Selection Criteria

The classical and Bayesian methods of estimation are compared using the model selection criteria, i.e., ML, AIC, AICC, BIC, and HQC, which are defined by

$$ML = -2\ln[L(\alpha, \theta_1, \theta_2)]$$

$$AIC = -2\ln[L(\alpha, \theta_1, \theta_2)] + 2k$$

$$AICC = -2\ln[L(\alpha, \theta_1, \theta_2)] + 2k(n/(n - k - 1))$$

$$BIC = -2\ln[L(\alpha, \theta_1, \theta_2)] + k\ln(n)$$

and

$$HQC = -2\ln[L(\alpha, \theta_1, \theta_2)] + 2k\ln[\ln(n)].$$

Here $\ln[L(\alpha, \theta_1, \theta_2)]$ denotes the log-likelihood, n denotes the number of observations and k denotes the number of parameters of the distribution under consideration. The smaller the values of these criteria are, the better the fit is. For more discussion on these criteria, see [52, 53]. The Maximum likelihood estimates, uninformative Bayes estimates and informative Bayes estimates along with their associated values of the model selection criteria are reported in Table 7.

Here we witness that the values of model selection criteria produced by Bayes method are less than those produced by the ML estimation method, which declare the Bayesian method more appropriate. Definitely, it is due to the distinct characteristic of the Bayesian methods that they incorporate the prior information related to the model parameters. However, it is pertinent to note that these results are sensitive to the selection of values of the hyperparameters. Hence a careful elicitation of the hyperparameters is demanding and earnest need of using the Bayesian methods. Carefully selected or elicited values of the hyperparameters may lead to even better estimates.

5 Summary and Conclusions

The bivariate generalized geometric distribution is believed to model reliability count datasets emerging from diverse phenomena. To understand the data generating phenomena, it is necessary to estimate the model parameters of the BGGD. To accomplish this, statistical theory offers two competing approaches, namely the frequentist and Bayesian approaches. The former approach is based on current data only; whereas, the later one utilizes prior information in addition to the current dataset produced by system or phenomenon. This study offers a comparison between the frequentist and Bayesian estimation approaches. To elaborate the frequentist approach, different descriptive measures and the maximum likelihood estimates are evaluated. The Bayesian estimation approach has also been illustrated by using uninformative and informative priors. We have worked out

Table 7 Model selection criteria for classical and Bayes estimation methods for the data of the example

Estimation approaches	Estimates of		log-likelihood	ML	AIC	AICC	BIC	HQC
	α	θ_1 θ_2						
MLEs	2.288	0.676 0.652	-432.957	865.914	871.914	872.164	879.729	868.968
Uninformative Bayes estimates	25.12	0.499 0.5	-408.902	817.805	823.805	824.055	831.62	822.386
Informative Bayes estimates	6.671	0.572 0.666	-420.122	840.245	846.245	846.495	854.060	844.826

the posterior summaries of the parameters comprising posterior means, standard errors, medians, credible intervals and predictions for both types of the priors using the MCMC simulation technique. Correlated bivariate count dataset on counts of surface and interior faults is used for the illustration purpose. Comparison of the two estimation methods has been made using different model selection criteria. It is proved by working out all the estimates that all the model selection criteria including ML, AIC, AICC, BIC, and HQC have proved that the Bayesian approach outperforms the competing ML approach across the board. It has also been observed that the results may coincide if the information contained in the prior distribution and the datasets agree. However, the improved prior information may improve the results. As a future study, it is recommended that the Bayesian analysis of datasets may be done by using the formally elicited hyperparameters of the priors instead of values chosen by the experimenter.

Appendix A1

See Table 8.

Table 8 Data set: Counts of surface (X) and interior faults (Y) in 100 lenses

X	Y												
	0	1	2	3	4	5	6	7	9	10	12	14	
0	1	1	4	0	0	0	0	0	0	0	0	0	1
1	3	2	6	2	5	0	2	0	0	0	0	0	0
2	1	2	4	3	2	1	1	1	1	0	1	0	0
3	0	5	1	2	2	3	2	0	0	0	0	0	0
4	1	2	2	5	3	1	1	0	0	0	0	0	0
5	1	2	1	2	1	2	0	0	0	1	0	0	0
6	2	2	0	1	1	0	0	1	0	0	0	0	0
7	1	3	0	1	0	0	0	0	0	0	0	0	0
8	0	2	0	0	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0	0	0
12	0	1	0	0	0	0	0	0	0	0	0	0	0

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Kemp, C., Papageorgiou, H.: Bivariate hermite distributions. *Indian J. Stat. Ser. A* **44**, 269–280 (1982)
2. Johnson, N.L., Kotz, S., Balakrishnan, N.: *Discrete Multivariate Distributions*, 165th edn. Wiley, New York (1997)
3. Lai, C.-D.: Constructions of discrete bivariate distributions. In: *Advances in Distribution Theory, Order Statistics, and Inference*, pp. 29–58. Springer (2006)
4. Basu, A.P., Dhar, S.: Bivariate geometric distribution. *J. Appl. Stat. Sci.* **2**(1), 3–44 (1995)
5. Sarhan, A.M., Balakrishnan, N.: A new class of bivariate distributions and its mixture. *J. Multivar. Anal.* **98**(7), 1508–1527 (2007)
6. Jamalizadeh, A., Kundu, D.: Weighted Marshall-Olkin bivariate exponential distribution. *Statistics* **47**(5), 917–928 (2013)
7. Kundu, D., Gupta, A.K.: On bivariate Weibull-geometric distribution. *J. Multivar. Anal.* **123**, 19–29 (2014)
8. Sankaran, P., Kundu, D.: A bivariate Pareto model. *Statistics* **48**(2), 241–255 (2014)
9. Gómez-Déniz, E., Ghitany, M., Gupta, R.C.: A bivariate generalized geometric distribution with applications. *Commun. Stat. Theory Methods* **46**(11), 5453–5465 (2017)
10. Gumbel, E.J.: Bivariate exponential distributions. *J. Am. Stat. Assoc.* **55**(292), 698–707 (1960)
11. Freund, J.E.: A bivariate extension of the exponential distribution. *J. Am. Stat. Assoc.* **56**(296), 971–977 (1961)
12. Marshall, A.W., Olkin, I.: A generalized bivariate exponential distribution. *J. Appl. Probab.* **4**(2), 291–302 (1967)
13. Marshall, A.W., Olkin, I.: A multivariate exponential distribution. *J. Am. Stat. Assoc.* **62**(317), 30–44 (1967)
14. Downton, F.: Bivariate exponential distributions in reliability theory. *J. R. Stat. Soc.* **32**(3), 408–417 (1970)
15. Hawkes, A.G.: A bivariate exponential distribution with applications to reliability. *J. R. Stat. Soc.* **11**, 129–131 (1972)
16. Block, H.W., Basu, A.: A continuous, bivariate exponential extension. *J. Am. Stat. Assoc.* **69**(348), 1031–1037 (1974)
17. Hougaard, P.: A class of multivariate failure time distributions. *Biometrika* **73**(3), 671–678 (1986)
18. Sarkar, S.K.: A continuous bivariate exponential distribution. *J. Am. Stat. Assoc.* **82**(398), 667–675 (1987)
19. Arnold, B.C., Strauss, D.: Bivariate distributions with exponential conditionals. *J. Am. Stat. Assoc.* **83**(402), 522–527 (1988)
20. Hanagal, D.D.: Bivariate Weibull regression model based on censored samples. *Stat. Pap.* **47**(1), 137–147 (2006)
21. de Oliveira, R.P., Achcar, J.A.: Basu-Dhar's bivariate geometric distribution in presence of censored data and covariates: some computational aspects. *Electron. J. Appl. Stat. Anal.* **11**(1), 108–136 (2018)
22. Arnold, B.C.: A characterization of the exponential distribution by multivariate geometric compounding. *Indian J. Stat. Ser. A* **37**, 164–173 (1975)
23. Nair, K.M., Nair, N.U.: On characterizing the bivariate exponential and geometric distributions. *Ann. Inst. Stat. Math.* **40**(2), 267–271 (1988)
24. Sun, K., Basu, A.P.: A characterization of a bivariate geometric distribution. *Statist. Probab. Lett.* **23**(4), 307–311 (1995)

25. Dhar, S.K., Balaji, S.: On the characterization of a bivariate geometric distribution. *Commun. Stat. Theory Methods* **35**(5), 759–765 (2006)
26. Krishna, H., Pundir, P.S.: A bivariate geometric distribution with applications to reliability. *Commun. Stat. Theory Methods* **38**(7), 1079–1093 (2009)
27. Li, J., Dhar, S.K.: Modeling with bivariate geometric distributions. *Commun. Stat. Theory Methods* **42**(2), 252–266 (2013)
28. Davarzani, N., et al.: Bivariate lifetime geometric distribution in presence of cure fractions. *J Data Sci.* **13**(4), 755–770 (2015)
29. Adams, E.S.: Bayesian analysis of linear dominance hierarchies. *Anim. Behav.* **69**(5), 1191–1201 (2005)
30. Foreman-Mackey, D., et al.: emcee: the MCMC hammer. *Publ. Astron. Soc. Pac.* **125**(925), 306 (2013)
31. Lunn, D., et al.: *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press (2012)
32. Kruschke, J.: *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press/Elsevier (2011)
33. Barry, J.: *Doing Bayesian data analysis: A tutorial with R and BUGS*. *Eur. J. Psychol.* **7**(4), 778 (2011)
34. Carlin, B.P., Louis, T.A.: *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC (2010)
35. Achcar, J.A., Santander, L.A.M.: Use of approximate Bayesian methods for the Block and Basu bivariate exponential distribution. *J. Ital. Stat. Soc.* **2**(3), 233–250 (1993)
36. dos Santos, C.A., Achcar, J.A.: A Bayesian analysis for the Block and Basu bivariate exponential distribution in the presence of covariates and censored data. *J. Appl. Stat.* **38**(10), 2213–2223 (2011)
37. Achcar, J., Davarzani, N., Souza, R.: Basu-Dhar bivariate geometric distribution in the presence of covariates and censored data: a bayesian approach. *J. Appl. Stat.* **43**(9), 1636–1648 (2016)
38. Davarzani, N., et al.: A Bayesian analysis for the bivariate geometric distribution in the presence of covariates and censored data. *J. Stat. Manag. Syst.* **20**(1), 1–16 (2017)
39. Scollnik, D.P.M.: Bayesian inference for three bivariate beta binomial models. *Open Stat. Probab. J.* **8**(1), 27 (2017)
40. de Oliveira, R.P., et al.: Discrete and continuous bivariate lifetime models in presence of cure rate: a comparative study under Bayesian approach. *J. Appl. Stat.* **46**(3), 449–467 (2019)
41. Najarzadegan, H., et al.: Weighted bivariate geometric distribution: simulation and estimation. *Commun. Stat. Simul. Comput.* (2018). <https://doi.org/10.1080/03610918.2018.1520870>
42. Abbas, N.: On examining complex systems using the q-weibull distribution in classical and bayesian paradigms. *J. Stat. Theory Appl.* **19**(3), 368–382 (2020)
43. Gelfand, A.E., Smith, A.F.: Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**(410), 398–409 (1990)
44. Chib, S., Greenberg, E.: Understanding the metropolis-hastings algorithm. *Am. Stat.* **49**(4), 327–335 (1995)
45. Kruschke, J.: *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press (2014)
46. Fisher, R.A.: On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond.* **222**(594–604), 309–368 (1922)
47. Aitchison, J., Ho, C.: The multivariate Poisson-log normal distribution. *Biometrika* **76**(4), 643–653 (1989)
48. Karlis, D., Meligkotsidou, L.: Finite mixtures of multivariate Poisson distributions with application. *J. Stat. Plan. Inference* **137**(6), 1942–1960 (2007)
49. Sarabia, J.M., Gómez-Déniz, E.: Construction of multivariate distributions: a review of some recent results. (invited article with discussion: M. del Carmen Pardo and Jorge Navarro). *SORT-Statistics and Operations Research Transactions*, pp. 3–36 (2008)
50. Bayes, T.: Bayes An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc.* **53**, 376–418 (1763)
51. Garthwaite, P.H., Kadane, J.B., O’Hagan, A.: Statistical methods for eliciting probability distributions. *J. Am. Stat. Assoc.* **100**(470), 680–701 (2005)
52. Fang, Y.: Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. *J. Data Sci.* **9**(1), 15–21 (2011)
53. Burnham, K.P., Anderson, D.R.: Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**(2), 261–304 (2004)