

Multilevel Bayesian Regression Model with some Selected Horseshoe Flat Priors for Count and Continuous Responses

Rasaki Olawale Olanrewaju^{1,*} and Sodiq Adejare Olanrewaju²

¹Africa Business School (ABS), Mohammed VI Polytechnic University (UM6P), X4JH+QJR, Avenue Mohamed Ben Abdellah Regragui, Rabat 10112, Morocco

²Department of Statistics, University of Ibadan, Ibadan, 900001, Nigeria

Received: 30 Apr. 2023, Revised: 30 May 2023, Accepted: 25 Jun. 2023.

Published online: 1 Sep. 2023.

Abstract: This article provides both the theoretical and experimental analysis for Bayesian multilevel regression model via fully specified random and fixed effects, coupled with variance error. The dependent variables for the multilevel Bayesian regression that are usually supported by distributions that are regarded as horseshoe priors for Bayesian multilevel regression analysis. The classes of horseshoe priors for multilevel regression used in this research are referred to as horseshoe flat priors for count and continuous regression multilevel responses that might be via non-linear and linear multilevel regression models with options for covariance structures, autocorrelation of covariates or responses. Among the horseshoe flat prior used to estimate the corresponding posteriors regression means and predictive checks are Binomial, Poisson, and Negative-Binomial; and Gamma, Lognormal for count and continuous responses respectively.

Keywords: Bayesian Multilevel Regression, horseshoe flat prior, random effects, fixed effects.

1 Introduction

Hierarchical models otherwise known as multilevel models do provide immense flexibility, dynamics, and diversity to regression analysis. In a different term, they are models that do allow observations to be measured on separate levels at the same time interval [1, 2]. However, when the observations are collected or collated in hierarchical or clumping structures, the accordant generalization would be multilevel models. Multilevel models usually account for correlations among or between observations within a given group by integrating cluster-specific random and fixed effects, such that the random and fixed effects are nested [3].

The response variable must be analyzed as rank, hierarchical, clump, or multilevel analysis with appropriate probability function that suits it, in comparison with cluster-specified parameters across clusters [4]. Many at times, it is considerably essential in many problems to use the observed data to estimate some hidden aspects of the population. Hierarchical regression models usually use parameters to fit to data with the use of observational and marginal distributions to structure-out dependency into parameters in order to avoid problem of parsimony [5]. For instance, in a biochemistry laboratory with population of different cadre molecular biologists (undergraduate, masters, PhD, and Postdoc candidates) that perform the same task. This appears naturally in accordance to model such as hierarchical structure to derive estimate for each cadre, provide cadre based predictions of future observations, and make conclusion about general population of the cadre biologists [6, 7].

Bayesian methods of inference have widely gained ground in range of scientific and medical fields — meteorology, ecology, and biostatistics due to its adaptability to yield probabilistic-oriented deductions that usually outstripped the deterministic approach. Among the advantages of Bayesian methods are, to lean on the strength of harmonizing pre-knowledge about the model coefficients with grounded information from the data; describe data generating process, derive boundaries for confidence intervals of estimated parameters; and to make predictions [8, 9]. Bayesian inference has been described as intractable in deriving model parameters in hierarchical models, but became tractable when the data to be analyzed often have a hierarchical structure [10, 11]. Employing Bayesian analysis for solving multilevel models makes it possible to pliability specify hierarchical structure parameters using priors; ability to manage small sample size and over-parameterization of likelihoods and information criteria via well-chosen priors; and ability to

*Corresponding author e-mail: olanrewaju_rasaq@yahoo.com

spontaneously derive and interpret results of credible intervals [12, 13]. These highlighted features of Bayesian methods for shaping multilevel regression analysis can be estimated via Markov Chain Monte Carlo (MCMC), Gibbs-sampling, or Metropolis-Hastings algorithms. Among the problems of these algorithms are they usually possessed slow rate of convergence when dealing with high and ultra-dimensional models with influence of collinearity effects among or between parameters; requirement of the prior to be conjugate of the likelihood of the parameter among others [14, 15]. It was in line of the lacunas of the aforementioned algorithms that Stan implemented Hamiltonian Monte Carlo called “Stan” algorithm — the algorithm was known for its quick convergence for Bayesian high and ultra-dimensional models regardless of whether a conjugate or non-informative priors were specified [16, 17]. In line with the usefulness of the Stan algorithm, the Stan algorithm will be employed as alternative algorithm for the mentioned algorithms for the Bayesian multilevel regression model in this work [18].

The class of priors to be chosen for Bayesian multilevel regression models might be difficult to ascertain due to the fact covariate(s) and response measurements must be supported by appropriate probabilistic functions (either Probability Density Function (PDF) or Probability Mass Function (PMF)). According to [19], most of the probabilistic functions are always carved-out in line with the distribution of the response variable with the support of added parameters like location, scale, shape, hurdle, Beta etc. These supported distributions are most regarded as the priors for Bayesian multilevel regression analysis. According to [20, 21], the class of priors for multilevel regression analyses are sometimes referred to as horseshoe prior. However, changing of priors based on response observations of the hierarchical regression has been partially criticized in order to avoid the problem of too much informative default priors [22]. The problematic aspect is that not all researchers are well educated about reasonable default priors for Bayesian methods. In fact, in some instances priors have to be test run, especially when one is not certain about the most robust prior in a class of appropriate distributions for response variable [23]. [24] adopted Bayesian approach in estimating Maximum a Posteriori (MAP) via Schuster method for making deductive inference about rates of heartbeats of newly born babies’ variability at exactly an hour and two hours after birth. Instead of adopting either a count or continuous likelihood with appropriate class of prior, an alternate periodogram and time axe Bayesian inferential statistics via known and unknown white noise was adopted, but pinpointed that it might fail in the presence hierarchical dataset. [25] and [26] juxtaposed evidence based standard approach and Bayes factor approach of modeling multilevel Bayesian regression that involves complex data with different levels of stratification. They employed simplest linear-type models with unrealistic priors, but came to term that direct statistical computation is unrealistic. They came into conclusion that sequential Markov Chain Monte Carlo sampling algorithm yielded estimates that are more consistent, but not very clear how well the technique performs particularly on data with hierarchical structure.

[27] and [28] juxtaposed valid inference from Bayesian multilevel regression and post stratification regression of hierarchical sampling scheme from non-probabilistic surveys. They adopted prior distributions of specified variance parameter and pinpointed that choice of prior distribution will be more challenging when observations are from highly selective sampling stage. They evaluated the choice of biasedness effect estimation of different distribution via Monte Carlo simulation and concluded that biasedness will be effective for proportion based sampling scheme. [29] concluded that high autocorrelation usually occur frequently in hierarchical Bayesian model models, especially when model simplifications or informative priors are often needed. He also affirmed that group-specific effects normally occur at different hierarchical levels anytime nested is needed, with an assumption that groups usually vary randomly with Gaussian distribution as a priori. He now affirmed that some special type of priors might be needed.

The core objectives of this research work are: (1) A full sketch Bayesian multilevel model with complete conditional posterior distribution for fixed and random effects will be specified with their associated deductive inference via the use of horseshoe flat prior. (2) The complete conditional posterior distribution of the variance-covariance matrix of the random effect will be derived as well. (3) Objectives one and two will be derived for some selected count and continuous likelihoods like — Negative-Binomial, Lognormal, Gamma, Poisson. (4) The hierarchical clustering of three varieties of grains — Kama wheat, Rosa wheat, and Canadian wheat; and readings of rate of heartbeats of newly born babies will be subjected to estimated solutions in (1), (2), and (3) for deductive inference.

Motivation for the study: Multilevel Bayesian model has received little attention, because of the uncertainty surrounding the appropriate class of prior to be employed that will yield reliable and deductive posterior coefficients and well-informed information criteria that will be effective for large and small sample sizes’ scenarios. However, Markov Chain Monte Carlo (MCMC) and Gibbs-sampling algorithms have once been used to mend-up some conjugate and non-informative priors for multilevel Bayesian regression, but they possessed slow rate of convergence. It is because of the slow rate of convergence by these mentioned algorithms that we carried out this research with rare horseshoe flat prior for some selected count and continuous likelihood via Stan implemented Hamiltonian Monte Carlo algorithm called “Stan”, in order to improve the slow rate of convergence.

In this light, this work shall widen the scope of distributional priors to a wide range of probabilistic functions for multilevel regression models. Among the wide range of distributions and link functions to be subjected to horseshoe flat prior multilevel regression models are Binomial, Poisson, and Negative-Binomial for count responses in order to estimate the associated posterior-predictive checks, performance criteria, credible interval, and leave-one-out cross

validation. For nonlinear and linear multilevel regression models with continuous-type related responses and options for covariance structures, autocorrelation of covariates horseshoe flat prior specification via Gamma, Lognormal, etc. will be employed to estimate the associated posterior-predictive checks, performance criteria, credible interval, and leave-one-out cross validation.

2 Model Specifications

Let " y " be a response of linear combination of predictors, say it can be transformed by an inverse link function say " g " that assumes certain distribution for the response variable " y ". Then,

$$y_i \sim D(g(\eta_i), \beta) \quad (1)$$

i^{th} is the data point. " D " is usually regarded as the "family". The parameter β delineates supplementary family of stipulated coefficients that do not vary across data points, such that coefficients can be standard deviation σ associated to the Gaussian models, shape coefficient in Gamma, "p" in Binomial, or Negative-Binomial distributions. It is to be noted that the distribution of the response variable " D " can be any type of distribution. The predictor can be written in matrix form as:

$$\eta = X\beta + Z\mu \quad (2)$$

The population-level and group-level coefficients are β and μ respectively, while X and Z are their corresponding design matrices for the response variable " y " that make-up the data. β and μ are fixed and random effects are model parameter to be estimated respectively.

2.1 Bayesian Multilevel Model Specification

Assuming a two-level normal model of y_{ij} of the response variable measured for each observation j in clusters i with mean π_{ij} , then

$$g(\pi_{ij}) = x_{ij}^T \beta + z_{ij}^T \delta_i + \mu_{ij} \quad (3)$$

$$\delta_i \sim N(0, \Theta) \quad (4)$$

$$\mu_{ij} \sim N(0, \sigma_\mu^2) \quad (5)$$

$\ni x_{ij}$ is $(p \times 1)$ and z_{ij} is $(k \times 1)$ design vectors for fixed effects β and random effects δ_i ; p and k are numbers of fixed and random effects. δ_i is a $(k \times 1)$ vector of random effects for cluster i . $g(\pi_{ij})$ is the link function. Θ and σ_μ^2 stand for the model covariance matrix of the random effects and error variance respectively. The prior specification for Θ , and the marginal priors for Θ and σ_μ^2 components are of interest.

2.2 Complete Derivation for Posterior Distributions

Let X be $(n \times p)$ design matrix for X_i ; β be the fixed effects and Z_i be $(j \times k)$ design matrix for random effects $\ni k$ is the design vector for random effects vector β_i .

2.3 The Complete Conditional Posterior for Fixed Effects (β)

The complete conditional posterior for fixed effects is the fixed effects proportional to product of the likelihood and fixed effects of horseshoe prior. That is, when the normalizing factor has been factored-out.

$$P(\beta / y, Z, X, \delta_i, \Theta, \sigma_\mu^2) \propto P(y / \beta, Z, X, \delta_i, \Theta, \sigma_\mu^2) P(\beta) \quad (6)$$

Where,

$$P(y/\beta, Z, X, \delta_i, \Theta, \sigma_\mu^2) = \prod_{i=1}^n \prod_{j=1}^J P(y_{ij} / \beta^T x_{ij} + \delta_i^T z_{ij}, \Theta, \sigma_\mu^2) \quad (7)$$

$$\propto \exp\left(-\frac{1}{2\sigma_\mu^2} RSS\right) \quad (8)$$

Such that, Residual Sum of Square

$$RSS = \sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \beta^T x_{ij} - \delta_i^T z_{ij}) \quad (9)$$

Let,

$$\bar{y}_{ij} = y_{ij} - \delta_i^T z_{ij}$$

$$\text{Then } RSS = \sum_{i=1}^n \sum_{j=1}^J (\bar{y}_{ij} - \beta^T x_{ij})^2$$

Writing RSS in a matrix form

$$RSS = (\bar{y} - X\beta)^T (\bar{y} - X\beta) \quad (10)$$

$$RSS = \bar{y}^T \bar{y} - 2\beta^T X \bar{y} + \beta^T X^T X \beta \quad (11)$$

It implies that,

$$P(y/\beta, Z, X, \delta_i, \Theta, \sigma_\mu^2) \propto \exp\left[-\frac{1}{2\sigma_\mu^2} (\bar{y}^T \bar{y} - 2\beta^T X \bar{y} + \beta^T X^T X \beta)\right] \quad (12)$$

According to some authors, reasonable non-informative horseshoe prior distribution of $P(\beta) \propto 1$ and $P(\sigma_\mu^2) \propto \sigma_\mu^{-2}$ recommended to have significant effect on making inferences on random and fixed effects.

Combining equation (12) with $P(\beta) \propto 1$ gives:

$$P(y/\beta, Z, X, \delta_i, \Theta, \sigma_\mu^2) \propto \exp\left[-\beta^T \left(\frac{X^T \bar{y}}{\sigma_\mu^2}\right) - \frac{1}{2} \beta^T \left(\frac{X^T X}{\sigma_\mu^2}\right) \beta\right] \quad (13)$$

$$\beta \sim N\left[\left(\frac{X^T X}{\sigma_\mu^2}\right)^{-1} \frac{X^T \bar{y}}{\sigma_\mu^2}, \left(\frac{X^T X}{\sigma_\mu^2}\right)^{-1}\right] \quad (14)$$

2.4 The Complete Conditional Posterior for Random Effects (δ_i)

$$P(\delta_i / y, Z, X, \delta_i, \Theta, \sigma_\mu^2) = P(y_i / \beta, Z, X, \delta_i, \Theta, \sigma_\mu^2) P(\delta_i) \quad (15)$$

Having ascertained that,

$$P(y_i / (\bullet)) = \prod P(y_{ij} / \beta^T x_{ij} + \delta_i^T z_{ij}, \sigma_\mu^2) \propto \exp\left(-\frac{1}{2\sigma_\mu^2} RSS\right) \quad (16)$$

$$RSS = \sum_{j=1}^J (y_{ij} - \beta^T x_{ij} - \delta_i^T z_{ij})$$

$$\text{Letting } \bar{y}_{ij} = y_{ij} - \beta^T x_{ij}$$

$$RSS = \sum_{j=1}^J (\bar{y}_{ij} - \delta_i^T z_{ij})^2 = (\bar{y} - \delta_i^T Z_i)^T (\bar{y} - \delta_i^T Z_i) \quad (17)$$

$$= \bar{y}^T \bar{y} - 2\delta_i^T Z_i \bar{y} + \delta_i^T Z_i^T Z_i \delta_i \quad (18)$$

$$P(y_i / (\bullet)) \propto \exp\left[-\frac{1}{2\sigma_\mu^2} (\bar{y}^T \bar{y} - 2\delta_i^T Z_i \bar{y} + \delta_i^T Z_i^T Z_i \delta_i)\right] \quad (19)$$

It is also known that the prior is of 2-stage equation

$$\delta_i \sim N(0, \Theta)$$

$$P(\delta_i) \propto \exp\left[-\frac{1}{2}(\delta_i^T \Theta \delta_i - 2\delta_i^T \Theta^{-1} 0)\right] \quad (20)$$

Combining (19) and (20), that is, the combination of the sampling model and horseshoe prior

$$\delta_i \sim N(0, \Theta)$$

$$P(\delta_i) \propto \exp\left[-\frac{1}{2\sigma_\mu^2}(\delta_i^T Z_i^T \delta_i - 2\delta_i^T Z_i \bar{y}) - \frac{1}{2}(\delta_i^T \Theta^{-1} \delta_i - 2\delta_i^T \Theta^{-1} 0)\right] \quad (21)$$

$$= \exp\left[-\frac{1}{2}\delta_i^T \left(\Theta^{-1} + \frac{Z_i^T Z_i}{\sigma_0^2}\right) \delta_i + \delta_i^T \left(\Theta^{-1} 0 + \frac{Z_i^T \bar{y}_i}{\sigma_0^2}\right)\right] \quad (22)$$

$$\delta_i / (\bullet) \sim N\left[\left(\Theta^{-1} + \frac{\delta_i^T \delta_i}{\sigma_\mu^2}\right)^{-1} \left(\frac{Z_i^T \bar{y}_i}{\sigma_\mu^2}\right), \left(\Theta^{-1} + \frac{Z_i^T Z_i}{\sigma_\mu^2}\right)^{-1}\right] \quad (23)$$

2.5 The Complete Posterior Distribution for the Error Variance (σ_μ^2)

Expressing σ_μ^2 complete conditional posterior as:

$$P(\sigma_\mu^2 / y, Z, X, \delta_i, \beta, \Theta) = P(y / \beta, \delta_i, Z, X, \delta_i, \Theta, \sigma_\mu^2) P(\sigma_\mu^2) \quad (24)$$

The sampling model for the complete conditional distribution of β

$$P(y / \beta, Z, X, \Theta, \sigma_\mu^2) = \prod_{i=1}^n \prod_{j=1}^J P(y_{ij} / \beta^T x_{ij} + \delta_i^T z_{ij}, \Theta, \sigma_\mu^2) \quad (25)$$

$$= \prod_{i=1}^n \prod_{j=1}^J (2\pi\sigma_\mu^2)^{-1} \exp\left[-\frac{(y_{ij} - \beta^T x_{ij} - \delta_i^T z_{ij})^2}{2\sigma_\mu^2}\right] \quad (26)$$

Since our interest is on σ_μ^2 , then

$$P(y / \beta, Z, X, \Theta, \sigma_\mu^2) \propto (\sigma_\mu^2)^{-\frac{N}{2}} \exp\left[-\frac{\sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \beta^T x_{ij} - \delta_i^T z_{ij})^2}{2\sigma_\mu^2}\right] \quad (27)$$

$$\propto (\sigma_\mu^2)^N \exp\left(-\frac{RSS}{2\sigma_\mu^2}\right) \quad (28)$$

$\ni n_1 + n_2 + \dots + n_n = \sum_{i=1}^n n_i = N$ is the entire sample size, (That is, the complete observations within specified clusters).

The total horseshoe prior is $P(\sigma_\mu^2) \propto \sigma_\mu^2$, the complete conditional posterior is as:

$$P(\sigma_\mu^2 / y, Z, X, \delta_i, \beta, \Theta) = (\sigma_\mu^2)^{-\frac{N}{2}-1} \exp\left(-\frac{1}{2\sigma_\mu^2} RSS\right) \quad (29)$$

$$\sigma_\mu^2 \sim IG\left(\frac{N}{2}, \frac{RSS}{2}\right) \quad (30)$$

IG is the Inverse Gamma.

2.6 The Complete Conditional Posterior Distribution of the Random Effect Variance- Covariance Matrix

$$P(\delta_i / \Theta) \propto |\Theta|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} tr(V_b \Theta^{-1})\right) \quad (31)$$

Where $V_b = \sum_i \delta_i \delta_i^T$, where $\bar{V}_b = \frac{1}{n} \sum_i \delta_i \delta_i^T$ is the empirical variance-covariance, "tr" stands for the trace matrix.

Given a horseshoe flat prior

$$P(\Theta) \propto 1$$

The complete conditional posterior for Θ is

$$P(\Theta / \bullet) \propto |\Theta|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \text{tr}(V_b \Theta^{-1})\right) \quad (32)$$

$$\propto IW\left(n-k-1, \sum_i \delta_i \delta_i^T\right) \quad (33)$$

IW stands for the Inverse-Wishart distribution

Given the Inverse-Wishart horseshoe prior

$$P(\Theta) \propto IW\left(v, \sum_0 \delta_0 \delta_0^T\right) \quad (34)$$

$$\propto |\Theta|^{-\frac{(v+k+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}(V_b \Theta^{-1})\right) \quad (35)$$

The complete conditional posterior for Θ is

$$P(\Theta / \bullet) \propto |\Theta|^{-\frac{(n+v+k+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (36)$$

$$\propto IW\left(n+v, \sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) \quad (37)$$

Given the Negative-Binomial horseshoe prior

$$P(\Theta) \propto NB\left(r, \sum_0 \delta_0 \delta_0^T\right) \quad (38)$$

$$\propto |\Theta|^{-\frac{(r+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (39)$$

The complete conditional posterior for Θ is

$$P(\Theta / \bullet) \propto |\Theta|^{-\frac{(n+r+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (40)$$

$$\propto B\left(\alpha + rn, \beta + \left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right)\right) \quad (41)$$

Given the Lognormal horseshoe prior

$$P(\Theta) \propto LN\left(rv, \sum_0 \delta_0 \delta_0^T\right) \quad (42)$$

$$\propto |\Theta|^{-\frac{(v+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (43)$$

The complete conditional posterior for Θ is

$$P(\Theta / \bullet) \propto |\Theta|^{-\frac{(n+v+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (44)$$

$$\propto LN\left(\log(v+n), \log\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right)\right) \quad (45)$$

Given the Gamma horseshoe prior

$$P(\Theta) \propto \text{Gamma}\left(\alpha + \beta, \sum_0 \delta_0 \delta_0^T\right) \quad (46)$$

$$\propto |\Theta|^{-\frac{(\alpha+\beta+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}(V_b \Theta^{-1})\right) \quad (47)$$

The complete conditional posterior for Θ is

$$P(\Theta / \bullet) \propto |\Theta|^{-\frac{(n+\alpha+\beta+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (48)$$

$$\propto \text{Gamma}\left(n\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) + \alpha, n + \beta\right) \quad (49)$$

Given the Poisson horseshoe prior

$$P(\Theta) \propto \text{Poi}\left(\lambda, \sum_0 \delta_0 \delta_0^T\right) \quad (50)$$

$$\propto |\Theta|^{-\frac{(\lambda+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(V_b \Theta^{-1}\right)\right) \quad (51)$$

The complete conditional posterior for Θ is

$$P(\Theta / \bullet) \propto |\Theta|^{-\frac{(n+\lambda+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) \Theta^{-1}\right) \quad (52)$$

$$\propto \text{Poi}\left(n\left(\sum_0 \delta_0 \delta_0^T + \sum_i \delta_i \delta_i^T\right) + \lambda\right) \quad (53)$$

3 Numerical Analysis

The hierarchical clustering of grain samples was worked-out with in this section. The grains are in three varieties — Kama wheat, Rosa wheat, and Canadian wheat, such that their measurements were provided in terms of area and perimeter of land planted; the spatial space between / among planted varieties (that is the compactness); the length and width of the planted grains; the occupying space (occupying space) and the long narrow furrow cut by a natural process, such as erosion (groove). The sample of the grain varieties with aforementioned measurements are of 210 sample points. The dataset can be obtained from the website- <https://github.com/benjaminwilson/python-hierarchical-clustering-exercises/blob/master/datasets/seeds-less-rows.csv>. The three varieties of grains make it to be a three level of hierarchical dataset with associated measurements that defined there outcome.

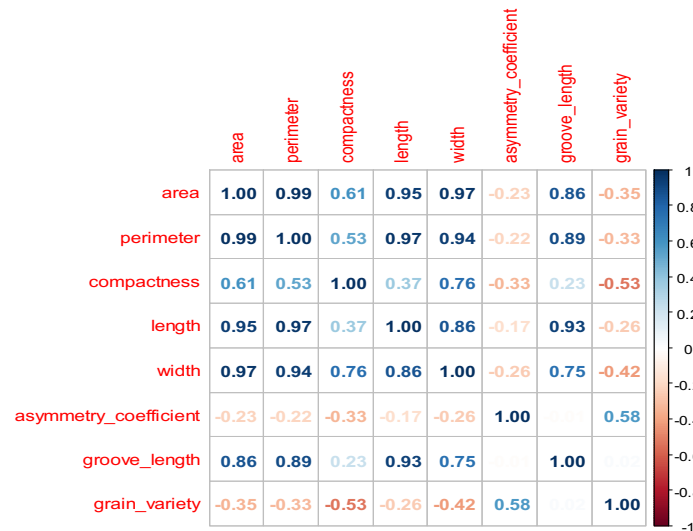


Fig.1: Correlation Plot of the Measurements of the Grains.

There is a strong positive correlation between the area of the soil the varieties of grains were planted in and its perimeter, compactness, length, width, and groove length, where there is a weak negative correlation between the area of the soil the varieties of grains planted in and its asymmetry. There is a strong positive correlation between perimeter of the soil the varieties of the grains were planted in and its compactness, length, width and its groove; whereas there is a weak negative correlation between its asymmetry coefficient. In a similar vein, the spatial space of grains possessed a positive correlation between its area, perimeter, length, width, and groove; whereas a negative correlation exists between compactness and its asymmetry coefficient. Same positive association goes for length and width with other measurements expect for asymmetry (See fig. 1 above).

Table 1: Family of Prior Distributions for the Grains Measurements.

Criteria	Binomial (B)	Zero inflated Binomial	Poisson	Zero inflated Poisson	Negative Binomial (NB)	Zero Inflated NB
elpd_waic	-115.1 (8.3)	-107.3 (8.7)	-207.4 (8.0)	-208.4 (8.0)	-208.5 (8.1)	-209.5 (8.1)
p_waic	4.8	5.9	2.4	2.4	2.4	2.4

	(0.7)	(1.0)	(0.3)	(0.3)	(0.3)	(0.3)
waic	230.2 (16.5)	214.6 (17.3)	414.9 (16.0)	416.8 (16.1)	417.0 (16.1)	418.9 (16.2)
elpd_loo	-115.1 (8.3)	-107.3 (8.7)	-207.5 (8.0)	-208.4 (8.0)	-208.5 (8.1)	-209.5 (8.1)
p_loo	4.8 (0.7)	6.0 (1.0)	2.4 (0.3)	2.4 (0.3)	2.4 (0.3)	2.4 (0.3)
looic	230.2 (16.5)	214.6 (17.4)	414.9 (16.0)	416.8 (16.1)	417.0 (16.1)	419.0 (16.2)

In brackets are standard errors of estimates. Keys: Zero Inflated Binomial=ZIB; Zero Inflated Poisson=ZIP; Negative Binomial=NB; Zero Inflated=ZI

Table 2: Posterior Regression Coefficients of the Population-Level Effects of the Grains.

Population-Level Effects:	Est.	Est. Error	1-95 % CI	μ -95 %CI	Bulk_ESS	Tail_ESS
Intercept	39.93	17.17	6.85	73.72	1850	2171
Groove length	3.38	0.62	2.19	4.63	2438	2504
Width	-0.79	2.11	-4.96	3.22	2678	2832
Length	-3.71	1.22	-6.11	-1.32	2704	2702
Compactness	-21.38	12.78	-46.23	3.54	2179	2423
Perimeter	-2.53	1.23	-5.00	-0.12	1781	2192
Area	1.34	0.62	0.17	2.55	1892	2292
Asymmetry coefficient	0.15	0.06	0.04	0.26	2925	2565

Keys: Est. = Estimate; ESS=Estimated Sample Size

Table 3: Zero-Inflated Binomial Family Specific Parameters.

Criteria	Est.	Est. Error	1-95% CI	μ -95% CI	Bulk_ESS	Tail_ESS
Shape	104.88	78.63	22.11	313.78	3453	2831
z_i	0.86	0.01	0.00	0.03	2643	1515

Table 1 to table 3, Widely Applicable Information Criterion (WAIC), otherwise known as Watanabe-Akaike Information Criteria (WAIC) viewed as an improvement for Deviance Information Criterion (DIC) for Bayesian models, as well as Leave-One-Out Information Criterion (LOOIC) for cross-validation will be adopted to measure or juxtapose reduced posterior error performance. WAIC is fully for Bayesian and closely approximates Bayesian cross-validation. Unlike DIC, WAIC is an invariant to parameterization and works for singular models as well. Expected Log Point-Wise Predictive Density for observations (ELPD) for WAIC and LOO are specified as elpd_waic and elpd_loo respectively, while their p-values are p_waic and p_loo respectively. Zero-Inflated Binomial prior produced the same minimum reduced error performance estimate for WAIC and LOOIC, valued at 214.6 among the discrete distributional priors studied. However, according to [30], whenever p_waic estimates is greater than 0.4, loo is usually recommended, but in this scenario the P-value of WAIC (that is p-waic) is far greater than 0.4, that is, $5.9 > 0.4$, yet WAIC and LOOIC still coincides (See Table 1 to 3).

The positive posterior mean estimates of 3.38, 1.34, and 0.15 with standard errors of 0.62, 0.62, and 0.06 for Groove length, Area and Asymmetry coefficient covariates implies that they contribute positively to the chance of a grain being either Kama wheat, Rosa wheat, or Canadian wheat keeping other factors constant. While negative posterior mean estimates of -0.79, -3.71, -21.38, and -2.53 with standard errors of 2.11, 1.22, 12.78, and 1.23 for Width, Length, Compactness, and Perimeter covariates implies they do not contribute towards ascertaining whether a grain is either Kama wheat, Rosa wheat, or Canadian wheat. The two-sided posterior of the means at 95% credible intervals (1-95% CI and μ -95% CI) based on quantiles confirmed that they are true estimates, because they fall within the intervals. The Bulk Estimated Sample Size (Bulk ESS) and Tail Estimated Sample Size (Tail ESS) values are estimation of the effective large and small sample sizes that the posterior distribution would needed independently to produce the same standard error of the posterior mean as if the dependent samples return by MCMC algorithm. The zero-inflated probability z_i is very much large with mean of 86%. This connote that there is 0.14% chance of the measurements not contributing to the chance of not having any of the grains (See Figure 2 below).

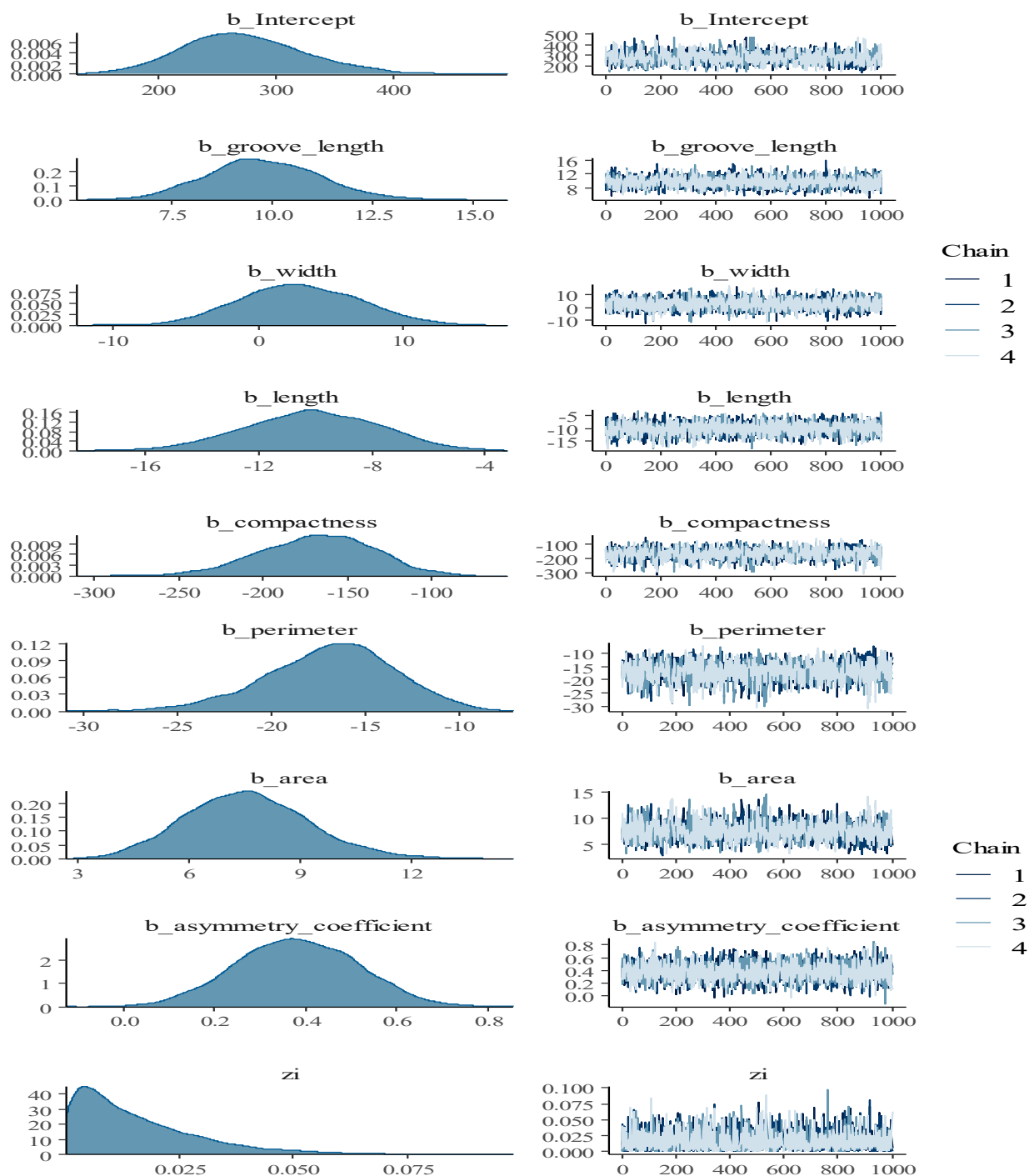


Fig. 2: Posterior Distributions of the Covariates Contributions to Varieties of Grains.

From figure 2 above, the posterior distributions of covariate contributions of multilevel regression intercept, groove length, width, length, area, and asymmetry coefficients for the hierarchical clustering of grain samples — Kama wheat, Rosa wheat, and Canadian wheat with family of prior distributions possessed nearly perfect symmetry distributions respectively. While distributions of the perimeter and the parameter of the Zero-Inflated Binomial coefficients possessed right and left skewed distribution respectively.

It is to be noted that the posterior predictive distribution of the Zero-Inflated Binomial possessed a three distributive switching non-asymmetry distributions (See figure 5 of the Appendix).

3.1 Section 2

In this section, a secondary dataset of readings of rate of heartbeats of just born babies in a teaching hospital, Lagos University Teaching Hospital (LUTH) would be used. These rates of heartbeats were recorded in successive hours of one, two, and three immediately after delivery (that is, after birth) with the use of Electrocardiogram (ECG). The readings were recorded for nine hundred and fifty (950) babies, such that the readings of heartbeats after three hours (HR_3) depend on the successive hours of the first (HR_1) and second (HR_2) hours of the readings.

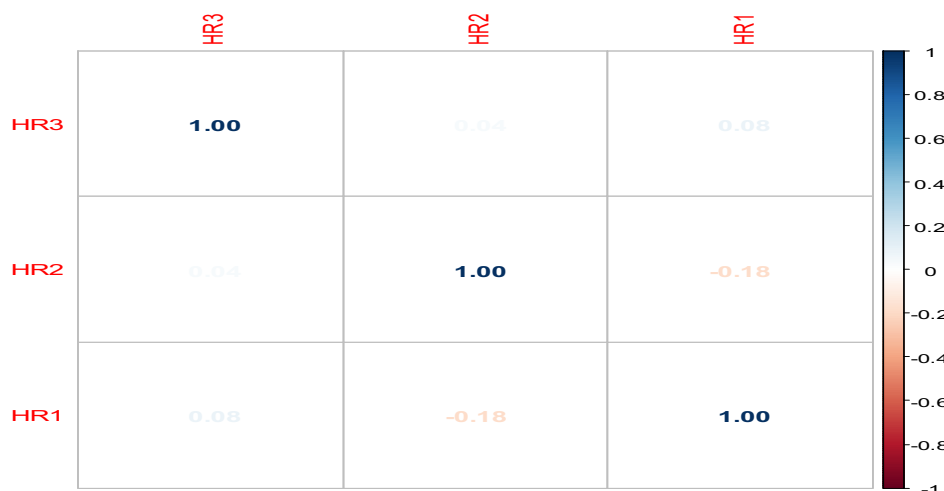


Fig.3: Correlation Plot of the three Successive Hours after Birth.

There is a weak negative correlation between heartbeats of babies recorded after an hour (HR_1) and two hours (HR_2) after birth. There is a weak positive correlation between heartbeats of babies recorded after three hours (HR_3) and an hour (HR_1) after birth. There is also a weak positive correlation between heartbeats of babies recorded after three hours (HR_3) and two hours (HR_2) after birth (See figure 3 above).

Table 4: Family of Prior Distributions for the Hours of Heartbeats.

Criteria	Lognormal	Gaussian	Asym Laplace	Ex-Gaussian	Exponential	Fréchet	Hurdle Gamma
elpd_waic	-2476.5 (36.4)	-2513.5 (39.8)	-2431.1 (28.6)	-2396.6 (27.8)	-4819.9 (1.8)	-2402.2 (25.8)	-2488.9 (37.2)
p_waic	5.3 (0.7)	5.6 (0.8)	6.2 (0.2)	4.8 (0.3)	0.0000 (0.000)	4.1 (0.4)	5.2 (0.6)
waic	4953.0 (72.7)	5026.9 (79.7)	4862.3 (57.1)	4793.1 (55.7)	9639.8 (3.6)	4804.4 (51.6)	4977.7 (74.4)
elpd_loo	-2476.5 (36.4)	-2513.5 (39.8)	-2431.1 (28.6)	-2396.6 (27.8)	-4819.9 (1.8)	-2402.2 (25.8)	-2488.9 (37.2)
p_loo	5.4 (0.7)	5.6 (0.8)	6.2 (0.2)	4.8 (0.3)	0.000 (0.000)	4.1 (0.4)	5.2 (0.6)
looic	4953.0 (72.7)	5026.9 (79.7)	4862.3 (57.1)	4793.1 (55.7)	9639.8 (3.6)	4804.5 (51.6)	4977.7 (74.4)

Criteria	Hurdle Lognormal	Shifted Lognormal	Skew-Normal	Student-t	Weibull
elpd_waic	-2477.4 (36.4)	-2404.7 (27.4)	-2426.9 (29.3)	-2422.9 (32.0)	-2753.3 (42.0)
p_waic	5.2 (0.6)	5.0 (0.4)	5.6 (0.5)	3.8 (0.2)	10.5 (2.8)
waic	4954.7 (72.9)	4809.4 (54.9)	4953.0 (72.7)	4845.9 (64.0)	5506.7 (84.0)
elpd_loo	-2477.4	-2404.7	-2426.9	-2422.9	-2753.3

	(36.4)	(27.4)	(29.3)	(32.0)	(42.0)
p_loo	5.2 (0.6)	5.0 (0.4)	5.6 (0.5)	3.9 (0.2)	10.5 (2.8)
looic	4954.7 (72.9)	4809.4 (54.9)	4853.9 (58.6)	4845.9 (64.0)	5506.6 (83.9)

Keys: Asym._Laplace=Asymmetric Laplace; Ex-Gaussian=Exponentiated Gaussian;
Exp.=Exponential

Table 5: Posterior Regression Coefficients of the Population-Level Effects of the Hours of Heartbeats.

Population-Level Effects:	Est.	Est. Error	1-95% CI	μ -95% CI	Bulk_ESS	Tail_ESS
Intercept	50.34	2.36	45.98	55.06	4402	2496
HR ₂	0.03	0.02	0.00	0.06	4749	3226
HR ₁	0.05	0.02	0.02	0.09	4177	2820

Table 6: Ex-Gaussian Family Specific Parameters

Family Specific Parameters:	Est.	Est. Error	1-95% CI	μ -95% CI	Bulk_ESS	Tail_ESS
Sigma (σ)	1.86	0.09	1.68	2.04	3995	2957
Beta (β)	2.65	0.15	2.37	2.94	3365	2621

Keys: Est. Error= Estimated Error

Table 4 to table 6, Weibull distribution produced the reduced-error model performance for elpd_waic valued at -2753.3, but its p-value for the p-waic = 10.5 is greater than four, that is, $10.5 > 0.4$. Since its elpd's p-value is greater than four (4), its model performance is not reliable and not recommended. However, the LOO model performance will be switched to, among the pool of continuous distributions for LOO; Ex-Gaussian produced the barely minimum reduced-error performance with its Leave-One-Out Information Criterion (LOOIC) valued to be 4793.1. The positive posterior mean estimates of 0.03 and 0.05 with coincide standard error of 0.02. The near posterior mean estimate for (HR₁) and (HR₂) and coincide standard error of 0.02 makes it reasonable that the successive record of heartbeats after an hour and two hours after birth will constitute and influence the impulse readings after three hours of birth. The estimated sigma of 1.86 for the family of the Ex-Gaussian parameter connotes a not too far variation among the observations, while the estimated Beta value of 2.65 speaks volume of a relatively low volatility (See Table 4 to 6).

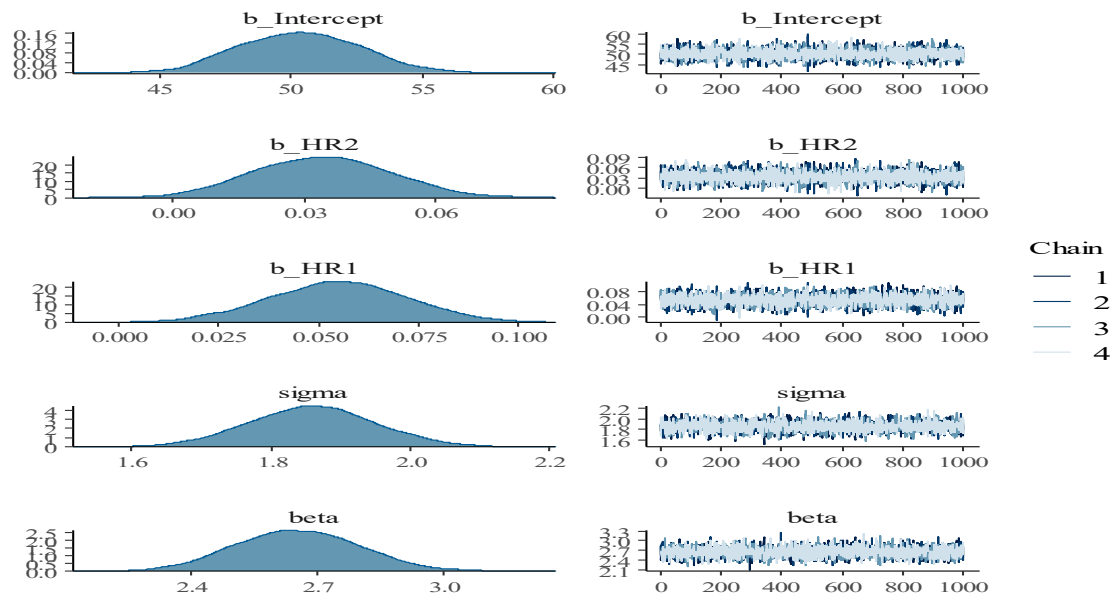


Fig. 4: Posterior Distributions of the Successive Hours of Heartbeats after Birth.

From figure 4 above, the posterior distributions of heartbeats of newly born babies of the intercept, one and two hours after birth for multilevel regression coefficients, as well as the sigma (σ) and beta (β) coefficients of the posterior distribution of the Ex-Gaussian family specific parameters possessed a bell-shape distribution for all.

It is to be noted that the posterior predictive distribution of the Ex-Gaussian family specific parameters possessed a negative skewed distribution (See figure 6 of the Appendix).

4 Conclusions

In conclusion, Bayesian multilevel hierarchical regression model was proposed in its full specifications with associated complete conditional posterior for fixed effects (β), complete conditional posterior for random effects (δ_i) and complete posterior distribution for the error variance (σ_μ^2). Additionally, the Residual Sum of Squares (RSS) for the model was factored-out and used to assign horseshoe flat prior distributions to the fixed and random effects, as well to the variance error (σ_μ^2). Among the specified posteriors for the model are Inverse-Wishart distribution, Gamma, Lognormal; and Negative-Binomial, Geometric, and Poisson for continuous and discrete observations respectively. Three levels of grains clustering samples were worked-out with Binomial, Zero-Inflated Binomial, Poisson, Zero-Inflated Poisson, Negative-Binomial, and Zero-Inflated Negative-Binomial. Zero-Inflated Binomial gave a reduced coincide WAIC and LOOIC, valued at 214.6 to best represent the generalization. Additionally, Ex-Gaussian posterior gave a perfect representation of babies impulses recorded in a succession of an hour, two hours, and three hours after birth.

Acknowledgments

We will like to acknowledge the support of Lagos University Teaching Hospital (LUTH) for releasing the dataset of newborn babies' heartbeats used in this project. This piece of intellectual property is written in honour of late Prof. (Associate) Johnson Funminiyi Ojo, of the Department of Statistics, University of Ibadan, Ibadan, Oyo state, Nigeria.

Conflicts of Interest

We the authors certify that there is no conflict of interest whatsoever with any affiliation, or involvement with any organization, financial and non-financial entity.

References

- [1] H. Brown, and R. Prescott, *Applied Mixed Models in Medicine*, John Wiley & Sons (2015).
- [2] J. Kim, D. Marcusson-Clavertz, and H.A. Park, Practical Guide to Analyzing Time-Varying Associations between Physical Activity and Affect using Multilevel Modeling, *Computational and Mathematical Methods in Medicine*, 1-11, (2018) doi.org/10.1155/2018/8652034.
- [3] C.A. Kitabo and E.T. Damtie, Bayesian Multilevel Analysis of Utilization of Antenatal Care Services in Ethiopia, *Computational and Mathematical Methods in Medicine*, **11**(3), 874-975 (2020) <https://doi.org/10.1155/2020/8749753>.
- [4] C. Charlton, J. Rasbash, W. Browne, M. Healy, and B. Cameron, *MLwiN In: Centre for Multilevel Modeling*, University of Bristol, Version 3.05, (2020).
- [5] M. O'Malley, I. Brigandt, A. Love, J. Crawford, J. Gilbert, R. Knight, S. Mitchell, and F. Rohwer, *Multilevel Research Strategies and Biological Systems*, Philos. Sci., **81**(5), 811-828 (2014) <https://doi.org/10.1086/677889>.
- [6] A. Gelman, Multilevel (Hierarchical) Modeling: What it Can and Cannot Do, *Technometrics*, **48**, 432-435 (2006) <https://doi.org/10.1198/004017005000000661>.
- [7] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*, Vol. 2, Taylor & Francis (2014).
- [8] J. Hox, and K. Roberts, *Handbook of Advanced Multilevel Analysis*, Wiley Series in Probability and Statistics, New York: Taylor and Francis Group, LLC, Routledge Handbooks Online (2011).
- [9] Y. Chung, S. Rabe-Hesketh, V. Dorie, A. Gelman, and J. Liu, A Non-degenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models, *Psychometrika*, **78**(4), 685-709 (2013) <http://gllamm.org/>.

- [10] S. Watanabe, Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory, *The Journal of Machine Learning Research*, **16**, 3571-3594 (2010).
- [11] A.R. Hassan, R.O. Olanrewaju, Q.C. Chukwudum, S.A. Olanrewaju, and S.E. Fadugba, Comparison Study of Generative and Discriminative Models Classification of Classifiers, *International Journal of Mathematics and Computer Simulation*, **16**(12), 76-87 (2022) [doi:10.46300/9102.2022.16.12](https://doi.org/10.46300/9102.2022.16.12).
- [12] M. Hoffman and A. Gelamn, The No-U-Turn Sampler:Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, *The Journal of Machine Learning Research*, **15**(1), 1593-1623 (2014).
- [13] R.O. Olanrewaju and S.A. Olanrewaju, Multinomial Naïve Bayes Classifier: Bayesian versus Nonparametric Classifier Approach, *European Journal of Statistics*, **2**(8), 1-14 (2022) [doi:10.28924/ada/stat.2.8](https://doi.org/10.28924/ada/stat.2.8).
- [14] J. Rasbash and W.J. Browne, *Hierarchical Multilevel Models in Handbook of Multilevel Analysis*, in J. Leeuw and E. Meijer, Eds., 301-334, Springer Science (2008).
- [15] J.F. Ojo, R.O. Olanrewaju and S. Folorunsho, Bayesian Logistic Regression Using Gaussian Naive Bayes (GNB), *Journal of Medical and Applied Biosciences*, **9**(2), 1-18 (2017).
- [16] M. Betancourt, S. Byrne, S. Livingstone and M. Girolami, *The Geometric Foundations of Hamiltonian Monte Carlo* (2014) [arXiv preprint arXiv:1410.5110](https://arxiv.org/abs/1410.5110).
- [17] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M.A. Brubaker, J. Guo, P. Li, and A. Ridell, Stan: A Probabilistic Programming Language, *Journal of Statistical Software*, **12**(3), (2017).
- [18] Bürkner, P., brms: An R Package for Bayesian Multilevel Models using Stan, *Journal of Statistical Software*, **80**(1), 1-28 (2017) [doi:10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01).
- [19] C.M. Carvalho, N. Polson, and J. Scott, The Horseshoe Estimator for Sparse Signals, *Biometrika*, **12**, 1-16 (2010).
- [20] A. Vehtari, A. Gelman and J. Gabry, Practical Bayesian Model Evaluation using Leave-One-Out Cross-Validation and WAIC, *Statistics and Computing*, **27**(5), 1413-1432 (2017) <https://doi.org/10.1007/s11222-016-9696-4>.
- [21] L. Kanga, X. Kangb, X. Dengc and R. Jind, A Bayesian Hierarchical Model for Quantitative and Qualitative Responses, *Journal of Quality Technology*, **50**(3), 290-308 (2018) <https://doi.org/10.1080/00224065.2018.1489042>.
- [22] B.R. Saville, Bayesian Multilevel Models and Medical Applications, *Journal of Applied Statistics*, **16**(3), 210-217 (2017).
- [23] M. Khane, and W. Shaw, Multilevel Analysis of HIV/AIDS Testing, *Journal of Computational Sciences*, **3**(2), 132-145 (2011).
- [24] R.O. Olanrewaju, Bayesian Approach: An Alternative to Periodogram and Time Axes Estimation for Known and Unknown White Noise, *International Journal of Mathematical Sciences and Computing*, **2**, 22-33 (2018) [doi:10.5815/ijmsc.2018.02.03](https://doi.org/10.5815/ijmsc.2018.02.03).
- [25] T. Edinburgh, A. Ercole, and S. Eglen, Bayesian model selection for multilevel models using integrated Likelihoods, *PLoS ONE*, **18**(2) (2023) <https://doi.org/10.1371/journal.pone.0280046>.
- [26] R.O. Olanrewaju, On the Efficiency and Robustness of Commingle Wiener and L'evy Driven Processes for Vasciek Model, *World Academy of Science, Engineering, and Technology International Journal of Mathematical and Computational Sciences*, **12**(11), 228-233 (2018) waset.org/Publication/10009776.
- [27] C. Bruch and B. Felderer, Prior Choice for the Variance Parameter in the Multilevel Regression and Post-Stratification Approach for Highly Selective Data. A Monte Carlo Simulation Study, *Austrian Journal of Statistics*, **51**, 76-95 (2022) [doi:10.1773/ajs.v51i4.1361](https://doi.org/10.1773/ajs.v51i4.1361). <http://www.ajs.or.at/>.
- [28] S.Y. Lee, Bayesian Nonlinear Models for Repeated Measurement Data: An Overview, Implementation, and Applications, *Mathematics*, **10**, 898-906 (2022) <https://doi.org/10.3390/math10060898>.
- [29] Y. Marchenko, *Bayesian multilevel modeling using Stata*, Vice President, Statistics, and Data Science StataCorp LLC. 2022 UK Stata Conference (2022).

[30] P. Burkner, Advanced Bayesian Multilevel Modeling with the R Package brms, *The R Journal*, **10**(1), 395-411 (2018)

[doi:10.32614/RJ-2018-017](https://doi.org/10.32614/RJ-2018-017).

Appendix

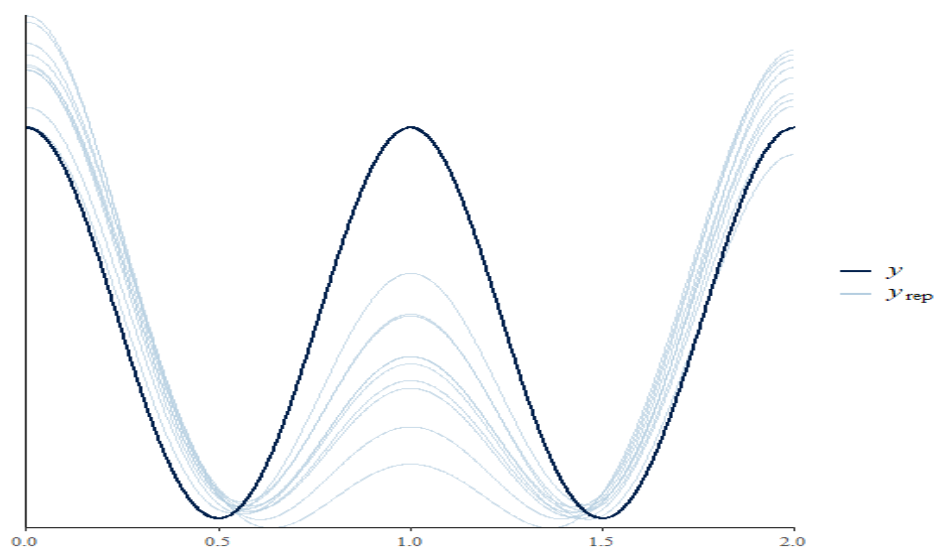


Fig 5: The Predictive Distributions of the Measurements for Varieties of Grains.

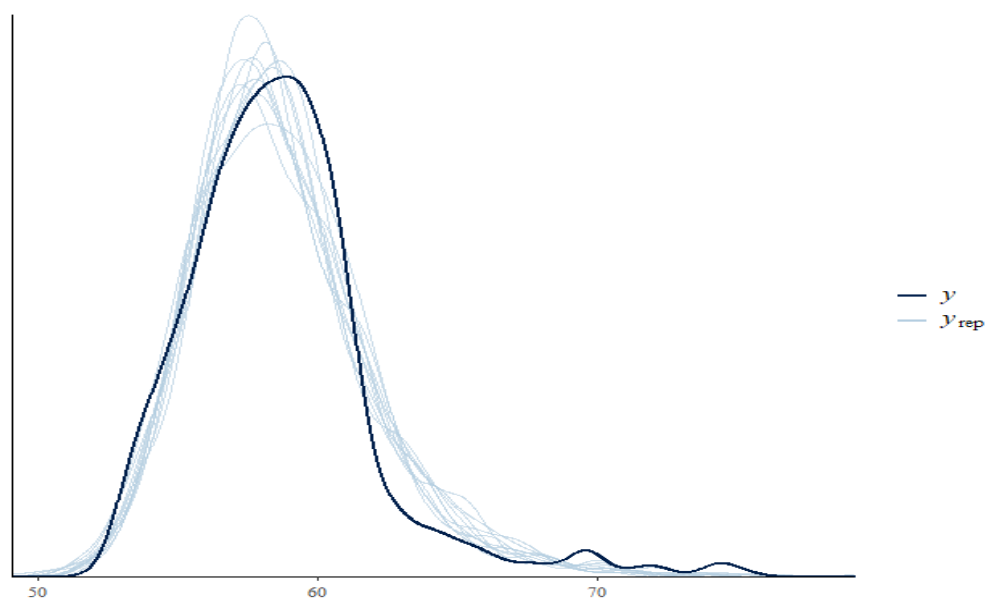


Fig. 6: The Predictive Distributions of the Heartbeats.