



---

## BAYESIAN ANALYSIS OF MIXED EFFECT MODELS AND ITS APPLICATIONS IN AGRICULTURE

Nageena Nazir\*, Abdul Hamid Mir, Athar Ali Khan

Department of Agri. Statistics, University of Agricultural Sciences & Technology of Kashmir, India

Received 12 January 2011; Accepted 28 May 2011

Available online 14 October 2011

**Abstract:** *The mixed effect models has been discussed and implemented from Bayesian viewpoint. In this paper we have made Bayesian analysis of mixed effect models and illustrated its application in agriculture. We focus on linear mixed models with a random intercept and fixed slope. The basic idea behind this approach is to model the phenomenon under study in stages and analyze that model in Bayesian framework. Advancement in the computational power of high speed computers has aided the application part. Suitable illustrations have been proposed on real data set generated on potato crop in year 2005-2006 at five different locations with twelve genotypes including both Yield and Growth attributing characters (tuber weight and Average tuber No.). The models used in this paper have been fitted by lme (fixed, data, random) of nlme library by [12] and it was observed on BIC(Bayesian information criteria) that we should treat locations as random and not as fixed.*

**Keywords:** *Bayesian analysis, Bayesian statistics, mixed effect models, linear models, hierarchical models.*

### 1. Introduction

Bayesian statistics is an approach to statistics, which formally seeks use of prior information and Baye's theorem provides the basis for making use of this information in a formal manner. Bayesian statistics is an excellent alternative to be more reasonable for moderate and especially for small sample sizes when non Bayesian procedures do not work see, for example, [1]. When data are collected from many similar units such as animals, cities, and experimental locations,

---

\* Corresponding author. E-mail: [nazir.nageena@gmail.com](mailto:nazir.nageena@gmail.com)

the problem in such situations is how to combine the data generated at different units to draw inference about phenomenon under study. For example, in a plant breeding trial several genotypes are planted at different location so that locational or environmental effects could be evaluated in respect of the genotypes. This helps in selecting the best location and genotype in respect of certain quantitative and qualitative traits. Also [8] and [11] are the pioneering papers dealing with such problems in linear model framework. The basic idea behind this approach is to model the phenomenon under study in stages and analyze that model in Bayesian framework, such a model is known as hierarchical model see, for example, [3] and [4].

Henderson's mixed model or best linear unbiased prediction (BLUP) theory (see, for example [6], [7] ) is a non-Bayesian example of analyzing such situations. Deficiencies of Henderson's approach in the form of negative estimates of variance components and conceptual difference between fixed and random effects are removed by [5] using complete Bayesian framework in the context of animal breeding theory. However [10] develop a methodology in which they have used t-distribution for fixed and random effects.

In this paper a description of mixed effects models in Bayesian framework in reference to multilocational trial of potato breeding has been given. The computations have been done using lme() function which is a part of nlme library of SPLUS and R- software packages of [12].

## 2. Description of the Model

In the spirit of [8] we consider the following Bayesian model:

**Stage1:** For  $i$ th location, conditional on  $X_i, Z_i, \beta = (\beta_1, \beta_2, \dots, \beta_p)^T$  and  $b_i = (b_1, b_2, \dots, b_q)^T$  let  $y \sim N(X_i\beta + Z_i b_i, \Sigma)$ , where  $y_i$  is a  $(n \times 1)$  vector of responses and  $X_i, Z_i$ , are known design matrices relating effects  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$  and  $b_i = (b_1, b_2, \dots, b_q)^T$  to  $y_i$ , and  $\Sigma$  is a positive definite matrix.

**Stage2:** The effects  $\beta$  and  $b_i$  are independent and have multivariate normal distribution i.e.,  $\beta \sim N(\beta_0, \Sigma_0)$  and  $b_i \sim N(0, D)$ . If  $\Sigma_0^{-1}$  approaches to null matrix then the distribution of  $\beta$  can be treated as approximately uniform distribution.

A non-Bayesian formulation of the proposed model is:

$$y_i = X_i\beta + Z_i b_i + e_i, \quad i = 1, \dots, k \quad (1)$$

where  $e_i$  and  $b_i$  and mutually independent with  $e_i \sim N(0, \sigma^2 I)$  and  $b_i \sim N(0, D)$ . It may be noted that  $\beta$  which follows normal distribution with  $\Sigma_0^{-1}$  approaching to null matrix, is termed as fixed effect, whereas  $b_i$  which follows normal distribution is termed as random effect. Since both these effects are present in the model (1) hence it is termed as mixed effect model.

### 3. Bayesian analysis of mixed effect models

Consider model (1), the first stage that is likelihood can be defined as:

$$L(\beta, \theta, \sigma^2 | y) = p(y | \beta, \theta, \sigma^2) \quad (2)$$

In the second stage:

$$p(\beta, b_i | \theta, \sigma^2) \propto p(\beta | \theta, \sigma^2) p(b_i | \theta, \sigma^2) \quad (3)$$

Thus joint posterior density of  $\beta, \sigma^2, \theta$  can be written as:

$$p(\beta, \theta, \sigma^2 | y) = \prod_{i=1}^k p(y_i | \beta, \theta, \sigma^2) = \prod_{i=1}^k \int p(y_i | b_i, \beta, \sigma^2) p(b_i | \theta, \sigma^2) db_i \quad (4)$$

The expression (3) can be used directly in an optimization routine to calculate the posterior mode for  $\theta, \beta$  and  $\sigma^2$ . However, the optimization is much simpler if we first concentrate or approximate the posterior density so that it is a function of  $\theta$  only. That is, we calculate the conditional posterior mode  $\hat{\beta}(\theta)$  and  $\hat{\sigma}^2(\theta)$  as the value that maximizes  $p(\beta, \theta, \sigma^2 | y)$  for a given  $\theta$ .

It is possible to describe this density as a normal distribution with mean zero and patterned variance covariance matrix  $\Sigma_i$ , a representation that is often used to derive the posterior density for the parameters in a linear mixed effect model. That is:

$$p(\beta, \theta, \sigma^2 | y_i) = (2\pi\sigma^2)^{-n_i/2} \exp\left(\frac{(y_i - X_i\beta)^T \Sigma_i^{-1} (y_i - X_i\beta)}{-2\sigma^2}\right) |\Sigma_i|^{-1/2} \quad (5)$$

For the purpose of studying posterior mode of the approximate posterior density (4) the method of QR decomposition is used by [12], thus defining the augmented matrix in QR decomposition form:

$$\begin{bmatrix} Z_i & X_i & y_i \\ \Delta & 0 & 0 \end{bmatrix} = Q^{(i)} \begin{bmatrix} R_{11(i)} & R_{10(i)} & C_{1(i)} \\ 0 & R_{00(i)} & C_{0(i)} \end{bmatrix} \quad (6)$$

which results posterior mode of  $\hat{\beta}(\theta)$  and  $\hat{\sigma}^2(\theta)$  for given  $\theta$  as:

$$\hat{\beta}(\theta) = R_{00}^{-1} C_0 \quad \text{and} \quad \hat{\sigma}^2(\theta) = \frac{\|C_{-1}\|^2}{N} \quad (7)$$

where  $C_0$  and  $C_{-1}$  are the matrices obtained after applying another QR decomposition on the rectangular matrix of (5).

Approximate posterior density of  $\theta$  for given  $\hat{\beta}(\theta)$  and  $\hat{\sigma}^2(\theta)$  can be defined as:

$$p(\theta | y) = \left( \frac{N}{2\pi \|C_{-1}\|^2} \right)^{N/2} \exp\left( \frac{N}{2} \prod_{i=1}^k \text{abs} \left( \frac{|\Delta|}{|R_{11(i)}|} \right) \right) \quad (8)$$

or log-posterior :

$$l(\theta | y) = \log p(\theta | y) \\ = \frac{N}{2} [\log N - \log(2\pi) - 1] - N \log \|C_{-1}\| + \sum \log \text{abs} \left( \frac{|\Delta|}{|R_{11(i)}|} \right) \quad (9)$$

This approximate log-posterior density (8) is maximized with respect to  $\theta$ , to get the posterior mode  $\hat{\theta}$ . The posterior mode of  $\hat{\beta}(\theta)$  and  $\hat{\sigma}^2(\theta)$  are obtained by substituting  $\theta = \hat{\theta}$ . This estimate is termed as ML estimate by [12].

There are two methods of optimization to get posterior mode  $\hat{\theta}$ . The most common is Newton-Raphson iteration scheme and second EM algorithm. It may note that EM algorithm consists of two main steps expectation, E and maximization, M. It may be noted that Newton-Raphson is fast converging algorithm whereas EM is quite slow. However, in first few iterations of EM estimate is very close to the optimum but near optimum it takes many iterations to reach the optimum value. Also [12] have implemented a hybrid of these two in their lme() function of nlme library in which first 25 iterations are EM followed by Newton-Raphson.

Posterior density of  $\beta, \theta$  and  $\sigma^2$  can be approximated by multivariate normal densities as:

$$p(\beta | y) \cong N(\hat{\beta}, \hat{\sigma}^2 [R_{00}^{-1} R_{00}^T]) \\ p(\theta, \log \sigma | y) \cong N \left( \begin{bmatrix} \hat{\theta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_k \\ \log \hat{\sigma} \end{bmatrix}, I^{-1}(\hat{\theta}_1, \dots, \hat{\theta}_k, \hat{\sigma}) \right) \quad (10)$$

where  $I^{-1}(\hat{\theta}_1, \dots, \hat{\theta}_k, \hat{\sigma})$  is the negative Hessian matrix computed at posterior mode. These posterior densities will be used for construction of credible regions and hypothesis testing. Bayesian parallel of likelihood ratio test can also be used for hypothesis testing and for construction of credible region. Moreover, for model selection this technique is very useful provided models are hierarchical in nature, i.e., one model is particular case of the other model. [12] have implemented this strategy in their function anova().

#### 4. Bayesian analysis for fixed and random effects

To perform Bayesian analysis of fixed effect  $\beta$  an approximate posterior density will be constructed for  $\beta$  assuming that  $\hat{\theta}$  is known, i.e., posterior density of  $\beta_j$ , the  $j$ th component of  $\beta$  is approximately:

$$p(\beta_j | \hat{\theta}) \approx t_1(\hat{\beta}_j(\hat{\theta}), \hat{\sigma}_R^2 (R_{00}^{-1} R_{00}^{-T})_{jj}, df_j) \quad (11)$$

where  $\hat{\beta}_j(\hat{\theta})$  is the posterior mode of  $\beta$  for given  $\hat{\theta}$ .

This results into an approximate  $(1-\alpha)\%$  credible region of  $\beta_j$  as:

$$\hat{\beta}_j + t_{df_j}(1-\alpha/2)\hat{\sigma}_R \sqrt{[R_{00}^{-1} R_{00}^{-T}]_{jj}} \quad (12)$$

Bayesian inference about  $\sigma$  can be obtained from its posterior density defined as:

$$p(\log \sigma | \hat{\theta}) \approx N(\log \hat{\sigma}(\hat{\theta}), [I^{-1}]_{\hat{\sigma}\hat{\sigma}}) \quad (13)$$

This results into  $(1-\alpha)\%$  credible region for  $\sigma$  as:

$$\left( \hat{\sigma} \exp(-z(1-\alpha/2)\sqrt{[I^{-1}]_{\hat{\sigma}\hat{\sigma}}}), \hat{\sigma} \exp(z(1-\alpha/2)\sqrt{[I^{-1}]_{\hat{\sigma}\hat{\sigma}}}) \right) \quad (14)$$

where  $z(1-\alpha/2)$  denotes the quantile of the standard normal distribution.

#### 5. Numerical Illustration

The methodologies discussed in the earlier sections have been implemented by [12] in the nlme library which is a recommended software package in R an S-PLUS. Details of this library can be obtained from internet <http://nlme.stat.wisc.edu> or <http://cm.belllabs.com/stat/NLME>. This library contains important function lme() meant for analysis of linear mixed effects models. The details of lme() function follows:

```
lme(fixed, data, random)
```

where fixed stands for fixed effect part of the model i.e., the part of the model which is common in all the locations. Bayesian term it that part of the model which contains non-informative prior for  $\beta$ . The argument data stands for the data frame which contains the data for which model is to be fitted. The argument random stands for the random effect part  $b_i$  of the model which is commonly termed by Bayesians a part of the model with informative prior.

This function returns estimate of regression coefficient  $\beta$ ,  $b_i$  and variance components for  $b_i$  and  $e_i$ . Details of output includes model selection criteria like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and loglikelihood computed at the posterior mode along with the number of iterations required during the optimization. It may be noted that:

$$AIC = -2\log Lik + 2n_{par}$$

$$BIC = -2\log Lik + n_{par} \log(N)$$

where logLik stands for logposterior compute at its mode. The  $n_{par}$  stands for number of parameters involved in the model and N stands for the number of rows in the data matrix see, for example, [2] and [14].

The multilocational data on potato recorded at five different research station of SKUAST-k (j & k) in the year 2005-2006 is used. This analysis is helpful for plant breeders as several genotypes planted at different locations are evaluated for Locational and environmental effects in respect of genotypes. This helps in selecting the best Location in respect of certain qualitative and quantitative traits.

**Table 1. Linear Mixed Effects Model Fit by REML for the Potato data.**

AIC	BIC	logLik			
1621.529	1637.41	-805.7646			
Random effects:					
Formula: ~1   Location					
(Intercept) Residual					
StdDev: 0.001098937 21.57763					
Fixed effects: Yield ~ TbrWeight + AvTuberNo.					
	Value	Std.Error	DF	t-value	p-value
(Intercept)	-253.39373	11.394762	173	-22.23774	0
TbrWeight	4.38574	0.077889	173	56.30790	0
AvTuberNo.	48.35250	1.471885	173	32.85073	0
Correlation:					
	(Intr)	TbrWgh			
TbrWeight	-0.635				
AvTuberNo.	-0.878	0.214			
Standardized Within-Group Residuals:					
	Min	Q1	Med	Q3	Max
	-5.1149084	-0.1928404	0.0459367	0.3990114	3.6062606
Number of Observations: 180					
Number of Groups: 5					

BIC when model is fitted as simple linear model (lm fit) where locations are treated as fixed and not as random:

```
BIC(fit.lm)
[1] 1653.772
```

Since the BIC value is the model selection criterion, lesser the value of the BIC better is the model fit. Thus as can be seen for the values of BIC obtained by fitting the model by two different methods the best fit is the lme as the BIC value is less in that fit. Thus the mixed effect model is better fit as compared to the lm fit. In the above fit model we have treated both tuber weight and average tuber number as fixed effects and the locations as random whereas as in the lm fit the locations are treated as fixed resulting in the higher values of BIC thus it can be concluded from the above results that the location should be treated as random and not as fixed.

**Table 2. Random effects when the model is fitted as mixed effect model**

Locations	(Intercept)
Gulmarg	6.418289e-07
Gurez	-5.872515e-07
Larnoo	-2.294771e-07
Shalimar	-9.148833e-08
Tangmarg	2.663879e-07

The above table represents the random effects. As can be seen from the values, Gulmarg is the better location for the cultivations of potato followed by Tangmarg. This result is different from the one obtained in the linear model fit because their only the yield was taken into consideration however here in addition to the yield the variance and environmental effects are also taken into consideration. Thus this is a better result.

## 6. Conclusion

It is worth to mention that mixed model methodology lacks theoretical and philosophical grounds and Bayesian approach is the only remedy as has been discussed by [11], [5] and [13]. From the above output it can be concluded that among locations Gulmarg is the best location followed by Tangmarg and Shalimar is not the good location for the cultivation of potato among the given locations. This is in contrary to non-Bayesian approach where Gurez the best Location. This is because there the ranking is based on the averages only. However, in Bayesian approach the combined effects of location averages, overall average, variation between the locations and within the location are taken in to consideration see, for example, [9].

## References

- [1]. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, second edition*. New York: Springer-Verlag.
- [2]. Chambers, J. M and Hastie, T. J. (1992). *Statistical Models*. New York: S. Chapman and Hall.
- [3]. Gelman, A., Carlin, J. B., Stern, S. H. and Rubin, D. B. (2003). *Bayesian Data Analysis. Second edition*. New York: S. Chapman and Hall.
- [4]. Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, 6(1), 55- 93.

- [5]. Gianola, D. and Fernando, R.L. (1986). Bayesian methods in animal breeding theory. *Journal of Animal Science*, 63, 217- 244.
- [6]. Henderson, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics*, 21, 309-383.
- [7]. Henderson, C. R. (1984). *Application of Linear Models in Animal Breeding Theory*. University of Guelf, Guelf Ontario.
- [8]. Kass, R.E., Staffey, D. (1989). Approximate Bayesian inference in conditional independent hierarchical models (parametrical empirical Bayes model). *Journal of American Statistical Association*, 84, 717- 726.
- [9]. Khan, A.A., Bhat, B.A., Ahmad, N. (2000). Bayesian statistical approach for comparing yield of capsicum in Kashmir. *Applied Biological Research*, 2(1): 1-6.
- [10]. Khan, A.A., Bhat, B.A. (2002). Bayesian analysis of mixed linear models using t-distribution for fixed and random effects. *SKUAST Journal of Research*, 4, 51-56.
- [11]. Lindley, D.V and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). *Journal of Royal Statistical Society B*, 34, 1- 41.
- [12]. Pinheiro, J.C. and Bates, M.D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.
- [13]. Robinson, G.K. (1991). That BLUP is a good thing: estimation of random effects (with discussion). *Statistical Science*, 6, 15- 51.
- [14]. Venables, W.N., Ripley, B.D. (2004). *Modern Applied Statistics with S-PLUS, 4<sup>th</sup> Edition*. New York: Springer-Verlag.