Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 8, Issue, 11, pp. 21536-21546, November, 2017

ISSN: 0976-3031

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

APPLICATIONS OF BAYESIAN SIMULATION TOOLS FOR REGRESSION MODELING OF AGRICULTURAL DATA

Shagufta Yasmeen* and Athar Ali Khan

Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh-202002, India

DOI: http://dx.doi.org/10.24327/ijrsr.2017.0811.1091

ARTICLE INFO

Article History:

Received 17th August, 2017
Received in revised form 21th September, 2017
Accepted 28th October, 2017
Published online 28th November, 2017

Key Words:

Bayesian inference, Hamiltonian Monte Carlo, independent Metropolis, JAGS, Metropolis within Gibbs, R, sampling importance resampling, Stan.

ABSTRACT

A large part of applied statistical analysis is based on linear regression modeling technique which is one of the most widely used statistical tools in agriculture. These regression models are of particular interest of agriculturists for a variety of inferential tasks such as prediction, parameter estimation and data description. The theory of least squares is widely used to analyse the agricultural field experiments. In this paper an attempt has been made to implement parallel Bayesian methods of deterministic as well as simulation tools to regression models. Implementations have been made using R, JAGS and Stan packages.

INTRODUCTION

In probability theory, linear regression model is one of the most commonly used statistical models that have received greatest attention due to their allowance for quantifying uncertainty and making predictions. Indeed these models became a much more flexible instrument in agriculture and in many areas of research through Fisher's introduction of regression models. It is the geocentric model of applied statistics which is concerned with describing how the average value of a numerical outcome variable vary over to subpopulations defined by linear functions of predictors. Numerous texts have been written on the linear regression model; see, for example, Graybill (1961) and Searle (1971) for the sampling theory viewpoint. Carlin and Forbes (2004) provide an excellent introduction to the concepts of linear modeling and regression, Neter *et al.* (1996) and Weisberg (1985) provide accessible introduction to regression, Ramsey and Schafer (2001) is a good complement, with a focus on issues such as model understanding, graphical display, and experimental design. Gomez and Gomez (1984), Snedecor and Cochran (1989), and Welham *et al.* (2015) provide regression modeling to agricultural data. However, the introduction of Bayes' theorem attracts researcher towards Bayesian inference to overcome the issues encountered in classical one. For example, Lindley (1971) rejects many sampling theory techniques because they violate the likelihood principle and thus violate the axioms of utility and probability. For more about the advantages and disadvantages of Bayesian inference and other inferential theories, Barnett (1973) gives an unbiased view. The general Bayesian inferential procedures as well as those methods which are peculiar to the linear models have seen in Jeffreys (1961) with vague prior distribution and Lindley (1965) with improper vague prior distribution. For details of Bayesian linear regression, see Zellner (1971), Box and Tiao (1973), and, for a more informally Bayesian treatment, see, Broemeling (1985), Gelman and Hill (2007) and Gelman *et al.* (2014).

The aim of this paper is to cover the Bayesian framework of general linear models which includes, as special cases, the fixed models, which are used for regression analysis. In the Bayesian normal linear model framework, the key statistical modeling issues are (i) defining the predictor (x) and response (y) variable so that the conditional expectation of y is reasonably linear as a function of the columns of X (matrix of predictors) with approximately normal errors, and (ii) setting up a prior distribution on the model parameters that accurately reflects substantive knowledge (Gelman *et al.* 2014). Let β be a $(p+1) \times 1$ vector of real parameters,$y = (y_1, \dots, y_n)'$ a $n \times 1$

Corresponding author:* **Shagufta Yasmeen
Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh-202002, India

vector of observations, X a $n \times (p+1)$ known model matrix. Then the general linear model is,

$$y = X\beta + \epsilon$$

where $\epsilon \sim N(0_n, \tau^{-1} I_n)$ and τI_n is the precision matrix of ϵ, which has covariance matrix $\sigma^2 I_n (\sigma^2 = 1/\tau > 0)$. To make above model a Bayesian linear model we require a prior for parameters βandτ. The simplest approach is to assume that all parameters are a priori independent having the structure

$$p(\beta, \tau) = \prod_{j=0}^{p} p\,(\beta_j) p(\tau)$$

$$\beta_j \sim N\left(\mu_{\beta_j}, c_j^2\right) \; for \; j = 0, \dots, p \; and \;\; \tau \sim U(a, b)$$

Once prior information is represented by a probability density function, then the Bayes theorem combines this information with the information contained in the data. Thus, the joint posterior density of $\theta = [\beta, \tau]$ is

$$p(\beta, \tau | y, X) \propto p(y|\beta, \tau, X) \times p(\beta, \tau)$$

With an equality sign, the posterior density is

$$p(\beta, \tau | y, X) = K \times p(y|\beta, \tau, X) \times p(\beta, \tau)$$
$$\beta \in R, \qquad \tau > 0$$

where K is the normalizing constant and is given by

$$K^{-1} = \int_0^{\infty} \int_R p\,(y|\beta, \tau, X) \times p(\beta, \tau) d\beta d\tau$$

defines the marginal likelihood of y, or the prior predictive distribution of y. It can be used to predict an outcome given a linear function of these predictors, and regression coefficients can be thought of as comparisons across predicted values or as comparisons among averages in the data. A large part of applied statistical analysis is based on linear regression techniques that can be thought of as Bayesian posterior inference based on a weakly-informative prior distribution for the parameters of the normal linear model. In this paper, we outline, from a Bayesian perspective, the regression models for agricultural data with one and multiple predictors.

Note that Bayes' rule provides a rational method for updating our beliefs in the light of new information. It does not tell us what our beliefs should be; it tells us how they should change after seeing new information. The prior distribution is important in Bayesian inference since it influences the posterior. When no information is available, we need to specify a prior which will not influence the posterior distribution$p(\theta|y)$. Such priors are called weakly-informative or non-informative or vague priors. This type of priors will be used throughout this paper. Usually a conjugate prior distribution for an unknown parameter leads to a posterior distribution is used for which there is a simple formulae for posterior means and variances. But we have situations in which obtaining exact value from posterior quantity may be difficult or impossible; however, if we generate random sample values of the parameters from the posterior distribution then we can get exact values of that posterior quantity of interest. Cases in which conjugate priors are considered to be unrealistic or are unavailable, either asymptotic approximation such as Laplace approximation (see, for example Tierney and Kadane 1986, Tierney, Kass and Kadane 1989, Erkanli 1994) or numerical integration techniques (see for example Evans and Swartz 1996) can be used. Another appealing alternative is the usage of simulation based techniques. By simulation, we mean summarizing inferences by generating random samples from posterior distribution. Simulation techniques based on Markov chain Monte Carlo (MCMC) methods enable statisticians to use highly complicated models and estimate the corresponding posterior distributions with accuracy. In this paper main focus is given on how to implement Bayesian methods using simulations tools as an alternative to least square theory. Moreover, asymptotic analytic tools using Laplace approximation have also been implemented to cross verify the simulation results. The variants of MCMC such as independence Metropolis (proposed by Hastings, 1970 and popularized by Tierney, 1994.), Metropolis within Gibbs, and Hamiltonian Monte Carlo (Duane *et al.*, 1987 and Neal, 1994) are performed to the Bayesian linear regression analysis of agricultural data. Computational and graphical aspects of Bayesian analysis have been implemented via R2jags, LaplacesDemon and rethinking packages of R.

Simple linear regression

We begin with the simplest case of linear regression in which only one predictor is involved. For Bayesian analysis of this simple linear model, we adopt Diploid wheat data from Welham *et al.* (2015) (also in Jing *et al.*, 2007) in which all the concepts and computations have been discussed in classical viewpoint. In this dataset, several morphological traits were measured for 190 seeds selected at random from a line of diploid wheat, *Triticum monococcum*, with the aim of identifying variables associated with differences in seed weight. The variables measured are weight (mg), diameter (mm), length (mm), moisture content (%) and endosperm hardness (single-kernel characterization system index value). Seed size, as measured by length, is expected to be a major contributor to differences in seed weight, and so, we start by examining the relationship between seed weight and seed length. The header part of the data is

DSeed	Weight	Length	Diameter	Moisture	Hardness
1.	30.15	3.27	2.09	10.27	-16.63
2.	35.51	3.65	2.34	10.61	-8.27
3.	29.16	3.36	2.15	10.27	-21.45
4.	16.82	2.77	1.79	11.05	4.13
5.	23.42	2.78	1.80	10.02	-2.05
6.	31.77	3.37	2.15	10.34	-41.78

Bayesian analysis of diploid wheat data

To make regression model, a Bayesian model, specification of the prior parameters is required. The full Bayesian model for the diploid wheat data is expressed as

$$y_i \sim N(\mu_i, \sigma^2) \;\; for \;\; i = 1, \dots, 190$$
$$\mu_i = \beta_0 + \beta_1 x_i$$

In matrix notation, this can be reexpressed as

$$y \sim N(X\beta, \sigma^2 I)$$

where $y = (y_1, \dots, y_n)^T$ is a 190×1 vector of seed weights, x_i the values of explanatory variable seed length for individuals

$i = 1, \ldots, 190$, X the 190×2 model matrix, and $\beta = (\beta_0, \beta_1)^T$ is the vector of regression coefficients. The next job is to choose appropriate priors for the parameters β and τ. For this, some weakly informative priors have been taken. Prior distribution for the parameters (β_0, β_1) are taken to be independent normal distribution with zero mean which corresponds to the assumption of no effect of X on y and high variance to make prior flat (Ntzoufras, 2009). Further, for the parameterσ, we consider uniform distribution as a low-information prior distribution since it a priori gives the same probability to any interval of the same range. Notationally, prior probabilities for β and $\sigma(= \tau^{-1/2})$ are specified as

$$\beta_j \sim N(0,1000) \quad for \quad j = 0,1$$
$$\sigma \sim U(0,100)$$

By using Bayes' theorem, the joint posterior density of β and σ can be defined as

$$p(\beta, \sigma|y, X) = \frac{p(y|\beta, \sigma, X) \times \prod_{j=0}^{1} p\,(\beta_j) \times p(\sigma)}{\int\int\int p\,(y|\beta, \sigma, X) \times \prod_{j=0}^{1} p\,(\beta_j) \times p(\sigma) d\beta_0 d\beta_1 d\sigma}$$

The corresponding marginal posterior probability densities can be obtained as,

$$p(\beta_0|y, X) = \int\int p\,(y|\beta, \sigma, X) \times \prod_{j=0}^{1} p\,(\beta_j) \times p(\sigma) d\beta_1 d\sigma$$
$$p(\beta_1|y, X) = \int\int p\,(y|\beta, \sigma, X) \times \prod_{j=0}^{1} p\,(\beta_j) \times p(\sigma) d\beta_0 d\sigma$$
$$p(\sigma|y, X) = \int\int p\,(y|\beta, \sigma, X) \times \prod_{j=0}^{1} p\,(\beta_j) \times p(\sigma) d\beta_0 d\beta_1$$

These posterior densities are not in closed form, hence, one has to use either analytic approximation or simulation tools. The simplest is to use normal approximation to the posterior density, essentially a Bayesian version of the Central limit theorem (Carlin and Louis, 2009). More complicated asymptotic techniques, such as Laplace's method (Tierney and Kadane, 1986), enable more accurate, possibly asymmetric posterior approximations. However, when approximate methods are intractable or result in insufficient accuracy, we must resort to simulation techniques such as MCMC, the output of which corresponds to a sample from the joint posterior density, provide more complete information and are comparatively easy to program, even for very high dimensional models. A variant of MCMC known as independence Metropolis is implemented to approximate the joint posterior density $p(\beta, \sigma|y, X)$, through the package LaplacesDemon in R and also the Metropolis within Gibbs sampling and Hamiltonian Monte Carlo are introduced through the package R2jags and package rethinking respectively in R.

Analysis with Laplaces Demon

The implementation has been done through two main functions of the package Laplaces Demon which are Laplace Approximation and Laplaces Demon. The Laplace Approximation is used for analytic approximation whereas, for simulation Laplaces Demon is used. First, we implement the analytic tool to approximate the posterior density and use the results as starting values in the MCMC algorithm.

Creation of diploid wheat data in R

The LaplacesDemon package requires data in a listed form. For the diploid data set, the individual's observations of seed weight are entered as a vector of y. The model matrix X has two columns which is denoted by J, the vector of 1's has been inserted for intercept and second column is for regress or variable seed length. The wheat is a data frame which contains all the information of data and is created from the data file TRITICUM.DAT from the homepage of the book.

```
wheat<-read.table("TRITICUM", header=TRUE)
y<-wheat$Weight
x<-wheat$Length
X<-cbind(1,x)
J<-ncol(X)
mon.names<-c("LP")
parm.names<-as.parm.names(list(beta=rep(0,J),sigma=0))
pos.beta<-grep("beta",parm.names)
pos.sigma<-grep("sigma",parm.names)
MyData<-
list(X=X,mon.names=mon.names,parm.names=parm.names,po
s.beta=pos.beta,pos.sigma=pos.sigma,y=y)
```

The three parameters beta[1], beta[2], and sigma are organized in the vector parm.names with the function as.parm.names. The function mon.names is used to monitor the log posterior (LP). In the end, all the data variables are combined in a listed form which is assigned to an object and named it MyData.

Model specification

To continue with the Bayesian analysis of the diploid wheat data, the logarithm of the unnormalized joint posterior distribution is used, which is the sum of the log-likelihood and prior distributions.

$$log\ p(\beta, \sigma|y, X) \propto log\ p(y|\beta, \sigma, X) + \sum_{j=0}^{1} l\ ogp(\beta_j) + log\ p(\sigma)$$

```
Model <- function(parm, Data)
# Parameters
beta <- parm[Data$pos.beta]
sigma <-interval(parm[Data$pos.sigma], 1e-100, Inf)
parm[Data$pos.sigma] <- sigma
### Log-Prior
beta.prior<-dnormv(beta,0,1000,
log=TRUE)
sigma.prior<-dunif(sigma,0,100,
log=TRUE)
# Log-Likelihood
mu <- tcrossprod(beta, Data$X)
LL <- sum(dnorm(Data$y, mu, sigma,
log=TRUE))
# Log-Posterior
LP <- LL + sum(beta.prior)+sigma.prior
Modelout <- list(LP=LP, Dev=-2*LL,
Monitor=LP,yhat=rnorm(length(mu), mu,
sigma), parm=parm)
return (Modelout)}
```

The Model function is evaluated and the logarithm of the unnormalized joint posterior density is calculated as LP, and returned in a list called Model out, along with the deviance(Dev), a vector (Monitor) of any variables desired to be monitored in addition to the parameters, y^{rep} (yhat) or replicates of y, and the parameter vector parm. The σ parameter must be positive-only, and so it is constrained to be positive in the interval function. The algorithm, outside of the Model function needs to be aware that σ has been constrained, so the parm vector is updated with the constrained value.

Initial Values

The functions Laplace Approximation and Laplaces Demon in Laplaces Demon package require a vector of initial values for the parameters. Each initial value is a starting point for the estimation of a parameter. The order of the elements of the vector of initial values must match the order of the parameters associated with each element of parm passed to the Model function. In this case, parameter beta is given to value zero and parameter scale equal to 1.

```
Initial.Values <- c(rep(0,J), 1)
```

Asymptotic approximation

The function Laplace Approximation is used to approximate integrals, which is a family of asymptotic techniques. It deterministically maximizes the logarithm of the unnormalized joint posterior density with one of several optimization algorithms. Here, trust region (TR) algorithm of Nocedal and Wright (1999) is used due to its efficiency than other algorithms as it attempts to reach its objective in the fewest number of iterations. In TR algorithm, the Hessian is approximated each iteration, making it best suited to models with small to medium dimensions.

```
fit.TR <- Laplace          Approximation (Model=Model, Initial.
Values, MyData, Method=  "TR", Iterations=500)
```

The two posterior summaries obtained by fit.TR object are reported in Table 1 and Table 2.

Table 1 Marginal posterior densities summaries of the parameters using the function Laplace Approximation.

Parameter	Mode	SD	LB	UB
beta[1]	-27.77	2.19	-32.15	-23.39
beta[2]	17.13	0.66	15.80	18.45
sigma	2.91	0.15	2.61	3.21

The units of the intercept and slope here are mg and mg/mm, respectively, and an increase of 1 mm in seed length is expected to produce an increase of 17.13 mg in seed weight. The intercept represents the estimated average weight for seeds of length zero (i.e. -27.77 mg). Biologically, this is a startling value for two reasons: we clearly cannot have negative seed weights, and we expect a seed with zero length to have zero weight.

Table 2 Summaries of the posterior samples drawn with sampling importance resampling (SIR), given the point-estimated posterior modes and the covariance matrix with the bounds that constitute a 95% probability interval.

Parameter	Mode	SD	LB	Median	UB
beta[1]	-27.61	2.19	-31.73	-27.67	-23.08
beta[2]	17.08	0.67	15.70	17.09	18.32
sigma	2.94	0.15	2.66	2.94	3.26
Deviance	948.47	2.48	945.62	947.95	954.73
LP	-488.11	1.24	-491.29	-487.83	-486.67

This means we need to check that the model is appropriate for the data, but it does not necessarily mean that the model is inappropriate. Both intercept and slope are statistically significant.

MCMC simulation

To initialize markov chains, the above asymptotic posterior summaries are used as the starting values. The function LaplacesDemon is used for this purpose which maximizes the logarithm of the unnormalized joint posterior density with MCMC and provides samples of the marginal posterior distributions, deviance, and other monitored variables. Independence Metropolis (IM) proposed by Tierney (1994) is used in which the proposal distribution does not depend on the previous state.

```
fit.Demon<-
LaplacesDemon(Model, Data =MyData, Initial.Values,Covar=
fit.TR$Covar,Iterations =5000,
Status=100,Thinning=10,Algorithm="IM",Specs=list(mu=fit.T
R$Summary1[1:length(Initial.Values),1]))
```

The two summary matrices of the marginal posterior distributions of parameters can be obtained, one calculated over all the samples and the other calculated only on the stationary samples. Here, we report only the posterior summaries calculated on the stationary samples.

Table 3 Simulated marginal posterior summary obtained by independence Metropolis algorithm over stationary samples.

Parameter	Mean	SD	LB	Median	UB
beta[1]	-27.80	1.26	-30.41	-27.79	-25.47
beta[2]	17.13	0.38	16.45	17.12	17.92
sigma	2.91	0.09	2.74	2.91	3.10
Deviance	946.40	0.87	945.45	946.19	948.65
LP	-487.09	0.44	-488.24	-486.99	-486.62

It can be seen from Table 3 that the posterior summaries based on simulation with IM come out with lower standard deviation as compared to that based on Laplace approximation. This is because of two reasons. Firstly, the simulation technique summarizes posterior on the basis of samples directly drawn from it, whereas, in Laplace's method, it is approximated asymptotically and thus, does not capture the true picture of the posterior density. Secondly, with independence-Metropolis algorithm, posterior is summarized more precisely, when the proposal is a good approximation of the true posterior (Ntzoufras, 2009).

Analysis with JAGS

JAGS (Plummer, 2003) is designed for inference on Bayesian models using Markov chain Monte Carlo (MCMC) simulation. JAGS can evaluate the integrals of full conditional distribution of the parameters by using Metropolis within Gibbs algorithm which is simply a component wise Metropolis-Hastings algorithm in which some components of the parameter vector are directly generated from the corresponding full conditional posterior distribution (Ntzoufras, 2009). To make the posterior analysis comparatively easier, the package R2jags (Su and Yajima, 2015) have been used in which data is created in R, simulation is done in JAGS and finally result is reported in R.

Setting up diploid wheat data in R

The vector y is the individual observations of seed weight of length n. The model matrix with two columns, one is for intercept and the other is for regressor seed length, is denoted by X. The object j.dat assembles data in a form of list.

```
y<-wheat$Weight
n<-length(y)
x<-wheat$Length
X<-cbind(1,x)
J<-ncol(X)
j.dat<-list(y=y,n=n,X=X,J=J)
```

Model definition

The model definition consists of a series of relations inside a block delimited by curly brackets and preceded by the keyword model.

```
cat("model{
for(i in 1:n){
y[i]~dnorm(mu[i],tau)
mu[i]<-inprod(X[i,],beta[])
}
for(j in 1:J){
beta[j]~ dnorm(0.0,1.0E-04)}
tau<-1/(sigma*sigma)
sigma~dunif(0,10)
}",file="wheat.txt")
```

The first three lines preceded by the keyword model define the data level model. The function inprod is used to compute the inner product of matrix X and vector beta. We assign a weakly-informative normal prior for thecoefficients β with mean 0 and standard deviation 100. This states, roughly, that we expect these coefficients to be in the range (-100,100) and if the estimates are in this range, the prior distribution providing very little information in the inference. In JAGS, the normal distribution is specified in terms of precision (inverse of variance) parameter ($\tau = 1/\sigma^2$) rather than the usual variance parameter.

Initial values and parameters

We supply the initial values (using random numbers) for the parameter σ. When initial values are not specified, JAGS generates them itself, however, BUGS often crashes when using its self-generated initial values (Lunn *et al*. 2013). The object j.ini is used for initial values and desired parameters are saved to object j.params.

```
j.ini<-function(){list(sigma=
runif(1,0,10))}
j.params<-c("beta","sigma")
```

Calling JAGS from R

After setting up all the codes, function jags is used to run the model. Gelman and Hill (2007) assess convergence by checking whether the distributions of the different simulated chains mix; thus at least two chains must be simulated. We simulate three chains for checking the convergence. Thus, jagsran three sequences, each with 2000 iterations, with the first 1000 from each sequence discarded.

```
model.jags<-
jags(data=j.dat,inits=j.ini,j.params,model.file="wheat.txt",
n.chains=3, n.iter=2000, progress.bar=NULL)
```

Table 4 Posterior parameter estimates for jags model of diploid wheat data.

Parameter	mean	sd	2.5%	50%	97.5%	Rhat	n.eff
beta[1]	-27.95	2.33	-32.42	-28.00	-23.36	1.001	3000
beta[2]	17.18	0.70	15.81	17.19	18.54	1.001	3000
sigma	2.96	0.15	2.67	2.95	3.26	1.001	2000
Deviance	948.48	2.52	945.59	947.83	954.80	1.001	3000

The first five columns of Table 4 give inferences for the model parameters. Column first named as mean denotes posterior mean of the parameters and column second named as sd denotes their respective posterior standard deviations. The intercept beta[1] has a mean estimate of -27.95 and a standard error of 2.33. The median estimate of beta[1] is -28, with a 95% interval of [-32.42, -23.36]. Moving to the bottom of the table, the 95% interval for the seed weight coefficient, beta[2], is [15.81,18.54]. Both intercept and coefficient of seed height are statistically significant as zero doesn't lie in their 95% credible regions. The second right most column (Rhat) gives informationabout the convergence of the algorithm. The Gelman-Rubin diagnostic test (Rhat) should be less than 1.1 for all parameters to have approximately converged algorithm. The final column, n.eff is the effective sample size of the simulations.

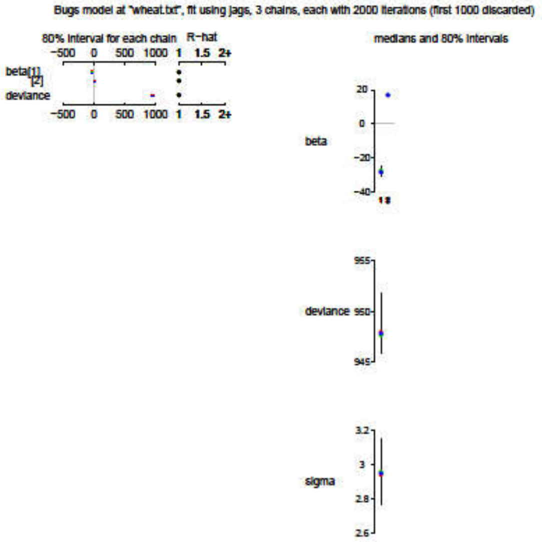

Figure 1 Graphical representation of posterior summaries. Rhat is near one for all parameters indicating good convergence, and right side shows the posterior inference for each parameter and the deviance.

Analysis with Stan

Stan is similar to BUGS: a program that draws random samples from the joint posterior distribution of the model parameters given a model, the data, prior distributions, and initial values. To do so, it uses the no-U-turn sampler, which is a type of Hamiltonian Monte Carlo(HMC) simulation (Hoffman and Gelman, 2013; Betancourt, 2013), and optimization-based point estimation. Stan can handle large data sets and complex models in less computing time than BUGS. Specifically, the number of effective samples compared to the total number of iterations is substantially higher in Stan compared to BUGS, since Stan uses more efficient MCMC algorithms and is implemented using a language focusing on efficiency and stability. The rethinking (McElreath, 2015) package provides a convenient interface, map2stan, to compile lists of formulas, into Stan HMC code. To see how it's done, let's revisit the diploid wheat data.

Model definition

```
model.stan<- map2stan(
alist(
Weight~ dnorm( mu , sigma) ,
mu <- a + b*Length ,
a ~ dnorm( 0, 100 ) ,
b ~ dnorm( 0 , 10 ) ,
sigma ~ dunif( 0 , 50 )
) ,data=wheat)
```

After executing this code, map2stan returns an object that contains a bunch of summary information , as well as samples from the posterior distribution of all parameters.

Table 5 Posterior parameter estimates for Stan model of diploid wheat data.

Parameters	Mean	StdDev	lower 0.95	upper 0.95	n_eff	Rhat
a	-27.75	2.08	-31.63	-23.38	286	1
b	17.12	0.63	15.89	18.37	279	1
sigma	2.94	0.15	2.64	3.21	476	1

The estimates in Table 5 are very similar to the estimates obtained through LaplacesDemon and JAGS. The interval boundaries are highest posterior density interval (HPDI).

Figure 2 shows all the simulated values from the joint posterior distribution of the three model parameters. It clearly shows that the intercept (a) and regression coefficient for seed length (b) are most perfectly negatively correlated. It justmeans that these two parameters carry the same information-as we change the slope of the line, the best intercept changes to match it. But in more complex models, strong correlations like this can make it difficult to fit the model to the data. It is possible to avoid it with center the predictors (McElreath, 2015). Centering is the procedure of subtracting the mean of a variable from each value. To create a centered version of the Length variable:

```
wheat$cLength <- (wheat$Length-mean(
wheat$Length))
```

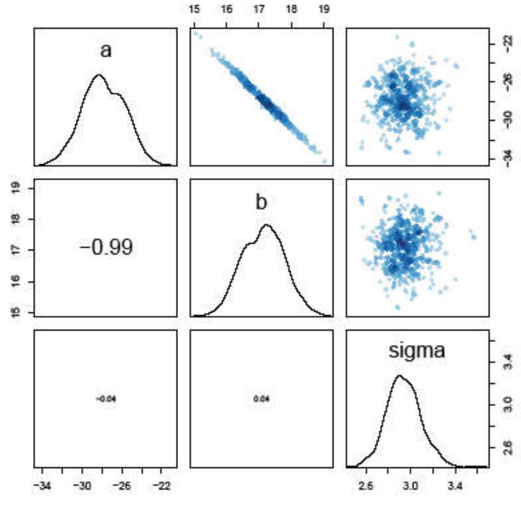

Figure 2 Pairs plot of the samples produced by Stan. The diagonal shows a density for each parameter. Below the diagonal, correlations between parameters are shown.

The refitted model is

```
model.stan1<- map2stan(
alist(
Weight~ dnorm( mu, sigma ),
mu <- a + b*cLength,
a ~ dnorm( 0, 50 ),
b ~ dnorm( 0 , 10 ),
sigma ~ dunif( 0 , 10)
),data=wheat)
```

The above code just replaces Length with cLength, the new variable. Now for the new estimates:

Table 6 Posterior parameter estimates for Stan model of diploid wheat data after centering the predictor Length.

Parameters	Mean	StdDev	lower 0.95	upper 0.95	n_eff	Rhat
a	28.66	0.22	28.23	29.11	853	1
b	17.12	0.66	15.81	18.34	886	1
sigma	2.95	0.16	2.68	3.30	943	1

The estimates for β_1 (b) and σ are unchanged, but the estimate for β_0 (a) is now the same as the average weight value in the data. And the correlations among parameters are almost zero (see, Figure 3). The estimate for the intercept, a, still means the same thing it did before: the expected value of the outcome variable, when the predictor variable is equal to zero. But now the mean value of the predictor is also zero. So the intercept also means: the expected value of the outcome, when the predictor is at its average value. This makes interpreting the intercept a lot easier.

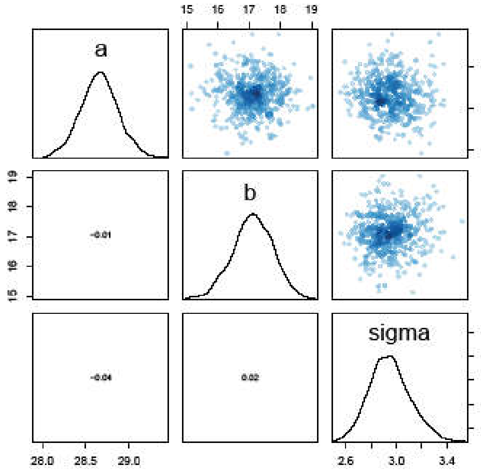

Figure 3 Pairs plot of the samples after centering the predictor variable. It is evident that the correlations among parameters are almost zero.

Figure 4 shows the 95% prediction intervals for actual seed weights. It encounters both uncertainty in parameter values and uncertainty in sampling process as it is known that prediction in a Bayesian context takes the parameter uncertainty directly into account, contrary to classical inference which does not take into account the sampling variability of $\hat{\theta}$.

Figure 4.95% prediction interval for seed weights. The solid line is the posterior estimate of the mean weight at each length. The two shaded regions show different 95% possible regions. The narrow shaded interval around the line is the distribution of μ. The wider shaded region represents the region within which the model expects to find 95% of actual seed weights in the population, at each seed length.

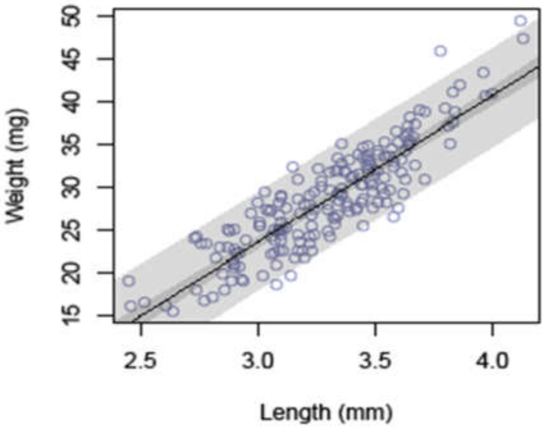

Figure 4

Multiple linear regression

Multiple regression model describes the relationship between a single response variable and two or more variates. Regression coefficients are more complicated to interpret with multiple predictors because the interpretation for any given coefficient is, in part, contingent on the other variables in the model. The implementation of multiple regression or multivariate regression (McElreath, 2015) with LaplacesDemon, JAGS, and Stan has been done in a similar way as simple regression. We just add all the predictor variables in the formula notation. To show how asymptotic approximation and simulation based Bayesian study works for multivariate regression, again diploid wheat data is used with all the predictors. As suggested by McElreath (2015), all predictors are centered and scaled. Centering typically improves the interpretation of main effects in the presence of interactions, and dividing by the standard deviation puts all predictors on a common scale.

Analysis with Laplaces Demon

The function Laplace Approximation of package Laplaces Demon is used for implementing the Laplace's method. Again TR algorithm is used due to its efficiency and fast convergence.

Creation of data

Data creation is as usual as in simple linear regression, only the difference is in the creation of model matrix (X). Since all the predictors are involved so that X is now a 190×5 model matrix. The model matrix with all the centered and scaled predictors are specified as

```
x1<-(wheat$Length-mean(wheat$Length))/sd(wheat$Length)
x2<-(wheat$Diameter-
mean(wheat$Diameter))/sd(wheat$Diameter)
x3<-(wheat$Moisture-
mean(wheat$Moisture))/sd(wheat$Moisture)
x4<-(wheat$Hardness-
mean(wheat$Hardness))/sd(wheat$Hardness)
X<-cbind(1,x1,x2,x3,x4)
```

Model specification

To specify a model, let us consider a linear regression model, which is often denoted as:

$$y \sim N(\mu, \sigma^2 I)$$
$$\mu = X\beta$$

The dependent variable, y, is normally distributed according to expectation vector μ and scalar variance $\sigma^2 I$, and the expectation vector μ is equal to the inner product of model matrix X and transposed parameter vector β. Prior probablities are specified for β and σ as

$$\beta_j \sim \mathcal{N}(0,1000), \qquad j = 1, \ldots, J$$

$$\sigma \sim \mathcal{HC}(25)$$

Each of the $J\beta$ paramters is assigned to a vague prior probability distribution with zero mean and large variance which indicates a lot of uncertainty about each β. The residual standard deviation σ is half-Cauchy with scale=25 as a weakly informative prior distribution which is the recommendation of Polson and Scott (2012). To specify the model, a function Model is created. All the codes would remain same as in simple regression model except for the prior density of sigma. Here, half-Cauchy with scale=25 is used rather than the uniform distribution.

```
sigma.prior <- dhalfcauchy(sigma, 25, log=TRUE)
```

Asymptotic approximation

To approximate the posterior densities with the optimization technique TR, some initial values are specified for parameters β and σ which is the requirement of LaplacesDemon package and then model is fitted with function LaplaceApproximation.

```
Initial. Values <- c(rep(0,J), 1)
fit.TR2 <-
LaplaceApproximation (Model=Model, Initial.Values, MyData,
Method="TR", Iterations=500)
```

Table 7 Marginal posterior densities summaries of the parameters using the function Laplace Approximation

Parameter	Mode	SD	LB	UB
beta[1]	28.66	0.20	28.26	29.05
beta[2]	-12.20	4.49	-21.18	-3.21
beta[3]	17.61	4.49	8.62	26.59
beta[4]	-0.39	0.20	-0.79	0.01
beta[5]	-0.66	0.20	-1.06	-0.26
sigma	2.72	0.14	2.44	3.00

Table 8 Summmaries of the posterior samples drawn with sampling importance resampling.

Parameter	Mode	SD	LB	Median	UB
beta[1]	28.65	0.21	28.21	28.65	29.03
beta[2]	-12.18	4.85	-22.00	-12.22	-2.93
beta[3]	17.58	4.86	8.63	17.52	27.50
beta[4]	-0.39	0.21	-0.78	-0.39	0.03
beta[5]	-0.66	0.21	-1.07	-0.66	-0.24
sigma	2.78	0.16	2.49	2.77	3.09
Deviance	925.57	3.76	920.38	924.96	934.70
LP	-489.00	1.90	-493.52	-488.65	-486.44

The two posterior summaries obtained from Laplace Approximation are reported in Table 7 and Table 8.

MCMC simulation with independence Metropolis

The function Laplaces Demon with independence Metropolis is used for simulation.

```
fit2.demon <-
 LaplacesDemon (Model, Data=MyData, Initial.Values,Covar=
fit.TR2$Covar, Iterations=5000, Status=100, Thinning=10,Alg
orithm="IM",
Specs=list(mu=fit.TR2$Summary1[1:
length(Initial.Values),1]))
```

Table 9 Simulated marginal posterior distributions of the parameters, deviance, and monitored variables over stationary samples. The closeness of posterior mean and posterior median of the parameters exhibits the symmetry in their posterior densities.

Parameter	Mean	SD	LB	Median	UB
beta[1]	28.65	0.12	28.42	28.65	28.87
beta[2]	-12.27	2.81	-17.32	-12.39	-6.88
beta[3]	17.66	2.81	12.23	17.76	22.77
beta[4]	-0.39	0.12	-0.62	-0.39	-0.16
beta[5]	-0.67	0.11	-0.89	-0.66	-0.44
sigma	2.73	0.08	2.58	2.72	2.87
Deviance	921.22	1.25	919.42	920.99	924.11
LP	-486.81	0.63	-488.29	-486.67	-485.92

When there are many explanatory variates, the selection of important one are highly preferred to agriculturists. Moreover, it is one of the main aim of regression analysis that provide a good description of the response. The Importance function of LaplacesDemon package is used for this purpose which considers variable importance (or predictor importance) to be the effect that the variable has on replicates y^{rep} when the variable is removed from the model by setting it equal to zero. Here, variable importance is considered in terms of the comparison of posterior predictive checks. This may be considered to be a form of sensitivity analysis, and can be useful for model revision, variable selection, and model interpretation.

```
Importance(fit2.demon, Model, MyData, Discrep="Chi-
Square", CPUs=1)
```

Table 10 Variable importance as the impact of each variable in design matrix $\boldsymbol{X}$ on $\boldsymbol{y^{rep}}$, when the variable is removed**.**

Model	BPIC
Full	922.784
Without Intercept	1705593.261
Without Length	3147891.369
Without Diameter	6565398.532
Without Moisture	926.114
Without Hardness	934.419

The results from Table 10 show the impact of sequentially removing each predictor. The criterion for variable importance is the Bayesian Predictive Information Criterion (BPIC), introduced by Ando (2007). BPIC is a variation of the Deviance Information Criterion (DIC) that has been modified for predictive distributions. With BPIC, variable importance has a positive relationship, such that larger values indicate a more important variable, because removing that variable resulted in a worse fit to the data. The best model has the lowest BPIC. In this way, it is evident that Length and Diameter are the most important variables for seed weight.

Analysis with JAGS

Bayesian analysis for diploid wheat data with multiple predictors has been done via package R2jags, interface of R and JAGS, which simulates the samples from posterior densities and approximate the results using Metropolis-within-Gibbs algorithm.

Creation of data in R

Here, each regressor is centered and scaled as per the recommendation of McElreath (2015). The model matrix X contains the column of one for intercept and and columns of all modified explanatory variables. The object jdat assembles all the data in a listed form.

```
y<-wheat$Weight
x1<-(wheat$Length-mean(wheat$Length))/sd(wheat$Length)
x2<-(wheat$Diameter-
mean(wheat$Diameter))/sd(wheat$Diameter)
x3<-(wheat$Moisture-
mean(wheat$Moisture))/sd(wheat$Moisture)
x4<-(wheat$Hardness-
mean(wheat$Hardness))/sd(wheat$Hardness)
X<-cbind(1,x1,x2,x3,x4)
J<-ncol(X); n<-length(y)
jdat<-list(y=y, J=J, n=n, X=X)
```

Model definition

A JAGS model is defined in a text file using a dialect of the BUGS language (Lunn *et al*., 2012). The model is same as defined in SLR model only the difference is in the specification of weakly informative prior distribution for scale parameter (sigma). Here, half-Cauchy (scale=25) prior is used instead of uniform distribution. When degrees of freedom, $\nu = 1$ of the half-t distribution, the density is proportional to a proper half-Cauchy distribution.

```
cat("model{
for(i in 1:n){
y[i]~dnorm(mu[i],tau)
mu[i]<-inprod(X[i,],beta[])}
for(j in 1:J){
beta[j]~ dnorm(0.0,1.0E-04)}
tau<-1/(sigma*sigma)
sigma~ dt(0, 25, 1) T(0,)
}",file="wheat.txt")
```

Initial values and parameters

Before a model can be run, it must be initialized. The user may supply initial value files, one for each chain, containing initial values for the model parameters. Initial values may not be supplied for logical or constant nodes. The object j.ini is used for initial values and j.params is used for monitored variables.

```
j.ini<-function(){list(sigma=runif(
1,0,10))}
j.params<-c("beta","sigma")
```

Running JAGS from R

The R2jags package offers a single interface to JAGS that carries out all the steps of running the model, with reasonable default values. The interface function, jags, is used to perform the simulations in JAGS and finally the results are assigned to an R object model.jags.

```
model.jags<-
 jags (data=jdat, inits=j.ini, j.params, model.file="wheat.txt",n.i
ter=3000, progress.bar=NULL)
```

The summaries of the posterior distribution are reported in Table 11 with the function print.

Table 11 Posterior parameter estimates for jags model of diploid wheat data with three chains, each with 3000 iterations (first 1500 discarded).

Parameter	mean	sd	2.5%	50%	97.5%	Rhat	n.eff
beta[1]	28.65	0.20	28.25	28.65	29.05	1.001	4500
beta[2]	-12.73	4.65	-21.93	-12.73	-3.77	1.002	1700
beta[3]	18.14	4.65	9.14	18.16	27.36	1.002	1900
beta[4]	-0.39	0.20	-0.78	-0.39	0.01	1.001	4500
beta[5]	-0.66	0.20	-1.06	-0.66	-0.24	1.001	4400
sigma	2.76	0.15	2.48	2.75	3.07	1.003	3600
deviance	925.16	3.56	920.29	924.44	933.58	1.002	970

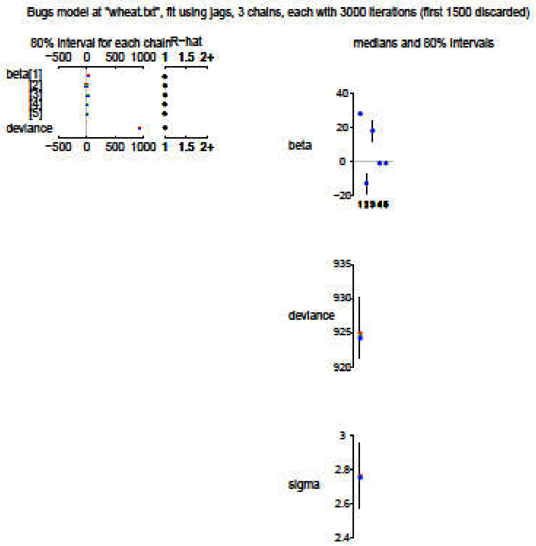

Figure 5

Figure 5. Graphical representation of posterior summaries. The upper plot displayed in the right panel shows the significance, medians, and 80% intervals of all the parameters. Rhat is near one for all parameters indicating good convergence, and right side shows the posterior inference for each parameter and the deviance.

Analysis with Stan

To analyze diploid wheat data with Hamiltonian Monte Carlo algorithm, again map2stan is used. Here, Ls, Ds, Ms, and Hs are defined as same as x1, x2, x3, and x4 defined in JAGS and LaplacesDemon respectively. All the coefficients of centered and scaled predictors are assigned to a weakly informative normal prior and a half-Cauchy prior is used for standard deviation (sigma).

```
model.stan<- map2stan(
alist(
 Weight ~ dnorm( mu , sigma ),
 mu <-
 Intercept + beta1*Ls + beta2*Ds +beta3*Ms + beta4*Hs,
  Intercept ~ dnorm(0,50),
  beta1 ~ dnorm(0,10),
  beta2 ~ dnorm(0,50),
  beta3 ~  dnorm(0,1),
  beta4 ~  dnorm(0,1),
  sigma ~ dcauchy(0,2)
),data=wheat,iter=4000)
```

Table 12 Posterior parameter estimates for Stan model of diploid wheat data with all the centered and scaled regressors.

Parameters	Mean	StdDev	lower 0.95	upper 0.95	n_eff	Rhat
Intercept	28.66	0.20	28.27	29.04	2000	1
beta1	-10.55	4.26	-8.63	-1.83	955	1
beta2	15.97	4.26	7.34	24.09	954	1
beta3	-0.37	0.19	-0.74	0.02	2000	1
beta4	-0.63	0.20	-1.03	-0.26	1871	1
sigma	2.76	0.14	2.50	3.05	1560	1

Figure 6 is the counterfactual plot which shows how the predictions change as we change only predictor at a time. This means holding the values of all predictors constant, except for a single predictor of interest.

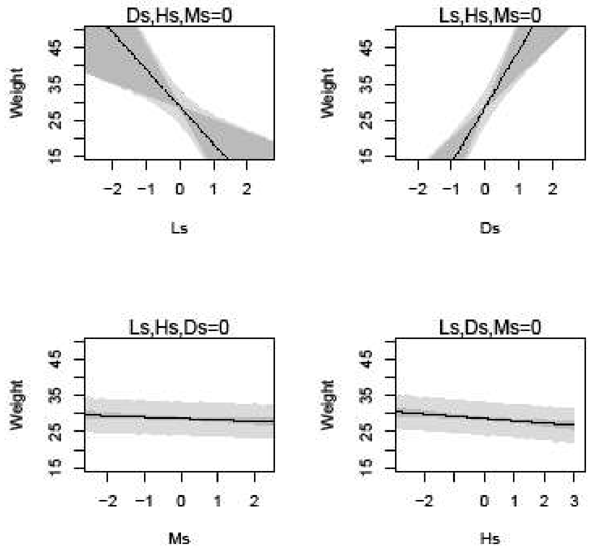

Figure 6

Figure 6. Counterfactual plots for the multivariate seed weight model,model.stan. Each plot shows the change in predicted mean across values of a single predictor, holding the other predictors constant as its mean value (zero in all cases). Shaded regions show 95% percentile intervals of the mean (dark, narrow) and 95% prediction intervals (light, wide)

CONCLUSION

Analysis of linear regression model with one and more predictors under the framework of Bayesian inference has been studied. For the Bayesian implementation, asymptotic technique such as Laplace approximation and simulation with sampling importance resampling, independent Metropolis,

Metropolis within Gibbs sampling and Hamiltonian Monte Carlo have been performed. It is concluded that the independence Metropolis implemented in LaplacesDemon function provides the lowest standard deviations as compared to other simulation techniques. The simulation results yielded by Laplace Approximation with SIR and R2jags with MWG have come out to be much closer for the parameters. We have also described the strategies for variable selection in terms of multiple linear regression model consists of a set of explanatory variates and concluded that Length and Diameter are the most important regressors for the seed weight variable.

References

1. Barnett, V. (1973). *Comparative statistical inference*, New York: Wiley
2. Betancourt, M. J. (2013). Generalizing the no-U-turn sampler to Riemannian manifolds. http://arxiv.org/abs/1304.1920
3. Betancourt, M. J. and Girolami, M. (2013). Hamiltonian Monte Carlo for hierarchical models. arXiv:1312.0906.
4. Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics.
5. Broemeling, L. D., (1985). *Bayesian analysis of linear models,* New York: Dekker.
6. Carlin, B., P., and Louis, T., A. (2009) *Bayesian Methods for Data Analysis* (3rd ed.), New York: Chapman and Hall.
7. Carlin, J. B., and Forbes, A. (2004). *Linear Models and Regression*. Melbourne: Biostatistics Collaboration of Australia.
8. Davis, P. and Rabinowitz, P. (1975). *Methods of numerical integration*. Academic Press, Waltham, MA.
9. Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* 195, 216-222.
10. Erkanli, A. (1994). Laplace approximations for posterior expectation when the model occurs at the boundary of the parameter space. *Journal of the American Statistical Association*, 89, 205-258
11. Evans, M., and Swartz, T. (1996). Discussion of methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science*, 11, 54-64
12. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd ed.), Chapman and Hall/CRC, New York.
13. Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, New York: Cambridge University Press.
14. Gomez, K. A., and Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research* (2nd ed.). John Wiley & Sons.
15. Graybill, F.A. (1961). *An Introduction to Linear Statistical Models.* Vl. I, McGraw Hill, Ney York
16. Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97-109.
17. Hoff, P. D. (2010). *A First Course in Bayesian Statistical Methods.* University of Washington, USA 98195-4322
18. Hoffman, M., and Gelman, A. (2013). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research.*
19. Jeffreys, H. (1961). *Theory of Pmbability.* Third edition. Oxford: University Press. Third edition
20. Jing, H.-C., Kornyukhin, D., Kanyuka, K., Orford, S., Zlatska, A., Mitrofanova, O.P., Koebner, R. and Hammond-Kosack, K. (2007). Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in *Triticum monococcum*. *Journal of Experimental Botany*, 58, 3749â€“3764.
21. Korner-Nievergelt, F., Roth, T., Felten, S., Guelat, J., Almasi, B., Korner-Nievergelt, P. (2015). *Bayesian Data Analysis in Ecology using R, BUGS and Stan*. Elsevier, New York.
22. Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian View point*, twovolumes. Cambridge University Press.
23. Lindley, D. V. (1971). Bayesian Statistics, a Review, *Regional Conference Series in Applied Malhematics, S.f.A.M.*
24. Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press.
25. McElreath, R. (2015). Statistical Rethinking: *A Bayesian Coursewith Examples in R and Stan.* Chapman and Hall/CRC
26. Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics* 111, 194–203.
27. Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied Linear Statistical Models,* fourth edition. Burr Ridge, Ill.: Richard D. Irwin, Inc.
28. Nocedal, J. and Wright, S. (1999). *Numerical Optimization.* Springer-Verlag, New York.
29. Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS.* John Wiley & Sons, New York.
30. Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, 20-22.
31. R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austri (2011), ISBN 3-900051-07-0, http://www.R-project.org.
32. Ramsey, F. L., and Schafer, D. W. (2001). *The Statistical Sleuth*, second edition. Pacific Grove, Calif.: Duxbury.
33. Searle, S. R. (1971). *Linear Models*. New York: John Wiley

34. Snedecor, G. W., and Cochran, W. G. (1989). *Statistical Methods*, eighth edition. Ames: Iowa State University Press.
35. Statisticat LLC (2013). LaplacesDemon: Complete environment for Bayesian inference. R package version 13.09.01, URL http://www.bayesian-inference.com/software
36. Su, Y.-S., and Yajima, M. (2015). Using R to Run 'JAGS', Version 0.5-7
37. Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701{1762. With discussion and a rejoinder by the author.
38. Tierney, L., and Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81, 82-86
39. Tierney, L., Kass, R., and Kadane, J. (1989), Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84, 710-716
40. Weisberg, S. (1985). *Applied Linear Regression*, second edition. New York: Wiley.
41. Welham, S. J., Gezan, S.A., Clark, S. J., and Mead, A. (2015). *Statistical methods in biology: Design and Analysis of Experiments and Regression*. Boca Raton, FL: Chapman and Hall/CRC Press.
42. Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

How to cite this article:

Shagufta Yasmeen and Athar Ali Khan.2017, Applications of Bayesian Simulation Tools for Regression Modeling of Agricultural Data. *Int J Recent Sci Res.* 8(11), pp. 21536-21546.
DOI: http://dx.doi.org/10.24327/ijrsr.2017.0811.1091

****