Bayesian Analysis of Experimental Data

Sanjay Kumar Prasad
Indian Agricultural Statistics Research Institute
Library Avenue, PUSA, New Delhi
skpiasri@gmail.com

Lalmohan Bhar
Indian Agricultural Statistics Research Institute
Library Avenue, PUSA, New Delhi
lmbhar@gmail.com

Abstract

Analysis of experimental data from Bayesian point of view has been considered. Appropriate methodology has been developed for application into designed experiments. Normal-Gamma distribution has been considered for a prior distribution. Developed methodology has been applied to real experimental data taken from long term fertilizer experiments.

Keywords: Bayesian analysis, Normal-gamma prior, Posterior distribution, Experimental design, Long term fertilizer experiments.

1. Introduction

Bayesian inference is an approach to statistics in which all forms of uncertainty are expressed in terms of probability. A Bayesian approach to a problem starts with the formulation of a model that is adequate to describe the situation of interest. As opposed to the point estimators used by classical statistics, Bayesian statistics is concerned with generating the posterior distribution of the unknown parameters given both the data and appropriate prior density for the parameters. As such, Bayesian statistics provides a much more complete picture of the uncertainty in the estimation of the unknown parameters, especially after the confounding effects of nuisance parameters are removed. Bayesian statistics has the advantage, in comparison to traditional statistics, of being easily established and derived. Intuitively, methods become apparent which in traditional statistics give the impression of arbitrary computational rules. For more details one may refer to Savage (1972), Raiffa and Schlaifer (1961), DeGroot (1970) and Berger (1985).

Bayesian ideas have also been introduced in recent literature on design of experiments. Bayesian design is an exciting and fast-developing area of research. Like most areas of Bayesian statistics, Bayesian experimental design has gained popularity in the past two decades. In classical theory, the experimental data are analyzed on the basis of observed data. Inferences are drawn after performing analysis of variance. If however some information in the form of distribution about the parameters of the model is available that is not generally utilized in the classical analysis of variance method. Bayesian approach gives an opportunity to exploit this information to improve the overall conclusions. This prior information may be extracted from the past experiments. The situation where experiments are conducted over the years, analysis of the data can be improved by using the previous year's data. A Bayesian approach to design gives a mechanism for formally

incorporating such information into the design process. By including prior data in the current analysis, the researcher avails himself of additional degrees of freedom that can reduce inference error risk in the current experiment and increase the precision with which results can be reported. That is to say, such a strategy has the potential to significantly reduce uncertainty, thereby improving the quality of the final result.

A Bayesian approach to design gives a mechanism for formally incorporating prior information into the design process. The subject design of experiments has two major components; first component deals with designing the experiment and the second component deals with analysis of the data generated from design of experiments. A lot of literature is available on designing the experiments from Bayesian point of view. Most of the research work centers around obtaining optimal designs. For an excellent review on Bayesian design of experiments one may refer to Chaloner and Verdinelli (1995). It seems that not much attention has been paid to the analysis component. Apart from Flournoy (1993), there are no "true case studies" that we know of where Bayesian ideas have been formally applied to the design of an actual scientific experiment before it is done. The Bayesian analysis of experimental data is considered by Broemeling (1985) and Box and Tiao (1973). The method of Box and Tiao (1973) uses numerical integration to isolate the marginal posterior distribution of each of the variance components, where as method of Broemeling (1985) is based on analytical approximation. Tsutakawa (1972) argued that when Bayesian inference is considered appropriate, it may be desirable to use two separate priors, one for constructing designs and the other for subsequent inference. Etzione and Kadane (1993) and Lindley and Singpurwalla (1991) considered the use of informative priors for design and noninformative priors for the subsequent statistical analysis. Wang and Hsu (2006) gave Bayesian analysis of the additive mixed model for Randomized Block Designs. This paper deals with the Bayesian analysis of the additive mixed model experiments.

In the present study an attempt has been taken to explore the use of Bayesian techniques for analyzing experimental data. Appropriate methodologies for analyzing experimental data from Bayesian point of view have been developed. Before considering the analysis of experimental data, the concept of Bayesian inference has been introduced. Bayesian concept is first explained then its extension to general linear model context is discussed. Then how this concept is used in designed experiments has been explained. Some examples are provided to explain the concept of Bayesian analytical technique for experimental data. In Section 2 concept of Bayesian inference has been introduced. Application of Bayesian inference to general linear model has also been discussed in this Section. Appropriate modification has been done for application into designed experiments. This is the subject matter of Section 3. In Section 4, application of the Bayesian methodologies is considered. Some real experimental data from long term fertilizer experiments have been taken for this purpose. The paper ended with a Section on Discussion.

2. Bayesian Inference

When one does Bayesian statistical inference, one is using prior information and sample data in order to find the values of the parameters in the model which generated the data. The components of statistical inference consist of the prior information, the sample data,

calculation of the posterior density of the parameters and sometimes calculation of the predictive distribution of future observations. The prior information is expressed by a probability density $\xi(\theta)$, of the parameter θ of the model $f(\mathbf{x}\,|\,\theta), \theta \in \Omega, \mathbf{x} \in \mathbf{s}$, where f is density of a random variable x, $\mathbf{s}$ is the sample space and Ω is the parametric space. The information in the data $\mathbf{x} = (x_1, x_2, ..., x_n)'$, where $x_1, x_2, \ldots, x_n$, is a random sample from a population with density f is contained in likelihood function $L(\theta\,|\,\mathbf{x})$, which is the joint density of the sample data. Then this is combined with the prior density of θ, by Bayes' theorem, and gives the posterior density of θ. Thus one may describe inference problem in terms of $(S, \Omega, \xi(\theta), f(\mathbf{x}\,|\,\theta))$ and this problem is solved once the posterior density

$$\xi(\theta\,|\,\mathbf{x}) \propto L(\theta\,|\,\mathbf{x})\xi(\theta) \qquad \theta \in \Omega \tag{2.1}$$

is calculated.

From the posterior density one may make inference for θ by examining the posterior density. Some prefer to give estimates of θ either in point or interval estimates which are computed from the posterior distribution. If θ is one dimensional, a plot of its posterior density tells one the story about θ. But if θ is multidimensional one must be able to isolate those components of $\boldsymbol{\theta}$ (now it is a vector of parameters) in which one is interested. Posterior inference mainly involves estimation, tests of hypothesis and prediction of future observation. In design of experiments, we are mainly interested in testing the significance difference between a pair of treatment effects. Thus our main concern here is tests of hypothesis. Regardless of what particular activity is contemplated, one must first find the posterior density of $\boldsymbol{\theta}$.Often all the components of $\boldsymbol{\theta}$ are of interest, but sometimes some of these components $\boldsymbol{\theta}_1$(say) are regarded as nuisance parameters and the remaining $\boldsymbol{\theta}_2$ are of primary interest. How should one make inferences about $\boldsymbol{\theta}_2$? The Bayesian will use

$$\xi_2(\boldsymbol{\theta}_2\,|\,\mathbf{x}) = \int_{\Omega_1} \xi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,|\,\mathbf{x}) d\boldsymbol{\theta}_1 \qquad \boldsymbol{\theta}_2 \in \Omega_2 \qquad \Omega = (\Omega_1 \cup \Omega_2) \tag{2.2}$$

ξ_2 is called the marginal posterior density of $\boldsymbol{\theta}_2$, and as with any posterior density function it may be used to estimate and test hypothesis concerning $\boldsymbol{\theta}_2$.

2.1 Testing of Hypothesis

To test hypothesis about the parameter $\boldsymbol{\theta}_2$, one perhaps would find a $(1-\lambda)$ Highest Posterior Density (HPD), $0 \le \lambda \le 1$, region $R_\lambda(\boldsymbol{\theta}_2)$ from $\xi_2(\boldsymbol{\theta}_2\,|\,\mathbf{x})$. Such a region has a property that if $\boldsymbol{\theta}'_2 \in R_\lambda(\boldsymbol{\theta}_2)$ and $\boldsymbol{\theta}''_2 \notin R_\lambda(\boldsymbol{\theta}_2)$, then $\xi_2(\boldsymbol{\theta}'_2\,|\,\mathbf{x}) > \xi_2(\boldsymbol{\theta}''_2\,|\,\mathbf{x})$, that is parameter values inside the region have larger posterior probability density than those excluded from the region which must satisfy

$$1-\lambda = \int_{R_\lambda(\boldsymbol{\theta}_2) \subseteq \Omega_2} \xi_2(\boldsymbol{\theta}_2\,|\,\mathbf{x}) d\boldsymbol{\theta}_2, \tag{2.3}$$

that is the HPD region has posterior probability content $1-\lambda$. To test the hypothesis $H_0: \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$ versus $H_1: \boldsymbol{\theta}_2 \ne \boldsymbol{\theta}_{20}$ one rejects H_0 if $\boldsymbol{\theta}_{20}$ is excluded from the region.

2.2 Bayesian Analysis in General Linear Model [Broemeling, 1985]

Let $\boldsymbol{\theta}$ be a $p\times 1$ vector of real parameters, $\mathbf{Y}=(y_1,y_2,...,y_n)'$ a $n\times 1$ vector of observations, $\mathbf{X}$ a $n\times p$ known design matrix. Then the general linear model is

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\theta}+\mathbf{e} \tag{2.4}$$

where $\mathbf{e}\sim N(0,\tau^{-1}\mathbf{I}_n)$ and $\tau\mathbf{I}_n$ is the precision matrix of $\mathbf{e}$, which has covariance matrix $\sigma^2\mathbf{I}_n$, and $\sigma^2=\tau^{-1}>0$ is unknown.

Here our objective is to provide inference for $\boldsymbol{\theta}$ and τ after observing $\mathbf{s}=(y_1,y_2,...,y_n)'$. In Bayesian analysis all inferences are based on the posterior distribution of $\boldsymbol{\theta}$. Suppose one's prior information about $\boldsymbol{\theta}$ is represented by a probability density function $\xi(\boldsymbol{\theta},\tau),\boldsymbol{\theta}\in\mathbf{R}^p,\tau>0,$ then Bayes' theorem combines this information with the information contained in the sample. The likelihood functions for $\boldsymbol{\theta}$ and τ is

$$L(\boldsymbol{\theta},\tau\,|\,\mathbf{s})\propto\tau^{n/2}\exp-\frac{\tau}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta}), \tag{2.5}$$

The likelihood function is one's sample information about the parameters and is the conditional density function of the sample random variables given $\boldsymbol{\theta}$ and τ. Bayes' theorem gives the conditional density of $\boldsymbol{\theta}$ given $\mathbf{s}$ as

$$\xi(\boldsymbol{\theta},\tau\,|\,\mathbf{s})\propto L(\boldsymbol{\theta},\tau\,|\,\mathbf{s})\xi(\boldsymbol{\theta},\tau), \tag{2.6}$$

The posterior density of $\boldsymbol{\theta}$ is $\xi(\boldsymbol{\theta},\tau\,|\,\mathbf{s})$ and represents one's knowledge of $\boldsymbol{\theta}$ and τ after observing the sample $\mathbf{s}$. On the other hand our information about $\boldsymbol{\theta}$ and τ before $\mathbf{s}$ is observed is contained in the prior density.

From (2.6) the posterior density can be written as

$$\xi(\boldsymbol{\theta},\tau\,|\,\mathbf{s})=\mathrm{K}.\mathrm{L}(\boldsymbol{\theta},\tau\,|\,\mathbf{s})\xi(\boldsymbol{\theta},\tau), \tag{2.7}$$

where K is a normalizing constant and is given by

$$\mathrm{K}^{-1}=\int_0^\infty\int_{\boldsymbol{R}^p}L(\boldsymbol{\theta},\tau\,|\,\boldsymbol{s})\xi(\boldsymbol{\theta},\tau)\,d\boldsymbol{\theta}\,d\tau, \tag{2.8}$$

which is the marginal probability density of $\mathbf{Y}$.

2.2.1 Normal-Gamma Prior Density

The prior information about $\boldsymbol{\theta}$ and τ can be given in many ways. In the present study we consider the case when $\xi(\boldsymbol{\theta},\tau)$ is Normal-Gamma prior density, *i.e.*, prior distribution of $\boldsymbol{\theta}$ is Normal and that of τ is Gamma, *i.e.*,

$$\xi(\boldsymbol{\theta},\tau)=\xi_1(\boldsymbol{\theta}\,|\,\tau)\xi_2(\tau), \tag{2.9}$$

where

$$\xi_1(\boldsymbol{\theta}\,|\,\tau)\propto\tau^{p/2}\exp-\frac{\tau}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\theta}-\boldsymbol{\mu}), \tag{2.10}$$

$\boldsymbol{\mu}$ is a $p \times 1$ given vector and $\mathbf{P}$ is known $p \times p$ positive definite matrix. Thus ξ_1 is conditional density of $\boldsymbol{\theta}$ given τ and is normal with mean vector $\boldsymbol{\mu}$ and precision matrix $\tau\mathbf{P}$. The marginal prior density of τ is gamma with parameters $\alpha > 0$ and $\beta > 0$.

$$\xi_2(\tau) \propto \tau^{\alpha-1} \exp -\tau\beta, \tag{2.11}$$

Since (2.9) is the joint prior density of $\boldsymbol{\theta}$ and τ, the marginal density of $\boldsymbol{\theta}$ is

$$\begin{aligned}
\xi_1(\boldsymbol{\theta}) &\propto \int_0^\infty \xi(\boldsymbol{\theta}, \tau)\, \mathrm{d}\tau \\
&\propto \int_0^\infty \tau^{((p+2\alpha)/2-1)} \exp -\frac{\tau}{2}[2\beta + (\boldsymbol{\theta}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\theta}-\boldsymbol{\mu})] d\tau \\
&\propto [2\beta + (\boldsymbol{\theta}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\theta}-\boldsymbol{\mu})]^{-(p+2\alpha)/2},
\end{aligned} \tag{2.12}$$

which is multidimensional *t* density with $p + 2\alpha$ degrees of freedom, location vector $\boldsymbol{\mu}$ and precision matrix $(2\alpha)(2\beta)^{-1}\mathbf{P}$.

With regard to the information about τ

$$\tau = (\sigma^2)^{-1} \quad \text{with} \quad \mathrm{E}(\tau) = \frac{\alpha}{\beta} \text{ and } \mathrm{V}(\tau) = \frac{\alpha}{\beta^2} \tag{2.13}$$

These two equation together with $E(\boldsymbol{\theta}) = \boldsymbol{\mu}$ and $D(\boldsymbol{\theta}) = 2\beta\mathbf{P}^{-1}(n + 2\alpha - 2)^{-1}$, which are the mean vector and dispersion matrix of $\boldsymbol{\theta}$ assist one in choosing the four hyperparameters for the prior distribution of $\boldsymbol{\theta}$ and τ.

By using the Normal-Gamma density as prior for the parameters, one cannot stipulate one's prior information about $\boldsymbol{\theta}$ separately from that of τ.The parameters of the marginal distribution of $\boldsymbol{\theta}$ involves α and β, which are parameters of the prior distribution of τ, but the marginal prior density of τ doesn't involve parameters of the marginal of $\boldsymbol{\theta}$. The parameter $\boldsymbol{\mu}$ is one's prior mean for $\boldsymbol{\theta}$. Actually Normal-Gamma prior density is a member of a conjugate class of distributions. The conjugate families have the advantage that one has a scale by which to judge the amount of information added by the sample, beyond the amount given a priori.

2.2.2 Posterior Analysis

Using Bayes Theorem and using the Normal-Gamma prior density (2.10), the posterior density of $\boldsymbol{\theta}$ and τ is given by

$$\xi(\boldsymbol{\theta}, \tau \mid \mathbf{s}) \propto \tau^{((n+2\alpha+p)/2-1)} \exp -\frac{\tau}{2}[2\beta + (\boldsymbol{\theta}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\theta}-\boldsymbol{\mu}) + (\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})].$$

Now, completing the square on $\boldsymbol{\theta}$ gives

$$\xi(\boldsymbol{\theta},\tau\,|\,\mathbf{s}) \propto \tau^{((n+2\alpha)/2-1)} \exp-\frac{\tau}{2}\left[\beta+\frac{\mathbf{Y'Y-(X'Y+P\mu)'(X'X+P)^{-1}(X'Y+P\mu)}}{2}\right]$$

$$\times\tau^{p/2} \exp-\frac{\tau}{2}[\boldsymbol{\theta}-\mathbf{(X'X+P)^{-1}(X'Y+P\mu)}]'(\mathbf{X'X+P})\times\ [\boldsymbol{\theta}-\mathbf{(X'X+P)^{-1}(X'Y+P\mu)}], \quad (2.14)$$

which is Normal-Gamma density, hence the marginal posterior density of τ is gamma with parameters

$$(n+2\alpha)/2 \quad \text{and} \quad \beta+\frac{\mathbf{Y'Y-(X'Y+P\mu)'(X'X+P)^{-1}(X'Y+P\mu)}}{2} \quad (2.15)$$

The marginal posterior density of $\boldsymbol{\theta}$ is found by integrating (2.14) with respect to τ and yields

$$\xi_1(\boldsymbol{\theta}\,|\,\mathbf{s}) \propto \left\{\begin{array}{l} 2\beta+\mathbf{Y'Y}-(\mathbf{X'Y+P\mu})'(\mathbf{X'X+P})^{-1}(\mathbf{X'Y+P\mu}) \\ +[\boldsymbol{\theta}-(\mathbf{X'X+P})^{-1}(\mathbf{X'Y+P\mu})]'(\mathbf{X'X+P}) \\ \times[\boldsymbol{\theta}-(\mathbf{X'X+P})^{-1}(\mathbf{X'Y+P\mu})] \end{array}\right\}^{-(n+2\alpha+p)/2} \quad (2.16)$$

which is *p*-dimensional *t* density with $n+2\alpha$ degrees of freedom with location vector

$$\boldsymbol{\mu}^* = (\mathbf{X'X+P})^{-1}(\mathbf{X'Y+P\mu}) \quad (2.17)$$

and precision matrix

$$\mathbf{D^*(\boldsymbol{\theta}\,|\,s)} = \mathbf{(X'X+P)}(n+2\alpha)[2\beta+\mathbf{Y'Y}-(\mathbf{X'Y+P\mu})'(\mathbf{X'X+P})^{-1}(\mathbf{X'Y+P\mu})]^{-1} \quad (2.18)$$

2.2.3 Testing of Hypothesis

Suppose we are interested in testing the linear combinations of parameters and consider the following hypotheses

$$H_0: \boldsymbol{U(\theta)} = \mathbf{A\theta} = \mathbf{b} \quad \text{versus} \quad H_1: \boldsymbol{U(\theta)} = \mathbf{A\theta} \neq \mathbf{b};$$

where $\mathbf{A\theta}$ represents a set of contrasts, $\mathbf{A}$ is the matrix of desired contrast of order $m\times p$. The approach is taken here is based on the Highest Posterior Density (HPD) region for $\boldsymbol{\theta}$ as discussed earlier. Since H_0 is given in terms of $\boldsymbol{U(\theta)} = \mathbf{A\theta} = \mathbf{b}$, the distribution of $U(\boldsymbol{\theta})$ is denoted by

$$U(\boldsymbol{\theta}) \sim t_m[u(\boldsymbol{\theta}); n+2\alpha, \mathbf{A\mu^*}, (\mathbf{AD^*A'})^{-1}], \quad (2.19)$$

which is an *m*-dimensional *t* distribution. Since $U(\boldsymbol{\theta})$ has a *t* distribution, the random variable

$$G[U(\boldsymbol{\theta})] = m^{-1}[U(\boldsymbol{\theta})-\mathbf{A\mu^*}]'(\mathbf{AD^*A})^{-1}[U(\boldsymbol{\theta})-\mathbf{A\mu^*}] \quad (2.20)$$

has an F distribution with *m* and $n+2\alpha$ degrees of freedom. Thus $(1-\lambda)$ HPD region for $U(\boldsymbol{\theta}) = u$ is given by

$$R_{1-\lambda}(u) = \{u: G(u) \leq F_{\lambda; m, n+2\alpha}\}, \quad (2.21)$$

where $F_{\lambda;m,n+2\alpha}$ is the upper 100λ% point of the F distribution with *m* and $n+2\alpha$ degrees of freedom. The null hypothesis is rejected if $b \notin R_{1-\lambda}(u)$ or when $G(u) > F_{\lambda;m,\ n+2\alpha}$.

3. Bayesian Analysis in Designed Experiments

Let us consider an additive model for a block design d (say) as

$$y_{ij} = \pi + \gamma_i + \delta_j + e_{ij}, \tag{3.1}$$

where y_{ij} is the response corresponding to i^{th} treatment in j^{th} block, π is the general mean, γ_i is the i^{th} treatment effect, δ_j is the j^{th} block effect, e_{ij} is the error term which follows normal distribution with mean 0 and variance τ^{-1}, $i = 1, 2, \ldots, t$ and $j = 1, 2, \ldots, b$.

Equivalently the model (3.1) can be written as

$$\mathbf{Y} = \pi\mathbf{1} + \mathbf{X}_1\boldsymbol{\gamma} + \mathbf{X}_2\boldsymbol{\delta} + \mathbf{e}, \tag{3.2}$$

where $\mathbf{Y}$ is a $n\times 1$ vector of observations, $\mathbf{X}_1$ is $n \times t$ incidence matrix of treatments, $\boldsymbol{\gamma}$ is a $t\times 1$ vector of treatment effects, $\mathbf{X}_2$ is a $n \times b$ incidence matrix of blocks, $\boldsymbol{\delta}$ is a $b\times 1$ vector of block effects, $\mathbf{1}$ is a unit vector of order $n\times 1$ and $\mathbf{e}$ is a $n\times 1$ vector of errors.

We rewrite the model (3.2) as follows

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}, \tag{3.3}$$

where $\mathbf{X} = [\mathbf{1} \quad \mathbf{X}_1 \quad \mathbf{X}_2]$ and $\boldsymbol{\theta} = [\pi \quad \boldsymbol{\gamma}' \quad \boldsymbol{\delta}']'$ be a $p\times 1$ vector of parameters. It is easy to note that the Rank($\mathbf{X}$) = $k = t - b - 1$.

Since $\mathbf{X}$ is less than full column rank the Bayesian methodologies as described in Section 2 cannot be applied directly to designed experiments. We, therefore, reparameterized the model (3.3) to get full rank model as follows:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e} \tag{3.4}$$

where $\mathbf{Z}$ is $n\times k$ ($k < p$) of full rank, $\boldsymbol{\alpha}$ is $k\times 1$ namely, $\boldsymbol{\alpha} = \mathbf{U}\boldsymbol{\theta}$ where $\mathbf{U}$ is $k\times p$ known matrix, and as before $\mathbf{e} \sim N(0, \tau^{-1}\mathbf{I}_n)$. Thus the contrasts of interest $\boldsymbol{\alpha}$ can now easily be estimated from the model (3.4). To construct $\mathbf{Z}$ and $\mathbf{U}$ we followed reparameterization method proposed by Graybill (1961) as follows:

Since $\mathbf{X}'\mathbf{X}$ is $p\times p$ and symmetric positive semi-definite matrix then there exists a non-singular $p\times p$ matrix $\mathbf{W}^*$ such that $\mathbf{W}'^*\mathbf{X}'\mathbf{X}\mathbf{W}^* = \begin{bmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$, where $\mathbf{B}$ is a $k\times k$ matrix of rank *k*. If we write $\mathbf{W}^* = (\mathbf{W} \quad \mathbf{W}_1)$, where $\mathbf{W}$ is $p\times k$, then the matrix $\mathbf{B}$ would become

$\mathbf{W}'(\mathbf{X}'\mathbf{X})\mathbf{W}$. Now let $\mathbf{W}^{*-1} = \mathbf{U}^* = \begin{pmatrix}\mathbf{U}' & \mathbf{U}_1'\end{pmatrix}'$, where $\mathbf{U}$ is $k \times p$, then pre-multiplying $\mathbf{X}$ in (3.3) with $\mathbf{W}^*$ actually lead to estimation of $\boldsymbol{\alpha}$ in (3.4). Thus remarameterization takes the following form.

$$\mathbf{Z} = \mathbf{XW} \text{ and } \quad \boldsymbol{\alpha} = \mathbf{U}\boldsymbol{\theta}. \tag{3.5}$$

Consider the two way additive model (3.3) with $p = (t+b+1)$ parameters where $\mathbf{X}$ is of rank $k = p-2 = t+b-1$, then using the above reparametrization procedure we arrive at a full rank representation as given in (3.4). Then problem remains to choose $\boldsymbol{\alpha}$, $\mathbf{U}$ and $\mathbf{W}$ appropriately.

Since our interest is to compare various comparisons among treatment effects as well as block effects one may choose $\mathbf{U}$ and $\boldsymbol{\alpha}$ as follows:

$$\boldsymbol{\alpha} = \begin{pmatrix} \pi + \gamma_1 + \delta_1 \\ \gamma_1 - \gamma_2 \\ . \\ . \\ . \\ \gamma_{t-1} - \gamma_t \\ \delta_1 - \delta_2 \\ . \\ . \\ . \\ \delta_{b-1} - \delta_b \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \end{pmatrix}, \tag{3.6}$$

where $\alpha_1 = \pi + \gamma_1 + \delta_1$. This means $\mathbf{U}$ must be

$$\mathbf{U} = \begin{pmatrix} 1;1, & 0, & 0, & ...,0;1, & 0, ...,0 \\ 0;1, & -1, & 0, & ...,0;0, & 0, ...,0 \\ & & . & & \\ & & . & & \\ & & . & & \\ 0;0, & ... & & 1,-1, ... & 0 \\ & & . & & \\ & & . & & \\ & & . & & \\ 0;0, & ... & 0;0, & 0, & 1,-1 \end{pmatrix}, \tag{3.7}$$

which is a $k \times p$ matrix. Then we construct matrix $\mathbf{U}^*$ as $\mathbf{U}^* = \begin{pmatrix}\mathbf{U} \\ \mathbf{U}_1\end{pmatrix}$ such that $\mathbf{U}^*$ is of full rank, where $\mathbf{U}_1$ is $(p-k)\times p$ matrix. Now let $\mathbf{W}^* = \mathbf{U}^{*-1}$, where $\mathbf{W}^* = (\mathbf{W}, \mathbf{W}_1)$, then $\mathbf{Z} = \mathbf{XW}$ and the reparameterization is complete.

3.1 Prior: Normal-Gamma Prior Density

When prior density is Normal-Gamma, that is,

$$\xi(\boldsymbol{\alpha},\tau)=\xi_1(\boldsymbol{\alpha}\mid\tau)\xi_2(\tau), \tag{3.8}$$

where

$$\xi_1(\boldsymbol{\alpha}\mid\tau)\propto\tau^{k/2}\exp-\frac{\tau}{2}(\boldsymbol{\alpha}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\alpha}-\boldsymbol{\mu}),\ \boldsymbol{\alpha}\in R^k\,. \tag{3.9}$$

and $\boldsymbol{\mu}$ is a $k\times1$ given vector and $\mathbf{P}$ is known $k\times k$ positive definite matrix. Thus ξ_1 is conditional density of $\boldsymbol{\alpha}$ given τ and is normal with mean vector $\boldsymbol{\mu}$ and precision matrix $\tau\mathbf{P}$. The marginal prior density of τ is gamma with parameters $\alpha>0$ and $\beta>0$.

$$\xi_2(\tau)\propto\tau^{\alpha-1}\exp-\tau\beta,\quad\tau>0\,. \tag{3.10}$$

Following the procedures outlined for general linear model in Section 2, we obtain the marginal density of $\boldsymbol{\alpha}$ as

$$\propto[2\beta+(\boldsymbol{\alpha}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\alpha}-\boldsymbol{\mu})]^{-(k+2\alpha)/2}, \tag{3.11}$$

which is t density with $p+2\alpha$ degrees of freedom, location vector $\boldsymbol{\mu}$ and precision matrix $(2\alpha)(2\beta)^{-1}\mathbf{P}$.

Similarly information regarding τ is obtained as

$$\tau=(\sigma^2)^{-1}\ \mathrm{E}(\tau)=\frac{\alpha}{\beta}\quad\text{and}\quad\mathrm{V}(\tau)=\frac{\alpha}{\beta^2}\,.$$

These two equation together with $E(\boldsymbol{\alpha})=\boldsymbol{\mu}$ and $D(\boldsymbol{\alpha})=2\beta\mathbf{P}^{-1}(n+2\alpha-2)^{-1}$ are used in choosing the four hyperparameters for the prior distribution of $\boldsymbol{\alpha}$ and τ.

3.2 Posterior Analysis

Following the procedures as outlined in Section 2, density of $\boldsymbol{\alpha}$ and τ can easily be obtained as

$$\xi(\boldsymbol{\alpha},\tau\mid\mathbf{s})\propto\tau^{((\boldsymbol{n}+2\alpha+k)/2-1)}\exp-\frac{\tau}{2}[2\beta+(\boldsymbol{\alpha}-\boldsymbol{\mu})'\mathbf{P}(\boldsymbol{\alpha}-\boldsymbol{\mu})+(\mathbf{Y}-\boldsymbol{Z\alpha})'(\mathbf{Y}-\boldsymbol{Z\alpha})],$$

$$\xi(\boldsymbol{\alpha},\tau\mid\mathbf{s})\propto\tau^{((n+2\alpha)/2-1)}\exp-\frac{\tau}{2}\left[\beta+\frac{\mathbf{Y}'\mathbf{Y}-(\boldsymbol{Z}'\mathbf{Y}+\mathbf{P}\boldsymbol{\mu})'(\boldsymbol{Z}'\boldsymbol{Z}+\mathbf{P})^{-1}(\boldsymbol{Z}'\mathbf{Y}+\mathbf{P}\boldsymbol{\mu})}{2}\right]$$

$$\times\tau^{k/2}\exp-\frac{\tau}{2}[\boldsymbol{\alpha}-(\boldsymbol{Z}'\boldsymbol{Z}+\mathbf{P})^{-1}(\boldsymbol{Z}'\mathbf{Y}+\mathbf{P}\boldsymbol{\mu})]'(\boldsymbol{Z}'\boldsymbol{Z}+\mathbf{P})\times$$

$$[\boldsymbol{\alpha}-(\boldsymbol{Z}'\boldsymbol{Z}+\mathbf{P})^{-1}(\boldsymbol{Z}'\mathbf{Y}+\mathbf{P}\boldsymbol{\mu})] \tag{3.12}$$

which is Normal-Gamma density.

Hence the marginal posterior density of τ is gamma with parameters

$$(n+2\alpha)/2\ \text{ and }\ \beta+\frac{\mathbf{Y}'\mathbf{Y}-(\boldsymbol{Z}'\mathbf{Y}+\mathbf{P}\boldsymbol{\mu})'(\boldsymbol{Z}'\boldsymbol{Z}+\mathbf{P})^{-1}(\boldsymbol{Z}'\mathbf{Y}+\mathbf{P}\boldsymbol{\mu})}{2}\,.$$

The marginal posterior density of $\boldsymbol{\alpha}$ is found by integrating (3.12) with respect to τ and yields

$$\xi_1(\boldsymbol{\alpha}\,|\,\mathbf{s}) \propto \left\{ \begin{array}{l} 2\beta + \mathbf{Y}'\mathbf{Y} - (\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{Z}'\mathbf{Z} + \mathbf{P})^{-1}(\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}) \\ +[\boldsymbol{\alpha} - (\mathbf{Z}'\mathbf{Z} + \mathbf{P})^{-1}(\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})]'(\mathbf{Z}'\mathbf{Z} + \mathbf{P}) \\ \times[\boldsymbol{\alpha} - (\mathbf{Z}'\mathbf{Z} + \mathbf{P})^{-1}(\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})] \end{array} \right\}^{-(n+2\alpha+k)/2} \tag{3.13}$$

which is *k*-dimensional *t* density with $n + 2\alpha$ degrees of freedom with location vector $\boldsymbol{\alpha}^* = (\mathbf{Z}'\mathbf{Z} + \mathbf{P})^{-1}(\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})$ and precision matrix

$$D^*(\boldsymbol{\alpha}\,|\,\mathbf{s}) = (\mathbf{Z}'\mathbf{Z} + \mathbf{P})(n + 2\alpha)[2\beta + \mathbf{Y}'\mathbf{Y} - (\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{Z}'\mathbf{Z} + \mathbf{P})^{-1}(\mathbf{Z}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})]^{-1}$$

3.3 Testing of Hypotheses

We are interested to pair wise comparison of treatment means. Thus in particular, we are interested to test H_0: $\gamma_1 = \gamma_2 = ... = \gamma_t$, which is true for $\boldsymbol{\alpha}_2 = \mathbf{0}$. This can be tested using the HPD region as discussed in Section 2. Since the random variable

$$F(\boldsymbol{\alpha}_2\,|\,\mathbf{s}) = (t-1)^{-1}[\boldsymbol{\alpha}_2 - E(\boldsymbol{\alpha}_2\,|\,\mathbf{s})]'\mathbf{P}(\boldsymbol{\alpha}_2\,|\,\mathbf{s})[\boldsymbol{\alpha}_2 - E(\boldsymbol{\alpha}_2\,|\,\mathbf{s})] \tag{3.14}$$

has an F distribution with *t-1* and *rbt-(b+t-1)* degrees of freedom, $(1-\lambda)$ HPD region for $\boldsymbol{\alpha}_2 = 0$ is given by

$$\mathrm{R}_{1-\lambda}(\mathrm{u}) = \{\mathrm{u} : \mathrm{G}(\mathrm{u}) \le \mathrm{F}_{\lambda;\, m,\, n+2\alpha}\}, \tag{3.15}$$

Consequently

$$\mathrm{E}(\boldsymbol{\alpha}_2|\mathbf{y}) = (\boldsymbol{\varphi}_1, \mathbf{I}_{\mathrm{t}-1}, \boldsymbol{\varphi}_2)\boldsymbol{\alpha}^*$$

Where $\boldsymbol{\varphi}_1$ is $\mathrm{t}-1\times 1$ matrix of zeros, $\mathbf{I}_{\mathrm{t}-1}$ is the identity matrix of order t-1 and $\boldsymbol{\varphi}_2$ is a zero matrix of order $(\mathrm{t}-1)\times(\mathrm{b}-1)$. The precision matrix of $\boldsymbol{\alpha}_2$ is

$$\mathrm{P}(\boldsymbol{\alpha}_2|\mathbf{y}) = [(\boldsymbol{\varphi}_1, \mathbf{I}_{\mathrm{t}-1}, \boldsymbol{\varphi}_2)\mathbf{P}^{-1}(\boldsymbol{\alpha}\,|\,\mathbf{y})(\boldsymbol{\varphi}_1, \mathbf{I}_{\mathrm{t}-1}, \boldsymbol{\varphi}_2)']^{-1},$$

where $\mathrm{F}_{\lambda;\, m,\, n+2\alpha}$ is the upper $100\lambda\%$ point of the F distribution with *m* and $n+2\alpha$ degrees of freedom. The null hypothesis is rejected if $\mathrm{b} \notin \mathrm{R}_{1-\lambda}(\mathrm{u})$ or when $\mathrm{G}(\mathrm{u}) > \mathrm{F}_{\lambda;\, m,\ n+2\alpha}$.

4. Applications

In this Section we applied the Bayesian methodologies as developed for designed experiments to real experimental data.

Data description

The data pertaining to Long-Term Fertilizer Experiments conducted at Ranchi, India has been considered for the present study. The experiment was conducted using a

Randomized Complete Block (RCB) design in 10 treatments and 4 replications. The crop sequence is Soybeans–Wheat and the data is available from 1979 to 2003. The plot size adopted is 100 sq. m. (12.5 m × 8.0 m). The details of the treatments are as given in Table 1.

Table 1: Treatments for Long Term Fertilizer Experiments

Treatments	Treatment details
T1	50% Optimal NPK
T2	100% Optimal NPK
T3	150% Optimal NPK
T4	100% Optimal NPK + Hand Weeding
T5	100% Optimal NPK + Lime
T6	100% Optimal NP
T7	100% Optimal N
T8	100% Optimal NPK + FYM
T9	100% Optimal NPK (Sulphur free)
T10	Control

We now consider some examples based on the data generated through this experiment.

Example 1: Here yield data on wheat for 2003 data is taken for the Bayesian analysis which serves the purpose of providing the likelihood information and the 2002 data serves the purpose of providing the prior information. The analysis is done in SAS package (SAS, 1990) in Interactive Matrix Language (IML). SAS codes are available with the authors and can be obtained on request. First we carry out usual classical analysis and then Bayesian analysis is performed.

Table 2: Yield of wheat in quintal/hectare year

	2002				2003			
	Blocks				Blocks			
Treatments	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1	22.10	15.90	20.40	20.40	29.15	29.70	32.00	28.30
2	28.50	27.85	28.50	27.35	38.80	33.20	37.50	40.50
3	23.85	30.38	30.40	28.55	27.50	42.60	39.00	28.15
4	33.30	32.30	31.57	30.00	37.90	44.45	39.25	41.15
5	32.10	40.40	42.50	43.90	37.25	43.20	46.20	48.60
6	30.30	29.65	33.37	30.25	42.55	33.25	42.91	39.50
7	0.75	0.50	0.48	0.50	0.90	0.75	2.39	0.80
8	41.70	42.40	40.50	42.00	43.40	42.10	47.50	51.20
9	8.35	8.25	10.02	8.43	8.90	12.30	9.74	6.55
10	3.80	3.43	3.90	3.60	4.20	5.10	5.55	5.80

Classical Analysis

We first conducted usual (classical) analysis of the data. The analysis of variance table is given in Table 3.

Table 3: Analysis of variance with the original 2003 data

Source of variation	Degrees of freedom	Sum of Squares	Mean Sum of Squares	F-value	p-value
Treatment	9	10337.024	1148.558	79.80	<.0001
Block	3	50.872	16.957	1.18	0.3365
Error	27	388.621	14.393		
Total	39	10776.519			

From the table it is observed that the treatment effects are significant at 5% level of significance and block effects are not significant at 5% level of significance. We then performed the test for pair wise comparison, i.e., whether there is any significance difference between a pair of treatment effects or not. Under this pair wise treatment effects comparison it is found that there are 12 number of treatment pairs which are not significant and these are (1,3), (2,3), (2,4), (2,6), (3,6), (4,5), (4,6), (4,8), (5,6), (5,8), (7,10) and (9,10). For example, treatment 1 is not significantly different from treatment 3 for the treatment comparison (1, 3) and so on.

Bayesian analysis

Now we apply Bayesian method of analysis of designed experiment. The model for Bayesian analysis is as given by (3.1). Here our interest is to test whether treatment effects are significantly different or not, i.e., whether the parameters γ_i's in (3.1) are significantly different from each other or not. Similarly if we are interested in the block parameters, then we can perform similar test for the parameters δ_i's. Here we considered Normal-Gamma prior as prior distribution as described in Section 3. Since we are interested to test the treatment comparisons, we developed HPD region for the parameter vector $\boldsymbol{\alpha}_2$ as described in Section 3.3. Similarly for comparison of block effects HPD region is developed for parameter vector $\boldsymbol{\alpha}_3$. Firstly, for testing the overall significance of treatment effects the probabilities of HPD regions for $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ are calculated using F-value as described in Section 3.3. Then for testing individual treatment effects comparison, probabilities of HPD regions are computed by taking individual component form $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ one by one. For the analysis of the year 2003 data, the 2002 data is taken as prior, i.e., data of 2002 has been used to estimate the hyperparameters. From (3.15), we calculate the F-value for the comparisons of overall treatment effects and it is found to be 110.75. Then the probability of HPD region with $\lambda = 0.05$ has been calculated. As explained in Section 3.3, this is calculated through F-distribution. The corresponding degrees of freedom are 9 and 27. From the table this value is obtained as 2.25. Since the calculated value is greater than the tabulated value, the null hypothesis that there is no difference between any treatment effects is rejected and concluded that the treatment effects are significant at 5% level of significance. However, the block effects are found to be non significant, as the F-test value for the block effects is 1.99, while tabulated value

for F ($\lambda = 0.05, 3, 27$) is 2.96. We then carry out testing for individual treatment effects comparisons. It was found that 7 treatment pairs are non significant. These pairs are (2,3), (2,4), (2,6), (4,5), (4,6), (5,8), and (7,10).

There are 12 treatment pairs which are not significant in classical method of analysis. On performing Bayesian analysis of the same set of data and combining the related prior information in the procedure of analysis there are only 7 non-significant pairs of treatment effects under Normal-Gamma prior. Thus we find improvement over 5 treatment pairs. These pairs are **(1,3), (3,6), (4,8), (5,6)** and **(9,10),** *i.e.*, these treatment pairs are now significant under Bayesian analysis.

Example 2: Here the data for the year 1999 on wheat is considered as data providing prior information while the data for year 2000 is taken as current data for likelihood information.

Table 4: Yield of wheat in quintal/hectare

	1999				**2000**			
	Block				**Block**			
Treatment	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1	21.78	22.80	27.70	23.40	20.10	19.30	18.60	20.30
2	37.40	32.10	40.20	35.55	30.00	28.10	32.85	32.95
3	35.50	34.80	40.85	39.80	29.40	33.35	36.30	37.40
4	36.45	34.10	39.40	37.10	30.15	33.25	34.25	39.10
5	40.90	44.80	41.10	39.35	43.40	40.80	41.10	38.90
6	30.80	32.95	28.00	28.83	30.40	24.60	30.85	32.65
7	1.85	1.25	1.23	1.65	1.95	1.10	1.25	1.80
8	40.10	42.10	40.00	47.70	32.80	38.80	41.50	43.90
9	19.20	20.88	19.35	11.38	15.60	13.75	14.87	13.55
10	3.90	3.60	5.10	2.90	3.40	3.95	4.90	3.40

Table 5: Analysis of variance with the original year 2000 data

Source of variation	**Degrees of freedom**	**Sum of Squares**	**Mean Sum of Squares**	**F-value**	**p-value**
Treatment	9	7286.691	809.632	134.97	<.0001
Block	3	56.206	18.735	3.12	0.0423
Error	27	161.967	5.998		
Total	39	7504.865			

From the above table it is observed that the treatment effects and block effects both are significant at 5% level of significance. Under the comparison of various pair wise treatment contrasts it was found that treatment pairs (2,3), (2,4), (2,6), (3,4), (5,8) and (7,10) are not significant.

Bayesian analysis

We find that the F-test value in case of Normal-Gamma prior for the treatment effects is 199.2010, while tabulated value for F (λ=0.05, 9, 27) is 2.2501. Therefore treatment effects are significant at 5% level of significance. The block effects are also found to be significant, as the F-test value for the block effects is 5.6264, while tabulated value for F (λ=0.05, 3, 27) is 2.9603. Under pair wise comparison of treatment effects we found that pairs (2,3) and (2,4) have now become significant. However, pairs (2,6), (3,4), (5,8) and (7,10) are still not significant. Thus there is a significant change is observed in using Bayesian methods.

5. Discussion

In the present study we present the Bayesian methodologies for analyzing experimental data. Here we restricted ourselves to the case of conjugate prior distributions of the parameters. Bayesian methods are well established in general linear models, where the design matrix is of full column rank. Here the approach is taken as that of Broemeling (1985), who developed usual F-statistics and t-statistic for testing the parameters or a linear function of parameters. However, these methods cannot be applied directly to designed experiments, because of rank deficiency of its design matrix. Therefore, reparameterization has been done in order to obtain full rank model.

One point to note is that though we need prior information for the parameters, yet one may wonder how old this prior information should be. For application in the present study, we had a series of experiments. Analysis was done by taking the data for a particular year and the data corresponding to its previous year was taken as prior information. However, we have also analyzed a data set for a particular year by taking a long series of data of previous years as prior separately. But there is not much change in the results as compared to that obtained by taking just previous year's data as prior. This suggests that for analyzing data from Bayesian point of view we need only the previous information about the parameters, it does not matter how old that information is.

References

1. Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (Second edition). Springer–Verlag, New York:

2. Box, G. E .P. and Tiao, G.C. (1973). *Bayesian inference in statistical analysis.* Addison-Wesley Publishing Company, London.

3. Broemeling, L. D. (1985). *Bayesian analysis of linear models*. Marcel Dekker, New York.

4. Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review, *Statistical Science*, 10, 273 – 304.

5. DeGroot, M.H. (1986). Concept of information based on utility. In recent developments in the foundations of utility and risk theory, L. Daboni, A. Montesano and M. Lines eds., Reidel, Dordrecht, Holland.

6. Etzione, R. and Kadane, J. B. (1993). Optimal experimental design for another's analysis. *J. Amer. Statist. Assoc*., 88, 1404-1411.

7. Flournoy, N. (1993). A clinical experiment in bone marrow transplantation: estimating a percentage point of a quantal response curve. *In case studies in Bayesian Statistics,* 324-336, Springer, New York.

8. Graybill, F. A. (1961). *An introduction to linear statistical models.* McGraw Hill, New York.

9. Lindley, D. V. and Singpurwalla, N. D. (1991). On the evidence needed to reach agreed action between adversaries, with application to acceptance sampling, *J. A. S. A.*, 86**,** 933-937.

10. Raiffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory.* Harvard University Press. Cambridge.

11. SAS Institute Inc. (1989). *SAS/STAT User's Guide*, Vol.2, version 6, 4th ed. Cary, NC.

12. Savage, L.J. (1972). *The foundation of statistics.* Dover, Inc., New York.

13. Tsutkawa, R. K. (1972). Design of an experiment for Bio assay. *J. Amer. Statist. Assoc.* 67, 584-590.

14. Wang, J. and Hsu, J. S. J. (2006). Bayesian analysis of the linear mixed model for randomized block designs. *Aust. N. Z. J. Stat.*, 48, 225-236.