On Selecting the Best Treatment in a Generalized Linear Model

Naveen K. Bansal[a], Klaus J. Miescke[b]

[a]*Department of Mathematics, Statistics, and Computer Science, P.O. Box 1881, Marquette University, Milwaukee, WI 53201, U.S.A.*
naveen.bansal@marquette.edu
[b]*Department of Mathematics, Statistics, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607-7045, U.S.A.*
klaus@uic.edu

Abstract: The problem of selecting the best treatment is studied under generalized linear models. For certain balanced designs, it is shown that simple rules are Bayes with respect to any non-informative prior on the treatment effects under any monotone invariant loss. When the nuisance parameters such as block effects are assumed to follow a uniform (improper) prior or a normal prior, Bayes rules are obtained for the normal linear model under more suitable balanced designs, keeping the generality of the loss and the generality of the non-informativeness on the prior of the treatment effects. These results are extended to certain types of informative priors on the treatment effects. When the designs are unbalanced, algorithms based on the Gibbs sampler and the Laplace method are provided to compute the Bayes rules.

MSC: 62F07; 62J12
Keywords: Selection of populations; generalized linear model; Gibbs sampler; Laplace method; balanced and unbalanced designs; Bayes rule.

1. Introduction

Most of the selection problems in statistical literature deal with the selection of the best of k populations when the populations are from one-parameter or two-parameter families. There are basically three approaches to the selection problem. The first, the *indifference-zone* approach, is due to Bechhofer (1954). It deals with the construction of the design, i.e. the minimum sample size for each of the populations, so that a natural selection criteria selects the correct population with

a pre-specified probability under the *preference-zone*. The complement of the *preference-zone* is called the *indifference-zone*. The second approach, the *subset selection* approach, is due to Gupta (1956). The goal under subset selection approach is to select a subset of k populations so that the best population is included in the subset with a guaranteed pre-specified probability. Both of these are frequentist approaches and can be generalized to the problem of selecting the best treatment in general design models; see, for example, Gupta and Panchapakesan (1996) and Bechhofer, Santner and Goldsman (1995). The third approach is the decision theoretic approach, first introduced by Bahadur (1950) and Bahadur and Goodman (1952). Lehmann (1966) and Eaton (1967) gave a more precise framework to this problem. Most of the work in the decision theoretic approach deals with the optimality or non-optimality of the natural selection rule; see, for example, Gupta and Miescke (1984, 1988) and Abughalous and Miescke (1989). Bansal and Gupta (1997) used this approach for selecting the best treatment in the normal general linear model. The goal of this paper is to discuss the decision theoretical approach for selecting the best treatment for a generalized linear model where the distribution may not be restricted to the normal distribution.

Consider n independent random observations $Y_1, Y_2, ..., Y_n$, where for $i = 1, ..., k$, Y_i has the density, with respect to a σ-finite measure $\nu(\cdot)$,

$$f(y_i, \theta_i, \sigma^2) = \exp\{(y_i\,\theta_i - b(\theta_i))/\sigma^2 + c(y_i, \sigma^2)\}. \tag{1.1}$$

Here, θ_i is the natural parameter that belongs to a natural parameter space $\Theta = \{\theta \in \mathbb{R} : b(\theta) < \infty\}$, and $\sigma^2 > 0$ is a dispersion parameter (see McCullagh and Nelder, 1989). The functions b and c specify the particular parametric family (e.g., binomial, normal, or Poisson). For one-parameter families such as binomial or Poisson, $\sigma^2 = 1$, but to cover the overdispersion in these distributions, the dispersion parameter σ^2 is still justifiable, see McCullagh and Nelder (1989). for notational convenience, let $\mathbf{y} = (y_1, y_2, ..., y_n)^T$ in the sequel.

Let $\mu_i = b'(\theta_i)$ be the mean of Y_i. Suppose μ_i depends on the external factors that can be identified by the design vectors $\mathbf{x}_i = (x_{i,1}, ..., x_{i,p+q})^T$, a $p \times 1$ vector of treatment effects $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_p)^T$ and a $q \times 1$ vector of block effects $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_q)^T$. The mean response μ_i and the factors are associated by a link function

$$g(\mu_i) = \mathbf{x}_i^T \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\beta} \end{pmatrix}.$$

$g(.)$ is called the canonical link if $g(\mu_i) = \theta_i$. In this case the natural parameter

θ_i and the treatment and block effects are associated by

$$\theta_i = \sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j \ , \tag{1.2}$$

or in matrix notation, with $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)^T$,

$$\boldsymbol{\theta} = X_1\,\boldsymbol{\tau} + X_2\,\boldsymbol{\beta} \ , \tag{1.3}$$

where $X_1(n\times p)$ and $X_2\,(n\times q)$ are the design matrices with i^{th} row $(x_{i,1}, x_{i,2}, ..., x_{i,p})^T$ and $(x_{i,p+1}, x_{i,p+2}, ..., x_{i,p+q})^T$ respectively. Throughout this paper we will assume the canonical link (1.2). Note that in some cases, the natural parameter space Θ, although a convex set, is not necessarily the entire real line $\mathbb{R}$. In such cases the spaces of $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ are restricted so that θ_i, $i = 1, ..., n$, defined by (1.2), belong to the natural space Θ. In cases such as binomial, normal and Poisson, no restrictions are needed since in these cases $\Theta = \mathbb{R}$. The parameter space for the model is given by

$$\Omega = \{\boldsymbol{\omega} = (\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) : \theta_i = \mathbf{x}_i^T(\boldsymbol{\tau}^T, \boldsymbol{\beta}^T) \in \Theta,\ \sigma^2 > 0\}. \tag{1.4}$$

The problem under consideration is to select the treatment that is associated with the largest effect, i.e., $\tau_{[p]} = \max(\tau_1, ..., \tau_p)$. The special case of this problem where the density (1.1) is a normal density has been considered by Bansal and Gupta (1997), Bansal and Misra (1999), and Bansal and Miescke (2002). Selection of the best of k exponential families has been studied by Abughalous and Bansal (1995), where, although the discussion was confined to exponential families with a quadratic variance function, many of the results mentioned there can be extended to general one-parameter exponential families.

Our main goal is the construction of designs under which a selection rule can be obtained so that it is Bayes under any monotone permutation invariant loss $L(\boldsymbol{\tau}, j)$ with respect to any prior that has some general structure, where $L(\boldsymbol{\tau}, j)$ denotes the loss function for selecting the jth component of $\boldsymbol{\tau}$ as the best, i.e. that is associated with $\tau_{[p]}$. The objective is to attain optimality in generality with respect to the loss as well as the prior.

A loss function $L(\boldsymbol{\tau}, j)$ is called monotone invariant loss if it satisfies the two conditions

$$L(\boldsymbol{\tau}, j_1) \le L(\boldsymbol{\tau}, j_2) \ \text{ for } \tau_{j_1} \ge \tau_{j_2} \ \text{ and } \ j_1 \ne j_2 \ , \tag{1.5}$$

and

$$L(\boldsymbol{\tau}, \rho(j)) = L(\rho(\boldsymbol{\tau}), j) \ \text{ for any } \rho(\boldsymbol{\tau}) = (\tau_{\rho(1)}, ..., \tau_{\rho(p)}) \ , \tag{1.6}$$

where $(\rho(1), ..., \rho(p))$ denotes a permutation of $(1, ..., p)$.

A special case of such a loss function is the traditional so-called "0-1" loss,

$$L^{0,1}(\boldsymbol{\tau}, j) = \begin{cases} 1 & \text{if} \quad \tau_j \neq \tau_{[p]} \\ 0 & \text{if} \quad \tau_j = \tau_{[p]} \end{cases} \tag{1.7}$$

under which a non-randomized rule $d(\mathbf{y})$ yields the expected loss $E_\omega[L(\boldsymbol{\tau}, d(\mathbf{Y}))]$ as the probability of incorrect selection, i.e., $P_\omega\{d(\mathbf{Y}) \neq (p)\}$, where $\tau_{(p)}$ denotes the component of $\boldsymbol{\tau}$ that is associated with $\tau_{[p]}$, assuming here for simplicity that it is unique.

Another special case of a monotone permutation invariant loss that is often considered is the so-called linear loss,

$$L^{lin}(\boldsymbol{\tau}, j) = \tau_{[p]} - \tau_j \,. \tag{1.8}$$

The article is organized as follows. In Section 2, we will consider a non-informative prior on $\boldsymbol{\tau}$ with a general structure, and Section 3 will be devoted to an informative prior with a general structure derived from exponential tilting. The goal of these two sections are to obtain some balanced designs and Bayes selection rules in a very general setting, both in terms of the loss function and in the prior information. Section 4 will be devoted to the Bayes rule with respect to the conjugate prior under any design. We will provide algorithms using the Gibbs sampler and the Laplace method to compute Bayes rules.

2. Bayes rules with respect to non-informative priors

Let $\phi(\mathbf{y}) = (\phi_1(\mathbf{y}), ..., \phi_p(\mathbf{y}))$ denotes a randomized decision rule, where $\phi_j(\mathbf{y})$ is the probability of selecting the jth component as the best, $j = 1, ..., p$. The expected loss of $\phi(\mathbf{y})$ is given by

$$\sum_{j=1}^{p} E^\omega E^{\mathbf{y}|\omega}[L(\boldsymbol{\tau}, j)\phi_j(\mathbf{Y})],$$

where the first expectation is with respect to the conditional distribution of $\mathbf{Y}$ given $\boldsymbol{\omega} = \omega$, and the second with respect to the prior distribution on Ω. It is easy to see, due to the monotonicity of the loss function, that the Bayes rule is given by $\phi^B(\mathbf{y}) = (\phi_1^B(\mathbf{y}), ..., \phi_p^B(\mathbf{y}))$, with

$$\phi_j^B(\mathbf{y}) = \begin{cases} |N(\mathbf{y})|^{-1} & \text{if} \quad j \in N(\mathbf{y}) \\ 0 & \text{if} \quad j \notin N(\mathbf{y}) \end{cases} \tag{2.1}$$

where

$$N(\mathbf{y}) = \{k : E^{\tau|\mathbf{y}} L(\boldsymbol{\tau}, k) = \min_{j=1,\ldots,p} E^{\tau|\mathbf{y}} L(\boldsymbol{\tau}, j)\}. \tag{2.2}$$

$|N(\mathbf{y})|$ denotes the cardinality of $N(\mathbf{y})$, and $E^{\tau|\mathbf{y}}$ denotes the expectation with respect to the posterior distribution of $\boldsymbol{\tau}$ given $\mathbf{Y} = \mathbf{y}$. Note that the posterior expected loss is proportional to

$$r_j(\mathbf{y}) = \int L(\boldsymbol{\tau}, j)\, f(\mathbf{y}; \boldsymbol{\omega})\, d\pi(\boldsymbol{\omega}) \tag{2.3}$$

where $f(\mathbf{y}; \boldsymbol{\omega})$ is the joint density of $\mathbf{Y}$ given $\boldsymbol{\omega} = \omega$, and $\pi(\cdot)$ is the prior on Ω. Thus, from (2.2),

$$N(\mathbf{y}) = \{k : r_k(\mathbf{y}) = \min_{j=1,\ldots,p} r_j(\mathbf{y})\} \tag{2.4}$$

The prior $\pi(\cdot)$ on Ω is called permutation invariant with respect to the components of $\boldsymbol{\tau}$ if

$$\pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) = \pi(\rho(\boldsymbol{\tau}), \boldsymbol{\beta}, \sigma^2), \text{ for all } (\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) \in \Omega, \tag{2.5}$$

for all permutations $(\rho(1), ..., \rho(p))$ of $(1, ..., p)$. We call this prior to be non-informative for the selection problem because (2.5) implies that apriori no one treatment is preferred over any other treatment. Note that this is the most general form of a non-informative prior for the selection problem because if there is no apriori information about the order of the components of $\boldsymbol{\tau}$, then the prior must satisfy (2.5).

It can be seen (see Bansal and Gupta 2000) that $r_{j_1}(\mathbf{y}) \geq r_{j_2}(\mathbf{y})$ for any $j_1 \neq j_2$, if

$$f(\mathbf{y}; \boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) \geq f(\mathbf{y}; \boldsymbol{\tau}^{(j_1,j_2)}, \boldsymbol{\beta}, \sigma^2) \tag{2.6}$$

whenever $\tau_{j_1} \geq \tau_{j_2}$. Here $\boldsymbol{\tau}^{(j_1,j_2)}$ denotes the vector $\boldsymbol{\tau}$ with components τ_{j_1} and τ_{j_2} interchanged. The proof of this essentially follows along the lines of proof of Theorem 4.1 of Eaton (1967).

From (1.1) and (1.2),

$$\begin{aligned} f(\mathbf{y}; \boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) &= \exp\{\frac{1}{\sigma^2}[\sum_{i=1}^{n}\sum_{j=1}^{p} y_i x_{i,j}\tau_j + \sum_{i=1}^{n}\sum_{j=1}^{q} y_i x_{i,p+j}\beta_j \\ &\quad - \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j)] + \sum_{i=1}^{n} c(y_i, \sigma^2)\}. \end{aligned} \tag{2.7}$$

Denoting

$$\begin{aligned} t_j &= \sum_{i=1}^{n} y_i x_{i,j}\,, \quad j=1,...,p, \quad \text{and} \\ b_j &= \sum_{i=1}^{n} y_i x_{i,p+j}\,,\ j=1,...,q, \end{aligned}$$

we get

$$\begin{aligned} f(\mathbf{y};\boldsymbol{\tau},\boldsymbol{\beta},\sigma^2) &= \exp\{\frac{1}{\sigma^2}[\sum_{j=1}^{p} t_j\tau_j + \sum_{j=1}^{q} b_j\beta_j \\ &\quad -\sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j)] + \sum_{i=1}^{n} c(y_i,\sigma^2)\}. \end{aligned} \tag{2.8}$$

From (2.6) and (2.8), we get that $r_{j_1}(\mathbf{y}) \geq r_{j_2}(\mathbf{y})$ if

$$\begin{aligned} (t_{j_1}-t_{j_2})(\tau_{j_1}-\tau_{j_2}) &\geq \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j) \\ &\quad -\sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j^{(j_1,j_2)} + \sum_{j=1}^{q} x_{i,p+j}\beta_j) \end{aligned} \tag{2.9}$$

whenever $\tau_{j_1} \geq \tau_{j_2}$ for $j_1 \neq j_2$. Here $\tau_j^{(j_1,j_2)}$ denotes the jth component of $\tau^{(j_1,j_2)}$.

This implies that if the design is such that the right hand side of (2.9) is 0, for all $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$, then $r_{j_1}(\mathbf{y}) \geq r_{j_2}(\mathbf{y})$ if $t_{j_1} \geq t_{j_2}$. Thus comparing r_{j_1} and r_{j_2} for all $j_1 \neq j_2$, from (2.1) and (2.4), we get the following result.

Theorem 2.1. *If the design is such that, for all $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ in the parameter space,*

$$\sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j) = \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j^{(j_1,j_2)} + \sum_{j=1}^{q} x_{i,p+j}\beta_j) \tag{2.10}$$

for all $j_1 \neq j_2$, where $\tau_j^{(j_1,j_2)}$ is the jth component of $\boldsymbol{\tau}^{(j_1,j_2)}$, then the Bayes rule with respect to any permutation invariant prior (2.5) under any monotone invariant loss is

$$\tilde{\phi}_j^B(\mathbf{y}) = \begin{cases} \left|\widetilde{N}^B(\mathbf{y})\right|^{-1} & \text{if} \quad j \in \widetilde{N}^B(\mathbf{y}) \\ 0 & \text{if} \quad j \notin \widetilde{N}^B(\mathbf{y}) \end{cases} \tag{2.11}$$

where $\widetilde{N}^B(\mathbf{y}) = \{k : t_k = \max(t_1, t_2, ..., t_p)\}$.

To see how this theorem works, consider an example of a randomized block design with q number of blocks, and with p treatments assigned randomly to each block with replications. Denote n_{ij} as the number of replications for the jth treatment in i^{th} block, y_{ijk} as the response of the kth replication due to the jth treatment in the ith block. Assume that the probability density of y_{ijk} belongs to the exponential dispersion family (1.1) with the canonical parameter

$$\theta_{ij} = \tau_j + \beta_i$$

Writing $\mathbf{y} = (y_{111}, ..., y_{11n_{11}}, ..., y_{1p1}, ..., y_{1pn_{1p}}, ..., y_{i11}, ..., y_{i1n_{i1}}, ...)^T$, and $\boldsymbol{\theta} = (\theta_{11}, ..., \theta_{1p}, ..., \theta_{i1}, ..., \theta_{ip}, ...)^T$, it can be seen that the joint density of $\mathbf{y}$ is of the form (2.8) with

$$\begin{aligned} t_j &= \sum_{i=1}^{q} n_{ij}\overline{y}_{ij}\,,\ j = 1, ..., p \\ b_i &= n_{i.}\overline{y}_{i.}\,, i = 1, ..., q \end{aligned} \tag{2.12}$$

and

$$\boldsymbol{\theta} = X_1\boldsymbol{\tau} + X_2\boldsymbol{\beta}$$

with

$$[X_1 : X_2] = \left[\begin{pmatrix} diag(\mathbf{1}_{n_{11}}, ..., \mathbf{1}_{n_{1p}}) \\ diag(\mathbf{1}_{n_{21}}, ..., \mathbf{1}_{n_{2p}}) \\ ..... \\ diag(\mathbf{1}_{n_{q1}}, ..., \mathbf{1}_{n_{qp}}) \end{pmatrix} : \begin{pmatrix} \mathbf{1}_{n_{1.}}, \mathbf{0}, ..., \mathbf{0} \\ \mathbf{0}, \mathbf{1}_{n_{2.}}, ..., \mathbf{0} \\ ..... \\ \mathbf{0}, ..., \mathbf{0}, \mathbf{1}_{n_{q.}} \end{pmatrix} \right]$$

where $\mathbf{1}_m$ denotes a vector of $1s$ of dimension m, $\mathbf{0}$ is a vector of $0s$, $\overline{y}_{ij}$ is the mean response corresponding to the jth treatment in the ith block, and $\overline{y}_{i.}$ is the mean response corresponding to the ith block, and $n_{i.}$ is the ith block size.

For this design, the design condition (2.10) of Theorem 2.1 implies that

$$\sum_{i=1}^{q}\sum_{j=1}^{p} n_{ij}b(\tau_j + \beta_i) = \sum_{i=1}^{q}\sum_{j=1}^{p} n_{ij}b(\tau_j^{(j_1,j_2)} + \beta_i)$$

for all $j_1 \neq j_2$. Note that this condition holds if $n_{i1} = n_{i2} = \cdots = n_{ip} = n_i^*$, say, i.e. if the number of replications within each block are the same for all treatments. Thus, from Theorem 2.1 and (2.12), if the number of replications within each block is the same for all the treatments, then the Bayes rule with respect to any

permutation invariant prior (2.5) under any monotone permutation invariant loss selects the best treatment according to the largest value of $\sum_{i=1}^{q} n_i^* \overline{y}_{ij}$ with ties broken randomly.

Remark 1. *For any fixed $\boldsymbol{\tau}$, $\boldsymbol{\beta}$, and $\sigma^2 > 0$, by taking the prior which puts mass of $\frac{1}{p!}$ at each of $(\rho(\boldsymbol{\tau}), \boldsymbol{\beta}, \sigma^2)$, where ρ varies over all $p!$ permutations, Theorem (2.1) implies that under the design condition (2.10), $\tilde{\phi}^B(y)$ minimizes the average risk $\frac{1}{p!}\sum_\rho R((\rho(\boldsymbol{\tau}), \boldsymbol{\beta}, \sigma^2), \phi)$ among all the selection rules ϕ. Thus, from Lemma 1 of Lehmann (1966), Theorem (2.1) implies that under the design condition (2.10), $\tilde{\phi}^B(y)$ minimizes the risk $R(\boldsymbol{\omega}, \phi)$ uniformly in $\boldsymbol{\omega} \in \Omega$ among all selection rules that are invariant in the components of $\mathbf{t}$.*

Note that the condition (2.10) is satisfied by a randomized complete block design, but not by an incomplete balanced design. For example, the following incomplete block design with ($p = 3$, $q = 3$) does not satisfy this condition.

block 1	block 2	block 3
trt 1	trt 2	trt 1
trt 2	trt 3	trt 3

(2.13)

For this design, the left hand side of (2.10) is

$$b(\tau_1 + \beta_1) + b(\tau_2 + \beta_1) + b(\tau_2 + \beta_2) + b(\tau_3 + \beta_2) + b(\tau_1 + \beta_3) + b(\tau_3 + \beta_3) \ ,$$

which is not same as the right hand side of (2.10) if τ_1 and τ_2 are interchanged, which is

$$b(\tau_1 + \beta_1) + b(\tau_2 + \beta_1) + b(\tau_1 + \beta_2) + b(\tau_3 + \beta_2) + b(\tau_2 + \beta_3) + b(\tau_3 + \beta_3) \ ,$$

unless $b(\cdot)$ is a linear function, which is not true since $b''(\cdot) > 0$.

One way to improve the above result is to use the principle of conditioning (see McCullagh and Nelder 1989). Since b_j, $j = 1, ..., q$ are sufficient for $\boldsymbol{\beta}$, conditioning on $\mathbf{b} = (b_1, ..., b_q)^T$ would make the conditional distribution of $\mathbf{y}$ independent of $\boldsymbol{\beta}$. Thus

$$f(\mathbf{y}; \boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) = f_1(\mathbf{t}|\mathbf{b}; \boldsymbol{\tau}, \sigma^2)\, f_2(\mathbf{b}; \boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2)\, g(\mathbf{y}; \sigma^2), \tag{2.14}$$

where f_1 is the conditional density of $\mathbf{t}$ given $\mathbf{b}$, f_2 is the marginal density of $\mathbf{b}$, and g is the condition density of $\mathbf{y}$ given $(\mathbf{t}, \mathbf{b})$ which is independent of $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$. We consider the prior of the following form, writing in terms of the product measure,

$$d\,\pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) = d\,\pi_1(\boldsymbol{\tau}|\sigma^2) \times d\,\pi_2(\boldsymbol{\beta}|\sigma^2) \times d\,\pi_3(\sigma^2), \tag{2.15}$$

where $\pi_1(\cdot|\sigma^2)$ is permutation invariant for every $\sigma^2 > 0$.

To illustrate how the conditioning principle works, consider the example of a normal linear model,

$$\mathbf{Y} \sim N_n(X_1\boldsymbol{\tau} + X_2\boldsymbol{\beta},\, \sigma^2).$$

In this case,

$$\mathbf{t} \;=\; X_1^T\mathbf{y}, \qquad \mathbf{b} = X_2^T\mathbf{y} \tag{2.16}$$

We assume without loss of generality that X_2 is of full rank. It can be seen that the conditional distribution of $\mathbf{t}$ given $\mathbf{b}$ is $N(M\,\boldsymbol{\tau} + X_1^TX_2(X_2^TX_2)^{-1}\mathbf{b},\, \sigma^2 M)$ and that the marginal distribution of $\mathbf{b}$ is $N(X_2^TX_1\boldsymbol{\tau} + X_2^TX_2\boldsymbol{\beta},\, \sigma^2X_2^TX_2)$, where

$$M = X_1^TX_1 - X_1^TX_2(X_2^TX_2)^{-1}X_2^TX_1. \tag{2.17}$$

Thus, by the factorization (2.14), we get

$$\begin{aligned} f_1(\mathbf{t}|\mathbf{b};\boldsymbol{\tau},\sigma^2) \;\propto\; & \exp\{-\frac{1}{2\sigma^2}(\mathbf{t} - X_1^TX_2(X_2^TX_2)^{-1}\mathbf{b} - M\,\boldsymbol{\tau})^T\overline{M} \\ & \times (\mathbf{t} - X_1^TX_2(X_2^TX_2)^{-1}\mathbf{b} - M\,\boldsymbol{\tau})\}, \end{aligned}$$

and

$$\begin{aligned} f_2(\mathbf{b};\boldsymbol{\tau},\boldsymbol{\beta},\sigma^2) \;\propto\; & \exp\{-\frac{1}{2\sigma^2}(\mathbf{b} - X_2^TX_1\boldsymbol{\tau} - X_2^TX_2\boldsymbol{\beta})^T(X_2^TX_2)^{-1} \\ & \times (\mathbf{b} - X_2^TX_1\boldsymbol{\tau} - X_2^TX_2\boldsymbol{\beta})\}, \end{aligned}$$

where $\overline{M}$ is a generalized inverse of M.

Since $\boldsymbol{\beta}$ is a nuisance parameter, the following uniform prior (improper) may be considered.

$$d\,\pi(\boldsymbol{\beta}|\sigma^2) \propto d\,\boldsymbol{\beta} \tag{2.18}$$

Thus, under prior (2.15), from (2.3) after integrating with respect to $\boldsymbol{\beta}$, we get

$$\begin{aligned} r_j(\mathbf{y}) \;=\; & \int L(\boldsymbol{\tau},\, j)\exp\{-\frac{1}{2\sigma^2}(\mathbf{t}_* - M\,\boldsymbol{\tau})^T\,\overline{M}\,(\mathbf{t}_* - M\,\boldsymbol{\tau})\} \\ & \times g_2(\mathbf{y};\sigma^2)\,d\,\pi_1(\boldsymbol{\tau}|\sigma^2)\,d\,\pi_3(\sigma^2), \end{aligned} \tag{2.19}$$

where $\mathbf{t}_* = \mathbf{t} - X_1^TX_2(X_2^TX_2)^{-1}\mathbf{b}$. Substituting $\mathbf{t}$ and $\mathbf{b}$ from (2.16), we get

$$\mathbf{t}_* = X_1^T\,Q_{X_2}\,\mathbf{y} \tag{2.20}$$

where $Q_{X_2} = I - X_2(X_2^T X_2)^{-1} X_2^T$ is the projection matrix orthogonal to the column space of X_2.

It can be seen that $r_{j_1}(\mathbf{y}) \geq r_{j_2}(\mathbf{y})$ if

$$f_1(\mathbf{t}|\mathbf{b}; \boldsymbol{\tau}, \sigma^2) \geq f_1(\mathbf{t}|\mathbf{b}; \boldsymbol{\tau}^{(j_1,j_2)}, \sigma^2) \tag{2.21}$$

whenever $\tau_{j_1} \geq \tau_{j_2}$ for $j_1 \neq j_2$. The latter holds (see Eaton 1967) if and only if M is of permutation symmetric structure, i.e., if M is of the form

$$M = \begin{pmatrix} a & b & \dots & b \\ b & a & \dots & b \\ . & . & \dots & . \\ b & b & \dots & a \end{pmatrix} \tag{2.22}$$

for some $a > 0$ and b such that M is a non-negative definite matrix. Moreover, if a design satisfies this condition, then the (2.21) holds if and only if $t^*_{j_1} \geq t^*_{j_2}$ for all $j_1 \neq j_2$, where t^*_j is the jth component of $\mathbf{t}_*$.

Thus, if a design satisfy the (2.22) condition, then the Bayes rule with respect to any permutation invariant prior of the form (2.15) with the uniform prior on $\boldsymbol{\beta}$ under any monotone invariant loss selects the best treatment in terms of the largest of the components of $\mathbf{t}_*$ with ties broken randomly, where $\mathbf{t}_*$ is defined by (2.20). Note that this is a much stronger result than the corresponding one in Bansal and Miescke (2002) if the loss due to estimation is ignored. In Bansal and Miescke (2002), the prior on both $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ was considered to be uniform (improper). Here, only the prior on $\boldsymbol{\beta}$ is considered to be uniform (improper), whereas the prior on $\boldsymbol{\tau}$ can be any permutation invariant prior.

Designs that satisfy (2.22) are called balanced designs which are covered in most of the design textbooks, as for example in Dey (1986). The incomplete block design (2.13) satisfies this condition. Note that for this design

$$X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad X_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix M for this design is of form (2.22) with $a = 1$, and $b = -\frac{1}{2}$. The vector

t_* for this design is given by

$$\mathbf{t}_* = \frac{1}{2}\begin{pmatrix} y_1 - y_2 + y_5 - y_6 \\ y_2 - y_1 + y_3 - y_4 \\ y_4 - y_3 + y_6 - y_5 \end{pmatrix}.$$

If, instead of the uniform prior, the prior $\pi_2(\cdot|\sigma^2)$ in (2.15) is $N(\boldsymbol{\mu}_0, \sigma^2\Sigma_0)$, as it could be the case when blocks are randomly selected from a normal population, then by integrating with respect to $\boldsymbol{\beta}$, we get from (2.3) that

$$\begin{aligned} r_j(\mathbf{y}) &= \int L(\boldsymbol{\tau},\, j)\exp\{-\frac{1}{2\sigma^2}(\mathbf{t}_* - M\,\boldsymbol{\tau})^T\,\overline{M}\,(\mathbf{t}_* - M\,\boldsymbol{\tau})\} \\ &\quad \times\exp\{-\frac{1}{2\sigma^2}(\mathbf{b} - X_2^T X_1\boldsymbol{\tau} - X_2^T X_2\boldsymbol{\mu}_0)^T(X_2^T X_2 +\, X_2^T X_2\Sigma_0 X_2^T X_2)^{-1} \\ &\qquad\qquad \times(\mathbf{b} - X_2^T X_1\boldsymbol{\tau} - X_2^T X_2\boldsymbol{\mu}_0)\} \\ &\quad \times g_3(\mathbf{y};\sigma^2)\,d\,\pi_1(\boldsymbol{\tau}|\sigma^2)\,d\,\pi_3(\sigma^2)\ , \end{aligned}$$

where $g_3(\cdot;\sigma^2)$ is some function of $\mathbf{y}$. It can be seen after some simplification that this is equivalent to

$$\begin{aligned} r_j(\mathbf{y}) &= \int L(\boldsymbol{\tau},\, j)\exp\{-\frac{1}{2\sigma^2}(\widetilde{\mathbf{t}}_* -\, \boldsymbol{\tau})^T\,V^{-1}\,(\widetilde{\mathbf{t}}_* -\, \boldsymbol{\tau})\} \\ &\qquad \times\,\widetilde{g}_3(\mathbf{y};\sigma^2)\,d\,\pi_1(\boldsymbol{\tau}|\sigma^2)\,d\,\pi_3(\sigma^2), \end{aligned} \tag{2.23}$$

where $\widetilde{g}_3(\cdot;\sigma^2)$ is some function of $\mathbf{y}$,

$$\begin{aligned} \widetilde{\mathbf{t}}_* &= V^{-1}(M\overline{M}\,\mathbf{t}_* + X_1^T X_2(X_2^T X_2 +\, X_2^T X_2\Sigma_0 X_2^T X_2)^{-1} \\ &\quad \times(\mathbf{b} - X_2^T X_2\boldsymbol{\mu}_0)), \quad \text{and} \\ V &= M + X_1^T X_2(X_2^T X_2 +\, X_2^T X_2\Sigma_0 X_2^T X_2)^{-1}X_2^T X_1. \end{aligned} \tag{2.24}$$

Simpler expressions of $\widetilde{\mathbf{t}}_*$ and V are presented below in (2.25) and their proofs are given in the Appendix.

As before, it can be seen again here that $r_{j_1}(\mathbf{y}) \geq r_{j_2}(\mathbf{y})$ if, for all $j_1 \neq j_2$,

$$\exp\{-\frac{1}{2\sigma^2}(\widetilde{\mathbf{t}}_* -\, \boldsymbol{\tau})^T\,V^{-1}\,(\widetilde{\mathbf{t}}_* -\, \boldsymbol{\tau})\} \geq \exp\{-\frac{1}{2\sigma^2}(\widetilde{\mathbf{t}} -\, \boldsymbol{\tau}^{(j_1,j_2)})^T\,V^{-1}\,(\widetilde{\mathbf{t}} -\, \boldsymbol{\tau}^{(j_1,j_2)})\}$$

whenever $\tau_{j_1} \geq \tau_{j_2}$. This holds if and only if V is of the same structure as M in (2.22), and if $\widetilde{t}^*_{j_1} \geq \widetilde{t}^*_{j_2}$ for all $j_1 \neq j_2$, where $\widetilde{t}^*_j$ is the jth component of $\widetilde{\mathbf{t}}_*$.

Thus, if a design is such that V is of the same structure as M in (2.22) then the Bayes rule with respect to any permutation invariant prior of the form (2.15), with $\pi_2(\cdot \mid \sigma^2)$ as $N(\mu_0, \sigma^2\Sigma_0)$, under any monotone invariant loss, selects the best treatment in terms of the largest component of $\widetilde{\mathbf{t}}_*$ with ties broken at random. The simpler expressions of V and $\widetilde{\mathbf{t}}_*$, as shown in the Appendix, are

$$\begin{aligned} V &= X_1^T X_1 - X_1^T X_2(\Sigma_0^{-1} + X_2^T X_2)^{-1} X_2^T X_1 \\ \widetilde{\mathbf{t}}_* &= X_1^T[I - X_2(\Sigma_0^{-1} + X_2^T X_2)^{-1} X_2^T](\mathbf{y} - X_2\boldsymbol{\mu}_0). \end{aligned} \tag{2.25}$$

It can be seen that if $\Sigma_0 = c\,I$ for some $c > 0$, then for the incomplete block design (2.13) V is of same structure as of M in (2.22), and in this case the Bayes rule selects in terms of the largest of $\widetilde{\mathbf{t}}_*$ with ties broken at random.

3. Bayes rules with respect to informative priors

A prior is informative for the selection problem if apriori at least one treatment is preferred over the other treatments. To construct such a prior, consider the prior on $\boldsymbol{\tau}$ depending on the parameter vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)^T$ denoted by $\pi_1(\cdot; \boldsymbol{\alpha})$. Suppose $\pi_1(\cdot; \boldsymbol{\alpha})$ is absolutely continuous with respect to a σ-finite measure $\nu(\cdot)$ which is permutation invariant. Let the ν-density of $\pi_1(\cdot; \boldsymbol{\alpha})$ be denoted by $f_{\pi_1}(\cdot; \boldsymbol{\alpha})$. Suppose $f_{\pi_1}(\cdot; \boldsymbol{\alpha})$ is such that for all $\alpha_{j_1} \geq \alpha_{j_2}$, $j_1 \neq j_2$,

$$f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha}) \geq f_{\pi_1}(\boldsymbol{\tau}^{(j_1,j_2)}; \boldsymbol{\alpha}) \text{ whenever } \tau_{j_1} \geq \tau_{j_2} \tag{3.1}$$

If $\tau_1, \tau_2, ..., \tau_p$ are independent with density, say, $\widetilde{f}_{\pi_1}(\cdot; \alpha_j)$ of τ_j, $j = 1, ..., p$, then it can be seen that the monotone likelihood ratio property of $\widetilde{f}_{\pi_1}(\cdot; \alpha)$ is equivalent to (3.1) for the joint density of $\tau_1, \tau_2, ..., \tau_p$. Also, note that, if $f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha}) = f_{\pi_1}(\rho(\boldsymbol{\tau}); \rho(\boldsymbol{\alpha}))$ for all permutations ρ, then (3.1) is equivalent to the property M of Eaton(1967), i.e.,

$$f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha}) \geq f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha}^{(j_1,j_2)}) \text{ whenever } \tau_{j_1} \geq \tau_{j_2} \text{ and } \alpha_{j_1} \geq \alpha_{j_2},\ j_1 \neq j_2. \tag{3.2}$$

We now demonstrate that densities that satisfy (3.1) allow, in a natural way, to convey the informative characteristics of the prior for the selection problem. In particular, we show that if $\alpha_{j_1} \geq \alpha_{j_2}$, then for any constant c,

$$P(\tau_{j_1} \geq c) \geq P(\tau_{j_2} \geq c), \tag{3.3}$$

i.e. that τ_{j_1} is stochastically larger than τ_{j_2}. Toward this end, we first prove the following lemma.

Lemma 3.1. *Suppose that $f(\boldsymbol{\tau};\boldsymbol{\alpha})$, a joint density of $\boldsymbol{\tau}$ with respect to a permutation symmetric σ-finite measure, satisfies (3.1), and let $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ be a monotonically increasing function. Then*

$$E[\psi(\tau_{j_1})] \geq E[\psi(\tau_{j_2})]$$

for all $j_1 \neq j_2$ with $\alpha_{j_1} \geq \alpha_{j_2}$.

Proof: For simplicity of notation, let us consider only the case of $j_1 = 1$ and $j_2 = 2$ with $\alpha_1 \geq \alpha_2$. The other cases of j_1 and j_2 follow analogously.

$$\begin{aligned} E[\psi(\tau_1) - \psi(\tau_2)] &= \int_{\tau_1 > \tau_2} [\psi(\tau_1) - \psi(\tau_2)]\, f(\boldsymbol{\tau};\boldsymbol{\alpha})\, d\,\nu(\boldsymbol{\tau}) \\ &\quad + \int_{\tau_1 < \tau_2} [\psi(\tau_1) - \psi(\tau_2)]\, f(\boldsymbol{\tau};\boldsymbol{\alpha})\, d\,\nu(\boldsymbol{\tau}). \end{aligned}$$

By interchanging τ_1 and τ_2 in the second integral, we see that

$$E[\psi(\tau_1) - \psi(\tau_2)] = \int_{\tau_1 > \tau_2} [\psi(\tau_1) - \psi(\tau_2)][\, f(\boldsymbol{\tau};\boldsymbol{\alpha}) \, - f(\boldsymbol{\tau}^{(1,2)};\boldsymbol{\alpha})] d\,\nu(\boldsymbol{\tau}).$$

The result now follows from the monotonicity of ψ and from (3.1). ■

If $\psi(\cdot) = I_{(c,\infty)}(\cdot)$, where $I_A(\cdot)$ denotes the indicator function, then $\psi(\cdot)$ is monotonically increasing. Thus, by Lemma 3.1, (3.3) holds for any $j_1 \neq j_2$ with $\alpha_{j_1} \geq \alpha_{j_2}$. This indicates that if the prior satisfies (3.1), then it can be used to describe the preference of one treatment over the other in terms of the corresponding parameters.

Another consequence of (3.1) is that if $\alpha_{j_1} \geq \alpha_{j_2}$, then

$$P(\tau_{j_1} \geq \tau_{j_2}) \geq \frac{1}{2}. \tag{3.4}$$

To see this, note that if $\alpha_1 \geq \alpha_2$, then from (3.1), using the permutation invariance of $\nu(\cdot)$, we get

$$\begin{aligned} P(\tau_1 > \tau_2) &= \int_{\tau_1 > \tau_2} f_{\pi_1}(\boldsymbol{\tau};\boldsymbol{\alpha})\, d\,\nu(\boldsymbol{\tau}) \\ &= \int_{\tau_1 < \tau_2} f_{\pi_1}(\boldsymbol{\tau}^{(1,2)};\alpha)\, d\,\nu(\boldsymbol{\tau}) \\ &\geq \int_{\tau_1 < \tau_2} f_{\pi_1}(\boldsymbol{\tau};\boldsymbol{\alpha})\, d\,\nu(\boldsymbol{\tau}) \\ &= P(\tau_1 < \tau_2). \end{aligned}$$

This implies that for $\alpha_1 \geq \alpha_2$, $P(\tau_1 \geq \tau_2) \geq P(\tau_1 < \tau_2)$ and thus (3.4) holds. By using the same argument, one can also see that if $\alpha_{j_*} = max\{\alpha_1, ..., \alpha_p\}$ then

$$P(\tau_{j_*} \geq \tau_j \text{ for all } j \neq j_*) \geq P(\tau_{j'} \geq \tau_j \text{ for all } j \neq j') \text{ for any } j' \neq j_*.$$

These quantities describe that the preference of one treatment over the others can be depicted by the corresponding parametric values. Note also that if $\alpha_1 = \cdots = \alpha_p$, then it can be seen, from (3.1), that for any permutation ρ, $f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha}) = f_{\pi_1}(\rho(\boldsymbol{\tau}); \boldsymbol{\alpha})$, i.e., $\pi_1(\cdot; \boldsymbol{\alpha})$ is permutation invariant and thus noninformative as it can be expected.

To construct priors that satisfies (3.1), we use the idea of exponential tilting (see Pace and Salvan 1997). To start, let $\pi_1^0(\cdot)$ be a permutation invariant measure of $\boldsymbol{\tau}$. Let

$$m(\boldsymbol{\alpha}) = \int \exp(\boldsymbol{\alpha}^T \boldsymbol{\tau}) d\, \pi_1^0(\boldsymbol{\tau}) \tag{3.5}$$

denote the moment generating function of $\pi_1^0(\cdot)$, and let $k(\boldsymbol{\alpha}) = \log m(\boldsymbol{\alpha})$ denote the cumulant generating function, defined over a space wherever it exists. Define the prior measure on $\boldsymbol{\tau}$ as

$$\pi_1(C; \boldsymbol{\alpha}) = \int_C \exp\{\boldsymbol{\alpha}^T \boldsymbol{\tau} - k(\boldsymbol{\alpha})\}\, d\, \pi_1^0(\boldsymbol{\tau}), \tag{3.6}$$

where C is a measurable set. If $\pi_1^0(\cdot)$ is absolutely continuous with respect to a permutation invariant σ-finite measure $\nu(\cdot)$, then $\pi_1(\cdot)$ defined above is also absolutely continuous with respect to $\nu\,(\cdot)$, with density

$$f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha}) = \exp\{\boldsymbol{\alpha}^T \boldsymbol{\tau} - k\,(\boldsymbol{\alpha})\} f^0(\boldsymbol{\tau}) \tag{3.7}$$

where $f^0(\cdot)$ is the ν-density of $\pi_1^0\,(\cdot)$.

Note that since $f^0(\boldsymbol{\tau})$ is permutation invariant in the components of $\boldsymbol{\tau}$, it is easy to see that the prior density $f_{\pi_1}(\boldsymbol{\tau}; \boldsymbol{\alpha})$ satisfies (3.1). Thus, as we discussed about the informative character of the densities that satisfy (3.1), (3.6) or (3.7) defines an informative prior in terms of the parametric values of $\boldsymbol{\alpha}$. It should also be noted, by taking $\alpha_1 = \alpha_2 = \cdots = \alpha_p$, that the prior density generated this way as in (3.7) for any $f^0(\cdot)$ satisfies (3.1) only if $f^0(\cdot)$ is permutation invariant. One can also see that (3.7) satisfies (3.2), since, from (3.5) by permuting the components of $\boldsymbol{\tau}$, $k(\cdot)$ is permutation invariant.

Now, we define a prior on Ω as (2.15) with a ν-density of $\pi_1(\cdot|\sigma^2)$ given by

$$f_{\pi_1}(\tau|\sigma^2) = \exp\{\frac{1}{\sigma^2}(\boldsymbol{\alpha}^T \boldsymbol{\tau} - k(\boldsymbol{\alpha})\} f^0(\boldsymbol{\tau}|\sigma^2), \tag{3.8}$$

where $f^0(\cdot|\sigma^2)$ is some permutation invariant ν-density for every $\sigma^2 > 0$.

Similarly as in Section 2, it can be seen that the Bayes rule under this prior is given by (2.4) with

$$\begin{aligned} r_j(\mathbf{y}) &= \int L(\boldsymbol{\tau}, j) \exp\{\frac{1}{\sigma^2}[\sum_{j=1}^{p}(t_j+\alpha_j)\,\tau_j + \sum_{j=1}^{q} b_j\,\beta_j - k(\boldsymbol{\alpha}) \\ &\quad - \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_i + \sum_{j=1}^{q} x_{i,p+j}\beta_j)] + \sum_{i=1}^{n} c(y_i,\sigma^2)\} \\ &\quad \times f^0(\boldsymbol{\tau}|\sigma^2)\, d\nu(\boldsymbol{\tau})\, d\pi_2(\boldsymbol{\beta}|\sigma^2) d\pi_3(\sigma^2). \end{aligned}$$

From this it can be seen, similarly as in Section 2, that $r_{j_1}(\mathbf{y}) \geq r_{j_2}(\mathbf{y})$ for any $j_1 \neq j_2$ if

$$f^*(\mathbf{y};\boldsymbol{\tau},\boldsymbol{\beta},\sigma^2) \geq f^*(\mathbf{y};\boldsymbol{\tau}^{(j_1,j_2)},\boldsymbol{\beta},\sigma^2)$$

whenever $\tau_{j_1} \geq \tau_{j_2}$, where

$$\begin{aligned} f^*(\mathbf{y};\boldsymbol{\tau},\boldsymbol{\beta},\sigma^2) &= \exp\{\frac{1}{\sigma^2}[\sum_{j=1}^{p}(t_j+\alpha_j)\,\tau_j + \sum_{j=1}^{q} b_j\,\beta_j - k(\boldsymbol{\alpha}) \\ &\quad - \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_i + \sum_{j=1}^{q} x_{i,p+j}\beta_j)] + \sum_{i=1}^{n} c(y_i,\sigma^2)\}. \end{aligned}$$

From this we get the following result.

Theorem 3.2. *If the design is such that for all $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ in the parameter space,*

$$\sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j) = \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j^{(j_1,j_2)} + \sum_{j=1}^{q} x_{i,p+j}\beta_j)$$

for all $j_1 \neq j_2$, where $\tau_j^{(j_1,j_2)}$ is the jth component of $\boldsymbol{\tau}^{(j_1,j_2)}$, then the Bayes rule with respect to prior (2.15) with density of $\pi_1(\cdot|\sigma^2)$ defined by (3.8), where $f^0(\cdot|\sigma^2)$ is any permutation invariant ν-density, under any monotone permutation invariant loss is

$$\phi_j^{*B}(\mathbf{y}) = \begin{cases} \left|N_*^B(\mathbf{y})\right|^{-1} & \text{if} \quad j \in N_*^B(\mathbf{y}) \\ 0 & \text{if} \quad j \notin N_*^B(\mathbf{y}) \end{cases}$$

where $N_*^B(\mathbf{y}) = \{k : t_k + \alpha_k = \max(t_1+\alpha_1, t_2+\alpha_2, ..., t_p+\alpha_p)\}$.

The principle of conditioning can also be applied here. For the normal linear model, if the prior of $\boldsymbol{\beta}$ is the uniform prior (improper), then the Bayes rule selects in terms of the largest of the components of $\mathbf{t}_* + \boldsymbol{\alpha}$ if the matrix $M = X_1^T X_1 - X_1^T X_2 (X_2^T X_2)^{-1} X_2^T X_1$ has the permutation symmetric structure (2.22). This generalizes the result of Bansal and Gupta (1997) where the prior $\pi_1(\cdot \mid \sigma^2)$ of $\boldsymbol{\tau}$ was considered to be normal. Similarly, if the prior $\pi_2(\cdot \mid \sigma^2)$ of $\boldsymbol{\beta}$ is $N(\boldsymbol{\mu}_0, \sigma^2 \Sigma_0)$, then the Bayes rule selects in terms of the largest of the components of $\tilde{\mathbf{t}}_* + \boldsymbol{\alpha}$ if the design is such that V defined in (2.25) is of same structure as M in (2.22).

It should be noted that the strength of these results lies in the generality of the prior on $\boldsymbol{\tau}$. For example, for the normal linear model there is no need to consider the normal prior on $\boldsymbol{\tau}$. Thus, one does not have to be too specific about the prior information on the components of $\boldsymbol{\tau}$.

4. Bayes rules for unbalanced designs

Sections 2 and 3 were devoted to the aspect of the designs so that Bayes rules can be obtained in generality in terms of the loss function and the prior. If the design is unbalanced, this generality is lost. When specific loss functions and specific priors are considered, the Bayes rules can be obtained from (2.1) and (2.2). The main difficulty, however, is the computation of the posterior risk. For example, if the loss function is the "0-1" loss, then the Bayes rule is given by $\phi^{0,B}(y) = (\phi_1^{0,B}(\mathbf{y}), ..., \phi_p^{0,B}(\mathbf{y}))$ with

$$\phi_j^{0,B}(\mathbf{y}) = \begin{cases} \left|N^{0,B}(\mathbf{y})\right|^{-1} & \text{if} \quad j \in N^{0,B}(\mathbf{y}) \\ 0 & \text{if} \quad j \notin N^{0,B}(\mathbf{y}) \end{cases}, \tag{4.1}$$

where

$$N^{0,B}(\mathbf{y}) = \{k : P(\tau_k = \tau_{[p]} | \mathbf{Y} = \mathbf{y}) = \max_{j=1,...,p} P(\tau_j = \tau_{[p]} | \mathbf{Y} = \mathbf{y})\}. \tag{4.2}$$

If the linear loss (1.8) is considered, then the Bayes rule is given by $\phi^{Lin,B}(\mathbf{y}) = (\phi_1^{Lin,B}(\mathbf{y}), ..., \phi_p^{Lin,B}(\mathbf{y}))$ with

$$\phi_j^{Lin,B}(\mathbf{y}) = \begin{cases} \left|N^{Lin,B}(\mathbf{y})\right|^{-1} & \text{if} \quad j \in N^{Lin,B}(\mathbf{y}) \\ 0 & \text{if} \quad j \notin N^{Lin,B}(\mathbf{y}) \end{cases}, \tag{4.3}$$

where

$$N^{Lin,B}(\mathbf{y}) = \{k : E(\tau_k | \mathbf{Y} = \mathbf{y}) = \max_{j=1,...,p} E(\tau_j | \mathbf{Y} = \mathbf{y})\}. \tag{4.4}$$

For the remainder of this section, the focus will be on the computational aspects of the posterior quantities such as in (4.2) and (4.4). The dispersion parameter σ^2 will be assumed to be known. We utilize here the conjugate prior from Diaconis and Ylvisaker (1979),

$$\pi_c(\boldsymbol{\tau}, \boldsymbol{\beta}; \boldsymbol{\alpha}_t, \boldsymbol{\alpha}_b, \gamma_0) \propto \exp\{\frac{\gamma_0}{\sigma^2}[\boldsymbol{\alpha}_t^T \boldsymbol{\tau} + \boldsymbol{\alpha}_b^T \boldsymbol{\beta} - \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_i + \sum_{j=1}^{q} x_{i,p+j}\beta_j)]\}, \quad (4.5)$$

where $\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_b$, and $\gamma_0 > 0$ are the hyperparameters of the prior. It can be seen, from (2.7) and (4.5) that the posterior density turns out to be

$$\tilde{\pi}_c(\boldsymbol{\tau}, \boldsymbol{\beta}; \frac{\gamma_0 \boldsymbol{\alpha}_t + \mathbf{t}}{1+\gamma_0}, \frac{\gamma_0 \boldsymbol{\alpha}_b + \mathbf{b}}{1+\gamma_0}, 1+\gamma_0). \quad (4.6)$$

In order to obtain the Bayes rules (4.1) and (4.3), one needs to compute the posterior probabilities in the (4.2) and the posterior expectation in the (4.4). The main difficulties here lie in the computations of these posterior quantities. One method of dealing with them is to use the Gibbs sampler (see Tanner 1996, Smith and Roberts 1993). We propose to use the Gibbs sampler algorithm of Meyer and Laud (2002) that is based on uniform auxiliary variables. Consider independent uniform random variables $U_1, U_2, ..., U_n$ such that the posterior density of $\mathbf{U} = (U_1, ..., U_n)^T, \boldsymbol{\tau}$ and $\boldsymbol{\beta}$ is proportional to

$$\begin{aligned} \prod_{i=1}^{n} I(0 \;<\; & u_i < \exp\{-\frac{1+\gamma_0}{\sigma^2} b(\sum_{j=1}^{p} x_{i,j}\tau_i + \sum_{j=1}^{q} x_{i,p+j}\beta_j)\}) \quad (4.7) \\ & \times \prod_{j=1}^{p} \exp\{\frac{1+\gamma_0}{\sigma^2}\tau_j \tilde{t}_j\} \prod_{j=1}^{q} \exp\{\frac{1+\gamma_0}{\sigma^2}\beta_j \tilde{b}_j\} \end{aligned}$$

with

$$\tilde{t}_j = \frac{\gamma_0 \alpha_j^t + t_j}{1+\gamma_0} \quad \text{and} \quad \tilde{b}_j = \frac{\gamma_0 \alpha_j^b + b_j}{1+\gamma_0},$$

where α_j^t is the jth component of $\boldsymbol{\alpha}_t$, and α_j^b is the jth component of $\boldsymbol{\alpha}_b$. It is easy to see that the marginal posterior density of $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ is (4.6). The conditional distribution of $U_i|\mathbf{U}_{-i}, \boldsymbol{\tau}, \boldsymbol{\beta}$ is uniform on $(0, \exp\{-\frac{1+\gamma_0}{\sigma^2} b(\sum_{j=1}^{p} x_{i,j}\tau_i + \sum_{j=1}^{q} x_{i,p+j}\beta_j)\})$, and $\tau_j|\boldsymbol{\tau}_{-j}, \boldsymbol{\beta}, \mathbf{U}$ and $\beta_j|\boldsymbol{\beta}_{-j}, \boldsymbol{\tau}, \mathbf{U}$ are truncated exponentials, where $\boldsymbol{\tau}_{-j}$ denotes the vector $\boldsymbol{\tau}$ excluding the j^{th}component, and $\boldsymbol{\beta}_{-j}$ and $\mathbf{U}_{-i}$

are defined analogously. The truncated bounds of the exponentials are obtained from the set

$$\{\boldsymbol{\tau} \in \mathbb{R}^p, \boldsymbol{\beta} \in \mathbb{R}^q : \quad b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j) < -\frac{\sigma^2}{1+\gamma_0} \log u_i, \ i = 1, ..., n\}. \tag{4.8}$$

The Gibbs sampling scheme can now be described as follows: Given the starting point $(\tau_1^0, ..., \tau_p^0, \beta_1^0, ..., \beta_q^0, u_1^0, ..., u_n^0)$, at the ith step,

(i) Generate $\tau_1^{(i+1)}$ from the truncated exponential with the bounds subject to (4.8), with all the $\tau_i s$ except τ_1, all the $\beta_j s$ and all the $u_i s$ substituted by the corresponding generated values from the ith step.

(ii) Generate $\tau_2^{(i+1)}$ from the truncated exponential with the bounds subject to (4.8) with τ_1 substituted by $\tau_1^{(i+1)}$, and all the other parameters except τ_2 substituted by the corresponding generated values from the ith step.

$\vdots$

(iii) Generate $\tau_p^{(i+1)}$ from the truncated exponential with the bounds subject to (4.8) with $\tau_1, ..., \tau_{p-1}$substituted by $\tau_1^{(i+1)}, ..., \tau_{p-1}^{(i+1)}$, and all the $\beta_j s$ and $u_i s$ substituted by the corresponding generated values from the ith step.

(iv) Generate $\beta_1^{(i+1)}$ from the truncated exponential with the bounds subject to (4.8) with all the $\tau_j s$ substituted by $\tau_j^{(i+1)} s$, all the $\beta_j s$ except β_1 and all the $u_i s$ substituted by the corresponding generated values from the ith step.

$\vdots$

(v) Generate $\beta_q^{(i+1)}$ from the truncated exponential with the bounds subject to (4.8) with all the $\tau_j s$, $\beta_1, ..., \beta_{q-1}$ substituted by $\tau_j^{(i+1)} s$ and $\beta_1^{(i+1)}, ..., \beta_{q-1}^{(i+1)}$ respectively, and all the $u_i s$ substituted by the corresponding generated values from the ith step.

(vi) Generate $u_1^{(i+1)}, u_2^{(i+1)}, ..., u_n^{(i+1)}$ from the independent uniform with the bounds for $u_m^{(i+1)}$, $(0, \exp\{-\frac{1+\gamma_0}{\sigma^2} b(\sum_{j=1}^{p} x_{m,j}\tau_j^{(i+1)} + \sum_{j=1}^{q} x_{m,p+j}\beta_j^{(i+1)})\}, m = 1, ..., n$.

Now, given the Gibbs sample $\{\boldsymbol{\tau}^{(i)} = (\tau_1^{(i)}, ..., \tau_p^{(i)})^T, i = 1, ..., B\}$, the posterior probabilities in (4.2) and posterior expectations in (4.4) can be approximated, respectively, by $\frac{1}{B}\sum_{i=1}^{B} I(\tau_j^{(i)} = \tau_{[p]}^{(i)})$, and $\frac{1}{B}\sum_{i=1}^{B} \tau_j^{(i)}$, $j = 1, ..., p$. In general, for a loss function, $L(\boldsymbol{\tau}, j)$, the posterior expected loss can be approximated by $\frac{1}{B}\sum_{i=1}^{B} L(\boldsymbol{\tau}^{(i)}, j)$.

Note that the Gibbs sampler can also be applied after first integrating out $\boldsymbol{\beta}$ of the posterior density. This can be done in a simple manner, if the sample size n is large, using the Laplace method as in Tierney and Kadane (1986). Based on this, the posterior marginal of $\boldsymbol{\tau}|\mathbf{Y} = \mathbf{y}$ can be approximated by

$$\begin{aligned}\hat{\pi}_c(\boldsymbol{\tau}|\mathbf{Y} &= \mathbf{y}) \propto (\det(\hat{\Sigma}(\boldsymbol{\tau})))^{-1/2} \exp\{\frac{1+\gamma_0}{\sigma^2}[\tilde{\mathbf{t}}^T\boldsymbol{\tau} + \tilde{\mathbf{b}}^T\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}) \\ &\quad - \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\hat{\beta}_j(\boldsymbol{\tau}))\},\end{aligned} \tag{4.9}$$

where $\tilde{\mathbf{t}} = (\tilde{t}_1, ..., \tilde{t}_p)^T$, $\tilde{\mathbf{b}} = (\tilde{b}_1, ..., \tilde{b}_q)^T$, and $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}) = (\hat{\beta}_1(\boldsymbol{\tau}), ..., \hat{\beta}_q(\boldsymbol{\tau}))^T$ is the solution of $\boldsymbol{\beta}$, for any fixed $\boldsymbol{\tau}$, of

$$\sum_{i=1}^{n} b'(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j)\, x_{i,p+j} = \tilde{b}_j,\ j = 1, ..., q \tag{4.10}$$

and

$$\hat{\Sigma}(\boldsymbol{\tau}) = \frac{1}{n}\sum_{i=1}^{n} b''(\sum_{j=1}^{p} x_{i,j}\tau_j + \sum_{j=1}^{q} x_{i,p+j}\beta_j)\, \mathbf{x}_i^{(2)}\mathbf{x}_i^{(2)T}, \tag{4.11}$$

where $\mathbf{x}_i^{(2)} = (x_{i,p+1}, ..., x_{i,p+q})^T$.

It should be noted that for the normal distribution this gives the exact solution, and the posterior distribution of $\boldsymbol{\tau}|\mathbf{Y} = \mathbf{y}$ turns out to be $N(\overline{M}M(\tilde{t} - X_1^T X_1(X_2^T X_2)^{-1}\tilde{b}), \frac{\sigma^2}{1+\gamma_0}M)$. From this the exact posterior probabilities in (4.2) and posterior expectations in (4.4) can be computed directly.

For non-normal distributions, the Gibbs sampler and the auxiliary uniform random variables can be used to find the posterior quantities. Note that (4.9) is proportional to the marginal of the joint density that is proportional to

$$\begin{aligned}\prod_{i=1}^{n} I(0 &< u_i < (\det(\hat{\Sigma}(\boldsymbol{\tau})))^{-1/2n} \exp\{-\frac{1+\gamma_0}{\sigma^2}[\frac{1}{n}\sum_{j=1}^{q}\tilde{b}_j\hat{\beta}_j(\boldsymbol{\tau}) \\ &\quad -b(\sum_{j=1}^{p} x_{i,j}\tau_i + \sum_{j=1}^{q} x_{i,p+j}\hat{\beta}_j(\boldsymbol{\tau})]\}\prod_{j=1}^{p}\exp\{\frac{1+\gamma_0}{\sigma^2}\,\tilde{t}_j\,\tau_j\}.\end{aligned}$$

The above discussion was confined to the conjugate prior (4.5). One problem with this prior is that it depends on the design. Thus, a suitable design must be used so that (4.5) gives a suitable informative prior for the components of $\boldsymbol{\tau}$. However, as shown in Meyer and Laud (2002), other priors such as normal priors can also be used while the Gibbs sampling scheme as described above can still be performed.

5. Appendix

The derivations of V and $\widetilde{\mathbf{t}}_*$ in (2.25) are as follows**.** From (2.24) we get

$$\begin{aligned} V &= X_1^T X_1 - X_1^T X_2 (X_2^T X_2)^{-1} X_2^T X_1 \\ &\quad + X_1^T X_2 (X_2^T X_2 + X_2^T X_2 \Sigma_0 X_2^T X_2)^{-1} X_2^T X_1 \\ &= X_1^T X_1 - X_1^T X_2 (X_2^T X_2)^{-1} [X_2^T X_2 - ((X_2^T X_2)^{-1} + \Sigma_0)^{-1}] \\ &\quad \times (X_2^T X_2)^{-1} X_2^T X_1 \\ &= X_1^T X_1 - X_1^T X_2 [(X_2^T X_2)^{-1} + \Sigma_0 - (X_2^T X_2)^{-1}] \\ &\quad \times ((X_2^T X_2)^{-1} + \Sigma_0)^{-1} (X_2^T X_2)^{-1} X_2^T X_1 \\ &= X_1^T X_1 - X_1^T X_2 (\Sigma_0^{-1} + X_2^T X_2)^{-1} X_2^T X_1 \end{aligned}$$

and

$$\widetilde{\mathbf{t}}_* = V^{-1}(M\,\overline{M}\,\mathbf{t}_* + X_1^T X_2 (X_2^T X_2 + X_2^T X_2 \Sigma_0 X_2^T X_2)^{-1} (\mathbf{b} - X_2^T X_2 \boldsymbol{\mu}_0). \tag{5.1}$$

Note that, from (2.20) and (2.17),

$$M\,\overline{M}\,\mathbf{t}_* = (X_1^T Q_{X_2} X_1)(X_1^T Q_{X_2} X_1)^- X_1^T Q_{X_2} \mathbf{y}. \tag{5.2}$$

Now, it can be seen that

$$\begin{aligned} H &= (X_1^T Q_{X_2} X_1)(X_1^T Q_{X_2} X_1)^- X_1^T Q_{X_2} - X_1^T Q_{X_2} \\ &= \mathbf{0} \end{aligned}$$

by noticing that $HH^T = \mathbf{0}$. Here $(X_1^T Q_{X_2} X_1)^-$ can be taken to be a symmetric generalized inverse of $X_1^T Q_{X_2} X_1$. Thus, from (5.1) and (5.2),

$$\begin{aligned}
\widetilde{\mathbf{t}}_* &= X_1^T Q_{X_2}\mathbf{y} + X_1^T X_2 (X_2^T X_2)^{-1}((X_2^T X_2)^{-1} + \Sigma_0)^{-1}(X_2^T X_2)^{-1} X_2^T \\
&\quad \times(\mathbf{y} - X_2\boldsymbol{\mu}_0) \\
&= X_1^T Q_{X_2}\mathbf{y} + X_1^T X_2 (X_2^T X_2)^{-1} X_2^T (\mathbf{y} - X_2\boldsymbol{\mu}_0) \\
&\quad - X_1^T X_2 \Sigma_0((X_2^T X_2)^{-1} + \Sigma_0)^{-1}(X_2^T X_2)^{-1} X_2^T (\mathbf{y} - X_2\boldsymbol{\mu}_0) \\
&= X_1^T(\mathbf{y} - X_2\boldsymbol{\mu}_0) - X_1^T X_2 (\Sigma_0^{-1} + X_2^T X_2)^{-1} X_2^T (\mathbf{y} - X_2\boldsymbol{\mu}_0) \\
&= X_1^T[I - X_2(\Sigma_0^{-1} + X_2^T X_2)^{-1} X_2^T](\mathbf{y} - X_2\boldsymbol{\mu}_0)
\end{aligned}$$

REFERENCES

Abughalous, M.M., and Bansal, N.K. (1995). "On selecting the best natural exponential families with quadratic variance function," *Statist. Probab. Lett.,* 25, 341-349.

Abughalous, M.M., and Miescke, K.J. (1989). "On selecting the largest success probability under unequal sample sizes," *J. Statist. Plan. Inference*, 21, 53-68.

Bahadur, R.R. (1950). "On the problem in the theory of k populations," *Ann. Math. Statist.*, 20, 362-375.

Bahadur, R.R., and Goodman, L.A. (1952). "Impartial decision rules and sufficient statistics," *Ann. Math. Statist.*, 23, 553-562.

Bansal, N.K., and Gupta, S. (1997). "On the natural rule in general linear models," *Metrika* 46, 59-69.

Bansal, N.K., and Gupta, S. (2000). "On the role of the natural decision rule in ranking and selecting the best component in multivariate populations," in Proceedings of the National Seminar on Bayesian Statistics and its Applications, eds. S.K. Upadhyay and U. Singh, Banaras Hindu University, Varanasi, 49-57.

Bansal, N.K., and Miescke, K.J. (2002). "Simultaneous selection and estimation in general linear models," *J. Statist. Plan. Inference,* 104, 377-390.

Bansal, N.K., and Misra, N.(1999). "On ranking treatment effects in a general linear model," *J. Statist. Plan. Inference,* 78, 1-11.

Bechhofer, R.E. (1954). "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Ann. Math. Statist.*, 25, 16-39.

Bechhofer, R.E., Santner, T.J., and Goldsman, D.M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons.* New York: Wiley.

Dey, A. (1986). *Theory of Block Designs.* New York: Wiley.

Diaconis, P., and Ylvisaker, S. (1979). "Conjugate priors for exponential families," *Ann. Statist.*, 7, 269-281.

Eaton, M.L. (1967). "Some optimum properties of ranking procedures," *Ann. Statist.,* 38, 124-137.

Gupta, S.S. (1956). "On a decision rule for a problem in ranking means," Mimeo, Ser. 150, Institute of Statistics, University of North Carolina, Chapel Hill, NC.

Gupta, S.S., and Miescke, K.J. (1984). "On two-stage Bayes selection procedures," *Sankhyā Ser. B*, 46, 123-134.

Gupta, S.S., and Miescke, K.J. (1984). "On the problem of finding the largest normal mean under heteroscedasticity*,"* in *Statistical Decision Theory and Related Topics IV, Vol. 2*, eds. S.S. Gupta and J.O. Berger, New York: Springer, 37-49.

Gupta, S.S., and Panchapakesan, S. (1996). "Design of experiments with selection and ranking goals," in *Handbook of Statistics: Design and Analysis of Experiments,* Vol. 13, eds. S. Ghosh and C.R. Rao, Amsterdam: Elsevier, 555-585.

Lehmann, E.L. (1966). "On a theorem of Bahadur and Goodman," *Ann. Statist.*, 37, 1-6.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Meyer, M.C., and Laud, P.W. (2002). "Predictive variable selection in generalized linear models," *J. Amer. Statist. Assoc.*, 97, 859-871.

Pace, L., and Salvan, S. (1997). *Principles of Statistical Inference*, Singapore: World Scientific.

Smith, A.F.M., and Roberts, G.O. (1993). "Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods," *J. Amer. Statist. Assoc.*, 55, 3-23.

Tanner, M.A. (1996). *Tools for Statistical Inference* (3rd ed.), New York: Springer.

Tierney, L., and Kadane, J.B. (1986). "Accurate approximations for posterior moments and marginal densities," *J. Amer. Statist. Assoc.*, 81, 82-86.