The Korean Communications in Statistics Vol. 15, No. 1, 2008, pp. 13–26

Multiple Comparison for the One-Way ANOVA with the Power Prior†

Rena Bae[1], Yunhee Kang[1], Minyoung Hong[1], Seong W. Kim[2]

Abstract

Inference on the present data will be more reliable when the data arising from previous similar studies are available. The data arising from previous studies are referred as historical data. The power prior is defined by the likelihood function based on the historical data to the power a_0, where $0 \leq a_0 \leq 1$. The power prior is a useful informative prior for Bayesian inference such as model selection and model comparison. We utilize the historical data to perform multiple comparison in the one-way ANOVA model. We demonstrate our results with some simulated datasets under a simple order restriction between the treatments.

Keywords: Bayes factor; historical data; Markov Chain Monte Carlo; order restricted inference; power prior.

1. Introduction

When we compare several treatment means, most of methodologies only conclude that either all means are same or at least one mean is different from other means. For instance, when the one-way ANOVA model is used, small P-values indicate that the treatment means are not equal. However, it is more interesting to explore inherent relations between the treatment means. Multiple comparison is one of useful statistical techniques to detect these relations. In some applications like carcinogenicity or dose response studies, researchers may believe that there exists a certain order restriction.

Order restricted inference should be employed when treatment means are assumed to be ordered. In a dose response model, one might believe that the mean response is nondecreasing or nonincreasing as the dose level increases. That is, we may assume

† This work was supported by Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund)(KRF-2006-003-C00050).

1) Graduate Student, Division of Applied Mathematics, Hanyang University, 1271 Sa 1-Dong, Sangnok-Gu, Ansan 426-791, Korea.

2) Professor, Division of Applied Mathematics, Hanyang University, 1271 Sa 1-Dong, Sangnok-Gu, Ansan 426-791, Korea. Correspondence: seong@hanyang.ac.kr

that $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$, where μ_i is the treatment mean for $i = 1, \ldots, k$. This is called a simple order restriction. When a tree ordering is imposed, we assume that $\mu_0 \leq \mu_1, \ldots, \mu_k$, where μ_0 is the mean for the control group. In this case, the order of treatment means is unknown.

It is known that frequentist approaches for multiple comparison have been well established under the simple-order assumption. There are two likelihood ratio tests such as the $\bar{\chi}^2$ test and the $\bar{E}^2$ test of Bartholomew (1959a, 1959b, 1961a, 1961b) to test the homogeneity of means against ordered alternatives. Bohrer and Francis (1972) found some useful results on simultaneous bounds for restricted contrasts. For pairwise comparison in the simple-order, Hayter (1990) performed the single-sided Studentized-range test. Liu (2001) developed the single sided the multiple comparison procedure for the case of a simple-order with known variance.

A considerable amount of work has been developed for order restricted inference in Bayesian perspectives. Gelfand *et al.* (1990) computed Bayes estimates of ordered normal means with any variance. Gopalan and Berry (1998) proposed both-sided multiple comparison procedures with Dirichlet-process prior. Pauler *et al.* (1999) utilized Bayes factors to test hypotheses with inequality constraints. In life testing models, Kim and Sun (2000) made use of intrinsic priors and Bayes factors in multiple comparison for exponential means. Kim and Kim (2001) performed a multiple test for multivariate normal observations. Son and Kim (2005) considered multiple comparison problems with a single change point.

Prior elicitation perhaps plays a very important role in Bayesian inference. In principle, priors formally represent available information but in practice noninformative and improper priors are often used. Nevertheless, they cannot be used in some situations such as model selection or hypothesis testing (*cf.* Berger and Pericchi, 1996). In these cases a proper prior on the parameters is needed making Bayesian inference plausible. Moreover, noninformative priors do not make use of real prior information that one may need for a specific situation. Thus, when we have real prior information, it is possible to make posterior inference quite accurate. This often occurs when the current study is similar to the previous study in measuring the response and covariates.

The data arising from previous studies are referred as 'historical data'. In carcinogenicity studies, for example, large historical databases exist for the control animals from previous experiments. In all of these situations, it is natural to incorporate the historical data into the current study by quantifying it with a suitable prior distribution on the model parameters. One method of constructing an informative prior based on the historical data is the power prior of Ibrahim and Chen (2000). The power prior is defined by the likelihood function based on the historical data, raised to a power a_0, where a_0 $(0 \leq a_0 \leq 1)$ is a scalar parameter that controls the influence of the historical data on the current study.

In this article we incorporate the historical data to perform multiple comparison in the one-way ANOVA model. The rest of this article is organized as follows. In Section 2, we propose the model and the prior distributions and derive full conditional distributions. In Section 3, we present computation schemes with Markov Chain Monte Carlo (MCMC) methodologies. We also explain the concept of the Bayes factor as a model selection criterion. In Section 4, we show numerical results with some simulated datasets. Finally, we provide concluding remarks in Section 5.

2. Methodologies

2.1. The model and hypotheses

The conventional one-way fixed ANOVA model can be characterized as

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad \text{for } i = 1, 2, \ldots, k, \ j = 1, 2, \ldots, m, \tag{2.1}$$

where y_{ij} denotes the observed response on the j^{th} subject under the i^{th} treatment, μ_i is a fixed treatment effect, and ϵ_{ij} is an error term. Note that y_{ij}, μ_i and ϵ_{ij} are all scalars and the overall sample size is $m \times k$.

We assume that the ϵ_{ij}'s are independent and normally distributed with mean 0 and variance σ^2. We rewrite the model in (2.1) with a more concise format. That is,

$$y_j = X\mu^* + \epsilon_j, \ j = 1, 2, \ldots, m, \tag{2.2}$$

where y_j, μ^* and ϵ_j are vectors and X is a $k \times k$ identity matrix. Specifically, y_j denotes a $k \times 1$ vector of observed responses on the j^{th} subject, μ^* is the $k \times 1$ vector with each component being μ_i for $i = 1, \ldots, k$, and ϵ_j is a $k \times 1$ error vector distributed as normal distribution with mean vector 0 and variance-covariance matrix $\sigma^2 I$.

Since k normal means are assumed to be under a simple order, we stay the smallest mean with itself and reparameterize remaining each mean with the difference between the preceding mean and itself. Thus, the second mean will be denoted as $\mu_1 + \delta_1$ and the i^{th} mean will be denoted as $\mu_1 + \delta_1 + \delta_2 + \cdots + \delta_{i-1}$. Therefore, for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, m$, the distribution of the individual response observation y_{ij} will satisfy the followings:

$$\begin{aligned}
y_{1j}|\mu_1, \sigma^2 &\overset{iid}{\sim} N(\mu_1, \sigma^2), \\
y_{2j}|\mu_1, \sigma^2, \delta_1 &\overset{iid}{\sim} N(\mu_1 + \delta_1, \sigma^2), \\
&\ \vdots \\
y_{kj}|\mu_1, \sigma^2, \delta_1, \cdots, \delta_{k-1} &\overset{iid}{\sim} N(\mu_1 + \delta_1 + \delta_2 + \cdots + \delta_{k-1}, \sigma^2).
\end{aligned}$$

In this article we consider the hypotheses for successive pairwise comparisons of the means in the model, $M_{0i} : \mu_i = \mu_i + \delta_i$ versus $M_{1i} : \mu_i < \mu_i + \delta_i$ for $i = 1, 2, \ldots, k-1$. If the null hypothesis is rejected, then it is concluded that there are significant differences in the population means.

2.2. Prior specifications

We consider the power prior of Ibrahim and Chen (2000). Let $y = (y_1, \ldots, y_n)$ be the response observation data from the current study. Suppose we have historical data $y_0 = (y_{01}, \ldots, y_{0n})$ from similar previous study. Let $L(\theta|y)$ denote the likelihood function for the current study, where θ is a vector of parameters. Further, let $\pi_0(\theta|\cdot)$ denote the prior distribution for θ, which is called the initial prior. This initial prior is assumed before the historical data y_0 is observed. We define the joint power prior distribution of (θ, a_0) for the current study as

$$\pi(\theta, a_0|y_0) \propto [L(\theta|y_0)]^{a_0}\pi_0(\theta|c_0)\pi(a_0|\gamma_0), \tag{2.3}$$

where c_0 is a specified hyperparameter for the initial prior and γ_0 is a specified hyperparameter for the prior distribution of a_0. The parameter c_0 controls the impact of the initial prior $\pi_0(\theta|c_0)$, and the parameter a_0 is a precision parameter for the historical data. The parameter a_0 controls heaviness of the tails of the prior for θ. As a_0 becomes smaller, the tails of (2.3) become heavier. Such control may be important when there is heterogeneity between the previous and the current study or the sample sizes of two studies are quite different. It is reasonable that the range of a_0 is restricted to be between 0 and 1, and thus it is natural that the distribution for $\pi(a_0|\gamma_0)$ is chosen to be a beta distribution. That is, $\pi(a_0) \propto a_0^{\eta_0-1}(1-a_0)^{\lambda_0-1}$ for $0 \le a_0 \le 1$.

We describe the initial priors for μ_1, δ_i, σ^2. We assume a normal prior for μ_1 with mean μ_0 and variance σ_0^2 and an inverse gamma prior for σ^2 with hyperparameters a and b. Note that these priors are (marginal) conjugate priors for μ_1 and σ^2 respectively. Since δ_i has either zero or a positive value, we assume a mixture of an exponential distribution and a discrete distribution with its entire mass at $\delta_i = 0$. So, the prior for δ_i is

$$p(\delta_i) = \begin{cases} \rho_i, & \text{if } \delta_i = 0, \\ (1-\rho_i)\frac{1}{\xi}\exp\left\{-\frac{\delta_i}{\xi}\right\}, & \text{if } \delta_i > 0, \\ 0, & \text{if } \delta_i < 0, \end{cases} \tag{2.4}$$

where ξ is assumed to be a known hyperparameter. Finally, we assume a beta distribution for ρ_i with hyperparameter α_0 and β_o.

2.3. Full conditional distributions

Let $Y_0 = (y'_{01}, \ldots, y'_{0m_0})'$ be the historical data from the similar study. Let

$$\begin{cases} \delta_0 = 0, & Y = (y'_1, \ldots, y'_m)', \\ \bar{y}_i = \displaystyle\sum_{j=1}^{m} \frac{y_{ij}}{m}, & \bar{y}_{0i} = \displaystyle\sum_{j=1}^{m_0} \frac{y_{0ij}}{m_0}, \\ s = \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{m}\left(y_{ij} - \mu_1 - \sum_{l=0}^{i-1}\delta_l\right)^2, & s_0 = \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{m_0}\left(y_{0ij} - \mu_1 - \sum_{l=0}^{i-1}\delta_l\right)^2, \\ q = \mu_1 + \delta_1 + \delta_2 + \cdots + \delta_{k-1}. \end{cases}$$

Let $\theta = (\mu_1, \sigma^2, \{\delta_i\}, \{\rho_i\}, a_0)$. The joint posterior with the power prior is then

$$\begin{aligned}
p(\theta|Y, Y_0) \propto & \left[\prod_{j=1}^{m}(2\pi)^{-k/2}|\sigma^2 I|^{-1/2}\exp\left\{-\frac{1}{2}(y_j - X\mu^*)'(\sigma^2 I)^{-1}(y_j - X\mu^*)\right\}\right] \\
& \times \left[\prod_{j=1}^{m_0}(2\pi)^{-k/2}|\sigma^2 I|^{-1/2}\exp\left\{-\frac{1}{2}(y_{0j} - X\mu^*)'(\sigma^2 I)^{-1}(y_{0j} - X\mu^*)\right\}\right]^{a_0} \\
& \times \frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left\{-\frac{(\mu_1-\mu_0)^2}{2\sigma_0^2}\right\}\frac{1}{\Gamma(a)b^a}\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left\{-\frac{1}{b\sigma^2}\right\} \\
& \times \prod_{i=1}^{k-1}\left[\rho_i I_{\{\delta_i=0\}} + (1-\rho_i)\frac{1}{\xi}\exp\left\{-\frac{\delta_i}{\xi}\right\}I_{\{\delta_i>0\}}\right] \\
& \times \prod_{i=1}^{k-1}\left[\frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)}\rho_i^{\alpha_0-1}(1-\rho_i)^{\beta_0-1}\right]a_0^{\eta_0-1}(1-a_0)^{\lambda_0-1},
\end{aligned}$$

where

$$X\mu^* = \begin{pmatrix} \mu_1 \\ \mu_1 + \delta_1 \\ \vdots \\ \mu_1 + \delta_1 + \cdots + \delta_{k-1} \end{pmatrix}.$$

In order to use the MCMC methods, we need the following full conditional distributions of each parameter.

1. The full conditional distribution of μ_1

Since

$$\begin{aligned}
& p(\mu_1|Y, \sigma^2, \{\delta_i\}, \rho_i, a_0) \\
& \propto \exp\left\{-\frac{s}{2\sigma^2}\right\}\exp\left\{-\frac{a_0 s_0}{2\sigma^2}\right\}\exp\left\{-\frac{(\mu_1-\mu_0)^2}{2\sigma_0^2}\right\} \\
& = \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{1j}-\mu_1)^2\right\}\cdots\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{kj}-q)^2\right\} \\
& \times \exp\left\{-\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(y_{01j}-\mu_1)^2\right\}\cdots\exp\left\{-\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(y_{0kj}-q)^2\right\}\exp\left\{-\frac{(\mu_1-\mu_0)^2}{2\sigma_0^2}\right\} \\
& \propto \exp\left[-\frac{\mu_1^2}{2}\left\{\frac{km}{\sigma^2}+\frac{ka_0m_0}{\sigma^2}+\frac{1}{\sigma_0^2}\right\}+\mu_1\left\{\frac{m}{\sigma^2}\left(\sum_{i=1}^{k}\bar{y}_i-\sum_{l=1}^{k-1}(k-l)\delta_l\right)\right.\right. \\
& \left.\left.+\frac{a_0m_0}{\sigma^2}\left(\sum_{i=1}^{k}\bar{y}_{0i}-\sum_{l=1}^{k-1}(k-l)\delta_l\right)+\frac{\mu_0}{\sigma_0^2}\right\}\right],
\end{aligned}$$

the full conditional distribution of μ_1 follows a normal distribution with mean u/v and variance $1/v$, where

$$u = \frac{m}{\sigma^2}\left[\sum_{i=1}^{k}\bar{y}_i - \sum_{l=1}^{k-1}(k-l)\delta_l\right] + \frac{a_0 m_0}{\sigma^2}\left[\sum_{i=1}^{k}\bar{y}_{0i} - \sum_{l=1}^{k-1}(k-l)\delta_l\right] + \frac{\mu_0}{\sigma_0^2}$$

and

$$v = \frac{km}{\sigma^2} + \frac{ka_0 m_0}{\sigma^2} + \frac{1}{\sigma_0^2}.$$

2. The full conditional distribution of σ^2

Since

$$\begin{aligned}
& p(\sigma^2|Y, \mu_1, \{\delta_i\}, \{\rho_i\}, a_0) \\
& \propto \frac{1}{(\sigma^2)^{\frac{km}{2}+\frac{ka_0 m_0}{2}+a+1}} \exp\left\{-\frac{s}{2\sigma^2}\right\} \exp\left\{-\frac{a_0 s_0}{2\sigma^2}\right\} \exp\left\{-\frac{1}{b\sigma^2}\right\} \\
& = \frac{1}{(\sigma^2)^{\frac{km}{2}+\frac{ka_0 m_0}{2}+a+1}} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{s}{2} + \frac{a_0 s_0}{2} + \frac{1}{b}\right]\right\},
\end{aligned}$$

the full conditional distribution of σ^2 is

$$p(\sigma^2|Y, \mu_1, \{\delta_i\}, \{\rho_i\}, a_0) \sim IG\left(\frac{km}{2} + \frac{ka_0 m_0}{2} + a, \left[\frac{s}{2} + \frac{a_0 s_0}{2} + \frac{1}{b}\right]^{-1}\right).$$

3. The full conditional distribution of δ_i

Note that

$$\begin{aligned}
& p(\delta_1|Y, \mu_1, \sigma^2, \{\delta_i, i = 2, 3, \ldots, k-1\}, \rho_1, a_0) \\
& \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{1j} - \mu_1)^2\right\} \cdots \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{kj} - q)^2\right\} \\
& \times \exp\left\{-\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(y_{01j} - \mu_1)^2\right\} \cdots \exp\left\{-\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(y_{0kj} - q)^2\right\} \\
& \times \left(\rho_1 I_{\{\delta_1=0\}} + (1-\rho_1)\frac{1}{\xi}\exp\left\{-\frac{\delta_1}{\xi}\right\} I_{\{\delta_1>0\}}\right).
\end{aligned}$$

Consider $\delta_1 > 0$. Then we have

$$
\begin{aligned}
&p(\delta_1|Y, \mu_1, \sigma^2, \{\delta_i, i=2,3,\ldots,k-1\}, \rho_1, a_0) \\
&\propto (1-\rho_1)\frac{1}{\xi}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(\delta_1-(y_{2j}-\mu_1))^2\right\}\cdots \\
&\quad\times\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(\delta_1-(y_{kj}-q+\delta_1))^2\right\}\exp\left\{-\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(\delta_1-(y_{02j}-\mu_1))^2\right\}\cdots \\
&\quad\times\exp\left\{-\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(\delta_1-(y_{0kj}-q+\delta_1))^2\right\}\exp\left\{-\frac{\delta_1}{\xi}\right\} \\
&=(1-\rho_1)\frac{1}{\xi}\exp\left\{\delta_1^2\left(-\frac{(k-1)(m+a_0m_0)}{2\sigma^2}\right)\right\} \\
&\quad\times\exp\left\{2\delta_1\left(\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{2j}-\mu_1)+\cdots+\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{kj}-q+\delta_1)\right.\right. \\
&\quad\left.\left.+\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(y_{02j}-\mu_1)+\cdots+\frac{a_0}{2\sigma^2}\sum_{j=1}^{m_0}(y_{0kj}-q+\delta_1)-\frac{1}{2\xi}\right)\right\} \\
&=(1-\rho_1)\frac{1}{\xi}\exp\left\{-\frac{(k-1)(m+a_0m_0)}{2\sigma^2}\left[\delta_1^2-\frac{2\sigma^2}{(k-1)(m+a_0m_0)}\left(\frac{m}{\sigma^2}(\overline{y}_2-\mu_1)\right.\right.\right. \\
&\quad\left.\left.\left.+\cdots+\frac{m}{\sigma^2}(\overline{y}_k-q+\delta_1)+\frac{a_0m_0}{\sigma^2}(\overline{y}_{02}-\mu_1)+\cdots+\frac{a_0m_0}{\sigma^2}(\overline{y}_{0k}-q+\delta_1)-\frac{1}{\xi}\right)\delta_1\right]\right\} \\
&=(1-\rho_1)\frac{1}{\xi}\exp\left\{-\frac{(k-1)(m+a_0m_0)}{2\sigma^2}\left[\delta_1-\frac{\sigma^2}{(k-1)(m+a_0m_0)}\left(\frac{m}{\sigma^2}(\overline{y}_2-\mu_1)\right.\right.\right. \\
&\quad\left.\left.\left.+\cdots+\frac{m}{\sigma^2}(\overline{y}_k-q+\delta_1)+\frac{a_0m_0}{\sigma^2}(\overline{y}_{02}-\mu_1)+\cdots+\frac{a_0m_0}{\sigma^2}(\overline{y}_{0k}-q+\delta_1)-\frac{1}{\xi}\right)\right]^2\right\} \\
&\quad\times\exp\left\{\frac{\sigma^2}{2(k-1)(m+a_0m_0)}\left(\frac{m}{\sigma^2}(\overline{y}_2-\mu_1)+\cdots+\frac{m}{\sigma^2}(\overline{y}_k-q+\delta_1)\right.\right. \\
&\quad\left.\left.+\frac{a_0m_0}{\sigma^2}(\overline{y}_{02}-\mu_1)+\cdots+\frac{a_0m_0}{\sigma^2}(\overline{y}_{0k}-q+\delta_1)-\frac{1}{\xi}\right)^2\right\}.
\end{aligned}
$$

If $\delta_1 = 0$, the prior probability of δ_1 is ρ_1 and the term $1/\xi$ drops out. We apply the identical method for $\delta_2 > 0$. Thus,

$$
\begin{aligned}
&p(\delta_2|Y,\mu_1,\sigma^2,\{\delta_i, i=1,3,\ldots,k-1\},\rho_2,a_0)\\
&\propto (1-\rho_2)\frac{1}{\xi}\exp\left\{-\frac{(k-2)(m+a_0m_0)}{2\sigma^2}\left[\delta_2-\frac{\sigma^2}{(k-2)(m+a_0m_0)}\left(\frac{m}{\sigma^2}(\bar{y}_3-\mu_1-\delta_1)\right.\right.\right.\\
&\quad\left.\left.\left.+\cdots+\frac{m}{\sigma^2}(\bar{y}_k-q+\delta_2)+\frac{a_0m_0}{\sigma^2}(\bar{y}_{03}-\mu_1-\delta_1)+\cdots+\frac{a_0m_0}{\sigma^2}(\bar{y}_{0k}-q+\delta_2)-\frac{1}{\xi}\right)\right]^2\right\}\\
&\quad\times\exp\left\{\frac{\sigma^2}{2(k-2)(m+a_0m_0)}\left(\frac{m}{\sigma^2}(\bar{y}_3-\mu_1-\delta_1)+\cdots+\frac{m}{\sigma^2}(\bar{y}_k-q+\delta_2)\right.\right.\\
&\quad\left.\left.+\frac{a_0m_0}{\sigma^2}(\bar{y}_{03}-\mu_1-\delta_1)+\cdots+\frac{a_0m_0}{\sigma^2}(\bar{y}_{0k}-q+\delta_2)-\frac{1}{\xi}\right)^2\right\}.
\end{aligned}
$$

If $\delta_2=0$, the prior probability of δ_2 is ρ_2 and $1/\xi$ drops out. Therefore, the full conditional distribution of δ_i can be expressed by a mixture of a discrete part and a continuous part. So,

$$
p(\delta_i|\cdot)=\begin{cases} c\rho_i h(\delta_i) & \text{if } \delta_i=0\\ c(1-\rho_i)\frac{1}{\xi}h(\delta_i) & \text{if } \delta_i>0\\ 0 & \text{if } \delta_i<0,\end{cases}
$$

where

$$
\begin{aligned}
h(\delta_i) &= \frac{1}{\sqrt{2\pi a_i}}\exp\left\{-\frac{1}{2a_i}\left[\delta_i-a_i(g_i-\frac{\Delta_i}{\xi})\right]^2\right\}\exp\left\{\frac{a_i(g_i-\Delta_i/\xi)^2}{2}\right\},\\
\Delta_i &= I_{\{\delta_i>0\}},\\
a_i &= \frac{\sigma^2}{(k-i)(m+a_0m_0)}
\end{aligned}
$$

and

$$
g_i=\sum_{p=i+1}^{k}\left[\frac{m}{\sigma^2}\left(\bar{y}_p-q+\delta_i\right)+\frac{a_0m_0}{\sigma^2}\left(\bar{y}_{0p}-q+\delta_i\right)\right]. \tag{2.5}
$$

Here, the normalizing constant c is

$$
c=\frac{1}{\rho_i h(0)+(1-\rho_i)\frac{1}{\xi}\int_0^\infty h(\delta_i)d\delta_i}.
$$

4. The full conditional distribution of ρ_i

$$
\begin{aligned}
p(\rho_i|\delta_i) &\propto \rho_i^{\alpha_0-1}(1-\rho_i)^{\beta_0-1}\left[\rho_i I_{\{\delta_i=0\}}+(1-\rho_i)\frac{1}{\xi}\exp\left\{-\frac{\delta_i}{\xi}\right\}I_{\{\delta_i>0\}}\right]\\
&\propto\begin{cases}\rho_i^{(\alpha_0+1)-1}(1-\rho_i)^{\beta_0-1}, & \text{if } \delta_i=0,\\ \rho_i^{\alpha_0-1}(1-\rho_i)^{(\beta_0+1)-1}, & \text{if } \delta_i>0.\end{cases}
\end{aligned}
$$

With $\alpha_0 = \beta_0 = 1$ (uniform), the full conditional distribution of ρ_i is

$$p(\rho_i|\delta_i) = \begin{cases} \text{BETA}(2,1), & \text{if } \delta_i = 0, \\ \text{BETA}(1,2), & \text{if } \delta_i > 0. \end{cases}$$

5. The full conditional distribution of a_0

$$p(a_0|Y, \mu_1, \sigma^2, \{\delta_i\}) \propto (2\pi\sigma^2)^{-\frac{km_0 a_0}{2}} \exp\left\{-\frac{a_0 s_0}{2\sigma^2}\right\} a_0^{\eta_0 - 1}(1-a_0)^{\lambda_0 - 1}.$$

3. Computation Schemes

Suppose two models M_1 and M_2 are under consideration. Under M_i, the data X has a parametric distribution with density $f_i(X|\theta_i)$. Let Θ_i be the parameter space for θ_i. Let $\pi(\theta_i)$ be the prior density for θ_i under M_i. Then the Bayes factor B_{21} of model M_1 to model M_2, is defined by

$$B_{21} = \frac{m_2(X)}{m_1(X)} = \frac{\int_{\Theta_2} f_2(X|\theta_2)\pi(\theta_2)d\theta_2}{\int_{\Theta_1} f_1(X|\theta_1)\pi(\theta_1)d\theta_1},$$

where $m_i(X)$ is the marginal or predictive density of X under M_i.

We consider hypotheses:

$$M_{0i} : \delta_i = 0 \quad \text{versus} \quad M_{1i} : \delta_i > 0.$$

The Bayes factor is then

$$bf = \frac{\Pr(\delta_i > 0|Y)}{\Pr(\delta_i = 0|Y)}, \quad i = 1, 2, \ldots, k-1. \tag{3.1}$$

We use the Gibbs sampler to compute the Bayes factor in (3.1)

Step 1

(1) Set $\delta_i^{(1)} = 0$ for all $1 \le i \le k-1$ as initial values.

(2) Sample $\rho_i^{(1)}|\delta_i^{(1)}$ for $1 \le i \le k-1$ from BETA(2,1).

(3) Set $\mu_1^{(1)} = 0$ as initial values.

(4) Sample $\sigma^{2(1)}|Y, \mu_1^{(1)}, \{\delta_i^{(1)}\}$ from

$$IG\left(\frac{km}{2} + \frac{ka_0 m_0}{2} + a, \left[\frac{s}{2} + \frac{a_0 s_0}{2} + \frac{1}{b}\right]^{-1}\right).$$

Step t

(1) First, compute the conditional posterior probability of $\delta_i^{(t)} = 0$ for $1 \le i \le k-1$ by

$$\lambda_i^{(t)} = P(\delta_i^{(t)} = 0|\cdot) = \frac{\rho_i^{(t-1)} h(0; a_i^{(t)}, g_i^{(t)})}{\rho_i^{(t-1)} h(0; a_i^{(t)}, g_i^{(t)}) + (1-\rho_i^{(t-1)})\frac{1}{\xi}\int_0^\infty h(\delta_i^{(t-1)}; a_i^{(t)}, g_i^{(t)}) d\delta_i^{(t-1)}},$$

where

$$a_i^{(t)} = \frac{(k-i)(m + a_0 m_0)}{\sigma^2}$$

and

$$g_i^{(t)} = \sum_{p=i+1}^{k} \left[\frac{m}{\sigma^2}\left(\bar{y}_p - q + \delta_i\right) + \frac{a_0 m_0}{\sigma^2}\left(\bar{y}_{0p} - q + \delta_i\right)\right].$$

Here, h$(\cdot)$ is defined in (2.5). Next, sample $B_i^{(t)}$ from a Bernoulli trial with the probability $\lambda_i^{(t)}$. If $B_i^{(t)} = 1$, set $\delta_i^{(t)}$. If $B_i^{(t)} = 0$, sample $\delta_i^{(t)}|\cdot$ from a truncated normal distribution with mean $g_i^{(t)}/a_i^{(t)}$ and variance $1/a_i^{(t)}$ for $0 < \delta_i^{(t)} < \infty$.

(2) Sample $\rho_i^{(t)}|\delta_i^{(t)}$ for $1 \le i \le k-1$ from

$$\begin{cases} \text{BETA}(2,1), & \text{if } \delta_i = 0, \\ \text{BETA}(1,2), & \text{if } \delta_i > 0. \end{cases}$$

(3) Sample $\mu_1^{(t)}|Y, \sigma^{2(t-1)}, \left\{\delta_i^{(t)}\right\}$ from

$$N\left(\frac{u^{(t-1)}}{v^{(t-1)}}, \frac{1}{v^{(t-1)}}\right),$$

where

$$u = \frac{m}{\sigma^2}\left[\sum_{i=1}^{k} \bar{y}_i - \sum_{l=1}^{k-1}(k-l)\delta_l\right] + \frac{a_0 m_0}{\sigma^2}\left[\sum_{i=1}^{k} \bar{y}_{0i} - \sum_{l=1}^{k-1}(k-l)\delta_l\right] + \frac{\mu_0}{\sigma_0^2}$$

and

$$v = \frac{km}{\sigma^2} + \frac{k a_0 m_0}{\sigma^2} + \frac{1}{\sigma_0^2}.$$

(4) Sample $\sigma^{2(t)}|Y, \mu_1^{(t)}, \{\delta_i^{(t)}\}$ from

$$IG\left(\frac{km}{2} + \frac{k a_0 m_0}{2} + a, \left[\frac{s}{2} + \frac{a_0 s_0}{2} + \frac{1}{b}\right]^{-1}\right).$$

Then, repeat Step t, for $t = 2, 3, \ldots,$ and continue.

Remark 3.1 We use a Metropolis-Hastings algorithm for estimating a_0 because the full conditional distribution of a_0 does not have a well-known specific form.

4. A Simulation Study

We conducted a simulation study to investigate the performance of the proposed methods. Let μ_0^* and μ_1^* denote three dimensional mean vectors for historical and current data respectively. We consider two different simulated data sets:

$$\text{data set } 1: \quad \mu_0^* = (3,3,3) \ \text{ and } \ \mu_1^* = (3,3,3),$$
$$\text{data set } 2: \quad \mu_0^* = (3,4,5) \ \text{ and } \ \mu_1^* = (3,4,5).$$

In our simulations we use $\sigma^2 = 1$, and the hyperparameters (μ_0, σ_0) and (a, b) are $(1, 10)$ and $(3, 1)$ respectively. We use $\xi = 10$ in (2.4). We fix the sample size as $m = m_0 = 50$ in each population for both historical and current data. We use the Metropolis-Hastings algorithm within the Gibbs sampler. The algorithm was run for 5,500 iterations after

Table 4.1: Parameter estimates and Bayes factors with the data set 1

				parameter				
(η_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$m = m_0$		μ_1	σ	a_0	δ_1	δ_2
(1,9)	(0.1,0.090)	20	*Mean*	2.9971	0.9872	0.5670	0.0064	0.0052
			S.D	0.1080	0.1463	0.2116	0.0489	0.0380
			Bayes factor				0.0879	0.1216
		30	*Mean*	3.1426	0.9675	0.7161	0.0035	0.0034
			S.D	0.0824	0.1103	0.1457	0.0300	0.0281
			Bayes factor				0.0658	0.0908
		50	*Mean*	2.9527	1.0169	0.8493	0.0018	0.0025
			S.D	0.0619	0.0858	0.0636	0.0168	0.0211
			Bayes factor				0.0552	0.0763
(30,30)	(0.5,0.064)	20	*Mean*	2.9965	0.9882	0.5920	0.0051	0.0057
			S.D	0.1063	0.1391	0.0694	0.0375	0.0403
			Bayes factor				0.0929	0.1296
		30	*Mean*	3.1462	0.9651	0.6330	0.0033	0.0034
			S.D	0.0848	0.1109	0.0674	0.0279	0.0261
			Bayes factor				0.0669	0.0964
		50	*Mean*	2.9536	1.0185	0.7096	0.0027	0.0030
			S.D	0.0665	0.0905	0.0581	0.0215	0.0230
			Bayes factor				0.0555	0.0781
(3,1)	(0.75,0.194)	20	*Mean*	2.9876	0.9935	0.9561	0.0059	0.0054
			S.D	0.0963	0.1279	0.0522	0.0417	0.0368
			Bayes factor				0.0790	0.1143
		30	*Mean*	3.1385	0.9645	0.9703	0.0026	0.0025
			S.D	0.0768	0.1016	0.0338	0.0221	0.0215
			Bayes factor				0.0590	0.0867
		50	*Mean*	2.9508	1.0201	0.9846	0.0015	0.0022
			S.D	0.0588	0.0813	0.0158	0.0145	0.0200
			Bayes factor				0.0529	0.0761

Table 4.2: Parameter estimates and Bayes factors with the data set 2

				parameter				
(η_0,λ_0)	(μ_{a_0},σ_{a_0})	$m=m_0$		μ_1	σ	a_0	δ_1	δ_2
(1,9)	(0.1,0.090)	20	*Mean*	2.9977	0.7492	0.4056	0.9923	1.0213
			S.D	0.1666	0.1180	0.2448	0.2346	0.2308
			Bayes factor				2.33×10^{16}	1.45×10^{9}
		30	*Mean*	3.0783	1.0893	0.7601	1.0130	1.0003
			S.D	0.1401	0.1223	0.1092	0.1993	0.2048
			Bayes factor				5.93×10^{23}	3.97×10^{11}
		50	*Mean*	3.0290	0.8445	0.7984	1.0243	0.9981
			S.D	0.0973	0.0732	0.1043	0.1376	0.1350
			Bayes factor				1.99×10^{44}	9.19×10^{19}
(30,30)	(0.5,0.064)	20	*Mean*	2.9949	0.7499	0.5605	0.9944	1.0209
			S.D	0.1526	0.1091	0.0717	0.2213	0.2222
			Bayes factor				2.84×10^{19}	1.76×10^{10}
		30	*Mean*	3.0813	1.0900	0.6502	1.0050	1.0075
			S.D	0.1461	0.1307	0.0653	0.2091	0.2092
			Bayes factor				1.25×10^{20}	1.11×10^{10}
		50	*Mean*	3.0247	0.8445	0.6740	1.0176	1.0078
			S.D	0.0987	0.0754	0.0641	0.1438	0.1415
			Bayes factor				1.60×10^{44}	2.51×10^{21}
(3,1)	(0.75,0.194)	20	*Mean*	2.9844	0.7453	0.9208	0.9991	1.0219
			S.D	0.1377	0.0997	0.1063	0.1940	0.1950
			Bayes factor				7.36×10^{24}	2.89×10^{13}
		30	*Mean*	3.0908	1.0903	0.9753	0.9946	1.0057
			S.D	0.1370	0.1142	0.0259	0.1896	0.1881
			Bayes factor				2.45×10^{25}	7.97×10^{12}
		50	*Mean*	3.0429	0.8420	0.9786	1.0075	1.0087
			S.D	0.0921	0.0690	0.0246	0.1294	0.1282
			Bayes factor				1.90×10^{46}	7.60×10^{22}

the initial 500 iterations were discarded as a burn-in. We estimate the parameters and compute the Bayes factors in each data set. We use two different hyperparameters for (η_0,λ_0) to see behavior of the precision parameter a_0. They are (1,9), (30,30) and (3,1). All numerical values are reported in Tables 4.1 and 4.2.

Table 4.1 is the simulation results for the data set 1. All estimates of parameters are close to true values regardless of the change of hyperparameters. Since the Bayes factors δ_1 and δ_2 have both very small values, the results are congruent with what we would expect from the data set. For fixed m and m_0, as the estimate of a_0 increases, the Bayes factor decreases. For a given set of hyperparameters, as the sample size increases, the estimate of a_0 increases resulting in small Bayes factors. These are due to the effect of historical data.

In general, as μ_{a_0} increases, we expect the estimate of a_0 increases. However, we can see a little bit awkward result. For example, when we change the hyperparameters from (1,9) to (30,30), the estimate of a_0 decreases.

Table 4.2 is the result for the data set 2. All estimates are quite close to true values. The Bayes factors are quite big as expected. For fixed m and m_0, as the estimate of a_0 increases, the Bayes factor increases in most of cases. In particular, as μ_{a_0} increases, the estimate of a_0 increases when $m = m_0 = 20$. For a given set of hyperparameters, as the sample size increases, the estimate of a_0 increases resulting in large Bayes factors except for a few cases.

5. Concluding Remarks

In this article we use the power prior under the simple order alternative to perform hypothesis testing in the one-way ANOVA fixed model. We only consider the balanced model in the sense that the sample sizes of each population are the same. We can also extend this result to the unbalanced case, which is not presented in the text due to complexity of notation. When the historical data are available, the power prior is practical in the sense that more information could be incorporated. Future research directions include analysis of other order restricted alternatives such as the tree and umbrella orderings.

References

Bartholomew, D. J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika*, **46**, 36–48.

Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives. II. *Biometrika*, **46**, 328–335.

Bartholomew, D. J. (1961a). A test of homogeneity of means under restricted alternatives (with discussions). *Journal of the Royal Statistical Society*, Ser. B, **23**, 239–281.

Bartholomew, D. J. (1961b). Ordered tests in the analysis of variance. *Biometrika*, **48**, 325–332.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91**, 109-122.

Bohrer, R. and Francis, G. K. (1972). Sharp one-sided confidence bounds over positive regions. *The Annals of Mathematical Statistics*, **43**, 1541–1548.

Ibrahim, J. G. and Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science*, **15**, 46–60.

Gelfand, A. E., Hills, S. E., Racine-poon, A. and Smith, A. F. M. (1900). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**, 972–985.

Gopalan, R. and Berry, D. A. (1998). Bayesian multiple comparisons using Dirichlet process priors. *Journal of the American Statistical Association*, **93**, 1130–1139.

Hayter, A. J. (1990). A one-sided studentized range test for testing against a simple ordered alternative. *Journal of the American Statistical Association*, **85**, 778–785.

Kim, S. W. and Sun, D. (2000). Intrinsic priors for model selection using an encompassing model with applications to censored failure time data. *Lifetime Data Analysis*, **6**, 251–269.

Kim, H. J. and Kim, S. W. (2001). Two-sample Bayesian tests using intrinsic Bayes factors for multivariate normal observations. *Communications in Statistics-Computation and Simulation*, **30**, 426–436.

Liu, L. (2001). Simultaneous statistical inference for monotone dose-response mean. *Doctoral dessertation, Memorial University of Newfoundland*, St. John's, Canada.

Pauler, D. K., Wakefield, J. C. and Kass, R. E. (1999). Bayes factors and approximations for variance component models. *Journal of the American Statistical Association*, **94**, 1242–1253.

Son, Y. and Kim, S. W. (2005). Bayesian single change point detection in a sequence of multivariate normal observations. *Journal of Theoretical and Applied Statistics*, **39**, 373–387.

[Received May 2007, Accepted October 2007]