

A Bayesian Approach to Successive Comparisons

A. Aghamohammadi¹, M. R. Meshkani² and M. Mohammadzadeh¹
¹Tarbiat Modares University and ²Shahid Beheshti University

Abstract: The present article discusses and compares multiple testing procedures (MTPs) for controlling the family wise error rate. Machekano and Hubbard (2006) have proposed empirical Bayes approach that is a resampling based multiple testing procedure asymptotically controlling the familywise error rate. In this paper we provide some additional work on their procedure, and we develop resampling based step-down procedure asymptotically controlling the familywise error rate for testing the families of one-sided hypotheses. We apply these procedures for making successive comparisons between the treatment effects under a simple-order assumption. For example, the treatment means may be a sequences of increasing dose levels of a drug. Using simulations, we demonstrate that the proposed step-down procedure is less conservative than the Machekano and Hubbard's procedure. The application of the procedure is illustrated with an example.

Key words: Bayesian inference, familywise error rate, MCMC, multiple comparisons.

1. Introduction

In some studies in which the effects of k treatments on a response variable are examined, the researcher may believe before observing the data that the true means are simply ordered ($\mu_1 \leq \dots \leq \mu_k$). For example, the treatments may be a sequence of increasing dose level of a drug. In the framework of one way analysis of variance (ANOVA) the problem of testing the null hypothesis $H_0 : \mu_1 = \dots = \mu_k$ against the simply ordered alternative hypothesis $H_1 : \mu_1 \leq \dots \leq \mu_k$ with at least one strict inequality has received considerable attention in the statistical literature. Barlow *et al.* (1972) and Robertson, Wright and Dykstra (1988). Recently, Hayter (1990) developed a one-sided studentized range test which provides simultaneous one-sided lower confidence bounds for the ordered pairwise comparisons $\mu_i - \mu_j, 1 \leq j < i \leq k$. These and other simultaneous inference procedures are discussed in Miller (1981), Hochberg and Tamhane (1987) and Hsu (1996).

However, under the ordering assumption $\mu_1 \leq \dots \leq \mu_k$ an experimenter may want to look directly at only the successive difference between the treatment effects, namely the subset of pairwise differences

$$\mu_2 - \mu_1, \dots, \mu_k - \mu_{k-1}.$$

Liu, Miwa and Hayter (2000) derived the simultaneous confidence intervals for the set of successive differences. Liu and Somerville (2004) developed a step-wise multiple testing procedure for successive comparisons of treatment effects which is uniformly more powerful than the simultaneous confidence interval procedure in terms of rejection of the null hypotheses. These procedures are not based on the assumption that μ_i 's follow a simple ordering. However, if there is some prior information about the order of μ_i 's, i.e., such as $\mu_1 \leq \dots \leq \mu_k$ for treatment effects, then it can be used to improve the precision of inference. On the other hand, the computation of critical values is also more complicated for large k with respect to controlling the familywise error rate. Recognizing such limitations, we propose a Bayesian method of multiple testing for successive comparisons of ordered treatment effects. This Bayesian development incorporates ordering information using some suitable prior distribution for the parameters. The Bayesian approach to various statistical problems has become popular in recent years due to the advancement in computational power. In the context of comparing treatments, Machezano and Hubbard (2006) proposed a single-step procedure, with a common cut-off value, based on empirical Bayes approach to control familywise error rate for the multiple testing problem. It is performed by estimating the test statistics null distribution over the true nulls. The null distribution comes either from the distribution of the maximum of the k correlated test statistics or the minimum of the p-values.

In the present study, we consider testing whether the two means in a pairwise comparison in the one-way ANOVA are equal ($\mu_i = \mu_j$) or satisfy increasing ($\mu_i < \mu_j$) order. In fact, we develop a procedure for testing the families of hypotheses

$$H_{i0} : \mu_{i+1} - \mu_i = 0 \quad vs \quad H_{i1} : \mu_{i+1} - \mu_i > 0, \quad i = 1, \dots, k-1$$

from Bayesian view point. When a test of hypothesis is to be conducted in a Bayesian framework and the parameter space under null hypothesis consists of a single point, a point-mass prior, which is typically a discrete-continuous mixture, may be used. On the successive differences of the means, we place independent prior distributions that are mixtures of a truncated normal distribution and a discrete distribution with entire mass at zero. For simultaneous testing of multiple point null hypotheses against one-sided alternatives, we propose a new step-down multiple testing procedure that also controls the familywise error rate. In

fact this procedure is step-down version of Machekano and Hubbard's single-step procedure. According to the proposed new step-down procedure, the ordered test statistics or the associated p-values are compared with a set of critical values in a stepwise fashion toward identifying the set of true and false null hypotheses. The critical value used in each step incorporates the decision made in the preceding steps. Therefore, it provides better control of the familywise error rate, hence compared to a single-step procedure where decision about true and false null hypotheses is made by simply comparing each test statistic or the corresponding p-value with a single critical value, it is more powerful.

In the remainder of this article, we describe the hierarchical model, prior specifications, and posterior computations. We then discuss the findings of a simulation study, comparing our procedure with other procedures that controls the familywise error rate for the multiple tests of successive comparisons of treatment effects. We demonstrate through this study that our proposed procedure provide much better control of familywise error rate than those that are currently available in the literature.

2. Multiple Testing Framework

Suppose the treatments are a sequence of increasing dose levels of a drug. An experimenter may want to look directly at only the successive differences between the treatment effects. In order to assess how the treatment effects change with increasing dose levels, we consider to test the following family of hypotheses,

$$H_{i0} : \mu_{i+1} - \mu_i = 0 \quad vs \quad H_{i1} : \mu_{i+1} - \mu_i > 0, \quad i = 1, \dots, k-1. \quad (2.1)$$

Let T_1, \dots, T_{k-1} be the corresponding test statistics and let p_1, \dots, p_{k-1} be the associated p-values. In single-step hypotheses testing, H_{j0} is rejected if $T_j > c_j(\alpha)$, or if $p_j \leq \alpha$, for some chosen $\alpha \in (0, 1)$, $j = 1, \dots, k-1$. It is often assumed that if H_{j0} is true, T_j comes from some standard theoretical distribution (e.g. normal or t-distribution), from which we get the critical values $c_j(\alpha)$.

A step-down test, which is not a single-step procedure, based on $k-1$ critical constants $d_1(\alpha) \leq \dots \leq d_{k-1}(\alpha)$, proceeds as follows. Denote the ordered T_i -values as $T_{(1)} \leq \dots \leq T_{(k-1)}$ and the corresponding null hypotheses as $H_{(1)0}, \dots, H_{(k-1)0}$. Start with the largest T -value, $T_{(k-1)}$. If $T_{(k-1)} \leq d_{k-1}(\alpha)$, then stop and accept all the null hypotheses, otherwise reject $H_{(k-1)0}$ and go to the next step. In general, if testing has continued to the i th step ($1 \leq i \leq k-1$) and if $T_{(k-i)} \leq d_{k-i}(\alpha)$, then stop testing and accept all the remaining hypotheses, $H_{(1)0}, \dots, H_{(k-i)0}$, otherwise reject $H_{(k-i)0}$ and go to the $(i+1)$ th step. Define S to be the set of all null hypotheses, S_0 the set of all true null hypotheses and let S_1 represent the set of true non-null hypotheses. In a multiple testing problem, we would like to accurately estimate the subset S_0 , thus its complement S_1 , while

controlling probabilistically some error rate under an assumed significance level α . In this article, the critical constants are chosen so that the familywise error rate (FWER), the probability of making at least one error by rejecting a true null hypothesis among all tests, is controlled at a pre-specified level α . It is obvious that choosing the rejection regions, $c_j(\alpha)$ and $d_j(\alpha), j = 1, \dots, k - 1$, are challenging, because the theoretical joint null distribution for the $k - 1$ test-statistics is not always obvious. Moreover, the difference between the theoretical null and empirical null distribution affects simultaneous inference.

3. Bayes Procedure

Single step maxT multiple testing procedure, with common cut-off, uses a null distribution based on the joint distribution of the test statistics. The null distribution comes either from the distribution of the maximum of the $k - 1$ correlated test statistics or minimum of the p-values. This null distribution is used to define the rejection regions as well as the p-values. Common cut-offs that guarantee control of FWER are chosen from the maximum test statistic distribution. Machekano and Hubbard (2006) introduced a method for choosing the common cut-off $c(\alpha)$ which involves controlling the tail probability of a random variable \tilde{V} defined as

$$\tilde{V} = \sum_{i=1}^{k-1} I(\tilde{T}_i > c(\alpha), i \in S_{0n}).$$

This random variable represents a guessed number of false positives among rejections, defined by drawing $\tilde{T} = (T_1, \dots, T_{k-1})$ from a null distribution for the test-statistic vector and independently drawing S_{0n} which represents a guess of the set of true null hypotheses S_0 . The distribution of S_{0n} and null distribution of \tilde{T} are chosen so that \tilde{V} asymptotically dominates in distribution the true number of false positives V where $V = \sum_{i=1}^{k-1} I(T_i > c(\alpha), i \in S_0)$. The Bayes utilizes a Bayesian method to estimate the guessed set of null hypotheses S_{0n} . Machekano and Hubbard (2006) used empirical Bayes approach to estimate S_{0n} . In their procedure, the critical value $c(\alpha)$ is given by the $\max\tilde{T}$ distribution only over the guessed set of null hypotheses. We shall refer to this procedure as maxT procedure. In the next subsection we will use a hierarchical Bayes approach to estimate the guessed set of null hypotheses, and describe how the null distribution of \tilde{T} is drawn. Under this hierarchical Bayes model, we define the maxT procedure and then present the new step-down procedure.

3.1 Hierarchical model

Consider the balanced one-way analysis of variance model

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad 1 \leq i \leq k, \quad 1 \leq j \leq n$$

where the ε_{ij} are independent $N(0, \sigma^2)$ random variables. Let \bar{X}_i , $1 \leq i \leq k$ be the i th sample mean based upon n observations, X and μ be the vector of \bar{X}_i and μ_i , $1 \leq i \leq k - 1$, respectively and Σ_1 the covariance matrix of X . We define A to be a matrix such that $AX = Y = (\bar{X}_2 - \bar{X}_1, \dots, \bar{X}_k - \bar{X}_{k-1})$ and $A\mu = \delta = (\mu_2 - \mu_1, \dots, \mu_k - \mu_{k-1})$. The distribution of Y will thus be as

$$f(y|\delta, \sigma^2) \propto \frac{1}{(\sigma^2)^{k-1/2}} \exp(-1/2((y - \delta)'(A\Sigma_1 A')^{-1}(y - \delta))).$$

The prior for the difference parameter, $\delta_i = \mu_{i+1} - \mu_i$ $i = 1, \dots, k - 1$, is a mixture of a truncated normal distribution and a discrete distribution with its entire mass at $\delta_i = 0$, in particular,

$$[\delta_i|\sigma^2, \rho] \sim \pi(\delta_i|\sigma^2, \rho) = \rho I_{\{\delta_i=0\}} + (1 - \rho) \frac{2}{(2\pi\sigma^2)^{1/2}} \exp(-\frac{\delta_i^2}{2\sigma^2}) I_{\{\delta_i>0\}}$$

with $1 \leq i \leq k - 1$, where $I_{\{\cdot\}}$ represents the ordinary indicator function.

The hyperparameter ρ , which represents $\Pr(\delta_i = 0)$, has a beta hyperprior distribution, $\rho \sim BETA(\alpha, \beta)$. Of course, setting $\alpha = \beta = 1$ makes this hyperprior $UNIF(0, 1)$. For σ^2 , we consider Jeffreys' reference prior (Jeffreys, 1946), which would be $\sigma^2 \sim \pi(\sigma^2) \propto 1/\sigma^2$.

In the next section for performing multiple comparisons we will use the posterior probability of δ_i , ($\Pr(\delta_i = 0|y)$), to define a random set S_{0n} which represents a guessed set of true null hypotheses S_0 . To compute the above probability, the Gibbs sampling is a special technique of implementing the Markov Chain Monte Carlo (MCMC) method, which generates observations from the target posterior distribution through iterative sampling. To use the Gibbs sampling, we must obtain the full conditional posterior distributions of all parameters and hyperparameters.

Due to conjugacy, the conditional posterior distribution of ρ and σ^2 can be

easily obtained as

$$\begin{aligned}\pi(\rho|\alpha, \beta, \delta_1, \dots, \delta_{k-1}) &= \text{BETA} \left(k-1 + \alpha + \sum_{i=1}^{k-1} \Delta_i, \sum_{i=1}^{k-1} \Delta_i + \beta \right) \\ \pi(\sigma^2|y, \delta_1, \dots, \delta_{k-1}) &= \text{IG} \left[(k-1 + \sum_{i=1}^{k-1} \Delta_i) / 2, \left((y - \delta)' (A \Sigma A')^{-1} \right. \right. \\ &\quad \left. \left. (y - \delta) \right) / 2 + \sum_{i=1}^{k-1} \Delta_i \delta_i^2 \right]\end{aligned}$$

with $\Delta_i = I_{\{\delta_i > 0\}}$, $1 \leq i \leq k-1$, $\Sigma = \sigma^{-2} \Sigma_1$ and $\text{IG}(a, b) \equiv \text{Inverse Gamma}(a, b)$.

The conditional posterior distribution of δ_i , is a mixture of a point mass at zero and a normal distribution truncated at zero, for $1 \leq i \leq k-1$, it can be shown that

$$\begin{aligned}\pi(\delta_i|y, \sigma^2, \rho, \delta_{-i}) &= s_i \left[\rho I_{\{\delta_i = 0\}} + (1 - \rho) \exp \left(\frac{d_i + 1}{2\sigma^2} p_i^2 \right) \sigma^{-2} \right. \\ &\quad \left. \times \exp \left(-\frac{d_i + 1}{2\sigma^2} (\delta_i - p_i)^2 \right) I_{\{\delta_i > 0\}} \right]\end{aligned}$$

where $d_i = \sum_{j=1}^i e_{ji}$, $p_i = \frac{d_i y_i + c_i}{d_i + 1}$, $c_i = \sum_{j=1}^i e_{ji} \sum_{m=j, m \neq i}^{k-1} (y_m - \delta_m) e_{jm}$ and e_{ji} is the (j, i) th element of an upper-triangular matrix which is the Choleski decomposition of $(A \Sigma A')^{-1}$. The normalizing constant s_i satisfies

$$s_i^{-1} = \rho + (1 - \rho) \exp \left(\frac{d_i + 1}{2\sigma^2} p_i^2 \right) \left(\frac{2\pi}{d_i + 1} \right)^{1/2} \left[1 - \Phi(0, p_i, \sigma(d_i + 1)^{-1/2}) \right]$$

where Φ is the normal cumulative distribution function with mean p_i and standard error $\sigma(d_i + 1)^{-1/2}$.

For our current hierarchical model, $\text{Pr}(\delta_i = 0|y)$ can be easily computed by counting the number of times $\delta_i = 0$ generated during the MCMC steps. We will use this posterior probability as the Bernoulli probability on H_{i0} being true. Now we define a random set S_{0n} which represents a guess of the set of true null hypotheses S_0 , as follows:

$$S_{0n} = \{i : C(i) = 1\}, \quad C(i) \sim \text{Bernoulli}(\text{Pr}(\delta_i = 0|y)).$$

As discussed in van der Laan, Brikner and Hubbard (2005), in order to control FWER, S_{0n} must contain the set of true null hypotheses i.e. $S_0 \subset S_{0n}$, due to the fact that if $i \in S_{0n}$ while in reality $i \in S_0^c$, the cut-off chosen will be too large. To avoid the above problem, they replace the single guess of the set of true

null hypotheses by a random guess from a distribution which is asymptotically degenerate at the set of true null hypotheses. In this study, we use the posterior probability, $Pr(\delta_i = 0|y)$, as the Bernoulli probability on H_{i0} being true which represents a random guess of the set of true null hypotheses. Our simulations show that using this method to generate S_{0n} instead of empirical Bayes method used by van der Laan, Brikner and Hubbard (2005), has no effects on the controlling of the familywise error rate at nominal level α .

\tilde{T} is computed by drawing a bootstrap sample X_1^*, \dots, X_{nk}^* from the empirical distribution P_n of the original sample X_1, \dots, X_{nk} , or from a model based estimate \tilde{P}_n of P such that P is the data generating distribution, and subsequently calculating the test statistics based on this bootstrap sample. This will be repeated B^* times and will result in an $(k-1) \times B^*$ matrix of test statistics vectors, each representing a draw from the test statistics vector under the empirical P_n or the model based estimate \tilde{P}_n . Subsequently, we compute the row means of the matrix, and the matrix is shifted (centered) by the respective means so that the row means after this shift are equal to the null-value. This matrix represents a sample of B^* draws from a null distribution. Each row of this matrix will specify a draw of $\tilde{T} = (\tilde{T}_j : j = 1, \dots, k-1)$. We shall refer to this matrix as \tilde{T}_0 . It can be also scaled columnwise so that the row-variances equal a null value (Dudoit, van der Laan and Pollard, 2004; Pollard and van der Laan, 2003). Machezano and Hubbard (2006) have used the single step maxT multiple testing procedure. This procedure is a single-step approach, with common cut-off, which uses a corrected test statistics null distribution as follows:

1. Generate $k-1$ Bernoulli random variables using probability $Pr(\delta_i = 0|y)$, $i = 1, \dots, k-1$.
2. Repeat 1, B^* times to form $(k-1) \times B^*$ matrix as \tilde{B}_0 .
3. Take the Hadamard product (entry-wise product) of \tilde{B}_0 and \tilde{T}_0 , then get the maximum in each column $T_1^{\max}, \dots, T_{B^*}^{\max}$ to get the maxT distribution only among the true null hypotheses.
4. The critical value $c(\alpha)$ is given as the $1-\alpha$ quantile of the maxT distribution. Reject H_{i0} if $T_i \geq c(\alpha)$, otherwise do not reject H_{i0} .

3.2 New step-down procedure

The families of hypotheses in (2.1) are non-hierarchical, that is, the truth or falsity of one hypothesis has no logical implication on the truth or falsity of any other hypotheses in the family. Liu (1996) discussed the determination of step-down and step-up critical constant for a family of non-hierarchical hypotheses.

As pointed out in Liu (1996), the step-down critical constants $d_1(\alpha) \leq \dots \leq d_{k-1}(\alpha)$ can be chosen to satisfy the following constraints in order that the familywise error rate is less than α :

$$\min Pr\{\max(T_{i_1}, \dots, T_{i_m}) \leq d_m(\alpha)\} \geq 1 - \alpha, \quad m = 1, \dots, k - 1$$

where the probability is calculated at $\mu_1 = \dots = \mu_k$, and the minimum is over all subsets $\{i_1, \dots, i_m\} \subset \{1, \dots, k - 1\}$ with cardinality m .

For the balanced one-sided situation, the minimum probability is attained at $\{i_1 = 1, \dots, i_m = m\}$ (Liu and Somerville, 2004).

We assume that the design is balanced, means that the sample sizes are equal for k treatments. Since in our study the hypotheses are one-sided by using the above step-down procedure, we propose the following new step-down procedure which entails the first two steps of Machekano and Hubbard's method and changes other steps to compute the critical values $d_i(\alpha), i = 1, \dots, k - 1$.

1. Generate $k - 1$ Bernoulli random variables using probability $P(\delta_i = 0|y), i = 1, \dots, k - 1$.
2. Repeat 1, B^* times to form $(k - 1) \times B^*$ matrix as \tilde{B}_0 .
3. Take the Hadamard product (entry-wise product) of \tilde{B}_0 and \tilde{T}_0 , then calculate

$$T_{ij}^{\max} = \max_{1 \leq h \leq i} \{V_{hj}\}, \quad j = 1, \dots, B^* \quad i = 1, \dots, k - 1$$

where V_{hj} is (h, j) th element of the matrix which is Hadamard product of \tilde{B}_0 and \tilde{T}_0 .

4. The critical value $d_i(\alpha), i = 1, \dots, k - 1$ is given as the $1 - \alpha$ quantile of the $T_{ij}^{\max}, j = 1, \dots, B^*$. Now in the i th step ($1 \leq i \leq k - 1$) if $T_{(k-i)} \leq d_{k-i}(\alpha)$, then stop testing and accept all the hypotheses, $H_{(1)0}, \dots, H_{(k-i)0}$, otherwise reject $H_{(k-i)0}$ and go to the $(i + 1)$ th step.

It is obvious from the computation of $d_i(\alpha), i = 1, \dots, k - 1$ in the new step-down procedure and of $c(\alpha)$ in the maxT procedure that $d_i(\alpha) \leq c(\alpha)$ for $1 \leq i \leq k - 1$. Consequently, the maxT procedure rejects a set of the hypotheses which is a subset of the hypotheses rejected by the new step-down procedure. Thus, the new step-down procedure is less conservative than the maxT procedure. However, we know that the step-down procedure introduced by Liu and Somerville (2004) controls FWER, so the new step-down procedure introduced above guarantees asymptotic control of FWER at level less than or equal to α , given the distribution of null-centered rescaled \tilde{T}' s dominates the distribution of observed test statistics

(see Machekano and Hubbard, 2006; van der Laan, Brikner and Hubbard, 2005 for details).

4. Simulation Studies

The simulation study compares the procedure outlined above, new step-down, with the maxT procedure presented in Machekano and Hubbard (2006) and step-down procedure presented in Liu and Somerville (2004).

In this simulation, the data are nk *i.i.d.* normally distributed $Y_{ij} = X_{(i+1)j} - X_{ij} \sim N(\mu_{i+1} - \mu_i, 2\sigma^2)$, $i = 1, \dots, k-1$ and $j = 1, \dots, n$. Recall that the prior for $\delta_i = \mu_{i+1} - \mu_i$, $i = 1, \dots, k-1$ is normal with mean zero and variance σ^2 and $\pi(\sigma^2) \propto 1/\sigma^2$. For $k = 5, 8, 10$ and 15 , we assumed $k-2$ hypotheses in the null group and for $k = 21$ we assumed 18 hypotheses in the null group, which means that the proportion of true null hypotheses approximately is 0.9. We observe that

$$Pr(H_{i0}) = Pr(\delta_i = 0) = E[E(I_{\{\delta_i=0\}}|\rho, \sigma^2)] = E(\rho) = \frac{\alpha}{\alpha + \beta}.$$

We select $\pi(\rho) = BETA(9, 1)$, this prior makes $Pr(\delta_i = 0) = 0.9$. σ^2 is generated uniformly from $(0, 10000)$ and we then draw δ_i and Y_{ij} , $i = 1, \dots, k-1$, $j = 1, \dots, n$, according to the above model. The above process is repeated N times. For each generated data set, we compute the numbers of type I error, $V = \sum_{i=1}^{k-1} \phi_i(1 - \theta_i)$, where

$$\theta_i = I_{\{H_{i0} \text{ is not true}\}}, \quad \phi_i = I_{\{H_{i0} \text{ is rejected}\}}, \quad i = 1, \dots, k-1.$$

We estimate the FWER for each procedure by computing the sample proportion of times the number of type I errors are greater than or equal to 1 in the simulations.

Using the steps described in Section 2.2, 100,000 MCMC samples are generated using the Gibbs sampling after discarding the first 3,000 as burn-in. Since the convergence is not slow, the number of burn-in samples is taken to be 3,000.

The above simulation was run for $n = 13$, $n = 20$, $\alpha = 0.05$ and $N = 8,000$. For bootstrap estimation of the null distribution of test statistics, we use $B^* = 3,000$ and non-parametric bootstrap method. That is, samples are drawn at random, with replacement from the observed data. The results of the simulations for $n = 13$ are presented in Table 1.

Table 1: FWER for three procedures, $n = 13$

k	5	8	10	15	21
step-down	0.0471	0.0468	0.0440	-	-
maxT	0.0451	0.0407	0.0350	0.0302	0.0351
New step-down	0.0501	0.0441	0.0463	0.0415	0.0451

It is clear from this Table all three procedures control the FWER at nominal level of $\alpha = 0.05$ for various k . Clearly, as we mentioned previously, the new step-down procedure is less conservative than the maxT procedure. We also see the FWER of the new step-down procedure is closed to the FWER of the step-down procedure. In many cases the FWER of the new step-down method is almost equal to the nominal Type I error rate, which is ideal for a multiple testing procedure.

Table 2 contains simulations results for above configurations for the case of $n = 20$. In the situation denoted by "–", we do not have computed FWER for the step-down procedure because in this case computation of the critical value for the step-down procedure is too complicated. As mentioned earlier our goal mostly is to compare the maxT procedure with the new step-down procedure and particularly showing the performance of the new step-down procedure relative to the maxT procedure.

Table 2: FWER for three procedures, $n = 20$

k	5	8	10	15	21
step-down	0.0470	0.4520	0.0486	-	-
maxT	0.0450	0.0384	0.0410	0.0369	0.0391
New step-down	0.0483	0.0463	0.0478	0.0441	0.0472

5. Application

We take a problem from White and Froeb (1980) for our example. The effect of smoking on pulmonary health is investigated in a retrospective study in which subjects who had been evaluated during a physical fitness profile were assigned, based on their smoking habits, to one of five groups: non-smoker (NS), passive smoker (PS), light smoker (LS), moderate smoker (MS) and heavy smoker (HS). A sample of 1000 female subjects, 200 from each group, were selected and data on pulmonary function (forced vital capacity, FVC) were recorded. It is reasonable to suspect that, for these variables, the mean values decreases from NS to PS, from PS to LS, and so on. Reversing the order, we assume that, with 1=HS, 2=MS, 3=LS, 4=PS and 5=NS, the means are simply ordered: $\mu_1 \leq \dots \leq \mu_5$. The sample averages are as

$$\bar{x}_1 = 2.55, \bar{x}_2 = 2.80, \bar{x}_3 = 3.15, \bar{x}_4 = 3.23, \bar{x}_5 = 3.35$$

and a pooled estimate $S = 0.458$ was obtained with $v = 995$ degrees of freedom. For this example, the one-sided step-down critical constants at $\alpha = 0.05$ are found from Liu and Somerville (2004). The maxT procedure yields the decisions

as follows:

$$\mu_1 < \mu_2, \mu_2 < \mu_3, \mu_3 = \mu_4, \mu_4 < \mu_5.$$

which means that the hypotheses $H_{10} : \mu_2 - \mu_1 = 0$, $H_{20} : \mu_3 - \mu_2 = 0$ and $H_{40} : \mu_5 - \mu_4 = 0$ are rejected. On the other hand, using the new step-down procedure and the step-down procedure, all four hypotheses are rejected. The new step-down procedure therefore rejects more hypotheses than the maxT procedure.

6. Concluding Remarks

This paper has introduced a new step-down procedure for multiple testing of simply ordered means in balanced situation. This technique uses the generally valid null-value shifted re-sampling based null distribution for the test statistics, as generally proposed by Pollard and van der Laan (2003) and Dudoit, van der Laan and Pollard (2004). It is important that the set of guessed true null hypotheses must be contain the set of true null hypotheses to control the FWER at nominal level α by the maxT and new step-down procedure. van der Laan, Brikner and Hubbard (2005) have not recommended using the single guess of the true null hypotheses and they replace the single guess of the true null hypotheses by a random guess from a distribution which is asymptotically degenerate at the set of true null hypotheses. Our method uses Bayesian model to generate random guesses of the set of true null hypotheses, which asymptotically converges to mixture model that is used by van der Laan, Brikner and Hubbard (2005). Our simulation studies show that the new step-down procedure is less conservative compared to the maxT procedure but still gives proper control ($\text{FWER} \leq 0.05$). Posterior distributions and consequently the posterior probabilities can be sensitive to the choice of the priors for unknown parameters, especially variance parameter in our studies. Gelman (2006) investigate various priors for variance parameters and recommended a uniform distribution on the standard deviation. In our study, we have also explored $\sigma \sim \text{UNIF}(0, a_0)$, with various value for a_0 . The results were close to those presented in tables 1 and 2.

Acknowledgments

The authors wish to express their appreciation to an anonymous referee for careful reading and valuable suggestions which served to improve the original version of this manuscript.

References

- Barlow, R. E., Bartholomew, D. J., Bermmer, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley.
- Dudoit, S., van der Lann, M. J. and Pollard, K. S. (2004). Multiple testing: Part I, Single step procedures for control of general type I error rates. *Statistical Applications in Genetics and Molecular Biology* **3**, 1-71.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceeding of the Royal Society of London* **A186**, 773-795.
- 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.
- Hochberg, Y., Tamhane, A. C. (1987). *Multiple Comparisons Procedures*. Wiley.
- Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. Chapman and Hall.
- Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis* **3**, 515-533.
- Liu, W. (1996). Multiple tests of a non-hierarchical finite family of hypotheses. *Journal of the Royal Statistics Society* **B58**, 455-461.
- Liu, W., Miwa, T. and Hayter, A. J. (2000). Simultaneous confidence interval estimation for successive comparisons of ordered treatment effects. *Journal of Statistical Planning and Inference* **88**, 75-86
- Liu, W. and Somerville, P. N. (2004). Stepwise multiple test for successive comparisons of treatment effects. *Computational Statistics and Data Analysis* **46**, 189-199.
- Machekano, R. N. and Hubbard, A. E. (2006). Empirical Bayes approach to controlling familywise error: An application to HIV resistance data. Division of Biostatistics, University of California, Berkeley. *Berkeley Electronic Press, Paper 205*.
- Miller, R. G. (1981). *Simultaneous Statistical Inference*, 2nd Edition. Springer.
- Pollard, K. S., and Van der Lann, M. J. (2003). Resampling-based multiple testing: Asymptotic control of type I error and applications to gene expression data. Division of Biostatistics, University of California, Berkeley, *Berkeley Electronic Press, Paper 121*.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley.
- van der Lann, M. J., Brikner, M. D. and Hubbard, A. E. (2005). Empirical Bayes and resampling based multiple testing procedure controlling tail probability of the proportion of false positive. *Statistical Applications in Genetics and Molecular Biology* **4**, Article 29.
- White, J. R. and Froeb, H. F. (1980). Small-airways dysfunction in nonsmokers chronically exposed to tobacco smoke. *The New England journal of Medicine* **302**, 720-723.

Received January 19, 2009; accepted May 6, 2009.

A. Aghamohammadi Department of Statistics
Tarbiat Modares University
P. O. Box 14115-175, Tehran, Iran
Ali_gha@modares.ac.ir

M. R. Meshkani
Department of Statistics
Shahid Beheshti University
Tehran, Iran
mrmeshkani@gmail.com

M. Mohammadzadeh
Department of Statistics
Tarbiat Modares University
Tehran, Iran
mohsen_m@modares.ac.ir