Journal of Statistical Planning and Inference 142 (2012) 878–895

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

A fiducial approach to multiple comparisons

Damian V. Wandler [a,*], Jan Hannig [b]

[a] Department of Statistics, Colorado State University, Fort Collins, CO 80523, United States
[b] Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States

ARTICLE INFO

Article history:
Received 16 May 2010
Received in revised form
26 May 2011
Accepted 21 October 2011
Available online 6 November 2011

Keywords:
Fiducial inference
Multiple comparisons
Importance sampling
Fiducial probability
Model selection

ABSTRACT

Comparing treatment means from populations that follow independent normal distributions is a common statistical problem. Many frequentist solutions exist to test for significant differences amongst the treatment means. A different approach would be to determine how likely it is that particular means are grouped as equal. We developed a fiducial framework for this situation. Our method provides fiducial probabilities that any number of means are equal based on the data and the assumed normal distributions. This methodology was developed when there is constant and non-constant variance across populations. Simulations suggest that our method selects the correct grouping of means at a relatively high rate for small sample sizes and asymptotic calculations demonstrate good properties. Additionally, we have demonstrated the flexibility in the methods ability to calculate the fiducial probability for any number of equal means. This was done by analyzing a simulated data set and a data set measuring the nitrogen levels of red clover plants that were inoculated with different treatments.

1. Introduction

Treatment means are commonly compared to each other to determine their relationship. A variety of problems compare treatment means. For example, comparing the effectiveness of multiple drugs in a pharmaceutical setting is a common practice. Other areas of application include agriculture, finance, production industries, etc.

In this scenario there are observations $\mathbf{X}_i = (X_{i1}, \ldots X_{in_i})$ for populations $i = 1, \ldots, k$. The k populations follow independent normal distributions with means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$ and variance $\boldsymbol{\eta}$. The multiple comparison problem (MCP) attempts to perform inference on the groupings of the individual means within $\boldsymbol{\mu}$ from the observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$.

There are several frequentist solutions for multiple comparison problems. Using frequentist methods, analysis of variance (ANOVA) is used to test for a significant treatment effect. There are several tests for differences among treatments. Some are, Fisher's least significant difference (LSD), Tukey's honest significant difference (HSD), Sheffe's pairwise differences, Duncan's multiple range test, etc. These solutions control the comparisonwise or experimentwise error rate for some α. However, these solutions cannot determine a likelihood that particular means are equal or unequal.

A Bayesian procedure for MCP has been developed in Gopalan and Berry (1998). This method uses a Dirichlet process prior to decide between competing groupings of $\boldsymbol{\mu}$. The final posterior probabilities are used to discern amongst the groupings for different priors.

We have developed methodology for this scenario using an extension of R.A. Fisher's fiducial inference. We use generalized fiducial inference as developed in Hannig (2009b) to determine the likelihood of grouping particular means as

* Corresponding author.
E-mail address: dvwandler@gmail.com (D.V. Wandler).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.10.011

equal or unequal. A model selection technique was used to determine, based on the data, the likely model(s). This is developed for $\boldsymbol{\eta}=(\eta,\ldots,\eta)$ (constant variance) and $\boldsymbol{\eta}=(\eta_1,\eta_2,\ldots,\eta_k)$ (non-constant variance). Simulation results suggest that our method selects the correct grouping at a high rate for small sample sizes. We have also proven that our method will asymptotically select the correct grouping of means.

In addition to simulation results and theoretical calculations, we analyzed a simulated data set and a data set measuring nitrogen levels of red clover plants that were inoculated with different treatments. The analyses were conducted assuming both constant and non-constant variance; the results from the red clover data set were compared with those of the Bayesian method (which assumes constant variance). Both the fiducial and Bayesian methods produce something of a posterior probability for each possible grouping.

2. Generalized fiducial inference

2.1. Overview

Fisher (1930) did not support the Bayesian idea of assuming a prior distribution on the parameters when there is limited information available. As a result, he developed fiducial inference to offset this perceived shortcoming. Fiducial inference did not garner approval when some of Fisher's claims were found to be untrue in Lindley (1958) and Zabell (1992). More recently, Weeranhandi (1993) has developed generalized inference and the work of Hannig et al. (2006) established a link between fiducial and generalized inference. Hannig (2009b) and references within provide a thorough background on fiducial inference and its properties.

The principle idea of generalized fiducial inference is similar to the likelihood function and "switches" the role of the data, $\mathbf{X}$, and the parameter(s) ξ. To formally introduce fiducial inference we assume that a relationship, called the *structural equation*, between the data, $\mathbf{X}$, and the parameter(s), ξ, exists in the form

$$\mathbf{X}=G(\xi,\mathbf{U}), \tag{1}$$

where $\mathbf{U}$ is a random vector with a completely known distribution and independent of any parameters. After observing $\mathbf{X}$ we use the known distribution of $\mathbf{U}$ and the relationship from the structural equation to infer a distribution on ξ. This allows us to define a probability measure on the parameter space, Ξ. If (1) can be inverted the inverse will be written as $G^{-1}(\cdot,\cdot)$. For an observed $\mathbf{x}$ and $\mathbf{u}$ we can calculate ξ from

$$\xi=G^{-1}(\mathbf{x},\mathbf{u}). \tag{2}$$

From this inverse relationship we can generate a random sample of $\mathbf{u}'_1,\mathbf{u}'_2,\ldots,\mathbf{u}'_M$ and obtain a random sample for $\xi:\xi'_1=G^{-1}(\mathbf{x},\mathbf{u}'_1),\xi'_2=G^{-1}(\mathbf{x},\mathbf{u}'_2),\ldots\xi'_M=G^{-1}(\mathbf{x},\mathbf{u}'_M)$. This sample is called a *fiducial sample* and can be used to calculate estimates and confidences intervals for the true parameter(s), ξ_0.

Hannig and Lee (2009) address two potential times that $G^{-1}(\cdot,\cdot)$ may not exist. They are when (i) there is no ξ that satisfies (2) or (ii) there is more than one ξ that satisfies (2). From Hannig (2009b) we will handle situation (i) by eliminating such $\mathbf{u}$'s and re-normalizing the sampling probabilities. This is reasonable because we know our data was generated using ξ_0 and $\mathbf{u}_0$ so at least one solution for (2) exists. We will only consider the $\mathbf{u}$'s that allow for $\mathbf{G}^{-1}(\cdot,\cdot)$ to exist. Hannig (2009b) suggests that situation (ii) is handled by selecting an ξ by some, possibly random, rule that satisfies the inverse in (2).

A more rigorous definition of the inverse is the set valued function of

$$Q(\mathbf{x},\mathbf{u})=\{\xi:\mathbf{x}=G(\xi,\mathbf{u})\}. \tag{3}$$

We know that our observed data was generated using some ξ_0 and $\mathbf{u}_0$. We also know the distribution of $\mathbf{U}$ and that $Q(\mathbf{x},\mathbf{u}_0)\neq\emptyset$. Coupling these facts we can compute the *generalized fiducial distribution* from

$$V(Q(\mathbf{x},\mathbf{U}^\star))\,\big|\,\{Q(\mathbf{x},\mathbf{U}^\star)\neq\emptyset\}, \tag{4}$$

where $\mathbf{U}^\star$ is an independent copy of $\mathbf{U}$ and $V(S)$ is a random element for any measurable set, S, with support on the closure of S, $\overline{S}$. Essentially, $V(\cdot)$ is the random rule for picking the possible ξ's. We will refer to a the random element that comes from (4) as $\mathcal{R}_\xi$. For a more detailed discussion of the derivation of the generalized fiducial distribution see Hannig (2009b).

From the structural equation the *generalized fiducial density* is calculated as proposed in Hannig (2009b) and justified theoretically in Hannig (2009a). Let $G=(g_1,\ldots,g_n)$ such that $X_i=g_i(\xi,\mathbf{U})$ for $i=1,\ldots,n$. ξ is a $p\times 1$ vector and denote $\mathbf{X_i}=G_{0,\mathbf{i}}(\xi,\mathbf{U_i})$ where $\mathbf{X_i}=(X_{i_1},\ldots,X_{i_p})$ and $\mathbf{U_i}=(U_{i_1},\ldots,U_{i_p})$ for all possible combinations of the indexes $\mathbf{i}=(i_1,\ldots,i_p)$. Furthermore, assume that the functions $\mathbf{G}_{0,\mathbf{i}}$ are one-to-one and differentiable. Under some technical assumptions in Hannig (2009a) this will produce the generalized fiducial density of

$$f_{\mathcal{R}_\xi}(\xi)=\frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_\Xi f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')\,d\xi'}, \tag{5}$$

where

$$J(\mathbf{x},\xi)=\binom{n}{p}^{-1}\sum_{\mathbf{i}=(i_1,\ldots,i_p)}\left|\frac{\det\left(\frac{d}{d\xi}\mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)}{\det\left(\frac{d}{d\mathbf{x}_{\mathbf{i}}}\mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)}\right| \tag{6}$$

is the average of all subsets where $1\le i_1<\cdots<i_p\le n$ and the determinants in (6) are the appropriate Jacobians.

3. Main results

3.1. Structural equation with constant variance

In a multiple comparison problem we have k populations with means $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_k)$. Data, which follows an independent normal distribution, is of the form $\mathbf{X}_i=(X_{i1},\ldots X_{in_i})$ for all $i=1,\ldots,k$ where $\mathbf{X}_i$ is independent of $\mathbf{X}_j$ for all i and j. We are interested in the k treatment means. We would like to make some judgement on the equality or inequality of the means within competing models.

For example if $\mathbf{X}_i=(X_{i1},\ldots X_{in_i})$ is an independent random sample from a $N(\mu_i,\eta)$ distribution for i=1,2 then the appropriate models would either assume $\mu_1=\mu_2$ or $\mu_1\neq\mu_2$. The structural equations in this case could be

$$X_{1j}=(\mu_1+\sqrt{\eta}Z_{1j})I_{\mu_1=\mu 2}+(\mu_1+\sqrt{\eta}Z_{1j})I_{\mu_1\neq\mu_2},$$

$$X_{2j}=(\mu_2+\sqrt{\eta}Z_{2j})I_{\mu_1=\mu 2}+(\mu_2+\sqrt{\eta}Z_{2j})I_{\mu_1\neq\mu_2},$$

where Z_{ij} are independent random variables from the N(0, 1) distribution. From these structural equations the generalized fiducial density in (5) can be calculated for each model ($\mu_1=\mu_2$ and $\mu_1\neq\mu_2$).

To simplify notation we will use $J=U_1|U_2|\ldots|U_{t_J}$ where U_i is a collection of indexes of the means that are equal. The means indexed by U_i and U_j separated by a vertical bar "|" are unequal. For example when k=3, if $J=123$ then $U_1=123$ signifies $\mu_1=\mu_2=\mu_3=\mu_1^*$. If $J=1\ 2|3$ then $U_1=1\ 2$ and $U_2=3$ signify $\mu_1=\mu_2=\mu_1^*$ and $\mu_3=\mu_2^*$ where $\mu_1^*\neq\mu_2^*$. Note that there are u_i equal means in group U_i, t_J total groupings in J, and the unique means are $(\mu_1^*,\mu_2^*,\ldots,\mu_t^*)$.

In general, if $X_{i1},\ldots,X_{in_i}$ is an independent random sample from a $N(\mu_i,\eta)$ distribution for $i=1,\ldots,k$ then a structural equation is

$$X_{ij}=\sum_{J\in\{J_1,\ldots,J_H\}}(\mu_i+\sqrt{\eta}Z_{ij})I_J, \tag{7}$$

where $I_J=1$ if grouping J is selected and 0 otherwise, the equality of $\mu_i=\mu_j$ follow the grouping in J for all possible groupings $\{J_1,\ldots J_H\}$, and Z_{ij} are independent random variables from the N(0, 1) distribution.

As explained below, the fiducial distribution in (4) based on the structural equations above, will favor the model with the most free means (all unequal means). To compensate for this we need to introduce additional structural equations that are independent of those in (7). These structural equations will allow us to introduce a weight function that down-weights the models with many free means.

From Eq. (4) we can see that the generalized fiducial distribution is calculated by taking p (number of parameters) structural equations and conditioning on the fact that the remaining equations occurred. As a result, when there are more parameters there are less equations that will be part of the conditioning or, equivalently, less conditions have to be satisfied. In this case we have N structural equations ($N=\sum_{i=1}^{k}n_i$). If all of the means are different ($J=1|2|3|\ldots|k$) then $p=k+1$ ($\xi=(\mu_1,\ldots,\mu_k,\eta)$) and we condition on $N-(k+1)$ events. If all of the means are equal ($J=123\ldots k$) then $p=2(\xi=(\mu,\eta))$ and we condition on $N-2$ events. Clearly as more means are grouped together there are more conditions that need to be satisfied. In order to offset this unbalanced conditioning we will introduce additional structural equations that are independent of our original structural equations as proposed in Hannig and Lee (2009). These additional structural equations will balance out the number of conditions that need to be met for each selected J.

As noted, adding additional structural equations allows us to down-weight the models with more free means to increase the likelihood of grouping several means together. Additionally, we used the weight function introduced by the additional structural equations to make the fiducial distribution more scale invariant. Attempting to make the method scale invariant in this fashion is rather ad hoc but seemed to work well in simulations and we can show that our method is asymptotically scale invariant.

The additional structural equations are:

$$W\left(\frac{\overline{MSX}N}{2\pi}\right)=\beta_i+P_i\quad\text{if } i\ge t_J,$$

$$W\left(\frac{\overline{MSX}N}{2\pi}\right)=P_i\quad\text{if } i<t_J, \tag{8}$$

where $\overline{MSX} = k^{-1}\sum_{i=1}^{k} MSX_i$, MSX_i is the maximum likelihood estimate of the variance for group i, P_i is an independent $\chi^2(1)$ random variable for all i, $W(z)$ is the Lambert W function, and t_J is the number of groupings in a given J. Because of the independence these structural equations will not affect the distribution of $\mathbf{X}$ but they will affect the conditional distribution in (4). When inverting the structural equations in (8), if $i \geq t_J$ we can choose a β_i for any P_i so that the equation is satisfied. Thus, conditioning on this equation will not effect the conditional distribution. If $i < t_J$ then $P_i = W((\overline{MSX}N)/(2\pi))$ which creates an additional condition to be met. Combining the additional conditions with the original structural equations there will always be $N-2$ conditions regardless of the grouping of the means. This will define the weight function as

$$w_J(\mathbf{x}) = \prod_{i<t_J} f(P_i) = \left(\frac{1}{\overline{MSX}N}\right)^{(t_J-1)/2},$$

where f is the density of the $\chi^2(1)$ distribution.

Using the original structural equations, combined with the additional structural equations, the generalized fiducial distribution (4) has a density given by

$$f(\xi) \propto \sum_{J\in\{J_1,\ldots,J_H\}} \tilde{f}_J(\xi) w_J(\mathbf{x}) I_J,$$

where $\tilde{f}_J(\xi)$ is the numerator in (5) and will be computed for all groupings. This numerator for a grouping, J, is

$$\tilde{f}_J(\xi) = \frac{V_{x,J}}{\eta}\frac{1}{(2\pi)^{n_1/2}\eta^{n_1/2}}\exp\left\{-\frac{1}{2\eta}\sum_{j=1}^{n_1}(x_{1j}-\mu_1)^2\right\}\cdots\times\frac{1}{(2\pi)^{n_k/2}\eta^{n_k/2}}\exp\left\{-\frac{1}{2\eta}\sum_{j=1}^{n_k}(x_{kj}-\mu_k)^2\right\}$$
$$= V_{x,J}\frac{\eta^{-N/2-1}}{(2\pi)^{N/2}}\exp\left\{-\frac{1}{2\eta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2\right\} = V_{x,J}\frac{\eta^{-N/2-1}}{(2\pi)^{N/2}}\exp\left\{-\frac{1}{2\eta}\sum_{i=1}^{t_J} n_i'(\mu_i^*-\overline{x}_i')^2\right\}\exp\left\{-\frac{1}{2\eta}\sum_{i=1}^{t_J} n_i' MSX_i'\right\}, \quad (9)$$

where

$$J_J(\mathbf{x},\xi) = C_{N,J}^{-1}\begin{cases}\dfrac{\sum_{l=1}^{t_J}\sum_{i_1,i_2\in U_l, i_1<i_2}\sum_{j_{i_1},j_{i_1}}|x_{i_1 j_1}-x_{i_2 j_2}|}{2\eta}, & t_J<k,\\[2ex] \dfrac{\sum_{l=1}^{k}\sum_{j_1<j_2}|x_{l j_1}-x_{l j_2}|}{2\eta}, & t_J=k,\end{cases} = \frac{V_{x,J}}{\eta},$$

$$n_i' = \sum_{l\in U_i} n_l, \quad \overline{x}_i' = \frac{\sum_{l\in U_i}\sum_{j=1}^{n_l} x_{lj}}{n_i'},$$

$$MSX_i' = \frac{\sum_{l\in U_i}\sum_{j=1}^{n_l}(x_{lj}-\overline{x}_i')^2}{n_i'}, \quad N = \sum_{i=1}^{k} n_i$$

and $C_{N,J}$ is the number of Jacobian terms to average over.

If we recognize that $\mu_i^*|\eta$ follows a normal distribution for all i and η follows an inverse gamma distribution then we can integrate $\tilde{f}_J(\xi)$ over the parameter space of each grouping Ξ_J. Thus,

$$p_J = \int_{\Xi_J} f_J(\xi) w_J(\mathbf{x})\, d\xi = \frac{V_{x,J} w_J(\mathbf{x}) 2^{N/2}\pi^{t_J/2}\Gamma\left(\frac{N-t_J}{2}\right)}{(2\pi)^{N/2}\left(\sum_{i=1}^{t_J} n_i' MSX_i'\right)^{(N-t_J)/2}\prod_{i=1}^{t_J}\sqrt{n_i'}}.$$

We can find the probability that any J is correctly grouping the means by

$$P(J) = \frac{p_J}{\sum_{\tilde{J}} p_{\tilde{J}}}. \quad (10)$$

Clearly, when a particular J is correctly grouping the means we would like $P(J)$ to be large.

3.2. Structural equation with non-constant variance

Similar to the previous setup, if $X_{i1},\ldots,X_{in_i}$ is an independent random sample from a $N(\mu_i,\eta_i)$ distribution for $i=1,\ldots,k$ then a structural equation is

$$X_{ij} = \sum_{J\in\{J_1,\ldots,J_H\}} (\mu_i + \sqrt{\eta_i} Z_{ij}) I_J$$

for groupings $\{J_1,\ldots,J_H\}$ where Z_{ij} are independent random variables from the $N(0,1)$ distribution.

Like the previous section, the numerator of (5) is

$$\tilde{f}_J(\xi) = \frac{V_{x,J}}{\prod_{i=1}^k \eta_i} \frac{1}{(2\pi)^{n_1/2}\eta_1^{n_1/2}} \exp\left\{-\frac{1}{2\eta_1}\sum_{j=1}^{n_1}(x_{1j}-\mu_1)^2\right\} \cdots \times \frac{1}{(2\pi)^{n_k/2}\eta_k^{n_k/2}} \exp\left\{-\frac{1}{2\eta_k}\sum_{j=1}^{n_k}(x_{kj}-\mu_k)^2\right\}$$
$$= V_{x,J} w_J(\mathbf{x}) \frac{\prod_{i=1}^k \eta_i^{-n_i/2-1}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^k n_i \frac{((\mu_i-\overline{x}_i)^2+MSX_i)}{\eta_i}\right\}, \tag{11}$$

where

$$J_J(\mathbf{x},\xi) = C_{N,J}^{-1} \frac{\sum_{z=1}^k \sum_{j_{1,z}<j_{2,z}\le n_z} \sum_{(\mathbf{i}_1\ldots,\mathbf{i}_{t_J})} \sum_{j=(j_{1,z},j_{2,z})} |T|}{2^k \prod_{i=1}^k \eta_i} = \frac{V_{x,J}}{\prod_{i=1}^k \eta_i},$$

$\mathbf{i}_l = \{i_1,\ldots,i_{u_l-1}\} \subset U_l$ is the set of $1 \le i_1 < i_2 < \cdots < i_{u_l-1} \le u_l$, $C_{N,J}$ is the number Jacobian terms to average over,

$$T = \prod_{l=1}^{t_J}\left[\prod_{i\in\mathbf{i}_l}(x_{i,j}-\mu_l^*)\right](x_{i_{u_l},j_{1,z}}-x_{i_{u_l},j_{2,z}}),$$

$$\overline{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \quad \text{and} \quad MSX_i = \frac{\sum_{j=1}^{n_i}(x_{lj}-\overline{x}_i)^2}{n_i}.$$

As an example of the Jacobian, if $J=1|2|3$ then we average over

$$\frac{|(x_{1,j_{1,1}}-x_{1,j_{2,1}})(x_{2,j_{1,2}}-x_{2,j_{2,2}})(x_{3,j_{1,3}}-x_{3,j_{2,3}})|}{2^k \prod_{i=1}^k \eta_i}$$

for all $j_{1,z} < j_{2,z} < n_z$ combinations $(z=1,2,3)$. If $J=12|3$ then we average over

$$\frac{|(x_{1,j_{1,1}}-\mu_1^*)(x_{2,j_{1,2}}-x_{2,j_{2,2}})(x_{3,j_{1,3}}-x_{3,j_{2,3}})|}{2^k \prod_{i=1}^k \eta_i} + \frac{|(x_{1,j_{1,1}}-x_{1,j_{2,1}})(x_{2,j_{1,2}}-\mu_1^*)(x_{3,j_{1,3}}-x_{3,j_{2,3}})|}{2^k \prod_{i=1}^k \eta_i}$$

for all of the appropriate $j_{1,z}$ and $j_{2,z}$ combinations.

The weight function is derived akin to the previous explanation. Again, the weight function needed to be incorporated to offset the lack of scale invariance and to down-weight the models with many free means. The additional structural equations for each J are

$$W\left(\frac{\left(\sum_{i=1}^k \frac{b_i}{MSX_i}\right)^{1/(t_J-1)} N}{\left(\prod_{j=1}^{t_J}\sqrt{\sum_{i\in U_j}\frac{b_i}{MSX_i}}\right)^{2/(t_J-1)}}\right) = \beta_i + P_i \quad \text{if } i \ge t_J,$$

$$W\left(\frac{\left(\sum_{i=1}^k \frac{b_i}{MSX_i}\right)^{1/(t_J-1)} N}{\left(\prod_{j=1}^{t_J}\sqrt{\sum_{i\in U_j}\frac{b_i}{MSX_i}}\right)^{2/(t_J-1)}}\right) = P_i \quad \text{if } i < t_J$$

and the weight function is

$$w_J(\mathbf{x}) = \prod_{i<t_J} f(P_i) = \frac{\prod_{j=1}^{t_J}\sqrt{\sum_{i\in U_j}\frac{b_i}{MSX_i}}}{\sqrt{\sum_{i=1}^k \frac{b_i}{MSX_i}}N^{(t_J-1)/2}},$$

where $b_i = n_i/\max_j(n_j)$, MSX_i is the maximum likelihood estimate of the variance for group i, P_i is an independent $\chi^2(1)$ random variable for all i, $W(z)$ is the Lambert W function, t_J is the number of groupings in a given J, and f is the density of the $\chi^2(1)$ distribution. We can find the probability that J is correctly grouping the means, $P(J)$, using (10). However, in this case p_J cannot be calculated in closed form.

3.3. Simulations

Ideally we would like this inference method to identify the correct model at a high rate. When we assume constant variance for all of the k groups we can calculate the probabilities directly. When the variance is not assumed to be constant we used a Monte Carlo approach to generate a sample from the generalized fiducial density. We used the importance sampling algorithm in Appendix A to sample from (11) and calculate $P(J)$ for all possible groupings. Our simulation used 1000 data sets and an effective sample size of 5000 when the variance was not assumed to be constant.

3.3.1. Constant variance

Looking at a few interesting cases will help us assess the validity of the method. Fig. 1 illustrates that the correct grouping, $J=123$, is selected at a high rate. Also, the magnitude of the variance does not effect the selection probability.

Difficulties arise when the true means are relatively close together. For instance, when $\boldsymbol{\mu}_0=(1,1.5,1.5)$ or $\boldsymbol{\mu}_0=(1,1.5,2)$ the correct model is selected at a higher rate as the sample size increases. As expected, at small samples sizes our method attempts to incorrectly group means as equal. Figs. 2 and 3 reflect this.

The easiest case occurs when the means are very different. Fig. 4 demonstrates $P(J)$ when $\boldsymbol{\mu}_0=(1,3,5)$ and $\eta_0=1$.

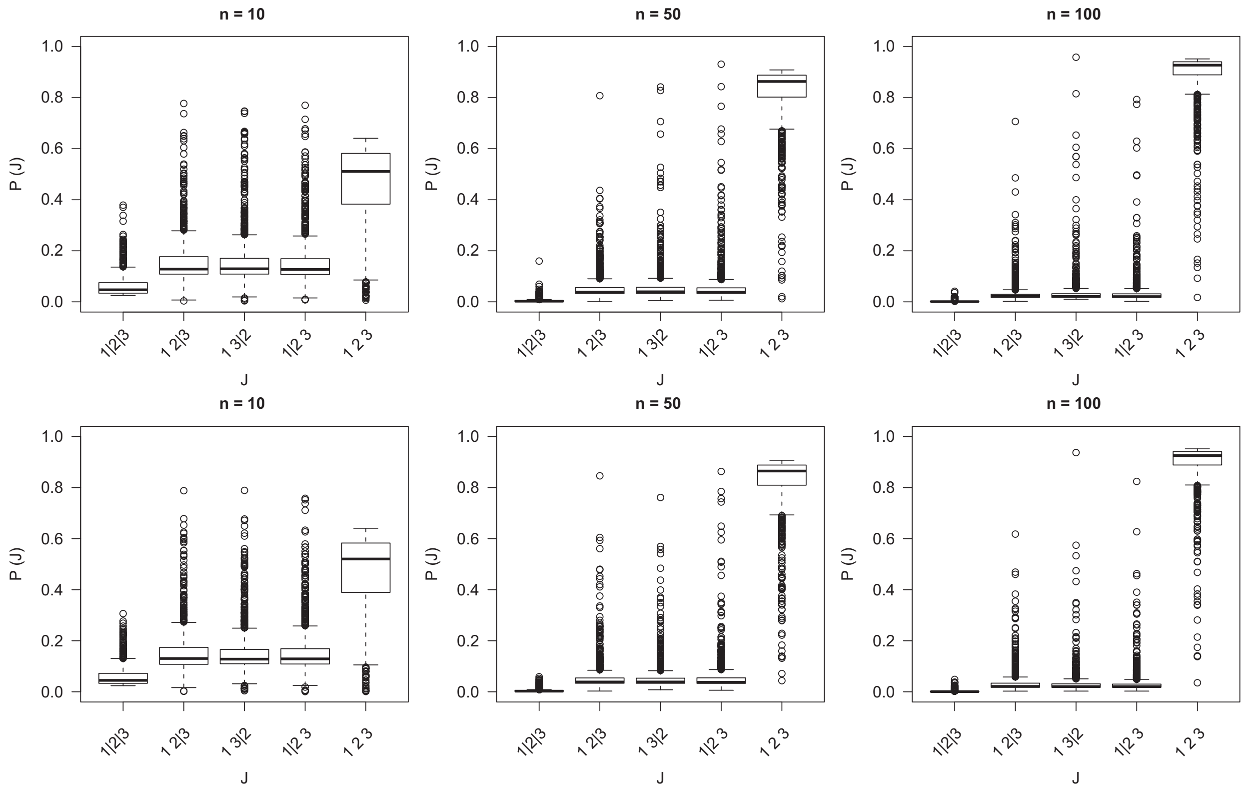

Fig. 1. $P(J)$ for $\boldsymbol{\mu}_0=(1,1,1)$ and $\eta_0=1$ and 100 for top and bottom rows respectively.

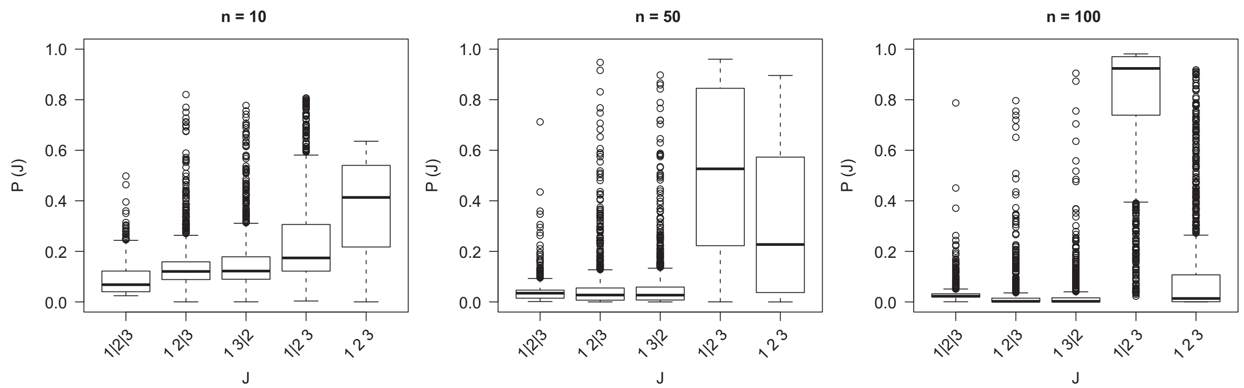

Fig. 2. $P(J)$ for $\boldsymbol{\mu}_0=(1,1.5,1.5)$ and $\eta_0=1$.

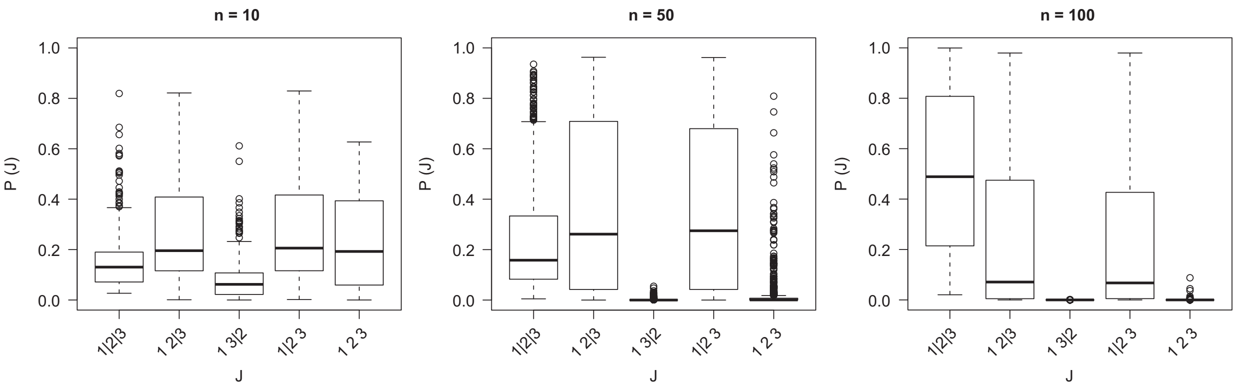

Fig. 3. $P(J)$ for $\boldsymbol{\mu}_0=(1,1.5,2)$ and $\eta_0=1$.

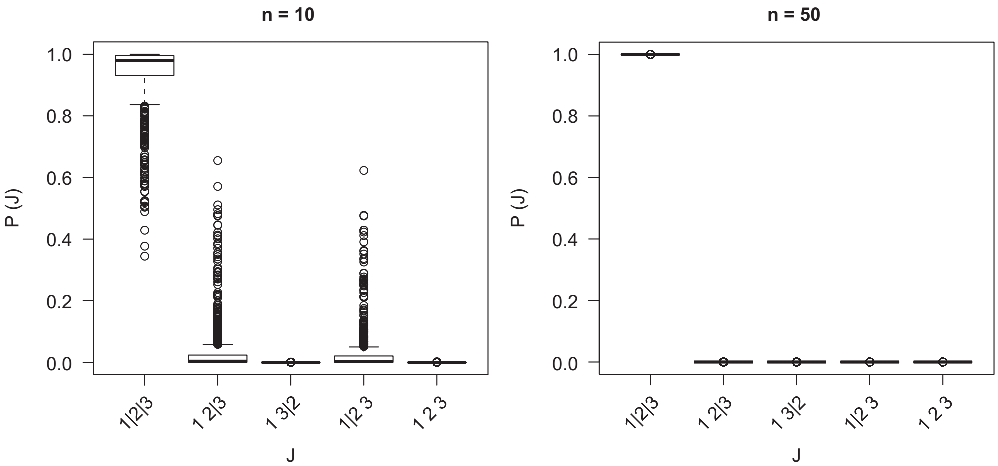

Fig. 4. $P(J)$ for $\boldsymbol{\mu}_0=(1,3,5)$ and $\eta_0=1$.

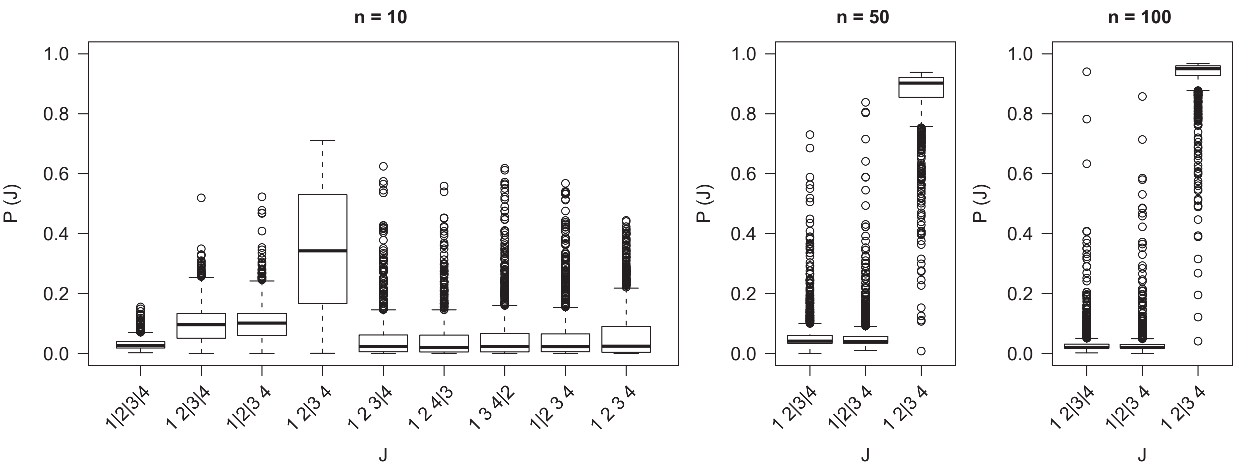

Fig. 5. $P(J)$ for $\boldsymbol{\mu}_0=(1,1,2,2)$ and $\eta_0=1$.

Similar analysis can be done at higher dimensions. Again, when $k=4$, $\boldsymbol{\mu}_0=(1,1,2,2)$, and $\eta_0=1$ our method is selecting the correct model at a high rate as the sample size increases. Fig. 5 reflects this. The omitted groupings in the figures had median probability, $P(J)$, of less than 0.02.

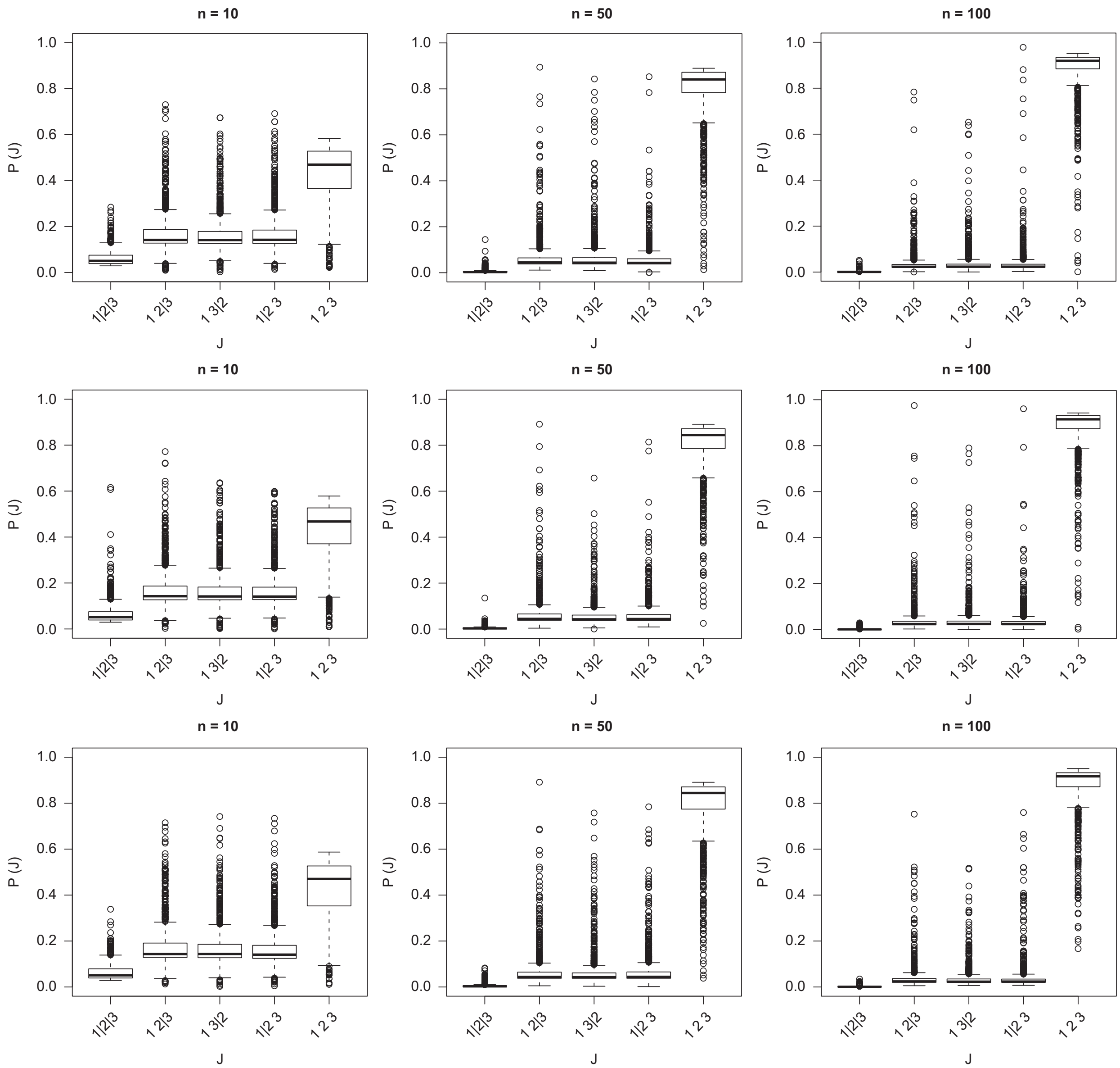

Fig. 6. $P(J)$ for $\boldsymbol{\mu}_0 = (1,1,1)$ and $\boldsymbol{\eta}_0 = (1,1,1)$, $(1,2,3)$, and $(100,100,100)$ for top, middle, and bottom rows respectively.

3.3.2. *Non-constant variance*

When variance is not assumed to be constant similar results follow. Highlighting a few we can see that the variance does not effect the probability of selecting the correct model. This is reflected in Fig. 6.

Again the easy case is when the means are very different from each other. Fig. 7 is reflective of this.

In the four dimensional simulation we can see that the correct model is being selected at a relatively high rate for all of the sample sizes. This is illustrated in Fig. 8 for all J where the median probability is greater than 0.02.

4. Asymptotic results

As defined in Eq. (10) we can calculate the probability that each J is the correct grouping. In this section we will prove that our method will asymptotically select the correct model.

Assumption 1. X_{ij} is an independent random variable from a $N(\mu_i,\eta_i)$ distribution.

Assumption 2. There exists $0 < b_i < \infty$ such that $n_i = b_i n$ for all $i = 1,\dots,k$.

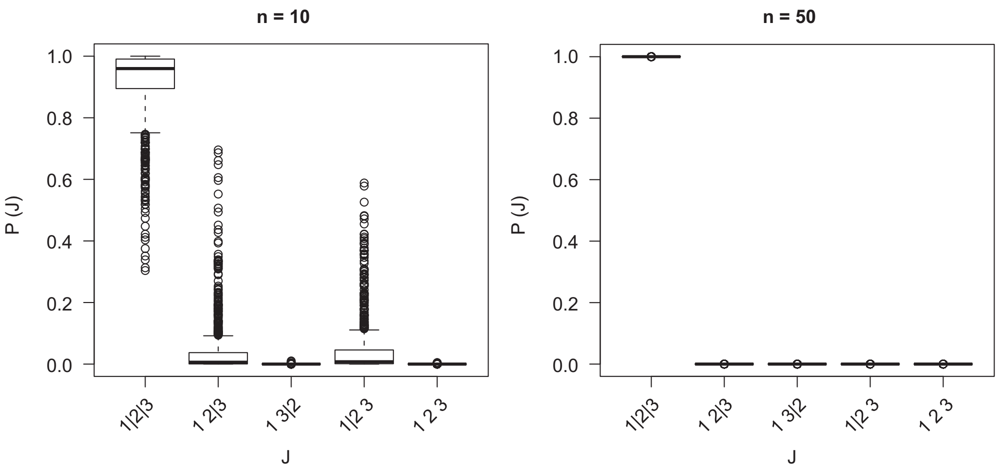

Fig. 7. $P(J)$ for $\boldsymbol{\mu}_0=(1,3,5)$ and $\boldsymbol{\eta}_0=(1,1,1)$.

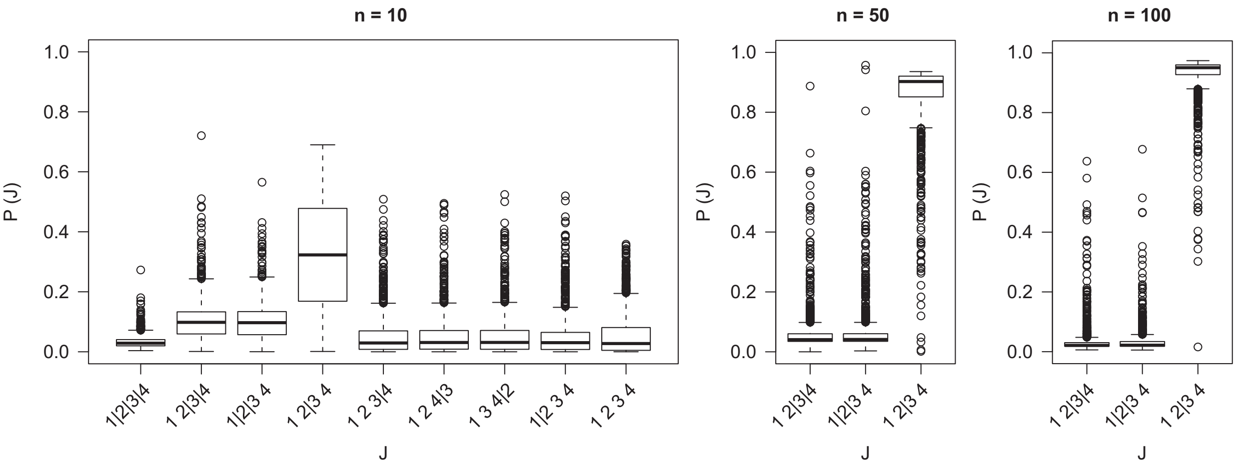

Fig. 8. $P(J)$ for $\boldsymbol{\mu}_0=(1,1,2,2)$ and $\eta_0=(1,1,1,1)$.

Theorem 1. *If J correctly groups the means then* $P(J)\to 1$ *almost surely*.

To prove this we will show that $p_{\tilde{J}}/p_J\to 0$ for any $\tilde{J}\neq J$. There are two cases that will be observed. First, when $\tilde{J}$ incorrectly groups means as equal. In this case $p_{\tilde{J}}/p_J$ will converge to zero exponentially as $n\to\infty$. The second case is when $\tilde{J}$ does not incorrectly group the means but there are too many groups. This will result in $p_{\tilde{J}}/p_J$ converging to zero polynomially as $n\to\infty$. The proof was done assuming both constant and non-constant variance. The details are relegated to Appendix B.

5. Examples

5.1. Simulated data

To further demonstrate the ability of our method we analyzed a simulated data set. This allows us to know what the true treatment means are. The sample mean and variance of the data is

$$\bar{\mathbf{x}}=(0.69,1.65,1.80,1.84)$$

and

$$\mathbf{s}^2=(1.56,1.35,1.61,2.13).$$

This data set was generated from independent normal distributions with $\boldsymbol{\mu}_0=(1,2,2,2)$, $\boldsymbol{\eta}_0=(2,2,2,2)$, and a simple size of $n=20$ for each treatment. Table 1 reflects grouping probabilities when $P(J)>0.03$. Both the constant and non-constant

Table 1
Multiple comparison *P(J)* for the simulated example.

J	*P(J)*
Constant variance	
1|23|4	0.049
1|24|3	0.051
1|2|34	0.062
12|34	0.044
1|234	0.663
1 2 3 4	0.060
Non-constant variance	
1|23|4	0.061
1|24|3	0.058
1|2|34	0.065
12|34	0.041
1|234	0.604
1 2 3 4	0.071

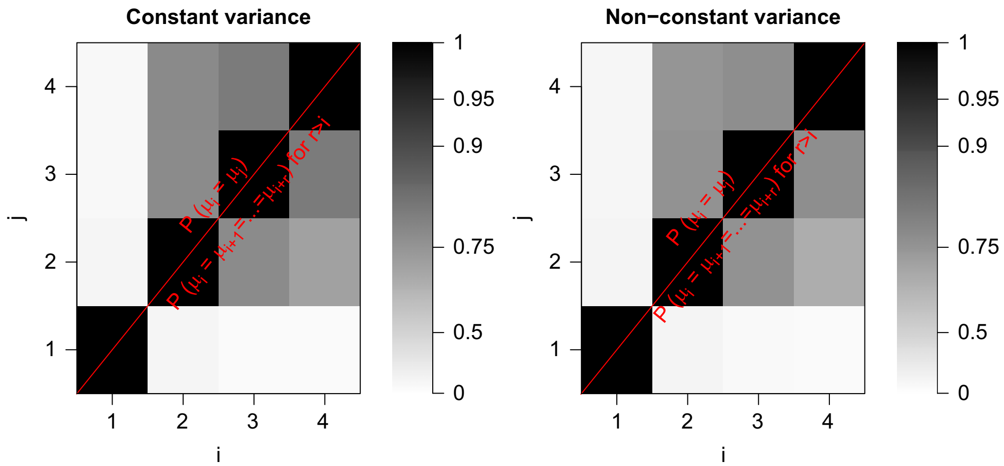

Fig. 9. $P(\mu_i=\mu_j)$ and $P(\mu_i=\mu_{i+1}=\cdots=\mu_{i+r})$ with constant and non-constant variance for the simulated example.

variance methods select the correct grouping at a high rate ($P(J)=0.663$ and $P(J)=0.604$ for $J=1|234$ when the variance is assumed to be constant and non-constant respectively).

In addition to finding the probability for each grouping the fiducial method can also find the fiducial probability of any number of means being equal. For instance, we can find the fiducial probability that any two means are equal ($\mu_i=\mu_j$) or the probability that any sequence of means are equal ($\mu_i=\mu_{i+1}=\cdots=\mu_{i+r}$). This is done by adding up probabilities for the models that $\mu_i=\mu_j$ or $\mu_i=\mu_{i+1}=\cdots=\mu_{i+r}$,

$$P(\mu_i=\mu_j)=\sum_{J\in\{J_1,\ldots,J_H\}} P(J)I_{\{J:\mu_i=\mu_j\}} \tag{12}$$

and

$$P(\mu_i=\mu_{i+1}=\cdots=\mu_{i+r})=\sum_{J\in\{J_1,\ldots,J_H\}} P(J)I_{\{J:\mu_i=\mu_{i+1}=\cdots=\mu_{i+r}\}}. \tag{13}$$

Fig. 9 pictorially represent these probabilities for the simulated example. As the pictures show it is very reasonable that $\mu_1\neq\mu_2=\mu_3=\mu_4$.

In comparison to a common frequentist method, Tukey's HSD test could not find significant differences in the means (1, 2) and (2, 3, 4) controlling the experimentwise error rate at $\alpha=0.05$. Tukey's HSD is commonly known to be rather conservative which makes it difficult to detect differences. A method described in Abdel-Karim (2005) uses a similar Tukey approach but allows for unequal variance across the treatments. This method could not find significant differences between the means (1, 2), (2, 3, 4), and (1, 4).

5.2. Clover plant data

A data set from Steele and Torrie (1980) measured the nitrogen content (in mg) of red clover plants inoculated with cultures of *Rhizobium trifolli* and the addition of *Rhizobium meliloti* strains. As discussed in Gopalan and Berry (1998), the *R. trifolli* was tested with a composite of five alpha strains (3DOk1, 3DOk4, 3DOk5, 3DOk7, 3DOk13), *R. meliloti*, and a composite of the alpha strains. There were six treatments in all. The goal of the experiment was to measure the nitrogen levels for the different treatments. The data can be seen in Table 2.

We analyzed this data set using both the constant and non-constant variance methods. The grouping probabilities are seen in Table 3 when $P(J)>0.03$. If we assume that the variance is constant $J=12|34|5|6$ is the most likely scenario. If we do not assume that the variance is constant the most likely grouping is $J=12|34|56$. Looking at the sample means and standard deviations both of these results seem very reasonable.

The Bayesian method described in Gopalan and Berry (1998) analyzed this data set with the constant variance assumption. Prior distributions were selected for the parameters using various distributions; the groupings used a Dirichlet process prior. Table 4 illustrates a few highlighted posterior probabilities. They claim, if the posterior probabilities are large in comparison to the prior probabilities for all values of M (Dirichlet process prior parameter) then these are likely groupings of the means. The resulting groupings in Table 4 are their recommended groupings.

Similarities between our analysis and theirs exist. $J=12|34|5|6$ and $12|34|56$ are common to all of the methods as likely groupings of the means.

Table 2
Clover plant data.

	Treatments					
	1 3DOk13	2 3DOk4	3 Composite	4 3DOk7	5 3DOk5	6 3DOk1
	14.3	17.0	17.3	20.7	17.7	19.4
	14.4	19.4	19.4	21.0	24.8	32.6
	11.8	9.1	19.1	20.5	27.9	27.0
	11.6	11.9	16.9	18.8	25.2	32.1
	14.2	15.8	20.8	18.6	24.3	33.0
Mean	13.26	14.64	18.70	19.92	23.98	28.82
SD	1.43	4.12	1.60	1.13	3.78	5.80

Table 3
Multiple comparison $P(J)$ for the red clover example.

J	$P(J)$
Constant variance	
1|2|3|4|5|6	0.037
12|3|4|5|6	0.100
1|2|34|5|6	0.052
1|2|3|45|6	0.030
12|34|5|6	0.196
12|3|45|6	0.063
12|3|4|56	0.043
1|2|34|56	0.036
12|34|56	0.078
12|345|6	0.051
Non-constant variance	
1|2|3|4|5|6	0.041
12|3|4|5|6	0.097
1|2|34|5|6	0.052
1|2|3|4|56	0.050
12|34|5|6	0.115
12|3|45|6	0.049
12|3|4|56	0.102
1|2|34|56	0.058
12|34|56	0.139
12|345|6	0.042

Table 4
Posterior probabilities of select *J* for the red clover example.

J	*M*								
	0.334	0.733	1.373	1.956	2.605	3.462	4.909	9.13	19.88
	Posterior probabilities								
1234|5|6	0.032	0.059	0.057	0.046	0.041	0.034	0.026	0.013	0.005
12|345|6	0.186	0.211	0.202	0.189	0.171	0.148	0.112	0.061	0.020
12|34|56	0.144	0.178	0.167	0.149	0.137	0.116	0.094	0.049	0.016
12|34|5|6	0.037	0.086	0.152	0.199	0.229	0.257	0.276	0.281	0.199

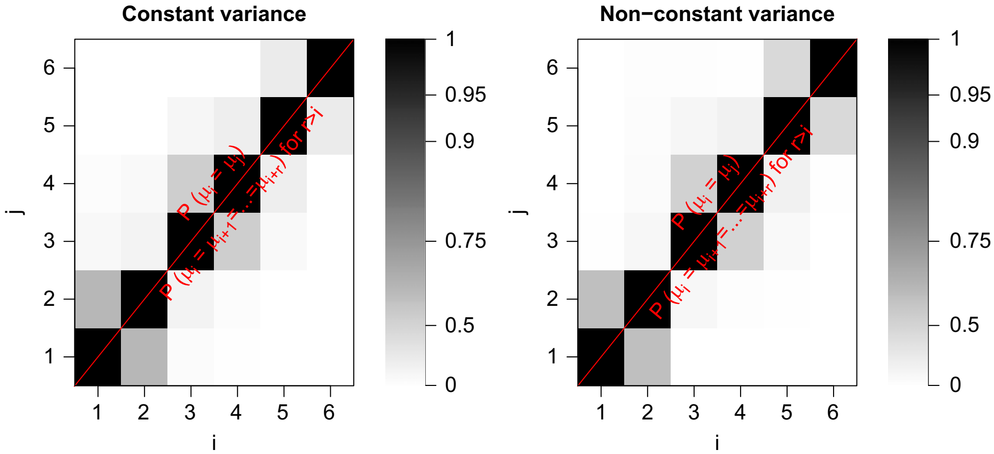

Fig. 10. $P(\mu_i=\mu_j)$ and $P(\mu_i=\mu_{i+1}=\cdots=\mu_{i+r})$ with constant and non-constant variance for the red clover example.

Fig. 10 pictorially represent the probabilities in Eqs. (12) and (13) for the red clover plant example. As the pictures show it is very reasonable that $\mu_1=\mu_2$, $\mu_3=\mu_4$, and possibly $\mu_5=\mu_6$. Tukey's HSD test could not find significant differences in the means (1, 2, 3, 4), (3, 4, 5), or (5, 6) and the method in Abdel-Karim (2005) could not detect differences in (1, 2) or (2, 3, 4, 5, 6) using an experimentwise error rate of $\alpha=0.05$.

6. Conclusion

Frequentist solutions for multiple comparison problems can test for a treatment effect or find differences among treatments. However, they cannot make a determination as to how reasonable it is that particular means are grouped together as equal.

Using a fiducial inference approach we have developed a method to determine the likelihood of grouping means together. Based on simulation results, our method selects the correct grouping at a relatively high rate for small sample sizes.

We analyzed a simulated data set and a data set that measured the nitrogen levels of red clover plants that were inoculated with six different treatments. The analysis of the simulated data set yielded a high probability for the correct model $(\mu_1\neq\mu_2=\mu_3=\mu_4)$ regardless of the variance assumptions. When analyzing the red clover example under the assumption of constant variance we found that $J=12|34|5|6$ was the most likely grouping of the means ($P(J)=0.196$). The Bayesian solution also found that grouping to be reasonable, however, no discernible probability could be assigned to it. Additionally, our method found that $J=12|34|56$ was the most likely grouping if the variance was not assumed to be constant ($P(J)=0.139$).

The fiducial method is an interesting solution to the multiple comparison problem. The intuitive feel of the fiducial probability for each model makes the interpretation very straightforward and the asymptotic properties and simulation results assure high confidence in the analysis.

Acknowledgments

The authors are thankful to the anonymous referees whose thoughtful comments have led to a great improvement in the exposition of the manuscript. The author's research was supported in part by the National Science Foundation grants DMS 0707037 and DMS 1007543.

Appendix A. Importance sampling algorithm

The following steps were implemented in order to obtain a fiducial sample for ξ.

1. For a particular J, start by generating $\mu_i^* = \overline{\overline{x}}_i + \sqrt{\min_{j\in U_i} MSX_j/(\overline{n}_i-1)}T_{df}$ where $T_{df} \sim t(\overline{n}_i-1)$, $\overline{\overline{x}}_i = u_i^{-1}(\sum_{j=1}^{u_i} \overline{x}_j)$, and $\overline{n}_i = u_i^{-1}(\sum_{j=1}^{u_i} n_j)$ for all $i=1,\dots,t_J$.
2. Note, that $J = U_1|U_2|\dots|U_{t_J}$ where $U_i = r_{i1}\, r_{i2} \dots r_{iu_i}$ are the indexes of equal means

$$\boldsymbol{\mu} = (\mu_{r_{11}},\dots\mu_{r_{1u_1}},\mu_{r_{21}},\dots\mu_{r_{2u_2}},\dots,\mu_{r_{t1}},\dots\mu_{r_{2u_{t_J}}}) = \left(\underbrace{\mu_1^*,\dots\mu_1^*}_{u_1 \text{ replictes}}, \underbrace{\mu_2^*,\dots\mu_2^*}_{u_2 \text{ replictes}},\dots,\underbrace{\mu_{t_J}^*,\dots\mu_{t_J}^*}_{u_{t_J} \text{ replictes}} \right).$$

   Generate $(\eta_l|\mu_l) = W$ where $W \sim Inv\text{-}Gamma(n_l/2,(n_l((\mu_l-\overline{x}_l)^2+MSX_l))/2)$ for $l=1,\dots,k$.
3. Calculate weights of each generated sample with,

$$w_J = \frac{f_J(\xi)}{(\prod_{i=1}^{t_J} g_i(\mu_i^*)h_i(\eta_i))},$$

   where $f_J(\xi)$ is the generalized fiducial density for the model with groupings J and $g_i(\mu_i)$ and $h_i(\eta_i)$ are the densities from distributions described in steps 1, and 2.
4. This process was repeated until we achieved the effective sample size calculated by $ESS_J = n_J(1+(s_{w_J}^2)\overline{w}_J^{-2})^{-1}$ where n_J is the sample size for model J, $s_{w_J}^2$ is the sample variance of the weights, and $\overline{w}_J$ is the sample mean of the weights.
5. Lastly the weights were divided by the ESS_J.
6. This process was repeated for all J.

Appendix B. Proof of Theorem 1

With constant variance: This proof will be done with the assumption of constant variance (i.e. $\eta_i = \eta_j$ for all i and j) and without using the weight function, $w_J(\mathbf{X})$. To prove Theorem 1 we will show that $p_{\tilde{J}}/p_J \to 0$ for any $\tilde{J} \neq J$. We will observe two cases. First, when $\tilde{J}$ incorrectly groups means as equal. Second, when $\tilde{J}$ does not incorrectly group the means but there are too many groups.

For the first case let J_2 incorrectly group the means and J_1 is the correct grouping. Thus, there are t_1 groups in J_1 and t_2 groups in J_2. At least one of the means in J_2 is incorrectly grouped. The subscript in the following calculations note the association with J_1 or J_2.

$$\frac{p_{J_2}}{p_{J_1}} \propto \frac{\Gamma\left(\frac{N-t_2}{2}\right)(\sum_{i=1}^{t_1} n'_{1i}MSX'_{1i})^{(N-t_1)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\Gamma\left(\frac{N-t_1}{2}\right)(\sum_{i=1}^{t_2} n'_{2i}MSX'_{2i})^{(N-t_2)/2}\prod_{i=1}^{t_2}\sqrt{n'_{2i}}} \quad \text{using Stirling's formula}$$

$$\le \frac{(2e)^{(t_2-t_1)/2}N^{(t_1-t_2)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\prod_{i=1}^{t_2}\sqrt{n'_{2i}}} \frac{(\sum_{i=1}^{t_1} n'_{1i}MSX'_{1i})^{(N-t_1)/2}}{(\sum_{i=1}^{t_2} n'_{2i}MSX'_{2i})^{(N-t_2)/2}} \quad \text{WLOG assume } U_1 \in J_2 \text{ is an incorrect grouping}$$

$$\le \frac{(2e)^{(t_2-t_1)/2}N^{(t_1-t_2)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\prod_{i=1}^{t_2}\sqrt{n'_{2i}}} \frac{\eta_0^{(t_2-t_1)/2}\left(\sum_{i=1}^{t_1} n'_{1i}\frac{\eta_0(1+O(1))}{\eta_0}\right)^{(N-t_1)/2}}{\left(n'_{21}\frac{\eta^*(1+O(1))}{\eta_0} + \sum_{i=2}^{t_2} n'_{2i}\frac{\eta_0(1+O(1))}{\eta_0}\right)^{(N-t_2)/2}}$$

where $\eta^* > \eta$ because of the incorrect grouping

$$\le \frac{(2e)^{(t_2-t_1)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\prod_{i=1}^{t_2}\sqrt{n'_{2i}}} \frac{\eta_0^{(t_2-t_1)/2}\left(1+\frac{r}{4}\left(\frac{\eta^*}{\eta_0}-1\right)\right)^{(N-t_1)/2}}{\left(\left[1+r\left(\frac{\eta^*}{\eta_0}-1\right)\right]c\right)^{(N-t_2)/2}} \quad \text{Eventually a.s.}$$

$$\to 0 \quad \text{a.s.}$$

for

$$c=\frac{1+\left(1+r\left(\frac{\eta^*}{\eta_0}-1\right)\right)}{2\left(1+r\left(\frac{\eta^*}{\eta_0}-1\right)\right)}$$

and $0<r<\sum_{i\in U_1} b_i(\sum_{i=1}^k b_i)^{-1}<1$.

The second case when J_2 is a valid model with too many groups and J_1 is the correct grouping. Thus, there are t_1 groups in J_1, t_2 groups in J_2 and $t_2>t_1$. Let

$$J_1=U_{11}|U_{12}|\cdots|U_{1t_1},$$

$$J_2=U_{21}|U_{22}|\cdots|U_{2t_2},$$

where

$$U_{1i}=\bigcup_{k\in K_i} U_{2k}$$

and $K_i\subset\{1,\ldots,t_2\}$ for at least one U_{1i}.

$$\begin{aligned}
\frac{p_{J_2}}{p_{J_1}} &\propto \frac{\Gamma\left(\frac{N-t_2}{2}\right)(\sum_{i=1}^{t_1} n'_{1i}MSX'_{1i})^{(N-t_1)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\Gamma\left(\frac{N-t_1}{2}\right)(\sum_{i=1}^{t_2} n'_{2i}MSX'_{2i})^{(N-t_2)/2}\prod_{i=1}^{t_2}\sqrt{n'_{2i}}}\\
&\le \frac{(2e)^{(t_2-t_1)/2}N^{(t_1-t_2)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\prod_{i=1}^{t_2}\sqrt{n'_{2i}}}\frac{(\sum_{i=1}^{t_1} SSX'_{1i})^{(N-t_1)/2}}{(\sum_{i=1}^{t_2} SSX'_{2i})^{(N-t_2)/2}}\\
&\le \frac{(2e)^{(t_2-t_1)/2}(2\eta)^{(-t_1+t_2)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\prod_{i=1}^{t_2}\sqrt{n'_{2i}}}\left(1-\frac{\left(\sum_{i=1}^{t_1}\sum_{k\in K_i} n_{2k}\frac{16\eta\ \log\ \log N}{n_{2k}}\right)}{N\eta}\right)^{-N/2}\\
&\quad\text{eventually a.s. using the law of iterated logarithms}\\
&\le \frac{(2e)^{(t_2-t_1)/2}(2\eta)^{(-t_1+t_2)/2}\prod_{i=1}^{t_1}\sqrt{n'_{1i}}}{\prod_{i=1}^{t_2}\sqrt{n'_{2i}}}\left(1-\frac{R\log\log N}{N}\right)^{-N/2}\\
&\quad\text{WLOG assume that } U_{1t_1}=U_{2(t_2-1)}\cup U_{2t_2} \text{ and } U_{1i}=U_{2i} \text{ for all other } i\\
&\le \frac{(2e)^{(t_2-t_1)/2}(2\eta)^{(-t_1+t_2)/2}b}{\sqrt{n'_{2t_2}}}\left(1-\frac{R\,\mathrm{loglog}\,N}{N}\right)^{-N/2}\\
&\to 0 \quad \text{a.s.}
\end{aligned}$$

for some $R>1$ and $b>0$.

Therefore, we have shown that $p_{\tilde{J}}/p_J\to 0$ for any $\tilde{J}\neq J$ where J is the correct grouping. This completes the proof. □

With non-constant variance: This proof will not assume constant variance. Additionally, the proof will be done without the use of the weight function. The generalized fiducial density for any J without the weight function is

$$\begin{aligned}
\tilde{f}_J(\xi) &= \frac{V_{xJ}}{\prod_{i=1}^k \eta_i}\frac{1}{(2\pi)^{n_1/2}\eta_1^{n_1/2}}\ \exp\left\{-\frac{1}{2\eta_1}\sum_{j=1}^{n_1}(x_{1j}-\mu_1)^2\right\}\cdots\times\frac{1}{(2\pi)^{n_k/2}\eta_k^{n_k/2}}\ \exp\left\{-\frac{1}{2\eta_k}\sum_{j=1}^{n_k}(x_{kj}-\mu_k)^2\right\}\\
&= V_{xJ}\frac{\prod_{i=1}^k \eta_i^{-n_i/2-1}}{(2\pi)^{N/2}}\ \exp\left\{-\frac{1}{2}\sum_{i=1}^k n_i\frac{((\mu_i-\overline{x}_i)^2+MSX_i)}{\eta_i}\right\}.
\end{aligned}$$

If we could integrate this function we could calculate the probabilities, $P(J)$, directly. However, we cannot fully integrate it so we will apply different techniques. Note, that $J=U_1|U_2|\cdots|U_{t_J}$ where $U_i=r_{i1}\ r_{i2}\cdots r_{iu_i}$ are the indexes of equal means

$$\boldsymbol{\mu}=(\mu_{r_{11}},\ldots\mu_{r_{1u_1}},\mu_{r_{21}},\ldots\mu_{r_{2u_2}},\ldots,\mu_{r_{t1}},\ldots\mu_{r_{2u_{t_J}}})=\left(\underbrace{\mu_1^*,\ldots\mu_1^*}_{u_1\ \text{replictes}},\underbrace{\mu_2^*,\ldots\mu_2^*}_{u_2\ \text{replictes}},\ldots,\underbrace{\mu_{t_J}^*,\ldots\mu_{t_J}^*}_{u_{t_J}\ \text{replictes}}\right).$$

Without loss of generality we will assume $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_k)=(\boldsymbol{\mu}_1^*,\ldots,\boldsymbol{\mu}_{t_J}^*)=\boldsymbol{\mu}^*$. Notice that

$$p_J=\int_{\Xi} f_J(\xi)\,d\xi=\pi^{-N/2}\prod_{i=1}^k\left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]\int_{\mathbb{R}^{t_J}} V_{xJ}\prod_{i=1}^k((\mu_i-\overline{x}_i)^2+MSX_i)^{-n_i/2}\ d\boldsymbol{\mu}^*.$$

Because V_{xJ} is dependent on μ_i^* we will bound this value. It is clear that $V_{xJ}>c_1$ for some $c_1>0$. We could re-write V_{xJ} as

$$V_{xJ}=\left|\left(\prod_{i=1}^{t_J}\mu_i^{*(u_i-1)}\right)V_{1,1}+\sum_{j=1}^{t_J}\left(\mu_j^{*(u_j-2)}\prod_{i=1,i\neq j}^{t_J}\mu_i^{*(u_i-1)}\right)V_{2,j}+\cdots+V_{z,1}\right|,$$

where $V_{i,j}$ are averages over a function of the data. If u_i is even then $|\mu_i^{*(u_i-1)}| \le \mu_i^{*u_i}+1$ and if u_i is odd then $|\mu_i^{*(u_i-1)}| = \mu_i^{*(u_i-1)}$. Regardless of the u_i the same technique will be used. Thus, without loss of generality we will assume that u_i is odd for all i:

$$V_{x,J} \le \left(\prod_{i=1}^{t_J} \mu_i^{*(u_i-1)}\right)|V_{1,1}| + \sum_{j=1}^{t_J}\left((\mu_j^{*(u_j-1)}+1)\prod_{i=1,i\neq j}^{t_J}\mu_i^{*(u_i-1)}\right)|V_{2,j}| + \cdots + |V_{z,1}|$$

$$\le \left(\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{(u_i-1)/2}\right)|V^{(1)}|+|V^{(2)}|,$$

where $V^{(1)}$ and $V^{(2)}$ are averages over the data, $\overline{x}_i$, and MSX_i. Thus $V^{(1)}$ and $V^{(2)}$ will converge to some constant almost surely by the strong law of large numbers.

A lower bound for p_J is

$$p_J \ge p_J^{\blacktriangledown} = c_1\pi^{-N/2}\prod_{i=1}^{k}\left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]\int_{\mathbb{R}^{t_J}}\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{-n_i/2}d\boldsymbol{\mu}^*.$$

An upper bound for p_J is

$$p_J \le \pi^{-N/2}\prod_{i=1}^{k}\left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]\int_{\mathbb{R}^{t_J}}\frac{\left(\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{(u_i-1)/2}\right)|V^{(1)}|+|V^{(2)}|}{\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{n_i/2}}\,d\boldsymbol{\mu}^*$$

$$\le \pi^{-N/2}\prod_{i=1}^{k}\left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]\int_{\mathbb{R}^{t_J}}\frac{|V^{(1)}|}{\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{(n_i-u_i-1)/2}}+\frac{|V^{(2)}|}{\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{n_i/2}}d\boldsymbol{\mu}^*$$

$$\le \pi^{-N/2}\prod_{i=1}^{k}\left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]\int_{\mathbb{R}^{t_J}}\frac{c_2|V^{(1)}|}{\prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{(n_i-u_i-1)/2}}\,d\boldsymbol{\mu}^* = p_J^{\blacktriangle}$$

for some $c_2>0$.

Because we cannot integrate with respect to $\boldsymbol{\mu}^*$ we observe

$$g_J(\boldsymbol{\mu}^*) = \prod_{i=1}^{k}((\mu_i-\overline{x}_i)^2+MSX_i)^{-n_i/2}$$

with the transformations of

$$\mu_i^* = \frac{m_i^*}{\sqrt{n}}+\mu_{i0}^* \quad \text{for } i=1,\ldots,t_J$$

and the substitutions of

$$\overline{x}_i = \mu_{i0}+\frac{Z_{i1}}{\sqrt{n_i}} \quad \text{for } i=1,\ldots,k$$

and

$$MSX_i = \eta_{i0}+\frac{Z_{i2}}{\sqrt{n_i}} \quad \text{for } i=1,\ldots,k,$$

where μ_{i0} and η_{10} are the true mean and variance for treatment i and $(Z_{i1},Z_{i2})\sim N(\mathbf{0},\Sigma)$. Thus,

$$g_J(\mathbf{m}^*) = n^{-t_J/2}\prod_{i=1}^{k}\left(\left(\frac{m_i}{\sqrt{n}}+\Delta_i-\frac{Z_{i1}}{\sqrt{n_i}}\right)^2+\eta_{i0}+\frac{Z_{i2}}{\sqrt{n_i}}\right)^{-n_i/2},$$

where $\mathbf{m}$ and $\mathbf{m}^*$ follows the same structure as $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$ previously stated and $\Delta_i = \mu_{j0}^*-\mu_{i0}$ for $i\in U_j$. We will see that m_i^* converges point-wise to a normal distribution.

Taylor expanding $\log(g_J(\mathbf{m}^*))$ we will get

$$\log(g_J(\mathbf{m}^*)) = -\frac{t_J}{2}\log(n) + \sum_{i=1}^{k}\left[-\frac{b_i n \log(\eta_{i0}+\Delta_i^2)}{2} - \frac{b_i\sqrt{n}\left(2\Delta_i\left(m_i - \frac{Z_{i1}}{\sqrt{b_i}}\right) + \frac{Z_{i2}}{\sqrt{b_i}}\right)}{2(\eta_{i0}+\Delta_i^2)}\right.$$

$$\left. + \frac{b_i\left(2\Delta_i\left(m_i - \frac{Z_{i1}}{\sqrt{b_i}}\right) + \frac{Z_{i2}}{\sqrt{b_i}}\right)^2}{4(\eta_{i0}+\Delta_i^2)^2} - \frac{b_i\left(m_i - \frac{Z_{i1}}{\sqrt{b_i}}\right)^2}{(2\eta_{i0}+\Delta_i^2)} + O(n^{-1/2})\right].$$

Clearly if J is correctly grouping the means then $\Delta_i = 0$. Otherwise we will select μ_{j0}^* such that

$$\sum_{i \in U_j} \frac{b_i \Delta_i}{(\eta_{i0}+\Delta_i^2)} = 0$$

for all $j = 1,\dots,t_J$. Thus,

$$\log(g_J(\mathbf{m}^*)) = -\frac{t_J}{2}\log(n) + \sum_{i=1}^{k}\left[-\frac{b_i n \log(\eta_{i0}+\Delta_i^2)}{2} + \frac{\sqrt{b_i n}(2\Delta_i Z_{i1} - Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right.$$

$$\left. + \frac{b_i\left(2\Delta_i\left(m_i - \frac{Z_{i1}}{\sqrt{b_i}}\right) + \frac{Z_{i2}}{\sqrt{b_i}}\right)^2}{4(\eta_{i0}+\Delta_i^2)^2} - \frac{b_i t\left(m_i - \frac{Z_{i1}}{\sqrt{b_i}}\right)^2}{(2\eta_{i0}+\Delta_i^2)} + O(n^{-1/2})\right].$$

Or

$$g_J(\mathbf{m}^*) = \frac{1}{n^{t_J/2}\prod_{i=1}^{k}(\eta_{i0}+\Delta_i^2)^{b_i n/2}} \exp\left\{\sum_{i=1}^{k}\frac{\sqrt{b_i n}(2\Delta_i Z_{i1} - Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right\}\exp\left\{-\sum_{i=1}^{t_J}\frac{1}{2\sigma_{\mathbf{Z},i}^2}(m_i^* - \zeta_{\mathbf{Z},i})^2 + C_{\mathbf{Z}} + O(n^{-1/2})\right\}, \quad \text{(B.1)}$$

where $\zeta_{\mathbf{Z},i}$ and $\sigma_{\mathbf{Z},i}^2$ are the appropriate mean and variance of m_i^* dependent on the Z_{ij} values and $C_{\mathbf{Z}}$ is the constant used in completing the square. Clearly $g_J(\mathbf{m}^*)C_n$ converges to a normal density for the appropriate normalizing constant, C_n.

Lemma 1. *Let*

$$h_{J,n}(\mathbf{m}^*) = C_n g_J(\mathbf{m}^*)$$

for the previously described C_n, then $h_{J,n}(\mathbf{m}^*) \le k_J(\mathbf{m}^*)$ where k is integrable.

Proof. First, square the function, g, without the n power:

$$\prod_{i=1}^{k}\left(\left(\frac{m_i}{\sqrt{n}} + \Delta_i - \frac{Z_{i1}}{\sqrt{n_i}}\right)^2 + \eta_{i0} + \frac{Z_{i2}}{\sqrt{n_i}}\right)^{-b_i}.$$

For each U_j we are looking at a function in m_j^* that has at most u_j peaks (local maximum) and at most $u_j - 1$ valleys (local minimum). For instance, for U_1 we are looking at the function

$$\prod_{i \in U_1}\left(\left(\frac{m_1^*}{\sqrt{n}} + \Delta_i - \frac{Z_{i1}}{\sqrt{n_i}}\right)^2 + \eta_{i0} + \frac{Z_{i2}}{\sqrt{n_i}}\right)^{-b_i}.$$

For large enough n this function has a unique global maximum with probability 1. Our μ_{j0}^* is close to this maximum (within $c_{j0}n^{-1/2}$ where the c_{j0} depends on the Z's). Next, we re-scale the function so that the global maximum is 1.

The other local maximums will be $c_{jl}n^{1/2}$ away where $l = 1,\dots,(u_j - 1)$. Here c_{jl} depends on the distances between maximums.

The fraction between the value of the function's local and global maximums is either a constant (<1 if $\eta_{s0} \ne \eta_{r0}$ for all $s,r \in U_j$) or $1 - c_{jl}n^{-1/2}$ if $\eta_s = \eta_r$ for all $s,r \in U_j$, in which case the difference comes from the Z's.

Finally, if we raise the function to the power n. The global maximum goes to 1. At the local maximum we have a height of $\exp\{-c_{jl}n^{1/2}\}$, i.e., the local maximum is located at the point $(c_{jl}n^{1/2}, \exp\{-c_{jl}n^{1/2}\})$ which is well below the Cauchy density of $(c_{jl}n^{1/2}, c(1 + c_{jl}^2 n)^{-1})$ for some constant c.

Finally, notice that if there was a point for which our function was larger than a bounding Cauchy it would be at the local maximum. This is because the function decays from its local and global maxima faster than the Cauchy distribution.

From Eq. (B.1) we can see m_i^* converges point-wise to a normal distribution and Lemma 1 allows us to bound $g_J(\mathbf{m}^*)$ for all n. Therefore, we can calculate the asymptotic behavior of $p_J^{\blacktriangle}$ and $p_J^{\blacktriangledown}$. Observe,

$$\begin{aligned} p_J^{\blacktriangledown} &= \frac{\pi^{-N/2}\prod_{i=1}^k \left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]c_1}{n^{t_J/2}(\eta_{i0}+\Delta_i^2)^{b_i n/2}} \exp\left\{\sum_{i=1}^k \frac{\sqrt{b_i n}(2\Delta_i Z_{i1}-Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right\} \\ &\times \int_{\mathbb{R}^{t_J}} \exp\left\{\frac{b_i\left(2\Delta_i\left(m_i-\frac{Z_{i1}}{\sqrt{b_i}}\right)+\frac{Z_{i2}}{\sqrt{b_i}}\right)^2}{4(\eta_{i0}+\Delta_i^2)^2}-\frac{b_i\left(m_i-\frac{Z_{i1}}{\sqrt{b_i}}\right)^2}{(2\eta_{i0}+\Delta_i^2)}+O(n^{-1/2})\right\}d\mathbf{m}^* \\ &= \frac{\pi^{-N/2}\prod_{i=1}^k \left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]c_1}{n^{t_J/2}(\eta_{i0}+\Delta_i^2)^{b_i n/2}} \exp\left\{\sum_{i=1}^k \frac{\sqrt{b_i n}(2\Delta_i Z_{i1}-Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right\}B_{1,n} \end{aligned}$$

and a similar calculation produces

$$\begin{aligned} p_J^{\blacktriangle} &= \frac{\pi^{-N/2}\prod_{i=1}^k \left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]c_2 V^{(1)}}{n^{t_J/2}(\eta_{i0}+\Delta_i^2)^{(b_i n-u_i-1)/2}} \exp\left\{\sum_{i=1}^k \frac{\sqrt{b_i n}(2\Delta_i Z_{i1}-Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right\} \\ &\times \int_{\mathbb{R}^{t_J}} \exp\left\{\frac{b_i\left(2\Delta_i\left(m_i-\frac{Z_{i1}}{\sqrt{b_i}}\right)+\frac{Z_{i2}}{\sqrt{b_i}}\right)^2}{4(\eta_{i0}+\Delta_i^2)^2}-\frac{b_i\left(m_i-\frac{Z_{i1}}{\sqrt{b_i}}\right)^2}{(2\eta_{i0}+\Delta_i^2)}+O(n^{-1/2})\right\}d\mathbf{m}^* \\ &= \frac{\pi^{-N/2}\prod_{i=1}^k \left[n_i^{-n_i/2}\Gamma\left(\frac{n_i}{2}\right)\right]c_2 V^{(1)}}{n^{t_J/2}(\eta_{i0}+\Delta_i^2)^{(b_i n-u_i-1)/2}} \exp\left\{\sum_{i=1}^k \frac{\sqrt{b_i n}(2\Delta_i Z_{i1}-Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right\}B_{2,n}, \end{aligned}$$

where $B_{i,n}$ is the constant that comes from integration of the normal density and $B_{i,n}\to B_i$ by Lemma 1.

To prove that $P(J)\to 1$ as $n\to\infty$ we will observe $p_{\tilde{J}}/p_J \le p_{\tilde{J}}^{\blacktriangle}/p_J^{\blacktriangledown}\to 0$ for any $\tilde{J}\neq J$. Like the previous proof there are two cases. First, when $\tilde{J}$ incorrectly groups means as equal. Second, when $\tilde{J}$ does not incorrectly group the means but there are too many groups.

For the first case let J_2 incorrectly group the means and J_1 is the correct grouping. Thus, there are t_1 groups in J_1 and t_2 groups in J_2. At least one of the means in J_2 is incorrectly grouped and at least one of the $\Delta_i\neq 0$. Equivalently $\Delta_i=0$ for the grouping in J_1.

$$\frac{p_{J_2}^{\blacktriangle}}{p_{J_1}^{\blacktriangledown}} = \frac{c_2 V^{(1)}(\eta_{i0})^{b_i n/2}}{c_1(\eta_{i0}+\Delta_i^2)^{(b_i n-u_i-1)/2}} \frac{\exp\left\{\sum_{i=1}^k \frac{\sqrt{b_i n}(2\Delta_i Z_{i1}-Z_{i2})}{2(\eta_{i0}+\Delta_i^2)}\right\}B_{2,n}}{\exp\left\{\sum_{i=1}^k -\frac{\sqrt{b_i n}Z_{i2}}{2\eta_{i0}}\right\}B_{1,n}} \to 0 \quad \text{a.s.}$$

The second case when J_2 is a valid model with too many groups and J_1 is the correct grouping. Thus, there are t_1 groups in J_1, t_2 groups in J_2 and $t_2>t_1$.

$$\begin{aligned} \frac{p_{J_2}^{\blacktriangle}}{p_{J_1}^{\blacktriangledown}} &= \frac{n^{t_1/2}c_2 V^{(1)}(\eta_{i0})^{b_i n/2}}{n^{t_2/2}c_1(\eta_{i0})^{(b_i n-u_i-1)/2}} \frac{\exp\left\{\sum_{i=1}^k -\frac{\sqrt{b_i n}Z_{i2}}{2(\eta)}\right\}B_{2,n}}{\exp\left\{\sum_{i=1}^k -\frac{\sqrt{b_i n}Z_{i2}}{2\eta_{i0}}\right\}B_{1,n}} \\ &= \frac{c_2 V^{(1)}(\eta_{i0})^{b_i n/2}}{n^{(t_2-t_1)/2}c_1(\eta_{i0})^{(b_i n-u_i-1)/2}}\frac{B_{2,n}}{B_{1,n}} \to 0 \quad \text{a.s.} \end{aligned}$$

Thus we have shown that $P(J)\to 1$.

The same convergence results for both constant a non-constant variance hold if the weight function is included. □

References

Abdel-Karim, A.H., 2005. Applications of Generalized Inference.

Fisher, R.A., 1930. Inverse probability. Proceedings of the Cambridge Philosophical Society xxvi, 528–535.

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

Hannig, J., 2009a On Asymptotic Properties of Generalized Fiducial Inference for Discretized Data. Technical Report UNC/STOR/09/02, Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill.

Hannig, J., 2009b. On generalized fiducial inference. Statistica Sinica 19, 491–544.

Hannig, J., Iyer, H.K., Patterson, P., 2006. Fiducial generalized confidence intervals. Journal of the American Statistical Association 101, 254–269.

Hannig, J., Lee, T.C., 2009. Generalized fiducial inference for wavelet regression. Biometrika 96, 847–860.
Lindley, D.V., 1958. Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society Series B 20, 102–107.
Steele, R.D., Torrie, J.H., 1980. Principles and Procedures of Statistics—A Biomedical Approach, second ed. McGraw Hill, New York.
Weeranhandi, S., 1993. Generalized confidence intervals. Journal of the American Statistical Association 88, 899–905.
Zabell, S.L., 1992. R.A. Fisher and the fiducial argument. Statistical Science 7, 369–387.