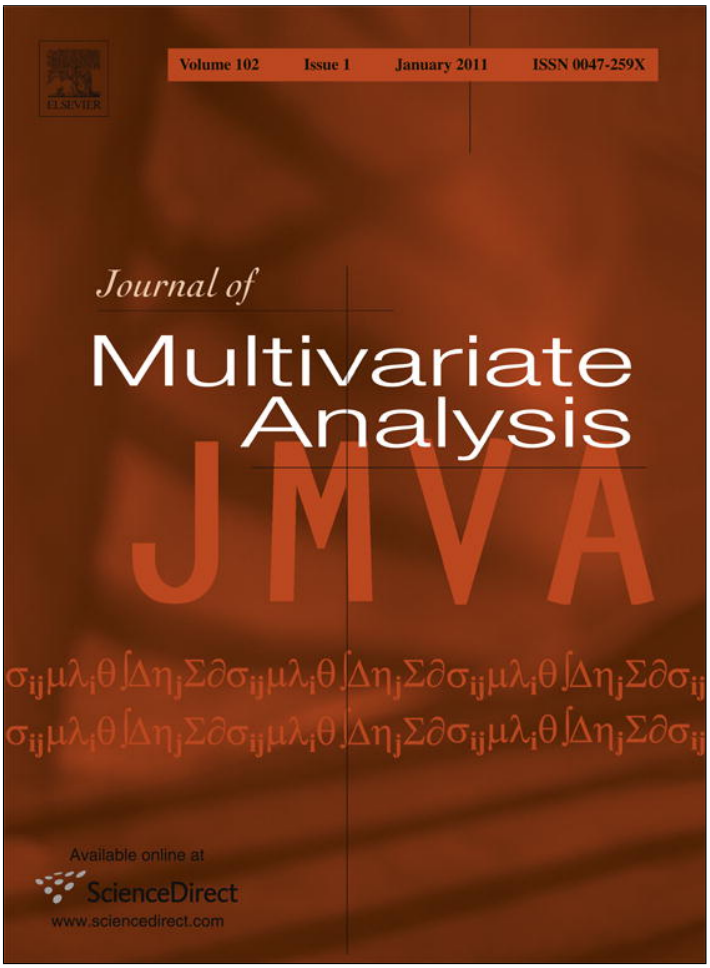

Journal of Multivariate Analysis 102 (2011) 87–104

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Fiducial inference on the largest mean of a multivariate normal distribution

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

[a] Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA

[b] Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 27 February 2010
Available online 21 August 2010

AMS subject classifications:
62H12
62H15
62F10
62F05
62F25
62F15
62F03
62F12

Keywords:
Fiducial inference
Largest mean
Asymptotic consistency
Importance sampling

ABSTRACT

Inference on the largest mean of a multivariate normal distribution is a surprisingly difficult and unexplored topic. Difficulties arise when two or more of the means are simultaneously the largest mean. Our proposed solution is based on an extension of R.A. Fisher's fiducial inference methods termed generalized fiducial inference. We use a model selection technique along with the generalized fiducial distribution to allow for equal largest means and alleviate the overestimation that commonly occurs. Our proposed confidence intervals for the largest mean have asymptotically correct frequentist coverage and simulation results suggest that they possess promising small sample empirical properties. In addition to the theoretical calculations and simulations we also applied this approach to the air quality index of the four largest cities in the northeastern United States (Baltimore, Boston, New York, and Philadelphia).

1. Introduction

Many applications attempt to find the best or worst treatment when dealing with data from a correlated multivariate normal distribution (e.g. pollution studies, drug trials, studies measuring the QT interval, investments, etc.). This is equivalent to finding the largest mean(s) of the normal distribution. Namely, if $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$ and Σ is an unstructured covariance matrix, we are attempting inference on the largest mean, $\theta = \max_i \mu_i$.

While most inference problems for the general multivariate normal distribution are well studied, interval estimation for the largest mean is still relatively unexplored. Obvious solutions tend to grossly overestimate the largest mean when several of the individual means are equal or close to the largest mean. We propose a new method using fiducial inference and demonstrate the empirical coverage of the intervals using a novel approach seen in [8,10]. In addition to using the general fiducial approach we use a model selection technique to allow multiple means to be the equal largest mean. By allowing for some or all of the means to be the largest mean our method will asymptotically select the correct model. This model selection will help to alleviate the common overestimation problem and allow our confidence intervals for the largest mean to have asymptotically correct coverage.

A naive upper tailed confidence interval, seen in the literature, is based on the intersection–union method that constructs t-intervals for each of the k dimensions then uses the maximum upper bound as the upper tailed confidence interval for

* Corresponding author.
E-mail addresses: dvwandler@gmail.com (D.V. Wandler), hannig@mail.unc.edu (J. Hannig).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.08.003

the largest mean. Another technique, seen in [4], approximates the maximum with a smooth function and uses the delta method with a bias adjustment to create a one sided upper-tailed interval. Additionally, Boos et al. [1] uses a linear model to account for the variation and the bias of the largest mean. This method produced two intervals using a normal approximation and intervals using a parametric and nonparametric bias adjusted bootstrap technique. We compared our method to those of the intersections–union, [4], and the bias adjusted method using a normal approximation from Boos et al. [1]. All of the methods from Boos et al. [1] had similar coverage rates, and the one discussed here created the shortest upper tailed interval. It seems reasonable that a Bayesian solution could also provide a viable method. However, we were not able to find a Bayesian solution in the literature.

The most common method in practice, the intersection–union, has a major downfall when some or all of the k means are equal or close to being equal. When there are multiple equal largest means, or means that are very close to being equal, the construction of the t-intervals will systematically overestimate the largest mean. This produces very conservative upper tailed confidence intervals. When two or more means are equal the intersection–union method produces confidence intervals that do not have asymptotically correct coverage, c.f., [3]. The other two methods attempt to account for the bias of the largest mean but have other shortcomings. Simulations suggest that our upper tailed fiducial interval tends to be shorter than the interval created by Eaton et al. [4] and, when the sample size is small, slightly longer than the interval from Boos et al. [1]. When the sample size is large the fiducial interval is shorter than all the competing methods. Furthermore, our method is the only method that also produces a lower tailed or two-tailed interval.

The fiducial approach was also applied to an interesting data set from the Environmental Protection Agency (EPA). This data set measured the air quality for each of the cities of Baltimore, Boston, New York, and Philadelphia. We attempt to find out if the cities experience adequate average air quality. This is analogous to finding out if the city with the worst air quality (largest average value) still has adequate air. In addition to finding out if they all have adequate air quality we would also like information as to which city or cities have the worst average air quality. The fiducial approach will accomplish both and we will compare our results for this example to the competing intervals stated previously. The analysis shows that the fiducial interval is shortest and Baltimore, New York, and Philadelphia are likely to have equally bad air quality, though all of the cities have adequate air quality.

2. Generalized fiducial inference

Fiducial inference was first introduced by Fisher [6]. He proposed the idea in an effort to overcome what he perceived as a deficiency in the Bayesian approach. Fisher opposed the Bayesian approach of assuming a prior distribution when there was not substantial information available about the parameters.

Opposition to fiducial inference arose when researchers discovered that this inference technique did not possess some of the properties that Fisher had originally claimed [11,14]. Recently, there has been somewhat of a resurgence in fiducial inference following the introduction of generalized inference by Weerahandi [13] and the establishment of a link between fiducial and generalized inference in [9]. Further information on the asymptotic and empirical properties and a thorough survey of the fiducial literature can be found in [8].

The basis of generalized fiducial inference, similar to the likelihood function, "switches" the role of the data, $\mathbf{X}$, and the parameter(s) ξ. Fiducial inference uses the model and the observed data, $\mathbf{X}$, to define a probability measure on the parameter space, Ξ. This is understood as a summary of the known information about the parameters, similar to a Bayesian posterior distribution. In the rest of this section we will formally introduce this approach.

First, we assume that a relationship between the $\mathbf{X}$ and 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. With this relationship, called the *structural equation*, the parameter ξ and the random vector $\mathbf{U}$ will determine the distribution of $\mathbf{X}$. After observing $\mathbf{X}$ we can use the relationship in (1) and what we know of the distribution of $\mathbf{U}$ to infer a distribution on ξ.

We define the inverse of the structural equation as the set valued function:

$$Q(\mathbf{x}, \mathbf{u}) = \{\xi : \mathbf{x} = G(\xi, \mathbf{u})\}. \tag{2}$$

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

$$V(Q(\mathbf{x}, \mathbf{U}^\star)) \mid \{Q(\mathbf{x}, \mathbf{U}^\star) \neq \emptyset\} \tag{3}$$

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, $\bar{S}$. Essentially, $V(\cdot)$ is a (possibly random) rule for discerning among the values of the inverse $Q(\mathbf{x}, \mathbf{U}^\star)$. We will refer to a random element with the distribution given by (3) as $\mathcal{R}_\xi$. For a detailed discussion of the derivation of the generalized fiducial distribution see [8].

We calculate the *generalized fiducial density* as proposed in [8] and justified theoretically in [7]. Let $G = (g_1, \ldots, g_n)$ be such that $X_i = g_i(\xi, \mathbf{U})$ for $i = 1, \ldots, n$. Note that 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 indices $\mathbf{i} = (i_1, \ldots, i_p)$. Assume that the functions $\mathbf{G_{0,i}}$ are one-to-one and differentiable. Under some technical assumptions in [7] this will produce the generalized fiducial density:

Table 1
Coverage for a 95% upper tailed interval when $\Sigma = I$ for the naive method.

$\boldsymbol{\mu}_0 \setminus n$	30	100	1000
$(5, 5)^T$	99.6	99.9	99.8
$(1, 5)^T$	92.3	94.9	95.3

$$f_{\mathcal{R}_\xi}(\xi) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x}, \xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x}, \xi')\mathrm{d}\xi'} \tag{4}$$

where

$$J(\mathbf{x}, \xi) = \binom{n}{p}^{-1} \sum_{\mathbf{i}=(i_1,\ldots,i_p)} \left| \frac{\det\left(\frac{\mathrm{d}}{\mathrm{d}\xi}\mathbf{G}^{-1}_{0,\mathbf{i}}(\mathbf{x_i}, \xi)\right)}{\det\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{u_i}}\mathbf{G}^{-1}_{0,\mathbf{i}}(\mathbf{x_i}, \xi)\right)} \right| \tag{5}$$

is the mean of all subsets, where $1 \le i_1 < \cdots < i_p \le n$ and the determinants in (5) are the appropriate Jacobians. For illustrative examples of this fiducial recipe see [8].

3. Main results

3.1. Two dimensional case

Before tackling the largest mean problem in full generality, we first consider the two dimensional case. Assume $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is an independent random sample from the $N(\boldsymbol{\mu}, \Sigma)$ distribution, where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and

$$\Sigma = \begin{bmatrix} \eta_1 & \rho_{12}\sqrt{\eta_1\eta_2} \\ \rho_{12}\sqrt{\eta_1\eta_2} & \eta_2 \end{bmatrix}. \tag{6}$$

The structural equation in (1) is given by

$$\mathbf{X}_i = G(\xi, \mathbf{U}_i) = \boldsymbol{\mu} + \mathbf{V}\mathbf{U}_i$$

where $\mathbf{U}_i \sim N(\mathbf{0}, \mathbf{I}_k)$ and independent for all i, $\mathbf{V}$ is the lower triangle Cholesky decomposition of Σ, and $\xi = (\mu_1, \mu_2, \eta_1, \eta_2, \rho_{12})^T$. The inverted structural equations are:

$$U_{1j} = \frac{X_{1j} - \mu_1}{V_{11}} \quad \text{and} \quad U_{2j} = \frac{X_{2j} - \mu_2 - V_{21}U_{1j}}{V_{22}}$$

for any individual $j = 1, \ldots, n$. Since there are five parameters, we need the same number of equations to define each of the terms in (5). When we differentiate the function $\mathbf{U}_0 = \mathbf{G}_0^{-1}(\mathbf{X}_0, \xi)$ we could, for example, use the following five equations from the first three individuals,

$$\begin{matrix} U_{11} & U_{21} \\ U_{12} & U_{22} \\ & U_{23}. \end{matrix}$$

As it would be arbitrary to choose those five equations we average over all possible selections. These computations will result in the generalized fiducial density:

$$\begin{aligned} f_1(\xi) \propto\ & \frac{1}{(2\pi)^{kn/2}} (\det \Sigma)^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} \\ & \times \binom{n}{2, 1, n-3}^{-1} \sum_{\mathbf{i}} \frac{|g(\mathbf{x_i})|}{2^2 \eta_1^{3/2} \eta_2^{1/2} (1 - \rho_{12}^2)} \end{aligned} \tag{7}$$

where $g(\mathbf{x_i})$ is a function of the data and will be explained later. From (7) we could draw a fiducial random sample of $\theta = \max(\mu_1, \mu_2)$ by generating a sample of $\boldsymbol{\mu}$ and taking the maximum. We calculate a 95% upper tailed interval by taking the estimated 0.95 quantile from this sample. This would be a naive solution and would overestimate the true largest mean, θ_0, when the true means, $\mu_0^{(1)}$ and $\mu_0^{(2)}$, are equal or relatively close to equal. For example, Table 1 shows the coverage for a 95% upper tailed interval of θ_0. When $\mu_0^{(1)}$ and $\mu_0^{(2)}$ are different the coverage is reasonable.

On the other hand, if $\mu_1 = \mu_2$ we can proceed in a similar fashion. If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is an independent random sample from the $N(\boldsymbol{\mu}, \Sigma)$ distribution, where $\boldsymbol{\mu} = (\mu, \mu)^T$, our inverted structural equations change to:

$$U_{1j} = \frac{X_{1j} - \mu}{V_{11}} \quad \text{and} \quad U_{2j} = \frac{X_{2j} - \mu - V_{21}U_{1j}}{V_{22}}.$$

Table 2
Coverage for a 95% upper tailed interval when $\Sigma = I$ when $\mu_1 = \mu_2$ is assumed.

$\boldsymbol{\mu}_0 \setminus n$	30	100	1000
$(5, 5)^T$	94.4	94.4	94.9
$(1, 5)^T$	0.8	0	0

We now have four parameters, so only four equations are necessary to define each of the terms in (5). The generalized fiducial density for this case is then computed as:

$$f_2(\xi) \propto \frac{1}{(2\pi)^{kn/2}} (\det \Sigma)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\right\} \times \binom{n}{2, n-2}^{-1} \sum_{\mathbf{i}} \frac{|g(\mathbf{x_i}, \mu)|}{2^2 \eta_1^{3/2} \eta_2^{1/2} (1 - \rho_{12}^2)}. \tag{8}$$

Comparing (7) and (8) we can see that the difference comes in the Jacobian calculation and the $\boldsymbol{\mu}$ vector. The Jacobian in (8) depends on μ, the equal mean, and (7) is only dependant on the data. This behavior is common to all fiducial densities that have an equal mean. Table 2 shows that the coverage is much better when the true means are actually equal and much worse in the other case.

Clearly we would like to use (7) when the means are different and (8) when the means are equal. We will achieve this by incorporating a model selection step into our problem. The model selection step is the important factor in reducing the overestimation that occurs when the true means are equal, and drives the asymptotic correctness of the intervals.

The largest mean could come from μ_1 or μ_2 by themselves, or $\mu_1 = \mu_2$, in which case they would both be the largest mean. As we have seen, naively assuming that $\mu_1 \neq \mu_2$ will overestimate the true largest mean when $\mu_0^{(1)} = \mu_0^{(2)}$ and the intervals are not asymptotically correct. In order to fix this deficiency we propose an alternative that allows for $\mu_1 = \mu_2$. The structural equation for this scenario is

$$\mathbf{X} = \left(\binom{\mu_1}{\mu_2} + \mathbf{V}_1\mathbf{U}\right) I_{\{\mu_1 > \mu_2\}} + \left(\binom{\mu_1}{\mu_2} + \mathbf{V}_2\mathbf{U}\right) I_{\{\mu_2 > \mu_1\}} + \left(\binom{\mu}{\mu} + \mathbf{V}_1\mathbf{U}\right) I_{\{\mu_1 = \mu_2 = \mu\}} \tag{9}$$

where $\mathbf{V}_1$ is the lower triangle Cholesky decomposition of Σ and $\mathbf{V}_2$ is obtained by permuting the Σ matrix in (6) so that $\mathbf{X} = (X_2, X_1)^T$, taking the lower triangular Cholesky decomposition and re-permuting the resulting matrix back to the original order. This permutation is done because the formulas simplify if the largest mean is in the first position. Notice that, when generating the data, $\mathbf{X}$, only one of the three terms in (9) is in effect at any given time. As a result, there is a model selection aspect to this problem. To simplify subsequent notation, let J be the indexes corresponding to the largest mean, i.e., $J = \{1\}$ if $\mu_1 > \mu_2$, $J = \{2\}$ if $\mu_2 > \mu_1$, and $J = \{1, 2\}$ if $\mu_1 = \mu_2$.

In this two dimensional case we calculate the generalized fiducial distribution by taking p (4 or 5) equations from our structural equation and condition on the fact that the remaining equations occurred $(2n - p)$. 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. For example, when $\mu_1 \neq \mu_2$ there are $2n - 5$ equations to condition on and if $\mu_1 = \mu_2$ there are $2n - 4$ equations to condition on. The conditional distribution will favor the model with less conditions or with more parameters. We can combat this problem by introducing additional structural equations that are independent of our original structural equations as proposed in [10]. These additional structural equations will balance out the number of conditions that need to be met for each selected set of equal means.

Adding additional structural equations will also allow us to introduce a weight function. The weight function will serve two purposes. First, the weight function will down-weight the models with more free means to increase the likelihood of grouping several means together as the largest mean. Secondly, our procedure was not scale invariant so we used the weight function to make it more scale invariant. Attempting to make the method scale invariant in this fashion is rather ad hoc but seemed to work well in simulations. The actual values that were incorporated into the weight function were obtained when we observed the asymptotic behavior of the probability of selecting a particular model.

The additional structural equations are:

$$\frac{1}{2}\log\left(\frac{2^2}{(MSX_1 + MSX_2 - 2MX_1X_2)\, n}\right) = \beta_i + P_i \quad \text{if } i \in J$$
$$\frac{1}{2}\log\left(\frac{2^2}{(MSX_1 + MSX_2 - 2MX_1X_2)\, n}\right) = P_i \quad \text{if } i \in J^c \tag{10}$$

where MSX_i and MX_1X_2 are the maximum likelihood estimates of the variance and covariance respectively and P_i are independent $Exp(1)$ random variables for all i. Because of the independence these structural equations will not affect the

distribution of $\mathbf{X}$ but they will affect the conditional distribution in (3). When inverting the structural equations in (10), if $i \in J$ we can choose a β_i for any P_i so that the equation is satisfied. Therefore, conditioning on this equation does not affect the conditional distribution. On the other hand, when $i \in J^c$ then $P_i = 2^{-1}\log\left(2^2\left[(MSX_1 + MSX_2 - 2MX_1X_2)\,n\right]^{-1}\right)$, which creates an additional condition to be met. Combining the additional condition and the conditions that need to be met from (9) there will now always be $2n-4$ conditions regardless of the number of equal means.

Adding the model selection component, we compute the generalized fiducial density as:

$$f(\xi) \propto f_{\{1\}}(\xi) + f_{\{2\}}(\xi) + f_{\{1,2\}}(\xi)$$

where

$$f_{\{1\}}(\xi) = w(\mathbf{x}) f_1(\xi) I_{\{\mu_1>\mu_2\}}, \qquad f_{\{2\}}(\xi) = w(\mathbf{x})\tilde{f}_1(\xi) I_{\{\mu_2>\mu_1\}}, \qquad f_{\{1,2\}}(\xi) = f_2(\xi) I_{\{\mu_1=\mu_2\}}$$

with $f_1(\xi)$ given by (7), $\tilde{f}_1(\xi)$ by (7) with η_1 and η_2 switched, $f_2(\xi)$ by (8), and the weight function

$$w(\mathbf{X}) = \frac{2}{(MSX_1 + MSX_2 - 2MX_1X_2)^{1/2}\sqrt{n}}.$$

Again, the weight function is a result of having the additional structural equations and its value is chosen by us.

3.2. General case

The same derivation can be applied to a k dimensional problem. Keeping with the same notation $J \subseteq \{1,\ldots,k\} = J_k$ as the index of the equal means, u ($u \geq 1$) is the number of elements in J, and v ($v \geq 0$) is the number of elements in J^c (note, $u+v=k$). In general, for a k dimensional problem, the structural equation is

$$\mathbf{X} = \sum_{J\subseteq\{1,\ldots,k\}} \left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} + \mathbf{V}_J \mathbf{U} \right) I_{\{\mu_i=\mu_j:,i,j\in J\}} I_{\{\mu_j>\mu_l:j\in J,l\notin J\}} \tag{11}$$

where $\mathbf{U} \sim N(\mathbf{0}, \mathbf{I}_k)$ is a $k \times 1$ vector. To calculate $\mathbf{V}_J$, permute the matrix

$$\Sigma = \begin{bmatrix} \eta_1 & \rho_{12}\sqrt{\eta_1\eta_2} & \cdots & \rho_{1k}\sqrt{\eta_1\eta_k} \\ \rho_{12}\sqrt{\eta_1\eta_2} & \eta_2 & & \\ \vdots & & \ddots & \\ \rho_{1k}\sqrt{\eta_1\eta_k} & & & \eta_k \end{bmatrix}$$

such that the X_i's with equal means are the first u variables, take the lower triangle Cholesky decomposition of Σ then re-permute $\mathbf{V}_J$ back to the original order.

Just as before we introduce the additional structural equations akin to (10), obtaining the weight function

$$w_J(\mathbf{X}) = \prod_{i\in J^c} f(P_i) = \frac{\det\left(M_{J_k}\right)^{(v-1)/2} M_J^{*v/2} k^v}{\det\left(M_J\right)^{(v-1)/2} M_{J_k}^{*v/2} n^{v/2}}$$

where $J \subseteq \{1,\ldots,k\} = J_k$ and M and M^* are defined in Appendix A. This defines the generalized fiducial density as:

$$f(\xi) = \frac{\sum_{J\subseteq\{1,\ldots,k\}} f_J(\xi)}{\int_\Xi \sum_{J\subseteq\{1,\ldots,k\}} f_J(\xi')\mathrm{d}\xi'} \tag{12}$$

where ξ is a vector of the $\boldsymbol{\mu}$ and Σ variables. For each particular $J \subseteq \{1,\ldots,k\}$ the generalized fiducial density with the weight function, $w_J(\mathbf{X})$, is

$$\begin{aligned} f_J(\xi) \propto\ & \frac{w_J(\mathbf{X})}{(2\pi)^{kn/2}} (\det\Sigma)^{-n/2} \exp\left\{ -\frac{1}{2}\sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) \right\} \\ & \times \binom{n}{C_{j,n}}^{-1} \sum_{i_1,\ldots,i_k} \frac{|g(\mathbf{X_i}, \mu)|}{\det\Sigma \prod_{j=1}^{k-1} \frac{\det\Sigma_{1,j}}{\eta_j^*}} I_{\{\mu_j>\mu_l:j\in J,l\notin J\}} I_{\{\mu_i=\mu_j:,i,j\in J\}} \\ =\ & \frac{w_J(\mathbf{X}) J_x}{(2\pi)^{kn/2}} \frac{(\det\Sigma)^{-n/2-1}}{\prod_{j=1}^{k-1} \frac{\det\Sigma_{1,j}}{\eta_j^*}} \exp\left\{ -\frac{1}{2} tr\left(S\Sigma^{-1}\right) \right\} I_{\{\mu_j>\mu_l:j\in J,l\notin J\}} I_{\{\mu_i=\mu_j:,i,j\in J\}} \end{aligned} \tag{13}$$

where

$$S = \left[\sum_l (X_{i,l}-\mu_i)(X_{j,l}-\mu_j)\right]_{i,j=1,\ldots k}$$
$$= n\left[((\mu_i-\bar{X}_{i,n})(\mu_j-\bar{X}_{j,n})+MX_iX_j)\right]_{i,j=1\ldots k},$$
$$MX_iX_j = \frac{\sum_l \left(X_{i,l}-\bar{X}_{i,n}\right)\left(X_{j,l}-\bar{X}_{j,n}\right)}{n}, \qquad J_x = \binom{n}{C_{J,n}}^{-1} \sum_{i_1,\ldots,i_k} |g(\mathbf{X_i},\mu)|,$$

and the rest of the terms are defined in Appendix A.

After we calculate the generalized fiducial density we can also compute the fiducial probability that any J is the index corresponding to the true largest means. This is done by calculating

$$P(J) = \frac{p_J}{\sum_{\tilde{J}\subseteq\{1,\ldots,k\}} p_{\tilde{J}}} \tag{14}$$

where $p_J = \int_{\mathcal{Z}} f_J(\xi)\mathrm{d}\xi$. This added information can help to determine the mean or means that are most likely to be the largest. We would like this value to be large when J correctly indexes the largest mean(s). A later asymptotic discussion will show that this model selection technique results in $P(J) \to 1$, as $n \to \infty$, when J is the correct index. After proving that we will asymptotically select the correct model we can apply previous results to prove that the confidence intervals for the largest mean are asymptotically correct.

3.3. Confidence intervals and coverage

Using the fiducial density in (13) we propose confidence intervals for the largest mean, $\theta_0 = \max_{1\le i\le k}\mu_0^{(i)}$. The intervals constructed using fiducial inference are analogous to the construction of Bayesian credible intervals. We define one dimensional one tailed intervals as (c_1,∞) and $(-\infty,c_2)$ for the lower and upper tailed intervals respectively, where c_1 and c_2 are the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the generalized fiducial distribution. A two tailed $(1-\alpha)100\%$ confidence interval is obtained by combining the two one tailed intervals as (c_1,c_2). When calculating the confidence intervals for the largest mean we do not select the model that is most likely and then get the confidence interval based on that. Instead we average the fiducial distribution for the largest mean according to the fiducial probability ($P(J)$) of each of the models.

As is often the case, we cannot integrate Eq. (13) explicitly so we use a Monte Carlo approach to generate a sample from the generalized fiducial density and calculate the estimated quantiles from that sample. We used the importance sampling algorithm in Appendix C to draw a sample $\{\theta_1',\ldots,\theta_E'\}$ from (12), where E is the number of samples needed in order achieve some pre-specified effective sample size.

Classically, the way to check the coverage of confidence intervals is to choose a desired confidence level (e.g. 95%), simulate data, and check the frequency in which the true parameter is captured by the constructed interval. By preference, we check the coverage rate at all confidence levels simultaneously using a graphic device demonstrated in [8]. To achieve this, set $C(\mathbf{X},\theta_0) = P(\mathcal{R}_\theta < \theta_0|\mathbf{X})$. $C(\mathbf{X},\theta_0)$ (which can be thought of as a p-value) is the lowest coverage level necessary for an upper tailed confidence interval to contain the true value, θ_0.

If the confidence interval for θ_0 were exact at all confidence levels, then $C(\mathbb{X},\theta_0)$ would follow the $U(0,1)$ distribution. Using QQ-plots, we can evaluate how closely $C(\mathbb{X},\theta_0)$ follows the $U(0,1)$ distribution. We plot the nominal p-values (desired coverage probability) vs. actual p-values (actual coverage probability). Fig. 1 is an example of the QQ-plots. If the coverage is exact for all confidence levels, the p-values ($C(\mathbf{X},\theta_0)$) would follow the diagonal line. Due to randomness of the simulation we also provided 95% confidence bands (dashed lines). The p-values (simulated line) cannot be distinguished from a sample of the $U(0,1)$ distribution if they stay within the 95% confidence bands. When this occurs we claim good coverage properties.

To check the coverage of our intervals, look at the QQ-plots at the nominal p-value and note the corresponding actual p-value that coincides with the simulated line. For example, the dotted line in the first plot in Fig. 1 shows that a 0.95 upper tailed interval has actual coverage of 0.987. The dotted line in the second plot shows our method has 0.129 probability in the lower tail or, equivalently, the 0.95 lower tailed interval has an actual coverage rate of 0.871.

Our simulation used 1000 data sets and an effective sample size of 10 000. We will highlight a few difficult and interesting cases before discussing the full simulation study. Fig. 1 illustrates the generated QQ-plots when $n = 30$, $\boldsymbol{\mu}_0 = (5,5,5)^T$, $\Sigma_0 = I$ and $25I$ for the top and bottom rows respectively. The fiducial method tends to overestimate the largest mean when the true means are equal. This overestimation is common to all solutions and our method tends to have a smaller overestimation problem than the competitors. The overestimation leads to conservative upper tailed and liberal lower tailed intervals, as seen in Fig. 1. When the correlation is positive the upper and lower tailed intervals are closer to exact. As the correlation becomes negative the upper and lower tailed intervals become even more conservative and liberal respectively. These plots also illustrate that changing the magnitude of the variance does not seem to affect the coverage. This behavior was seen in the other $\boldsymbol{\mu}_0$ and Σ_0 configurations as well.

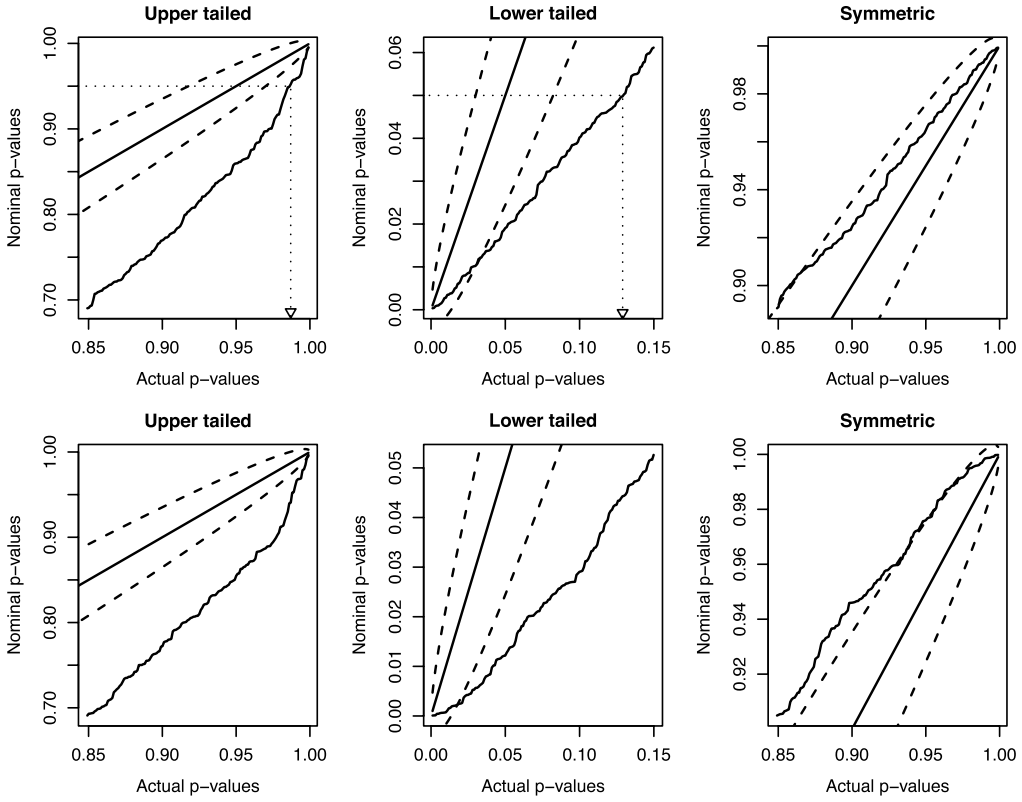

Fig. 1. QQ-plots when $n = 30$, $\mu_0 = (5, 5, 5)^T$, $\Sigma_0 = I$ and $25I$ for the top and bottom rows respectively.

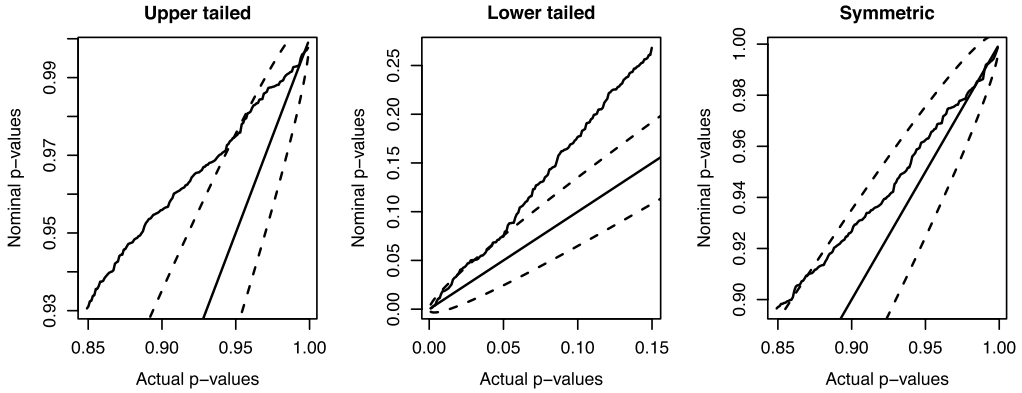

Fig. 2. QQ-plots when $n = 30$, $\mu_0 = (3, 4, 5)^T$ and $\Sigma_0 = 25I$.

When the true means are close, but not equal, our method will tend to select the model where the means are equal. This results in underestimating the largest mean, producing liberal upper tailed and conservative lower tailed intervals. This is seen in Fig. 2 when $n = 30$, $\mu_0 = (3, 4, 5)^T$ and $\Sigma_0 = 25I$.

Cases where the largest mean is much different from the other means results in confidence intervals that are, expectedly, close to exact. This is reflected in Fig. 3 when $n = 30$, $\mu_0 = (1, 3, 5)^T$ and $\Sigma_0 = I$.

When the sample size is dramatically increased to $n = 10\,000$ in the case where $\mu_0 = (3, 4, 5)^T$ and $\Sigma_0 = 25I$, the QQ-plots in Fig. 4 reflect close to exact coverage. This motivates our later discussion on the asymptotics of our intervals.

3.4. Simulation results and discussion

In addition to the select configurations that were previously highlighted, we also performed an extensive simulation study for two and three dimensional data. We looked at all combinations of the parameter values listed in Tables 3 and 4. Each $\rho_0^{(ij)}$ in the last configurations were randomly generated values from the $U(0, 1)$ distribution, where the resulting Σ_0 matrix was positive definite.

The coverage and length of the upper tailed fiducial interval was compared to the upper tailed intervals produced by the intersection–union method (t), [4] (Eaton), and the bias adjusted interval based on a normal approximation from Boos

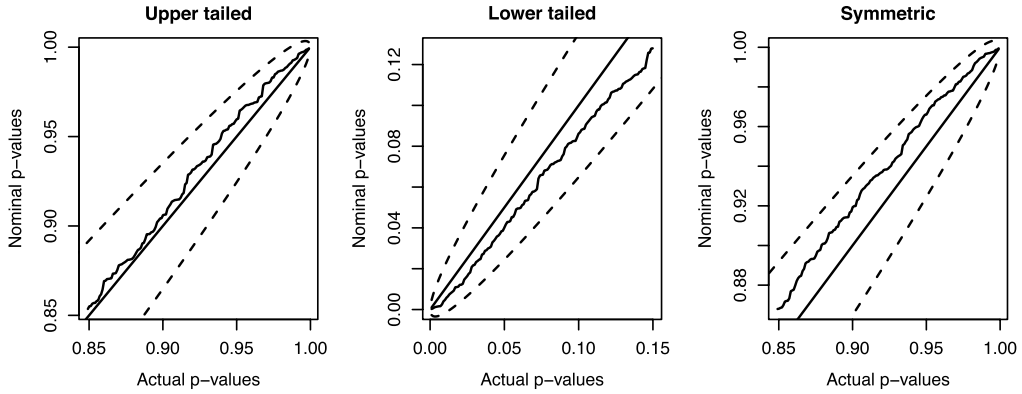

Fig. 3. QQ-plots when $n = 30$, $\mu_0 = (1, 3, 5)^T$ and $\Sigma_0 = I$.

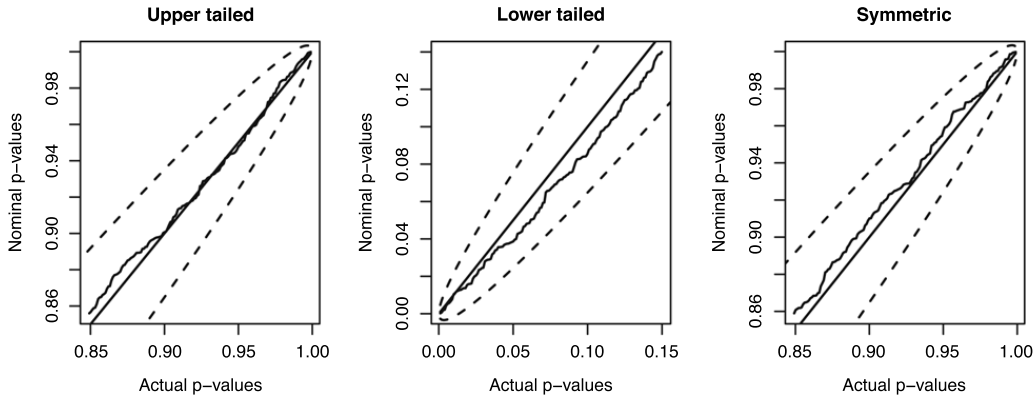

Fig. 4. QQ-plots when $n = 10\,000$, $\mu_0 = (3, 4, 5)^T$ and $\Sigma_0 = 25I$.

Table 3
Simulation combinations in two dimensions.

μ_0	$(2, 5)^T$	$(4, 5)^T$	$(5, 5)^T$
η_0	$(1, 1)$	$(25, 25)$	
$\rho_0^{(1,2)}$	0	± 0.4	± 0.9
n	30	100	

Table 4
Simulation combinations in three dimensions.

μ_0	$(1, 3, 5)^T$ $(4, 5, 5)^T$	$(3, 4, 5)^T$ $(5, 5, 5)^T$	$(1, 5, 5)^T$
η_0	$(1, 1, 1)$	$(25, 25, 25)$	
$\rho_0^{(i,j)}$	0 $\begin{bmatrix} 1 & & \\ 0.051 & 1 & \\ -0.392 & 0.754 & 1 \end{bmatrix}$	± 0.4 $\begin{bmatrix} 1 & & \\ 0.104 & 1 & \\ -0.528 & -0.710 & 1 \end{bmatrix}$	0.9 $\begin{bmatrix} 1 & & \\ -0.133 & 1 & \\ 0.844 & -0.154 & 1 \end{bmatrix}$
n	30	100	

et al. [1] (Boos). As previously noted, Boos et al. [1] also introduced an interval without a bias adjustment and two bias adjusted intervals using a bootstrap approach. The code for the bootstrap methods was proprietary, so we attempted to recreate the intervals described by the authors. Based on our attempt, the interval we compared with performed the best in terms of length and coverage. This agrees with results reported in [1].

The coverage for the 95% and 99% intervals when the data is two dimensional can be seen in Figs. 5 and 7. The coverage when the covariance matrix was randomly generated is seen in Fig. 9. When the correlation is positive, as is the likely case in practice, the upper tailed fiducial interval is close to the nominal coverage level. At the larger sample size, the median coverage appears to be very close to the nominal coverage rate for the fiducial method. The very liberal outliers, seen in Fig. 7, occur when the true means are relatively close together and there is negative correlation. For instance, when $n = 100$, $\mu_0 = (4, 5)$, $\eta_0 = (25, 25)$ and $\rho^{(1,2)} = -0.9$, the 0.95 upper tailed interval has an actual coverage rate of 0.60. As the sample size increases these intervals will converge to the exact coverage level, but they seem to converge slower than the cases with

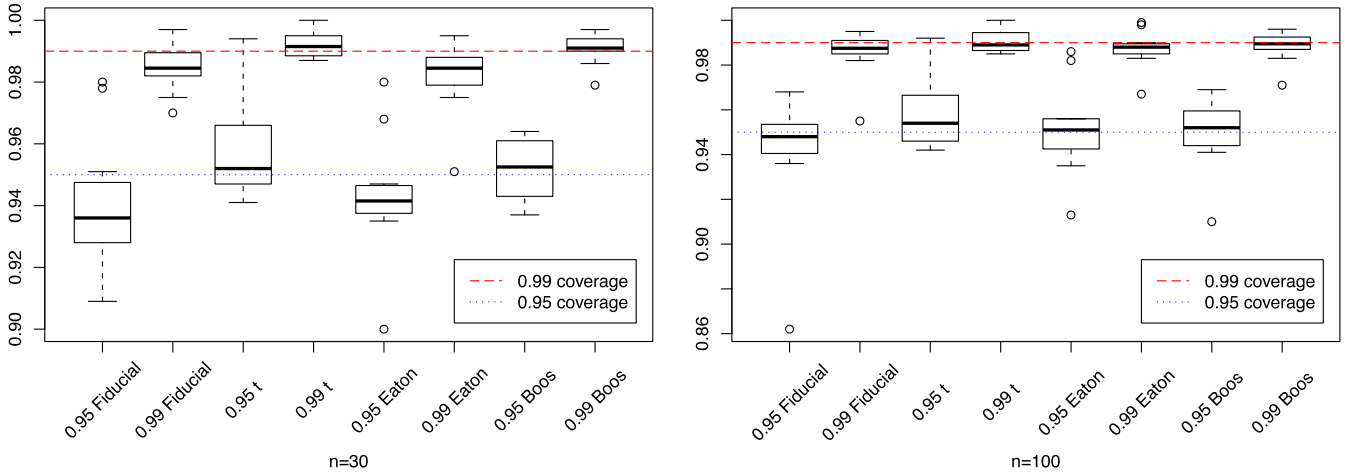

Fig. 5. Coverage when the data is two dimensional and $\rho^{(1,2)} > 0$.

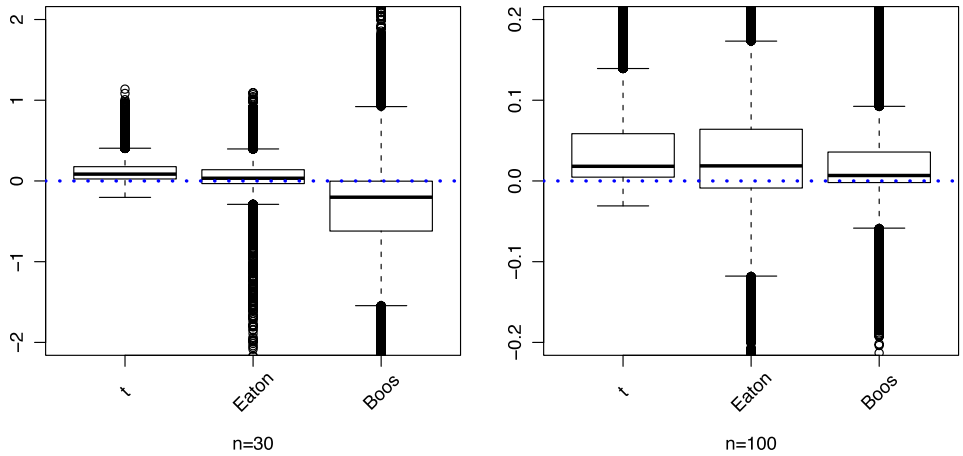

Fig. 6. Length of the upper tailed intervals relative to the fiducial interval when data is two dimensional and $\rho^{(1,2)} > 0$.

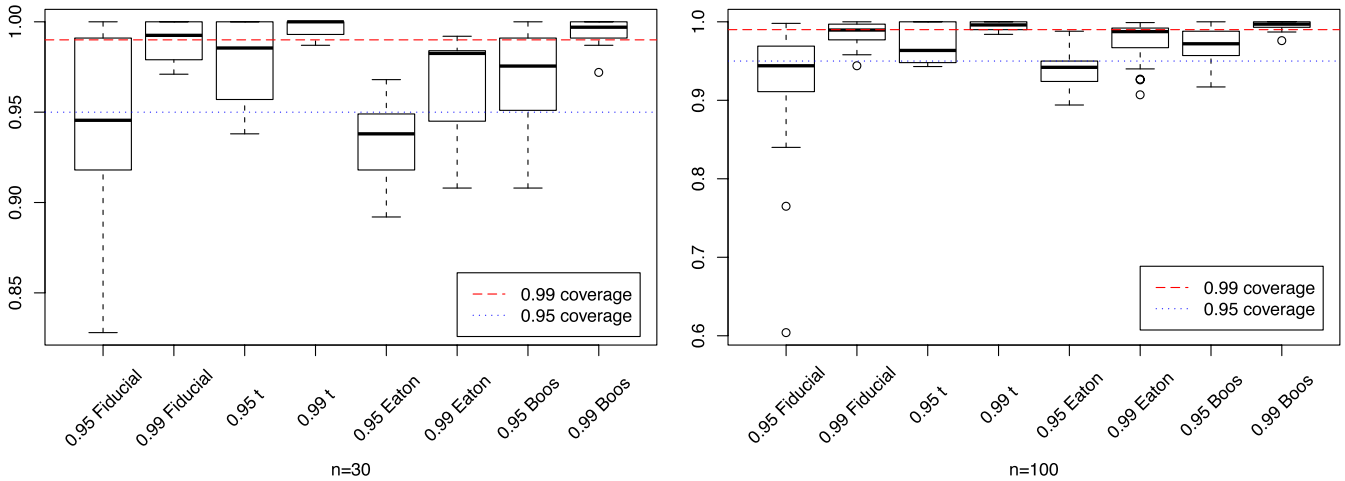

Fig. 7. Coverage when the data is two dimensional and $\rho^{(1,2)} \leq 0$.

positive correlation. Similar behavior also occurred with the three dimensional data. Additional plots are available from the authors.

Figs. 6, 8 and 10 compare the length of the upper tailed fiducial interval to its competitors. The median fiducial interval is shorter than the intervals created by the intersection–union (t) and Eaton. At the small sample size our method is slightly longer than the Boos interval. At the larger sample size our interval is the shortest. Additionally, the Boos interval assumes that the data is equicorrelated and equivariant, where we allow for a totally unstructured covariance matrix.

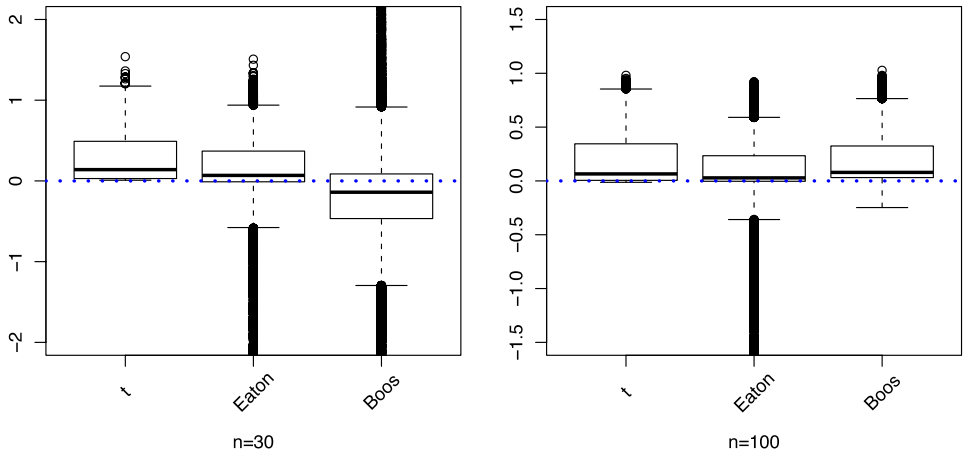

Fig. 8. Length of the upper tailed intervals relative to the fiducial interval when data is two dimensional and $\rho^{(1,2)} \leq 0$.

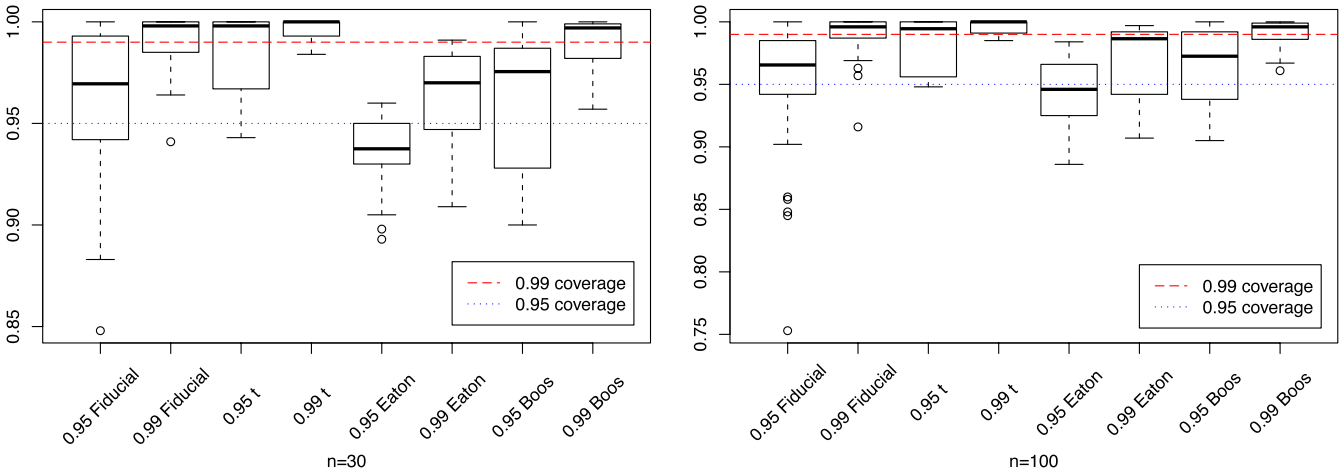

Fig. 9. Coverage when the data is three dimensional and $\rho^{(i,j)}$ is random.

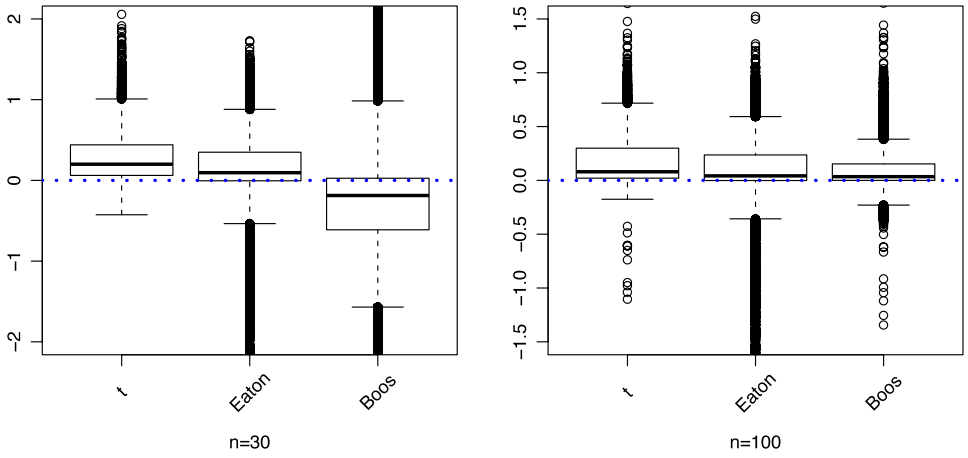

Fig. 10. Length of the upper tailed intervals relative to the fiducial interval when data is three dimensional and $\rho^{(i,j)}$ is random.

4. Asymptotic results

In this section we will prove that the coverage rate for this method is asymptotically correct. We prove that the fiducial probability of the correct model goes to 1 and, therefore, for large n the inference will be almost entirely based on the unknown correct model. We assume the following:

Assumption 1. $\mathbf{X}_i$ is an independent random variable from the $N\left(\left(\mu_0^{(1)}, \ldots, \mu_0^{(k)}\right)^T, \Sigma_0\right)$ distribution.

Table 5
AQI range of health effects.

AQI	Air quality
0–50	Good
51–100	Moderate
101–150	Unhealthy for sensitive groups
151–200	Unhealthy
201–300	Very unhealthy
301–500	Hazardous

Assumption 2. $\eta_0^{(i)}$ is in a compact set of the positive values of $\mathbb{R}^1$ for all i.

Assumption 3. $\rho_0^{(ij)}$ is in a compact set of the interval $(-1, 1)$ for all i, j.

The probability that the means from a particular J are equal to the largest mean is given in (14). If J is the index corresponding to the true largest means and $P(J)$ is large, then the method is selecting the correct model at a high rate. The following proof will show that this method will asymptotically select the correct model. This result does not follow from previous theory on generalized fiducial inference.

Theorem 1. *If $\boldsymbol{\mu}_0$ is the true mean and J correctly identifies the equal largest means then $P(J) \to 1$ in probability.*

We could not integrate (13), so we bound it from above and below using the last two assumptions to show $p_{\tilde{J}}/p_J \to 0$ for any $\tilde{J} \neq J$. If $\tilde{J}$ incorrectly identified the largest means, then $p_{\tilde{J}}/p_J$ converged to zero exponentially. On the other hand, if $\tilde{J}$ was a valid model but had too many free means, then $p_{\tilde{J}}/p_J$ converged to zero polynomially as $n \to \infty$. This was proved without the use of the weight function. The details of the proof are relegated to Appendix B.

Theorem 2. *The confidence intervals for $\theta_0 = \max_{1\leq i\leq k} \mu_0^{(i)}$ are asymptotically correct.*

Proof. Using Theorem 1 we will asymptotically select the correct model. By asymptotically selecting the correct model it follows from the standard methods in [8] to prove that the confidence intervals are asymptotically correct. □

5. Air quality example

The EPA measures the air quality to help inform people of the daily air conditions. This measurement is called the air quality index (AQI) and is calculated from the five major air pollutants regulated by the Clean Air Act. The AQI ranges from 0 to 500, where the higher the value the greater the level of pollution and the greater the health risk. Table 5 breaks down the air quality for different AQI values. Overall, an AQI value of 100 or less is the standard at which the EPA determined as satisfactory.

We obtained monthly AQI measurements from January 1, 1998 to October 1, 2008 from the EPA [5] to perform inference on the largest mean AQI for the cities of Baltimore, Boston, New York, and Philadelphia. The data set is available upon request from the authors. Because of the proximity of these cities there is clearly spatial correlation.

Monthly data was used to eliminate the temporal correlation between successive data points. It was determined from auto correlation plots that data points that were a month apart were reasonably uncorrelated. Using normal QQ-plots we determined that the transformed data of $X' = \log(X)$, where X is the original data and X' is the transformed data, appeared to be reasonably normal. Thus, all analysis was done on the transformed data. The sample mean and covariance for X' is are:

$$\bar{\mathbf{x}}' = \begin{bmatrix} 3.959 \\ 3.845 \\ 3.928 \\ 3.963 \end{bmatrix} \quad \text{and} \quad \hat{\Sigma}' = \begin{bmatrix} 0.251 & 0.123 & 0.161 & 0.194 \\ 0.123 & 0.161 & 0.132 & 0.137 \\ 0.161 & 0.132 & 0.193 & 0.172 \\ 0.194 & 0.137 & 0.172 & 0.230 \end{bmatrix}. \tag{15}$$

If we were to assume that those were the true values for $\boldsymbol{\mu}_0$ and Σ_0, the coverage for θ_0 is shown in Fig. 11. Our method appears to be slightly liberal for the upper tailed interval and slightly conservative for the lower tailed interval. The equal tailed interval appears to have exact coverage. The coverage of the competing methods can be seen in Fig. 12. Because the means are relatively close together, two of the methods tend to overestimate the maximum and produce conservative intervals.

At the 99% level the fiducial method has an approximate coverage rate of 98.3% (based on the Fig. 11) and a back transformed upper limit of 56.8. The intersection–union (t) interval has a coverage rate and upper limit of 99.9% and 58.1, the Eaton interval has a coverage rate and upper limit of 99.5% and 58.0, and the Boos interval has coverage rate and upper limit of 99.3% and 57.5. Even though all the methods produce intervals with upper tails in the moderate range, our method's interval is the shortest and allows for a two tailed interval when needed.

The fiducial method also provides added information as to which city or cities are likely to have the largest average AQI. Table 6 illustrates these fiducial probabilities for any J when $P(J) > 0$ (note, 1 = Baltimore, 2 = Boston, 3 = New York, and

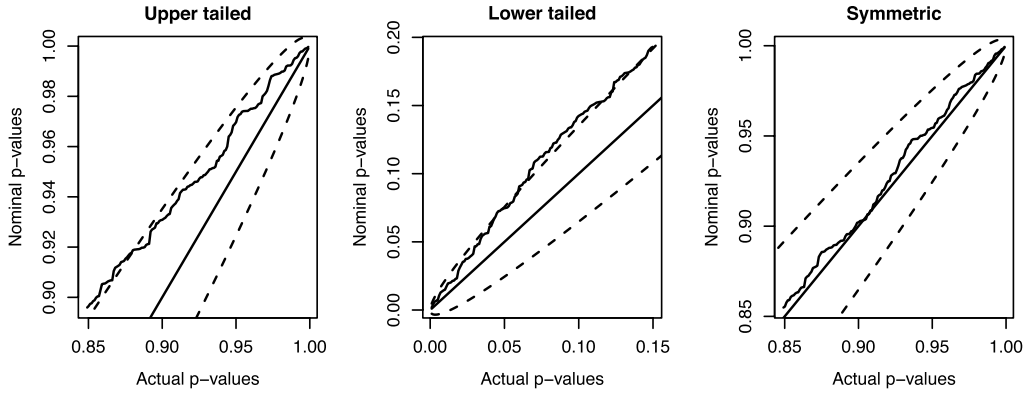

Fig. 11. Fiducial method QQ-plots using μ_0 and Σ_0 from (15).

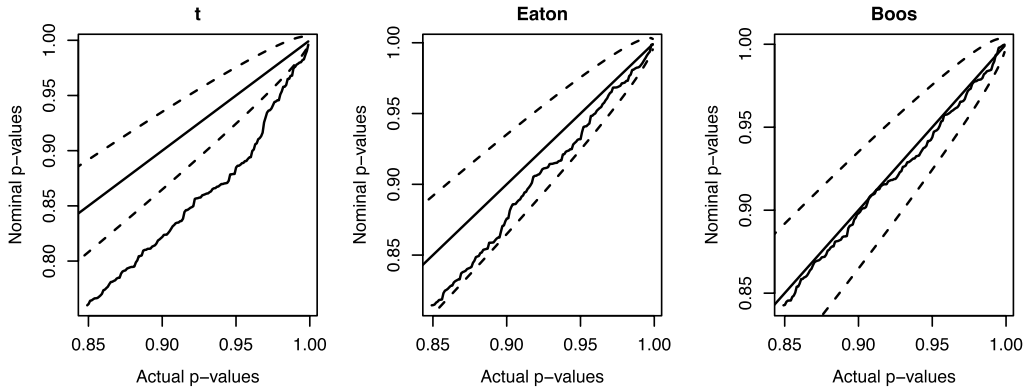

Fig. 12. Competing methods QQ-plots using μ_0 and Σ_0 from (15).

Table 6
Probabilities for each model, J.

J	{1}	{3}	{4}	{1, 3}
$P(J)$	0.009	0.002	0.012	0.014
J	{1, 4}	{3, 4}	{1, 3, 4}	{1, 2, 3, 4}
$P(J)$	0.271	0.095	0.595	0.001

4 = Philadelphia). The probability that $J = \{1, 3, 4\}$ is 0.595, which reflects the likelihood that Baltimore, New York, and Philadelphia are the equal largest mean. This information could be used in an effort to clean up the worst polluting cities.

6. Conclusion

The application of inference on the largest mean of a multivariate normal distribution is wide reaching. There are applications in drug trials, stock returns, agriculture, pollution (as seen in this paper), etc. Clearly, a viable inference technique for the largest mean is necessary.

We proposed a method based on fiducial inference that possesses many nice qualities. First, from simulation results, the empirical coverage for the one and two tailed intervals is close to exact in small sample sizes when the correlation is positive. Second, the upper tailed interval is shorter than two of the competitors when the sample size is small and shorter than all of them when the sample size is large. Unlike the other methods, we have proven our intervals are asymptotically correct. Lastly, this method also provides information as to how likely any of the means are to be the equal largest mean. This could serve as valuable resource management information when any sort of action is taken with the group(s) that have the largest mean.

Our illustrative example examined the air quality of the four largest northeastern cities in the United States. The fiducial approach produced a shorter 99% upper tailed interval than the competitors and it provided information as to which cities had the worst air quality. If a reclamation project were to take place, it would be reasonable to focus the efforts on Baltimore, New York, and Philadelphia.

Acknowledgments

The corresponding author would like to thank the Department of Statistics and Operations Research at The University of North Carolina at Chapel Hill for providing hospitality during the spring of 2009. This research was supported in part by the National Science Foundation under grant-0707037.

Appendix A. Definitions

We derived the generalized fiducial density to be

$$f_J(\xi) \propto \frac{w_J(\mathbf{X})J_x}{(2\pi)^{kn/2}} \frac{(\det \Sigma)^{-n/2-1}}{\prod_{j=1}^{k-1} \frac{\det \Sigma_{1,j}}{\eta_j^*}} \exp\left\{-\frac{1}{2} tr\left(S\Sigma^{-1}\right)\right\} I_{\{\mu_j>\mu_l:j\in J,l\notin J\}} I_{\{\mu_i=\mu_j:,i,j\in J\}}.$$

We will now define all of the components of the equation. First, $\Sigma_{1,j}$ is the upper left $j \times j$ minor of Σ and $\eta_j^* = \prod_{z=1}^{j} \eta_z$. As noted before, $J_x = \binom{n}{C_{u,n}}^{-1} \sum_{i_1,\ldots,i_k} |g(\mathbf{X_i}, \mu)|$, where $g(\mathbf{X_i}, \mu) = \frac{1}{2^k} \det(\mathbb{X}_2) \prod_{j=3}^{u+1} \left|\mathbb{X}_j^*\right| \prod_{j=u+2}^{k+1} \left|\mathbb{X}_j\right|$, $\mathbf{i} = (i_1, \ldots, i_{k+1})$,

$$\mathbb{X}_j = \begin{bmatrix} 1 & X_{i_1,1} & \cdots & X_{i_1,j-1} \\ 1 & X_{i_2,1} & \cdots & X_{i_2,j-1} \\ \vdots & & & \\ 1 & X_{i_j,1} & \cdots & X_{i_j,j-1} \end{bmatrix}, \qquad \mathbb{X}_j^* = \begin{bmatrix} 1 & \mu & \cdots & \mu \\ 1 & X_{i_1,1} & \cdots & X_{i_2,j-1} \\ \vdots & & & \\ 1 & X_{i_{j-1},1} & \cdots & X_{i_j,j-1} \end{bmatrix},$$

$$C_{u,n} = \begin{cases} 2, 1, \ldots, 1, n-k-1 & u = 1 \\ 2, 2, 1, \ldots, 1, n-k-1 & u = 2, \ldots, k-1 \\ 2, 1, \ldots, 1, n-k & u = k, \end{cases}$$

and u is the number of elements in J or the number of equal means ($u \geq 1$). This method was not scale invariant, so we attempted to reduce the scale dependence through our weight function. In order to add a weight function the additional structural equations

$$\begin{aligned} \frac{1}{2v} \log\left(\frac{\det\left(M_{J_k}\right)^{(v-1)} M_J^{*v} k^{2v}}{\det\left(M_J\right)^{(v-1)} M_{J_k}^{*v} n^v}\right) &= \beta_i + P_i \quad \text{if } i \in J \\ \frac{1}{2v} \log\left(\frac{\det\left(M_{J_k}\right)^{(v-1)} M_J^{*v} k^{2v}}{\det\left(M_J\right)^{(v-1)} M_{J_k}^{*v} n^v}\right) &= P_i \quad \text{if } i \in J^c \end{aligned} \tag{A.1}$$

were used, where P_i are independent random variables from the $Exp(1)$ distribution. Thus, the weight function is

$$w_J(\mathbf{X}) = \frac{\det\left(M_{J_k}\right)^{(v-1)/2} M_J^{*v/2} k^v}{\det\left(M_J\right)^{(v-1)/2} M_{J_k}^{*v/2} n^{v/2}}$$

where v is the number elements in J^c or the number of unequal means ($u + v = k$), $J \subseteq \{1, \ldots, k\} = J_k$, $J = \{j_1, j_2, \ldots, j_u\}$, $M_J = \left[MX_iX_j\right]_{i,j\in J}$, and $M_J^* = \det\left(1 + M_J\right) - \det\left(M_J\right)$.

Appendix B. Proof of Theorem 1

Preliminary work for the proof of Theorem 1

To find the fiducial probability that we select a particular model, we will be calculating the probabilities $P(J) = p_J \left(\sum_{\tilde{J}\subseteq\{1,\ldots,k\}} p_{\tilde{J}}\right)^{-1}$. Since we cannot integrate f_J, upper and lower bounds were used to observe the asymptotic behavior of $P(\cdot)$. With the addition of Assumptions 2 and 3 the upper and lower bounds are:

$$f_J(\xi) \leq \frac{J_x^{(\cdot)}}{b_\rho (2\pi)^{kn/2}} (\det \Sigma)^{-n/2-1} \exp\left\{-\frac{1}{2} tr\left(S\Sigma^{-1}\right)\right\} = f_J^{\blacktriangle}(\xi) \tag{B.1}$$

and

$$f_J(\xi) \geq f_J^{\dagger}(\xi) I_{\{\mu_j>\mu_l:j\in J,l\notin J\}} I_{\{-r\leq\rho_{ij}\leq r:\forall i,j\}} I_{\{w_1\leq\eta_i\leq w_2:\forall i\}} = f_J^{\blacktriangledown}(\xi) \tag{B.2}$$

where $f_J^{\dagger}(\xi) = b(2\pi)^{-kn/2}(\det\Sigma)^{-n/2-1}\exp\left\{-\frac{1}{2}tr\left(S\Sigma^{-1}\right)\right\}I_{\{\mu_i=\mu_j:,i,j\in J\}}$. Additionally, $\left|\rho_0^{(ij)}\right| \leq r < 1$ and $0 < w_1 \leq \eta_0^{(i)} \leq w_2 < \infty$ for all i and j, $b_\rho > 0$ is a lower bound for $\prod_{j=1}^{k-1}\det\Sigma_j\left(\eta_j^*\right)^{-1}$, and $b > 0$ is a lower bound for J_x. When proving the asymptotic consistency we will assume, without loss of generality, that $E(\mathbf{X}) = (\mu_1, \mu_2, \ldots, \mu_v, \mu, \ldots, \mu)^T$. That is saying that $X_{v+1}, \ldots, X_k$ share the common mean, μ.

If we notice that Σ^{-1} and μ_i follow a Wishart and t distribution for all i, the resulting integration for the upper bound is

$$\begin{aligned}
p_J^{\blacktriangle} &= \int_{\Xi} f_J^{\blacktriangle}(\xi)\mathrm{d}\xi \\
&= \frac{1}{b_\rho}\frac{2^{k(k+3)/2}\pi^{k(k-2n-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2}s_1^{(n+k+3)/2}s_2^{(n+k+2)/2}\cdots s_v^{(n+k+4-v)/2}} \\
&\quad\times\frac{\pi^{v/2}\Gamma\left(\frac{n+k+2}{2}\right)\Gamma\left(\frac{n+k+1}{2}\right)\cdots\Gamma\left(\frac{n+k+3-(k-u)}{2}\right)}{\Gamma\left(\frac{n+k+3}{2}\right)\Gamma\left(\frac{n+k+2}{2}\right)\cdots\Gamma\left(\frac{n+k+4-(k-u)}{2}\right)} \\
&\quad\times\int_{\mathbb{R}}\frac{J_x^{(\cdot)}}{(\sigma_v^2-\zeta_v^2)^{(n+k+3-v)/2}}\mathrm{d}\mu
\end{aligned}$$

where $J_x^{(\cdot)}$ is the appropriate Jacobian, which will be discussed later. From this point we will break the integration into 3 cases. When $u = 1$, $u > 1$ and even, and $u > 1$ and odd. The first case, when $u = 1$ (all means are different denoted J_1) results in

$$p_{J_1}^{\blacktriangle} = \frac{J_x^{(0)}}{b_\rho}\frac{2^{k(k+3)/2}\pi^{k(k-2n-1)/4}\pi^{k/2}\prod_{i=1}^{k}\Gamma\left(\frac{n+k+3-i}{2}\right)}{n^{k(n+k+3)/2}s_1^{(n+k+3)/2}s_2^{(n+k+2)/2}\cdots s_k^{(n+4)/2}}\frac{1}{(\sigma_k^2-\zeta_k^2)^{(n+3)/2}}.$$

When $u > 1$ and odd,

$$\begin{aligned}
p_{J_u}^{\blacktriangle} &= \frac{1}{b_\rho}\frac{2^{k(k+3)/2}\pi^{k(k-2n-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2}s_1^{(n+k+3)/2}s_2^{(n+k+2)/2}\cdots s_v^{(n+k+4-v)/2}\tilde{s}_u^{(n+u+3)/2}} \\
&\quad\times\frac{\pi^{(v+1)/2}\prod_{i=1}^{v}\Gamma\left(\frac{n+k+3-i}{2}\right)}{\prod_{i=1}^{v}\Gamma\left(\frac{n+k+4-i}{2}\right)}\sum_{i=1}^{(u+1)/2}\frac{J_{x_i}^{(1)}}{(\tilde{\sigma}_u^2-\tilde{\zeta}_u^2)^{(n+1+2i)/2}}\frac{\Gamma\left(\frac{n+2i+1}{2}\right)}{\Gamma\left(\frac{n+2i+2}{2}\right)}.
\end{aligned}$$

When $u > 1$ and even,

$$\begin{aligned}
p_{J_u}^{\blacktriangle} &= \frac{1}{b_\rho}\frac{2^{k(k+3)/2}\pi^{k(k-2n-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2}s_1^{(n+k+3)/2}s_2^{(n+k+2)/2}\cdots s_v^{(n+k+4-v)/2}\tilde{s}_u^{(n+u+3)/2}} \\
&\quad\times\frac{\pi^{(v+1)/2}\prod_{i=1}^{v}\Gamma\left(\frac{n+k+3-i}{2}\right)}{\prod_{i=1}^{v}\Gamma\left(\frac{n+k+4-i}{2}\right)}\sum_{i=1}^{u/2+1}\frac{J_{x_i}^{(2)}}{(\tilde{\sigma}_u^2-\tilde{\zeta}_u^2)^{(n+2i)/2}}\frac{\Gamma\left(\frac{n+2i}{2}\right)}{\Gamma\left(\frac{n+2i+1}{2}\right)}.
\end{aligned}$$

Integrating the lower bound $f_J^{\blacktriangledown}(\xi)$ will result in a very similar calculation. First notice that from Eq. (B.2),

$$\begin{aligned}
p_J^{\blacktriangledown} &= \int_{\Xi} f_J^{\blacktriangledown}(\xi)\mathrm{d}\xi \\
&= \int_{\Xi} f_J^{\dagger}(\xi)I_{\{\mu_j>\mu_l:j\in J,l\notin J\}}I_{\{-r\leq\rho_{ij}\leq r:\forall i,j\}}I_{\{w_1\leq\eta_i\leq w_2:\forall i\}}\mathrm{d}\xi \\
&= \int_{\Xi} f_J^{\dagger}(\xi)\mathrm{d}\xi + L_n
\end{aligned}$$

where

$$L_n = -\int_{\Xi} f_J^{\dagger}(\xi)\left(1-I_{\{\mu_j>\mu_l:j\in J,l\notin J\}}\right)\mathrm{d}\xi - \int_{\Xi} f_J^{\dagger}(\xi)\left(1-I_{\{-r\leq\rho_{ij}\leq r:\forall i,j\}}\right)\mathrm{d}\xi$$

$$
\begin{aligned}
&- \int_{\Xi} f_J^{\dagger}(\xi)\left(1 - I_{\{w_1 \le \eta_i \le w_2 : \forall i\}}\right) \mathrm{d}\xi \\
&+ \int_{\Xi} f_J^{\dagger}(\xi)\left(1 - I_{\{\mu_j > \mu_l : j \in J, l \notin J\}}\right)\left(1 - I_{\{-r \le \rho_{ij} \le r : \forall i,j\}}\right) \mathrm{d}\xi \\
&+ \int_{\Xi} f_J^{\dagger}(\xi)\left(1 - I_{\{\mu_j > \mu_l : j \in J, l \notin J\}}\right)\left(1 - I_{\{w_1 \le \eta_i \le w_2 : \forall i\}}\right) \mathrm{d}\xi \\
&+ \int_{\Xi} f_J^{\dagger}(\xi)\left(1 - I_{\{-r \le \rho_{ij} \le r : \forall i,j\}}\right)\left(1 - I_{\{w_1 \le \eta_i \le w_2 : \forall i\}}\right) \mathrm{d}\xi \\
&- \int_{\Xi} f_J^{\dagger}(\xi)\left(1 - I_{\{\mu_j > \mu_l : j \in J, l \notin J\}}\right)\left(1 - I_{\{-r \le \rho_{ij} \le r : \forall i,j\}}\right)\left(1 - I_{\{w_1 \le \eta_i \le w_2 : \forall i\}}\right) \mathrm{d}\xi . \qquad \text{(B.3)}
\end{aligned}
$$

We do not need explicitly integrate L_n, because this term will converge to 0. Using the transformation

$$\sqrt{n}\left(\mathcal{R}_{\xi} - \xi_0\right) = \mathbf{p} \qquad \text{(B.4)}$$

where $\mathcal{R}_{\xi}$ is a vector of the fiducial random variables for ξ, ξ_0 is a vector of the values of $\boldsymbol{\mu}_0$ and Σ_0, and $\mathbf{p}$ is a vector of the transformed variables. Observing the indicators from (B.3) with the transformation in (B.4), we can see that they converge to 1 in probability. Therefore, as $n \to \infty$ the convergence of $L_n \to 0$ in probability follows, by comparison with the Wishart and t densities. Thus,

$$
\begin{aligned}
p_J^{\blacktriangledown} &= \int_{\Xi} f_J^{\dagger}(\xi)\mathrm{d}\xi \left(1 + \frac{L_n}{\int_{\Xi} f_J^{\dagger}(\xi)\mathrm{d}\xi}\right) \\
&= b\frac{2^{k(k+3)/2}\pi^{k(k-2n-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2}s_1^{(n+k+3)/2}s_2^{(n+k+2)/2}\ldots s_v^{(n+k+4-v)/2}\tilde{s}_u^{(n+k+3-v)/2}} \\
&\quad \times \frac{\pi^{(v+1)/2}\prod_{i=1}^{v+1}\Gamma\left(\frac{n+k+3-i}{2}\right)}{\prod_{i=1}^{v+1}\Gamma\left(\frac{n+k+4-i}{2}\right)} \frac{1}{(\tilde{\sigma}_u^2 - \tilde{\zeta}_u^2)^{(n+k+2-v)/2}}\left(1 + \frac{L_n}{\int_{\Xi} f_J^{\dagger}(\xi)\mathrm{d}\xi}\right)
\end{aligned}
$$

where $\left(1 + L_n\left(\int_{\Xi} f_J^{\dagger}(\xi)\mathrm{d}\xi\right)^{-1}\right) \to 1$ in probability. For future reference we use the notation $I_J = \left(1 + L_n\left(\int_{\Xi} f_J^{\dagger}(\xi)\mathrm{d}\xi\right)^{-1}\right)$.

To define the Jacobian term, $J_x^{(\cdot)}$, a lower bound can be $J_x > b$ for some $b > 0$. If $u = 1$, meaning all of the means are different, then J_x is independent of μ and we can explicitly write $J_x = J_x^{(0)}$, where an upper bound is unnecessary. For all other cases we will bound J_x from above as follows. First,

$$
\begin{aligned}
J_x &= \left|J_{x_1}\mu^{u-1} + J_{x_2}\mu^{u-2} + \cdots + J_{x_u}\right| \\
&\le \left|J_{x_1}\right|\left|\mu^{u-1}\right| + \left|J_{x_2}\right|\left|\mu^{u-2}\right| + \cdots + \left|J_{x_u}\right| .
\end{aligned}
$$

We can bound $\left|\mu^t\right| < \mu^{t+1} + 1$ when t is odd. If u is odd the equation above is bounded by

$$
\begin{aligned}
J_x &\le \left|J_{x_{2,1}}\right|\mu^{u-1} + \left|J_{x_{2,2}}\right|\left(\mu^{u-1} + 1\right) + \cdots + \left|J_{x_{2,u}}\right| \\
&\le J_{x_1}^{(1)}\left(\left(\mu - \tilde{\zeta}_u\right)^2 + \tilde{\sigma}_u^2 - \tilde{\zeta}_u^2\right)^{(u-1)/2} + J_{x_2}^{(1)}\left(\left(\mu - \tilde{\zeta}_u\right)^2 + \tilde{\sigma}_u^2 - \tilde{\zeta}_u^2\right)^{(u-3)/2} \\
&\quad + \cdots + J_{x_{\frac{u+1}{2}}}^{(1)} + J_{x_{\frac{u+1}{2}+1}}^{(1)}\left(\left(\mu - \tilde{\zeta}_u\right)^2 + \tilde{\sigma}_u^2 - \tilde{\zeta}_u^2\right)^{(u-2)/2}(\mu - \tilde{\zeta}_u) \\
&\quad + J_{x_{\frac{u+1}{2}+2}}^{(1)}\left(\left(\mu - \tilde{\zeta}_u\right)^2 + \tilde{\sigma}_u^2 - \tilde{\zeta}_u^2\right)^{(u-4)/2}(\mu - \tilde{\zeta}_u) + \cdots + J_{x_u}^{(1)}(\mu - \tilde{\zeta}_u).
\end{aligned}
$$

Note that $\left(\mu - \tilde{\zeta}_u\right)^2 + \tilde{\sigma}_u^2 - \tilde{\zeta}_u^2 = \frac{\sigma_v^2 - \zeta_v^2}{\tilde{s}_u}$, then

$$
\begin{aligned}
J_x &\le \sum_{i=1}^{(u+1)/2} J_{x_i}^{(1)}\left(\frac{\sigma_v^2 - \zeta_v^2}{\tilde{s}_u}\right)^{(u-2i-1)/2} + \sum_{i=1}^{(u+1)/2} J_{x_{\frac{u+1}{2}+i}}^{(1)}\left(\frac{\sigma_v^2 - \zeta_v^2}{\tilde{s}_u}\right)^{(u-2i)/2}(\mu - \tilde{\zeta}_u) \\
&= J_x^{(1)}.
\end{aligned}
$$

Using the exact same logic when u is even, we can show that

$$J_x \le \sum_{i=1}^{u/2+1} J_{x_i}^{(2)} \left(\frac{\sigma_v^2-\zeta_v^2}{\tilde{s}_u}\right)^{(u-2i-2)/2} + \sum_{i=1}^{u/2} J_{x_{\frac{u}{2}+1+i}}^{(2)} \left(\frac{\sigma_v^2-\zeta_v^2}{\tilde{s}_u}\right)^{(u-2i)/2} (\mu-\tilde{\zeta}_u)$$
$$= J_x^{(2)}$$

where $\tilde{\zeta}_u$ and $\tilde{\sigma}_u^2-\tilde{\zeta}_u^2$ are the mean and a scale factor of μ. They are functions of $(\bar{X}_{v+1},\ldots,\bar{X}_k)$ and the values from M_J. Likewise, $J_{x\cdot}$ is an average of some combinations of $\mathbf{X}$. Thus, $J_{x\cdot}^{(1)}$ and $J_{x\cdot}^{(2)}$ are combinations of $\tilde{\zeta}_u$, $\tilde{\sigma}_u^2$ and $J_{x\cdot}$, and will converge to some value by the strong law of large numbers.

Now we will introduce some notation for this scenario:

$$\bar{\mathbf{X}}_{d,v} = \left(\bar{X}_{v+1}-\bar{X}_{v+2}, \bar{X}_{v+1}-\bar{X}_{v+3},\ldots,\bar{X}_{v+1}-\bar{X}_k, \bar{X}_{v+2}-\bar{X}_{v+3},\ldots,\bar{X}_{k-1}-\bar{X}_k\right)^T$$

is the difference in the sample mean of the variables that share a common mean, μ. Next, $\bar{\mathbf{X}}_h = \left(\bar{X}_h,\ldots,\bar{X}_k\right)^T$, $\bar{\mathbf{X}}_{\mu,h} = \left(\bar{X}_h, \bar{X}_{h+1}-\mu_{h+1}, \bar{X}_{h+2}-\mu_{h+2},\ldots,\bar{X}_k-\mu_k\right)^T$,

$$M_{j,k} = \begin{cases} M_J & \text{for } J=\{j,\ldots,k\} \\ 1 & \text{otherwise,}\end{cases} \qquad M_{u,k}^* = \det\left(1+M_{v+1,k}\right) - \det\left(M_{v+1,k}\right),$$

$$s_j = \begin{cases} \det\left(M_{2,k}\right) & \text{for } j=1 \\ \dfrac{\det\left(M_{j+1,k}\right)\det\left(M_{j-1,k}\right)}{\det\left(M_{j,k}\right)^2} & \text{for } j=2,\ldots,k, \end{cases} \quad \text{and}$$

$$\tilde{s}_u = \begin{cases} \dfrac{M_{u,k}^*\det\left(M_{v,k}\right)}{\det\left(M_{v+1,k}\right)^2} & \text{for } u<k \\ M_{k,k}^* & \text{for } u=k. \end{cases}$$

The next two terms $\sigma_h^2-\zeta_h^2$ and ζ_h can be thought of as a scale and mean of μ_h conditional on $\mu_{h+1},\ldots,\mu_k,\mu$. They are defined as:

$$\sigma_h^2-\zeta_h^2 = \frac{\det\left(M_{h,k}\right)}{\det\left(M_{h+1,k}\right)}\left(1+\frac{\bar{\mathbf{X}}_{\mu,h}^T D_h \bar{\mathbf{X}}_{\mu,h}}{\det\left(M_{h+1,k}\right)}\right), \quad \text{and} \quad \zeta_h = \frac{\bar{\mathbf{X}}_{\mu,h}^T D_h \mathbf{1}_0}{\det\left(M_{h+1,k}\right)}$$

for $h=1,\ldots,v$ where $\mathbf{1}_0=(1,0,\ldots,0)^T$. Likewise, $\tilde{\sigma}_u^2-\tilde{\zeta}_u^2$ and $\tilde{\zeta}_u$ can be thought of as a scale and mean of μ, the equal mean.

$$\tilde{\sigma}_u^2-\tilde{\zeta}_u^2 = \frac{\det\left(M_{v+1,k}\right)}{M_{u,k}^*}\left(1+\frac{\bar{\mathbf{X}}_{d,v}^T D \bar{\mathbf{X}}_{d,v}}{M_{u,k}^*}\right), \quad \text{and} \quad \tilde{\zeta}_u = \frac{\bar{\mathbf{X}}_{v+1}^T D\mathbf{1}}{M_{u,k}^*}$$

where $\mathbf{1}=(1,\ldots,1)^T$. The matrix D_h is the 1st adjugate matrix of $M_{h,k}$ and D is the 2nd adjugate matrix or $M_{v+1,k}$.

First and second adjugate matrix

Constructing a first order adjugate matrix will be done in a similar manner as laid out in [2]. $C^1(A)$ with elements $c_{ij}^{(1)}$ is the first adjugate of A, an $m\times m$ matrix. $c_{ij}^{(1)}$ is calculated by removing row i and column j from A, taking a determinant of the resulting minor, and multiplying it by $(-1)^{i+j}$. Specifically, $c_{ij}^{(1)} = (-1)^{i+j}\det(A_{-i,-j})$, where $A_{-i,-j}$ is the $(m-1)\times(m-1)$ minor of A with row i and column j removed.

The second order adjugate matrix of A, denoted $C^2(A)$ with elements $c_{ij}^{(2)}$, is calculated in a similar fashion. First, let $a^{(1)}, a^{(2)},\ldots,a^{(r)}$ be the $\binom{m}{2}$ pairs of the series $1,\ldots,m$ in lexicographic order. Now, to calculate $c_{ij}^{(2)}$ we will remove rows $a^{(i)}$ and columns $a^{(j)}$ from A, take the determinant of the resulting minor, and multiply it by $(-1)^{\sum_{l=1}^2 a_l^{(i)}+a_l^{(j)}}$. Specifically, $c_{ij}^{(2)} = (-1)^{\sum_{l=1}^2 a_l^{(i)}+a_l^{(j)}}\det(A_{-a^{(i)},-a^{(j)}})$, where $A_{-a^{(i)},-a^{(j)}}$ is the $(m-2)\times(m-2)$ minor of A with rows $a^{(i)}$ and columns $a^{(j)}$ removed, and $\sum_{l=1}^2 a_l^{(i)}+a_l^{(j)}$ is the sum of the rows and columns that were removed from A.

Proof of Theorem 1. After integrating the bounding equations we can now look at the asymptotic behavior of $P(\cdot)$. We will break this into 2 cases: first, when the correct model has only one largest mean, when $J=\{i: i=1,\ldots,k\}=J_1$ or equivalently $u=1$. The second case is when there are multiple largest means, $J\subseteq\{1,\ldots,k\}$ and $u\ge 2$. We will show that if J correctly indexes the largest mean(s) then $P(J)=p_J\left(\sum_{\tilde{J}\subseteq\{1,\ldots,k\}} p_{\tilde{J}}\right)^{-1}\to 1$, or equivalently $p_{\tilde{J}}/p_J\to 0$ for any $\tilde{J}\neq J$.

Before completing the proof we need to observe the following. Using Stirling's formula:

$$\frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+k+2-v}{2}\right)} \le \frac{2^{(k+1-v)/2}}{n^{(k-1-v)/2}} \quad \text{eventually a.s.}$$

Also, because the matrix D is the 2nd adjugate matrix, or $M_{v+1,k}$, and we know that $M_{v+1,k}$ is positive definite, then D is also a positive definite matrix. This comes from a result obtained by Rados [12], which states that the eigenvalues of D are a product of the $k-2$ eigenvalues of $M_{v+1,k}$ with different indices (e.g. the first eigenvalue of D is $\lambda_1^D = \lambda_3, \lambda_4, \ldots, \lambda_u$, where λ_i is the ith eigenvalue of $M_{v+1,k}$).

When J correctly indexes the largest means note that $\bar{\mathbf{X}}_{d,v} \to 0$. As a result, using the fact that D is positive definite along with the law of iterated logarithms, then we can write:

$$\frac{n+u+2}{2}\log\left(1+\frac{\bar{\mathbf{X}}_{d,v}^T D\bar{\mathbf{X}}_{d,v}}{M_{u,k}^*}\right) \le c\log\log n \quad \text{for some } c>0.$$

This means that $\left(1+\bar{\mathbf{X}}_{d,v}^T D\bar{\mathbf{X}}_{d,v}\left(M_{u,k}^*\right)^{-1}\right)^{(n+u+2)/2} \le (\log n)^c$ eventually a.s. For any $\tilde{J}$ that incorrectly indexes the largest mean(s) then $\left(1+\bar{\mathbf{X}}_{d,v}^T D\bar{\mathbf{X}}_{d,v}\left(M_{u,k}^*\right)^{-1}\right)^{(n+u+2)/2}$ grows exponentially.

Now we have the tools to show that $p_{\tilde{J}}/p_J \le p_{\tilde{J}}^{\blacktriangle}/p_J^{\blacktriangledown} \to 0$ in probability when J indexes the largest means and $J \neq \tilde{J}$. For example, if J_u is the correct index for the largest means, then

$$\begin{aligned}
\frac{p_{J_1}^{\blacktriangle}}{p_{J_u}^{\blacktriangledown}} &= \frac{J_x^{(0)}}{I_{J_u}b_\rho b}\frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+k+2-v}{2}\right)} \times \frac{\pi^{k/2}\tilde{s}_u^{(n+k+4-(v+1))/2}(\tilde{\sigma}_u^2-\tilde{\zeta}_u^2)^{(n+k+3-(v+1))/2}}{\pi^{(v+1)/2}s_{v+1}^{(n+k+3-v)/2}s_{v+2}^{(n+k+2-v)/2}\cdots s_k^{(n+4)/2}s_{k+1}^{(n+3)/2}} \\
&= \frac{J_x^{(0)}}{I_{J_u}b_\rho b}\frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+k+2-v}{2}\right)}\pi^{(u-1)/2}M_{u,k}^{*1/2}\left(1+\frac{\bar{\mathbf{X}}_{d,v}^T D\bar{\mathbf{X}}_{d,v}}{M_{u,k}^*}\right)^{(n+2+u)/2} \\
&\le \frac{J_x^{(0)}}{I_{J_u}b_\rho b}\pi^{(u-1)/2}M_{u,k}^{*1/2}\frac{2^{(u+1-v)/2}}{n^{(u-1)/2}}(\log n)^c \quad \text{eventually a.s.} \\
&\xrightarrow{\mathcal{P}} 0.
\end{aligned}$$

We can see that this converges polynomially. A similar calculation when J_1 correctly indexes the largest mean shows exponential convergence when u is odd:

$$\begin{aligned}
\frac{p_{J_u}^{\blacktriangle}}{p_{J_1}^{\blacktriangledown}} &\le \frac{1}{I_{J_1}bb_\rho}\frac{4n^{u/2}}{\pi^{(u-1)/2}}\sum_{i=1}^{(u+1)/2}\frac{J_{x_i}^{(1)}\det\left(M_{v+1,k}\right)^{(u+1-2i)/2}}{M_{u,k}^{*(u+2-2i)/2}\left(1+\frac{\bar{\mathbf{X}}_d^T D\bar{\mathbf{X}}_d}{M_{u,k}^*}\right)^{(n+1+2i)/2}} \quad \text{eventually a.s.} \\
&\xrightarrow{\mathcal{P}} 0.
\end{aligned}$$

In a similar fashion, we have shown that $p_{J_u}^{\blacktriangle}/p_{J_1}^{\blacktriangledown} \to 0$ in probability when u is even, and $p_{J_{u_1}}^{\blacktriangle}/p_{J_{u_2}}^{\blacktriangledown} \to 0$ in probability for some J_{u_2} that correctly indexes the largest means and $u_1, u_2 \ge 2$.

These calculations have shown that for any $J \neq \tilde{J}$ then $p_{\tilde{J}}/p_J \le p_{\tilde{J}}^{\blacktriangle}/p_J^{\blacktriangledown} \to 0$ in probability when J correctly indexes largest mean(s). Specifically, we have shown, $P(J) = p_J\left(\sum_{\tilde{J}\subseteq\{1,\ldots,k\}} p_{\tilde{J}}\right)^{-1} \to 1$ in probability. This completes the proof of Theorem 1. □

Appendix C. Importance sampling algorithm

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

1. Start by generating $\mu = \tilde{\zeta}_u + \sqrt{\frac{\tilde{\sigma}_u^2-\tilde{\zeta}_u^2}{n+u+2}}T_u$, where $T_u \sim t(n+u+2)$.
2. Generate $(\mu_h|\mu_{h+1},\ldots,\mu_v,\mu) = \zeta_h + \sqrt{\frac{\sigma_h^2-\zeta_h^2}{n+k+3-h}}T_h$, where $T_h \sim t(n+k+3-h)$ for all $h=1,\ldots,v$.
3. Generate $\left(\Sigma^{-1}|\boldsymbol{\mu}\right) = W$, where $W \sim \mathit{Wishart}(n+k+3, S^{-1})$
4. Calculate weights of each generated sample with

$$w_J = \frac{f_J(\xi)}{g(\mu)\left(\prod_{i=1}^{v} g_i(\mu_i)\right)h(\Sigma^{-1})}$$

where $f_J(\xi)$ is the generalized fiducial density for the model with index J, and $g(\mu)$, $g_i(\mu_i)$ and $h(\Sigma^{-1})$ are the densities from distributions described in steps 1, 2 and 3 respectively.

5. This process was repeated until we achieved the effective sample size calculated by $ESS_J = n_J \left(1 + \left(s^2_{w_J}\right) \bar{w}_J^{-2}\right)^{-1}$, where n_J is the sample size for model J, $s^2_{w_J}$ is the sample variance of the weights, and $\bar{w}_J$ is the sample mean of the weights.
6. Lastly the weights were divided by the ESS_J, and all samples that did not meet the condition of $\mu > \max_{i \le v} \mu_i$ were eliminated.
7. This process was repeated for all possible models $J \subseteq \{1, \ldots, k\}$.

References

[1] D.D. Boos, D. Hoffman, R. Kringle, J. Zhang, New confidence bounds for QT studies, Statistics in Medicine 26 (2007) 3801–3817.
[2] N.G.D. Bruijn, Inequalities concerning minors and eigenvalues, Nieuw Archief voor Wiskunde IV (1956) 18–35.
[3] A.P. Dawid, Multivariate Analysis and Its Applications, in: Lecture Notes-Monograph Series Selection Paradoxes of Bayesian Inference, vol. 24, Institute of Mathematical Statistics, Hayward, CA, 1994, pp. 211–220.
[4] M.L. Eaton, R.J. Muirhead, J.Y. Mancuse, S. Kolluri, A confidence interval for maximal mean QT interval change caused by drug effect, Drug Information Journal 40 (3) (2006) 267–271.
[5] EPA, 2009. http://www.epa.gov/air/data/geosel.html.
[6] R.A. Fisher, Inverse probability, Proceedings of the Cambridge Philosophical Society (1930) xxvi:528–535.
[7] J. Hannig, 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, 2009.
[8] J. Hannig, On generalized fiducial inference, Statistica Sinica 19 (2009) 491–544.
[9] J. Hannig, H.K. Iyer, P. Patterson, Fiducial generalized confidence intervals, Journal of the American Statistical Association 101 (2006) 254–269.
[10] J. Hannig, T.C. Lee, (2009). Generalized fiducial inference for wavelet regression, Biometrika (in press).
[11] D.V. Lindley, Fiducial distributions and Bayes' theorem, Journal of the Royal Statistical Society Series B 20 (1958) 102–107.
[12] G. Rados, Zur theorie der adjungirten substitutionen, Mathematische Annalen 48 (3) (1896) 417–424.
[13] S. Weerahandi, Generalized confidence intervals, Journal of the American Statistical Association 88 (1993) 899–905.
[14] S.L. Zabell, R.A. Fisher and the fiducial argument, Statistical Science 7 (1992) 369–387.