# DISSERTATION

# A FIDUCIAL APPROACH TO EXTREMES AND MULTIPLE COMPARISONS

Submitted by Damian V. Wandler Department of Statistics

In partial fulfillment of the requirements for the Degree of Doctor of Philosophy Colorado State University Fort Collins, Colorado Spring 2010

# COLORADO STATE UNIVERSITY

February 12, 2010

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY DAMIAN V. WANDLER ENTITLED A FIDUCIAL AP-PROACH TO EXTREMES AND MULTIPLE COMPARISONS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Haonan Wang

Edwin Chong

Advisor: Jan Hannig

Co-Advisor: Hari Iyer

Department Chair: F. Jay Breidt

# ABSTRACT OF DISSERTATION

# A FIDUCIAL APPROACH TO EXTREMES AND MULTIPLE COMPARISONS

Generalized fiducial inference is a powerful tool for many difficult problems. Based on an extension of R. A. Fisher's work, we used generalized fiducial inference for two extreme value problems and a multiple comparison procedure.

The first extreme value problem is dealing with the generalized Pareto distribution. The generalized Pareto distribution is relevant to many situations when modeling extremes of random variables. We use a fiducial framework to perform inference on the parameters and the extreme quantiles of the generalized Pareto. This inference technique is demonstrated in both cases when the threshold is a known and unknown parameter. Simulation results suggest good empirical properties and compared favorably to similar Bayesian and frequentist methods.

The second extreme value problem pertains to the largest mean of a multivariate normal distribution. Difficulties arise when two or more of the means are simultaneously the largest mean. Our solution uses a generalized fiducial distribution and allows for equal largest means to alleviate the overestimation that commonly occurs. Theoretical calculations, simulation results, and application suggest our solution possesses promising asymptotic and empirical properties.

Our solution to the largest mean problem arose from our ability to identify the correct largest mean(s). This essentially became a model selection problem. As a result, we applied a similar model selection approach to the multiple comparison problem. We allowed for all possible groupings (of equality) of the means of k independent normal distributions. Our resulting fiducial probability for the groupings

of the means demonstrates the effectiveness of our method by selecting the correct grouping at a high rate.

> Damian V. Wandler Department of Statistics Colorado State University Fort Collins, Colorado 80523 Spring 2010

### ACKNOWLEDGEMENTS

I would like to thank the department of Statistics and my committee members for all of their assistance. This opportunity to complete a PhD would have not been possible without the generosity of the department and the help of my committee. In my time at CSU, Jan Hannig has been a great teacher, advisor, and friend. He has always been available to help and offer support and encouragement throughout this process. Both Hari Iyer and Jan were instrumental in exposing me to the fiducial methods and perking my interest in this novel inference technique. Though no longer part of my committee, Dan Cooley was an incredible resource to help me understand extreme value concepts. I am also grateful for the involvement of Edwin Chong and Haonan Wang. I appreciated the contribution of all my committee members and I am very thankful to you all.

I have received funding from various sources that I am grateful for. The CSU PRIMES program served as a valuable opportunity to collaborate with other scientist and fund a portion of my research under the National Science Foundation's IGERT grant #DGE-0221595. Part of this research was also supported by National Science Foundation under grant #DMS- 0707037. I would also 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 and 2010.

I also owe a thank you to my parents and brother. Both my parents and brother have taught me valuable lessons about hard work and perseverance. Without their support this accomplishment would have been much tougher. Lastly, I would like to thank Jacque for everything she has done for me. Her patience, love, and support helped keep me focused and motivated through this trying and rewarding journey.

# DEDICATION

To Jacque and our future.

# CONTENTS

1	Intr	oduction to Extremes and Multiple Comparisons	1
	1.1	Introduction to the generalized Pareto	1
	1.2	Introduction to the largest mean of a multivariate normal distribution	3
	1.3	Introduction to multiple comparisons	4
<b>2</b>	Ger	neralized Fiducial Confidence Intervals for Extremes	6
	2.1	Introduction	6
	2.2	Generalized Fiducial Inference	8
		2.2.1 Overview	8
	2.3	Main Results	10
		2.3.1 Structural equation	10
		2.3.2 Confidence intervals, coverage, and point estimates when $a$ is	
		known	12
		2.3.3 Confidence intervals and coverage when $a$ is unknown	18
	2.4	Nasdaq 100 data set	23
	2.5	Conclusion	25
3	Fid	ucial Inference on the Largest Mean of a Multivariate Normal	L
	Dist	tribution	<b>27</b>
	3.1	Introduction	27
	3.2	Generalized Fiducial Inference	29
	3.3	Main Results	31
		3.3.1 Two dimensional case	31
		3.3.2 General case	36
		3.3.3 Confidence intervals and coverage	37
		3.3.4 Simulation results and discussion	41
	3.4	Asymptotic Results	44
	3.5	Air quality example	45
	3.6	Conclusion	48
<b>4</b>	Fid	ucial Approach to Multiple Comparisons	<b>58</b>
	4.1	Introduction	58
	4.2	Generalized Fiducial Inference	59
		4.2.1 Overview	59
	4.3	Main Results	61
		4.3.1 Structural equation with constant variance	61
		4.3.2 Structural equation with non-constant variance	64

		4.3.3 Simulations	66
	4.4	Asymptotic results	72
	4.5	Examples	72
	4.6	Conclusion	76
<b>5</b>	Con	clusion and Future Work	87
	5.1	Conclusion and future work for the generalized Pareto	87
	5.2	Conclusion and future work for the largest mean of a multivariate nor-	
		mal distribution	88
	5.3	Conclusion and future work for multiple comparisons $\ldots \ldots \ldots$	89
Re	eferei	nces	91

# LIST OF TABLES

2.1	Simulation values when $a$ is known.	14
2.2	Distributions of the simulated data	18
2.3	Estimates and confidence intervals for the 0.99-quantile of the Nasdaq $100$	
	data set.	24
3.1	Coverage for a 95% upper tailed interval when $\Sigma = I$ for the naive method.	32
3.2	Coverage for a 95% upper tailed interval when $\Sigma = I$ when the $\mu_1 = \mu_2$ is	
	assumed.	33
3.3	Simulation combinations in two dimensions	41
3.4	Simulation combinations in three dimensions	42
3.5	AQI range of health effects	46
3.6	Largest mean probabilities for each model, $J$	47
4.1	Multiple comparison $P(J)$ for the simulated example	73
4.2	Rhizobium Data	75
4.3	Multiple comparison $P(J)$ for the red clover example	75
4.4	Posterior probabilities of select $J$ for the red clover example	76

# LIST OF FIGURES

2.1	Equal tailed confidence region for $(\gamma_0, \sigma_0)$	13
2.2	QQ-plots when $\gamma_0 = -0.2$ , $\sigma_0 = 1$ , and $n = 50$	14
2.3	QQ-plots when $\gamma_0 = 0$ and $\sigma_0 = 1$ , and $n = 50 \dots \dots \dots \dots \dots$	15
2.4	QQ-plots when $\gamma_0 = 0.4$ and $\sigma_0 = 1$ , and $n = 50 \dots \dots \dots \dots \dots$	15
2.5	Coverage for the 0.99-quantile with various $\gamma_0$ values and $\sigma_0 = 1$	16
2.6	Length of the two tailed fiducial, Bayesian, and profile log-likelihood in-	
	tervals for the 0.99-quantile when $\gamma_0 > 0$	16
2.7	Length of the two tailed fiducial, Bayesian, and profile log-likelihood in-	
	tervals for the 0.99-quantile when $\gamma_0 \leq 0$	17
2.8	Absolute bias and standard deviation of the point estimates	17
2.9	QQ-plots for the 0.99 and 0.999 quantile when $X \sim Exp(1)$ using the	
	fiducial method	19
2.10	QQ-plots for the 0.99 and 0.999 quantile when $X \sim t(5) + 10$ using the	
	fiducial method	19
2.11	QQ-plots for the 0.99 and 0.999 quantile when $X \sim N(10, 100)$ using the	
	fiducial method	20
2.12	QQ-plots for the 0.99 and 0.999 quantile when $X \sim t(5) + 10$ using the	
	Bayesian method with a nonparametric central model	21
2.13	QQ-plots for the 0.99 and 0.999 quantile when $X \sim Exp(1)$ using the	
	Bayesian method with a normal central model	21
2.14	QQ-plots for the 0.99 and 0.999 quantile when $X \sim N(10, 100)$ using the	
	Bayesian method with a normal central model	22

2.15	QQ-plots for the 0.99 and 0.999 quantile when $X \sim t(10) + 10$ using the	
	Bayesian method with a normal central model	22
2.16	Length of the two tailed fiducial and Bayesian intervals for the 0.99-	
	quantile when the data was generated from $Exp(1)$ , $Normal(10, 100)$ ,	
	and $t(5) + 10$ respectively	23
3.1	QQ-plots when $n = 30$ , $\boldsymbol{\mu}_0 = (5, 5, 5)^T$ , $\boldsymbol{\Sigma}_0 = I$ and $25I$ for the top and	
	bottom rows respectively	39
3.2	QQ-plots when $n = 30$ , $\mu_0 = (3, 4, 5)^T$ , and $\Sigma_0 = 25I$	40
3.3	QQ-plots when $n = 30$ , $\mu_0 = (1, 3, 5)^T$ , and $\Sigma_0 = I$	40
3.4	QQ-plots when $n = 10000$ , $\mu_0 = (3, 4, 5)^T$ , and $\Sigma_0 = 25I$	40
3.5	Coverage when the data is two dimensional and $\rho^{(1,2)} > 0$	42
3.6	Length of the upper tailed intervals relative to the fiducial interval when	
	data is two dimensional and $\rho^{(1,2)} > 0$ .	42
3.7	Coverage when the data is two dimensional and $\rho^{(1,2)} \leq 0.$	43
3.8	Length of the upper tailed intervals relative to the fiducial interval when	
	data is two dimensional and $\rho^{(1,2)} \leq 0$ .	43
3.9	Coverage when the data is three dimensional and $\rho^{(i,j)}$ is random	44
3.10	Length of the upper tailed intervals relative to the fiducial interval when	
	data is three dimensional and $\rho^{(i,j)}$ is random.	44
3.11	Fiducial method QQ-plots using $\boldsymbol{\mu}_0$ and $\Sigma_0$ from (3.15).	47
3.12	Competing methods QQ-plots using $\mu_0$ and $\Sigma_0$ from (3.15)	47
4.1	$P(J)$ for $\boldsymbol{\mu}_0 = (1,1,1)$ and $\eta_0 = 1$ and 100 for top and bottom row	
	respectively.	67
4.2	$P(J)$ for $\mu_0 = (1, 1.5, 1.5)$ and $\eta_0 = 1.$	68
4.3	$P(J)$ for $\mu_0 = (1, 1.5, 2)$ and $\eta_0 = 1$ .	68
4.4	$P(J)$ for $\mu_0 = (1, 3, 5)$ and $\eta_0 = 1$ .	69
4.5	$P(J)$ for $\mu_0 = (1, 1, 2, 2)$ and $\eta_0 = 1$ .	69

4.6	$P(J)$ for $\mu_0 = (1, 1, 1)$ and $\eta_0 = (1, 1, 1), (1, 2, 3),$ and $(100, 100, 100)$ for	
	top, middle, and bottom rows respectively. $\ldots \ldots \ldots \ldots \ldots$	70
4.7	$P(J)$ for $\mu_0 = (1, 3, 5)$ and $\eta_0 = (1, 1, 1)$ .	71
4.8	$P(J)$ for $\mu_0 = (1, 1, 2, 2)$ and $\eta_0 = (1, 1, 1, 1)$ .	71
4.9	$P(\mu_i = \mu_j)$ and $P(\mu_i = \mu_{i+1} = \cdots = \mu_{i+r})$ with constant and non-constant	
	variance for the simulated example.	74
4.10	$P(\mu_i = \mu_j)$ and $P(\mu_i = \mu_{i+1} = \cdots = \mu_{i+r})$ with constant and non-constant	
	variance for the red clover example.	76

# Chapter 1

# INTRODUCTION TO EXTREMES AND MULTIPLE COMPARISONS

Throughout this dissertation we will investigate two extreme value problems and a multiple comparison problem. The first extreme value problem performs inference on the parameters and extreme quantiles of the generalized Pareto distribution. This is a classic extreme value problem that has been investigated using various approaches. The second extreme value problem deals with inference on the largest mean of a correlated multivariate normal distribution. The largest mean is easily recognized as an extreme parameter. Estimating and calculating confidence intervals for the largest mean pose numerous overestimation problems when there are multiple means that are the equal largest means. Grouping several means together as the largest mean will help to alleviate such problems. Grouping the means together becomes a model selection problem as there are different combinations of possible groupings of the largest means.

A natural extension to the model selection of the largest mean problem is the development of a solution to multiple comparison problems. The multiple comparison problem will estimate which means of independent normal data are equal and unequal. Again, this will involve a model selection technique amongst the potential groupings.

#### 1.1 Introduction to the generalized Pareto

The generalized Pareto distribution (GPD) was introduced by Pickands (1975). It was used in situations that an exponential distribution may be appropriate but robustness may be required with heavy or light tailed alternatives. In particular, it is known that data of the peaks over a high threshold variety are appropriately modeled by the generalized Pareto distribution. These situations arise in economics, hydrology, environmental science, insurance, etc.

As defined by Pickands (1975), if  $X \sim F$  then the limiting distribution of X - a conditional on X > a as  $a \to \omega_F$  where  $\omega_F$  is the right-hand endpoint of the distribution follows a generalized Pareto distribution. The density for the GPD is defined as:

$$f(x) = \begin{cases} \frac{1}{\sigma} \left( 1 + \frac{\gamma(x-a)}{\sigma} \right)^{-1/\gamma - 1} & \gamma \neq 0\\ \frac{1}{\sigma} \exp\left\{ - \left( x - a \right) / \sigma \right\} & \gamma = 0 \end{cases}$$

Here  $\gamma$  is the shape parameter,  $\sigma$  is the scale parameter, and a is the threshold parameter. The generalized Pareto distribution poses numerous interesting problems. Estimating  $\gamma$ ,  $\sigma$ , a, and the high quantile (return level) can produce many difficulties.

In many cases the threshold, a, is considered a known value and estimators for  $\gamma$ ,  $\sigma$ , and the  $\beta$ -quantile are calculated under this assumption. This is the case in Smith (1984, 1985), Davison (1984), Monfort and Witter (1985), Davison (1984), Hosking and Wallis (1987), and Castellanos and Cabras (2005). Estimators have included the use of maximum likelihood, method of moments, L-moments, and Bayesian techniques. The maximum likelihood estimators were found to be asymptotically normal and consistent when  $\gamma > -1/2$  in Smith (1984). L-moments were found to be equivalent to probability weighted moments and have good qualities in Hosking (1990). Castellanos and Cabras (2005) used a Bayesian technique with Jeffereys prior to calculate estimators.

Work has also been done when the threshold, a, is an unknown parameter. When the threshold is unknown the interest lies in the estimation and confidence intervals for the  $\beta$ -quantile. In the past, a threshold value was fixed and inference was done using the fixed threshold. When the threshold is not set as a fixed value additional inference problems arise. Some methods test how well the generalized Pareto fits a data set for different threshold values (Choulakian and Stephens (2001) and Dupuis (1999)). Another approach, seen in Guillou and Hall (2001), is to model a fixed number of the largest order statistics with the GPD. Weighting schemes have been incorporated with a mixture of a Weibull and GPD in Frigessi et al. (2002). Cabras and Castellanos (2009) and Tancredi et al. (2006) used a Bayesian approach with a mixture model to incorporate the variability of the unknown threshold. Using a mixture model allows the data to select likely threshold values and build confidence intervals for the  $\beta$ -quantile with this added uncertainty.

The use of a mixture model as done in Cabras and Castellanos (2009) and Tancredi et al. (2006) is a novel idea. The majority of the distribution can be modeled by some distribution and the tail can be modeled using the GPD. However, as the underlying distribution of the data can be anything the appropriate distribution needs to be fit below the threshold to have any sort success in estimating the high quantile in the upper tail of the data.

#### 1.2 Introduction to the largest mean of a multivariate normal distribution

The largest mean is defined by  $\theta = \max_i \mu_i$  when  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \ \boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$ , and

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

Inference on  $\theta$  is a wide reaching and difficult problem. There are applications to this problem in pollution studies, drug trials, studies measuring the QT interval, investments, etc.

As the parameter of interest is the largest mean there is a very difficult aspect to this problem. If some or all of the means are equal or close to being equal the largest mean,  $\theta$ , is oftentimes overestimated. As a result, some of the current inference techniques have very conservative upper tailed confidence intervals. There are very few potential solutions to this problem. The most common, the intersection-union method, constructs an upper tailed *t*-interval for all of the kdimensions. The largest upper bound is the upper tailed confidence interval for  $\theta$ . Another proposed technique is seen in Eaton et al. (2006). This solution approximates the maximum with a smooth function which allows for the delta method to make a normal approximation and a bias adjustment. Boose et al. (2007) uses a linear model to estimate the bias and variation of the estimated largest mean. This technique allows for intervals to be calculated with and without a bias adjustment using a normal approximation. All of these methods create upper-tailed intervals but do not propose a solution for a lower-tailed or two tailed interval.

A model selection technique that would allow for equal largest means would help to offset the overestimation and the need for a bias adjustment. Furthermore, allowing for equal largest means will provide good asymptotic properties.

### **1.3** Introduction to multiple comparisons

Multiple comparisons allow for judgements to be made about the means of independent normal distributions. Namely,  $\mathbf{X}_i = (X_{i1}, \dots, X_{in_i})$  for all  $i = 1, \dots, k$ are observed from independent normal distributions with mean  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$ and variance  $\boldsymbol{\eta}$ . The interesting problem is to perform inference on the groupings of the individual means within  $\boldsymbol{\mu}$ . For example, it may be of interest to determine if  $\mu_1 = \mu_2 = \dots = \mu_k$  from the observations  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k$ .

Some of the multiple comparison problems have been addressed using frequentist solutions. In the previous example, analysis of variance (ANOVA) would be used to test for a treatment affect. Likewise, to test for differences among the individual means some of the frequentist solutions consist of Fisher's least significant difference (LSD), Tukey's honest significant difference (HSD), Sheffe's pairwise differences, Duncan's multiple range test, etc. While these methods can detect differences while controlling for the comparisonwise or experimentwise error rate they do not necessarily lead to a conclusion of the likely groupings of the means within  $\mu$ .

A Bayesian procedure for the multiple comparison problem has been developed in Gopalan and Berry (1998). The Bayesian procedure will produce a posterior probability for the likely groupings within  $\mu$ . This method assigns prior and hyperprior distributions to the parameters of  $\mu$  and  $\eta$ . Furthermore, a Dirichlet process prior is used to decide between competing groupings of  $\mu$ . Using this method posterior probabilities are compared to the prior probabilities to decide amongst competing groupings for different priors.

The Bayesian method does address some of the shortcomings in the frequentist solutions. However, the posterior probabilities can be dramatically affected by the prior distributions.

### Chapter 2

# GENERALIZED FIDUCIAL CONFIDENCE INTERVALS FOR EXTREMES

#### 2.1 Introduction

Extreme value theory is of practical interest in a variety of different fields (e.g. economics, hydrology, environmental science, insurance, etc.). It is well known that modeling data over a high threshold with the generalized Pareto distribution (GPD) is appropriate, Davison and Smith (1990). As stated by Hosking and Wallis (1987) the applications of the GPD include analysis in extreme events like the modeling of large insurance claims and in situations that an exponential distribution might be used but robustness is required with heavy or light tailed alternatives.

The generalized Pareto was first introduced by Pickands (1975). Later Smith (1984, 1985), Davison (1984), and Monfort and Witter (1985) all showed interest in its application and theoretical properties. If  $X \sim F$ , Pickands (1975) showed that the limiting distribution of (X - a) conditional on X > a as  $a \to \omega_F$  where  $\omega_F$  is the right-hand endpoint of the distribution follows a generalized Pareto distribution. The density for the GPD is defined as:

$$f(x) = \begin{cases} \frac{1}{\sigma} \left( 1 + \frac{\gamma(x-a)}{\sigma} \right)^{-1/\gamma - 1} & \gamma \neq 0\\ \frac{1}{\sigma} \exp\left\{ - \left( x - a \right) / \sigma \right\} & \gamma = 0 \end{cases}$$

Estimators of the parameters using maximum likelihood, method of moments, and L-moments have been explored in Davison (1984), Hosking and Wallis (1987), and Smith (1984). Smith (1984) found the maximum likelihood estimators to be asymptotically normal and consistent when  $\gamma > -1/2$ . Hosking (1990) showed that procedures using L-moments and probability weighted moments are equivalent. Furthermore, L-moments are more robust than method of moments and are often times more efficient than maximum likelihood estimates. A Bayesian solution to this problem has been explored in Castellanos and Cabras (2005).

We propose new confidence intervals for  $\gamma$ ,  $\sigma$ , and the  $\beta$ -quantiles for the cases when the threshold, a, is a known and unknown parameter. The proposed solution is based on generalized fiducial intervals of Hannig (2009b). Simulation results suggest that this inference technique performs well with small sample sizes, and, when a is known, the confidence intervals have asymptotically correct coverage. This fiducial method for calculating intervals for the extreme quantiles (return levels) of the GPD also compare favorably to the profile log-likelihood method described in Coles (2001) and the Bayesian method described in Castellanos and Cabras (2005). That is to say that the fiducial intervals have good empirical coverage and are often times shorter than the comparable profile log-likelihood and Bayesian intervals. Furthermore, the point estimates for  $\gamma$  and  $\sigma$  using this fiducial approach have smaller bias when compared with estimators calculated using maximum likelihood, L-moments, and the aforementioned Bayesian methods. The bias for the estimate of the  $\beta$ -quantile is smaller than the estimates based on maximum likelihood and L-moments but slightly larger than those calculated by the Bayesian method.

We also developed fiducial methods when the threshold is unknown. As seen in Coles (2001), the threshold is generally chosen by some *ad hoc* procedure of looking at plots and fixing the threshold for all subsequent calculations. Other methods will test whether the generalized Pareto fits the data for various different thresholds as seen in Choulakian and Stephens (2001) and Dupuis (1999). Guillou and Hall (2001) investigate this problem of choosing a threshold by using a fixed number of the largest order statistics and Frigessi et al. (2002) used a weighting scheme with a mixture of a Weibull distribution and the GPD to model the data. These methods do

not account for fact that the threshold is unknown in all practical applications. The unknown threshold will add variability to the estimates of the extreme quantiles. Our method, like the Bayesian methods developed in Cabras and Castellanos (2009) and Tancredi et al. (2006), assumes the threshold is another parameter that is unknown. As a result, the fiducial method will select likely values for the threshold based on the data. Using this method we performed a simulation study for data that was generated from various distributions that could be seen in real life settings. Based on the simulations the fiducial framework produced intervals for the  $\beta$ -quantile that had reasonable frequentist coverage for all of the distributions and compared favorably to the approach described in Cabras and Castellanos (2009).

This fiducial approach was also used to analyze a popular data set in both cases when the threshold is assumed to be known and unknown. The data set that was analyzed was the log-weekly losses of the Nasdaq 100 index. Our analysis produced fiducial intervals for the 0.99-quantile that were generally shorter than the intervals from the appropriate profile log-likelihood and Bayesian methods.

### 2.2 Generalized Fiducial Inference

#### 2.2.1 Overview

The original idea for fiducial inference was developed by Fisher (1930) in an attempt to overcome what he perceived as a deficiency in the Bayesian framework. Namely, he was opposed to assuming a prior distribution when there was little or no information about the parameters available. Opposition to the fiducial framework arose when it was later discovered that some of the properties that Fisher had originally claimed were not actually true, (Lindley, 1958; Zabell, 1992). In the more recent past fiducial inference has begun to gain more acceptance in the statistics community following the introduction of generalized inference by Weeranhandi (1993) and the work of Hannig et al. (2006) where a relationship between fiducial and generalized

inference was established. A further background of fiducial inference and discussion of the asymptotic and empirical properties can be found in Hannig (2009b).

The underlying principle of generalized fiducial inference uses the model and the observed data,  $\mathbf{X}$ , to define a probability measure on the parameter space,  $\Xi$ . The result produces a similar distribution to that of a Bayesian posterior distribution. To formally describe fiducial inference we assume that a relationship between  $\mathbf{X}$  and  $\boldsymbol{\xi}$  exists in the form of,

$$\mathbf{X} = G(\xi, \mathbf{U}). \tag{2.1}$$

This relationship is called the *structural equation* where **U** is a random vector with a completely known distribution and independent of any parameters. After **X** is observed we use the known distribution of **U** and the functional relationship from the structural equation to infer a distribution on  $\xi$ . The distribution of  $\xi$  is established by taking the inverse of the structural equations with respect to  $\xi$ .

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  $i = 1, \ldots, n$ .  $\xi$  is a  $p \times 1$  vector and denote  $\mathbf{X}_i = G_{0,i}(\xi, \mathbf{U}_i)$  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  $\mathbf{i} = (i_1, \ldots, i_p)$ . Assume that the functions  $\mathbf{G}_{0,i}$  are oneto-one and differentiable. Under some technical assumptions in Hannig (2009a) this will produce the generalized fiducial density of

$$f_{\mathcal{R}_{\xi}}(\xi) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')d\xi'}$$
(2.2)

where

$$J(\mathbf{x},\xi) = {\binom{n}{p}}^{-1} \sum_{\mathbf{i}=(i_1,\dots,i_p)} \left| \frac{\det\left(\frac{d}{d\xi} \mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)}{\det\left(\frac{d}{d\mathbf{u}_{\mathbf{i}}} \mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)} \right|$$
(2.3)

is the mean of all subsets where  $1 \le i_1 < \cdots < i_p \le n$  and the determinants in (4.5) are the appropriate Jacobians.

### 2.3 Main Results

#### 2.3.1 Structural equation

#### Known threshold

If  $X_1, X_2, \ldots, X_n$  are independent and identically distributed random variables from the GPD with parameters  $\gamma$  and  $\sigma$  then a structural equation when a is known can be defined as:

$$X_i = a + (U_i^{-\gamma} - 1)\frac{\sigma}{\gamma}$$

where  $U_i$  for i = 1, ..., n are independent random variables from the U(0, 1) distribution.

Following the recipe from (4.4) we were able to calculate the fiducial density for  $\xi = (\gamma, \sigma)$ . The generalized fiducial density when a is known is:

$$f_{\mathcal{R}_{\xi}}(\xi) \propto \frac{1}{\sigma^{n}} \prod_{i=1}^{n} \left( 1 + \frac{\gamma \left( x_{i} - a \right)}{\sigma} \right)^{-\frac{1}{\gamma} - 1} I_{(a,\infty)}(x_{(1)}) I_{(0,\infty)}(\sigma) I_{(-\frac{\sigma}{x_{(n)-a}},\infty)}(\gamma) \\ \times \binom{n}{2}^{-1} \frac{1}{\gamma^{2}} \sum_{i < j} \left| (x_{i} - a) \left( 1 + \frac{\gamma \left( x_{j} - a \right)}{\sigma} \right) \log \left( 1 + \frac{\gamma \left( x_{j} - a \right)}{\sigma} \right) \right| \\ - (x_{j} - a) \left( 1 + \frac{\gamma \left( x_{i} - a \right)}{\sigma} \right) \log \left( 1 + \frac{\gamma \left( x_{i} - a \right)}{\sigma} \right) \left| (2.4) \right|$$

where  $\mathcal{R}_{\xi} = (\mathcal{R}_{\gamma}, \mathcal{R}_{\sigma})$  is the fiducial random variable for  $(\gamma, \sigma)$ .

To find the fiducial density for the  $\beta$ -quantile (return level) a transformation is needed to be made on (2.4). Namely, we need to find the distribution of:

$$\mathcal{R}_q = a + \frac{\mathcal{R}_\sigma}{\mathcal{R}_\gamma} \left( (1 - \beta)^{-\mathcal{R}_\gamma} - 1 \right) \qquad \beta \in (0, 1)$$
(2.5)

where  $\mathcal{R}_q$  is the fiducial random variable associated with the  $\beta$ -quantile. The fiducial density for the  $\beta$ -quantile is:

$$f_{\mathcal{R}_q}(q) \propto \int f_{\mathcal{R}_{\gamma q}}(\gamma, q) \, d\gamma$$
 (2.6)

where  $f_{\mathcal{R}_{\gamma q}}(\gamma, q)$  is the joint density of  $(\mathcal{R}_{\gamma}, \mathcal{R}_{q})$  using the transformation in (2.5).

#### Unknown threshold

In most practical applications the threshold is unknown and the GPD does not fit the tail of the distribution exactly. As a result, we consider the following model as an approximation:

$$X_{i} = I_{(0,p)}(U_{i}) \left( aW_{i} \right) + I_{(p,1)}(U_{i}) \left( a + (W_{i}^{-\gamma} - 1)\frac{\sigma}{\gamma} \right)$$
(2.7)

where  $U_i$  and  $W_i$  for i = 1, ..., n are independent random variables from the U(0, 1)distribution. Furthermore, p is chosen so the density based on (2.7) is continuous at a.

As we are only interested in the data above the threshold the model below the threshold serves as a way of introducing a penalty. The penalty helps to ensure that the threshold is not forced far into the tail of the distribution. The uniform distribution, while not the correct distribution of the data below the threshold, seems to introduce the correct penalty for selecting the threshold as demonstrated by our simulation studies.

Using this structural equation, the recipe described earlier produces the generalized fiducial density as:

$$f(\xi) \propto \sum_{i=1}^{B} \left\{ \frac{I_{\left(x_{(i)}, x_{(i+1)}\right)}(a) J\left(\mathbf{x}_{i+1:n}, \xi\right)}{\left(a+\sigma\right)^{n}} \prod_{j=i+1}^{n} \left[ \left(1 + \frac{\gamma\left(x_{(j)}-a\right)}{\sigma}\right)^{-\frac{1}{\gamma}-1} \right] \right\}$$
(2.8)

where

$$J(\mathbf{x}_{i+1:n},\xi) = \binom{n-i-1}{3}^{-1} \frac{1}{\gamma^2} \sum_{j < k < l} \left| (x_j - x_l) \left( 1 + \frac{\gamma(x_k - a)}{\sigma} \right) \log \left( \frac{\gamma(x_k - a)}{\sigma} + 1 \right) - (x_k - x_l) \left( 1 + \frac{\gamma(x_j - a)}{\sigma} \right) \log \left( \frac{\gamma(x_j - a)}{\sigma} + 1 \right) - (x_j - x_k) \left( 1 + \frac{\gamma(x_l - a)}{\sigma} \right) \log \left( \frac{\gamma(x_l - a)}{\sigma} + 1 \right) \right|,$$

 $\mathbf{x}_{j:n} = (x_{(j)}, x_{(j+1)}, \dots, x_{(n)})$ , and B < n. Note that B forces a certain number of values in the tail to be fit with a GPD. Again, the  $(a + \sigma)^{-n}$  portion is a product of using the U(0, a) distribution below the threshold and down-weights the generalized fiducial density for large threshold values.

### 2.3.2 Confidence intervals, coverage, and point estimates when a is known

#### Confidence intervals and coverage

Based on the fiducial densities in equations (2.4) and (2.6) we defined point estimates and confidence intervals for the parameters. The point estimates are defined as the median of the marginal distributions of (2.4) and (2.6). We constructed the equal tailed confidence region for the true parameters ( $\gamma_0, \sigma_0$ ) and one and two sided intervals for the true high quantile,  $q_{\beta_0}$ . First, one sided lower and upper tailed intervals for  $q_{\beta_0}$  are defined as ( $c_1, \infty$ ) and ( $0, c_2$ ) respectively. The values  $c_1$  and  $c_2$  are the  $\alpha$  and  $1 - \alpha$  quantiles of (2.6). Two tailed intervals were calculated in two different ways. A symmetric ( $1 - \alpha$ ) 100% interval is obtained by combining two  $1 - \frac{\alpha}{2}$  one tailed intervals to get ( $c_1, c_2$ ). The second two tailed interval is defined as:

$$\left\{ (d_1, d_2) : \arg\min_{d_1, d_2} \left\{ (d_2 - d_1), \ \int_{d_1}^{d_2} f_{\mathcal{R}_Q}(q) dq = 1 - \alpha \right\} \right\}.$$
 (2.9)

The interval  $(c_1, c_2)$  will be referred to as the "fiducial symmetric interval" and the interval in (2.9) will be referred to as the "fiducial shortest interval". Likewise, we define the equal tailed joint confidence region for the true parameters  $(\gamma_0, \sigma_0)$  as:

$$C(\mathbf{X}) = \left\{ (\gamma, \sigma) : A_{1} = \int_{0}^{\infty} \int_{\frac{-\sigma}{x(n)-a}}^{d_{1}} f_{\mathcal{R}_{\xi}}(\gamma, \sigma) d\gamma d\sigma, \right. \\ A_{2} = \int_{0}^{d_{2}} \int_{\frac{-\sigma}{x(n)-a}}^{\infty} f_{\mathcal{R}_{\xi}}(\gamma, \sigma) d\gamma d\sigma, A_{3} = \int_{0}^{\infty} \int_{d_{3}}^{\infty} f_{\mathcal{R}_{\xi}}(\gamma, \sigma) d\gamma d\sigma, \\ A_{4} = \int_{d_{4}}^{\infty} \int_{\frac{-\sigma}{x(n)-a}}^{\infty} f_{\mathcal{R}_{\xi}}(\gamma, \sigma) d\gamma d\sigma, \text{ where } d_{1}, d_{2}, d_{3}, d_{4} \text{ satisfy} \\ A_{1} = A_{2} = A_{3} = A_{4}, \int_{d_{2}}^{d_{4}} \int_{d_{1}}^{d_{3}} f_{\mathcal{R}_{\xi}}(\gamma, \sigma) d\gamma d\sigma = 1 - \alpha \right\}.$$
(2.10)



Figure 2.1: Equal tailed confidence region for  $(\gamma_0, \sigma_0)$ 

See Figure 2.1 for illustration.

The traditional way to assess the coverage of confidence intervals is to choose a desired confidence level (e.g. 95%), simulate data, then check the frequency in which the true parameter(s) is/are captured by the constructed interval(s). Alternatively, we use a graphic device demonstrated in Hannig (2009b) which allows us to check the coverage at all confidence levels simultaneously. To accomplish this for the one tailed intervals for the  $\beta$ -quantile set  $Q_{\beta}(\mathbf{X}, q_{\beta_0}) = P(\mathcal{R}_q < q_{\beta_0} | \mathbf{X})$ . This is essentially the smallest coverage level of an upper tailed confidence interval that will contain the true quantile,  $q_{\beta_0}$ . If the confidence interval for  $q_{\beta_0}$  were exact at all confidence levels then  $Q_{\beta}(\mathbf{X}, q_{\beta_0})$  (which can be thought of as a p-value) would follow the U(0, 1) distribution.

We generated 1000 data sets from a generalized Pareto distribution with the parameter values seen in table 2.1. An MCMC algorithm was used to draw a sample from (2.4). Each generated data set produced one  $Q_{\beta}(\mathbf{X}, q_{\beta_0})$  value which we used to construct U(0, 1) QQ-plots. To assess the coverage look at the nominal p-value (desired coverage) and then note the corresponding actual p-value (actual coverage) that coincides with the simulated line. For example, following the dotted line in the

$\gamma_0$	-0.2	0	0.2	0.4	0.6
$\sigma_0$	1				
n	20	50	200		

Table 2.1: Simulation values when a is known.



Figure 2.2: QQ-plots when  $\gamma_0 = -0.2$ ,  $\sigma_0 = 1$ , and n = 50

first plot of Figure 2.2 reflects that the 0.95 upper tailed interval has actual coverage of 0.962. The dotted line in the second plot demonstrates that our method has 0.052 probability in the lower tail or, equivalently, the 0.95 lower tailed interval has an actual coverage rate of 0.948. Figures 2.3 and 2.4 are additional QQ-plots for specified parameter values. In general, intervals with exact coverage would follow the diagonal line in the plots. Because there is variation due to simulation we provide confidence bands (dashed lines). If the p-values (simulated line) stay within the dashed lines (95% confidence bands) then they cannot be distinguished from a sample of the U(0, 1) distribution and we claim good coverage properties.

A similar calculation can be done with the two tailed interval  $(c_1, c_2)$ , the interval in (2.9), and the joint confidence region in (2.10). The coverage for the intervals of the 0.99-quantile and the joint confidence region are seen in Figures 2.2, 2.3, and 2.4. The figures with the titles "Upper tailed", "Lower tailed", and "Symmetric" coincide with the previously described intervals of  $(0, c_2)$ ,  $(c_1, \infty)$ , and  $(c_1, c_2)$  respectively. The figures titled "Shortest" and "C(X)" coincide with the intervals defined in (2.9) and (2.10) respectively. As the plots illustrate all of the intervals are very close to achieving the nominal coverage rate. This behavior was also seen for different values of  $\gamma_0$  and sample sizes.



Figure 2.3: QQ-plots when  $\gamma_0 = 0$  and  $\sigma_0 = 1$ , and n = 50



Figure 2.4: QQ-plots when  $\gamma_0 = 0.4$  and  $\sigma_0 = 1$ , and n = 50

## Asymptotic properties

In this case, with a known threshold, we were able to prove that the intervals described previously have asymptotically correct coverage.

Using theorem 5.1 from Hannig (2009a) we could easily verify the conditions and show that the confidence region in (2.10) has asymptotically correct coverage. Similarly, because the  $\beta$ -quantile is a differentiable function of  $\gamma$  and  $\sigma$  it follows directly from the delta method that the intervals for the  $\beta$ -quantile are also asymptotically correct.

### Interval comparisons

Our assessment of the confidence intervals for the  $\beta$ -quantile also involved a comparison to intervals constructed using the profile log-likelihood described in Coles (2001) and the Bayesian approach using Jefferys prior described in Castellanos and Cabras (2005). The coverage for two tailed 95 and 99% intervals for the 0.99-quantile



Figure 2.5: Coverage for the 0.99-quantile with various  $\gamma_0$  values and  $\sigma_0 = 1$ .



Figure 2.6: Length of the two tailed fiducial, Bayesian, and profile log-likelihood intervals for the 0.99-quantile when  $\gamma_0 > 0$ .

are seen in Figure 2.5. Due to convergence problems the profile log-likelihood could not always be calculated for the small sample sizes. As the plots illustrate, the fiducial intervals are very close to the desired coverage rate while the Bayesian and profile log-likelihood methods tend to be slightly liberal. Figures 2.6 and 2.7 demonstrate the lengths of the intervals when  $\gamma_0 > 0$  and  $\gamma_0 \leq 0$  respectively. When  $\gamma_0 > 0$ the mean (denoted by the triangle) and the median length of the fiducial shortest interval described in (2.9) is less than its competitors for all sample sizes. When  $\gamma_0 \leq 0$  the mean and median length of the fiducial intervals are slightly longer than the competitors when n = 20 and 50. The fiducial shortest interval is shorter than its competitors when n = 200.



Figure 2.7: Length of the two tailed fiducial, Bayesian, and profile log-likelihood intervals for the 0.99-quantile when  $\gamma_0 \leq 0$ .



Figure 2.8: Absolute bias and standard deviation of the point estimates.

# Point estimates comparisons

Comparisons of the point estimates of our method, the Bayesian estimates, the MLE estimates, and estimates based on L-moments were also performed. The results were similar for all  $\gamma_0$  values so, to save space, we report these comparisons in Figure 2.8. Again, the MLE estimates at small sample sizes occasionally had convergence

Exp(1)	$t(5){+}10$
Exp(5)	$t(10){+}10$
Gamma(5,1)	N(10, 100)

Table 2.2: Distributions of the simulated data

issues so we limit the MLE comparisons to the sample sizes of n = 50 and 200. The fiducial estimates for  $\gamma_0$  and  $\sigma_0$  have smaller bias than all of the competitors. The Bayesian estimate for the 0.99-quantile is slightly less biased. All of the methods have similar variability amongst the estimates.

#### 2.3.3 Confidence intervals and coverage when a is unknown

### Confidence intervals and coverage

A similar analysis can be done when the threshold, a, is unknown. We have derived the generalized fiducial density when a is treated as an unknown parameter in (2.8). We assume the data came from a mixture of the U(0, a) and the GPD beyond a. As a result, all computations were done on the transformed data set of  $X' = X - X_{(1)}$ where  $X_{(1)}$  is the minimum and then back transformed to the original scale. Using the transformation in (2.5) we can calculate the generalized fiducial density for the  $\beta$ -quantile of the X' data set. A Metropolis-Hastings algorithm allowed us to draw a sample from (2.8) and calculate intervals for the parameters in the same manner that was discussed earlier.

To assess the usefulness of our method we applied it to general data sets. We generated 1000 data sets from the distributions listed in table 2.2 with a sample size of 1000 and assessed the confidence intervals for the  $\beta$ -quantile. Figures 2.9,2.10, and 2.11 reflect the coverage for the 0.99 and 0.999 quantiles when the underlying distributions of the data are Exp(1), t(5) + 10, and N(10, 100). Our intervals are very close to achieving the nominal coverage rate in those scenarios with the exception of the two tailed intervals for the 0.99-quantile in Figure 2.10. Those intervals are slightly liberal. Similar results were seen when the data was generated from the other distributions.



Figure 2.9: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim Exp(1)$  using the fiducial method.



Figure 2.10: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim t(5) + 10$  using the fiducial method.



Figure 2.11: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim N(10, 100)$  using the fiducial method.

#### Comparisons

In addition to checking the coverage we also compared our method to a similar Bayesian method described in Cabras and Castellanos (2009). This method used a mixture of a normal, Weibull, or nonparametric model below the threshold and a GPD above the threshold. Obviously, this method depends on which central model is chosen. We found that the nonparametric model, often times, chose a threshold far into the tail and estimated the extreme quantiles using the prescribed nonparametric distribution. As a result, there was very little variation in the MCMC chain which caused very liberal confidence intervals that rarely contained the true quantile. When the data came from a t(5)+10 the nonparametric method does a better job but is still rather liberal in its two tailed intervals, seen in Figure 2.12. Using a normal central model worked well when the data was normal and when it was very different from a normal distribution, seen in Figures 2.13 and 2.14. When the data came from a t distribution the normal central model attempted to fit the bulk of the data and forced the threshold into the tail. This caused an underestimation of the quantiles and produced very liberal upper tailed and symmetric intervals, seen in Figure 2.15. The Weibull central model produced reasonable coverage when the data came from an exponential distribution. The other cases were not appropriate for a Weibull central model.



Figure 2.12: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim t(5) + 10$  using the Bayesian method with a nonparametric central model.



Figure 2.13: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim Exp(1)$  using the Bayesian method with a normal central model.



Figure 2.14: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim N(10, 100)$  using the Bayesian method with a normal central model.



Figure 2.15: QQ-plots for the 0.99 and 0.999 quantile when  $X \sim t(10) + 10$  using the Bayesian method with a normal central model.



Figure 2.16: Length of the two tailed fiducial and Bayesian intervals for the 0.99quantile when the data was generated from Exp(1), Normal(10, 100), and t(5) + 10 respectively.

Figure 2.16 illustrates the lengths of the two tailed intervals when the methods had reasonable coverage rates. The first plot reflects the Bayesian approach with a Weibull central model is the shortest. The fiducial shortest interval is only slightly longer in this case when the data came from an Exp(1). In the second plot the data was generated from a N(10, 100) and, as expected, the Bayesian method that fits a normal model is shorter than both of our fiducial methods. In the third plot the data came from a t(5) + 10 and our fiducial shortest interval was the best. The fiducial symmetric interval was only slightly longer than the Bayesian method using a nonparametric central model.

### 2.4 Nasdaq 100 data set

In the extreme value literature a popular data set is the log-weekly losses of the Nasdaq 100 index. The data consists of 1222 weeks from October 1985 to March 2009.

Using the ad hoc approaches suggested in Coles (2001) a fixed threshold ranging from 0.03 to 0.05 would be appropriate. The estimates and confidence intervals for the 0.99-quantile can be seen in table 2.3. The estimates for the 0.99-quantile are all very similar and the fiducial shortest interval is equivalent or shorter than the Bayesian interval and the profile log-likelihood interval. Considering that the Bayesian and
Method	a (fixed)	$q_{0.99}$ estimate	q <sub>0.99</sub> 95% CI
	0.03	0.109	(0.096, 0.128)
Fiducial symmetric interval	0.04	0.109	(0.097, 0.128)
	0.05	0.107	(0.095, 0.125)
	0.03	0.109	(0.095, 0.126)
Fiducial shortest interval	0.04	0.109	(0.096, 0.125)
	0.05	0.107	(0.094, 0.123)
	0.03	0.109	(0.097, 0.128)
Bayesian interval	0.04	0.110	(0.097, 0.129)
	0.05	0.108	(0.095, 0.125)
	0.03	0.108	(0.096, 0.126)
Profile log-likelihood interval	0.04	0.109	(0.097, 0.126)
	0.05	0.106	(0.094, 0.124)
Method	a (estimate)	$q_{0.99}$ estimate	q <sub>0.99</sub> 95% CI
Fiducial symmetric interval	0.038	0.110	(0.096, 0.129)
Fiducial shortest interval	0.038	0.110	(0.095, 0.126)
Bayesian normal central model	0.016	0.110	(0.099, 0.129)

Table 2.3: Estimates and confidence intervals for the 0.99-quantile of the Nasdaq 100 data set.

profile log-likelihood methods were both slightly liberal the fiducial intervals are much more attractive.

When the threshold is unknown the Bayesian method using the normal central model was slightly shorter than the fiducial intervals. As demonstrated in the simulations, when the data came from a normal distribution the Bayesian method with a normal central model produced the shortest intervals. However, as normal QQplots would suggest, it is not likely that this data came from a normal distribution. As a result, the coverage of the Bayesian interval with a normal central model may be called into question. The fiducial shortest interval was equal in length to the Bayesian method that used a nonparametric central model and the symmetric interval was slightly longer. When the threshold was unknown the fiducial method still produced confidence intervals for the  $\beta$ -quantile with good coverage. Thus, we can be confident that the reported intervals have good empirical properties.

## 2.5 Conclusion

There has been substantial interest in peaks over threshold modeling. Various estimation techniques have been developed and methods continue to improve. The challenge in modeling peaks over threshold data come from estimating the extreme quantiles and also estimating the threshold. We developed a fiducial approach to both problems.

When the threshold is assumed to be known our fiducial approach constructs intervals and point estimates for the true parameters ( $\gamma_0, \sigma_0$ ) and the  $\beta$ -quantile that have good small sample properties. Furthermore, all of the proposed intervals have asymptotically correct coverage. Comparisons with a Bayesian method and the profile log-likelihood approach suggest that the fiducial intervals were closer to achieving the nominal coverage rate for the 0.99-quantile and one of the fiducial intervals was shortest. Likewise, the point estimates for the shape and scale parameters using the fiducial method had the smallest bias. The point estimate for the 0.99-quantile was slightly better using the Bayesian approach.

When the threshold is unknown the proposed fiducial intervals for the  $\beta$ -quantile had good frequentist coverage. Our method worked well on all different data types that could be seen in real life. The competing Bayesian method did not universally work well for all data types. First, a central model had to be chosen from either a Weibull, normal, or nonparametric distribution. Even after choosing an adequate central model it was not assured that the coverage for the  $\beta$ -quantiles would be reasonably close to exact. When the Bayesian intervals had a reasonable coverage rate the recommended fiducial intervals were either close to the same length or shorter. We analyzed a data set from the log-weekly losses of the Nasdaq 100 index. Our analysis demonstrated that the intervals for the fiducial method were often times shorter than the competing methods in both cases where the threshold was known and unknown. Because we demonstrated that the coverage for the fiducial intervals were reasonably close to exact for any data type we can be confident that these intervals have adequate coverage.

Based on our findings the fiducial approach to the generalized Pareto distribution is a very viable alternative to modeling peaks over threshold data sets. The good small sample properties and the asymptotic results make this an attractive solution to a difficult problem. R code for our procedure can be downloaded at http://www.unc.edu/~hannig/.

# Chapter 3

# FIDUCIAL INFERENCE ON THE LARGEST MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

## 3.1 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  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$  and  $\boldsymbol{\Sigma}$  is an unstructured covariance matrix we are attempting inference on the largest mean,  $\boldsymbol{\theta} = \max_i \mu_i$ .

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 Hannig (2009b) and Hannig and Lee (2009). 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 allows 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 the largest mean. Another technique, seen in Eaton et al. (2006), 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, Boose et al. (2007) 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, Eaton et al. (2006), and the bias adjusted method using a normal approximation from Boose et al. (2007). All of the methods from Boose et al. (2007) 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., Dawid (1994). 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. (2006) and, when the sample size is small, slightly longer than the interval from Boose et al. (2007). 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.

#### 3.2 Generalized Fiducial Inference

Fiducial inference was first introduced by Fisher (1930). 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 (Lindley, 1958; Zabell, 1992). Recently, there has been somewhat of a resurgence in fiducial inference following the introduction of generalized inference by Weeranhandi (1993) and the establishment of a link between fiducial and generalized inference in Hannig et al. (2006). Further information on the asymptotic and empirical properties and a thorough survey of the fiducial literature can be found in Hannig (2009b).

The basis of generalized fiducial inference, similar to the likelihood function, "switches" the role of the data,  $\mathbf{X}$ , and the parameter(s)  $\xi$ . Fiducial inference uses the model and the observed data,  $\mathbf{X}$ , to define a probability measure on the parameter space,  $\Xi$ . 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 X and  $\xi$  exists in the form of,

$$\mathbf{X} = G(\xi, \mathbf{U}) \tag{3.1}$$

where **U** is a random vector with a completely known distribution and independent of any parameters. With this relationship, called the *structural equation*, the parameter  $\xi$  and the random vector **U** will determine the distribution of **X**. After observing **X** we can use the relationship in (4.1) and what we know of the distribution of **U** to infer a distribution on  $\xi$ .

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

$$Q(\mathbf{x}, \mathbf{u}) = \{\xi : \mathbf{x} = G(\xi, \mathbf{u})\}.$$
(3.2)

We know that our observed data was generated using some unknown  $\xi_0$  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:

$$V(Q(\mathbf{x}, \mathbf{U}^{\star})) \mid \{Q(\mathbf{x}, \mathbf{U}^{\star}) \neq \emptyset\}$$
(3.3)

where  $\mathbf{U}^*$  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 a (possibly random) rule for discerning among the values of the inverse  $Q(\mathbf{x}, \mathbf{U}^*)$ . We will refer to a random element with the distribution given by (4.3) as  $\mathcal{R}_{\xi}$ . For a detailed discussion of the derivation of the generalized fiducial distribution see Hannig (2009b).

We calculate the generalized fiducial density as proposed in Hannig (2009b) and justified theoretically in Hannig (2009a). Let  $G = (g_1, \ldots, g_n)$  be such that  $X_i = g_i(\xi, \mathbf{U})$  for  $i = 1, \ldots, n$ . Note that  $\xi$  is a  $p \times 1$  vector and denote  $\mathbf{X_i} = G_{0,i}(\xi, \mathbf{U_i})$ 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  $\mathbf{i} = (i_1, \dots, i_p)$ . 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

$$f_{\mathcal{R}_{\xi}}(\xi) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')d\xi'}$$
(3.4)

where

$$J(\mathbf{x},\xi) = \binom{n}{p}^{-1} \sum_{\mathbf{i}=(i_1,\dots,i_p)} \left| \frac{\det\left(\frac{d}{d\xi} \mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)}{\det\left(\frac{d}{d\mathbf{u}_{\mathbf{i}}} \mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)} \right|$$
(3.5)

is the mean of all subsets where  $1 \leq i_1 < \cdots < i_p \leq n$  and the determinants in (4.5) are the appropriate Jacobians. For illustrative examples of this fiducial recipe see Hannig (2009b).

### 3.3 Main Results

#### 3.3.1 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  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  distribution where  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$  and

$$\Sigma = \begin{bmatrix} \eta_1 & \rho_{12}\sqrt{\eta_1\eta_2} \\ \rho_{12}\sqrt{\eta_1\eta_2} & \eta_2 \end{bmatrix}.$$
 (3.6)

The structural equation in (4.1) 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  $\Sigma$ , and  $\xi = (\mu_1, \mu_2, \eta_1, \eta_2, \rho_{12})^T$ . The inverted structural equations are:

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

$oldsymbol{\mu}_0ackslash n$	30	100	1000
$(5,5)^T$	99.6	99.9	99.8
$(1,5)^T$	92.3	94.9	95.3

Table 3.1: Coverage for a 95% upper tailed interval when  $\Sigma = I$  for the naive method.

for any individual j = 1, ..., n. Since there are five parameters we need the same number of equations to define each of the terms in (4.5). 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,

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

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 of:

$$f_{1}(\xi) \propto \frac{1}{(2\pi)^{kn/2}} \left(\det \Sigma\right)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \left(\mathbf{x}_{i} - \boldsymbol{\mu}\right)^{T} \Sigma^{-1} \left(\mathbf{x}_{i} - \boldsymbol{\mu}\right)\right\} \\ \left(\binom{n}{2, 1, n-3}^{-1} \sum_{\mathbf{i}} \frac{|g(\mathbf{x}_{\mathbf{i}})|}{2^{2} \eta_{1}^{3/2} \eta_{2}^{1/2} (1 - \rho_{12}^{2})}\right)$$
(3.7)

where  $g(\mathbf{x_i})$  is a function of the data and will be explained later. From (3.7) we could draw a fiducial random sample of  $\theta = \max(\mu_1, \mu_2)$  by generating a sample of  $\mu$  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,  $\theta_0$ , when the true means,  $\mu_0^{(1)}$  and  $\mu_0^{(2)}$ , are equal or relatively close to equal. For example, Table 3.1 shows the coverage for a 95% upper tailed interval of  $\theta_0$ . 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 a  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  distribution where  $\boldsymbol{\mu} = (\mu, \mu)^T$  our inverted structural equations change to:

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

$\mu_0 \setminus n$	30	100	1000
$(5,5)^T$	94.4	94.4	94.9
$(1,5)^T$	0.8	0	0

Table 3.2: Coverage for a 95% upper tailed interval when  $\Sigma = I$  when the  $\mu_1 = \mu_2$  is assumed.

We now have four parameters so only four equations are necessary to define each of the terms in (4.5). The generalized fiducial density for this case is then computed as:

$$f_{2}(\xi) \propto \frac{1}{(2\pi)^{kn/2}} \left(\det \Sigma\right)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \left(\mathbf{x}_{i} - \boldsymbol{\mu}\right)^{T} \Sigma^{-1} \left(\mathbf{x}_{i} - \boldsymbol{\mu}\right)\right\} \\ \left(\frac{n}{2, n-2}\right)^{-1} \sum_{\mathbf{i}} \frac{|g(\mathbf{x}_{\mathbf{i}}, \boldsymbol{\mu})|}{2^{2} \eta_{1}^{3/2} \eta_{2}^{1/2} (1 - \rho_{12}^{2})}.$$
(3.8)

Comparing (3.7) and (3.8) we can see that the difference comes in the Jacobian calculation and the  $\mu$  vector. The Jacobian in (3.8) depends on  $\mu$ , the equal mean, and (3.7) is only dependent on the data. This behavior is common to all fiducial densities that have an equal mean. Table 3.2 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 (3.7) when the means are different and (3.8) 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 is occurring when the true means are equal and drives the asymptotic correctness of the intervals.

The largest mean could come from  $\mu_1$  or  $\mu_2$  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:

$$\mathbf{X} = \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \mathbf{V}_1 \mathbf{U} \right) I_{\{\mu_1 > \mu_2\}} + \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \mathbf{V}_2 \mathbf{U} \right) I_{\{\mu_2 > \mu_1\}}$$
(3.9)  
+ 
$$\left( \begin{pmatrix} \mu \\ \mu \end{pmatrix} + \mathbf{V}_1 \mathbf{U} \right) I_{\{\mu_1 = \mu_2 = \mu\}}$$

where  $\mathbf{V}_1$  is the lower triangle Cholesky decomposition of  $\Sigma$  and  $\mathbf{V}_2$  is obtained by permuting the  $\Sigma$  matrix in (3.6) so that  $\mathbf{X} = (X_2, X_1)^T$ , taking the lower triangular Cholesky decomposition and re-permuted 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 (3.9) 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 Hannig and Lee (2009). 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:

$$\frac{1}{2} \log \left( \frac{2^2}{(MSX_1 + MSX_2 - 2MX_1X_2)n} \right) = \beta_i + P_i \quad \text{if } i \in J$$
  
$$\frac{1}{2} \log \left( \frac{2^2}{(MSX_1 + MSX_2 - 2MX_1X_2)n} \right) = P_i \quad \text{if } i \in J^c \quad (3.10)$$

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 **X** but they will affect the conditional distribution in (4.3). When inverting the structural equations in (3.10), if  $i \in J$  we can choose a  $\beta_i$  for any  $P_i$  so that the equation is satisfied. Therefore, conditioning on this equation does not effect 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 (3.9) there will now always be 2n - 4 conditions regardless of 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

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

with  $f_1(\xi)$  given by (3.7),  $\tilde{f}_1(\xi)$  by (3.7) with  $\eta_1$  and  $\eta_2$  switched,  $f_2(\xi)$  by (3.8), and the weight function

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

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

## 3.3.2 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,  $u \ (u \ge 1)$  is the number of elements in J, and  $v \ (v \ge 0)$  is the number of elements in  $J^c$  (note, u + v = k). In general, for a k dimensional problem the structural equation is:

$$\mathbf{X} = \sum_{J \subseteq \{1,\dots,k\}} \left( \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} + \mathbf{V}_J \mathbf{U} \right) I_{\{\mu_i = \mu_j : , i, j \in J\}} I_{\{\mu_j > \mu_l : j \in J, l \notin J\}}$$
(3.11)

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  $\Sigma$  then re-permute  $\mathbf{V}_J$  back to the original order.

Just as before we introduce the additional structural equations akin to (3.10) 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 I. This defines the generalized fiducial density as:

$$f(\xi) = \frac{\sum_{J \subseteq \{1,...,k\}} f_J(\xi)}{\int_{\Xi} \sum_{J \subseteq \{1,...,k\}} f_J(\xi') d\xi'}$$
(3.12)

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

$$f_{J}(\xi) \propto \frac{w_{J}(\mathbf{X})}{(2\pi)^{kn/2}} (\det \Sigma)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{X}_{i} - \boldsymbol{\mu})^{T} \Sigma^{-1} (\mathbf{X}_{i} - \boldsymbol{\mu})\right\} \\ \left(\binom{n}{C_{j,n}}^{-1} \sum_{i_{1},...,i_{k}} \frac{|g(\mathbf{X}_{i},\boldsymbol{\mu})|}{\det \Sigma \prod_{j=1}^{k-1} \frac{\det \Sigma_{1,j}}{\eta_{j}^{*}}} I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}} I_{\{\mu_{i} = \mu_{j}:, i, j \in J\}} \\ = \frac{w_{J}(\mathbf{X}) J_{x}}{(2\pi)^{kn/2}} \frac{(\det \Sigma)^{-n/2-1}}{\prod_{j=1}^{k-1} \frac{\det \Sigma_{1,j}}{\eta_{j}^{*}}} \exp\left\{-\frac{1}{2} tr\left(S\Sigma^{-1}\right)\right\} \\ \times I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}} I_{\{\mu_{i} = \mu_{j}:, i, j \in J\}}$$
(3.13)

where

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

and the rest of the terms are defined in appendix I.

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,

$$P(J) = \frac{p_J}{\sum_{\tilde{J} \subseteq \{1,...,k\}} p_{\tilde{J}}}$$
(3.14)

where  $p_J = \int_{\Xi} f_J(\xi) d\xi$ . 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.

#### 3.3.3 Confidence intervals and coverage

Using the fiducial density in (3.13) 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  $(c_1, \infty)$  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 equation (3.13) 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 III to draw a sample  $\{\theta'_1, \ldots, \theta'_E\}$  from (3.12) 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 Hannig (2009b). 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 true value,  $\theta_0$ .

If the confidence interval for  $\theta_0$  were exact at all confidence levels then  $C(\mathbf{X}, \theta_0)$ would follow the U(0,1) distribution. Using QQ-plots we can evaluate how closely  $C(\mathbf{X}, \theta_0)$  follows the U(0,1) distribution. We plot the nominal p-values (desired coverage probability) vs. actual p-values (actual coverage probability). Figure 3.1 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 pvalue and note the corresponding actual p-value that coincides with the simulated line. For example, the dotted line in the first plot in Figure 3.1 shows that a 0.95



Figure 3.1: QQ-plots when n = 30,  $\mu_0 = (5, 5, 5)^T$ ,  $\Sigma_0 = I$  and 25*I* for the top and bottom rows respectively.

upper tailed interval has actual coverage of 0.987. The dotted line in 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 10000. We will highlight a few difficult and interesting cases before discussing the full simulation study. Figure 3.1 illustrates the generated QQ-plots when n = 30,  $\mu_0 = (5, 5, 5)^T$ ,  $\Sigma_0 = I$  and 25*I* 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 Figure 3.1. 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  $\mu_0$  and  $\Sigma_0$  configurations as well.



Figure 3.2: QQ-plots when n = 30,  $\mu_0 = (3, 4, 5)^T$ , and  $\Sigma_0 = 25I$ .



Figure 3.3: QQ-plots when n = 30,  $\mu_0 = (1, 3, 5)^T$ , and  $\Sigma_0 = I$ .

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 Figure 3.2 when n = 30,  $\mu_0 = (3, 4, 5)^T$ , and  $\Sigma_0 = 25I$ .

Cases where the largest mean is much different than the other means results in confidence intervals that are, expectantly, close to exact. This is reflected in Figure 3.3 when n = 30,  $\mu_0 = (1, 3, 5)^T$ , and  $\Sigma_0 = I$ .



Figure 3.4: QQ-plots when n = 10000,  $\mu_0 = (3, 4, 5)^T$ , and  $\Sigma_0 = 25I$ .

$oldsymbol{\mu}_0$	$(2,5)^T$	$(4,5)^T$	$(5,5)^T$
$oldsymbol{\eta}_0$	(1, 1)	(25, 25)	
$ ho_{0}^{(1,2)}$	0	$\pm 0.4$	$\pm 0.9$
n	30	100	

Table 3.3: Simulation combinations in two dimensions

When the sample size is dramatically increased to n = 10000 in the case where  $\mu_0 = (3, 4, 5)^T$  and  $\Sigma_0 = 25I$  the QQ-plots in Figure 3.4 reflect close to exact coverage. This motivates our later discussion on the asymptotics of our intervals.

## 3.3.4 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 3.3 and 3.4. Each  $\rho_0^{(ij)}$  in the last configurations were randomly generated values from the U(0,1) distribution where the resulting  $\Sigma_0$  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), Eaton et al. (2006) (Eaton), and the bias adjusted interval based on a normal approximation from Boose et al. (2007) (Boos). As previously noted, Boose et al. (2007) 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 Boose et al. (2007).

The coverage for the 95 and 99% intervals when the data is two dimensional can be seen in Figures 3.5 and 3.7. The coverage when the covariance matrix was randomly generated is seen in Figure 3.9. 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

$\mu_0$	$(1, 3, 5)^T$	$(3,4,5)^T$	$(1,5,5)^T$	
	$(4, 5, 5)^T$	$(5,5,5)^T$		
$oldsymbol{\eta}_0$	(1,1,1)	(25, 25, 25)		
$\rho_0^{(i,j)}$	0	$\pm 0.4$	0.9	
	$\begin{bmatrix} 1 \\ 0.051 & 1 \\ -0.392 & 0.754 & 1 \end{bmatrix}$	$\begin{bmatrix} 1\\ 0.104 & 1\\ -0.528 & -0.710 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -0.133 & 1 \\ 0.844 & -0.154 & 1 \end{bmatrix}$	
n	30	100		

Table 3.4: Simulation combinations in three dimensions



Figure 3.5: Coverage when the data is two dimensional and  $\rho^{(1,2)} > 0$ .



Figure 3.6: Length of the upper tailed intervals relative to the fiducial interval when data is two dimensional and  $\rho^{(1,2)} > 0$ .



Figure 3.7: Coverage when the data is two dimensional and  $\rho^{(1,2)} \leq 0$ .



Figure 3.8: Length of the upper tailed intervals relative to the fiducial interval when data is two dimensional and  $\rho^{(1,2)} \leq 0$ .

close to the nominal coverage rate for the fiducial method. The very liberal outliers, seen in Figure 3.7, occurs 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 positive correlation. Similar behavior also occurred with the three dimensional data. Additional plots are available from the authors.

Figures 3.6, 3.8, and 3.10 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



Figure 3.9: Coverage when the data is three dimensional and  $\rho^{(i,j)}$  is random.



Figure 3.10: Length of the upper tailed intervals relative to the fiducial interval when data is three dimensional and  $\rho^{(i,j)}$  is random.

shortest. Additionally, the Boos interval assumes that the data is equicorrelated and equivariant where we allow for a totally unstructured covariance matrix.

# 3.4 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.

**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 (4.8). 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  $\mu_0$  is the true mean and J correctly identifies the equal largest means then  $P(J) \to 1$  in probability.

We could not integrate (3.13) 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 II.

**Theorem 2.** The confidence intervals for  $\theta_0 = \max_{1 \le i \le k} \mu_0^{(i)}$  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 Hannig (2009b) to prove that the confidence intervals are asymptotically correct.  $\Box$ 

#### 3.5 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 3.5 breaks down the air quality for different AQI values. Overall, an AQI value of 100 or less is the standard in which the EPA has determined as satisfactory.

AQI	Air Quality
0 to 50	Good
51 to 100	Moderate
101 to 150	Unhealthy for Sensitive Groups
151 to 200	Unhealthy
201 to 300	Very Unhealthy
301 to 500	Hazardous

Table 3.5: AQI range of health effects

We obtained monthly AQI measurements from January 1, 1998 to October 1, 2008 from the EPA (2009) 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:

$$\bar{\mathbf{x}}' = \begin{bmatrix} 3.959\\ 3.845\\ 3.928\\ 3.963 \end{bmatrix} \text{ and } \hat{\Sigma}' = \begin{bmatrix} 0.251 & 0.123 & 0.161 & 0.194\\ 0.123 & 0.161 & 0.132 & 0.137\\ 0.161 & 0.132 & 0.193 & 0.172\\ 0.194 & 0.137 & 0.172 & 0.230 \end{bmatrix}.$$
(3.15)

If we were to assume that those were the true values for  $\mu_0$  and  $\Sigma_0$  the coverage for  $\theta_0$  is shown in Figure 3.11. 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 Figure 3.12. Because the means are relatively close together two of the methods tend to overestimate the maximum and produce conservative intervals.



Figure 3.11: Fiducial method QQ-plots using  $\mu_0$  and  $\Sigma_0$  from (3.15).



Figure 3.12: Competing methods QQ-plots using  $\mu_0$  and  $\Sigma_0$  from (3.15).

At the 99% level the fiducial method has an approximate coverage rate of 98.3% (based on the Figure 3.11) 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 3.6 illustrates these fiducial probabilities for any J when P(J) > 0 (note, 1=Baltimore, 2=Boston, 3=New York, and

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

Table 3.6: Largest mean probabilities for each model, J.

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.

## 3.6 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.

#### Appendix I

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  $\Sigma$  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  $g(\mathbf{X}_i,\mu) = \frac{1}{2^k} \det(\mathbb{X}_2) \prod_{j=3}^{u+1} |\mathbb{X}_j^*| \prod_{j=u+2}^{k+1} |\mathbb{X}_j|$ ,  $\mathbf{i} = (i_1,\ldots,i_{k+1})$ ,

$$\mathbb{X}_{j} = \begin{bmatrix} 1 & X_{i_{1},1} & \cdots & X_{i_{1},j-1} \\ 1 & X_{i_{2},1} & \cdots & X_{i_{2},j-1} \\ \vdots & & & \\ 1 & X_{i_{j},1} & \cdots & X_{i_{j},j-1} \end{bmatrix}, \ \mathbb{X}_{j}^{*} = \begin{bmatrix} 1 & \mu & \cdots & \mu \\ 1 & X_{i_{1},1} & \cdots & X_{i_{2},j-1} \\ \vdots & & \\ 1 & X_{i_{j-1},1} & \cdots & X_{i_{j},j-1} \end{bmatrix}$$
$$C_{u,n} = \begin{cases} 2, 1, \dots, 1, n-k-1 & u = 1 \\ 2, 2, 1, \dots, 1, n-k-1 & u = 2, \dots, k-1, \\ 2, 1, \dots, 1, n-k & u = k \end{cases}$$

and u is the number of elements in J or the number of equal means  $(u \ge 1)$ . 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

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

$$\frac{1}{2v} \log \left( \frac{\det (M_{J_k})^{(v-1)} M_J^{*v} k^{2v}}{\det (M_J)^{(v-1)} M_{J_k}^{*v} n^v} \right) = P_i \quad \text{if } i \in J^c \quad (3.16)$$

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

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

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

# Appendix II

#### 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  $P(J) = p_J \left(\sum_{\tilde{J} \subseteq \{1,...,k\}} p_{\tilde{J}}\right)^{-1}$ . 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:

$$f_J(\xi) \leq \frac{J_x^{(\cdot)}}{b_{\rho}(2\pi)^{kn/2}} \left(\det \Sigma\right)^{-n/2-1} \exp\left\{-\frac{1}{2} tr\left(S\Sigma^{-1}\right)\right\} = f_J^{\blacktriangle}(\xi) \qquad (3.17)$$

and

$$f_{J}(\xi) \geq f_{J}^{\dagger}(\xi) I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}} I_{\{-r \leq \rho_{ij} \leq r: \forall i, j\}} I_{\{w_{1} \leq \eta_{i} \leq w_{2}: \forall i\}} = f_{J}^{\bullet}(\xi)$$
(3.18)

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,  $\left|\rho_0^{(ij)}\right| \leq r < 1 \text{ and } 0 < w_1 \leq \eta_0^{(i)} \leq w_2 < \infty \text{ for all } i \text{ and } j, b_{\rho} > 0 \text{ is a lower bound for } \prod_{j=1}^{k-1} \det \Sigma_j \left(\eta_j^*\right)^{-1}$ , 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,  $\mu$ .

If we notice that  $\Sigma^{-1}$  and  $\mu_i$  follow a Wishart and t distribution for all i the resulting integration for the upper bound is:

$$\begin{split} p_{J}^{\bullet} &= \int_{\Xi} f_{J}^{\bullet}(\xi) d\xi \\ &= \frac{1}{b_{\rho}} \frac{2^{k(k+3)/2} \pi^{k(k-2n-1)/4} \prod_{i=1}^{k} \Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2} s_{1}^{(n+k+3)/2} s_{2}^{(n+k+2)/2} \dots s_{v}^{(n+k+4-v)/2}} \\ &\times \frac{\pi^{v/2} \Gamma\left(\frac{n+k+2}{2}\right) \Gamma\left(\frac{n+k+1}{2}\right) \dots \Gamma\left(\frac{n+k+3-(k-u)}{2}\right)}{\Gamma\left(\frac{n+k+3}{2}\right) \Gamma\left(\frac{n+k+2}{2}\right) \dots \Gamma\left(\frac{n+k+4-(k-u)}{2}\right)} \int_{\mathbb{R}} \frac{J_{x}^{(\cdot)}}{(\sigma_{v}^{2} - \zeta_{v}^{2})^{(n+k+3-v)/2}} d\mu \end{split}$$

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 > 1and odd. The first case, when u = 1 (all means are different denoted  $J_1$ ) results in:

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

When u > 1 and odd,

$$p_{J_{u}}^{\bullet} = \frac{1}{b_{\rho}} \frac{2^{k(k+3)/2} \pi^{k(k-2n-1)/4} \prod_{i=1}^{k} \Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2} s_{1}^{(n+k+3)/2} s_{2}^{(n+k+2)/2} \dots s_{v}^{(n+k+4-v)/2} \tilde{s}_{u}^{(n+u+3)/2}} \frac{\pi^{(v+1)/2} \prod_{i=1}^{v} \Gamma\left(\frac{n+k+3-i}{2}\right)}{\prod_{i=1}^{v} \Gamma\left(\frac{n+k+4-i}{2}\right)} \times \sum_{i=1}^{(u+1)/2} \frac{J_{x_{i}}^{(1)}}{(\tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2})^{(n+1+2i)/2}} \frac{\Gamma\left(\frac{n+2i+1}{2}\right)}{\Gamma\left(\frac{n+2i+2}{2}\right)}.$$

When u > 1 and even,

$$p_{J_{u}}^{\bullet} = \frac{1}{b_{\rho}} \frac{2^{k(k+3)/2} \pi^{k(k-2n-1)/4} \prod_{i=1}^{k} \Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2} s_{1}^{(n+k+3)/2} s_{2}^{(n+k+2)/2} \dots s_{v}^{(n+k+4-v)/2} \tilde{s}_{u}^{(n+u+3)/2}} \frac{\pi^{(v+1)/2} \prod_{i=1}^{v} \Gamma\left(\frac{n+k+3-i}{2}\right)}{\prod_{i=1}^{v} \Gamma\left(\frac{n+k+4-i}{2}\right)} \times \sum_{i=1}^{u/2+1} \frac{J_{x_{i}}^{(2)}}{(\tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2})^{(n+2i)/2}} \frac{\Gamma\left(\frac{n+2i}{2}\right)}{\Gamma\left(\frac{n+2i+1}{2}\right)}.$$

Integrating the lower bound  $f_J^{\checkmark}(\xi)$  will result in a very similar calculation. First notice that from equation (3.18),

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

where

$$\begin{split} L_{n} &= -\int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}}\right) d\xi - \int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{-r \leq \rho_{ij} \leq r: \forall i, j\}}\right) d\xi \\ &- \int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{w_{1} \leq \eta_{i} \leq w_{2}: \forall i\}}\right) d\xi \\ &+ \int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}}\right) \left(1 - I_{\{-r \leq \rho_{ij} \leq r: \forall i, j\}}\right) d\xi \\ &+ \int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}}\right) \left(1 - I_{\{w_{1} \leq \eta_{i} \leq w_{2}: \forall i\}}\right) d\xi \\ &+ \int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{-r \leq \rho_{ij} \leq r: \forall i, j\}}\right) \left(1 - I_{\{w_{1} \leq \eta_{i} \leq w_{2}: \forall i\}}\right) d\xi \\ &- \int_{\Xi} f_{J}^{\dagger}(\xi) \left(1 - I_{\{\mu_{j} > \mu_{l}: j \in J, l \notin J\}}\right) \left(1 - I_{\{w_{1} \leq \eta_{i} \leq w_{2}: \forall i\}}\right) d\xi \end{split}$$
(3.19)

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

$$\sqrt{n}\left(\mathcal{R}_{\xi} - \xi_0\right) = \mathbf{p}.\tag{3.20}$$

where  $\mathcal{R}_{\xi}$  is a vector of the fiducial random variables for  $\xi$ ,  $\xi_0$  is a vector of the values of  $\mu_0$  and  $\Sigma_0$ , and **p** is a vector of the transformed variables. Observing the indicators from (3.19) with the transformation in (3.20) 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,

$$p_{J}^{\bullet} = \int_{\Xi} f_{J}^{\dagger}(\xi) d\xi \left( 1 + \frac{L_{n}}{\int_{\Xi} f_{J}^{\dagger}(\xi) d\xi} \right)$$
  
$$= b \frac{2^{k(k+3)/2} \pi^{k(k-2n-1)/4} \prod_{i=1}^{k} \Gamma\left(\frac{n+k+4-i}{2}\right)}{n^{k(n+k+3)/2} s_{1}^{(n+k+3)/2} s_{2}^{(n+k+2)/2} \dots s_{v}^{(n+k+4-v)/2} \tilde{s}_{u}^{(n+k+3-v)/2}} \frac{\pi^{(v+1)/2} \prod_{i=1}^{v+1} \Gamma\left(\frac{n+k+3-i}{2}\right)}{\prod_{i=1}^{v+1} \Gamma\left(\frac{n+k+4-i}{2}\right)} \times \frac{1}{(\tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2})^{(n+k+2-v)/2}} \left( 1 + \frac{L_{n}}{\int_{\Xi} f_{J}^{\dagger}(\xi) d\xi} \right)$$

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

To define the Jacobian term,  $J_x^{(\cdot)}$  a lower bound can be  $J_x > b$  for some b > 0. If u = 1, all of the means are different, then  $J_x$  is independent of  $\mu$  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,

$$J_{x} = |J_{x_{1}}\mu^{u-1} + J_{x_{2}}\mu^{u-2} + \dots + J_{x_{u}}|$$
  
$$\leq |J_{x_{1}}||\mu^{u-1}|| + |J_{x_{2}}||\mu^{u-2}|| + \dots + |J_{x_{u}}||.$$

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

$$J_{x} \leq |J_{x_{2,1}}| \mu^{u-1} + |J_{x_{2,2}}| (\mu^{u-1} + 1) + \dots + |J_{x_{2,u}}|$$
  

$$\leq J_{x_{1}}^{(1)} \left( \left( \mu - \tilde{\zeta}_{u} \right)^{2} + \tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2} \right)^{(u-1)/2} + J_{x_{2}}^{(1)} \left( \left( \mu - \tilde{\zeta}_{u} \right)^{2} + \tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2} \right)^{(u-3)/2}$$
  

$$+ \dots + J_{x_{\frac{u+1}{2}}}^{(1)} + J_{x_{\frac{u+1}{2}+1}}^{(1)} \left( \left( \mu - \tilde{\zeta}_{u} \right)^{2} + \tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2} \right)^{(u-2)/2} (\mu - \tilde{\zeta}_{u})$$
  

$$+ J_{x_{\frac{u+1}{2}+2}}^{(1)} \left( \left( \mu - \tilde{\zeta}_{u} \right)^{2} + \tilde{\sigma}_{u}^{2} - \tilde{\zeta}_{u}^{2} \right)^{(u-4)/2} (\mu - \tilde{\zeta}_{u}) + \dots + J_{x_{u}}^{(1)} (\mu - \tilde{\zeta}_{u})$$

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 ,

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

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

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

where  $\tilde{\zeta}_u$  and  $\tilde{\sigma}_u^2 - \tilde{\zeta}_u^2$  are the mean and a scale factor of  $\mu$ . 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 **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:

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

is the difference in the sample mean of the variables that share a common mean,  $\mu$ . Next,  $\bar{\mathbf{X}}_h = (\bar{X}_h, \dots, \bar{X}_k)^T$ ,  $\bar{\mathbf{X}}_{\mu,h} = (\bar{X}_h, \bar{X}_{h+1} - \mu_{h+1}, \bar{X}_{h+2} - \mu_{h+2} \dots, \bar{X}_k - \mu_k)^T$ ,  $M_{j,k} = \begin{cases} M_J & \text{for } J = \{j, \dots, k\} \\ 1 & \text{otherwise} \end{cases}$ ,  $M_{u,k}^* = \det(1 + M_{v+1,k}) - \det(M_{v+1,k})$ ,

$$s_{j} = \begin{cases} \det(M_{2,k}) & \text{for } j = 1\\ \frac{\det(M_{j+1,k})\det(M_{j-1,k})}{\det(M_{j,k})^{2}} & \text{for } j = 2, \dots, k \end{cases}, \text{ and } \widetilde{s}_{u} = \begin{cases} \frac{M_{u,k}^{*}\det(M_{v,k})}{\det(M_{v+1,k})^{2}} & \text{for } u < k\\ M_{k,k}^{*} & \text{for } u = k \end{cases}$$

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

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

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

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

where  $\mathbf{1} = (1, \ldots, 1)^T$ . The matrix  $D_h$  is the  $1^{st}$  adjugate matrix of  $M_{h,k}$  and D is the  $2^{nd}$  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 Bruijn (1956).  $C^1(A)$  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  $J = \{i : i = 1, ..., k\} = J_1$  or equivalently u = 1. The second case is when there are multiple largest means,  $J \subseteq \{1, ..., k\}$  and  $u \ge 2$ . We will show that if J correctly indexes the largest mean(s) then  $P(J) = p_J \left(\sum_{\tilde{J} \subseteq \{1,...,k\}} p_{\tilde{J}}\right)^{-1} \to 1$ or equivalent to  $p_{\tilde{J}}/p_J \to 0$  for any  $\tilde{J} \ne J$ . Before completing the proof we need to observe the following. Using Stirling's formula:

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

Also, because the matrix D is the  $2^{nd}$  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 (1896) 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  $\lambda_i$  is the  $i^{th}$  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:

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

This means that,  $\left(1 + \bar{\mathbf{X}}_{d,v}^T D \bar{\mathbf{X}}_{d,v} \left(M_{u,k}^*\right)^{-1}\right)^{(n+u+2)/2} \leq (\log n)^c$  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  $p_{\tilde{J}}/p_J \leq p_{\tilde{J}}^{*}/p_J^{\vee} \to 0$  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

$$\begin{split} \frac{p_{J_{1}}^{\bullet}}{p_{J_{u}}^{\bullet}} &= \frac{J_{x}^{(0)}}{I_{J_{u}}b_{\rho}b} \frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+k+2-v}{2}\right)} \times \frac{\pi^{k/2}\tilde{s}_{u}^{(n+k+4-(v+1))/2} (\tilde{\sigma}_{u}^{2}-\tilde{\zeta}_{u}^{2})^{(n+k+3-(v+1))/2}}{\pi^{(v+1)/2}s_{v+1}^{(n+k+3-v)/2}s_{v+2}^{(n+k+2-v)/2} \dots s_{k}^{(n+4)/2}s_{k+1}^{(n+3)/2}} \\ &= \frac{J_{x}^{(0)}}{I_{J_{u}}b_{\rho}b} \frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+k+2-v}{2}\right)} \pi^{(u-1)/2} M_{u,k}^{*1/2} \left(1 + \frac{\bar{\mathbf{X}}_{d,v}^{T}D\bar{\mathbf{X}}_{d,v}}{M_{u,k}^{*}}\right)^{(n+2+u)/2} \\ &\leq \frac{J_{x}^{(0)}}{I_{J_{u}}b_{\rho}b} \pi^{(u-1)/2} M_{u,k}^{*1/2} \frac{2^{(u+1-v)/2}}{n^{(u-1)/2}} \left(\log n\right)^{c} \quad \text{eventually a.s.} \\ &\stackrel{\mathcal{P}}{\to} 0. \end{split}$$

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,

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

In a similar fashion, we have shown that  $p_{J_u}^{\bigstar}/p_{J_1}^{\bigstar} \to 0$  in probability when u is even and  $p_{J_{u_1}}^{\bigstar}/p_{J_{u_2}}^{\bigstar} \to 0$  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  $p_{\tilde{J}}/p_J \leq p_{\tilde{J}}^{\blacktriangle}/p_J^{\blacktriangledown} \to 0$  in probability when J correctly indexes largest mean(s). Specifically, we have shown,  $P(J) = p_J \left( \sum_{\tilde{J} \subseteq \{1,...,k\}} p_{\tilde{J}} \right)^{-1} \to 1$  in probability. This completes the proof of theorem 1.

# Appendix III

## Importance Sampling Algorithm

The following steps were implemented in order to obtain a fiducial sample for  $\xi$ .

- 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  $(\mu_h | \mu_{h+1}, \dots, \mu_v, \mu) = \zeta_h + \sqrt{\frac{\sigma_h^2 \zeta_h^2}{n+k+3-h}} T_h$  where  $T_h \sim t(n+k+3-h)$  for all  $h = 1, \dots, v$ .
- 3. Generate  $(\Sigma^{-1}|\boldsymbol{\mu}) = W$  where  $W \sim Wishart(n+k+3, S^{-1})$
- 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  $ESS_J = n_J \left(1 + (s_{w_J}^2) \bar{w}_J^{-2}\right)^{-1}$  where  $n_J$  is the sample size for model J,  $s_{w_J}^2$ 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  $J \subseteq \{1, \ldots, k\}$ .

# Chapter 4

## FIDUCIAL APPROACH TO MULTIPLE COMPARISONS

# 4.1 Introduction

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

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

There are several frequentist solutions for multiple comparison problems. Using frequentist methods, analysis of variance (ANOVA) is used to test for 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  $\alpha$ . However, these solutions do not determine any 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  $\mu$ . 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 illustrate the likelihood of grouping particular means as equal or unequal. We used a model selection technique to determine, based on the data, the likely model(s). This is developed for  $\boldsymbol{\eta} = (\eta, \dots, \eta)$  (constant variance) and  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_k)$  (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 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, and 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.

#### 4.2 Generalized Fiducial Inference

#### 4.2.1 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; 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.
To formally introduce fiducial inference we assume that a relationship, called the *structural equation*, between the data,  $\mathbf{X}$ , and the parameters,  $\xi$ , exists in the form

$$\mathbf{X} = G(\xi, \mathbf{U}) \tag{4.1}$$

where **U** is a random vector with a completely known distribution and independent of any parameters. After observing the data, **X**, we use the known distribution of **U** and the relationship from the structural equation to infer a distribution on  $\xi$ . This allows us to to define a probability measure on the parameter space,  $\Xi$ .

An inverse of the structural equation with respect to  $\xi$  is defined by the set valued function

$$Q(\mathbf{x}, \mathbf{u}) = \{\xi : \mathbf{x} = G(\xi, \mathbf{u})\}.$$
(4.2)

The generalized fiducial distribution is calculated by:

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

$$(4.3)$$

where  $\mathbf{U}^*$  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 a (possibly random) rule for discerning among the values when  $Q(\mathbf{x}, \mathbf{U}^*)$  has multiple solutions.

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  $i = 1, \ldots, n$ .  $\xi$  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  $\mathbf{i} = (i_1, \ldots, i_p)$ . Assume that the functions  $\mathbf{G}_{0,\mathbf{i}}$  are oneto-one and differentiable. Under some technical assumptions in Hannig (2009a) this will produce the generalized fiducial density of

$$f_{\mathcal{R}_{\xi}}(\xi) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')d\xi'}$$
(4.4)

where

$$J(\mathbf{x},\xi) = \binom{n}{p}^{-1} \sum_{\mathbf{i}=(i_1,\dots,i_p)} \left| \frac{\det\left(\frac{d}{d\xi} \mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)}{\det\left(\frac{d}{d\mathbf{u}_{\mathbf{i}}} \mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}},\xi)\right)} \right|$$
(4.5)

is the average of all subsets where  $1 \leq i_1 < \cdots < i_p \leq n$  and the determinants in (4.5) are the appropriate Jacobians.

#### 4.3 Main Results

#### 4.3.1 Structural equation with constant variance

In a multiple comparison problem we have k populations with means  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$ . Data, which follows an independent normal distribution, is of the form  $\mathbf{X}_i = (X_{i1}, \dots, X_{in_i})$  for all  $i = 1, \dots, k$  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  $\mathbf{X}_i = (X_{i1}, \dots, X_{in_i})$  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:

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

From these structural equations the generalized fiducial density from (4.4) can be calculated for each model ( $\mu_1 = \mu_2$  and  $\mu_1 \neq \mu_2$ ).

To simplify notation we will use  $J = U_1|U_2| \dots |U_t|$  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  $J = 1 \ 2 \ 3$  then  $U_1 = 1 \ 2 \ 3$ 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 total groupings in J, and the unique means are  $(\mu_1^*, \mu_2^*, \dots, \mu_t^*)$ .

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:

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

where 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.

Following the recipe in (4.5) the generalized fiducial density for a grouping, J, is:

$$f_{J}(\xi) \propto \frac{V_{x,J}w_{J}(\mathbf{x})}{\eta} \frac{1}{(2\pi)^{n_{1/2}} \eta^{n_{1/2}}} \exp\left\{-\frac{1}{2\eta} \sum_{j=1}^{n_{1}} (x_{1j} - \mu_{1})^{2}\right\}$$
$$\cdots \times \frac{1}{(2\pi)^{n_{k/2}} \eta^{n_{k/2}}} \exp\left\{-\frac{1}{2\eta} \sum_{j=1}^{n_{k}} (x_{kj} - \mu_{k})^{2}\right\}$$
$$= V_{x,J} \frac{\eta^{-N/2-1}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2\eta} \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} (x_{ij} - \mu_{i})^{2}\right\}$$
$$= V_{x,J} \frac{\eta^{-N/2-1}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2\eta} \sum_{i=1}^{t} n_{i}' \left(\mu_{i}^{*} - \bar{x}_{i}'\right)^{2}\right\} \exp\left\{-\frac{1}{2\eta} \sum_{i=1}^{t} n_{i}'MSX_{i}'\right\} 4.6$$

where

$$J_{J}(\mathbf{x},\xi) = C_{N,J}^{-1} \begin{cases} \frac{\sum_{l=1}^{t} \sum_{i_{1},i_{2} \in U_{l},i_{1} < i_{2}} \sum_{j_{1},j_{1}} |x_{i_{1},j_{1}} - x_{i_{2},j_{2}}|}{2\eta} & t < k \\ \frac{\sum_{l=1}^{k} \sum_{j_{1} <,j_{2}} |x_{l,j_{1}} - x_{l,j_{2}}|}{2\eta} & t = k \end{cases}$$
$$= \frac{V_{x,J}}{\eta},$$
$$n'_{i} = \sum_{l \in U_{i}} n_{l}, \ \bar{x}'_{i} = \frac{\sum_{l \in U_{i}} \sum_{j=1}^{n_{l}} x_{lj}}{n'_{i}},$$
$$MSX'_{i} = \frac{\sum_{l \in U_{i}} \sum_{j=1}^{n_{l}} (x_{lj} - \bar{x}'_{i})^{2}}{n'_{i}}, \ N = \sum_{i=1}^{k} n_{i},$$

 $C_{N,J}$  is the number Jacobian terms to average over, and  $w_J(\mathbf{x})$  is a weight function that will be derived in the next paragraph. Using (4.6) the fiducial density can be calculated as:

$$f(\xi) \propto \sum_{J \in \{J_1, \dots, J_H\}} f_J(\xi) I_J.$$

From equation (4.3) 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. If all of the means are different (J = 1|2|3|...|k) then p = k + 1  $(\xi = (\mu_1, ..., \mu_k, \eta))$ and we condition on N - (k+1) events. If all of the means are equal  $(J = 1 \ 2 \ 3 ... k)$ 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.

Adding structural equations will also allow us to introduce a weight function. The weight function will allow us to down-weight the models with less groupings to increase the likelihood of grouping several means together. Additionally, this 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 and we can show that our method is asymptotically scale invariant.

The additional structural equations are:

$$\frac{1}{2} \log \left( \frac{1}{\overline{MSXN}} \right) = \beta_i + P_i \quad \text{if } i \ge t$$

$$\frac{1}{2} \log \left( \frac{1}{\overline{MSXN}} \right) = P_i \quad \text{if } i < t \quad (4.7)$$

where  $\overline{MSX} = k^{-1} \sum_{i=1}^{k} MSX_i$ ,  $MSX_i$  is the maximum likelihood estimate of the variance for group *i*, and  $P_i$  is an independent Exp(1) random variable for all *i*. Because of the independence these structural equations will not affect the distribution of **X** but they will affect the conditional distribution in (4.3). When inverting the structural equations in (4.7), if  $i \geq t$  we can choose a  $\beta_i$  for any  $P_i$  so that the equation is satisfied. Thus, conditioning on this equation will not effect the conditional distribution. If i < t then  $P_i = 2^{-1} \log \left( \left[ \overline{MSXN} \right]^{-1} \right)$  which creates an additional

condition to be met. Combining the additional condition with the original structural equations there will now always be N-2 conditions regardless of the grouping of the means. This will define the weight function as:

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

If we recognize that  $\mu_i^* | \eta$  follows a normal distribution for all *i* and  $\eta$  follows an inverse gamma distribution then we can integrate  $f_J(\xi)$  over the space  $\Xi$ ,

$$p_{J} = \int_{\Xi} f_{J}(\xi) d\xi$$
  
=  $\frac{V_{x,J} w_{J}(\mathbf{x}) 2^{N/2} \pi^{t/2} \Gamma\left(\frac{N-t}{2}\right)}{\left(2\pi\right)^{N/2} \left(\sum_{i=1}^{t} n'_{i} MSX'_{i}\right)^{(N-t)/2} \prod_{i=1}^{t} \sqrt{n'_{i}}}$ 

Thus, we can find the probability that each J is correctly grouping the means by:

$$P(J) = \frac{p_J}{\sum_{\tilde{J}} p_{\tilde{J}}}.$$
(4.8)

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

#### 4.3.2 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:

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

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

Following the recipe in (4.5) the generalized fiducial density is:

$$f_{J}(\xi) \propto \frac{V_{x,J}w_{J}(\mathbf{x})}{\prod_{i=1}^{k} \eta_{i}} \frac{1}{(2\pi)^{n_{1/2}} \eta_{1}^{n_{1/2}}} \exp\left\{-\frac{1}{2\eta_{1}} \sum_{j=1}^{n_{1}} (x_{1j} - \mu_{1})^{2}\right\}$$
$$\cdots \times \frac{1}{(2\pi)^{n_{k/2}} \eta_{k}^{n_{k/2}}} \exp\left\{-\frac{1}{2\eta_{k}} \sum_{j=1}^{n_{k}} (x_{kj} - \mu_{k})^{2}\right\}$$
$$= V_{x,J}w_{J}(\mathbf{x}) \frac{\prod_{i=1}^{k} \eta_{i}^{-n_{i/2}-1}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} n_{i} \frac{((\mu_{i} - \bar{x}_{i})^{2} + MSX_{i})}{\eta_{i}}\right\} (4.9)$$

where

$$J_{J}(\mathbf{x},\xi) = C_{N,J}^{-1} \frac{\sum_{z=1}^{k} \sum_{j_{1,z} < j_{2,z} \le n_{z}} \sum_{(\mathbf{i}_{1}...,\mathbf{i}_{t})} \sum_{j=(j_{1,z},j_{2,z})} |T|}{2^{k} \prod_{i=1}^{k} \eta_{i}} = \frac{V_{x,J}}{\prod_{i=1}^{k} \eta_{i}},$$

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

$$T = \prod_{l=1}^{t} \left[ \prod_{i \in \mathbf{i}_{l}} \left( x_{i,j} - \mu_{l}^{*} \right) \right] \left( x_{i_{u_{l}},j_{1,z}} - x_{i_{u_{l}},j_{2,z}} \right),$$
$$\bar{x}_{i} = \frac{\sum_{j=1}^{n_{i}} x_{ij}}{n_{i}}, \ MSX_{i} = \frac{\sum_{j=1}^{n_{i}} \left( x_{lj} - \bar{x}_{i} \right)^{2}}{n_{i}}$$

and  $w_J(\mathbf{x})$  is the weight function. As an example of the Jacobian, if J = 1|2|3 then we average over

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

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

$$\frac{\left|\left(x_{1,j_{1,1}}-\mu_{1}^{*}\right)\left(x_{2,j_{1,2}}-x_{2,j_{2,2}}\right)\left(x_{3,j_{1,3}}-x_{3,j_{2,3}}\right)\right|}{2^{k}\prod_{i=1}^{k}\eta_{i}} + \frac{\left|\left(x_{1,j_{1,1}}-x_{1,j_{2,1}}\right)\left(x_{2,j_{1,2}}-\mu_{1}^{*}\right)\left(x_{3,j_{1,3}}-x_{3,j_{2,3}}\right)\right|}{2^{k}\prod_{i=1}^{k}\eta_{i}}$$

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

This 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:

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

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

and the weight function is:

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

where  $b_i = n_i / \max_j(n_j)$ . We can find P(J) using (4.8) where

$$p_J = \int_{\Xi} f_J(\xi) d\xi.$$

In this case  $p_J$  can not be calculated in closed form.

#### 4.3.3 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 I to sample from (4.9) 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.

## **Constant Variance**

Looking at a few interesting cases will help us assess the validity of the method. Figure 4.1 illustrates that the correct grouping,  $J = 1 \ 2 \ 3$ , 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  $\mu_0 = (1, 1.5, 1.5)$  or  $\mu_0 = (1, 1.5, 2)$  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. Figures 4.2 and 4.3 reflect this.

The easiest case occurs when the means are very different. Figure 4.4 demonstrates P(J) when  $\mu_0 = (1, 3, 5)$  and  $\eta_0 = 1$ .



Figure 4.1: P(J) for  $\mu_0 = (1, 1, 1)$  and  $\eta_0 = 1$  and 100 for top and bottom row respectively.



Figure 4.2: P(J) for  $\mu_0 = (1, 1.5, 1.5)$  and  $\eta_0 = 1$ .



Figure 4.3: P(J) for  $\mu_0 = (1, 1.5, 2)$  and  $\eta_0 = 1$ .

Similar analysis can be done at higher dimensions. Again, when k = 4,  $\mu_0 = (1, 1, 2, 2)$ , and  $\eta_0 = 1$  our method is selecting the correct model at a high rate as the sample size increases. Figure 4.5 reflects this. The omitted groupings, J, in the figures had median probability, P(J), of less than 0.02.

## 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 Figure 4.6.

Again the easy case is when the means are very different from each other. Figure 4.7 is reflective of this.



Figure 4.4: P(J) for  $\mu_0 = (1, 3, 5)$  and  $\eta_0 = 1$ .



Figure 4.5: P(J) for  $\mu_0 = (1, 1, 2, 2)$  and  $\eta_0 = 1$ .



Figure 4.6: P(J) for  $\mu_0 = (1, 1, 1)$  and  $\eta_0 = (1, 1, 1)$ , (1, 2, 3), and (100, 100, 100) for top, middle, and bottom rows respectively.



Figure 4.7: P(J) for  $\boldsymbol{\mu}_0 = (1, 3, 5)$  and  $\boldsymbol{\eta}_0 = (1, 1, 1)$ .



Figure 4.8: P(J) for  $\mu_0 = (1, 1, 2, 2)$  and  $\eta_0 = (1, 1, 1, 1)$ .

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 Figure 4.8 for all J where the median probability is greater than 0.02.

#### 4.4 Asymptotic results

As defined in equation (4.8) 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 4.**  $X_{ij}$  is an independent random variable from a  $N(\mu_i, \eta_i)$  distribution.

Assumption 5. There exists  $0 < b_i < \infty$  such that  $n_i = b_i n$  for all  $i = 1, \ldots, k$ .

**Theorem 3.** 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 II.

# 4.5 Examples

#### Simulated data

To 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 the data is:

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

and

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

Constant	Variance	Non-Constant Variance			
J	P(J)	J	P(J)		
1 2 3 4	0.049	1 2 3 4	0.061		
1 2 4 3	0.051	$1 2 \ 4 3$	0.058		
1 2 3 4	0.062	1 2 3 4	0.065		
$1\ 2 3\ 4$	0.044	$1\ 2 3\ 4$	0.041		
$1 2 \ 3 \ 4$	0.663	$1 2 \ 3 \ 4$	0.604		
$1\ 2\ 3\ 4$	0.060	$1\ 2\ 3\ 4$	0.071		

Table 4.1: Multiple comparison P(J) for the simulated example.

This data set was generated from independent normal distributions with  $\mu_0 = (1, 2, 2, 2)$ ,  $\eta_0 = (2, 2, 2, 2)$ , and a simple size of n = 20 for each treatment. Table 4.1 reflects grouping probabilities when P(J) > 0.03. Both the constant and nonconstant variance methods select the correct grouping at a high rate (P(J) = 0.663and P(J) = 0.604 for J = 1|2|3|4 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 sequential 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  $\mu_i = \mu_{i+1} = \cdots = \mu_{i+r}$ ,

$$P(\mu_i = \mu_j) = \sum_{J \in \{J_1, \dots, J_H\}} P(J) I_{\{J: \mu_i = \mu_j\}}$$
(4.10)

and

$$P(\mu_i = \mu_{i+1} = \dots = \mu_{i+r}) = \sum_{J \in \{J_1, \dots, J_H\}} P(J) I_{\{J: \mu_i = \mu_{i+1} = \dots = \mu_{i+r}\}}.$$
(4.11)

Figure 4.9 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



Figure 4.9:  $P(\mu_i = \mu_j)$  and  $P(\mu_i = \mu_{i+1} = \cdots = \mu_{i+r})$  with constant and non-constant variance for the simulated example.

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).

# 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 4.2.

We analyzed this data set using both the constant and non-constant variance methods. The grouping probabilities are seen in Table 4.3 when P(J) > 0.03. If we assume that the variance is constant J = 1 2/3 4/5/6 is the most likely scenario. If we do not assume that the variance is constant the most likely grouping is J = 1 2/3 4/5 6. Looking at the sample means and standard deviations both of these results seem very reasonable.

	Treatments						
	1	2	3	4	5	6	
	3DOk13	3DOk4	Composite	3DOk7	3DOk5	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 4.2: Rhizobium Data

Constant V	ariance	Non-Constant Variance			
J	P(J)	J	P(J)		
1 2 3 4 5 6	0.037	1 2 3 4 5 6	0.041		
$1 \ 2 3 4 5 6$	0.100	$1\ 2 3 4 5 6$	0.097		
1 2 3 4 5 6	0.052	1 2 3 4 5 6	0.052		
$1 2 3 4\ 5 6$	0.030	1 2 3 4 5 6	0.050		
$1\ 2 3\ 4 5 6$	0.196	$1\ 2 3\ 4 5 6$	0.115		
$1\ 2 3 4\ 5 6$	0.063	$1\ 2 3 4\ 5 6$	0.049		
$1\ 2 3 4 5\ 6$	0.043	$1\ 2 3 4 5\ 6$	0.102		
1 2 3 4 5 6	0.036	1 2 3 4 5 6	0.058		
$1\ 2 3\ 4 5\ 6$	0.078	$1\ 2 3\ 4 5\ 6$	0.139		
$1\ 2 3\ 4\ 5 6$	0.051	$1\ 2 3\ 4\ 5 6$	0.042		

Table 4.3: Multiple comparison P(J) for the red clover example.

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 and the groupings using a Dirichlet process prior. Table 4.4 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 4.4 are their recommended groupings.

Similarities between our analysis and theirs exist.  $J = 1 \ 2|3 \ 4|5|6$  and  $1 \ 2|3 \ 4|5|$ 6 are common to all of the methods as likely groupings of the means.

Figure 4.10 pictorially represent the probabilities in equations (4.10) and (4.11) for the red clover plant example. As the pictures show it is very reasonable that

					M				
	0.334	0.733	1.373	1.956	2.605	3.462	4.909	9.13	19.88
J	posterior probabilities								
$1\ 2\ 3\ 4 5 6$	0.032	0.059	0.057	0.046	0.041	0.034	0.026	0.013	0.005
$1\ 2 3\ 4\ 5 6$	0.186	0.211	0.202	0.189	0.171	0.148	0.112	0.061	0.020
$1\ 2 3\ 4 5\ 6$	0.144	0.178	0.167	0.149	0.137	0.116	0.094	0.049	0.016
$1\ 2 3\ 4 5 6$	0.037	0.086	0.152	0.199	0.229	0.257	0.276	0.281	0.199

Table 4.4: Posterior probabilities of select J for the red clover example.



Figure 4.10:  $P(\mu_i = \mu_j)$  and  $P(\mu_i = \mu_{i+1} = \cdots = \mu_{i+r})$  with constant and non-constant variance for the red clover example.

 $\mu_1 = \mu_2$ ,  $\mu_3 = \mu_4$ , 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  $\alpha = 0.05$ .

## 4.6 Conclusion

Frequentist solutions for multiple comparison problems can test for a treatment effect or find differences among treatments. However, they can not 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. Analyzing the red clover example under the assumption of constant variance we found that J = 1 2|3 4|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=1 2|3 4|5 6 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 straight forward and the asymptotic properties and simulation results assure high confidence in the analysis.

# Appendix I

## Importance Sampling Algorithm

The following steps were implemented in order to obtain a fiducial sample for  $\xi$ .

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

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

Generate  $(\eta_l | \mu_l) = W$  where  $W \sim Inv - Gamma\left(\frac{n_l}{2}, \frac{n_l((\mu_l - \bar{x}_l)^2 + MSX_l)}{2}\right)$  for  $l = 1, \dots, k$ .

3. Calculate weights of each generated sample with,

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

where  $f_J(\xi)$  is the generalized fiducial density for the model with groupings Jand  $g_i(\mu_i)$  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  $ESS_J = n_J \left(1 + \left(s_{w_J}^2\right) \bar{w}_J^{-2}\right)^{-1}$  where  $n_J$  is the sample size for model J,  $s_{w_J}^2$ is the sample variance of the weights, and  $\bar{w}_J$  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 II

#### Proof of theorem 3 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 3 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  $J_2$ .

$$\begin{split} \frac{p_{J_2}}{p_{J_1}} &\propto \frac{\Gamma\left(\frac{N-t_2}{2}\right) \left(\sum_{i=1}^{t_1} n'_{1i} MSX'_{1i}\right)^{(N-t_1)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\Gamma\left(\frac{N-t_1}{2}\right) \left(\sum_{i=1}^{t_2} n'_{2i} MSX'_{2i}\right)^{(N-t_2)/2} \prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \\ & \text{using Stirling's formula} \\ &\leq \frac{(2e)^{(t_2-t_1)/2} N^{(t_1-t_2)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \frac{\left(\sum_{i=1}^{t_1} n'_{1i} MSX'_{1i}\right)^{(N-t_1)/2}}{\left(\sum_{i=1}^{t_2} n'_{2i} MSX'_{2i}\right)^{(N-t_2)/2}} \\ & \text{WLOG assume } U_1 \in J_2 \text{ is an incorrect grouping} \\ &\leq \frac{(2e)^{(t_2-t_1)/2} N^{(t_1-t_2)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \frac{\eta_0^{(t_2-t_1)/2} \left(\sum_{i=1}^{t_1} n'_{1i} \frac{\eta_0^{(1+O(1))}}{\eta_0}\right)^{(N-t_1)/2}}{\left(n'_{21} \frac{\eta^*(1+O(1))}{\eta_0} + \sum_{i=2}^{t_2} n'_{2i} \frac{\eta_0(1+O(1))}{\eta_0}\right)^{(N-t_2)/2}} \\ & \text{where } \eta^* > \eta \text{ because of the incorrect grouping} \\ &\leq \frac{(2e)^{(t_2-t_1)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \frac{\eta_0^{(t_2-t_1)/2} \left(1 + \frac{r}{4} \left(\frac{\eta^*}{\eta_0} - 1\right)\right)^{(N-t_1)/2}}{\left(1 + (\eta^* - 1)^2\right)} \text{ Eventually a.s.} \end{split}$$

$$\leq \frac{(2e)^{(t_2-t_1)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \frac{\eta_0^{(t_2-t_1)/2} \left(1 + \frac{r}{4} \left(\frac{\eta^*}{\eta_0} - 1\right)\right)^{(N-t_1)/2}}{\left(\left[1 + r \left(\frac{\eta^*}{\eta_0} - 1\right)\right] c\right)^{(N-t_2)/2}} \text{ Eventually a.s.} \\ \to 0 \text{ a.s.}$$

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

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

$$J_1 = U_{11}|U_{12}|\dots|U_{1t_1}$$
$$J_2 = U_{21}|U_{22}|\dots|U_{2t_2}$$

where

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

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

$$\frac{p_{J_2}}{p_{J_1}} \propto \frac{\Gamma\left(\frac{N-t_2}{2}\right) \left(\sum_{i=1}^{t_1} n'_{1i} M S X'_{1i}\right)^{(N-t_1)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\Gamma\left(\frac{N-t_1}{2}\right) \left(\sum_{i=1}^{t_2} n'_{2i} M S X'_{2i}\right)^{(N-t_2)/2} \prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \\
\leq \frac{\left(2e\right)^{(t_2-t_1)/2} N^{(t_1-t_2)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \frac{\left(\sum_{i=1}^{t_1} S S X'_{1i}\right)^{(N-t_1)/2}}{\left(\sum_{i=1}^{t_2} S S X'_{2i}\right)^{(N-t_2)/2}} \\
\leq \frac{\left(2e\right)^{(t_2-t_1)/2} \left(2\eta\right)^{(-t_1+t_2)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \left(1 - \frac{\left(\sum_{i=1}^{t_1} \sum_{k \in K_i} n_{2k} \frac{16\eta \log \log N}{n_{2k}}\right)}{N\eta}\right)^{-N/2} \\$$
eventuall a.s. using the law of iterated logarithms

$$\leq \frac{(2e)^{(t_2-t_1)/2} (2\eta)^{(-t_1+t_2)/2} \prod_{i=1}^{t_1} \sqrt{n'_{1i}}}{\prod_{i=1}^{t_2} \sqrt{n'_{2i}}} \left(1 - \frac{R \log \log N}{N}\right)^{-N/2}$$
  
WLOG assume that  $U_{1t_1} = U_{2(t_2-1)} \cup U_{2t_2}$  and  $U_{1i} = U_{2i}$  for all other  $i \leq \frac{(2e)^{(t_2-t_1)/2} (2\eta)^{(-t_1+t_2)/2} b}{\sqrt{n'_{2t_2}}} \left(1 - \frac{R \log \log N}{N}\right)^{-N/2}$   
 $\rightarrow 0$  a.s.

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.

## Proof of theorem 3 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:

$$f_{J}(\xi) \propto \frac{V_{x,J}}{\prod_{i=1}^{k} \eta_{i}} \frac{1}{(2\pi)^{n_{1/2}} \eta_{1}^{n_{1/2}}} \exp\left\{-\frac{1}{2\eta_{1}} \sum_{j=1}^{n_{1}} (x_{1j} - \mu_{1})^{2}\right\}$$
$$\cdots \times \frac{1}{(2\pi)^{n_{k/2}} \eta_{k}^{n_{k/2}}} \exp\left\{-\frac{1}{2\eta_{k}} \sum_{j=1}^{n_{k}} (x_{kj} - \mu_{k})^{2}\right\}$$
$$= V_{x,J} \frac{\prod_{i=1}^{k} \eta_{i}^{-n_{i/2} - 1}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} n_{i} \frac{((\mu_{i} - \bar{x}_{i})^{2} + MSX_{i})}{\eta_{i}}\right\}$$

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  $J = U_1 | U_2 | \dots | U_t$  where  $U_i = r_{i1} r_{i2} \dots r_{iu_i}$  are the indexes of equal means

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

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

$$p_{J} = \int_{\Xi} f_{J}(\xi) d\xi$$
  
=  $\pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] \int_{\mathbb{R}^{t}} V_{x,J} \prod_{i=1}^{k} \left( (\mu_{i} - \bar{x}_{i})^{2} + MSX_{i} \right)^{-n_{i}/2} d\mu^{*}$ 

Because  $V_{x,J}$  is dependent on  $\mu_i^*$  we will bound this value. It is clear that  $V_{x,J} > c_1$ for some  $c_1 > 0$ . We could re-write  $V_{x,J}$  as:

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

where  $V_{i,j}$  are averages over a function of the data. If  $u_i$  is even then  $\left|\mu_i^{*(u_i-1)}\right| \leq \mu_i^{*u_i} + 1$  and if  $u_i$  is odd then  $\left|\mu_i^{*(u_i-1)}\right| = \mu_i^{*(u_i-1)}$ . 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.

$$V_{x,J} \leq \left(\prod_{i=1}^{t} \mu_i^{*(u_i-1)}\right) |V_{1,1}| + \sum_{j=1}^{t} \left(\left(\mu_j^{*(u_j-1)} + 1\right) \prod_{i=1, i \neq j}^{t} \mu_i^{*(u_i-1)}\right) |V_{2,j}| + \dots + |V_{z,1}|$$
  
$$\leq \left(\prod_{i=1}^{k} \left((\mu_i - \bar{x}_i)^2 + MSX_i\right)^{(u_i-1)/2}\right) |V^{(1)}| + |V^{(2)}|$$

where  $V^{(1)}$  and  $V^{(2)}$  are averages over the data,  $\bar{x}_i$ , 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:

$$p_{J} \geq p_{J}^{\bullet}$$
  
=  $c_{1}\pi^{-N/2}\prod_{i=1}^{k} \left[n_{i}^{-n_{i}/2}\Gamma\left(\frac{n_{i}}{2}\right)\right] \int_{\mathbb{R}^{t}}\prod_{i=1}^{k} \left((\mu_{i}-\bar{x}_{i})^{2}+MSX_{i}\right)^{-n_{i}/2} d\mu^{*}.$ 

An upper bound for  $p_J$  is:

$$p_{J} \leq \pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] \int_{\mathbb{R}^{t}} \frac{\left( \prod_{i=1}^{k} \left( (\mu_{i} - \bar{x}_{i})^{2} + MSX_{i} \right)^{(u_{i} - 1)/2} \right) |V^{(1)}| + |V^{(2)}|}{\prod_{i=1}^{k} \left( (\mu_{i} - \bar{x}_{i})^{2} + MSX_{i} \right)^{n_{i}/2}} d\mu^{*}$$

$$\leq \pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] \int_{\mathbb{R}^{t}} \frac{|V^{(1)}|}{\prod_{i=1}^{k} \left( (\mu_{i} - \bar{x}_{i})^{2} + MSX_{i} \right)^{(n_{i} - u_{i} - 1)/2}} + \frac{|V^{(2)}|}{\prod_{i=1}^{k} \left( (\mu_{i} - \bar{x}_{i})^{2} + MSX_{i} \right)^{n_{i}/2}} d\mu^{*}$$

$$\leq \pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] \int_{\mathbb{R}^{t}} \frac{c_{2} |V^{(1)}|}{\prod_{i=1}^{k} \left( (\mu_{i} - \bar{x}_{i})^{2} + MSX_{i} \right)^{(n_{i} - u_{i} - 1)/2}} d\mu^{*}$$

$$= p_{J}^{A}.$$

for some  $c_2 > 0$ .

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

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

with the transformations of:

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

and the substitutions of

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

and

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

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

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

where **m** and **m**<sup>\*</sup> follows the same structure as  $\boldsymbol{\mu}$  and  $\boldsymbol{\mu}^*$  above 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:

$$\log(g_{J}(\mathbf{m}^{*})) = -\frac{t}{2}\log(n) + \sum_{i=1}^{k} \left[ -\frac{b_{i}n\log(\eta_{i0} + \Delta_{i}^{2})}{2} - \frac{b_{i}\sqrt{n}\left(2\Delta_{i}\left(m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}}\right) + \frac{Z_{i2}}{\sqrt{b_{i}}}\right)}{2(\eta_{i0} + \Delta_{i}^{2})} + \frac{b_{i}\left(2\Delta_{i}\left(m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}}\right) + \frac{Z_{i2}}{\sqrt{b_{i}}}\right)^{2}}{4(\eta_{i0} + \Delta_{i}^{2})^{2}} - \frac{b_{i}\left(m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}}\right)^{2}}{(2\eta_{i0} + \Delta_{i}^{2})} + O\left(n^{-1/2}\right) \right].$$

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

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

for all  $j = 1, \ldots, t$ . Thus,

$$\log(g_J(\mathbf{m}^*)) = -\frac{t}{2}\log(n) + \sum_{i=1}^{k} \left[ -\frac{b_i n \log(\eta_{i0} + \Delta_i^2)}{2} + \frac{\sqrt{b_i n} (2\Delta_i Z_{i1} - Z_{i2})}{2(\eta_{i0} + \Delta_i^2)} + \frac{b_i \left( 2\Delta_i \left( m_i - \frac{Z_{i1}}{\sqrt{b_i}} \right) + \frac{Z_{i2}}{\sqrt{b_i}} \right)^2}{4(\eta_{i0} + \Delta_i^2)^2} - \frac{b_i \left( m_i - \frac{Z_{i1}}{\sqrt{b_i}} \right)^2}{(2\eta_{i0} + \Delta_i^2)} + O(n^{-1/2}) \right].$$

Or

$$g_{J}(\mathbf{m}^{*}) = \frac{1}{n^{t/2} \prod_{i=1}^{k} (\eta_{i0} + \Delta_{i}^{2})^{b_{i}n/2}} \exp\left\{\sum_{i=1}^{k} \frac{\sqrt{b_{i}n} (2\Delta_{i}Z_{i1} - Z_{i2})}{2(\eta_{i0} + \Delta_{i}^{2})}\right\}$$
$$\times \exp\left\{-\sum_{i=1}^{t} \frac{1}{2\sigma_{\mathbf{Z},i}^{2}} (m_{i}^{*} - \zeta_{\mathbf{Z},i})^{2} + C_{\mathbf{Z}} + O(n^{-1/2})\right\}$$
(4.12)

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 4. Let

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

for the previously described  $C_n$ , then  $h_{J,n}(\mathbf{m}^*) \leq k_J(\mathbf{m}^*)$  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  $\mu_{j0}^*$  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  $l = 1, ..., (u_j - 1)$ . 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  $\eta_s \neq \eta_r$  for all  $s, r \in U_j$ ) or  $1 - c_{jl}n^{-1/2}$  if  $\eta_s \neq \eta_r$  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  $\exp\left\{-c_{jl}n^{1/2}\right\}$ , i.e., the local maximum is located at the point  $(c_{j0}n^{1/2}, \exp\left\{-c_{jl}n^{1/2}\right\})$  which is well below the Cauchy density of  $(c_{j0}n^{1/2}, c\left(1 + c_{jl}^2n\right)^{-1})$  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.  $\Box$ 

From equation (4.12) we can see  $m_i^*$  converges point-wise to a normal distribution and lemma 4 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,

$$p_{J}^{\bullet} = \frac{\pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] c_{1}}{n^{t/2} (\eta_{i0} + \Delta_{i}^{2})^{b_{i}n/2}} \exp\left\{ \sum_{i=1}^{k} \frac{\sqrt{b_{i}n} (2\Delta_{i}Z_{i1} - Z_{i2})}{2 (\eta_{i0} + \Delta_{i}^{2})} \right\} \\ \times \int_{\mathbb{R}^{t}} \exp\left\{ \frac{b_{i} \left( 2\Delta_{i} \left( m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}} \right) + \frac{Z_{i2}}{\sqrt{b_{i}}} \right)^{2}}{4 (\eta_{i0} + \Delta_{i}^{2})^{2}} - \frac{b_{i} \left( m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}} \right)^{2}}{(2\eta_{i0} + \Delta_{i}^{2})} + O\left( n^{-1/2} \right) \right\} d\mathbf{m}^{*} \\ = \frac{\pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] c_{1}}{n^{t/2} (\eta_{i0} + \Delta_{i}^{2})^{b_{i}n/2}} \exp\left\{ \sum_{i=1}^{k} \frac{\sqrt{b_{i}n} (2\Delta_{i}Z_{i1} - Z_{i2})}{2 (\eta_{i0} + \Delta_{i}^{2})} \right\} B_{1,n}$$

and a similar calculation produces

$$p_{J}^{\bullet} = \frac{\pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] c_{2} V^{(1)}}{n^{t/2} (\eta_{i0} + \Delta_{i}^{2})^{(b_{i}n-u_{i}-1)/2}} \exp\left\{ \sum_{i=1}^{k} \frac{\sqrt{b_{i}n} (2\Delta_{i}Z_{i1} - Z_{i2})}{2(\eta_{i0} + \Delta_{i}^{2})} \right\} \\ \times \int_{\mathbb{R}^{t}} \exp\left\{ \frac{b_{i} \left( 2\Delta_{i} \left( m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}} \right) + \frac{Z_{i2}}{\sqrt{b_{i}}} \right)^{2}}{4(\eta_{i0} + \Delta_{i}^{2})^{2}} - \frac{b_{i} \left( m_{i} - \frac{Z_{i1}}{\sqrt{b_{i}}} \right)^{2}}{(2\eta_{i0} + \Delta_{i}^{2})} + O\left( n^{-1/2} \right) \right\} d\mathbf{m}^{*} \\ = \frac{\pi^{-N/2} \prod_{i=1}^{k} \left[ n_{i}^{-n_{i}/2} \Gamma\left(\frac{n_{i}}{2}\right) \right] c_{2} V^{(1)}}{n^{t/2} (\eta_{i0} + \Delta_{i}^{2})^{(b_{i}n-u_{i}-1)/2}} \exp\left\{ \sum_{i=1}^{k} \frac{\sqrt{b_{i}n} (2\Delta_{i}Z_{i1} - Z_{i2})}{2(\eta_{i0} + \Delta_{i}^{2})} \right\} B_{2,n}$$

where  $B_{i,n}$  is the constant that comes from integration of the normal density and  $B_{i,n} \to B_i$  by lemma 4.

To prove that  $P(J) \to 1$  as  $n \to \infty$  we will observe  $p_{\tilde{J}}/p_J \leq p_{\tilde{J}}^{\bullet}/p_J^{\bullet} \to 0$  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  $J_1$ .

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

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  $t_2 > t_1$ .

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

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

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

# Chapter 5

# CONCLUSION AND FUTURE WORK

This dissertation has used the fiducial framework to propose solutions to some very interesting problems. We used fiducial interference on the parameters and extreme quantiles of the generalized Pareto distribution, the largest mean of a correlated multivariate normal distribution, and the model selection of some multiple comparison problems. These solutions proved to have advantages over the current methods and have good asymptotic properties.

All of these problems have natural extensions that future work can address.

#### 5.1 Conclusion and future work for the generalized Pareto

We have demonstrated the fiducial solution to the generalized Pareto as a viable method in extreme value problems. We compared our method to some competing methods when the threshold is assumed to be known and unknown.

When the threshold is assumed to be known our method compared favorable to its competitors. Namely, the bias for the parameter estimates are less than the maximum likelihood, L-moment, and Bayesian estimates. The estimate for the high quantile is slightly less biased using the Bayesian solution from Castellanos and Cabras (2005). Also, the fiducial approach produced confidence intervals for the high quantile that are shorter than those calculated by the Bayesian and profile log-likelihood methods when  $\gamma_0 > 0$  for all sample sizes and when  $\gamma_0 \leq 0$  for large sample sizes.

A fiducial approach was also developed to handle the situation when the threshold is assumed to be an unknown parameter. There are very few competing methods in this case. Our method produced intervals with good empirical coverage regardless of the underlying distribution of the data. Likewise, the confidence intervals were very similar in length to the Bayesian intervals created in Cabras and Castellanos (2009) that had a reasonable coverage rate.

There is a variety of future open questions surrounding this problem. First, further investigation into the asymptotic properties when the threshold is unknown would be of interest. It could also be interesting to attempt to use fiducial inference when there are additional covariates that can be incorporated into the model. For example, as described in Coles (2001) if there are daily observations  $X_1, X_2, \ldots$  and let s(t) denote the season that observation t falls in then different thresholds can exist for each season. The seasonal component means that  $(X_t - a_{s(t)}|X_t > a_{s(t)})$  follows a generalized Pareto with seasonal parameters  $\gamma_{s(t)}$  and  $\sigma_{s(t)}$ . Another extension is attempting to use fiducial inference with a point process approach as stated in Smith (1989) and Coles (2001). Using a point process approach takes into account the probability of crossing the threshold.

# 5.2 Conclusion and future work for the largest mean of a multivariate normal distribution

The largest mean of correlated multivariate normal data posed numerous inference challenges. Our solution approached the problem assuming that there was an unstructured covariance matrix. We were able to develop confidence intervals that had good empirical properties when the correlation was positive. We also proved that our intervals were asymptotically correct. In comparison to other methods, the fiducial method produced upper tailed intervals that were generally shorter with better coverage rates. This methodology was developed for a single sample when  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$ , and

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

It could be of practical interest to extend this problem to two samples. Namely, if  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  and  $\mathbf{Y} \sim N(\boldsymbol{\tau}, \Sigma)$  the parameter of interest would be  $\theta = \max_i \mu_i - \tau_i$ . This should be a relatively straight forward extension but would have implications in drug trials where a treatment is compared to a placebo.

Another interesting extension would be when  $\Sigma$  was assumed to be a structured covariance matrix. Boose et al. (2007) assumed that the covariance matrix was structured as

$$\Sigma = \begin{bmatrix} \eta & \sigma & \cdots & \sigma \\ \sigma & \eta & & \vdots \\ \vdots & & \ddots & \sigma \\ \sigma & \cdots & \sigma & \eta \end{bmatrix}$$

This assumption should help us to get even better empirical properties at small sample sizes.

## 5.3 Conclusion and future work for multiple comparisons

Many solutions to multiple comparison problems allow for a test of the treatment affect or to test differences among individual means. Our fiducial solution to the multiple comparison problem allows us to calculate the fiducial probability of any grouping of the means. Using the data, we can come up with a very intuitive fiducial probability of the equalities and inequalities of the means. This information could be used to test specific hypotheses or to test equality of any number of means.

This method has been developed when the variance is assumed to be both constant and non-constant across the treatments. We have demonstrated, through simulation, that the method can select the correct grouping at a relatively high rate at all sample sizes and asymptotically selects the correct grouping with probability 1. A similar Bayesian method developed in Gopalan and Berry (1998) needs further assumptions about the parameters and produces posterior probabilities for the groupings that are very dependent on the prior distributions.

Additional work with different covariance structures would be very beneficial to this problem. As stated, we have developed the method for independent data. It could be of interest to further development to situations where there is correlation in the data. For example, if  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$ , and

$$\Sigma = \begin{bmatrix} \eta & \sigma & \cdots & \sigma \\ \sigma & \eta & & \vdots \\ \vdots & & \ddots & \sigma \\ \sigma & \cdots & \sigma & \eta \end{bmatrix}.$$

Other covariance structures may also be of interest to this problem (e.g. unstructured covariance).

Further improvement to the computational methods also need to be investigated. The number of models increase at an extremely fast rate as k increases. As a result, the computation time increases dramatically. Using a different Monte Carlo approach may prove to be more efficient and a better option.

# References

Abdel-Karim, A. H. (2005). Applications of generalized inference.

Boose, D. D., Hoffman, D., Kringle, R., and Zhang, J. (2007). New confidence bounds for qt studies. *Statistics in Medicine*, 26:3801–3817.

Bruijn, N. G. D. (1956). Inequalities concerning minors and eigenvalues. *Nieuw* Archief v. Wiskunde, IV(18-35).

Cabras, S. and Castellanos, M. E. (2009). An objective bayesian approach for threshold estimation in the peaks over the threshold model.

Castellanos, M. E. and Cabras, S. (2005). A default bayesian procedure for the generalized pareto distribution. *Journal of Statistical Planning and Inference*, 137:473 – 483.

Choulakian, V. and Stephens, M. A. (2001). Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, 43(478-484).

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.

Davison, A. C. (1984). *Statistical Extremes and Applications*, chapter Modelling Excesses Over High Thresholds With an Application, pages 461–482. Reidel, Dordrecht.

Davison, A. C. and Smith, R. L. (1990). Models for excedances over high thresholds. *Journal of the Royal Statistical Society Series B*, 52:393–442.

Dawid, A. P. (1994). *Multivariate Analysis and Its Applications*, volume 24, lecture notes-monograph series Selection paradoxes of Bayesian inference, pages 211–220. Institute of Mathematical Statistics, Hayward, CA.

Dupuis, D. J. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1(251-261).

Eaton, M. L., Muirhead, R. J., Mancuse, J. Y., and Kolluri, S. (2006). A confidence interval for maximal mean qt interval change caused by drug effect. *Drug Information Journal*, 40(3):267–271.

EPA (2009). http://www.epa.gov/air/data/geosel.html.

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

Frigessi, A., Haug, O., and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5:219–235.

Gopalan, R. and Berry, D. A. (1998). Bayesian multiple comparisons using dirichlet process priors. *Journal of the American Statistical Association*, 93(443):1130–1139.

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme values theory. *Journal of the Royal Statistical Society Series B*, 63:293–305.

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., and Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101:254–269.

Hannig, J. and Lee, T. C. (2009). Generalized fiducial inference for wavelet regression. *To appear in Biometrika*.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B*, 52:105–124.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349.

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society Series B*, 20:102–107.

Monfort, M. A. J. V. and Witter, J. V. (1985). Testing exponentiality against generalized pareto distribution. *Journal of Hydrology*, 78:305–315.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131.

Rados, G. (1896). Zur theorie der adjungirten substitutionen. *Mathematische Annalen*, 48(3):417–424.

Smith, R. L. (1984). *Statistical Extremes and Applications*, chapter Threshold Methods for Sample Extremes, pages 621–638. Reidel, Dordrecht.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–90.

Smith, R. L. (1989). Extreme value theory based on the r largest annual events. *Statistical Science*, 4:367–393.

Steele, R. D. and Torrie, J. H. (1980). *Principles and procedures of statistics-A biomedical approch*. McGraw Hill, New York, 2nd edition.

Tancredi, A., Anderson, C., and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9:87–1006.

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.