Simultaneous Inference Procedures in the Presence of Heteroscedasticity

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Meng Li, M.A.

Graduate Program in Statistics

The Ohio State University

2017

Dissertation Committee:
Steven N. MacEachern, Advisor
Jason C. Hsu
Jennifer Sinnott
Abstract

Simultaneous inference has become an increasingly important statistical tool in handling many real world problems. In a frequentist analysis, performing simultaneous inference typically requires exploring the null distribution of the multivariate test statistic in order to control the resulting inference at the pre-determined nominal level. As the probability calculation of the true null distribution involves high-dimensional integration that was computationally too difficult before the prevalence of computing resources, simultaneous inferences were only available for some special designs with specific sample size patterns in the one-way analysis of variance setting.

With the dramatic rise in computing power, numerical calculation of the multivariate $t$ probability becomes accessible, enabling one to perform simultaneous analyses in the general linear model framework assuming homogeneous variance. In this case, the null distribution of the test statistic for any set of linear hypotheses is a multivariate $t$ distribution. In this dissertation, we devise a computationally efficient algorithm to calculate the tail areas and find the critical values of any arbitrary multivariate $t$ distribution. The performance of the algorithm is compared to the existing method with regard to the accuracy of inference and computing time. In particular, the restriction on integer degrees of freedom of the current practice is relaxed in the new algorithm. An additional benefit of the new algorithm is the allowance for the error probability for each individual comparison to vary.
However, in the presence of heteroscedasticity, the true null distribution is unknown and often replaced by a computationally tractable approximation that is instead used for inference. The performance of the resulting inferential procedures is impacted by the discrepancy between the true null distribution and its approximation. Suggested by the lack of control over the family-wise error rate caused by failing to account for the dependence structure among the denominators of the test statistic of the Plug-in procedure – a common approach for multiple inference in the presence of heteroscedasticity – this dissertation develops an inferential procedure that exploits the dependence structure of the components of the test statistic more precisely.

The suggested approach approximates the unknown true null distribution by a variation of the classical multivariate $t$ distribution, referred to as the generalized multivariate $t$ distribution. We propose an extension of the numerical algorithm for the classical multivariate $t$ probability calculation to accommodate the generalized multivariate $t$. In addition to the accuracy gained by incorporating the dependence structure among the denominators into the procedure compared to the Plug-in method, the computing time is considerably reduced.

With the remarkable developments in statistical models, it becomes increasingly important to consider the impact that modeling choices have on the subsequent parametric inferences. We investigate the effect of modeling strategies on the inference in a special setting where the data generating process is a blend of two qualitatively separate portions – one portion is stable and amenable to modeling while another portion is unstable. Together, the reliability of the resulting inference on the truth can be attributable to multiple sources including, but not limited to, the model as a description of the mechanism and the inference procedure as assessed in this work.
This is dedicated to my family.
Acknowledgments

With passion and curiosity, I decided to set out on the graduate school journey. Years later, as the graduate life comes to an end, I affirmatively say that this journey is undoubtedly a rewarding and life-enriching one. I am grateful for everyone whom I had the opportunity to meet and shared the journey with even for just a moment, and every experience life sent my way to learn from. It is true that joy feels better than tears, but it was not until now did I realize that it is the difficulties that granted me the permit to see the extraordinary sceneries and encounter the many wonderful people I would cherish for life.

First, I would like to express my deepest gratitude to my advisor, Professor Steven MacEachern, for taking me as his student. I certainly would never have come this far without his help over the years. I am forever grateful for his close guidance on my research work and the enormous amount of time he invested in my individual studies with him. Thank you for always making yourself available for the innumerous times I showed up in your office abruptly with questions. Thank you for your patience with me on my often less than mature research ideas and guiding me towards the right path. Your support and encouragement have been incredibly helpful for both my academic and personal growth. The impact he has on my life is beyond he could ever imagine. For this, I will remain humble and grateful for the rest of my life.
I would like to give my special thanks to Professor Jason Hsu, whose rigorous attitude towards research has been a great inspiration. I will never forget the many times he shared his academic and industrial insights enthusiastically with me over coffee. His spirit of ‘what we do really makes a difference’ will undoubtedly have great influence on my future career. I am particularly grateful for your sharing knowledge and wisdom with the world passionately and unreservedly. I am indebted to the rest of my committee members, Dr. William Notz, and Dr. Jennifer Sinnott, for the insightful questions during my candidacy exam.

My sincere thanks also go to Dr. Christopher Holloman for offering me the opportunity to become a member of the consulting team for a year. I would like to thank for having the chance to learn from every one of my clients and every project that they came to me with.

I would like to give my thanks to the rest of the Statistics department for this open and enjoyable learning and working environment. I want to thank the staff for their willingness to help the students, and also Malika for her effort in keeping the offices clean and organized. I also appreciate the alternative teaching arrangement the department made in accommodating me through my family emergency.

Finally, thank you to my family for your unconditional love and support all along my life journey. The degree I am about to earn does not just belong to me, but to you as well. I especially thank my dad for believing in me and encouraging me to think big and try with boldness. I miss the fun language only you and I spoke and the time we spent together. In the midst of the great memories, it has always been my regret for having caused you all the worries in your last days. Thank you, mom, for always telling me to follow my heart and dream and you will be in support of every decision
I make whole-heartedly. Thank you for encouraging me to stick through when things seem difficult. It means so much to me that you do not always seem to ‘stand on my side’, but view situations from a neutral standpoint. Thank you for helping me understand how difficulties are part of life and how to see them as opportunities.

This acknowledgements could go on and on. Junguang, thank you for the exceptional friendship and staying with me during the stormiest time of my life. I will never forget the night-long phone calls we had sharing laughers and tears. Thank you for making me realize how much stronger we are than we think. I am grateful for every friend and fellow student that I have shared moments with over the last few years, especially the group of amazing friends, Molly, Sherry, Daniel, Carrie, Yihan, Yong, Wenna, Leo and Lanlan for your persistence and ‘never leaving me alone’ on getting me to join the group for fun activities, when I felt overwhelmed with school work.

Thank you, all! Time flies, but the memory will last forever.
Vita

April 13, 1987 .......................... Born - Weifang, Shandoing, China

2009 .......................... B.S. Mathematical Sciences, Shandoing University

2010 .......................... M.A. Statistics, Columbia University

2011-present .......................... Graduate Teaching Associate,
                                  The Ohio State University.

Fields of Study

Major Field: Statistics
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Vita</td>
<td>ix</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td><strong>1. Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Multiple Testing</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1 Homogeneous Variances</td>
<td>3</td>
</tr>
<tr>
<td>1.1.2 Heterogeneous Variances</td>
<td>8</td>
</tr>
<tr>
<td>1.2 Separation of the Quickly Varying and Slowly Varying</td>
<td>10</td>
</tr>
<tr>
<td>1.3 Overview of Dissertation</td>
<td>13</td>
</tr>
<tr>
<td><strong>2. Multivariate ( t )-Distribution</strong></td>
<td>15</td>
</tr>
<tr>
<td>2.1 Classical Multivariate ( t ) Distribution</td>
<td>15</td>
</tr>
<tr>
<td>2.1.1 Table 0/1 Approach: Probability Calculation</td>
<td>26</td>
</tr>
<tr>
<td>2.1.2 Table 0/1 Approach: Quantile Calculation</td>
<td>28</td>
</tr>
<tr>
<td>2.1.3 Spherical Radial Transformation: Probability Calculation</td>
<td>35</td>
</tr>
<tr>
<td>2.1.4 Spherical Radial Transformation: Quantile Calculation</td>
<td>53</td>
</tr>
<tr>
<td>2.2 Generalized Multivariate ( t ) Distribution</td>
<td>60</td>
</tr>
<tr>
<td>2.2.1 Mathematical Formulation and Model Parameterization</td>
<td>63</td>
</tr>
<tr>
<td>2.2.2 Generation from a Generalized Multivariate ( t ) Random Vector</td>
<td>84</td>
</tr>
<tr>
<td>2.2.3 Probability Calculation with the Table 0/1 Approach</td>
<td>85</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1 A Simple Demonstration of the 0/1 Implementation of Multivariate $t$ Probability Calculation</td>
<td>28</td>
</tr>
<tr>
<td>Table 3.1 The Variances Specification for the 8 Treatment Groups for each Scenario Considered in Simulation Setting (1)</td>
<td>146</td>
</tr>
<tr>
<td>Table 3.2 The Sample Sizes for the 8 Treatment Groups under each Scenario Considered in Simulation Setting (1)</td>
<td>147</td>
</tr>
<tr>
<td>Table 3.3 The Estimated Family-wise Error Rate for Various Approximate Procedures Under each Scenario in Simulation Setting (1)</td>
<td>153</td>
</tr>
<tr>
<td>Table 3.4 The Standard Errors associated with the Estimated Family-wise Error Rate for Various Approximate Procedures Under each Scenario in Simulation Setting (1)</td>
<td>153</td>
</tr>
<tr>
<td>Table 4.1 The Battery Lifetime Data</td>
<td>168</td>
</tr>
<tr>
<td>Table 4.2 Traditional Approach: Simultaneous Inferences on the Family of Pairwise Comparisons on the Average Life per Unit Cost for the Four Types of Batteries</td>
<td>170</td>
</tr>
<tr>
<td>Table 4.3 The Simultaneous Inferences on the Battery Example using the Plug-in Approach (with Rounded Down Degrees of Freedom by Default)</td>
<td>174</td>
</tr>
<tr>
<td>Table 4.4 The Simultaneous Inferences on the Battery Example using the Plug-in Approach with Rounded Up Degrees of Freedom</td>
<td>175</td>
</tr>
<tr>
<td>Table 4.5 The Simultaneous Inferences on the Battery Example using the Plug-in Approach with Linearly Interpolated Degrees of Freedom</td>
<td>176</td>
</tr>
</tbody>
</table>
4.6 The Satterthwaite Degrees of Freedom for each Marginal Null Distribution of the Test Statistic................................. 176

4.7 The Simultaneous Inferences on the Battery Example using the RPI Approach................................................................. 182

4.8 The Simultaneous Inferences on the Battery Example using the NORTA Approach.............................................................. 183

4.9 The Simultaneous Inferences on the Battery Example using the Resampling Approach.......................................................... 183
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Scatterplots of bivariate $t$ Random Variates with the Same Correlation, yet different degrees of freedom. The top left: bivariate $t$ with correlation $\rho_T = 0.5$ and degrees of freedom $\nu = 2$; The other plots are generated from bivariate $t$ with the same correlation $\rho_T = 0.5$ and $\nu = 3$ (the top right), $\nu = 8$ (the bottom left), and $\nu = 10$ (the bottom right).</td>
</tr>
<tr>
<td>2.2</td>
<td>The Red (Blue) Points denote the Pattern of Discrepancy between the Actual Tail Probabilities and the Pre-specified Level $\alpha$, using the Rounding Down (Linearly Interpolated) Degrees of freedom</td>
</tr>
<tr>
<td>2.3</td>
<td>An Illustration on the use of the Spherical Radial Transformation in Variance Reduction in the Multivariate $t$ Probability Calculation</td>
</tr>
<tr>
<td>2.4</td>
<td>The Left Panel: A Numerical Comparison on the amount of Variance Reduction Achieved by the Use of the two Implementations of the Spherical Radial transformation and the Table 0/1 Evaluation with Different Selections of the Matrix Factorization Methods, as the Number of Comparisons Increase. The Right Panel: Contrasts on the Accuracy of the Probability Estimates</td>
</tr>
<tr>
<td>2.5</td>
<td>A Simple Illustration of the Updating Process in the Quantile Calculation for a Classical Multivariate $t$ Distribution Using the Spherical Radial Transformation Technique, with Colors Representing the Different Sizes of the Small Tables Used in Combination with the Bisection Algorithm in Phase 1</td>
</tr>
</tbody>
</table>
2.6 The Four Plots Correspond to Four Generalized Bivariate $t$ Distributions with Common Gaussian Correlation $\rho_X = 0.5$ and Common Degrees of Freedom $\nu = (3, 3)$, with Gamma Correlation varies as such: $\rho_W = 0$ (Top Left), 0.5 (Top Right), 0.8 (Bottom Left) and 1 (Bottom Right).

2.7 The Updating Procedure in the Quantile Calculation for a Generalized Multivariate $t$ Distribution Using the Spherical Radial Transformation Technique, with Colors Representing the Different Sizes of the Small Tables Used in Combination with the Bisection Algorithm in Phase 1.

3.1 The Estimated Family-wise Error Rate for the set of All Pairwise Comparisons Among 8 Treatment Groups with Various Approximate Procedures.

3.2 The Estimated Family-wise Error Rate Produced by Various Approximate Procedures under Simulation Settings 2, 3, and 4.

4.1 The Upper Left (right) is the Residual Plot for the Battery Life per Unit Cost (Battery Life) data. The Lower Left (right) panel is the Residual Plot for the Square Root (Log) Transformed Battery Life Data.

4.2 Multiplicity-adjusted Critical Values on the Battery Life Data using Various Approximate Procedures.

4.3 Multiplicity-adjusted Critical Values on the Battery Life Data using the PI and the NORTA Methods.
Chapter 1: Introduction

There have been remarkable developments in statistical methodology over the past several decades. Statistical modeling typically has two main goals in many research endeavors, consisting of the understanding of the true mechanism and the prediction of a future observation. The 'best' model is often determined based on some model selection criterion, the choice of which is guided by the intention of the analysis. In other words, the model that yields the best fit of the data as a descriptive summary of the underlying mechanism may differ from the model that yields the best predictive performance for future observations. This dissertation focuses on the study of the impact on the inferences concerning the data generating mechanism in the form of hypotheses test of two modeling techniques for a specific situation where the data generating mechanism is a mix of separate processes.

The validity of the statistical inference hinges on the reliability of various stages of the modeling process including, but not limited to, data sampling, model fitting and model selection, and inferences on the model parameters representing the questions of interest regarding the true data generating mechanism. Multiple inference arise from the need of assessing several hypotheses simultaneously with some joint probability error rate. Among the wide spectrum of possible causes that may compromise the reliability of the resulting inference, the aim of this dissertation is to consider (1) the
effect of inappropriate choice of the simultaneous inference procedure, particularly in
the presence of heteroscedasticity, and (2) the impact on, for simplicity of discussion,
simultaneous inferences following from two different modeling strategies. In addition,
we develop a new multiple inference procedure and discuss the merits of alternative
method to address the problem of multiple hypotheses testing for heterogeneous data.

This chapter begins with a brief review on the history of the development of mul-
tiple testing procedures, under both homogeneous variance assumption and heteroge-
neous variance assumptions. Next, we describe a special but commonly encountered
scenario, under which we study the effect of modeling techniques on subsequence
inference. In particular, we assess several merits of the new modeling strategy we
propose, specifically used for this scenario.

1.1 Multiple Testing

The multiple testing problem is frequently encountered in the application of statis-
tical techniques in a wide variety of research fields. In contrast to the simple univariate
test of a single hypothesis, multiple testings involves making joint inferences on a col-
lection of individual hypotheses in order to address more than one scientific question
simultaneously. Perhaps the most common use of multiple testing procedures is the
comparison of means of a set of treatments or the comparison of a set of population
means. When phrased as a hypothesis test, these comparisons may be as simple as
equality of means or they may involve contrasts or other linear combinations of means.
The focus of these procedures is preservation of the type I error rate for the family
of comparisons. This may either be done in conservative fashion, guaranteeing (un-
der assumptions) that the family-wise error rate does not exceed some pre-specified
bound, or it may be less conservative, taking the pre-specified error rate as a target to be approximated.

1.1.1 Homogeneous Variances

In the early 20th century, many simultaneous inference procedures were introduced into the statistics literature with an attempt to account for the multiplicity effect and to ensure that the family-wise error rate is maintained at some joint significance level, say $\alpha$. These procedures often take the name of their inventor. Bonferroni and Sidak tests are well-known, as are some stepwise procedures derived from them. Several classical single-step multiple comparison procedures (MCPs) were developed in succession, including Scheffe’s method \cite{Scheffe_1953} for all linear estimable functions, Tukey’s all pairwise comparisons \cite{Tukey_1949}, Dunnett’s treatments versus control comparisons \cite{Dunnett_1955} and Hsu’s method \cite{Hsu_1984} for multiple comparisons with the best. Each of these procedures is designed for a specific family of linear comparisons for the population means, in the context of a basic one-way fixed effect analysis of variance (ANOVA) model with homogeneous variances. Theoretical results for the procedures sometimes also relies on equal sample sizes.

Scheffe’s method provides an exact $F$ test for the family of all linear estimable functions, without regard to the pattern of sample sizes across the studied populations. Scheffe’s method is phrased with great generality for the linear model and is routinely used in regression modeling. However, as was pointed out by Toothaker \cite{Toothaker_1991}, Scheffe’s method is among the least used of all multiple comparison procedures for a set of means. The reluctant of practitioners to use the method stems from the mismatch between the technique and their questions. Scheffe treats each member of
the family of all linear estimable functions symmetrically, placing no special emphasis on any individual hypothesis. In practice, particular subsets of comparisons are of greater or lesser interest, and the methods developed for these restricted families of comparisons are very popular.

Tukey’s method (Tukey, 1949) was proposed to simultaneously assess the family of all possible pairwise differences across the set of population (treatment) means, under a balanced one way ANOVA model with homogeneous variances. Shortly after, in the same modeling environment, Dunnett’s method (Dunnett, 1955) was established to identify population (treatment) groups that differ, in a directional manner or not, from the predetermined reference (control) group. As with a single test, exact implementation of Tukey’s and Dunnett’s procedures relies on appropriate calculation of the multiplicity-adjusted critical vector, evaluated as a multivariate quantile of the associated reference distribution. Before the computer era, the multiplicity-adjusted critical values were primarily available through tables. The tables exist only for special situations, for example, the studentized range distribution, associated with Tukey’s pairwise comparisons with equal sample sizes, and multivariate $t$ distributions with special symmetric correlation structures, associated with Dunnett’s procedure with equal sample sizes across all groups except the control group.

In real world applications, data arising from unbalanced experimental design are common, giving rise to the natural question of how to modify traditional multiple comparison procedures to accommodate flexible sample size patterns. In the late 20th century, a considerable literature was devoted to the extension of Tukey’s and Dunnett’s procedures to the general setting of an unbalanced design. Much effort has been spent on exploring numerical approximations to the multiplicity-adjusted critical
vector corresponding to a multivariate $t$ distribution with any arbitrary correlation
structure. Three popular conservative procedures currently available in most ma-
jor statistical software packages are the Tukey-Kramer procedure, the Miller-Winer
procedure and the Hochberg GT2 procedures. [Toothaker (1991)] provides a com-
prehensive review of these approximate procedures.

More recent work taps into modern computational power and focuses on a general
simultaneous inference procedure suitable for any finite collection of linear estimable
functions under the homogeneous variances assumption. ‘All pairwise differences’ and
‘all treatments versus control’ comparisons based upon a one way ANOVA model are
special cases of this general procedure, and so computation has replaced tables. These
computations can be performed because of the availability of probability calculations
for multivariate $t$ distributions. [Hsu (1992)] is the first to discuss numerical approx-
imation of multivariate $t$ probabilities for simultaneous inferences and proposes the
factor analytic approach that is exact for correlation matrix possessing a one-factor
structure in the factor analytic sense. [Genz and Bretz (2002)] present three Monte
Carlo algorithms, namely an acceptance-rejection sampling algorithm, an approach
based on the spherical-radial transformation, and a separation-of-variables approach.

These computational advances have settled into the discipline. Hsu’s one-factor
approximation is implemented in the Proc Mixed procedure in SAS. Genz and Bretz
algorithms are available in R the ‘pmvtf’ function in the ‘mvtnorm’ package pub-
lished by [Genz et al. (2017)]. However, the R implementation is restricted to the
multivariate $t$ with integer degrees of freedom. This is adequate for a selection of
multiple comparisons procedures, in particular, those based on homogeneous vari-
ance models, but is inadequate for various other applications where fractional degrees
of freedom are the norm. Moreover, increased accuracy for these Monte Carlo or quasi Monte Carlo methods is usually achieved at the cost of a dramatic increase in computational effort, especially when the dimension of multivariate $t$ exceeds 20 (Genz and Bretz, 2002). In the standard multiple hypotheses testing problems, the multiplicity-adjusted critical value is usually obtained as the equal-coordinate quantile vector using the ‘qmvtd()’ function, yielding the same individual error rate for each comparison in the family. In practical settings, it may be desirable to allow individual error rates to be inversely linked to the potential losses incurred as a consequence of erroneous inferences, while maintaining the family-wise error rate at a pre-specified level $\alpha$. Together, these practical limitations suggest the need for a new efficient multivariate $t$ algorithm that accommodates any number of degrees of freedom and that allows for variation in individual tail probabilities.

The appropriateness of these general simultaneous inference procedures depends largely on the validity of the constant variance assumption across all population groups. When data are heterogeneous in nature, it is important to establish an appropriate modeling strategy that captures the variance structure well enough to make reliable inferences. The old strategy for dealing with heteroscedasticity (Box and Cox, 1964; Carroll and Ruppert, 1981b) was to find a variance stabilizing transformation for the response variable with the remaining analysis performed on the derived variable in the homogeneous variance framework. The main disadvantage of this approach is that inference typically does not translate to inference on the mean of the original response variables. For many problems we envision, the strategy is inadequate for handling the quickly changing aspects of the problem that we will investigate in Chapter 4 and the associated decision problems therein. As an example, the model
proposed by Carroll and Ruppert (1982) where the variance structure is a function of the mean with some extra structural parameters does not address all of our goals.

1.1.2 Heterogeneous Variances

Welch (1938) proposed the first generally applied adjustment for comparing two normally distributed population means with heterogeneous variance. He suggests approximating the associated test statistic $T$ by a Student $t$ distribution with the Welch-Satterthwaite approximated degrees of freedom. The degrees of freedom are approximated by matching the first two moments of $T$ (under the null hypothesis) to those of a $t$ distribution. Subsequently, Satterthwaite (1946) extended this idea to a linear combination of means from multiple populations. These approximations have been shown to work well for single comparisons, and the nonpooled $t$ test and associated confidence interval play a prominent role in applications.

To date, the Welch-Satterthwaite techniques do not fare as well in the multiple comparisons setting. Over the past 40 years, the classical multivariate $t$-based simultaneous inference procedures have been modified for the case where the population variances are independently estimated by their corresponding sample analogues in the basic one-way ANOVA model setting. Games and Howell (1976) suggests using comparison-specific critical values for any finite family of linear combinations of the means, calculated from a sequence of multivariate $t$ distributions with common correlation matrix, as in the homogeneous case, and each with its own degree of freedom according to Satterthwaite (1946). More recently, Hasler and Hothorn (2008) improves upon the Games-Howell procedure by replacing the homogeneous correlation
by a correlation structure with plugged-in sample variances. An alternative general-
ization is proposed by Herberich et al. (2010) that uses a heteroscedastic-consistent
sandwich matrix estimator.

Unlike the homogeneous situation for which computational difficulties hindered
our ability to develop a general simultaneous inference procedure before the availabil-
ity of routines to compute classical multivariate $t$ integrals, exact procedures are not
available for heterogeneous data, because the theoretical reference distribution of the
multivariate test statistic is unknown. This adds to the challenge of constructing the
multiplicity-adjusted critical vector that guarantees an overall family-wise error rate
at level $\alpha$. Instead, procedures are developed to achieve an approximate family-wise
error rate of $\alpha$ for the test.

We propose the use of a class of multivariate $t$ distributions, termed as the gen-
eralized multivariate $t$ distributions, that includes the classical multivariate $t$ as a
special case. The new class allows an arbitrary correlation structure for a vector of
Gaussian variates with standard normal marginal distributions and also allows an
arbitrary correlation structure and shape parameters for a vector of gamma (scaled
square root of chi-square) variates. The Gaussians provide the numerators; the gam-
mas the denominators, in the usual construction of the $t$ distribution (Student, 1908).

This class of $t$ distributions has the flexibility to accommodate dependence (or lack
thereof) among the linear combinations of means that define the family of inferences.
It also has the flexibility to capture dependence in the estimators of the variance of
these linear combinations.

In addition to novel use of the generalized multivariate $t$ distributions, we develop
an efficient numerical algorithm to compute the probability content of a convex set,
with particular attention to rectangles centered at the origin. This efficient probability calculation is then used to create a routine to find quantiles that go with the flexible multivariate $t$ distribution.

1.2 Separation of the Quickly Varying and Slowly Varying

Since the early 20th century, statisticians have developed a powerful array of statistical tools and procedures that could help understand the mechanism that produces the data, ranging from exploratory data analysis to selection of a family of models and parameter estimation to hypothesis tests. The preceding discussion, details of which will be presented in Chapter 2 and Chapter 3, shows how the validity of the statistical inference can be compromised using an inference procedure that is suboptimal for a given set of hypotheses of interest, in the context of a simple one way ANOVA setting with heterogeneous variances. In addition to the inappropriate choice of inference procedure, another source of unreliability of the resulting inference is due to possible unsuitable modeling strategy, on the basis of which the inference procedure will be performed.

Statistical inference is traditionally performed assuming the model has a known form with a set of working assumptions, for example, presuming that the relationship between an explanatory variable and the mean of the response variable is linear, that observations are independent and that the conditional distribution of the response given the explanatory variable is Gaussian that the variances of two (or more) Gaussian distributions are identical and so on. The assumptions that comprise the model are unlikely to be true, but they often provide a good approximation to reality and/or do little harm to subsequent inferential procedures. At other times, the conflict with
reality in such a way that inference suffers: parameter estimates may be badly bi-
ased, hypothesis tests may not hold their level and the actual coverage of confidence
intervals and prediction intervals may stray far from the nominal coverage.

Many researchers have studied the impact of different modeling strategies, as an
extension of the study of model misspecification, on statistical inference, in both
frequentist and Bayesian framework. Chapter 4 focuses on the investigation of the
influence on the ensuing inferences of different modeling strategies, when the data
generating process that yields the outcome variable of interest is, upon close inspec-
tion, a hybrid of two qualitatively separate portions. One portion is usually stable and
amenable to statistical modeling, while the other is unstable and subject to seemingly
erratic variation. The subsequent discussion is guided by the following research ques-
tion: How do the inferences differ by modeling the mixture of the stable and unstable
processes than modeling the stable process with unstable process being accounted for
in the inference?

As an example, consider a manufactured consumer product. The manufacturing
process is closely monitored with a variety of quality control techniques that are
designed to produce stable output – a stream of products for which measurements on
sample products resemble independent draws from some distribution. This portion
of the process is stable and amenable to statistical modeling. The resulting model
should be useful for a long time.

The consumer faces a decision problem which includes an additional important
element – price. The products price is set in a competitive environment. It is subject
to rapid fluctuations as the variety of vendors monitor and control their inventory,
seek to expand their business with lower profit margins and offer special pricing and

10
sale prices. These price fluctuations are fast and often substantial, and are the result of a dynamic environment. As a consequence, they are difficult to model and any model developed for them must be constantly monitored and modified to ensure that it remains relevant.

This dissertation advocates modeling for the stable portion of the problem, with full use of traditional tools and move the unstable portion of the problem to inference. This necessitates the development of novel inference strategist and it suggests a broader view of inference. Rather than a single inferential statement, a family of statements is made with inference available for each specific variation of the unstable portion of the problem. This provides the analyst with the ability to react quickly to a changing environment.

A classic data set on battery lifetime (Dean and Voss, 1999) will be analyzed to demonstrate the different inferential conclusions drawn from the two modeling strategies. The data consists of measurements on the useful lifetime of four types of batteries along with the prices for the batteries. The main question is which type of battery delivers the largest mean life per dollar. Previous analyses have blended manufacturing and pricing to produce a single response for analysis. In contrast to this approach, we propose to model the stable portion of the problem – and push the unstable portion of the problem – price – into the inference.

1.3 Overview of Dissertation

The rest of the dissertation is organized as follows. Chapter 2 provides a review of the classical multivariate t distribution, focusing primarily on the development of an efficient numerical integration algorithm for the probability calculation as well
as quantile evaluation of the classical multivariate $t$ distributions. The numerical algorithm allows the individual error rates to vary in a proportional manner, while controlling the joint tail area to be at a pre-determined rate. In addition, a new class of multivariate $t$ distributions (to be referred to as the generalized multivariate $t$ distributions) is developed that prepares us for the investigation of a multiple inference procedure in the presence of heterogeneity, as a modification of the established Plug-in procedure. A numerical algorithm for the generalized multivariate $t$ distributions is developed that extends the numerical algorithm for the classical $t$.

Chapter 3 reviews the simultaneous inference procedure under homogeneous variance assumption in the general linear model framework, along with a numerical implementation using our efficient probability and quantile calculation approach. Afterwards, an extensive review on some currently used multiple inference procedures in the presence of heterogeneity is provided, with emphasis on the Plug-in method of Hasler and Hothorn (2008) that is used as the benchmark for the new approach we propose in this chapter. In Chapter 4 we examine the impact on the statistical inference of different modeling strategies, in the special case where the underlying data generating process is known to be a hybrid of a stable process and an unstable process. The two modeling strategies, with different treatments of the unstable process, lead to notably different inferential conclusions. The ideas will be applied to the battery lifetime dataset introduced in Dean and Voss (1999). Chapter 5 provides a conclusion of the ideas presented in the previous chapters with discussion on the implications of this work in some more sophisticated statistical modeling paradigms. In this chapter, future work that is interesting to explore further on this topic will be outlined.
Chapter 2: Multivariate $t$-Distribution

2.1 Classical Multivariate $t$ Distribution

The classical univariate student $t$ distribution was introduced into the statistics literature by Student (1908). The original derivation of the student $t$ distribution, or simply the $t$ distribution, is to provide an appropriate reference distribution for a small-sample hypothesis test of a normal mean.

Assume that $\{X_1, \ldots, X_n\}$ is a random sample of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then, with $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$, $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ (2.1) follows a $t$ distribution with $n - 1$ degrees of freedom, denoted by $t_{n-1}$. For a test of $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$, the test statistic is given by (2.1), provided that $\mu_0$ is substituted for $\mu$. The $t_{n-1}$ distribution is used to find the $p$-value and reach a conclusion.

The $t$ distribution with $\nu$ degrees of freedom can be alternatively defined as the ratio of a standard normal variate and the square root of an independent chi-square variate with $\nu$ degrees of freedom over its degrees of freedom, as is suggested by a
factorization of the student \( t \) in (2.1) as

\[
T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S} = \frac{Z}{\sqrt{U/\nu}},
\]

(2.2)

where \( Z \) in (2.2) is a standard normal and \( U \) is a chi-square with \( \nu \) degrees of freedom – and also a gamma variable with mean \( \nu \) and variance \( 2\nu \). The probability density function (p.d.f.) for the \( t_\nu \) distribution follows from this representation and is

\[
f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{1/2}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.
\]

(2.3)

It can be easily proved that a \( t_\nu \) random variable \( T \) has mean 0, provided \( \nu > 1 \) and variance \( \frac{\nu}{\nu-2} \), provided \( \nu > 2 \).

Many variants of the \( t \) distribution exist. A typical example is the location-scale \( t \) distribution parameterized by three parameters - a location parameter \( \mu \), a scale parameter \( \sigma \) and the degrees of freedom \( \nu \). A location-scale \( t \) variate \( X \) is related to the standard \( t \) variate \( T \) through

\[
X = \mu + \sigma T,
\]

given \( T \sim t_\nu \). In addition, a two parameter non-central \( t \) distribution produced by replacing the standard normal \( Z \) in (2.2) with a normal variable with mean \( \mu \neq 0 \) and variance 1 is the second most commonly used variation of the \( t \) distribution. A non-central \( t \) variate \( X \) can be represented as

\[
X = \frac{Z + \mu}{\sqrt{U/\nu}}.
\]

Since its development, the univariate \( t \) distribution has found widespread use as a model for a heavy-tailed distribution and outlier-prone data. The \( t \) distribution, when used for errors, provides one approach to robust regression.
The multivariate $t$ distribution emerges as a natural generalization of the univariate student $t$ distribution. It is constructed as the ratio of a Gaussian random vector with standard normal marginal distributions scaled by an independent vector of gamma distributed scaling factors. The multivariate $t$ distribution exists in the literature in a number of different forms, depending upon the specifications of the underlying gamma scaling factors. Among the various available forms, the classical multivariate $t$ defined using a single gamma variable has received the most attention. As with the univariate $t$, the multivariate $t$ figures prominently in the literature on hypothesis testing when several univariate tests are performed simultaneously. The multivariate $t$ distribution has also been widely applied in classical as well as Bayesian statistical modeling frameworks. Analogous to the univariate case, the multivariate $t$ distribution captures thicker-tailed behavior than does its Gaussian counterpart.

The validity of the classical multivariate $t$ as a reference distribution for multiple hypothesis test relies on homogeneity of variances. Nevertheless, in many settings, populations show heterogeneity of variances, leading to such procedures as the Welch-Satterthwaite approximation and non-pooled $t$ test, which relies on an approximate univariate $t$ reference distribution. Heterogeneity of variances also occurs in the multiple testing setting and requires a suitable reference distribution — again, a version of a multivariate $t$ distribution, but not the classical $t$ distribution.

The next section describes the classical multivariate $t$ distribution in detail and shows how it is connected to the multiple testing problem. Computational algorithms are devised to compute critical values and to find tail areas. These algorithms are compared to existing computational methods. The subsequent section places special emphasis on broadening the scope of the classical $t$ to a more flexible class which
handles multiple testing in the presence of heterogeneous variances, along with the
development of effective computation for these more general \( t \) distributions. Com-
parisons of their use with methods based on the classical multivariate \( t \) are provided.

The joint probability density function of a \( p \)-variate classical \( t \) random variable
\( \mathbf{T} = (T_1, \ldots, T_p)' \) is given by,
\[
f(\mathbf{t}) = \frac{\Gamma\left(\frac{\nu + p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu \pi)^{p/2} \left|\Sigma_X\right|^{1/2}} \left(1 + \frac{1}{\nu}(\mathbf{t} - \mathbf{\mu})\Sigma_X^{-1/2}(\mathbf{t} - \mathbf{\mu})\right)^{-\frac{\nu + p}{2}} \\
\tag{2.4}
\]
where \( \mathbf{\mu} \in \mathbb{R}^p \) is the mean vector, \( \Sigma_X \) is a \( p \times p \) positive definite covariance matrix for
the associated Gaussian random vector and \( \nu \) is the degrees of freedom for the gamma
scaling variable. This form is for the location-scale family with location vector \( \mathbf{\mu} \) and
scale matrix \( \Sigma_X^{1/2} \).

Nonetheless, the classical multivariate \( t \) distribution is legitimately defined for
any non-negative definite covariance matrix \( \Sigma_G \geq 0 \), except that the p.d.f. is not
available when \( \Sigma_X \) is singular. The univariate student \( t \) is a special case with \( p = 1 \).
As was mentioned earlier, the multivariate \( t \) can be alternatively formulated as a
scaled mixture of a Gaussian random vector \( \mathbf{X} \) and a common gamma distributed
scaling factor \( W \). Suppose that \( \mathbf{X} \) has a non-negative definite covariance matrix \( \Sigma_G \)
with rank \( q \leq p \) factorized as
\[
\Sigma_X = \mathbf{C} \mathbf{C}^T,
\]
where \( \mathbf{C} \) is a \( p \times q \) matrix of full column rank. For positive definite correlation \( \Sigma_X \), the
Cholesky decomposition and spectral decomposition are commonly taken approaches
to find the matrix factor \( \mathbf{C} \). For a rank-deficient \( \Sigma_X \), the factorization can be achieved
through spectral decomposition
\[
\Sigma_X = \mathbf{P} \mathbf{\Omega} \mathbf{P}^T,
\]
16
where \( P \in \mathbb{R}^{p \times q} \) consists of the \( q \) leading eigenvectors corresponding to the \( q \) non-zero eigenvalues and \( \Omega \in \mathbb{R}^{q \times q} \) is a diagonal matrix with eigenvalues as its diagonal entries. Clearly, setting \( C = P\Omega^{1/2} \) leads to the decomposition \( \Sigma_X = CC^T \). Consequently, the original multivariate \( t \) can be constructed by means of the stochastic representation as follows

\[
T \overset{d}{=} \mu + \frac{CZ}{\sqrt{W/\nu}},
\]

(2.5)

where \( \overset{d}{=} \) denotes equal in distribution, \( Z = (Z_1, Z_2, \ldots, Z_p)^T \) is a \( q \)-dimensional random vector with \( Z_i, 1 \leq i \leq q \) being independent standard normal random variables, \( W \sim \chi^2_\nu \) independent of \( Z \) and \( \mu \in \mathbb{R}^p \) is the mean vector. The probability density function of \( T \) in (2.5) can be trivially shown to coincide with the original definition in (2.4) with tedious derivation. Moreover, a closer look at the stochastic representation in (2.4) reveals that the tail characteristics are predominantly governed by the degrees of freedom parameter \( \nu \) for the scaling variable \( W \), with larger \( \nu \) associated with less heavy tails. Also, the well established result that the multivariate \( t \) distribution converges to a Gaussian distribution, as the degrees of freedom \( \nu \) tends to infinity can be seen from (2.5). Figure 2.1 displays 50,000 independent draws of bivariate \( t \) distributions with common correlation \( \rho = 0.5 \) and four different degrees of freedom. The elliptical shape of the scatterplots, e.g., the principal direction is determined by the correlation structure of the two marginals. The tail behavior is primarily governed by the degrees of freedom with smaller \( \nu \) tied to heavier tails. The density plot starts to resemble the bivariate normal density as the degrees of freedom grows larger. The stochastic representation of the multivariate \( t \) provides a convenient way to investigate theoretical and numerical properties of the distribution.
The classical multivariate $t$ random vector $T$ has a covariance matrix $\Sigma_T$ related to the underlying Gaussian covariance matrix $\Sigma_X$ and the degrees of freedom $\nu$ in an explicit form

$$\Sigma_T = \frac{\nu}{\nu - 2} \Sigma_X,$$

(2.6)

provided $\nu > 2$. This is readily shown by taking iterated expectations in (2.5). In addition, when $\nu > 2$, a distinguishing feature of the classical $t$ is that the correlation matrix $R_T$ for the multivariate $t$ is identical to that of the underlying Gaussian $R_X$, without regard to the extra parameter for the gamma scaling random quantity, i.e., $R_T = R_X$. Kotz and Nadarajah (2004) describes many of the properties of the multivariate $t$.

Furthermore, the stochastic representation (2.5), together with matrix decomposition of $\Sigma_X$ allows for a direct sampling strategy of a classical $t$ variate, which can easily be implemented by constructing $T$ from independently generated standard Gaussian random vectors $Z$ and gamma random variable $W$.

A basic question that frequently arises in statistical analysis is how to compute the probability for a convex set $A$ under the multivariate $t$ distribution. This calculation lays the groundwork for the theoretical understanding and practical application of the distribution. For instance, the probability calculation over hyper-rectangles is heavily involved in its typical application in multiple testing, for both the multiplicity-adjusted critical value calculation and the power calculation. There has been a fair amount of research related to the evaluation of multivariate $t$ probabilities for various types of set $A$, e.g. ellipsoidal region, rectangular region, polygon and any general convex region. Nadarajah and Kotz (2008) provide a comprehensive review of both historical and contemporary results.
Figure 2.1: Scatterplots of bivariate $t$ Random Variates with the Same Correlation, yet different degrees of freedom. The top left: bivariate $t$ with correlation $\rho_T = 0.5$ and degrees of freedom $\nu = 2$; The other plots are generated from bivariate $t$ with the same correlation $\rho_T = 0.5$ and $\nu = 3$ (the top right), $\nu = 8$ (the bottom left), and $\nu = 10$ (the bottom right).
Computational strategies include approximate numerical integration and Monte Carlo integration. Several Monte Carlo algorithms are presented and carefully compared with a number of tests in Genz and Bretz (2002) for multivariate $t$ probabilities over rectangular regions. All methods face a trade off between approximation accuracy and computational speed, with the cost of a notably increased computational effort in exchange of higher accuracy when the dimension $p$ reaches 20 or beyond. Computer implementation in multivariate $t$ probability computation using the ‘pmvt’ function in ‘mvtnorm’ package contributed by Genz et al. (2017) is the most common practice. Yet, its use is restricted to multivariate $t$ distributions with integer degrees of freedom over rectangular regions. In particular, in the application of multivariate $t$ distributions in multiple hypothesis testing, a common choice is to replace the multivariate $t$ with fractional degrees of freedom by one with the degrees of freedom rounded down to the nearest integer. The accuracy of the resulting inferences with this approximation strategy is most troubling in situations where there are relatively few degrees of freedom. Figure 2.2 shows a systematic departure from the target level $\alpha$ of the tail probability of the quantile, calculated from the multivariate $t$ with the rounded down degrees of freedom, evaluated at a true multivariate $t$ with fractional degrees of freedom. As an example, the red dot shows that the tail probability of a multivariate $t$ with $\nu = 2.9$ corresponding to the 95th quantile of a multivariate $t$ with $\nu = 2$, significantly deviates from 0.05. This discrepancy is due to the inflated quantiles caused by the heavy tailed behavior of multivariate $t$ with smaller degrees of freedom, the impact of which is in turn transmitted to that on the tail probabilities.

The general trend depicts the diminishing effect of the rounding strategy as the degrees of freedom increase. As it varies between any two adjacent integers, a larger
discrepancy tends to occur at bigger step down to the nearest integer. An intuitive explanation behind this phenomenon is that the discrepancy between two multivariate $t$ distributions with degrees of freedom differ by a fixed amount becomes less discernible as the degrees of freedom grow larger, indicating that the rounding strategy is less concerning at larger degrees of freedom.

Figure 2.2 indicates that the ‘random adjustment’ is inadequate for many uses of the multivariate $t$ distribution, especially when small fractional degrees of freedom
are the norm. It provides a motivation for the development of a probability integral estimation strategy that allows for (1) any real-valued degrees of freedom over its feasible set \( \nu > 1 \), and (2) any convex set \( A \).

In the next section, we develop a basic acceptance-rejection sampling method for a crude multivariate \( t \) probability calculation over any convex set \( A \in \mathbb{R}^p \). The flexibility of this method comes at the cost of computational efficiency. In the subsequent section, a more efficient methodology for high accuracy computation is introduced. This strategy also extends effortlessly to other variations of the multivariate \( t \). For simplicity, this dissertation work is, without loss of generality, confined to the discussion of the central multivariate \( t \) with the mean vector set at the origin, i.e., \( \mu = 0 \), unless otherwise noted. Extensions to other forms of multivariate \( t \) are straightforward.

### 2.1.1 Table 0/1 Approach: Probability Calculation

To numerically approximate the probability content of a set \( A \) under the \( p \)-variate classical \( t \) distribution parameterized by \( \Sigma \) and \( \nu \), a direct Monte Carlo integration provides the simplest way. With this method, a large number of variates, say \( t_k, k = 1, \ldots, N \), are independently generated from the \( t \) distribution. Each draw from the \( t \) distribution either falls in the set \( A \) or does not. The estimate of \( P(A) \) is the fraction of generated variates that lie in \( A \). Formally,

\[
P(A) = \int_A f(t \mid \Sigma, \nu) dt
\]

and

\[
\hat{P}(A) = N^{-1} \sum_{k=1}^{N} I(t_k \in A),
\]

where \( I(\cdot) \) is the indicator function.
To set the stage for further development, we outline how the multivariate \( t \) variate can be generated from more basic distributions. The covariate matrix, \( \Sigma_T \), can be decomposed as \( \Sigma_T = CC^T \), where \( C \) is a \( p \times q \) matrix. The stochastic representation (2.5) has
\[
T = \frac{Cz}{\sqrt{W/\nu}},
\]
where the \( q \)-vector \( z \) consists of \( q \) independent standard Gaussian variates and \( W \) is an independent chi-square variate with \( \nu \) degrees of freedom.

When the set \( A \) is a hyper-rectangle with sides parallel to the coordinate axes, the integral in (2.7) can be rewritten. Let \( A = \{ t \mid a_i < t_i < b_i, \ i = 1, \ldots, p \} \). Then,
\[
P(T \in A) = P(a_i \leq T_i \leq b_i; \ i = 1, \ldots, p) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_p}^{b_p} f(t_1, t_2, \ldots, t_p) \, dt_1 \, dt_2 \cdots dt_p.
\]
(2.10)
The integral in (2.10) is estimated as,
\[
\hat{P}(A) = N^{-1} \sum_{k=1}^{N} I(t_k \in A)
\]
\[
= N^{-1} \sum_{k=1}^{N} \prod_{i=1}^{p} I(t_{ki} \in (a_i, b_i)).
\]
(2.11)
The standard error associated with this estimate is that of a binomial proportion.
\[
\hat{\sigma}_{Table,N}^2 = \frac{\hat{P}(A)(1 - \hat{P}(A))}{N}.
\]
(2.12)

We note that this direct Monte Carlo approach works for singular multivariate \( t \) distributions such as those which appear for Tukey’s family of all pairwise comparisons across \( l \) (\( l \geq 3 \)) treatment groups. In these cases, \( q = l - 1 \) and \( p = \binom{l}{2} \). Clearly, the correlation matrix \( \Sigma_T \) is not of full rank and the density in (2.10) does not exist. Nevertheless, the estimate in (2.11) is a valid estimate of \( P(A) \).
We find it helpful to envision the indicators arrayed in a two-way table. The table consists of \( p \) rows and \( N \) columns. The \( (i,k) \) entry is 0 if \( t_{ik} \not\in (a_i, b_i) \) and is 1 if \( t_{ik} \in (a_i, b_i) \). The \( k^{th} \) entry in the margin at the bottom of the table is 1 if \( t_k \in A \) and 0 if not. \( \hat{P}(A) \) is the fraction of ones in this margin.

As a simple example, the 0/1 table for a set of \( m = 3 \) comparisons is listed in Table 2.1. Each column, say \( k \), is created independently by performing a vectorized comparison of a random draw \( t_k \) with the \( k \)-dimensional integration region \( A \). For column \( k \), each entry \( i \) is set to be 1 if the \( i^{th} \) component of \( t_k \) is outside of the integration region along the \( i^{th} \) coordinate, i.e., \( t_{ik} \not\in (a_i, b_i) \), \( i = 1, 2, 3 \).

### 2.1.2 Table 0/1 Approach: Quantile Calculation

Calculation of quantiles of a distribution is a near dual to the problem of probability calculation. Instead of defining a set \( A \) and searching for \( P(A) \), one sets a target probability, say \( \alpha \), and searches for the set \( A \) lying in a well-defined class for which \( P(A) = \alpha \). For example, to obtain an exact set of level \( (1 - \alpha) \) - simultaneous confidence intervals for the family of all linear combinations of the component means \( \mu \in \mathbb{R}^p \), Scheffe’s method focuses on sets \( A \) defined by a quadratic form centered at
\( \mu \). Formally, the set is

\[
A_{\alpha} = \{ t \in \mathbb{R}^{p} | (t - \mu)^{T} \Sigma_{X}^{-1} (t - \mu) \leq p \hat{\sigma}^{2} Q(\alpha) \} \tag{2.13}
\]

where \( \mu \) is the center of the ellipsoidal contours.

In a univariate setting, the sets \( A \) are of the form \([-\infty, q]\) and are easily matched to the cumulative distribution function. Alternatively, with an eye toward use in hypothesis tests and confidence intervals, the set \( A \) may be of a different form, say \((\mu - q, \mu + q)\) for a \( t \) distribution in the location-scale family with center \( \mu \).

In this multivariate setting, we focus on families of sets tied to the family of all linear combinations \( l^{T} \mu, l \in \mathbb{R}^{p} \). The sets are chosen in this fashion because the multivariate \( t \) is a member of the class of elliptical distributions, for which the probability density function takes the form

\[
p(t|t \in \mathbb{R}^{p}) = f \left( (t - \mu)^{T} \Sigma_{X}^{-1} (t - \mu) \right) \tag{2.14}
\]

Furthermore, the probability density function \( f(\cdot) \) in (2.14) is decreasing in the quadratic form \((t - \mu)^{T} \Sigma_{X}^{-1} (t - \mu)\). This monotonicity leads to an easy search for the set \( A_{\alpha} \) and corresponding ‘quantile’ \( Q(\alpha) \) for which \( P(A_{\alpha}) = \alpha \).

Hochberg and Tamhane (1987) show that

\[
\sup_{l \in \mathcal{L}^{p}} \frac{l^{T}(\hat{\mu} - \mu)}{\hat{\sigma} \sqrt{l^{T} \Sigma_{\mu} l}} \stackrel{d}{=} \frac{(\hat{\mu} - \mu)^{T} \Sigma_{\mu}^{-1} (\hat{\mu} - \mu)}{p \hat{\sigma}^{2}} \sim F_{p,N-p}, \tag{2.15}
\]

provided that \( \mathcal{L}^{p} \) is any \( p \)-dimensional linear subspace. In addition, for each \( l_{k} \in \mathcal{L}^{p} \)

\[
\frac{l_{k}^{T}(\hat{\mu} - \mu)}{\hat{\sigma} \sqrt{l_{k}^{T} \Sigma_{\mu} l_{k}}} \overset{d}{=} \frac{X_{k}}{\sqrt{W/(N-p)}} \overset{d}{=} t_{N-p},
\]

where \( X_{k} \sim t_{N-p} \) and \( W \sim \chi_{N-p}^{2} \). Thus, the joint test statistic

\[
T_{\infty} = \begin{cases} 
  l_{k}^{T}(\hat{\mu} - \mu) \\
  \hat{\sigma} \sqrt{l_{k}^{T} \Sigma_{\mu} l_{k}} 
\end{cases} , \quad l_{k} \in \mathcal{L}^{p}
\]
is viewed as an infinite dimensional multivariate \( t \) random vector with pairwise correlation given by

\[
\rho_T(k, j) = \frac{l_k^T \hat{\Sigma} l_j}{\sqrt{l_k^T \hat{\Sigma} l_k \cdot l_j^T \hat{\Sigma} l_j}}
\]

We see that \( Q(\alpha) \) can be equivalently obtained by

\[
P \left( \frac{X^T \Sigma^{-1} X / p}{W / (N - p)} \leq \frac{Q(\alpha)}{p} \right) = \alpha. \tag{2.16}
\]

The standard result that \( \frac{X^T \Sigma^{-1} X / p}{W / (N - p)} \sim F(p, N - p) \) leads immediately to Scheffe’s result that the collection of \( t \in \mathbb{R}^p \) satisfying (2.16) forms an ellipsoid that contains 100\( \alpha \) percent of the probability mass of the multivariate \( t \).

In contrast to Scheffe’s method, the more recent multiple comparison procedures focus on a finite subset of inferences. The multivariate \( t \) is set up so that each component of the vector is linked to an individual comparison. This may yield a distribution with full rank \( \Sigma_X \) (and also \( \Sigma_T \)), as for Dunnett’s family of all treatments versus control, but more typically results in a singular \( \Sigma_X \) (and also \( \Sigma_T \)), as for Tukey’s family of all pairwise comparisons. In either event, each individual comparison is based on a univariate \( t \) and the sets \( A \) are hyper-rectangles.

Hyper-rectangles aligned with the coordinate axes allow for great flexibility in defining the family of sets \( A \). They allow one to ‘spend’ \( \alpha \) (\( \alpha \) being here the joint type I error probability) for early looks at the data, as is done in the sequential testing literature for clinical trials. They allow one to take the impact of type I (and type II) errors into account through adjustment of individual error rates in asymmetric fashion. But most commonly, they allow one to treat each comparison in the family symmetrically by setting all of the individual type I error rates to a common value.
In the general multivariate setting, we focus on the family of hyper-rectangles $A$ tied to a finite collection of linear combinations $l_k^T \mu$, $k = 1, \ldots, p$. Using the notation introduced earlier, the $100\alpha^{th}$ quantile of a $t$ distribution $T \sim MVT_p(\Sigma, \nu)$ is any $p$-variate vector $Q(\alpha)$ that satisfies,

$$\alpha = P_p(|T_i| < Q_i(\alpha), \ i = 1, \ldots, p)$$

(2.17)

where $Q(\alpha) = (Q_1(\alpha), Q_2(\alpha), \ldots, Q_p(\alpha))^T$ is a vector-valued function of some prespecified probability $\alpha$. We call each component $Q_i(\alpha)$ the $i^{th}$ coordinate quantile. Unlike the family of ellipsoidal-shaped sets $A$ for which the collection of quantile vectors $Q(\alpha)$ forms the boundary surface of certain confidence ellipsoids centered at the origin, the quantile vector $Q(\alpha)$ is not uniquely defined for the family of hyper-rectangles $A$. To ensure uniqueness, an additional set of constraints is added to the coordinate quantiles $Q_i(\alpha), i = 1, \ldots, p$, as

$$P(|T_1| > Q_1(\alpha)) = \kappa_i P(|T_i| > Q_i(\alpha)), \ i = 2, \ldots, p,$$

(2.18)

with given $\kappa_i > 0, \forall i$. Clearly, each of these $p - 1$ added constraints sets a monotonically increasing functional relationship between $Q_1(\alpha)$ and $Q_i(\alpha), \forall i = 2, \ldots, p$, leaving $Q_1(\alpha)$ the only unknown quantity to be determined for a target probability $\alpha$. Often all of the $\kappa_i$ are chosen to be $1, \ i = 2, \ldots, p$.

In the application to multiple comparisons, each constraint in (2.18) is interpreted as the control over the individual error rate for the $i^{th}$ comparison relative to that of the $1^{st}$ comparison. Setting the constraints this way provides a plausible way of more effectively controlling the overall loss by allocating less error rate to individual hypotheses associated with higher potential risk/loss.
Henceforth, the critical vector $Q(\alpha) \in \mathcal{R}^p$ satisfying (2.17) under the constraints stated in equation (2.18) is associated with the family-wise error rate at a pre-specified level $\alpha$, with relative magnitudes of individual type I error rates specified by $\kappa_i, i = 2, \ldots, p$. As the closed from expression of $Q_1(\alpha)$ is not theoretically accessible, our next effort is concentrated on an algorithmic search for $Q_1(\alpha)$.

Suppose that a random sample of size $N$ is generated according to the $t$ distribution and arranged into a $p \times N$ two way table $T_N$, the columns of which represent sample points $t_i \in \mathcal{R}^p, 1 \leq i \leq N$. For any given $\alpha$, let $c_1N(\alpha)$ denote the number of sample points $t_i$, whose 1st coordinate values exceed the unknown $Q_1(\alpha)$ in absolute terms, i.e., $|t_{i1}| > Q_1(\alpha)$. Formally,

$$c_1N(\alpha) = \sum_{i=1}^{N} I(|t_{i1}| > Q_1(\alpha)).$$

(2.19)

Our iterative algorithm aims at numerically solving for the $p$-dimensional quantile vector $Q(\alpha)$ satisfying (2.17) and (2.18) through searching for $c_1N(\alpha)$.

For any $c_1^kN(\alpha)$ produced in the $k^{th}$ iteration, the 1st coordinate quantile $Q_1^k(\alpha)$ is obtained as the $(c_1^kN(\alpha))^{th}$ largest element in the 1st row of the table $T_N$. In addition, the additional constraints in (2.18) yield

$$c_{iN}^k(\alpha) = \left[ \frac{c_{1N}^k(\alpha)}{\kappa_i} \right], \quad i = 2, \ldots, p,$$

(2.20)

where $[\cdot]$ symbolizes the operation of rounding to the nearest integer. The remaining coordinate quantiles $Q_i^k(\alpha)$ are, therefore, determined as the $(c_{iN}^k(\alpha))^{th}$ largest element in the $i^{th}$ row of the table $T_N$.

To start the algorithm, a reasonable initial guess for $c_1^0N(\alpha)$ can be picked in one of a number of appropriate ways. First, the Bonferroni idea assumes that the $p$ sets
\[ S_j = \{ i : |t_{ij}| > Q_j(\alpha) = 1 \} \] are mutually disjoint, i.e.,

\[ S_j \bigcap S_k = \emptyset, \quad \forall j, k. \quad (2.21) \]

The total number of sample points with at least one entry falling outside of the target hyper-rectangle \( Q^0(\alpha) \) is

\[ \sum_{i=1}^{p} \frac{c^0_{1N}(\alpha)}{\kappa_i}, \]

yielding

\[ c^0_{1N} = \left[ \frac{N(1 - \alpha)}{\sum_{i=1}^{p} 1/\kappa_i} \right]. \quad (2.22) \]

Alternatively, assuming independence across the \( p \) marginal variables leads to \( c^0_{1N}(\alpha) \),

\[ \alpha = \prod_{i=1}^{p} \left( 1 - \frac{c^0_{1N}(\alpha)}{\kappa_i N} \right). \quad (2.23) \]

Both initializations \( c^0_{1N}(\alpha) \) typically lead to the enclosure of \( Q^0(\alpha) \) with probabilities higher than the preset value \( \alpha \), i.e.,

\[ P \left( |T_i| < Q^0_i(\alpha), \varepsilon i = 1, \ldots, p \right) \geq \alpha. \quad (2.24) \]

The Bonferroni method guarantees that the empirical probability from the table is at least \( \alpha \), up to minor rounding adjustment in (2.22). No such guarantee exists for the independence method, as even a large table is merely a collection of vector \( t_i \) that can exhibit an unusual pattern.

For initializations with too large a quantile, a step down adjustment on \( c^0_{1N}(\alpha) \) is needed until the enclosure probability deviates from \( \alpha \) by no more than an absolute error tolerance parameter \( \varepsilon \). On the other hand, a conservative initialization can be set up by assuming maximum overlap over \( S_j, \ j = 1, \ldots, p \). In other words, the largest set \( S_j \) has size \( N(1 - \alpha) \) and includes all other \( p - 1 \) sets. This yields,

\[ c^0_{1N}(\alpha) = \left[ N(1 - \alpha) \min_{i=1,\ldots,p} \kappa_i \right]. \quad (2.25) \]
Accordingly, a step up adjustment is required.

To alleviate the computational burden, an updating formula for $c_{1N}^k(\alpha)$ with a large increment without overshooting the target $c_{1N}(\alpha)$ is given as

$$c_{1N}^k(\alpha) = c_{1N}^{k-1}(\alpha) + \frac{\sum_{i=1}^{N} \left( 1 - \prod_{j=1}^{p} I_{ij} (t_{ij} > Q_{ij}^{k-1}(\alpha)) \right) - N(1 - \alpha)}{\sum_{i=1}^{p} 1/\kappa_i}, \quad (2.26)$$

where $I_{ij}(\cdot)$ is an indicator function. This is a generic updating procedure without regard to the choice of the initial value. This updating strategy guarantees that the sequence $\{c_{1N}^k(\alpha) : k = 0, 1, \ldots\}$ approaches the target $c_{1N}(\alpha)$ in a monotonic way.

Formally, the numerical procedure for the quantile calculation of a $p$-variate $t$ distribution is carried out according to the following steps:

- **Initialization Stage:**
  1. Set up an initial value $c_{1N}^0(\alpha)$, using (2.22) or (2.25), and let $k = 0$;
  2. Calculate $c_{1N}^0(\alpha), i = 2, \ldots, p$, according to formula (2.20);
  3. Find the corresponding $Q_{0}(\alpha) = \{Q_{0}^i(\alpha), i = 1, \ldots, p\}$, where $Q_{0}^i(\alpha)$ is calculated as the $(c_{1N}^0(\alpha))^{th}$ largest element in the $i^{th}$ row of table $T_N$;
  4. Evaluate the probability $\hat{P}_T^0$ within the hyper-rectangle bounded by $Q_{0}^i(\alpha)$ using the Table 0/1 approach.

- **Updating Stage:**
  Repeatedly cycle through the following steps until the termination criterion is met. The criterion is governed by a tolerance error $\varepsilon_N$ with default value $p/N$:

$$\left| \hat{P}_T^k - \alpha \right| \leq \varepsilon_N. \quad (2.27)$$

  1. Set $k = k + 1$;
  2. Update $c_{1N}^k(\alpha)$ according to the updating formula (2.26);
  3. Calculate $c_{1N}^k(\alpha), i = 2, \ldots, p$, according to formula (2.20);
  4. Find the quantile $Q_{k}(\alpha)$;
  5. Evaluate the enclosure probability $\hat{P}_T^k$ within $Q_{k}(\alpha)$;
  6. Check the stopping criterion (2.27).
Output the final quantile estimate $Q^K(\alpha)$, provided the algorithm terminates at iteration $K$.

2.1.3 Spherical Radial Transformation: Probability Calculation

This section is devoted to the introduction of a variance reduction technique in the numerical approximation of a classical multivariate $t$ probability over any convex set $A \in \mathbb{R}^p$, as stated in (2.7). The probability estimate generated by this variance reduction procedure is anticipated to have increased accuracy along with more a rapid convergence rate, compared to the Table 0/1 method.

In the Table 0/1 procedure, each sample $T_i$, $i = 1, \ldots, N$, contributes to the sample proportion in (2.8) either as 0 or 1, which, in turn, is used as an approximation to the probability integral in (2.7). At the $\alpha$ quantile, the standard error of $\hat{P}(|T_i| < Q_i(\alpha), i = 1, \ldots, p)$, is $\sqrt{\alpha(1-\alpha)/N}$. A strategy that replaces the 0 or 1 from the table with a number between 0 and 1 without changing the mean will reduce the standard error. We develop such a strategy here.

The estimation strategy with uncertainty reduction is based on a Rao-Blackwell argument. The generation of $t_i$ is viewed as a two-stage process. In the first step, a direction is generated; in the second step, a specific value for $t_i$ is generated. The right panel of Figure 2.3 shows this two-stage procedure. The intuitive mechanism by which variance reduction is achieved is to view the 0/1 from the Table 0/1 approach as a single draw of $t_i$, conditional on this direction. The new approach replaces this single draw with an integral – equivalent to drawing an infinite number of $t_i$ conditional on this direction. The spherical-radial transformation method provides a strict mathematical treatment of this variance reduction approach. It reformulates
the probability integral regarding $\mathbf{T}$ in the coordinate system of the spherically transformed variable, say $\mathbf{Z}$. The left panel of Figure 2.3 shows a correlated bivariate $\mathbf{t}$ which is transformed from a random point on the unit sphere in the spherical coordinate system. The arrow gives a direction from the center of the $\mathbf{t}$ distribution. The point $\mathbf{t}_i$ may fall in or out of the rectangle $A$. This method replaces the binary 0 (out) and 1 (in) with an integral over the ray, leading to variance reduction.

In this section, we investigate the spherical-radial transformation method in the multivariate $\mathbf{t}$ probability calculation and its two Monte Carlo implementations, namely the MVT implementation and the MVN implementation. The MVT implementation makes direct use of the p.d.f. of the multivariate $\mathbf{t}$, whilst the MVN
implementation exploits the conditional Gaussian property of a t distribution on any given realizations of the gamma scaling variables, as is suggested in the stochastic representation [2,5]. The MVT implementation produces probability estimates with notably greater variance reduction, especially for multivariate t’s with small degrees of freedom, e.g., \( \nu < 20 \). However, it is feasible only for the class of classical t distributions and does not generalize to other varieties of multivariate t. On the contrary, the MVN implementation can be easily adapted to the probability evaluation of other forms of multivariate t, regardless of the existence of a well defined p.d.f.

Note that both implementations begin with a matrix factorization of the multivariate Gaussian covariance matrix \( \Sigma_X \) in the form of \( \Sigma_X = CC^T \), where the matrix factor \( C \in \mathcal{R}^{p \times q} (p \geq q) \) is of full rank. In addition to the several common factorization approaches described in the previous section, the matrix factor \( C \) can, under some circumstances, be promptly recognized by the construction of the underlying multivariate t random variable. We take multiple hypotheses testing for an example. Suppose that a finite collection of \( p \) linear hypotheses on \( l \) treatment groups is specified through a constant coefficient matrix \( K \in \mathcal{R}^{p \times l} \), i.e., \( \mathcal{H}_0 : K \mu = 0 \). The associated multivariate test statistic \( T_{\text{test}} \) has a classical multivariate t distribution, the Gaussian covariance structure of which decomposes directly as \( \Sigma_X = CC^T \), where \( C \in \mathcal{R}^{p \times l} \) is the constant coefficient matrix \( K \) column standardized by the square root of the sample size, i.e, \( C = K \cdot \text{diag}(\sqrt{n_1}, \ldots, \sqrt{n_l}) \).

The matrix factors \( C \) produced by different factorization methods often differ in column sizes. The number of columns of \( C \) produced by the most generic spectral decomposition approach is \( q \), the rank of \( \Sigma_X \). For positive definite \( \Sigma_X \), Cholesky decomposition yields a lower triangular matrix factor \( C \) with \( p \) columns. In addition,
the number of columns in \( C \) specific to the multiple testing framework is the number of treatment groups \( l \) involved in the multiple hypotheses.

Before proceeding to formally introduce the spherical-radial transformation, we comment that all theoretical results presented in the following subsections hold without regard to the choice of \( C \) satisfying \( \Sigma_X = CC^T \). Nevertheless, the amount of uncertainty reduction in the probability estimate is inversely related to the number of columns of \( C \), as was illustrated in Figure 2.2. As a consequence, we advocate the uniform use of optimal \( C \) that is constructed as the truncated spectral decomposition of \( \Sigma_X \) at the \( q \), i.e., the rank of \( \Sigma_X \), largest eigenvalues. The multivariate \( t \) probability estimate by the use of the spherical-radial transformation attains the maximum possible variance reduction at this optimal \( C \).

### The MVT Implementation

Any \( p \)-variate classical \( t \) random vector \( \mathbf{T} \sim MVT_p(\Sigma_X, \nu) \) can be written as a matrix transformation of a standard \( q \)-variate \( t \) random vector \( \mathbf{Y} \sim MVT_q(I, \nu) \) with identity covariance matrix and \( \nu \) degrees of freedom. Formally,

\[
\mathbf{T} = C \mathbf{Y},
\]

given that \( C \in \mathbb{R}^{p \times q} \) satisfies \( \Sigma_X = CC^T \). Thus, the probability content of \( \mathbf{T} \) over any convex set \( A \in \mathbb{R}^p \) is converted to that of \( \mathbf{Y} \) over the transformed convex set \( \mathcal{G} \in \mathbb{R}^q \) such that \( \mathcal{G} = \{ \mathbf{y} \in \mathbb{R}^q \mid C \mathbf{y} \in A \} \). The set \( \mathcal{G} \) is also a convex set, as linear transformation preserves convexity. Therefore,

\[
P(\mathbf{T} \in A) = P(C \mathbf{Y} \in A)
= \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu \pi)^{q/2}} \int_{C \mathbf{y} \in A} \left(1 + \frac{\mathbf{y}^T \mathbf{y}}{\nu}\right)^{-\frac{\nu+q}{2}} d\mathbf{y}.
\]

(2.29)
The spherical radial transformation method operates in the hyper-spherical coordinate system by taking an additional transformation such that \( y = r z \), where \( \| z \| = 1 \) and \( dy = r^{q-1}dz \). The radial direction \( Z \in R^q \) is uniformly distributed over the surface of the unit sphere in the \( q \) dimensional space \( U_q \). For a given random direction \( z \), let the radial distance \( r \) be the distance from the origin to the boundary of the convex set \( G \) in the transformed space. Accordingly, under the spherical coordinate system, the multivariate \( t \) probability in (2.29) becomes,

\[
P(T \in A) = \frac{\Gamma(\frac{\nu+q}{2})}{\Gamma(\frac{q}{2})(\nu \pi)^{q/2}} \int_{\|z\|=1} \int_{rCz \in A} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+q}{2}} r^{q-1}drdz. \tag{2.30}
\]

Adjusting for the surface area \( \Gamma(\frac{q}{2})/(2\pi^{\frac{q}{2}}) \) of the unit sphere \( U_q \), the probability (2.30) is further written as

\[
P(T \in A) = \frac{\Gamma(\frac{\nu+q}{2})}{\Gamma(\frac{q}{2})(\nu \pi)^{q/2}} \int_{\|z\|=1} \int_{rCz \in A} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+q}{2}} r^{q-1}drdz = \frac{\Gamma(\frac{q}{2})}{2\pi^{q/2}} \int_{\|z\|=1} F(A, z \nu, C)dz, \tag{2.31}
\]

where the \( F(A, z \nu, C) \) function, defined as

\[
F(A, z \nu, C) = \frac{2\Gamma(\frac{\nu+q}{2})}{\Gamma(\frac{q}{2})\Gamma(\frac{q}{2})^{\nu/2}} \int_{rCz \in A} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+q}{2}} r^{q-1}dr, \tag{2.32}
\]

is a simple 1-dimensional integral with respect to the radius \( r \) for any given direction \( z \in U_q \). The function \( F(A, z \nu, C) \) has an explicit expression which is revealed through integration by parts, as shown in the following recursive form.
\[ F(A, z, \nu, C) = \frac{2\Gamma(\frac{\nu+q}{2})}{\Gamma(\nu/2)\Gamma(\frac{\nu}{2})} F_0 \]

For \( i = 0, 1, \ldots \),
\[
\left\{ \begin{array}{ll}
\frac{q-3}{2} & \text{for } q \text{ odd} \\
\frac{q}{2} - 2 & \text{for } q \text{ even}
\end{array} \right.
\]

\[ F_i = \int_0^{r(z)} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+q}{2} + i} r^{q-1-2i} dr 
\]

\[ F_i = \frac{1}{2(i + 1 - \frac{\nu+q}{2})/\nu} F_i^{q-2(i+1)} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+q}{2} + i} r(z) \bigg|_0 \bigg| - \frac{1}{2(i + 1 - \frac{\nu+q}{2})/\nu} (q - 2(i + 1)) F_{i+1} 
\]

\[ F_{\frac{q-1}{2}} = \int_0^{r(z)} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+1}{2}} dr = \frac{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}}{\Gamma(\frac{\nu+1}{2})} (F_{T,\nu}(r(z)) - F_{T,\nu}(0)), \text{ for } q \text{ odd} 
\]

\[ F_{\frac{q-1}{2}} = \int_0^{r(z)} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+1}{2}} dr = 1 - \left( \frac{r(z)^2}{\nu} + 1 \right)^{-\frac{\nu}{2}}, \text{ for } q \text{ even}, 
\]

(2.33)

where \( F_{T,\nu} \) is the c.d.f. of the univariate \( t \) distribution with \( \nu \) degrees of freedom, for which accurate and efficient computational algorithms are available in many statistical packages. The evaluation is performed in a backward order.

Numerous random sampling schemes have been proposed in earlier works for the generation of samples from the surface of a \( q \)-dimensional unit sphere \( U_q \). Among them, a simple and commonly taken approach is to rescale a set of \( q \) i.i.d. standard Gaussian variates to have norm 1. That is, for \( Q_i \sim N(0, 1), i = 1, \ldots, q \),

\[ Q = \left( \begin{array}{c}
Q_i \\
\sqrt{\sum_{i=1}^q Q_i^2}
\end{array} \right), \quad i = 1, \ldots, q \]

(2.34)

The \( Q \) is a uniform random vector over the surface of \( U_q \). In our discussion, this standard sampling strategy is adopted for the purpose of randomly generating radial directions. See [Genz and Bretz (2002)] for a more sophisticated sampling scheme that
maps points from the unit hypercube to the surface of $U_q$, using the transformation described in [Fang et al. (1994)].

Depending on the inclusion or exclusion of the origin in the convex set $A$ and the randomly chosen direction $z \in U_q$, there are three possible scenarios. For a convex set $A$ that contains the origin, any randomly chosen direction $z \in U_q$ passes through the boundary of the convex set $G$ in the spherical transformed space at a single point. The radial distance $r$ is uniquely obtained as the distance from origin to the intersection point. The function $F(A, z \nu, C)$ is used to provide an unbiased estimate of $P(T \in A)$ via

$$E[F(A, Z \nu, C)] = P(T \in A),$$

where the expectation is over the random direction $Z$. For a convex set $A$ that does not contain the origin, a randomly chosen direction $z \in U_q$ either intersects the boundary of $G$ at two points or does not intersect at all. If the radial line intersects the boundary, then the contribution to the estimator of $P(T \in A)$ is given by

$$F(A, z \nu, C) = F_1(A, z \nu, C) - F_2(A, z \nu, C),$$

where $F_1(A, z \nu, C)$ is calculated according to the recursive formula (2.33) with the larger radial distance $r_1(z)$ and $F_1(A, Z \nu, C)$ with the smaller distance $r_2(z)$. If it does not intersect the set $A$, then the contribution to the estimator is 0. In a unified manner, the heart of an unbiased estimator for $P(T \in A)$ is given by

- **Scenario 1:** For $0 \in A$,

$$F(A, z \nu, C)$$

- **Scenario 2:** For $0 \notin A$,

$$F(A, z \nu, C) = \begin{cases} F_1(A, z \nu, C) - F_2(A, z \nu, C) & z \cap A \neq \emptyset \\ 0 & z \cap A \neq \emptyset \end{cases}.$$
The Monte Carlo approximation of the probability content $P(T \in A)$ is obtained by averaging over sufficiently many $F(A, z_i, \nu, C)$ values. Each $F(A, z_i, \nu, C)$ value is evaluated at a randomly generated direction $z_i \in U^q$. Hence,

$$\hat{P}(T \in A) = \frac{1}{N} \sum_{i=1}^{N} F(A, z_i, \nu, C).$$

(2.39)

The standard error $\widehat{\sigma}_{MVT,N}$ of the unbiased estimator $\hat{P}(T \in A)$, calculated as

$$\widehat{\sigma}_{MVT,N} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} \left( F(A, z_i, \nu, C) - \hat{P}(T \in A) \right)^2},$$

(2.40)

provides the uncertainty estimate for the MVT implementation of the spherical radial transformation technique. As argued previously, the error estimate may vary with different choices of the matrix factor $C$.

In the multiple testing framework, the probability calculation of a $p$-variate $t$ distribution over a hyper-rectangular region defined as $A = \{-\infty \leq a \leq z \leq b \leq \infty\}$ plays a distinctive role. The image of this set $A$ in the transformed space is an irregular convex polytope $G$, bounded by $2q$ hyperplanes. The two main tasks in multiple hypotheses test are the establishment of a joint rejection region that contains $100\alpha$ percent of the probability mass of the multivariate $t$ and the power calculation – that is, the probability mass of the joint rejection region under some non-central multivariate $t$ distribution. In the formulation of the decision rule, a central multivariate $t$ distribution is evaluated to obtain the hyper-rectangular non-rejection region with probability $(1 - \alpha)$ centered at the origin, in which case any random direction $z$ intersects $G$ only once. For a given random direction $z$, let $v = Cz$, a little basic algebra provides a closed form expression of the radial distance $r(z)$ as

$$r(z) = \max \left( 0, \min_{v_i > 0} \{b_i/v_i\}, \min_{v_i < 0} \{a_i/v_i\} \right)$$

(2.41)
The second scenario is often encountered in the power calculation when the non-central multivariate $t$ distribution with a location parameter in the alternative space lies in the rejection region.

**The MVN Implementation**

The MVT implementation of the spherical radial transformation method makes direct use of the well-defined p.d.f. of the classical multivariate $t$ distribution, $Y \sim MVT_p(I, \nu)$, which is linearly related to $T$ in the form $T = CY$. Nonetheless, a non-central $t$ distribution with non-centrality parameter $\mu$ represented by

$$T^{NC} = \frac{X + \mu}{\sqrt{W/\nu}} \quad \text{(2.42)}$$

is asymmetric and has a density function that is not ellipsoidal-shaped. Furthermore, the MVT implementation fails to work for the class of $p$-variate non-central $t$ distributions, as the $p$-dimensional space cannot be mapped down to a $q$-dimensional space through a matrix transformation.

In addition, numerous variations of the multivariate $t$ distribution exist, each of which is a modified form of the classical $t$ attempting to account for some special feature of the underlying data generating mechanism. We propose a flexible class of generalized multivariate $t$ distributions that accommodates heterogeneity across the gamma variables for the Gaussian random vector. Generally, new classes of multivariate $t$ distributions are proposed in terms of the stochastic representation, but probabilistic models are typically not available in neat form, which renders the MVT implementation inapplicable for the probability calculation for these modified multivariate $t$ classes.
To this end, we seek an alternative implementation of the spherical radial transformation method by the use of the stochastic representation of the classical $t$, which can be easily adapted to the probability calculation for other classes of multivariate $t$. This approach uses the conditional Gaussian property of the multivariate $t$ distribution to advantage, by which the original multivariate $t$ probability calculation proceeds as a Gaussian probability calculation with the gamma scaling factor handled separately. For this reason, we term it the MVN implementation of the spherical radial transformation method.

Consider a $p$-variate classical multivariate $t$ random vector $\mathbf{T} \sim \text{MVT}_p(\Sigma_X, \nu)$, centered at the origin $\mathbf{0}$. By virtue of the stochastic representation, it can be viewed as the ratio of a Gaussian random vector $\mathbf{X} \sim \text{MVN}_p(\mathbf{0}, \Sigma_X)$ and an independent gamma variable $W \sim \chi^2_{\nu}$,

$$\mathbf{T} = \frac{\mathbf{X}}{\sqrt{W/\nu}}.$$  

(2.43)

Conditional on any realized value $w$, $\mathbf{T}|(W = w)$ has a multivariate Gaussian distribution,

$$\mathbf{T}|(W = w) \sim \text{MVN}_p(0, \nu \Sigma_X/w).$$  

(2.44)

The conditional Gaussian property is employed to provide an alternative implementation of the spherical-radial transformation approach in multivariate $t$ probability calculations. In the classical $t$ setting, the amount of variance reduction in probability estimate achieved via the MVN implementation is noticeably less than that via the MVT implementation, with higher discrepancy observed for smaller degrees of freedom $\nu$. However, the MVT implementation is restricted merely to the classical $t$, and the superiority of the MVN implementation becomes apparent when it comes to other general forms of multivariate $t$.  

40
Using the conditional representation of $T$, the original probability calculation of $T$ over any convex set $A \in \mathcal{R}^p$, conditioning on a gamma sample $w$ is converted to one for the Gaussian random variable $X$. Formally,

$$P(T \in A | W = w) = P(X \in A_w),$$

(2.45)

$$A_w = \{x | x = t \sqrt{w/v}, t \in A\}.$$

(2.46)

The Gaussian random vector $X \in \mathcal{R}^p$ can be written as a linear transformation of a standard Gaussian random vector $Y \in \mathcal{R}^q$ in the form of $X = CY$, provided that $\Sigma_X = CC^T$ and $C \in \mathcal{R}^{p \times q}$. Thus,

$$P(X \in A_w) = P(CY \in A_w) = \frac{1}{(2\pi)^{q/2}} \int_{C_y \in A_w} e^{-\frac{r^2 y^2}{2}} dy.$$

(2.47)

Analogous to the spherical radial transformation used in the MVT implementation, letting $y = rz$, the $q$-fold integration in (2.47) with respect to $y$ becomes

$$P(X \in A_w) = \frac{1}{(2\pi)^{q/2}} \int_{\|z\|=1} \int_{rCz \in A_w} e^{-\frac{r^2 z^2}{2}} r^{q-1} dr dz.$$

(2.48)

Adjusting for the surface area $\Gamma(\frac{q}{2})/(2\pi^{\frac{q}{2}})$ of the unit sphere $U_q$, the probability (2.48) is further written as

$$P(X \in A_w) = \frac{1}{(2\pi)^{q/2}} \int_{\|z\|=1} \int_{rCz \in A_w} e^{-\frac{r^2 z^2}{2}} r^{q-1} dr dz = \frac{\Gamma(\frac{q}{2})}{2\pi^{q/2}} \int_{\|z\|=1} G(A_w, z, \nu, C) dz,$$

(2.49)

where the $G(A_w, z, \nu, C)$ at a given direction $z \in U_q$ and a sample point $w$ is defined as

$$G(A_w, z, \nu, C) = \frac{1}{2^{\nu-1} \Gamma(\frac{q}{2})} \int_{rCz \in A_w} e^{-\frac{r^2 z^2}{2}} r^{q-1} dr.$$

(2.50)

The $G(A_w, z, \nu, C)$ can be simplified to

$$G(A_w, z, \nu, C) = F_G \left( r_1^2(z, w) \right) - F_G \left( r_2^2(z, w) \right),$$

(2.51)
where $F_G$ is the gamma distribution function with shape parameter $q/2$ and scale 2, and $r_1^2(z, w)$ and $r_2^2(z, w)$ are the distances from the center $\mu_Y$ of the transformed variable $Y \in \mathcal{R}^q$ to the boundary of the convex set $\mathcal{G} = \{ y \in \mathcal{R}^q \mid Cy \in A_w \}$ in the spherical coordinate system. The transformed center $\mu_Y$ is the unique solution to the system of linear equations,

$$\mu_X = C \mu_Y$$

for the choice of a full rank matrix $C$. As with the MVT implementation, the radial line from the center $\mu_Y$ in the direction of $z$ intersects $\mathcal{G}$ in two possible ways, depending on the relative position of $\mu_Y$ in reference to $\mathcal{G}$.

- When $\mu_Y \in \mathcal{G}$, any radial line passes through the boundary of $\mathcal{G}$ at a single point. In this case, $r_2^2(z, w) = 0$.

- When $\mu_Y \notin \mathcal{G}$, any radial line either intersects $\mathcal{G}$ at two distinct points or does not intersect at all. In the former case, $r_1^2(z, w) > r_2^2(z, w) > 0$ and in the latter case, $r_1^2(z, w) = r_2^2(z, w) = 0$.

In particular, letting $v = Cz$, $c = a \sqrt{\frac{w}{\nu}}$ and $d = b \sqrt{\frac{w}{\nu}}$, the spherical limits for hyper rectangles $A = \{ -\infty \leq a \leq z \leq b \leq \infty \}$ are calculated as

$$r_1^2(z, w) = \max(0, \min\{d_i/v_i\}, \min\{c_i/v_i\})$$

$$r_2^2(z, w) = \max(0, \max\{c_i/v_i\}, \max\{d_i/v_i\}).$$

(2.53)

The Monte Carlo algorithm estimates the probability $P(T \in A)$ by averaging over a sufficiently large number of $G(A_{w_i}, z_i, \nu, C)$ values. Each is calculated at a randomly sampled radial direction $z_i \in U^q$ and gamma variable $w_i$,

$$\hat{P}(T \in A) = \frac{1}{N} \sum_{i=1}^{N} G(A_{w_i}, z_i, \nu, C).$$

(2.54)
The standard error $\hat{\sigma}_{MVN,N}$ of the unbiased estimator $\hat{P}(T \in A)$ is calculated as

$$\hat{\sigma}_{MVN,N} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} \left( G(A_{w_i}, z_i, \nu, C) - \hat{P}(T \in A) \right)^2}.$$  \hspace{1cm} (2.55)

The Monte Carlo standard error is used as a criterion to assess and compare relative performances of the several probability estimation approaches described in the preceding sections, namely the Table 0/1 approach, the MVT and MVN formulations of the spherical radial transformation technique. Furthermore, as is demonstrated in our later numerical work, a substantial amount of reduction in uncertainty estimates can be achieved by the exploitation of the spherical radial transformation technique in comparison to the Table 0/1 algorithm. An additional benefit of the spherical radial method over the Table 0/1 approach is in its reduced computational effort.

To highlight the difference between the two formulations of the spherical radial method, we recognize that the MVT implementation passes the randomness in $T \sim MVT(\Sigma_X, \nu)$ on to that in $Y \sim MVT_q(I, \nu)$ through their linear relation $T = CY$. However, in the MVN implementation, the randomness in $T$ is passed on to that in $Y$, which is further broken down into that of the multivariate Gaussian random vector $X \sim MVN_q(0, \Sigma_X)$ and the gamma variable $W$:

$$P(T \in A) = P(CY \in A)$$  \hspace{1cm} (2.56)

$$P(T \in A) = EW \left( P(X \in A_W) \right).$$  \hspace{1cm} (2.57)

The MVT formulation is bound to yield an estimate for $P(T \in A)$ with greater uncertainty reduction compared to the MVN formulation. A plausible justification for this phenomenon is the additional source of randomness caused by the separate treatment of $X$ and $W$ as two random quantities in the MVN formulation, as opposed
to treating their ratio $Y$ as a single random quantity. The slower convergence can be attributable to the randomness in $W$. Despite its relative deficiency of the MVN implementation, it opens up the possibility of applying this probability evaluation methodology to some more general classes of multivariate $t$ distributions, for which the MVT formulation does not apply.

In addition, the convergence of the univariate Student’s $t$ to the standard normal carries over to that of the multivariate $t$ distribution to a standard multivariate Gaussian, as $\nu \to \infty$. This convergence property yields the near-equivalence in the probability calculation using both the MVT and MVN implementations of the spherical radial transformation, for a classical $t$ distribution with sufficiently large degrees of freedom $\nu$. Moreover, we emphasize the generic use of MVN implementation in probability calculation over a convex set for any form of the multivariate $t$ that possess the conditional Gaussian property.

An interesting pattern that we find worth exploring deeper is the downward trend in the achievable uncertainty reduction, as the number of columns of $C$ satisfying $\Sigma_X = CC^T$ increases. This pattern suggests that minimizing the number of columns of $C$, or equivalently using $C$ with full column rank yields the greatest variance reduction with the spherical radial transformation. An intuitive explanation to this phenomenon is tied to the amount of randomness that is brought into the estimation mechanism. In the MVT implementation, the randomness in $T$ is entirely taken over by the randomness in $Y$ that satisfies $T = CY, C \in \mathcal{R}^{p\times q}$. The number of i.i.d. standard Gaussian quantities matches the number of columns of $C$. In the MVN implementation, the mechanism is governed by an additional gamma variable $W$,
apart from the \( q \) Gaussian variables. A rigorous argument can be inferred from the Rao-Blackwell theorem.

Figure 2.4 illustrates the inverse relation between the column size and achievable variance reduction, for both the MVT and the MVN implementations. The points are colored by the three implementation methods considered in this work, the Table 0/1 approach, the MVN and MVT implementations of the variance reduction technique via spherical radial transformation. Regardless of the choice of the matrix factorization method, the MVT implementation achieves more uncertainty reduction in the multivariate \( t \) probability estimation than the MVN implementation, which is in turn more effective than the Table method. What is more interesting is the contrast across the three symbols within each color. A comparison among the green symbols indicates that a smaller number of columns of \( C \) typically results in lower uncertainty in the probability estimate. Moreover, the superior performance of the smaller matrix becomes more apparent as the number of treatments gets larger.

2.1.4 Spherical Radial Transformation: Quantile Calculation

The previous section places special focus on the spherical radial transformation approach in probability calculation of a classical \( t \) distribution for any convex set \( A \). In particular, as noted earlier, the MVN implementation enhances the utility of the spherical radial method for the probability calculation in a variety of other classes of multivariate \( t \) distribution. Using the probability calculation established in the preceding section as a building block, this section focuses on the quantile evaluation of a \( p \)-variate \( t \) distribution characterized by an arbitrary correlation matrix \( \Sigma_X \) and degrees of freedom \( \nu > 1 \). Our practical algorithm employs the simplest
Figure 2.4: The Left Panel: A Numerical Comparison on the amount of Variance Reduction Achieved by the Use of the two Implementations of the Spherical Radial transformation and the Table 0/1 Evaluation with Different Selections of the Matrix Factorization Methods, as the Number of Comparisons Increase. The Right Panel: Contrasts on the Accuracy of the Probability Estimates.
bisection method to illustrate the procedure of searching for the \((100\alpha)^{th}\) quantile with any choice of the probability calculation strategy. Various more sophisticated root-finding algorithms, e.g., Brent’s method, can be used instead. However, the bisection method serves very well the purpose of quantile function evaluation in the sense that it yields moderately accurate estimates with noticeably reduced run-time compared to the naive Table 0/1 approach. To this end, we will not consider alternative search algorithms in the remainder of our work. Potential improvement is anticipated with other more efficient search algorithms.

Suppose that the MVN implementation of the spherical radial method is chosen to compute the multivariate \(t\) probability over a hyper-rectangle defined as \(A = [-Q, Q]\), where \(Q = \{Q_i, i = 1, \ldots, p\}\). Other available approaches work equally well. Denote the probability content within \(\prod_{i=1}^{p} [-Q_i, Q_i]\), subject to the constraints in (2.18) by \(\alpha_N(Q_1)\), i.e.,

\[
\alpha_N(Q_1) = P(|T_i| < Q_i, i = 1, \ldots, p)
\]  

(2.58)

based on \(N\) replications. Moreover, the Monte Carlo method uses the arithmetic mean of \(N\) independently sampled \(G(A_{w_i}, z_i, C, \nu)\) values as an approximation to the target probability integral. Each \(G(A_{w_i}, z_i, C, \nu)\) is a conditionally (or direction and scale) unbiased estimate of the target probability. However, higher estimation accuracy based on a large scale simulation inevitably comes at the cost of additional computational complexity, which increases with the simulation size \(N\) in a roughly linear fashion. In addition, to fulfill the principle goal of accurate quantile function calculation, probability estimation for any given hyper-rectangle \(A\) is regarded as a fundamental building unit, the estimation accuracy of which plays a deterministic
role in the achievable accuracy of the subsequent quantile estimation. With this motivation, it is desirable to seek a numerical algorithm that produces quantile estimate for any given $\alpha \in (0, 1)$ within arbitrarily specified precision, and yet with affordable computational effort.

To numerically solve for $Q_1(\alpha)$ satisfying equations (2.17) and (2.18), we start our iterative algorithm with a moderate scale simulation of size $N_s$ and initial guesses for the upper and lower bounds $Q_1^{u,0}(\alpha)$ and $Q_1^{l,0}(\alpha)$ such that $Q_1^{l,0}(\alpha) < Q_1(\alpha) < Q_1^{u,0}(\alpha)$ is guaranteed with probability 1. We recommend setting the upper bound $Q_1^{u,0}(\alpha)$ by assuming independence across the $p$ marginals $T_i \sim t_\nu, i = 1, \ldots, p$, i.e.,

$$P(|T_i| < Q_1^{u,0}(\alpha)) = \alpha^{1/p},$$

and setting the lower bound $Q_1^{l,0}(\alpha)$ assuming complete dependence among $T_i$, i.e

$$P(|T_i| < Q_1^{l,0}(\alpha)) = \alpha.$$  

With this initialization, the algorithm proceeds with the basic bisection search algorithm to obtain a rough quantile approximation $Q_1^{Temp}(\alpha)$, which usually requires about 10 iterations before meeting the termination criterion. The preliminary $Q_1^{Temp}(\alpha)$ is a coarse estimate of $Q_1(\alpha)$ based on a relatively small simulation of size $N_s$, the weak performance of which in accuracy is markedly compensated by its computational manageability.

Consider the probability function

$$f(Q_1) = P(|T_i| < Q_1, i = 1, \ldots, p)$$

satisfying the series of equations specified in (2.18). Within a sufficiently small interval around the target quantile $Q_1(\alpha)$, an application of the Taylor series expansion suggests a linear approximation for $f(Q_1)$ with negligible higher order remainder terms.
The tangent, usually obtained as \( f'(Q_1(\alpha)) \) is not analytically available for \( f(Q_1) \), due to the complexity of the function. Instead, numerical approximation \( \hat{f}'(Q_1(\alpha)) \) will be used by simply fitting a regression line to the set of points \( (Q_{\text{Bisection}}, \alpha_{N_s}(Q_{\text{Bisection}})) \) generated and stored along with the bisection search algorithm, i.e.,

\[
\alpha_{N_s}(Q_{\text{Bisection}}) = \gamma_0 + \gamma_1 Q_{\text{Bisection}} + \varepsilon. \tag{2.62}
\]

Another useful notion embraced in the algorithm is the subtle difference in tangent estimates as the simulation size varies. At the exit of the first step, the temporary quantile estimate \( Q_1^{\text{Temp}}(\alpha) \) together with the tangent estimate \( \hat{f}'(Q_1(\alpha)) \) will be carried into the second stage for better assessment of the quantile \( Q_1(\alpha) \).

The second step of the algorithm is concerned with enhancing the approximation accuracy for \( Q_1(\alpha) \) by means of a larger scale simulation of size \( N_b \), without requiring considerably added computational effort. This can be effectively done by sequentially assessing the probability \( \alpha_{N_b}(Q_1^{\text{Temp}}(\alpha)) \) at the tentatively favored \( Q_1^{\text{Temp}}(\alpha) \) and updating the quantile the \( Q_1^{\text{Updated}}(\alpha) \) as

\[
Q_1^{\text{Updated}}(\alpha) = \gamma_1 \times \left( \alpha - \alpha_{N_b}(Q_1^{\text{Temp}}(\alpha)) \right) + Q_1^{\text{Temp}}(\alpha), \tag{2.63}
\]

assuming that the tangent \( \gamma_1 \) based on \( N_s \) number of simulations is indiscernibly different from that based on \( N_b \). The second step terminates with the probability evaluation \( \alpha_{N_b}(Q_1^{\text{Updated}}) \).

Lastly, a final adjustment is made by a natural interpolation between the two points \( Q_1^{\text{Updated}}(\alpha) \) and \( Q_1^{\text{Temp}}(\alpha) \) evaluated according to the large simulation as follows

\[
Q_{1}^{\text{Final}}(\alpha) = Q_1^{\text{Updated}}(\alpha) + \left( \alpha - \alpha_{N_b}(Q_1^{\text{Updated}}(\alpha)) \right) \times \frac{Q_1^{\text{Updated}}(\alpha) - Q_1^{\text{Temp}}(\alpha)}{\alpha_{N_b}(Q_1^{\text{Updated}}(\alpha)) - \alpha_{N_b}(Q_1^{\text{Temp}}(\alpha))}. \tag{2.64}
\]
and our algorithm ends with $Q_{1}^{Final}(\alpha)$ being the final output. The algorithm consists of two steps: (1) an evaluation step based on a small scale simulation; and (2) a polishing step based on a large scale simulation. This two-step procedure enables us to obtain arbitrarily precise quantile approximation, with affordable computational complexity. We empirically illustrate the superior performance of our proposed quantile evaluation algorithm, in terms of both estimation accuracy and computational efficiency, over other currently available strategies in the simulation study section.

The algorithm is formally presented in two phases.

Phase 1:
Generate a random sample of moderate size $N_s$ according to $T \sim MVT_p(\Sigma_X, \nu)$ and arrange it into a $p \times N_s$ table consisting of $N_s$ columns. Each column represents a sample point $t_j \in \mathcal{R}^p$. Let $\varepsilon$ be any error tolerance parameter and for any $\alpha \in (0, 1),$

- Set $t = 0$ and initialize the upper and lower bounds $Q_{1}^{u,0}(\alpha)$ and $Q_{1}^{l,0}(\alpha)$; Set $\alpha_t = 1$, $Q_{Bisection} = \emptyset$ and $\alpha_{Bisection} = \emptyset$;
- Repeat until $|\alpha_t - \alpha| < \varepsilon$,
  - Set $t = t + 1$;
  - Let $Q_1 = (Q_1^{u,t-1} + Q_1^{l,t-1}) / 2$ and evaluate the probability $\alpha_t = \alpha_{N_s}(Q_1)$;
    - If $\alpha_t < \alpha$, set $Q_1^{u,t} = Q_1^t$; otherwise, set $Q_1^{l,t} = Q_1^t$;
  - Set $Q_{bisec} = (Q_{bisec}, Q_1^t)$ and $\alpha_{bisec} = (\alpha_{bisec}, \alpha_t)$;
  - Set $Q_1^{Temp}(\alpha) = \frac{Q_1^{u,t-1} + Q_1^{l,t-1}}{2}$;
- Obtain the regression coefficient $\gamma_1$ according to regression model formulated in [2.62];

Phase 2:
Generate another random sample of sufficiently large size $N_b$ according to $T \sim MVT_p(\Sigma_X, \nu)$. Based on this large table, carry out the following steps,

- Assess the probability integral $\alpha_{N_b}(Q_1^{Temp}(\alpha))$;
• Update the tentative first coordinate quantile according to the updating formula in (2.63) and obtain $Q_{Updated}^1(\alpha)$;

• Assess the probability $\alpha_{N_b}(Q_{Updated}^1)$ with the updated $Q_{Updated}^1(\alpha)$;

• Make a final adjustment by interpolating between the two points evaluated at the larger table corresponding to formula (2.64);

The algorithm terminates at the completion of $Q_{Final}^1(\alpha)$.

In the subsequent discussion, we conduct a simple numerical study to illustrate the implementation of the quantile evaluation algorithm outlined in this section, in conjunction of the MVN version of the probability calculation given in the preceding section. Figure 2.5 depicts three separate repetitions, all of which attempt to estimate the 95th quantile of the multivariate $t$ distribution with a certain correlation structure and degrees of freedom $\nu = 6$, with different table sizes for a coarse evaluation of the quantile in the first phase. The three independent implementations are represented by different colors. The ‘+’ symbolizes the points generated along the bisection search algorithm for a coarse evaluation of the first quantile, using the MVN implementation of the variance reduction technique for probability evaluation in the first step. The plot shows that the rate of change in the tail probability as a function of the first quantile value, constant within a sufficiently small interval guaranteed by the Taylor’s approximation, can be estimated equally well without regard to the choice of table sizes in the initial phase. With the tentatively ‘best’ quantile obtained from the evaluation step, the remaining quantile components are obtained by the explicit relationship between the tail probabilities set as constraints.

The solid ‘△’ symbol denotes the tail probability of the quantile vector obtained in the first phase, evaluated using a sufficiently large table based on the Table 0/1 method. The updated first quantile, according to the equation (2.63) is denoted by
the solid ‘□’ symbol. The final update, according to the updating formula (2.64) produces the output of the algorithm and is represented by a solid ‘◦.’ The three solid ‘◦’ symbols are layered on top of one another and overlap completely with the green and blue ‘□,’ the update in the second phase for the largest two initial tables.

The most appealing feature of the design in the updating procedure for quantile evaluation is the reduced computational effort in the iterative quantile evaluations required for the bisection search algorithm. Despite that the candidate value proposed by a small table is less satisfactory than that proposed by a bigger table in the initial
step, the accuracy of the first quantile estimated based on a small table catches up with that for a bigger table in the two updating steps, leaving the final output hardly distinguishable.

2.2 Generalized Multivariate $t$ Distribution

Over the past few decades, statistical techniques for modeling multivariate data have been greatly enriched with the development of the multivariate $t$ distributions. The classical multivariate $t$ distribution is generally viewed as more flexible in modeling multivariate data compared to the multivariate Gaussian, as a result of the extra degrees of freedom provided by the gamma scaling factor. Potential outliers that are difficult to handle with a Gaussian distribution can be presumably accounted for with a multivariate $t$ distribution, especially for data exhibiting homogeneous patterns of outliers across coordinates. Moreover, as noted in previous studies, quite a few other forms of multivariate $t$ distributions have been proposed to model data that possess/exhibit special characteristics, e.g., asymmetry, skewness and non-centrality. These distributions have been used in many applications. A good survey of a number of modifications and extensions of the classical multivariate $t$ distribution is provided by Kotz and Nadarajah (2004).

In the last section, our exploration of the classical multivariate $t$ distribution focuses primarily on its role as the reference distribution of an appropriate test statistic in the multiple hypotheses testing framework, with an equal variance assumption. In the presence of heterogeneous variances as is frequently the case in real data analysis, the multivariate statistic does not follow a well studied distribution with known theoretical properties. Instead, a $t$ distribution that captures the essence of the statistical
properties of the statistic is used as the reference distribution to calculate critical values, resulting in an approximate test. Empirical evidence shows that the classical multivariate \( t \) distribution is a poor choice for approximation, producing an actual family-wise error rate that differs greatly from the nominal rate. This inadequacy of the multivariate \( t \) approximation suggests the need for a new class of multivariate \( t \) distributions that provides better approximation to the prescribed family-wise error rate for simultaneous comparisons across treatment groups. This section is devoted to the introduction of this new class of distributions that we call the generalized multivariate \( t \) distribution. These distributions are used for simultaneous inferences in subsequent chapters.

We recall that the stochastic representation of the classical multivariate \( t \) distribution can be viewed as component-wise division of a multivariate Gaussian random vector and a gamma random vector \( \mathbf{W} \) consisting of identical elements. Formally,

\[
T = \frac{\mathbf{X}}{\sqrt{\mathbf{W}/\nu}} = \left( \frac{X_1}{\sqrt{W_1/\nu}}, \frac{X_2}{\sqrt{W_2/\nu}}, \ldots, \frac{X_p}{\sqrt{W_p/\nu}} \right)^\top. \tag{2.65}
\]

The constant gamma random vector \( \mathbf{W} = (W_1, W_2, \ldots, W_p)^\top \) can be viewed as a scalar repeated \( p \) times, or a \( p \)-dimensional random vector, where all pairs of elements have correlation one, leading to a \( p \times p \) correlation matrix of ones indicating maximum positive dependence. The rigid correlation structure places a restriction on the application of the classical multivariate \( t \) distribution in many contexts. It also provides a practical motivation of generating a new variation of multivariate \( t \) distributions by relaxing the perfect correlation structure. The subsequent discussion in this section provides a formal definition of the new class of multivariate \( t \) distributions that allows coordinate-specific gamma divisors \( \mathbf{W} \) with arbitrary achievable correlation structure for their corresponding Gaussian components \( \mathbf{X} \). This new class
can be added to the list of modifications and extensions of the classical $t$ distributions extensively discussed in Kotz and Nadarajah (2004).

Many multivariate analysis tasks require the formulation of the joint distribution of a set of $p$ marginally $t$-distributed variables, with potentially varying degrees of freedom across the $p$ components. This joint distribution has clear benefits over the classical multivariate $t$ distribution in that it provides a more flexible probabilistic model in (1) capturing heterogeneous tail behaviors across the $p$ variables; (2) multivariate outlier detection, especially when the projected outliers onto the $p$ coordinates exhibit heterogeneous patterns. Furthermore, other tasks may require the $p$ gamma variables in the denominators of the $p$ marginal $t$ variables to have less than perfect correlation structure. One application of this joint distribution of $p$ marginally $t$-distributed variables with distinct degrees of freedom and arbitrary correlation structure in the gamma random vector is outlier identification, where the presence of an outlier for one variable tends to be diminished, preserved or increased by that for some other variables.

We propose to expand the scope of the multivariate $t$ distribution by modifying the stochastic representation in (2.65) such that the set of gamma variables $W = \{W_i, i = 1, \ldots, p\}$ is treated as a gamma random vector with constant scale parameter 2, arbitrary shape parameters and any achievable correlation structure. This generalized multivariate $t$ distribution accommodates the classical multivariate $t$ distribution as a special case with constant shape parameter across all $p$ components and perfect correlation matrix composed of ones. Additionally, the class of multivariate $t$ with coordinate-specific independent gamma scaling factors, referred to as the alternative $t$ in Finegold and Drton (2014) is regarded as another special case of the generalized
multivariate $t$ distribution. With independence and perfect correlation being the two extreme cases of the gamma random vector, our proposal places a less restrictive condition on the correlation structure and allows any feasible correlation structure that falls in between the two extreme cases, and even slightly beyond. As argued above, a distinctive feature of this general class of multivariate $t$ distributions lies in its enhanced ability to account for different rescaling of Gaussian components, allowing for varying outlier patterns (or tail behavior) and interdependence of the occurrences of outliers across coordinates. Lastly, a simulation study demonstrating the higher capacity of the generalized $t$ distribution in capturing the truth of the underlying mechanism will be shown at the end of this chapter.

2.2.1 Mathematical Formulation and Model Parameterization

The generalized multivariate $t$ distribution is defined in terms of its stochastic representation as

$$
T = \frac{X}{\sqrt{W/\nu}} = \left( \frac{X_1}{\sqrt{W_1/\nu_1}}, \frac{X_2}{\sqrt{W_2/\nu_2}}, \ldots, \frac{X_p}{\sqrt{W_p/\nu_p}} \right)\top, \quad (2.66)
$$

where $X = (X_1, \ldots, X_p)\top$ denotes the Gaussian random vector in the numerator with mean $\mu$ and correlation structure $\Sigma_X$, and is independent of the gamma random vector $W = (W_1, \ldots, W_p)\top$ in the denominator with arbitrary achievable correlation matrix $\Sigma_W$. Each marginal $W_i$ is a gamma variate with constant scale parameter 2 and shape parameter $\nu_i/2$, $i = 1, \ldots, p$. We consider only achievable correlation matrices for $W$ as the pairwise linear correlation is bounded at a level that is determined by the shape parameters. Figure 2.6 shows simulated random draws from some example generalized $t$ distributions with varying gamma correlations. The principle direction
of the scatterplots is controlled by the Gaussian correlation, whereas the shape is
driven by the correlation between the gamma variates defining the denominators. The
two extreme cases are a diamond shape for complete independence and an elliptical
shape for complete dependence. Like the classical $t$, the degrees of freedom control
the concentration of the density around the center.

We remark that the marginals and correlation structure provide only an approx-
imate representation of the gamma random vector $W$ since the joint distribution is
not fully specified by this information. Generally, complete characterization of de-
dependent random variables requires much more information than the marginals and
correlation structure. The multivariate Gaussian case is a prominent exception where
the marginals and correlations determine the joint distribution. Even in this setting,
many marginal Gaussian distributions exist and the correlation matrix is not enough
to determine the joint distribution.

This stochastic representation provides a direct interpretation of the generalized
multivariate $t$ distribution as the ratio of two independent random vectors. As a
consequence, each marginal $T_i$ of the generalized $t$ random vector $T$ has a univariate
$t$ distribution with $\nu_i$ degrees of freedom. Alternatively, a traditional definition of $T$
by means of the probability density function (p.d.f.) can be derived from (2.66) using
a transformation theorem that is well presented in many textbooks on multivariate
statistics, for instance [Rao (2001)]. The explicit analytical expression for the pdf is
complex and not easy to interpret. Furthermore, the pdf, in a conventional sense, only
exists when both $\Sigma_X$ and $\Sigma_W$ are non-singular matrices. The generalized multivariate
$t$ distribution is characterized by the set of Gaussian parameters and the gamma
parameters, denoted by $\text{MVT}(\Sigma_X, \Sigma_W, \mu, \nu_1, \ldots, \nu_p)$. 

57
Figure 2.6: The Four Plots Correspond to Four Generalized Bivariate $t$ Distributions with Common Gaussian Correlation $\rho_X = 0.5$ and Common Degrees of Freedom $\nu = (3, 3)$, with Gamma Correlation varies as such: $\rho_W = 0$ (Top Left), 0.5 (Top Right), 0.8 (Bottom Left) and 1 (Bottom Right)
In alignment with our original intent of the construction of the generalized multivariate $t$, we concentrate our effort on the exploration of mathematical properties directly pertaining to its potential application in the context of multiple hypotheses testing with heterogeneous variance. It was shown in Section 2.1 that the correlation structure of the classical multivariate $t$ random vector matches that of the associated Gaussian vector. In other words, it is independent of the common gamma scaling variable. However, the generalized multivariate $t$ random vectors admit more complex dependence structures. A mild algebraic argument using the law of total expectation shows that, for any $T \sim MVT(\Sigma_X, \Sigma_W, \mu, \nu_1, \ldots, \nu_p)$, the pairwise correlation is,

$$\Sigma_T(i, j) = \text{Corr}(T_i, T_j) = \Sigma_X(i, j) \times \text{Corr} \left( \sqrt{\frac{\nu_i}{W_i}}, \sqrt{\frac{\nu_j}{W_j}} \right), \quad \forall (i, j),$$

(2.67)

provided, of course that the variances are finite. This occurs when $\nu_i$ and $\nu_j$ both exceed 2. Clearly, (2.67) indicates that the correlation structure for $T$ depends on the properties of both the associated Gaussian random vector $X$ and the gamma random vector $W$, which constitutes a major difference from the classical $t$ distribution. The multivariate test statistic used in multiple comparisons with unequal group variances usually has partially correlated scaling variables in the denominator, due to the appearance of common sample variance terms across individual comparisons. To improve the current solution to the multiple comparison problem, we adopt the traditional method of approximating the set of scaling variables by a gamma random vector. This general correlation structure renders the classical $t$ distribution an inadequate approximation to the distribution of the multivariate test statistic and so makes the classical $t$ distribution a poor choice for a reference distribution. Multiple testing is the main focus of the next chapter.
Relative few results are available that deal with simultaneous inferences in the presence of unequal group variances. We address the problem by developing a method based on the class of generalized multivariate $t$ distributions, which distinguishes itself from the currently practiced methods by accounting for the dependence structure of the set of scaling variables, in addition to their marginals. The class of generalized multivariate $t$-distributions is relevant in numerous ways in the context of the multiple testing problem, including the evaluation of decision rules that aim to control the family-wise type I error rate at a preset $\alpha$ as well as in power calculation. We concentrate our effort on the investigation of its role in reference distribution approximation, which in turn is used to suggest simultaneous decision rules and joint rejection regions, etc. For this purpose, we restrict our discussion to the class of central generalized multivariate $t$ distributions with $\mu = 0$. The power calculation is the probability evaluation of the rejection region according to some non-central generalized multivariate $t$ distribution.

As with the application of the classical multivariate $t$ in multiple comparisons with equal variance, the assessment of simultaneous decision rules for unequal variances is translated to the problem of quantile calculation of the generalized multivariate $t$ distribution with additional constraints set forth to regulate the relative magnitudes of the individual errors. Typically, quantile evaluation is carried out through a numerical search algorithm, e.g., bisection, with an available probability calculation method for at least hyper-rectangular regions. This suggests that the development of an efficient probability calculation algorithm for the generalized multivariate $t$ is the key to its application in multiple comparisons with unequal variance.
Analogous to the classical $t$, we propose to numerically evaluate the generalized multivariate $t$ probability over any hyper-rectangular region using either the Table 0/1 approach or the spherical radial transformation approach. Similarly, an efficient simulation algorithm to generate i.i.d. replicates from the target generalized $t$ random vector is a prerequisite for both approaches. Making use of the stochastic representation of the generalized $t$, a random draw from the generalized $t$ is constructed simply by a component-wise division of a pair of independently sampled multivariate Gaussian and multivariate gamma variates. While random sampling from multivariate Gaussian is common practice, sampling from a gamma random vector with specified marginal gamma distributions and correlation matrix is not routine. Description of the implementation of the numerical probability calculation approaches is deferred until a simulation procedure for gamma random vectors is established. The original task of probability evaluation is converted to the exploration of an efficient simulation algorithm for generating samples from the associated gamma random vector.

In the following section, we discuss a sampling strategy for a gamma random vector with given marginal distributions and an achievable correlation matrix. Two methods — the copula method (Sklar, 1959) and the NORmal To Anything (NORTA) method — will be investigated in the next two subsections.

The Copula Method

The copula method (Sklar, 1959) is developed to model and estimate the joint distribution of any high dimensional random vector $W \in \mathbb{R}^p$ by separately estimating its marginals and dependence structure. MacEachern (2000), among others extends the notion of copula for a finite collection of variates to copula process for an infinite
number of variates. Wilson and Ghahramani (2010) study the copula method’s modeling in non-Gaussian time series by formally defining the Copula Process. Sklar’s theorem (see Nelsen (1999) for proof) states that any multivariate joint distribution function of any finite set of random variables can be expressed in terms of univariate marginal distribution functions and a copula $C$, defined as a joint cumulative distribution function c.d.f. of a random vector $(U_1, U_2, \ldots, U_d)$ with uniform marginals. That is,

$$C(u_1, u_2, \ldots, u_p) = P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_p \leq u_p)$$

$$U_i \sim U(0, 1) \quad i = 1, \ldots, p. \quad (2.68)$$

The copula $C$ describes the entire dependence structure of the original random vector $W$, not only the correlations. It is a well known result that marginals and correlation structure contain full amount of information for a multivariate Gaussian random vector, but this does not generalize to other multivariate vectors.

Let $W = (W_1, \ldots, W_p)\top$ be a continuous random vector with marginal distribution functions $F_1, \ldots, F_p$ and a $p$-dimensional joint distribution function $H(x_1, \ldots, x_p)$. As a consequence of the probability integral transformation (PIT) applied to each component, for any $i = 1, \ldots, p$, the marginal c.d.f. $U_i = F_i(W_i)$ follows a uniform distribution over the interval of $(0,1)$. The joint distribution function $C(\mu_1, \ldots, \mu_n)$ of the $p$-variate vector $U = (U_1, \ldots, U_n)\top$ is the copula of $W = (W_1, \ldots, W_p)\top$. According to Sklar’s theorem, there exists a copula $C$ such that

$$H(x_1, \ldots, x_n) = P(W_1 \leq x_1, \cdots, W_n \leq x_n) = C(F_1(x_1), \cdots, F_n(x_n)) = C(\mu_1, \cdots, \mu_n), \quad (2.69)$$
for $\forall x_i \in \mathcal{R}, \; i = 1, \ldots, p$. Conversely, if $C$ is a $p$-copula and $F_1, \ldots, F_p$ are marginal distribution functions, then the function $H$ defined in (2.69) is a $p$-dimensional distribution function with margins $F_1, \ldots, F_p$. This theorem suggests the representation of any joint distribution function $H$ by a copula transformation on its marginals.

An example parametric copula model is the multivariate Gaussian copula defined over the unit cube $[0, 1]^p$ as

$$C(\mu_1, \cdots, \mu_p) = \Phi_R(\Phi^{-1}(F_1(x_1)), \cdots, \Phi^{-1}(F_p(x_p))), \tag{2.70}$$

where $\Phi$ denotes the standard (univariate) normal distribution function and $\Phi_R$ denotes the multivariate normal distribution function with correlation structure $R$. The term Gaussian copula originates from the fact that the dependence structure $R$ for the Gaussian copula $C$ in (2.70) goes with that for the multivariate Gaussian vector. The Gaussian copula density has an explicit form omitted here.

The Gaussian random vector is a well studied multivariate model for which we have a wealth of theoretical understanding. On the contrary, specification of any non-Gaussian multivariate random vectors is comparatively challenging, as information beyond the marginals and correlation structure is typically required to fully characterize the joint distribution. Importantly, the Gaussian copula transformation provides an implicit connection between a random vector with arbitrary marginals and joint distribution function and a Gaussian random vector, making it possible to analytically investigate a non-Gaussian random vector through the well established Gaussian model.

The copula transformation provides an attractive approach to jointly model any finite set of random variables $W = \{W_i, \; i = 1, \ldots, p\}$ with specified marginal distributions $F_i$ by construction of a multivariate Gaussian copula model. The Gaussian
copula attempts to estimate the dependence structure across the $W_i$ variables. Sklar’s theorem guarantees the full specification of the joint distribution $H$ of $W$ with matching marginals by the Gaussian copula model. Furthermore, a common application of the copula model is in generating random samples from a random vector $W$ by taking element-wise inverse $F_i^{-1}$ transformation of samples $U$ from its Gaussian copula. Usually, the Gaussian copula is empirically estimated from i.i.d. realizations of $W$. Strelen (2009) reviews a modeling technique for random vectors via copulas and an efficient sampling method of the random vectors. Strelen (2009) also presents the copula modeling and generation of random vectors for time series framework. Other related works include MacEachern (2000), Wilson and Ghahramani (2010), etc.

Despite that the copula model provides a convenient tool in modeling and generating samples from multivariate non-Gaussian random variables, it does not serve well as a solution to our original task as to how to generate random samples from a non-Gaussian random vector $W$ with specified marginals and joint distribution function $H$. Our problem is formulated in a slightly different manner than the usual copula framework. The main challenge lies in the difficulty of the analytical derivation of a Gaussian copula model that is tied to the target joint distribution $H$.

In the multiple testing framework, the joint distribution of the approximating gamma random vector is typically unspecified. We intend to improve the current test procedures by bringing the correlation across the set of gamma random variables into account. The distributions for individual components are approximately gamma variables, in accord with Satterthwaite’s approximation. To preserve the essence of the joint distribution, we propose the representation of the joint gamma random vector by its gamma marginals, a Gaussian copula and the correlation matrix of the
gamma variables. The correlation matrix is allowed to be any legitimate correlation matrix.

In the following subsection, we briefly describe an initial effort we made in attempting to generate samples from a gamma random vector with a feasible correlation matrix. It is based on the additive property of gamma variable with common scale parameters, yielding the idea of constructing correlated gamma variables by allowing them to share common building units. However, this method fails to work for a fair proportion of the class of feasible correlation matrices, even for \( p = 3 \). A general and more stable solution to sampling an arbitrary random vector is presented by the NORmal To Anything (NORTA) method, due to Cario and Nelson (1997).

**Partition of Degrees of freedom**

Suppose that we wish to generate i.i.d. samples of a gamma random vector \( \mathbf{W} = (W_1, W_2, \ldots, W_p) \) with prescribed gamma marginal distributions with shape parameters \( \nu_i > 0 \) for all \( i \) and a feasible correlation matrix. That is,

\[
W_i \sim \text{Gamma}(\nu_i/2, 2), \quad i = 1, \ldots, p;
\]

\[
\Sigma_X = \Sigma_X(i, j), \quad 1 \leq i, j \leq p.
\]

Unlike the multivariate Gaussian random vectors for which any correlation matrix, i.e., a symmetric and positive semidefinite matrix \( \Sigma \) with ones as diagonal elements, qualifies as a feasible correlation matrix in the sense that a Gaussian random vector with the prescribed correlations exists, the set of feasible correlation matrices for gamma random vectors constitutes only a subset of possible correlation matrices. In other words, there exists marginal gamma distributions and correlation matrices that do not match any gamma random vector with the desired quantities. Computer programs have been developed to examine the feasibility of a given correlation matrix.
for the marginal distributions. See Ghosh and Henderson (2003). For example, the achievable correlation $\rho$ between two gamma variables $W_1 \sim \text{Gamma}(\nu_1/2, 2)$ and $W_2 \sim \text{Gamma}(\nu_2/2, 2)$ is given by $|\rho| \leq \frac{\min(\nu_1, \nu_2)}{\sqrt{\nu_1 \nu_2}}$, ruling out extreme correlations when $\nu_1 \neq \nu_2$. Further restrictions apply to larger sets of variables. We focus our discussion on the set of feasible correlation matrices.

To construct a correlated gamma random vector, our first attempt makes use of the partition property that a gamma variable can be decomposed into independent gamma components with common scale parameters, given that shape parameters add up to that of the target gamma variable. This partition property suggests an easy construction of correlated gamma variables by allowing them to share common individual gamma components. As a simple example, a bivariate gamma variable $W_1 \sim \text{Gamma}(\nu_1/2, 2)$ and $W_2 \sim \text{Gamma}(\nu_2/2, 2)$ with a feasible correlation $\rho \geq 0$ can be split as

$$W_1 = W_c + W_{r,1}$$
$$W_2 = W_c + W_{r,2}.$$ 

(2.72)

The common term $W_c$ is a gamma random variable with shape parameter $\nu_c = \rho \sqrt{\nu_1 \nu_2}$ governing the dependence between $W_1$ and $W_2$, given that $|\rho| \leq \frac{\min(\nu_1, \nu_2)}{\sqrt{\nu_1 \nu_2}}$. And, $W_{r,1}$ and $W_{r,2}$ are the gammas with complementary shape parameters to $W_1$ and $W_2$. This partition method works as a simple strategy for producing correlated bivariate gamma samples for the set of non-negative feasible correlations. Nevertheless, one can easily imagine how dramatic an increase in complexity could arise, as the dimension of the gamma random vector with feasible correlation matrices grows. An example can be envisioned for a 3-variate gamma random vector $\mathbf{W} = (W_1, W_2, W_3)^\top$, for whom the partition of any variable, say $W_1$ to match its desired correlations with other variables $W_2$ and $W_3$ is implicitly compounded by the need to correlate $W_2$ and $W_3$. More
explicitly, a common occurrence is the shortage of free remaining shape parameters (or degrees of freedom) in $W_2$ and $W_3$ to fully accommodate their correlations with $W_1$ after those spent on matching $W_2$ and $W_3$. This lack of generalizability makes the partition method rather restrictive and suggests the need for a different model with more general applicability.

This idea of splitting the degrees of freedom for the construction of correlated gamma random variates was explored in Jones (2002) in the development of Jones’ bivariate $t$ distribution with distinct marginals degrees of freedom. Briefly, the dependent bivariate gamma random variables, each with its own arbitrary shape parameter, are constructed as

$$W_1 \sim \text{Gamma}(\nu_1/2, 2)$$

$$W_2 = W_1 + U_2,$$

where $U_2 \sim \text{Gamma}((\nu_2 - \nu_1)/2, 2)$, provided that $\nu_1 < \nu_2$. In addition, extension of the construction of dependent bivariate gamma random variables in (2.73) to the multivariate case is as follows

$$W_1 \sim \text{Gamma}(\nu_1/2, 2)$$

$$W_i = W_1 + U_i,$$

where $U_i$ are independent gamma random variables with shape parameters $(\nu_i - \nu_1)/2, \forall i$. The use of correlated gamma variates in Jones’ bivariate $t$ distribution mainly aims to allow the tail behaviors of different marginals to be better accounted for in the resulting model, as opposed to the recovery of a target correlation structure. Thus, the inherent lack of ability for the decomposition method to fully control the pairwise correlation is less involved in the context of Jones’ research problem than ours – sampling from a gamma random vector with a specified correlation structure.
The NORmal To Anything (NORTA) method described by Cario and Nelson (1997) provides a general procedure for generating random samples of a finite-dimensional random vector represented by arbitrary marginal distributions and a feasible correlation structure. The kernel of the NORTA method is an appropriate transformation that converts a multivariate normal random vector \( Z \), entirely described by its Gaussian marginals and correlation structure, into a random vector with other marginal distributions and a desired correlation. The advantage of this NORTA method lies in its utilization of the fact that multivariate normal random vectors can be quickly and accurately generated in many software packages. We provide a short presentation of the NORTA method, followed by the description of a moment-based numerical procedure for its practical implementation.

The NORTA algorithm

The NORTA method (Cario and Nelson 1997) is a computationally efficient and easy to implement random vector generation algorithm. It has been widely used to solve many practical problems. It is a two-stage procedure to generate correlated random vectors \( W \) with prescribed marginals \( W_i \sim F_i, \forall i \) and correlation structure \( \Sigma_W \). The first step involves transforming a Gaussian random vector \( Z \) with a known correlation structure \( \Sigma_Z \) as the base vector into a multivariate uniform vector \( U \). This is essentially the copula transformation, as the joint distribution of \( U \) is a copula. The multivariate uniform vector \( U \) is in turn transformed into the desired random vector \( W \) by conducting a series of inverse c.d.f. transformations \( F_i^{-1}(U_i), i = 1, \ldots, p \). However, the NORTA approach is fundamentally distinct from the copula method as to how the pairwise correlations \( \Sigma_W(i,j), \forall i,j \) are transmitted in the construction process, which constitutes the key to its success as a random vector sampling
algorithm. The unsuitability of the copula method in our particular setting is due to the inability of passing the given correlation structure for the random vector \( W \) to the base vector \( U \). On the other hand, the NORTA algorithm explicitly relates the pairwise correlation in \( W \) with that of the corresponding pairs in the Gaussian random vector \( Z \). This enables us to compute the correlation for the basis \( Z \) in a component-wise fashion.

To formalize, sampling from a \( p \)-variate correlated gamma random vector \( W = (W_1, \ldots, W_p)^T \) with properties stated in (2.71) is achieved by the NORTA method through the following procedure.

**Generation Procedure:**

- Calculate the \( p \times p \) correlation matrix \( \Sigma_Z \) of the Gaussian input random vector \( Z \) that yields the desired \( p \times p \) correlation matrix \( \Sigma_Z \) by repeatedly evaluating each entry according to the bivariate correlation matching procedure, described subsequently;

- Generate an input data sample \( z \) from the base vector \( Z \) with correlation matrix \( \Sigma_Z \);

- Obtain an output data sample \( x \) from the desired random vector \( X = (W_1, \ldots, W_p)^T \) via a sequence of \( p \) inverse c.d.f. transformations,

\[
W_i = F_i^{-1}(\Phi(Z_i)) \quad i = 1, \ldots, p, \tag{2.75}
\]

where \( F_i^{-1} \) is the inverse c.d.f. of the \( i^{th} \) marginal gamma distribution. The so-constructed random vector \( W \) is guaranteed to have the desired marginal distributions \( F_i \) as a consequence of the probability integral transformation. However, as an
analytical expression of $\Sigma_Z(i, j)$ in terms of the desired $\Sigma_W(i, j)$ is not generally available, except for some special cases discussed in Cario and Nelson (1997). A numerical search is required to find the $\Sigma_Z$ the leads to the desired correlation $\Sigma_W$. In fact, an appropriate specification of $\Sigma_Z$ is the most challenging step in the implementation of the NORTA method.

The correlation between any two variables $W_i$, $W_j$ in the desired random vector is a function of only the correlation between the corresponding $Z_i$ and $Z_j$. To see this,

$$\Sigma_W(i, j) = \text{Corr}(W_i, W_j) = \text{Corr}(F_i^{-1}(\Phi(Z_i)), F_j^{-1}(\Phi(Z_j))).$$  \quad (2.76)

This result suggests that the task of determining the desired $\Sigma_Z$ for a $p$-dimensional Gaussian vector $Z$ can be reduced to $p(p - 1)/2$ independent individual tasks of determining the correlation for a bivariate Gaussian vector $(Z_i, Z_j)$, for $1 \leq i < j \leq p$. For each pair $(i, j)$, we find the value $\Sigma_Z(i, j)$ for which $\Sigma^{NORTA}_X(i, j)$ produced by the NORTA algorithm matches $\Sigma_W(i, j)$ in the target correlation matrix $\Sigma_W$ to any desired accuracy level. Equation (2.76) also shows that the correlation in $W_i$, $W_j$ is a strictly increasing function of that in $Z_i$ and $Z_j$. This monotonic property provides effective direct guidance on the direction of adjustment for the Gaussian correlation when the gamma correlation is mismatched.

As mentioned above, the random vector $W$ produced by the NORTA procedure has marginal distributions $F_i$, $i = 1, \ldots, p$, as specified. Therefore, a numerical estimate of $\Sigma_Z(i, j)$ that solves equation (2.76) can be obtained by matching the 2nd order moments $E(W_i W_j)$. The marginal expectation $E(W_i), E(W_j)$ and marginal variances $\text{Var}(W_i), \text{Var}(W_j)$ are known from the marginal distributions $F_i, F_j$, since

$$\text{Corr}(W_i, W_j) = \frac{E(W_i W_j) - E(W_i)E(W_j)}{\sqrt{\text{Var}(W_i)\text{Var}(W_j)}}.$$  \quad (2.77)
In addition,

\[
E(W_iW_j) = E(F_i^{-1}(\Phi(Z_i))F_j^{-1}(\Phi(Z_j)))
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_i))F_j^{-1}(\Phi(z_j))\phi_{\rho_{ij}}(z_i, z_j)dz_idz_j, \tag{2.78}
\]

where \(\phi_{\rho_{ij}}(z_i, z_j)\) is the standard bivariate normal probability density function with correlation \(\rho_{ij}\).

In the special case where the \(p\) marginal variates \(W_i, i = 1, \ldots, p\), are identically distributed, i.e., \(W_i \sim F\), the ARTAFACTS software described in Cario and Nelson (1997) provides a direct implementation of the NORTA method. In general, an efficient numerical algorithm is required to obtain the pairwise correlation \(\Sigma_{Z(i,j)}\) for the bivariate Gaussian inputs \(Z_i, Z_j\) so that the produced \(W_i, W_j\) have the desired correlation \(\Sigma_{W(i,j)}\).

To numerically estimate the Gaussian correlation \(\Sigma_{Z(i,j)}\) that matches the gamma correlation \(\Sigma_{W(i,j)}\), we propose a moment-based algorithm combined with the bisection search method. The bivariate correlation matching procedure for each pair \((i,j)\) is carried out according to the ‘bivariate correlation matching algorithm,’ consisting of three separate procedures, as described below.

**Bivariate Correlation Matching Algorithm:**

**Procedure 1:**

**Output:** A matrix, each column of which represents the probability mass function (p.m.f.) of the discretized bivariate normal variable over the fine lattice at a unique correlation \(\rho_k\).

- Create a regularly spaced 2-dimensional fine lattice over the range \([-4, 4]\) with incremental length of 0.01 in both dimensions (probability outside of the range is neglected).
• Arrange the 2-dimensional lattice points into a data matrix, denoted by \( Z = (Z_1, Z_2) \). Each point on the lattice is denoted by \((z_{1k_1}, z_{2k_2})\), with \( k_1, k_2 \) tracking its position along the 1\(^{st}\) and 2\(^{nd}\) coordinate axis, respectively.

• For each \( \rho_k = 0.01 \times k, \ k = 1, \ldots, 100 \), the p.m.f. of the discretized bivariate normal variable at correlation \( \rho_k \) is evaluated as

\[
p_{\rho_k}(z_{1i}, z_{2j}) = \frac{\phi_{\rho_k}(z_{1i}, z_{2j})}{\sum_{k_1=1}^{801} \sum_{k_2=1}^{801} \phi_{\rho_k}(z_{1k_1}, z_{2k_2})} \tag{2.79}
\]

where \( \phi_{\rho_k}(\cdot, \cdot) \) is the p.d.f. of a central bivariate normal with correlation \( \rho_k \).

We mention here that procedure 1 needs to be performed only once. The output quantities are stored and repeatedly used for each pair \((i, j)\). For a \( m \)-variate gamma random vector, the Gaussian correlation matrix in the NORTA implementation is produced by \( \binom{m}{2} \) independent repeated implementation of procedures 2 and 3 for each pair \((i, j)\), \( 1 \leq i < j \leq m \).

**Procedure 2:**

**Input:** bivariate normal correlation \( \rho_k \);

**Output:** the estimated bivariate gamma correlation \( r_{ij}(\rho_k) \);

- Convert the discretized bivariate random vector into bivariate gamma random vector marginally by performing inverse c.d.f. transformation, i.e., \( W_i = F_i^{-1}(\Phi(Z_i)) \), \( i = 1, 2 \);

- Assess the correlation between the gamma components \( W_1 \) and \( W_2 \) as

\[
r_{ij}(\rho_k) = \frac{E(W_1 W_2) - E(W_1)E(W_2)}{\sqrt{Var(W_1)Var(W_2)}}, \tag{2.80}
\]

with \( E(W_i) = \nu_i \) and \( Var(W_i) = 2\nu_i \) and

\[
E(W_1 W_2) = \sum_{k_1=1}^{801} \sum_{k_2=1}^{801} F_1^{-1}(\Phi(z_{1k_1}))F_2^{-1}(\Phi(z_{2k_2}))p_{\rho_k}(z_{1k_1}, z_{2k_2}), \tag{2.81}
\]

where \( p_{\rho_k}(\cdot, \cdot) \) is the \( k^{th} \) column in the matrix obtained from procedure 1.
For each pair $\Sigma_W(i, j)$, the optimal bivariate normal correlation $\rho_k$ (to be referred to as $\Sigma_Z(i, j)$) that produces a bivariate gamma correlation $r_{ij}(\rho_k)$ matching the desired $\Sigma_W(i, j)$ within the tolerable error is searched through a bisection algorithm, described in Procedure 3.

**Procedure 3:**

<table>
<thead>
<tr>
<th>Input:</th>
<th>target bivariate gamma correlation $\Sigma_W(i, j)$;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output:</td>
<td>estimated bivariate normal correlation $\Sigma_Z(i, j)$;</td>
</tr>
</tbody>
</table>

Set $t = 0$ and initialize the upper and lower bounds $\rho_l$ and $\rho_u$ for the bivariate normal vector such that,

$$r_{ij}(\rho_l) \leq \Sigma_W(i, j) \leq r_{ij}(\rho_u). \quad (2.82)$$

Set $r_{ij} = r_{ij}(\rho_l)$;

**Repeat**

- $r_{ij} \leftarrow r_{ij}(\rho_l + \rho_u) / 2$ using Procedure 2;
- $\rho_l \leftarrow \rho_l + \rho_u / 2$ if $r_{ij} \leq \Sigma_W(i, j)$; Otherwise, $\rho_u \leftarrow \rho_l + \rho_u / 2$;
- $t \leftarrow t + 1$;

**Until** $|r_{ij} - \Sigma_W(i, j)| < \varepsilon$;

**Let** $\Sigma_Z(i, j) = \rho_l + \rho_u / 2$;

The kernel of the NORTA algorithm resides in the acquisition of the correlation matrix $\Sigma_Z$ of the multivariate normal random vector $Z$ that matches the desired correlation matrix $\Sigma_W$. In fact, the success of the algorithm in generating i.i.d. replicates of a random vector $W$ with given marginals and correlation matrix $\Sigma_W$ depends largely on the quality of the correlation matrix $\Sigma_Z$, which is the output matrix of the correlation matching procedure and input matrix of the generation procedure. Nonetheless, as the $\Sigma_Z$ matrix is calculated in a component-wise fashion without further adjustment to ensure its non-negativity, it is a common occurrence for the correlation matrix
$\Sigma_Z$ produced by the NORTA algorithm to be an invalid correlation matrix, failing to be non-negative definite. When the output $\Sigma_Z$ obtained through the correlation matching procedure produces an invalid covariance matrix, the NORTA generation procedure cannot proceed with $\Sigma_Z$ as the input correlation matrix. A simulation study by Ghosh and Henderson (2003) shows that an increasingly large proportion of feasible gamma correlation matrices cannot be matched using the NORTA method, as the dimension of the random vector increases. This indicates that a unified remedial action is required for the NORTA procedure to be successful as a general random sampling technique.

To overcome this potential issue, Ghosh and Henderson (2003) suggest a semi-definite program (SDP) that finds a matrix $\Sigma_{Z_{\text{sd}}}^\text{df}$ ”closest” to the invalid $\Sigma_Z$ and is non-negative definite. The matrix of $\Sigma_{Z_{\text{sd}}}^\text{df}$ used as the input matrix for the NORTA generation procedure. Formally, suppose that the input Gaussian correlation matrix $\Lambda_Z$ is viewed as the solution of the semi-definite programming problem with linear objective function.

$$\min d(\Sigma_Z, \Lambda_Z),$$

subject to $\Lambda_Z \succeq 0 \& \Lambda_Z(i, i) = 1$ for all $i$.

(2.83)

The metric $d(\cdot, \cdot)$ as a measure of dissimilarity between two matrices can be chosen in numerous ways. Common choices as suggested in Ghosh and Henderson (2003) are the $L_1$ metric,

$$d(\Sigma_Z, \Lambda_Z) = \sum_{i>j} |\Sigma_Z(i, j) - \Lambda_Z(i, j)|,$$

(2.84)

and the $L_\infty$ metric,

$$d(\Sigma_Z, \Lambda_Z) = \max_{i>j} |\Sigma_Z(i, j) - \Lambda_Z(i, j)|,$$

(2.85)
with \( \Sigma_Z(i, j) \) referring to the \((i, j)^{th}\) entry of the \(\Sigma_Z\) matrix. The performance of this adjustment strategy is empirically evaluated by examining the closeness of the NORTA correlation matrix of the generated random vector \( W \) to the desired correlation matrix. Numerical studies conducted in [Ghosh and Henderson (2003)] indicate that the NORTA sampling method produces random vectors with SDP correlation matrices that are not discernibly different from the desired ones, even in high dimensions.

Instead, we exploit an alternative commonly used approach in obtaining a non-negative definite matrix ‘close’ to \( \Sigma_Z \). The legitimate correlation matrix \( \Lambda_Z \) is simply produced by discarding the principle components associated with negative eigenvalues. Formally, suppose that a spectral decomposition of \( \Sigma_Z \) yields \( \Sigma_Z = P \Lambda \Sigma P^T \), the modified correlation matrix \( \Lambda = P_+ \Lambda_+ P_+^T \), where \( P_+ \) consists of the leading eigenvectors corresponding to positive eigenvalues, diagonal elements of \( \Lambda_+ \). The ‘closeness’ is defined in the sense of common principal directions as well as strength along each direction.

The NORTA algorithm coupled with a remedial action provides an efficient method of sampling from a gamma random vector with specified marginals and correlation matrix, defining the denominator of the desired \( T \). Compared to the classical multivariate \( t \), this generalized \( t \) class provides a more open environment for capturing the true mechanism that generates the data.

### 2.2.2 Generation from a Generalized Multivariate \( t \) Random Vector

We now return our attention to the generation of random samples from a central generalized multivariate \( t \) random vector \( T \in R^p \sim \text{MVT} (\Sigma_X, \Sigma_W, \mu = 0, \nu = (\nu_1, \ldots, \nu_p)) \).
Let \( x \in \mathbb{R}^p \) denote a realization of the multivariate Gaussian \( X \sim MVN(0, \Sigma_x) \), defining the numerator of the generalized \( t \). In addition, let \( w \in \mathbb{R}^p \) denote a sample of the multivariate gamma \( W \), approximately represented by the marginals \( W_i \sim \text{Gamma}(\nu_i/2, 2), \forall i \) and correlation matrix \( \Sigma_W \). Thus, a sample \( t \in \mathbb{R}^p \) from the target generalized \( t \) is simply constructed as

\[
t = \frac{x}{\sqrt{w/\nu}} \in \mathbb{R}^p,
\]

where the divisions are element-wise divisions. A random sample \( t_1, t_2, \ldots, t_N \) from the generalized \( t \) is obtained by \( N \) independent implementations of this procedure.

We remark that the ‘correlation matching procedure’ in the NORTA algorithm is executed only once to find the correlation matrix of the input Gaussian vector, for a definite set of gamma variables defining the denominator of the generalized multivariate \( t \). The generation procedure is carried out repeatedly.

### 2.2.3 Probability Calculation with the Table 0/1 Approach

The Table 0/1 approach is introduced as a basic numerical algorithm for the probability calculation of the \( p \)-variate classical multivariate \( t \) probability within any set \( A \in \mathbb{R}^p \). The target probability is estimated as the proportion of a large number of independently generated variates, say \( N \), from the \( t \) distribution that fall in the set \( A \). In fact, the notion of the Table 0/1 approach in both probability calculation and quantile evaluation carries over to the generalized multivariate \( t \) setting, provided that a random sample can be easily acquired.

Suppose that a random sample \( \{t_k \in \mathbb{R}^p, k = 1, \ldots, N\} \) is generated and arrayed in a two way table consisting of \( p \) rows and \( N \) columns. The generalized \( t \) probability
over a hyper-rectangle $A = \{ t \mid a_i \leq t_i \leq b_i, i = 1, \ldots, p \}$ is estimated as

$$\hat{P}(A) = N^{-1} \sum_{k=1}^{N} \prod_{i=1}^{p} I(a_i \leq t_{ik} \leq b_i).$$

### 2.2.4 Probability Calculation with Spherical Radial Transformation

Apart from the Table 0/1 approach, the MVN implementation of the spherical radial transformation method can be easily adapted to the generalized multivariate $t$ setting. However, the MVT implementation that makes transformation directly on the probability density function is not applicable to this general framework, as a nice closed form expression for the p.d.f. is not obtainable. The target probability is estimated by averaging over a sufficiently large number, say $N$, of independent $G$ values corresponding to distinct uniformly distributed radial directions $z_k \in U^p$ and the gamma sample $w_k$. That is,

$$\hat{P}(A) = N^{-1} \sum_{k=1}^{N} G(a, b, z_k, \nu, C, w_k),$$

where $G(a, b, z_k, \nu, C, w_k) = F_G(u^2(z_k, w_k)) - F_G(l^2(z_k, w_k))$, and spherical limits $u(z_k, w_k)$ and $l(z_k, w_k)$ are calculated as

$$u(z_k, w_k) = \max(0, \min_{s_i > 0} \{d_i/s_i\}, \min_{s_i < 0} \{c_i/s_i\})$$

and

$$l(z_k, w_k) = \max(0, \max_{s_i > 0} \{c_i/s_i\}, \max_{s_i < 0} \{d_i/s_i\}),$$

provided that $s = Cz_k$, $c = a\sqrt{w/\nu}$ and $d = b\sqrt{w/\nu}$.

This method yields an estimate with much less uncertainty compared to the basic Table 0/1 approach. We recommend the spherical radial transformation approach.
Figure 2.7: The Updating Procedure in the Quantile Calculation for a Generalized Multivariate $t$ Distribution Using the Spherical Radial Transformation Technique, with Colors Representing the Different Sizes of the Small Tables Used in Combination with the Bisection Algorithm in Phase 1

Later, we will conduct simulation study to demonstrate the improved efficiency of the spherical radial transformation approach over the Table 0/1 approach.
Chapter 3: Simultaneous Inference Procedure

As with the univariate test for the significance of a single hypothesis, multiple testing for the significance of more than one scientific hypotheses simultaneously has become an increasingly popular statistical technique with a wide array of applications in many research fields. For example, in agricultural field studies, it is a standard task to make joint inferences on the effectivenesses of several different types of fertilizers compared to a pre-determined control group. Furthermore, simultaneous assessment of multiple dose levels of a fertilizer to investigate the dose-response relationship is often of interest in order to identify an optimal dosage that produces the best trial results, e.g., the highest crop yield. Successful applications of multiple hypothesis test are demonstrated in many other scientific fields as well.

3.1 Definition of the Family-wise Error Rate

When multiple scientific questions are addressed simultaneously, multiple inference procedures need to be utilized to adjust for the multiplicity effect in order to control the type I error of the family of hypotheses at a prescribed level $\alpha$. However, the concept of the type I error rate, uniquely defined as the probability of rejecting the null hypothesis when it is true for the single hypothesis testing problem, is not extended uniquely to the multiple hypotheses case. In fact, various definitions of
overall type I error rate exist that measure the probability of making at least one erroneous inference of any true null hypothesis.

To provide formal mathematical definitions of the various measures of type I error rates, we consider simultaneously testing \( m \) individual hypotheses specified as

\[
H_0 : \cap_{i=1}^{m} H_{0i} \quad \text{vs.} \quad H_1 : \cup_{i=1}^{m} H_{1i} \tag{3.1}
\]

provided that each individual hypothesis \( H_{0i} \) is formulated in terms of appropriate model parameters \( \mu \) in the form of

\[
H_{0i} : \mu \in A_i \quad \text{v.s.} \quad H_{1i} : \mu \notin A_i \tag{3.2}
\]

The overall type I error rate is defined as the probability of falsely rejecting at least one true null hypotheses \( H_{0i} \). The family-wise error rate (FWER) is said to be controlled at the specified level \( \alpha \) in a weak sense, if the error probability is calculated under the global null hypothesis \( H_0 \) that all \( m \) null hypotheses \( H_{0i} \) are simultaneously true. Mathematically,

\[
\text{FWER} = \sup_{\{\mu \in \cap_{i=1}^{m} A_i\}} P(\text{reject at least one } H_{0i}, i = 1, \ldots, m) \tag{3.3}
\]

We note that the weak control of the FWER preserves the overall error rate under the restrictive condition that all null hypotheses are true. From a practical point of view, the assumption that all \( m \) null hypotheses are simultaneously true is somewhat restrictive and difficult to verify, which prompts a relaxation on this assumption to more appropriately preserve the overall Type I error rate at any arbitrary set of true null hypotheses. For instance, in situations where only a subset of the \( m \) null hypotheses is true, the type I error rate may be smaller than the intended FWER.
or larger than the intended FWER which makes the weak control of the FWER an
inadequate principle in protecting the overall error rate.

A relaxation on the assumption that all $m$ null hypotheses are true gives rise to a
more appealing and practical concept of the type I error rate. The strong control of
the FWER is said to be achieved at level $\alpha$, if the probability of incorrectly rejecting
at least one true null hypotheses $H_{0i}, i \in T$ regardless of which and how many other
hypotheses are true is kept under $\alpha$, i.e.,

$$\sup FWER = \max_T \sup_{\mu(T)} P(\text{reject at least one } H_{0i}, i \in T) \leq \alpha$$  \hspace{1cm} (3.4)

given that the maximum is taken over all $2^m$ possible configurations of the true null
hypotheses. For a particular configuration, let $T$ be the index set of the collection of
the true null hypotheses. Accordingly, the supremum is taken over the null parameter
space $\mu(T)$ formed by the set of the true null hypotheses $T$ and its complement
$T^C = \{1, 2, \ldots, m\} \setminus T$ consisting of the collection of false null hypotheses. Specifically,

$$\mu(T) = \{\mu : \mu \in \bigcap_{i \in T} A_i \text{ and } \mu \in \bigcap_{i \in T^C} A_i^C\}$$

It can be argued that, for any given decision rule (joint rejection region), the
following inequality holds generally,

$$\sup FWER \geq FWER.$$  \hspace{1cm} (3.5)

Consequently, it can be inferred from (5) that multiple inference procedures that
control the family-wise error rate in a strong sense automatically achieve the FWER
control in the weak sense.

The closed testing principle is generally used to construct multiple testing proce-
dures that control the FWER in the strong sense, taking the local test procedures
that weakly control the FWER as the building block. The test accuracy of the resulting procedure depends greatly on the appropriateness of the local test procedure. In the subsequent endeavor, we focus on the investigation of local procedures that attempt to closely preserve the nominal significance level $\alpha$ in the weak sense under the condition that all $m$ null hypotheses are simultaneously true.

3.2 Multiple Hypotheses Testing with Constant Variance

When multiple hypotheses are simultaneously evaluated as a family, multiplicity adjustment becomes an inherent problem that aims at preserving the FWER at some pre-specified level $\alpha$. Conventionally, the hypothesis test problem, i.e., the specification of the null and alternative statements, is explicitly formulated to guard against the false rejection of the null hypothesis unless sufficient evidence is provided to declare its falsity. In many empirical applications, falsely rejecting a true null hypothesis typically results in more severe consequences/losses to the decision maker than failing to reject a false null hypothesis. This implicit argument dictates the strong need for the multiplicity adjustment to be taken so that the FWER is maintained at size $\alpha$. Furthermore, if an approximate procedure needs to be used instead, it is more desirable to approximate it in a conservative fashion so that the actual error rate is well kept under the threshold $\alpha$ than an anti-conservative approximation with inflated error rate.

It is widely known that simply testing each individual hypothesis at level $\alpha$ usually produces rather poor simultaneous joint inferences, with greatly inflated FWER
ranging anywhere from $\alpha$ to $m\alpha$. Obviously, this brute force test procedure is, qualitatively, too liberal, with the actual error rate surpassing the preset level $\alpha$. Quantitatively, the actual magnitude of the excess error depends upon the number of the individual hypotheses and the degree of overlap among the decision regions for the $m$ individual hypotheses. To ensure that the type I error rate does not exceed the tolerable error level $\alpha$, numerous simultaneous inference procedures have been proposed in the multiple testing literature to adjust for the multiplicity effect.

3.2.1 Bonferroni Procedure

One of the earliest solutions to deal with the multiple hypotheses problem is the widely applied Bonferroni (1936) procedure, which tests each null hypothesis individually by an appropriate univariate test procedure at significance level $\alpha/m$. As a result of the Bonferroni inequality, the Bonferroni procedure is guaranteed to be conservative without regard to the joint distribution of the $m$ marginal test statistics. Thus, the Bonferroni method is also known as a non-parametric test procedure. Several Bonferroni-based stepwise procedures, e.g., Holm’s procedure (Holm, 1979) and Hochberg’s procedure (Hochberg, 1988), are conservative approaches that are uniformly more powerful than the single-step Bonferroni method. Despite the improvement in test power over the Bonferroni procedure, their performance can be inadequate when the number of hypotheses is large or the test statistics exhibit strong dependence, due to the ignorance of how the $m$ test statistics are jointly distributed.

3.2.2 Classical Multiple Testing Procedures

To alleviate the conservativeness and further elevate the test power, a number of classical multiple comparison procedures (MCPs) that account for the dependence
structure among the collection of individual test statistics were developed successively in the mid-twentieth century, including Scheffe’s procedure (Scheffe, 1953) for all linear estimable functions, Tukey’s all pairwise comparisons (Tukey, 1949), Dunnett’s treatments versus control comparisons (Dunnett, 1955) and Hsu’s method (Hsu, 1984) for multiple comparisons with the unknown best.

All of these classical procedures were originally proposed in the context of the basic one-way analysis of variance (ANOVA) setting with homogeneous variance assumption. In contrast to Scheffe’s procedure that is associated with a univariate $F$ test statistic, the multiplicity-adjusted critical values for Tukey’s and Dunnett’s methods were calculated as quantile vectors via numerical integration for the associated multivariate reference distributions. Nevertheless, before the advent of accelerated computation, the numerical probability calculation for multivariate distributions was restricted to only some special cases, i.e., the studentized range distribution, induced by Tukey’s pairwise comparisons with equal sample sizes, and multivariate $t$ distributions with the special compound symmetry correlation structure induced by the Dunnett’s all treatments versus control comparisons with equal sample sizes across all treatment groups, except for the control group. Tukey’s all pairwise comparisons and Dunnett’s all treatments versus control comparisons, under a balanced one-way design, exemplify the many common forms of the multiple hypothesis testing problem. With modern computing power, a general test procedure for any arbitrary set of linear hypotheses under the general linear model framework with the equal variance assumption, can be conceptually formulated with ease of calculation and preservation of the pre-specified significance level $\alpha$. 
This section reviews the frequently used classical multiple comparison procedures in the basic one-way layout setting with the homogeneous variance assumption. Moreover, extensions to the general linear regression framework that allows for arbitrary patterns of sample sizes and adjustments for covariates and/or other factors, under the homogeneous variance assumption, will be provided in the subsequent section.

Suppose that $Y_{ij}$ is the measurement taken on the $j^{th}$ subject from the $i^{th}$ treatment group and is assumed to be independently normally distributed with mean $\mu_i$ and constant variance $\sigma^2$, $i = 1, \ldots, \nu;$ $j = 1, \ldots, n_i$. Accordingly, $Y_{ij}$ can be equivalently represented by a basic one-way layout model as such,

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2). \tag{3.6}$$

The least squares estimates of the treatment means $\mu_i, i = 1, \ldots, \nu$, and the error variance $\sigma^2$ in model (3.6) are given by,

$$\hat{\mu}_i = \bar{Y}_i,$$

$$\hat{\sigma}^2 = \frac{1}{\nu \sum_{i=1}^{\nu} (n_i - 1) \sum_{i=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}.$$ 

Consider any collection of $m$ linear estimable functions specified through an $m \times \nu$ contrast coefficient matrix $L$, i.e., $L\mu = (l_1^T \mu, l_2^T \mu, \ldots, l_m^T \mu)^T$, where $l_k^T$ denotes the $k^{th}$ row of the $m \times \nu$ coefficient matrix $L$. It is a standard result that the least squares estimate $L\hat{\mu}$ follows an $m$-variate Gaussian distribution as

$$L\hat{\mu} \sim N(L\mu, \sigma^2 L \Sigma \mu L^T), \tag{3.7}$$

given that $\Sigma \mu$ is the variance-covariance matrix of the estimated mean vector $\hat{\mu}$. That is, $\Sigma \mu = \sigma^2 \text{diag}\{1/n_i, i = 1, \ldots, \nu\}$. The estimator $L\hat{\mu}$ is statistically independent
of the pooled variance estimate $\hat{\sigma}^2$, $\chi^2$ distributed with $df = \sum_{i=1}^{\nu}(n_i - 1)$ degrees of freedom,
\[ df \cdot \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{df}. \] (3.8)

These distributional properties provide the theoretical basis for the derivations of the classical multiple test procedures. Our traditional statistical way to examine any significant differences among the treatments consists of two steps: (1) Scheffe’s procedure is performed as a preliminary test for an overall significant difference across all treatment means, i.e., $H_0: \mu_0 = \mu_1 = \ldots = \mu_{\nu}$; and (2) if global heterogeneity of the treatment means is detected by Scheffe’s test at the pre-specified level $\alpha$, then an appropriately chosen multiple comparison procedure for a suitable family of comparisons of interest is carried out to further test local hypotheses, which provide more detailed information regarding the source of the overall significant difference.

1. Scheffe’s Procedure

Scheffe’s procedure (Scheffe, 1953) was designed to make simultaneous inferences on the family of all linear estimable functions of a set of treatment means. Suppose that $\mathcal{L}^m$ is the $m$-dimensional linear subspace spanned by the rows of an $m \times \nu$ full rank coefficient matrix $L$. Scheffe’s procedure yields an exact level $\alpha$ test for the set of all linear estimable functions formulated as $H_0: \hat{l}^T \mu = 0, \forall l \in \mathcal{L}^m$, and rejects each individual hypothesis $\hat{l}^T \mu = 0$ if
\[ \frac{|\hat{l}^T \mu|}{\sqrt{\hat{l}^T D \hat{l}}} > \sqrt{k \hat{\sigma}^2 F_{k, \nu-\nu, \alpha}} \]
for any arbitrary $l \in \mathcal{L}^m$, provided that $D = \text{diag}\{1/n_i, i = 1, \ldots, \nu\}$. 

86
Importantly, Scheffe’s procedure places equal importance on every individual hypothesis in the family of all linear estimable functions. Nonetheless, in practical applications, it is rarely the case that all linear estimable functions of the set of means are of equal interest to the analyst. On the contrary, it is more often the case that an explicit set of hypotheses that closely reflect the primary objectives of the study is formed a priori. Therefore, Scheffe’s test procedure which treats each member (or any subset) of the family of infinitely many hypotheses symmetrically often yields a conservative control of the type I error rate for the hypotheses of interest. This renders it an inadequate procedure for simultaneous inference on a finite set of individual comparisons. Thus, a specialized multiple test procedure that focuses on the pre-defined finite set of hypotheses is needed to reach conclusive decisions on each local test, while maintaining the joint error rate at $\alpha$.

Various multiple comparison procedures, each designed for a specific family of linear comparisons, were introduced soon after the development of Scheffe’s method. The two such procedures that have been most thoroughly studied in the multiple comparison literature are Tukey’s all pairwise comparisons and Dunnett’s treatments versus control comparisons. In addition, A more recently developed procedure that contrasts all treatment means with the unknown best by Hsu (1984) receives much attention as well. However, Hsu’s procedure does not fit into the general framework that we present here and needs to be considered separately from other approaches, due to the non-linearity of the hypotheses.

2. Tukey’s Procedure

Tukey’s procedure (Tukey 1949) was designed to simultaneously assess the family of all pairwise comparisons across a set of treatment means, in the context of model
(3.6), formulated as

\[ H_0 : \mu_i = \mu_j, \text{ for each pair } (i, j) \text{ versus } H_a : \mu_i \neq \mu_j, \text{ for at least one pair } (i, j) \]

(3.9)

with an additional constraint of equal sample size, i.e., \( n_i = n, i = 1, \ldots, \nu \). The standard multivariate test statistic associated with (3.9) is given by

\[ T = \begin{cases} T_{ij} = \frac{\hat{\mu}_i - \hat{\mu}_j}{\hat{\sigma} \sqrt{1/n_i + 1/n_j}}, & 1 \leq i < j \leq \nu \end{cases} \]

(3.10)

Assume that each member in the family of all pairwise comparisons is of equal importance in the sense that each individual hypothesis carries an equal portion of significance level \( \alpha \), the decision rule for each individual hypothesis \( H_{ij} : \mu_i = \mu_j \) is in the form that

\[ \frac{|\hat{\mu}_i - \hat{\mu}_j|}{\hat{\sigma} \sqrt{1/n_i + 1/n_j}} > Q_T(1 - \alpha) \]

for each pair \((i, j)\), given that the multiplicity-adjusted critical value \( Q_T(1 - \alpha) \) is calculated as the 100\(\alpha\)th equal coordinate quantile of the joint distribution of the test statistic \( T \) under \( H_0 \), which is a \((\nu)^2\)-variate \( t \) distribution with correlation structure identical to the correlation induced by the multivariate Gaussian vector and \( \sum_{i=1}^{\nu} (n_i - 1) \) degrees of freedom. With the establishment of the theoretical reference distribution for \( T \), exact implementation of the procedure depends on the appropriate calculation of the multiplicity-adjusted critical vector, evaluated as the multivariate quantile of the associated multivariate \( t \) distribution. We refer back to Section 2.1 in which an efficient numerical algorithm for the probability and quantile calculation of the multivariate \( t \) distribution is carefully investigated.

Prior to the recent work on efficient numerical computation of multivariate \( t \) probabilities, the multiplicity-adjusted critical vector associated with Tukey’s all pairwise
comparison at the significance level $\alpha$ was obtained mainly through tables, achievable only for the balanced one-way design with $n_i = n$, $i = 1, \ldots, \nu$. The resulting reference distribution is a studentized range distribution with parameters $\nu$ and $\sum_{i=1}^{\nu} (n_i - 1)$, a special case of the multivariate $t$ distribution. For a balanced design, Tukey’s procedure provides an exact level $\alpha$ test in the sense that the FWER is retained at the pre-specified level $\alpha$. Tukey’s procedure serves as a successful example of how the test accuracy can be enhanced by exploiting the joint distribution of the test statistics, which automatically captures the dependence structure among the test statistics.

In addition, [Gabriel (1970)] showed that Tukey’s procedure gives the shortest intervals for every pairwise difference among the procedures having joint confidence level $\geq (1 - \alpha)$. Furthermore, as discussed in [Kunte and Rattihalli (1984)], the joint confidence region produced by Tukey’s procedure has the minimum volume among all $(1 - \alpha)$-level simultaneous confidence regions for all pairwise differences.

3. Dunnett’s Procedure

Dunnett’s method [Dunnett (1955)] was established to compare the treatment effects, in a directional manner or not, with a pre-determined control group, in model (3.6), formulated as

$$H_0 : \mu_i = \mu_\nu, \ i = 1, \ldots, \nu - 1 \ vs. \ H_a : \mu_i \neq \mu_\nu, \ for \ at \ least \ one \ i = 1, \ldots, \nu - 1,$$

(3.11)

under the constraint that $n_i = n$, $i = 1, \ldots, \nu - 1$, assuming that $\mu_\nu$ is the mean of the control group. The standard multivariate test statistic associated with the multiple hypotheses (3.11) is given by

$$T = \left\{ T_i = \frac{\hat{\mu}_i - \hat{\mu}_\nu}{\hat{\sigma} \sqrt{1/n_i + 1/n_\nu}}, \ i = 1, \ldots, \nu - 1 \right\}.$$

(3.12)
The joint distribution of \( T \), under \( H_0 \), has a \( (\nu - 1) \)-variate \( t \) distribution with correlation structure \( R = \{ r_{ij}, 1 \leq i, j \leq \nu - 1 \} \) as the multivariate Gaussian random vector, where

\[
r_{ij} = \frac{1/n_\nu}{\sqrt{1/n_i + 1/n_\nu}} \sqrt{1/n_j + 1/n_\nu}
\]

and \( \sum_{i=1}^{\nu} (n_i - 1) \) degrees of freedom. Dunnett’s test rejects each null individual hypothesis if

\[
\frac{|\hat{\mu}_i - \hat{\mu}_\nu|}{\hat{\sigma} \sqrt{1/n_i + 1/n_\nu}} > Q_T(1 - \alpha),
\]

provided that \( Q_T(1 - \alpha) \) is the 100(1 - \( \alpha \))\( ^{th} \) equal coordinate quantile of the \( (\nu - 1) \)-variate multivariate \( t \) distribution with correlation structure \( R \) and \( \sum_{i=1}^{\nu} (n_i - 1) \) degrees of freedom. As with Tukey’s procedure, due to computational limitation, exact implementation of the procedure was feasible only for a special pattern of sample size — that is, equal sample sizes across all treatment groups except for the control group — considered by Dunnett (1955). The resulting multivariate \( t \) distribution with compound symmetry correlation structure is tabulated through numerical integration.

Under the restrictive experimental condition, Dunnett’s procedure provides an exact level \( \alpha \) test for the family of treatments versus control. In other words, the probability of concluding that any treatment differs significantly from the control when no difference is present is retained at \( \alpha \) within the allowable computational error.

It may be remarked that Tukey’s and Dunnett’s procedures, both considered the traditional MCPs along with Scheffe’s procedure, were specifically designed to make simultaneous inferences on specific families of linear comparisons in the simple one-way ANOVA framework with constant variance assumption. The added constraint on the sample size configuration — equal sample size for Tukey’s test and partially
equal for Dunnett’s test — was further enforced to cope with the computational difficulty in probability calculation of the associated reference distribution with general correlation structures, before better computation was available. Clearly, in spite of its strong theoretical performance and easy implementation through tables, the set of the conditions under which Tukey’s and Dunnett’s procedures were developed is rather restrictive, rendering both procedures of limited applicability to many practical situations.

In real data applications, departures from the set of working conditions could originate from a variety of possible sources, ranging from the experimental conditions related to the sample size allocation to the model construction that most adequately fits the data. The equal sample size assumption, in a complete or partial sense, was mainly imposed for the ease of computation, as opposed to the theoretical justification. In fact, the presence of unbalanced data is very common in modern research areas, even for studies that are originally planned to generate balanced data, due to a variety of possible reasons, e.g., subject dropout. Before the invention of efficient numerical algorithms for the multivariate $t$ probability calculation, a considerable amount of research was conducted by individual researchers, striving to modify Tukey’s and Dunnett’s procedures to accommodate arbitrary sample size patterns. Nevertheless, among the numerous conservative procedures that emerged in the 1980s, accurate extensions to unbalanced designs with exact numerical implementation was successful only for the special case of the family of all pairwise comparisons among three treatment groups. In recent years, with the dramatically increased computational power and numerical availability in the evaluation of the multivariate $t$ reference distributions with arbitrary parameters, the notion of unbalanced data, a stumbling
block in multiple testing prior to the computer era, is conceptually indistinguishable from that of balanced data in the multiple comparison practice in the present age. For the sake of unification, detailed discussion on the practical implementation of the traditional multiple comparison procedures will be deferred to the next section.

In addition, the fact that Tukey’s and Dunnett’s procedures were derived in the basic one-way ANOVA framework places an inherent limitation on their practical use. The one-way ANOVA model treats the treatment factor of interest as the sole source of variation for the measurement variability, without adjusting for the covariates/factors potentially related to the response variable. It is commonly agreed among statisticians that failing to include important variables in the model can seriously diminish both the descriptive and predictive power of the model and renders the subsequent inferential procedure untenable. In accordance with this viewpoint, we adopt the general principle to construct the model that most adequately describes the underlying data generating process, and formulate the hypotheses that best reflects the scientific questions of interest. This argument suggests the necessity of extending the traditional MCPs under the one-way layout to the general linear modeling framework that allows the adjustment for possible covariates and/or other controlling factors collected along with the response $Y$.

Moreover, Tukey’s and Dunnett’s procedures are specifically designed for two special families of linear comparisons — all pairwise comparisons for the former and all treatments versus control comparisons for the latter. The set of linear comparisons needs to be tailored to the specific study. We do not restrict our attention to such contrasts of means. In fact, the numerical availability of the multivariate $t$ probability calculation ensures that the scope of the applicability of the MCPs can be broadened
to any finite families of linear functions of the treatment means, under the general linear model framework.

### 3.2.3 Multiple Testing under the General Linear Model Framework

This section is devoted to the development of a simultaneous inference procedure for any finite collection of linear estimable functions of the model parameters, in the context of general linear modeling framework with the homogeneous variance assumption, which accommodates arbitrary sample size pattern automatically. The methodological framework is very general and thus applicable to a wide range of multiple testing problems, including Tukey’s and Dunnett’s procedures as special cases. This general multiple comparison approach accounts for the stochastic dependencies among the test statistics. Consider the general linear regression model,

\[
Y = X\beta + \varepsilon, \tag{3.14}
\]

where \(Y\) is the \(n\)-dimensional response vector, \(X\) is a \(n \times p\) fixed design matrix, \(\beta\) is a \(p \times 1\) unknown parameter vector, and the error term \(\varepsilon\) is an \(n \times 1\) random vector consisting of i.i.d. standard Gaussian variates with mean 0 and constant variance \(\sigma^2\). This general framework enables one to account for some important covariates in order to more precisely assess the quantities of primary concern, e.g., pairwise treatment differences and differences between treatments and control. In the absence of significant covariates, model (3.14) simplifies to the basic ANOVA specified in (3.6). The discussion below provides a unified treatment of multiple comparisons in the general linear model setting. To set the groundwork for the presentation of
the general test procedure, we review some basic statistical properties of the model parameter estimates.

Suppose that the design matrix $X$ is full rank, the least squares estimate for $\hat{\beta}$ is given by

$$\hat{\beta} = (X'X)^{-1}X'Y$$

and the unbiased estimate for the constant variance $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^T(Y - X\hat{\beta})}{n-p}.$$ 

It follows by standard arguments that the unbiased least squares estimate $\hat{\beta}$ has a $p$-variate Gaussian distribution given as

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

and is statistically independent of the estimated variance $\hat{\sigma}^2$, distributed as

$$(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}.$$

Suppose that a family of hypotheses of arbitrary size $m$ is formulated as linear functions of the model parameters $\beta$ through an $m \times p$ constant matrix $L$, with $\text{rank}(L) \leq \min(m, p)$. Let $l_i^T$ denote the $i^{th}$ row of $L$. We cast our discussion in terms of the two-sided testing problem. Similar results hold for one-sided testing problems.

The two-sided test is formally formulated as

$$H_0 : L\beta = 0 \quad \quad H_a : L\beta \neq 0.$$ 

The standard $m$-variate test statistic $T$ is given as

$$T = \left\{ T_i = \frac{l_i^T\hat{\beta}}{\hat{\sigma} \sqrt{l_i^T(X^TX)^{-1}l_i}}, i = 1, \ldots, m \right\}. \quad (3.15)$$
The joint distribution for $T$, under $H_0$, follows an $m$-variate $t$ distribution with $n-p$ degrees of freedom and the associated correlation matrix is given by $R = \{r_{ij}, 1 \leq i, j \leq m\}$, with

$$r_{ij} = \frac{l_i^T(X^TX)^{-1}l_j}{\sqrt{l_i^T(X^TX)^{-1}l_i} \sqrt{l_j^T(X^TX)^{-1}l_j}}.$$  

Moreover, each marginal test statistic $T_i$ associated with the null hypothesis $H_{0i}$: $l_i^T\beta = 0$ is a univariate $t$ with $n-p$ degrees of freedom. By making use of the joint distribution of the test statistics, simultaneous inference on the set of $m$ null hypotheses is ensured to be controlled at the pre-set level. The decision rules for the $m$ tests are jointly formed, taking the common form that the $i^{th}$ individual hypothesis is rejected, provided that the marginal test statistic $|t_i| > \xi_i$, $i = 1, \ldots, m$, where $\xi = \{\xi_i, i = 1, \ldots, m\}$ is the set of multiplicity-adjusted critical values associated with significance level $\alpha$, calculated as the quantile of the reference distribution. It is worthwhile to note that this general test procedure is theoretically exact with the accuracy determined by that of the quantile computation of $\xi$.

In a multiple hypotheses test, it is a common occurrence that the set of simultaneously assessed hypotheses is of unequal importance in the sense that erroneous rejection of any true null hypothesis would incur consequences that might differ from one another to a noticeable extent. Prompted by this need, we introduce the notion of ‘$\alpha$ allocation,’ denoted by $\kappa_i$ to quantify the relative importance of the $m$ hypotheses. Suppose that the local significance level for hypothesis $H_{0i}$ is denoted by $\alpha_i$ and is defined as

$$\alpha_i = P(|T_i| > \xi_i).$$

To maximize the test procedure’s capacity in protecting against false rejections on hypotheses that incur the most harmful consequences, it is a sensible strategy to
allow the individual error rate (local significance level) to be inversely related to the potential loss caused by an erroneous inference, while maintaining the family-wise error rate at level $\alpha$. Motivated by this need, we assume that the $m$ local error rates are deterministically related in the simple functional form

$$\alpha_1 = \kappa_i \alpha_i.$$  

Clearly, the relative importance measure $\kappa_i$ provides a practical way of bringing the loss information into an inference procedure to better facilitate the generation of effective and statistically more powerful decisions. Moreover, the specification of $\kappa_i$ is usually not an overwhelming task for the decision maker, especially with its physical interpretation as the relative importance of hypothesis $i$ compared with hypothesis 1.

An analogue to the notion of the $p$-value in the univariate case is the adjusted $p$-value, a key terminology in multiple hypothesis testing in addition to the critical vector. Like the $p$ value, the adjusted $p$-value tells directly the significance of the associated hypothesis, i.e., the hypothesis is inferred to be significant if the adjusted $p$-value is less than $\alpha$. However, in contrast to the univariate case, the adjusted $p$-value is unique to each individual hypothesis. The decision rule for each hypothesis $i$ formed by the use of the adjusted $p$-value is the same as that by the critical values, leading to the rejection of the hypothesis if the test statistic $t_i$ exceeds the critical value $\xi_i$. The critical vector satisfies

$$P(T < \xi) = 1 - \alpha$$

$$P(|T_1| > \xi_1) = \kappa_i P(|T_i| > \xi_i), \quad i = 2, \ldots, m.$$  

The adjusted $p$-value is calculated as the probability of the null distribution outside of the hyper-rectangle $[-t_1^{(i)}, t_1^{(i)}] \times [-t_2^{(i)}, t_2^{(i)}] \times \ldots \times [-t_m^{(i)}, t_m^{(i)}]$, where $t_k^{(i)}$ is the value that gives tail probability for the $k^{th}$ marginal distribution equal to that for the $i^{th}$
hypothesis corresponding to $|t_i|$. 

$$\kappa_i P(|T_i| > t_i) = \kappa_k P(|T_k| > t_k^{(i)}), \quad k = \{1, 2, \ldots, m\}$$

Explicitly, the adjusted $p$-value for hypothesis $i$ is given by

$$p_i = 1 - P_0(\max_{k=1,\ldots,m} |T_k| < t_k^{(i)}). \quad (3.16)$$

In addition, the adjusted $p$-value is also interpreted as the smallest significance level at which one would rejects the individual hypothesis $i$, while keeping the family-wise error rate maintained at $\alpha$.

In the early years (prior to 1992), the multivariate $t$ probability calculation relied heavily on tables, available for a selection of specific distributions as mentioned previously. Over the recent decades, the development of various fast numerical algorithms in probability calculation of the multivariate $t$ distributions with any arbitrary correlation structure has cleared out the computational blocks, making exact implementation of the general multiple testing procedures achievable. In addition, the theoretical and numerical studies of the multivariate $t$ distribution has greatly expanded our ability in the exploration of many other aspects of statistical applications, e.g., robust regression and outlier detection. [Hsu (1992)] proposed the factor analytic approach that requires approximating the correlation structure by the closest one-factor (or higher order) structure correlation matrix. [Genz and Bretz (2002)] presented several Monte Carlo algorithms, namely an acceptance-rejection sampling algorithm, a spherical-radial transformation approach and a separation-of-variables approach. Furthermore, in this dissertation work, we adopt the spherical-radial transformation as a variance reduction technique in the numerical approximation of the
classical multivariate $t$ probability over any convex set with slightly different implementation from Genz and Bretz (2002), as was thoroughly investigated in Section 2.1.3.

The separation-of-variables algorithm introduced by Genz and Bretz (2002) is routinely available in R in the ‘mvtnorm’ package, including, for example, the ‘pmvt()’ function that calculates the probability of any hyper-rectangles and the ‘qmvt()’ function that generates the equal coordinate quantile vector for any given probability $p \in (0, 1)$. However, the implementation is restricted to the multivariate $t$ distributions with only integer degrees of freedom, mismatching the intellectual argument provided by the authors that the algorithm works generally for the classical multivariate $t$ distribution with any number of degrees of freedom and an arbitrary correlation matrix. Despite its adequacy in multiple hypotheses testing with the homogeneous variance assumption, the under-delivery of the methodology in computer implementation shows a clear deficiency in many other applications, where fractional degrees of freedom are the norm, e.g., the heterogeneous variance case as will be presented shortly. In addition to the constraint on integer degrees of freedom, the ‘qmvt()’ function calculates the multiplicity-adjusted critical vector by evaluating the equal-coordinate quantile vector yielding the local significance level for each comparison in the family, which leaves the analyst with no flexibility to control the allocation of the significance level. To deal with the computational inadequacy of the ‘mvtnorm’ package, we recommend the multivariate $t$ algorithms sketched in Sections 2.1.3 and 2.1.4 that accommodate any number of degrees of freedom and allow arbitrary patterns of variation in individual tail probabilities.
As a special case, when the hypotheses are equally weighted in the sense that the resulting consequences of any false rejections are comparable, \( \kappa_i = 1 \), the set of multiplicity-adjusted critical vector \( \xi \) is obtained as the 100\( \alpha^{th} \) equal coordinate quantile of the reference distribution, a multivariate \( t \) distribution with correlation structure \( R \) and degrees of freedom \( n - p \).

In summary, a unified description of a generic simultaneous inference procedure for any pre-selected set of linear comparisons, in the general linear regression framework with homogeneous variance assumptions has been presented. Many classical multiple comparison procedures, e.g., Tukey’s all pairwise comparisons and Dunnett’s treatments versus control comparisons can be embedded into this framework. We end the discussion on homogeneous variance by remarking that the popularity of the multiple testing procedure stems largely from its ease of calculation and full preservation of correlation between the test statistics.

The appropriateness of these general simultaneous inference procedures described in this section depends largely on the validity of the homogeneous variance assumption across all population groups. A direct application of the standard procedure developed for homogeneous data to heteroscedastic data can result in invalid statistical inferences, even when the variation in treatment variances show moderate departures from homogeneity. Simulation results reported in Petrinovich and Hardyck (1969) indicate that the validity of the test procedures can be compromised in the presence of heterogeneous variances. In addition, failure of the equal variance assumption can lead to either conservative or anti-conservative control of the error rate, with the direction and magnitude of the discrepancy between the actual error rate and the nominal level \( \alpha \) depends partly on the sample sizes and error variance configuration.
The simulation study reveals that the classical test procedures, e.g., Scheffe’s $F$ test, Tukey’s, and Dunnett’s tests tend to yield inflated (deflated) error rates when the sample sizes are inversely (directly) paired with the error variances. Moreover, recent empirical studies demonstrate the importance of adequately taking into account the heterogeneity of the error variances. Two notable examples are presented by Shahbaba and Johnson (2013) and Bar et al. (2014).

Nevertheless, the homogeneity assumption on the error variances is often difficult to verify and seldom satisfied. For instance, in clinical trial studies intended for the assessment of the efficacy of a new drug, the placebo (negative control) or the known effective drug (active control) often exhibits a higher variability than the investigational treatments, for a variety of possible reasons ranging from the intrinsic traits of the drug to the administration of the clinical trials. Such unequal variance situations abound in practice and are hardly avoidable in many scientific fields. In addition, it is often a subjective matter as to whether the equality of variance assumption is satisfactory for the analysis to be performed. A standard Levene test (Levene, 1960) can be used to identify variance heterogeneity, whereas other analysts may find it more reliable to base the diagnosis on some coarse visual examination of some residual plots.

In light of the foregoing discussion, the concerns about the invalidity of the test procedure specifically designed for homogeneous data suggest the need for an appropriate inference procedure that accompanies the models for heterogeneous data. As a matter of fact, the development of a suitable test procedure that retains the test size $\alpha$ remains a challenging task in the presence of heteroscedasticity. In essence, the chief technical problems in the construction of an exact test procedure are twofold,
the analytical derivation of a proper pivotal test statistic that is independent of the unknown model parameters and the probability calculation of the reference distribution, under the global null hypothesis. In contrast to the homogeneous variance case in which the reference distribution is known, the true null distribution of the multivariate test statistic under heterogeneous variances is not theoretically available. To address the multiple hypotheses problems in the presence of heteroscedasticity, a number of approximate test procedures have been suggested in the multiple comparison literature with each striving to gain enhanced ability to keep the actual test size closer to the nominal significance level $\alpha$.

One of the early approaches, contributed by Carroll and Ruppert (1981a), attempts to find an appropriate variance stabilizing transformation for the response variable, with the remaining analysis performed on the derived variable in the homogeneous variance framework. Some commonly used transformations are logarithm, square root and reciprocal transformations with the best choice dependent upon the variance structure specific to the data at hand. Albeit its wide applicability and ease of implementation, the main disadvantage of this approach is that the non-linear transformation makes it difficult to translate the exact inference on the means of transformed variable, carried out in the homogeneous variance framework, back to that of the original scale. In addition, Carroll and Ruppert (1981b) suggested modeling the variance structure as a smooth function of the mean with some extra structural parameters. The model fitting is handled through either iteratively re-weighted least squares or restricted maximum likelihood estimation, both of which make the associated statistical inferences large sample asymptotic.
3.3 Multiple Hypotheses Testing with Heterogeneous Variances

Although it is our goal to derive an appropriate multiple testing procedure that accounts for the heterogeneity in error variances, it is instructive to begin with a brief discussion of the univariate test for the significance of a single hypothesis that compares the difference between two population means with unequal variances. This test is one of the most widely applied traditional statistical techniques and it has been used in innumerable fields of study.

3.3.1 Single Hypothesis

Suppose that $X_1 = \{X_{1j}, j = 1, \ldots, n_1\}$ and $X_2 = \{X_{2j}, j = 1, \ldots, n_2\}$ are independent random samples of sizes $n_1$ and $n_2$ from two normally distributed populations with means $\mu_1, \mu_2$, and variances $\sigma^2_1, \sigma^2_2$, respectively. Assume that variances of the two populations are unequal, i.e., $\sigma^2_1 \neq \sigma^2_2$. The finite-sample test for the equality of population means is formulated as

$$H_0 : \mu_1 - \mu_2 = \mu \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 \neq \mu \quad (3.17)$$

It is common practice to employ the non-pooled two sample $t$ test to assess the equality of the two population means. The test statistics is defined as

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \mu}{\sqrt{S^2_1/n_1 + S^2_2/n_2}}. \quad (3.18)$$

Let $t$ denote the observed test statistics. The numerator and denominator in (3.18) are stochastically independent. However, the analytical expression of the p.d.f. of $T$ is not theoretically tractable, as the denominator is a mixture of two independent $\chi^2$ variates with distinct scale factors and degrees of freedom. When the distribution of
the test statistic under the null hypothesis is unknown, one resorts to an approximate
procedure, the appropriateness of which is measured by the discrepancy between the
actual test error rate and the nominal level of the test, $\alpha$.

[Welch (1938)] suggests approximating the null distribution of $T$ by a univariate $t$
distribution with degrees of freedom,

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1(n_1-1)} + \frac{s_2^4}{n_2(n_2-1)}}.$$  

The approximation comes from matching the first two moments of the linear combina-
tions of the sample variances in the denominator in (3.18). The resulting approximate
test procedure is often referred to as the two sample non-pooled $t$ test or Welch’s test.

Numerous simulation studies have validated the use of the non-pooled $t$ test for compar-
ing two population means in the presence of heterogeneity in the sense that it
keeps the test size reasonably close to the desired level $\alpha$. Moreover, we note that,
as the sample size increases, the approximate $t$ distribution more closely resembles
the true null distribution, both approaching the standard normal distribution in the
limit.

In addition to Welch’s $t$ test, some non-parametric procedures, e.g., the Mann-
Whitney or Wilcoxon rank-sum test, are frequently used to test the equality of various
features of the two distributions. It is generally true that, unlike the $t$ test, the rank-
based procedures require only minimal distribution assumptions. A switch to non-
parametric tests, however, is not necessarily an effective solution to the hypothesis
test problem in (3.17) with unequal variances. The non-parametric procedures that
assume the location shift of the distributions of the two populations are rendered
invalid by a clear violation of the equal variance assumption.
When the true null distribution of the test statistic is unknown, the resampling-based test procedure, parametric or non-parametric as presented in Dmitrienko et al. (2009), provides a generic approach to the construction of a reference distribution that approximates the true null distribution. The parametric resampling scheme makes use of the distribution assumption on the outcome variable, estimates parameters of the distribution, and repeatedly generates samples from the estimated distribution. One built-in benefit of the resampling approach lies in its intrinsic capacity to capture the distinctive features, e.g., unequal variances, in the data.

In the univariate testing problem stated in (3.17), the parametric resampling-based procedure generates random i.i.d. replicates of the test statistic by repeatedly carrying out the following two steps for a sufficiently large number, say \( N \) of times,

- **Step 1:** Generate a resampled data set according to

  \[
  X_{ij}^* = 0 + \varepsilon_{ij}^* \quad i = 1, 2, \quad j = 1, \ldots, n_i,
  \]

  where the \( \varepsilon_{ij}^* \) are independently distributed normal error variables with mean 0 and variance \( s_i^2 \).

- **Step 2:** Compute the test statistic

  \[
  t^* = \frac{\bar{x}_{i1}^* - \bar{x}_{i2}^* - \mu}{\sqrt{s_{i1}^2/n_1 + s_{i2}^2/n_2}}
  \]

  where \( \bar{x}_{i}^* \) is the sample mean and \( s_i^2 \) is the sample variance computed from the resampled data set for \( i = 1, 2 \).

The symbol \( * \) is used to differentiate the resampled data from the observed data. The \( p \)-value for the parametric resampling-based test is approximated by the proportion
of the \( N \) replicates for which the generated test statistic \( t^* \) exceed the observed test statistic \( t \) in absolute value. In other words,

\[
P\text{-value} = \frac{2}{N} \sum_{k=1}^{N} I(|t^*_k| > |t|).
\]

The heterogeneous variance of the data is automatically accounted for by the construction of the test statistic \( t^* \). In spite of the generality of the methodology, we remark that for small to moderate sample sizes, the approximate reference distribution may differ substantially from the true null distribution. For large sample sizes, limiting arguments ensure the validity of the approximation.

We conclude this subsection by making the observation that the performance of the resampling-based procedure remains unchanged by a further modification of Step 1, that independently generates the mean \( \bar{X}^*_i \) and variance \( S^*_i \), \( i = 1, 2 \) directly according to,

\[
\bar{X}^*_i \sim N(0, s_i^2/n_i)
\]

\[
S^*_i \sim s_i^2/n_i - 1 \chi^2_{n_i - 1}
\]

as opposed to resampling individual observations as in Step 1.

### 3.3.2 Multiple Hypotheses

This section aims to extend the two sample \( t \) test problem presented in Section 3.3.1 to the general multiple hypotheses test problem with more than two population groups, in the context of a simple one-way ANOVA model with heteroscedastic errors. Consider the one-way ANOVA model

\[
Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_i^2)
\]

\[
i = 1, \ldots, \nu, \quad j = 1, \ldots, n_i.
\]
The least squares estimators for the means $\mu_i$ and the error variances $\sigma_i^2$ are given by

$$\hat{\mu}_i = \bar{Y}_i,$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2,$$

where $s_i^2$ is the sample variance for treatment group $i$ and is independent of $\hat{\mu}_i$, for all $i$. As was sketched in the homogenous variance case, we assume that a finite collection of $m$ linear estimable functions, specified through a contrast coefficient matrix $L \in \mathbb{R}^{m \times \nu}$ is simultaneously assessed. Specifically, the multiple test problem is set up as

$$H_0 : L\mu = 0 \quad \text{vs.} \quad H_1 : L\mu \neq 0. \quad (3.20)$$

It follows from the normal distribution theory that

$$L\hat{\mu} \sim N(L\mu, L\Sigma\hat{\mu}L'),$$

provided that $\Sigma\hat{\mu}$ is the variance-covariance matrix of the estimated mean vector $\hat{\mu}$. That is, $\Sigma\hat{\mu} = \text{diag}\{\sigma_i^2/n_i, i = 1, \ldots, \nu\}$. In addition, $L\hat{\mu}$ is statistically independent of any linear combination of the sample variances $S_i^2$. In particular, for any component $k$,

$$l_k^T \hat{\mu} \perp l_k^T \hat{\Sigma}\hat{\mu}l_k,$$

where $l_k = \{l_{kj}, j = 0, \ldots, \nu - 1\}$ refers to the contrast coefficient for the $k^{th}$ hypothesis, i.e., the $k^{th}$ row of the $L$ matrix, and $\hat{\Sigma}\hat{\mu}$ denotes the estimated covariance matrix of $\hat{\mu}$. The commonly used test statistic associated with the multiple test problem (3.20) is expressed as

$$T = \left\{ T_k = \frac{l_k^T \hat{\mu}}{\sqrt{l_k^T \hat{\Sigma}\hat{\mu}l_k}}, \ k = 1, \ldots, m \right\}. \quad (3.21)$$
We note that the marginal component $T_k$ is the test statistic for the $k^{th}$ elementary hypothesis, i.e., $H_{0k} : \mathbf{l}_k^T \mathbf{\mu} = 0$. With an identical argument as in the two sample problem discussed in Section 3.3.1 the true null distribution of $T_k$, and thus $\mathbf{T}$, is not theoretically available, raising the need for an approximate procedure with adequate control of the test size $\alpha$.

Satterthwaite (1946) extends Welch’s two sample approximation results and suggests approximating the linear combination of a finite number of independent sample variances $S^2_i$, $i = 1, \ldots, \nu$, for example,

$$\sum_{j=1}^{\nu} \frac{l_{kj}^2 S^2_j / n_j}{\rho}, \quad (3.22)$$

by a $\chi^2$ variate with the Satterthwaite degrees of freedom given as

$$d_{f_k} = \frac{\left( \sum_{j=0}^{\nu-1} \frac{l_{kj}^2 s_j^2 / n_j}{l_{kj}^4 s_j^4 / n_j(n_j-1)} \right)^2}{\sum_{j=0}^{\nu-1} \frac{l_{kj}^4 s_j^4 / n_j(n_j-1)}} \quad (3.23)$$

by matching the first two moments. In other words,

$$d_{f_k} \cdot \frac{\sum_{j=1}^{\nu} \frac{l_{kj}^2 S^2_j / n_j}{l_{kj}^2 \sigma^2_j / n_j}}{\rho} \approx \chi^2_{d_{f_k}}. \quad (3.24)$$

And hence, the unknown distribution of each individual test statistic $T_k$, as defined in (3.21), is marginally approximated by a univariate $t$ variate, say $T^*_k$, with its own degrees of freedom,

$$T_k \overset{d}{\approx} T^*_k = \frac{\mathbf{l}_k^T \hat{\mathbf{\mu}}}{\sqrt{\mathbf{l}_k^T \Sigma \hat{\mathbf{\mu}} \mathbf{l}_k \sqrt{W^*_k/d_{f_k}}}} \sim t_{d_{f_k}}, \quad k = 1, \ldots, m \quad (3.25)$$

provided that $W^*_k$ is a $\chi^2$ variate with $d_{f_k}$ degrees of freedom. As a consequence, we regard the resulting random vector $\mathbf{T}^* = \{T^*_k, k = 1, \ldots, m\}$ as an approximation to the original test statistics $\mathbf{T}$, as in (3.21), in the marginal sense.
However, the approach of simultaneously assessing a collection of hypotheses differs radically from the notion of a single hypothesis test, as the joint distribution is specified by the marginals along with the dependence structure across the set of individual test statistics $T_k$, $k = 1, \ldots, m$. The use of $T^*$ as an approximation to $T$ is ambiguous, unless it is embedded with a clearly defined dependence structure, at least in the pairwise sense.

We recall that, in the homogeneous case, the dependence structure of the multivariate test statistic $T$ in (3.17) matches that of the associated multivariate Gaussian random vector, without regard to the common gamma scaling variate in the denominator. Consequently, this pivotality condition ensures the theoretical exactness of the simultaneous test procedure. In the heterogeneous case, the null distribution for the individual test statistic $T_k$ depends on the vector of population variances, $\sigma_i^2, i = 1, \ldots, \nu$. In addition, the dependence structure of the multivariate $T$ is due both to the correlation among the $\mathbf{l}_k^T \hat{\mu}$ in the numerator and to the dependence among the denominators. In the approximation, this dependence arises from the dependence of the $W_k^*$. The lack of homogeneity in error variances means that no exact solution exists for testing a suitable family of linear hypotheses on the treatment means when the population variances are not homogeneous, even for the simplest univariate test for the equality of two means. Approximations also become more challenging.

Moreover, we note that, as the sample sizes $n_i$ increases for all population groups, i.e.,

$$\min_{i=1,\ldots,\nu} n_i \to \infty,$$
the limiting distribution of the test statistic $T$ is a multivariate Gaussian distribution $Z$ with each marginal $T_i$ approaching a standard normal variate and correlation structure depending on the rates of growth in sample sizes $n_i$ across the $\nu$ populations. Specifically,

$$T_k \overset{d}{\to} Z_k \sim N(0, 1), \quad k = 1, \ldots, m$$

and

$$\text{Corr}(Z_k, Z_g) = \frac{\sum_{i=1}^{\nu} l_{ki} l_{gi} k_{1i} \sigma_i^2}{\sqrt{\sum_{i=1}^{\nu} l_{ki}^2 k_{1i} \sigma_i^2} \sqrt{\sum_{i=1}^{\nu} l_{gi}^2 k_{1i} \sigma_i^2}}, \quad \text{given that} \quad \lim_{n_i \to \infty} \frac{n_1}{n_i} = k_{1i}.$$

The multiple comparisons literature contrasts a number of approximate multiple comparison methods designed to handle heterogeneity in the small sample scenario. In our view, the different approaches vary from one another essentially with regard to the reference distributions employed to approximate the joint distribution of the test statistic, under the null hypothesis. The marginal approximation by a univariate $t$ with Satterthwaite degrees of freedom given in (3.23) is a common ground for all the procedures considered including the new method we propose shortly, with the exception of the resampling-based method. With practically identical marginal approximations, we place our focus primarily on the second order characteristics of the approximating test statistic $T^*$, measuring the pairwise correlations across the set of marginals $T_k$. An algebraic argument shows that the pairwise correlation in (3.25) can be factorized as

$$\text{Corr}(T^*_k, T^*_g) = \text{Corr}(X^*_k, X^*_g) \times \text{Corr} \left( \sqrt{\frac{df_k}{W^*_k}}, \sqrt{\frac{df_g}{W^*_g}} \right), \quad \forall (k, g), \quad (3.26)$$
provided that $X^*$ is the Gaussian random vector that constitutes the numerator of $T^*$, i.e.,

$$X^*_k = \frac{\mu_k^T \hat{\mu}}{\sqrt{\mu_k^T \Sigma \mu_k}}, \quad k = 1, \ldots, m,$$

(3.27)

and $W^*_k$ is the gamma variate associated with the $k^{th}$ individual hypothesis. This indicates a clear separation of the two sources of correlation — the Gaussian random vector $X^*$ and the gamma random vector $W^*$, that contribute and work together to determine that for $T^*$.

With the above argument, the construction of an approximate procedure is simply converted to the development of a reference distribution that adequately reflects the correlation information in $T$. This can be handled through separate consideration of the Gaussian component and the gamma component, as shown in (3.26). To set the stage for a logically coherent discussion in the remainder of this section, we make an organizational remark that the approximate procedures considered here, which endeavor to capture the dependence in the test statistics $T$, will be presented in a sequential manner as to how and the amount of correlation for each component is accounted for. With the supporting notes in place, we are in a proper position to provide systematic descriptions of various existing approaches.

1. **Complete Independence Procedure**

For the sake of completeness, we start with a rarely used procedure that assumes complete independence across components of the test statistics $T$. The improper independence assumption on the multivariate test statistic $T$ leads to a poor procedure, yielding an actual test size that deviates considerably from the nominal level $\alpha$. We complete the description of the procedure in the context of the set of $m$ hypotheses.
formulated in (3.20) by stating that each null hypothesis, say the $k^{th}$ is rejected at level $\alpha^{1/m}$, if

$$|t_k| \geq q_t \left( \frac{\alpha^{1/m}}{2}, df_k \right) \quad k = 1, \ldots, m, \quad (3.28)$$

where $q_t \left( \frac{\alpha^{1/m}}{2}, df_k \right)$ refers to the $100 \left( \frac{\alpha^{1/m}}{2} \right)^{th}$ quantile of a univariate $t$ distribution with $df_k$ degrees of freedom. This leads to the nominal probability of rejecting at least one null hypothesis of $1 - (1 - \alpha^{1/m})^m$. In addition, it clearly suggests that the amount of conservativeness can be somewhat alleviated by relaxing the strong independence assumption and considering, instead, a reference distribution that accounts for the correlation among components of $T$.

2. Partial Independence Procedure

In contrast to the complete independence procedure, the partial independence procedure successfully incorporates the Gaussian correlation structure of the numerator in (3.21) into the approximating reference distribution. The independence assumption of the gamma random vector in the denominator is left unchanged. It is easy to show that the Gaussian correlation $R_{X^*} = \{r_{kg}^X : 1 \leq k, g \leq m\}$ is given by

$$r_{kg}^X = \frac{l_k^T \Sigma_{\mu} l_g}{\sqrt{(l_k^T \Sigma_{\mu} l_k) (l_g^T \Sigma_{\mu} l_g)}} = \frac{\sum_{i=1}^\nu l_{ki} l_{gi} \sigma_i^2 / n_i}{\sqrt{\left( \sum_{i=1}^\nu l_{ki}^2 \sigma_i^2 / n_i \right) \left( \sum_{i=1}^\nu l_{gi}^2 \sigma_i^2 / n_i \right)}}. \quad (3.29)$$

Thus, the partial Independence procedure approximates the true null distribution of the test statistic $T$ by the reference distribution of the approximating test statistic $T^*$ built by the use of the Gaussian correlation $R_{X^*}$, provided that the sample variances $s_i^2$ are substituted for the unknown variances $\sigma_i^2$, and the independent
gamma random vector. Referring back to the class of generalized multivariate t distributions described in Section 2.2.1, the reference distribution suggested by the partial independence procedure is a generalized multivariate t distribution, denoted by $MVT_{m}(R_{X^*}, I_m, df_1, \ldots, df_m)$. Accordingly, this procedure rejects the $k^{th}$ individual null hypothesis if

$$|t_k| \geq Q^k_{T^*} (1 - \alpha) \quad k = 1, \ldots, m,$$

(3.30)

where $Q^k_{T^*} (1 - \alpha)$ refers to the $k^{th}$ entry of the $(1 - \alpha)^{th}$ quantile vector of the generalized multivariate t distribution with Gaussian correlation $R_X$, independence for the gamma vector and degrees of freedom $df_k$ for each univariate t marginal $T^*_k$.

The quantile vector is uniquely defined with the set of constraints in the form of,

$$P(|T_1| > Q^1_{T^*}) = \kappa_k P(|T_k| > Q^k_{T^*})$$

(3.31)

One expects that accounting for the Gaussian correlation in the reference distribution should increase the overall accuracy of test procedure compared to the complete independence procedure. This is borne out by simulation studies. The discrepancy between the actual family-wise type I error rate and the designated level $\alpha$ tends to shrink with reference to the former procedure, due to the consideration of the Gaussian correlation.

The inappropriate use of the independence assumption on the associated gamma random vector in the partial independence procedure inevitably compromises the validity of statistical inferences. Inference under this procedure tends to be conservative with an actual family-wise error rate lower than the nominal rate. The conservativeness can become increasingly alarming, as the discrepancy between the gamma correlation structure and the assumed independence structure gets larger. It is helpful
to envision changes in the actual error rate as the correlation of the set of gamma variates moves towards the other extreme, that is when the gamma variates exhibit the strongest possible correlation. For strongly correlated variates, the use of a procedure that assumes complete dependence across the gamma vector could be a viable alternative that provides more effective control of the overall error rate. The relevant procedure that matches this alternative strategy was developed by Hasler and Hothorn (2008), modifying the procedure proposed by Games and Howell (1976).

We comment that the Games & Howell procedure fits into the framework mainly as a stepping stone for the Plug-In (PI) method, for which we rigorously analyze the theoretical as well as the numerical performance. In addition, the PI procedure and the partial independence procedure will be used as benchmarks to illustrate the improvement of the new procedure we propose over the standard choices.

3. The Games & Howell Procedure

The Games & Howell (GH) procedure (Games and Howell, 1976) is recognized to be the first multiple testing procedure specifically designed for heteroscedastic data. Analogous to many other approximate procedures, the GH procedure makes use of the Satterthwaite approximation for each marginal test statistic $T_k^*$. Nevertheless, unlike the partial independence structure and other procedures we will present shortly, the Gaussian correlation structure used in the GH procedure is given by $R_{X^*}^{GH} = \{r_{X^*}^{GH}(k,g) : 1 \leq k, g \leq m\}$, where

$$r_{X^*}^{GH}(k,g) = \frac{\sum_{i=1}^\nu l_{ki}l_{gi}/n_i}{\sqrt{\left(\sum_{k=1}^\nu l_{ki}^2/n_i\right)\left(\sum_{g=1}^\nu l_{gi}^2/n_i\right)}}.$$
It is easy to note that the approximate Gaussian correlation structure $R_{X^\star}^{GH}$ matches the Gaussian correlation structure that would have been derived, were the homogeneous variance across the treatment groups an appropriate assumption. On the other hand, the gamma random vector is, as in the homogeneous variance case, assumed to have the maximum possible correlation in a conceptual, yet not technically precise sense, due to the major distinguishing feature of the variation of the degrees of freedom parameters across the $m$ univariate marginals. The moment-based correlation between two gamma variables with different degrees of freedom is bounded at a certain level. We use the term ‘maximum possible correlation’ or ‘nearly perfect correlation’ in place of the ‘perfect correlation,’ due to the constraint on the achievable maximum correlation.

The implementation of the GH procedure requires numerous evaluations of the multivariate $t$ distributions with common correlation matrix $R_{X^\star}^{GH}$ and distinct degrees of freedom. Specifically, the GH procedure calculates the comparison-specific critical value $\xi_k$ for each elementary hypothesis $H_{0k}$ separately, as the $(1 - \alpha)^{th}$ equal coordinate quantile of a classical $m$-variate $t$ distribution with the common correlation matrix $R_{X^\star}^{GH}$ and its own Satterthwaite degree of freedom $d_{f_k}$. Using probability notation, $\xi_k$ satisfies,

$$P(|T_{1k}| \leq \xi_k, |T_{2k}| \leq \xi_k, \ldots, |T_{mk}| \leq \xi_k) = 1 - \alpha,$$

where $T^k = \{T_{ik}, i = 1, \ldots, m\}$ is a multivariate $t$ random vector with correlation $R_{X^\star}^{GH}$ and degrees of freedom $d_{f_k}$. Accordingly, the simultaneous test decisions are made by comparing each test statistic value $t_k$ to its own comparison-specific critical
value $\xi_k$, that is, the $k^{th}$ individual null hypothesis is rejected if

$$|t_k| \geq \xi_k, \ k = 1, \ldots, m. \quad (3.33)$$

We note that the heterogeneous nature of the data is reflected in the GH procedure in a subtle manner. Seemingly, it differs from the simultaneous inference procedure designed for homogenous data in the sense that the Satterthwaite approximation needs to be performed on each univariate marginal $T_k$ as the first step. Apart from the heterogeneity conveyed in the marginal approximations, the two correlation factors are captured in essentially the same way in the GH procedure as in the standard procedure for homogeneous data. That is, the Gaussian correlation is accounted for in an identical manner and the maximum pairwise correlation is assumed for the gamma component $W^*$ in the GH procedure and the constant gamma vector $W$ in the standard procedure. To accommodate the nonconformity of the degrees of freedom parameters across the $m$ coordinates of $T$, the GH procedure requires repeated quantile calculations of a sequence of classical multivariate $t$ distributions, parameterized by the correlation matrix $R_{X^*}^{GH}$ for the Gaussian component under homogeneity and comparison-specific degrees of freedom.

The use of the Gaussian correlation structure under the homogeneity assumption can be unsuitable when population variances are markedly different. This suggests a possibility of improving the test accuracy by considering an alternative Gaussian correlation that involves the individual variance estimators $S_k^2$, which matches the idea of the Plug-In method proposed by Hasler and Hothorn (2008), and described by Hasler (2016) to be, thus far, the most promising test procedure among all available approximate methods for heterogeneous data.
4. The Plug-in Procedure

The original intention of the plug-in (PI) procedure proposed by Hasler and Hothorn (2008) was to address the concern that the GH procedure makes inadequate use of the heterogeneity in the data, particularly in the construction of the Gaussian correlation of the approximating test statistic. The PI procedure replaces the Gaussian correlation matrix $R_{X}^{GH}$ by the correlation matrix $R_{X}$, given in (3.29) that involves the individual variance estimates $S_{k}^{2}$. Accordingly, the implementation of the PI procedure resembles that of the GH procedure in terms of the component-wise critical value calculation. The PI method approximates comparison-specific critical values $\xi_{k}$ for the individual comparisons $H_{0k}$ by repeatedly evaluating the $(1 - \alpha)^{th}$ equal coordinate quantile of a sequence of multivariate $t$ distributions with a common correlation structure $R_{X}$ and distinct Satterthwaite degrees of freedom $d_{f_{k}}$, $k = 1, \ldots, m$ for each coordinate. In other words, the $k^{th}$ individual hypothesis is rejected if,

$$|t_{k}| \geq \xi_{k} \quad k = 1, \ldots, m, \quad (3.34)$$

where $\xi_{k}$ satisfies,

$$P(|T_{1}^{k}| \leq \xi_{k}, |T_{2}^{k}| \leq \xi_{k}, \ldots, |T_{m}^{k}| \leq \xi_{k}) = 1 - \alpha, \quad (3.35)$$

where $T^{k} = \{T_{i}^{k}, i = 1, \ldots, m\}$ is a multivariate $t$ random vector with correlation $R_{X}$ and degrees of freedom $d_{f_{k}}$.

Additionally, the adjusted $p$-value for the $i^{th}$ hypothesis based on the PI procedure for is defined as,

$$p_{i} = 1 - P_{i}(\max_{k = 1, \ldots, m} |T_{k}| < |t_{i}|), \quad (3.36)$$

where $P_{i}$ refers to the multivariate $t$ probability with some correlation structure and unique Satterthwaite degrees of freedom $d_{f_{i}}$. 

116
Comparisons between the Partial Independence and the PI Procedures

Before moving forward to the new procedure, we take a moment to make an explicit connection between the partial independence procedure and the PI procedure. The Gaussian components in both procedures are equipped with identical correlation structure \( R_{X^*} \), defined in (3.29), that depends on the individual variances for the \( \nu \) population groups. The two procedures differ significantly from one another in terms of how the correlation structure of the gamma component \( W^* \) appears in the scaling of the random vector in \( T^* \). Specifically, the partial independence procedure assumes complete independence across the individual gamma variates, whereas the PI procedure assumes the maximum possible positive dependence. A special technical adjustment is used to enforce the core principle of the PI procedure with distinct component-wise degrees of freedom.

The statistical validity of both procedures depends heavily on the true correlation structure of \( W \), the collection of the linear combinations of the independent sample variances \( S_k^2, \ k = 1, 2, \ldots, \nu \), in the test statistic \( T \). The PI procedure is anticipated to be a suitable procedure, when all pairwise correlations of \( W \) are strong. However, as the correlation structure moves toward independence, the partial independence procedure becomes better than the PI method. When the correlation structure falls in the middle ground, both options may perform poorly. Furthermore, the pattern of correlations may be uneven, with some being very strong and others weak. In this case, neither procedure would seem to be appropriate.

The structure of the set of comparisons dictates the success of the approximations. Take, for example the family of all treatments versus control comparisons. The appearance of the sample variance for the control group as a common term in all linear
combinations suggests that the partial independence procedure will not work well, especially when the control group exhibits much larger variability than the remaining groups. On the other hand, the family of all pairwise comparisons among $\nu$ distinct treatment groups provides a different correlation structure, especially when the number of groups $\nu$ is large. For instance, among the collection of $K = \binom{\nu}{2}$ pairwise comparisons among $\nu$ distinct population groups, the number of pairs of individual test statistics $(T_k, T_g)$ that have independent denominators, i.e., linear combinations of disjoint sets of independent sample variances can be analytically calculated to be $\binom{K}{2} - \nu \binom{\nu - 1}{2}$, out of $\binom{K}{2}$ pairs in total. The magnitude of the correlations among the remaining pairs depends on both group variances and sample sizes. Taking a limit as $\nu \to \infty$, the fraction of pairs that are independent tends to 1. In this case, the PI procedure is a reckless choice.

Examination of various scenarios has shown that the PI method is generally either anti-conservative or approximate, yielding an actual overall type I error rate that may be biased upward. The discrepancy between the actual error rate and the pre-specified level is impacted by the deviation of the true correlation structure of $W$ from the maximum correlation assumed by the PI procedure. It is worthwhile to note that the discrepancy could also be attributed to many other aspects, e.g., the validity of the univariate $t$ representation of each individual test statistic by Satterthwaite approximation and the suitability of the associated Gaussian correlation. Moreover, the accuracy of the sample variances as estimates for the unknown group variances serves as the driving force behind the appropriateness of the test procedure.

In contrast to the PI procedure, the partial independence procedure is either conservative or approximate for a typical instance. Its adequacy is impacted by the
closeness of the correlation of $W$ to the independence condition. As before, when one has no knowledge of the true correlation structure, very little is known about how the overall significance level $\alpha$ is preserved with any choice of the available procedures.

**A Few Shortcomings of the Plug-in Procedure**

The plug-in procedure by Hasler and Hothorn (2008) is perhaps the most commonly used approach to multiple inferences in the presence of heteroscedasticity. The popularity of the PI method stems largely from its ability to capture the correlation of the numerators of the test statistics. Nevertheless, an inherent problem with the PI method is its ignorance of the dependence structure among the collection of linear combinations of the independence sample variances as the denominator vector $W$. Thus, the accuracy and statistical power of the PI procedure in conducting joint inferences is contingent upon the proximity of the true correlation structure for $W$ and the maximum possible dependence. Exact analytic and numerical evidence shows that the statistical validity of the PI method can be compromised even when a moderate departure between the true and assumed ‘nearly perfect’ correlation structures arises. Likewise, a parallel argument justifies the advantages and tradeoffs of the alternative partial independence procedure. We omit the reasoning for brevity.

The PI method, despite its use of the ‘nearly complete’ pairwise correlation among the set of gamma variables of the test statistics as in the standard procedure for homogeneous data, is computationally much more involved than the standard procedure, due to its distinctive feature of varying numbers of degrees of freedom across the marginal distributions. Unlike the standard procedure that is often paired with a classical multivariate $t$ distribution as the reference distribution, the implementation of the PI procedure for a collection of $m$ linear hypotheses requires repeated quantile
calculations of a sequence of classical multivariate $t$ distributions with a certain common correlation structure and coordinate-specific Satterthwaite degrees of freedom. The standard calculation is repeated for each linear hypothesis with its unique degrees of freedom. Considering the schizophrenic treatment of the different hypotheses, the PI method is not fully tied to any single reference distribution, as each comparison-specific critical value is calculated as the $(1 - \alpha)^{th}$ quantile of a different multivariate $t$ distribution.

As was shown in the simulation study in Section 2.2.1, numerical calculation of multivariate $t$ probabilities becomes less accurate and computationally more intensive, when the number of comparisons $m$ grows large. Moreover, the fact that multivariate $t$ distributions with different degrees of freedoms are evaluated repeatedly for assessing a set of $m$ simultaneous hypotheses slows the numerical implementation of the PI method for large $m$. The frequently tested family of all pairwise comparisons provides a simple example of the computational cost of this procedure for a large number of treatment groups, e.g., $\nu$ treatments lead to $m = \binom{\nu}{2}$ comparisons. In practice, even a moderately large number of treatment levels $\nu$ may pose a problem for the PI procedure.

The computer implementation of the PI procedure is available through the ‘SimComp’ package published by Hasler (2014) in R, which relies on the ‘qmvt’ function in the ‘mvtnorm’ package by Genz and Bretz (2002) for the quantile calculation of multivariate $t$ distribution. However, the ‘qmvt’ function for equal coordinate quantile calculation is restricted to only multivariate $t$ distributions with an integer degrees of freedom, and so the Satterthwaite degrees of freedom for each coordinate is dropped down to the next integer in the implemented PI procedure.
There is one more potential shortcoming of the PI procedure. The PI procedure as described in Section 3.3.2 presumes that each individual comparison has the same significance level. In practice, the various hypotheses may not be of equal interest, and the analyst may choose to ‘spend’ the family-wise error rate $\alpha$ in an asymmetric fashion. The spend is often driven by the relative importance of the hypotheses. The PI procedure can be extended to handle differing individual error rates. This practical limitation is handled automatically through both of the new formulations of the test procedure, namely (to be referred to as) the NORTA procedure and the rank procedure. Similar adjustments can be used with the PI procedure.

**Our Proposals to Overcome the Shortcomings**

5. The Rank-based PI Procedure

As a general principle, it is attractive for a test procedure to be computationally efficient. With the aim of reducing the computational complexity of the PI procedure especially for large $m$, we suggest a computationally more favorable procedure that follows closely the core philosophical principles of the PI procedure. Maximal positive dependence of the denominators is captured in the PI procedure by assuming a single degrees of freedom and proceeding with classical multivariate $t$ calculations. In the rank-based plug-in (RPI) procedure, we capture this maximal positive dependence by coupling the vector of gamma variates with a single uniform variate. This avoids the need to recompute for each comparison with its unique degrees of freedom.

The dependence of the gamma random vector is measured by the rank correlation matrix defined by,

$$
\tau^{W*}_{kg} = \frac{Cov(F_k(W_k^*), F_g(W_g^*))}{\sqrt{Var(F_k(W_k^*))Var(F_g(W_g^*))}}.
$$

(3.37)
The random sampling of the perfectly rank correlated gamma random vector with varying marginal parameters is accomplished by the use of the probability integral transformation or distribution function inversion method. More specifically, a random sample of size $N$ can be generated by repeatedly implementing the following simple procedure $N$ times.

- Generate $U_i \sim \text{Uniform}(0, 1)$;
- Compute $W_i = (W_{i1}, W_{i2}, \ldots, W_{im})$, where $W_{ik} = F_k^{-1}(U_i)$ with $F_k^{-1}$ being the inverse gamma c.d.f. with shape parameter $df_k/2$ and scale parameter 2. Equivalently, $W_{ik}$ can be obtained as the $(100U_i)^{th}$ quantile of the distribution $F_k$.

The rank procedure accounts for the major features of the test statistics $T$, namely the marginals, the Gaussian correlation and the gamma correlation in a manner similar to the PI procedure. Its performance should be similar to that of the PI procedure.

The numerical implementation of the rank procedure is essentially the same as the quantile calculation of the generalized multivariate $t$ distribution, with the slight difference arising from the construction of the $m \times N$ table of the test statistic. The use of the distribution function inversion method in the RPI procedure replaces other joint distributions on the gamma vectors, as in the NORTA method. With the tables set up, the same search algorithm is used to calculate the quantile vector, subject to a set of constraints in the form of (3.31). Therefore, the computational effort required for the implementation of the RPI procedure can be reduced to a manageable level, regardless of the number of comparisons.
6. Linear Interpolation of the Degrees of Freedom

The use of the integer portion of the Satterthwaite degrees of freedom for each component $T_k$ does not align well with Hasler and Hothorn’s original description of the PI method. The impact of the computational constraint passes to the ability of the PI procedure to retain the overall error rate for the given set of multiple hypotheses. To ensure that the computer implementation is in closer agreement with the methodological sketch, we suggest a local linear interpolation on the degrees of freedom to find comparison-specific multiplicity-adjusted critical values. We anticipate that these adjusted critical values will be more compatible with the PI method as described by Hasler and Hothorn (2008).

Suppose that the multiplicity-adjusted critical values $\xi_k$ for the $k^{th}$ elementary hypothesis are denoted by $\xi_{k}^{\text{down}}$ and $\xi_{k}^{\text{up}}$ respectively, when the Satterthwaite degrees of freedom $d_{f_k}$ is rounded down $[d_{f_k}]$ and rounded up to the integer $[d_{f_k}] + 1$. The linearly interpolated $\xi_k$ is

$$\xi_k = \xi_{k}^{\text{down}} + (\xi_{k}^{\text{up}} - \xi_{k}^{\text{down}})(d_{f_k} - [d_{f_k}]), \quad k = 1, \ldots, m.$$ 

The monotonically decreasing relationship between the critical value and the tail probability outside of the hyper-rectangle formed by the critical values ensures that the linear interpolation on the multiplicity-adjusted critical values for all the $m$ comparisons will produce an overall type I error rate between those for the rounded down and rounded up degrees of freedom. We again note that the PI procedure is generally liberal with an inflated overall type I error rate, whereas the size of the anti-conservativeness (excess error rate) is strictly larger when the rounded down degrees of freedom is used compared to the rounded up version. The effect of the rounding
strategy is the most profound when the Satterthwaite degrees of freedom is small, say less than 10. The linear interpolation serves to recalibrate the PI procedure, generally improving its performance.

As will be shown in the simulation study, for many families of comparisons, the PI procedure implemented with the linear interpolation adjustment is essentially indistinguishable from the rank-based procedure, as the main features are captured in virtually the same manner.

7. The NORTA Procedure

The methodological deficiency of the PI procedure suggests the need for the development of a flexible simultaneous inference procedure that more effectively exploits the correlation information across the collection of linear combinations of the independent sample variances in the test statistics. With this aim, this section is intended for the development of a variation of the PI method, that allows one to associates the gamma random vector with a correlation structure that is more reflective of the true correlation pattern. This variation is to be referred to as the NORTA procedure, due to its use of the NORTA algorithm in the numerical implementation.

For the sake of convenience, we reiterate the test statistic $T$, established in (3.20) for the multiple hypotheses formulated in (3.21). That is,

$$
T = \left\{ T_k = \frac{l_k^T \hat{\mu}}{\sqrt{l_k^T \hat{\Sigma} \hat{\mu} l_k}}, \ k = 1, \ldots, m \right\}.
$$

(3.38)

As before, let $W$ denote the collection of linear combinations of the independent sample variances with components $W_k = l_k^T \hat{\Sigma} \hat{\mu} l_k$. According to the Satterthwaite approximation \cite{Satterthwaite1946}, each individual component $W_k$ is approximated by a scaled $\chi^2$ variate with degrees of freedom parameter $d f_k$ calculated by matching
the first two moments. However, much less attention has been placed on the correlation structure of $W$. The PI procedures and the partial independence procedures represent the extremes of the maximum possible dependence among all components of $W$ and independence among all components. Typical sets of comparisons fall between these extremes. In this work, we investigate this middle ground.

The pairwise product-moment covariance between any two individual linear combinations of the sample variances, say $W_k$ and $W_g$ is,

$$\text{Cov} \left( l_k^T \hat{\Sigma}_\mu l_k, l_g^T \hat{\Sigma}_\mu l_g \right) = \sum_{i=1}^{\nu} \frac{2l_k^2 l_g^2 \sigma_i^4}{n_i^2(n_i - 1)}. $$

That is to say, the gamma random vector $W^\star = (W_1^\star, W_2^\star, \ldots, W_m^\star)$ has joint correlation structure $\Sigma_{W^\star} = (r_{W^\star}^W)_{1 \leq k < g \leq m}$, given that

$$ r_{W^\star}^{W^\star} = \frac{\sum_{i=1}^{\nu} \frac{l_k^2 l_g^2 \sigma_i^4}{n_i^2(n_i - 1)}}{\sqrt{\left( \sum_{i=1}^{\nu} \frac{l_k^4 \sigma_i^4}{n_i^2(n_i - 1)} \right) \left( \sum_{i=1}^{\nu} \frac{l_g^4 \sigma_i^4}{n_i^2(n_i - 1)} \right)}}. \quad (3.39) $$

In practice, $r_{W^\star}^{W^\star}$ is estimated by plugging in the sample variances $s_i^2$ for the unknown group variances $\sigma_i^2$, $i = 1, \ldots, \nu$. The expression in (3.39) clearly indicates that the moment-based correlation $\hat{\Sigma}_W$ is a function of the sample variances, sample sizes, and the contrast coefficient matrix $L$. It is desirable to allow the data to automatically inform the inference procedure about the correlation structure among the gamma variates. This can be accomplished by a modification of the PI procedure – we call the NORTA procedure.

The NORTA procedure approximates the true null joint distribution of the test statistics by a generalized multivariate $t$ distribution with the Gaussian and gamma correlation parameters specified by $R_X$, (3.29) and $R_W$, (3.39) respectively, and the degrees of freedom $df = (df_1, \ldots, df_m)$. The multiplicity-adjusted critical vector

125
\( \xi \in \mathcal{R}^m \) of the NORTA procedure is calculated as the \((1 - \alpha)^{th}\) quantile vector of the associated reference distribution \( T^* \), satisfying

\[
P_{T^*}(\|T_k^*\| \leq \xi_k, k = 1, \ldots, m) = 1 - \alpha, \tag{3.40}
\]

subject to a set of constraints pertaining to the weighted importance of the \( m \) hypotheses,

\[
P(\|T_1^*\| > \xi_1) = \kappa_k P(\|T_k^*\| > \xi_k), \tag{3.41}
\]

for known, pre-specified constants \( \kappa_k, k = 2, \ldots, m \).

For hypothesis \( i \), let \( t_k^{(i)} \) be the value, for component \( k \), that gives tail probability for the \( k^{th} \) marginal distribution that relates to the tail probability beyond \( |t_i| \) for the \( i^{th} \) hypothesis according to,

\[
\kappa_i P(\|T_i\| > t_i) = \kappa_k P(\|T_k\| > t_k^{(i)}), \quad k = 1, 2, \ldots, m.
\]

The adjusted \( p \)-value for the \( i^{th} \) hypothesis based on the NORTA procedure for is defined as

\[
p_i = 1 - P_0(\|T_k\| < t_k^{(i)}, k = 1, 2, \ldots, m), \tag{3.42}
\]

where \( P_0 \) refers to the generalized multivariate \( t \) probability with Gaussian correlation structure \( R_X^* \) and gamma correlation structure \( R_W^* \) and Satterthwaite degrees of freedom \( df \).

The theoretical formulation of the NORTA procedure that makes full use of the resources obtainable from the data has been left unconsidered until this work. This gap in the literature can be largely attributed to the technical difficulties in computing the multivariate \( t \) probabilities with varying marginal parameters. The development of the generalized multivariate \( t \) distributions along with the numerical approximation
algorithms of the probability calculation and closely related quantile function calculation in Section 2.2, makes it possible to consider adding any arbitrary correlation structure for the gamma random vector. The NORTA procedure is another procedure, in addition to the partial independence procedure, that benefits the formulation of the class of generalized multivariate $t$ distributions as described in Section 2.2. The numerical implementation of the NORTA procedure is described in Section 2.2. The name of the NORTA procedure originates from the use of the NORTA algorithm in generating i.i.d. replicates from the gamma random vector.

The resulting procedure interpolates naturally between the two extreme cases – ‘nearly perfect’ dependence imposed by the PI method and the complete independence imposed by the partial independence procedure by allowing the data to govern the adjustment for the complex gamma correlation. As the NORTA procedure has the capacity to capture the intrinsic correlation structure of the test statistics, it is anticipated to markedly improve the performance of the standard choices with a better approximation to the nominal significance level $\alpha$.

In addition to the conceptual advantage of the NORTA approach, the NORTA procedure also has a number of subsidiary benefits over the PI procedure. The computing time in the calculation of the comparison-specific critical values is approximately $(1/m)^{th}$ of that required by its PI competitor, regardless of the size $m$ of the family of the linear hypotheses. Also, the NORTA procedure, as with the PI procedure, can be customized to allow variation in individual significance levels.

**The Resampling-based Procedure**

The multiple test procedures described so far all approximate each marginal test statistic by a univariate $t$ variable with degrees of freedom given by the Satterthwaite
approximation method. However, the $\chi^2$ approximation in Satterthwaite’s method can be poor, especially when the true variances vary substantially. To examine the impact of this approximation, we explore an alternative resampling-based approach to construct the reference distribution.

The resampling-based test approach for the two-sample problem can be extended straightforwardly to more than two treatment groups. This method is very easy to implement.

In the context of the one-way ANOVA model with unequal variances, the parametric resampling scheme makes use of the distribution assumption on the outcome variable and repeatedly generates samples from a model with parameters estimated from the data. As in the single testing problem stated in (3.17), the parametric resampling-based procedure starts out with the creation of the $m \times N$ test statistic table, by implementing the following two steps $N$ times,

- **Step 1:** Generate a resampled data set according to

  \[ X_{ij}^* = 0 + \varepsilon_{ij}^* \quad i = 1, 1, \ldots, \nu, \quad j = 1, \ldots, n_i, \]

  where the error terms $\varepsilon_{ij}^*$ are independently distributed normal error variables with mean 0 and variance $s_i^2$.

- **Step 2:** Compute the resampled multivariate test statistics, $t^* = \{t_k^*, k = 1, \ldots, m\}$, with

  \[ t_k^* = \frac{l_k^T \hat{\mu}}{\sqrt{l_k^T \hat{\Sigma} \hat{\mu} l_k}}, \]

  given that $l_k$ is the contrast coefficient vector specifying the $k^{th}$ individual comparison, $\hat{\mu}$ denotes the sample mean vector based on the resampled dataset from
Step 1; and \( \hat{\Sigma}_\mu \) refers to the estimated variance-covariance matrix of the mean estimator \( \hat{\mu} \) with the sample variances plugged in.

Let \( T \) be the \( m \times N \) test statistics table consisting of \( N \) columns, each of which represents an independent sample from the plug-in test statistic \( T^\ast \). As with the other procedures, the resampling-based procedure is completed with the calculation of the multiplicity-adjusted critical vector \( \xi \), which is performed with the search algorithm described for the quantile evaluation of the generalized multivariate \( t \) distribution.

We conclude this subsection with the observation that resampling individual observations according to Step 1 is equivalent to generating the mean \( \bar{X}_i^\ast \) and variance \( S_i^\ast^2 \) independently according to

\[
\hat{\mu}_i \sim N(0, s_i^2/n_i)
\]
\[
S_i^\ast^2 \sim \frac{s_i^2}{n_i - 1} \chi_{n_i-1}^2,
\]

which will be carried into the Step 2 for the construction of the test statistics table.

### 3.4 Simulation Study

In this section, we conduct a simulation study to demonstrate the behavior of the various multiple comparison procedures that deal with heteroscedasticity investigated in this work. The comparison among the procedures focuses on their abilities to preserve the pre-specified family-wise error rate \( \alpha = 0.05 \). The two well-known families of comparisons – all pairwise and all treatments vs. control are considered with different numbers of treatment groups. Two sided comparisons are performed under four simulation settings. They are,

1. All pairwise comparisons with \( \nu = 8 \) treatment groups;
2. All pairwise comparisons with $\nu = 7$ treatment groups;

3. All treatments vs. control with $\nu = 8$ treatment groups;

4. All treatments vs. control with $\nu = 10$ treatment groups;

For each simulation setting, various scenarios with different group variances and sample sizes are arbitrarily generated, satisfying the condition that the maximum variance ratio among the treatment groups is controlled under 6. Specially, for each scenario under each simulation setting, the variance for the first group $\sigma_1^2$ is randomly drawn from the uniform$(5, 20)$ distribution. The variances for the remaining groups are created by randomly sampling their ratios to that of the first group, according to uniform$(0.25, 4)$. The qualification of the set of variances for simulation analysis is dictated by whether the maximum pairwise ratio is below 6 or not. The constraint on variance ratios arises from the concern that treatment means alone may no longer be of interest, in circumstances where the variances are enormously different. In addition, the sample sizes are randomly generated, with equal probability on all integers ranging from 4 to 20. Throughout the study, the treatment means are assumed to be the same $\mu = 0$.

<table>
<thead>
<tr>
<th>The parameter generating process:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output</strong>: A suitable variance vector $\sigma^2 = {\sigma_i^2, i = 1, \ldots, \nu}$ and sample sizes.</td>
</tr>
<tr>
<td>• Generate $n_i \sim \text{Discrete Uniform}(4, 5, \ldots, 20), i = 1, \ldots, \nu$;</td>
</tr>
<tr>
<td>• Generate $\sigma_1^2 \sim \text{Uniform (5, 20)}$;</td>
</tr>
<tr>
<td><strong>Repeated until</strong>: $\max_i \sigma_i^2 / \min_i \sigma_i^2 \leq 6$</td>
</tr>
<tr>
<td>• Generate $\lambda_i \sim \text{Uniform}(0.25, 4)$;</td>
</tr>
<tr>
<td>• Calculate the remaining group variances as $\sigma_i^2 = \sigma_1^2 \cdot \lambda_i, \ i = 2, \ldots, \nu$;</td>
</tr>
</tbody>
</table>

The number of scenarios considered under each simulation setting is 8, 13, 16, and 130.
Table 3.1: The Variances Specification for the 8 Treatment Groups for each Scenario Considered in Simulation Setting (1)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>9.28</td>
</tr>
<tr>
<td>2</td>
<td>11.03</td>
</tr>
<tr>
<td>3</td>
<td>10.23</td>
</tr>
<tr>
<td>4</td>
<td>6.08</td>
</tr>
<tr>
<td>5</td>
<td>10.09</td>
</tr>
<tr>
<td>6</td>
<td>16.54</td>
</tr>
<tr>
<td>7</td>
<td>14.11</td>
</tr>
<tr>
<td>8</td>
<td>15.36</td>
</tr>
</tbody>
</table>

22, respectively. The following discussion focuses on the first simulation setting with 8 scenarios. The remaining ones will be included in the appendix with detailed discussion omitted for brevity.

The variance and sample size parameters for the eight scenarios considered in simulation setting (1) are given in Tables 3.1 and 3.2, respectively. For each scenario, the estimated family-wise error rate for the pairwise differences on treatment means, given in Figure 3.1, is obtained based on 500 simulation runs, for the various approximate procedures including: (1) The PI procedure; (2) the PI procedure with linear interpolated degrees of freedom; (3) the RPI method; (4) The NORTA method; and (5) the resampling method. In addition, the estimated and their associated standard errors are given in Tables 3.3 and 3.4, respectively.

Figure 3.1 shows that the RPI (blue curve) procedure that makes the least effort to account for the correlations among the \( \binom{8}{2} = 28 \) denominators of the test statistic \( T \) yields the most inflated family-wise error rate. In addition, the use of the floored
Table 3.2: The Sample Sizes for the 8 Treatment Groups under each Scenario Considered in Simulation Setting (1)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>16</td>
<td>12</td>
<td>17</td>
<td>7</td>
<td>14</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>8</td>
<td>17</td>
<td>13</td>
<td>20</td>
<td>15</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>18</td>
<td>17</td>
<td>7</td>
<td>5</td>
<td>16</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>16</td>
<td>13</td>
<td>4</td>
<td>17</td>
<td>5</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>17</td>
<td>9</td>
<td>7</td>
<td>14</td>
<td>12</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>7</td>
<td>19</td>
<td>17</td>
<td>4</td>
<td>18</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>19</td>
<td>20</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>17</td>
<td>20</td>
<td>11</td>
<td>14</td>
</tr>
</tbody>
</table>

degrees of freedom in the PI method, as suggested in Hasler and Hothorn (2008), lessens the gap between the estimated FWER based on the RPI procedure and the target level $\alpha$, as indicated by the light blue curve. In fact, the flooring of the degrees of freedom can be thought of as an adjustment that counteracts the anti-conservativeness caused by the failure to account for the positive gamma correlations. However, the haphazard adjustment may or may not work well for a particular data, depending on the size of the degrees of freedom as well as that of the fractional part.

Opposite to the PI method that assumes complete dependence among the denominators of the test statistic $T$ is the partial independence procedure that assumes complete independence among the denominators. The yellow curve in Figure 3.1 shows that the partial independence procedure is a conservative test procedure for a typical situation. Its performance is impacted by the closeness of the gamma correlation to independence.
In this simulation setting, the family of all pairwise comparisons among 8 treatment groups consists of 28 pairwise differences. Furthermore, the 28-dimensional test statistic $T$ has 210 pairs of independent marginals out of the 378 possible pairs, to name a few, 1 vs. 2 and 3 vs. 4, 1 vs. 2 and 3 vs. 5, and 1 vs. 2 and 7 vs. 8, etc. Independent pairs are associated with hypotheses on disjoint sets of treatment groups. The fact that 56% of the pairs are independent clearly indicates that neither complete dependence nor complete independence among the denominators is a reasonable choice.

With the PI and partial independence procedures being the two extreme cases in modeling the gamma dependence, the NORTA procedure proposed in Section 3.3.2 falls in between by exploiting the positive dependence among the gamma variates in the denominators with the sample variances substituting for the unknown true variances. The green curve shows that the NORTA procedure reduces the amount of conservativeness induced by the PI method to a noticeable degree and is considered a superior procedure for this particular setting, as compared to the PI method.

The orange curve appears to be the ‘optimal’ choice for virtually all simulation scenarios considered in this setting, in the sense of its ability to preserve the family-wise error rate. In contrast to the NORTA procedure that adjusts for the complex gamma dependence, a simplistic adjustment for the correlation in the denominators can be derived by a coordinate-wise linear interpolation between the partial independent (the orange curve) and the PI (the blue curve). Therefore, it is referred to as the correlation interpolation approach.

Assume that the estimated correlation between the denominator of the $i^{th}$ marginal component of the test statistic $T$ and that of all other components is denoted by
\( \gamma_{i,j}^{(i)}, j \neq i \). The \( i^{th} \) critical value for the correlation interpolation approach is achieved via a linear interpolation between the \( i^{th} \) critical values obtained using the RPI method and the partial independence method as

\[
\xi_{i}^{\text{Corr}} = (\xi_{i}^{RPI} - \xi_{i}^{\text{Indep}}) \times w_i + \xi_{i}^{\text{Indep}}
\]

(3.44)
given that \( w_i \) is the averaged estimated correlation calculated as,

\[
w_i = \frac{1}{(\nu) - 1} \sum_{j \neq i} \gamma_{i,j}^{(i)}
\]

(3.45)

Among the various procedures that were in place before this work and a number of new methods we developed for either accuracy improvement or computational convenience, our general recommendation is to adopt the NORTA procedure, that accommodates the partial independence procedure and the PI procedure as extreme cases, and that accounts for the dependence feature of the gamma variables in a more flexible manner. Similarly, the linear interpolation of the comparison-specific correlation between the partial independence procedure and the PI procedure is demonstrated to yield actual FWER closer to the target level \( \alpha \), revealing that it could be an alternative approach for consideration with enhanced performance, as compared to either the partial independence procedure and the PI procedures.

The simulation results in other simulation settings depicted in the beginning of this section are included in the end of the section. As similar patterns are exhibited across all simulation settings set up in this simulation study, detailed comparisons are omitted here.
Figure 3.1: The Estimated Family-wise Error Rate for the set of All Pairwise Comparisons Among 8 Treatment Groups with Various Approximate Procedures
Figure 3.2: The Estimated Family-wise Error Rate Produced by Various Approximate Procedures under Simulation Settings 2, 3, and 4
Table 3.3: The Estimated Family-wise Error Rate for Various Approximate Procedures Under each Scenario in Simulation Setting (1)

<table>
<thead>
<tr>
<th>Procedures</th>
<th>Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PI (DF rounded down)</td>
<td>0.0776</td>
</tr>
<tr>
<td>PI (DF rounded up)</td>
<td>0.0731</td>
</tr>
<tr>
<td>PI (DF Interpolated)</td>
<td>0.0524</td>
</tr>
<tr>
<td>RPI</td>
<td>0.0674</td>
</tr>
<tr>
<td>NORTA</td>
<td>0.0796</td>
</tr>
<tr>
<td>Resampling</td>
<td>0.0559</td>
</tr>
</tbody>
</table>

Table 3.4: The Standard Errors associated with the Estimated Family-wise Error Rate for Various Approximate Procedures Under each Scenario in Simulation Setting (1)

<table>
<thead>
<tr>
<th>Procedures</th>
<th>Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PI (DF rounded down)</td>
<td>0.00374</td>
</tr>
<tr>
<td>PI (DF rounded up)</td>
<td>0.00369</td>
</tr>
<tr>
<td>PI (DF Interpolated)</td>
<td>0.00313</td>
</tr>
<tr>
<td>RPI</td>
<td>0.00342</td>
</tr>
<tr>
<td>NORTA</td>
<td>0.00369</td>
</tr>
<tr>
<td>Resampling</td>
<td>0.00322</td>
</tr>
</tbody>
</table>
Chapter 4: Separation of Modeling and Inference

4.1 Introduction

A common goal of many data analysis tasks is to understand the true mechanism that produces the data. In principle, a full representation of the mechanism requires an infinitely large number of samples from the underlying population, which is practically impossible due to both time and budget constraints. In practice, a given set of observed data, experimental or observational, carries only a finite amount of information about the true mechanism. Hence, it is often an unachievable goal to acquire a complete understanding of the true data generating mechanism based on a finite sample, regardless of the selection of data analysis tools. Moreover, an appropriate statistical model, often regarded as an idealized version of the true reality, is used to capture as much information about the truth contained in the sample data as possible, and in turn provides a reasonable basis for the statistician to make formal inferences about the data generating mechanism. As such, the intent of data analysis is often thought of as an effort to seek a parsimonious model that approximates the truth based on the observed data, rather than an attempt to fully identify the underlying true mechanism.
It is a general philosophy of statistics that statistical modeling and inferences go hand in hand. The effectiveness of the resulting inferences concerning a predetermined hypothesis or a set of hypotheses is closely tied to the appropriateness of the associated statistical model. It is a frequent occurrence that one statistical model may give support to a particular hypothesis, whilst an alternative model does not. Seemingly, the potentially contradictory inferences may leave the utility of the hypothesis test open to contention. In fact, it is a widely accepted notion that reliable inferences rest upon a solid statistical model that adequately describes the underlying true mechanism. There has been a vast literature on investigating the consequences of model misspecification on the resulting inferences, in both frequentist and Bayesian settings.

On the philosophical level, the best possible model, among many candidate models, is often identified a priori in a separate procedure from the statistical inferences to be made. One convention in selecting the best model is to go with a parsimonious model that adequately fits the data, as is the principle behind many model selection criteria. On the other hand, the scientific questions of the researcher’s interest concerning the truth are often expressed in the form of hypotheses. Once the best model is established, the ensuing inferences on the set of hypotheses can be performed accordingly.

When creating the idealized model, the statistician will often turn to common families of statistical models, formulated under certain working assumptions. For instance, the linear regression model presumes that the relationship between an explanatory variable and the mean of the response variable is linear, that observations
are independent and that the conditional distribution of the response given the explanatory variables is Gaussian, that the variances of two (or more) Gaussian distributions are identical and so on. The appropriateness of the modeling techniques depends largely upon the validity of the model assumptions for the data being analyzed. Moreover, a variety of validation methods, ranging from graphical visualizations to formal statistical tests, exist to aid the study the suitability of the model assumptions.

The assumptions that comprise the idealized model are unlikely to be precisely true, but they often provide a good approximation to reality and do little harm to the subsequent inferential procedures. At other times, the assumptions may conflict with the reality to such a degree, that the subsequent inference suffers in the sense that the parameter estimates may be badly biased, hypothesis tests may not hold the pre-determined level and the actual coverage of confidence intervals and prediction intervals may stray far from the nominal coverage. In addition, it is worthwhile to note that the impact of violations of model assumptions may also depend on the types of inferences to be made based on the resulting model. For instance, the importance of the normality assumption may be relatively weak for inferences concerning the mean responses of a number of treatments when sample sizes are sufficiently large; the homogeneous variance assumption, however, shows strong relevance for the same set of inferences.

This chapter focuses on the investigation of the influence on subsequent inferences of different modeling strategies when the data generating process that yields the outcome variable of interest is, upon close inspection, a hybrid of two qualitatively separate portions. One portion is usually stable and amenable to statistical modeling, while the other is unstable and subject to seemingly erratic variation. The subsequent
discussion is guided by the following research question: How do the inferences differ when one models both the stable and unstable processes as opposed to modeling the stable process with the unstable process being accounted for in the inference?

As an example, we consider a manufactured consumer product. The manufacturing process is closely monitored with a variety of effective quality control techniques that are designed to produce stable output — a stream of products for which measurements on sample products resemble independent draws from some distribution. Moreover, the sample to sample variability could be due to unavoidable small fluctuations of the operating conditions in the manufacturing facilities, such as temperature, moisture, etc., and in common with many other data collections, random measurement error. As a consequence of the stability of the manufacturing process guaranteed by the routine quality control procedure, the measurements taken directly on the sampled product units can be described by a probability model with time-invariant parameters. In other words, the mean and the variability of the measurements on sample products should be constant over time, assuming that the product line does not need a major change. We refer to this stable manufacturing process as the slowly varying component of the analysis.

In today’s competitive market, a wide selection of goods with similar attributes are offered for a specific purpose, even though the products not perfect substitutes because all consumers do not place the same valuations on all attributes of the goods, each product has a utility to the consumer. The consumer’s purchase intention is affected by the quality of the goods, along with an additional main consideration – the cost of the product. For the remainder of the chapter, we make an economically sensible assumption that a rational consumer seeks to purchase the goods, among a
large number of closely related goods, that maximizes the utility to price ratio, i.e.,
the benefit gained per dollar spent. In addition, the unit cost for any particular good
is assumed to be constant, independent of the number of units being purchased. In
contrast to the quality of a product (and hence the utility for a given consumer)
strictly maintained by the quality control process, the price of a product is subject
to rapid fluctuations due to various reasons, e.g., the variety of vendors monitor and
control their inventory, seek to expand their business with lower profit margins and
other special pricing and sale prices. The price fluctuations are often substantial and
are the result of a dynamic competitive environment. It is thus referred to as the
quickly varying component of the analysis.

The traditional approach for modeling such a composite system proceeds by blend-
ing the stable (manufacturing) and unstable (pricing) portions together to produce
a single response variable that is reflective of the unified stochastic process of the
two separate portions. For the quality price ratio, the derived variable is simply
constructed by standardizing the measurements taken directly on the product units
with respect to their corresponding prices. Accordingly, the derived variables provide
information about the mean level of the price-adjusted response. The remaining anal-
ysis is performed on the derived variable in the usual sense, using the broad spectrum
of conventional statistical techniques.

Despite the wide application and easy interpretation of the traditional modeling
approach, its versatility suffers from the rapid variation of the quickly varying com-
ponent. In other words, any inferences drawn based on a suitable model developed
for a particular realization of the unstable process must be constantly monitored and
modified to ensure that it remains relevant as the unstable process varies. The validity of the inferences does not hold when the unstable process evolves in an irregular fashion.

Another drawback of the traditional approach arises from its lack of interpretability with regard to the hybrid system. Suppose that the variance \( \sigma_i^2, i = 1, \ldots, \nu \), of the original measurement \( Y_{ij}, i = 1, \ldots, n_j, j = 1, \ldots, r \), taken directly on the product units has the functional form,

\[
\sigma_i^2 = \sigma^2 f(\mu_i, \theta). \tag{4.1}
\]

The variance function \( f \) may or may not depend on the mean response \( u_i \), and \( \theta \) is the structural parameter. The variance \( \sigma_i^2, i = 1, 2 \), is intrinsically related to the manufacturing fluctuations, measuring the variability of the manufactured products. Excessive variation reveals an increased risk of manufacturing defective product. Thus, it serves as an important performance index in measuring the stability of a manufacturing process. The quality control department of a company aims to reduce variability to ensure the quality of a manufactured product is consistent with a defined set of quality criteria through routine quality inspections and analysis. The significance of an accurate understanding of the variance structure resides not only in the subsequent inferences, but also in the routine production processes.

The traditional approach works with the appropriately standardized response variable with respect to the quickly varying quantity \( p \), denoted by

\[
W = \left\{ W_{ij} = \frac{Y_{ij}}{p_i}, i = 1, \ldots, n_j, \; j = 1, \ldots, \nu \right\}
\]
It can be asserted that the variance structure of $W_{ij}$ depends greatly on the quickly varying variable $p$. Mathematically, $\sigma_{W,i}^2$ can be expressed as

$$\sigma_{W,i}^2 = \sigma^2 f(u_i, \theta)/p_i^2.$$ \hspace{1cm} (4.2)

The expression shows that the variance may be greatly inflated or reduced depending on the magnitude of the quantity $p$. Therefore, the systematic structure of the variance, e.g., homogeneous variance or variance as a function of the mean, is highly likely to be obscured by the incorporation of the quickly varying quantity $p$ into the model. This makes it a challenging task to build a proper model that describes the variance structure sufficiently well. This, together with the fact that the model built at a particular price vector $p$ is only of tentative use, compromises the attractiveness of the common approach.

In contrast to this approach, we advocate modeling for the stable portion of the problem, with full use of traditional tools and push the unstable portion of the problem — price — into the inference. Both conceptual and empirical evidence as to why the separation of quickly varying component into the inference part is considered more statistically favorable will be provided in detail.

In addition to the ease of modeling of the error structure, the resulting model as a description of the stable manufacturing process remains valid over time, provided that the manufacturing procedure does not need to be advanced in major aspects. Nevertheless, this necessitates the development of novel inference strategies and it suggests a broader view of inference. Rather than a single inferential statement, a family of statements is made with inference available for each specific variation of the unstable portion of the problem. This provides the analyst with the ability to react quickly to a changing environment.
An additional question that can provide tremendous business value is the partition of the price space into disjoint regions, each of which leads to a unique conclusion as to the selection of the best commodities. This research question arises from the urge for retailers to exploit the possibilities for a pricing policy to elevate the competitiveness of their products. An effective pricing strategy can appreciably impact the customers’ purchase tendency towards a product with slightly compromised quality, while it may seem implausible or overly costly for the manufacturer to improve through modification/replacement of the product line. An efficacious pricing strategy can act as a substitute for product quality. The goal of identifying the price that generates the maximum profit can be accomplished by first identifying the set of prices, in a relative sense with reference to other alternative goods, that maintains or improves the competitive capability of the underlying product according to the criterion of quality price ratio.

The use of the new data analysis technique that builds a versatile model for the stable process, i.e., product quality, and moves the unstable process, i.e., cost, to inference provides a unified appealing way to assess and compare the cost-adjusted performance of various related products, and in particular to set the pricing strategy. On the contrary, the traditional approach treats each feasible price vector as an individual case and requires that one model the cost-adjusted response variable catering to each specific price vector.

To summarize, the two analyses, one widely used as the current practice and the other one promoted in this work, differ from one another in many aspects. The notion of the moving the unstable process into the inference is the basis of the new approach. To make a distinction with the new approach, we term the analysis on the derived
variable the traditional approach. Apart from the conceptual advantages of the new approach as justified earlier, we illustrate, in the subsequent section, the contrast of the two modeling approaches using a classic dataset on battery lifetime (Dean and Voss, 1999).

4.2 Formulation of the Two Approaches

This section investigates the impact on the inference of the two analyses — the inclusion or exclusion of the quickly varying component in the model in a simple one-way analysis of variance (ANOVA) setting. As usual, the main goal is to determine the significance of the difference between the various products in terms of the quality price ratio. We first sketch the theoretical results on the statistical inference in relation to the two analyses. Shortly, the battery lifetime dataset (Dean and Voss, 1999) will be analyzed to empirically illustrate the difference.

4.2.1 The Traditional Approach

The traditional modeling strategy considers the two qualitatively separate portions — manufacturing and pricing — of the composite stochastic mechanism as a whole, by blending the two components into a single derived response variable for analysis. Suppose that $Y_{ij}, i = 1, \ldots, \nu; j = 1, \ldots, n_i$, denotes the measurement on the $j^{th}$ product unit in the batch of $n_i$ randomly selected samples from the $i^{th}$ consumer product. In addition, the unit cost for product $i$, the quickly varying component of the analysis, is denoted by $p_i, i = 1, \ldots, \nu$. 
Let $W_{ij}$ denote the cost-adjusted product quality on the $j^{th}$ sample product of type $i$, i.e., $W_{ij} = \frac{Y_{ij}}{p_i}$. Consider the one-way ANOVA model,

$$W_{ij} = \alpha_i + \gamma_{ij} \quad i = 1, \ldots, \nu; \quad j = 1, \ldots, n_i$$

$$\gamma_{ij} \sim N(0, \tau_i^2).$$

To make reliable inferences on the means across the $\nu$ products, it is essential to model the error variances $\tau_i^2$ adequately. We note that the incorporation of the quickly varying portion into the model inevitably disturbs the systematic structure of the error variances, resulting from the inherently erratic variation of the unstable process. Therefore, the ‘best’ model that would be used for inferences, as the unstable process evolves, needs to be altered to adapt to the new environment.

The families of linear comparisons can take many different forms, the choice of which is typically guided by the original research questions of interest. We focus on families of linear hypotheses specified through a linear coefficient matrix $L \in \mathbb{R}^{m \times \nu}$ as

$$H_0 : L\alpha = 0 \quad v.s. \quad H_a : L\alpha \neq 0,$$

where the coefficient matrix $L^T$, for example, for pairwise comparisons on the mean quality price ratio across, in the instance of $\nu = 4$, types of products is formulated as

$$L = \begin{bmatrix}
1 & -1 & 0 & 0 \\
1 & 0 & -1 & 0 \\
1 & 0 & 0 & -1 \\
0 & 1 & -1 & 0 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & -1
\end{bmatrix}$$

The simultaneous inference procedure for the family of linear comparisons in (4) is driven by the specification of the error variance structure. In the case of homogeneous variance, a general procedure is presented in Section 3.2.1 is the standard
choice that ensures effective control of the designated nominal size $\alpha$. Furthermore, when the ratios of error variances for the various populations are known, the classical procedure designed for the homogeneous data can be easily modified to accommodate the variance heterogeneity.

However, in the case of heterogeneous variances, no exact solution for linear hypotheses exists. A commonly used statistical technique is to find a variance stabilizing transformation that equalizes the error variances of the transformed data. The remaining analysis thus proceeds with the transformed data that appear to meet the homogeneity criterion, using the conventional modeling techniques in the homogeneous framework. However, inferences of interest, typically on the original measure, require backward transformation. One disadvantage of the transformation approach, if an appropriate one can be identified, that may hinder the analyst’s ability in making proper inference is the difficulty in translating the exact inference on the means of transformed variable back to that of the original scale, except for some special transformations. For this reason, one may turn to an approximate procedure specifically designed for unequal variance data that works with the original data and attempts to preserve the pre-specified level $\alpha$, as presented in Section 3.2.2.

4.2.2 The New Approach

We promote the use of the new approach that models the stable portion of the process and pushes the time-varying portion into the inference. The avoidance of the quickly varying component in the model is potentially advantageous in the sense that the variance structure of the stable process is amenable to statistical modeling,
which in turn is favorable for the subsequent inference. More importantly, as mentioned earlier, error variability often serves as an important performance indicator pertinent to the product design and quality control, apart from the mean product quality. Lastly, the resulting model remains valid and provides continual guidance on performing inferences as the unstable process changes.

Recall that $Y_{ij}$ represents the original measurement taken on the $j^{th}$ sample unit of product $i$. Consider the one-way ANOVA model,

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1, \ldots, \nu; \quad j = 1, \ldots, n_i$$

$$\varepsilon_{ij} \sim N(0, \sigma_i^2)$$

The variability $\sigma_i^2$ in the manufactured products from the stable manufacturing processes is often a result of random causes. In contrast to the traditional approach, the multiple hypotheses on the mean quality price ratio are formulated as

$$H_0 : L_p \mu = 0 \quad v.s. \quad H_a : L_p \mu \neq 0,$$

where the linear comparisons matrix $L_p^T$, for pairwise comparisons in the instance of $\nu = 4$, is given by

$$L_p = \begin{bmatrix}
\frac{1}{p_1} & -\frac{1}{p_2} & 0 & 0 \\
\frac{1}{p_1} & 0 & -\frac{1}{p_3} & 0 \\
\frac{1}{p_1} & 0 & 0 & -\frac{1}{p_4} \\
0 & \frac{1}{p_2} & -\frac{1}{p_3} & 0 \\
0 & \frac{1}{p_2} & 0 & -\frac{1}{p_4} \\
0 & 0 & \frac{1}{p_3} & -\frac{1}{p_4}
\end{bmatrix}.$$

As an aside, many other linear comparisons may be of interest, in addition to the pairwise comparisons. For instance, the experimenter may find it interesting to decide between the store brand and name brand batteries. The hypothesis can be formulated as an equality test on the mean lifetime per dollar averaging over the
heavy duty and alkaline types, i.e.,

\[ H_0 : \frac{\alpha_1 + \alpha_3}{2} = \frac{\alpha_2 + \alpha_4}{2} \quad \text{vs.} \quad H_a : \frac{\alpha_1 + \alpha_3}{2} \neq \frac{\alpha_2 + \alpha_4}{2}. \]

In the succeeding section, the battery lifetime data (Dean and Voss, 1999) is analyzed to empirically demonstrate the differences between the two methodologies in performing statistical inferences on the pairwise comparisons across the four types of batteries in terms of mean lifetime per cost.

### 4.3 The Battery Lifetime Example

The battery life experiment was run by Dean and Voss (1999) and used in their text *Design and Analysis of Experiments*. It was designed as a completely randomized design with two treatment factors, each having two levels. These are battery ‘duty’ (level 1 = alkaline, level 2 = heavy duty) and ‘brand’ (level 1 = name brand, level 2 = store brand). This gives four treatment combinations recoded as battery types 1 to 4. In the battery experiment, the experimenter was interested in comparing the average lifetime per unit cost of each battery type with that of each of the other battery types. The main goal is to identify the type of the non-rechargeable battery that delivers the largest mean lifetime per unit cost. The data consists of measurements on the useful lifetime of four types of non-rechargeable batteries along with the prices for the batteries. The data appears in Table 4.1.

The Traditional Approach

The traditional approach, as described in Dean and Voss (1999), performs the analysis on the battery life per unit cost data. The residual plot versus the fitted values, shown in the upper left panel of Figure 4.1, indicates constant error variances.
Table 4.1: The Battery Lifetime Data

<table>
<thead>
<tr>
<th>Battery Type</th>
<th>Life (min)</th>
<th>Unit cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkaline, Name Brand (Type 1)</td>
<td>602, 529, 534, 585</td>
<td>0.985</td>
</tr>
<tr>
<td>Alkaline, Store Brand (Type 2)</td>
<td>863, 743, 773, 840</td>
<td>0.935</td>
</tr>
<tr>
<td>Heavy Duty, Name Brand (Type 3)</td>
<td>232, 255, 200, 215</td>
<td>0.520</td>
</tr>
<tr>
<td>Heavy Duty, Store Brand (Type 4)</td>
<td>235, 282, 238, 228</td>
<td>0.495</td>
</tr>
</tbody>
</table>

across the four types of batteries. Therefore, the model inferences will be conducted under the homogeneous variance assumption, with the pooled variance being used as an estimate of the constant error variance.

Using notation established in Section 3.2.1, the test statistic $T_W$ for the family of all pairwise comparisons, as stated in (4), is given as

$$T_W = \left\{ \frac{\bar{W}_i - \bar{W}_j}{\sigma_W \sqrt{1/n_i + 1/n_j}}, 1 \leq i < j \leq 4 \right\}.$$  

The null distribution of the test statistic $T_W$ is a classical multivariate $t$ distribution with 12 degrees of freedom and correlation structure $R_W = \{r_{W,ij}, 1 \leq i, j \leq 4\}$, where $r_{W,ij} = \frac{r_{ij}^W}{\sqrt{l_i^Wl_j^W}}$ for balanced sample sizes. The set of multiplicity-adjusted critical values associated with the family-wise error rate $\alpha = 0.05$ is obtained as the 95$^{th}$ percentile of the null distribution. Assuming equal individual error rates to account for equal importance of all pairwise comparisons, the use of the quantile calculation algorithm for classical multivariate $t$ distributions developed in Section 2.1.4 yields the critical vector $\xi$ with constant entry $\xi_i = 2.97$. Table 4.2 shows the simultaneous inference results using the classical procedure presented in Section 3.2.1 for homogeneous data.
Figure 4.1: The Upper Left (right) is the Residual Plot for the Battery Life per Unit Cost (Battery Life) data. The Lower Left (right) panel is the Residual Plot for the Square Root (Log) Transformed Battery Life Data.
Table 4.2: Traditional Approach: Simultaneous Inferences on the Family of Pairwise Comparisons on the Average Life per Unit Cost for the Four Types of Batteries.

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Observed Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.41</td>
<td>2.97</td>
<td>5.08e-06</td>
</tr>
<tr>
<td>1 – 3</td>
<td>3.99</td>
<td>2.97</td>
<td>0.0086</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.17</td>
<td>2.97</td>
<td>0.19</td>
</tr>
<tr>
<td>2 – 3</td>
<td>12.40</td>
<td>2.97</td>
<td>2.8e-08</td>
</tr>
<tr>
<td>2 – 4</td>
<td>10.57</td>
<td>2.97</td>
<td>8.62e-07</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.82</td>
<td>2.97</td>
<td>0.31</td>
</tr>
</tbody>
</table>

As with the univariate hypothesis, the decision rules for the 6 pairwise comparisons are jointly formed equivalently by (1) comparing the magnitudes of the observed test statistic with the critical value, i.e., the $i^{th}$ individual hypothesis is rejected if the marginal test statistic $|t_i| > \xi_i$; or (2) comparing the adjusted $p$-value with the pre-specified significance level $\alpha$, i.e., we reject the $i^{th}$ individual hypothesis if the adjusted $p$-value $p_i < \alpha$.

The traditional modeling approach leads to the conclusion that all pairwise differences are inferred to be significant, except for pairs 1, 4 and 3, 4 where the means are indistinguishable from one another. In addition, the probability of concluding that some indistinguishable pair of battery types are different is preserved at level $\alpha = 0.05$.

One intrinsic weakness of the traditional approach lies in the fact that the response variable involves the unstable process. The constant variance assumption on the derived variable at one particular snapshot of the dynamic system may not be appropriate as the system progresses. Consequently, for the battery example, the
suitable model developed for the specific price recorded by the experimenter provides no guidance on how to perform pairwise comparisons when a change in price occurs. As such, we propose an alternative approach that models the stable process while pushing information about the unstable process into the inference.

The New Approach

The new approach builds the model to be used for inference based on the original measure — battery life. The residual plot in the upper right panel of Figure 4.1 shows that the error variability tends to increase modestly with the mean battery lifetime. Dean & Voss suggest using a transformation to stabilize the error variance. The lower left (right) residual plot in Figure 4.1 is based on the square root (log) transformed battery life. Both residual plots are consistent with constant variance with no noticeable patterns, suggesting both transformations are sensible to use.

Log Transformation: The model is expressed as,

\[ W_{ij} = \log(Y_{ij}) = \beta_i + \varepsilon_{ij}, \ i = 1, \ldots, \nu; \ j = 1, \ldots, n_i \]

\[ \varepsilon_{ij} \sim N(0, \sigma^2_{\log}). \]

To fulfill the goal of making comparisons of the four types of batteries in terms of the average battery life per unit cost, it is necessary to derive the mean battery life \( E(Y_{ij}) \) from the model fitted to the log transformed battery life data, i.e., \( \log(Y_{ij}) \). The mean lifetime is given by \( E(Y_{ij}) = E(e^{W_{ij}}) = e^{\beta_i + \sigma^2_{\log}} \). Hence, the family of all pairwise comparisons on the mean battery life per dollar can be set up as

\[ H_0 : \frac{e^{\beta_i + \sigma^2_{\log}}}{p_i} = \frac{e^{\beta_j + \sigma^2_{\log}}}{p_j}, \ \forall 1 \leq i < j \leq 4, \]

which can be simplified to

\[ H_0 : \beta_i - \beta_j = \log(p_i/p_j), \ \forall 1 \leq i < j \leq 4. \]
One advantage of the use of the log transformation, in contrast to other transformations, as a variance stabilization transformation is that any set of linear hypotheses of interest concerning the original mean measure, i.e., \( E(Y_{ij}) \) can be translated to that on the means of the transformed variables, i.e., \( E(\log(Y_{ij})) = \beta_i \). The set of linear hypotheses on the transformed data is performed under the homogeneous variance assumption, for which an exact procedure with easy implementation is available.

**Square Root Transformation:** The model is specified as

\[
U_{ij} = \sqrt{Y_{ij}} = \gamma_i + \varepsilon_{ij} \quad i = 1, \ldots, \nu; \quad j = 1, \ldots, n_i
\]

\[
\varepsilon_{ij} \sim N(0, \sigma^2_{sqrt}).
\]

Following from the observation that

\[
E(Y_{ij}) = E(U_{ij}^2) = Var(U_{ij}) + (E(U_{ij}))^2 = \gamma_i^2 + \sigma^2_{sqrt}
\]

Hence, the family of all pairwise comparisons on the mean battery life per dollar is set up as

\[
H_0 : \frac{\gamma_i^2 + \sigma^2_{sqrt}}{p_i} = \frac{\gamma_j^2 + \sigma^2_{sqrt}}{p_j}, \quad \forall 1 \leq i < j \leq 4.
\]

Unlike the log transformation, the set of hypotheses concerning the means of \( Y \) is re-expressed in terms of both the mean \( \gamma_i \) and constant variance \( \sigma^2_{sqrt} \) of the square root transformed data. Indeed, the appearance of the error variance in the hypothesis leaves it a non-standard inference problem. The square root transformation serves as an example of one of the major drawbacks of the transformation approach in dealing with heterogeneity. That is, the exact inference drawn on the means of the transformed data does not extend to that of the original data. As was commented by Dean & Voss (Page 137), ‘any more detailed interpretation of the results is muddled by use of the transformation.’
This drawback provides a motivation to turn to an approximate simultaneous
inference procedure designed for non-constant variances. The development of a new
approximate inferential procedure in the presence of heterogeneity has been the main
focus of Chapter 3. In what follows, various approximate procedures presented in
Chapter 3 will be applied to the battery life data.

Table 4.3 gives the multiplicity-adjusted critical vector and the adjusted \( p \)-value for
the family of pairwise comparisons on the cost-adjusted average battery life, specified
through the coefficient matrix \( L_p \), based on the Plug-in (PI) method developed by
Hasler and Hothorn (2008). To cater for the needs of the computational algorithm
in the multivariate \( t \) probability calculation developed by Hasler and Hothorn (2008)
that accompanies the PI procedure, the authors suggest the implementation of the
PI procedure with the degrees of freedom dropped down to the next integers. The
discrepancy between the critical vector that matches the PI procedure described by
the authors and that resulting from the rounding strategy for the computational
feasibility tends to be larger for small degrees of freedom. Moreover, the faithfulness
of the implementation algorithm to the PI procedure is also driven by the closeness
of the fractional degrees of freedom to the rounded down integer. The PI procedure
is the most compromised for small fractional degrees of freedom reaching toward the
next integers.

As such, in addition to the rounding strategy on the Satterthwaite degrees of
freedom suggested by the authors, the critical vectors along with the adjusted \( p \)-values
are also provided in Tables 4.4 and 4.5 respectively, based on the use of Satterthwaite
degrees of freedom rounded up to the next integer and the linear interpolation on the
degrees of freedom, as described in Section 2.2. Comparing across Tables 4.3 4.4

156
### Table 4.3: The Simultaneous Inferences on the Battery Example using the Plug-in Approach (with Rounded Down Degrees of Freedom by Default)

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.208</td>
<td>4.060</td>
<td>0.00404</td>
</tr>
<tr>
<td>1 – 3</td>
<td>4.688</td>
<td>3.678</td>
<td>0.0196</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.411</td>
<td>3.678</td>
<td>0.190</td>
</tr>
<tr>
<td>2 – 3</td>
<td>11.344</td>
<td>3.678</td>
<td>0.000213</td>
</tr>
<tr>
<td>2 – 4</td>
<td>9.357</td>
<td>3.677</td>
<td>0.000874</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.869</td>
<td>3.677</td>
<td>0.343</td>
</tr>
</tbody>
</table>

and [4.5], it clearly shows that the critical value for the 1st hypothesis testing the equality of batteries 1 and 2 is largely affected by the rounding strategy, due to the dropping from 4.997 to 4 of $d_f_1$, as shown in Table [4.6]. The consequence of neglecting the fractional portion of the degrees of freedom for the multivariate $t$ probability calculation was investigated in Section [2.1]. The lack of ability in precise calculation of the multivariate $t$ probability and quantile is inevitably transmitted to the inference procedures, for which the critical values are constructed using multivariate $t$. The linearly interpolated critical value in Table [4.5] matches closely the PI methodology outlined in Hasler and Hothorn (2008).

The critical vectors and the adjusted $p$-values for the family of pairwise differences in average battery life per unit cost, based on the various approximate procedures proposed in this work in Section 3.2.2, namely the rank-based procedure (RPI) and the NORTA procedure and the resampling-based approach, are provided in Table [4.7], [4.8] and [4.9] respectively. Figure [4.2] shows the critical values for various procedures investigated in this section.
Table 4.4: The Simultaneous Inferences on the Battery Example using the Plug-in Approach with Rounded Up Degrees of Freedom

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.208</td>
<td>3.685</td>
<td>0.00182</td>
</tr>
<tr>
<td>1 – 3</td>
<td>4.688</td>
<td>3.452</td>
<td>0.0133</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.411</td>
<td>3.451</td>
<td>0.173</td>
</tr>
<tr>
<td>2 – 3</td>
<td>11.344</td>
<td>3.456</td>
<td>0.000219</td>
</tr>
<tr>
<td>2 – 4</td>
<td>9.357</td>
<td>3.455</td>
<td>0.000497</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.869</td>
<td>3.451</td>
<td>0.329</td>
</tr>
</tbody>
</table>

Table 4.5: The Simultaneous Inferences on the Battery Example using the Plug-in Approach with Linearly Interpolated Degrees of Freedom

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.208</td>
<td>3.686</td>
<td>0.00182</td>
</tr>
<tr>
<td>1 – 3</td>
<td>4.688</td>
<td>3.504</td>
<td>0.0148</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.411</td>
<td>3.553</td>
<td>0.181</td>
</tr>
<tr>
<td>2 – 3</td>
<td>11.344</td>
<td>3.547</td>
<td>0.000293</td>
</tr>
<tr>
<td>2 – 4</td>
<td>9.357</td>
<td>3.502</td>
<td>0.000566</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.869</td>
<td>3.461</td>
<td>0.329</td>
</tr>
</tbody>
</table>

Table 4.6: The Satterthwaite Degrees of Freedom for each Marginal Null Distribution of the Test Statistic

<table>
<thead>
<tr>
<th>Individual Comparisons</th>
<th>1 – 2</th>
<th>1 – 3</th>
<th>1 – 4</th>
<th>2 – 3</th>
<th>2 – 4</th>
<th>3 – 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satterthwaite Degrees of Freedom</td>
<td>4.997</td>
<td>5.769</td>
<td>5.558</td>
<td>5.586</td>
<td>5.791</td>
<td>5.955</td>
</tr>
<tr>
<td>Degrees of Freedom used in PI</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure 4.2: Multiplicity-adjusted Critical Values on the Battery Life Data using Various Approximate Procedures
The implementation of all these procedures is carried out using the variance reduction technique through the spherical radial transformation (SRT). The estimated gamma correlation, the key feature that distinguishes the NORTA algorithm from other procedures, is given by

\[
\hat{\Sigma}_W = \begin{bmatrix}
1.00 & 0.198 & 0.174 & 0.811 & 0.772 & 0.000 \\
0.198 & 1.000 & 0.271 & 0.413 & 0.000 & 0.535 \\
0.174 & 0.271 & 1.000 & 0.000 & 0.491 & 0.668 \\
0.811 & 0.413 & 0.000 & 1.000 & 0.718 & 0.319 \\
0.772 & 0.000 & 0.491 & 0.718 & 1.000 & 0.431 \\
0.000 & 0.535 & 0.668 & 0.319 & 0.431 & 1.000
\end{bmatrix}.
\]

The rank plug-in (RPI) procedure conforms to the same set of principles established by the PI procedure, with the slight difference that the assumed maximum positive dependence of the denominators is measured in the rank sense, as opposed to the conventional moment-based correlation as used in the PI procedure. This conceptual equivalence between the two approaches is validated by the empirical results that the critical vectors in Table 4.5 and Table 4.7 are close together, in comparison to their closenesses to other procedures. In addition, the uncertainty measure on the adjusted \(p\)-value, shown in the last column of Table 4.7, illustrates the stability of the numerical algorithm in the calculation of the generalized multivariate \(t\) probability. Moreover, a major advantage of the rank-based implementation, as compared to the PI method adjusted by the linear interpolation, lies in the fact that the number of required computations is drastically reduced via the need to evaluate one table, rather than \(m\) tables, one for each comparison.

However, one can expect that the critical value for each comparison is underestimated due to the use of complete dependence on the denominators, when some pairs are independent (e.g., 1 vs. 2 and 3 vs. 4). The critical vector obtained by the use
of the NOTRA procedure is uniformly larger than the two implementations of the Plug-in method in accordance with the authors’ conceptual argument — the RPI and the Plug-in with degrees of freedom adjusted by linear interpolation. The intuition behind this phenomenon is that the assumption on the maximum positive dependence for the denominators causes the positive correlation across the components of the multivariate test statistic to be overly accounted for, yielding an estimated critical vector less than what would be produced were the true null distribution known and used for inference.

As is shown in Figure 4.2, the Plug-in method with the degrees of freedom dropped down to the nearest integers produces a critical vector with more resemblance with that from the NORTA algorithm, in contrast to other variations of the Plug-in method. However, the comparability results mainly from the adjustment made on the Satterthwaite degrees of freedom, rather than the inherent similarity of the procedures.

The PI method adjusts for ‘too small’ critical values by flooring the Satterthwaite degrees of freedom. Figure 4.3 shows how the adjustment works by comparing the critical vectors produced by the PI and NORTA methods, plotted against the fractional portion of the degrees of freedom (the step-down size). As a general trend, larger adjustment on the degrees of freedom tends to yield smaller discrepancy between the two critical values for each individual comparison, i.e., comparisons 1 and 6 (corresponding to the last two points in Figure 4.3) that drop from almost the next integer. The flooring of the degrees of freedom happens to work pretty well for the battery life data. However, this adjustment is arbitrary, depending entirely on the dataset being analyzed. That is to say, the closeness of the NORTA and the Plug-in
Figure 4.3: Multiplicity-adjusted Critical Values on the Battery Life Data using the PI and the NORTA Methods

with degrees of freedom dropped down to the nearest integers suggested by Hasler & Hothorn is not a repeated pattern, leaving the performance of the Plug-in method unstable. Moreover, as is revealed in Figure 4.3, the adjustment makes essentially no difference when the degrees of freedom are moderate or large.

In summary, as conveyed in Figure 4.2, the linear interpolated PI yields multiplicity-adjusted critical values that lie between the ones resulting from the PI method implemented with dropped down and rounded up integer degrees of freedom. Moreover, the closeness of each individual critical value to the two extremes depends on the
non-integer portion of the Satterthwaite degrees of freedom. The RPI and the linear interpolated PI produce critical values fairly close together, as they approximate the true null distribution to essentially the same degree, with the slight difference stemming from the choice of the measure for the pairwise correlation. The NORTA method is generally recommended for its ability to flexibly accommodate the major features of the true null distribution with minimum assumptions. For this particular example, the PI with rounded down degrees of freedom is roughly equally calibrated as the NORTA method, due to the haphazard adjustment for the critical values attributed to flooring the degrees of freedom. Nevertheless, we do not recommend it as a general solution. Lastly, Figure 4.2 also suggests that the resampling approach performs reasonably well, for this battery data, indicated by its likeness to the NORAT results. However, it suffers from the fact that it does not generalize Welch’s $t$-test for one pair of means. Welch’s $t$ test is the standard approach in the non-constant variance setting.

The adjusted $p$-value provides direct guidance on the decision rule in such a way that any individual hypotheses associated with adjusted $p$-values less than $\alpha$ are declared significant. Equivalently, a simple check on the magnitude of each test statistic as compared to the corresponding critical value leads to the inference, i.e., the $i^{th}$ hypothesis is inferred to be non-significant if $|t_i| > \xi_i$ for all $i$, while keeping the family-wise error rate approximating the target level $\alpha$. Comparing across the Tables, all approximate procedures for the new approach lead to the rejection of the overall hypothesis and conclude that battery type 4 (heavy duty, store brand) does not differ significantly from either type 3 (heavy duty, name brand) or type 1 (alkaline, name brand), while the rest of the pairwise differences are insignificant.

163
Table 4.7: The Simultaneous Inferences on the Battery Example using the RPI Approach

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted $p$-value</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.208</td>
<td>3.711</td>
<td>0.00105</td>
<td>3.655e-05</td>
</tr>
<tr>
<td>1 – 3</td>
<td>4.688</td>
<td>3.493</td>
<td>0.00847</td>
<td>1.0237e-04</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.411</td>
<td>3.546</td>
<td>0.115</td>
<td>3.382e-04</td>
</tr>
<tr>
<td>2 – 3</td>
<td>11.344</td>
<td>3.539</td>
<td>0.000103</td>
<td>1.175e-05</td>
</tr>
<tr>
<td>2 – 4</td>
<td>9.357</td>
<td>3.488</td>
<td>0.000254</td>
<td>1.777e-05</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.869</td>
<td>3.451</td>
<td>0.215</td>
<td>4.264e-04</td>
</tr>
</tbody>
</table>

In summary, based on the intellectual argument and the numerical evidence presented in this work, we recommend the use of the NORTA algorithm for simultaneous inferences in the presence of the heterogeneous variances.

Tables 4.2 and 4.8 show the analysis results based on the traditional approach (modeling the stable and unstable processes) and the new approach (modeling the stable process only) coupled with the NORTA procedure for this lifetime and prices data. The qualitative conclusions drawn from the two analyses are the same. However, there are big differences in the adjusted $p$-values. For the 1 vs. 3 comparison, the traditional approach differs from both the PI method and the NORTA methods at the $\alpha = 0.01$ level.
Table 4.8: The Simultaneous Inferences on the Battery Example using the NORTA Approach

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted $p$-value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.208</td>
<td>3.997</td>
<td>0.00134</td>
<td>3.655e-05</td>
</tr>
<tr>
<td>1 – 3</td>
<td>4.688</td>
<td>3.725</td>
<td>0.0104</td>
<td>1.024e-04</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.411</td>
<td>3.787</td>
<td>0.131</td>
<td>3.382e-04</td>
</tr>
<tr>
<td>2 – 3</td>
<td>11.344</td>
<td>3.778</td>
<td>0.000156</td>
<td>1.175e-05</td>
</tr>
<tr>
<td>2 – 4</td>
<td>9.357</td>
<td>3.719</td>
<td>0.000320</td>
<td>1.777e-05</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.869</td>
<td>3.677</td>
<td>0.239</td>
<td>4.264e-04</td>
</tr>
</tbody>
</table>

Table 4.9: The Simultaneous Inferences on the Battery Example using the Resampling Approach

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Test Statistic</th>
<th>Critical Value</th>
<th>Adjusted $p$-value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2</td>
<td>-8.208</td>
<td>3.958</td>
<td>0.00101</td>
<td>3.288 e-05</td>
</tr>
<tr>
<td>1 – 3</td>
<td>4.688</td>
<td>3.699</td>
<td>0.00936</td>
<td>9.676 e-05</td>
</tr>
<tr>
<td>1 – 4</td>
<td>2.411</td>
<td>3.759</td>
<td>0.130</td>
<td>3.355 e-04</td>
</tr>
<tr>
<td>2 – 3</td>
<td>11.344</td>
<td>3.751</td>
<td>0.000101</td>
<td>1.00494 e-05</td>
</tr>
<tr>
<td>2 – 4</td>
<td>9.357</td>
<td>3.693</td>
<td>0.000230</td>
<td>1.549e-05</td>
</tr>
<tr>
<td>3 – 4</td>
<td>-1.869</td>
<td>3.651</td>
<td>0.237</td>
<td>4.253e-04</td>
</tr>
</tbody>
</table>
Chapter 5: Discussion and Future Work

Multiplicity is an intrinsic problem of any simultaneous inference procedure. A general simultaneous inference procedure for any finite collection of linear estimable functions of the model parameters is presented in Section 3.2 in the context of the linear fixed effect model setting with the homogeneous variance assumption. The ability of the test procedure to control the nominal significance level is guaranteed by the availability of the true null distribution of the multivariate test statistic – a multivariate $t$ distribution. The implementation of the test procedure is available mainly through the ‘multcomp’ and ‘SimComp’ packages in the $R$ statistical package.

When heteroscedasticity arises, the true null distribution of the test statistic under this circumstance is typically unknown, suggesting the need for the development of an approximate procedure that produces adequate control of the test size. As a matter of fact, the investigation of multiple testing procedures for heterogeneous data has been a long standing statistics problem. The Plug-in (PI) procedure by Hasler and Hothorn (2008) is the state of the art approach for heterogeneous data, in the basic one-way ANOVA setting. Its success stems largely from its ability to capture the correlation among the numerators of components of the test statistic. Nevertheless, as extensively sketched earlier in the work, an inherent problem with the PI method lies in its lack of effort in attempting to account for the dependence structure among
the denominators, another major feature of the null distribution. In addition to this design deficiency, the PI procedure becomes computationally heavy, when the number of comparisons grows large. Moreover, its reliance on the ‘qmvt’ function in the ‘mvtnorm’ package contributed by Genz & Bretz (2001) for the multivariate $t$ probability and quantile calculation adds to the list of drawbacks of the PI method, as rounding the degrees of freedom down to the nearest integer inevitably introduces another source of inaccuracy that causes lack of control over the actual significance level in the PI method.

Our contribution to the multiple comparisons community is twofold. First, we presented the NORTA procedure that aims to incorporate the dependence among the gamma variates into the inference procedure so that the reference distribution that produces the multiplicity-adjusted critical vector resembles the true null distribution closely in the sense that all major features of the test statistics, namely the marginals, the Gaussian correlation and the gamma correlation are accounted for. The performance of the approximate procedures is quantified by the discrepancy between the actual family-wise error rate and the designated level $\alpha$. The superior performance of the NORTA procedure in reference to the PI method is numerically illustrated through a Monte Carlo simulation study on various simulation setting. The improvement is a result of adding the gamma correlation.

In addition, an efficient computational algorithm for the multivariate $t$ probability calculation was established, using the spherical radial transformation as a variance reduction technique. We conducted a small simulation study to show the amount of reduction in uncertainty that can be achieved with the variance reduction technique compared to the plain Table 0/1 approach. In addition, the greatest reduction is
attained with the choice of matrix factor with the number of columns matching the rank of the Gaussian correlation matrix. Coupling with the probability calculation algorithm, we developed a computationally efficient algorithm for the evaluation of the critical value functions of the multivariate $t$. The computing time is dramatically reduced through the ‘small table evaluation and large table polishing up’ strategy.

In the future, we can provide the multiple comparisons community with a package that accompanies the various approximate procedures we investigated in this work. For simultaneous inferences on heterogeneous data, our general recommendation is to choose the NORTA approach as a default selection, resulting from its flexible ability of making inferences based upon a reference distribution that preserves the marginal and dependence features of the unknown true null distribution better than other methods developed to date. The user may have the ability to switch from the default to other possible procedures, judging from the strength of correlations among the components of the test statistic for the particular dataset being analyzed.
Bibliography


