

DISSERTATION

OUTLIER DISCORDANCY TESTS BASED ON SADDLEPOINT APPROXIMATIONS

Submitted by

Andrew D. Sleeper

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2019

Doctoral Committee:

Advisor: Louis Scharf

Duane Boes

Jay Breidt

Anura Jayasumana

Copyright 2019 by Andrew D. Sleeper

All Rights Reserved

ABSTRACT

OUTLIER DISCORDANCY TESTS BASED ON SADDLEPOINT APPROXIMATIONS

When testing for the discordancy of a single observed value, a test based on large values of the maximum absolute studentized residual (MASR) or maximum squared studentized residual (MSSR) is known to be optimal, by maximizing the probability of correctly identifying an outlying value, while controlling the risk of a false identification to α . The exact distribution of MASR and MSSR is not known. In its place, the first Bonferroni bound on the distribution of these statistics is commonly used as an outlier test; see Grubbs (1950). We present new approximations to the distribution of MASR or MSSR, based on saddlepoint approximations of the density of statistics calculated from truncated normal random variables. These approximations are developed in three settings: a one-sample case, univariate regression, and multivariate regression. In comparisons with three versions of Bonferroni bounds and a Monte Carlo simulation, the saddlepoint approximations are shown to perform well in a wide range of situations, especially at larger sample size. The saddlepoint approximations also calculate faster than the improved versions of Bonferroni bounds.

ACKNOWLEDGEMENTS

Ronald Butler, my first advisor on this project, has been a long-term mentor and advisor to me, even after he left CSU. I am deeply indebted to Ron for his continued support and advice, especially during the challenging final stages of this project.

I am also grateful to my advisor, Louis Scharf, and my doctoral committee for their graciousness and time invested in this very non-traditional student. I also thank Department Chair Don Estep, Kristin Stephens, the CSU Statistics Department, and the CSU Graduate School for allowing me the opportunity to complete this work. I would like to especially recognize Dean Mary Stromberger, who counseled me, and who found a path for me to move forward and through a difficult situation in my program.

My employer, Hach Company, has granted me flexibility in work scheduling as required to complete this degree, and I am very grateful for their support.

On a personal level, many people have encouraged and supported me along this journey, which has made it possible for me to complete the work involved. In particular, Edward Jaeck and Chris Nachtsheim, who took me to dinner in 2016 and figuratively kicked my ass, made a huge difference in my attitude and ability to move forward. The encouragement and pushiness of Edward, especially when I was not in the mood to hear it, helped me to find strength in times of discouragement. I must also recognize my dear friends Terry Traub and the late Sidney Traub, MD, who have always told me what I needed to hear.

My late parents, Harold Sleeper, MD, and Ann Sleeper instilled in me a love of learning and curiosity to explore the unexplored boundaries.

Lastly and firstly, my wife, Julie, and my children, Pascal, Minh, and Kim, have been always supportive and helpful, even as they sacrificed time with me and handled things I did not have time for, during the toughest parts of this project. I am forever grateful to my loving family.

DEDICATION

*This work is dedicated to my loving wife Julie,
who never lost faith in my ability*

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
DEDICATION	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
Chapter 1 Introduction	1
1.1 Definition of Outlier	2
1.2 Outlier Handling Practices	3
1.3 Brief History of Outlier Discordancy Tests	5
1.4 The Window of Opportunity	9
1.5 Recommendations	16
1.6 Our Analytical Approach	18
Chapter 2 One-Sample Outlier Discordancy Test	19
2.1 Setting	19
2.1.1 Null model setting	19
2.2 Saddlepoint Approximation	21
2.2.1 Reformulation of $F_M(x)$	21
2.2.2 First-Order Approximation	25
2.2.3 Second-Order Approximation	26
2.2.4 Calibration	27
2.3 Results	29
2.4 Cumulant Generating Function $\mathcal{K}(s, t)$ and its Derivatives	34
2.4.1 Limiting Forms	38
2.5 Bonferroni Bounds	39
2.5.1 Bonferroni Inequalities	39
2.5.2 Distribution Theory	41
Chapter 3 Outlier Discordancy in Univariate Regression	44
3.1 Setting	44
3.1.1 Null model setting	44
3.1.2 Optimality	46
3.2 Saddlepoint Approximation	47
3.2.1 Reformulation of $F_M(x)$	47
3.2.2 First-Order Approximation	51
3.2.3 Second-Order Approximation	52
3.2.4 Calibration	54
3.3 Results	55
3.4 Cumulant Generating Function $\mathcal{K}(s, t)$ and its Derivatives	62
3.4.1 Limiting Forms	66

3.5	Bonferroni Bounds	67
3.5.1	Bonferroni Inequalities	67
3.5.2	Distribution Theory	67
Chapter 4	Outlier Discordancy in Multivariate Regression	70
4.1	Setting	70
4.1.1	Null model setting	70
4.1.2	Optimality	72
4.2	Saddlepoint Approximation	73
4.2.1	Reformulation of $F_M(x)$	73
4.2.2	Second-Order Approximation	78
4.2.3	Calibration	80
4.3	Results	80
4.4	Cumulant Generating Function $\mathcal{K}(\mathbf{S}, \mathbf{T})$ and its Derivatives	88
4.5	Special Case: i.i.d.	104
4.6	Bonferroni Bounds	112
4.6.1	Bonferroni Inequalities	113
4.6.2	Distribution Theory	113
References		115
Appendix A	R Code for Calculations	119
A.1	Description of PMASR Function	121
A.2	Description of PMASR_Bonf Function	123

LIST OF TABLES

2.1	Saddlepoint and Bonferroni p -values for Four Examples	29
3.1	Saddlepoint and Bonferroni p -values for Four Regression Examples	56
4.1	Saddlepoint and Bonferroni p -values for Four Multivariate i.i.d. Examples	81
4.2	Saddlepoint and Bonferroni p -values for Four Multivariate Regression Examples	82

LIST OF FIGURES

1.1	Alpha from Outlier Test Based on Rule of 3	6
1.2	Discordancy Test Critical Values with $\alpha = 0.05$	7
1.3	Joint Density of Two Residuals when $n = 6$	10
1.4	Joint Density of Two Absolute Residuals when $n = 6$	11
1.5	Joint Density of Absolute Residuals when $n = 6$ and $m = 2$	13
1.6	Example of Bivariate Data with Outlier	15
2.1	Comparison of p -values when $n = 6$	30
2.2	Comparison of p -values when $n = 18$	31
2.3	Comparison of p -values when $n = 30$	32
2.4	Comparison of p -values when $n = 100$	33
3.1	Comparison of p -values for a 10×10 Balanced Design	58
3.2	Comparison of p -values for <code>airquality</code> Dataset	59
3.3	Comparison of p -values for a Plackett-Burman 12-run Design	60
3.4	Comparison of p -values for a Definitive Screening Design	61
4.1	Comparison of p -values for i.i.d. case, $n = 30$ and $m = 2$	83
4.2	Comparison of p -values for i.i.d. case, $n = 30$ and $m = 5$	84
4.3	Comparison of p -values for i.i.d. case, $n = 100$ and $m = 2$	85
4.4	Comparison of p -values for i.i.d. case, $n = 100$ and $m = 5$	86
4.5	Comparison of p -values for DSD example, $n = 51$ and $m = 2$	87

Chapter 1

Introduction

The problem of detecting outliers in a set of data is centuries old, yet still not fully solved. In specific settings, the optimal statistic for detecting a single outlying observation is well known, but its exact distribution is not. As methods evolve, a series of ever sharper approximations are proposed, generally in the form of Bonferroni bounds. For testing the discordancy of a single outlying observation from among independent and identically distributed (i.i.d.) normally distributed variates, a new and better approximation is presented here. As will be seen, the same method works as well in a broader class of problems with left-orthogonally invariant error distributions.

Following this introductory chapter, this outlier discordancy test is presented in three settings of increasing complexity:

- Chapter 2 describes an outlier test applied to a one-sample case, where a set of univariate data is assumed to be independent and identically normally distributed, except possibly for one value. This setting is the most common application for outlier tests.
- Chapter 3 considers the univariate regression setting, where a linear model has been fitted to explain the effects of independent variables, and then the studentized residual vector is examined for outlying observations, using the maximum absolute studentized residual (MASR) as a metric.
- Chapter 4 examines the multivariate regression setting, in which each observation is a vector of m dimensions. After fitting a multivariate linear model, the residuals can be considered for the presence of outliers in m dimensions. In multivariate literature, the maximum Mahalanobis distance is often used for this purpose, and for balanced designs, this is an optimal metric, providing the best probability of detecting a single outlying value. However, for both balanced and unbalanced designs, an optimal metric is the maximum squared studentized

residual (MSSR). Chapter 4 presents approximate distributions for MSSR, comparing them to Bonferroni bounds and a Monte Carlo simulation.

- Appendix A describes the R code developed for this project. The code itself will be available from the author. In the code, one top-level function `PMASR` calculates discordancy test p -values for all settings considered, based on saddlepoint approximations. `PMASR_Bonf` provides similar calculations based on Bonferroni bounds. Full comments are included in the code.

1.1 Definition of Outlier

What is an outlier? Even this simple question has no simple answer. Here are definitions written by three prominent authors in the field:

"An outlying observation, or 'outlier,' is one that appears to deviate markedly from other members of the sample in which it occurs." - Grubbs (1969)

"The intuitive definition of an outlier would be 'an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism.'" - Hawkins (1980)

"We shall define an outlier in a set of data to be *an observation (or subset of observations) which appears to be inconsistent from the remainder of that set of data.*"
- Barnett and Lewis (1994)

Each of these definitions is about feelings and appearances, but not math. Without the specification of a mathematical model presumed to explain the data, these are the best available definitions. Once a null model is specified to describe the outlier-free case, an alternative model incorporating one or more outlying values can be evaluated probabilistically. To establish a mathematical method for an outlier discordancy test, each of the following chapters defines a more mathematically precise definition for the potentially outlying observations to be tested.

1.2 Outlier Handling Practices

For as long as data has been collected and explanatory models applied, the responsible handling of outliers has been a serious concern eliciting divergent practices and strong opinions from different analysts. In a publication laying the groundwork for maximum likelihood estimation, Daniel Bernoulli (1777) expresses ambivalence about these practices:

"I see no way of drawing a dividing line between those (observations) that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others. Nevertheless, I do not condemn in every case the principle of rejecting one or other of the observations, indeed I approve it, whenever in the course of observation an accident occurs which in itself raises an immediate scruple in the mind of the observer, before he has considered the event and compared it with the other observations. If there is no such reason for dissatisfaction I think each and every observation should be admitted whatever its quality, as long as the observer is conscious that he has taken every care.

In modern statistical practice, the divergent handling of suspected discrepant values remains unresolved. Practices regarding outliers may be broadly assigned to these three categories:

- Discordancy testing, in which the dataset is subjected to a procedure to identify whether one or more observations can be proven to be outlying, beyond a reasonable doubt, as quantified by a known and specified risk level α .
- Accommodation, in which a modified descriptive or modeling procedure attaches greater weight to more central observations than to more outlying ones, thus reducing the bias introduced by outliers, should they be present.
- Ignorance, in which the existence, the causes, and the potential impact of outliers are not considered.

Understandably, other scholarly works on outliers omit the third practice, ignorance, from discussion, but it is real and widespread. With computers, it is far easier to apply statistical methods than to understand the theory and assumptions behind them. Because of the proliferation of Six Sigma and related business initiatives, legions of business leaders have become accustomed to making data-based decisions from rote summaries comprising mean, standard deviation, and other statistics derived only from those two. Such summaries rarely account for the bias induced by outliers, should they exist. Unless individual observations are evaluated thoughtfully, the impact of contaminating outliers is omitted from the decision process.

Accommodation takes many forms. From Chauvenet's method of 1863, discussed below, to today's sophisticated robust modeling tools, accommodation procedures algorithmically handle outliers by diminishing their influence on the conclusions reached from the analysis. The application of an outlier accommodation method requires no thought about the potentially denigrated outliers or the causes behind them, other than to select a particular procedure to apply. Since each accommodation procedure may perform differently than other procedures on similar data, this selection process is inevitably biased by the practitioner's experience, and likely an *a priori* examination of the data. As a result, the conclusions reached may depend as much on who analyzes the data as on the process that produced the data.

Only discordancy testing, based on an optimal test statistic, provides a repeatable assessment of the existence of outliers, free from the analyst's biases. Once an outlier is identified, all options are available. One may explore the potential causes behind the discrepant observations, to the extent they can be known. After investigation, one can rationally decide whether to accept the outlier into the data, reject it, remeasure it, or attempt to replicate the trial, depending on the situation. After handling the outlier in a scientific and transparent manner, the model applied to the data may be presented with greater confidence.

The remainder of this work will only consider discordancy tests based on an optimal test statistic and applied to a dataset presumed to follow specified types of models.

1.3 Brief History of Outlier Discordancy Tests

In an appendix to an astronomy text published in 1863, Chauvenet advises this method of handling outliers: "Any result of a series containing n ... observations shall be rejected when the magnitude of its deviation from the mean of all measurements is such that the probability of occurrence of all deviations as large or larger is less than $1/2n$." In effect, Chauvenet's method rejects one-half an observation, on average, from every dataset. Barnett and Lewis (1994) aptly describe the statistical weaknesses of this algorithm, especially its high likelihood of rejecting an otherwise reasonable observation, while also observing that Chauvenet's method lives on through citations in modern textbooks. Chauvenet's method is notable as one of the first proposals to adopt a probabilistic basis for identifying an outlier, flawed as it may be. Because of the recommendation to reject selected observations, this may be regarded as both a detection and an accommodation algorithm.

In 1884, Wright recommended rejecting any observation deviating from the mean by more than 3 times the standard deviation, while making no allowance for sample size. As will be seen, sample size and model complexity have an enormous effect on the distribution of the maximum absolute residual. The *ad hoc* value of 3, like any other single value, cannot perform well in all situations, and it does not maintain a reasonable risk level α . Therefore, it ought not to be recommended. This conclusion is clear to anyone who has studied the distribution of residuals, but even today, such methods persist. Montgomery, in his widely used textbook on the design and analysis of experiments (2005, p. 396), defines the i^{th} standardized residual as $d_i = e_i/\hat{\sigma}$, where e_i is the raw residual and $\hat{\sigma}$ is the estimated residual standard deviation. Montgomery continues with this advice: "Most of the standardized residuals should lie in the interval $-3 \leq d_i \leq 3$, and any observation with a standardized residual outside of this interval is potentially unusual with respect to its observed response." Montgomery is not alone in offering similar advice, based on the maximum absolute residual value of 3. A web search for "outlier residual three," or similar keywords, finds many authors and instructors training practitioners in this way. The same value, 3, is even applied to different varieties of residuals with clearly different distributions. Figure 1.1

illustrates the impact of applying this rule of 3 to standardized residuals as Montgomery advises. The graph plots the risk of a Type 1 error α , that is, identifying a non-discordant observation as outlying, as a function of sample size. This figure uses the best methods from Chapter 2, with a conversion from studentized to standardized residuals.

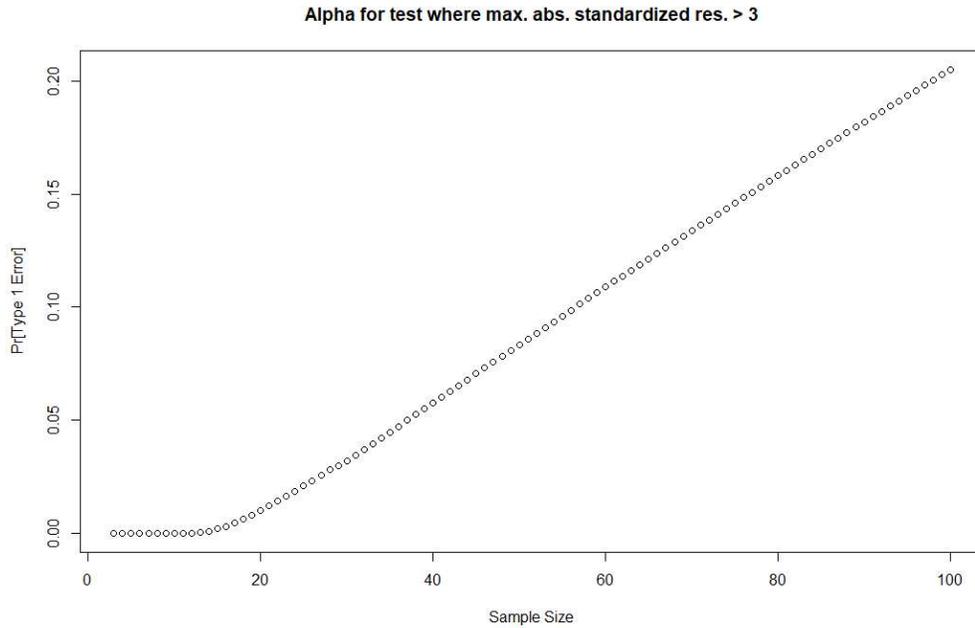


Figure 1.1: Alpha from Outlier Test Based on the Maximum Absolute Standardized Residual Exceeding 3

To control type 1 error to $\alpha = 0.05$ or any other value, one must consider both sample size and the model applied to the data. For the mean-only model setting of Chapter 2, Figure 1.2 plots critical values for an outlier discordancy test based on maximum absolute residuals, of both standardized and studentized varieties. In Figure 1.2, the type 1 error risk is controlled to $\alpha = 0.05$. A program `CritValue`, written for this work, calculates this value using the best available methods.

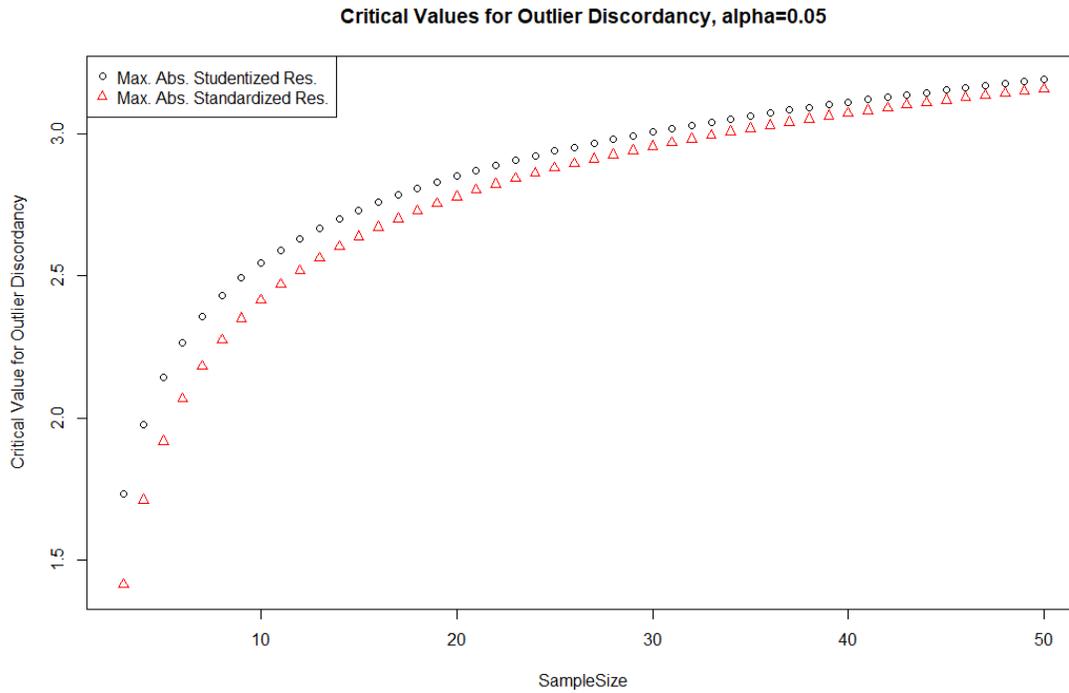


Figure 1.2: Critical Values for a Discordancy Test of a Single Outlier, with a Fixed Alpha Risk of 0.05

In the time of Wright, there was no effort to distinguish between the sample and population standard deviation. If the only information available was derived from the sample at hand, its standard deviation was assumed to be correct. Following the work of William Gosset ("Student," 1908) and others, the distributions of sample statistics from Gaussian (normal) data gained acceptance. After inference procedures for means and variances were well established, attention eventually turned to outlier identification. William Thompson (1935) published a distribution for what would become known as studentized residuals. E. S. Pearson and Chandra Sekar (1936) applied Thompson's result to the rejection of outlying observations by observing the maximum absolute deviation from the mean and rejecting the null hypothesis when large. Pearson and Chandra Sekar include a graph of critical values quite similar to Figure 1.2.

Notably for this work, Pearson and Chandra Sekar identify a value within the support of the maximum absolute residual beyond which it is not possible to observe more than one absolute residual. We label this value M_2 . When the observed maximum absolute residual exceeds this value, a simple and exact formula for the distribution of the maximum absolute residual is available

from the first Bonferroni inequality. Grubbs (1950) developed this prior work into tests for outliers in normally distributed samples based on the maximum absolute studentized residual. This method has become widely known as Grubbs' test, and this test is available in the "Outlier Test" menu item in Minitab 18 statistical software (2017). Grubbs' test provides an exact p value for an outlier test only when the observed value exceeds M_2 , and an upper bound on the p value otherwise. Srikantan (1961) extended this work to selected regression settings. Stefansky (1969, 1971, and 1972) applied this method to balanced experimental designs and outlined a procedure by which higher-order Bonferroni bounds could be calculated. Butler (1984) applied this concept to the selection of a best-fitting regressor variable in the general linear regression setting, for which the outlier identification problem turns out to be a special case. Hunter (1978) and Worsley (1982) proposed an improvement to the first Bonferroni bound by using spanning trees. Applied to outlier testing, the Hunter / Worsley procedure sharpens the upper bound to the p value for residuals in all applications where Grubbs' test does not provide an exact value.

The popular box plot, or box-and-whisker plot, of John Tukey (1977) often includes an outlier test by which observations deviating beyond the first or third quartile by more than 1.5 times the interquartile range are shown with a distinctive symbol. As an outlier test, this algorithm suffers from the same lack of risk control as the rule of 3 discussed above. For normal data without outliers, the probability of incorrectly identifying one or more observations as outlying varies from 0 to almost 1, depending on sample size.

Multivariate literature on the detection of outliers is both more sparse and more varied in the approach to outliers. Wilks (1963) proposed an outlier test for a single outlier in multivariate normal data. In this paper, Wilks showed that when the dimensionality $m \geq 2$, it is possible to have multiple observations with distances from the mean up to the maximum possible value. This concept will be graphically illustrated below. Srivastava and Van Rosen (1998) published a distribution for one residual and a joint distribution of two such residuals from a multivariate regression setting. But we have found no publication of an exact distribution for the most outlying

multivariate residual. Therefore, the available choices for an optimal outlier test include Bonferroni bounds and the approximate methods we developed here.

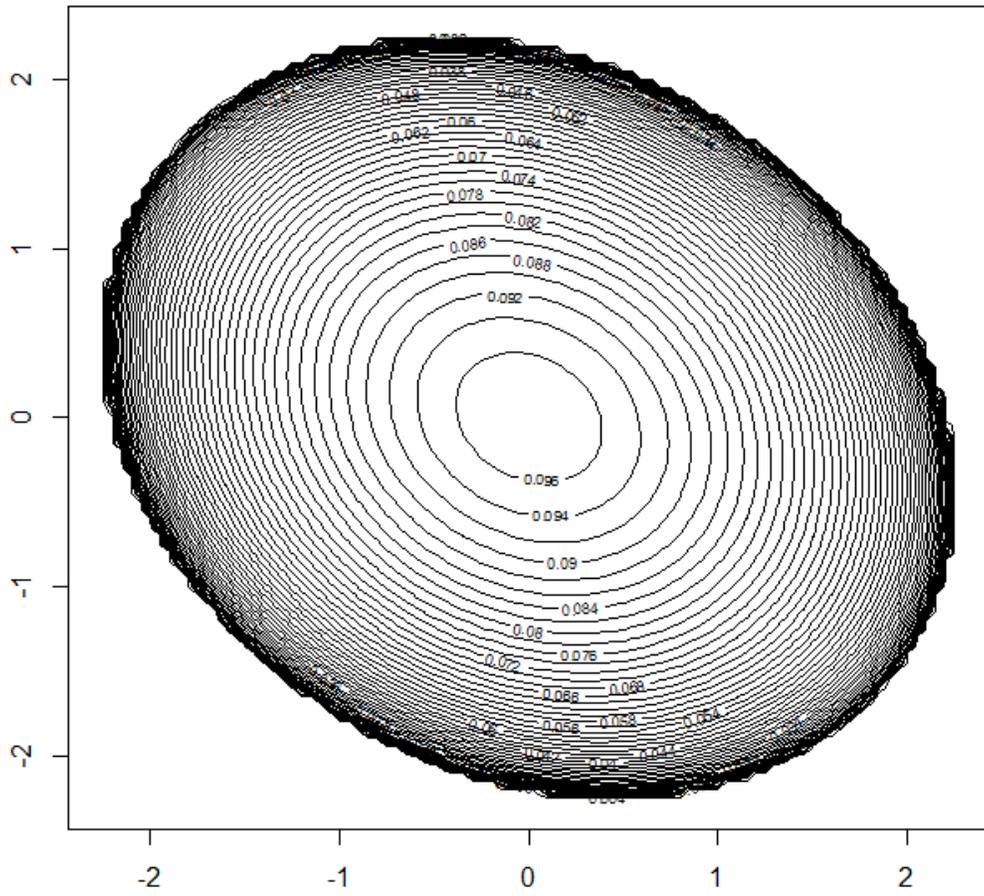
A survey of articles published in 2017 and 2018 on "multivariate outliers" found them evenly divided between discordancy testing and accommodation methods. Also, there is no consensus in the literature on which metric of deviation from the centroid ought to be employed. While many authors use versions of the Mahalanobis distance, many other metrics, with and without modifications for "robustness" are being employed, and new variations continue to be published.

Although Chapters 2 and 3 discuss an absolute residual metric, we have chosen to follow the lead of most authors in the multivariate literature by using a squared residual metric for the multivariate case. In Chapter 4, the optimality of the maximum squared studentized residual (MSSR) as a test statistic for a single outlying value in multivariate regression is discussed. Any other metric monotonically related to MSSR is also optimal. Mahalanobis distance, as generally defined, does not include an adjustment for the variance of individual fitted values, calculated from the design matrix X . Therefore, in balanced designs when all fitted values have equal variance, Mahalanobis distance is equivalent to MSSR. In more general regression cases, it is not equivalent.

1.4 The Window of Opportunity

With an easily calculated Bonferroni bound offering an exact p value in a portion of the right tail of the MASR and MSSR statistics, is there any need for an improved method? This section illustrates why a better method is needed.

Srivastava and Van Rosen (1998) provides a joint density function for two studentized residuals in a multivariate regression setting. A simplified version of this density was used to produce the contour plot in Figure 1.3. In this example, the sample size $n = 6$, there is only $p = 1$ parameter, the mean, and, since $m = 1$, the observations are univariate. The vector of studentized residuals, \mathbf{a} , is defined in Chapter 2. The upper end of the support for each a_i is $\sqrt{5}$, which is the outer boundary of the probability contours in this figure.



Since the optimal outlier test is based on the maximum absolute studentized residual (MASR), Figure 1.4 shows the joint density function of $|a_i|$ and $|a_j|$, for the same example. The red lines on Figure 1.4 are located at M_2 , which is $\sqrt{3}$ for this example. It is possible for one residual to have a value in the range $|a_i| > \sqrt{3}$, but the probability is 0 that $|a_i| > \sqrt{3}$ and $|a_j| > \sqrt{3}$ for $i \neq j$.

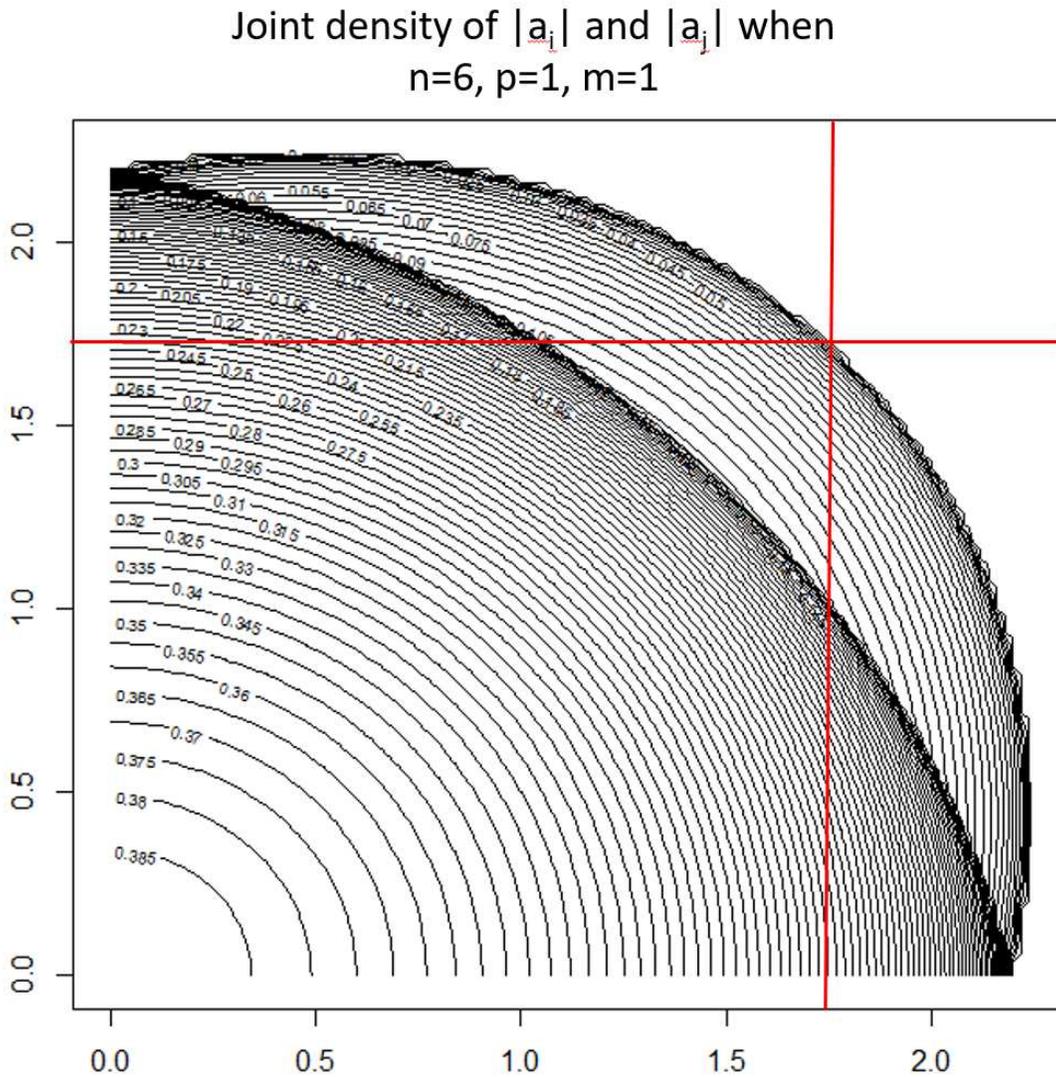


Figure 1.4: Joint density function of $|a_i|$ and $|a_j|$ when $n = 6$, $p = 1$, and $m = 1$, with lines representing M_2

The first Bonferroni inequality provides an upper bound on the distribution of $\max_i |a_i|$, based on the marginal density of a single residual, multiplied by the sample size. When any $|a_i|$ exceeds

M_2 , at most one value can be in that range, so this lower bound is the exact distribution of $\max_i |a_i|$. When $\max_i |a_i| < M_2$, the first Bonferroni bound diverges rapidly from the true distribution.

The second Bonferroni inequality is more difficult to calculate, but it provides an exact distribution of $\max_i |a_i|$ down to $M_3 < M_2$, where the M_3 is the smallest value such that there is a zero probability that 3 values $|a_i|$ could exceed M_3 . The formulas for this calculation are provided in Chapters 2 and 3, and a program `PMASR_Bonf` written for this project and described in Appendix A calculates these bounds, as well as the values of M_2 and M_3 for any setting considered here.

Now consider what happens in the multivariate setting. Chapter 4 discusses the distribution of $\max_i a_i^2$, the maximum squared residual, but for easier comparison, Figure 1.5 plots the joint density of a_i and a_j in the bivariate case where sample size $n = 6$, there is only $p = 1$ parameter, the mean, and the observations have $m = 2$ dimensions. Notice that the general shape of the density in Figure 1.5 is similar to Figure 1.4, except in the upper right corner. There is a positive and significant probability that more than one residual a_i could exceed the value $\sqrt{3}$. Because of this fact, there is no region of values for which the first Bonferroni inequality provides an exact distribution for $\max_i a_i^2$.

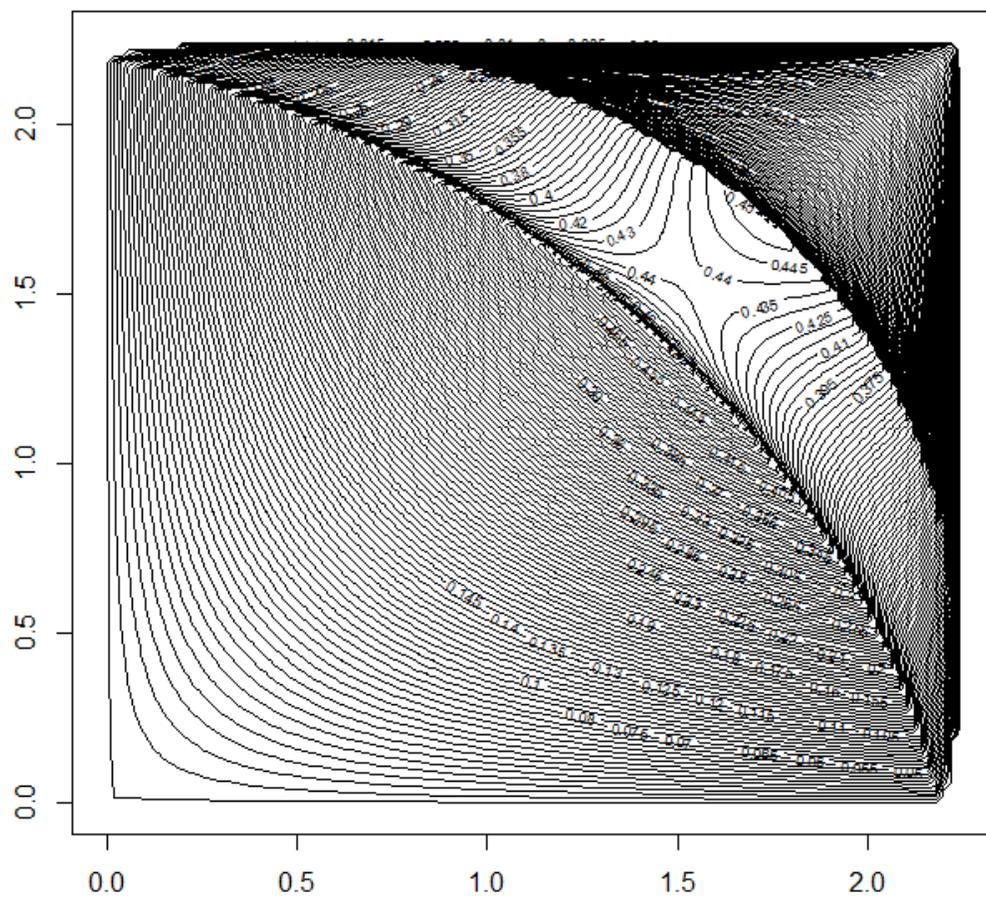


Figure 1.5: Joint density function of a_1 and a_2 when $n = 6$, $p = 1$, and $m = 2$

Outside of academic papers, we have not seen any second or higher-order Bonferroni bound applied to outlier discordancy testing. Each higher bound is significantly harder to calculate, so this is understandable. In this work, we present approximate distributions for $\max_i |a_i|$ based on saddlepoint methods. These new approximations require one univariate root finding operation followed by a set of formulas that calculate quickly in all but the most complicated situations. These formulas have been coded in the `PMASR` function described in Appendix A.

We offer one more example to show the importance of the multivariate approach. Figure 1.6 is a scatter plot of randomly generated bivariate data, with $n = 30$ observations. The positive correlation is obvious. But where is the outlier? The observation represented by the red diamond appears to be outlying, but it is not. At least two models are reasonable here. If a univariate regression model were fit to this data, the observation represented in red would have high leverage on the model, but it is very close to a regression line fit through the data, so it is not an outlier. Using a bivariate model where only the mean and the correlation are estimated, the red observation is not significantly far from the bivariate mean, considering the variation along that vector. However, in either model, the observation represented by the purple diamond has an unusually large studentized residual value. Applying the methods from Chapter 4 to this bivariate dataset, the MSSR of the purple observation is 11.413, with a p -value of 0.033 generated by saddlepoint methods, so it is a significantly outlying value. But if any univariate outlier test were applied to either variable individually, this outlier would be undetectable.

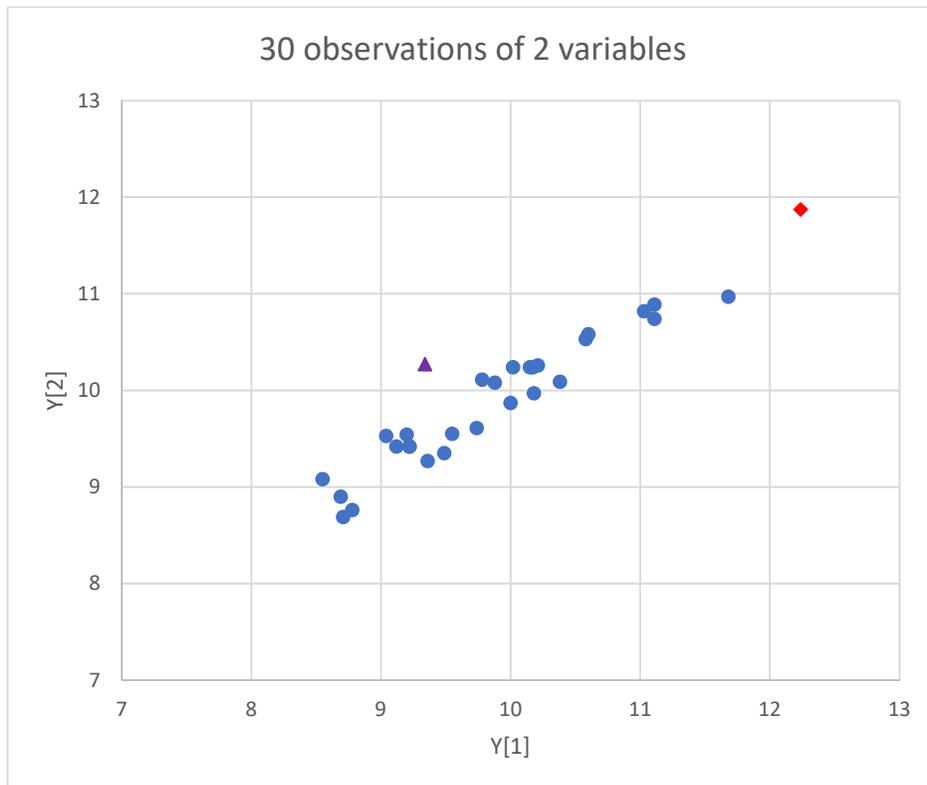


Figure 1.6: Example Scatter Plot of Bivariate Observations with One Outlier

1.5 Recommendations

To summarize this research, here are our recommendations to practitioners on the best way to perform tests of discordancy of a single possibly outlying observation.

- We discourage the use of robust methods or other techniques designed to accommodate possible outlying values. An outlier has a story to tell. After applying an accommodation method, that story will never be heard.
- Use a discordancy test based on the MASR or MSSR, as this is the optimal test for a single outlying observation, minimizing the probability of making a testing error.
- Always calculate the first Bonferroni bound, a.k.a. Grubbs' test, because this provides an upper bound on the p -value. If this upper bound is less than the chosen error risk α , the outlier is significantly discordant, and no further calculation is required.
- For univariate problems, when only one y variable is observed:
 - Calculate M_2 . If the observed value of MASR, $x \geq M_2$, then the first Bonferroni bound (Grubbs' test) provides the exact p -value. When $x < M_2$, this calculation provides a bound that may not settle the question of whether the suspected outlying value is significantly discordant.
 - When $x < M_2$, calculate the calibrated second-order exponential saddlepoint approximation for the p -value, represented as $1 - \bar{F}_{2e}(x)$ in Chapters 2 and 3. Use this value to make the best decision about outlier discordancy.
 - The second Bonferroni bound and the improved first Bonferroni bound often provide tight upper and lower bounds on the p -value, but these require numerical integration which may consume more time than the saddlepoint method. This is especially true when regressing empirical data, in which each pair of observations may have different correlation, and therefore require a different numerical integration.
- For multivariate problems, when $m > 1$ variables are observed:

- Always analyze the dataset as a multivariate dataset, looking for an outlier in m -dimensional space.
- The calibrated second-order exponential saddlepoint approximation for the p -value, represented as $1 - \bar{F}_{2e}(x)$ in Chapter 4, is generally the best saddlepoint alternative, especially for larger sample size n . If calculation takes too long, the calibrated first-order saddlepoint approximation may be used instead.
- The second Bonferroni bound and the improved first Bonferroni bound provide lower and upper bounds on the p -value. In multivariate applications, these bounds require a numerical triple integral for each pair of observations with a different correlation. In all of the multivariate examples we have examined, this Bonferroni calculation consumes significantly more computing time than any of the saddlepoint methods.

1.6 Our Analytical Approach

We conclude this chapter by briefly discussing the novel approach to the analysis of this problem, proposed in this dissertation. Some portions of this approach were previously applied to the approximation of other less challenging distributions in Butler and Sutton (1998). These distributions include gamma, Poisson, binomial, and negative binomial. Full details are presented in the subsequent chapters, but here are the major steps.

- The statistic of interest, $M = \max_i |a_i|$ is an ancillary statistic, which allows the parameters of the underlying noise distribution to be set to any convenient values.
- By Basu's theorem, ancillary statistic M is independent of any complete, sufficient statistic $\hat{\theta}$, so any probability statement about M can be conditioned on any convenient value for $\hat{\theta}$.
- Applying Bayes theorem, the probability statement about M can be reexpressed in terms of a conditional density of $\hat{\theta}$, given a probability statement about M , combined with other factors that are simple to calculate.
- The probability statement about M is reexpressed in terms of an intersection of events concerning the observed data. In this work, the event $\max_i |a_i| \leq x$ is the intersection of events $E_i = \{|a_i| \leq x\}$ for all i .
- If the statistic $\hat{\theta}$ is calculated from a distribution truncated in such a way that $\Pr[E_i] = 1$ for all i , then the conditioning event becomes certain and disappears from the probability statement.
- The unconditional density of $\hat{\theta}$ can then be calculated from truncated distributions using a saddlepoint approximation from Butler (2007).

Chapter 2

One-Sample Outlier Discordancy Test

This chapter examines the most common application for an outlier test, a univariate sample with no model applied, other than the mean. This application can be referred to as the independent and identically distributed (i.i.d.) case, the mean-only case, or the one-sample case. The discordancy test based on saddlepoint approximation requires finding the root of one equation, then using that root to calculate the approximate probability that the maximum absolute studentized value (MASV) M exceeds the observed value x . In this chapter, the saddlepoint approximation is compared to Grubbs' test, which is also the first Bonferroni bound, two improved Bonferroni bounds, and a Monte Carlo simulation.

2.1 Setting

In this section, we define the setting for the outlier discordancy test and discuss the test as an optimal test, from among a large class of potential tests.

2.1.1 Null model setting

The null model assumes that there are no outliers in the data. The null model is simply

$$\mathbf{y} = \varepsilon, \tag{2.1}$$

in which $\mathbf{y} = (y_1, \dots, y_n)^T$ is an $n \times 1$ vector of observed values. The random vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is typically assumed to be i.i.d. normally distributed, but as detailed in Chapters 3 and 4, these methods work equally well over a larger class of distributions referred to as left-orthogonally invariant (l.o.i.).

We define y_j to be an outlier when this alternative model holds:

$$\mathbf{y} = \theta \mathbf{e}_j + \varepsilon, \quad (2.2)$$

when $\theta \in \mathfrak{R}$, $\theta \neq 0$, and \mathbf{e}_j is the j^{th} column of \mathbf{I}_n , for $j = 1, \dots, n$.

When $\varepsilon \stackrel{iid}{\sim} \mathbb{N}(\mu, \sigma^2)$, the usual unbiased parameter estimates are $\hat{\mu} = \bar{y} = n^{-1} \sum_{j=1}^n y_j$ and $\hat{\sigma}^2 = (n-1)^{-1} \sum_{j=1}^n (y_j - \bar{y})^2$. The vector of studentized values is $\mathbf{a} = (a_1, \dots, a_n)^T$, where

$$a_j = \frac{y_j - \bar{y}}{\hat{\sigma}} \sqrt{\frac{n}{n-1}} \quad j = 1, \dots, n. \quad (2.3)$$

The more customary "standardized" values are $(y_j - \bar{y}) / \hat{\sigma}$, but this chapter retains the $\sqrt{n/(n-1)}$ factor so that a_j has a consistent definition throughout this dissertation.

The maximum absolute studentized value (MASV) is the statistic $M = \max_j |a_j|$ whose cumulative distribution function (CDF) under the null model is denoted as $F_M(x) = \Pr(M \leq x)$. A typical test for the discordancy of a single outlier caused by a change in distribution for one component of ε rejects the null model when M is large. Thus, if x is the observed value of M , then the p -value is $1 - F_M(x)$. Since the exact distribution F_M is intractable, the purpose of this chapter is to derive an approximation for the distribution of M , using saddlepoint methods.

Alternatively, upper and lower Bonferroni bounds are available. In limited ranges at the upper end of the support of M , these bounds can be exact. The upper Bonferroni bound on the p -value has become known as Grubbs' test. The details of these bounds, and the ranges within which they are exact, are presented in Section 2.5. This chapter will show where the Bonferroni bounds are adequate or inadequate and show the opportunity for an improved method to be provided by the saddlepoint approximations.

In Chapter 3, based on Butler (1986), we discuss that the outlier discordancy test defined here as the optimal test among a large class of tests, in that no other test can have a higher probability of making a correct decision about whether $\theta \neq 0$ in (2.2), and about which observation j is discordant.

2.2 Saddlepoint Approximation

An approximation for $F_M(x)$ is derived by first reexpressing this CDF as a conditional probability to which Bayes' theorem can be applied. A reversal of the probabilities in Bayes' theorem allows approximation to the CDF to be reexpressed in terms of a saddlepoint density approximation. Using first- and various second-order saddlepoint approximations for this multivariate density leads to various CDF approximations $\hat{F}_M(x)$ with second-order approximations providing greater accuracy.

Statistic M has bounded support, and owing to the nature of saddlepoint approximations, this can lead to a loss in accuracy for the various $\hat{F}_M(x)$. Since the least upper bound on the support of M is $\sqrt{n-1}$, these approximations can be calibrated by using a known exact value at the upper end of the support of M , as described in Section 2.2.4. For small sample sizes, a correction may also be needed at the lower end of the support. This calibration procedure is shown to improve accuracy.

2.2.1 Reformulation of $F_M(x)$

The studentized vector \mathbf{a} has a joint distribution which does not depend on the parameters (μ, σ^2) , so that \mathbf{a} is ancillary. Thus, there is no loss in generality in assuming $\mu = 0$ and $\sigma^2 = 1$. Since estimates $(\hat{\mu}, \hat{\sigma}^2)$ form a complete sufficient statistic for (μ, σ^2) , $(\hat{\mu}, \hat{\sigma}^2)$ and \mathbf{a} are independent by Basu's theorem. Thus, the distribution of \mathbf{a} may be conditioned on any convenient value for $(\hat{\mu}, \hat{\sigma}^2)$, say $(0, 1)$. The net effect of such conditioning leads to

$$\begin{aligned} F_M(x) &= \Pr \left(\max_j |a_j| \leq x \mid \hat{\mu} = 0, \hat{\sigma}^2 = 1 \right) \\ &= \Pr \left(\max_j |y_j| \sqrt{\frac{n}{n-1}} \leq x \mid \hat{\mu} = 0, \hat{\sigma}^2 = 1 \right) \end{aligned} \quad (2.4)$$

By defining the events $E_j = \left\{ |y_j| \leq x \sqrt{(n-1)/n} \right\}$, so that $\{M \leq x\} = \bigcap_{j=1}^n E_j$, then Bayes' theorem for densities is applied to the conditional probability in (2.4) to give

$$F_M(x) = \frac{f_{\hat{\mu}, \hat{\sigma}^2} \left(0, 1 \mid \bigcap_{j=1}^n E_j \right) \Pr \left(\bigcap_{j=1}^n E_j \right)}{f_{\hat{\mu}, \hat{\sigma}^2} (0, 1)}, \quad (2.5)$$

where $f_{\hat{\mu}, \hat{\sigma}^2} (0, 1)$ denotes the marginal density of $\hat{\mu}, \hat{\sigma}^2$ at $(0, 1)$, and $f_{\hat{\mu}, \hat{\sigma}^2} (0, 1 \mid \bigcap_{j=1}^n E_j)$ denotes the conditional density of $\hat{\mu}, \hat{\sigma}^2$ at $(0, 1)$ given event $\bigcap_{j=1}^n E_j$. The denominator density is simple and explicit. With $\mu = 0$ and $\sigma^2 = 1$, then $\hat{\mu} \sim \mathbb{N}\{0, 1/n\}$ and $\hat{\sigma}^2 (n-1) \sim \chi_{n-1}^2$ are independent, so that

$$\begin{aligned} f_{\hat{\mu}, \hat{\sigma}^2} (0, 1) &= f_{\hat{\mu}} (0) f_{\hat{\sigma}^2} (1) = \sqrt{\frac{n}{2\pi}} \times (n-1) f_{\chi_{n-1}^2} (n-1) \\ &= \sqrt{\frac{n}{2\pi}} \times (n-1) \frac{(n-1)^{(n-1)/2-1} \exp\{- (n-1)/2\}}{2^{(n-1)/2} \Gamma\{(n-1)/2\}} \\ &= \frac{\sqrt{n}}{2\pi} \sqrt{\frac{n-1}{2}} \frac{\hat{\Gamma}\{(n-1)/2\}}{\Gamma\{(n-1)/2\}} =: g, \end{aligned} \quad (2.6)$$

where $\hat{\Gamma}(v) = \sqrt{2\pi} v^{v-1/2} e^{-v}$ denotes Stirling's approximation for $\Gamma(v)$.

Since $\{E_j\}$ are independent events and y_1, \dots, y_n are i.i.d. $\mathbb{N}(0, 1)$,

$$\Pr \left(\bigcap_{j=1}^n E_j \right) = \prod_{j=1}^n \Pr(E_j) = \prod_{j=1}^n \operatorname{erf} \left(\frac{\tau}{\sqrt{2}} \right) = \left\{ \operatorname{erf} \left(\frac{\tau}{\sqrt{2}} \right) \right\}^n, \quad (2.7)$$

where $\operatorname{erf}(x) := (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$, and $\tau := x \sqrt{(n-1)/n}$.

Only the conditional density in the numerator of (2.5) offers any difficulty. Note that $\hat{\mu}$ is linear in $\{y_j\}$, and

$$(n-1) \hat{\sigma}^2 = \sum_{j=1}^n y_j^2 - n \hat{\mu}^2,$$

is linear in $\{y_j^2\}$ apart from its dependence on $\hat{\mu}$. To obtain a multivariate density in which all components are linear in $\{y_j\}$ and $\{y_j^2\}$, write the conditional density of $(\hat{\mu}, \hat{\sigma}^2)$ given $\bigcap_{j=1}^n E_j$ in terms of the conditional density of $(n\hat{\mu}, v) = (n\hat{\mu}, \mathbf{y}^T \mathbf{y})$ given $\bigcap_{j=1}^n E_j$ through a one-to-one transformation. The Jacobian of this transformation is $n(n-1)$ and

$$f_{\hat{\mu}, \hat{\sigma}^2} \left(0, 1 \mid \bigcap_{j=1}^n E_j \right) = n(n-1) f_{n\hat{\mu}, v} \left(0, n-1 \mid \bigcap_{j=1}^n E_j \right).$$

Next, note that the net effect of conditioning on the event $\bigcap_{j=1}^n E_j$ is to truncate the normal random variables so that y_j is restricted to the range $|y_j| < \tau$ for $j = 1, \dots, n$. Conditioning on $\bigcap_{j=1}^n E_j$ affects the distribution theory in such a way that y_j is independent of $\bigcap_{i \neq j} E_i$ and is only affected by the conditioning on E_j . Thus $\{y_j | E_j\}$ are independent with $y_j | E_j$ having a standard normal distribution truncated to $\{|y_j| < \tau_j\}$. Let z_j denote such a truncated random variable with density

$$f_{z_j}(x) = \frac{\phi(x)}{\text{erf}(\tau/\sqrt{2})} \quad |x| < \tau,$$

where $\phi(\cdot)$ is the standard normal density function, and let $(n \times 1)$ random vector \mathbf{z} have independent elements z_j . In terms of \mathbf{z} ,

$$n\hat{\mu} \left| \bigcap_{j=1}^n E_j \stackrel{D}{=} \sum_{j=1}^n z_j = n\bar{z} =: b, \text{ and} \quad (2.8)$$

$$v \left| \bigcap_{j=1}^n E_j = \mathbf{y}^T \mathbf{y} \left| \bigcap_{j=1}^n E_j \stackrel{D}{=} \mathbf{z}^T \mathbf{z} =: u. \quad (2.9)$$

The conditional distributions on the left-hand side of (2.8) and (2.9) are the same as the unconditional distributions on the right-hand side using truncated random variables, thus effectively removing the conditioning required in the distribution theory. From (2.5) we may write

$$F_M(x) = f_{b,u}(0, n-1) \frac{n(n-1)}{g} \left\{ \text{erf} \left(\frac{\tau}{\sqrt{2}} \right) \right\}^n. \quad (2.10)$$

In (2.10), the density $f_{b,u}(0, n-1)$ cannot be readily computed but a saddlepoint approximation can be constructed for it and used in its place to calculate an approximation for $F_M(x)$, as detailed in the next section. This will be called a saddlepoint approximation for $F_M(x)$.

Saddlepoint Density Function

A saddlepoint approximation for the density function of 2-dimensional (b, u) , as defined in (2.8) and (2.9), is provided in Butler (2007, eqn. 3.2). Let 2-dimensional (s, t) index the moment generating function (MGF) of (b, u) denoted as

$$\mathcal{M}(s, t) = \mathbb{E} \left(e^{sb+tu} \right).$$

The corresponding cumulant generating function (CGF) is $\mathcal{K}(s, t) = \ln \mathcal{M}(s, t)$. Since the density $f_{b,u}$ is to be evaluated at $(0, n - 1)$, a joint first-order saddlepoint density for $f_{b,u}(0, n - 1)$ is

$$\hat{f}_1(0, n - 1) := \frac{\exp \left\{ \mathcal{K}(\hat{s}, \hat{t}) - \hat{t}(n - 1) \right\}}{2\pi \left| \mathcal{K}''(\hat{s}, \hat{t}) \right|^{\frac{1}{2}}}, \quad (2.11)$$

where saddlepoint (\hat{s}, \hat{t}) simultaneously solves the equations

$$\mathcal{K}'_s(\hat{s}, \hat{t}) = \left. \frac{\partial \mathcal{K}(s, t)}{\partial s} \right|_{s=\hat{s}, t=\hat{t}} = 0 \quad (2.12)$$

$$\mathcal{K}'_t(\hat{s}, \hat{t}) = \left. \frac{\partial \mathcal{K}(s, t)}{\partial t} \right|_{s=\hat{s}, t=\hat{t}} = n - 1 \quad (2.13)$$

and matrix $\mathcal{K}''(\hat{s}, \hat{t})$ is the 2×2 Hessian matrix of $\mathcal{K}(s, t)$ evaluated at (\hat{s}, \hat{t}) .

Much of the analysis needed to complete the details of this expression is provided in a later section. Section 2.4 gives formulas for $\mathcal{K}(\hat{s}, \hat{t})$ and its derivatives and shows that $\hat{s} = 0$. The remaining saddlepoint component \hat{t} can be found numerically as a 1-dimensional root of (2.13), $\mathcal{K}'_t(0, \hat{t}) = n - 1$, in two steps. Since $\mathcal{K}(0, \hat{t})$ and its derivatives have a removable discontinuity at $\hat{t} = 1/2$, corresponding to the MASV value $x = \sqrt{3}$, the first step in the solution process is to determine on which side of the discontinuity the solution lies, according to x . Here is the decision rule:

$$\hat{t} \begin{matrix} \leq \\ > \end{matrix} \frac{1}{2} \quad \iff \quad x \begin{matrix} \geq \\ < \end{matrix} \sqrt{3}. \quad (2.14)$$

Let $q := 1 - 2t$ and $\hat{q} := 1 - 2\hat{t}$. Then saddlepoint \hat{t} is the solution to

$$\frac{n}{\hat{q}} \{1 - R(\hat{t})\} = n - 1,$$

where

$$R(\hat{t}) = \frac{\sqrt{\frac{2\hat{q}}{\pi}} \tau \exp\left(\frac{-\hat{q}\tau^2}{2}\right)}{\operatorname{erf}\left(\tau\sqrt{\frac{\hat{q}}{2}}\right)}.$$

We now present different versions of the saddlepoint approximation: first-order, second-order, and second-order exponential denoted respectively as \hat{f}_1 , \hat{f}_2 , and \hat{f}_{2e} . Upon substitution into (2.10) to approximate $F_M(x)$, a calibration procedure may also be applied, which is indicated by changing the circumflex to a macron. This leads to the possibility of six approximations of $F_M(x)$, denoted by \hat{F}_1 , \bar{F}_1 , \hat{F}_2 , \bar{F}_2 , \hat{F}_{2e} , and \bar{F}_{2e} .

2.2.2 First-Order Approximation

The CGF $\mathcal{K}(s, t)$ and its derivatives are listed in Section 2.4, for $\hat{t} \neq 1/2$, corresponding to the almost certain case where $x \neq \sqrt{3}$. From (2.26),

$$\exp\{\mathcal{K}(0, \hat{t})\} = \hat{q}^{-n/2} \left(\frac{\operatorname{erf}\left(\tau\sqrt{\hat{q}/2}\right)}{\operatorname{erf}\left(\tau/\sqrt{2}\right)} \right)^n. \quad (2.15)$$

When $\hat{q} < 0$, corresponding to $x < \sqrt{3}$, note that $\operatorname{erf}(\cdot)$ in the numerator has a complex argument, which is ok.

In the denominator of (2.11), $|\mathcal{K}''(0, \hat{t})|$ is the determinant of a 2×2 matrix of second derivatives. From (2.30), $\hat{\mathcal{K}}''_{st}(0, \hat{t}) = 0$. This separates the determinant into

$$\begin{aligned} |\mathcal{K}''(0, \hat{t})| &= \mathcal{K}''_{s^2}(0, \hat{t}) \mathcal{K}''_{t^2}(0, \hat{t}) \\ &= \frac{n^2}{\hat{q}^3} R_1(\hat{t}) R_2(\hat{t}), \end{aligned} \quad (2.16)$$

where

$$R_1(\hat{t}) = 1 - R(\hat{t})$$

and

$$R_2(\hat{t}) = 2 - (1 + \hat{q}\tau^2)R(\hat{t}) - R^2(\hat{t}).$$

Now all the pieces are in place to complete the approximate distribution of MASV M . Substituting the expressions from (2.11) and (2.15) into (2.10) and simplifying gives this result for the first-order approximation:

$$\hat{F}_1(x) = \frac{(n-1) \exp\{-\hat{t}(n-1)\}}{2\pi\hat{q}^{(n-3)/2}g\sqrt{R_1(\hat{t})R_2(\hat{t})}} \left\{ \operatorname{erf}\left(\tau_j\sqrt{\frac{\hat{q}}{2}}\right) \right\}^n, \quad (2.17)$$

where g is defined in (2.6).

2.2.3 Second-Order Approximation

The saddlepoint density is an asymptotic expression, of which the first-order approximation is only the leading term. Including the second term often improves the accuracy of the approximation. From Butler (2007, eqn. 3.37), the second-order saddlepoint density approximation of $f(x)$ is

$$\hat{f}_2(x) = \hat{f}_1(x) (1 + O) \quad (2.18)$$

$$\text{where } O = \{\hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24\}.$$

The κ terms are sums of higher-order cumulants, defined later in section 2.4.

In the one-sample setting of this chapter, O can be calculated by this function of polynomials in $R = R(\hat{t})$:

$$O = \frac{1}{24n} \left(-6\frac{R_2(\hat{t})}{R_1^2(\hat{t})} + 3\frac{R_4(\hat{t})}{R_2^2(\hat{t})} - 5\frac{R_3(\hat{t})}{R_2^3(\hat{t})} - 6 \right), \quad (2.19)$$

where

$$R_3(\hat{t}) := 8 - (\hat{q}^2\tau^4 + 2\hat{q}\tau^2 + 3)R - 3(1 + \hat{q}\tau^2)R^2 - 2R^3,$$

and

$$\begin{aligned} R_4(\hat{t}) := & 48 - (\hat{q}^3\tau^6 + 3\hat{q}^2\tau^4 + 9\hat{q}\tau^2 + 15)R \\ & - (7\hat{q}^2\tau^4 + 14\hat{q}\tau^2 + 15)R^2 - 12(1 + \hat{q}\tau^2)R^3 - 6R^4. \end{aligned}$$

An alternative second-order exponential approximation, based on McCullagh (1987, § 6.3) is

$$\hat{f}_{2e}(x) = \hat{f}_1(x) e^O, \quad (2.20)$$

where O is the same as in (3.17). In many cases we have evaluated, including the examples shown below, this $2e$ approximation is the best of all.

Therefore, the second-order saddlepoint approximations for $F(x)$ are

$$\hat{F}_2(x) = \hat{F}_1(x) (1 + O) \quad (2.21)$$

and

$$\hat{F}_{2e}(x) = \hat{F}_1(x) e^O \quad (2.22)$$

Section 2.4 provides more details on the derivation of \mathcal{K} and its derivatives required for this application.

2.2.4 Calibration

Calibration is a simple procedure recommended to correct for known inaccuracies with the saddlepoint approximation at the ends of the support of M . We denote the support of M as the interval $[M_L, M_U]$, where $M_L = 1$ when n is even, or $\{n/(n-1)\}^{1/2}$ when n is odd, and $M_U = (n-1)^{1/2}$. These limits can be derived from the joint distribution for residuals, as derived by Raiffa and Schlaifer (1961, p. 259).

We use only the first- and second-order saddlepoint density approximations, which do not integrate to 1, so they may be scaled too high or too low. For large values of M , where p -values for a discordancy test are small, uncalibrated saddlepoint p -values could be negative or significantly too high. Fortunately, this discrepancy is easily corrected. There are three options for where to calibrate, specifically, $M_* \in \{M_U, M_2, M_3\}$. As defined in Section 2.5, M_2 and M_3 are the points above which it is not possible to have two or three, respectively, of the absolute residuals $|a_j|$. Once the value of M_* is selected, we use the saddlepoint formula from (2.17), (2.21), or 2.22) to

calculate $\hat{F}_\bullet(x)$, $\hat{F}_\bullet(M_L)$ and $\hat{F}_\bullet(M_*)$, where the \bullet subscript represents either $1, 2$, or $2e$. With these values, the calibrated saddlepoint value is

$$\bar{F}_\bullet(x) = \frac{F_M(M_*) \{ \hat{F}_\bullet(x) - \hat{F}_\bullet(M_L) \}}{\hat{F}_\bullet(M_*) - \hat{F}_\bullet(M_L)}. \quad (2.23)$$

To calibrate at $M_* = M_U = (n - 1)^{1/2}$, set $F(M_U) = 1$.

To calibrate at $M_* = M_2$, Section 2.5 shows that

$$M_2 = \sqrt{\frac{n}{2}}. \quad (2.24)$$

Since no more than one absolute residual may ever exceed M_2 , the exact value of $F_M(M_2)$ is calculated from the tail probability of the marginal distribution of a single residual and the sample size n :

$$F_M(M_2) = 1 - 2n \{ T_{n-2}(\sqrt{n}) \},$$

where $T_\nu(\cdot)$ is the right tail probability of the Student's t distribution with ν degrees of freedom.

To calibrate at $M_* = M_3$ is a lot more work, and uses an algorithm from Butler (1984). Section 2.5 provides the calculation algorithms for both M_3 and $F_M(M_3)$, using the second Bonferroni bound.

When $n > 11$, $\hat{F}_\bullet(M_L)$ is so small as to be irrelevant, and the calibration at the low end can be ignored, by setting $\hat{F}_\bullet(M_L) = 0$.

As a compromise between accuracy and computational difficulty, we recommend calibrating at $M_* = M_2$, and when $M > M_2$, using the first Bonferroni bound, an exact solution. This choice assures a smooth transition between calibrated saddlepoint and Bonferroni methods at M_2 . This is the method used to generate the calibrated p -values in the tables below. Comparing calibrated to uncalibrated values in the tables shows the importance of calibration in providing a usable and reliable discordancy test procedure.

Table 2.1: Saddlepoint and Bonferroni p -values for Four Examples

	$n = 6$	$n = 18$	$n = 30$	$n = 100$
n	6	18	30	100
M value x	1.996	2.577	2.790	3.220
Simulated p -value	0.100	0.100	0.100	0.100
Saddlepoint approx.:				
$1 - \hat{F}_1(x)$	-0.015	0.048	0.069	0.090
$1 - \hat{F}_2(x)$	0.112	0.099	0.100	0.100
$1 - \hat{F}_{2e}(x)$	0.104	0.097	0.099	0.100
$1 - \bar{F}_1(x)$	0.097	0.104	0.099	0.099
$1 - \bar{F}_2(x)$	0.116	0.100	0.100	0.100
$1 - \bar{F}_{2e}(x)$	0.114	0.101	0.100	0.100
Bonferroni bounds:				
1 st upper (Grubbs)	0.100	0.100	0.101	0.102
1 st upper (Worsley)	0.100	0.100	0.101	0.102
2 nd lower	0.100	0.100	0.100	0.100
Limits on M:				
M_U	2.236	4.123	5.385	9.950
M_2	1.732	3.000	3.873	7.071
M_3	1.195	2.423	3.143	5.763

2.3 Results

We present four examples, which are compared using the same outlier discordancy p -value of approximately 0.1, corresponding to $F_M(x) = 0.9$. The examples are all one-sample problems with sample sizes $n = 6$, $n = 18$, $n = 30$, and $n = 100$.

Table 2.1 lists p -values for all versions of saddlepoint and Bonferroni methods discussed in this article. Also included is a simulated p -value, based on sufficient trials that the width of a 99% confidence interval is less than the least-significant digit listed.

For smaller sample sizes, the Bonferroni upper bound (Grubbs) performs very well, but it starts to diverge for higher sample sizes. The saddlepoint methods are inaccurate at smaller sample sizes, but perform very well at larger sample sizes. The 2nd Bonferroni lower bound performs well at all sample sizes tested, but this is more work to calculate than even the most complex saddlepoint method tested.

For the four examples discussed here, Figures 2.1 through 2.4 below show all approximations of $1 - F_M(x)$ for 100 values of x in the support of M . Each of these four-panel graphs is organized in a similar way. The top left panel shows all saddlepoint approximations of $1 - F_M(x)$, plus a Monte Carlo simulation. The top right panel shows all Bonferroni bounds on $1 - F_M(x)$, plus a Monte Carlo simulation. The bottom two panels show errors by subtracting the best available estimate of $1 - F_M(x)$ from each approximation in the top two panels. The best available estimate is the Bonferroni lower bound when $x \geq M_3$, and the Monte Carlo simulation otherwise. When $x < M_3$, these graphs appear jittery because of the uncertainty in the Monte Carlo values.

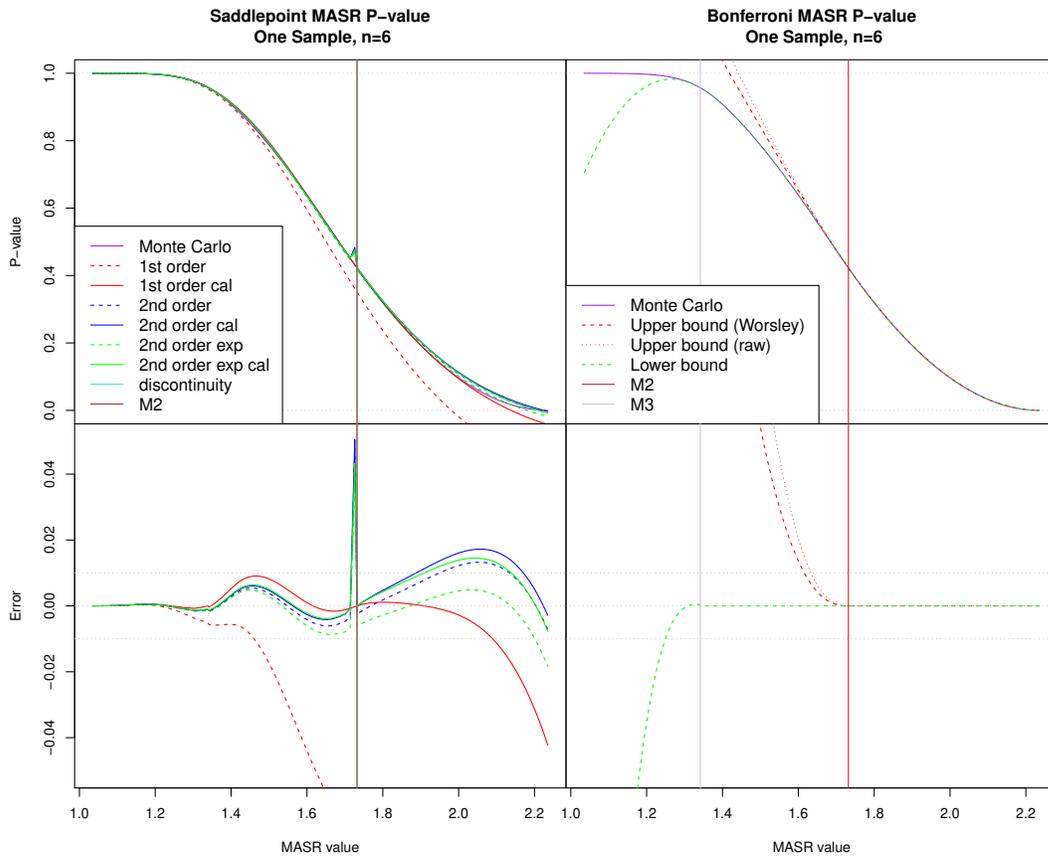


Figure 2.1: Comparison of p -values when $n = 6$

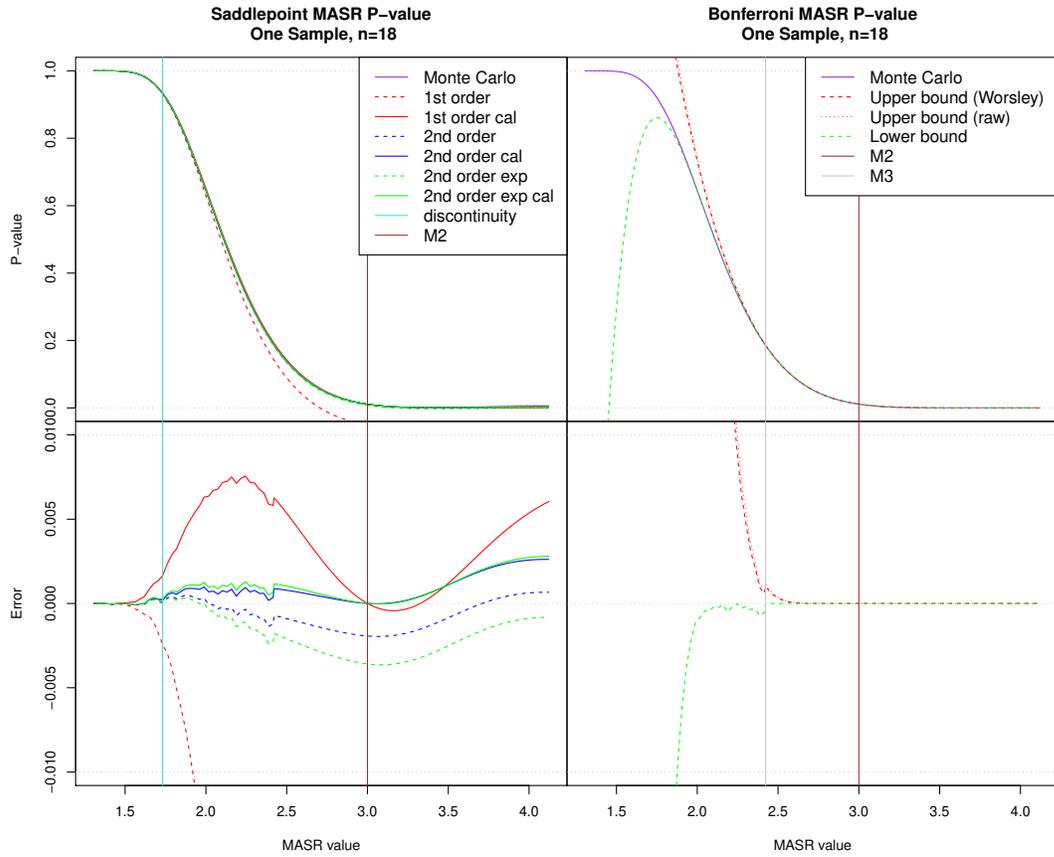


Figure 2.2: Comparison of p -values when $n = 18$

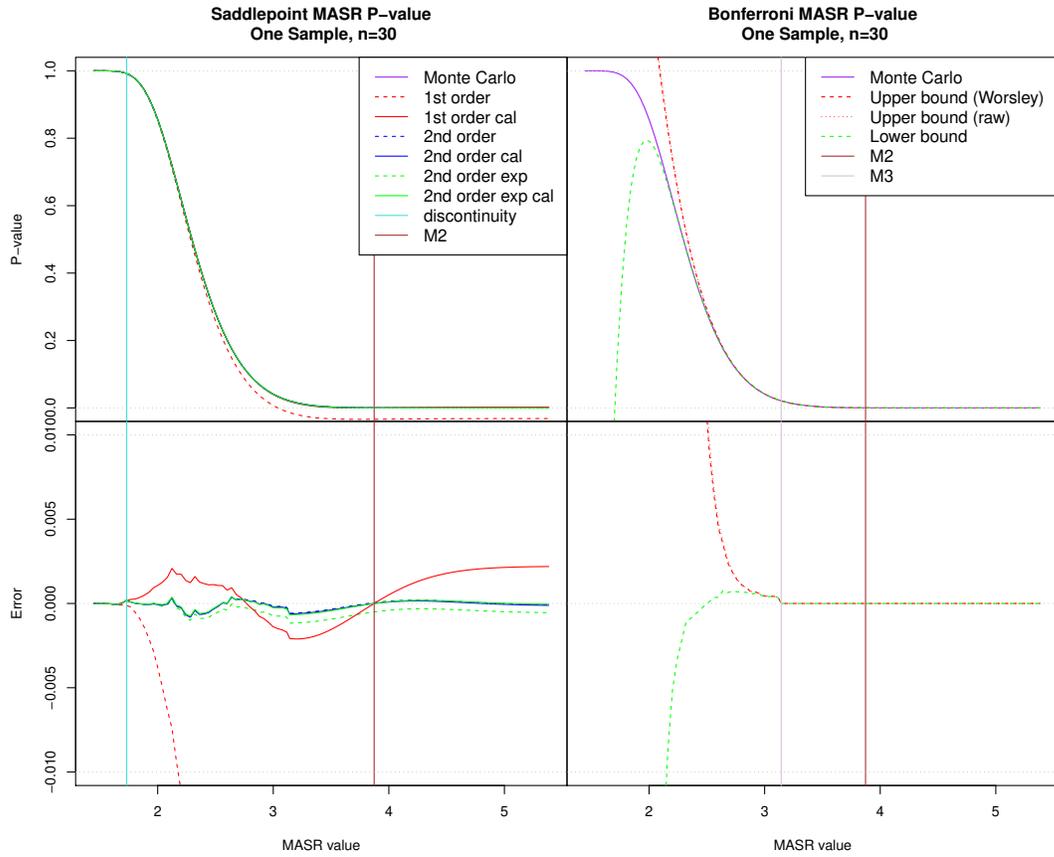


Figure 2.3: Comparison of p -values when $n = 30$

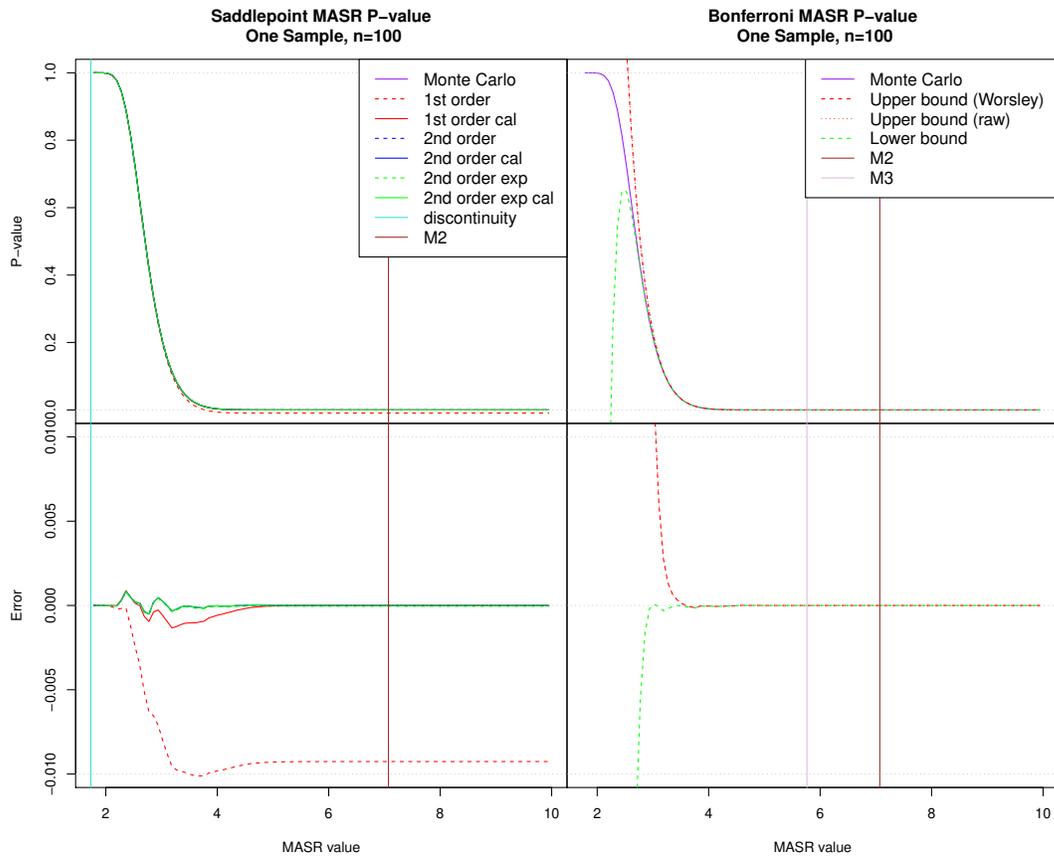


Figure 2.4: Comparison of p -values when $n = 100$

2.4 Cumulant Generating Function $\mathcal{K}(s, t)$ and its Derivatives

In this section, the joint CGF $\mathcal{K}(s, t)$ of $b = n\bar{z}$ and $u = \mathbf{z}^T \mathbf{z}$ is developed into a form for calculation at the saddlepoint (\hat{s}, \hat{t}) , along with its derivatives required for the saddlepoint approximations. With $q = 1 - 2t$, define the function

$$\begin{aligned} \mathcal{M}_1(s, t) &= \mathbb{E} \left\{ \exp(sz_1 + tz_1^2) \right\} \\ &= \frac{1}{\sqrt{2\pi} \operatorname{erf}(\tau/\sqrt{2})} \int_{-\tau}^{\tau} \exp\left(sz - \frac{qt}{2}z^2\right) dz. \end{aligned}$$

After completing the square separately for $q \geq 0$, let $\varsigma = \operatorname{sig}(q) := q/|q|$, so that

$$\mathcal{M}_1(s, t) = \frac{\exp(s^2/2q)}{\sqrt{2\pi} \operatorname{erf}(\tau/\sqrt{2})} \int_{-\tau}^{\tau} \exp\left\{-\varsigma \left(\sqrt{\frac{|q|}{2}}z - \frac{v}{\sqrt{2|q|}}\right)^2\right\} dz,$$

for $q \neq 0$. Now we change the variable of integration and substitute $\operatorname{erf}(x) := (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$, which will have an imaginary argument when $q < 0$, so that

$$\begin{aligned} \mathcal{M}_1(s, t) &= \frac{\exp(s^2/2q)}{2\sqrt{|q|} \operatorname{erf}(\tau/\sqrt{2})} \times \\ &\left[\operatorname{erf}\left\{\sqrt{\varsigma} \left(\sqrt{\frac{|q|}{2}}\tau - \frac{s}{\sqrt{2|q|}}\right)\right\} - \operatorname{erf}\left\{\sqrt{\varsigma} \left(-\sqrt{\frac{|q|}{2}}\tau - \frac{s}{\sqrt{2|q|}}\right)\right\} \right]. \end{aligned} \quad (2.25)$$

When $q = 0$, at the boundary between real and complex arithmetic, \mathcal{M}_j has a removable singularity, as do all other formulas in this section. Relevant limiting forms are presented in Section 2.4.1.

The joint MGF of (b, u) is

$$\begin{aligned} \mathcal{M}(s, t) &= \mathbb{E} \left\{ \exp(sn\bar{z} + t\mathbf{z}^T \mathbf{z}) \right\} \\ &= \mathbb{E} \left\{ \exp\left(\sum_j \{sz_j + tz_j^2\}\right) \right\} = \{\mathcal{M}_1(s, t)\}^n \end{aligned}$$

and is convergent for all $(s, t) \in \mathbb{R}^2$. Therefore, $\mathcal{K}(s, t) = n \ln \mathcal{M}_1(s, t)$.

Note that since $z_j \in (-\tau, \tau)$, for $j = 1, \dots, n$, then $b \in (-n\tau, n\tau)$. Also,

$$u = \mathbf{z}^T \mathbf{z} < n\tau^2 = x^2(n-1)$$

and $u \in (0, x^2(n-1))$. Thus the support is bounded in \mathbb{R}^2 .

To show that $\hat{s} = 0$, note

$$\begin{aligned} 0 &= \mathcal{K}'_s(\hat{s}, \hat{t}) = \frac{n}{\mathcal{M}_1(\hat{s}, \hat{t})} \left. \frac{\partial \mathcal{M}_1(s, t)}{\partial s} \right|_{s=\hat{s}} \\ &= \frac{n}{\mathcal{M}_1(\hat{s}, \hat{t})} \mathbb{E} [z_1 \exp(s z_1 + \hat{t} z_1^2)]. \end{aligned}$$

For any value of \hat{t} , $\hat{s} = 0$ solves the equation since $E[z_1 \exp\{\hat{t} z_1^2\}] = 0$. If, for $\hat{s} = 0$ a solution \hat{t} can be determined for $\mathcal{K}'_t(0, \hat{t}) = n - 1$, then $(0, \hat{t})$ must be the saddlepoint, since there is only one solution to both equations. (Note that as $t \uparrow \infty$ then $\mathcal{K}'_t(0, \hat{t})$ must exceed $n - 1$). The finding that $\hat{s} = 0$ induces symmetry into the evaluation of \mathcal{K} and its derivatives at saddlepoint $(0, \hat{t})$. At the saddlepoint, $\exp\{\mathcal{K}(0, \hat{t})\}$, required by (2.11), simplifies to

$$\exp\{\mathcal{K}(0, \hat{t})\} = \{\mathcal{M}_1(\mathbf{0}, \hat{t})\}^n = \hat{q}^{-n/2} \left(\frac{\operatorname{erf}(\tau \sqrt{\hat{q}/2})}{\operatorname{erf}(\tau/\sqrt{2})} \right)^n, \quad (2.26)$$

where $\hat{q} = 1 - 2\hat{t}$.

The saddlepoint \hat{t} is the solution to

$$\begin{aligned} n - 1 &= \mathcal{K}'_t(0, \hat{t}) = \frac{n}{\mathcal{M}_1(0, \hat{t})} \left. \frac{\partial \mathcal{M}_1(0, t)}{\partial t} \right|_{t=\hat{t}} = \frac{n}{\mathcal{M}_1(0, \hat{t})} \mathbb{E}\{z_1^2 \exp(\hat{t} z_1^2)\} \\ &= \frac{n}{\hat{q}} \left(1 - \frac{\sqrt{2\hat{q}/\pi} \tau e^{-\hat{q}\tau^2/2}}{\operatorname{erf}(\tau \sqrt{\hat{q}/2})} \right) = \frac{n}{\hat{q}} \{1 - R(\hat{t})\} \end{aligned} \quad (2.27)$$

The derivatives of $\mathcal{K}(s, t)$ presented below are shown only for $s = \hat{s} = 0$.

First derivatives:

$$\begin{aligned}\hat{\mathcal{K}}'_s(0, t) &= \left. \frac{\partial}{\partial s} \mathcal{K}(s, t) \right|_{s=0} = 0 \quad \forall t \\ \hat{\mathcal{K}}'_t(0, \hat{t}) &= \left. \frac{\partial}{\partial t} \mathcal{K}(s, t) \right|_{s=0, t=\hat{t}} = \frac{nR_1(\hat{t})}{\hat{q}}\end{aligned}$$

where

$$R_1(\hat{t}) := 1 - R(\hat{t}) \quad (2.28)$$

$$R(\hat{t}) := \frac{\sqrt{2\hat{q}/\pi}\tau e^{-\hat{q}\tau^2/2}}{\operatorname{erf}\left(\tau\sqrt{\hat{q}/2}\right)} \quad (2.29)$$

Second derivatives:

$$\begin{aligned}\hat{\mathcal{K}}''_{s^2}(0, \hat{t}) &= \left. \frac{\partial^2}{\partial s^2} \mathcal{K}(s, t) \right|_{s=0, t=\hat{t}} = \frac{nR_1(\hat{t})}{\hat{q}} \\ \hat{\mathcal{K}}''_{st}(0, \hat{t}) &= \left. \frac{\partial^2 \mathcal{K}(s, t)}{\partial s \partial t} \right|_{s=0, t=\hat{t}} = \frac{\partial \mathcal{K}'_s(0, t)}{\partial t} = 0 \quad \forall t \\ \hat{\mathcal{K}}''_{t^2}(0, \hat{t}) &= \left. \frac{\partial^2 \mathcal{K}(s, t)}{\partial t^2} \right|_{s=0, t=\hat{t}} = \frac{nR_2(\hat{t})}{\hat{q}^2}\end{aligned} \quad (2.30)$$

where

$$\hat{R}_2(\hat{t}) := 2 - (1 + \hat{q}\tau^2)R(\hat{t}) - R^2(\hat{t}) \quad (2.31)$$

Third derivatives:

$$\begin{aligned}\hat{\mathcal{K}}^{(3)}_{s^3}(0, \hat{t}) &= \left. \frac{\partial^3}{\partial s^3} \mathcal{K}(s, t) \right|_{s=0, t=\hat{t}} = 0 \quad \forall t \\ \hat{\mathcal{K}}^{(3)}_{s^2t}(0, \hat{t}) &= \left. \frac{\partial^3}{\partial s^2 \partial t} \mathcal{K}(s, t) \right|_{s=0, t=\hat{t}} = \frac{nR_2(\hat{t})}{\hat{q}^2} \\ \hat{\mathcal{K}}^{(3)}_{st^2}(0, \hat{t}) &= \left. \frac{\partial^3 \mathcal{K}(s, t)}{\partial s \partial t^2} \right|_{s=0, t=\hat{t}} = 0 \quad \forall t \\ \hat{\mathcal{K}}^{(3)}_{t^3}(0, \hat{t}) &= \left. \frac{\partial^3 \mathcal{K}(s, t)}{\partial t^3} \right|_{s=0, t=\hat{t}} = \frac{nR_3(\hat{t})}{\hat{q}^3}\end{aligned}$$

where

$$R_3(\hat{t}) := 8 - (\hat{q}^2\tau^4 + 2\hat{q}\tau^2 + 3) R(\hat{t}) - 3(1 + \hat{q}\tau^2) R^2(\hat{t}) - 2R^3(\hat{t}) \quad (2.32)$$

Fourth derivatives:

$$\begin{aligned} \hat{\mathcal{K}}_{s^4}^{(4)}(0, \hat{t}) &= \left. \frac{\partial^4 \mathcal{K}(s, t)}{\partial s^4} \right|_{s=0, t=\hat{t}} = \frac{n \{R_2(\hat{t}) - 2R_1^2(\hat{t})\}}{\hat{q}^2} \\ \hat{\mathcal{K}}_{s^3t}^{(4)}(0, \hat{t}) &= \left. \frac{\partial^4 \mathcal{K}(s, t)}{\partial s^3 \partial t} \right|_{s=0} = 0 \quad \forall t \\ \hat{\mathcal{K}}_{s^2t^2}^{(4)}(0, \hat{t}) &= \left. \frac{\partial^4 \mathcal{K}(s, t)}{\partial s^2 \partial t^2} \right|_{s=0, t=\hat{t}} = \frac{nR_3(\hat{t})}{\hat{q}^3} \\ \hat{\mathcal{K}}_{st^3}^{(4)}(0, \hat{t}) &= \left. \frac{\partial^4 \mathcal{K}(s, t)}{\partial s \partial t^3} \right|_{s=0} = 0 \quad \forall t \\ \hat{\mathcal{K}}_{t^4}^{(4)}(0, \hat{t}) &= \left. \frac{\partial^4 \mathcal{K}(s, t)}{\partial t^4} \right|_{s=0, t=\hat{t}} = \frac{nR_4(\hat{t})}{\hat{q}^4} \end{aligned}$$

where

$$\begin{aligned} R_4(\hat{t}) &:= 48 - (\hat{q}^3\tau^6 + 3\hat{q}^2\tau^4 + 9\hat{q}\tau^2 + 15) R(\hat{t}) \\ &\quad - (7\hat{q}^2\tau^4 + 14\hat{q}\tau^2 + 15) R^2(\hat{t}) - 12(1 + \hat{q}\tau^2) R_j^3(\hat{t}) - 6R^4(\hat{t}) \end{aligned} \quad (2.33)$$

From (2.18), the second-order term in the saddlepoint approximation is $O = \{\hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24\}$, where the $\hat{\kappa}$ terms are now defined. Using the Einstein summation convention as in McCullagh (1987, §1.2), where summation is implied over any subscript appearing once in a subscript and once in a superscript, the three new $\hat{\kappa}$ terms in the O formula are defined as

$$\begin{aligned} \hat{\kappa}_4 &:= \hat{\mathcal{K}}_{ijkl} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kl} \\ \hat{\kappa}_{23}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ir} \hat{\mathcal{K}}^{js} \hat{\mathcal{K}}^{kt} , \\ \hat{\kappa}_{13}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kr} \hat{\mathcal{K}}^{st} \end{aligned} \quad (2.34)$$

where the subscripted $\hat{\mathcal{K}}$ terms are derivatives of \mathcal{K} with respect to the subscripted components, evaluated at the saddlepoint $(s, t) = (0, \hat{t})$, and the superscripted $\hat{\mathcal{K}}$ terms are components of the inverse of the Hessian matrix $\mathcal{K}''(0, \hat{t})$. When any index is 1, the referenced derivative is

by s ; when the index is 2, the derivative is by t . For example, when $i = j = 1$ and $k = 2$, $\hat{\mathcal{K}}_{112} = \partial^3 \mathcal{K}(s, t) / \partial s^2 \partial t|_{(s,t)=(0,\hat{t})}$. In more complex applications, as in Chapters 3 and 4, these sums must be calculated and summed term by term. However, in this case, O can be simplified into this combination of the R functions defined above in (2.28), (2.31), (2.32), and (2.33).

$$O = \frac{1}{24n} \left(-6 \frac{R_2(\hat{t})}{R_1^2(\hat{t})} + 3 \frac{R_4(\hat{t})}{R_2^2(\hat{t})} - 5 \frac{R_3^2(\hat{t})}{R_2^3(\hat{t})} - 6 \right) \quad (2.35)$$

2.4.1 Limiting Forms

As noted, the CGF $\mathcal{K}(s, t)$, its derivatives, and all saddlepoint distributions for M have a removable discontinuity at $t = 0.5$, corresponding to the M value $x = \sqrt{3}$. Although this is a set of probability 0, with theoretically no significance, it is important in programming these methods. When calculating approximations of F , using typical double-precision arithmetic, there is a neighborhood of the discontinuity within which these calculations become unstable. In this neighborhood, it is advisable to switch to a Taylor series approximation, built using these limiting forms, to present a well-behaved function for the user. Here are the limiting values of $\mathcal{K}(0, 1/2)$, and its non-zero derivatives.

$$\exp \left\{ \mathcal{K} \left(0, \frac{1}{2} \right) \right\} = \left(\frac{2\tau}{\sqrt{2\pi} \operatorname{erf}(\tau/\sqrt{2})} \right)^n \quad (2.36)$$

$$\mathcal{K}'_t \left(0, \frac{1}{2} \right) = \mathcal{K}''_{s^2} \left(0, \frac{1}{2} \right) = \frac{n\tau^2}{3} \quad (2.37)$$

$$\mathcal{K}''_{t^2} \left(0, \frac{1}{2} \right) = \mathcal{K}^{(3)}_{s^2 t} \left(0, \frac{1}{2} \right) = \frac{4n\tau^4}{45} \quad (2.38)$$

$$\mathcal{K}^{(3)}_{t^3} \left(0, \frac{1}{2} \right) = \mathcal{K}^{(4)}_{s^2 t^2} \left(0, \frac{1}{2} \right) = \frac{16n\tau^6}{945} \quad (2.39)$$

$$\mathcal{K}^{(4)}_{s^4} \left(0, \frac{1}{2} \right) = \frac{-2n\tau^4}{15} \quad (2.40)$$

$$\mathcal{K}^{(4)}_{t^4} \left(0, \frac{1}{2} \right) = \frac{-32n\tau^8}{4725} \quad (2.41)$$

2.5 Bonferroni Bounds

A Bonferroni expansion provides upper and lower bounds for the distribution of M . In fact, the series of Bonferroni inequalities provides ever-sharper alternating bounds for the true value. If all the terms in the Bonferroni inequalities could be calculated, the exact distribution of M would be known. Only the first bound is easily calculated, and the second bound requires numerical integrations. Additional terms become increasingly difficult to calculate.

However, each Bonferroni bound has a region of M values for which it provides the exact answer. When $M \geq M_2$, the first Bonferroni bound is exact; when $M \in [M_3, M_2)$, the second bound is exact. Since the saddlepoint method is only an approximation, it is reasonable to use the exact Bonferroni method when it is available and easier to calculate. We recommend calibrating the saddlepoint method to provide a smooth transition point to the exact Bonferroni bound, either at M_2 or at M_3 .

Grubbs (1950) and others have used the first Bonferroni bound to estimate p -values for an outlier test. Table 2.1 lists these values as 1st upper (Grubbs) Bonferroni bounds on the p -value. Kounias (1968), Hunter (1976), and Worsley (1982) developed an improved Bonferroni first bound that is sharper. Table 2.1 lists these values as 1st upper (Worsley) bounds.

2.5.1 Bonferroni Inequalities

As defined earlier, $E_j = \{|a_j| \leq x\} = \{|y_j| \leq x\sqrt{(n-1)/n}\}$, so $\{M \leq x\} = \bigcap_{j=1}^n E_j$, and $1 - F_M(x) = \Pr(M > x) = \Pr\left\{\bigcup_{j=1}^n \overline{E}_j\right\}$. Now, define sums of probabilities of collections of k events as follows:

$$\begin{aligned} S_1 &= \sum_{j=1}^n \Pr(\overline{E}_j) \\ S_2 &= \sum_{1 \leq i < j \leq n} \Pr(\overline{E}_i \cap \overline{E}_j) \\ &\quad \vdots \\ S_k &= \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \Pr(\overline{E}_{i_1} \cap \overline{E}_{i_2} \cap \dots \cap \overline{E}_{i_k}) \end{aligned}$$

From elementary probability theory, as in Ross (1988), $1 - F_M(x) = \Pr\left(\bigcup_{j=1}^n \overline{E}_j\right) = S_1 - S_2 + S_3 - \dots + (-1)^{n+1} S_n$. Noting that the probability of an intersection cannot be larger than the probability of the component events, this leads to the following series of inequalities:

$$\begin{aligned}
S_1 - S_2 &\leq 1 - F_M(x) \leq S_1 \\
S_1 - S_2 + S_3 - S_4 &\leq 1 - F_M(x) \leq S_1 - S_2 + S_3 \\
&\vdots \\
S_1 - S_2 + \dots - S_{2r} &\leq 1 - F_M(x) \leq S_1 - S_2 + \dots + S_{2r-1} \\
&\text{for } r = 1, \dots, \lfloor \frac{n+1}{2} \rfloor
\end{aligned} \tag{2.42}$$

From this point on, we focus only on the first line of (2.42), comprising the first Bonferroni bound S_1 and the second bound $S_1 - S_2$.

Kounias (1968), Hunter (1976), and Worsley (1982) derived an improvement on the upper bound S_1 , using a spanning tree on the graph of the set of intersections. From Worsley's Theorem 1, represent events $\overline{E}_1, \dots, \overline{E}_n$ as vertices of a graph G , where vertices \overline{E}_i and \overline{E}_j are joined by an edge e_{ij} if and only if $\Pr(\overline{E}_i \cap \overline{E}_j) > 0$. Let T be a subgraph of G . Then,

$$\Pr\left(\bigcup_{j=1}^n \overline{E}_j\right) \leq S_1 - \sum_{i,j:e_{ij} \in T} \Pr(\overline{E}_i \cap \overline{E}_j) \tag{2.43}$$

if and only if T is a tree, defined as an acyclic connected graph, in which any two vertices are connected by exactly one path. T is defined to be a spanning tree of G if every vertex in G is included in T . It follows that to achieve the sharpest upper bound on $1 - F_M(x)$ by this method, one should find the spanning tree T that maximizes $\sum_{i,j:e_{ij} \in T} \Pr(\overline{E}_i \cap \overline{E}_j)$ and subtract this sum from S_1 in (2.43). Denoting that sum of intersection probabilities over the maximal spanning tree as S_{2*} , the improved first two Bonferroni bounds are

$$S_1 - S_2 \leq 1 - F(x) \leq S_1 - S_{2*} \tag{2.44}$$

For each Bonferroni bound, there is a value of M , say M_q for the q^{th} bound, such that if $M \geq M_q$, then all q -way intersections have zero probability. In other words, $M_q = \inf \{M : S_q = 0\}$. When $M \geq M_2$, $S_j = 0$ for $j \geq 2$, and the first bound is exact, i.e., $1 - F(x) = S_1$ when $x \geq M_2$. When $M \geq M_3$, $S_j = 0$ for $j \geq 3$, and the second bound is exact, i.e., $1 - F(x) = S_1 - S_2$ when $x \geq M_3$.

For the one-sample problem providing the setting for this chapter, all $\Pr(\overline{E}_i \cap \overline{E}_j)$ are equal for any $i \neq j$, and therefore, the probability S_{2*} of any spanning tree is the same. Therefore, only one calculation $\Pr(\overline{E}_1 \cap \overline{E}_2)$ is required. From this probability, $S_{2*} = (n - 1) \Pr(\overline{E}_1 \cap \overline{E}_2)$ and $S_2 = \{n(n - 1) / 2\} \Pr(\overline{E}_1 \cap \overline{E}_2)$.

2.5.2 Distribution Theory

The theory behind the computation of M_2 , M_3 , S_1 , S_2 , and S_{2*} derives from the joint distribution of the studentized residuals. Butler (1984) presented this as Lemma 3.1, in a slightly different setting. For more generality, Butler uses a \mathbf{Z} matrix, here set to the identity matrix for outlier identification; Butler also scales the \mathbf{r} residual vector to a maximum value of 1. To adapt to our setting, let $\mathbf{r} = \mathbf{a}(n - 1)^{-1/2}$.

Let $\mathcal{S} = \{i_1, \dots, i_q\}$ be any q -tuple subset of $\{1, \dots, n\}$. Let $\mathbf{R}_{\mathcal{S}}$ be the $q \times q$ correlation matrix of $(r_j : j \in \mathcal{S})$, namely $\mathbf{R}_{\mathcal{S}} = (\rho_{ij})$. In this setting, $\rho_{ij} = -1/(n - 1)$ for all $i \neq j$ and $\rho_{ij} = 1$ for $i = j$. The q -variate density of $\mathbf{r}_{\mathcal{S}}$ is

$$f(\mathbf{r}) = \frac{\Gamma\{(n-1)/2\}}{\Gamma\{(n-1-q)/2\}} \pi^{-q/2} |\mathbf{R}_{\mathcal{S}}|^{-1/2} (1 - \mathbf{r}^T \mathbf{R}_{\mathcal{S}}^{-1} \mathbf{r})^{(n-q-3)/2} \quad \text{if } \mathbf{r}^T \mathbf{R}_{\mathcal{S}}^{-1} \mathbf{r} \leq 1 \quad (2.45)$$

To calculate M_2 , Butler provides the following formulas in Lemma 4.1, here including the $\sqrt{n - 1}$ adjustment to this setting.

$$\Pr(|a_i| > x, |a_j| > x) = 0 \quad \text{iff } x \geq \sqrt{\frac{(n - 1)(1 + |\rho_{ij}|)}{2}}.$$

From this, it directly follows that:

$$M_2 = \sqrt{\frac{n}{2}}.$$

The calculation of M_3 depends on the bounds of the ellipsoid defined by $\mathbf{r}^T \mathbf{R}_S^{-1} \mathbf{r} = 1$, the boundary of the density function in (2.45). The calculation procedure is based on Lemma 4.2 of Butler (1984), with $q = 3$. This algorithm could be simplified for the one-sample setting. The `PMASR_Bonf` code described in Appendix A uses the same algorithm for all settings analyzed in this dissertation.

The sum $S_1 = \sum_{j=1}^n \Pr(|a_j| > x)$ is calculated by using the marginal distribution from (2.45), which has been shown by Grubbs (1950) and others to be related to Student's t distribution through this identity:

$$a_j \sqrt{\frac{n-2}{n-1-a_j^2}} \sim t_{n-2} \quad (2.46)$$

Therefore, in terms of the t distribution:

$$S_1 = 2n \left\{ T_{n-2} \left(x \sqrt{\frac{n-2}{n-1-x^2}} \right) \right\},$$

where $T_\nu(\cdot)$ is the right tail probability of the Student's t distribution with ν degrees of freedom.

The sums S_2 and S_{2*} comprise probabilities of intersections $\Pr[|a_i| > x, |a_j| > x]$. Calculating the intersection probability $\Pr(\overline{E}_i \cap \overline{E}_j)$ is an integration of the bivariate marginal density of $\mathbf{r} = (r_i, r_j)^T$, given as:

$$f(\mathbf{r}) = \frac{n-3}{2\pi\sqrt{1-\rho_{ij}^2}} (1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r})^{\frac{n-5}{2}} \quad \text{if } \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \leq 1 \quad (2.47)$$

where

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix},$$

and $\rho_{ij} = -1/(n-1)$

Calculation of $\Pr(|a_i| > x, |a_j| > x) = \Pr(|r_i| > c, |r_j| > c)$, where $c = x(n-1)^{-1/2}$, requires bivariate integration over four regions of the ellipse defined by $\mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \leq 1$. By symmetry,

this can be reduced from four to two regions. For each of the two regions, using the analysis of Butler (1982), we simplify the bivariate into univariate integration, computed using the `integrate` function of R. Note that for the one-sample setting of this chapter, all $\Pr(\overline{E}_i \cap \overline{E}_j)$ are equal, so only $\Pr(\overline{E}_1 \cap \overline{E}_2)$ needs to be calculated,

One special case, when $n = 3$, requires special handling in the algorithm. For this small sample size, the joint density of two residuals becomes degenerate. In this case, use (2.46) to calculate joint probabilities to avoid a calculation error.

Chapter 3

Outlier Discordancy in Univariate Regression

In the univariate regression setting, the optimal test for a single outlying observation is provided by testing for large values of the maximum absolute studentized residual (MASR) M . Since the exact distribution of M is intractable, we present approximations to the distribution based on a saddlepoint approximation. These are compared to first and second Bonferroni bounds and to a Monte Carlo simulation.

3.1 Setting

In this section, we define the setting for the outlier discordancy test and discuss the optimality of the test.

3.1.1 Null model setting

The null model assumes that there are no outliers in the data. The data derive from a standard univariate linear model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{3.1}$$

in which $\mathbf{y} = (y_1, \dots, y_n)^T$ is an $n \times 1$ vector of observed values; $\mathbf{X} = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)^T$ is the $n \times p$ design matrix of full rank $p < n$, in which the first column is a vector of ones and β is an unknown $p \times 1$ parameter vector, which may consist of both fixed and random effects. The studentized residuals, defined shortly, are ancillary to any potential random effect parameters and hence are independent of the corresponding random effect distributions. Hence these outlier results apply to all fixed, random, or mixed models. Note that for residuals to have a non-degenerate distribution, this requires $p < n - 1$.

The random error vector ε in (3.1) is assumed to be left-orthogonally invariant (l.o.i.) with uncorrelated elements. This means that its distribution is invariant to premultiplication by any $n \times n$

orthogonal matrix. The most prominent members of this class are independent and identically distributed (i.i.d.) normal components and the class of spherically symmetric multivariate t distributions; see Butler (1986) and Dawid (1977) for details. From Dempster (1969, Section 13.3), the distribution of ε is fully determined by that of $\varepsilon^T \varepsilon$. Also, from Butler (1986), any statistic which is a function of the least square residuals will have the same distribution regardless of which l.o.i. distribution is assumed for ε . Therefore, the following results in this chapter assume that $\varepsilon \sim \mathbb{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, and the same results also apply to any model (3.1) where ε may follow any l.o.i. distribution.

We define y_j to be an outlier when this alternative model holds:

$$\mathbf{y} = \mathbf{X}\beta + \theta \mathbf{e}_j + \varepsilon, \quad (3.2)$$

when $\theta \in \mathfrak{R}$, $\theta \neq 0$, and \mathbf{e}_j is the j^{th} column of \mathbf{I}_n , for $j = 1, \dots, n$.

Define $\sigma^2 := \mathbb{E}(\varepsilon^T \varepsilon) / n$. Denote the hat matrix as $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = (h_{ij})$ and the residual projection matrix as $\mathbf{I} - \mathbf{H}$. The usual parameter estimates are $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ and $\hat{\sigma}^2 = \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y} / (n - p) = \hat{\varepsilon}^T \hat{\varepsilon} / (n - p)$, where $\hat{\varepsilon} = (\mathbf{I} - \mathbf{H}) \mathbf{y} = (\hat{\varepsilon}_j)$ and $\hat{\mathbf{y}} = \mathbf{H} \mathbf{y}$. The vector of (internally) studentized residuals is $\mathbf{a} = (a_1, \dots, a_n)^T$, where

$$a_j = \frac{\hat{\varepsilon}_j}{\hat{\sigma} \sqrt{1 - h_{jj}}} = \frac{y_j - \mathbf{x}_j^T \hat{\beta}}{\hat{\sigma} \sqrt{1 - h_{jj}}} \quad j = 1, \dots, n. \quad (3.3)$$

The maximum absolute studentized residual (MASR) is the statistic $M = \max_j |a_j|$ whose cumulative distribution function (CDF) under the null model is denoted as $F_M(x) = \Pr(M \leq x)$. The distribution F_M is the same, regardless of the assumed l.o.i. distribution for ε . A typical test for the discordancy of a single outlier caused by a change in distribution for one component of ε rejects the null model when M is large. Thus, if x is the observed value of M , then the p -value is $1 - F_M(x)$. The exact distribution F_M is intractable.

Alternatively, upper and lower Bonferroni bounds are available. In limited ranges at the upper end of the support of M , these bounds can be exact. The upper Bonferroni bound on the p -value

has become known as Grubbs' test. The details of these bounds, and the ranges within which they are exact, are presented in Section 3.5. This chapter will show where the Bonferroni bounds are adequate or inadequate and show the opportunity for an improved method to be provided by the saddlepoint approximations.

The approach is also applicable when some of the columns of \mathbf{X} are random. In this instance, the distribution theory is conditional upon the attained values of these random columns. Furthermore, the approach applies to classical linear mixed models in which some components of β represent random effects. Conditional upon the values these random effects, the distribution of M does not depend upon these values; hence neither does the unconditional distribution. However, the design matrix for the accompanying random effects must be a part of the matrix \mathbf{X} .

3.1.2 Optimality

The optimality of this outlier test occurs when the distribution of ε is left orthogonally invariant and 1-extendable with the latter property now defined. Error vector ε is said to be 1-extendable if ε is the marginal distribution of some $(n + 1) \times 1$ random vector which is also l.o.i. This condition is still satisfied by normal, spherically symmetric multivariate t , and many other error distributions. Butler (1986) proved that the two-hypothesis F test in ANOVA is uniformly most powerful invariant (UMPI) for all such 1-extendable l.o.i. error distributions.

Butler (1986) also showed that the test based on large $M = \max_j |a_j|$ is optimal within a multiple $(n + 1)$ -hypotheses framework with H_0 : no outliers vs. H_j : the j^{th} element of \mathbf{y} is an outlier for $j = 1, \dots, n$, as defined in (3.2). Let ϕ_j denote the probability of selecting H_j , for $j = 0, 1, \dots, n$ and choose a loss function that is the indicator of a correct decision. Then the Bayes risk with respect to a prior weighting of the hypotheses is the probability of an incorrect decision.

Subject to group invariant conditions on the tests allowed, Butler (1986, thm. 4.1) shows that the test defined by $\phi_{j^*} = 1$ if $|a_{j^*}| = M > c_1$, and $\phi_0 = 1$ otherwise is the Bayes decision rule that minimizes the probability of an incorrect decision uniformly in $\theta \in \mathfrak{R}$ as given by the Bayes risk.

The tests over which the optimality applies are restricted as follows: (i) invariant to $\mathbf{y} \rightarrow \mathbf{y} + \mathbf{X}\gamma$ for any $\gamma \in \mathfrak{R}^p$, i.e. translation of \mathbf{y} by an arbitrary $(n \times 1)$ vector in the column space of \mathbf{X} , (ii) invariant to $\mathbf{y} \rightarrow \mathbf{y}/\xi$ for any $\xi > 0$, i.e. rescaling of \mathbf{y} by an arbitrary constant, (iii) with level $0 \leq \alpha \leq 1$, so that $E\left(\sum_{j=1}^n \phi_j | H_0\right) = 1 - E(\phi_0 | H_0) \leq \alpha$, and (iv) based upon an equal prior weighting of $0 < \pi < n^{-1}$ given to H_1, \dots, H_n , where π ensures the level α in (iii). The consequence of this is that any other test for a single outlier that meets the given criteria can not have a higher probability of making a correct decision than the test we consider here, i.e. the test based upon large values for $M = \max_j |a_j|$.

3.2 Saddlepoint Approximation

An approximation for $F_M(x)$ is derived by first reexpressing this CDF as a conditional probability to which Bayes' theorem can be applied. A reversal of the probabilities in Bayes' theorem allows approximation to the CDF to be reexpressed in terms of a saddlepoint density approximation. Using first- and various second-order saddlepoint approximations for this multivariate density leads to various CDF approximations $\hat{F}_M(x)$ with second-order approximations providing greater accuracy.

Statistic M has bounded support, and owing to the nature of saddlepoint approximations, this can lead to a loss in accuracy for the various $\hat{F}_M(x)$. Since the least upper bound on the support of M is $\sqrt{n-p}$, these approximations can be calibrated by using a known exact value at the upper end of the support of M , as described in Section 3.2.4. This procedure is shown to improve accuracy.

3.2.1 Reformulation of $F_M(x)$

The studentized residual vector \mathbf{a} has a joint distribution which does not depend on the parameters (β, σ^2) , so that \mathbf{a} is ancillary. Thus, there is no loss in generality in assuming $\beta = \mathbf{0}$ and $\sigma^2 = 1$. Since estimates $(\hat{\beta}, \hat{\sigma}^2)$ form a complete sufficient statistic for (β, σ^2) , $(\hat{\beta}, \hat{\sigma}^2)$ and \mathbf{a} are independent by Basu's theorem. Thus, the distribution of \mathbf{a} may be conditioned on any convenient

value for $(\hat{\beta}, \hat{\sigma}^2)$, say $(\mathbf{0}, 1)$. The net effect of such conditioning leads to

$$\begin{aligned} F_M(x) &= \Pr \left(\max_j |a_j| \leq x \mid \hat{\beta} = \mathbf{0}, \hat{\sigma}^2 = 1 \right) \\ &= \Pr \left(\max_j \frac{|y_j|}{\sqrt{1 - h_{jj}}} \leq x \mid \hat{\beta} = \mathbf{0}, \hat{\sigma}^2 = 1 \right). \end{aligned} \quad (3.4)$$

By defining the events $E_j = \{|y_j| \leq x\sqrt{1 - h_{jj}}\}$, so $\{M \leq x\} = \bigcap_{j=1}^n E_j$, then Bayes' theorem for densities is applied to the conditional probability in (3.4) to give

$$F_M(x) = \frac{f_{\hat{\beta}, \hat{\sigma}^2}(\mathbf{0}, 1 \mid \bigcap_{j=1}^n E_j) \Pr \left(\bigcap_{j=1}^n E_j \right)}{f_{\hat{\beta}, \hat{\sigma}^2}(\mathbf{0}, 1)}, \quad (3.5)$$

where $f_{\hat{\beta}, \hat{\sigma}^2}(\mathbf{0}, 1)$ denotes the marginal density of $\hat{\beta}, \hat{\sigma}^2$ at $(\mathbf{0}, 1)$, and $f_{\hat{\beta}, \hat{\sigma}^2}(\mathbf{0}, 1 \mid \bigcap_{j=1}^n E_j)$ denotes the conditional density of $\hat{\beta}, \hat{\sigma}^2$ at $(\mathbf{0}, 1)$ given event $\bigcap_{j=1}^n E_j$.

The denominator density is simple and explicit. With $\beta = \mathbf{0}$ and $\sigma^2 = 1$, then $\hat{\beta} \sim \mathbb{N}_p\{\mathbf{0}, (\mathbf{X}^T \mathbf{X})^{-1}\}$ and $\hat{\sigma}^2 (n - p) \sim \chi_{n-p}^2$ are independent, so that

$$\begin{aligned} f_{\hat{\beta}, \hat{\sigma}^2}(\mathbf{0}, 1) &= f_{\hat{\beta}}(\mathbf{0}) f_{\hat{\sigma}^2}(1) = \frac{|\mathbf{X}^T \mathbf{X}|^{1/2}}{(2\pi)^{p/2}} \times (n - p) f_{\chi_{n-p}^2}(n - p) \\ &= \frac{|\mathbf{X}^T \mathbf{X}|^{1/2}}{(2\pi)^{p/2}} \times (n - p) \frac{(n - p)^{(n-p)/2-1} \exp\{-(n - p)/2\}}{2^{(n-p)/2} \Gamma\{(n - p)/2\}} \\ &= \frac{|\mathbf{X}^T \mathbf{X}|^{1/2}}{(2\pi)^{(p+1)/2}} \frac{\hat{\Gamma}\{(n - p)/2\}}{\Gamma\{(n - p)/2\}} \sqrt{(n - p)/2} =: g, \end{aligned} \quad (3.6)$$

where $\hat{\Gamma}(v) = \sqrt{2\pi} v^{v-1/2} e^{-v}$ denotes Stirling's approximation for $\Gamma(v)$.

Since $\{E_j\}$ are independent events and y_1, \dots, y_n are i.i.d. $\mathbb{N}(0, 1)$,

$$\Pr \left(\bigcap_{j=1}^n E_j \right) = \prod_{j=1}^n \Pr(E_j) = \prod_{j=1}^n \operatorname{erf} \left(\tau_j / \sqrt{2} \right),$$

where $\operatorname{erf}(x) := (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$, and $\tau_j := x\sqrt{1 - h_{jj}}$ for $j = 1, \dots, n$.

Only the conditional density in the numerator of (3.5) offers any difficulty. Note that $\hat{\beta}$ is linear in $\{y_j\}$, and

$$(n - p) \hat{\sigma}^2 = \mathbf{y}^T \mathbf{y} - \hat{\beta}^T \mathbf{X}^T \mathbf{X} \hat{\beta},$$

is linear in $\{y_j^2\}$ apart from its dependence on $\hat{\beta}$. To obtain a multivariate density in which all components are linear in $\{y_j\}$ and $\{y_j^2\}$, write the conditional density of $(\hat{\beta}, \hat{\sigma}^2)$ given $\bigcap_{j=1}^n E_j$ in terms of the conditional density of $(n\hat{\beta}, v) = (n\hat{\beta}, \mathbf{y}^T \mathbf{y})$ given $\bigcap_{j=1}^n E_j$ through a one-to-one transformation. The Jacobian of this transformation is $n^p (n - p)$ and

$$f_{\hat{\beta}, \hat{\sigma}^2}(\mathbf{0}, 1 \mid \bigcap_{j=1}^n E_j) = n^p (n - p) f_{n\hat{\beta}, v}(\mathbf{0}, n - p \mid \bigcap_{j=1}^n E_j).$$

Next, note that the net effect of conditioning on the event $\bigcap_{j=1}^n E_j$ is to truncate the normal random variables so that y_j is restricted to the range $|y_j| < \tau_j$ for $j = 1, \dots, n$. Conditioning on $\bigcap_{j=1}^n E_j$ affects the distribution theory in such a way that y_j is independent of $\bigcap_{i \neq j} E_i$ and is only affected by the conditioning on E_j . Thus $\{y_j \mid E_j\}$ are independent with $y_j \mid E_j$ having a standard normal distribution truncated to $\{|y_j| < \tau_j\}$. Let z_j denote such a truncated random variable with density

$$f_{z_j}(x) = \frac{\phi(x)}{\text{erf}(\tau_j/\sqrt{2})} \quad |x| < \tau_j,$$

where $\phi(\cdot)$ is the standard normal density function, and let $(n \times 1)$ random vector \mathbf{z} have independent elements z_j . In terms of \mathbf{z} ,

$$n\hat{\beta} \mid \bigcap_{j=1}^n E_j \stackrel{D}{=} n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z} =: \mathbf{b}, \text{ and} \quad (3.7)$$

$$v \mid \bigcap_{j=1}^n E_j = \mathbf{y}^T \mathbf{y} \mid \bigcap_{j=1}^n E_j \stackrel{D}{=} \mathbf{z}^T \mathbf{z} =: u. \quad (3.8)$$

The conditional distributions on the left-hand side of (3.7) and (3.8) are the same as the unconditional distributions on the right-hand side using truncated random variables, thus effectively removing the conditioning required in the distribution theory. From (3.5) we may write

$$F_M(x) = f_{\mathbf{b},u}(\mathbf{0}, n-p) \frac{n^p(n-p)}{g} \prod_{j=1}^n \operatorname{erf}\left(\tau_j/\sqrt{2}\right). \quad (3.9)$$

In (3.9), the density $f_{\mathbf{b},u}(\mathbf{0}, n-p)$ cannot be readily computed but a saddlepoint approximation can be constructed for it and used in its place to calculate an approximation for $F_M(x)$, as detailed in the next section. This will be called a saddlepoint approximation for $F_M(x)$.

Saddlepoint Density Function

A saddlepoint approximation for the density function of $(p+1)$ -dimensional (\mathbf{b}, u) , as defined in (3.7) and (3.8), is provided in Butler (2007, eqn. 3.2). Let $(p+1)$ -dimensional (\mathbf{s}^T, t) , with $\mathbf{s} = (s_1, \dots, s_p)^T$, index the moment generating function (MGF) of (\mathbf{b}, u) denoted as

$$\mathcal{M}(\mathbf{s}, t) = \mathbb{E}\left(e^{\mathbf{s}^T \mathbf{b} + tu}\right).$$

The corresponding cumulant generating function (CGF) is $\mathcal{K}(\mathbf{s}, t) = \ln \mathcal{M}(\mathbf{s}, t)$. Since the density $f_{\mathbf{b},u}$ is to be evaluated at $(\hat{\mathbf{s}}, \hat{t})$, this determines a joint first-order saddlepoint density for $f_{\mathbf{b},u}(\mathbf{0}, n-p)$ as

$$\hat{f}_1(\mathbf{0}, n-p) := \frac{\exp\{\mathcal{K}(\hat{\mathbf{s}}, \hat{t}) - \hat{t}(n-p)\}}{(2\pi)^{(p+1)/2} |\mathcal{K}''(\hat{\mathbf{s}}, \hat{t})|^{1/2}}, \quad (3.10)$$

where saddlepoint $(\hat{\mathbf{s}}, \hat{t})$ simultaneously solves the $(p+1)$ -dimensional equation

$$\mathcal{K}'_{\mathbf{s}}(\hat{\mathbf{s}}, \hat{t}) = \left. \frac{\partial \mathcal{K}(\mathbf{s}, t)}{\partial \mathbf{s}} \right|_{\mathbf{s}=\hat{\mathbf{s}}, t=\hat{t}} = \mathbf{0} \quad (3.11)$$

$$\mathcal{K}'_t(\hat{\mathbf{s}}, \hat{t}) = \left. \frac{\partial \mathcal{K}(\mathbf{s}, t)}{\partial t} \right|_{\mathbf{s}=\hat{\mathbf{s}}, t=\hat{t}} = n-p \quad (3.12)$$

and matrix $\mathcal{K}''(\hat{\mathbf{s}}, \hat{t})$ is the $(p+1) \times (p+1)$ Hessian matrix of $\mathcal{K}(\mathbf{s}, t)$ evaluated at $(\hat{\mathbf{s}}, \hat{t})$.

Much of the analysis needed to complete the details of this expression is provided in a later section. Section 3.4 gives formulas for $\mathcal{K}(\hat{\mathbf{s}}, \hat{t})$ and its derivatives and shows that $\hat{\mathbf{s}} = \mathbf{0} = (0, \dots, 0)^T$. The remaining saddlepoint component \hat{t} can be found numerically as a 1-dimensional root of (3.12), $\mathcal{K}'_t(\mathbf{0}, \hat{t}) = n-p$, in two steps. Since $\mathcal{K}(\mathbf{0}, \hat{t})$ and its derivatives have a removable dis-

continuity at $\hat{t} = 1/2$, corresponding to the MASR value $x = \sqrt{3}$, the first step in the solution process is to determine on which side of the discontinuity the solution lies, according to x . Here is the decision rule:

$$\hat{t} \begin{matrix} \leq \\ \geq \end{matrix} \frac{1}{2} \iff x \begin{matrix} \geq \\ \leq \end{matrix} \sqrt{3}. \quad (3.13)$$

Let $q_t := 1 - 2t$ and $\hat{q} := 1 - 2\hat{t}$. Then saddlepoint \hat{t} is the solution to

$$\sum_{j=1}^n \frac{1}{\hat{q}} \{1 - R_j(\hat{t})\} = n - p,$$

where

$$R_j(\hat{t}) = \frac{\sqrt{\frac{2\hat{q}}{\pi}} \tau_j \exp\left(\frac{-\hat{q}\tau_j^2}{2}\right)}{\operatorname{erf}\left(\tau_j \sqrt{\frac{\hat{q}}{2}}\right)}.$$

We now present different versions of the saddlepoint approximation: first-order, second-order, and second-order exponential denoted respectively as \hat{f}_1 , \hat{f}_2 , and \hat{f}_{2e} . Upon substitution into (3.9) to approximate $F_M(x)$, a calibration procedure may also be applied, which is indicated by changing the circumflex to a macron. This leads to the possibility of six approximations of $F_M(x)$, denoted by \hat{F}_1 , \bar{F}_1 , \hat{F}_2 , \bar{F}_2 , \hat{F}_{2e} , and \bar{F}_{2e} .

3.2.2 First-Order Approximation

The CGF $\mathcal{K}(\hat{s})$ and its derivatives are listed in Section 3.4, for $\hat{t} \neq 1/2$, corresponding to the almost certain case where $x \neq \sqrt{3}$. From (3.24),

$$\exp\{\mathcal{K}(\mathbf{0}, \hat{t})\} = \prod_{j=1}^n \frac{\operatorname{erf}\left(\tau_j \sqrt{\frac{\hat{q}}{2}}\right)}{\sqrt{\hat{q}} \operatorname{erf}\left(\frac{\tau_j}{\sqrt{2}}\right)}. \quad (3.14)$$

When $\hat{q} < 0$, corresponding to $x < \sqrt{3}$, note that $\operatorname{erf}(\cdot)$ in the numerator has a complex argument.

In the denominator of (3.10), $|\mathcal{K}''(\mathbf{0}, \hat{t})|$ is the determinant of a $(p+1) \times (p+1)$ matrix of second derivatives, some of which are zero. From Section 3.4, $\mathcal{K}''_{s,t}(\mathbf{0}, \hat{t}) = \mathbf{0}$. This separates the

determinant into

$$\begin{aligned}
|\mathcal{K}''(\mathbf{0}, \hat{t})| &= \mathcal{K}''_{t^2}(\mathbf{0}, \hat{t}) \times |\mathcal{K}''_{\text{ss}^T}(\mathbf{0}, \hat{t})| \\
&= \left(\sum_{j=1}^n \frac{R_{2j}(\hat{t})}{\hat{q}^2} \right) \left| \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{R_{1j}(\hat{t})}{\hat{q}} \right|,
\end{aligned} \tag{3.15}$$

where $R_{1j}(\hat{t}) = (1 - R_j(\hat{t}))$, $R_{2j}(\hat{t}) = 2 - (1 + \hat{q}\tau_j^2)R_j(\hat{t}) - R_j^2(\hat{t})$, and $\mathbf{C} = n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \{\mathbf{c}_{ij}\} = (\mathbf{c}_1, \dots, \mathbf{c}_n)$.

Now all the pieces are in place to complete the approximate distribution of MASR M . Substituting the expressions from (3.10) and (3.14) into (3.9) and simplifying gives this result for the first-order approximation:

$$\hat{F}_1(x) = \frac{n^p(n-p) \exp\{-\hat{t}(n-p)\}}{g(2\pi)^{(p+1)/2} \hat{q}^{n/2} |\mathcal{K}''(\hat{\mathbf{s}}, \hat{t})|^{\frac{1}{2}}} \prod_{j=1}^n \text{erf}\left(\tau_j \sqrt{\frac{\hat{q}}{2}}\right), \tag{3.16}$$

where g is defined in (3.6).

3.2.3 Second-Order Approximation

The saddlepoint density is an asymptotic expression, of which the first-order approximation is only the leading term. Including the second term often improves the accuracy of the approximation. From Butler (2007, eqn. 3.37), the second-order saddlepoint density approximation of $f(x)$ is

$$\begin{aligned}
\hat{f}_2(x) &= \hat{f}_1(x) (1 + O) \\
\text{where } O &= \{\hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24\}.
\end{aligned} \tag{3.17}$$

Using the Einstein summation convention as in McCullagh (1987, §1.2), where summation is implied over any subscript appearing once in a subscript and once in a superscript, the three new $\hat{\kappa}$ terms in the O formula are defined as

$$\begin{aligned}
\hat{\kappa}_4 &:= \hat{\mathcal{K}}_{ijkl} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kl} \\
\hat{\kappa}_{23}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ir} \hat{\mathcal{K}}^{js} \hat{\mathcal{K}}^{kt} , \\
\hat{\kappa}_{13}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kr} \hat{\mathcal{K}}^{st}
\end{aligned} \tag{3.18}$$

where the subscripted $\hat{\mathcal{K}}$ terms are derivatives of \mathcal{K} with respect to the subscripted components, evaluated at the saddlepoint $(\mathbf{s}, t) = (\mathbf{0}, \hat{t})$, and the superscripted $\hat{\mathcal{K}}$ terms are components of the inverse of the Hessian matrix $\mathcal{K}''(\mathbf{0}, \hat{t})$. For example, when $i, j, k \leq p$, $\hat{\mathcal{K}}_{ijk} = \partial^3 \mathcal{K}(\mathbf{s}, t) / \partial s_i \partial s_j \partial s_k |_{(\mathbf{s}, t) = (\mathbf{0}, \hat{t})}$. In (3.18), the subscripts $i, j, k, l, r, s, t \in \{1, 2, \dots, p+1\}$, and include t as the $(p+1)^{\text{th}}$ argument of \mathcal{K} .

An alternative second-order exponential approximation, based on McCullagh (1987, § 6.3) is

$$\hat{f}_{2e}(x) = \hat{f}_1(x) e^O, \tag{3.19}$$

where O is the same as in (3.17). In many cases we have evaluated, including the examples shown below, this $2e$ approximation is the best of all.

Putting all this together, the second-order approximate distributions of MASR M are one of the following:

$$\hat{F}_2(x) = \hat{F}_1(x) \{1 + \hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24\} \tag{3.20}$$

$$\hat{F}_{2e}(x) = \hat{F}_1(x) \exp \{ \hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24 \}. \tag{3.21}$$

In some special cases, as in Chapter 2, the $\hat{\kappa}$ functions might be simplified, but in general, the four-way and six-way sums must be calculated and summed term by term. Section 3.4 lists all the third- and fourth-order derivatives of $\mathcal{K}(\mathbf{s}, t)$ required to make these calculations.

3.2.4 Calibration

Calibration is a simple procedure recommended to correct for known inaccuracies with the saddlepoint approximation at the upper end of the support of M . We denote the support of M as the interval $[M_L, M_U]$, where $M_L = 1$ when n is even, or $\{n/(n-1)\}^{1/2}$ when n is odd, and $M_U = (n-p)^{1/2}$. These limits can be derived from the joint distribution for residuals, as derived by Raiffa and Schlaifer (1961, p. 259).

We use only the first- or second-order saddlepoint density approximation, which generally does not integrate to 1. Thus, a calibration procedure is required. There are three options for where to calibrate, specifically, $M_* \in \{M_U, M_2, M_3\}$. As defined in Section 3.5, M_2 and M_3 are the points above which it is not possible to have two or three, respectively, of the absolute residuals $|a_j|$. Once the value of M_* is selected, the calibration formula is

$$\bar{F}_\bullet(x) = \frac{F_M(M_*) \hat{F}_\bullet(x)}{\hat{F}_\bullet(M_*)}, \quad (3.22)$$

where $F_M(M_*)$ is the known exact value of F_M at M_* and $\hat{F}_\bullet(x)$ and $\hat{F}_\bullet(M_*)$ are the (uncalibrated) saddlepoint approximations for F_M at x and M_* , respectively, where the \bullet subscript represents either 1, 2, or 2e. In Chapter 2, for the one-sample case, the calibration formula includes an additional element $\hat{F}_\bullet(M_L)$ at the low end of the support and for very small values of n . Because of its negligible impact in regression examples, we ignore this complication here.

To calibrate at $M_* = M_U = (n-p)^{1/2}$, set $F(M_U) = 1$.

To calibrate at $M_* = M_2$, Section 3.5 shows that

$$M_2 = \sqrt{\frac{n-p}{2} \left(1 + \max_{i < j} |\rho_{ij}|\right)}, \quad (3.23)$$

where $\rho_{ij} = -h_{ij} / \sqrt{(1-h_{ii})(1-h_{jj})}$ is the correlation between residuals a_i and a_j . Since no more than one absolute residual may ever exceed M_2 , the exact value of $F_M(M_2)$ is calculated from the tail probability of the marginal distribution of a single residual and the sample size n :

$$F_M(M_2) = 1 - 2n \left\{ T_{n-p-1} \left(M_2 \sqrt{\frac{n-p-1}{n-p-M_2^2}} \right) \right\},$$

where $T_\nu(\cdot)$ is the right tail probability of the Student's t distribution with ν degrees of freedom.

To calibrate at $M_* = M_3$ is a lot more work. Section 3.5 provides the calculation algorithms for both M_3 and $F_M(M_3)$, using the second Bonferroni bound. In general, this calculation requires many numerical integrations.

As a compromise between accuracy and computational difficulty, we recommend calibrating at $M_* = M_2$, and then using first Bonferroni bound, an exact solution, for any problems where $M \geq M_2$. This is the method used to generate the calibrated p -values in the tables below. Comparing calibrated to uncalibrated values in the tables shows the importance of calibration in providing a usable and reliable discordancy test procedure.

3.3 Results

We present four examples, compared with the same outlier discordancy p -value of 0.1, corresponding to $F_M(x) = 0.9$. The examples are:

- A balanced design with $n = 100$ observations, $p = 10$ treatments, and 10 replications per treatment.
- A subset of the `airquality` dataset distributed with R (2019). The fitted model with $p = 3$ terms, expressed in R, is `Ozone ~ Temp + Wind`, and this model is fitted to the $n = 116$ rows without missing values in any of the three involved variables.
- A 12-run Plackett-Burman (1946) design, with 1 replication and $p = 8$ terms in the model, including a mean and 7 main effects. This example is notable because of the failure of the Grubbs' test. At least two of the residuals have a correlation $\rho_{ij} = 1$, with the consequence that more than one residual may assume any value up to the maximum value, with positive probability.

Table 3.1: Saddlepoint and Bonferroni p -values for Four Regression Examples

	10 × 10 balanced	airquality	PB 12-run	DSD 17-run
n	100	116	12	51
p	10	3	8	9
M value x	3.213	3.445	1.904	2.971
Simulated p -value	0.100	0.100	0.100	0.100
Saddlepoint approx.:				
$1 - \hat{F}_1(x)$	-0.229	0.068	-3.603	-0.506
$1 - \hat{F}_2(x)$	0.140	0.100	3.382	0.222
$1 - \hat{F}_{2e}(x)$	0.090	0.100	-0.009	0.071
$1 - \bar{F}_1(x)$	0.101	0.099	0.236	0.108
$1 - \bar{F}_2(x)$	0.099	0.100	0.262	0.099
$1 - \bar{F}_{2e}(x)$	0.100	0.100	0.222	0.110
Bonferroni bounds:				
1 st upper (Grubbs)	0.102	0.103	0.149	0.101
1 st upper (Worsley)	0.102	0.103	0.100	0.101
2 nd lower	0.100	0.100	0.100	0.100
Limits on M:				
M_U	9.487	10.63	2.000	6.481
M_2	7.071	7.843	2.000	5.180
M_3	5.669	6.376	1.732	4.061

- A definitive screening design (DSD) after Jones and Nachtsheim (2011), with 17 runs and 3 replications, for $n = 51$. This \mathbf{X} matrix comes from the DSD example packaged with Minitab 18 (2017), and the fitted model includes $p = 9$ terms for the mean, A, C, D, E, G, GG, AE, and CE.

Table 3.1 lists p -values for all versions of saddlepoint and Bonferroni methods discussed in this article. Also included is a simulated p -value, based on sufficient trials that the width of a 99% confidence interval is less than the least-significant digit listed. Determining these p -values required at least 2.4 million trials for each case.

Note that in three of the four examples, the first Bonferroni upper bound, also known as Grubbs' test, performs well at the upper end of the distribution of M , where p -values are small. For the PB 12-run design, the Grubbs p -value is substantially too high. The Worsley improvement to the upper bound, or the lower bound, are significantly more accurate for this case, although each

requires more computational time. The calibrated saddlepoint approximations also perform well, when sample size is larger. Saddlepoint approximations generally work best at larger sample sizes. For the same PB 12-run example, with only $n - p = 4$ degrees of freedom, none of the saddlepoint approximations return accurate p -values. In general, when degrees of freedom are few, neither the first Bonferroni bound (Grubbs test) nor the saddlepoint approximations work particularly well. The two improvements to the Bonferroni bound evaluated here, the Worsley improvement using spanning trees, or the second Bonferroni bound, both show improved performance at the upper end of the support of M , at the expense of additional computation.

For the four examples discussed here, Figures 3.1 through 3.4 show all approximations of $1 - F_M(x)$ for 100 values of x in the support of M . Each of these four-panel graphs is organized in a similar way. The top left panel shows all saddlepoint approximations of $1 - F_M(x)$, plus a Monte Carlo simulation. The top right panel shows all Bonferroni bounds on $1 - F_M(x)$, plus a Monte Carlo simulation. The bottom two panels show errors by subtracting either the Bonferroni lower bound, when $x \geq M_3$, or the Monte Carlo simulation otherwise, from each approximation in the top two panels.

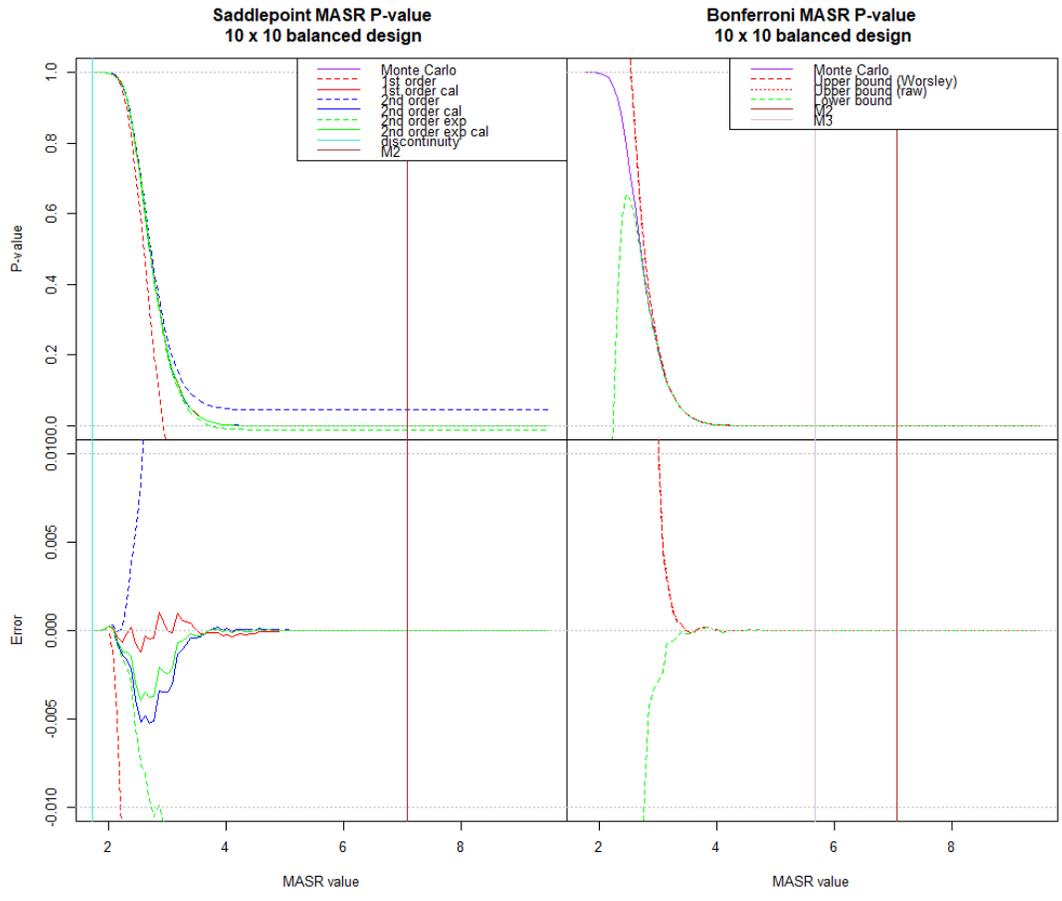


Figure 3.1: Comparison of p -values for a 10×10 Balanced Design

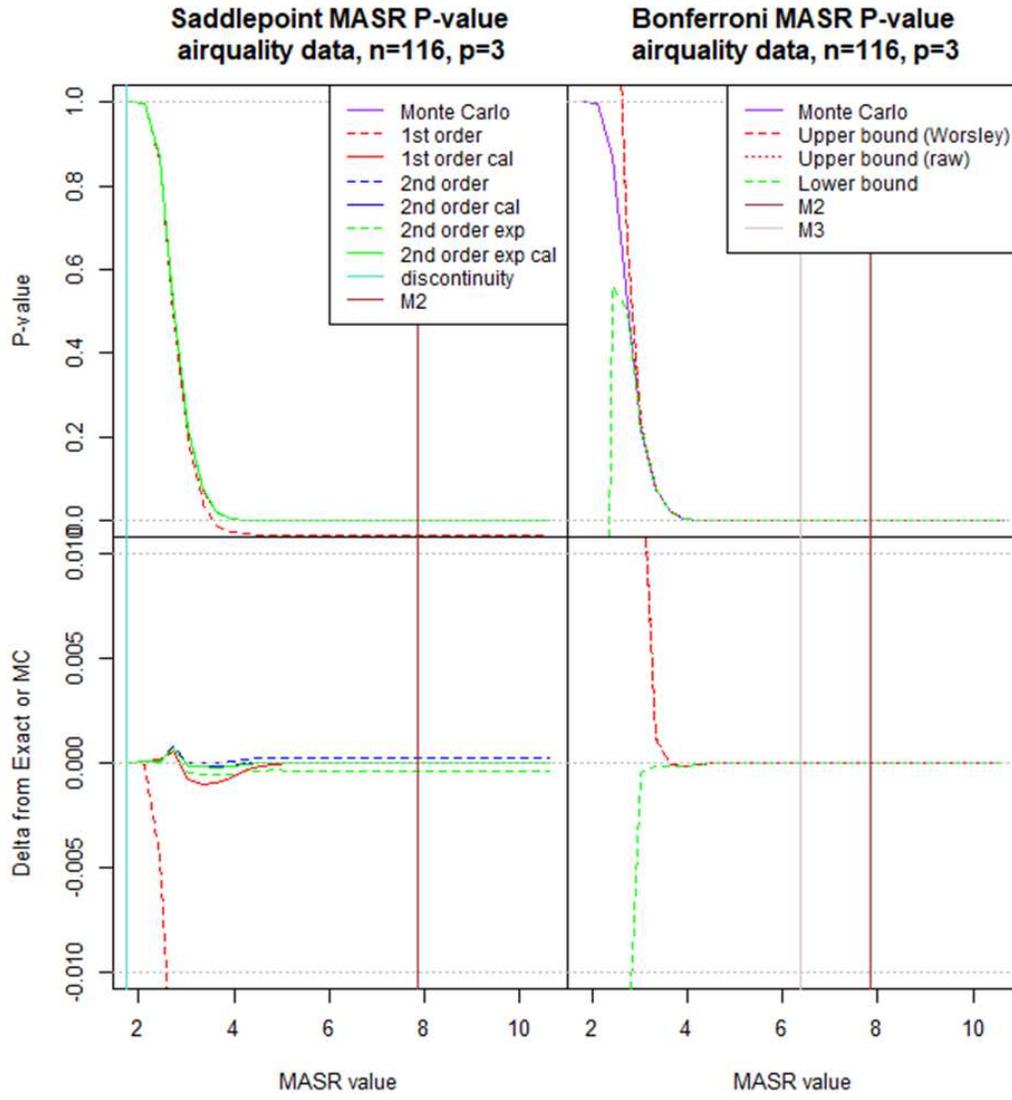


Figure 3.2: Comparison of p -values for airquality Dataset

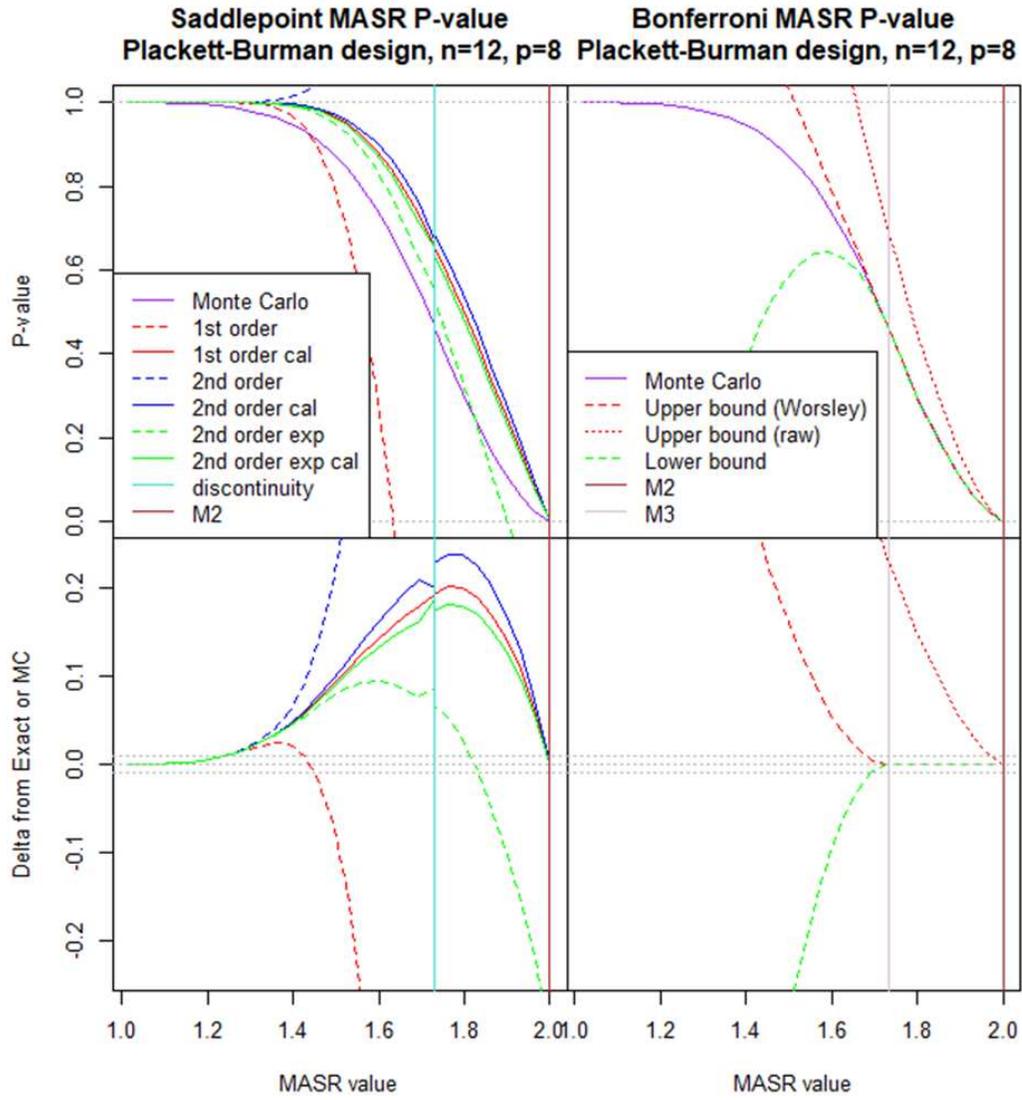


Figure 3.3: Comparison of p -values for a Plackett-Burman 12-run Design

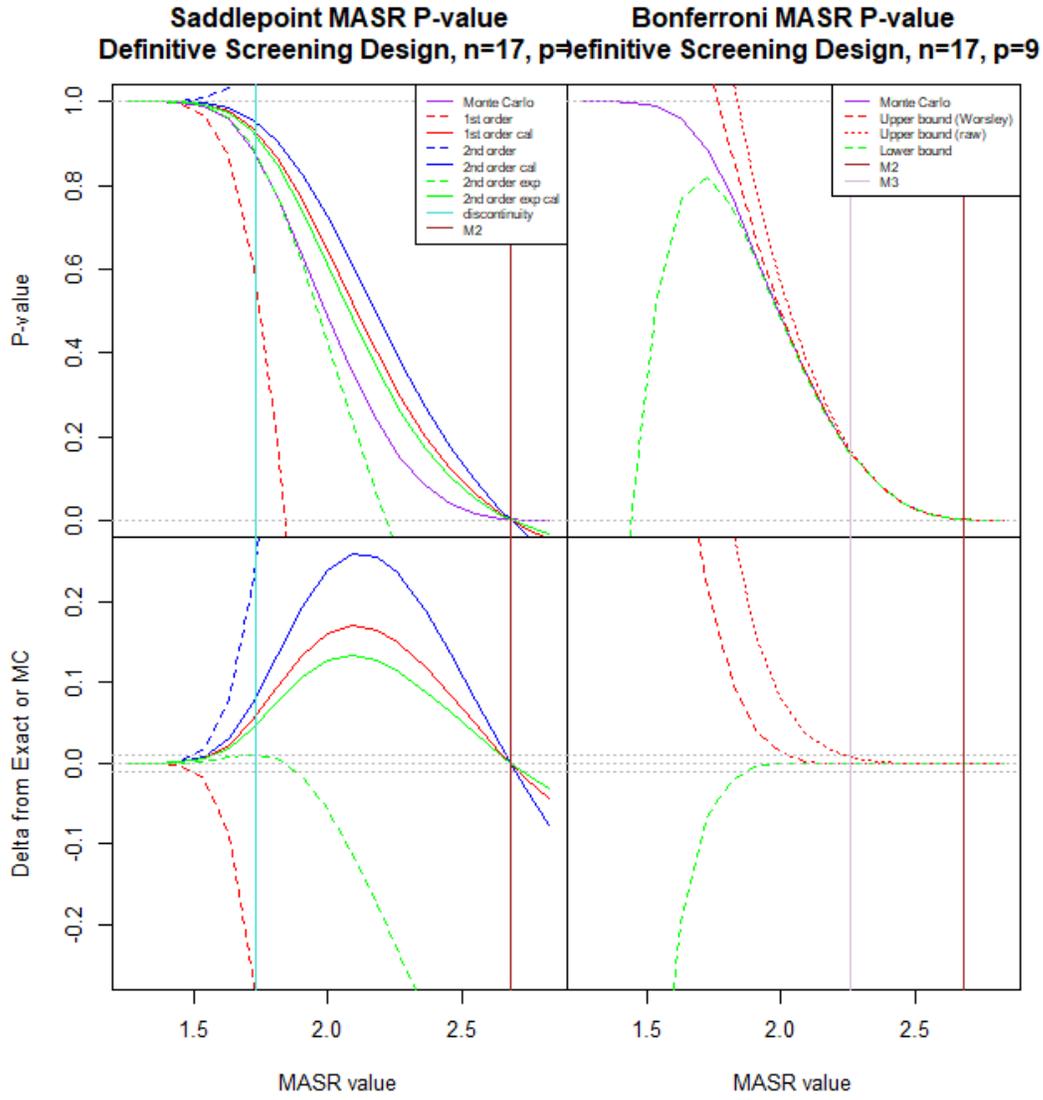


Figure 3.4: Comparison of p -values for a Definitive Screening Design

3.4 Cumulant Generating Function $\mathcal{K}(\mathbf{s}, t)$ and its Derivatives

In this section, the joint CGF $\mathcal{K}(\mathbf{s}, t)$ of $\mathbf{b} = n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z}$ and $u = \mathbf{z}^T \mathbf{z}$ is developed into a form for calculation at the saddlepoint $(\hat{\mathbf{s}}, \hat{t})$, along with its derivatives required for the saddlepoint approximations. Let

$$\mathbf{C} := n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = (\mathbf{c}_1, \dots, \mathbf{c}_n) = (c_{ij}),$$

so that $\mathbf{b} = \mathbf{Cz}$. Also, with $q_t = 1 - 2t$, define the function

$$\begin{aligned} \mathcal{M}_j(v, t) &= \mathbb{E} \left\{ \exp(vz_j + tz_j^2) \right\} \\ &= \frac{1}{\sqrt{2\pi} \operatorname{erf}(\tau_j/\sqrt{2})} \int_{-\tau_j}^{\tau_j} \exp\left(vz - \frac{q_t}{2}z^2\right) dz. \end{aligned}$$

After completing the square separately for $q_t \geq 0$, let $\varsigma_t = \operatorname{sig}(q_t) := q_t/|q_t|$, so that

$$\mathcal{M}_j(v, t) = \frac{\exp(v^2/2q_t)}{\sqrt{2\pi} \operatorname{erf}(\tau_j/\sqrt{2})} \int_{-\tau_j}^{\tau_j} \exp\left\{-\varsigma_t \left(\sqrt{\frac{|q_t|}{2}}z - \frac{v}{\sqrt{2|q_t|}}\right)^2\right\} dz,$$

for $q_t \neq 0$. Now we change the variable of integration and substitute $\operatorname{erf}(x) := (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$, which will have an imaginary argument when $q_t < 0$, so that

$$\begin{aligned} \mathcal{M}_j(v, t) &= \frac{\exp(v^2/2q_t)}{2\sqrt{|q_t|} \operatorname{erf}(\tau_j/\sqrt{2})} \times \\ &\quad \left[\operatorname{erf}\left\{\sqrt{\varsigma_t} \left(\sqrt{\frac{|q_t|}{2}}\tau_j - \frac{v}{\sqrt{2|q_t|}}\right)\right\} - \operatorname{erf}\left\{\sqrt{\varsigma_t} \left(-\sqrt{\frac{|q_t|}{2}}\tau_j - \frac{v}{\sqrt{2|q_t|}}\right)\right\} \right]. \end{aligned}$$

When $q_t = 0$, at the boundary between real and complex arithmetic, \mathcal{M}_j has a removable singularity, as do all other formulas in this section. Relevant limiting forms are presented in Section 3.4.1.

The joint MGF of (\mathbf{b}, u) is

$$\begin{aligned}\mathcal{M}(\mathbf{s}, t) &= \mathbb{E} \left\{ \exp \left(\mathbf{s}^T \mathbf{C} \mathbf{z} + t \|\mathbf{z}\|^2 \right) \right\} \\ &= \mathbb{E} \left\{ \exp \left(\sum_j \left\{ (\mathbf{s}^T \mathbf{c}_j) z_j + t z_j^2 \right\} \right) \right\} = \prod_{j=1}^n \mathcal{M}_j(\mathbf{s}^T \mathbf{c}_j, t)\end{aligned}$$

and is convergent for all $(\mathbf{s}, t) \in \mathbb{R}^{p+1}$. Therefore, $\mathcal{K}(\mathbf{s}, t) = \sum_{j=1}^n \ln \mathcal{M}_j(\mathbf{s}^T \mathbf{c}_j, t)$.

Note that since $\mathbf{z} \in \otimes_{j=1}^n (-\tau_j, \tau_j)$, then \mathbf{b} is in a bounded subset of the column space of \mathbf{C} .

Also,

$$u = \|\mathbf{z}\|^2 < \sum_j \tau_j^2 = x^2 \sum_j (1 - h_{jj}) = x^2(n - p)$$

and $u \in (0, x^2(n - p))$. Thus the support is bounded in \mathbb{R}^{p+1} .

To show that $\hat{\mathbf{s}} = \mathbf{0}$, note

$$\begin{aligned}\mathbf{0} = \mathcal{K}'_{\mathbf{s}}(\hat{\mathbf{s}}, \hat{t}) &= \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\hat{\mathbf{s}}^T \mathbf{c}_j, \hat{t})} \frac{\partial \mathcal{M}_j(\mathbf{s}^T \mathbf{c}_j, t)}{\partial \mathbf{s}} \Big|_{\mathbf{s}=\hat{\mathbf{s}}} = \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\hat{\mathbf{s}}^T \mathbf{c}_j, \hat{t})} \frac{\partial \mathcal{M}_j(v, \hat{t})}{\partial v} \Big|_{v=\hat{\mathbf{s}}^T \mathbf{c}_j} \mathbf{c}_j \\ &= \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\hat{\mathbf{s}}^T \mathbf{c}_j, \hat{t})} \mathbf{c}_j \mathbb{E} \left[z_j \exp \left\{ (\hat{\mathbf{s}}^T \mathbf{c}_j) z_j + \hat{t} z_j^2 \right\} \right].\end{aligned}$$

For any value of \hat{t} , $\hat{\mathbf{s}} = \mathbf{0}$ solves the equation since $E[z_j \exp\{\hat{t} z_j^2\}] = 0$ for all j . If, for $\hat{\mathbf{s}} = \mathbf{0}$, a solution \hat{t} can be determined for $\mathcal{K}'_t(\mathbf{0}, \hat{t}) = n - p$, then $(\mathbf{0}, \hat{t})$ must be the saddlepoint, since there is only one solution to both equations. (As $t \uparrow \infty$ then $\mathcal{K}'_t(\mathbf{0}, t)$ must exceed $n - p$). Saddlepoint $\hat{\mathbf{s}} = \mathbf{0}$ induces symmetry into the evaluation of \mathcal{K} and its derivatives at saddlepoint $(\mathbf{0}, \hat{t})$. At the saddlepoint, $\exp\{\mathcal{K}(\mathbf{0}, \hat{t})\}$, required by (3.10), simplifies to

$$\exp\{\mathcal{K}(\mathbf{0}, \hat{t})\} = \prod_{j=1}^n \mathcal{M}_j(\mathbf{0}, \hat{t}) = \prod_{j=1}^n \frac{\operatorname{erf}\left(\tau_j \sqrt{\frac{\hat{q}}{2}}\right)}{\sqrt{\hat{q}} \operatorname{erf}\left(\frac{\tau_j}{\sqrt{2}}\right)}, \quad (3.24)$$

where $\hat{q} = q_{\hat{t}} = 1 - 2\hat{t}$.

The saddlepoint \hat{t} is the solution to

$$\begin{aligned}
n - p &= \mathcal{K}'_t(\mathbf{0}, \hat{t}) = \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\mathbf{0}, \hat{t})} \frac{\partial \mathcal{M}_j(\mathbf{0}, t)}{\partial t} \Big|_{t=\hat{t}} = \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\mathbf{0}, \hat{t})} \mathbb{E} \{ z_j^2 \exp(\hat{t} z_j^2) \} \\
&= \sum_{j=1}^n \frac{1}{\hat{q}} \left(1 - \frac{\sqrt{\frac{2\hat{q}}{\pi}} \tau_j e^{-\frac{1}{2}\hat{q}\tau_j^2}}{\operatorname{erf}\left(\tau_j \sqrt{\frac{\hat{q}}{2}}\right)} \right) = \sum_{j=1}^n \frac{1}{\hat{q}} \{1 - R_j(\hat{t})\}, \tag{3.25}
\end{aligned}$$

where $R_j(\hat{t})$ is given in (3.26).

The derivatives of $\mathcal{K}(\mathbf{s}, t)$ presented below are shown only for $\mathbf{s} = \hat{\mathbf{s}} = \mathbf{0}$.

First derivatives:

$$\begin{aligned}
\hat{\mathcal{K}}'_s(\mathbf{0}, t) &= \frac{\partial}{\partial \mathbf{s}} \mathcal{K}(\mathbf{s}, t) \Big|_{\mathbf{s}=\mathbf{0}} = \mathbf{0} \quad \forall t \\
\hat{\mathcal{K}}'_t(\mathbf{0}, \hat{t}) &= \frac{\partial}{\partial t} \mathcal{K}(\mathbf{s}, t) \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \frac{R_{1j}(\hat{t})}{\hat{q}}
\end{aligned}$$

where

$$\begin{aligned}
R_{1j}(\hat{t}) &:= 1 - R_j(\hat{t}) \\
R_j(\hat{t}) &:= \frac{\sqrt{\frac{2\hat{q}}{\pi}} \tau_j \exp\left(-\frac{\tau_j^2 \hat{q}}{2}\right)}{\operatorname{erf}\left(\tau_j \sqrt{\frac{\hat{q}}{2}}\right)} \tag{3.26}
\end{aligned}$$

Second derivatives:

$$\begin{aligned}
\hat{\mathcal{K}}''_{\mathbf{s}\mathbf{s}^T}(\mathbf{0}, \hat{t}) &= \frac{\partial^2}{\partial \mathbf{s} \partial \mathbf{s}^T} \mathcal{K}(\mathbf{s}, t) \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{R_{1j}(\hat{t})}{\hat{q}} \\
\hat{\mathcal{K}}''_{\mathbf{s}, t}(\mathbf{0}, \hat{t}) &= \frac{\partial^2 \mathcal{K}(\mathbf{s}, t)}{\partial \mathbf{s} \partial t} \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \frac{\partial \mathcal{K}'_s(\mathbf{0}, t)}{\partial t} = \mathbf{0} \quad \forall t \\
\hat{\mathcal{K}}''_{t^2}(\mathbf{0}, \hat{t}) &= \frac{\partial^2 \mathcal{K}(\mathbf{s}, t)}{\partial t^2} \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \frac{R_{2j}(\hat{t})}{\hat{q}^2}
\end{aligned}$$

where

$$\hat{R}_{2j}(\hat{t}) := 2 - (1 + \hat{q}\tau_j^2)\hat{R}_j - \hat{R}_j^2$$

Third derivatives: Let $\hat{\mathcal{K}}_{klm}^{(3)} = \partial^3 \mathcal{K} / \partial s_k \partial s_l \partial s_m |_{\mathbf{s}=\mathbf{0}, t=\hat{t}}$, for $k, l, m \leq p$. Then

$$\begin{aligned}\hat{\mathcal{K}}_{klm}^{(3)}(\mathbf{0}, \hat{t}) &= 0 \quad \forall t \\ \hat{\mathcal{K}}_{\mathbf{ss}^T, t}^{(3)}(\mathbf{0}, \hat{t}) &= \frac{\partial^3}{\partial \mathbf{s} \partial \mathbf{s}^T \partial t} \mathcal{K}(\mathbf{s}, t) \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{R_{2j}(\hat{t})}{\hat{q}^2} \\ \hat{\mathcal{K}}_{\mathbf{s}, t^2}^{(3)}(\mathbf{0}, \hat{t}) &= \frac{\partial^3 \mathcal{K}(\mathbf{s}, t)}{\partial \mathbf{s} \partial t^2} \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \mathbf{0} \quad \forall t \\ \hat{\mathcal{K}}_{t^3}^{(3)}(\mathbf{0}, \hat{t}) &= \frac{\partial^3 \mathcal{K}(\mathbf{s}, t)}{\partial t^3} \Big|_{\mathbf{s}=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \frac{R_{3j}(\hat{t})}{\hat{q}^3}\end{aligned}$$

where

$$R_{3j}(\hat{t}) := 8 - (\hat{q}^2 \tau_j^4 + 2\hat{q} \tau_j^2 + 3) R_j - 3(1 + \hat{q} \tau_j^2) R_j^2 - 2R_j^3$$

Fourth derivatives: Let $\hat{\mathcal{K}}_{klmr}^{(4)} = \partial^4 \mathcal{K} / \partial s_k \partial s_l \partial s_m \partial s_r |_{\mathbf{s}=\mathbf{0}, t=\hat{t}}$, for $k, l, m, r \leq p$. Then

$$\begin{aligned}\hat{\mathcal{K}}_{klmr}^{(4)}(\mathbf{0}, \hat{t}) &= \sum_{j=1}^n \frac{\{R_{2j}(\hat{t}) - 2R_{1j}^2(\hat{t})\}}{\hat{q}^2} c_{kj} c_{lj} c_{mj} c_{rj} \\ \hat{\mathcal{K}}_{klm, t}^{(4)}(\mathbf{0}, \hat{t}) &= \frac{\partial^4 \mathcal{K}(\mathbf{s}, t)}{\partial s_k \partial s_l \partial s_m \partial t} \Big|_{\mathbf{s}^*=\mathbf{0}} = 0 \quad \forall t \text{ and } k, l, m \in \{1, \dots, p\} \\ \hat{\mathcal{K}}_{\mathbf{ss}^T, t^2}^{(4)}(\mathbf{0}, \hat{t}) &= \frac{\partial^4 \mathcal{K}(\mathbf{s}, t)}{\partial \mathbf{s} \partial \mathbf{s}^T \partial t^2} \Big|_{\mathbf{s}^*=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{R_{3j}(\hat{t})}{\hat{q}^3} \\ \hat{\mathcal{K}}_{\mathbf{s}, t^3}^{(4)}(\mathbf{0}, \hat{t}) &= \frac{\partial^4 \mathcal{K}(\mathbf{s}, t)}{\partial \mathbf{s} \partial t^3} \Big|_{\mathbf{s}=\mathbf{0}} = \mathbf{0} \quad \forall t \\ \hat{\mathcal{K}}_{t^4}^{(4)}(\mathbf{0}, \hat{t}) &= \frac{\partial^4 \mathcal{K}(\mathbf{s}, t)}{\partial t^4} \Big|_{\mathbf{s}^*=\mathbf{0}, t=\hat{t}} = \sum_{j=1}^n \frac{R_{4j}(\hat{t})}{\hat{q}^4}\end{aligned}$$

where

$$\begin{aligned}R_{4j}(\hat{t}) &:= 48 - (\hat{q}^3 \tau_j^6 + 3\hat{q}^2 \tau_j^4 + 9\hat{q} \tau_j^2 + 15) R_j \\ &\quad - (7\hat{q}^2 \tau_j^4 + 14\hat{q} \tau_j^2 + 15) R_j^2 - 12(1 + \hat{q} \tau_j^2) R_j^3 - 6R_j^4\end{aligned}$$

3.4.1 Limiting Forms

As noted, the CGF $\mathcal{K}(\mathbf{s}, t)$, its derivatives, and all saddlepoint distributions for M have a removable discontinuity at $t = 0.5$, corresponding to the \bar{M} value $x = \sqrt{3}$. Although this is a set of probability 0, with theoretically no significance, it is important in programming these methods. When calculating approximations of F , using typical double-precision arithmetic, there is a neighborhood of the discontinuity within which these calculations become unstable. In this neighborhood, it is advisable to switch to a Taylor series approximation, built using these limiting forms, to present a well-behaved function for the user. Here are the limiting values of $\mathcal{K}(\mathbf{0}, 1/2)$, and its non-zero derivatives.

$$\exp \left\{ \mathcal{K} \left(\mathbf{0}, \frac{1}{2} \right) \right\} = \prod_{j=1}^n \frac{2\tau_j}{\sqrt{2\pi} \operatorname{erf}(\tau_j/\sqrt{2})}$$

$$\mathcal{K}'_t \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n \frac{\tau_j^2}{3}$$

$$\mathcal{K}''_{\mathbf{ss}^T} \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{\tau_j^2}{3}$$

$$\mathcal{K}''_{t^2} \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n \frac{4\tau_j^4}{45}$$

$$\mathcal{K}^{(3)}_{\mathbf{ss}^T, t} \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{4\tau_j^4}{45}$$

$$\mathcal{K}^{(3)}_{t^3} \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n \frac{16\tau_j^6}{945}$$

$$\mathcal{K}^{(4)}_{klmr} \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n -c_{kj} c_{lj} c_{mj} c_{rj} \frac{2\tau_j^4}{15}$$

$$\mathcal{K}^{(4)}_{\mathbf{ss}^T, t^2} \left(\mathbf{0}, \frac{1}{2} \right) = \sum_{j=1}^n \mathbf{c}_j \mathbf{c}_j^T \frac{16\tau_j^6}{945}$$

$$\mathcal{K}_{t^4}^{(4)}\left(\mathbf{0}, \frac{1}{2}\right) = \sum_{j=1}^n \frac{-32\tau_j^8}{4725}$$

3.5 Bonferroni Bounds

Bonferroni bounds and their application to outlier discordancy testing were explained in section 2.5. Only the details that are different for the regression setting are listed here.

3.5.1 Bonferroni Inequalities

In the regression setting of this chapter, $E_j = \{|a_j| \leq x\} = \{|y_j| \leq x\sqrt{1-h_{jj}}\}$, so $\{M \leq x\} = \bigcap_{j=1}^n E_j$. Reexpressing $F_M(x)$ in terms of a union of events, $1 - F_M(x) = \Pr(M > x) = \Pr\left\{\bigcup_{j=1}^n \overline{E_j}\right\}$. With this definition of events E_j , refer to section 2.5.1 for the remaining details.

3.5.2 Distribution Theory

The theory behind the computation of ρ_{ij} , M_2 , M_3 , S_1 , S_2 , and S_{2*} derives from the joint distribution of the studentized residuals. Butler (1984) presented this as Lemma 3.1, in a slightly different setting. For more generality, Butler uses a \mathbf{Z} matrix, here set to the identity matrix for outlier identification; Butler also scales the \mathbf{r} residual vector to a maximum value of 1. To adapt to our setting, let $\mathbf{r} = \mathbf{a}(n-p)^{-1/2}$.

Let $\mathcal{S} = \{i_1, \dots, i_q\}$ be any q -tuple subset of $\{1, \dots, n\}$. Let $\mathbf{R}_{\mathcal{S}}$ be the $q \times q$ correlation matrix of $(r_j : j \in \mathcal{S})$, namely $\mathbf{R}_{\mathcal{S}} = (\rho_{ij})$, where $\rho_{ij} = -h_{ij}/\sqrt{(1-h_{ii})(1-h_{jj})}$. The q -variate density of $\mathbf{r}_{\mathcal{S}}$ is

$$f(\mathbf{r}) = \frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n-p-q}{2}\right)} \pi^{-\frac{q}{2}} |\mathbf{R}_{\mathcal{S}}|^{-\frac{1}{2}} (1 - \mathbf{r}^T \mathbf{R}_{\mathcal{S}}^{-1} \mathbf{r})^{\frac{n-p-q-2}{2}} \quad \text{if } \mathbf{r}^T \mathbf{R}_{\mathcal{S}}^{-1} \mathbf{r} \leq 1 \quad (3.27)$$

To calculate M_2 , Butler provides the following formulas in Lemma 4.1, here including the $\sqrt{n-p}$ adjustment to this setting.

$$\Pr(|a_i| > x, |a_j| > x) = 0 \text{ iff } x \geq \sqrt{\frac{(n-p)(1+|\rho_{ij}|)}{2}}.$$

From this, it directly follows that:

$$M_2 = \sqrt{\frac{(n-p)(1+\max_{i<j} |\rho_{ij}|)}{2}}.$$

In the case where $|\rho_{ij}| = 1$ for any pair of residuals a_i and a_j , indicating perfect correlation, observe that $M_2 = \sqrt{n-p}$, equaling the upper end of the support of M . In effect, this perfect correlation means there is no region of positive probability in which the Bonferroni upper bound is exact. In the test data sets evaluated for this work, some experiments, such as a 12-run Plackett-Burman design with more than 6 factors, exhibit this pattern for some but not all pairs of residuals.

The calculation of M_3 depends on the bounds of the ellipsoid defined by $\mathbf{r}^T \mathbf{R}_S^{-1} \mathbf{r} = 1$, the boundary of the density function in (3.27). The calculation procedure is based on Lemma 4.2 of Butler (1984), with $q = 3$.

The sum $S_1 = \sum_{j=1}^n \Pr(|a_j| > x)$ is calculated by using the marginal distribution from (3.27), which has been shown by Grubbs (1950) and others to be related to Student's t distribution through this identity:

$$a_j \sqrt{\frac{n-p-1}{n-p-a_j^2}} \sim t_{n-p-1} \quad (3.28)$$

Therefore, in terms of the t distribution,

$$S_1 = 2n \left\{ T_{n-p-1} \left(x \sqrt{\frac{n-p-1}{n-p-x^2}} \right) \right\},$$

where $T_\nu(\cdot)$ is the right tail probability of the Student's t distribution with ν degrees of freedom.

The sums S_2 and S_{2*} comprise probabilities of intersections $\Pr[|a_i| > x, |a_j| > x]$. Calculating each intersection probability is an integration of the bivariate marginal density of $\mathbf{r} = (r_i r_j)^T$, given as:

$$f(\mathbf{r}) = \frac{n-p-2}{2\pi\sqrt{1-\rho_{ij}^2}} (1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r})^{\frac{n-p-4}{2}} \quad \text{if } \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \leq 1 \quad (3.29)$$

where

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix}.$$

Calculation of $\Pr(|a_i| > x, |a_j| > x) = \Pr(|r_i| > c, |r_j| > c)$, where $c = x(n-p)^{-1/2}$, requires bivariate integration over four regions of the ellipse defined by $\mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \leq 1$. By symmetry, this can be reduced from four to two regions. For each of the two regions, using the analysis of Butler (1982), we simplify the bivariate into univariate integration, computed using the `integrate` function of R.

Two special cases requires special handling in the algorithm. First, for any pair of residuals a_i and a_j with $|\rho_{ij}| = 1$, the bivariate distribution of a_i and a_j degenerates into univariate. In this case, the intersection probability is calculated this way:

$$\Pr[|a_i| > x, |a_j| > x] = 2 \left[T_{n-p-1} \left(x \sqrt{\frac{n-p-1}{n-p-x^2}} \right) \right] \quad \text{when } |\rho_{ij}| = 1. \quad (3.30)$$

The second special case is when $n - p = 2$. In this minimal case, the joint density of two residuals becomes degenerate, so again we use (3.30).

With all $\binom{n}{2}$ intersection probabilities calculated, S_2 is their sum. S_{2*} requires selecting a spanning tree of the non-zero intersection probabilities that maximizes the summed probability. The Kruskal (1956) algorithm, modified to find a maximal spanning tree, works for the task, although certainly faster algorithms are available.

Chapter 4

Outlier Discordancy in Multivariate Regression

In the multivariate regression setting, the optimal test for a single outlying observation is provided by testing for large values of the maximum squared studentized residual (MSSR) M . Since the exact distribution of M is intractable, we present approximations to the distribution based on a saddlepoint approximation. These are compared to first and second Bonferroni bounds and to a Monte Carlo simulation.

4.1 Setting

In this section, we define the setting for the outlier discordancy test and discuss the optimality of the test. There is a significant difference in notation between earlier chapters and this one. In Chapters 2 and 3, M is MASR. Here, M is MSSR, the square of MASR. Most multivariate literature focuses on squared distance metrics such as Mahalanobis distance, so in this chapter, we are consistent with that convention.

4.1.1 Null model setting

The null model assumes that there are no outliers in the data. The data derive from a standard multivariate linear model

$$\mathbf{Y}_{(n \times m)} = \mathbf{X}_{(n \times p)} \mathbf{B}_{(p \times m)} + \mathbf{E}_{(n \times m)}, \quad (4.1)$$

in which $\mathbf{Y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$ is the matrix of observed values, $\mathbf{X} = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)^T$ is the design matrix of full rank $p < n - m$ in which the first column is a vector of ones and the remaining columns may be fixed or random, and \mathbf{B} is an unknown parameter matrix, which may comprise both fixed and random effects. The studentized residuals, defined shortly, are ancillary to any potential random effect parameters and hence are independent of the corresponding random effect distributions. Hence, these results apply to fixed, random, or mixed models.

The random error matrix \mathbf{E} in (3.1) is assumed to be left-orthogonally invariant (l.o.i.) with uncorrelated rows, as discussed in chapter 3. It suffices this chapter to assume that $\text{vec}(\mathbf{E}) \sim \mathbb{N}_{nm}(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$, where $\text{vec}(\mathbf{E})$ is the vector of stacked columns of \mathbf{E} , that is, the rows of \mathbf{E} are i.i.d. $\mathbb{N}_m(\mathbf{0}, \Sigma)$, since the same results apply to any model (4.1), in which \mathbf{E} is a l.o.i. distribution.

We define y_j to be an outlier when the following alternative model holds in which

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{e}_j\theta^T + \mathbf{E}, \quad (4.2)$$

when $\theta \in \mathfrak{R}^m$, $\theta \neq \mathbf{0}$, and \mathbf{e}_j is the j^{th} column of \mathbf{I}_n , for $j = 1, \dots, n$.

Define $\Sigma := \mathbb{E}(\mathbf{E}^T\mathbf{E})/n$. Denote the hat matrix as $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = (h_{ij})$ and the residual projection matrix as $\mathbf{I} - \mathbf{H}$. The usual parameter estimates are $\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ and $\hat{\Sigma} = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}/(n - p) = \hat{\mathbf{E}}^T\hat{\mathbf{E}}/(n - p)$, where $\hat{\mathbf{E}} = (\mathbf{I} - \mathbf{H})\mathbf{Y} = (\hat{\varepsilon}_1^T, \dots, \hat{\varepsilon}_n^T)^T$ and $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$. A vector of (internally) studentized residuals is $\mathbf{a} = (a_1, \dots, a_n)^T$, where

$$a_j^2 = \frac{\hat{\varepsilon}_j \hat{\Sigma}^{-1} \hat{\varepsilon}_j^T}{1 - h_{jj}} = \frac{(\mathbf{y}_j - \mathbf{x}_j \hat{\mathbf{B}}) \hat{\Sigma}^{-1} (\mathbf{y}_j - \mathbf{x}_j \hat{\mathbf{B}})^T}{1 - h_{jj}} \quad j = 1, \dots, n. \quad (4.3)$$

The maximum squared studentized residual (MSSR) is the statistic $M = \max_j a_j^2$ whose cumulative distribution function (CDF) under the null model is denoted as $F(x) = \Pr(M \leq x)$. A typical test for the discordancy of a single outlier caused by a change in distribution for one row of \mathbf{E} rejects the null model when M is large. Thus, if x is the observed value of M , then the p -value is $1 - F(x)$. The exact distribution F is intractable.

Alternatively, upper and lower Bonferroni bounds are available. The upper Bonferroni bound on the p -value has become known as Grubbs' test in univariate applications. Unlike the univariate settings of earlier chapters, there is no range of values for which Grubbs' test provides an exact p -value, as shown by Wilks (1963). The details of these bounds are presented in Section 4.6. This chapter will show where the Bonferroni bounds are adequate or inadequate and show the opportunity for an improved method to be provided by saddlepoint approximations.

4.1.2 Optimality

The optimality of this outlier test occurs when the distribution of \mathbf{E} is left orthogonally invariant and m -extendable with the latter property now defined. Error matrix \mathbf{E} is said to be m -extendable if \mathbf{E} is the marginal distribution of some $(n + m) \times m$ random matrix which is also l.o.i., using a $(n + m) \times (n + m)$ orthogonal transformation. This condition is still satisfied by normal, spherically symmetric matrix-variate t , and many other error distributions. Butler (1986) proved that Hotelling's T^2 test in MANOVA is uniformly most powerful invariant (UMPI) amongst all such m -extendable l.o.i. error distributions.

Butler (1986) also showed that rejecting the null hypothesis of no outliers when $M = \max_j a_j^2$ is large is the optimal test for a single outlying observation within the same class of error distributions. The context is that of a multi-decision problem with H_0 : no outliers vs. H_j : the j^{th} row of \mathbf{Y} has slipped by $\theta \in \mathfrak{R}^m$, with ϕ_j denoting the probability of selecting H_j , for $j = 0, 1, \dots, n$. Let the loss in deciding H_i when H_j is true be the indicator function of $i \neq j$, so the Bayes risk, with respect to a prior weighting of the hypotheses is the probability of an incorrect decision. Subject to group invariant conditions on the tests allowed, Butler (1986, thm. 4.1) shows that the test defined by $\phi_{j^*} = 1$ if $a_{j^*}^2 = M > c_1$, and $\phi_0 = 1$ otherwise is the Bayes decision rule that minimizes the probability of an incorrect decision uniformly in $\theta \in \mathfrak{R}$ as given by the Bayes risk. The optimality applies amongst tests restricted as follows: (i) invariant to translation of \mathbf{Y} by any $n \times m$ matrix whose columns are in the column space of \mathbf{X} , (ii) invariant to rescaling of \mathbf{Y} by post-multiplication with any $m \times m$ nonsingular matrix, (iii) with level α , $0 \leq \alpha \leq 1$, so that $E(\sum_{j=1}^n \phi_j | H_0) = 1 - E(\phi_0 | H_0) \leq \alpha$, and (iv) based upon an equal prior weighting of $\pi \in (0, n^{-1})$ given to H_1, \dots, H_n , where π ensures level α in (iii).

The practical consequence of this is that any other α -level test for a single outlier that meets the given criteria must have a lower probability of making a correct decision than the α -level test considered here, i.e. the test based upon large values for $M = \max_j a_j^2$.

4.2 Saddlepoint Approximation

An approximation for $F_M(x)$ is derived by first reexpressing this CDF as a conditional probability to which Bayes' theorem can be applied. A reversal of the probabilities in Bayes' theorem allows approximation to the CDF to be reexpressed in terms of a saddlepoint density approximation. Using first- and various second-order saddlepoint approximations for this multivariate density leads to various CDF approximations $\hat{F}_M(x)$ with second-order approximations providing greater accuracy.

Statistic M has bounded support, and owing to the nature of saddlepoint approximations, this can lead to a loss in accuracy for the various $\hat{F}_M(x)$. Since the least upper bound on the support of M is $n - p$, these approximations can be calibrated by using the fact that $F_M(x) = 1$, as described in Section 4.2.3. This procedure is shown to improve accuracy.

4.2.1 Reformulation of $F_M(x)$

The studentized residual vector $\mathbf{a} = (a_1, \dots, a_n)^T$, where a_j^2 is defined in (4.3), has a joint distribution which does not depend on the parameters (\mathbf{B}, Σ) , so that \mathbf{a} is ancillary. Thus, there is no loss in generality in assuming $\mathbf{B} = \mathbf{0}$ and $\Sigma = \mathbf{I}_m$. Since estimates $(\hat{\mathbf{B}}, \hat{\Sigma})$ form a complete sufficient statistic for (\mathbf{B}, Σ) , $(\hat{\mathbf{B}}, \hat{\Sigma})$ and \mathbf{a} are independent by Basu's theorem. Thus, the distribution of \mathbf{a} may be conditioned on any convenient value for $(\hat{\mathbf{B}}, \hat{\Sigma})$, say $(\mathbf{0}, \mathbf{I}_m)$. The net effect of such conditioning leads to

$$\begin{aligned} F_M(x) &= \Pr \left(\max_j a_j^2 \leq x \mid \hat{\mathbf{B}} = \mathbf{0}, \hat{\Sigma} = \mathbf{I} \right) \\ &= \Pr \left(\max_j \frac{\mathbf{y}_j^T \mathbf{y}_j}{1 - h_{jj}} \leq x \mid \hat{\mathbf{B}} = \mathbf{0}, \hat{\Sigma} = \mathbf{I} \right). \end{aligned} \quad (4.4)$$

By defining the events $E_j = \{\mathbf{y}_j^T \mathbf{y}_j \leq x(1 - h_{jj})\}$, so $\{M \leq x\} = \bigcap_{j=1}^n E_j$, Bayes' theorem for densities can be applied to the conditional probability in (4.4) to give

$$F_M(x) = \frac{f_{\hat{\mathbf{B}}, \hat{\Sigma}}(\mathbf{0}, \mathbf{I} \mid \bigcap_{j=1}^n E_j) \Pr \left(\bigcap_{j=1}^n E_j \right)}{f_{\hat{\mathbf{B}}, \hat{\Sigma}}(\mathbf{0}, \mathbf{I})}, \quad (4.5)$$

where $f_{\hat{\mathbf{B}}, \hat{\Sigma}}(\mathbf{0}, \mathbf{I})$ denotes the marginal density of $\hat{\mathbf{B}}, \hat{\Sigma}$ at $(\mathbf{0}, \mathbf{I})$, and $f_{\hat{\mathbf{B}}, \hat{\Sigma}}(\mathbf{0}, \mathbf{I} | \bigcap_{j=1}^n E_j)$ denotes the conditional density of $\hat{\mathbf{B}}, \hat{\Sigma}$ at $(\mathbf{0}, \mathbf{I})$ given event $\bigcap_{j=1}^n E_j$.

The denominator density in (4.5) is explicit. With $\mathbf{B} = \mathbf{0}$ and $\Sigma = \mathbf{I}$, then $\hat{\mathbf{B}}$ and $\hat{\Sigma}$ are independent, with $\text{vec}(\hat{\mathbf{B}}) \sim \mathbb{N}_{pm} \{ \mathbf{0}, \mathbf{I}_m \otimes (\mathbf{X}^T \mathbf{X})^{-1} \}$, and $\hat{\Sigma}(n-p) = \hat{\mathbf{E}}^T \hat{\mathbf{E}} \sim \mathbb{W}_m(\mathbf{I}, n-p)$, a Wishart distribution. With a Jacobian of $(n-p)^{m(m+1)/2}$,

$$f_{\hat{\Sigma}}(\mathbf{I}) = (n-p)^{m(m+1)/2} f_{\mathbb{W}_m(\mathbf{I}, n-p)} \{ (n-p) \mathbf{I} \},$$

so that

$$\begin{aligned} f_{\hat{\mathbf{B}}, \hat{\Sigma}}(\mathbf{0}, \mathbf{I}) &= f_{\hat{\mathbf{B}}}(\mathbf{0}) f_{\hat{\Sigma}}(\mathbf{I}) = \frac{|\mathbf{X}^T \mathbf{X}|^{m/2}}{(2\pi)^{mp/2}} \times (n-p)^{m(m+1)/2} f_{\mathbb{W}_m(\mathbf{I}, n-p)} \{ (n-p) \mathbf{I}_m \} \\ &= |\mathbf{X}^T \mathbf{X}|^{m/2} \frac{\left(\frac{n-p}{2}\right)^{m(n-p)/2} \exp\{-m(n-p)/2\}}{(2\pi)^{mp/2} \Gamma_m\{(n-p)/2\}} =: g, \end{aligned} \quad (4.6)$$

where $\Gamma_m(x)$ is the multivariate gamma function, with

$$\Gamma_m(x) = \pi^{m(m-1)/4} \prod_{j=1}^m \Gamma\{x + (1-j)/2\}.$$

Since $\{E_j\}$ are independent events and $\mathbf{y}_1, \dots, \mathbf{y}_n$ are i.i.d. $\mathbb{N}_m(\mathbf{0}, \mathbf{I}_m)$, $\{\mathbf{y}_j^T \mathbf{y}_j\}$ are i.i.d. χ_m^2 , and

$$\Pr\left(\bigcap_{j=1}^n E_j\right) = \prod_{j=1}^n \Pr(E_j) = \prod_{j=1}^n F_{\chi_m^2}(\tau_j),$$

where $\tau_j := x(1 - h_{jj})$ for $j = 1, \dots, n$. Define

$$P_m(\tau) := \int_{\mathbf{z}^T \mathbf{z} < \tau} (2\pi)^{-m/2} \exp(-\mathbf{z}^T \mathbf{z}/2) d\mathbf{z} = \Pr(\chi_m^2 < \tau) = F_{\chi_m^2}(\tau), \quad (4.7)$$

so that $P_m(\tau)$ is the readily available incomplete gamma function. With this substitution,

$$\Pr\left(\bigcap_{j=1}^n E_j\right) = \prod_{j=1}^n P_m(\tau_j).$$

Only the conditional density in the numerator of (4.5) offers any difficulty. Note that $\hat{\mathbf{B}}$ is linear in $\{\mathbf{y}_j\}$, and

$$(n - p) \hat{\boldsymbol{\Sigma}} = \mathbf{Y}^T \mathbf{Y} - \hat{\mathbf{B}}^T \mathbf{X}^T \mathbf{X} \hat{\mathbf{B}},$$

is linear in $\{\mathbf{y}_j \mathbf{y}_j^T\}$ apart from its dependence on $\hat{\mathbf{B}}$. To obtain a multivariate density in which all components are linear in $\{\mathbf{y}_j\}$ and $\{\mathbf{y}_j \mathbf{y}_j^T\}$, write the conditional density of $(\hat{\mathbf{B}}, \hat{\boldsymbol{\Sigma}})$ given $\bigcap_{j=1}^n E_j$ in terms of the conditional density of $(n\hat{\mathbf{B}}, \mathbf{Y}^T \mathbf{Y})$ given $\bigcap_{j=1}^n E_j$ through a one-to-one transformation. The Jacobian of this transformation is $n^{pm} (n - p)^{m(m+1)/2}$ and therefore,

$$f_{\hat{\mathbf{B}}, \hat{\boldsymbol{\Sigma}}} \left(\mathbf{0}, \mathbf{I} \mid \bigcap_{j=1}^n E_j \right) = n^{pm} (n - p)^{m(m+1)/2} f_{n\hat{\mathbf{B}}, \mathbf{Y}^T \mathbf{Y}} \left(\mathbf{0}, (n - p) \mathbf{I} \mid \bigcap_{j=1}^n E_j \right).$$

Next, note that the net effect of conditioning on the event $\bigcap_{j=1}^n E_j$ is to truncate the m -variate normal random variables so that \mathbf{y}_j is restricted to the range $\mathbf{y}_j^T \mathbf{y}_j < \tau_j$ for $j = 1, \dots, n$. Conditioning on $\bigcap_{j=1}^n E_j$ affects the distribution theory in such a way that \mathbf{y}_j is independent of $\bigcap_{i \neq j} E_i$ and is only affected by the conditioning on E_j . Thus $\{\mathbf{y}_j \mid E_j\}$ are independent and $\mathbf{y}_j \mid E_j$ has a standard m -variate normal distribution that has been spherically truncated to $\{\mathbf{y}_j^T \mathbf{y}_j < \tau_j\}$. Denote this truncated variable as \mathbf{z}_j with density

$$f_{\mathbf{z}_j}(\mathbf{x}) = \frac{\exp(-\mathbf{x}^T \mathbf{x} / 2)}{(2\pi)^{m/2} P_m(\tau_j)}, \quad \mathbf{x}^T \mathbf{x} < \tau_j. \quad (4.8)$$

Let $(n \times m)$ random matrix \mathbf{Z} have independent rows \mathbf{z}_j^T and elements z_{ji} for $j = 1, \dots, n$, and $i = 1, \dots, m$. In terms of \mathbf{Z} ,

$$n\hat{\mathbf{B}} \Big| \bigcap_{j=1}^n E_j \stackrel{D}{=} n (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z} =: \mathbf{V} \quad (4.9)$$

$$\mathbf{Y}^T \mathbf{Y} \Big| \bigcap_{j=1}^n E_j \stackrel{D}{=} \mathbf{Z}^T \mathbf{Z} =: \mathbf{U} \quad (4.10)$$

The conditional distributions on the left-hand side of (4.9) and (4.10) are the same as the unconditional distributions on the right-hand side using truncated random variables, thus effectively

removing the conditioning required in the distribution theory. From (4.5) we may write

$$F_M(x) = f_{\mathbf{V}, \mathbf{U}}(\mathbf{0}, (n-p)\mathbf{I}) \frac{n^{pm} (n-p)^{m(m+1)/2}}{g} \prod_{j=1}^n P_m(\tau_j), \quad (4.11)$$

where g is defined in (4.6).

In (4.11), the density $f_{\mathbf{V}, \mathbf{U}}(\mathbf{0}, (n-p)\mathbf{I})$ cannot be readily computed, but a saddlepoint approximation can be constructed for it and used in its place to calculate an approximation for $F_M(x)$, which we refer to as a saddlepoint approximation for $F(x)$.

Saddlepoint Density Function

A saddlepoint approximation for the density function of (\mathbf{V}, \mathbf{U}) is provided in Butler (2007, eqn. 3.2). Since \mathbf{V} is $(p \times m)$ and $\mathbf{U} = \{u_{ik}\}$ is $(m \times m)$ symmetric, (\mathbf{V}, \mathbf{U}) is treated here as a random vector with $m\{p + (m+1)/2\}$ elements. Let (\mathbf{S}, \mathbf{T}) index the moment generating function (MGF), where \mathbf{S} is $(p \times m)$. Since \mathbf{U} is symmetric, with only $m(m+1)/2$ distinct elements, we define \mathbf{T} as an $(m \times m)$ upper triangular matrix, with elements $\{t_{ik}\}$, $i \leq k$. The MGF of (\mathbf{V}, \mathbf{U}) is denoted as

$$\mathcal{M}(\mathbf{S}, \mathbf{T}) = \mathbb{E} \left[\exp \left\{ \text{tr}(\mathbf{S}^T \mathbf{V}) + \text{tr}(\mathbf{T}^T \mathbf{U}) \right\} \right],$$

where $\mathbf{S} \in \mathbb{R}^{pm}$ and $\mathbf{T}^U \in \mathbb{R}^{m(m+1)/2}$, with \mathbf{A}^U defined as the upper triangular portion of square matrix \mathbf{A} . The corresponding cumulant generating function (CGF) is $\mathcal{K}(\mathbf{S}, \mathbf{T}) = \ln \mathcal{M}(\mathbf{S}, \mathbf{T})$. Since the density $f_{\mathbf{V}, \mathbf{U}}$ is to be evaluated at $(\hat{\mathbf{S}} = \mathbf{0}, \hat{\mathbf{T}} = \hat{t}\mathbf{I}_m)$, this determines a joint first-order saddlepoint density for $f_{\mathbf{V}, \mathbf{U}}(\mathbf{0}, (n-p)\mathbf{I})$ as

$$\hat{f}_1(\mathbf{0}, (n-p)\mathbf{I}) := \frac{\exp \left\{ \mathcal{K}(\hat{\mathbf{S}}, \hat{\mathbf{T}}) - (n-p)m\hat{t} \right\}}{(2\pi)^{m\{p+(m+1)/2\}/2} \left| \mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}}) \right|^{1/2}}, \quad (4.12)$$

where saddlepoint $(\hat{\mathbf{S}}, \hat{\mathbf{T}} = \hat{t}\mathbf{I}_m)$ simultaneously solves the $m\{p + (m+1)/2\}$ -dimensional equation

$$\left. \frac{\partial}{\partial \mathbf{S}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \right|_{\mathbf{S}=\hat{\mathbf{S}}, \mathbf{T}=\hat{\mathbf{T}}^U} = \mathbf{0} \quad (4.13)$$

$$\left. \frac{\partial}{\partial \mathbf{T}^U} \mathcal{K}(\mathbf{S}, \mathbf{T}) \right|_{\mathbf{S}=\hat{\mathbf{S}}, \mathbf{T}=\hat{\mathbf{T}}^U} = (n-p) \mathbf{I}_m^U. \quad (4.14)$$

Matrix $\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})$ is the Hessian matrix of $\mathcal{K}(\mathbf{S}, \mathbf{T})$ evaluated at $(\hat{\mathbf{S}}, \hat{\mathbf{T}})$.

As shown in Section 4.4, the saddlepoint equation is solved by $\hat{\mathbf{S}} = \mathbf{0}$ and $\hat{\mathbf{T}} = \hat{t} \mathbf{I}_m$. From (4.32), the scalar value \hat{t} is the solution to

$$n-p = \sum_{j=1}^n \frac{P_{m+2}(\hat{q}\tau_j)}{\hat{q}P_m(\hat{q}\tau_j)}, \quad (4.15)$$

where $\hat{q} = 1 - 2\hat{t}$, and P_m is defined in (4.7). It is important to note that the range of possible values for \hat{t} is $(-\infty, 1/2)$, corresponding to the right tail of M , the MSSR statistic. All these formulas have a removable discontinuity at $\hat{t} = 1/2$, and the form of the equations changes on the other side of the discontinuity, although these alternate forms when $\hat{t} \geq 1/2$ are not presented here. At the saddlepoint, from (4.29),

$$\exp \left\{ \mathcal{K}(\mathbf{0}, \hat{\mathbf{T}}) \right\} = \prod_{j=1}^n \frac{P_m(\hat{q}\tau_j)}{\hat{q}^{m/2} P_m(\tau_j)}.$$

The Hessian matrix is a potentially large, block diagonal matrix, with determinant $|\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})|$ given by (4.41). Putting these pieces together, the first-order saddlepoint approximation for the density of \mathbf{V}, \mathbf{U} is

$$\hat{f}_1(\mathbf{0}, (n-p)\mathbf{I}) := \frac{\exp \left\{ -(n-p)m\hat{t} \right\}}{(2\pi)^{m\{p+(m+1)/2\}/2} |\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})|^{1/2}} \prod_{j=1}^n \frac{P_m(\hat{q}\tau_j)}{\hat{q}^{m/2} P_m(\tau_j)}. \quad (4.16)$$

With this approximate density for $f_{\mathbf{V}, \mathbf{U}}(\mathbf{0}, (n-p)\mathbf{I})$, substituted into (4.11), the first-order saddlepoint approximation for the CDF of M , the maximum squared studentized residual, is

$$\hat{F}_1(x) = \frac{n^{pm} (n-p)^{m(m+1)/2} \exp \left\{ -m\hat{t}(n-p) \right\}}{g \hat{q}^{nm/2} (2\pi)^{m\{p+(m+1)/2\}/2} |\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})|^{1/2}} \prod_{j=1}^n P_m(\hat{q}\tau_j), \quad (4.17)$$

where g is defined in (4.6). After substituting g and simplifying,

$$\hat{F}_1(x) = \frac{2^{m(m+1)/2} n^{pm} \left(\frac{n-p}{2}\right)^{m(m-n+p+1)/2} \exp\left\{m\hat{q}\left(\frac{n-p}{2}\right)\right\} \Gamma_m\left\{\frac{n-p}{2}\right\}}{\hat{q}^{nm/2} (2\pi)^{m(m+1)/4} |\mathbf{X}^T \mathbf{X}|^{m/2} |\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})|^{1/2}} \prod_{j=1}^n P_m(\hat{q}\tau_j) \quad (4.18)$$

The $\hat{\cdot}$ on $\hat{F}_1(x)$ denotes the use of a saddlepoint density approximation to approximate $F_M(x)$, and the 1 subscript denotes the first-order saddlepoint approximation.

We now present different versions of the saddlepoint approximation: first-order, second-order, and second-order exponential denoted respectively as \hat{f}_1 , \hat{f}_2 , and \hat{f}_{2e} . Upon substitution into (4.11) to approximate $F_M(x)$, a calibration procedure may also be applied, which is indicated by changing the circumflex to a macron. This leads to the possibility of six approximations of $F_M(x)$, denoted by \hat{F}_1 , \bar{F}_1 , \hat{F}_2 , \bar{F}_2 , \hat{F}_{2e} , and \bar{F}_{2e} .

4.2.2 Second-Order Approximation

The saddlepoint density is an asymptotic expression, of which the first-order approximation is only the leading term. Including the second term often improves the accuracy of the approximation. From Butler (2007, eqn. 3.37), the second-order saddlepoint density approximation of $f(x)$ is

$$\hat{f}_2(x) = \hat{f}_1(x) (1 + O) \quad (4.19)$$

$$\text{where } O = \{\hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24\}.$$

Using the Einstein summation convention as in McCullagh (1987, §1.2), where summation is implied over any subscript appearing once in a subscript and once in a superscript, the three new $\hat{\kappa}$ terms in the O formula are defined as

$$\begin{aligned} \hat{\kappa}_4 &:= \hat{\mathcal{K}}_{ijkl} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kl} \\ \hat{\kappa}_{23}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ir} \hat{\mathcal{K}}^{js} \hat{\mathcal{K}}^{kt}, \\ \hat{\kappa}_{13}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kr} \hat{\mathcal{K}}^{st} \end{aligned} \quad (4.20)$$

where the subscripted $\hat{\mathcal{K}}$ terms are derivatives of \mathcal{K} with respect to the subscripted components, evaluated at the saddlepoint $(\mathbf{S}, \mathbf{T}) = (\mathbf{0}, \hat{\mathbf{t}}_m)$, and the superscripted $\hat{\mathcal{K}}$ terms are components of

the inverse of the Hessian matrix $\mathcal{K}''(\mathbf{0}, \hat{\mathbf{I}}_m)$. Since \mathbf{S} is a $(p \times m)$ matrix and \mathbf{T} is a $(m \times m)$ upper triangular matrix, the vectorized version of (\mathbf{S}, \mathbf{T}) has $m\{p + (m + 1)/2\}$ elements. Therefore, in (4.20), the suffixes i, j, k, r, s, t range from 1 to $m\{p + (m + 1)/2\}$.

In programming this calculation, the inverse Hessian matrix, third- and fourth-order derivatives are stored into two-, three-, and four-dimensional arrays of size $m\{p + (m + 1)/2\}$ in each dimension. It is convenient to assign rows $1, \dots, mp$ to elements of \mathbf{S} , rows $mp + 1, \dots, m(p + 1)$ to diagonal elements of \mathbf{T} , and rows $m(p + 1) + 1, \dots, m\{p + (m + 1)/2\}$ to the off-diagonal elements of \mathbf{T} , in each dimension. These arrays may be sparse, with many zero elements, but also many non-zero elements. Except for the i.i.d. special case discussed in Section 4.5, it is necessary to store and calculate with the full arrays. Formulas for each derivative of $\hat{\mathcal{K}}$ are provided in section 4.4.

An alternative second-order exponential approximation, based on McCullagh (1987, § 6.3) is

$$\hat{f}_{2e}(x) = \hat{f}_1(x) e^O, \quad (4.21)$$

where O is the same as in (4.19). In many cases we have evaluated, including the examples shown below, this exponential approximation is the best of all.

Putting all this together, the second-order approximate distributions of MASR M are one of the following:

$$\hat{F}_2(x) = \hat{F}_1(x) \{1 + \hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24\} \quad (4.22)$$

$$\hat{F}_{2e}(x) = \hat{F}_1(x) \exp \{ \hat{\kappa}_4/8 - (2\hat{\kappa}_{23}^2 + 3\hat{\kappa}_{13}^2)/24 \}. \quad (4.23)$$

In some special cases, the $\hat{\kappa}$ functions might be simplified, but in general, the four-way and six-way sums must be calculated and summed term by term. Section 4.4 lists all the third- and fourth-order derivatives of $\mathcal{K}(\mathbf{S}, \mathbf{T})$ required to make these calculations.

4.2.3 Calibration

Calibration is a simple procedure recommended to correct for known inaccuracies with the saddlepoint approximation. The upper end of the support of M is $M_U = n - p$, so that $F_M(M_U) = 1$. Whichever version of the saddlepoint approximation is used, it may not meet $\hat{F}_\bullet(M_U) = 1$, where here the \bullet subscript represents either $_1$, $_2$, or $_e$. To improve the accuracy of p -values for the discordancy test, it is recommended to calculate saddlepoint approximations for both $\hat{F}_\bullet(M_U)$ and $\hat{F}_\bullet(x)$, and then calibrate the approximation by

$$\bar{F}_\bullet(x) = \frac{\hat{F}_\bullet(x)}{\hat{F}_\bullet(M_U)}, \quad (4.24)$$

where the macron over \bar{F} denotes the calibrated saddlepoint approximation.

In Chapter 3, with the univariate regression setting, options for calibrating at M_2 or M_3 were described. However, in the multivariate setting, with nonzero probabilities of two or more residual values all the way up to M_U , these calibration alternatives are not available.

4.3 Results

In this section, we apply the methods of this chapter to eight examples. All examples are compared with a value of M determined by simulation to correspond to a survival probability of approximately 0.1. Four examples of the multivariate i.i.d. case are summarized in Table 4.1, using the streamlined calculation formulas from section 4.5. Table 4.2 summarizes multivariate versions of regression examples first analyzed in Section 3.3 in the univariate case. Following the tables are graphs comparing survival curves for some of these examples.

From the graphs and tables, but also from the experience of running these calculations, we make these observations:

- The only saddlepoint approximation that consistently approximates the survival probability determined by simulation is the calibrated, second-order, exponential version. This version of the saddlepoint method is recommended for all applications.

Table 4.1: Saddlepoint and Bonferroni p -values for Four Multivariate i.i.d. Examples

n	30	30	100	100
p	1	1	1	1
m	2	5	2	5
M value x	9.969	14.519	13.084	19.058
Saddlepoint approx.:				
$1 - \hat{F}_1(x)$	-0.011	-1.05	0.069	-0.139
$1 - \hat{F}_2(x)$	0.085	-1.40	0.096	-0.199
$1 - \hat{F}_{2e}(x)$	0.081	-1.43	0.095	-0.201
$1 - \bar{F}_1(x)$	0.091	0.059	0.097	0.090
$1 - \bar{F}_2(x)$	0.099	0.098	0.100	0.102
$1 - \bar{F}_{2e}(x)$	0.099	0.105	0.100	0.102
Bonferroni bounds:				
1 st upper (Grubbs)	0.102	0.104	0.103	0.105
1 st upper (Worsley)	0.102	0.104	0.103	0.104
2 nd lower	0.100	0.100	0.100	0.100
Limits on M:				
M_U	29	29	99	99

- The Bonferroni bounds listed in these tables are deceptively encouraging for the Bonferroni method. It is worth noting two important weaknesses of the Bonferroni approach to this problem:
 - Only the 1st upper (Grubbs) bound is fast and easy to calculate. Both the improved Worsley upper bound and the 2nd lower bound require calculation of all joint probabilities of pairs of residuals, and each pair of residuals with a different correlation requires a separate numerical triple integral. In the `airquality` example, the 116 observations require the calculation of 6,670 such triple integrals.
 - In the examples tabulated here, the 1st upper bound looks quite good, but in some other examples, this bound could be substantially higher than the true value, with no way to determine when this is happening, other than to calculate the lower bound.

Table 4.2: Saddlepoint and Bonferroni p -values for Four Multivariate Regression Examples

	DSD 17-run	DSD 17-run	airquality	airquality
n	51	51	116	116
p	9	9	3	3
m	2	5	2	4
M value x	11.224	16.319	13.420	17.68
Saddlepoint approx.:				
$1 - \hat{F}_1(x)$	0.785	0.992	0.069	1.000
$1 - \hat{F}_2(x)$	1.000	-1.40	1.000	1.000
$1 - \hat{F}_{2e}(x)$	0.921	-1.43	0.095	1.000
$1 - \bar{F}_1(x)$	0.104	0.059	0.097	0.093
$1 - \bar{F}_2(x)$	1.146	0.098	0.100	0.103
$1 - \bar{F}_{2e}(x)$	0.113	0.105	0.100	0.102
Bonferroni bounds:				
1 st upper (Grubbs)	0.102	0.102	0.104	0.104
1 st upper (Worsley)	0.101	0.102	0.104	0.104
2 nd lower	0.099	0.099	0.100	0.100
Limits on M:				
M_U	42	42	113	113

MSSR Outlier P-value, $n=30$ $p=1$ $m=2$

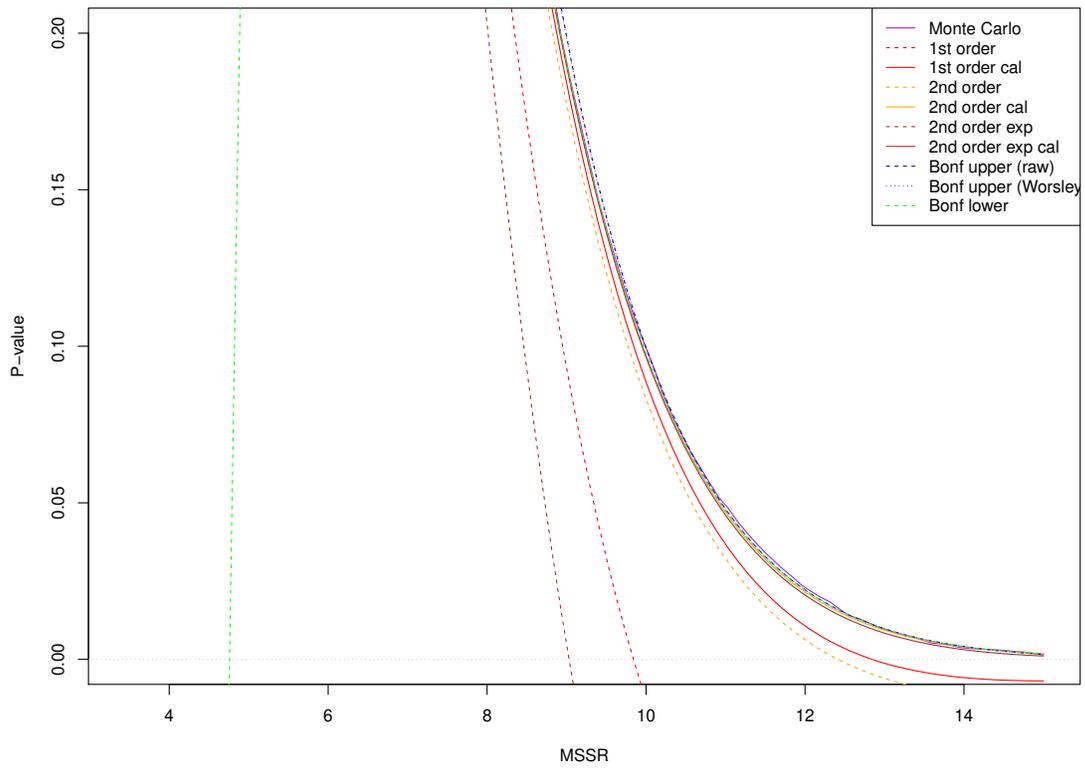


Figure 4.1: Comparison of p -values for i.i.d. case, $n = 30$ and $m = 2$

MSSR Outlier P-value, $n=30$ $p=1$ $m=5$

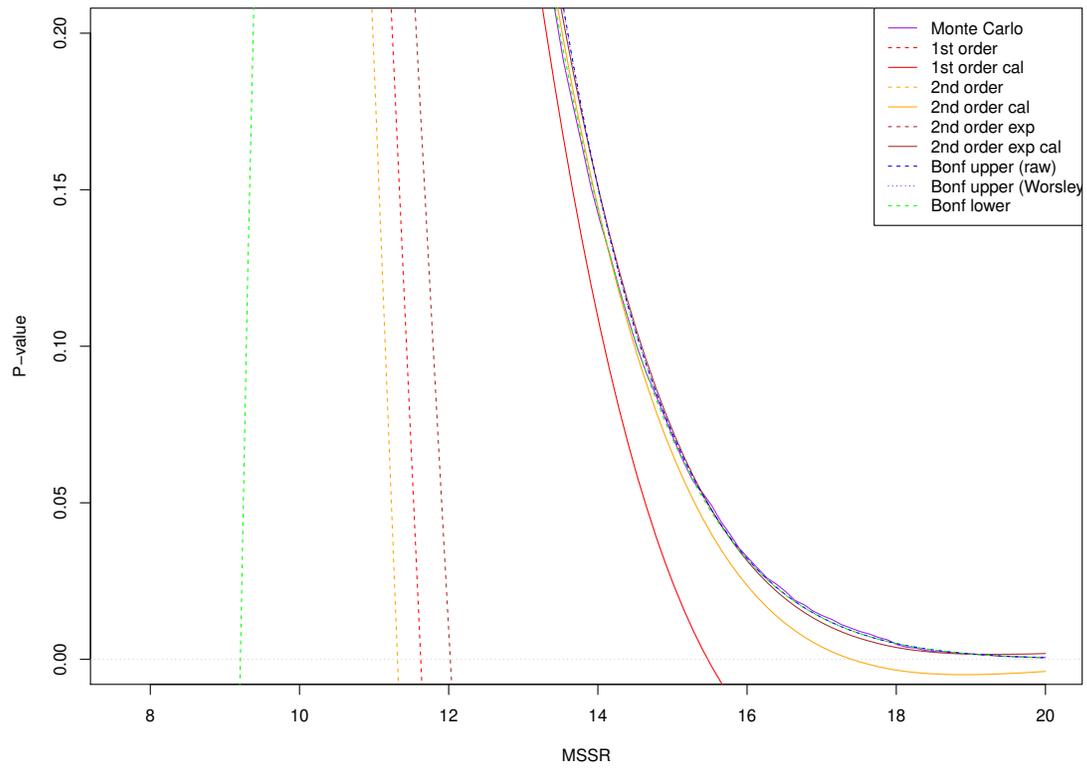


Figure 4.2: Comparison of p -values for i.i.d. case, $n = 30$ and $m = 5$

MSSR Outlier P-value, $n=100$ $p=1$ $m=2$

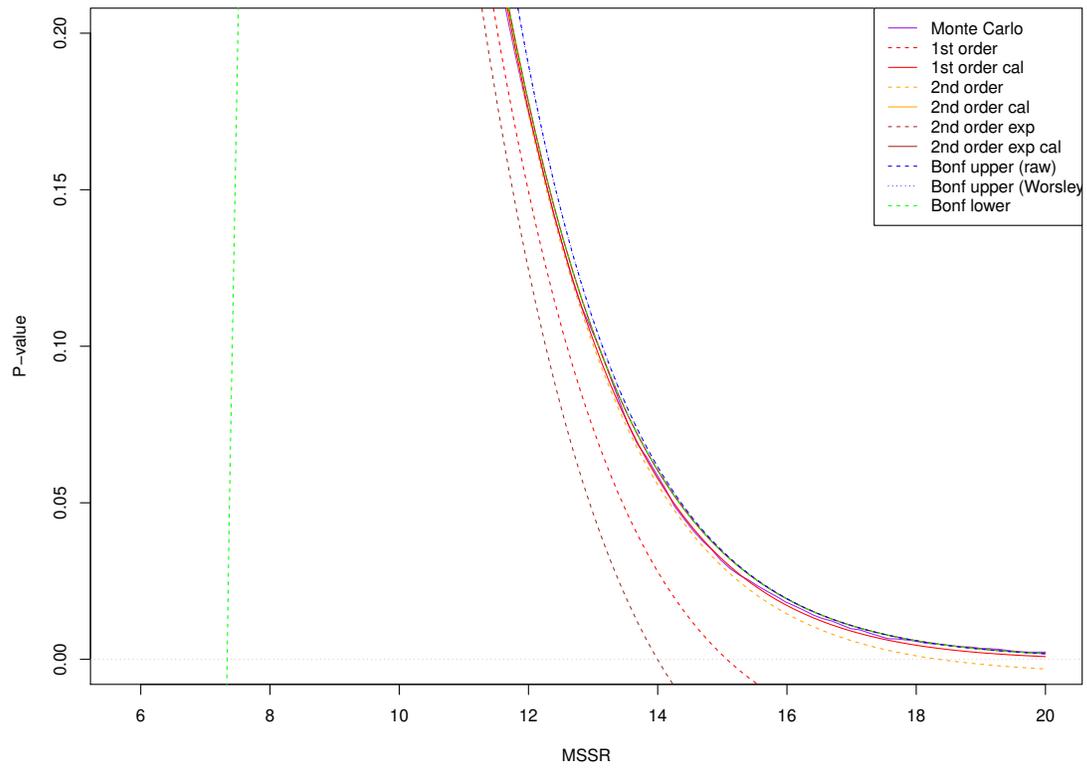


Figure 4.3: Comparison of p -values for i.i.d. case, $n = 100$ and $m = 2$

MSSR Outlier P-value, $n=100$ $p=1$ $m=5$

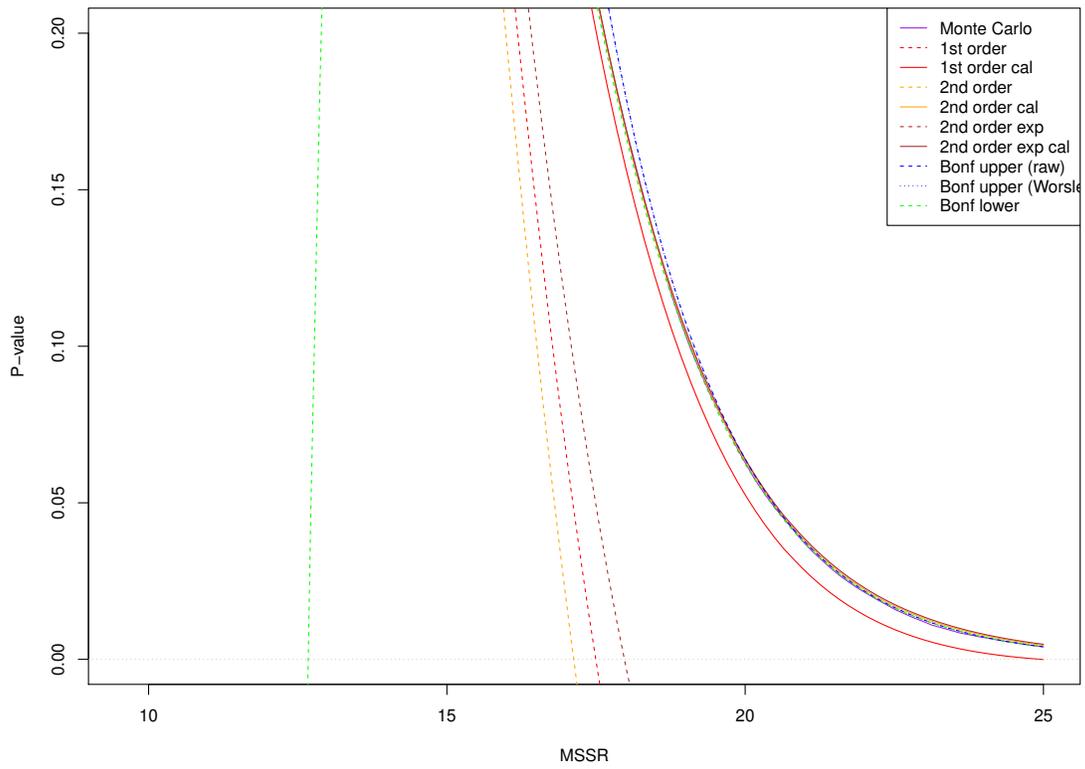


Figure 4.4: Comparison of p -values for i.i.d. case, $n = 100$ and $m = 5$

MSSR Outlier P-value, $n=51$ $p=9$ $m=2$

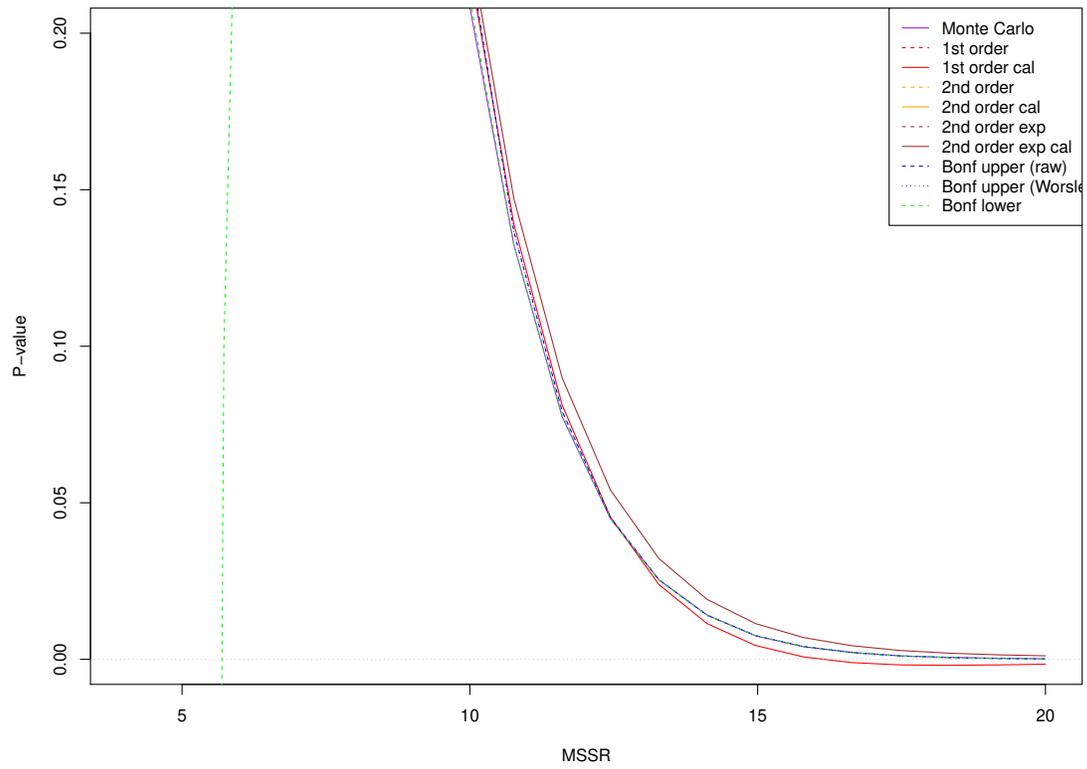


Figure 4.5: Comparison of p -values for DSD example, $n = 51$ and $m = 2$

4.4 Cumulant Generating Function $\mathcal{K}(\mathbf{S}, \mathbf{T})$ and its Derivatives

In this section, the joint CGF $\mathcal{K}(\mathbf{S}, \mathbf{T})$ of $\mathbf{V} = n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z}$ and $\mathbf{U} = \mathbf{Z}^T \mathbf{Z}$ is developed into a form for calculation at the saddlepoint $(\hat{\mathbf{S}}, \hat{\mathbf{T}})$, along with its derivatives required for the saddlepoint approximations. Let

$$\mathbf{C} := n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = (\mathbf{c}_1, \dots, \mathbf{c}_n) = (c_{ij})$$

be $p \times n$. Then $\mathbf{V} = \mathbf{CZ} = \sum_{j=1}^n \mathbf{c}_j \mathbf{z}_j^T$ is $p \times m$. Also, $\mathbf{U} = \mathbf{Z}^T \mathbf{Z} = \sum_{j=1}^n \mathbf{z}_j \mathbf{z}_j^T$ is $m \times m$

The joint MGF of (\mathbf{V}, \mathbf{U}) with index variables \mathbf{S} , $(p \times m)$, and upper triangular \mathbf{T} , $(m \times m)$, is

$$\begin{aligned} \mathcal{M}(\mathbf{S}, \mathbf{T}) &= \mathbb{E} \left[\exp \left\{ \text{tr}(\mathbf{S}^T \mathbf{V}) + \text{tr}(\mathbf{T} \mathbf{U}) \right\} \right] \\ &= \mathbb{E} \left[\exp \left\{ \text{tr} \left(\mathbf{S}^T \sum_{j=1}^n \mathbf{c}_j \mathbf{z}_j^T \right) + \text{tr} \left(\mathbf{T} \sum_{j=1}^n \mathbf{z}_j \mathbf{z}_j^T \right) \right\} \right] \\ &= \mathbb{E} \left[\prod_{j=1}^n \exp \left\{ \text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}_j^T) + \text{tr}(\mathbf{T} \mathbf{z}_j \mathbf{z}_j^T) \right\} \right]. \end{aligned}$$

Since the $\{\mathbf{z}_j\}$, the rows of \mathbf{Z} , are uncorrelated,

$$\mathcal{M}(\mathbf{S}, \mathbf{T}) = \prod_{j=1}^n \mathbb{E} \left[\exp \left\{ \text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}_j^T) + \text{tr}(\mathbf{T} \mathbf{z}_j \mathbf{z}_j^T) \right\} \right] := \prod_{j=1}^n \mathcal{M}_j(\mathbf{S}, \mathbf{T}).$$

Since \mathbf{z}_j is spherically truncated standard normal,

$$\begin{aligned} \mathcal{M}_j(\mathbf{S}, \mathbf{T}) &= \frac{1}{(2\pi)^{m/2} P_m(\tau_j)} \times \\ &\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr} \left\{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T \right\} \right] d\mathbf{z}, \end{aligned} \quad (4.25)$$

where $P_m(\tau)$ is defined in (4.7). The integrand in (4.25) is bounded and τ_j is finite, therefore $\mathcal{M}_j(\mathbf{S}, \mathbf{T})$ is convergent for all real, finite (\mathbf{S}, \mathbf{T}) . The cumulant generating function (CGF) is

$\mathcal{K}(\mathbf{S}, \mathbf{T}) = \ln \mathcal{M}(\mathbf{S}, \mathbf{T}) = \sum_{j=1}^n \mathcal{K}_j(\mathbf{S}, \mathbf{T})$, where $\mathcal{K}_j(\mathbf{S}, \mathbf{T}) = \ln \mathcal{M}_j(\mathbf{S}, \mathbf{T})$. Therefore, $\mathcal{K}(\mathbf{S}, \mathbf{T})$ is also convergent.

To show that $\hat{\mathbf{S}} = \mathbf{0}$, note that

$$\begin{aligned} \mathcal{K}'_{\mathbf{S}}(\hat{\mathbf{S}}, \hat{\mathbf{T}}) &:= \left. \frac{\partial \mathcal{K}}{\partial \mathbf{S}} \right|_{\mathbf{s}=\hat{\mathbf{s}}} = \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\hat{\mathbf{S}}, \hat{\mathbf{T}})} \mathcal{M}_{j,\mathbf{s}}(\mathbf{S}, \hat{\mathbf{T}}) \Big|_{\mathbf{s}=\hat{\mathbf{s}}} \\ &= \sum_{j=1}^n \frac{\frac{\partial}{\partial \mathbf{S}} \mathbb{E} \left[\exp \left\{ \text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}_j^T) + \text{tr} \left\{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z}_j \mathbf{z}_j^T \right\} \right\} \right]}{(2\pi)^{m/2} P_m(\tau_j) \mathcal{M}_j(\hat{\mathbf{S}}, \hat{\mathbf{T}})} \Big|_{\mathbf{s}=\hat{\mathbf{s}}} \\ &= \sum_{j=1}^n \frac{\mathbf{c}_j \mathbb{E} \left[\mathbf{z}_j^T \exp \left\{ \text{tr}(\hat{\mathbf{S}}^T \mathbf{c}_j \mathbf{z}_j^T) + \text{tr} \left\{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z}_j \mathbf{z}_j^T \right\} \right\} \right]}{(2\pi)^{m/2} P_m(\tau_j) \mathcal{M}_j(\hat{\mathbf{S}}, \hat{\mathbf{T}})}. \end{aligned}$$

For any value of $\hat{\mathbf{T}}$, the saddlepoint equation $\mathcal{K}'_{\mathbf{S}}(\hat{\mathbf{S}}, \hat{\mathbf{T}}) = \mathbf{0}$ is solved by $\hat{\mathbf{S}} = \mathbf{0}$, since

$$\mathbb{E} \left[\mathbf{z}_j^T \text{etr} \left\{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z}_j \mathbf{z}_j^T \right\} \right] = \mathbf{0}$$

for all j , where $\text{etr}(\mathbf{A}) = \exp \{ \text{tr}(\mathbf{A}) \}$. If, for $\hat{\mathbf{S}} = \mathbf{0}$, a solution $\hat{\mathbf{T}}$ can be determined for $\mathcal{K}'_{\mathbf{T}}(\hat{\mathbf{S}}, \hat{\mathbf{T}}) = (n-p) \mathbf{I}$, then $(\hat{\mathbf{S}}, \hat{\mathbf{T}})$ must be the saddlepoint, since there can be only one solution to both equations in the solution space for all real, finite (\mathbf{S}, \mathbf{T}) .

Saddlepoint $\hat{\mathbf{S}} = \mathbf{0}$ induces symmetry into the evaluation of \mathcal{K} and its derivatives at the saddlepoint $(\mathbf{0}, \hat{\mathbf{T}})$. The density of \mathbf{Z} is symmetric in that it is invariant to replacement of any element of \mathbf{Z} , say z_{ij} , with $-z_{ij}$. As a consequence, $\mathcal{K}(\mathbf{S}, \mathbf{T})$ must be an even function of each element of \mathbf{S} , and therefore, all odd-order derivatives of \mathcal{K} with respect to any element of \mathbf{S} are equal to $\mathbf{0}$ at $(\mathbf{0}, \hat{\mathbf{T}})$.

To show that the off-diagonal elements of $\hat{\mathbf{T}}$ are 0, at the saddlepoint $(\mathbf{0}, \hat{\mathbf{T}})$, consider any t_{ik} , where $1 \leq i < k \leq m$. Note that

$$\begin{aligned}
\frac{\partial}{\partial t_{ik}} \mathcal{K}(\mathbf{0}, \mathbf{T}) &= \sum_{j=1}^n \frac{1}{\mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}})} \left. \frac{\partial \mathcal{M}_j(\mathbf{0}, \mathbf{T})}{\partial t_{ik}} \right|_{\mathbf{T}=\hat{\mathbf{T}}} \\
&= \sum_{j=1}^n \frac{\frac{\partial}{\partial t_{ik}} \mathbb{E} [\text{tr} \{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z}_j \mathbf{z}_j^T \}]}{(2\pi)^{m/2} P_m(\tau_j) \mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}})} \Big|_{\mathbf{T}=\hat{\mathbf{T}}} \\
&= \sum_{j=1}^n \frac{\mathbb{E} [z_{ji} z_{jk} \text{tr} \{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z}_j \mathbf{z}_j^T \}]}{(2\pi)^{m/2} P_m(\tau_j) \mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}})}
\end{aligned}$$

The saddlepoint equation (4.14) requires that $\partial \mathcal{K}(\mathbf{0}, \mathbf{T}) / \partial t_{ik} = 0$ at the saddlepoint. Given the spherically symmetric nature of \mathbf{z}_j as given in (4.8), the saddlepoint must have $\hat{t}_{ik} = 0$ for $i \neq k$.

To show that $\hat{\mathbf{T}} = \hat{t} \mathbf{I}_m$, observe that \mathbf{Z} is right orthogonally invariant, so that interchanging columns of \mathbf{Z} has no effect on the distribution. Therefore, $\mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}})$ must also be invariant to interchanging diagonal elements of $\hat{\mathbf{T}}$. Therefore, $\hat{\mathbf{T}} = \hat{t} \mathbf{I}_m$.

We now define $q := 1 - 2t$ and $\hat{q} := 1 - 2\hat{t}$. Therefore, at the saddlepoint, (4.25) simplifies to

$$\begin{aligned}
\mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}}) &= \frac{1}{(2\pi)^{m/2} P_m(\tau_j)} \int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left\{ \left(\hat{t} - \frac{1}{2} \right) \sum_{i=1}^m z_i^2 \right\} d\mathbf{z} \\
&= \frac{1}{(2\pi)^{m/2} P_m(\tau_j)} \int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left(\sum_{i=1}^m \frac{-\hat{q} z_i^2}{2} \right) d\mathbf{z}.
\end{aligned}$$

Now change variables with $\mathbf{y} = (y_1, \dots, y_m)^T = (\sqrt{q} z_1, \dots, \sqrt{q} z_m)^T$. This leads to

$$\mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}}) = \frac{(2\pi\hat{q})^{-m/2}}{P_m(\tau_j)} \int_{\mathbf{y}^T \mathbf{y} < \hat{q}\tau_j} \exp \left(\sum_{i=1}^m \frac{-y_i^2}{2} \right) d\mathbf{y}. \quad (4.26)$$

Re-expressing the integral in spherical coordinates,

$$\mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}}) = \frac{(2\pi\hat{q})^{-m/2}}{P_m(\tau_j)} \int_0^{\sqrt{\hat{q}\tau_j}} \int_{S^{m-1}(r)} e^{-r^2/2} dA dr$$

where $S^{m-1}(r)$ is an $(m-1)$ -sphere of radius r , an $(m-1)$ -dimensional manifold in \mathbb{R}^m , and dA is an area element on that sphere. Since $\exp(-r^2/2)$ is rotationally invariant, the integral separates.

It is well-known that the surface area of $S^{m-1}(r)$ is

$$\int_{S^{m-1}(r)} dA = \frac{2\pi^{m/2} r^{m-1}}{\Gamma(m/2)}, \quad (4.27)$$

and therefore, with χ_m being the χ distribution with m degrees of freedom,

$$\begin{aligned} \mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}}) &= \frac{(2\pi\hat{q})^{-m/2} 2\pi^{m/2}}{P_m(\tau_j) \Gamma(m/2)} \int_0^{\sqrt{\hat{q}\tau_j}} r^{m-1} e^{-r^2/2} dr \\ &= \frac{\hat{q}^{-m/2}}{P_m(\tau_j) 2^{m/2-1} \Gamma(m/2)} \int_0^{\sqrt{\hat{q}\tau_j}} r^{m-1} e^{-r^2/2} dr \\ &= \frac{\hat{q}^{-m/2}}{P_m(\tau_j)} \Pr \left[\chi_m < \sqrt{\hat{q}\tau_j} \right] \\ &= \frac{\hat{q}^{-m/2}}{P_m(\tau_j)} \Pr \left[\chi_m^2 < \hat{q}\tau_j \right] \\ \mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}}) &= \hat{q}^{-m/2} \frac{P_m(\hat{q}\tau_j)}{P_m(\tau_j)}. \end{aligned} \quad (4.28)$$

Therefore,

$$\begin{aligned} \mathcal{K}(\mathbf{0}, \hat{\mathbf{T}}) &= \sum_{j=1}^n \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}) = \sum_{j=1}^n \ln \mathcal{M}_j(\mathbf{0}, \hat{\mathbf{T}}) \\ &= \sum_{j=1}^n \ln \left(\hat{q}^{-m/2} \frac{P_m(\hat{q}\tau_j)}{P_m(\tau_j)} \right). \end{aligned} \quad (4.29)$$

At the saddlepoint, $\hat{\mathbf{T}} = \hat{t}\mathbf{I}_m$, where \hat{t} is the solution to $n - p = \sum_{j=1}^n \partial \mathcal{K}_j(\mathbf{0}, \mathbf{T}) / \partial t_{11}$. To find this derivative, start from this formula, where $t_{ik} = 0$ when $i < k$, but all t_{ii} are distinct:

$$\mathcal{K}_j(\mathbf{0}, \mathbf{T}) = \ln \left[\frac{1}{(2\pi)^{m/2} P_m(\tau_j)} \int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left\{ \sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i^2 \right\} dz \right].$$

The first derivative by t_{11} is

$$\frac{\partial \mathcal{K}_j(\mathbf{0}, \mathbf{T})}{\partial t_{11}} = \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^2 \exp \left[\sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i^2 \right] dz}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left[\sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i^2 \right] dz}. \quad (4.30)$$

Now, allowing $\hat{t} = t_{ii}$ for all i and $\hat{q} = 1 - 2\hat{t}$,

$$\frac{\partial \mathcal{K}_j(\mathbf{0}, \mathbf{T})}{\partial t_{11}} = \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^2 \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z}}.$$

Using the fact that $P_m(x)$ is the χ_m^2 CDF,

$$\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z} = (2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j),$$

and

$$\frac{\partial \mathcal{K}_j(\mathbf{0}, \mathbf{T})}{\partial t_{11}} = \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^2 \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z}}{(2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j)}$$

To make these formulas more compact, we define this new notation:

$$\begin{aligned} I_j \{g(\mathbf{z})\} &:= \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} g(\mathbf{z}) \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z}} \\ &= \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} g(\mathbf{z}) \exp \{ -\hat{q} \mathbf{z}^T \mathbf{z} / 2 \} d\mathbf{z}}{(2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j)}, \end{aligned} \quad (4.31)$$

so that

$$\frac{\partial \mathcal{K}_j(\mathbf{0}, \mathbf{T})}{\partial t_{11}} = I_j \{z_1^2\}.$$

$I_j \{z_1^2\}$ can be derived from the identity that $I_j \{\mathbf{z}^T \mathbf{z}\} = m I_j \{z_1^2\}$, because of the interchangeability of columns of \mathbf{Z} . Then,

$$I_j \{\mathbf{z}^T \mathbf{z}\} (2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j) = \int_{\mathbf{z}^T \mathbf{z} < \tau_j} \mathbf{z}^T \mathbf{z} \exp [-\hat{q} \mathbf{z}^T \mathbf{z} / 2] d\mathbf{z}.$$

Since the integrand is rotationally invariant, we apply the same change of variables as in (4.26) and the surface area of $S^{m-1}(r)$ from (4.27) so that

$$I_j \{\mathbf{z}^T \mathbf{z}\} (2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j) = \left(\frac{2\pi}{\hat{q}} \right)^{m/2} \frac{m}{\hat{q}} P_{m+2}(\hat{q}\tau_j),$$

and

$$I_j \{z_1^2\} = \frac{\partial \mathcal{K}_j(\mathbf{0}, \mathbf{T})}{\partial t_{11}} = \frac{P_{m+2}(\hat{q}\tau_j)}{\hat{q}P_m(\hat{q}\tau_j)}.$$

Therefore, the saddlepoint equation is the solution to

$$n - p = \sum_{j=1}^n \frac{P_{m+2}(\hat{q}\tau_j)}{\hat{q}P_m(\hat{q}\tau_j)}. \quad (4.32)$$

The Hessian matrix $\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})$ is a possibly large matrix, with $m(p + (m + 1)/2)$ rows and columns, each corresponding to a unique element of \mathbf{S} and \mathbf{T} . Fortunately for computation, the matrix is block diagonal at the saddlepoint, with this form:

$$\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}}) = \begin{matrix} & mp & m & \frac{m(m-1)}{2} \\ \begin{matrix} mp \\ m \\ \frac{m(m-1)}{2} \end{matrix} & \left[\begin{array}{ccc} \mathcal{K}_{\mathbf{SS}}''(\mathbf{0}, \hat{\mathbf{T}}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}_{t_{ii}}''(\mathbf{0}, \hat{\mathbf{T}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{K}_{t_{i < k}}''(\mathbf{0}, \hat{\mathbf{T}}) \end{array} \right] \end{matrix}$$

In the top left of the Hessian, $\mathcal{K}_{\mathbf{SS}}''(\mathbf{0}, \hat{\mathbf{T}}) = \partial^2 \mathcal{K}(\mathbf{S}, \mathbf{T}) / \partial \text{vec}(\mathbf{S}) \partial \text{vec}(\mathbf{S})^T |_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}}$. Because $\partial \mathcal{K}(\mathbf{S}, \mathbf{T}) / \partial s_{ik} = 0$ at the saddlepoint for every element of \mathbf{S} , $\mathcal{K}_{\mathbf{SS}}''(\mathbf{0}, \hat{\mathbf{T}})$ is a diagonal matrix. Starting from (4.25),

$$\begin{aligned} \mathcal{M}_j(\mathbf{S}, \mathbf{T}) &= \frac{1}{(2\pi)^{m/2} P_m(\tau_j)} \times \\ &\quad \int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp [\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr} \{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T \}] dz \\ \mathcal{K}_j(\mathbf{S}, \mathbf{T}) &= \ln \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp [\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr} \{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T \}] dz}{(2\pi)^{m/2} P_m(\tau_j)} \\ \frac{\partial \mathcal{K}_j}{\partial \mathbf{S}} &= \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \mathbf{c}_j \mathbf{z}^T \exp [\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr} \{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T \}] dz}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp [\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr} \{ (\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T \}] dz} \end{aligned}$$

$$\frac{\partial^2 \mathcal{K}_j}{\partial \mathbf{S}^2} = \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \text{vec}(\mathbf{c}_j \mathbf{z}^T) \text{vec}(\mathbf{c}_j \mathbf{z}^T)^T \exp[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr}\{(\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T\}] d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr}\{(\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T\}] d\mathbf{z}} - \left(\frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \text{vec}(\mathbf{c}_j \mathbf{z}^T) \exp[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr}\{(\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T\}] d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr}\{(\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T\}] d\mathbf{z}} \right) \times \left(\frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \text{vec}(\mathbf{c}_j \mathbf{z}^T) \exp[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr}\{(\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T\}] d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp[\text{tr}(\mathbf{S}^T \mathbf{c}_j \mathbf{z}^T) + \text{tr}\{(\mathbf{T} - \frac{1}{2} \mathbf{I}) \mathbf{z} \mathbf{z}^T\}] d\mathbf{z}} \right)^T$$

At the saddlepoint, where $\hat{\mathbf{S}} = \mathbf{0}$ and $\hat{\mathbf{T}} = \hat{t} \mathbf{I}$, $\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_j \exp \sum_{i=1}^m (\hat{t} - \frac{1}{2}) z_i d\mathbf{z} = 0$ for all j .

Based on this fact,

$$\left. \frac{\partial^2 \mathcal{K}_j}{\partial s_{ik} \partial s_{lo}} \right|_{\mathbf{S}=\hat{\mathbf{S}}, \mathbf{T}=\hat{t} \mathbf{I}} = 0$$

when $\{i, k\} \neq \{l, o\}$, and

$$\left. \frac{\partial^2 \mathcal{K}_j}{\partial s_{ik}^2} \right|_{\mathbf{S}=\hat{\mathbf{S}}, \mathbf{T}=\hat{t} \mathbf{I}} = c_{ij}^2 \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_k^2 \exp \left\{ \sum_{i=1}^m (t_{ii} - \frac{1}{2}) z_i^2 \right\} d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left\{ \sum_{i=1}^m (t_{ii} - \frac{1}{2}) z_i^2 \right\} d\mathbf{z}} = c_{ij}^2 \left. \frac{\partial \mathcal{K}_j}{\partial t_{kk}} \right|_{\mathbf{S}=\hat{\mathbf{S}}, \mathbf{T}=\hat{t} \mathbf{I}}$$

Therefore,

$$\left. \frac{\partial^2 \mathcal{K}}{\partial s_{ik}^2} \right|_{\mathbf{S}=\hat{\mathbf{S}}, \mathbf{T}=\hat{t} \mathbf{I}} = \sum_{j=1}^n c_{ij}^2 I_j \{z_1^2\},$$

for $i = 1, \dots, p$ and $k = 1, \dots, m$. Of the mp diagonal elements of $\mathcal{K}_{\text{SS}}''(\mathbf{0}, \hat{\mathbf{T}})$, there are no more than p distinct values, each with multiplicity m . For a balanced design, all diagonal elements of $\mathcal{K}_{\text{SS}}''(\mathbf{0}, \hat{\mathbf{T}})$ will be the same. The determinant of this portion of the Hessian is

$$\left| \mathcal{K}_{\text{SS}}''(\mathbf{0}, \hat{\mathbf{T}}) \right| = \left(\prod_{i=1}^p \sum_{j=1}^n c_{ij}^2 I_j \{z_1^2\} \right)^m. \quad (4.33)$$

In the center of the Hessian, $\mathcal{K}_{t_{ii}}''(\mathbf{0}, \hat{\mathbf{T}})$ is the second derivative by the diagonal elements of \mathbf{T} at the saddlepoint. This matrix is not diagonal, but it has a pattern $a\mathbf{I} + b\mathbf{1}\mathbf{1}^T$ for some a and b , owing to the interchangeability of the m rows of \mathbf{Z} . This means that all diagonal elements of $\mathcal{K}_{t_{ii}}''(\mathbf{0}, \hat{\mathbf{T}})$ have one value, and all off-diagonal elements have the same different value.

To calculate $\partial^2 \mathcal{K}_j(\mathbf{0}, \mathbf{T}) / \partial t_{11}^2$, start from (4.30) and take one more derivative to get

$$\frac{\partial^2 \mathcal{K}_j(\mathbf{0}, \mathbf{T})}{\partial t_{11}^2} = \frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^4 \exp \left\{ \sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i \right\} d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left\{ \sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i^2 \right\} d\mathbf{z}} - \left(\frac{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^2 \exp \left\{ \sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i^2 \right\} d\mathbf{z}}{\int_{\mathbf{z}^T \mathbf{z} < \tau_j} \exp \left\{ \sum_{i=1}^m \left(t_{ii} - \frac{1}{2} \right) z_i^2 \right\} d\mathbf{z}} \right)^2.$$

Now, allowing $\hat{t} = t_{ii}$ for all i and $\hat{q} = 1 - 2\hat{t}$,

$$\frac{\partial^2 \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}})}{\partial t_{11}^2} = I_j \{z_1^4\} - (I_j \{z_1^2\})^2$$

Therefore,

$$\frac{\partial^2 \mathcal{K}(\mathbf{0}, \hat{\mathbf{T}})}{\partial t_{11}^2} = \sum_{j=1}^n I_j \{z_1^4\} - (I_j \{z_1^2\})^2. \quad (4.34)$$

For second and higher derivatives, it will be useful to have a general formula for $I_j \{z_1^\nu\}$, where ν is an even integer. This expression can be simplified in this way:

$$\begin{aligned} I_j \{z_1^\nu\} &= \frac{1}{(2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j)} \int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^\nu \exp \left\{ -\frac{\hat{q}\mathbf{z}^T \mathbf{z}}{2} \right\} d\mathbf{z} \\ &= \frac{1}{(2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j)} \int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z_1^\nu e^{-\hat{q}z_1^2/2} \int_{\mathbf{z}_{-1}^T \mathbf{z}_{-1} < \tau_j - z_1^2} \exp \left\{ -\frac{\hat{q}\mathbf{z}_{-1}^T \mathbf{z}_{-1}}{2} \right\} d\mathbf{z}_{-1} dz_1, \end{aligned}$$

where $\mathbf{z}_{-1} = (z_2, \dots, z_m)^T$. Substituting P_{m-1} for the inner integral,

$$\begin{aligned} I_j \{z_1^\nu\} &= \frac{1}{(2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j)} \int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z_1^\nu e^{-\hat{q}z_1^2/2} \left(\frac{2\pi}{\hat{q}} \right)^{(m-1)/2} P_{m-1} \{ \hat{q}(\tau_j - z_1^2) \} dz_1 \\ &= \frac{1}{(2\pi/\hat{q})^{1/2} P_m(\hat{q}\tau_j)} \int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z_1^\nu e^{-\hat{q}z_1^2/2} P_{m-1} \{ \hat{q}(\tau_j - z_1^2) \} dz_1. \end{aligned} \quad (4.35)$$

For a given value of m , it is possible to evaluate the remaining univariate integral into an integral-free formula, but not a simple one. For practical reasons, we chose to numerically integrate the remaining integral using the R function `integrate` in the stats package. Despite the incomplete gamma function P_{m-1} in the integrand, the integral evaluates rapidly using `pchisq` for P_{m-1} .

The remaining terms in the Hessian matrix require this expression:

$$\begin{aligned}
I_j \{z_1^2 z_2^2\} &= \frac{1}{(2\pi/\hat{q})^{m/2} P_m(\hat{q}\tau_j)} \int_{\mathbf{z}^T \mathbf{z} < \tau_j} z_1^2 z_2^2 \exp\left\{\frac{-\hat{q}\mathbf{z}^T \mathbf{z}}{2}\right\} d\mathbf{z} \\
&= \frac{1}{(2\pi/\hat{q})^{1/2} P_m(\hat{q}\tau_j) (m-1)} \times \\
&\quad \left\{ \left(\frac{2\pi}{\hat{q}}\right)^{1/2} \frac{(m+2)P_{m+4}(q\tau_j)}{\hat{q}^2} - \int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z^4 e^{-\hat{q}z^2/2} P_{m-1}\{\hat{q}(\tau_j - z^2)\} dz \right\}.
\end{aligned} \tag{4.36}$$

This formula derives from the identity that

$$I_j \{(\mathbf{z}^T \mathbf{z})^2\} = m I_j \{z_1^4\} + m(m-1) I_j \{z_1^2 z_2^2\}, \tag{4.37}$$

a convenience induced by the interchangeability of columns in \mathbf{Z} . The left-hand side in (4.37) can be evaluated, again using the $(m-1)$ -sphere surface area substitution, so that

$$I_j \{(\mathbf{z}^T \mathbf{z})^2\} = \frac{m(m+2)}{\hat{q}^2} \frac{P_{m+4}(\hat{q}\tau_j)}{P_m(\hat{q}\tau_j)}. \tag{4.38}$$

Also, $I_j \{z_1^4\}$ is calculated according to (4.35). From these formulas, $I_j \{z_1^2 z_2^2\}$ can be shown to have the formula in (4.36).

The off-diagonal elements of $\mathcal{K}_{\mathbf{t}_{ii}}''(\mathbf{0}, \hat{\mathbf{T}})$ are all equal to $\partial^2 \mathcal{K}(\mathbf{0}, \hat{\mathbf{T}})/\partial t_{11} \partial t_{22}$. With $\hat{\mathbf{S}} = \mathbf{0}$,

$$\left. \frac{\partial^2}{\partial t_{11} \partial t_{22}} \mathcal{K}_j(\mathbf{0}, \mathbf{T}) \right|_{\mathbf{T}=\hat{\mathbf{T}}} = I_j \{z_1^2 z_2^2\} - (I_j \{z_1^2\})^2.$$

After summing over j ,

$$\left. \frac{\partial^2}{\partial t_{11} \partial t_{22}} \mathcal{K}(\mathbf{0}, \mathbf{T}) \right|_{\mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^2 z_2^2\} - (I_j \{z_1^2\})^2. \tag{4.39}$$

Because of the pattern in $\mathcal{K}_{\mathbf{t}_{ii}}''(\mathbf{0}, \hat{\mathbf{T}})$, its determinant is formulaic, based on the fact that

$$|a\mathbf{I}_m + b\mathbf{1}\mathbf{1}^T| = a^m \left(1 + m\frac{b}{a}\right).$$

Using the derivative forms from (4.34) for $a + b$, and (4.39) for b ,

$$\begin{aligned} \left| \mathcal{K}''_{\mathbf{t}_{ii}}(\mathbf{0}, \hat{\mathbf{T}}) \right| &= \left(\sum_{j=1}^n I_j \{z_1^4\} - I_j \{z_1^2 z_2^2\} \right)^m \times \\ &\quad \left(1 + m \frac{\sum_{j=1}^n I_j \{z_1^2 z_2^2\} - (I_j \{z_1^2\})^2}{\sum_{j=1}^n I_j \{z_1^4\} - I_j \{z_1^2 z_2^2\}} \right). \end{aligned} \quad (4.40)$$

In the bottom right of the Hessian, $\mathcal{K}''_{\mathbf{t}_{i < k}}(\mathbf{0}, \hat{\mathbf{T}})$ is the second derivative by all $m(m-1)/2$ unique off-diagonal elements of \mathbf{T} . Because the solution to the saddlepoint equation, $\hat{\mathbf{T}}$, is a diagonal matrix, it follows that the off-diagonal elements of $\mathcal{K}''_{\mathbf{t}_{i < k}}(\mathbf{0}, \hat{\mathbf{T}})$, representing $\partial^2 \mathcal{K}_j(\mathbf{0}, \mathbf{T}) / \partial t_{ik} \partial t_{lh}$, where $(i, k) \neq (l, h)$, are all 0. Also, the diagonal elements are all equal, because of the interchangeability of the m rows of \mathbf{Z} . At the saddlepoint, the diagonal elements of this matrix are

$$\left. \frac{\partial^2}{\partial t_{12}^2} \mathcal{K}_j(\mathbf{0}, \mathbf{T}) \right|_{\mathbf{T}=\hat{\mathbf{T}}} = I_j \{z_1^2 z_2^2\}$$

Given this value,

$$\left. \frac{\partial^2}{\partial t_{12}^2} \mathcal{K}(\mathbf{0}, \mathbf{T}) \right|_{\mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^2 z_2^2\},$$

and

$$\left| \mathcal{K}''_{\mathbf{t}_{i < k}}(\mathbf{0}, \hat{\mathbf{T}}) \right| = \left(\sum_{j=1}^n I_j \{z_1^2 z_2^2\} \right)^{m(m-1)/2}.$$

With these pieces calculated, the determinant of the Hessian matrix is

$$\begin{aligned} \left| \mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}}) \right| &= \left(\prod_{i=1}^p \sum_{j=1}^n c_{ij}^2 I_j \{z_1^2\} \right)^m \left(\sum_{j=1}^n I_j \{z_1^4\} - I_j \{z_1^2 z_2^2\} \right)^m \times \\ &\quad \left(1 + m \frac{\sum_{j=1}^n I_j \{z_1^2 z_2^2\} - (I_j \{z_1^2\})^2}{\sum_{j=1}^n I_j \{z_1^4\} - I_j \{z_1^2 z_2^2\}} \right) \left(\sum_{j=1}^n I_j \{z_1^2 z_2^2\} \right)^{(m-1)/2}. \end{aligned} \quad (4.41)$$

For the second-order saddlepoint approximation, all third- and fourth-order derivatives of $\mathcal{K}(\mathbf{S}, \mathbf{T})$ are required at the saddlepoint. Many of these derivatives are 0, because of the symmetry in $\mathcal{K}(\mathbf{S}, \mathbf{T})$. Any odd-order derivative of any element of \mathbf{S} or of any off-diagonal element of \mathbf{T} is zero. The non-zero derivatives are listed here, without details of their derivations. Terms

of the form $I_j \{z_1^\nu\}$ are evaluated using (4.35), and $I_j \{z_1^2 z_2^2\}$ is given in (4.36). A useful tool for verifying these formulas is the polar integration function `adaptIntegrateBallPolar` in the `SphericalCubature` library, Nolan (2017), for R.

$$I_j \{z_1^\nu \mathbf{z}^T \mathbf{z}\} = \frac{\int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z^{\nu+2} e^{-\frac{qz^2}{2}} P_{m-1}(\hat{q}(\tau_j - z^2)) dz + \frac{m-1}{\hat{q}} \int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z^\nu e^{-\frac{qz^2}{2}} P_{m+1}(\hat{q}(\tau_j - z^2)) dz}{(2\pi/\hat{q})^{1/2} P_m(\hat{q}\tau_j)}$$

$$I_j \{(\mathbf{z}^T \mathbf{z})^\nu\} = \left(\frac{2}{\hat{q}}\right)^\nu \frac{\Gamma(m/2 + \nu)}{\Gamma(m/2)} \frac{P_{m+2\nu}(\hat{q}\tau_j)}{P_m(\hat{q}\tau_j)}$$

$$I_j \{z_1^\nu z_2^2\} = \frac{1}{m-1} [I_j \{z_1^\nu \mathbf{z}^T \mathbf{z}\} - I_j \{z_1^{\nu+2}\}]$$

$$I_j \{z_1^2 z_2^2 z_3^2\} = \frac{I_j \{(\mathbf{z}^T \mathbf{z})^3\}}{m(m-1)(m-2)} - \frac{3I_j \{z_1^4 z_2^2\}}{m-2} - \frac{I \{z_1^6\}}{(m-1)(m-2)}$$

$$I_j \{z_1^4 z_2^4\} = \begin{cases} \frac{9P_{10}(\hat{q}\tau_j)}{\hat{q}^4 P_2(\hat{q}\tau_j)}, & m = 2 \\ \frac{3\hat{q}}{128P_m(\hat{q}\tau_j)} \int_0^{\sqrt{\tau_j}} \rho^9 e^{-\hat{q}\rho^2/2} P_{m-2}(\hat{q}(\tau_j - \rho^2)) d\rho, & m > 2 \end{cases}$$

$$I_j \{z_1^4 z_2^2 z_3^2\} = \left(\frac{2\pi}{\hat{q}}\right)^{-1/2} \frac{(m+1)}{\hat{q}^2 (m-2) P_m(\hat{q}\tau_j)} \times \int_{-\sqrt{\tau_j}}^{\sqrt{\tau_j}} z^4 e^{-\hat{q}z^2/2} P_{m+3}(\hat{q}(\tau_j - z^2)) dz - \frac{I_j \{z_1^4 z_2^4\}}{(m-2)}$$

$$I_j \{z_1^2 z_2^2 z_3^2 z_4^2\} = \frac{I_j \{(\mathbf{z}^T \mathbf{z})^4\}}{m(m-1)(m-2)(m-3)} - \frac{I_j \{z_1^8\}}{(m-1)(m-2)(m-3)} - \frac{4I_j \{z_1^6 z_2^2\} + 3I_j \{z_1^4 z_2^4\}}{(m-2)(m-3)} - \frac{6I_j \{z_1^4 z_2^2 z_3^2\}}{m-3}$$

$$\frac{\partial^3}{\partial s_{ik}^2 \partial t_{kk}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \left\{ I_j \{z_1^4\} - (I_j \{z_1^2\})^2 \right\}$$

for $i = 1, \dots, p$ and $k = 1, \dots, m$

$$\frac{\partial^3}{\partial s_{ik}^2 \partial t_{ll}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \left\{ I_j \{z_1^2 z_2^2\} - (I_j \{z_1^2\})^2 \right\}$$

for $i = 1, \dots, p$, and $k, l = 1, \dots, m, k \neq l$

$$\frac{\partial^3}{\partial t_{ii}^3} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^6\} - 3I_j \{z_1^4\} I_j \{z_1^2\} + 2(I_j \{z_1^2\})^3$$

for $i = 1, \dots, m$

$$\frac{\partial^3}{\partial t_{ii}^2 \partial t_{kk}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^4 z_2^2\} - (2I_j \{z_1^2 z_2^2\} + I_j \{z_1^4\}) I_j \{z_1^2\} + 2(I_j \{z_1^2\})^3$$

for $i, k = 1, \dots, m, i \neq k$

$$\frac{\partial^3}{\partial t_{ii} \partial t_{kk} \partial t_{ll}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^2 z_2^2 z_3^2\} - 3I_j \{z_1^2 z_2^2\} I_j \{z_1^2\} + 2(I_j \{z_1^2\})^3$$

for $i, k, l = 1, \dots, m, i \neq k \neq l \neq i$

$$\begin{aligned} \frac{\partial^3}{\partial t_{ik}^2 \partial t_{ii}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} &= \frac{\partial^3}{\partial t_{ik}^2 \partial t_{kk}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} \\ &= \sum_{j=1}^n I_j \{z_1^4 z_2^2\} - I_j \{z_1^2 z_2^2\} I_j \{z_1^2\} \end{aligned}$$

for $i, k = 1, \dots, m, i \neq k$

$$\frac{\partial^3}{\partial t_{ik}^2 \partial t_{ll}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^2 z_2^2 z_3^2\} - I_j \{z_1^2 z_2^2\} I_j \{z_1^2\}$$

for $i, k, l = 1, \dots, m, i \neq k \neq l \neq i$

$$\frac{\partial^4}{\partial s_{ik}^4} \mathcal{K}_j(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^4 \left[I_j \{z_1^4\} - 3 (I_j \{z_1^2\})^2 \right]$$

for $i = 1, \dots, p$ and $k = 1, \dots, m$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial s_{lk}^2} \mathcal{K}_j(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 c_{lj}^2 \left[I_j \{z_1^4\} - 3 (I_j \{z_1^2\})^2 \right]$$

for $i, l = 1, \dots, p, i \neq l$, and $k = 1, \dots, m$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial s_{lo}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 c_{lj}^2 \left[I_j \{z_1^2 z_2^2\} - (I_j \{z_1^2\})^2 \right]$$

for $i, l = 1, \dots, p$ and $k, o = 1, \dots, m, k \neq o$

$$\frac{\partial^4}{\partial t_{ii}^4} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^8\} - 4 I_j \{z_1^6\} I_j \{z_1^2\} - 3 (I_j \{z_1^4\})^2 +$$

$$12 I_j \{z_1^4\} (I_j \{z_1^2\})^2 - 6 (I_j \{z_1^2\})^4$$

for $i = 1, \dots, m$

$$\begin{aligned} \frac{\partial^4}{\partial t_{ii}^3 \partial t_{kk}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} &= \sum_{j=1}^n I_j \{z_1^6 z_2^2\} - 3I_j \{z_1^4 z_2^2\} I_j \{z_1^2\} - \\ & 3I_j \{z_1^2 z_2^2\} I_j \{z_1^4\} + 6I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2 + \\ & 6I_j \{z_1^4\} (I_j \{z_1^2\})^2 - I_j \{z_1^6\} I_j \{z_1^2\} - 6(I_j \{z_1^2\})^4 \end{aligned}$$

for $i, k = 1, \dots, m, i \neq k$

$$\begin{aligned} \frac{\partial^4}{\partial t_{ii}^2 \partial t_{kk}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} &= \sum_{j=1}^n I_j \{z_1^4 z_2^4\} - 4I_j \{z_1^4 z_2^2\} I_j \{z_1^2\} - 2(I_j \{z_1^2 z_2^2\})^2 + \\ & 8I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2 - (I_j \{z_1^4\})^2 + \\ & 4I_j \{z_1^4\} (I_j \{z_1^2\})^2 - 6(I_j \{z_1^2\})^4 \end{aligned}$$

for $i, k = 1, \dots, m, i \neq k$

$$\begin{aligned} \frac{\partial^4}{\partial t_{ii}^2 \partial t_{kk} \partial t_{ll}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} &= \sum_{j=1}^n I_j \{z_1^4 z_2^2 z_3^2\} - 2I_j \{z_1^4 z_2^2\} I_j \{z_1^2\} - 2I_j \{z_1^2 z_2^2 z_3^2\} I_j \{z_1^2\} - \\ & 2(I_j \{z_1^2 z_2^2\})^2 - I_j \{z_1^2 z_2^2\} I_j \{z_1^4\} + \\ & 10I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2 + 2I_j \{z_1^4\} (I_j \{z_1^2\})^2 - 6(I_j \{z_1^2\})^4 \end{aligned}$$

for $i, k, l = 1, \dots, m, i \neq k \neq l \neq i$

$$\begin{aligned} \frac{\partial^4}{\partial t_{ii} \partial t_{kk} \partial t_{ll} \partial t_{oo}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} &= \sum_{j=1}^n I_j \{z_1^2 z_2^2 z_3^2 z_4^2\} - 4I_j \{z_1^2 z_2^2 z_3^2\} I_j \{z_1^2\} - \\ & 3(I_j \{z_1^2 z_2^2\})^2 + 12I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2 - 6(I_j \{z_1^2\})^4 \end{aligned}$$

for $i, k, l, o = 1, \dots, m$, all distinct

$$\frac{\partial^4}{\partial s_{ik}^2 \partial t_{kk}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \frac{\partial^3}{\partial t_{kk}^3} \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}), \text{ for } i = 1, \dots, p \text{ and } k = 1, \dots, m$$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial t_{kk} \partial t_{ll}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \frac{\partial^3}{\partial t_{kk}^2 \partial t_{ll}} \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}), \text{ for } i = 1, \dots, p \text{ and } k, l = 1, \dots, m$$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial t_{ll}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \frac{\partial^3}{\partial t_{kk} \partial t_{ll}^2} \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}), \text{ for } i = 1, \dots, p \text{ and } k, l = 1, \dots, m$$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial t_{ll} \partial t_{oo}} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \frac{\partial^3}{\partial t_{kk} \partial t_{ll} \partial t_{oo}} \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}),$$

for $i = 1, \dots, p$ and $k, l, o = 1, \dots, m, k \neq l \neq o \neq k$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial t_{kl}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \frac{\partial^3}{\partial t_{kk} \partial t_{kl}^2} \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}),$$

for $i = 1, \dots, p$, and $k, l = 1, \dots, m, k \neq l$

$$\frac{\partial^4}{\partial s_{ik}^2 \partial t_{lo}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n c_{ij}^2 \frac{\partial^3}{\partial t_{kk} \partial t_{lo}^2} \mathcal{K}_j(\mathbf{0}, \hat{\mathbf{T}}),$$

for $i = 1, \dots, p$, and $k, l, o = 1, \dots, m, k \neq l \neq o \neq k$

$$\frac{\partial^4}{\partial t_{ik}^4} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^4 z_2^4\} - 3 (I_j \{z_1^2 z_2^2\})^2$$

for $i, k = 1, \dots, m, i \neq k$

$$\begin{aligned} \frac{\partial^4}{\partial t_{ii}^2 \partial t_{ik}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} &= \sum_{j=1}^n I_j \{z_1^6 z_2^2\} - 2I_j \{z_1^4 z_2^2\} I_j \{z_1^2\} - \\ &I_j \{z_1^2 z_2^2\} I_j \{z_1^4\} + 2I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2 \end{aligned}$$

for $i, k = 1, \dots, m, i \neq k$

$$\frac{\partial^4}{\partial t_{ii} \partial t_{kk} \partial t_{ik}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^4 z_2^4\} - 2I_j \{z_1^4 z_2^2\} I_j \{z_1^2\} -$$

$$(I_j \{z_1^2 z_2^2\})^2 + 2I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2$$

for $i, k = 1, \dots, m, i \neq k$

$$\frac{\partial^4}{\partial t_{ii}^2 \partial t_{kl}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^4 z_2^2 z_3^2\} - 2I_j \{z_1^2 z_2^2 z_3^2\} I_j \{z_1^2\} -$$

$$(I_j \{z_1^2 z_2^2\})^2 + 2I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2$$

for $i, k, l = 1, \dots, m; i \neq k \neq l \neq i$

$$\frac{\partial^4}{\partial t_{ii} \partial t_{kk} \partial t_{il}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^4 z_2^2 z_3^2\} - I_j \{z_1^2 z_2^2 z_3^2\} I_j \{z_1^2\} -$$

$$I_j \{z_1^4 z_2^2\} I_j \{z_1^2\} - (I_j \{z_1^2 z_2^2\})^2 +$$

$$2I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2$$

for $i, k, l = 1, \dots, m; i \neq k \neq l \neq i$

$$\frac{\partial^4}{\partial t_{ii} \partial t_{kk} \partial t_{lo}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^2 z_2^2 z_3^2 z_4^2\} - 2I_j \{z_1^2 z_2^2 z_3^2\} I_j \{z_1^2\} -$$

$$(I_j \{z_1^2 z_2^2\})^2 + 2I_j \{z_1^2 z_2^2\} (I_j \{z_1^2\})^2$$

for $i, k, l, o = 1, \dots, m; i, k, l, o$ all distinct

$$\frac{\partial^4}{\partial t_{ik}^2 \partial t_{il}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^4 z_2^2 z_3^2\}$$

for $i, k, l = 1, \dots, m, i \neq k \neq l \neq i$

$$\frac{\partial^4}{\partial t_{ik}^2 \partial t_{lo}^2} \mathcal{K}(\mathbf{S}, \mathbf{T}) \Big|_{\mathbf{S}=\mathbf{0}, \mathbf{T}=\hat{\mathbf{T}}} = \sum_{j=1}^n I_j \{z_1^2 z_2^2 z_3^2 z_4^2\}$$

for $i, k, l, o = 1, \dots, m; i, k, l, o$ all distinct

4.5 Special Case: i.i.d.

In the common special case where the observed data \mathbf{Y} are independent and identically distributed, the design matrix $\mathbf{X} = \mathbf{1}_n$. This has the effect of streamlining the calculations required for the second-order saddlepoint approximation. The algorithm in the previous section requires a six-way nested loop to calculate $\hat{\kappa}_{23}^2$ and $\hat{\kappa}_{13}^2$. This calculation time is of order $O(m^{12})$ and is impractical for larger m . In this section, single formulas are presented for $\hat{\kappa}_4$, $\hat{\kappa}_{23}^2$, and $\hat{\kappa}_{13}^2$ only for this special case. Although the formulas are lengthy, their use shortens calculation time dramatically for this situation. For brevity in this section, $I\{g(\mathbf{z})\} = I_1\{g(\mathbf{z})\}$, since $I_j\{g(\mathbf{z})\}$ is constant for all $j = 1, \dots, n$.

In this special case, the Hessian matrix is

$$\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}}) = \begin{bmatrix} nI\{z_1^2\} \mathbf{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & n(I\{z_1^4\} - I\{z_1^2 z_2^2\}) \mathbf{I}_m + \\ & n(I\{z_1^2 z_2^2\} - I\{z_1^2\}^2) \mathbf{1}_m \mathbf{1}_m^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & nI\{z_1^2 z_2^2\} \mathbf{I}_{m(m-1)/2} \end{bmatrix}$$

with this inverse:

$$\{\mathcal{K}''(\hat{\mathbf{S}}, \hat{\mathbf{T}})\}^{-1} = \begin{bmatrix} (nI\{z_1^2\})^{-1} \mathbf{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}''_{t_{ii}}(\hat{\mathbf{S}}, \hat{\mathbf{T}})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (nI\{z_1^2 z_2^2\})^{-1} \mathbf{I}_{m(m-1)/2} \end{bmatrix}$$

where

$$\mathcal{K}''_{t_{ii}} \left(\hat{\mathbf{S}}, \hat{\mathbf{T}} \right)^{-1} = [n (I \{z_1^4\} - I \{z_1^2 z_2^2\})]^{-1} \mathbf{I}_m + \frac{I \{z_1^2\}^2 - I \{z_1^2 z_2^2\}}{n (I \{z_1^4\} - I \{z_1^2 z_2^2\}) (I \{z_1^4\} + (m-1) I \{z_1^2 z_2^2\} - m I \{z_1^2\}^2)} \mathbf{1}_m \mathbf{1}_m^T$$

From (4.20),

$$\begin{aligned} \hat{\kappa}_4 &:= \hat{\mathcal{K}}_{ijkl} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kl} \\ \hat{\kappa}_{23}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ir} \hat{\mathcal{K}}^{js} \hat{\mathcal{K}}^{kt} \\ \hat{\kappa}_{13}^2 &:= \hat{\mathcal{K}}_{ijk} \hat{\mathcal{K}}_{rst} \hat{\mathcal{K}}^{ij} \hat{\mathcal{K}}^{kr} \hat{\mathcal{K}}^{st} \end{aligned}$$

In the formulas that follow, factors of the form $\hat{\mathcal{K}}^{ij}$ from the center part of the inverse Hessian will be denoted as follows:

$$\begin{aligned} \hat{\mathcal{K}}^{t_{11}} &:= \frac{I \{z_1^4\} + (m-2) I \{z_1^2 z_2^2\} - (m-1) I \{z_1^2\}^2}{(I \{z_1^4\} - I \{z_1^2 z_2^2\}) (I \{z_1^4\} + (m-1) I \{z_1^2 z_2^2\} - m I \{z_1^2\}^2)} \\ \hat{\mathcal{K}}^{t_{12}} &:= \frac{I \{z_1^2\}^2 - I \{z_1^2 z_2^2\}}{(I \{z_1^4\} - I \{z_1^2 z_2^2\}) (I \{z_1^4\} + (m-1) I \{z_1^2 z_2^2\} - m I \{z_1^2\}^2)} \end{aligned}$$

Note that these formulas exclude the n in the denominator.

We also use the following symbols to represent various derivatives of \mathcal{K} :

$$\begin{aligned} \mathcal{D}_{s_1^2 t_{11}}^3 &:= I \{z_1^4\} - I \{z_1^2\}^2 \\ \mathcal{D}_{s_1^2 t_{22}}^3 &:= I \{z_1^2 z_2^2\} - I \{z_1^2\}^2 \\ \mathcal{D}_{t_{11}^3}^3 &:= I \{z_1^6\} - 3I \{z_1^4\} I \{z_1^2\} + 2I \{z_1^2\}^3 \\ \mathcal{D}_{t_{11}^2 t_{22}}^3 &:= I \{z_1^4 z_2^2\} - 2I \{z_1^2 z_2^2\} I \{z_1^2\} - I \{z_1^4\} I \{z_1^2\} + 2I \{z_1^2\}^3 \\ \mathcal{D}_{t_{11} t_{22} t_{33}}^3 &:= I \{z_1^2 z_2^2 z_3^2\} - 3I \{z_1^2 z_2^2\} I \{z_1^2\} + 2I \{z_1^2\}^3 \\ \mathcal{D}_{t_{12}^2 t_{11}}^3 &:= I \{z_1^4 z_2^2\} - I \{z_1^2 z_2^2\} I \{z_1^2\} \\ \mathcal{D}_{t_{12}^2 t_{33}}^3 &:= I \{z_1^2 z_2^2 z_3^2\} - I \{z_1^2 z_2^2\} I \{z_1^2\} \end{aligned}$$

$$\begin{aligned}
\mathcal{D}_{s_1^4}^4 &:= I \{z_1^4\} - 3I \{z_1^2\}^2 \\
\mathcal{D}_{s_1^2 s_2^2}^4 &:= I \{z_1^2 z_2^2\} - I \{z_1^2\}^2 \\
\mathcal{D}_{s_1^2 t_{11}}^4 &:= I \{z_1^6\} - 3I \{z_1^4\} I \{z_1^2\} + 2I \{z_1^2\}^3 \\
\mathcal{D}_{s_1^2 t_{11} t_{22}}^4 &:= \mathcal{D}_{s_1^2 t_{22}}^4 := I \{z_1^4 z_2^2\} - 2I \{z_1^2 z_2^2\} I \{z_1^2\} - I \{z_1^4\} I \{z_1^2\} + 2I \{z_1^2\}^3 \\
\mathcal{D}_{s_1^2 t_{22} t_{33}}^4 &:= I \{z_1^2 z_2^2 z_3^2\} - 3I \{z_1^2 z_2^2\} I \{z_1^2\} + 2I \{z_1^2\}^3 \\
\mathcal{D}_{s_1^2 t_{12}}^4 &:= I \{z_1^4 z_2^2\} - I \{z_1^2 z_2^2\} I \{z_1^2\} \\
\mathcal{D}_{s_1^2 t_{23}}^4 &:= I \{z_1^2 z_2^2 z_3^2\} - I \{z_1^2 z_2^2\} I \{z_1^2\} \\
\mathcal{D}_{t_{11}^4}^4 &:= I \{z_1^8\} - 4I \{z_1^6\} I \{z_1^2\} - 3I \{z_1^4\}^2 + 12I \{z_1^4\} I \{z_1^2\}^2 - 6I \{z_1^2\}^4 \\
\mathcal{D}_{t_{11}^4 t_{22}}^4 &:= I \{z_1^6 z_2^2\} - 3I \{z_1^4 z_2^2\} I \{z_1^2\} - 3I \{z_1^2 z_2^2\} I \{z_1^4\} + 6I \{z_1^2 z_2^2\} I \{z_1^2\}^2 + \\
&\quad 6I \{z_1^4\} I \{z_1^2\}^2 - I \{z_1^6\} I \{z_1^2\} - 6I \{z_1^2\}^4 \\
\mathcal{D}_{t_{11}^4 t_{22}^2}^4 &:= I \{z_1^4 z_2^4\} - 4I \{z_1^4 z_2^2\} I \{z_1^2\} - 2I \{z_1^2 z_2^2\}^2 + 8I \{z_1^2 z_2^2\} I \{z_1^2\}^2 - \\
&\quad I \{z_1^4\}^2 + 4I \{z_1^4\} I \{z_1^2\}^2 - 6I \{z_1^2\}^4 \\
\mathcal{D}_{t_{11}^4 t_{22} t_{33}}^4 &:= I \{z_1^4 z_2^2 z_3^2\} - 2I \{z_1^4 z_2^2\} I \{z_1^2\} - 2I \{z_1^2 z_2^2 z_3^2\} I \{z_1^2\} - 2I \{z_1^2 z_2^2\}^2 - \\
&\quad I \{z_1^2 z_2^2\} I \{z_1^4\} + 10I \{z_1^2 z_2^2\} I \{z_1^2\}^2 + 2I \{z_1^4\} I \{z_1^2\}^2 - 6I \{z_1^2\}^4 \\
\mathcal{D}_{t_{11}^4 t_{22} t_{33} t_{44}}^4 &:= I \{z_1^2 z_2^2 z_3^2 z_4^2\} - 4I \{z_1^2 z_2^2 z_3^2\} I \{z_1^2\} - 3I \{z_1^2 z_2^2\}^2 + \\
&\quad 12I \{z_1^2 z_2^2\} I \{z_1^2\}^2 - 6I \{z_1^2\}^4 \\
\mathcal{D}_{t_{12}^4}^4 &:= I \{z_1^4 z_2^4\} - 3I \{z_1^2 z_2^2\}^2 \\
\mathcal{D}_{t_{11}^4 t_{12}^2}^4 &:= I \{z_1^6 z_2^2\} - 2I \{z_1^4 z_2^2\} I \{z_1^2\} - I \{z_1^2 z_2^2\} I \{z_1^4\} + 2I \{z_1^2 z_2^2\} I \{z_1^2\}^2 \\
\mathcal{D}_{t_{11}^4 t_{22} t_{12}^2}^4 &:= I \{z_1^4 z_2^4\} - 2I \{z_1^4 z_2^2\} I \{z_1^2\} - I \{z_1^2 z_2^2\}^2 + 2I \{z_1^2 z_2^2\} I \{z_1^2\}^2 \\
\mathcal{D}_{t_{11}^4 t_{23}^2}^4 &:= I \{z_1^4 z_2^2 z_3^2\} - 2I \{z_1^2 z_2^2 z_3^2\} I \{z_1^2\} - I \{z_1^2 z_2^2\}^2 + 2I \{z_1^2 z_2^2\} I \{z_1^2\}^2 \\
\mathcal{D}_{t_{11}^4 t_{22} t_{13}^2}^4 &:= I \{z_1^4 z_2^2 z_3^2\} - I \{z_1^2 z_2^2 z_3^2\} I \{z_1^2\} - I \{z_1^4 z_2^2\} I \{z_1^2\} - I \{z_1^2 z_2^2\}^2 + \\
&\quad 2I \{z_1^2 z_2^2\} I \{z_1^2\}^2 \\
\mathcal{D}_{t_{11}^4 t_{22} t_{34}^2}^4 &:= I \{z_1^2 z_2^2 z_3^2 z_4^2\} - 2I \{z_1^2 z_2^2 z_3^2\} I \{z_1^2\} - I \{z_1^2 z_2^2\}^2 + 2I \{z_1^2 z_2^2\} I \{z_1^2\}^2 \\
\mathcal{D}_{t_{12}^4 t_{13}^2}^4 &:= I \{z_1^4 z_2^2 z_3^2\} \\
\mathcal{D}_{t_{12}^4 t_{34}^2}^4 &:= I \{z_1^2 z_2^2 z_3^2 z_4^2\}
\end{aligned}$$

In the formulas that follow, some terms will only become active when $m \geq 3, 4, 5$ or 6 .

Indicator functions of the form $1_{\{m \geq 3\}}$ denote these groups of terms.

$$\begin{aligned}
\hat{\kappa}_4 = & \frac{m}{n} \left[\mathcal{D}_{s_1^4}^4 I \{z_1^2\}^{-2} + 2\mathcal{D}_{s_1^2 t_{11}^2}^4 I \{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{11}} + \mathcal{D}_{t_{11}^4}^4 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \right] + \\
& \frac{m!}{(m-2)!n} \times \\
& \left[\mathcal{D}_{s_1^2 s_2^2}^4 I \{z_1^2\}^{-2} + \mathcal{D}_{s_1^2 t_{11} t_{22}}^4 I \{z_1^2\}^{-1} \left(2\hat{\mathcal{K}}^{t_{11}} + 4\hat{\mathcal{K}}^{t_{12}} \right) + \right. \\
& 2\mathcal{D}_{s_1^2 t_{12}^2}^4 I \{z_1^2\}^{-1} I \{z_1^2 z_2^2\}^{-1} + 4\mathcal{D}_{t_{11}^3 t_{22}}^4 \hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + \mathcal{D}_{t_{11}^2 t_{22}^2}^4 \left\{ \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 + 2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \right\} + \\
& \left. 2\mathcal{D}_{t_{11}^2 t_{12}^2}^4 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{11}} + 2\mathcal{D}_{t_{11} t_{22} t_{12}^2}^4 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} + \mathcal{D}_{t_{12}^4}^4 I \{z_1^2 z_2^2\}^{-2} \right] + \\
& 1_{\{m \geq 3\}} \frac{m!}{(m-3)!n} \times \\
& \left[\mathcal{D}_{s_1^2 t_{23}^2}^4 I \{z_1^2\}^{-1} I \{z_1^2 z_2^2\}^{-1} + 2\mathcal{D}_{s_1^2 t_{22} t_{33}}^4 I \{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} + \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^4 \left(2\hat{\mathcal{K}}^{t_{11}} + 4\hat{\mathcal{K}}^{t_{12}} \right) \hat{\mathcal{K}}^{t_{12}} + \right. \\
& \left. \mathcal{D}_{t_{11}^2 t_{23}^2}^4 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} + 4\mathcal{D}_{t_{11} t_{22} t_{13}^2}^4 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} + \mathcal{D}_{t_{12}^2 t_{13}^2}^4 I \{z_1^2 z_2^2\}^{-2} \right] + \\
& 1_{\{m \geq 4\}} \frac{m!}{(m-4)!n} \times \\
& \left[\mathcal{D}_{t_{11} t_{22} t_{33} t_{44}}^4 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \mathcal{D}_{t_{11} t_{22} t_{34}^2}^4 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} + \frac{1}{4} \mathcal{D}_{t_{12}^2 t_{34}^2}^4 I \{z_1^2 z_2^2\}^{-2} \right]
\end{aligned}$$

$$\begin{aligned}
\hat{\kappa}_{23}^2 &= \frac{m}{n} \left[3 \left(\mathcal{D}_{s_1^2 t_{11}}^3 \right)^2 I \{ z_1^2 \}^{-2} \hat{\mathcal{K}}^{t_{11}} + \left(\mathcal{D}_{t_{11}^3}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^3 \right] + \\
&\quad \frac{m!}{(m-2)!n} \times \\
&\quad \left[6 \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{s_1^2 t_{22}}^3 I \{ z_1^2 \}^{-2} \hat{\mathcal{K}}^{t_{12}} + 3 \left(\mathcal{D}_{s_1^2 t_{22}}^3 \right)^2 I \{ z_1^2 \}^{-2} \hat{\mathcal{K}}^{t_{11}} + 6 \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} + \right. \\
&\quad 3 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^3 + 6 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 6 \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \\
&\quad \left(\mathcal{D}_{t_{11}^3}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + 3 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + 6 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} + \\
&\quad \left. 3 \left(\mathcal{D}_{t_{12}^2 t_{11}}^3 \right)^2 I \{ z_1^2 z_2^2 \}^{-2} \left(\hat{\mathcal{K}}^{t_{11}} + \hat{\mathcal{K}}^{t_{12}} \right) \right] + \\
&\quad 1_{\{m \geq 3\}} \frac{m!}{(m-3)!n} \times \\
&\quad \left[3 \left(\mathcal{D}_{s_1^2 t_{22}}^3 \right)^2 I \{ z_1^2 \}^{-2} \hat{\mathcal{K}}^{t_{12}} + \right. \\
&\quad \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left\{ 3 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} + 21 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 12 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 \right\} + \\
&\quad 6 \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 6 \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + \\
&\quad 12 \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \left\{ \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} \right\} + \\
&\quad \left(\mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \right)^2 \left\{ \left(\hat{\mathcal{K}}^{t_{11}} \right)^3 + 3 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 \right\} + \\
&\quad \left. 6 \mathcal{D}_{t_{12}^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{ z_1^2 z_2^2 \}^{-2} \hat{\mathcal{K}}^{t_{12}} + \frac{3}{2} \left(\mathcal{D}_{t_{12}^2 t_{33}}^3 \right)^2 I \{ z_1^2 z_2^2 \}^{-2} \hat{\mathcal{K}}^{t_{11}} \right] + \\
&\quad 1_{\{m \geq 4\}} \frac{m!}{(m-4)!n} \times \\
&\quad \left[\frac{3}{2} \left(\mathcal{D}_{t_{12}^2 t_{33}}^3 \right)^2 I \{ z_1^2 z_2^2 \}^{-2} \hat{\mathcal{K}}^{t_{12}} + 2 \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + \right. \\
&\quad \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \left\{ 18 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + 18 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \right\} + 9 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + \\
&\quad \left. \left(\mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \right)^2 \left\{ 6 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 3 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} + 9 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 \right\} \right] \\
&\quad 1_{\{m \geq 5\}} \frac{m!}{(m-5)!n} \times \\
&\quad \left[6 \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + \left(\mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \right)^2 \left\{ 3 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 6 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 \right\} \right] + \\
&\quad 1_{\{m \geq 6\}} \frac{m!}{(m-6)!n} \left(\mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3
\end{aligned}$$

$$\begin{aligned}
\hat{\kappa}_{13}^2 &= \frac{m}{n} \left[\left(\mathcal{D}_{s_1^2 t_{11}}^3 \right)^2 I \{ z_1^2 \}^{-2} \hat{\kappa}^{t_{11}} + \left(\mathcal{D}_{t_{11}^3}^3 \right)^2 \left(\hat{\kappa}^{t_{11}} \right)^3 + 2 \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11}^3}^3 I \{ z_1^2 \}^{-1} \left(\hat{\kappa}^{t_{11}} \right)^2 \right] + \\
&\quad \frac{m!}{(m-2)!n} \times \\
&\quad \left[\left(\mathcal{D}_{s_1^2 t_{11}}^3 \right)^2 I \{ z_1^2 \}^{-2} \hat{\kappa}^{t_{12}} + \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{s_1^2 t_{22}}^3 I \{ z_1^2 \}^{-2} \left(2\hat{\kappa}^{t_{12}} + 2\hat{\kappa}^{t_{11}} \right) + \right. \\
&\quad 2 \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11}^3}^3 I \{ z_1^2 \}^{-1} \hat{\kappa}^{t_{11}} \hat{\kappa}^{t_{12}} + \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 I \{ z_1^2 \}^{-1} \hat{\kappa}^{t_{11}} \left(6\hat{\kappa}^{t_{12}} + 2\hat{\kappa}^{t_{11}} \right) + \\
&\quad 4 \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 I \{ z_1^2 \}^{-1} \left(\hat{\kappa}^{t_{12}} \right)^2 + \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{ z_1^2 \}^{-1} I \{ z_1^2 z_2^2 \}^{-1} \left(2\hat{\kappa}^{t_{11}} + 2\hat{\kappa}^{t_{12}} \right) + \\
&\quad \left(\mathcal{D}_{s_1^2 t_{22}}^3 \right)^2 I \{ z_1^2 \}^{-2} \left(\hat{\kappa}^{t_{11}} + \hat{\kappa}^{t_{12}} \right) + \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11}^3}^3 I \{ z_1^2 \}^{-1} \hat{\kappa}^{t_{11}} \left(2\hat{\kappa}^{t_{11}} + 2\hat{\kappa}^{t_{12}} \right) + \\
&\quad \left. + \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 I \{ z_1^2 \}^{-1} \left\{ 4 \left(\hat{\kappa}^{t_{12}} \right)^2 + 6\hat{\kappa}^{t_{11}} \hat{\kappa}^{t_{12}} + 2 \left(\hat{\kappa}^{t_{11}} \right)^2 \right\} + \right. \\
&\quad \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{ z_1^2 \}^{-1} I \{ z_1^2 z_2^2 \}^{-1} \left(2\hat{\kappa}^{t_{12}} + 2\hat{\kappa}^{t_{11}} \right) + \left(\mathcal{D}_{t_{11}^3}^3 \right)^2 \left(\hat{\kappa}^{t_{11}} \right)^2 \hat{\kappa}^{t_{12}} + \\
&\quad \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \left\{ 2 \left(\hat{\kappa}^{t_{11}} \right)^3 + 6 \left(\hat{\kappa}^{t_{11}} \right)^2 \hat{\kappa}^{t_{12}} \right\} + \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{ z_1^2 z_2^2 \}^{-1} \hat{\kappa}^{t_{11}} \left\{ 2\hat{\kappa}^{t_{11}} + 2\hat{\kappa}^{t_{12}} \right\} \\
&\quad \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\kappa}^{t_{11}} \right)^3 + 8 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \hat{\kappa}^{t_{11}} \left(\hat{\kappa}^{t_{12}} \right)^2 + 4 \mathcal{D}_{t_{11}^3}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \hat{\kappa}^{t_{11}} \left(\hat{\kappa}^{t_{12}} \right)^2 + \\
&\quad 4 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\kappa}^{t_{12}} \right)^3 + 5 \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\kappa}^{t_{11}} \right)^2 \hat{\kappa}^{t_{12}} \\
&\quad + \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{ z_1^2 z_2^2 \}^{-1} \left\{ 2 \left(\hat{\kappa}^{t_{11}} \right)^2 + 6\hat{\kappa}^{t_{11}} \hat{\kappa}^{t_{12}} + 4 \left(\hat{\kappa}^{t_{12}} \right)^2 \right\} \\
&\quad \left. \left(\mathcal{D}_{t_{12}^2 t_{11}}^3 \right)^2 I \{ z_1^2 z_2^2 \}^{-2} \left(\hat{\kappa}^{t_{11}} + \hat{\kappa}^{t_{12}} \right) \right] +
\end{aligned}$$

continued below

$$\begin{aligned}
& 1_{\{m \geq 3\}} \frac{m!}{(m-3)!n} \times \\
& \left[2\mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{s_1^2 t_{22}}^3 I\{z_1^2\}^{-2} \hat{\mathcal{K}}^{t_{12}} + \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 I\{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(2\hat{\mathcal{K}}^{t_{11}} + 4\hat{\mathcal{K}}^{t_{12}} \right) + \right. \\
& 2\mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I\{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} I\{z_1^2 z_2^2\}^{-1} + \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I\{z_1^2\}^{-1} I\{z_1^2 z_2^2\}^{-1} \left(\hat{\mathcal{K}}^{t_{11}} + 2\hat{\mathcal{K}}^{t_{12}} \right) + \\
& \left. \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 I\{z_1^2\}^{-1} \left\{ 2\hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + 4 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \right\} + \right. \\
& \left(\mathcal{D}_{s_1^2 t_{22}}^3 \right)^2 I\{z_1^2\}^{-2} \left(\hat{\mathcal{K}}^{t_{11}} + 3\hat{\mathcal{K}}^{t_{12}} \right) + 2\mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11}^2}^3 I\{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + \\
& \left. \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 I\{z_1^2\}^{-1} \left\{ 2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 + 12 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 10\hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} \right\} + \right. \\
& \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I\{z_1^2\}^{-1} I\{z_1^2 z_2^2\}^{-1} \left(6\hat{\mathcal{K}}^{t_{12}} + 2\hat{\mathcal{K}}^{t_{11}} \right) + \\
& \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I\{z_1^2\}^{-1} I\{z_1^2 z_2^2\}^{-1} \left(4\hat{\mathcal{K}}^{t_{12}} + 2\hat{\mathcal{K}}^{t_{11}} \right) + \\
& \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 I\{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left\{ 8\hat{\mathcal{K}}^{t_{12}} + 4\hat{\mathcal{K}}^{t_{11}} \right\} + \\
& \mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} + \\
& \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left\{ 7 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} + 16\hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 12 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 \right\} + \\
& \mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \left\{ 4\hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} \right\} + \mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \left\{ 6\hat{\mathcal{K}}^{t_{12}} + 4\hat{\mathcal{K}}^{t_{11}} \right\} + \\
& 2\mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I\{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + \mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I\{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{11}} + 2\hat{\mathcal{K}}^{t_{12}} \right) + \\
& \left(\mathcal{D}_{t_{11}^2 t_{22}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^3 + \\
& \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \left\{ 16 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + 16\hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 4 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 \hat{\mathcal{K}}^{t_{12}} \right\} + \\
& \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I\{z_1^2 z_2^2\}^{-1} \left\{ 2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 + 10\hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + 12 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \right\} + \\
& \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I\{z_1^2 z_2^2\}^{-1} \left\{ 8\hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + 2 \left(\hat{\mathcal{K}}^{t_{11}} \right)^2 + 8 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \right\} \\
& \left(\mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \right)^2 \left\{ \left(\hat{\mathcal{K}}^{t_{11}} \right)^3 + 2\hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + 4 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 \right\} + \\
& \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I\{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(8\hat{\mathcal{K}}^{t_{12}} + 4\hat{\mathcal{K}}^{t_{11}} \right) + \\
& \mathcal{D}_{t_{11}^2 t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I\{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(4\hat{\mathcal{K}}^{t_{12}} + 2\hat{\mathcal{K}}^{t_{11}} \right) +
\end{aligned}$$

continued below

$$\begin{aligned}
& 4\mathcal{D}_{t_{12}^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-2} \hat{\mathcal{K}}^{t_{12}} + \left(\mathcal{D}_{t_{12}^2 t_{33}}^3\right)^2 I \{z_1^2 z_2^2\}^{-2} \left(\frac{1}{2}\hat{\mathcal{K}}^{t_{11}} + \hat{\mathcal{K}}^{t_{12}}\right) + \\
& 2\mathcal{D}_{t_{12}^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-2} \hat{\mathcal{K}}^{t_{11}} + \left(\mathcal{D}_{t_{12}^2 t_{11}}^3\right)^2 I \{z_1^2 z_2^2\}^{-2} \left(\hat{\mathcal{K}}^{t_{11}} + 3\hat{\mathcal{K}}^{t_{12}}\right) \Big] + \\
& 1_{\{m \geq 4\}} \frac{m!}{(m-4)!n} \times \\
& \left[\mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} I \{z_1^2 z_2^2\}^{-1} + \mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 I \{z_1^2\}^{-1} \left(\hat{\mathcal{K}}^{t_{12}}\right)^2 + \right. \\
& \left. \left(\mathcal{D}_{s_1^2 t_{22}}^3\right)^2 I \{z_1^2\}^{-2} \hat{\mathcal{K}}^{t_{12}} + \right. \\
& \left. \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11}^2 t_{22}}^3 I \{z_1^2\}^{-1} \left\{4\left(\hat{\mathcal{K}}^{t_{12}}\right)^2 + 2\hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}}\right\} + \right. \\
& \left. \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2\}^{-1} I \{z_1^2 z_2^2\}^{-1} \left(5\hat{\mathcal{K}}^{t_{12}} + \hat{\mathcal{K}}^{t_{11}}\right) + \right. \\
& \left. \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 I \{z_1^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(10\hat{\mathcal{K}}^{t_{12}} + 2\hat{\mathcal{K}}^{t_{11}}\right) + \right. \\
& \left. \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{z_1^2\}^{-1} I \{z_1^2 z_2^2\}^{-1} 2\hat{\mathcal{K}}^{t_{12}} + \right. \\
& \left. 2\mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}}\right)^2 + \mathcal{D}_{t_{11}^2}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + \right. \\
& \left. \left(\mathcal{D}_{t_{11}^2 t_{22}}^3\right)^2 \left\{4\hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}}\right)^2 + 4\left(\hat{\mathcal{K}}^{t_{12}}\right)^3 + \left(\hat{\mathcal{K}}^{t_{11}}\right)^2 \hat{\mathcal{K}}^{t_{12}}\right\} + \right. \\
& \left. \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(2\hat{\mathcal{K}}^{t_{11}} + 4\hat{\mathcal{K}}^{t_{12}}\right) + \right. \\
& \left. \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-1} \left\{7\hat{\mathcal{K}}^{t_{11}} \hat{\mathcal{K}}^{t_{12}} + \left(\hat{\mathcal{K}}^{t_{11}}\right)^2 + 10\left(\hat{\mathcal{K}}^{t_{12}}\right)^2\right\} + \right. \\
& \left. \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \left\{20\left(\hat{\mathcal{K}}^{t_{12}}\right)^3 + 14\hat{\mathcal{K}}^{t_{11}} \left(\hat{\mathcal{K}}^{t_{12}}\right)^2 + 2\left(\hat{\mathcal{K}}^{t_{11}}\right)^2 \hat{\mathcal{K}}^{t_{12}}\right\} + \right. \\
& \left. \left(\mathcal{D}_{t_{12}^2 t_{11}}^3\right)^2 I \{z_1^2 z_2^2\}^{-2} \hat{\mathcal{K}}^{t_{12}} + \right. \\
& \left. \left(\mathcal{D}_{t_{11} t_{22} t_{33}}^3\right)^2 I \{z_1^2 z_2^2\}^{-2} \left(4\hat{\mathcal{K}}^{t_{11}} + 14I \{z_1^2 z_2^2\}^{-1}\right) + \right. \\
& \left. \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(10\hat{\mathcal{K}}^{t_{12}} + 2\hat{\mathcal{K}}^{t_{11}}\right) + \right. \\
& \left. \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(14\hat{\mathcal{K}}^{t_{12}} + 4\hat{\mathcal{K}}^{t_{11}}\right) + \right. \\
& \left. \mathcal{D}_{t_{12}^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-2} \left(5\hat{\mathcal{K}}^{t_{12}} + \hat{\mathcal{K}}^{t_{11}}\right) + \right. \\
& \left. \left(\mathcal{D}_{t_{12}^2 t_{33}}^3\right)^2 I \{z_1^2 z_2^2\}^{-2} \left(\hat{\mathcal{K}}^{t_{11}} + \frac{7}{2}\hat{\mathcal{K}}^{t_{12}}\right) \Big]
\end{aligned}$$

continued below

$$\begin{aligned}
& 1_{\{m \geq 5\}} \frac{m!}{(m-5)!n} \times \\
& \left[\mathcal{D}_{s_1^2 t_{11}}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 I \{z_1^2\}^{-1} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \right. \\
& \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2\}^{-1} I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} + \\
& 2 \mathcal{D}_{s_1^2 t_{22}}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 I \{z_1^2\}^{-1} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \\
& \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \left(4 \hat{\mathcal{K}}^{t_{12}} + 2 \hat{\mathcal{K}}^{t_{11}} \right) + \\
& \mathcal{D}_{t_{11}^2 t_{22}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(\hat{\mathcal{K}}^{t_{11}} + 2 \hat{\mathcal{K}}^{t_{12}} \right) + \\
& \left(\mathcal{D}_{t_{11} t_{22} t_{33}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 \left(\hat{\mathcal{K}}^{t_{11}} + 8 \hat{\mathcal{K}}^{t_{12}} \right) + \\
& \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{11}}^3 I \{z_1^2 z_2^2\}^{-1} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \\
& \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-1} \hat{\mathcal{K}}^{t_{12}} \left(8 \hat{\mathcal{K}}^{t_{11}} + \hat{\mathcal{K}}^{t_{12}} \right) + \\
& \mathcal{D}_{t_{12}^2 t_{11}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-2} \hat{\mathcal{K}}^{t_{12}} + \\
& \left. \left(\mathcal{D}_{t_{12}^2 t_{33}}^3 \right)^2 I \{z_1^2 z_2^2\}^{-2} \left(\frac{1}{4} \hat{\mathcal{K}}^{t_{11}} + 2 \hat{\mathcal{K}}^{t_{12}} \right) \right] +
\end{aligned}$$

$$\begin{aligned}
& 1_{\{m \geq 6\}} \frac{m!}{(m-6)!n} \times \\
& \left[\left(\mathcal{D}_{t_{11} t_{22} t_{33}}^3 \right)^2 \left(\hat{\mathcal{K}}^{t_{12}} \right)^3 + \right. \\
& \mathcal{D}_{t_{11} t_{22} t_{33}}^3 \mathcal{D}_{t_{12}^2 t_{33}}^3 I \{z_1^2 z_2^2\}^{-1} \left(\hat{\mathcal{K}}^{t_{12}} \right)^2 + \\
& \left. \frac{1}{4} \left(\mathcal{D}_{t_{12}^2 t_{33}}^3 \right)^2 I \{z_1^2 z_2^2\}^{-2} \hat{\mathcal{K}}^{t_{12}} \right]
\end{aligned}$$

4.6 Bonferroni Bounds

Bonferroni bounds and their application to outlier discordancy testing were explained in section 2.5. Only the details that are different for the regression setting are listed here.

4.6.1 Bonferroni Inequalities

In the multivariate regression setting of this chapter, $E_j = \{\mathbf{y}_j^T \mathbf{y}_j \leq x(1 - h_{jj})\}$, so $\{M \leq x\} = \bigcap_{j=1}^n E_j$. With this definition of events E_j , refer to section 2.5.1 for the remaining details.

4.6.2 Distribution Theory

The theory behind the computation of ρ_{ij} , S_1 , S_2 , and S_{2*} derives from the joint distribution of the studentized residuals. Srivastava and Van Rosen (1998) derive the density of T_j and the joint density of T_i and T_j , where $T_j = a_j^2/(n - p)$. After converting into our notation, the density of T_j is given by

$$\frac{n - p - m}{m} \frac{T_j}{1 - T_j} \sim F_{m, n-p-m}.$$

Therefore,

$$S_1 = n \left\{ 1 - F_{F_{m, n-p-m}} \left(\frac{(n - p - m)x}{m(n - p - x)} \right) \right\},$$

where x is the observed value of M , and $F_{F_{\mu, \nu}}$ is the CDF of the F distribution with μ and ν degrees of freedom.

The bivariate density of T_i and T_j is

$$\begin{aligned} & |\mathbf{R}|^{-\frac{n-p-3}{2}} \frac{\prod_{i=1}^m \left(\frac{n-p-1-i}{2} \right)}{\sqrt{\pi} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{m-1}{2}\right)} \times \\ & \int_a^b (T_i T_j - T_{ij})^{\frac{m-3}{2}} \left\{ (1 - T_i)(1 - T_j) - (\rho_{ij} - T_{ij})^2 \right\}^v dT_{ij}, \end{aligned} \quad (4.42)$$

where

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix} \\ \rho_{ij} &= \frac{-h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}} \\ a &= \max \left\{ -\sqrt{T_i T_j}, \rho_{ij} - \sqrt{(1 - T_i)(1 - T_j)} \right\} \\ b &= \min \left\{ \sqrt{T_i T_j}, \rho_{ij} + \sqrt{(1 - T_i)(1 - T_j)} \right\} \\ v &= \frac{n-p-m-3}{2} \end{aligned}$$

The sums S_2 and S_{2*} comprise probabilities of intersections $\Pr [a_i^2 > x, a_j^2 > x]$. Given the integral in (4.42), this is a triple integral over the region where $b > a$. The integral must be calculated separately for each distinct value of ρ_{ij} . For some balanced design where all values ρ_{ij} are equal, only one calculation is required, but for a general regression problem where each ρ_{ij} may be different, this calculation can consume a lot of time.

Some problems may have partially degenerate residual distributions, where $|\rho_{ij}| = 1$. For general-purpose applications, the code must detect these situations and substitute the univariate distribution $\Pr [a_i^2 > x]$ for $\Pr [a_i^2 > x, a_j^2 > x]$, as required.

References

- Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd ed. New York: John Wiley and Sons
- Bernoulli, D. (1777). "The most probable choice between several discrepant observations and the formation therefrom of the most likely indication," translated from Latin by C. G. Allen (1961). *Biometrika*, **48**, 1/2, 3-13
- Bjornstad, J. F. and Butler, R. W. (1988). "The Equivalence of Backward Elimination and Multiple Comparisons." *J Amer. Stat. Assoc.* **83**:401 (March) 136-144
- Butler, R. W. (1982). "Bounds on the Significance Attained by the Best-Fitting Regressor Variable." *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **31**(3), 290-292
- Butler, R. W. (1984). "The Significance Attained by the Best-Fitting Regressor Variable." *J Amer. Stat. Assoc.* **79**:386, (June) 341-348
- Butler, R. W. (1986). "Extendability and the Optimality of F, T_2 and Forward Variable Selection." *Scandinavian Journal of Statistics*, **13**, 4, pp 257-262
- Butler, R. W. and Sutton, R. K. (1998). "Saddlepoint Approximation for Multivariate Cumulative Distribution Functions and Probability Computations in Sampling Theory and Outlier Testing." *J Amer. Stat. Assoc.* **93**:442, 596-604
- Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge: Cambridge University Press
- Chauvenet, W. (1863). "Method of Least Squares," appendix to *Manual of Spherical and Practical Astronomy*, vol 2. Philadelphia: Lippincott. reprinted (1960) New York: Dover
- Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall
- Daniels, H. E. (1954). "Saddlepoint Approximations in Statistics." *Ann. Math. Stat.* **25**, 631-650

- Dempster, A. P. (1969). *Elements of Continuous Multivariate Analysis*. Reading, MA: Addison-Wesley
- Dawid, A. P. (1977). "Spherical Matrix Distributions and a Multivariate Model." *Journal of the Royal Statistical Society, Series B (Methodological)*, **39**, 2, 254-261
- Ferguson, T. S. (1961) "On the Rejection of Outliers." *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 253-287. Univ. of California.
- Gossett, W. S. "Student" (1908). "The Probable Error of a Mean." *Biometrika*, **6**(1), 1-25
- Grubbs, F. E. (1950). "Sample Criteria for Testing Outlying Observations." *Annals of Mathematical Statistics*, **21**, 27-58
- Grubbs, F. E. (1969). "Procedures for Detecting Outlying Observations in Samples." *Technometrics*, **11**, 1-21
- Hawkins, D. M. (1980). *Identification of Outliers*. London: Chapman and Hall
- Hossain, A. and Naik, D. N. (1989). "Detection of influential observations in multivariate regression." *Journal of Applied Statistics*, **16**, 25-37
- Huber, P. (1977). "Robust Statistical Procedures." No. 27, Regional conference series in applied mathematics. Philadelphia: SIAM
- Hunter, D. (1976). "An Upper Bound for the Probability of a Union." *Journal of Applied Probability*, **13**, 3 (Sep.), 597-603
- Jones, B., and Nachtsheim, C. J. (2011), "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects," *Journal of Quality Technology*, **43**, 1-15
- Kounias, E. G. (1968). "Bounds for the Probability of a Union, with Applications." *The Annals of Mathematical Statistics*, **39**, 6 (Dec.), 2154-2158
- McCullagh, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall/CRC
- Minitab 18.1 Statistical Software (2017) [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com)
- Montgomery, D. C. (2005). *Design and Analysis of Experiments*, 6th ed. Hoboken, NJ: John Wiley & Sons

- Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*. Hoboken, NJ: John Wiley & Sons
- Neter, J., Wasserman, W., and Kunter, M., (1985). *Applied Linear Statistical Models*. Homewood, IL: Irwin
- Nolan, J. P. (2017). SphericalCubature package for R [Computer software]. <https://cran.r-project.org/web/packages/SphericalCubature/index.html>
- Park, J. T. (2004). "Some Properties of a Certain Patterned Matrix." *J. Appl. Math & Computing*, Vol. 15, No. 1-2, pp 485-493
- Pearson, E. S. and Chandra Sekar, C. (1936). "The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations." *Biometrika* **28**, 308-320
- Plackett, R. L. and Burman, J. P. (1946). "The Design of Optimum Multifactorial Experiments." *Biometrika* **33** (4), pp. 305-25, June 1946 doi:10.1093/biomet/33.4.305
- Quesenberry, C. P. and David, H. A. (1961). "Some Tests for Outliers." *Biometrika* **48**, 379-390
- Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Cambridge, MA: MIT Press
- R Core Team (2019). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Ross, S. (1988). *A First Course in Probability*. New York: MacMillan Publishing
- Srikantan, K. S. (1961). "Testing for the Single Outlier in a Regression Model." *Sankhya: The Indian Journal of Statistics, Series A*, **23** (3), 251-260
- Srivastava, M. S. and von Rosen, D. (1998). "Outliers in multivariate regression models." *Journal of Multivariate Analysis*, **65**, 195-208
- Stefansky, W. (1969). "On the rejection of outliers by maximum normed residual." Ph.D. Thesis. Berkeley: University of California.
- Stefansky, W. (1971). "Rejecting outliers by maximum normed residual." *Ann. Math. Stat.* **42**, 35-45
- Stefansky, W. (1972). "Rejecting Outliers in Factorial Designs." *Technometrics* **14**, 469-479

Thompson, W. R. (1935). "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of Deviation to Sample Standard Deviation." *The Annals of Mathematical Statistics*, **6** (4), 214-219

Timm, N. H. (2002). *Applied Multivariate Analysis*. Springer

Tukey, J. E. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, pp 39-43

Wilks, S. S. (1963) "Multivariate Statistical Outliers." *Sankhya: The Indian Journal of Statistics, Series A*, **25** (4), 407-426

Worsley, K. J. (1982). "An Improved Bonferroni Inequality and Applications." *Biometrika*, **69**, 2, 297-302

Wright, T. W. (1884). *A Treatise on the Adjustment of Observations by the Method of Least Squares*. New York: Van Nostrand

Appendix A

R Code for Calculations

This appendix describes the R code written to perform the calculations for this project. A file containing the code is available on request from the author at AndrewDSleeper@gmail.com. The code has been tested and runs properly with R version 3.5.3 (2019-3-11) "Great Truth". Here are the top-level functions in the code, along with a description of required parameters. Note that only minimal error-checking of parameters is performed.

The following R libraries are required for this code to run:

- `NORMT3` provides the error function used for univariate calculations
- `cubature` is required for the Bonferroni bounds, when dimension $m > 1$
- `tictoc` is required for the optional reporting of calculation time

These discordancy test functions do not calculate the MASR or MSSR statistics. The regression, calculation of residuals, and determination of the maximum residual must be calculated prior to running these functions. The studentized residuals are calculated using equation (2.3) for the one-sample case, (3.3) for the univariate regression case, or (4.3) for multivariate regression. The maximum absolute or squared studentized residual is the first parameter required for these functions.

The second required parameter is the design matrix \mathbf{X} . In the one-sample case, the sample size n may be provided as a scalar value instead of matrix \mathbf{X} .

Example: In a set of $n = 30$ observations from an assumed normal distribution, the most outlying value lies at $\bar{y} + 3.00\hat{\sigma}$. Is this value significantly discordant?

First, note that 3.00 is a standardized residual, which must be converted to a studentized residual by equation (2.3). $M = \max_j |a_j| = 3.00 \times \sqrt{30/29} = 3.05$.

In R, calculate the p -value for a discordancy test this way:

```
> PMASR(3.05, 30) $P
```

```
[1] 0.03242239
```

Since the p -value of 0.032 is less than 0.05, conclude that the value is significantly discordant.

Omitting the `$P` from the function call returns a list of information about the calculation, as described below. By default, `PMASR` uses the Bonferroni method when it is exact, or the saddlepoint method otherwise. In this example, saddlepoint was used. To see the Bonferroni bounds, enter this:

```
> PMASR_Bonf(3.05, 30)
```

```
$Upper_Bound
```

```
[1] 0.03310819
```

```
$Upper_Bound_Raw
```

```
[1] 0.0331091
```

```
$Lower_Bound
```

```
[1] 0.03309549
```

Example: A 2^4 full factorial design (FFD) with 16 total runs is performed. In the analysis with only the 4 main effects and no interactions, the maximum absolute studentized residual is 2.50. Is this significantly discordant?

For this problem, the `PMASR` function requires the \mathbf{X} design matrix as the second parameter. Matrices may be entered into R in different ways. If the data has been analyzed using the `lm` function in R, the `x` member of the `lm` object is the design matrix. Another way to create the matrix directly is to source this code from a text file:

```
X.FFD16 <- cbind( rep(1,16),  
  c(rep(-1, 8), rep(1, 8)),  
  rep(c(rep(-1, 4), rep(1, 4)), 2),  
  rep(c(-1, -1, 1, 1), 4),  
  rep(c(-1, 1), 8))
```

After creating the matrix, inspect it for correctness:

```

> X.FFD16

[,1] [,2] [,3] [,4] [,5]
[1,] 1 -1 -1 -1 -1
[2,] 1 -1 -1 -1 1
[3,] 1 -1 -1 1 -1
...
[16,] 1 1 1 1 1

```

Then, calculate the discordancy test p -value this way:

```

> PMASR(2.5, X.FFD16) $P
[1] 0.07410689

```

With a p -value greater than 0.05, the MASR of 2.5 is not significantly discordant.

A.1 Description of PMASR Function

`PMASR(value, X)` performs all calculations involving saddlepoint approximations. When an exact Bonferroni bound is available, this exact solution is substituted automatically, using the default settings. The function returns a list including a PDF value for MASR and a p -value for the outlier discordancy test, plus other information. Here are the parameters

- `value` (required) is the observed value x of MASR or MSSR. A vector of values is allowed, and the function will calculate PDF and p -values for each element in the vector
- `X` (required) is the $n \times p$ design matrix for a regression setting. For a one-sample problem, when \mathbf{X} is a single column of 1's, `X` can be an integer value representing the sample size n
- `m` (optional, default = 1) is the dimension m , representing the number of columns in the observed data \mathbf{Y}
- `eps` (optional, default = 10^{-5}) is a convergence criterion used for some sub-algorithms
- `order2` (optional, default = TRUE) is logical, FALSE for first-order saddlepoint, TRUE for second-order

- `o2exp` (optional, default = TRUE) is logical, FALSE for the $1 + O$ form of second-order saddlepoint, TRUE for the e^O form
- `calibrated` (optional, default=TRUE) is logical, TRUE for calibrated, FALSE for not
- `calpoint` (optional, default = 2) is either 1, 2, or 3 for which point to calibrate at. 1 is the upper end of the support, 2 is M_2 , 3 is M_3 .
- `method` (optional, default = "best") is text, either "saddlepoint", "Bonferroni", or "best". Only the first two characters are required
- `quiet` (optional, default = TRUE) for debugging only. FALSE produces lots of intermediate results
- `squared` (optional, default = FALSE) FALSE for maximum absolute studentized residual MASR, or TRUE for maximum squared studentized residual MSSR
- `timing` (optional, default=FALSE) TRUE will measure and print out the time required for the calculation.

The value returned by the `PMASR` function is a list containing the following values.

- `$Error` is a string listing errors in the input parameters
- `$P` is the p -value for an outlier discordancy test, which is $1 - F_M(x)$.
- `$CDF` is the approximated value of $F_M(x)$.
- `$MASR` is the value of the MASR or MSSR statistic passed to the function in the first parameter.
- `$squared` echoes the parameter value passed to the function.
- `$n` is the number of observations, represented by the number of rows in the \mathbf{X} matrix.
- `$p` is the number of parameters, represented by the number of columns in the \mathbf{X} matrix.

- `$m` is the number of dimensions, passed to the function
- `$order2` echoes the value passed to the function.
- `$o2exp` echoes the value passed to the function.
- `$calibrated` echoes the value passed to the function.
- `$calpoint` echoes the value passed to the function.
- `$method` echoes the value passed to the function, which may be interpreted from only two provided characters.
- `$M2` is M_2 , the calculated limit beyond which it is not possible for two residuals to have values, when $m = 1$.
- `$M3` is M_3 , the calculated limit beyond which it is not possible for three residuals to have values, when $m = 1$.
- `$eps` is the value passed to the function.

A.2 Description of `PMASR_Bonf` Function

`PMASR_Bonf` performs all Bonferroni-based calculations, including the limits M_2 and M_3 . Values are returned in a list, including all available bounds and other information. Here are the parameters

- `value` (required) is the observed value x of MASR or MSSR. A vector of values is allowed, and the function will calculate PDF and p -values for each element in the vector. Note: The MSSR option (`squared=TRUE`) is not available for this function. Enter MASR or the square root of MSSR.
- `X` (required) is the $n \times p$ design matrix for a regression setting. For a one-sample problem, when \mathbf{X} is a single column of 1's, `X` can be an integer value representing the sample size n

- `mv` (optional, default = 1) is the dimension m , representing the number of columns in the observed data \mathbf{Y}
- `eps` (optional, default = 10^{-4}) is a convergence criterion used for some sub-algorithms
- `quiet` (optional, default = TRUE) for debugging only. FALSE produces lots of intermediate results
- `status` (optional, default = FALSE) when runs take a long time, set `status = TRUE` to get progress reports.
- `maxeval` (optional, default = 10^5) When `mv > 1`, this code may run a lot of triple numerical integrals, and this may take time. To limit the number of iterations in these integrals, set this parameter lower.
- `timing` (optional, default=FALSE) TRUE will measure and print out the time required for the calculation.

The value returned by the `PMASR_Bonf` is a list containing these elements.

- `$Error` is a string returning an Error message, if any
- `$Upper_Bound` is the first Bonferroni bound, an upper bound on the p -value, including the spanning tree improvement after articles by Kounias, Hunter, and Worsley
- `$Upper_Bound_Raw` is the first Bonferroni bound, an upper bound on the p -value, without any improvements. This is identical to the Grubbs test, as generally calculated.
- `$Lower_Bound` is the second Bonferroni bound, a lower bound on the p -value.
- `$Upper_Exact` is a logical value, TRUE if the upper bound is exact.
- `$Lower_Exact` is a logical value, TRUE if the lower bound is exact.
- `$M2` is the value of M_2 , above which no more than two residuals cannot occur.

- \$M3 is the value of M_3 , above which no more than three residuals cannot occur.

The code module contains numerous other subroutines and debugging functions with names that start with . and will not appear in default directory listings. Each function is commented in the code listing.