ASYMPTOTIC TECHNIQUES OF SELECTIVE INFERENCE

By

DEWEI ZHONG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Statistics

Written under the direction of

John Kolassa

and approved by

New Brunswick, New Jersey

October, 2020

ABSTRACT OF THE DISSERTATION

Asymptotic Techniques of Selective Inference

by DEWEI ZHONG

Dissertation Director: John Kolassa

Asymptotic approaches are widely used in statistics. Generally, I recognize two applications of asymptotics.

First, asymptotics can solve some problems which cannot be solved exactly in mathematics. For example, density and mass functions and distribution functions of some statistics often cannot be found exactly. Asymptotic approaches will be used for finding the asymptotic density and mass functions and distribution functions under such circumstances. The error between asymptotic methods and truth is controlled within tolerance, like $O(1/n)$ or something else. Chapter 1 presents this kind of problem. The two-stage Mann-Whitney statistic has known mass and distribution functions. But these exact representations are given only recursively, and the recursion is complicated. It means that we cannot express them mathematically. With the help of an asymptotic method, the Edgeworth expansion, we can express the distribution functions. Moments and cumulants are necessary for the Edgeworth expansion and I focus on the calculation of them in Chapter 1.

The second use of asymptotics is to compare two different methods or functions and find how they are close. When various methods are proposed to

approximate something, one may just determine whether they are asymptotically correct. If asymptotically, the methods are correct, the error between them should be determined. Furthermore, how close they are to the truth must be determined. Chapter 2 is a typical example of this kind of problem. The traditional approach is called the studentized bootstrap and the new one is the tilted bootstrap. We compare the two approaches in multi-dimension and conclude the difference between their p-values is $o(1)$ based on some assumptions.

Chapters 3 and 4 discuss a significance test to perform a variable selection for regression. The test is called the covariance test. The test is based on the exponential distribution, but the statistic does not follow it exactly but asymptotically. We investigate the properties of the test statistic and propose another covariance test based on the gamma distribution. This topic is a combination of the two problems mentioned above. We compare all available methods and provide an alternative better approach.

Chapter 5 presents a method for calculating the order of error numerically. It is derived from Chapters 3 and 4. We have to find the order of error numerically when it is too hard to find it analytically. Many examples are illustrated to demonstrate effectiveness.

Acknowledgements

I sincerely express the deepest gratitude to my supervisor and trusted friend John Kolassa for his continuous support during the recent five years. He did not only teach me statistics, but also life and career experience. Thanks a lot for his help in my growth. It is my fortune to meet him in Rutgers.

I could not achieve this level without the support from my supervisor and other faculty and students in the department. I would like to thank Dr. Cabrera, Dr. Hoover and Dr. Seifu for serving on my dissertation committees.

I declare that the thesis has been composed by myself and that the work has not be submitted for any degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included.

Dedication

This dissertation is dedicated to my wife Weiying Wang, as well as my parents Huaihong Zhong and Ling Liu.

Contents

List of Tables

List of Figures

Chapter 1

Moments and Cumulants of the Two-Stage Mann-Whitney Statistic

In the chapter, an asymptotic technique is used for the calculation of critical values and powers for nonparametric tests in group-sequential clinical trials.

A multiple-stage design is often preferable in early phase clinical trials to investigate the activity of a new treatment [7]. Such a design is able to protect patients better as compared to the traditional one-stage design, by allowing a trial to be stopped earlier when the new treatment is indeed ineffective. For this reason, early stopping for futility is allowed in these trials. Among multiple-stage designs, a two-stage design is widely used in phase II clinical trials [5]. The expected whose sample size is smaller relative to that in the following phase III trial to confirm the effectiveness of the new treatment. [1]

Mann and Whitney [2] presented the two-sample one-stage rank-sum test which is widely used in many areas of application, including randomized Phase II trials, and its statistical properties are well studied. Castagliola [3] proposed a faster algorithm for computation of the exact distribution of the Mann-Whitney statistic $U_{m,n}$ under the null hypothesis. Spurrier and Hewett [4] proposed a two-stage test based on sample sizes m_1, n_1 in the first stage and another m_2, n_2 added at the second stage. Although an exact recurrence relation is discussed, the authors had not been able to find an effective method for computing the joint probability distribution of the first and second stage statistics. Kolassa [10] provided an efficient approach to compute the joint distribution and critical values for them. But the mixed moments and cumulants of the first and second stage statistics are still necessary.

In this Chapter, the mixed moments of cumulants of the first and second statistics are found through extensive calculations. I introduce the topic in Section 1.1, calculate the mixed moments in Section 1.2, and find the cumulants in Section 1.3.

1.1 Introduction to The Mann-Whitney Statistic and Two-Stage Test

The Mann-Whitney U test is often used to test a difference in the responses of two groups. Suppose that $X_1, ...X_M$ represent measurements from the control group, $Y_1, ...Y_N$ represent measurements from the treatment group. The Mann-Whitney statistic is defined as

$$U = \sum_{j=1}^{N} \sum_{i=1}^{M} I_{ij}, \quad \text{for} \quad I_{ij} = I(X_i < Y_j) = \begin{cases} 1, & X_i < Y_j \\ 0, & X_i \geq Y_j \end{cases}. \tag{1.1}$$

We select c as the critical value, defined so that if U is larger than c, the two populations are claimed different. The value c depends on the significance level. This is one-stage test proposed by Mann and Whitney [2].

Due to ethcal concerns and resource management, common designs allow for early stoppage in the presence of early and convincing proof. Spurrier and Hewett [4] provide a two-stage test based on the Mann-Whitney statistic. The two-stage test has two critical values. Denote them as c_1 and c_2. First, gather m observations from the control group and n observations from the treatment group. Define

$$U_1 = \sum_{j=1}^{n} \sum_{i=1}^{m} I_{ij}, \tag{1.2}$$

for I_{ij} defined as in (1.1). Calculate U_1, and if it is larger than or equal to c_1, stop the trial early to declare the treatment group superior to the control group. If U_1 is less than the first critical value c_1, gather $M - m$ observations from the control group and $N - n$ observations from the treatment group, where $m \leq M, n \leq N$. Define

$$U_2 = \sum_{j=1}^{N} \sum_{i=1}^{M} I_{ij}. \tag{1.3}$$

Then calculate U_2. If it is larger than or equal to c_2, we claim the treated is superior to the controls.

The most difficult task in performing this test is to find two critical values. The critical values of Mann-Whitney statistic in one dimension, and hence the critical values of one-stage test, can be easily calculated. We focus on the critical values for two-stage test. Due to the complexity of the mass function for two-dimensional Mann-Whitney statistics, the computation required to get exact critical values is intensive. Therefore, the asymptotic critical values within tolerable error will be substituted. The Cornish-Fisher expansion provides the method to obtain critical values [2]. In order to use the Cornish-Fisher expansion, the cumulants are necessary. So next, we will find the moments for two-stage test statistics U_1 and U_2 in Section 1.2 and get the cumulants in Section 1.3. The univariate moments are currently known, and my focus are mixed moments.

1.2 Moments of Two-Stage Statistic

First, assume $X_1 \cdots X_M$ and $Y_1 \cdots Y_M$ are jointly independent and identically distributed. Under the null hypthesis, all of the observations belong to the same population. In the first subsection, we give some hyphenated transition. In the second, third, fourth and fifth subsections, we find the first, second, third and fourth moments repectively under both the general case and the null hypothesis.

1.2.1 Probability Definition

Moment calculations will require the expectations of products of indicators as in (1.1).

Under the null hypothesis:

$$
\begin{gathered}
E(I_{ij}) = \frac{1}{2} \qquad E(I_{ij}I_{kj}) = \frac{1}{3} \qquad E(I_{ij}I_{il}) = \frac{1}{3} \qquad E(I_{ij}I_{kl}) = \frac{1}{4} \\
E(I_{ij}I_{il}I_{it}) = \frac{1}{4} \qquad E(I_{ij}I_{kj}I_{sj}) = \frac{1}{4} \qquad E(I_{ij}I_{kj}I_{kl}) = \frac{5}{24} \\
E(I_{ij}I_{kj}I_{sj}I_{st}) = \frac{3}{20} \qquad E(I_{ij}I_{kj}I_{it}I_{st}) = \frac{2}{15} \qquad E(I_{ij}I_{kj}I_{sj}I_{pj}) = \frac{1}{5} \\
E(I_{ij}I_{kj}I_{iq}I_{it}) = \frac{3}{20} \qquad E(I_{ij}I_{kj}I_{il}I_{kt}) = \frac{1}{6} \qquad E(I_{ij}I_{kj}I_{iq}I_{kl}) = \frac{2}{15} \\
E(I_{ij}I_{il}I_{kj}) = \frac{5}{24} \qquad E(I_{ij}I_{il}I_{iq}I_{st}) = \frac{1}{8} \qquad E(I_{ij}I_{kj}I_{pj}I_{st}) = \frac{1}{8} \\
E(I_{ij}I_{il}I_{pq}I_{pt}) = \frac{1}{9} \qquad E(I_{ij}I_{kj}I_{pq}I_{sq}) = \frac{1}{9} \qquad E(I_{ij}I_{il}I_{pq}I_{sq}) = \frac{1}{9} \\
E(I_{ij}I_{kj}I_{il}I_{st}) = \frac{1}{9} \qquad E(I_{ij}I_{kj}I_{pq}I_{st}) = \frac{1}{12} \qquad E(I_{ij}I_{kl}I_{st}I_{sq}) = \frac{1}{12} \\
E(I_{ij}I_{kl}I_{st}I_{pq}) = \frac{1}{16} \qquad E(I_{ij}I_{il}I_{it}I_{iq}) = \frac{1}{5} \qquad E(I_{ij}I_{kj}I_{iq}I_{st}) = \frac{5}{48}.
\end{gathered}
\tag{1.4}
$$

Here we intend i, k, s, p in the same expression are pairwise unequal, as are j, l, q, t. For example, in $I_{ij}I_{kl}I_{pq}I_{st}$, $i \neq k, i \neq p, i \neq s, k \neq p, k \neq s, p \neq s$ and $j \neq l, j \neq t, j \neq q, l \neq q, l \neq t, q \neq t$. In the general case, outside of the null hypothesis, define the probabilities π_i below:

$$
\begin{gathered}
\pi_0 = E(I_{ij}) \qquad \pi_1 = E(I_{ij}I_{kj}) \qquad \pi_2 = E(I_{ij}I_{kj}I_{sj}) \qquad \pi_3 = E(I_{ij}I_{kj}I_{sj}I_{st}) \\
\pi_4 = E(I_{ij}I_{kj}I_{kl}) \qquad \pi_5 = E(I_{ij}I_{kj}I_{it}I_{st}) \qquad \pi_6 = E(I_{ij}I_{kj}I_{sj}I_{pj}) \\
\pi_7 = E(I_{ij}I_{kj}I_{iq}I_{it}) \qquad \pi_8 = E(I_{ij}I_{kj}I_{il}I_{kl}) \qquad \pi_9 = E(I_{ij}I_{il}) \\
\pi_{10} = E(I_{ij}I_{kj}I_{il}I_{kt}) \qquad \pi_{12} = E(I_{ij}I_{il}I_{it}) \qquad \pi_{13} = E(I_{ij}I_{il}I_{it}I_{iq}).
\end{gathered}
\tag{1.5}
$$

Here, $I_{ij}I_{il}I_{kj} = 1$ implies that all of $X_i < Y_j$, $X_i < Y_l$, $X_k < Y_j$ hold. The equation $I_{ij}I_{il}I_{kj} = 0$ implies at least one of them does not hold.

1.2.2 First Moments

In general, by (1.2), (1.3) and (1.5),

$$E(U_1) = \sum_{j=1}^{n}\sum_{i=1}^{m} E(I_{ij}) = mn\pi_0, \tag{1.6}$$

$$E(U_2) = MN\pi_0. \tag{1.7}$$

Under the null hypothesis, by (1.2), (1.3) and (1.4),

$$E(U_1) = \sum_{j=1}^{n}\sum_{i=1}^{m} E(I_{ij}) = \frac{mn}{2}, \tag{1.8}$$

$$E(U_2) = \frac{MN}{2}. \tag{1.9}$$

1.2.3 Second Moments

$$\begin{aligned} U_1^2 &= (\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij})^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1,k\neq i}^{m} I_{ij}I_{kj} + \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{il} + \\ &\quad + \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{l=1,l\neq j}^{n}\sum_{k=1,k\neq i}^{m} I_{ij}I_{kl}. \end{aligned} \tag{1.10}$$

Substituting the probability values in (1.10) by (1.5)

$$E(U_1^2) = mn\pi_0 + m(m-1)n\pi_1 + mn(n-1)\pi_9 + m(m-1)n(n-1)\pi_0^2. \tag{1.11}$$

By the same reasoning,

$$\begin{aligned} E(U_2^2) = MN\pi_0 + M(M-1)N\pi_1 + MN(N-1)\pi_9 + \\ M(M-1)N(N-1)\pi_0^2. \end{aligned} \tag{1.12}$$

While calculating mixed moments, conditional expectations are usually utilized. Here

$$\begin{aligned} E(U_1U_2) &= E(E(U_1U_2|U_2)) \\ &= E(U_2E(U_1|U_2)) \\ &= E(U_2\frac{mn}{MN}U_2) \\ &= \frac{mn}{MN}E(U_2^2). \end{aligned} \tag{1.13}$$

Thus, by (1.12) and (1.13)

$$E(U_1U_2) = mn\pi_0 + mn(M-1)\pi_1 + mn(N-1)\pi_9 + mn(M-1)(N-1)\pi_0^2. \tag{1.14}$$

Under the null hypothesis, by (1.4) and (1.14):

$$\begin{aligned} E(U_1^2) = &\frac{1}{2}mn + \frac{1}{3}m(m-1)n + \frac{1}{3}mn(n-1) + \frac{1}{4}m(m-1)n(n-1) \\ = &\frac{m^2n^2}{4} + \frac{m^2n}{12} + \frac{mn^2}{12} + \frac{mn}{12}, \end{aligned} \tag{1.15}$$

By the same reasoning, under the null hypothesis,

$$E(U_2^2) = \frac{M^2N^2}{4} + \frac{M^2N}{12} + \frac{MN^2}{12} + \frac{MN}{12}. \tag{1.16}$$

The results (1.6)-(1.16) are the properties of the one-stage Mann-Whitney statistic [2]. The following are mixed moments as well as two-stage moments. Again under the null hypothesis,

$$\begin{aligned} E(U_1U_2) &= E(E(U_1U_2|U_2)) \\ &= E(U_2E(U_1|U_2)) \\ &= E(U_2\frac{mn}{MN}U_2) \\ &= \frac{mn}{MN}E(U_2^2). \end{aligned} \tag{1.17}$$

Substituting (1.16) in (1.17),

$$E(U_1U_2) = \frac{mnMN}{4} + \frac{mnM}{12} + \frac{mnN}{12} + \frac{mn}{12}. \tag{1.18}$$

1.2.4 Third Moments

We extend calculations for lower orders to find it.

$$
\begin{aligned}
E(U_1^3) = {} & \sum_{j=1}^{n}\sum_{i=1}^{m} I_{ij} + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m} I_{ij}I_{ij}I_{kj} + 3\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{il} \\
& + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{kl} + 6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{il}I_{kj} \\
& + \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{kj}I_{sj} + \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{il}I_{it} \\
& + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{kl}I_{sj} \\
& + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{kl}I_{it} \\
& + \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{kl}I_{st}.
\end{aligned}
\tag{1.19}
$$

Substituting (1.5) in (1.19),

$$
\begin{aligned}
E(U_1^3) = {} & mn\pi_0 + 3m(m-1)n\pi_1 + 3mn(n-1)\pi_9 + 3m(m-1)n(n-1)\pi_0^2 \\
& + 6m(m-1)n(n-1)\pi_4 + m(m-1)(m-2)n\pi_2 \\
& + mn(n-1)(n-2)\pi_{12} + 3m(m-1)(m-2)n(n-1)\pi_0\pi_1 \\
& + 3m(m-1)n(n-1)(n-2)\pi_0\pi_9 \\
& + m(m-1)(m-2)n(n-1)(n-2)\pi_0^3.
\end{aligned}
\tag{1.20}
$$

Simplifying,

$$
\begin{aligned}
E(U_1^3) = {} & mn\pi_0 + 3m(m-1)n\pi_1 + 3mn(n-1)\pi_9 + 3m(m-1)n(n-1)\pi_0^2 \\
& + 6m(m-1)n(n-1)\pi_4 + m(m-1)(m-2)n\pi_2 \\
& + mn(n-1)(n-2)\pi_{12} + 3m(m-1)(m-2)n(n-1)\pi_0\pi_1 \\
& + 3m(m-1)n(n-1)(n-2)\pi_0\pi_9 \\
& + m(m-1)(m-2)n(n-1)(n-2)\pi_0^3.
\end{aligned}
\tag{1.21}
$$

By the same reasoning,

$$
\begin{aligned}
E(U_2^3) = {} & MN\pi_0 + 3M(M-1)N\pi_1 + 3MN(N-1)\pi_9 \\
& + 3M(M-1)N(N-1)\pi_0^2 + 6M(M-1)N(N-1)\pi_4 \\
& + M(M-1)(M-2)N\pi_2 + MN(N-1)(N-2)\pi_{12} \\
& + 3M(M-1)(M-2)N(N-1)\pi_0\pi_1 \\
& + 3M(M-1)N(N-1)(N-2)\pi_0\pi_9 \\
& + M(M-1)(M-2)N(N-1)(N-2)\pi_0^3.
\end{aligned}
\tag{1.22}
$$

The above (1.19)-(1.22) are moments of the one-stage Mann-Whitney statistic [2]. Consider the mixed moments for the two-stage statistic.

The conditional expectations are used to get $E(U_1U_2^2)$ and $E(U_1^2U_2)$.

$$E(U_1U_2^2) = E[U_2^2E(U_1|U_2)] = E(U_2^2\frac{mn}{MN}U_2) = \frac{mn}{MN}E(U_2^3). \tag{1.23}$$

Substituting (1.22) in (1.23),

$$
\begin{aligned}
E(U_1U_2^2) = {} & mn\pi_0 + 3mn(M-1)\pi_1 \\
& + 3mn(N-1)\pi_9 + 3mn(M-1)(N-1)\pi_0^2 \\
& + 6mn(M-1)(N-1)\pi_4 \\
& + mn(M-1)(M-2)\pi_2 + mn(N-1)(N-2)\pi_{12} \\
& + 3mn(M-1)(M-2)(N-1)\pi_0\pi_1 \\
& + 3mn(M-1)(N-1)(N-2)\pi_0\pi_9 \\
& + mn(M-1)(M-2)(N-1)(N-2)\pi_0^3.
\end{aligned}
\tag{1.24}
$$

The $E(U_1^2U_2)$ is a little difficult to find. It is simplified in (1.25) and conquered part by part.

$$\begin{aligned}
E(U_1^2U_2) =& E(\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n} I_{ij}I_{kl}U_2) \\
=& E(\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{kl}U_2 + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1,k\neq i}^{m} I_{ij}I_{kj}U_2 \\
&+ \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n} I_{ij}I_{il}U_2 + \sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}U_2) \\
=& E(\frac{m(m-1)n(n-1)}{M(M-1)N(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M}\sum_{l=1,l\neq j}^{N} I_{ij}I_{kl}U_2) \\
&+ E(\frac{m(m-1)n}{M(M-1)N}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M} I_{ij}I_{kj}U_2) \\
&+ E(\frac{mn(n-1)}{MN(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{l=1,l\neq j}^{N} I_{ij}I_{il}U_2) + E(\frac{mn}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} I_{ij}U_2) \\
=& E(\frac{m(m-1)n(n-1)}{M(M-1)N(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{N} I_{ij}I_{kl}U_2) \\
+& E((\frac{m(m-1)n}{M(M-1)N} - \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)})\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M} I_{ij}I_{kj}U_2) \\
+& E((\frac{mn(n-1)}{MN(N-1)} - \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)})\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{l=1,l\neq j}^{N} I_{ij}I_{il}U_2) \\
&+ E((\frac{mn}{MN} - \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)})\sum_{i=1}^{M}\sum_{j=1}^{N} I_{ij}U_2) \\
=& \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)}E(U_2^3) \\
&+ (\frac{mn}{MN} - \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)})E(U_2^2) \\
+& (\frac{m(m-1)n}{M(M-1)N} - \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)})E(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M} I_{ij}I_{kj}U_2) \\
+& (\frac{mn(n-1)}{MN(N-1)} - \frac{m(m-1)n(n-1)}{M(M-1)N(N-1)})E(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{l=1,l\neq j}^{N} I_{ij}I_{il}U_2).
\end{aligned} \tag{1.25}$$

We can find

$$
\begin{aligned}
E(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M} I_{ij}I_{kj}U_2) =& E(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M}\sum_{l=1,l\neq j}^{N} I_{ij}I_{kj}I_{il} \\
&+ \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M} I_{ij}I_{kj}I_{kl} \\
&+ \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M}\sum_{s=1,s\neq i,s\neq k}^{M}\sum_{l=1,l\neq j}^{N} I_{ij}I_{kj}I_{sl} \\
&+ \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M}\sum_{s=1,s\neq i,s\neq k}^{M} I_{ij}I_{kj}I_{sj} \\
&+ \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1,k\neq i}^{M} (I_{ij}I_{kj}I_{ij} + I_{ij}I_{kj}I_{kj})) \\
=& \pi_0\pi_1 M(M-1)(M-2)N(N-1) \\
&+ 2\pi_4 M(M-1)N(N-1) \\
&+ \pi_2 M(M-1)(M-2)N \\
&+ 2\pi_1 M(M-1)N.
\end{aligned}
$$

By the same reasoning,

$$
\begin{aligned}
E(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{l=1,l\neq j}^{N} I_{ij}I_{il}U_2) =& \pi_0\pi_9 M(M-1)N(N-1)(N-2) \\
&+2\pi_4 M(M-1)N(N-1) \\
&+\pi_{12} MN(N-1)(N-2) + 2\pi_9 MN(N-1).
\end{aligned}
\tag{1.26}
$$

It is easy to find (1.25) by above expressions as well as (1.21) and (1.11). *Under the null hypothesis*, we can just use the values in (1.4) to substitute the $\pi_0 \ldots \pi_{13}$ in (1.21), (1.22), (1.24) and (1.25) .

Another Method for Third Moments

We will show another method to get the third moments.

The distribution of U_1 is symmetric about 0 under the null hypothesis. Hence

$$E[U_1 - E(U_1)]^3 = 0, \qquad \text{and so}$$

$$E(U_1^3) = [E(U_1)]^3 - 3E(U_1)[E(U_1)]^2 + 3E(U_1^2)E(U_1). \tag{1.27}$$

Substituting (1.8) and (1.15), then simplifying (1.27), (1.28) is obtained.

$$E(U_1^3) = \frac{m^3n^3}{8} + \frac{m^3n^2}{8} + \frac{m^2n^3}{8} + \frac{m^2n^2}{8}, \tag{1.28}$$

The same way for U_2 to find (1.29) and the same conditional expectations to (1.30).

$$E(U_2^3) = \frac{M^3N^3}{8} + \frac{M^3N^2}{8} + \frac{M^2N^3}{8} + \frac{M^2N^2}{8}, \tag{1.29}$$

$$E(U_1U_2^2) = \frac{mnM^2N^2}{8} + \frac{mnM^2N}{8} + \frac{mnMN^2}{8} + \frac{mnMN}{8}. \tag{1.30}$$

There is a property under the null hypothesis,

$$\begin{aligned} E(U_2|U_1) =& \frac{M+N+1}{m+n+1}U_1 + \frac{1}{2}\Big[\frac{(M-m)n(n+1) + (N-n)m(m+1)}{m+n+1} \\ &+ (M-m)(N-n)\Big]. \end{aligned}$$

And thus,

$$
\begin{aligned}
E(U_1^2U_2) &= E[U_1^2E(U_2|U_1)] \\
&= E\big[U_1^2(\frac{M+N+1}{m+n+1}U_1 + \frac{1}{2}[\frac{(M-m)n(n+1)+(N-n)m(m+1)}{m+n+1} \\
&\quad + (M-m)(N-n)])\big] \\
&= \frac{M+N+1}{m+n+1}E(U_1^3) + \frac{1}{2}\big[\frac{(M-m)n(n+1)+(N-n)m(m+1)}{m+n+1} \\
&\quad + (M-m)(N-n)\big]E(U_1^2) \\
&= \frac{M+N+1}{m+n+1}\frac{m^2n^2}{8}(mn+m+n+1) \\
&\quad + \frac{mn}{24}\frac{(M-m)n(n+1)+(N-n)m(m+1)}{m+n+1}(3mn+m+n+1) \\
&\quad + \frac{1}{12}(M-m)(N-n))(3m^2n^2+m^n+mn^2+mn).
\end{aligned}
$$

The result is the same as (1.25) under the null hypothesis.

1.2.5 Fourth Moments

We extend calculations for lower order moments to find it,

$$
\begin{aligned}
U_1^4 &= (\sum_{j=1}^{n}\sum_{i=1}^{m} I_{ij}I_{ij}I_{ij}I_{ij})^4 \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m} I_{ij}I_{ij}I_{ij}I_{ij} + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m} I_{ij}I_{ij}I_{kj}I_{kj} + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{il}I_{il} \\
&+ 4\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m} I_{ij}I_{ij}I_{ij}I_{kj} + 4\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{ij}I_{il} \\
&+ 6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{ij}I_{il}I_{it} + 6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{ij}I_{kj}I_{sj} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{kj}I_{kl} + 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{kj}I_{il} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{kl}I_{il} + 6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{il}I_{kj}I_{kl} \\
&+ 4\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{ij}I_{kl} + 3\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n} I_{ij}I_{ij}I_{kl}I_{kl} \\
&+ 6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{ij}I_{kl}I_{kt} \\
&+ 6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{ij}I_{kl}I_{sl} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{il}I_{it}I_{kj} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{sj}I_{il}I_{kj} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{ij}I_{il}I_{kt} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{kj}I_{il}I_{sl} \\
&+ 12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{kj}I_{il}I_{kt}
\end{aligned}
$$

$$
\begin{aligned}
&+12\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{ij}I_{kj}I_{sl}\\
&+\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{p=1,p\neq i,p\neq k,p\neq s}^{m} I_{ij}I_{kj}I_{sj}I_{pj}\\
&+\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{q=1,q\neq j,q\neq l,q\neq t}^{n} I_{ij}I_{iq}I_{it}I_{il}\\
&+4\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{q=1,q\neq j,q\neq l,q\neq t}^{n} I_{ij}I_{iq}I_{it}I_{kl}\\
&+4\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{p=1,p\neq i,p\neq k,p\neq s}^{m} I_{ij}I_{kj}I_{sj}I_{pl}\\
&+3\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{q=1,q\neq j,q\neq l,q\neq t}^{n} I_{ij}I_{il}I_{kt}I_{kq}\\
&+3\sum_{j=1}^{n}\sum_{l=1,l\neq j}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{p=1,p\neq i,p\neq k,p\neq s}^{m} I_{ij}I_{kj}I_{sl}I_{pl}\\
&+6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{t=1,t\neq j,t\neq l}^{n} I_{ij}I_{il}I_{kt}I_{st}\\
&+6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{ij}I_{kl}I_{st}\\
&+24\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{s=1,s\neq i,s\neq k}^{m} I_{ij}I_{kj}I_{il}I_{st}\\
&+6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{p=1,p\neq i,p\neq k,p\neq s}^{m} I_{ij}I_{kj}I_{pl}I_{st}\\
&+6\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{p=1,p\neq i,p\neq k,p\neq s}^{m} I_{ij}I_{il}I_{kq}I_{st}\\
&+\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1,k\neq i}^{m}\sum_{l=1,l\neq j}^{n}\sum_{t=1,t\neq j,t\neq l}^{n}\sum_{s=1,s\neq i,s\neq k}^{m}\sum_{p=1,p\neq i,p\neq k,p\neq s}^{m}\sum_{q=1,q\neq j,q\neq l,q\neq t}^{n} I_{ij}I_{kl}I_{pq}I_{st},
\end{aligned}
\tag{1.31}
$$

Substituting the values in (1.5),

$$\begin{aligned}
E(U_1^4) = {} & \pi_0 mn + 3\times\pi_1 m(m-1)n + 3\times\pi_9 mn(n-1) \\
& + 4\times\pi_1 m(m-1)n + 4\times\pi_9 mn(n-1) \\
& + 6\times\pi_{12} mn(n-1)(n-2) + 6\times\pi_2 m(m-1)(m-2)n \\
& + 12\times\pi_4 m(m-1)n(n-1) + 12\times\pi_4 m(m-1)n(n-1) \\
& + 12\times\pi_4 m(m-1)n(n-1) + 6\times\pi_8 m(m-1)n(n-1) \\
& + 4\times\pi_0^2 m(m-1)n(n-1) + 3\times\pi_0^2 m(m-1)n(n-1) \\
& + 6\times\pi_0\pi_9 m(m-1)n(n-1)(n-2) \\
& + 6\times\pi_0\pi_1 m(m-1)(m-2)n(n-1) \\
& + 12\times\pi_7 m(m-1)n(n-1)(n-2) \\
& + 12\times\pi_3 m(m-1)(m-2)n(n-1) \\
& + 12\times\pi_0\pi_9 m(m-1)n(n-1)(n-2) \\
& + 12\times\pi_0\pi_1 m(m-1)(m-2)n(n-1) \\
& + 12\times\pi_5 m(m-1)(m-2)n(n-1) \\
& + 12\times\pi_{10} m(m-1)n(n-1)(n-2) \\
& + \pi_6 mn(m-1)(m-2)(m-3) + \pi_{13} mn(n-1)(n-2)(n-3) \\
& + 4\times\pi_0\pi_{12} m(m-1)n(n-1)(n-2)(n-3) \\
& + 4\times\pi_0\pi_2 m(m-1)(m-2)(m-3)n(n-1) \\
& + 3\times\pi_9^2 m(m-1)(m-2)(m-3)n(n-1) \\
& + 3\times\pi_1^2 m(m-1)n(n-1)(n-2)(n-3) \\
& + 6\times\pi_1\pi_9 m(m-1)(m-2)n(n-1)(n-2) \\
& + 6\times\pi_0^3 m(m-1)(m-2)n(n-1)(n-2) \\
& + 24\times\pi_0\pi_4 m(m-1)(m-2)n(n-1)(n-2)
\end{aligned} \tag{1.32}$$

$$+ 6 \times \pi_0^2\pi_1 m(m-1)(m-2)n(n-1)(n-2)(n-3)$$

$$+ 6 \times \pi_9\pi_0^2 m(m-1)(m-2)(m-3)n(n-1)(n-2)$$

$$+ \pi_0^4 m(m-1)(m-2)(m-3)n(n-1)(n-2)(n-3).$$

Simplifying,

$$\begin{aligned}
E(U_1^4) = {} & \pi_0 mn + 7 \times \pi_1 m(m-1)n + 7 \times \pi_9 mn(n-1) \\
& + 6 \times \pi_{12} mn(n-1)(n-2) + 6 \times \pi_2 m(m-1)(m-2)n \\
& + 36 \times \pi_4 m(m-1)n(n-1) + 6 \times \pi_8 m(m-1)n(n-1) \\
& + 7 \times \pi_0^2 m(m-1)n(n-1) \\
& + 6 \times \pi_0\pi_9 m(m-1)n(n-1)(n-2) \\
& + 6 \times \pi_0\pi_1 m(m-1)(m-2)n(n-1) \\
& + 12 \times \pi_7 m(m-1)n(n-1)(n-2) \\
& + 12 \times \pi_3 m(m-1)(m-2)n(n-1) \\
& + 12 \times \pi_0\pi_9 m(m-1)n(n-1)(n-2) \\
& + 12 \times \pi_0\pi_1 m(m-1)(m-2)n(n-1) \\
& + 12 \times \pi_5 m(m-1)(m-2)n(n-1) \\
& + 12 \times \pi_{10} m(m-1)n(n-1)(n-2) \\
& + \pi_6 mn(m-1)(m-2)(m-3) + \pi_{13} mn(n-1)(n-2)(n-3) \\
& + 4 \times \pi_0\pi_{12} m(m-1)n(n-1)(n-2)(n-3) \\
& + 4 \times \pi_0\pi_2 m(m-1)(m-2)(m-3)n(n-1) \\
& + 3 \times \pi_9^2 m(m-1)(m-2)(m-3)n(n-1) \\
& + 3 \times \pi_1^2 m(m-1)n(n-1)(n-2)(n-3) \\
& + 6 \times \pi_1\pi_9 m(m-1)(m-2)n(n-1)(n-2)
\end{aligned} \tag{1.33}$$

$$+ 6 \times \pi_0^3 m(m-1)(m-2)n(n-1)(n-2)$$

$$+ 24 \times \pi_0\pi_4 m(m-1)(m-2)n(n-1)(n-2)$$

$$+ 6 \times \pi_0^2\pi_1 m(m-1)(m-2)n(n-1)(n-2)(n-3)$$

$$+ 6 \times \pi_9\pi_0^2 m(m-1)(m-2)(m-3)n(n-1)(n-2)$$

$$+ \pi_0^4 m(m-1)(m-2)(m-3)n(n-1)(n-2)(n-3).$$

As with U_1

$$\begin{aligned}
E(U_2^4) = {} & \pi_0 MN + 7 \times \pi_1 M(M-1)N + 7 \times \pi_9 MN(N-1) \\
& + 6 \times \pi_{12} MN(N-1)(N-2) \\
& + 6 \times \pi_2 M(M-1)(M-2)N + 36 \times \pi_4 M(M-1)N(N-1) \\
& + 6 \times \pi_8 M(M-1)N(N-1) + 7 \times \pi_0^2 M(M-1)N(N-1) \\
& + 6 \times \pi_0\pi_9 M(M-1)N(N-1)(N-2) \\
& + 6 \times \pi_0\pi_1 M(M-1)(M-2)N(N-1) \\
& + 12 \times \pi_7 M(M-1)N(N-1)(N-2) \\
& + 12 \times \pi_3 M(M-1)(M-2)N(N-1) \\
& + 12 \times \pi_0\pi_9 M(M-1)N(N-1)(N-2) \\
& + 12 \times \pi_0\pi_1 M(M-1)(M-2)N(N-1) \\
& + 12 \times \pi_5 M(M-1)(M-2)N(N-1) \\
& + 12 \times \pi_{10} M(M-1)N(N-1)(N-2) \\
& + \pi_6 MN(M-1)(M-2)(M-3) + \pi_{13} MN(N-1)(N-2)(N-3) \\
& + 4 \times \pi_0\pi_{12} M(M-1)N(N-1)(N-2)(N-3) \\
& + 4 \times \pi_0\pi_2 M(M-1)(M-2)(M-3)N(N-1)
\end{aligned} \tag{1.34}$$

$$+ 3 \times \pi_9^2 M(M-1)(M-2)(M-3)N(N-1)$$

$$+ 3 \times \pi_1^2 M(M-1)N(N-1)(N-2)(N-3)$$

$$+ 6 \times \pi_1\pi_9 M(M-1)(M-2)N(N-1)(N-2)$$

$$+ 6 \times \pi_0^3 M(M-1)(M-2)N(N-1)(N-2)$$

$$+ 24 \times \pi_0\pi_4 M(M-1)(M-2)N(N-1)(N-2)$$

$$+ 6 \times \pi_0^2\pi_1 M(M-1)(M-2)N(N-1)(N-2)(N-3)$$

$$+ 6 \times \pi_9\pi_0^2 M(M-1)(M-2)(M-3)N(N-1)(N-2)$$

$$+ \pi_0^4 M(M-1)(M-2)(M-3)N(N-1)(N-2)(N-3).$$

The conditional expectations are used to find mixed moments.

$$E(U_1U_2^3) = E(U_2^3E(U_1|U_2)) = E(\frac{mn}{MN}U_2^4) = \frac{mn}{MN}E(U_2^4),$$

$$E(U_1U_2^3) = \pi_0 mn + 7\times\pi_1 mn(M-1) + 7\times\pi_9 mn(N-1)$$
$$+ 6\times\pi_{12}mn(N-1)(N-2) + 6\times\pi_2 mn(M-1)(M-2)$$
$$+ 36\times\pi_4 mn(M-1)(N-1) + 6\times\pi_8 mn(M-1)(N-1)$$
$$+ 7\times\pi_0^2 mn(M-1)(N-1) + 6\times\pi_0\pi_9 mn(M-1)(N-1)(N-2)$$
$$+ 6\times\pi_0\pi_1 mn(M-1)(M-2)(N-1) + 12\times\pi_7 mn(M-1)(N-1)(N-2)$$
$$+ 12\times\pi_3 mn(M-1)(M-2)(N-1)$$
$$+ 12\times\pi_0\pi_9 mn(M-1)(N-1)(N-2)$$
$$+ 12\times\pi_0\pi_1 mn(M-1)(M-2)(N-1)$$
$$+ 12\times\pi_5 mn(M-1)(M-2)(N-1) + 12\times\pi_{10}mn(M-1)(N-1)(N-2)$$
$$+ \pi_6 mn(M-1)(M-2)(M-3) + \pi_{13}mn(N-1)(N-2)(N-3)$$
$$+ 4\times\pi_0\pi_{12}mn(M-1)(N-1)(N-2)(N-3)$$
$$+ 4\times\pi_0\pi_2 mn(M-1)(M-2)(M-3)(N-1)$$
$$+ 3\times\pi_9^2 mn(M-1)(M-2)(M-3)(N-1)$$
$$+ 3\times\pi_1^2 mn(M-1)(N-1)(N-2)(N-3)$$
$$+ 6\times\pi_1\pi_9 mn(M-1)(M-2)(N-1)(N-2)$$
$$+ 6\times\pi_0^3 mn(M-1)(M-2)(N-1)(N-2)$$
$$+ 24\times\pi_0\pi_4 mn(M-1)(M-2)(N-1)(N-2)$$
$$+ 6\times\pi_0^2\pi_1 mn(M-1)(M-2)(N-1)(N-2)(N-3)$$
$$+ 6\times\pi_9\pi_0^2 mn(M-1)(M-2)(M-3)(N-1)(N-2)$$
$$+ \pi_0^4 mn(M-1)(M-2)(M-3)(N-1)(N-2)(N-3). \tag{1.35}$$

1.3 Cumulants of Two-Stage Mann-Whitney Statistic

In this section, we will recount the relation between the cumulants and moments. So we can use the moments to calculate the cumulants which we use to substitute in Cornish Fisher expansion to get the critical values. As the first step, we express derivatives of the cumulant-generating function in terms of derivatives of the moment generating function. The approach for the moment-cumulant conversion is well defined in [16], and I apply it on the two-stage Mann-Whitney statistic.

I will calculate the first to fourth moments in the first to fourth sections and provide the expression of cumulants in the fifth section. The moment generating function and cumulant-generating function of (U_1, U_2) are

$$M(t_1, t_2) = E(e^{U_1 t_1 + U_2 t_2}) \quad \text{and} \quad K(t_1, t_2) = \log E(e^{t_1 U_1 + t_2 U_2}).$$

For an abbreviated notation, we denote

$$M = E(e^{U_1 t_1 + U_2 t_2}) \quad \text{and} \quad S = e^{t_1 U_1 + t_2 U_2}.$$

1.3.1 First derivatives

$\frac{\partial K(t_1,t_2)}{\partial t_1} = \frac{\partial logM}{\partial t_1} = M^{-1}\frac{\partial M}{\partial t_1} = M^{-1}E(U_1 S),$

$\frac{\partial K(t_1,t_2)}{\partial t_2} = \frac{\partial logM}{\partial t_2} = M^{-1}\frac{\partial M}{\partial t_2} = M^{-1}E(U_2 S).$

1.3.2 Second derivatives

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_1^2} = M^{-1}\frac{\partial E(U_1S)}{\partial t_1} + E(U_1S)\frac{\partial M^{-1}}{\partial t_1} = M^{-1}E(U_1^2S) - M^{-2}E(U_1S)E(U_1S),$$

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_2^2} = M^{-1}\frac{\partial E(U_2S)}{\partial t_2} + E(U_2S)\frac{\partial M^{-1}}{\partial t_2} = M^{-1}E(U_2^2S) - M^{-2}E(U_2S)E(U_2S),$$

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_1 t_2} = M^{-1}\frac{\partial E(U_1S)}{\partial t_2} + E(U_1S)\frac{\partial M^{-1}}{\partial t_2} = M^{-1}E(U_1U_2S) - M^{-2}E(U_1S)E(U_2S).$$

1.3.3 Third derivatives

$$\begin{aligned}\frac{\partial^3 K(t_1,t_2)}{\partial t_1^3} &= M^{-1}\frac{\partial E(U_1^2S)}{\partial t_1} + E(U_1^2S)\frac{\partial M^{-1}}{\partial t_1} - M^{-2}\frac{\partial E(U_1S)E(U_1S)}{\partial t_1} \\ &\quad - E(U_1S)E(U_1S)\frac{\partial M^{-2}}{\partial t_1} \\ &= M^{-1}E(U_1^3S) - M^{-2}E(U_1^2S)E(U_1S) - 2M^{-2}E(U_1S)E(U_1^2S) \\ &\quad + 2M^{-3}(E(U_1S))^3,\end{aligned}$$

$$\begin{aligned}\frac{\partial^3 K(t_1,t_2)}{\partial t_2^3} &= M^{-1}\frac{\partial E(U_2^2S)}{\partial t_2} + E(U_2^2S)\frac{\partial M^{-1}}{\partial t_2} - M^{-2}\frac{\partial E(U_2S)E(U_2S)}{\partial t_2} \\ &\quad - E(U_2S)E(U_2S)\frac{\partial M^{-2}}{\partial t_2} \\ &= M^{-1}E(U_2^3S) - M^{-2}E(U_2^2S)E(U_2S) - 2M^{-2}E(U_2S)E(U_2^2S) \\ &\quad + 2M^{-3}(E(U_2S))^3,\end{aligned}$$

$$\begin{aligned}\frac{\partial^3 K(t_1,t_2)}{\partial t_1^2 \partial t_2} &= M^{-1}\frac{\partial E(U_1^2S)}{\partial t_2} + E(U_1^2S)\frac{\partial M^{-1}}{\partial t_2} - M^{-2}\frac{\partial E(U_1S)E(U_1S)}{\partial t_2} \\ &\quad - E(U_1S)E(U_1S)\frac{\partial M^{-2}}{\partial t_2} \\ &= M^{-1}E(U_1^2U_2S) - M^{-2}E(U_1^2S)E(U_2S) \\ &\quad - 2M^{-2}E(U_1U_2S)E(U_1S) \\ &\quad + 2M^{-3}(E(U_1S))^2E(U_2S),\end{aligned}$$

$$\frac{\partial^3 K(t_1,t_2)}{\partial t_2^2 \partial t_1} = M^{-1}\frac{\partial E(U_2^2 S)}{\partial t_1} + E(U_2^2 S)\frac{\partial M^{-1}}{\partial t_1} - M^{-2}\frac{\partial E(U_2 S)E(U_2 S)}{\partial t_1}$$

$$- E(U_2 S)E(U_2 S)\frac{\partial M^{-2}}{\partial t_1}$$

$$= M^{-1}E(U_2^2 U_1 S) - M^{-2}E(U_2^2 S)E(U_1 S)$$

$$- 2M^{-2}E(U_1 U_2 S)E(U_2 S)$$

$$+ 2M^{-3}(E(U_2 S))^2 E(U_1 S).$$

1.3.4 Fourth derivatives

$$\frac{\partial^4 K(t_1,t_2)}{\partial t_1^4} = M^{-1}\frac{\partial E(U_1^3 S)}{\partial t_1} + E(U_1^3 S)\frac{\partial M^{-1}}{\partial t_1} - M^{-2}\frac{\partial E(U_1^2 S)E(U_1 S)}{\partial t_1}$$

$$- E(U_1^2 S)E(U_1 S)\frac{\partial M^{-2}}{\partial t_1} - 2M^{-2}\frac{\partial E(U_1 S)E(U_1^2 S)}{\partial t_1}$$

$$- 2E(U_1 S)E(U_1^2 S)\frac{\partial M^{-2}}{\partial t_1} + 2M^{-3}\frac{\partial (E(U_1 S))^3}{\partial t_1}$$

$$+ 2(E(U_1 S))^3\frac{\partial M^{-3}}{\partial t_1}$$

$$= \quad M^{-1}E(U_1^4 S) - M^{-2}E(U_1 S)E(U_1^3 S)$$

$$- M^{-2}[E(U_1^3 S)E(U_1 S) + (E(U_1^2 S))^2]$$

$$+ 2M^{-3}(E(U_1 S))^2 E(U_1^2 S)$$

$$- 2M^{-2}[(E(U_1^2 S)^2) + E(U_1 S)E(U_1^3 S)]$$

$$+ 4M^{-3}(E(U_1 S))^2 E(U_1^2 S) + 2M^{-3}[3(E(U_1 S))^2 E(U_1^2 S)]$$

$$- 6M^{-4}(E(U_1 S))^4,$$

$$\begin{aligned}
\frac{\partial^4 K(t_1,t_2)}{\partial t_1^3 \partial t_2} =& M^{-1}\frac{\partial E(U_1^3 S)}{\partial t_2} + E(U_1^3 S)\frac{\partial M^{-1}}{\partial t_2} - M^{-2}\frac{\partial E(U_1^2 S)E(U_1 S)}{\partial t_2} \\
&- E(U_1^2 S)E(U_1 S)\frac{\partial M^{-2}}{\partial t_2} - 2M^{-2}\frac{\partial E(U_1 S)E(U_1^2 S)}{\partial t_2} \\
&- 2E(U_1 S)E(U_1^2 S)\frac{\partial M^{-2}}{\partial t_2} + 2M^{-3}\frac{\partial (E(U_1 S))^3}{\partial t_2} \\
&+ 2(E(U_1 S))^3\frac{\partial M^{-3}}{\partial t_2} \\
=& \quad M^{-1}E(U_1^3 U_2 S) - M^{-2}E(U_2 S)E(U_1^3 S) \\
&- M^{-2}[E(U_1^2 U_2 S)E(U_1 S) + E(U_1 U_2 S)E(U_1^2 S)] \\
&+ 2M^{-3}E(U_2 S)E(U_1 S)E(U_1^2 S) - 2M^{-2}[E(U_1 U_2 S)E(U_1^2 S) \\
&+ E(U_1^2 U_2 S)E(U_1 S)] + 4M^{-3}E(U_1 S)E(U_2 S)E(U_1^2 S) \\
&+ 6M^{-3}(E(U_1 S))^2 E(U_1 U_2 S) - 6M^{-4}(E(U_1 S))^3 E(U_2 S),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4 K(t_1,t_2)}{\partial t_1^2 t_2^2} =& M^{-1}\frac{\partial E(U_1 U_2^2 S)}{\partial t_1} + E(U_1 U_2^2 S)\frac{\partial M^{-1}}{\partial t_1} - M^{-2}\frac{\partial E(U_2^2 S)E(U_1 S)}{\partial t_1} \\
&- E(U_2^2 S)E(U_1 S)\frac{\partial M^{-2}}{\partial t_1} - 2M^{-2}\frac{\partial E(U_1 U_2 S)E(U_2 S)}{\partial t_1} \\
&- 2E(U_1 U_2 S)E(U_2 S)\frac{\partial M^{-2}}{\partial t_1} + 2M^{-3}\frac{\partial E(U_1 S)(E(U_2 S))^2}{\partial t_1} \\
&+ 2E(U_1 S)(E(U_2 S))^2\frac{\partial M^{-3}}{\partial t_1} \\
&= M^{-1}E(U_1^2 U_2^2 S) - M^{-2}E(U_1 U_2^2 S)E(U_1 S) \\
&- M^{-2}[E(U_1 U_2^2 S)E(U_1 S) + E(U_1^2 S)E(U_2^2 S)] \\
&+ 2M^{-3}E(U_1 S)E(U_1 S)E(U_2^2 S) - 2M^{-2}[E(U_1^2 U_2 S)E(U_2 S) \\
&+ E(U_1 U_2 S)E(U_1 U_2 S)] \\
&+ 4M^{-3}E(U_1 S)E(U_2 S)E(U_1 U_2 S) + 2M^{-3}[E(U_1^2 S)(E(U_2 S))^2 \\
&+ 2E(U_2 S)E(U_1 U_2 S)E(U_1 S)] - 6M^{-4}(E(U_1 S))^2(E(U_2 S))^2,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4 K(t_1,t_2)}{\partial t_2^3 \partial t_1} &= M^{-1}\frac{\partial E(U_2^3 S)}{\partial t_1} + E(U_2^3 S)\frac{\partial M^{-1}}{\partial t_1} - M^{-2}\frac{\partial E(U_2^2 S)E(U_2 S)}{\partial t_1} \\
&- E(U_2^2 S)E(U_2 S)\frac{\partial M^{-2}}{\partial t_1} - 2M^{-2}\frac{\partial E(U_2 S)E(U_2^2 S)}{\partial t_1} \\
&- 2E(U_2 S)E(U_2^2 S)\frac{\partial M^{-2}}{\partial t_1} + 2M^{-3}\frac{\partial (E(U_2 S))^3}{\partial t_1} \\
&+ 2(E(U_2 S))^3\frac{\partial M^{-3}}{\partial t_1} \\
&= M^{-1}E(U_2^3 U_1 S) - M^{-2}E(U_1 S)E(U_2^3 S) - M^{-2}[E(U_2^2 U_1 S)E(U_2 S) \\
&+ E(U_2 U_1 S)E(U_2^2 S)] + 2M^{-3}E(U_1 S)E(U_2 S)E(U_2^2 S) \\
&- 2M^{-2}[E(U_1 U_2 S)E(U_2^2 S) + E(U_2^2 U_1 S)E(U_2 S)] \\
&+ 4M^{-3}E(U_2 S)E(U_1 S)E(U_2^2 S) + 6M^{-3}(E(U_2 S))^2 E(U_2 U_1 S) \\
&- 6M^{-4}(E(U_2 S))^3 E(U_1 S),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4 K(t_1,t_2)}{\partial t_2^4} &= M^{-1}\frac{\partial E(U_2^3 S)}{\partial t_2} + E(U_2^3 S)\frac{\partial M^{-1}}{\partial t_2} - M^{-2}\frac{\partial E(U_2^2 S)E(U_2 S)}{\partial t_2} \\
&- E(U_2^2 S)E(U_2 S)\frac{\partial M^{-2}}{\partial t_2} - 2M^{-2}\frac{\partial E(U_2 S)E(U_2^2 S)}{\partial t_2} \\
&- 2E(U_2 S)E(U_2^2 S)\frac{\partial M^{-2}}{\partial t_2} + 2M^{-3}\frac{\partial (E(U_2 S))^3}{\partial t_2} + 2(E(U_2 S))^3\frac{\partial M^{-3}}{\partial t_2} \\
&= M^{-1}E(U_2^4 S) - M^{-2}E(U_2 S)E(U_2^3 S) - M^{-2}[E(U_2^3 S)E(U_2 S) \\
&+ (E(U_2^2 S))^2] + 2M^{-3}(E(U_2 S))^2 E(U_2^2 S) - 2M^{-2}[(E(U_2^2 S)^2) \\
&+ E(U_2 S)E(U_2^3 S)] + 4M^{-3}(E(U_2 S))^2 E(U_2^2 S) \\
&+ 2M^{-3}[3(E(U_2 S))^2 E(U_2^2 S)] - 6M^{-4}(E(U_2 S))^4.
\end{aligned}$$

1.3.5 Cumulants

Let k_1, k_2, k_3 and k_4 represent the first, second, third ,fourth cumulant respectively. Then the first cumulant is given by

$$\frac{\partial K(t_1,t_2)}{\partial t_1}|_{t_1=0,t_2=0} = E(U_1),$$

$$\frac{\partial K(t_1,t_2)}{\partial t_2}|_{t_1=0,t_2=0} = E(U_2),$$

Hence, the first cumulant is done.

The second cumulant is given by

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_1^2}|_{t_1=0,t_2=0} = E(U_1^2) - (E(U_1))^2,$$

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_1 \partial t_2}|_{t_1=0,t_2=0} = E(U_1U_2) - E(U_1)E(U_2),$$

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_2^2}|_{t_1=0,t_2=0} = E(U_2^2) - (E(U_2))^2,$$

$$k_2 = E(U_1^2) - (E(U_1))^2 + E(U_1U_2) - E(U_1)E(U_2) + E(U_2^2) - (E(U_2))^2.$$

The second cumulant is done.

$$\frac{\partial^3 K(t_1,t_2)}{\partial t_1^3}|_{t_1=0,t_2=0} = E(U_1^3) - 3E(U_1^2)E(U_1) + 2(E(U_1))^3,$$

$$\frac{\partial^3 K(t_1,t_2)}{\partial t_2^3}|_{t_1=0,t_2=0} = E(U_2^3) - 3E(U_2^2)E(U_2) + 2(E(U_2))^3,$$

$$\begin{aligned}\frac{\partial^3 K(t_1,t_2)}{\partial t_1^2 \partial t_2}|_{t_1=0,t_2=0} &= E(U_1U_2^2) - E(U_2^2)E(U_1) - 2E(U_1U_2)E(U_2) \\ &\quad +2(E(U_2))^2E(U_1),\end{aligned}$$

$$\begin{aligned}\frac{\partial^3 K(t_1,t_2)}{\partial t_1 \partial t_2^2}|_{t_1=0,t_2=0} &= E(U_1^2U_2) - E(U_1^2)E(U_2) - 2E(U_1U_2)E(U_1) \\ &\quad +2(E(U_1))^2E(U_2).\end{aligned}$$

$$\begin{aligned}k_3 = {} & E(U_1^3) + E(U_2^3) + E(U_1U_2^2) + E(U_1^2U_2) - 3E(U_1^2)E(U_1) - 3E(U_2^2)E(U_2) \\ & - E(U_1)E(U_2^2) - E(U_1^2)E(U_2) - 2E(U_1)E(U_1U_2) - 2E(U_2)E(U_1U_2) \\ & + 2E(U_1)(E(U_2))^2 + 2(E(U_1))^2E(U_2) + 2(E(U_1))^3 + 2(E(U_2))^3.\end{aligned}$$

The third cumulant is done.

$$\frac{\partial^4 K(t_1,t_2)}{\partial t_1^4}|_{t_1=0,t_2=0} = E(U_1^4) - E(U_1)E(U_1^3) - E(U_1^3)E(U_1) - (E(U_1^2))^2$$
$$+ 2(E(U_1))^2E(U_1^2) - 2[(E(U_1^2))^2 + E(U_1)E(U_1^3)]$$
$$+ 4(E(U_1))^2E(U_1^2) + 6(E(U_1))^2E(U_1^2) - 6(E(U_1))^4,$$

$$\frac{\partial^4 K(t_1,t_2)}{\partial t_2^4}|_{t_1=0,t_2=0} = E(U_2^4) - E(U_2)E(U_2^3) - E(U_2^3)E(U_2) - (E(U_2^2))^2$$
$$+ 2(E(U_2))^2E(U_2^2) - 2[(E(U_2^2))^2 + E(U_2)E(U_2^3)]$$
$$+ 4(E(U_2))^2E(U_2^2) + 6(E(U_2))^2E(U_2^2) - 6(E(U_2))^4,$$

$$\frac{\partial^4 K(t_1,t_2)}{\partial t_1^3 \partial t_2}|_{t_1=0,t_2=0} = E(U_1^3U_2) - E(U_1^3)E(U_2) - E(U_1)E(U_1^2U_2)$$
$$- E(U_1^2)E(U_1U_2) + 2E(U_1)E(U_2)E(U_1^2)$$
$$- 2E(U_1^2)E(U_1U_2) - 2E(U_1)E(U_1^2U_2)$$
$$+ 4E(U_1)E(U_2)E(U_1^2) + 6(E(U_1))^2E(U_1U_2)$$
$$- 6(E(U_1))^3E(U_2),$$

$$\frac{\partial^4 K(t_1,t_2)}{\partial t_1 \partial t_2^3}|_{t_1=0,t_2=0} = E(U_1U_2^3) - E(U_1)E(U_2^3) - E(U_2)E(U_2^2U_1)$$
$$- E(U_2^2)E(U_1U_2) + 2E(U_1)E(U_2)E(U_2^2)$$
$$- 2E(U_2^2)E(U_1U_2) - 2E(U_2)E(U_1U_2^2)$$
$$+ 4E(U_1)E(U_2)E(U_2^2) + 6(E(U_2))^2E(U_1U_2)$$
$$- 6(E(U_2))^3E(U_1),$$

$$\begin{aligned}\frac{\partial^4 K(t_1,t_2)}{\partial t_1^2 \partial t_2^2}|_{t_1=0,t_2=0} &= E(U_1^2U_2^2) - 2E(U_1)E(U_1U_2^2) - E(U_1^2)E(U_2^2) \\ &+ 2(E(U_1))^2E(U_2^2) - 2E(U_2)E(U_1^2U_2) \\ &- 2E((U_1U_2))^2 + 4E(U_1)E(U_2)E(U_1U_2) \\ &+ 2E(U_1^2)(E(U_2))^2 + 4E(U_1)E(U_2)E(U_1U_2) \\ &- 6(E(U_1))^2(E(U_2))^2,\end{aligned}$$

The fourth cumulant is done.

1.4 Conclusion

These calculations were used in [10], in conjunction with the bivariate two-dimensional Cornish-Fisher approximation to produce Table 1 of approximate and exact test levels and powers.

Chapter 2

Relative Accuracy of Multivariate Bootstrap Procedures

The last chapter utilized asymptotic technique to solve a problem which cannot be handled by exact calculation. Chapter 2 describes and solves a different asymptotic problem: finding the relative accuracy of two values from two methods. In other words, find how close they are.

Concretely, I demonstrate that the p-values of the multivariate studentized bootstrap and tilted bootstrap are very close and show that the relative accuracy is $O(\frac{1}{n})$ for $\bar{x} = O(n^{-1/3})$ under the null hypothesis $H_0 : \mu = 0$. I also prove the two bootstrap approximations have the approximately equal covariance matrices.

Section 2.1 introduces the saddlepoint test. Section 2.2 focuses on two kinds of bootstrap methods, studentized and tilted. Section 2.3 finds the relative accuracy of them and Section 2.4 provides the simulation for the conclusion.

2.1 Introduction to Saddlepoint Test

This section gives a brief introduction to the saddlepoint test [11].

Let $X_1, ...X_n$ be identically and independently distributed with the distribution F on the sample space $\mathbf{R}^d$. Let $\mathrm{E}(X_1) = \mu$ and $\mathrm{Var}(X_1) = \Sigma$. Consider a test of the null hypothesis $H_0 : \mu = 0$ in $\mathbf{R}^d$ versus the alternative hypothesis that $\mu \neq 0$. Let $x_1, ..., x_n$ be the observed values of $X_1, ..., X_n$. The empirical mean and covariance matrix are

$$\bar{x} = \sum_{i=1}^{n} x_i/n, \tag{2.1}$$

and

$$\hat{\Sigma} = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T/(n-1). \tag{2.2}$$

Assume that the density and cumulant-generating function of $X_1, ...X_n$ exist, so that the saddlepoint approximation [6] for the density of $\bar{X}$ is

$$f_{\bar{X}}(x) = (2\pi/n)^{d/2} e^{-nh(x)} |K''(\hat{\tau})|^{-1/2} (1 + O(n^{-1})).$$

Let $K'(\tau)$ and $K''(\tau)$ denote the first and second derivative of the known cumulant-generating function $K(\tau)$ with respect to τ, and $K(\tau)$ is finite for $|\tau| < c$ for some $c > 0$. The saddlepoint $\hat{\tau}$ is the solution of

$$K'(\tau) = \bar{x}. \tag{2.3}$$

We define $h(x)$ as

$$h(x) = \hat{\tau}^T x - K(\hat{\tau}). \tag{2.4}$$

Section 4 in [9] defines the p-value based on the statistic $h(\bar{X})$, to be

$$p = P(h(\bar{X}) > h(\bar{x})), \tag{2.5}$$

and he derives the approximation

$$p = \bar{Q}_d(2nh(\bar{x}))(1 + O((1 + 2nh(\bar{x}))/n)), \tag{2.6}$$

in equation (4.2) of his paper, where $Q_d(x)$ is the Chi-square distribution function and $\bar{Q}_d(z) = P(\chi^2_d > z)$. Given the distribution F and the sample, one might find the cumulant function, and solve $\hat{\tau}, h(\bar{x})$ and u. Next, we consider the case with unknown distribution F.

2.2 The Studentized and Tilted Bootstrap

When the distribution F is unknown, the probability and cumulant-generating functions cannot be determined, but the bootstrap can be used to approximate the p-value of form (2.5).

The bootstrap is a random resampling method with replacement. Suppose we have samples $x_1, x_2 ... x_n$. When bootstrapping, these samples are treated as the population.

In studentized bootstrap, every sample has the same probability. In order to make a sampling distribution compatible with the null hypothesis H_0, we introduce the tilted bootstrap, where every observation has a weighted probability. The probability of selecting observation i is

$$q_i = \frac{e^{\beta^T x_i}}{\sum_{j=1}^{n} e^{\beta^T x_j}}, \tag{2.7}$$

where β is determined by

$$\frac{\sum_{i=1}^{n} x_i e^{\beta^T x_i}}{\sum_{i=1}^{n} e^{\beta^T x_i}} = \mu = 0. \tag{2.8}$$

Recall that $\mu = 0$ under the null hypothesis. The bootstrap distribution, then, is the distributions of n observations $X_1^*, \ldots, X_n^*$ selected with replacement, where each observation is independent and identically distributed, and has probability q_i of begin x_i. Given the samples $x_1 ... x_n$, the conditional studentized mean is $EX_s^* = \bar{x}$. The expectation of the tilted bootstrap is

$$E(X_t^*) = \sum_{i=1}^{n} q_i x_i = \sum_{i=1}^{n} \frac{e^{\beta^T x_i}}{\sum_{j=1}^{n} e^{\beta^T x_j}} x_i = \mu, \tag{2.9}$$

which gives an expectation for the weighted distribution consisted with the null hypothesis H_0. That justifies the introduction of the tilted bootstrap. Let Σ_s be the covariance matrix of Studentized bootstrap and Σ_t be that of tilted bootstrap. By the definition of population variance, the studentized bootstrap covariance matrix is

$$\Sigma_s = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T. \tag{2.10}$$

The tilted bootstrap covariance matrix is defined as

$$\Sigma_t = \sum_{j=1}^{n} q_j (x_j - \bar{x})(x_j - \bar{x})^T. \tag{2.11}$$

where q_j is defined in(2.7).

A lot of comparisons between the one-dimensional studentized and tilted bootstraps are performed in [8], but no inference between them in multi-dimensions.

Lemma 2.1 The covariance matrix of studentized bootstrap Σ_s is asymptotically equal to that of tilted bootstrap Σ_t as

$$\Sigma_t = \Sigma_s + O_P(||\bar{x}||). \tag{2.12}$$

Proof:

The covariance matrix of studentized bootstrap denoted as is

$$\begin{aligned}\Sigma_s =&\mathrm{E}(\frac{1}{n}\sum_{j=1}^{n}(X_i^* - \bar{X^*})(X_i^* - \bar{X^*})^T|x_1...x_n)\\ =&\frac{1}{n}\sum_{j=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T\\ =&\frac{n-1}{n}\hat{\Sigma}\end{aligned}$$

The tilted bootstrap covariance matrix is

$$\begin{aligned}\Sigma_t =&\mathrm{Var}(X_t^*|x_1...x_n)\\ =&\sum_{j=1}^{n} q_j(x_j - \bar{x})(x_j - \bar{x})^T,\end{aligned} \tag{2.13}$$

where q_j is defined as

$$q_j = \frac{e^{x_j^T\beta}}{\sum_{i=1}^{n} e^{x_i^T\beta}}.$$

Using a Taylor series,

$$q_j = \frac{1}{n}(1 + \beta^T(x_j - \bar{x})) + O(\beta^2). \tag{2.14}$$

By (2.14) substitution in (2.13), we obtain

$$\begin{aligned}\Sigma_t =&\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})^T + \frac{1}{n}\sum_{j=1}^{n}\beta^T(x_j - \bar{x})(x_j - \bar{x})(x_j - \bar{x})^T\\ =&\Sigma_s + \frac{1}{n}\sum_{j=1}^{n}\beta^T(x_j - \bar{x})(x_j - \bar{x})(x_j - \bar{x})^T.\end{aligned} \tag{2.15}$$

Consider an element of the second term. Let $[\quad]_{kl}$ represent the component k, l.

$$\begin{aligned}&\frac{1}{n}\Big[\sum_{j=1}^{n}\beta^T(x_j - \bar{x})(x_j - \bar{x})(x_j - \bar{x})^T\Big]_{kl}\\ =&\frac{1}{n}\Big[\sum_{i=1}^{n}\beta_i\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)(x_{jl} - \bar{x}_l)\Big]\\ =&\sum_{i=1}^{n}\beta_i k_{ikl}\\ =&O_P(||\beta||).\end{aligned} \tag{2.16}$$

The k_{ikl} is the component i,k,l of the third cumulant, and it is bounded.

$$\mu = K'(\beta) = K'(0) + K''(0)\beta + \beta^T K'''(0)\beta + O_P(||\beta||^3),$$

$$0 = \bar{x} + \Sigma\beta + \beta^T K'''(0)\beta + O_P(||\beta||^3), \tag{2.17}$$

simplifying (2.17),

$$\beta = -\Sigma^{-1}\bar{x}. + O_P(||\beta||^2).$$

The covariance matrix Σ is bounded. So we get

$$O_P(||\beta||) = O_P(||\bar{x}||). \tag{2.18}$$

Thus combining (2.15), (2.16) and (2.18), we prove

$$\Sigma_t = \Sigma_s + O_P(||\bar{x}||). \tag{2.19}$$

Q.E.D.

These results will be used later.

Let τ_t to denote the solutions with respect to τ in (2.20),

$$K_t'(\tau) = \bar{x}, \tag{2.20}$$

where $K_t(\tau) = \log(\sum_{i=1}^n q_i e^{\tau^T x_i})$ is the tilted bootstrap cumulant-generating function. The prime denotes the first derivative.

We use τ_t and $K_t'(\tau)$ to substitute $\hat{\tau}$ and $K'(\tau)$ in (2.2), to obtain

$$h_t(\bar{x}) = 2(\tau_t^T\bar{x} - K_t(\tau_t)). \tag{2.21}$$

Robinson [9] proved that

$$p_s^* = \bar{Q}_d(nh_s(\bar{x}))(1 + O((1 + 2nh_s(\bar{x}))/n)), \tag{2.22}$$

and that the tilted bootstrap p-value is

$$p_t^* = \bar{Q}_d(nh_t(\bar{x}))(1 + O((1 + 2nh_t(\bar{x}))/n)). \tag{2.23}$$

Both $h_s(\bar{x})$ and $h_t(\bar{x})$ are infinitely differentiable functions in a neighborhood of origin so they are analytic functions. We can write

$$h_t(\bar{x}) = C_{t0} + C_{t1}\bar{x} + \bar{x}^T C_{t2}\bar{x} + O(||\bar{x}||^3), \tag{2.24}$$

where the C_{t0} is a scalar, the C_{t1} is a vector, the C_{t2} is a matrix, and the $||\cdot||$ is Euclidean norm. Using the value when $\bar{x} = 0$, we obatin $C_{t0} = 0$. Let $v_t = 2(\tau_t^T\bar{x} - K_t(\tau_t))$, the first derivative with respect to $\bar{x}$ is

$$v_t' = 2\tau_t.$$

The first coefficient is

$$C_{t_1} = v_t'(0) = 2\tau_t(0) = 0.$$

The second derivative is

$$v_t'' = 2\tau_t', \quad v_t''(0) = 2\tau_t'(0).$$

By the equation

$$K_t'(\tau_t) = \bar{x}, \quad K_t''(\tau_t)\tau_t' = I,$$

Then obtain,

$$\tau_t' = K_t''(\tau_t)^{-1}, \quad \tau_t'(0) = K_t''(0)^{-1}.$$

Thus $C_{t2} = \frac{1}{2!}2v_t''(0) = \hat{\Sigma}^{-1}$.

we established that $C_{t0} = 0$, $C_{t1} = 0$ and $C_{t2} = \hat{\Sigma}^{-1}$. By substitution in (2.24), we find

$$h_t(\bar{x}) = \bar{x}^T \hat{\Sigma}^{-1} \bar{x} + O(||\bar{x}||^3). \tag{2.25}$$

By the same reasoning, obtain

$$h_s(\bar{x}) = \bar{x}^T \Sigma_s^{-1} \bar{x} + O(||\bar{x}||^3) = \frac{n}{n-1} \bar{x}^T \hat{\Sigma}^{-1} \bar{x} + O(||\bar{x}||^3).$$

The above results are vital to find the relative accuracy.

2.3 Relative Accuracy of the Studentized and Tilted Bootstrap

The lemma similar to Mill's ratio [15] for chi-squared distribution is proven as follow.

Lemma 2.2: The **pdf.** of Chisqaure distribution with d degree of freedom is

$$f(x) = \frac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1} e^{-x/2}. \tag{2.26}$$

The **cdf** is denoted as $F(x)$. We define

$$R_x = \frac{1 - F(x)}{f(x)}. \tag{2.27}$$

The limiting value is

$$\lim_{x \to \infty} R_x = 2. \tag{2.28}$$

Proof: Simplifying (2.27),

$$R_x = x^{1-d/2} e^{x/2} \int_x^\infty t^{d/2-1} e^{-t/2} dt. \tag{2.29}$$

The derivative can be easily shown to be,

$$\frac{dR_x}{dx} = R_x(\frac{1}{2} + \frac{1}{x} - \frac{d}{2x}) - 1. \tag{2.30}$$

The second derivative is

$$\frac{d^2R_x}{dx^2} = \frac{dR_x}{dx}(\frac{1}{2} + \frac{1}{x} - \frac{d}{2x}) + R_x(-\frac{1}{x^2} + \frac{d}{2x^2}); \tag{2.31}$$

simplifying,

$$\frac{d^2R_x}{dx^2} = R(\frac{1}{4} + \frac{d^2}{4x^2} + \frac{1}{x} - \frac{d}{2x} - \frac{d}{2x^2}) - (\frac{1}{2} + \frac{1}{x} - \frac{d}{2x}). \tag{2.32}$$

The third derivative is

$$\frac{d^3R_x}{dx^3} = \frac{d^2R_x}{dx^2}(\frac{1}{2} + \frac{1}{x} - \frac{d}{2x}) + 2\frac{dR_x}{dx}(-\frac{1}{x^2} + \frac{d}{2x^2}) + R_x(\frac{2}{x^3} - \frac{d}{x^3}). \tag{2.33}$$

By (2.27), it is easy to know

$$R_x > 0, \tag{2.34}$$

and

$$\lim_{x\to\infty} \frac{dR_x}{dx} = 0. \tag{2.35}$$

The d is the degrees of freedom. Consider $d > 2$. The idea is to prove (2.30) is less than 0 and (2.31) is greater than 0.

(i) If $\exists x_1$ such that

$$\left.\frac{dR_x}{dx}\right|_{x=x_1} > 0, \tag{2.36}$$

because $(\frac{1}{2} + \frac{1}{x} - \frac{d}{2x})$ is increasing with respect to x, both R_x and $\frac{dR_x}{dx}$ will increase with x. Thus, the R_x and $\frac{dR_x}{dx}$ are divergent.The conclusion contracts (2.35). So $\forall x > 0$,

$$\frac{dR_x}{dx} < 0. \tag{2.37}$$

By (2.30) and (2.37), we get

$$R_x < \frac{2x}{x+2-d} = 2 + \frac{2d-4}{x+2-d}. \tag{2.38}$$

(ii) If $\exists x_2$ such that

$$\frac{d^2 R_x}{dx^2}\Big|_{x=x_2} < 0,$$

by (2.33), we obtain

$$\frac{d^3 R_x}{dx^3}\Big|_{x=x_2} < 0.$$

Thus both $\frac{d^2 R_x}{dx^2}$ and $\frac{d^3 R_x}{dx^3}$ are negative in $[x_2, +\infty]$. The $\frac{d^2 R_x}{dx^2}$ will be divergent. The conclusion contracts (2.35). It follows that (2.32) is positive. By (2.32),

$$R_x > \frac{x+2-d}{2x} \cdot \frac{4x^2}{x^2+d^2+4x-2xd-2d}.$$

Simplifying

$$R_x > 2 + \frac{2d-4+\frac{4d-2d^2}{x}}{x+4-2d+\frac{d^2-2d}{x}}. \tag{2.39}$$

According to (2.38) and (2.39), it is easy to see

$$\lim_{x\to\infty} R_x = 2.$$

Q.E.D.

Lemma 2.2 is used for the following calculation.

The relative accuracy is define to be

$$\frac{p_t^*}{p_s^*} = \frac{\bar{Q}_d(nh_t(\bar{x}))(1+O((1+nh_t(\bar{x}))/n))}{\bar{Q}_d(nh_s(\bar{x}))(1+O((1+nh_s(\bar{x}))/n))}. \tag{2.40}$$

where $\bar{Q}_d(x) = P(\chi_d^2 > x)$. Based on the assumption

$$\frac{1+O((1+nh_t(\bar{x}))/n)}{1+O((1+nh_s(\bar{x}))/n)} = 1 + O(\frac{1}{n}),$$

and by simple arithmetic, we obtain

$$\frac{p_t^*}{p_s^*} = \frac{\bar{Q}_d(nh_t(\bar{x}))}{\bar{Q}_d(nh_s(\bar{x}))}(1+O(\frac{1}{n})),$$
$$\frac{p_t^*}{p_s^*} = \frac{\bar{Q}_d(nh_t(\bar{x})) + \bar{Q}_d(nh_s(\bar{x})) - \bar{Q}_d(nh_s(\bar{x}))}{\bar{Q}_d(nh_s(\bar{x}))}(1+O(\frac{1}{n})),$$
$$\frac{p_t^*}{p_s^*} = 1 + O_P(\frac{\bar{Q}_d(nh_t(\bar{x})) - \bar{Q}_d(nh_s(\bar{x}))}{\bar{Q}_d(nh_s(\bar{x}))}).$$

By the mean value theorem,

$$\frac{p_t^*}{p_s^*} = 1 + O_P(\frac{n(h_t(\bar{x}) - h_s(\bar{x}))}{\bar{Q}_d(nh_s(\bar{x}))}q_d(\xi)),$$

where $q_d(x)$ is the density function of chi-squared distribution with d.f. d. The $\xi \in (\min(nh_t(\bar{x}), nh_s(\bar{x})), \max(nh_t(\bar{x}), nh_s(\bar{x})))$.

By Lemma 2.1 $\lim\limits_{x\to\infty} \frac{q_d(x)}{\bar{Q}_d(x)} = \frac{1}{2}$, and we obtain

$$\frac{p_t^*}{p_s^*} = 1 + O_p(n(h_t(\bar{x}) - h_s(\bar{x}))). \tag{2.41}$$

Recall the formulas,

$$h_s(\bar{x}) = \frac{n}{n-1}\bar{x}^T\hat{\Sigma}^{-1}\bar{x}$$
$$h_t(\bar{x}) = \bar{x}^T\Sigma_t^{-1}\bar{x} + O(||\bar{x}||^3) = \bar{x}^T\hat{\Sigma}^{-1}\bar{x} + O_p(||\bar{x}||^3).$$

Then (2.41) is

$$\frac{p_t^*}{p_s^*} = 1 + O_P(n||\bar{x}||^3) + O_P(||\bar{x}||^2). \tag{2.42}$$

Consequently, for $||\bar{x}|| = o(n^{-1/3})$, (2.42) shows that although the studentized and tilted bootstraps use different resampling methods, the p-values will be approximately equal when sample size is large.

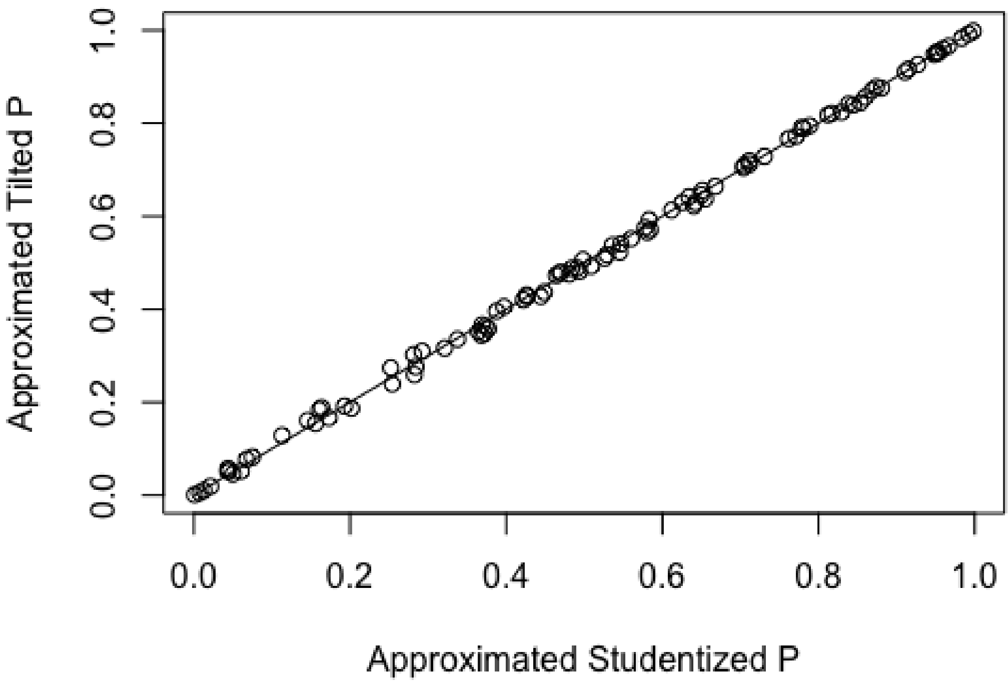

Figure 2.1: Comparison between two Approximated p-values

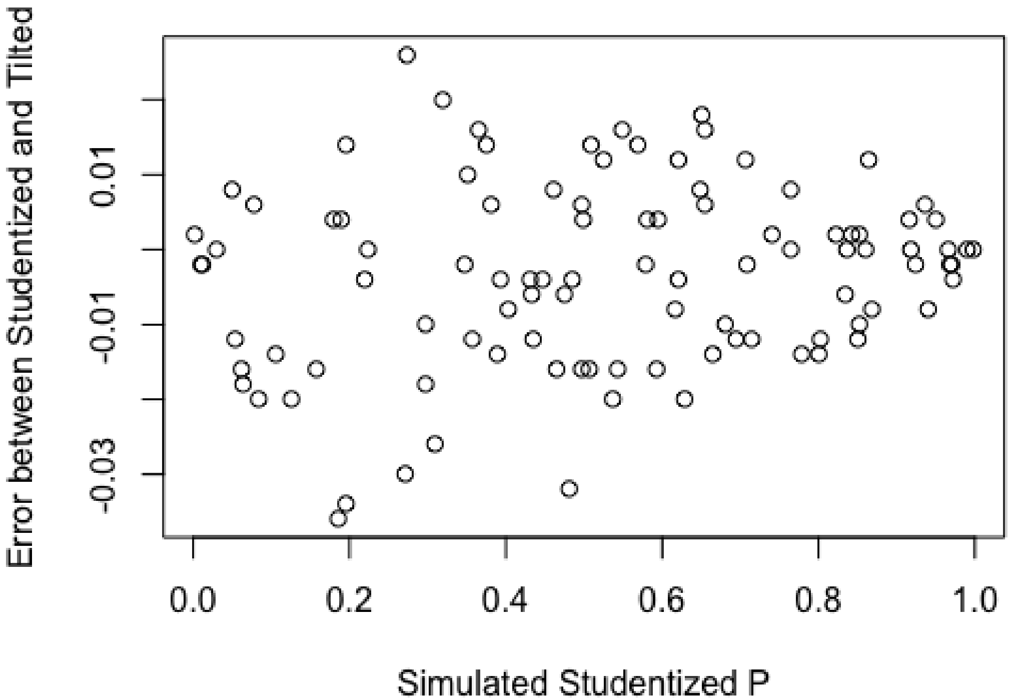

Figure 2.2: Comparison between Approximated and Simulated Tilted p-values

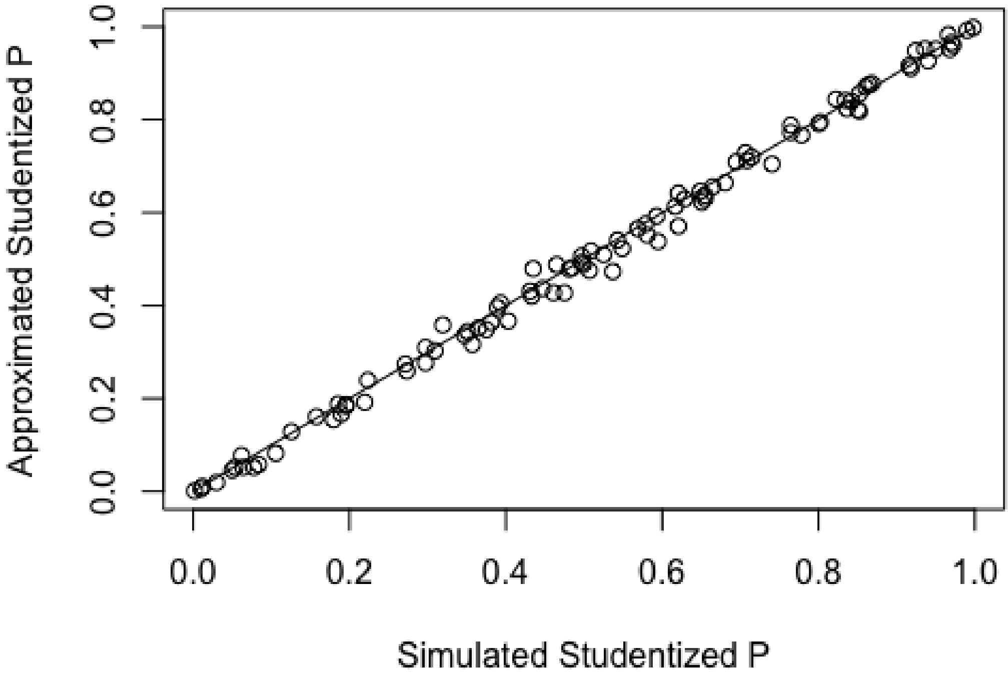

Figure 2.3: Comparison between Approximated and Simulated Studentized p-values

2.4 Simulation and Conclusion

We compare four p-values in the section. Two of them are theoretically asymptotic p-values for Studentized and tilted bootstrap, as discussed before, defined as

$$\tilde{p_s} = 1 - Q(nh_s(\bar{x})),$$

$$\tilde{p_t} = 1 - Q(nh_t(\bar{x})).$$

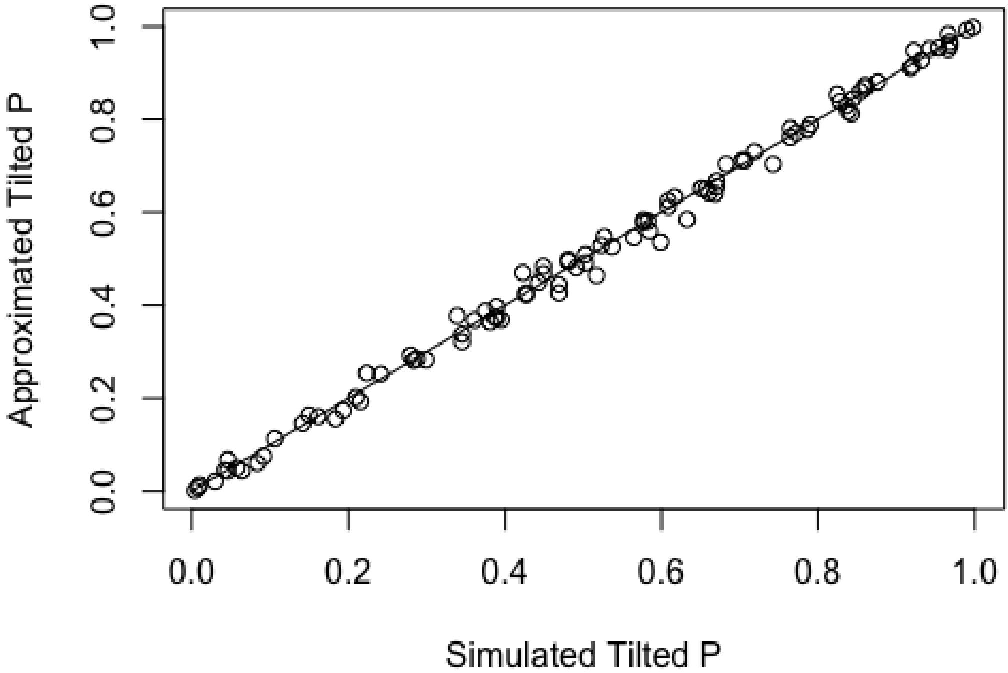

Figure 2.4: Comparison between Approximated and Simulated Tilted p-values

The other ones are simulated results by

$$p_s^* = P(h_s^*(\bar{x^*}) \geq h_s(\bar{x})),$$

$$p_t^* = P(h_t^*(\bar{x^*}) \geq h_t(\bar{x})).$$

For each dataset, we simulated 500 times for the Studentized and tilted respectively and get 500 $h_s^*(\bar{x^*})$ and $h_t^*(\bar{x^*})$ to compare with $h_s(\bar{x})$ and $h_t(\bar{x})$ to get the simulated values p_s^* and p_t^*. Furthermore, 100 datasets are used to do the comparison.

The comparison between theoretically approximated Studentized bootstrap $\tilde{p_s}$ and tilted $\tilde{p_t}$ are showed in Figure 1. To further investigate the error, we show the difference between them in Figure 2 and the difference is close to 0.

The comparison and error justify our conclusion. We also compare $\tilde{p}_s$ and p_s^* in Figure 3, $\tilde{p}_t$ and p_t^* in Figure 4. All of them are close. The phenomenon shows Studentized and tilted bootstrap methods have very similar performance.

Chapter 3

The Covariance Test

Chapter 3 and 4 focus on the selective inference of a significance test, covariance test. Chapter 3 introduces the background and reviews of the covariance test, as well as my simulation investigation. Chapter 4 provides my new results for the covtest.

3.1 Review of Lasso Regression

Consider the usual linear regression model [13], for an outcome vector $y \in R^n$ and matrix of predictor variables $X \in R^{n \times p}$:

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I); \tag{3.1}$$

here $\beta \in R^p$ are unknown coefficients to be estimated. The lasso estimator [14] is

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2} ||y - X\beta||_2^2 + \lambda ||\beta||_1, \tag{3.2}$$

where λ is a tuning parameter. The solution $\hat{\beta}$ is a continuous and piecewise linear function of λ. When $\lambda = \infty$, all of variables in $\hat{\beta}(\infty)$ are inactive. A

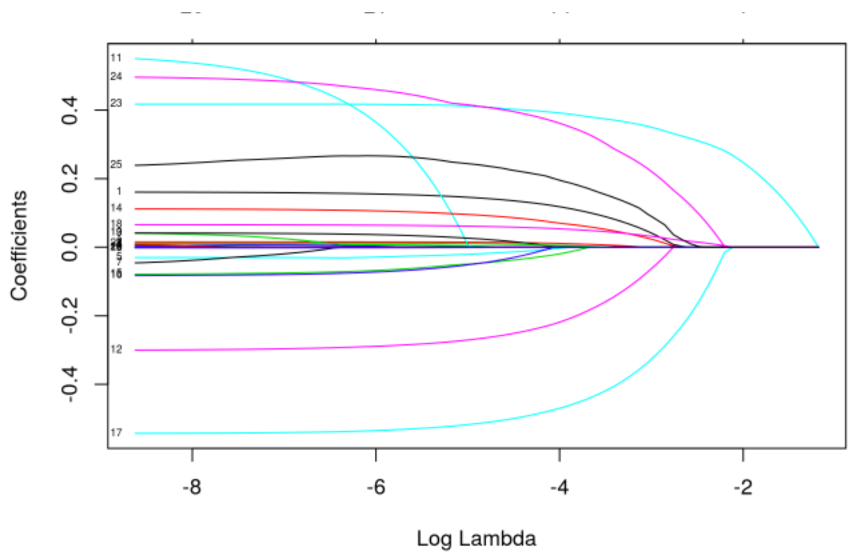

Figure 3.1: Lasso Path

variable is called inactive if its coefficient is zero. When $\lambda = 0$, lasso regression linear regression and all variables are active. When λ decreases from ∞ to 0, each λ_k marks the entry of a variable. For example, when $\lambda = \lambda_1$, the first variable enters the active set. The coefficient of the active variable is nonzero and the coefficients of the rest variables are zeroes. We use A to denote the active set before k. The X_A is the matrix with active predictors. For instance, when $k = 0$, $\lambda_{k+1} = \lambda_1$ is the entry of the first variable and A is an empty set as well as X_A. When $k = 4$, $\lambda_{k+1} = \lambda_5$ marks the entry of the fifth variable. The active set A includes 4 variables and the matrix X_A has 4 columns. Figure 3.1 illustrates the Lasso path containing 25 variables. From right to left, with the λ decreases, more variables become active.

There has been a considerable amount of recent work dedicated to the lasso problem. I give a short summary of what I cite: [12], [20], [19], [18] [17] and [13].

3.2 Problems in Model Selection and Significance Testing

When I choose different λ, different models are selected. The first problem is how to choose the active variables. Once the model is determined, the second problem is inference after selection.

3.2.1 Model Selection

The Akaike information criterion and Bayesian information criterion are often utilized to do the variable selection. This section reviews these before presenting a testing-based approach.

Consider a statistical model for some data. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}). \tag{3.3}$$

A set of candidate models are constructed, and then the corresponding AIC values are determined. The smaller the AIC value is, the better the model is. Although the method is widely used in model selection, it is still heuristic. Simlarly, the Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models. It is based, on the likelihood function and it is closely related to the Akaike information criterion (AIC). The BIC is defined as

$$\mathrm{BIC} = k\ln(n) - 2\ln(\hat{L}), \tag{3.4}$$

where $\hat{L}$ and k are defined as those in AIC. The n is the number of observations.

Both of the methods are empirical methods based on the rule of thumb. In comparison, the covariance test is tied to a statistical test with a known level.

3.2.2 Model Selection via Significance Testing

The traditional method to do significance testing is to find the residual sum of square and then do Chi-square test with the known σ^2 or t-test with the unknown σ^2.

For example, the set A contains $j-1$ variables. We try adding the variable j. The sum of squares statistic is

$$R_j = RSS_A - RSS_{A \cup j}. \tag{3.5}$$

If σ^2 is known, $R_j/\sigma^2 \sim \chi_1^2$, otherwise, $\sqrt{R_j}/s \sim t_{n-1}$. When R_j is great enough, I claim the variable is significant. The approach does not adjust for model selection.

The lasso procedure takes a different approach. The R_j is the jth largest sum of square rather than a random one so the χ_1^2 or t_{n-1} does not hold.

For example, when $A = \emptyset$, the predictor j provides the largest RSS difference. Then $R_j/\sigma^2 \sim \chi_1^2$ does not follow χ_1^2 or t_{n-1} distribution.

3.3 Covariance Test

Lockhart [12] introduced the covtest statistic (abbreviated as covtest). Make the following assumptions: (1) The linear model is correct; (2) The variance is constant; (3) The errors are normally distributed; (4) The true parameter vector is sparse; (5) The design matrix has weak collinearity.

The covariance test statistic is constructed based on the lasso fitted path and defined by:

$$T_k = (\langle y, X\hat{\beta}(\lambda_{k+1})\rangle - \langle y, X_A\tilde{\beta}_A(\lambda_{k+1})\rangle)/\sigma^2. \tag{3.6}$$

The $\hat{\beta}(\lambda_{k+1})$ is the lasso solution at λ_{k+1} while the $\tilde{\beta}_A(\lambda_{k+1})$ is the solution at λ_{k+1} with only the active predictors X_a, defined as the following:

$$\tilde{\beta}_A(\lambda_{k+1}) = \underset{\beta_A \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{2}||y - X_A\beta_A||_2^2 + \lambda_{k+1}||\beta_A||_1. \tag{3.7}$$

Then covariance statistic in (3.6) is a function of the difference between $X\hat{\beta}$ and $X_A\tilde{\beta}_A$, the fitted value by adding the jth variable into the active set and not. The (3.6) may be written as

$$T_k = (\langle y, X\hat{\beta}(\lambda_{k+1}) - X_A\tilde{\beta}_A(\lambda_{k+1}\rangle/\sigma^2. \tag{3.8}$$

This statistic is expected to be distributed approximately exponentially with mean 1 under the null hypothesis that the current lasso model contains all truly active variables [12].

Example: Let the coefficients $\beta_i \neq 0, \text{for} \quad i = 1, 2, 3, 4$ and $\beta_j = 0 \quad \text{for} \quad i = 5, 6...k$. Then the statistic

$$T_4 = (\langle y, X\hat{\beta}(\lambda_5)\rangle - \langle y, X_A\tilde{\beta}_A(\lambda_5)\rangle)/\sigma^2 \sim \text{Exp}(1). \tag{3.9}$$

The statistic is complicated and seems amazing to simply follow a basic exponential distribution. In fact, the statistic does not exactly follow Exp(1), but asymptotically follows it.

Figure 3.2 shows the CDFs of covtest statistic and exponential distribution with mean 1. They seem very close but there is still a gap between them. Figure 3.3 indicates the difference between the two CDFs.

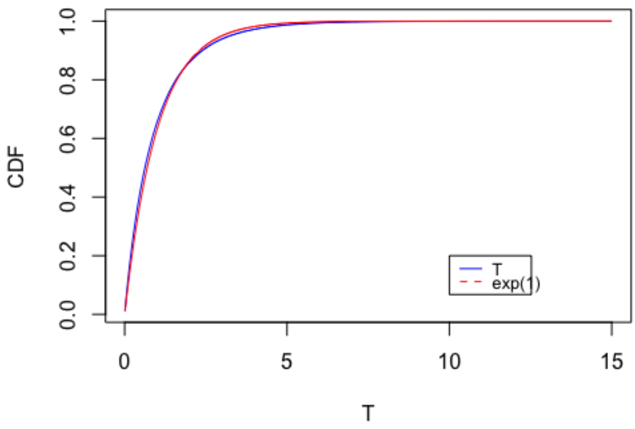

Figure 3.2: CDFs of Covtest Statisttic vs Exp(1)

The phenomenon leads to a series of questions. Why does the discrepancy appear, how accurate the test is? The following sections and chapter will provide a detailed analysis aimed at the properties of the covariance test. Specifically, the next sections start from the simplest case and reveal the inherent cause of covtest from the aspect of statistical theory. The next section talks about the moments and cumulants of the statistic and then find the asymptotic CDF. Furthermore, extend the test from the normal theory linear model to the general linear model with non-Gaussian error terms.

3.4 Distributional Features of the Covtest Statistics

In this section, I investigate the covtest statistic and find the relation between it and exponential distribution.

To get the solution of general lasso regression in (3.2), differentiate it and

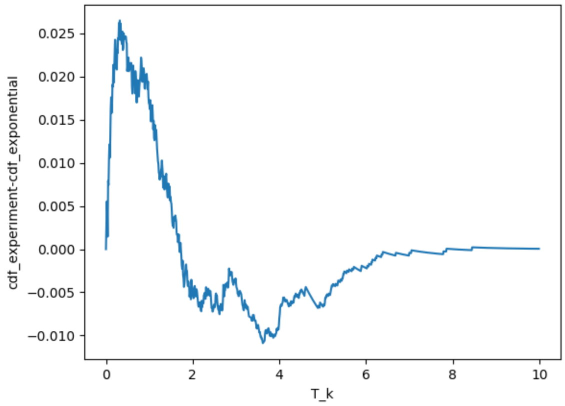

Figure 3.3: Difference between CDFs of Covtest Statisttic vs Exp(1)

obtain

$$-X^T(y - X\beta) + \lambda \text{sgn}(\beta) = 0, \tag{3.10}$$

where sgn is the sign function. Simplifying,

$$\hat{\beta}(\lambda) = (X^T X)^{-1} X^T y - \lambda (X^T X)^{-1} \text{sgn}(\beta). \tag{3.11}$$

By definition, the variables outside the active set A have zero coefficients, i.e. $\beta_j = 0$ if $j \notin A$. Hence

$$X\hat{\beta}(\lambda_k) = X_A \hat{\beta}_A(\lambda_k) = X_A \tilde{\beta}_A(\lambda_k), \tag{3.12}$$

and similarly for $A \cup \{j\}$,

$$X\hat{\beta}(\lambda_{k+1}) = X_{A\cup j}\hat{\beta}_{A\cup j}(\lambda_{k+1}). \tag{3.13}$$

When the active set is A or $A \cup \{j\}$ at the knot λ_{k+1},

$$\begin{aligned} X_A\tilde{\beta}_A(\lambda_{k+1}) &= X_A\hat{\beta}_A(\lambda_{k+1}) \\ &= X_A(X_A^T X_A)^{-1} X_A^T y - \lambda_{k+1} X_A (X_A^T X_A)^{-1} \text{sgn}(\hat{\beta}_A), \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} X\hat{\beta}(\lambda_{k+1}) &= X_{A\cup j}\hat{\beta}_{A\cup j}(\lambda_{k+1}) \\ &= X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1} X_{A\cup j}^T y - \lambda_{k+1}(X_{A\cup j}^T X_{A\cup j})^{-1}\text{sgn}(\beta_{A\cup j}). \end{aligned} \tag{3.15}$$

We substitute (3.14) and (3.15) into (3.8) to obtain,

$$\begin{aligned} T_k = y^T\big[X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1} X_{A\cup j}^T - X_A(X_A^T X_A)^{-1}X_A^T\big]y/\sigma^2 \\ -\lambda_{k+1} y^T\big[X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1}\text{sgn}(\beta_{A\cup j}) - X_A(X_A^T X_A)^{-1}\text{sgn}(\hat{\beta}_A)\big]/\sigma^2. \end{aligned} \tag{3.16}$$

At the knot λ_k, the j coefficient is zero, so obtain

$$X_A\hat{\beta}_A(\lambda_k) = X_{A\cup j}\hat{\beta}_{A\cup j}(\lambda_k). \tag{3.17}$$

By substitution,

$$\begin{aligned} X_A(X_A^T X_A)^{-1}X_A^T y - \lambda_k X_A(X_A^T X_A)^{-1}\text{sgn}(\hat{\beta}_A) = \\ X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1} X_{A\cup j}^T y - \lambda_k X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1}\text{sgn}(\beta_{A\cup j}). \end{aligned} \tag{3.18}$$

Simplifying,

$$\begin{aligned} y^T\big[X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1} X_{A\cup j}^T - X_A(X_A^T X_A)^{-1}X_A^T\big] = \\ \lambda_k\big[X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1}\text{sgn}(\beta_{A\cup j}) - X_A(X_A^T X_A)^{-1}\text{sgn}(\hat{\beta}_A)\big]. \end{aligned} \tag{3.19}$$

Using the identity of $X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1} X_{A\cup j}^T - X_A(X_A^T X_A)^{-1}X_A^T$, it is an orthogonal projection. Raking Euclidean norms of both sides in (3.19),

$$\begin{aligned} y^T\big[X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1} X_{A\cup j}^T - X_A(X_A^T X_A)^{-1}X_A^T\big]y = \\ \lambda_k^2 ||X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1}\text{sgn}(\hat{\beta}_{A\cup j}) - X_A(X_A^T X_A)^{-1}\text{sgn}(\hat{\beta}_A)||_2^2. \end{aligned} \tag{3.20}$$

Replacing (3.19) and (3.20) in (3.16), get

$$T_k = C(A, j, \text{sgn}(\hat{\beta}_A), \text{sgn}(\hat{\beta}_{A\cup j}))(\lambda_k^2 - \lambda_k\lambda_{k+1})/\sigma^2, \tag{3.21}$$

where

$$C(A, j, \text{sgn}(\hat{\beta}_A), \text{sgn}(\hat{\beta}_{A\cup j})) =$$

$$||X_{A\cup j}(X_{A\cup j}^T X_{A\cup j})^{-1}\text{sgn}(\hat{\beta}_{A\cup j}) - X_A(X_A^T X_A)^{-1}\text{sgn}(\hat{\beta}_A)||_2^2. \quad (3.22)$$

We use a specific example to uncover the distributional characteristics of covtest. Consider the simplest case in which the design matrix X is orthogonal, $\sigma^2 = 1$ and no variable is in the active set. It means $y = \epsilon \sim N(0, 1)$. Under the orthogonal condition, able to prove

$$C(A, j, \text{sgn}(\hat{\beta}_A), \text{sgn}(\hat{\beta}_{A\cup j})) = 1.$$

By the orthogonal property $X^T X = I$, obtain

$$||y - X\beta||_2^2 = ||X||_2^2 \cdot ||y - X\beta||_2^2 = ||X^T y - \beta||_2^2 + C, \quad (3.23)$$

for a constant C. Every column X_j corresponds to a coefficient β_j. To minimize (3.23), we have the close form solution for β_j

$$\hat{\beta}_j(\lambda) = \begin{cases} x - \lambda, & \text{if} \quad x > \lambda \\ 0, & \text{if} \quad -\lambda \le x \le \lambda \\ x + \lambda, & \text{if} \quad x < \lambda \end{cases}$$

The values of $X_j^T y, j = 1...p$ are the knots where the corresponding λ become nonzero. Consider X orthogonal matrix and $y \sim N(0, 1)$, so $X_j^T y \sim N(0, 1)$. We get a sequence of variables $U_j = X_j^T y \sim N(0, 1), j = 1, ...p$, and the λ are

$$\lambda_1 = |U_{(1)}|,$$

$$\lambda_2 = |U_{(2)}|,$$

$$\vdots$$

$$\lambda_p = |U_{(p)}|.$$

where $U_{(1)}, ...U_{(p)}$ are order statistics such that $U_{(1)} \geq U_{(2)} \geq ... \geq U_{(p)}$.

Letting $V_j = |U_j|, j = 1, ...p$, it is straightforward that $V_j \sim \chi_1$ (χ_1 is square root of χ_1^2 distribution). We summarize the properties of the covtest statistic as the following statement: given the i.i.d variables $V_1...V_p \sim \chi_1$, in which the largest and second statistics are denoted as $V_{(1)}$ and $V_{(2)}$, then

$$T_1 = V_1(V_1 - V_2) \xrightarrow{d} \text{Exp}(1), \text{as} \quad p \to \infty. \tag{3.24}$$

The distribution of T_1 is closer to Exp(1) if p becomes greater. Thus, the error of covtest depends on the value of p.

3.5 Simulation Investigation of the Influence of p

The last section proposes that the values of p affect the error of covtest. We investigate the pattern via simulation based on a variety of p values. We restate that the p is the column number of design matrix X for lasso regression, and the sample size of $V_1...V_p$ when I sample χ_1 variables.

First, let the value $p = 100$. We get samples $v_1, v_2...v_{100}$ from χ_1 distribution and calculate the value $t = v_{(1)}(v_{(1)} - v_{(2)})$, where $v_{(1)}$ and $v_{(2)}$ are the largest two samples.We do the process 5000 times and obtain 5000 t values.

We summarize the 5000 t values as an empirical distribution function F_t and plot it in Figure 3.4. The blakc line is the distribution function of F_t and the blue line is exponential distribution with mean 1. They are very close and I cannot find obvious difference in the graph. We plot the difference $F_t(x) - F_{exp}(x)$ versus x in Figure 3.5. We find $F_t(x) > F_{exp}(x)$ in the interval

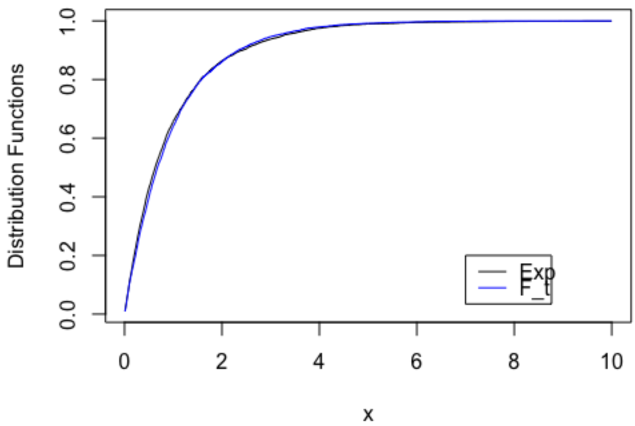

Figure 3.4: Distribution functions of F_t and Exp(1)

$[0, 2]$ and $F_t(x) < F_{exp}(x)$ when $x > 2$. Eventually, the difference between them converge to 0 when x is large enough.

We repeat the above process when $p = 20$, and get Figures 3.6 and 3.7. The empirical distribution function does not fit as well as before. The difference in Figure 3.5 is apparently smaller than that in Figure 3.7. The result coincides (3.24), greater p, better fit.

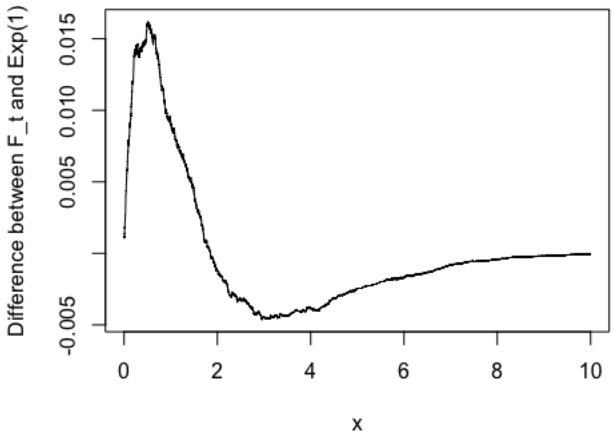

Figure 3.5: Difference between F_t and Exp(1)

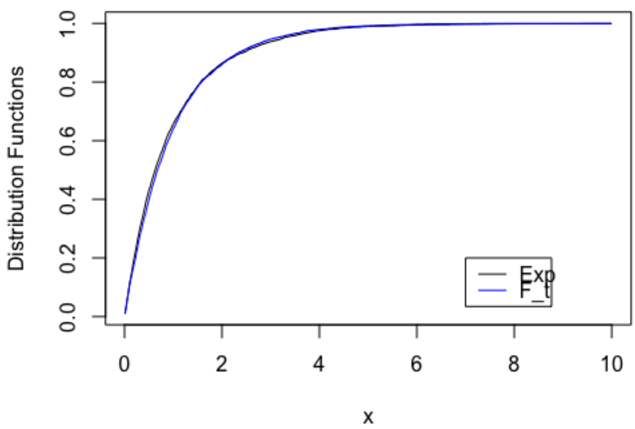

Figure 3.6: Distribution functions of F_t and Exp(1)

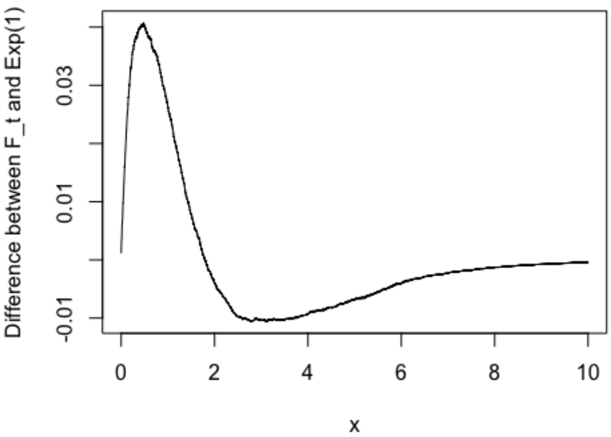

Figure 3.7: Difference between F_t and Exp(1)

Chapter 4

Inference of Covtest Statistic

In this chapter, new results on using the covtest for inference are provided. The moments and cumulants are found and discussed in the first section. The second section introduces the Edgeworth series approximation to covtest and its defects. In the third section, the gamma distribution is discussed as a better approximation than the exponential one for the covtest. The fourth section investigates the covtest without normality assumption.

4.1 Moments of Covtest Statistic

In this section, find the moments of covtest statistic.

$$\mathbf{E}(T_1) = \mathbf{E}\big[V_{(1)}(V_{(1)} - V_{(2)})\big]. \tag{4.1}$$

Consider the density function

$$\phi(x) = \frac{\exp(-\frac{x^2}{2})}{\sqrt{2\pi}}.$$

The distribution and density funcitons of $V_{(1)}$ are

$$F_{V_{(1)}}(x) = (2\Phi(x) - 1)^n,$$

and

$$f_{V_{(1)}}(x) = 2n\phi(x)(2\Phi(x) - 1)^{n-1}.$$

The joint distribution function and density function of $V_{(1)}$ and $V_{(2)}$ are

$$F_{V_{(1)},V_{(2)}}(x, y) = n(2\Phi(x) - 1)(2\Phi(y) - 1)^{n-1}, \tag{4.2}$$

and

$$f_{V_{(1)},V_{(2)}}(x, y) = 4(n-1)n\phi(x)\phi(y)(2\Phi(y) - 1)^{n-2}. \tag{4.3}$$

4.1.1 First moment of T_1

With the density function, the first moment of T_1 is

$$\begin{aligned} E(T_1) &= \int_0^\infty \int_y^\infty x(x-y)4(n-1)n\phi(x)\phi(y)(2\Phi(y)-1)^{p-2}dxdy \\ &= \int_0^\infty 4p(p-1)\phi(y)(1-\Phi(y))(2\Phi(y)-1)^{p-2}dy \\ &=1. \end{aligned} \tag{4.4}$$

In this case, the exact moments exactly equal the asymptotic limit.

4.1.2 Second moments

Furthermore, calculate the second moment by evaluating the integral

$$\begin{aligned} E(T_1^2) &= E[U_1^2(U_1 - U_2)^2] \\ &= (p-1)\int_0^\infty 4n\big(\phi(y)(y^2+3)(1-\Phi(y)) - y\phi(y)^2\big)(2\Phi(y)-1)^{p-2}dy, \end{aligned} \tag{4.5}$$

and find

$$E(T_1^2) = 2 + O_p(\exp(-\sqrt{\log(p)})).$$

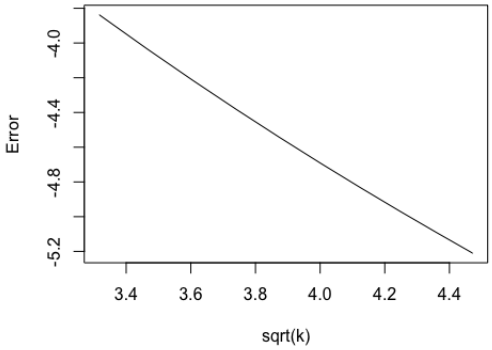

Figure 4.1: Log Difference between Second Moment $E(T^2)$ and 2 v.s. Squared Root of k for Normal Covtest

The simulation procedures are as follow:

Let $k = 10, 11, 12, ..., 70$, $p = a^k$, for some positive constant a.

We choose $a = 1.1$ to avoid exceeding the maximum positive value in computing.

Denote the logarithm of the error as $d(p) = E(T_1^2(p)) - 2$, where the expected values of covtest statistic and exponential statistic are $E(T_1^2(p))$ and 2. Figure 5.1 shows the strong linear relation by $\sqrt{k}$ in the horizontal axis and $\log(d)$ in the vertical axis. Thus

$$\log(d) = O_p(-\sqrt{k})$$

$$d = O_p(\exp(-\sqrt{k}))$$

$$d = O_p(\exp(-\sqrt{\log(p)})).$$

The linear relation is determined from a graph of numerically obtained values rather than the error bound calculated analytically.

Higher moments can be found by the same calculation approach.

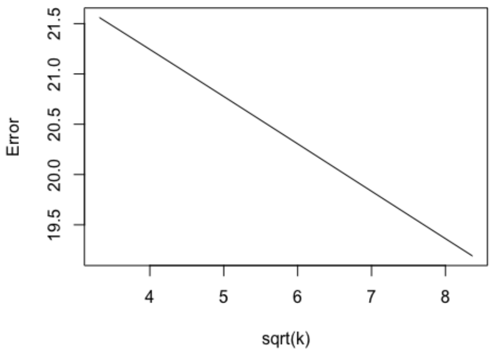

Figure 4.2: Log Difference between Third Moment $E(T^3)$ and Exp Dist. v.s. Squared Root of k for Normal Covtest

4.1.3 Higher Moments

The above method is used to find the third moment. Let $k = 10, 11, 12, ..., 70$, $p = 1.1^k$, and denote the logarithm of error as $d_3(p) = E(T_1^3(p)) - 6$. Figure 5.2 shows a strong linear relation and infer

$$d_3 = O_p(\exp(-\sqrt{\log(p)})).$$

More similar results could be found by the approach.

4.2 Edgeworth Expansions for Covtest

4.2.1 Intrduction to Edgeworth Series

The Edgeworth series are series that approximate a probability distribution in terms of its cumulants[11]. The following formal construction is due to [11] and [18]. Let X and Y be two random variables, such that Y has density f_Y and X has density f_X. Suppose further that X and Y can be constructed on a common probability space such that $Z = X - Y$ is independent of Y.

Conditionally on $Z = z$, X has density $f_Y(x - z)$, and expanding f_Y as a power series about x, $f_Y(x-z) = \sum_{j=0}^{\infty} f_Y^{(j)}(-z)^j/j!$. Hence the unconditional density of X is

$$f_X(x) = \sum_{j=0}^{\infty} f_Y^{(i)}(x)(-1)^j \mu_j^*/j!,$$

where μ_j^* are the moments of Z, again assuming that such a construction is possible. Writing

$$h_j(x) = (-1)^j f_Y^{(j)}(x)/f_Y(x),$$

Observe that

$$f(x) = f_Y(x) \sum_{j=0}^{\infty} h_j(x)\mu_j^*/j!. \tag{4.6}$$

The "moments" μ_j^* are the moments of whatever distribution is necessary to add to Y to get X. The cumulant of order j associated with these moments is the cumulant of order j associated with X minus the corresponding cumulant for Y. The functions h_i are ratios of the derivatives of the baseline density to the density itself. In the case a normal baseline, these are polynomials. After substituting in the expressions for pseudo-moments in terms of pseudo-cumulants, and collecting terms according to their power in n, the Edgeworth density and cumulative distribution functions are

$$\begin{aligned} f(x) =& \phi(x)\big[1 + k_3^n h_2(x)/6 + (k_4^n h_3(x)/24 + 10k_3^{n2} h_5(x)/720) \\ & +(k_5^n h_4(x)/120 + 35k_3^n k_4^n h_6(x)/5040 + 280k_3^n 3h_8(x/362880) + \cdots\big]. \end{aligned}$$

and

$$\begin{aligned} F(x) =& \Phi(x) - \phi(x)\big[k_3^n h_2(x)/6 + (k_4^n h_3(x)/24 + 10k_3^{n2} h_5(x)/720) \\ & +(k_5^n h_4(x)/120 + 35k_3^n k_4^n h_6(x)/5040 + 280k_3^n 3h_8(x/362880) + \cdots\big]. \end{aligned}$$

The Hermite polynomials and the issue are discussed in the next section.

4.2.2 Defect from the Edgeworth approximation

The Edgeworth approximation is usually better than a baseline approximation. For example, the Edgeworth approximation with the normal baseline function is generally better than the normal approximation. But the gamma distribution is not a good baseline function, so the Edgeworth approximation with the gamma is not feasible.

The following explains why it is not a good baseline function.

The Hermite polynomials under the gamma baseline function are investigated.

Denote the PDF of the gamma distribution as

$$f(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}. \tag{4.7}$$

where α and β are the shape and rate. The first derivative is

$$\frac{df(x)}{dx} = (\alpha-1)\frac{\beta^\alpha x^{\alpha-2}}{\Gamma(\alpha)} e^{-\beta x} - \beta\frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} = \left(\frac{\alpha-1}{x} - \beta\right) f(x). \tag{4.8}$$

The second derivative is

$$\begin{aligned} \frac{d^2 f(x)}{dx^2} = (\frac{\alpha-1}{x} - \beta)\frac{df(x)}{dx} - \frac{\alpha-1}{x^2} f(x) = \\ \left(\frac{\alpha^2-3\alpha+2}{x^2} - \frac{2(\alpha-1)\beta}{x} + \beta^2\right) f(x). \end{aligned} \tag{4.9}$$

The first two Hermite polynomials are

$$\begin{aligned} h_1(x) &= -1\frac{df(x)}{dx}/f(x) = -\frac{\alpha-1}{x} - \beta, \\ h_2(x) &= \frac{\alpha^2-3\alpha+2}{x^2} - \frac{2(\alpha-1)\beta}{x} + \beta^2. \end{aligned}$$

When $x \to 0$, both of the Hermite polynomials are $\to \infty$. That's why the gamma baseline function is not a good baseline function and the Edgeworth gamma approximation does not work.

Table 4.1: $E(T^2)$ with an increasing p the number

p	50	100	200	500	1000
$E(T^2)$	2.2374	2.2049	2.1802	2.1554	2.1407

4.3 An improved distribution approximation for the covtest

4.3.1 Motivations

The second moment of Exp(1) is 2. In Section 4.1, the second moment is always greater than 2. Table 4.1 shows the values of $E(T^2)$ when p increases. The value decrease with an increasing p. But it decreases more slowly with a greater p. The value changes 0.0325 when p changes from 50 to 100, but alters only 0.0147 when p increases from 500 to 1000. The speed of convergence $O(1/\log(p))$ is too slow.

The errors in the second moment are not trivial, and cannot be reduced significantly by increasing sample size p. This error is indicative of a failure of the exponential distribution to model the Covtest statistic distribution.

4.3.2 A Solution from the Gamma Distribution

To fix the mismatch between moments of the true and target distributions, approximate the target distribution by a family and a variety of mean and variance calculations. We pick the gamma distribution with the first moment 1 and the second moment $E(T^2)$, equal to that of the covtest statistic. The gamma distribution has two parameters, allowing adjustment of both the first and second moments, so to exactly match the first two moments of the target distribution. The gamma distribution has the shape

$$\alpha = \frac{E(T_1)^2}{E(T_1^2) - E(T_1)^2},$$

and rate

$$\beta = \frac{E(T_1)}{E(T_1^2) - E(T_1)^2},$$

which include the mean $E(T_1)$ and the second moment $E(T_1^2)$. We approximate $f_{T_1}(t, p)$ by

$$g(t) = \frac{t^{\alpha-1}\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta t}. \tag{4.10}$$

The $E(T_1)$ is exactly 1 under the normal error term. Substituting,

$$\alpha = \frac{1}{E(T_1^2) - 1},$$
$$\beta = \frac{1}{E(T_1^2) - 1}.$$

The Exp(1) is exactly $\Gamma(1, 1)$, a special case of the gamma distribution. But the exponential distribution has only one parameter while the gamma distribution has two. So an opportune member of the gamma family is likely to fit better than the exponential.

Figure 4.3 compares the distribution functions among the simulation cov-test statistic, the Exp(1), and the adjusted gamma distribution. The blue curve, representing the gamma distribution, is closer to the black curve, the simulated Covtest statistic distribution, than the blue Exp(1). Figure 4.4 compares the gamma approximation error and exponential approximation error. Figure 4.5 compares the relative gamma approximation error and relative exponential approximation error.

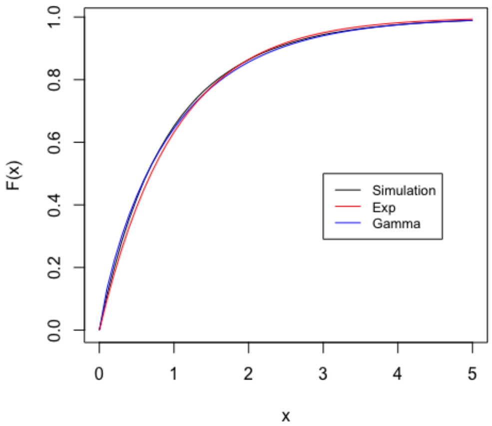

Figure 4.3: Comparison Among Simulation, Exponential Approximation and Gamma Approximation for Normal Covtest

The Kolmogorov-Smirnov test is used to decide the goodness of fit between the approximate and true distributions. The null hypothesis of K-S test is that the samples are drawn from the same distribution. The results are shown in Table 4.2. We choose the number of variables $p = 100, 500 and 1000$. The gamma K-S statistic is always smaller than the exponential K-S statistic, and the gamma p-value is always much larger than that of the exponential distribution. The results show the Gamma curve fits much better than the exponential curve. Please note that the variable p is the number of variables and p-value is for the tests. With the increasing number of variables p, the test p-value increases. It indicates the Gamma curve fits better when the number of variables increases.

The gamma distribution has the first and second moments matching the covtest statistic. But the higher moments are not matched.

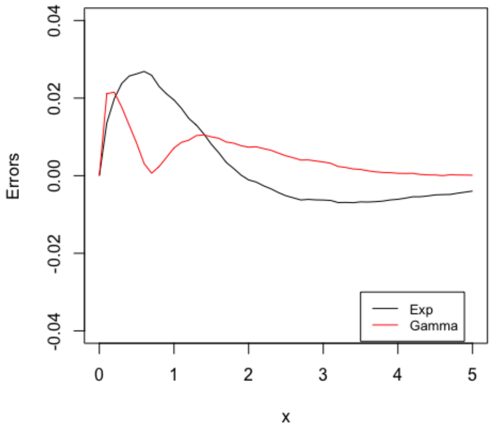

Figure 4.4: Comparison of Differences between Simulation and Exponential Approximation vesus Simulation and Gamma Approximation for Normal Covtest

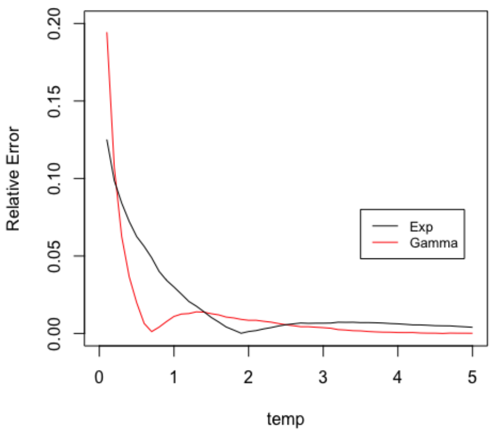

Figure 4.5: Comparison of Relative Errors between Simulation and Exponential Approximation vesus Simulation and Gamma Approximation for Normal Covtest

Table 4.2: K-S Testing for Covtest fitted by Exp. Approximation and fitted by Gamma Approximation

# of Variables p	Exp D	Exp p-value	Gamma D	Gamma p-value
100	0.073852	0.1301	0.035928	0.9029
500	0.073852	0.1301	0.027944	0.9897
1000	0.073852	0.1301	0.025948	0.9959

4.4 Covtest Statistic With Laplace Error Term

The Covtest is expected to test the significance of a single variable in regression. Existing results for this distribution require that ϵ must follow the normal distribution. We extend the method to a general situation, without the normality assumption. We first consider the error terms following the Laplace distribution.

Suppose the Laplace distribution with mean 0 and variance 1. Let $V_j = |U_j|$, so $V_1, ..., V_p \sim \text{Exp}(\sqrt{2})$. As before, denote the largest and second largest statistics by $V_{(1)}$ and $V_{(2)}$. Then

$$T_1 = V_{(1)}(V_{(1)} - V_{(2)}) \sim \text{Exp}(1), \text{as} \quad p \to \infty. \tag{4.11}$$

The exponential distribution approximation has the rate $1/E(T)$, meaning the expected value $E(T)$. The gamma distribution has the shape

$$\alpha = \frac{E(T)^2}{E(T^2) - E(T)^2},$$

and rate

$$\beta = \frac{E(T)}{E(T^2) - E(T)^2},$$

which corresponds to the mean $E(T)$ and the second moment $E(T^2)$.

Reconsider $V_1, ...V_p$. The largest $V_{(1)}$ and the second largest value $V_{(2)}$ have the joint distribution

$$f_{V_{(1)},V_{(2)}}(x,y) = (n-1)n\sqrt{2}\text{exp}(-\sqrt{2}x)\sqrt{2}\text{exp}(-\sqrt{2}y)(1-\text{exp}(-\sqrt{2}y))^{p-2}. \tag{4.12}$$

The first moment of T_1 is

$$\begin{aligned} E(T_1) =& E\big[V_{(1)}(V_{(1)} - V_{(2)})\big] \\ =& \int_0^\infty \int_y^\infty x(x-y)p(p-1)2\text{exp}(-\sqrt{2}(x+y)) \\ & \cdot (1-\text{exp}(-\sqrt{2}y))^{p-2}dxdy \\ =& \int_0^\infty p(p-1)\text{exp}(-2\sqrt{2}y)(\sqrt{2}+y)(1-\text{exe}(-\sqrt{2}y))^{p-2}dy. \end{aligned} \tag{4.13}$$

Continue simplifying (4.13) by substituting $u = 1-\text{exp}(-\sqrt{2}y)$ and $G(u) = g(y(u)) = 1 + y/\sqrt{2}$ to obtain

$$E(T_1) = p(p-1)\int_0^1 u^{p-1}(1-u)G(u)du. \tag{4.14}$$

It is too difficult to integrate the expression in (4.14) directly. We expand $G(u)$ in

$$\hat{u} = \frac{p}{p+2}.$$

and integrate the expressions termwise. The expansion of (4.14) is

$$\begin{aligned} E(T_1) = p(p-1)\big[G(\hat{u})\int_0^1 u^{p-1}(1-u)du + G'(\hat{u})\int_0^1 u^{p-1}(1-u)(u-\hat{u})du \\ + G''(u)\int_0^1 u^{p-1}(1-u)(u-\hat{u})^2du + \cdots\big]. \end{aligned} \tag{4.15}$$

Integrating by term, the first term in (4.15) is

$$\int_0^1 u^{p-1}(1-u)du = \frac{\Gamma(p)\Gamma(2)}{\Gamma(p+2)} = \frac{1}{p(p+1)},$$

the second term in (4.15) is

$$\int_0^1 u^{p-1}(1-u)(u-\hat{u})du = \int_0^1 u^p(1-u)du - \hat{u}\int_0^1 u^{p-1}(1-u)du$$
$$= \frac{\Gamma(p+1)\Gamma(2)}{\Gamma(p+3)} - \frac{p}{p+2}\frac{\Gamma(p)\Gamma(2)}{\Gamma(p+2)} = 0,$$

and the third term (4.15) is

$$\begin{aligned}\int_0^1 u^{p-1}(1-u)(u-\hat{u})^2 du &= \int_0^1 u^{p+1}(1-u) - 2\hat{u}u^p(1-u) + \hat{u}^2 u^{p-1}(1-u)du \\ &= \frac{1}{(p+3)(p+2)} - 2\frac{p}{p+2}\frac{1}{(p+2)(p+1)} \\ &\quad + (\frac{p}{p+2})^2\frac{1}{p(p+1)} \\ &= \frac{2}{(p+3)(p+2)^2(p+1)}.\end{aligned}$$

Simplifying (4.15) and obtaining

$$\begin{aligned}E(T_1) =& p(p-1)\big[\frac{G(\hat{u})}{p(p+1)} + 0 + \frac{G''(\hat{u})}{(p+2)^2(p+1)(p-1)} + o(1/p^4)\big] \\ =& \frac{p-1}{p+1}G(\hat{u}) + O(1/p^2) \\ =& G(\hat{u}) - \frac{2}{p+1}G(\hat{u}) + O(1/p^2) \\ =& 1 - \frac{1}{2}\log 2 + \frac{1}{2}\log(p+2) + O_p(1/p).\end{aligned}$$

Mathematica says that $E(T_1)$ can be given in terms of the harmonic numbers, which are approximate $\log(p)$.

The result is confirmed by numerical integration. Calculate (4.13) numerically, and plot the relation between $\log(p)$ and $E(T_1)$ in Figure 4.6. When p increases exponentially, $E(T)$ will increase linearly. It confirms

$$E(T_1) = O_p(\log(p)). \tag{4.16}$$

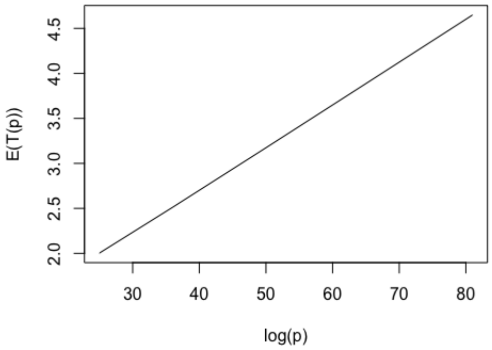

Figure 4.6: Numerical Calculation for Equation (4.16)

The second moment is

$$\begin{aligned} E(T_1^2) =& E\big[V_{(1)}^2(V_{(1)} - V_{(2)})^2\big] \\ =& \int_0^\infty \int_y^\infty x^2(x-y)^2 p(p-1)2\exp(-\sqrt{2}(x+y)) \\ & \cdot (1-\exp(-\sqrt{2}y))^{p-2} dxdy. \end{aligned} \tag{4.17}$$

It is too difficult to simplify and calculate (4.17) analytically. Numerical calculation is adopted to calculate the value $E(T_1^2)$ alone. The $E(T_1^2)$ is calculated numerically by (4.17) under different number of variables p. We find: When $\log(p)$ increases exponentially, $E(T_1^2)$ does not increase linearly, as shown in Figure 4.7 where the horizontal axis is $\log(p)$ and the vertical axis is $E(T_1^2)$. But $\sqrt{E(T_1^2)}$ will increase linearly when $\log(p)$ increase exponentially as shown in Figure 4.8. The graphic facts imply

$$E(T_1^2) = O_p(\log^2(p)). \tag{4.18}$$

Please note that the above results are obtained numerical integration and approximately linearity.

Our purpose is to apply covtest to Laplace distribution. First, the above values $E(T_1)$ and $E(T_1^2)$ from the integrals can produce a Gamma distribution.

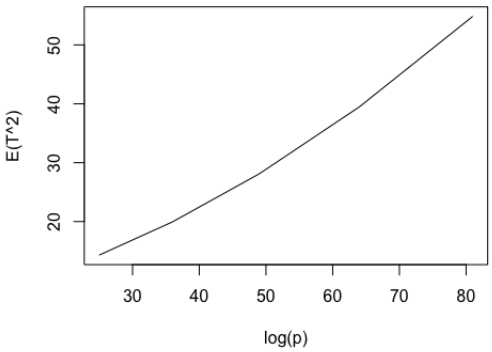

Figure 4.7: Simulation for Equation (4.18)

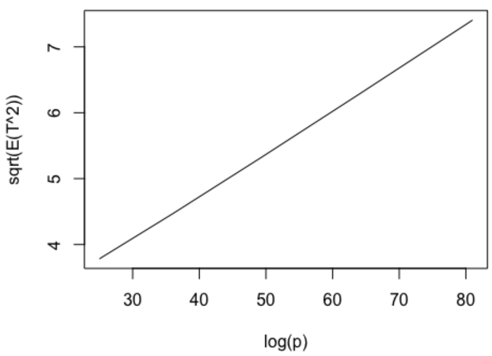

Figure 4.8: Simulation for Equation (4.18)

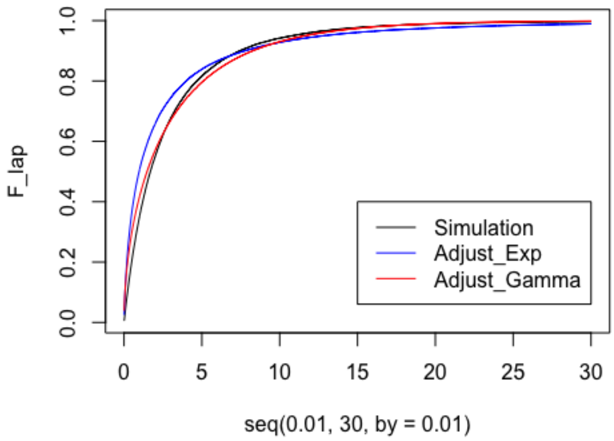

Figure 4.9: Comparison between Gamma Approximation, Exponential Approximation and Empirical Distribution under Laplace Error Terms

Second, I can also get the empirical distribution: sampling p random values from Exp($\sqrt{2}$), finding the two largest values, calculating the value T_1 numerically by (4.13), and repeating the above procedures. The empirical distribution of T_1 is obtained from the samples. The sampling and integration are based on $p = 1024$.

Figure 4.9 shows the comparison of the distribution functions among the empirical distribution(black line), the adjusted exponential distribution(blue line) with mean $E(T_1)$ and the adjusted gamma approximation(red line) with mean $E(T_1)$ and second moment $E(T_1^2)$. It shows that the adjusted gamma approximation fits well, and better than the adjusted exponential approximation.

4.5 Other Exponential Family Distributions

Consider the following density function

$$f(x) = k\exp(-\frac{x^p}{p}),$$

where k is a constant and p is a positive number, $x \in [0, \infty)$. The Covtest may apply to all distributions like this. The following is an example for $p = 3$. The density function is

$$f(x) = \frac{3^{2/3}}{\Gamma(1/3)} \exp(-x^3/3). \tag{4.19}$$

We obtain n samples from the distribution and find t by $T = U_1(U_1 - U_2)$ where U_1, U_2 are the greastest and second greatest samples. The distribution of T is asymptotically close to Exp(1/mean(T)) and the adjusted Gamma approximation(defined in 4.3.2). The following simulation figures demonstrate the conclusion. Figure 4.10 shows the CDF of (4.19), the adjusted exponential distribution and adjusted gamma distribution. They are very close. Figure 4.11 shows the difference between the adjusted exponential CDF and empirical distribution (black line), as well as the difference between the adjusted gamma CDF and empirical one(red line). The red line is closer to the X-axis than the black. It indicated that the gamma approximation fits better.

The general density function may be written as

$$f(x) = \exp(x^p/p)/(p^{\frac{1}{p}}\mathrm{Gamma}(1 + \frac{1}{p})). \tag{4.20}$$

We infer that any PDF function in the form of (4.20) will be applied to the covtest. Concretely, If the response variable in regression has an error term such that PDF follows (4.20), the covtest is a valid significance test for the regression.

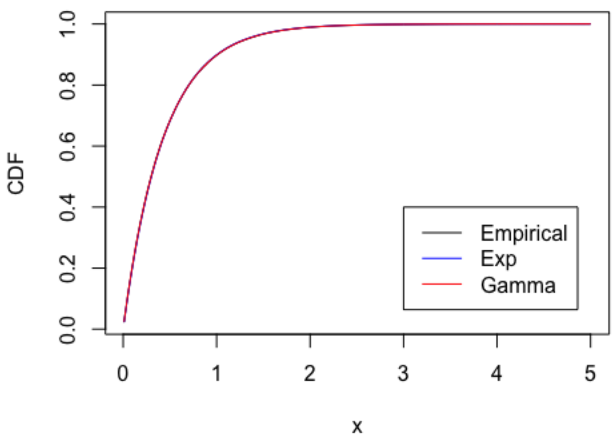

Figure 4.10: Comparison between Gamma Approximation, Exponential Approximation and Empirical Distribution under PDF (4.19)

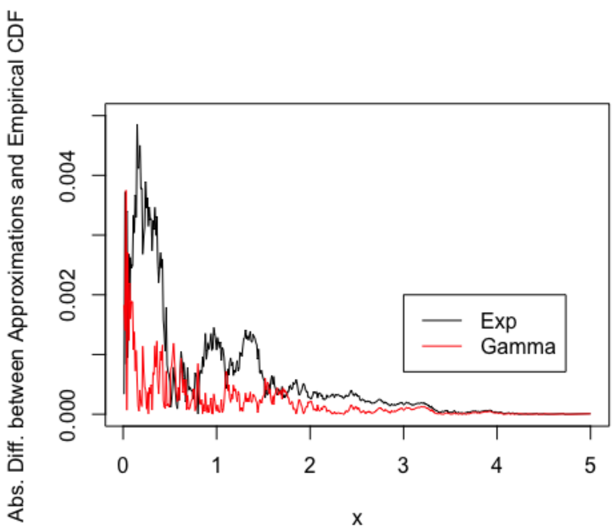

Figure 4.11: Absolute CDF Differences Between Gamma Approximation and Empirical Distribution, versus Exponential Approximation and Empirical Distribution under PDF (4.19)

4.6 Discussion

The covariance test is a significance test based on the normal error assumption and has an approximately exponential distribution. I improve both the error assumption and the approximate distribution. The normal error is extended to the error in the exponential distribution family. The gamma distribution is proposed as a better approximation than the exponential. Thanks to this progress, the covtest can be widely used in the significance testing of different kinds of linear regressions.

Chapter 5

Numerical Approach for Finding Error of Order

It is often not easy to find the order of error when approximating complex functions or statistics. The expansion or factorization of complicated functions is difficult, so the analysis is difficult. This numerical approach is utilized several times to find the order of error in the above chapters. I summarize the approach in this chapter.

5.1 Introduction by an Example

In the section, a simple example to find the order of error, which is very easy to understand, is provided.

Consider the function

$$f(n) = \frac{1}{1 - 1/n}. \tag{5.1}$$

Let

$$g(n) = 1 + \frac{1}{n}. \tag{5.2}$$

This chapter explores graphical heuristic techniques that indicate the order of error, when the $g(x)$ is used to approximate $f(x)$.

Let $n_k = 2^k, k = 1, 2, 3, \cdots, 20$, and then the error term is expressed as

$$d(n_k) = f(n_k) - g(n_k), k = 1, 2, \cdots, 20.$$

We choose to approximate the logarithmic error,

$$\log_2 d(n_k) = \log_2\big(f(n_k) - g(n_k)\big), k = 1, 2, \cdots, 20.$$

Figure 5.1 contains a plot of k versus $\log_2 d(n_k)$. The graph shows the obvious linearity. Noting that the slope is -2, we infer that the order of error is $O(1/n^2)$.

We justify this assessment using the following argument, under the assumption

$$d(n) = O(\frac{1}{n^a}),$$

where a is unknown. So we have

$$d(n) = \frac{C}{n^a} + o(\frac{1}{n^a}).$$

where C is a constant. Then we find

$$\begin{aligned} \log_2(d(n_k)) =& \log(C) - a\log_2(n) \\ =& \log(C) - ak. \end{aligned} \tag{5.3}$$

The linear equation (5.3) of k corresponds to Figure 5.1, so $-a$ corresponds to the slope -2, resulting in

$$a = 2.$$

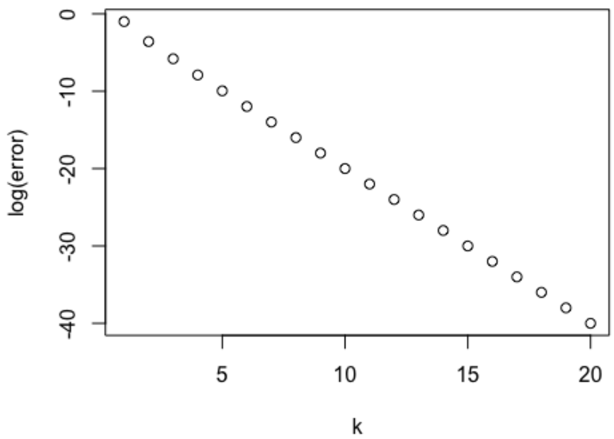

Figure 5.1: log(error) versus k

When we execute the method, n is supposed to be an exponent expressed as something to the kth power like 2^k. We choose 2 in the example so we find $\log_2(d)$ in the last step. The natural exponent can be chosen instead, then the ordinate of the plot is the natural logarithm. Because of the limit of 64-bit floating-point numbers in most computers and programming languages, we recommend a smaller base like $2, 1.5$ or $\sqrt{2}$.

5.2 Asymptotic Same Order

Section 5.1 presents an ideal instance to illustrate this method. But it is often interfered in practical application. The following two functions are $O(1/n)$, but they are not the exact expression $\frac{1}{n}$.

$$\begin{aligned} f(n) &= \frac{1}{n-4}, \\ g(n) &= \frac{1}{n} + \frac{1}{n^2}. \end{aligned} \tag{5.4}$$

We summarize the pattern as

$$O(\frac{1}{n^a}) = \frac{C}{n^a + \Omega(n^a)} + o(\frac{1}{n^a}), \tag{5.5}$$

which is easy to prove.

5.3 The Method Applied to Covtest Moments

The problem of covtest moments is used as an example to illustrate the approach.

5.3.1 The Second Moment

The second moment of the covtest statistic is

$$E(T_1^2) = E(U_1^2(U_1 - U_2)^2), \tag{5.6}$$

and it is a function of sample size n.

The error term is defined as

$$d(n) = E(T_1^2) - 2. \tag{5.7}$$

The method described in Section 5.2 is utilized. We denote

$$n_k = 2^k, k = 1, 2, 3, \cdots, 20. \tag{5.8}$$

The logarithmic errors are plotted in Figure 5.2. It is apparently not a linear function.

Similar results may be obtained using

$$n_k = 2^{\frac{k}{2}} = (\sqrt{2})^k, \text{for} \quad k = 1, 2, 3, \cdots, 20, \tag{5.9}$$

and employing the same techniques, plotting k versus $\log_{\sqrt{2}}(d(n_k))$ in Figure 5.3. It seems no significant linearity in the graph. Perhaps we need to revisit the assumption that the error is of this form of $O(\frac{1}{n^a})$ and consider a is not a positive integer.

When $a = 1/2$, the plot of $\sqrt{k}$ versus the logarithmic errors, it shows linearity to some extent in Figure 5.4. The black line represents the relationship between sample size and error, and the blue line connects the head and tail of the black line by a straight line to show the linearity. In order to confirm the linearity, we change $k = 1, 2, ..., 30$ to $k = 1, 2, ..., 200$, and adjust the logarithmic base to 1.1 to avoid 64-bit numeral limit in the computer. Then we get Figure 5.5. They overlap in the tail and the linearity is indicated. Then we have

$$\log(d(1.1^k)) = -a\sqrt{k} + c,$$

$$d = O(\exp(\sqrt{\log(n)})).$$

5.3.2 The Third Moment

The third moment of the covtest statistic is

$$E(T_1^3) = E(U_1^3(U_1 - U_2)^3) \tag{5.10}$$

The error is defined as

$$d_3 = E(T_1^3) - 6, \tag{5.11}$$

where 6 is the third moment of Exp(1). We do the same process and plot $k^{1/2}$ versus $\log_{1.1}(d_3(n_k))$ in Figure 5.6. The result shows

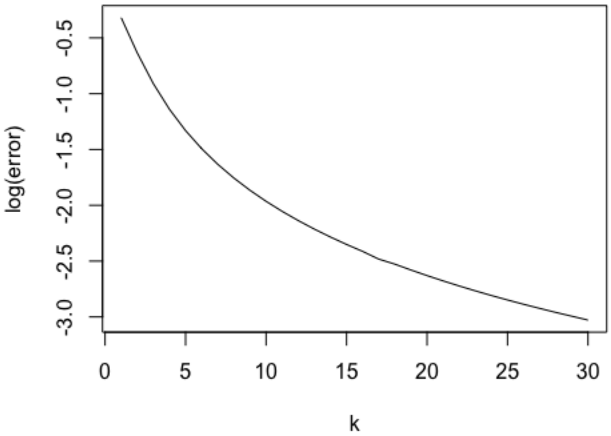

Figure 5.2: k versus log(error)

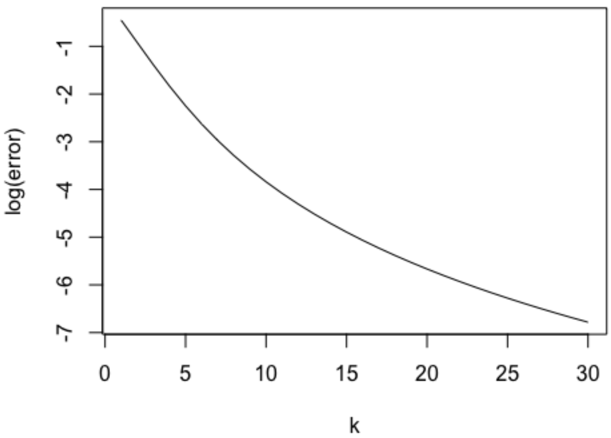

Figure 5.3: k versus log(error)

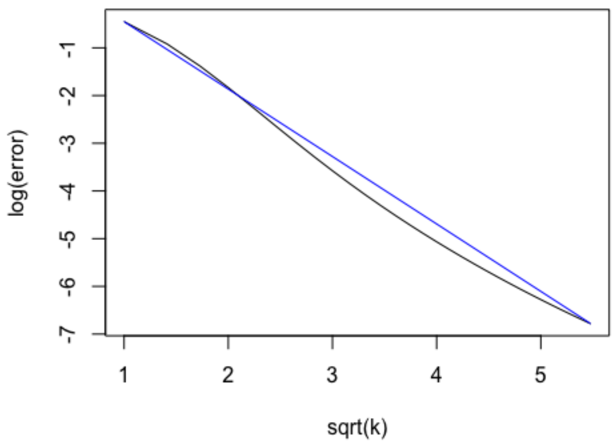

Figure 5.4: $\sqrt{k}$ versus log(error)

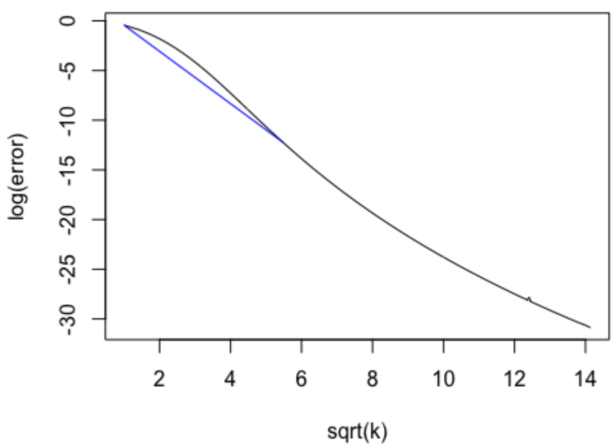

Figure 5.5: $\sqrt{k}$ versus log(error)

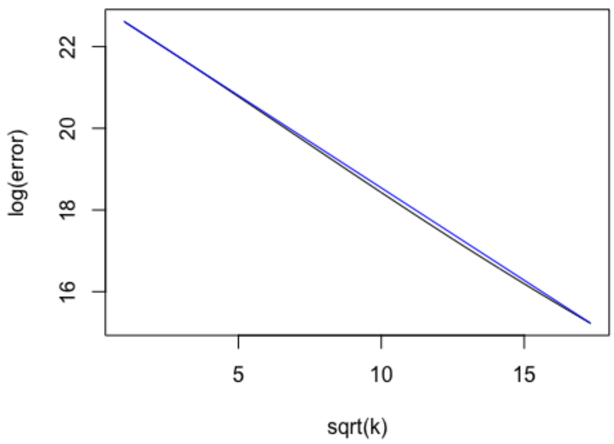

Figure 5.6: $\sqrt{k}$ versus log(error)

$$d_3 = O(\exp(\sqrt{\log(n)})).$$

5.4 A Useful Investigation on $O(\frac{1}{\Phi(1-\frac{1}{n})})$

We obtain the result $O(1/\Phi(1-1/n)) = O(1/\sqrt{\log n})$ by simulation rather than theoretical deduction by the above approach.

Let $n = 2^m$ for $m = 11, 12...70$, calculate

$$X = \sqrt{m}, \tag{5.12}$$

$$Y = \Phi^{-1}(1 - \frac{1}{2^m}). \tag{5.13}$$

and then plot X versus Y in Figure 5.7. The linearity is apparent. Thus we find

$$\Phi^{-1}(1 - \frac{1}{2^m}) = c_1\sqrt{m} + c_2, \tag{5.14}$$

where c_1 and c_2 are constants. We substitute m with n and get

$$\Phi^{-1}(1 - \frac{1}{n}) = c_1\sqrt{\log n} + c_2, \tag{5.15}$$

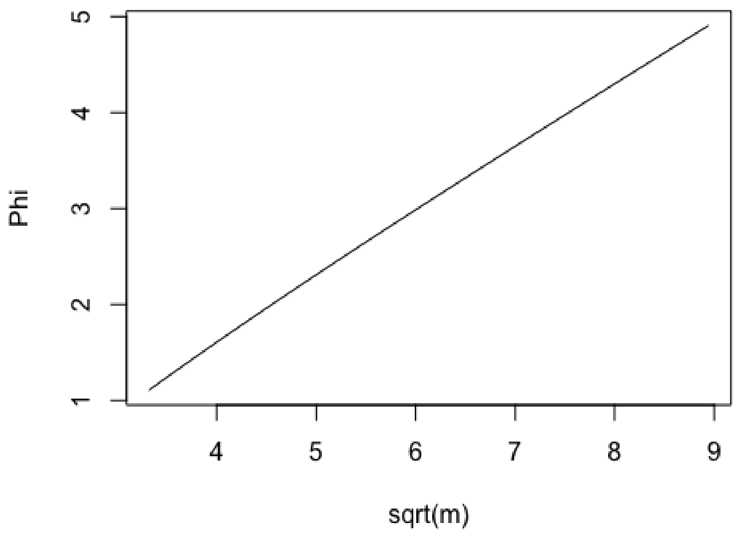

Figure 5.7: $\sqrt{k}$ vs $\Phi^{-1}(1 - \frac{1}{2^m})$

which means

$$O(\frac{1}{\Phi^{-1}(1 - \frac{1}{n})}) = O(\frac{1}{\sqrt{\log n}}). \tag{5.16}$$

5.5 Discussion

The approach, using the numeric method to find the error of order, is a heuristic and exploratory method. Many papers discuss numerical methods by the Euler method or ODE. But they cannot solve the problem with exponential or logarithmic order of accuracy.

My approach focuses on the issue of accuracy with the exponential or logarithmic order. Although it is exploratory and lacks a set of systematic procedures, it provides a creative approach to solve the hard issue, finding the order of accuracy.

Bibliography

[1] Shan, G., Zhang, H. Two-stage optimal designs with survival endpoint when the follow-up time is restricted. *BMC Med Res Methodol* 19, 74 (2019).

[2] H. B. Mann and D. R. Whitney. On a test whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 1850-60, 1947

[3] Castagliola P. (1996) Optimized Algorithms for Computing Wilcoxon's Tn, Wilcoxon's Wm,n and Ansari-Bradley's Am,n Statistics when m and n are Small. *Journal of Applied Statistics 23(1):41-58*

[4] Spurrier J., and Hewett J., Two-Stage Wilcoxon Tests of Hypotheses. *Journal of the American Statistical Association* Vol. 71, No. 356 (Dec., 1976), pp. 982-987

[5] Wilding G., Shan G., Hutson A., Exact two-stage designs for phase ii activity trials with rank-based endpoints. *Contemporary Clinical Trials 33(2012)332-341*

[6] Kolassa J., A comparison of size and power calculations for the Wilcoxon Statistic for ordered categorical data. *Statistics in Medicine.* VOL. 14, 1577-1581, 1995.

[7] O'Brien P., Fleming T., A Multiple Testing Procedure for Clinical Trials. *Biometrics* Vol. 35, No. 3, (Sep., 1979), pp. 549-556

[8] Feuerverger, A., Robinson, J., Wong, A. (1999) On the relative accuracy of certain bootstrap procedures,*The Canadian Journal of Statistics*, Vol. 27, No. 2, Jun., 1999

[9] Kolassa, J., Robinson, J. (2011) Saddlepoint approximation for likelihood ratio like statistics with applications to permutation tests,*The Annals of Statistics*, Vol. 39, No. 6, 3357-3368

[10] Kolassa, J., Chen, X. Power Calculations and Critical Values for Two-Stage Nonparametric Testing Regimes. Preparing

[11] Robinson, J. (2004) Multivariate tests based on empirical saddlepoint approximations, *International Journal of Statistics*, 2004, vol. LXII,n.1, pp. 1-14

[12] R. Lockhart, J. Taylor, R. J. Tibshirani (2014), and R. Tibshirani. A significance test for the lasso. *Ann. Statist.* Volume 42, Number 2 (2014), 413-468.

[13] H. L. Seal (1967) , The historical development of the Gauss linear model *Biometrika.* 54 (1/2): 124.

[14] R. Tibshirani (1996), Regression Shrinkage and Selection via the lasso *Journal of the Royal Statistical Society.* Series B (methodological). Wiley. 58 (1): 26788.

[15] G. Grimmett and S. Stirzaker (2001), *Probability Theory and Random Processes.* Cambridge. p. 98. ISBN 0-19-857223-9.

[16] J. Kolassa (2006), *Series Approximation Methods in Statistics.* Springer ISBN 978-0-387-94277-3.

[17] Regularization Paths for Generalized Linear Models via Coordinate Descent *Journal of Statistical Software.* 2010;33(1):1-22.

[18] AW Davis (1976),Statistical distributions in univariate and multivariate Edgeworth populations. *Biometrika* Volume 63, Issue 3

[19] Wasserman L. (2014) Discussion: A significance test for the lasso *Ann. Statist.* Volume 42, Number 2 (2014), 501-508.

[20] Jason D. Lee, Dennis L. Sun, Yuekai Sun, Jonathan E. Taylor *Exact post-selection inference, with application to the lasso*, Annals of Statistics 2016, Vol. 44, No. 3, 907-927