MORE ACCURATE TWO SAMPLE COMPARISONS FOR SKEWED POPULATIONS

by

BO TONG

M.S., Kansas State University of Kansas, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the

requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2016

Abstract

Various tests have been created to compare the means of two populations in many scenarios and applications. The two-sample t-test, Wilcoxon Rank-Sum Test and bootstrap-t test are commonly used methods. However, methods for skewed two-sample data set are not well studied. In this dissertation, several existing two sample tests were evaluated and four new tests were proposed to improve the test accuracy under moderate sample size and high population skewness.

The proposed work starts with derivation of a first order Edgeworth expansion for the test statistic of the two sample t-test. Using this result, new two-sample tests based on Cornish Fisher expansion (TCF tests) were created for both cases of common variance and unequal variances. These tests can account for population skewness and give more accurate test results. We also developed three new tests based on three transformations (T_i test, $i = 1, 2, 3$) for the pooled case, which can be used to eliminate the skewness of the studentized statistic.

In this dissertation, some theoretical properties of the newly proposed tests are presented. In particular, we derived the order of type I error rate accuracy of the pooled two-sample t-test based on normal approximation (TN test), the TCF and T_i tests. We proved that these tests give the same theoretical type I error rate under skewness. In addition, we derived the power function of the TCF and TN tests as a function of the population parameters. We also provided the detailed conditions under which the theoretical power of the two-sample TCF test is higher than the two-sample TN test. Results from extensive simulation studies and real data analysis were also presented in this dissertation. The empirical results further confirm our theoretical results. Comparing with commonly used two-sample parametric and nonparametric tests, our new tests (TCF and T_i) provide the same empirical type I error rate but higher power.

MORE ACCURATE TWO SAMPLE COMPARISONS FOR SKEWED POPULATIONS

by

BO TONG

M.S., Kansas State University of Kansas, 2012

A DISSERTATION

submitted in partial fulfillment of the

requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2016

Approved by:

Major Professor

Haiyan Wang

Abstract

Various tests have been created to compare the means of two populations in many scenarios and applications. The two-sample t-test, Wilcoxon Rank-Sum Test and bootstrap-t test are commonly used methods. However, methods for skewed two-sample data set are not well studied. In this dissertation, several existing two sample tests were evaluated and four new tests were proposed to improve the test accuracy under moderate sample size and high population skewness.

The proposed work starts with derivation of a first order Edgeworth expansion for the test statistic of the two sample t-test. Using this result, new two-sample tests based on Cornish Fisher expansion (TCF tests) were created for both cases of common variance and unequal variances. These tests can account for population skewness and give more accurate test results. We also developed three new tests based on three transformations (T_i test, $i = 1, 2, 3$) for the pooled case, which can be used to eliminate the skewness of the studentized statistic.

In this dissertation, some theoretical properties of the newly proposed tests are presented. In particular, we derived the order of type I error rate accuracy of the pooled two-sample t-test based on normal approximation (TN test), the TCF and T_i tests. We proved that these tests give the same theoretical type I error rate under skewness. In addition, we derived the power function of the TCF and TN tests as a function of the population parameters. We also provided the detailed conditions under which the theoretical power of the two-sample TCF test is higher than the two-sample TN test. Results from extensive simulation studies and real data analysis were also presented in this dissertation. The empirical results further confirm our theoretical results. Comparing with commonly used two-sample parametric and nonparametric tests, our new tests (TCF and T_i) provide the same empirical type I error rate but higher power.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many scenarios and applications, people are interested in doing hypothesis testing concerning differences between means of two populations. For example, based on proper samples we may wish to decide whether medicine A can be as effective as medicine B, or based on a survey we may want to decide whether the average weekly income of families in one city exceeds that in another city and etc. These comparisons of two independent population means are very common in many research fields.

Various tests have been created to do the comparisons under different scenarios. The two-sample T test is the most commonly used approach. It is a test to check the equality of means, and derived under the assumption that the two populations follow normal distribution. In many research areas, the assumption of normality is often violated. Skewed data are common. For example, several well-known variables are known to be markedly skewed, such as the survival time of a product following Weibull distribution; the pharmacokinetics parameters often following log-normal distribution; the bacterial growth rate following exponential distribution and etc. When the normality assumption does not hold, the nonparametric test such as Wilcoxon-Mann-Whitney test or the test based on resampling method such as bootstrap-t test can be used. These tests, however, were not developed specially for skewed data. There are situations that comparison of means or total in skewed samples is of interest. One example is about profit in farm animals, such as cattle. The weight gain of some animals may be heavily right skewed due to some diet additives. To compare the profit based on weight gain of animals with the additives vs. those without the additives, it is necessary to compare the mean of possibly skewed populations. Figure 1.1 shows a real two-sample data set from exercise 6.17 of the textbook Ott and Longnecker (2008). The data from both control and treatment groups show a skewed population. To make the question more challenging, the sample size is not very large. It can be clearly seen from the boxplot that the medians differs by quite a bit. However, neither the two-sample t-test nor the Wilcoxon Rank-Sum Test can give significant result.

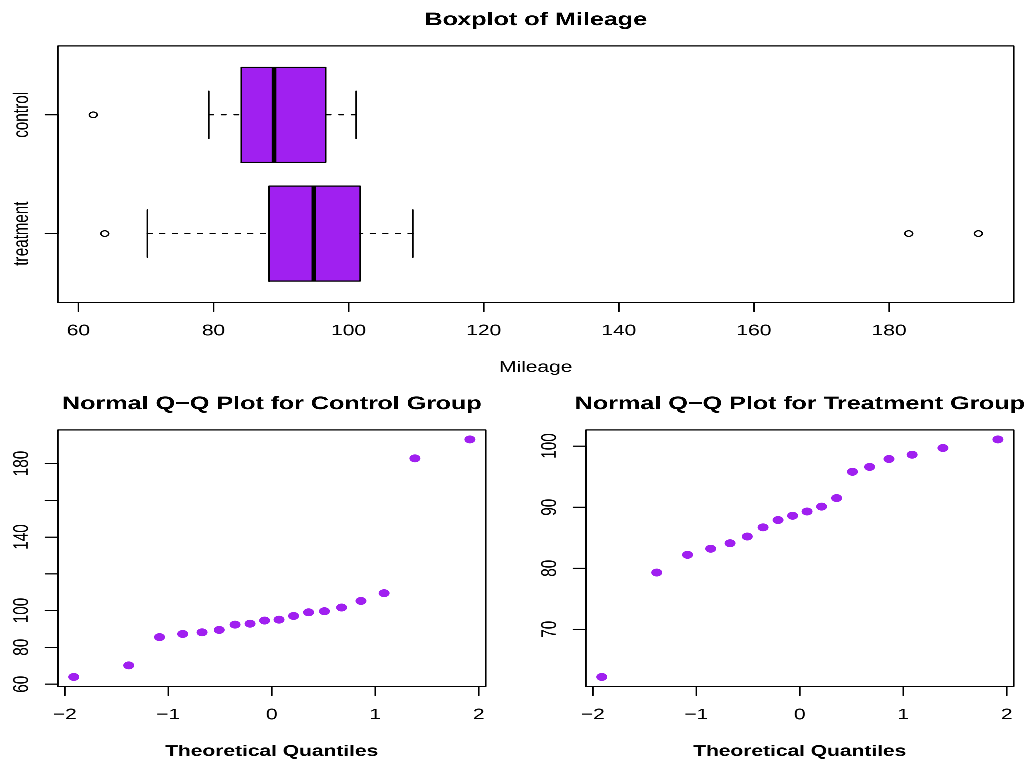

Figure 1.1: *Descriptive Statistics*

In this dissertation, we will evaluate several existing two sample tests and propose new tests that improve the test accuracy under moderate sample size and high population skewness. Some of our proposed new tests did find significant difference for this example. The organization of the dissertation is as follows: Chapter 2 reviews several existing two sample tests, which include ordinary two sample t-test, some nonparametric tests and tests based on resampling method. Since the methodology development utilizes Edgeworth expansion, we will also review the principle of Edgeworth expansion in Chapter 2. Chapter 3 and Chapter 4 introduce the methodology to derive four new test based on Edgeworth expansion theory. Chapter 4 gives three new tests obtained with transformations. Simulation study are presented for comparing the type I error rate and power of several existing tests with the new tests. We expand this research and apply the results to do the two-sample comparison for skewed populations with unequal variance in Chapter 5. Appendix A provides the proofs of several theorems in Chapter 3 and Chapter 4.

Chapter 2

Literature review

Suppose $Y_{1,1}, ., Y_{1,n_1}$ and $Y_{2,1}, ., Y_{2,n_2}$ are two simple random samples from two independent populations, with sample size n_1 , n_2, population mean μ_1 , μ_2 and variance σ_1^2 , σ_2^2 respectively. Let N denote the total sample size with $N = n_1 + n_2$. We are interested in testing wether the two populations have equal means or not. In this section, we will review several commonly used tests in practice under the above settings, including ordinary two sample t-test, modifications of two sample t-test with non-normal data, some nonparametric tests and the test based on resampling method. A review of each test is given as follows.

2.1 Commonly used tests

2.1.1 Two-sample t-test

The most commonly used test is Student's t-test, proposed by Gosset (Mankiewicz, 2004; Box, 1987). There are various versions of student t-test depending on the context of problem. Under the settings above, if variances of two populations are different, unpooled two sample t-test is used. The test statistic is

$$T = \frac{\bar{Y_1} - \bar{Y_2} - (\mu_1^0 - \mu_2^0)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \tag{2.1.1}$$

where μ_i^0 is the true population mean under null hypothesis, $\bar{Y_i} = \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $S_i^2 = \frac{1}{n_i-1}\sum_{i=1}^{n_i}(Y_i - \bar{Y_i})^2$ with $i = 1, 2$. Under the null hypothesis, when data are normally distributed the test statistic follows a t-distribution with degree of freedom

$$DF = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}.$$

If the variance of two populations are the same, the pooled two sample t-test should be used with the test statistic

$$T = \frac{\bar{Y_1} - \bar{Y_2} - (\mu_1^0 - \mu_2^0)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{2.1.2}$$

where $Sp = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$. Under the null hypothesis, when data are normally distributed, the test statistic follows a t-distribution with degree of freedom $n_1 + n_2 - 2$.

Both versions of the two sample t-test presented above require certain assumptions. One of the most important assumptions is that the data follow normal distribution. This assumption ensures that the test statistics in equations (2.1.1) and (2.1.2) follow t-distribution with corresponding degrees of freedom. If the normality assumption is violated, the exact distribution of test statistic becomes unknown. T-test is known to be robust to modest departures from the normality assumption (Mankiewicz, 2004). However, it still has its limitations depending on the magnitude of departure from normal distribution.

Skewness is one of the most commonly used statistic to quantify the magnitude to which the data are asymmetry. Let γ denote the value of skewness of the population calculated as $\gamma_i = E[(\frac{Y_{ij}-\mu_i}{\sigma_i})^3]$. The sample skewness is $\hat{\gamma}_i = \frac{n_i}{(n_i-1)(n_i-2)}\sum_{j=1}^{n_i}(\frac{Y_{ij}-\bar{Y_i}}{S_i})^3$ with $i = 1, 2$ in Zhou and Philip (2005). Several investigations have been conducted to figure out the impact of skewness on Student's t-test (Chen, 1995; Gayen, 1949; Johnson, 1978). These studies found that the performance of t-test can be poor when the data are skewed (Barrett and Goldsmith, 1976; Boos and Hughes, 2000).

If the data are skewed, the most commonly used method to get around skewness is based on Central Limit Theorem (CLT). It has been justified that normal distribution can be used to approximate the distribution of test statistic in equations (2.1.1) and (2.1.2) with skewed data. When the sample size is large this approximation is shown to have order of accuracy $O(n^{-1/2})$ by Hall (1992a). That is,

$$P(T \leq x) = \Phi(x) + O(n^{-1/2}), \tag{2.1.3}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. The limitation for application of this approximation is that, for a specific data set, a reasonable size of n used for approximation in (2.1.3) is unknown. The size of n should depend on the skewness (Barrett and Goldsmith, 1976; Boos and Hughes, 2000).

Another commonly used method under skewness does a transformation on the observed data, such that the transformed data follow normal distribution. A logarithm transformation is usually applied to the original data, followed by the two sample t-test on the transformed data. Finally, inferences will be made on the mean of the transformed data. By log transformation, skewness problem is avoided on the two sample t-test. However, the test on the two sample mean difference on the log-transformed data do not always reveal the relationship of two population means of the original skewed data, due to the fact that $E(log(X)) \leq log(E(X))$ as a result of the Jensen's Inequality.

2.1.2 Wilcoxon Rank-Sum Test

The two-sample t test is based on three important assumptions: independent sampling, normality, and equal variances. When the conditions of normality and equal variances are not valid but the sample sizes are large, using a t test is approximately correct. In this case, the Wilcoxon Rank-Sum test provides an alternative test procedure that requires less stringent conditions for comparing two independent samples. It replaces the normality assumption with that the two samples are taken from identical distribution. It does not require that the populations have normal distribution. The other conditions, equal variances and independence of the random samples, are still required for the Wilcoxon Rank-sum test.

In Wilcoxon Rank-sum test, the ranks of all observations are first obtained from the combined samples. Let W_i be the sum of the ranks of the observations from sample i, here $i = 1, 2$. The Wilcoxon Rank-Sum test is a two-sample permutation test based on W_i.

Assume that n_1 observations are from sample 1 and n_2 observations from sample 2. And no two observations have the same value, so that the ranks are distinct. The procedure of the rank sum test consists of the following steps in below:

- Combine the $n_1 + n_2$ observations and rank them from smallest to largest. Find the observed rank sum W_i of sample i, $i = 1, 2$.

- Find all the possible permutation of the ranks in which n_1 ranks are assigned to sample 1 and n_2 ranks are assigned into sample 2 separately.

- For each permutation of the ranks, find the sum of the ranks for sample 1 (or sample 2).

- Determine the upper-tail, lower-tail, or two-sided p-value. For an upper-tail test, the p-value is:

$$P_{upper\ \ tail} = \frac{\text{number of rank sums} \geq \text{observed rank sum } W}{\binom{n_1+n_2}{n_1}}$$

The rank sum of either sample can be used. The choice of sample 1 is arbitrary. Instead of using the sum of the ranks, the test could also be based on the difference of mean ranks. Let W_1 be the Wilcoxon sum rank for sample 1. Since we have total $N = n_1 + n_2$ observations, the sum of all ranks is: $T = 1 + 2 + \ldots + N = N(N+1)/2$. And the difference of mean ranks is defined as:

$$\text{Difference of mean ranks} = W_1 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) - \frac{N(N+1)}{2n_2}$$

This implies that the statistical test based on the sum of ranks of one of the treatments will have the same p-value as just using the first method discussed above. And either to use upper-tail or lower-tail test is determined by the context of the specific problem. When there are ties among observations, all the tied observations are grouped together and the average rank to tied values in that group is calculated and assigned to them. These ranks are called mid-ranks. So the Wilcoxon rank-sum test adjusted for ties becomes:

- Compute the mid-ranks

- Perform the permutation test using the mid-ranks, where the test statistic is the sum of mid-ranks for sample 1 (or sample 2).

2.1.3 Mann-Whitney Test

The Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed. The Mann-Whitney U test is often viewed as the nonparametric equivalent of Student's t-test. The major difference between the Mann-Whitney Test and Student's t-Test is that the former one does not require a normal distribution of the data from the sample. Based on our two sample settings, assuming the data have no ties so a given observation is either strictly less than or strictly larger than any other observation. The Mann-Whitney statistic, denoted as U, is defined as:

$$U = \text{number of pairs } (Y_{1i}, Y_{2j}) \text{ for which } Y_{1i} < Y_{2j}.$$

A large U value means that the larger observations tend to occur with sample 2 and a small U value means that the larger observations tend to occur with sample 1. Lower-tail U values and upper-tail U values under the null hypothesis (the distributions of Y_1 and Y_2 are the same) are related as: $U_{upper} = n_1 n_2 - U_{lower}$. The procedure of the Mann-Whitney U test consists of the following steps as follows:

- Assign numeric ranks to all observations, starting with 1 for the lowest rank. The observations with tied values are assigned a mid-rank.
- Calculate the sum of the rank W_1 for sample 1 and W_2 for sample 2.
- U_i is defined as: $U_i = W_i - \frac{n_i(n_i+1)}{2}$. We have $U_1 + U_2 = n_1 n_2$, $i = 1, 2$.
- The smaller U_i is used to consult significance tables.

The Wilcoxon Rank Sum Test and the Mann-Whitney Test are equivalent. In fact, the test is often call the Mann-Whitney-Wilcoxon Test (or more commonly called the MWW Test). These two tests are equivalent in the sense that one is a linear combination of the other.

2.2 Test bases on resampling methods

2.2.1 Permutation test

The method of permutation, also called randomization, is a very general approach to testing statistical hypotheses. Permutation test can be traced back to at least Fisher (1935) and Pitman (1937). Permutation test provides an efficient alternative when the data do not follow normal distribution. It is applicable to very small samples without specifying the parametric form of the underlying distribution. The speed of modern computers allow us to perform many statistical test using the permutation method. The advantage is that one does not have to worry about distribution assumptions of classical testing procedures. In a two-sample case, let $F_i(x)$ be the cdf of population i, $i = 1, 2$. The two-sided hypothesis of permutation test is:

$$H_0 : F_1(x) = F_2(x) \qquad H_a : F_1(x) \leq F_2(x) \ or \ F_1(x) \geq F_2(x) \ for \ all \ x,$$

where strictly inequality happens for at least one x. The alternative hypothesis indicates that the observations for population 1 tend to be larger or smaller than the observations for population 2. With the same two-sample setting, under the above null hypothesis, any permutation of the observations between the two populations have the same chance to happen as any other permutations. The steps for a two-sample permutation test are as follows:

- Compute the mean difference between the two samples, denote as D_{obs};
- Permute the N observations from the combined samples, so that there are n_1 observations for sample 1 and n_2 observations for sample 2. Obtain all the possible permutations with a total number $\binom{N}{n_1}$;
- For each permutation of the data, compute the mean difference, denoted as D;
- If the population mean μ_1 is bigger than μ_2, under H_1 compute the p-value as the

proportion of D's greater than or equal to D_{obs}, i.e.

$$P_{upper\ \ tail} = \frac{number\ of\ D's \geq D_{obs}}{\binom{N}{n_1}}$$

- If the level of significance is α, then reject H_0 if the p-value $\leq \alpha$.

This permutation test is very flexible. One can choose a test statistic fitted to the context of the question. Besides testing the mean difference between two populations as a test statistic, one can also use the sum of each sample. And these two methods will reach the same conclusion.

2.2.2 Bootstrap tests

Bootstrap is a well-known method to derive asymptotic approximations for carrying out inference. The basic idea of bootstrapping is that inference about a population from sample can be modeled by resampling the sample and performing inference on those resamples. More formally, given the original data, the bootstrap works by treating inference of the true probability distribution f, as being analogous to inference of the empirical distribution of $\hat{f}$, given the resampled data. The accuracy of inferences regarding $\hat{f}$ using the resampled data can be assessed because $\hat{f}$ is known. If $\hat{f}$ is a reasonable approximation to f, then the quality of inference on f can in turn be inferred. This technique allows estimation of the sampling distribution of almost any statistic using random sampling methods.

2.2.2.1 Basic Bootstrap

The permutation tests described in Section 2.2.1 are special nonparametric resampling tests, in which the resampling process is done without replacement. In this section we discuss the resampling procedure with replacement rather than without. Basic bootstrap test is based on the resampling procedure with replacement, which may apply to much wider areas including hypothesis testing.

When doing significance tests, the probability calculation under the null hypothesis is crucial. With the same two-sample settings about Y_1 and Y_2, we are interested in comparing two populations means. So the test statistic of basic bootstrap test is defined as $T = \bar{Y_1} - \bar{Y_2}$. The steps for a basic bootstrap test are listed below:

- Randomly draw n_i bootstrap samples from sample i with replacement, $i = 1, 2$. Repeat this process B times.

- Calculate the bootstrap sample mean $\bar{Y_i}^*$ for sample i, and then the bootstrap test statistic $T^* = \bar{Y_1^*} - \bar{Y_2^*}$.

- Under $H_0 : \mu_1 = \mu_2$, compute the p-value as the proportion of $T^* \geq T$, which is

$$P\text{-value} = \frac{1 + \ \textit{number of } T^* \geq \ T}{B + 1}$$

2.2.2.2 Tilted Bootstrap

Based on the same two-sample example of basic bootstrap test above, generally we might tilt the empirical distribution of T^* by sampling with weight $p_i = (p_{i1}, \ldots, p_{in_i})$, attached to the data values $y_{i1}, \ldots, y_{in_i}$, with $i = 1, 2$. For the sampling procedure of basic bootstrap, the corresponding weight $p_i = n_i^{-1}(1, \ldots, 1)$. Of course, the p_i's form a multinomial distribution with $p_i > 0$ and $\sum_i p_i = 1$. The Tilted bootstrap test follows the same steps of the basic bootstrap test except sampling with weight p_i, which is listed as follows:

- Randomly draw with replacement n_i bootstrap samples from sample i with weighs $(\hat{p}_{i1}, \ldots, \hat{p}_{in_i})$, here $i = 1, 2$ and $\hat{p}_{ij}$ is the weight of jth observation from sample i. Repeat this process B times.

- Calculate the bootstrap sample mean $\bar{Y_i}$ for sample i, then the bootstrap test statistic is $T^* = \bar{Y_1^*} - \bar{Y_2^*}$.

- Under $H_0 : \mu_1 = \mu_2$, compute the p-value as the proportion of $T^* \geq T$, which is

$$P\text{-value} = \frac{1 + \ \textit{number of } T^* \geq \ T}{B + 1}$$

Tilting is used in many contemporary generalizations of the bootstrap, such as empirical likelihood and the weighted or biased bootstrap.

2.2.2.3 Studentized Bootstrap

The studentized bootstrap test is generally referred as bootstrap-t test, which is the test that works similarly as the usual Student's t-test, but replaces the quantiles derived from the normal or Student t-distribution approximation by the quantiles from the bootstrapped distribution of the Student's t-test. The estimated quantiles from bootstrap distribution are theoretically demonstrated to give more accurate asymptotic approximation under skewness (Hall, 1992a).

To illustrate the procedure of Bootstrap-t, consider the same two sample settings about Y_1 and Y_2 and testing whether the two populations have equal means or not. The test statistic of pooled two sample t-test T is defined in (2.1.2). The Bootstrap-t test uses resampling method to estimate the quantiles of test statistic T by its bootstrapped distribution. The main principle is illustrated as follows:

- Draw B bootstrap samples of size $N = n_1 + n_2$ with replacement from the original two samples respectively, with sample of size n_1 from original sample one and sample of size n_2 from original sample two.

- For each bootstrap sample, compute

$$T_b^* = \frac{\bar{Y}_{1b}^* - \bar{Y}_{2b}^* - \bar{Y}_{1n} + \bar{Y}_{2n}}{S_{p_b}^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

  where $\bar{Y}_{1b}^*$ and $\bar{Y}_{2b}^*$ are the sample means of the b^{th} bootstrap sample from sample 1 and sample 2 respectively; $\bar{Y}_{1n}$ and $\bar{Y}_{2n}$ are the sample means of original sample 1 and original sample 2 respectively; $S_{p_b}^* = \sqrt{\frac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}}$ is the pooled two sample standard deviation of bth bootstrap sample; $S_i^* = \sqrt{\frac{1}{n_i-1}\sum_{j=1}^{n_i}(Y_{ij}^* - \bar{Y}_{ib}^*)^2}$, $i = 1, 2$ is the b^{th} bootstrap sample standard deviation for each sample.

- Estimate the p-value as the proportion of $T_b^* \geq T$, which is

$$P\text{-value} = \frac{1 + \ number \ of \ T_b^* \geq \ T}{B + 1}$$

2.2.2.4 Other Bootstrap methods

Several bootstrap methods are specifically created for constructing confidence intervals. In this section, we will review two of them: bias-corrected and accelerated bootstrap (BCa) by Efron (1987) and t-Pivot by Fisher and Hall (1989).

The bias-corrected and accelerated bootstrap (BCa) is an improved version of bootstrap method, which adjusts for both bias and skewness in the bootstrap distribution. This method has been widely used in constructing confidence intervals. And it has been shown to give better intervals in terms of higher coverage accuracy, narrower width and less computation requirements.

These good properties of BCa bootstrap intervals will be demonstrated by the following example. Considering the location model $X = \mu + \varepsilon$, where $E(\varepsilon) = 0$, the interest is to construct a confidence interval for μ. With an i.i.d. sample of $X : \{X_1, ..., X_n\}$, a standard way to construct the confidence interval is based on an asymptotic normal approximation as follows:

$$\frac{\bar{X}_n - \mu}{s/\sqrt{n}} \sim N(0,1).$$

Its $1 - 2\alpha$ standard confidence interval is

$$[\bar{X}_n + z_\alpha s/\sqrt{n} \ \ , \ \ \bar{X}_n + z_{1-\alpha} s/\sqrt{n}], \tag{2.2.1}$$

where z_α and $z_{1-\alpha}$ are the α^{th} and $(1-\alpha)^{th}$ percentiles of standard normal distribution. When the population distribution is heavily skewed, the above confidence interval (2.2.1) can be greatly improved by replacing $\bar{X}_n$ and μ by some monotone transformation $g(.)$ with $\hat{\phi} = g(\bar{X}_n)$ and $\phi = g(\mu)$, some bias constant z_0, and some acceleration constant a. Now the asymptotic normal approximation becomes:

$$\begin{aligned} (\hat{\phi} - \phi)/\tau &\sim N(-z_0\sigma_\phi, \sigma_\phi^2) \\ \sigma_\phi &= 1 + a\phi, \end{aligned} \tag{2.2.2}$$

where $\sigma_\phi > 0$ and τ is the standard error of $\hat{\phi}$. In addition, it has been shown that (2.2.2) can always be reduced to the case with $\tau = 1$ by (Efron, 1987). Denote $\hat{G}^{-1}(\alpha)$ to be the

α^{th} percentile of the bootstrap distribution. Then the $(1-2\alpha)^{th}$ BCa bootstrap confidence interval for μ becomes

$$[\hat{G}^{-1}(\Phi(z_{[\alpha]})), \hat{G}^{-1}(\Phi(z_{[1-\alpha]}))],$$

where

$$z_{[\alpha]} = z_0 + \frac{(z_0 + z_\alpha)}{1 - a(z_0 + z_\alpha)},$$

and $z_{[1-\alpha]}$ is similarly defined. In addition, the constant a and z_0 are estimated by:

$$z_0 \approx \Phi^{-1}(\hat{G}(\bar{X}_n))$$
$$a \approx \frac{1}{6} SKEW_{\mu=\bar{X}_n}(\dot{I}_\mu),$$

where $SKEW_{\mu=\bar{X}_n}(X)$ is the skewness of a random variable X, evaluated at parameter $\mu = \bar{X}_n$, and $\dot{I}$ is the score function of the family $f_\mu(\bar{X}_n)$:

$$\dot{I}_\mu(\bar{X}_n) = \partial/\partial\mu \; log \; f_\mu(\bar{X}_n). \tag{2.2.3}$$

Luckily, the bootstrap process will automatically transfer $\bar{X}_n$ to normal. Therefore one does not need to compute the exact form of transformation function $g(.)$ beforehand.

Besides bias-corrected and accelerated bootstrap (BCa) method, pivoting bootstrap is another improved version of bootstrap method for constructing confidence intervals. The pivoting bootstrap confidence intervals were first created by Fisher and Hall (1989). They found that the bootstrap confidence intervals based on pivotal statistics have higher coverage accuracy than the ones derived from nonpivotal statistics. In our two-sample case, with test statistic T defined in (2.1.2), the distribution of T depends only on the distribution of the error term ε. To demonstrate this, replacing Y_{ij} in (2.1.2) with $\mu_i + \varepsilon_{ij}$ gives

$$T_\varepsilon = \frac{\bar{\varepsilon}_1 - \bar{\varepsilon}_2}{S_{\varepsilon p}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\bar{\varepsilon}_i$ is the sample mean of error of sample i, and $S^2_{\varepsilon p}$ is

$$S^2_{\varepsilon p} = \sum_{i=1}^{2}\sum_{j=1}^{n_i} \frac{(\varepsilon_{ij} - \bar{\varepsilon}_i)^2}{n_1 + n_2 - 2}.$$

If ε_{ij} follows normal distribution, then T and T_ε are equivalent under the null hypothesis and both follow t-distribution with degrees of freedom $n_1 + n_2 - 2$. So the 95% confidence interval for $\mu_1 - \mu_2$ can be obtained by solving the following inequality

$$-t_{0.975} \;<\; T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \;<\; t_{0.975},$$

where $t_{0.975}$ is the 97.5^{th} percentile of the t-distribution with degrees of freedom $n_1 + n_2 - 2$. If ε_{ij}'s do not follow normal distribution, then t-Pivot bootstrap can be used to approximate the distribution of T_ε and obtain a bootstrap confidence interval for the difference of the means $\mu_1 - \mu_2$. The steps are as follows:

- Calculate the observed error $\hat{\varepsilon}_{ij} = Y_{ij} - \bar{y}_i$ and combine the $\hat{\varepsilon}_{ij}$ of two samples together;
- Randomly draw n_i errors with replacement from the set of all errors and assign them to sample i. Denote these errors as ε^*_{ij} then compute T_ε based on ε^*_{ij} and denote it as T^*_ε.
- Repeat the previous step B times to get $T^*_{\varepsilon.b}$, $b = 1, 2, \ldots, B$;
- Find 2.5^{th} and 97.5^{th} sample percentiles of $T^*_{\varepsilon.b}$ and denote them as $t_{e,.025}$ and $t_{e,.975}$. Then the 95% bootstrap confidence interval of $\mu_1 - \mu_2$ based on t-Pivot can be obtained by solving the following inequality:

  $$t_{e,.025} \;<\; T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \;<\; t_{e,.975},$$

  with

  $$\bar{Y}_1 - \bar{Y}_2 - t_{e,.975} Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < \bar{Y}_1 - \bar{Y}_2 - t_{e,.025} Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

2.3 Edgeworth expansion and Cornish Fisher expansion

Edgeworth expansion by Hall (1992a) is a well known asymptotic expansion theory. It is used for investigating the behaviour of asymptotically normally distributed random variables such as sum of independent variables. In some cases, Edgeworth expansion theory helps find the convolution integral or sum explicitly and derive any asymptotic expansion from an explicit form. By the end, we can get an explicit function to approximate the distribution of the statistics. Cornish Fisher expansion is an inverse form of the Edgeworth expansion that gives an asymptotic expansion of percentiles, so it is also called Cornish Fisher inversion. Cornish and Fisher (1938) and Fisher and Cornish (1960) studied the Cornish Fisher expansion of the sums of independent variables. Generally, Cornish Fisher expansion can be derived from Edgeworth expansion. In this section, we will first introduce the general idea of these two expansions in Section 2.3.1. Then an example of Edgeworth expansion is provided to construct confidence interval for two sample mean difference from Zhou and Philip (2005) is given in Section 2.3.2.

2.3.1 General idea of the two expansions

We will demonstrate these two expansions by constructing the asymptotic expansion for $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$, where X_i are independently and identically distributed random variables from $F(x)$ with population mean $\mu = 0$ and variance $\sigma^2 = 1$. Denoting $\gamma = E(X_i^3)$ and $\tau = E(X_i^4)$ with $\tau < \infty$, γ and τ are the population skewness and kurtosis respectively. By Central limit theorem, for each X:

$$P(S_n \leq x) \to \Phi(x).$$

The Edgeworth expansion of S_n is built on the above approximation, but gives a better approximation of $P(S_n \leq x)$ than $\Phi(x)$. The main principle is illustrated as follows:

1. Compute the characteristic function of S_n:

$$\Psi_{S_n}(t) = E\, exp\{(it/\sqrt{n}) \sum_i X_i\} = [\Psi_X(t/\sqrt{n})]^n.$$

2. Apply Taylor expansion on $exp\{itX/\sqrt{n}\}$, as $n \to \infty$, then

$$\begin{aligned}\Psi_X(\frac{t}{\sqrt{n}}) &= E\{1 + \frac{itX}{\sqrt{n}} + \frac{(it)^2X^2}{2n} + \frac{(it)^3X^3}{6n\sqrt{n}} + \frac{(it)^4X^4}{24n^2}\} + o(\frac{1}{n^2}) \\ &= (1 - \frac{t^2}{2n}) + \frac{(it)^3\gamma}{6n\sqrt{n}} + \frac{(it)^4\tau}{24n^2} + o(\frac{1}{n^2}).\end{aligned}$$

Increasing $exp\{itX/\sqrt{n}\}$ to its nth power gives:

$$\Big[\Psi_X(\frac{t}{\sqrt{n}})\Big]^n = \Big[(1-\frac{t^2}{2n})^n + (1-\frac{t^2}{2n})^{n-1}(\frac{(it)^3\gamma}{6\sqrt{n}} + \frac{(it)^4\tau}{24n}) + (1-\frac{t^2}{2n})^{n-2}\frac{(n-1)(it)^6\gamma^2}{72n^2}\Big] + o(\frac{1}{n}).$$

By binomial theory (Bag, 1966), $\Psi_{S_n}(t)$ in step 1 becomes:

$$\begin{aligned}\Psi_{S_n}(t) &= e^{-t^2/2}\Big[1 - \frac{t^4}{8n} + \frac{(it)^3\gamma}{6\sqrt{n}} + \frac{(it)^4\tau}{24n} + \frac{(it)^6\gamma^2}{72n}\Big] + o(\frac{1}{n}) \\ &= e^{-t^2/2}\Big[1 + \frac{(it)^3\gamma}{6\sqrt{n}} + \frac{(it)^4(\tau-3)}{24n} + \frac{(it)^6\gamma^2}{72n}\Big] + o(\frac{1}{n}).\end{aligned}$$

3. Do a Fourier Transformation (Bochner and Chandrasekharan, 1949) on the approximated characteristic function of S_n in step 1 to get the following probability density function $g(x)$ as an approximation of the distribution of S_n:

$$g(x) = \frac{1}{2\pi}\Big(\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}dt + \frac{\gamma}{6\sqrt{(n)}}\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}(it)^3dt + \frac{\tau-3}{24n}\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}(it)^4dt + \frac{\gamma^2}{72n}\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}(it)^6dt\Big).$$

Simplify the integrals in this equation to yield

$$g(x) = \phi(x)\Big(1 + \frac{\gamma H_3(x)}{6\sqrt{n}} + \frac{(\tau-3)H_4(x)}{24n} + \frac{\gamma^2 H_6(x)}{72n}\Big),$$

where $H_j(x)$ is the jth Hermite polynomials (Fedoryuk, 2001).

4. Integrate $g(x)$ to get its cumulative distribution function of $G(x)$:

$$\begin{aligned}G(x) &= \Phi(x) - \phi(x)\Big(\frac{\gamma H_2(x)}{6\sqrt{n}} + \frac{(\tau-3)H_3(x)}{24n} + \frac{\gamma^2 H_5(x)}{72n}\Big) \\ &= \Phi(x) - \phi(x)\Big(\frac{\gamma(x^2-1)}{6\sqrt{n}} + \frac{(\tau-3)(x^3-3x)}{24n} + \frac{\gamma^2(x^5-10x^3-15x)}{72n}\Big).\end{aligned} \tag{2.3.1}$$

The $G(x)$ function in equation (2.3.1) is the second-order Edgeworth expansion of distribution of S_n. And the first-order Edgeworth expansion of distribution of S_n is:

$$P(S_n \leq x) = G(x)' + O(n^{-1}) = \Phi(x) - \phi(x)\Big(\frac{\gamma(x^2-1)}{6\sqrt{n}}\Big) + O(n^{-1}). \tag{2.3.2}$$

From the above two expansions in (2.3.1) and (2.3.2) we can see that:

1. If distribution $F(x)$ is symmetric and $\gamma = 0$, then zero order central limit theorem approximation $\Phi(x)$ becomes first-order accurate.

2. The distribution approximation of S_n based on the first order Edgeworth expansion in (2.3.2) can adjust to the population skewness, which provide a more accurate approximation than the one from Central limit theorem with skewed population.

From the approximated distribution $G(x)$ in (2.3.2), the α^{th} percentiles of $F(x)$ can be derived by Cornish Fisher expansion. By inverting the equation (2.3.2), we can show that the solution $x = \mu_\alpha$ of the equation $P(S_n \leq x) = \alpha$ admits an expansion as follows:

$$\mu_\alpha = z_\alpha + \frac{\gamma(x^2 - 1)}{6\sqrt{n}}, \tag{2.3.3}$$

where z_α is the α^{th} percentile of standard normal distribution given by $\Phi(z_\alpha) = \alpha$. The inverse formula in (2.3.3) is called first-order Cornish Fisher expansion (Hall, 1992a). The percentile of μ_α in (2.3.3) adjusts the approximation by taking into account of the population skewness, which provides a more accurate approximation under high skewness.

2.3.2 Application of Edgeworth expansion in two-sample case

So far, the Edgeworth expansion theory has been used to construct confidence intervals under skewness in many studies. The corresponding confidence intervals or tests give more accurate results than the ordinary ones (Hall, 1992a). We will first review the way to construct confidence intervals by three transformations depending on the final form of Edgeworth expansion. Section 2.3.2.2 reviews the work on constructing confidence intervals by Edgeworth expansion in a two-sample comparison scenario.

2.3.2.1 Confidence intervals based on three transformations

Considering the same settings from Section 2.3.1, X_i are independently and identically distributed random variables from $F(x)$ with population mean μ and variance σ^2. Usually, the interval of population mean μ is based on the one-sample t-statistic:

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

where $\bar{X}$ is the sample mean, S is the sample standard deviation and n is the sample size. The corresponding t-statistic based confidence interval of population mean μ is:

$$\left(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right),$$

where $t_{\alpha/2,n-1}$ is the $\frac{\alpha}{2}^{th}$ percentile of a t-distribution with degrees of freedom $n-1$. This confidence interval is known to have exact $1-\alpha$ coverage when $F(x)$ follows normal distribution. But when $F(x)$ does not follow normal distribution and highly skewed, the coverage accuracy of the above t-statistic based confidence interval can be poor.

Hall (1992b) and Zhou and Philip (2005) proposed three transformations to set up the confidence intervals of Studentized statistics under skewness. These three transformations can eliminate skewness from the distribution of a Studentized statistic. Let us denote these three transformations as T_i, $i = 1, 2, 3$, and they are listed as follows:

$$T_1 = T_1(U) = U + a\hat{\gamma}U^2 + \frac{1}{3}a^2\hat{\gamma}^2U^3 + n^{-1}b\hat{\gamma} \tag{2.3.4}$$

$$T_2 = T_2(U) = (2an^{-1/2}\hat{\gamma})^{-1}\{exp(2an^{-1/2}\hat{\gamma}U) - 1\} + n^{-1}b\hat{\gamma} \tag{2.3.5}$$

$$T_3 = T_3(U) = U + U^2 + \frac{1}{3}U^3 + n^{-1}b\hat{\gamma}, \tag{2.3.6}$$

where the values of a, b and γ depend on the final form of the statistic derived from Edgeworth expansion. For example, under the settings in Section 2.3.1, X_i are independently and identically distributed random variables from $F(x)$ with population mean $\mu = 0$ and variance $\sigma^2 = 1$. Applying T_1 transformations to $U = \frac{1}{\sqrt{n}}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the values of a, b and γ in (2.3.4) depend on the Edgeworth expansion of S_n in (2.3.2):

$$\begin{aligned} P(S_n \leq x) = G(x) + O(n^{-1}) =& \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left[\gamma(\frac{x^2}{6} - \frac{1}{6})\right] + O(n^{-1}) \\ =& \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left[\gamma(ax^2 + b)\right] + O(n^{-1}) \end{aligned}$$

We have $a = \frac{1}{6}$, $b = -\frac{1}{6}$ and γ equals to the population skewness of X_i.

In addition, these three transformations have two properties:

1. After transformation, the Studentized statistic under skewness becomes virtually symmetric and approximately normal, with property:

$$P(\sqrt{n}T_i(U) \leq x) = \Phi(x) + O(n^{-1}). \quad (2.3.7)$$

2. The three transformations are monotone and have simple, explicit inversion formulae as follows:

$$T_1^{-1}(t) = (a\hat{\gamma})^{-1}\{1 + 3a\hat{\gamma}(t - b\hat{\gamma}/n)\}^{1/3} - (a\hat{\gamma})^{-1} \quad (2.3.8)$$

$$T_2^{-1}(t) = (2an^{-1/2}\hat{\gamma})^{-1}log\{2an^{-1/2}\hat{\gamma}(t - n^{-1}b\hat{\gamma}) + 1\} \quad (2.3.9)$$

$$T_3^{-1}(t) = \{1 + 3(t - n^{-1}b\hat{\gamma})\}^{1/3} - 1. \quad (2.3.10)$$

According to the above two properties, the procedure to construct t-statistic based confidence intervals under skewness by each of the three transformations is:

- Denote the t-statistic as T. Do a T_i transformation on the $U = \frac{1}{\sqrt{n}}T$ to derive a new statistic. The distribution of the transformed statistic $\sqrt{n}T_i(U)$ is virtually symmetric and approximate standard normal

- Use the percentile of standard normal distribution, say $z_{\alpha/2}$, to approximate the percentile of the transformed variable $\sqrt{n}T_i(U)$

- Now, $T_i(U)_{\alpha/2} = \frac{z_{\alpha/2}}{\sqrt{n}}$ is the $\frac{\alpha}{2}^{th}$ percentile of $T_i(U)$. Plugging in $T_i(U)_{\alpha/2}$ into $T_i^{-1}(U)$ to get the corresponding $\frac{\alpha}{2}^{th}$ percentile of U

- Since $U = \frac{1}{\sqrt{n}}T$, the $\frac{\alpha}{2}^{th}$ percentile of the original t-test statistic is $\sqrt{n}T_i^{-1}\left(\frac{z_{\alpha/2}}{\sqrt{n}}\right)$

By the above procedure, we can get confidence interval of μ from each of the three transformations as follows:

$$\left(\bar{X} - \sqrt{n}T_i^{-1}\left(\frac{z_{1-\alpha/2}}{\sqrt{n}}\right)\frac{S}{\sqrt{n}}, \bar{X} - \sqrt{n}T_i^{-1}\left(\frac{z_{\alpha/2}}{\sqrt{n}}\right)\frac{S}{\sqrt{n}}\right),$$

where $i = 1, 2, 3$ and $z_{\alpha/2}$ and $z_{1-\alpha/2}$ is the $\alpha/2$th and $1 - \alpha/2$th percentiles of standard normal distribution respectively. The confidence intervals based on three transformations above have better coverage accuracy than the ones from t-statistic under skewness.

2.3.2.2 Confidence intervals in two sample case by Edgeworth expansion

Zhou and Philip (2005) applied the above idea to construct confidence intervals in a two-sample scenario. Suppose $Y_{1,1}, ., Y_{1,n_1}$ and $Y_{2,1}, ., Y_{2,n_2}$ are two simple random samples from two independent populations, with sample sizes n_1 , n_2, population means μ_1 , μ_2, variances σ_1^2 , σ_2^2 and skewness γ_1^2 , γ_2^2, respectively. Let N denote the total sample size, i.e., $N = n_1 + n_2$. We are interested in constructing confidence intervals for $\mu_1 - \mu_2$.

By Edgeworth expansion theory, Zhou and Philip (2005) derived an approximation distribution for the test statistic of unpooled two sample t-test defined in (2.1.1) as follow: Let $\lambda_N = n_1/(n_1 + n_2) = n_1/N$. Assume $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$. Under regularity conditions in Hall (1992a) Appendix A, the distribution of the unpooled two sample t-statistic T given in (2.1.1) has the following expansion:

$$P(T \leq x) = \Phi(x) + \frac{A}{6\sqrt{N}}(2x^2 + 1)\phi(x) + O(N^{-min(1,r+1/2)}), \tag{2.3.11}$$

where $\phi(x)$ is the probability density function of the standard normal distribution, $\Phi(x)$ is the cumulative distribution function of the standard normal distribution and

$$A = \left\{\frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1-\lambda}\right\}^{-3/2}\left\{\frac{\sigma_1^3\gamma_1}{\lambda^2} - \frac{\sigma_2^3\gamma_2}{(1-\lambda)^2}\right\}. \tag{2.3.12}$$

By the same idea in Section 2.3.2.1, Zhou and Philip (2005) provided three $(1-\alpha)100\%$ transformation-based confidence intervals as follows:

$$\left(\bar{Y_1} - \bar{Y_2} - \sqrt{n}T_i^{-1}\left(\frac{z_{1-\alpha/2}}{\sqrt{n}}\right)S, \bar{X} - \sqrt{n}T_i^{-1}\left(\frac{z_{\alpha/2}}{\sqrt{n}}\right)S\right),$$

where T_i^{-1} is one of the three inverse transformations, $i = 1, 2, 3$ and $S = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$.

Zhou and Philip (2005) proved that the coefficient of $A/\sqrt{N}$ in equation (2.3.11) can represent the extent of skewness in a two-sample scenario. When $A/\sqrt{N}$ is small (< 0.3), the coverage accuracy of t-based confidence interval is good; Meanwhile, when the $A/\sqrt{N}$ is large (≥ 0.3), the t-based confidence intervals can be improved by T_1 and T_3 transformation-based confidence intervals.

So far, we have reviewed four confidence intervals of the difference of two sample means, one based on Cornish Fisher expansion and the other three based on transformations. The confidence intervals can account for population skewness. We would like to consider similar ideas to hypothesis tests and hope that the derived tests would have better test properties such as Type I error rate and power. In the next Section 2.4, we will review factors related to the performance of a test.

2.4 Factors related to type I error rate and power of the test

Type I error rate and power are two of the most important properties of a test. The value of type I error rate is defined as the probability to reject a true null hypothesis. Under our setting, the type I error is the incorrect rejection of a true null hypothesis. It happens when we conclude that the two populations have different means while their true population means are equal.

$$\text{Type I Error Rate} = P(\text{reject null hypothesis} \mid \text{null hypothesis is true}).$$

The maximum Type I error rate is defined as the level of significance, denoted as α. The value of α is often predetermined before data collection and usually $\alpha = 0.05$.

The power of a test is defined as the probability to reject a false null hypothesis, which under our setting is the probability of rejecting null hypothesis when the true population means are unequal.

$$Power = P(\text{reject null hypothesis} \mid \text{alternative hypothesis is true}).$$

For every test, we want to increase the power as much as possible, while maintain the type I error rate to be a small value. The power of a test depends on several parameters, which are listed as follows:

- the distribution of the test statistic
- the significance level;
- the sample size from each population;
- the effect size δ, defined as the difference between true and hypothetical value of the parameter of interest (Cohen, 1988);
- the two population variances;
- the two population skewness (if populations are skewed).

In two-sample comparison, the effect size equals the absolute value of the true population mean difference minus the hypothetical mean difference, as $\delta = (\mu_1 - \mu_2) - \text{hypothetical}(\mu_1 - \mu_2)$.

There are several ways to increase the power of a test. First, one can define a larger value of significance level, which will enlarge the area of rejection region. As a result, the probability of rejecting the null hypothesis will increase, so does the power of test. The second thing is to increase the sample size. The greater the sample size, the more information for the population. Then the test will have bigger chance to reject the null hypothesis when the two populations have different means. A third way is to magnify the effect size. The effect size actually reflects how much the 'true' value of the parameter is away from the one specified in the null hypothesis. In other words, the greater the difference between the 'true' value of the parameter and the value specified in the null hypothesis, the greater the power of the test.

On the contrary, either increasing the population variance or skewness decreases the power of a test. A big population variance increases the amount of sampling error inherent in a test result. This reduces the probability of rejecting the null hypothesis when the two populations have different means and hence reduces the power of the test. In addition, the population skewness has been proved to have effect on the power of one sample t-test by (Chaffin and Rhiel, 1993; Reineke et al., 2003). When data are skewed, the test statistic of t-test does not follow t-distribution any more. So it is no longer appropriate to use t-distribution to approximate the distribution of the test statistic. Furthermore, using an inaccurate test statistic distribution will not only decrease the accuracy of a test but also reduce the power of a test.

Although Zhou and Philip (2005) already gave the explicit form of Edgeworth expansion of the unpooled two sample test statistic, it is still important to find the form of Edgeworth expansion of the pooled two sample test statistic. As we know, with same population variances and unbalanced sample size, the pooled two sample test gives higher power than the unpooled two sample test. This increment of power of the test is a crucial factor to increase the test accuracy when the sample size is small and limited. In phase-one study of pharmaceutical industry, the researchers often need to test if there is a significantly difference in population means of some pharmaceutical dynamics parameters between two groups of healthy volunteers. In most of the studies, the two group sample sizes are often limited. In this case, the pooled two sample test is preferred when the two groups have same population variance, in that the pooled two sample test gives higher power and provides bigger chance to reject the null hypothesis when the two populations have different means. In the following Chapter 3 and Chapter 4, we will derived the explicit form of Edgeworth expansion of the pooled two sample test statistic and provide four new two sample tests under skewness, one based on Cornish Fisher expansion and three based on transformation.

Chapter 3

Proposed two-sample test using Cornish Fisher expansion

In this chapter, we will derive a better approximation of the distribution for pooled two sample t-statistics based on Edgeworth expansion theorem. The new approximated distribution can account for population skewness and gives more accurate test results. The theoretical results are given in Section 3.1 and 3.2. Section 3.3 presents a simulation study to investigate and compare the type I error rate and power of different two-sample tests under skewness.

3.1 Edgeworth expansion of the pooled two sample t-statistic

In this section, we derive the Edgeworth expansion of the pooled two sample t-statistics to achieve a better approximation of the distribution of the test statistics, which we expect to account for the effect of skewness.

Suppose $X_{1,1}, ., X_{1,n_1}$ and $X_{2,1}, ., X_{2,n_2}$ are two simple random samples from two independent populations with mean μ_1, μ_2 and common variance σ^2. The two population distributions are possibly from different families which are distinct from each other with population skewness γ_1 and γ_2 respectively. Denote N as the total sample size $N = n_1 + n_2$ and the population parameters are summarized in Table 3.1: We know that, the test statistic for

	mean	variance	skewness
population 1	μ_1	σ	γ_1
population 2	μ_2	σ	γ_2

Table 3.1: *Parameters of the two populations*

the pooled two sample t-test is as follows,

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1^0 - \mu_2^0)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3.1.1}$$

where μ_i^0 is the true population mean under null hypothesis, $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$, $S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$ and $S_i^2 = \frac{1}{n_i-1}\sum_{i=1}^{n_i}(X_i - \bar{X}_i)^2$ with $i = 1, 2$.

Let $Y_{ij}^* = \frac{X_{ij}-\mu_i^0}{\sigma}$, $\bar{Y}_i^* = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}^*$ and $S_i^{*^2} = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(Y_{ij}^* - \bar{Y}_i^*)^2$, for $i = 1, 2$ and $j = 1, ..., n_i$. Using these newly defined equations to replace the original statistics in pooled two sample t-statistic, we have

$$T = \frac{\sigma\bar{Y}_1^* - \sigma\bar{Y}_2^*}{\sqrt{\frac{(n_1-1)\sigma^2 S_1^{*^2}+(n_2-1)\sigma^2 S_2^{*^2}}{n_1+n_2-2}\frac{(n_1+n_2)}{n_1 n_2}}} = \sqrt{N}\frac{\bar{Y}_1^* - \bar{Y}_2^*}{\sqrt{\frac{(n_1-1)S_1^{*^2}+(n_2-1)S_2^{*^2}}{(N-2)\lambda_N(1-\lambda_N)}}}, \tag{3.1.2}$$

where $\lambda_N = n_1/N = n_1/(n_1 + n_2)$.

Furthermore, let $\mathbf{X} = (X_1, X_2, X_3, X_4)$, where

$$X_1 = \bar{Y}_1^* \ , \ X_2 = \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j}^{*^2} \ , \ X_3 = \bar{Y}_2^* \ , \ X_4 = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j}^{*^2}.$$

Now plugging $\mathbf{X}$ into (3.1.2) to further transform the pooled two sample t-statistic, finally we can write the test statistic T as a function of $\mathbf{X}$ with $T = \sqrt{N}g(\mathbf{X})$, which has the form as follows:

$$T = \sqrt{N}g(\mathbf{X}) = \frac{\sqrt{N}(X_1 - X_3)}{k(X)^{1/2}},$$

where, $k(\mathbf{X}) = \frac{(n_1-1){S_1^*}^2+(n_2-1){S_2^*}^2}{(N-2)\lambda_N(1-\lambda_N)} = \frac{(n_1-1)(X_2-X_1^2)+(n_2-1)(X_4-X_3^2)}{(N-2)\lambda_N(1-\lambda_N)}$. Next, applying Taylor expansion to $g(\mathbf{X})$ at $E\mathbf{X} \equiv U \equiv (U_1, U_2, U_3, U_4) = (0, 1, 0, 1)$ gives:

$$g(\mathbf{X}) = g(U) + \frac{\partial g(U)}{\partial U}'(\mathbf{X} - U) + \frac{1}{2}(\mathbf{X} - U)' \frac{\partial^2 g(U)}{\partial U^2}(\mathbf{X} - U) + \ldots,$$

note that $g(U) = 0$. so

$$T = \sqrt{N}\left\{\frac{\partial g(U)}{\partial U}'(\mathbf{X} - U) + \frac{1}{2}(\mathbf{X} - U)' \frac{\partial^2 g(U)}{\partial U^2}(\mathbf{X} - U) + \ldots\right\}.$$

If we let

$$W_N = \sqrt{N}\left\{\frac{\partial g(U)}{\partial U}'(\mathbf{X} - U) + \frac{1}{2}(X - U)' \frac{\partial^2 g(U)}{\partial U^2}(\mathbf{X} - U)\right\}.$$

Under regularity conditions, we can show that

$$T = W_N + O_p(N^{-1}).$$

Corollary 3.1.1. *Assuming $EY_{ij}^6 < \infty$. The first three moments of W_N are as follows:*

$$E(W_N) = -\frac{1}{2}N^{-1/2}[\lambda(1-\lambda)]^{1/2}(\gamma_1 - \gamma_2) + O(N^{-min(1,r+\frac{1}{2})}), \tag{3.1.3}$$

$$E(W_N^2) = 1 + O(N^{-1}), \tag{3.1.4}$$

$$E(W_N^3) = -\left[\frac{\lambda(1-\lambda)}{4N}\right]^{1/2}\left(\frac{11\lambda - 2}{\lambda}\gamma_1 - \frac{9 - 11\lambda}{1-\lambda}\gamma_2\right) + O(N^{-min(1,r+\frac{1}{2})}), \tag{3.1.5}$$

where $\lambda_N = \frac{n_1}{n_1+n_2} = \frac{n_1}{N}$. Assume $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$. $\gamma_i = E[(\frac{Y_{ij}-\mu_i}{\sigma_i})^3]$ is the population skewness, $i = 1, 2$.

The proof of Corollary 3.1.1 is given in Appendix A.1.1.

Corollary 3.1.2. *Let K_{1N}, K_{2N} and K_{3N} be the first three cumulants of W_N. The following results hold:*

$$K_{1N} = -\frac{1}{2\sqrt{N}}[\lambda(1-\lambda)]^{1/2}(\gamma_1 - \gamma_2) + O(N^{-min(1/2,r+1/2)}), \tag{3.1.6}$$

$$K_{2N} = 1 + O(N^{-min(1,r+1/2)}), \tag{3.1.7}$$

$$K_{3N} = -\left[\frac{\lambda(1-\lambda)}{4N}\right]^{1/2}\left[(\frac{8\lambda - 2}{\lambda})\gamma_1 - (\frac{6 - 8\lambda}{1-\lambda})\gamma_2\right] + O(N^{-min(1/2,r+1/2)}), \tag{3.1.8}$$

where λ is given in Corollary 3.1.1.

The proof of Corollary 3.1.2 is given in Appendix A.1.2.

Let $\chi_N(t)$ be the characteristic function of W_N. Then:

$$\chi_N(t) = \exp\left\{K_{1N}(it) + K_{2N}\frac{(it)^2}{2} + K_{3N}\frac{(it)^3}{6} + \dots\right\}$$

Based on the properties of cumulants from (Good, 1977), all cumulants of order $r \geq 3$ of the standardized sum tend to zero, which is a demonstration of central limit theorem. Since W_N is a function of the standardized sum, then we have the following results:

$$\begin{aligned}\chi_N(t) &= \exp\left\{K_{1N}(it) + K_{2N}\frac{(it)^2}{2} + K_{3N}\frac{(it)^3}{6} + O(N^{-min(1,r+1/2)})\right\}\\ &= \exp\left(-\frac{t^2}{2}\right) exp\left\{N^{-1/2}\left(-A(it) - B(it)^3\right) + O(N^{-min(1,r+1/2)})\right\},\end{aligned} \tag{3.1.9}$$

where A and B are defined as

$$\begin{aligned}A &= [\lambda(1-\lambda)]^{1/2}(\gamma_1 - \gamma_2)/2,\\ B &= [\lambda(1-\lambda)]^{1/2}\left(\frac{8\lambda - 2}{\lambda}\gamma_1 - \frac{6 - 8\lambda}{1-\lambda}\gamma_2\right)/12.\end{aligned} \tag{3.1.10}$$

By Taylor expansion, we have

$$\chi_N(t) = exp\left(-\frac{t^2}{2}\right)\left\{1 + N^{-1/2}\left(-A(it) - B(it)^3\right) + O(N^{-min(1,r+1/2)})\right\}. \tag{3.1.11}$$

Using Fourier Transformation, the probability density function of W_N can be obtained with:

$$\begin{aligned}&f_{W_N}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\chi_N(t)dt\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} exp\left(-\frac{t^2}{2}\right)\left\{1 + N^{-1/2}\left(-A(it) - B(it)^3\right)\right\}dt + O(N^{-min(1,r+1/2)}).\end{aligned}$$

Based on the properties of Hermite polynomials of k^{th} order

$$H_k(x)\phi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}(it)^k dt.$$

We can write

$$\begin{aligned}f_{W_N}(x) &= \phi(x) - N^{-1/2}AH_1(x)\phi(x) - N^{-1/2}BH_3(x)\phi(x) + O(N^{-min(1,r+1/2)})\\ &= \phi(x)[1 + N^{-1/2}(3B - A)x - N^{-1/2}Bx^3] + O(N^{-min(1,r+1/2)}).\end{aligned}$$

where $H_1(x) = x$ and $H_3(x) = x^3 - 3x$. To get the cumulative distribution function of W_N, we can use another property of the Hermite polynomial

$$\frac{d}{dx}[H_k(x)\phi(x)] = -H_{k+1}(x)\phi(x).$$

Then

$$\begin{aligned}
P(W_N \leq x) =& \int_{-\infty}^{x} f_{W_N}(x)dx + O(N^{-min(1,r+1/2)}) \\
=& \int_{-\infty}^{x} \left[\phi(x) - N^{-1/2}AH_1(x)\phi(x) - N^{-1/2}BH_3(x)\phi(x)\right] dx + O(N^{-min(1,r+1/2)}) \\
=& \Phi(x) + N^{-1/2}AH_0(x)\phi(x) + N^{-1/2}BH_2(x)\phi(x) + O(N^{-min(1,r+1/2)}) \\
=& \Phi(x) + N^{-1/2}[A + B(x^2 - 1)]\phi(x) + O(N^{-min(1,r+1/2)}),
\end{aligned} \tag{3.1.12}$$

where $H_0(x) = 1$ and $H_2(x) = x^2 - 1$. Since $T = W_N + O_p(N^{-1})$, we have the following result for the distribution of T:

Theorem 3.1.3. *Let $\lambda_N = n_1/(n_1 + n_2) = n_1/N$. Assuming $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$, under regularity conditions (in Appendix A), the distribution of the pooled two sample t-statistic T given in (2.1.2) has the following expansion:*

$$F_T^{(1)}(x) = P(T \leq x) = \Phi(x) + N^{-1/2}[A + B(x^2 - 1)]\phi(x) + O(N^{-min(1,r+1/2)}), \tag{3.1.13}$$

where $\phi(x)$ is the probability density function of the standard normal distribution, $\Phi(x)$ is the cumulative distribution function of the standard normal distribution and A, B are defined in (3.1.10).

The proof of Theorem 3.1.3 is given in Appendix A.1.

The right hand side of equation (3.1.13) is the first order Edgeworth expansion of the pooled two sample t-statistic, which can account for population skewness by A and B in the second term of this expansion. The Edgeworth expansion theory can be applied to construct confidence intervals by giving better approximation of the percentiles of the test statistic. Following this idea, we construct a new two-sample test with rejection region derived from the Edgeworth expansion theory. The details are given in Section 3.2. This rejection region is expected to be more accurate than the ones from normal approximation under skewness. We expect the new test to provide a better power in detecting the true two population mean difference.

3.2 Cornish Fisher expansion and the new test based on Cornish Fisher expansion

3.2.1 A two-sample test based on Cornish Fisher expansion

One new test can be constructed by Cornish Fisher expansion theory. From Section 2.3.1, we know that the Cornish Fisher expansion can be used to compute the percentiles of the distribution derived from Edgeworth expansion. In this case, the percentiles of the distribution in (3.1.13) admits a Cornish Fisher expansion, which has the form as follows

Corollary 3.2.1. *Let η_α denote the α^{th} percentile of distribution $F_T^{(1)}(t)$ in (3.1.13). Then based on Cornish Fisher expansion theory, the value of η_α admits an expansion with the form below:*

$$\eta_\alpha = z_\alpha - N^{-1/2}[A + B(z_\alpha^2 - 1)] + O(N^{-min(1,r+1/2)}), \tag{3.2.1}$$

where z_α is the α^{th} percentile of the standard normal distribution and A , B are defined in (3.1.10).

This corollary is a direct result of the theory for Fisher expansion from Hall (1992a). Hence we omit the proof.

Now define $\hat{t}_\alpha^{cf} = z_\alpha - N^{-1/2}[\hat{A} + \hat{B}(z_\alpha^2 - 1)]$, where

$$\hat{A} = [\lambda_N(1-\lambda_N)]^{1/2}(\hat{\gamma_1} - \hat{\gamma_2})/2$$
$$\hat{B} = [\lambda_N(1-\lambda_N)]^{1/2}\left(\frac{8\lambda_N - 2}{\lambda_N}\hat{\gamma_1} - \frac{6 - 8\lambda_N}{1-\lambda_N}\hat{\gamma_2}\right)/12,$$

and

$$\hat{\gamma}_i = \frac{n_i}{(n_i-1)(n_i-2)} \sum_{j=1}^{n_i} \left\{ \frac{X_{ij} - \bar{X}_i}{S_i} \right\}^3$$
$$\lambda_N = \frac{n_1}{(n_1+n_2)} \qquad , i = 1, 2.$$

With the test statistic T defined in equation (2.1.2), we have:

1. The rejection region for two-sided test with hypothesis $H_0 : \mu_1 - \mu_2 = \mu_{10} - \mu_{20}$ vs. $H_a : \mu_1 - \mu_2 \neq \mu_{10} - \mu_{20}$ is

$$T \leq \hat{t}^{cf}_{\alpha/2} \text{ or } T \geq \hat{t}^{cf}_{1-\alpha/2} \tag{3.2.2}$$

2. The rejection region for one-sided upper tail test with hypothesis $H_0 : \mu_1 - \mu_2 = \mu_{10} - \mu_{20}$ vs. $H_a : \mu_1 - \mu_2 > \mu_{10} - \mu_{20}$ is

$$T \geq \hat{t}^{cf}_{1-\alpha}. \tag{3.2.3}$$

3. The rejection region for one-sided lower tail test with hypothesis $H_0 : \mu_1 - \mu_2 = \mu_{10} - \mu_{20}$ vs. $H_a : \mu_1 - \mu_2 < \mu_{10} - \mu_{20}$ is

$$T \leq \hat{t}^{cf}_{\alpha}. \tag{3.2.4}$$

We reject the null hypothesis if T falls into the rejection regions for corresponding alternative hypothesis. In the further discussions, we will refer the two sample test based on Cornish Fisher expansion as "TCF" and the two sample test based on normal approximation as "TN".

3.2.2 Type I error rate of the two-sided test based on Cornish Fisher expansion

In this section, we calculate the order of approximation to type I error rate for the two-sided test with rejection region in (3.2.2) from the first order Cornish Fisher expansion.

Under the two-sample setting in above Subsection 3.2.1, the distribution of the test statistic T is $F_T^{(1)}(x)$ defined in Theorem 3.1.3, and the rejection region of the test is defined in formula (3.2.2). Denote the two cutoffs in formula (3.2.2) as

$$\begin{aligned} \hat{t}^{cf}_{\alpha/2} &= z_{\alpha/2} - N^{-1/2}[\hat{A} + \hat{B}(z^2_{\alpha/2} - 1)] \triangleq z_{\alpha/2} + \hat{\Delta}_{N,\alpha/2} \\ \hat{t}^{cf}_{1-\alpha/2} &= z_{1-\alpha/2} - N^{-1/2}[\hat{A} + \hat{B}(z^2_{1-\alpha/2} - 1)] \triangleq z_{1-\alpha/2} + \hat{\Delta}_{N,1-\alpha/2}, \end{aligned} \tag{3.2.5}$$

where

$$\begin{aligned} \hat{\Delta}_{N,\alpha/2} &= -N^{-1/2}[\hat{A} + \hat{B}(z^2_{\alpha/2} - 1)] \\ \hat{\Delta}_{N,1-\alpha/2} &= -N^{-1/2}[\hat{A} + \hat{B}(z^2_{1-\alpha/2} - 1)]. \end{aligned}$$

Note that the two cutoffs in (3.2.2) are not symmetric about zero. Instead, they are a shift version of the cutoff from the large sample normal test. From Bai and NG (2005), under standard regularity conditions we have

$$\hat{\gamma}_i = \gamma_i + O_p(N^{-1/2}),$$

and as defined previously $\lambda_N = \lambda + O(N^{-r})$. Then we have the following results

$$\begin{aligned} \hat{A} &= [\lambda_N(1-\lambda_N)]^{1/2}(\hat{\gamma}_1 - \hat{\gamma}_2)/2 \\ &= [(\lambda(1-\lambda))^{1/2} + O(N^{-r})][\frac{\gamma_1 - \gamma_2}{2} + O_p(N^{-1/2})] \\ &= (\lambda(1-\lambda))^{1/2}\frac{\gamma_1 - \gamma_2}{2} + O_p(N^{-min(r,\frac{1}{2})}) \\ &= A + O_p(N^{-min(r,\frac{1}{2})}), \end{aligned}$$

similarly

$$\begin{aligned} \hat{B} &= [\lambda_N(1-\lambda_N)]^{1/2}\left(\frac{8\lambda_N - 2}{\lambda_N}\hat{\gamma}_1 - \frac{6 - 8\lambda_N}{1-\lambda_N}\hat{\gamma}_2\right)/12 \\ &= [(\lambda(1-\lambda))^{1/2} + O(N^{-r})]\left\{\frac{8\lambda - 2}{\lambda}\gamma_1 - \frac{6-8\lambda}{1-\lambda}\gamma_2 + O_p(N^{-min(r,\frac{1}{2})})\right\}/12 \\ &= (\lambda(1-\lambda))^{1/2}(\frac{8\lambda - 2}{\lambda}\gamma_1 - \frac{6-8\lambda}{1-\lambda}\gamma_2)/12 + O_p(N^{-min(r,\frac{1}{2})}) \\ &= B + O_p(N^{-min(r,\frac{1}{2})}). \end{aligned}$$

Based on the above results of $\hat{A}$ and $\hat{B}$, we can obtain

$$\begin{aligned}
&\hat{\Delta}_{N,\alpha/2} = -N^{-1/2}[\hat{A} + \hat{B}(z^2_{\alpha/2} - 1)] \\
&= -N^{-1/2}[(A + O_p(N^{-min(r,\frac{1}{2})})) + (B + O_p(N^{-min(r,\frac{1}{2})}))(z^2_{\alpha/2} - 1)] \\
&= -N^{-1/2}[A + B(z^2_{\alpha/2} - 1)] + O_p(N^{-min(1,r+1/2)}) \\
&= \Delta_{N,\alpha/2} + O_p(N^{-min(1,r+1/2)}),
\end{aligned}$$

and

$$\hat{\Delta}_{N,1-\alpha/2} = \Delta_{N,1-\alpha/2} + O_p(N^{-min(1,r+1/2)}).$$

Now we have

$$\begin{aligned}
\hat{t}^{cf}_{\alpha/2} &= t^{cf}_{\alpha/2} + O_p(N^{-min(1,r+1/2)}) \\
\hat{t}^{cf}_{1-\alpha/2} &= t^{cf}_{1-\alpha/2} + O_p(N^{-min(1,r+1/2)}),
\end{aligned}$$

and the following Lemma:

Lemma 3.2.2. *Let η_α denote the α^{th} percentile of distribution $F^{(1)}_T(t)$ in (3.1.13) and $\hat{\eta}_\alpha$ denote the estimate of η_α, which satisfies $\hat{\eta}_\alpha = \eta_\alpha + O_p(N^{-min(1,r+1/2)})$. Then under standard regularity conditions, the following result holds:*

$$\begin{aligned}
&P(T < \hat{\eta}_{\alpha/2}) + P(T > \hat{\eta}_{1-\alpha/2}) \\
&= 1 - F^{(1)}_T(\eta_{1-\alpha/2}) + F^{(1)}_T(\eta_{\alpha/2}) + O(N^{-min(1,r+1/2)}).
\end{aligned}$$

Recall that the distribution of $F^{(1)}_T(x)$ in Theorem 3.1.3 takes the following form

$$F^{(1)}_T(x) = P(T \leq x) = \Phi(x) + N^{-1/2}[A + B(x^2 - 1)]\phi(x) + O(N^{-min(1,r+1/2)}).$$

Hence, based on the result of Lemma 3.2.2, the type I error rate of the two-sample TCF test can be obtained as

$$\begin{aligned}
&P(\text{type I error of the two-sided TCF test}) \\
=&1 - F^{(1)}_T(t^{cf}_{1-\alpha/2}) + F^{(1)}_T(t^{cf}_{\alpha/2}) + O(N^{-min(1,r+1/2)}) \\
=&1 - \Phi(t^{cf}_{1-\alpha/2}) + \Phi(t^{cf}_{\alpha/2}) - N^{-1/2}[A + B((t^{cf}_{1-\alpha/2})^2 - 1)]\phi(t^{cf}_{1-\alpha/2}) \\
&+ N^{-1/2}[A + B((t^{cf}_{\alpha/2})^2 - 1)]\phi(t^{cf}_{\alpha/2}) + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{3.2.6}$$

Due to the fact that A and B in (3.1.10) are fixed, then we have

$$\Delta_{N,\alpha/2} = \Delta_{N,1-\alpha/2} = O(N^{-1/2}) \tag{3.2.7}$$

Note that

$$\begin{aligned}
&-N^{-1/2}[A + B((t^{cf}_{\alpha/2})^2 - 1)] \\
&= -N^{-1/2}[A + B((z_{\alpha/2} + \Delta_{N,\alpha/2})^2 - 1)] \\
&= -N^{-1/2}[A + B(z^2_{\alpha/2} + O(\Delta_{N,\alpha/2}) - 1)] \\
&= \Delta_{N,\alpha/2} + O(\Delta_{N,\alpha/2} \cdot N^{-1/2}) \\
&= \Delta_{N,\alpha/2} + O(N^{-1}).
\end{aligned} \tag{3.2.8}$$

Similarly,

$$\begin{aligned}
&-N^{-1/2}[A + B((t^{cf}_{1-\alpha/2})^2 - 1)] \\
&= \Delta_{N,1-\alpha/2} + O(N^{-1}) = \Delta_{N,\alpha/2} + O(N^{-1}).
\end{aligned} \tag{3.2.9}$$

Then apply Taylor expansion to $\Phi(t^{cf}_{1-\alpha/2})$, $\phi(t^{cf}_{1-\alpha/2})$ at $z_{1-\alpha/2}$ and to $\Phi(t^{cf}_{\alpha/2})$, $\phi(t^{cf}_{\alpha/2})$ at $z_{\alpha/2}$ correspondingly, we have

$$\begin{aligned}
&\Phi(t^{cf}_{1-\alpha/2}) = \Phi(z_{1-\alpha/2}) + \phi(z_{1-\alpha/2})\Delta_{N,1-\alpha/2} + O(\Delta^2_{N,1-\alpha/2}), \\
&\Phi(t^{cf}_{\alpha/2}) = \Phi(z_{\alpha/2}) + \phi(z_{\alpha/2})\Delta_{N,\alpha/2} + O(\Delta^2_{N,\alpha/2}), \\
&\phi(t^{cf}_{1-\alpha/2}) = \phi(z_{1-\alpha/2}) + \phi^{'}(z_{1-\alpha/2})\Delta_{N,1-\alpha/2} + O(\Delta^2_{N,1-\alpha/2}), \\
&\phi(t^{cf}_{\alpha/2}) = \phi(z_{\alpha/2}) + \phi^{'}(z_{\alpha/2})\Delta_{N,\alpha/2} + O(\Delta^2_{N,\alpha/2}).
\end{aligned}$$

Using these four Taylor expansions and (3.2.8), (3.2.9) to replace the terms in (3.2.6), we have:

$$
\begin{aligned}
& P(\text{type I error of the two-sided TCF test }) \\
= \quad & 1-\Phi(t^{cf}_{1-\alpha/2})+\Phi(t^{cf}_{\alpha/2})-N^{-1/2}[A+B((t^{cf}_{1-\alpha/2})^2-1)]\phi(t^{cf}_{1-\alpha/2}) \\
& +N^{-1/2}[A+B((t^{cf}_{\alpha/2})^2-1)]\phi(t^{cf}_{\alpha/2})+O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha/2-\phi(z_{1-\alpha/2})\Delta_{N,\alpha/2}+\alpha/2+\phi(z_{\alpha/2})\Delta_{N,\alpha/2}+O(\Delta^2_{N,\alpha/2}) \\
& +[\Delta_{N,\alpha/2}+O(N^{-1})][\phi(z_{1-\alpha/2})+\phi^{'}(z_{1-\alpha/2})\Delta_{N,\alpha/2}+O(\Delta^2_{N,\alpha/2})] \\
& -[\Delta_{N,\alpha/2}+O(N^{-1})][\phi(z_{\alpha/2})+\phi^{'}(z_{\alpha/2})\Delta_{N,\alpha/2}+O(\Delta^2_{N,\alpha/2})] \\
& +O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha+\Delta_{N,\alpha/2}\phi(z_{1-\alpha/2})-\Delta_{N,\alpha/2}\phi(z_{\alpha/2})+O(N^{-1})+O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha+O(N^{-min(1,r+1/2)}), \qquad (3.2.10)
\end{aligned}
$$

where (3.2.10) is due to the fact that $\phi(z_{1-\alpha/2}) = \phi(z_{\alpha/2})$ and $\Delta_{N,\alpha/2} = \Delta_{N,1-\alpha/2} = O(N^{-1/2})$. Then we have the follow theorem,

Theorem 3.2.3. *Under standard regularity conditions, when H_0 is true, the theoretical type I error rate of the two-sample TCF test, with level of significance α is*

$$
\begin{aligned}
& P(T<\hat{t}^{cf}_{\alpha/2})+P(T>\hat{t}^{cf}_{1-\alpha/2}) \\
& =\alpha+O(N^{-min(1,r+1/2)}).
\end{aligned}
$$

We can easily show that the approximated type I error rate of the pooled two-sample t-test based on normal approximation is $\alpha+O(N^{-min(1,r+1/2)})$. This means the type I error rate accuracy is the same order of $O(N^{-min(1,r+1/2)})$ for both the test based on Cornish Fisher expansion and the test based on normal approximation.

3.2.3 Power of the two-sided test based on Cornish Fisher expansion

Now consider data generated under $H_a : \mu_1-\mu_2 \neq \mu_{10}-\mu_{20}$. For power calculation, let $\delta = (\mu_1-\mu_2)-(\mu_{10}-\mu_{20})$. When H_a is true, the theoretical power equals the probability of rejecting the null hypothesis with formula:

$$P_{H_a}\left(T \geq \hat{t}^{cf}_{1-\alpha/2}\right) + P_{H_a}\left(T \leq \hat{t}^{cf}_{\alpha/2}\right).$$

Following the same Edgeworth expansion procedures in Section 3.1, the distribution of the test statistic T under H_a can be obtained:

Theorem 3.2.4. *Let $\lambda_N = n_1/(n_1+n_2) = n_1/N$. Assume $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$, and $\delta = O(N^{-1/2})$ under H_a. Then under regularity conditions (in Appendix A), the distribution of the pooled two sample t-statistic T given in (2.1.2) has the following expansion under H_a*

$$F_T^{(H_a)}(x) = P_{H_a}(T \leq x) = F_T^{(1)}(x-c_n) + \frac{q_n}{2}(x-c_n)\phi(x-c_n) + O(N^{-min(1,r+\frac{1}{2})}), \quad (3.2.11)$$

where $\phi(x)$ is the probability density function of the standard normal distribution, $\Phi(x)$ is the cumulative distribution function of the standard normal distribution, and $F_T^{(1)}(x)$ is the distribution of T under H_0 defined in Theorem 3.1.3. Here, $c_N = \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ and $q_N = \delta\sigma^{-1}[\lambda(1-\lambda)](\gamma_1-\gamma_2)$.

The proof of Theorem 3.2.4 is given in Appendix A.2. Now, denote

$$Q(x) = \frac{q_N}{2}(x-c_N)\phi(x-c_N). \quad (3.2.12)$$

Thus the equation (3.2.11) becomes

$$F_T^{(H_a)}(x) = P_{H_a}(T \leq x) = F_T^{(1)}(x-c_N) + Q(x) + O(N^{-min(1,r+1/2)}). \quad (3.2.13)$$

Let $t^{cf}_{1-\alpha/2}$ and $t^{cf}_{\alpha/2}$ be the $1-\alpha/2$ and $\alpha/2$ percentiles obtained from distribution $F_T^{(1)}(x)$. Denote $\hat{t}^{cf}_{1-\alpha/2}$ and $\hat{t}^{cf}_{\alpha/2}$ their sample estimate given in (3.2.5).

$$\begin{aligned}
\hat{L}^{cf}_u &= \hat{t}^{cf}_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} \\
&= z_{1-\alpha/2} + \hat{\Delta}_{N,1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} \\
&= U_{N,1-\alpha/2} + \hat{\Delta}_{N,1-\alpha/2} \\
&= U_{N,1-\alpha/2} + \Delta_{N,1-\alpha/2} + O_p(N^{-min(1,r+1/2)}) \\
&= L^{cf}_u + O_p(N^{-min(1,r+1/2)})
\end{aligned}$$

where $U_{N,1-\alpha/2} = z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ and $L_u^{cf} = U_{N,1-\alpha/2} + \Delta_{N,1-\alpha/2}$. Similarly

$$\begin{aligned}
\hat{L}_l^{cf} &= \hat{t}_{\alpha/2}^{cf} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} \\
&= z_{\alpha/2} + \hat{\Delta}_{N,\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} \\
&= U_{N,\alpha/2} + \hat{\Delta}_{N,\alpha/2} \\
&= U_{N,\alpha/2} + \Delta_{N,\alpha/2} + O_p(N^{-min(1,r+1/2)}) \\
&= L_l^{cf} + O_p(N^{-min(1,r+1/2)})
\end{aligned}$$

where $U_{N,\alpha/2} = z_{\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ and $L_l^{cf} = U_{N,\alpha/2} + \Delta_{N,\alpha/2}$. We have,

$$\hat{L}_u^{cf} = L_u^{cf} + O_p(N^{-min(1,r+1/2)})$$
$$\hat{L}_l^{cf} = L_l^{cf} + O_p(N^{-min(1,r+1/2)}).$$

Then based on the result of Lemma 3.2.2, under standard regularity conditions, the theoretical power of the two-sample TCF test can be obtained by

$$\begin{aligned}
& P_{H_a}\left(T \geq \hat{t}_{1-\alpha/2}^{cf}\right) + P_{H_a}\left(T \leq \hat{t}_{\alpha/2}^{cf}\right) \\
= \; & 1 - F_T^{(H_a)}(t_{1-\alpha/2}^{cf}) + F_T^{(H_a)}(t_{\alpha/2}^{cf}) + O(N^{-min(1,r+1/2)}) \\
= \; & 1 - \Phi(L_u^{cf}) - N^{-1/2}[A + B((L_u^{cf})^2 - 1)]\phi(L_u^{cf}) - Q(L_u^{cf}) \\
& + \Phi(L_l^{cf}) + N^{-1/2}[A + B((L_l^{cf})^2 - 1)]\phi(L_l^{cf}) + Q(L_l^{cf}) + O(N^{-min(1,r+1/2)}). \quad (3.2.14)
\end{aligned}$$

Under the local alternative, we consider the departure from the null hypothesis $\delta = \delta_N \to 0$ in the order of $O(N^{-1/2})$ i.e. $\delta_N = O(N^{-1/2})$. In this case, the order of $\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ is $O(1)$ and both $U_{N,\alpha/2}$, $U_{N,1-\alpha/2}$ have the same order of $O(1)$. While under the fixed alternative, δ equals a constant which is $O(1)$. Therefore the order of $\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ is $O(\sqrt{N})$ and both $U_{N,\alpha/2}$ and $U_{N,1-\alpha/2}$ have the same order of $O(\sqrt{N})$.

Note that the main purpose of this study is to improve the power of the two-sample test when the two population mean difference is small and the total sample size is limited, i.e., δ/σ is small and the total sample size N is not very big. Under this scenario, without loss of generality, we can assume:

$$z_{1-\alpha/2} > \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}.$$

For example, if $\alpha = 0.05$ we have:

$$
\begin{aligned}
& z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})} \; > \; 0 \\
\Rightarrow \qquad & 1.96 > \frac{\delta}{\sigma}\sqrt{\frac{n_1 n_2}{N}} \\
\Rightarrow \qquad & N < \frac{1.96^2\sigma^2}{\delta^2\lambda_N(1-\lambda_N)},
\end{aligned}
$$

The Figure 3.1 shows the upper bound of N that satisfies the above equation, with effect size $\frac{\delta}{\sigma}$ in $(0.1, 0.3)$. When $\frac{\delta}{\sigma} = 0.3$ the minimum upper bound of N is 170, which is reached at $\lambda_N = 0.5$. As long as N is smaller than the values on the plot with corresponding λ_N, we have:

$$
\begin{aligned}
& z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\tfrac{1}{n_1} + \tfrac{1}{n_2})} \; > \; 0 \\
\Rightarrow \qquad & \delta < z_{1-\alpha/2}\sigma\sqrt{\tfrac{n_1+n_2}{n_1 n_2}} \\
\Rightarrow \qquad & \delta = O(N^{-1/2}),
\end{aligned}
$$

Therefore, under the main focus of this study, the order of δ is more close to $O(N^{-1/2})$, which is the order under local alternative hypothesis. Thus, for further discussion, we will focus on the power of two-sample TCF test with the data generated under local alternative hypothesis.

3.2.3.1 Power of the two-sided TCF test under local alternative hypothesis

Recall that, under the local alternative, both $U_{N,\alpha/2}$ and $U_{N,1-\alpha/2}$ have the same order of $O(1)$. Therefore we have

$$
\begin{aligned}
& - N^{-1/2}[A + B((L_l^{cf})^2 - 1)] \\
= & - N^{-1/2}[A + B((U_{N,\alpha/2} + \Delta_{N,\alpha/2})^2 - 1)] \\
= & - N^{-1/2}[A + B(U_{N,\alpha/2}^2 + O(N^{-1/2}) - 1)] \\
= & - N^{-1/2}[A + B(U_{N,\alpha/2}^2 - 1)] + O(N^{-1}).
\end{aligned}
\tag{3.2.15}
$$

Similarly, $-N^{-1/2}[A + B((L_u^{cf})^2 - 1)] = -N^{-1/2}[A + B(U_{N,1-\alpha/2}^2 - 1)] + O(N^{-1})$. Then apply Taylor expansion to $\Phi(L_u^{cf})$, $\phi(L_u^{cf})$ at $U_{N,1-\alpha/2}$ and to $\Phi(L_l^{cf})$, $\phi(L_l^{cf})$ at $U_{N,\alpha/2}$

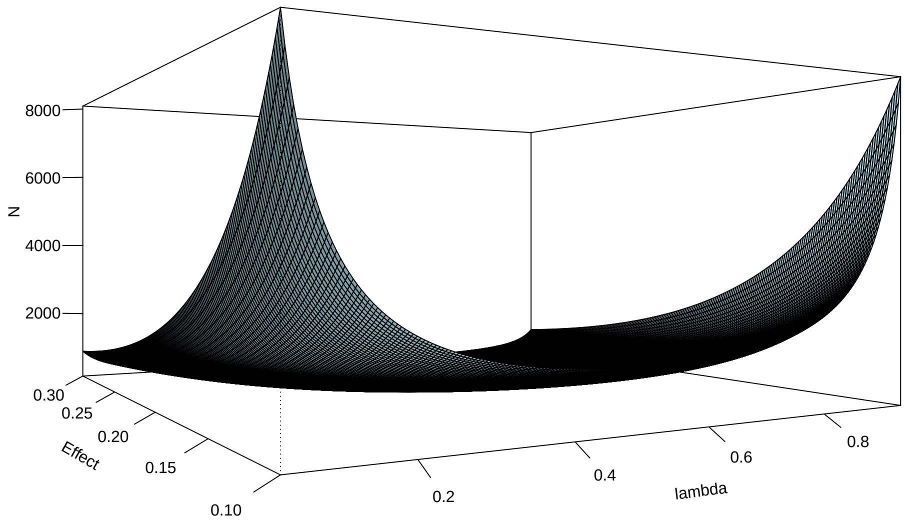

Figure 3.1: *Upper bound of total sample size N for local alternative. The upper bound is a function of the effect size $= \frac{\delta}{\sigma}$ and λ_N.*

correspondingly, we have

$$\begin{aligned}
\Phi(L_u^{cf}) &= \Phi(U_{N,1-\alpha/2}) + \phi(U_{N,1-\alpha/2})\Delta_{N,1-\alpha/2} + O(\Delta^2_{N,1-\alpha/2}),\\
\Phi(L_l^{cf}) &= \Phi(U_{N,\alpha/2}) + \phi(U_{N,\alpha/2})\Delta_{N,\alpha/2} + O(\Delta^2_{N,\alpha/2}),\\
\phi(L_u^{cf}) &= \phi(U_{N,1-\alpha/2}) + \phi^{'}(U_{N,1-\alpha/2})\Delta_{N,1-\alpha/2} + O(\Delta^2_{N,1-\alpha/2}),\\
\phi(L_l^{cf}) &= \phi(U_{N,\alpha/2}) + \phi^{'}(U_{N,\alpha/2})\Delta_{N,\alpha/2} + O(\Delta^2_{N,\alpha/2}).
\end{aligned} \tag{3.2.16}$$

Then (3.2.14) can be further calculated as

$$
\begin{aligned}
&\text{power of the two-sided TCF test} = 1 - F_T^{(H_a)}(t^{cf}_{1-\alpha/2}) + F_T^{(H_a)}(t^{cf}_{\alpha/2}) + O(N^{-min(1,r+1/2)}) \\
&= 1 - \Phi(U_{N,1-\alpha/2}) - \phi(U_{N,1-\alpha/2})\Delta_{N,\alpha/2} + \Phi(U_{N,\alpha/2}) + \phi(U_{N,\alpha/2})\Delta_{N,\alpha/2} + O(N^{-1}) - \\
&\left(N^{-1/2}[A + B(U^2_{N,1-\alpha/2} - 1)] + O(N^{-1})\right)\left(\phi(U_{N,1-\alpha/2}) + \phi^{'}(U_{N,1-\alpha/2})\Delta_{N,1-\alpha/2} + O(N^{-1})\right) \\
&+ \left(N^{-1/2}[A + B(U^2_{N,\alpha/2} - 1)] + O(N^{-1})\right)\left(\phi(U_{N,\alpha/2}) + \phi^{'}(U_{N,\alpha/2})\Delta_{N,\alpha/2} + O(N^{-1})\right) \\
&- Q(L^{cf}_u) + Q(L^{cf}_l) + O(N^{-min(1,r+1/2)}) \\
&= 1 - \Phi(U_{N,1-\alpha/2}) - N^{-1/2}[A + B(U^2_{N,1-\alpha/2} - 1)]\phi(U_{N,1-\alpha/2}) - Q(L^{cf}_u) + Q(L^{cf}_l) \\
&+ \Phi(U_{N,\alpha/2}) + N^{-1/2}[A + B(U^2_{N,\alpha/2} - 1)]\phi(U_{N,\alpha/2}) + L_{N,\gamma_1,\gamma_2,\lambda} + O(N^{-min(1,r+1/2)}),
\end{aligned}
\tag{3.2.17}
$$

where $L_{N,\gamma_1,\gamma_2,\lambda}$ has the form

$$
\begin{aligned}
L_{(N,\gamma_1,\gamma_2,\lambda)} =& - \phi(U_{N,1-\alpha/2})\Delta_{N,\alpha/2} + \phi(U_{N,\alpha/2})\Delta_{N,\alpha/2} + O(N^{-1}) \\
& - (N^{-1/2}[A + B(U^2_{N,1-\alpha/2} - 1)]\phi^{'}(U_{N,1-\alpha/2})\Delta_{N,1-\alpha/2} \\
& + (N^{-1/2}[A + B(U^2_{N,\alpha/2} - 1)]\phi^{'}(U_{N,\alpha/2})\Delta_{N,\alpha/2} \\
=& - \phi(U_{N,1-\alpha/2})\Delta_{N,\alpha/2} + \phi(U_{N,\alpha/2})\Delta_{N,\alpha/2} + O(N^{-1}) \\
=& \Delta_{N,\alpha/2}[\phi(U_{N,\alpha/2}) - \phi(U_{N,1-\alpha/2})] + O(N^{-1})
\end{aligned}
\tag{3.2.18}
$$

Note that, the approximated power of the same test based on standard normal approximation is:

$$
\begin{aligned}
&\text{power of the two-sided TN test} = 1 - F_T^{(H_a)}(U_{N,1-\alpha/2}) + F_T^{(H_a)}(U_{N,\alpha/2}) + O(N^{-min(1,r+1/2)}) \\
&= 1 - \Phi(U_{N,1-\alpha/2}) - N^{-1/2}[A + B(U^2_{N,1-\alpha/2} - 1)]\phi(U_{N,1-\alpha/2}) - Q(U_{N,1-\alpha/2}) + Q(U_{N,\alpha/2}) \\
&+ \Phi(U_{N,\alpha/2}) + N^{-1/2}[A + B(U^2_{N,\alpha/2} - 1)]\phi(U_{N,\alpha/2}) + O(N^{-min(1,r+1/2)}),
\end{aligned}
\tag{3.2.19}
$$

By comparing the two power functions in (3.2.17) and (3.2.19), we can prove the following result.

Corollary 3.2.5. *For the two-sided tests at level α,*

$$
\textit{Power of TCF test - Power of TN test} = L_{N,\gamma_1,\gamma_2,\lambda} + O(N^{-min(1,r+1/2)}). \tag{3.2.20}
$$

The proof of Corollary 3.2.5 is given in Appendix A.3. Clearly, the power of the test from Cornish Fisher expansion in (3.2.17) will be bigger than the power of the test based on normal approximation in (3.2.19) if $L_{(N,\gamma_1,\gamma_2,\lambda)}$ in (3.2.18) is bigger than zero.

Corollary 3.2.6. *The two-sided TCF test at level α is more powerful than the TN test, if and only if the following inequality holds*

$$(8 + c_\alpha - \frac{2}{\lambda})\gamma_1 > (8 + c_\alpha - \frac{2}{1-\lambda})\gamma_2, \tag{3.2.21}$$

where $c_\alpha = \frac{6}{z^2_{\alpha/2}-1}$.

The proof is given in Appendix A.5. Clearly the sign of $L_{N,\gamma_1,\gamma_2,\lambda}$ depends on the values of λ, γ_1 and γ_2. In real practice, the values of γ_1 and γ_2 are determined by the population. Then the value of $\lambda = \lambda_N + O(N^{-r}) = n_1/N + O(N^{-r})$ becomes the main factor to control the sign of $L_{N,\gamma_1,\gamma_2,\lambda}$. With a fixed value of γ_1 and γ_2, we can further rewrite the equation (3.2.21) into

$$\begin{aligned}&(8 + c_\alpha - \frac{2}{\lambda})\gamma_1 > (8 + c_\alpha - \frac{2}{1-\lambda})\gamma_2\\ \Leftrightarrow\ &(8 + c_\alpha)(\gamma_1 - \gamma_2)\lambda^2 + [(8 + c_\alpha)(\gamma_2 - \gamma_1) - 2(\gamma_1 + \gamma_2)]\lambda + 2\gamma_1 < 0. \end{aligned}\tag{3.2.22}$$

The left side of inequality (3.2.22) always has roots since

$$\begin{aligned}&[(8 + c)(\gamma_2 - \gamma_1) - 2(\gamma_1 + \gamma_2)]^2 - 4(8 + c)(\gamma_1 - \gamma_2)(2\gamma_1)\\ =&(c + 6)^2\gamma_1^2 + (c + 6)^2\gamma_2^2 - (56 + 24c + 2c^2)\gamma_1\gamma_2\\ =&[(c + 6)\gamma_1 - (c + 6)\gamma_2]^2 + 2(c + 6)^2\gamma_1\gamma_2 - (56 + 24c + 2c^2)\gamma_1\gamma_2\\ =&[(c + 6)\gamma_1 - (c + 6)\gamma_2]^2 + 16\gamma_1\gamma_2 \geq 0,\end{aligned}$$

with equality holds iff $\gamma_1 = \gamma_2 = 0$. Suppose ω_1 and ω_2 are the two roots of the left side of (3.2.22), without loss of generality let $\omega_1 \leq \omega_2$. Then we have the following solutions of λ for $L_{N,\gamma_1,\gamma_2,\lambda} > 0$:

- if $\gamma_1 > \gamma_2$ then $\omega_1 \leq \lambda \leq \omega_2$;

- if $\gamma_1 < \gamma_2$ then $\lambda \leq \omega_1$ or $\lambda \geq \omega_2$.

To demonstrate the relationship between λ and $L_{(N,\gamma_1,\gamma_2,\lambda)}$, consider one specific example. Suppose population 1 follows Gamma distribution and population 2 follows normal distribution. Let the two populations have common population variance σ^2 and population skewness $\gamma_1 > 0$ and $\gamma_2 = 0$ respectively. In addition, the total sample size $N = 40$ and the true two population mean difference $\mu_1 - \mu_2$ departs from H_0 in the amount of δ that satisfies $\delta/\sigma = 0.3$. With $\alpha = 0.05$, the TCF and TN tests were applied to test the two population mean difference. Since $\gamma_1 > \gamma_2$, by solving the inequations (3.2.22), the solutions can be obtained as $\omega_1 < \lambda < \omega_2$, where $\omega_1 = \frac{2}{8+c} = 0.1978$ and $\omega_2 = 1$. The power gains of TCF over TN test (i.e. $L_{(N,\gamma_1,\gamma_2,\lambda)}$) for $\gamma_1 = 2$ and $\gamma_2 = 6$ are shown in Figure 3.2.

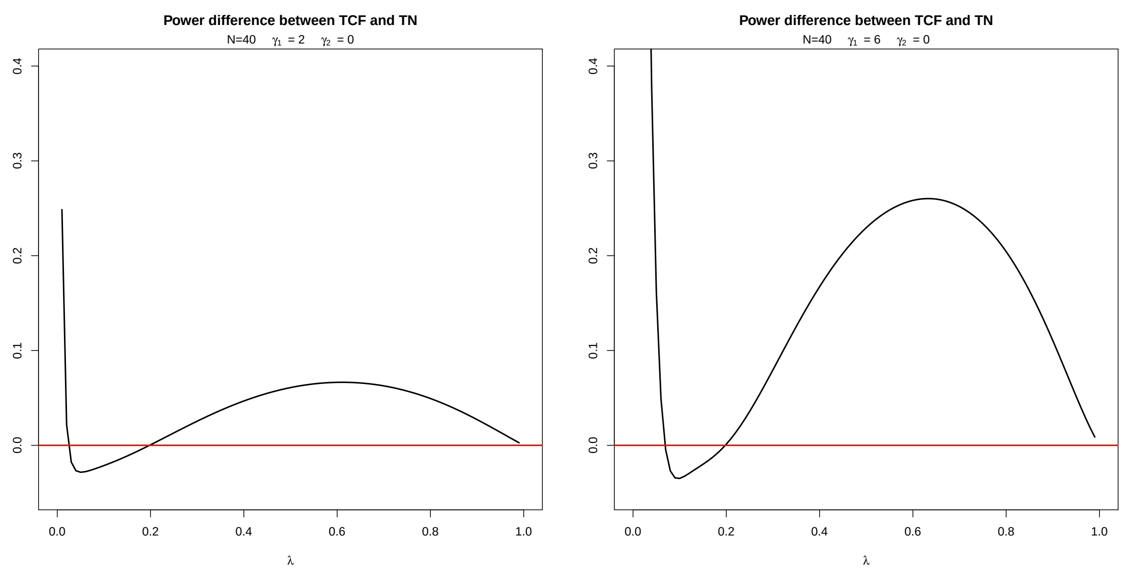

Figure 3.2: *Power difference between TCF test and pooled two-sample t-test*

From Figure 3.2, we can see that the power of TCF test is bigger than the pooled two-sample t-test when the value of λ is between 0.1978 and 1, which is the same solution from solving the inequations (3.2.22). In majority of real applications, the two sample sizes satisfy this condition of $0.19 < \lambda < 1$, which promises a higher power of TCF test than the pooled two-sample t-test. Note that we should ignore the phenomenon around $\lambda = 0$ in the plot. This is because the Edgeworth expansion theory was established for fixed $\lambda > 0$.

Discussion: the difference of power relies on the value γ_1, γ_2 critically. In practice, γ_1, γ_2

are unknown. We would need an estimate of them in order to see the difference in power. With larger sample size, the estimate $\hat{\gamma}_i$ is generally close to γ_1 (in the order of $O(N^{-1/2})$). But with small sample size, the estimated $\hat{\gamma}_i$ may be far from the γ_1. In this case the actual power difference from the tests can only achieve $L_{(N,\hat{\gamma}_1,\hat{\gamma}_2,\lambda_N)}$. For example, $N = 40$, $\lambda = 0.6$, (which gives $n_1 = 24$, $n_2 = 16$), $\gamma_1 = 6$, $\gamma_2 = 0$. In this setting with data from gamma distribution for population 1 and normal distribution for population 2, the average estimate of $\hat{\gamma}_1$ from 10000 runs is only 3.17. However the true parameter is $\gamma_1 = 6$. Even though $\hat{\gamma}_2$ is close to γ_2, the power difference is $L_{(N,\hat{\gamma}_1,\hat{\gamma}_2,\lambda_N)}$ which is around 0.13. If we had used $L_{(N,\gamma_1,\gamma_2,\lambda)}$, we would expect the difference of power to be 0.258. This would be unrealistic since the $\hat{\gamma}_1$ estimate used in the test did not get close enough to 6.

From this section, we know that the power difference between the two-sample TCF test and the pooled two-sample t-test depends on $L_{N,\gamma_1,\gamma_2,\lambda}$. Under local alternative, we can always arrange the λ_N to achieve a higher power of TCF test in real practice. In addition, the actual power increment depends on the estimation accuracy of the population skewness λ_i.

3.2.3.2 The difference of sample sizes needed to achieve certain power for TCF and TN tests

In Section 3.2.3.1, the theoretical power the of two-sample TCF test and TN test can be calculated based on the equations (3.2.17) and (3.2.19) respectively. Thus, with a given level of significance α and the desired power, we can compute the minimum sample size of two-sample TCF test and TN test by solving these two equations. For example, with given population parameters $(\gamma_1, \gamma_2, \lambda)$ and desired power, we can plug these values into the power equations (3.2.17) and (3.2.19) to solve for the total sample size N:

$$\text{power of the two-sided TCF test} = \text{desired power} \tag{3.2.23}$$

$$\text{power of the two-sided TN test} = \text{desired power.} \tag{3.2.24}$$

The equation is a nonlinear function of the total sample size N. The solution can be obtained numerically using known root find algorithms. Some examples are bisection method, secant method, Newton's method, fixed point iteration method, etc. We provide an R function to solve for the total sample size.

From Section 3.2.3, we know that the power of the two-sided TCF test is bigger than the power of the two-sided TN test if the condition in Corollary 3.2.6 holds. In this case, the two-sided TCF test will need smaller sample size than the two-sided TN test to achieve the same power under skewness. In this section, we will illustrate this property by an example.

Suppose population 1 follows gamma distribution and population 2 follows normal distribution, with common population variance σ^2. Suppose the population skewness is γ_1 and $\gamma_2 = 0$ respectively. In addition, the true two population mean difference $\mu_1 - \mu_2$ departs from H_0 in the amount of δ that satisfies $\delta/\sigma = 0.3$. Take $\lambda = 0.6$, that is $n_1 = 0.6N$. With $\alpha = 0.05$, we want to use TCF and TN tests to test the two population mean difference. Then the total sample sizes needed for each test to achieve power 0.8 are shown in Table 3.2.

γ_1	N_{TCF}	N_{TN}	N_{diff}	$\Delta_{N,\alpha/2}$
0	363	363	0	0.00
1	355	366	11	-0.04
2	347	369	22	-0.08
3	339	371	32	-0.13
4	331	374	43	-0.17
5	323	377	54	-0.22
6	314	380	66	-0.27
7	306	383	77	-0.32
8	297	386	89	-0.36
9	290	389	99	-0.42
10	281	392	111	-0.47

Table 3.2: *Sample size to achieve power=0.8 for TCF and TN tests*

In Table 3.2, γ_1 is the population 1 skewness that increases from 0 to 10. N_{TCF} and N_{TN} are the sample sizes needed to achieve 0.8 power for two-sample TCF and TN tests respectively. $N_{diff} = N_{TN} - N_{TCF}$ is the sample size difference between two-sample TCF and TN tests. $\Delta_{N,\alpha/2} = L_l^{cf} - U_{N,\alpha/2} = -N^{-1/2}[A + B(z_\alpha^2 - 1)]$ equals the amount of percentile correction based on Cornish Fisher expansion in equation (3.2.1) from Corollary 3.2.1. The amount of percentile correction $|\Delta_{N,\alpha/2}|$ gets bigger as the skewness γ_1 becomes larger. Clearly, the TN test needs more samples to achieve the same power 0.8 than the TCF test when $\gamma_1 > 0$. In addition, the sample size difference N_{diff} increases quickly as γ_1 increases.

Note that the sample size calculation of these two tests is based on the population skewness γ_1 and γ_2. In practice, the values of population skewness are unknown. We have to estimate the population skewness by sample skewness $\hat{\gamma}_1$ and $\hat{\gamma}_2$. Thus, the computation of N_{TCF} and N_{TN} relies on the estimation accuracy of γ_1 and γ_2 critically. In this case, the sample sized needed is the $max(N_\alpha, N_\beta)$, where N_α is the sample size needed to give satisfactory estimate of the population skewness and N_β refers to the sample size calculated from solving equations (3.2.23) and (3.2.24).

Recall that, if the condition in Corollary 3.2.6 holds, the power of the two-sided TCF test is higher than the power of the two-sided T test under skewness. This power increment of the test is a crucial factor to increase the test accuracy. Furthermore, when the sample is limited or very expensive to be obtained, the two-sample TCF test can not only achieve a higher power of the test but also save money in collecting the data comparing with the two-sample TN test.

3.2.4 Type I error rate of the one-sided test based on Cornish Fisher expansion

As is known in the Edgeworth expansion literature that one-sided tests may behave differently from two-sided tests. We also present the order of approximation to type I error rate for the one-sided TCF test with rejection regions given in (3.2.3) and (3.2.4) for upper and lower-tailed tests respectively under the same two-sample setting.

3.2.4.1 Type I error rate of the one-sided upper-tailed TCF test

Recall that the distribution of $F_T^{(1)}(x)$ in Theorem 3.1.3 takes the following form

$$F_T^{(1)}(x) = P(T \leq x) = \Phi(x) + N^{-1/2}[A + B(x^2 - 1)]\phi(x) + O(N^{-min(1,r+1/2)}).$$

Then the type I error rate of the one-side upper-tailed TCF test can be expressed as:

$$\begin{aligned} & P(\text{type I error of the one-sided upper TCF test}) \\ =& 1 - F_T^{(1)}(t_{1-\alpha}^{cf}) + O(N^{-min(1,r+1/2)}) \qquad (3.2.25) \\ =& 1 - \Phi(t_{1-\alpha}^{cf}) - N^{-1/2}[A + B((t_{1-\alpha}^{cf})^2 - 1)]\phi(t_{1-\alpha}^{cf}) + O(N^{-min(1,r+1/2)}). \end{aligned}$$

Earlier from (3.2.8), we can approve:

$$-N^{-1/2}[A + B((t_{1-\alpha}^{cf})^2 - 1)] = \Delta_{N,1-\alpha} + O(N^{-1}), \tag{3.2.26}$$

where $t_{1-\alpha}^{cf} = z_{1-\alpha} + \Delta_{N,1-\alpha}$ and $\Delta_{N,1-\alpha} = -N^{-1/2}[A + B(z_{1-\alpha}^2 - 1)]$. From previous results, we can also show:

$$\begin{aligned} \Phi(t_{1-\alpha}^{cf}) &= \Phi(z_{1-\alpha}) + \phi(z_{1-\alpha})\Delta_{N,1-\alpha} + O(\Delta_{N,1-\alpha}^2), \\ \phi(t_{1-\alpha}^{cf}) &= \phi(z_{1-\alpha}) + \phi'(z_{1-\alpha})\Delta_{N,1-\alpha} + O(\Delta_{N,1-\alpha}^2). \end{aligned} \tag{3.2.27}$$

Put (3.2.26) and (3.2.27) into (3.2.25), we have

$$\begin{aligned} & P(\text{type I error of the one-sided upper TCF test }) \\ = \;& 1 - \Phi(t_{1-\alpha}^{cf}) - N^{-1/2}[A + B((t_{1-\alpha}^{cf})^2 - 1)]\phi(t_{1-\alpha/2}^{cf}) + O(N^{-min(1,r+1/2)}) \\ = \;& \alpha - \phi(z_{1-\alpha})\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2) \\ & + [\Delta_{N,\alpha} + O(N^{-1})][\phi(z_{1-\alpha}) + \phi'(z_{1-\alpha})\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2)] \\ & + O(N^{-min(1,r+1/2)}) \\ = \;& \alpha + \Delta_{N,\alpha}\phi(z_{1-\alpha}) - \Delta_{N,\alpha}\phi(z_{\alpha}) + O(N^{-1}) + O(N^{-min(1,r+1/2)}) \\ = \;& \alpha + O(N^{-min(1,r+1/2)}), \qquad (3.2.28) \end{aligned}$$

Then we have the follow theorem,

Theorem 3.2.7. *Under standard regularity conditions, when H_0 is true, the theoretical type I error rate of the upper-tailed TCF test, with level of significance α is*

$$P(T > \hat{t}_{1-\alpha}^{cf}) = \alpha + O(N^{-min(1,r+1/2)}).$$

3.2.4.2 Type I error rate of the one-sided lower-tailed TCF test

Following the same procedures in Section 3.2.4.1, the type I error rate of the one-sided lower-tailed TCF test is:

$$\begin{aligned}
&P(\text{type I error of the one-sided lower TCF test})\\
=&F_T^{(1)}(t_\alpha^{cf}) + O(N^{-min(1,r+1/2)}) \qquad (3.2.29)\\
=&\Phi(t_\alpha^{cf}) + N^{-1/2}[A + B((t_\alpha^{cf})^2 - 1)]\phi(t_\alpha^{cf}) + O(N^{-min(1,r+1/2)}).
\end{aligned}$$

Again from (3.2.8), we can show:

$$-N^{-1/2}[A + B((t_\alpha^{cf})^2 - 1)] = \Delta_{N,\alpha} + O(N^{-1}), \qquad (3.2.30)$$

where $t_\alpha^{cf} = z_\alpha + \Delta_{N,\alpha}$ and $\Delta_{N,\alpha} = -N^{-1/2}[A + B(z_\alpha^2 - 1)]$. Furthermore,

$$\begin{aligned}
\Phi(t_\alpha^{cf}) &= \Phi(z_\alpha) + \phi(z_\alpha)\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2),\\
\phi(t_\alpha^{cf}) &= \phi(z_\alpha) + \phi^{'}(z_\alpha)\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2).
\end{aligned} \qquad (3.2.31)$$

Put (3.2.30) and (3.2.31) into (3.2.29), we have

$$\begin{aligned}
& P(\text{type I error of the one-sided lower TCF test })\\
= \quad & \Phi(t_\alpha^{cf}) + N^{-1/2}[A + B((t_\alpha^{cf})^2 - 1)]\phi(t_{\alpha/2}^{cf}) + O(N^{-min(1,r+1/2)})\\
= \quad & \alpha + \phi(z_\alpha)\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2)\\
& -[\Delta_{N,\alpha} + O(N^{-1})][\phi(z_\alpha) + \phi^{'}(z_\alpha)\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2)]\\
& +O(N^{-min(1,r+1/2)})\\
= \quad & \alpha + \Delta_{N,\alpha}\phi(z_\alpha) - \Delta_{N,\alpha}\phi(z_\alpha) + O(N^{-1}) + O(N^{-min(1,r+1/2)})\\
= \quad & \alpha + O(N^{-min(1,r+1/2)}), \qquad (3.2.32)
\end{aligned}$$

Then we have the follow theorem,

Theorem 3.2.8. *Under standard regularity conditions, when H_0 is true, the theoretical type I error rate of the lower-tailed TCF test, at significance level α is*

$$P(T < \hat{t}_\alpha^{cf}) = \alpha + O(N^{-min(1,r+1/2)}).$$

For the two-sample TCF test, both one-sided upper and lower-tailed tests have the same type I error rate approximation accuracy of $\alpha + O(N^{-min(1,r+1/2)})$. Moreover, we can show that the approximated type I error rates of the same one-sided tests based on normal approximation are

$$\begin{aligned}
& P(\text{type I error of the one-sided lower-tailed TN test }) \\
= \quad & \Phi(z^{cf}_{\alpha}) + N^{-1/2}[A + B((z^{cf}_{\alpha})^2 - 1)]\phi(z^{cf}_{\alpha/2}) + O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha + O(N^{-1/2}). \qquad (3.2.33)
\end{aligned}$$

$$\begin{aligned}
& P(\text{type I error of the one-sided upper-tailed TN test }) \\
= \quad & 1 - \Phi(z^{cf}_{1-\alpha}) - N^{-1/2}[A + B((z^{cf}_{1-\alpha})^2 - 1)]\phi(z^{cf}_{1-\alpha/2}) + O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha + O(N^{-1/2}). \qquad (3.2.34)
\end{aligned}$$

According to the above results, the type I error rate accuracy for the one-sided TCF test has higher approximation accuracy than the one-sided TN test, since $min(1, r + 1/2) > 1/2$ with $r > 0$ as defined in Corollary 3.1.1.

3.2.5 Power of the one-sided TCF test under local alternative hypothesis

In this section, we derive the power function of the one-sided upper and lower tail TCF and TN tests under local alternative hypothesis. We also provide the detailed condition under which the theoretical power of the one-sample TCF test is higher than the one-sample TN test.

3.2.5.1 Power of the one-sided upper-tailed TCF test under local alternative hypothesis

According to the previous two-sided results in Section 3.2.3.1, now we consider data generated under $H_a : \mu_1 - \mu_2 > \mu^0_1 - \mu^0_2$. When H_a is true, the theoretical power of the one-sided upper-tailed TCF test equals the probability of rejecting the null hypothesis:

$$P_{H_a}\left(T \geq \hat{t}^{cf}_{1-\alpha}\right),$$

where $\hat{t}^{cf}_{1-\alpha}$ is the estimated $(1-\alpha)^{th}$ percentile from distribution $F^{(1)}_T(x)$ following similar formula as (3.2.5). Denote

$$\begin{aligned}
\hat{L}^{cf}_{1-\alpha} &= \hat{t}^{cf}_{1-\alpha} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} \\
&= z_{1-\alpha} + \hat{\Delta}_{N,1-\alpha} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} \\
&= U_{N,1-\alpha} + \hat{\Delta}_{N,1-\alpha} \\
&= U_{N,1-\alpha} + \Delta_{N,1-\alpha} + O_p(N^{-min(1,r+1/2)}) \\
&= L^{cf}_{1-\alpha} + O_p(N^{-min(1,r+1/2)}),
\end{aligned}$$

where $U_{N,1-\alpha} = z_{1-\alpha} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$, and $L^{cf}_{1-\alpha} = U_{N,1-\alpha} + \Delta_{N,1-\alpha}$. We have

$$\hat{L}^{cf}_{1-\alpha} = L^{cf}_{1-\alpha} + O_p(N^{-min(1,r+1/2)}).$$

Then the power can be expressed as,

$$\begin{aligned}
& P_{H_a}\left(T \geq \hat{L}^{cf}_{1-\alpha}\right) \\
= \; & 1 - F^{(1)}_T(L^{cf}_{1-\alpha}) - Q(L^{cf}_{1-\alpha}) + O(N^{-min(1,r+1/2)}) \\
= \; & 1 - \Phi(L^{cf}_{1-\alpha}) - N^{-1/2}[A + B((L^{cf}_{1-\alpha})^2 - 1)]\phi(L^{cf}_{1-\alpha}) \\
& -Q(L^{cf}_{1-\alpha}) + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{3.2.35}$$

We have shown in equations (3.2.15) and (3.2.16) that

$$-N^{-1/2}[A + B((L^{cf}_{1-\alpha})^2 - 1)] = -N^{-1/2}[A + B(U^2_{N,1-\alpha} - 1)] + O(N^{-1}); \tag{3.2.36}$$

$$\begin{aligned}
\Phi(L^{cf}_{1-\alpha}) &= \Phi(U_{N,1-\alpha}) + \phi(U_{N,1-\alpha})\Delta_{N,1-\alpha} + O(\Delta^2_{N,1-\alpha}) \\
\phi(L^{cf}_{1-\alpha}) &= \phi(U_{N,1-\alpha}) + \phi^{'}(U_{N,1-\alpha})\Delta_{N,1-\alpha} + O(\Delta^2_{N,1-\alpha}).
\end{aligned} \tag{3.2.37}$$

Putting equations (3.2.36) and (3.2.37) back into (3.2.35), we have the

$$\begin{aligned}
&\text{power of one-sided upper tail TCF test} = 1 - F_T^{(H_a)}(L_{1-\alpha}^{cf}) + O(N^{-min(1,r+1/2)})\\
&= 1 - \Phi(U_{N,1-\alpha}) - \phi(U_{N,1-\alpha})\Delta_{N,1-\alpha} + O(\Delta_{N,1-\alpha}^2)\\
&- \left(N^{-1/2}[A + B(U_{N,1-\alpha}^2 - 1)] + O(N^{-1})\right)\left(\phi(U_{N,1-\alpha}) + \phi^{'}(U_{N,1-\alpha})\Delta_{N,1-\alpha}\right.\\
&\left.+ O(\Delta_{N,1-\alpha}^2)\right) - Q(L_{1-\alpha}^{cf}) + O(N^{-min(1,r+1/2)})\\
&= 1 - \Phi(U_{N,1-\alpha}) - N^{-1/2}[A + B(U_{N,1-\alpha}^2 - 1)]\phi(U_{N,1-\alpha}) - Q(L_{1-\alpha}^{cf})\\
&+ L_{N,\gamma_1,\gamma_2,\lambda}^u + O(N^{-min(1,r+1/2)}),
\end{aligned} \tag{3.2.38}$$

where $L_{N,\gamma_1,\gamma_2,\lambda}^u$ has the form

$$\begin{aligned}
&L_{(N,\gamma_1,\gamma_2,\lambda)}^u\\
&= -\phi(U_{N,1-\alpha})\Delta_{N,1-\alpha} - N^{-1/2}[A + B(U_{N,1-\alpha}^2 - 1)]\phi^{'}(U_{N,1-\alpha})\Delta_{N,1-\alpha}\\
&= -\phi(U_{N,1-\alpha})\Delta_{N,1-\alpha} + O(N^{-1})
\end{aligned} \tag{3.2.39}$$

The approximated power of the same T statistic based on standard normal approximation is:

$$\begin{aligned}
&\text{power of the one-sided upper tail TN test} = 1 - F_T^{(H_a)}(U_{N,1-\alpha}) + O(N^{-min(1,r+1/2)})\\
&= 1 - \Phi(U_{N,1-\alpha}) - N^{-1/2}[A + B(U_{N,1-\alpha}^2 - 1)]\phi(U_{N,1-\alpha}) - Q(U_{N,1-\alpha})\\
&+ O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{3.2.40}$$

From (3.2.38) and (3.2.40) the difference in power relies on $L_{N,\gamma_1,\gamma_2,\lambda}^u$. We write this result in the following Corollary.

Corollary 3.2.9. *For the one-sided upper-tailed tests at level α,*

$$\textit{Power of TCF test - Power of TN test} = L_{N,\gamma_1,\gamma_2,\lambda}^u + O(N^{-min(1,r+1/2)}). \tag{3.2.41}$$

The prove is similar to Corollary (3.2.5). Then we have the following result:

Corollary 3.2.10. *The one-sided upper-tailed TCF test at level α is more powerful than the TN test, if and only if the following inequality holds*

$$(8 + c_\alpha - \frac{2}{\lambda})\gamma_1 > (8 + c_\alpha - \frac{2}{1-\lambda})\gamma_2, \tag{3.2.42}$$

where $c_\alpha = \frac{6}{z_\alpha^2 - 1}$.

Comparing Corollary (3.2.10) with Corollary (3.2.6), the inequality conditions in (3.2.42) and (3.2.21) are the same. Therefore, the discussion on the sign of $L_{N,\gamma_1,\gamma_2,\lambda}$ in Section 3.2.3.1 still holds for the sign of $L^u_{N,\gamma_1,\gamma_2,\lambda}$. With given estimates of two population skewness γ_1 and γ_2, we can solve (3.2.42) to get a solution of λ as follows,

- if $\gamma_1 > \gamma_2$ then $\omega_1 \leq \lambda \leq \omega_2$;
- if $\gamma_1 < \gamma_2$ then $\lambda \leq \omega_1$ or $\lambda \geq \omega_2$,

where ω_1 and ω_2 are the two roots of the left side of (3.2.22), here without loss of generality assume $\omega_1 \leq \omega_2$.

3.2.5.2 Power of the lower-tailed TCF test under local alternative hypothesis

Following the same procedure in Section 3.2.5.1, now consider data generated under H_a : $\mu_1 - \mu_2 < \mu_1^0 - \mu_2^0$. When H_a is true, the theoretical power of the one-sided lower tail TCF test is:

$$P_{H_a}\left(T \leq \hat{t}^{cf}_\alpha\right),$$

where $\hat{t}^{cf}_\alpha$ is the estimated α^{th} percentile from distribution $F_T^{(1)}(x)$ using formula similar to (3.2.5). Denote

$$\begin{aligned}
\hat{L}^{cf}_\alpha &= \hat{t}^{cf}_\alpha - \delta/\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})} \\
&= z_\alpha + \hat{\Delta}_{N,\alpha} - \delta/\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})} \\
&= U_{N,\alpha} + \hat{\Delta}_{N,\alpha} \\
&= U_{N,\alpha} + \Delta_{N,\alpha} + O_p(N^{-min(1,r+1/2)}) \\
&= L^{cf}_\alpha + O_p(N^{-min(1,r+1/2)}),
\end{aligned}$$

where $U_{N,\alpha} = z_\alpha - \delta/\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}$, and $L^{cf}_\alpha = U_{N,\alpha} + \Delta_{N,\alpha}$. We have

$$\hat{L}^{cf}_\alpha = L^{cf}_\alpha + O_p(N^{-min(1,r+1/2)}).$$

Then the power can be expressed as

$$\begin{aligned}
& P_{H_a}\left(T \le \hat{L}_\alpha^{cf}\right) \\
= \quad & F_T^{(1)}(L_\alpha^{cf}) + Q(L_\alpha^{cf}) + O(N^{-min(1,r+1/2)}) \\
= \quad & \Phi(L_\alpha^{cf}) + N^{-1/2}[A + B((L_\alpha^{cf})^2 - 1)]\phi(L_\alpha^{cf}) + Q(L_\alpha^{cf}) + O(N^{-min(1,r+\frac{1}{2})})
\end{aligned} \tag{3.2.43}$$

We have shown in equations (3.2.15) and (3.2.16) that

$$-N^{-1/2}[A + B((L_\alpha^{cf})^2 - 1)] = -N^{-1/2}[A + B(U_{N,\alpha}^2 - 1)] + O(N^{-1}); \tag{3.2.44}$$

$$\begin{aligned}
\Phi(L_\alpha^{cf}) &= \Phi(U_{N,\alpha}) + \phi(U_{N,\alpha})\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2) \\
\phi(L_\alpha^{cf}) &= \phi(U_{N,\alpha}) + \phi'(U_{N,\alpha})\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2).
\end{aligned} \tag{3.2.45}$$

Putting equations (3.2.44) and (3.2.45) back into (3.2.43), we have the

$$\begin{aligned}
& \text{power of one-sided lower tail TCF test} = F_T^{(H_a)}(t_\alpha^{cf}) + O(N^{-min(1,r+1/2)}) \\
& = \Phi(U_{N,\alpha}) + \phi(U_{N,\alpha})\Delta_{N,\alpha} + O(\Delta_{N,\alpha}^2) \\
& + \left(N^{-1/2}[A + B(U_{N,\alpha}^2 - 1)] + O(N^{-1})\right)(\phi(U_{N,\alpha}) + \phi'(U_{N,\alpha})\Delta_{N,\alpha} \\
& + O(\Delta_{N,\alpha}^2)) + Q(L_\alpha^{cf}) + O(N^{-min(1,r+1/2)}) \\
& = \Phi(U_{N,\alpha}) + N^{-1/2}[A + B(U_{N,\alpha}^2 - 1)]\phi(U_{N,\alpha}) + Q(L_\alpha^{cf}) \\
& + L_{N,\gamma_1,\gamma_2,\lambda}^l + O(N^{-min(1,r+1/2)}),
\end{aligned} \tag{3.2.46}$$

where $L_{N,\gamma_1,\gamma_2,\lambda}^l$ is defined as

$$\begin{aligned}
& L_{(N,\gamma_1,\gamma_2,\lambda)}^l \\
= & \phi(U_{N,\alpha})\Delta_{N,\alpha} + N^{-1/2}[A + B(U_{N,\alpha}^2 - 1)]\phi'(U_{N,\alpha})\Delta_{N,\alpha} \\
= & \phi(U_{N,\alpha})\Delta_{N,\alpha} + O(N^{-1})
\end{aligned} \tag{3.2.47}$$

The approximated power of the same T statistic based on standard normal approximation is:

$$\begin{aligned}
& \text{power of the one-sided lower tail TN test} = F_T^{(H_a)}(U_{N,\alpha}) + O(N^{-min(1,r+1/2)}) \\
& = \Phi(U_{N,\alpha}) + N^{-1/2}[A + B(U_{N,\alpha}^2 - 1)]\phi(U_{N,\alpha}) + Q(U_{N,\alpha}) + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{3.2.48}$$

Again, the power difference depends on $L_{N,\gamma_1,\gamma_2,\lambda}^l$. We state this result in the following Corollary.

Corollary 3.2.11. *For the one-sided lower-tailed tests at level α,*

$$\textit{Power of TCF test - Power of TN test} = L^l_{N,\gamma_1,\gamma_2,\lambda} + O(N^{-min(1,r+1/2)}). \qquad (3.2.49)$$

The prove is similar to Corollary (3.2.5). Then we have the following result:

Corollary 3.2.12. *The one-sided lower-tailed TCF test at level α is more powerful than the TN test, if and only if the following inequality holds*

$$(8 + c_\alpha - \frac{2}{\lambda})\gamma_1 < (8 + c_\alpha - \frac{2}{1-\lambda})\gamma_2, \qquad (3.2.50)$$

where $c_\alpha = \frac{6}{z_\alpha^2 - 1}$.

Comparing Corollary (3.2.12) with Corollary (3.2.6), the inequality in (3.2.50) defines a set that is the complement of the set given in (3.2.21). Thus, the sign of $L_{N,\gamma_1,\gamma_2,\lambda}$ in Section 3.2.3.1 is opposite of the sign of $L^u_{N,\gamma_1,\gamma_2,\lambda}$. Given the two population skewness γ_1 and γ_2, we can solve (3.2.50) to get a solution of λ as follows,

- if $\gamma_1 < \gamma_2$ then $\omega_1 \leq \lambda \leq \omega_2$;
- if $\gamma_1 > \gamma_2$ then $\lambda \leq \omega_1$ or $\lambda \geq \omega_2$,

where $\omega_1 \leq \omega_2$ are the two roots of (3.2.22).

In summary, for two-sided test the power difference between the TCF test and the pooled two-sample t-test relies on $L_{N,\gamma_1,\gamma_2,\lambda}$. Under local alternative, we can always arrange the λ_N to achieve a higher power of TCF test in real practice. In addition, the actual power gain depends on the estimation accuracy of the population skewness λ_i. For one-sided tests, the power difference depends on $L^u_{N,\gamma_1,\gamma_2,\lambda}$ for the upper-tailed test and $L^l_{N,\gamma_1,\gamma_2,\lambda}$ for the lower-tailed test respectively. The way to arrange λ_N for achieving a higher power of TCF test is also presented.

As we have seen that the sets to ensure higher power for the TCF test than the TN test mutually exclusive for the upper-tailed and lower-tailed test. Therefore, only one of the one-sided TCF test will have higher power than the TN test.

3.3 Simulation study for two-sample TCF test

3.3.1 Main purpose of the simulation study

Recall that the main purpose of this study is to improve the power of the two-sample test under skewness, when the two population mean difference is small and the total sample size is limited. Thus, the main purpose of this simulation study is to compare the type I error rate and the power of our TCF test with commonly used tests under the above scenario. We consider the following four tests, pooled two-sample t-test, test by Cornish Fisher expansion, Bootstrap-t test (Davison and Hinkley, 1997; Efron and Tibshirani, 1993) and Wilcoxon Rank-Sum Test. We apply the four tests in this simulation study on testing the population mean difference between two independent populations with equal variances. This simulation study considers on one-sided upper-tailed test, one-sided lower-tailed test and two-sided test.

3.3.2 Detailed settings of a simulation study

This simulation study has two pairs of population settings. We let the first population be Gamma or Log-normal distribution and let the second population be normal distribution. Under the null hypothesis of equal means, the settings of each population parameters are listed in Table 3.3.

	Population1	Population2	γ_1
Pair1	$Gamma(\alpha = 0.1, \beta = 0.08)$	$N(\mu_1 = 1.25, \sigma_1^2 = 15.625)$	6.3
Pair2	$Lognormal(\mu_2 = 0, \sigma_2^2 = 1)$	$N(\mu_1 = e^{1/2}, \sigma_1^2 = e^1(e^1 - 1))$	2.9

Table 3.3: *Population parameters setting*

Under the alternative hypothesis of one-sided upper-tailed and two-sided tests, a constant in the amount of 0.3σ was added to the first population mean. That is the population mean of the first population is bigger than the second population, and the mean difference is 0.3σ.

Under the alternative hypothesis of one-sided lower-tailed test, a constant in the amount of 0.3σ was subtracted from the first population mean. The reason that we choose the value 0.3σ is according to Cohen (1988), the effect size is considered as small when $\delta/\sigma = 0.2$, medium when $\delta/\sigma = 0.5$, and large when $\delta/\sigma = 0.8$. In other words, the mean difference of the two population is relatively small, so that we can check how much power these four tests have to detect a small amount of population mean difference.

As we mentioned in Chapter 2, the power of the test depends on several factors. In order to figure out the effect of the skewness on the power of the test, we have to fix the other parameters. The parameter settings of each test and population are as follows:

- Significance level $\alpha = 0.05$;
- First population sample size $n_1 = 15, 25, \ldots, 150$;
- $\lambda_N = n_1/N = 0.6$, here $N = n_1 + n_2$;
- Second population sample size $n_2 = (1 - \lambda_N)N$;
- Effect size: $\delta = (\mu_1 - \mu_2) - Hypothesized(\mu_1 - \mu_2) = 0.3\sigma$, where σ^2 is the common variance.

Note that the population 1 is a right skewed distribution either from Gamma distribution or Log-normal distribution and population 2 is symmetric. The value of λ reflects the sample size ratio. When λ is close to 1, the majority samples are from population 1. When λ is close to 0, the majority samples are from population 2. Thus, under fixed population skewness and sample size n_1, the two-sample data sets with bigger λ are easier to detect departure from H_0 than the two-sample data sets with small λ. In addition, when we increase n_1, the total sample size N increase faster than n_1 with a smaller λ.

In order to achieve a small sampling variation for the four tests, we generate $10,000$ samples for each parameter setting and each sample size. For bootstrap tests, $1,000$ bootstrap samples are resampled from each generated data set.

3.3.3 Details of the simulation study for each test

For each simulation setting, the empirical type I error rate and power of the pooled two-sample t-test, the TCF test and the Wilcoxon Rank-Sum Test are calculated as follows:

1. Generate $10,000$ data sets, each contains a sample of size n_1, n_2 from the two populations respectively, with sample of size n_1 from population one and sample of size n_2 from population two.

2. For each generated data set, conduct the pooled two sample t-test using R built-in function t.test with var.equal=T and the Wilcoxon Rank-Sum Test using R function wilcox.test. For the TCF test, compute the test statistic T defined in equation (2.1.2) as
$$T = \frac{\bar{Y_1} - \bar{Y_2} - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$
Reject H_0 if the test statistic T falls into the rejection regions given by 3.2.2, 3.2.3 and 3.2.4 for two-sided, upper-tailed and lower-tailed test, respectively.

3. Repeat this testing process for each of the $10,000$ generated data set.

4. When the alternative hypothesis is true, the two populations having different means, the proportion of times rejecting the null hypothesis out of $10,000$ generated data set is the empirical power of the tests. When the null hypothesis is true, the two populations having equal means, the proportion of times rejecting the null hypothesis out of $10,000$ generated data set is the empirical type I error rate.

The empirical type I error rate and power of the two-sided bootstrap-t test can be computed by the following steps:

1. Draw $B = 1,000$ bootstrap samples of size $N = n_1 + n_2$ with replacement from each of the $10,000$ generated data set, with sample of size n_1 from sample one and sample of size n_2 from sample two.

2. For each bootstrap sample, compute

$$T_b^* = \frac{\bar{X}_{1b}^* - \bar{X}_{2b}^* - \bar{X}_{1n} + \bar{X}_{2n}}{S_{p_b}^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\bar{X}_{1b}^*$ and $\bar{X}_{2b}^*$ are the bth bootstrap sample means of sample one and sample two respectively; $\bar{X}_{1n}$ and $\bar{X}_{2n}$ are the sample means for original sample one and original sample two respectively; $S_{p_b}^* = \sqrt{\frac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}}$ is the pooled two sample standard deviation of bth bootstrap sample; $S_i^* = \sqrt{\frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}^* - \bar{X}_{ib}^*)^2}$, $i = 1, 2$ are the bth bootstrap sample standard deviation for sample i.

3. Estimate the $\alpha/2^{th}$ percentile of test statistic T by the value $\hat{t}_{\alpha/2}$ such that

$$B^{-1}\sum_{b=1}^{B} I(T_b^* \leq \hat{t}_{\alpha/2}) = \alpha/2$$

4. The rejection region of two-sided bootstrap-t test with significance level α is then:

$$T \leq \hat{t}_{\alpha/2} \;\; or \;\; T \geq \hat{t}_{1-\alpha/2}.$$

Reject H_0 if the test statistic T falls into the rejection region above.

5. When the alternative hypothesis is true, the two populations having different means, the proportion of times rejecting the null hypothesis out of $10,000$ generated data set is the empirical power of the test. When the null hypothesis is true, the two populations having equal means, the proportion of times rejecting the null hypothesis out of $10,000$ generated data set is the empirical type I error rate.

For the one-sided bootstrap-t test, the α^{th} percentile and $(1-\alpha)^{th}$ percentile can be estimated as the same way from step 3. Then the rejection region for one-sided upper-tailed bootstrap-t test is

$$T \geq \hat{t}_{1-\alpha},$$

and the rejection region for one-sided lower-tailed bootstrap-t test is

$$T \leq \hat{t}_{\alpha}.$$

In the following section, we will present the results of a simulation study with data generated from skewed distributions for first population and normal for second population.

3.3.4 Results of simulation study and discussion

Following the data generation settings description in Section 3.3.2, we run the simulation study with the procedure in Section 3.3.3. The simulation results are shown in the following three sections.

3.3.4.1 Simulation results for two-sided test

The simulation results of the four tests from each pair of population settings are presented in Figure 3.3, Table 3.4 and Table 3.5.

Figure 3.3 shows the empirical type I error rate and power from four tests including pooled two-sample t-test denoted as "T", two-sample Bootstrap-t test denoted as "$bootstrap$", Wilcoxon Rank-Sum Test denoted as "$Wilcox$" and the Cornish Fisher expansion based two sample TCF test denoted as "CF". The top two panels in Figure 3.3 give the empirical type I error rate of the four tests with $\lambda = 0.6$. The bottom two graphes provide their corresponding empirical powers. From Figure 3.3, when $n_1 < 15$, only the two-sample Bootstrap-t test gives a type I error rate below α. The empirical type I error rate of the other three tests are all bigger than 0.1. When $n_1 > 15$, the empirical type I error rate of all four tests reduces to α except for Wilcoxon Rank-Sum Test. The empirical type I error rate of Wilcoxon Rank-Sum Test keeps increasing as n_1 getting larger. The two-sample TCF test gives a constant higher empirical power than the other three tests. The two-sample Bootstrap-t test gives the second highest power and the Wilcoxon Rank-Sum Test provides the smallest power. The numerical results of the empirical power are presented in Tables 3.4 and 3.5.

This simulation study shows that the two sample TCF test has significantly better power than the other tests in testing the two population mean difference under skewness. Comparing with the two-sample T and bootstrap t-test, the two-sample TCF test provides the same amount of type I error rate but requires fewer sample size to reach the desired power. The Wilcoxon Rank-Sum Test gives the most rejection regardless of under our H_0 or H_a. That's because the assumption of the Wilcoxon Rank-Sum Test is violated in this simulation study. From Section 2.1.2, we know that the Wilcoxon Rank-Sum Test assumes the two samples are taken from identical distribution but the two populations in this simulation follow distinct distributions.

n_1	skewness	Tests							
		T-Test		Bootstrap-t		Wilcox		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.477	0.087	0.107	0.070	0.146	0.124	0.132	0.100	0.228
25	0.435	0.081	0.134	0.069	0.199	0.132	0.156	0.083	0.271
35	0.412	0.078	0.175	0.068	0.247	0.147	0.175	0.079	0.318
45	0.385	0.076	0.220	0.066	0.306	0.163	0.200	0.074	0.369
55	0.364	0.070	0.266	0.065	0.350	0.177	0.211	0.070	0.419
65	0.350	0.067	0.310	0.063	0.401	0.188	0.237	0.066	0.466
75	0.335	0.066	0.365	0.062	0.448	0.207	0.259	0.068	0.512
85	0.322	0.065	0.416	0.060	0.497	0.213	0.282	0.063	0.562
95	0.313	0.063	0.458	0.059	0.538	0.224	0.298	0.062	0.599
105	0.302	0.061	0.503	0.057	0.580	0.230	0.323	0.060	0.638
115	0.293	0.062	0.554	0.058	0.625	0.244	0.343	0.060	0.682
125	0.286	0.061	0.590	0.060	0.660	0.258	0.364	0.062	0.712
135	0.278	0.059	0.629	0.060	0.688	0.267	0.384	0.061	0.739
145	0.271	0.058	0.663	0.059	0.718	0.277	0.403	0.060	0.762

Table 3.4: *Proportion of rejections for two-sided tests when* $1st$ *population is Gamma*

3.3.4.2 Simulation results for one-sided upper-tailed test

Figure 3.4, Table 3.6 and Table 3.7 present the simulation results of one-sided upper-tail test. In Figure 3.4, when $n_1 > 15$ all four tests maintain their type I error rate below or close to α. When it comes to the empirical power, the one-side upper-tail TCF test gives consistently higher power than the other three tests, which is similar as the results of

n_1	skewness	Tests							
		T-Test		Bootstrap-t		Wilcox		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.302	0.073	0.104	0.067	0.143	0.086	0.104	0.090	0.186
25	0.283	0.068	0.139	0.064	0.194	0.097	0.126	0.077	0.233
35	0.273	0.067	0.175	0.060	0.238	0.105	0.144	0.072	0.278
45	0.259	0.064	0.222	0.058	0.288	0.116	0.167	0.066	0.327
55	0.248	0.062	0.268	0.058	0.336	0.125	0.185	0.066	0.378
65	0.240	0.065	0.312	0.059	0.381	0.135	0.203	0.066	0.415
75	0.231	0.063	0.368	0.057	0.429	0.147	0.230	0.066	0.468
85	0.223	0.060	0.416	0.058	0.476	0.155	0.252	0.061	0.515
95	0.219	0.058	0.454	0.056	0.514	0.164	0.272	0.061	0.554
105	0.213	0.057	0.501	0.058	0.558	0.173	0.294	0.061	0.596
115	0.207	0.057	0.547	0.056	0.599	0.185	0.321	0.059	0.636
125	0.203	0.058	0.586	0.058	0.634	0.196	0.341	0.061	0.670
135	0.199	0.054	0.622	0.056	0.668	0.200	0.359	0.060	0.703
145	0.195	0.054	0.655	0.056	0.696	0.211	0.378	0.060	0.729

Table 3.5: *Proportion of rejections for two-sided tests when* 1*st population is Lognormal*

empirical power in two-sided tests from Figure 3.3. The numerical values of the type I error and power are given in Tables 3.6 and 3.7.

For one-sided upper-tailed test, all four tests control their type I error rate well. The TCF test still provides higher empirical power than the other three tests under skewed populations.

3.3.4.3 Simulation results for two-sided lower-tailed test

The results for lower-tailed tests are given in Figure 3.5. They are quite different from the results for two-sided or upper-tailed test. The Wilcoxon Rank-Sum Test failed to maintain

n_1	skewness	Tests							
		T-Test		Bootstrap-t		Wilcox		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.477	0.034	0.161	0.061	0.212	0.043	0.174	0.074	0.288
25	0.435	0.030	0.222	0.058	0.289	0.033	0.205	0.061	0.351
35	0.412	0.029	0.280	0.056	0.344	0.029	0.237	0.056	0.410
45	0.385	0.028	0.343	0.056	0.410	0.022	0.267	0.058	0.470
55	0.364	0.029	0.399	0.056	0.461	0.018	0.288	0.056	0.520
65	0.350	0.030	0.452	0.055	0.513	0.018	0.315	0.054	0.566
75	0.335	0.030	0.507	0.057	0.562	0.017	0.343	0.056	0.614
85	0.322	0.031	0.561	0.057	0.610	0.014	0.373	0.054	0.663
95	0.313	0.030	0.606	0.057	0.647	0.012	0.392	0.056	0.696
105	0.302	0.032	0.647	0.054	0.688	0.012	0.417	0.051	0.730
115	0.293	0.031	0.690	0.056	0.730	0.010	0.442	0.052	0.766
125	0.286	0.032	0.722	0.056	0.756	0.009	0.464	0.052	0.794
135	0.278	0.034	0.754	0.058	0.783	0.010	0.484	0.054	0.818
145	0.271	0.031	0.778	0.056	0.808	0.008	0.508	0.052	0.839

Table 3.6: *Proportion of rejections of one-sided upper-tailed tests when 1st population is Gamma*

the type I error rate as in the two-sided case. The T test has elevated type I error rates even with a big n_1 sample size. In fact, for most of the n_1 sample size settings, the type I error of the T test is close to or more than 0.1, twice the intended level significance level $\alpha = 0.05$. When $n_1 > 15$, the type I error rate of TCF and bootstrap-t test are both smaller than the T test and approaches α as n_1 increases.

Wilcoxon Rank-Sum Test gives the most rejection regardless of under our H_0 or H_a. Among the remaining three tests, the T test provides the biggest power at the expense of doubling the type I error rate. On the contrary, bootstrap-t test gives the lowest power and smallest type I error rate. The TCF test is in the middle, not only keeps a reasonable type

n_1	skewness	Tests							
		T-Test		Bootstrap-t		Wilcox		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.302	0.036	0.163	0.060	0.210	0.033	0.143	0.064	0.253
25	0.283	0.033	0.224	0.058	0.279	0.025	0.184	0.058	0.317
35	0.273	0.032	0.283	0.056	0.335	0.020	0.214	0.055	0.373
45	0.259	0.033	0.343	0.053	0.395	0.018	0.245	0.053	0.434
55	0.248	0.030	0.400	0.052	0.449	0.015	0.267	0.050	0.484
65	0.240	0.031	0.447	0.052	0.492	0.014	0.293	0.051	0.529
75	0.231	0.035	0.499	0.055	0.542	0.013	0.325	0.056	0.576
85	0.223	0.032	0.554	0.056	0.591	0.011	0.354	0.053	0.624
95	0.219	0.035	0.594	0.054	0.629	0.012	0.377	0.053	0.660
105	0.213	0.036	0.639	0.054	0.671	0.009	0.400	0.052	0.702
115	0.207	0.036	0.679	0.052	0.707	0.008	0.430	0.050	0.733
125	0.203	0.038	0.711	0.055	0.735	0.007	0.450	0.054	0.761
135	0.199	0.039	0.741	0.058	0.762	0.007	0.475	0.054	0.787
145	0.195	0.037	0.767	0.055	0.788	0.006	0.498	0.052	0.812

Table 3.7: *Proportion of rejections of one-sided upper-tailed tests when* 1*st population is Lognormal*

I error rate but also gives a good power. These three tests clearly demonstrate the trade off between the type I error rate and power. The T test sacrifices the type I error rate but gives a higher power, while the bootstrap-t test is more conservative on the type I error rate which leads to a small power.

This simulation study shows that the two-sample TCF test has good property in testing the two population mean difference under all three hypotheses for skewed data. It can not only maintain a reasonable type I error rate close to α, but also provides a higher power than the other commonly used tests.

n_1	skewness	Tests							
		T-Test		Bootstrap-t		Wilcox		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.477	0.109	0.303	0.063	0.199	0.137	0.348	0.082	0.248
25	0.435	0.106	0.355	0.065	0.262	0.176	0.502	0.080	0.294
35	0.412	0.099	0.398	0.063	0.301	0.198	0.605	0.076	0.335
45	0.385	0.100	0.436	0.063	0.345	0.223	0.689	0.074	0.373
55	0.364	0.092	0.480	0.060	0.381	0.240	0.763	0.069	0.412
65	0.350	0.089	0.515	0.055	0.416	0.256	0.815	0.065	0.449
75	0.335	0.088	0.542	0.056	0.448	0.279	0.855	0.064	0.477
85	0.322	0.084	0.571	0.055	0.479	0.285	0.891	0.064	0.509
95	0.313	0.080	0.599	0.052	0.513	0.298	0.912	0.061	0.540
105	0.302	0.075	0.627	0.051	0.540	0.312	0.936	0.058	0.567
115	0.293	0.077	0.649	0.052	0.567	0.327	0.953	0.060	0.591
125	0.286	0.076	0.676	0.053	0.594	0.341	0.964	0.060	0.621
135	0.278	0.074	0.695	0.052	0.618	0.354	0.974	0.059	0.639
145	0.271	0.073	0.721	0.051	0.646	0.367	0.979	0.055	0.667

Table 3.8: *Proportion of rejections of one-sided lower-tailed tests when* 1*st population is Gamma*

3.4 Summary

In this chapter, we obtained the Edgeworth expansion of the test statistic of pooled two sample t-test to derive a new approximation of it under skewness. The explicit form of its first order expansion was given in Theorem 3.1.3. On the basis of this expansion, a new two-sample test from Cornish Fisher expansion theory was constructed. We proved that the new two-sample test based on Cornish Fisher expansion (TCF test) can not only control the type I error rate but give a higher power of the test comparing with the pooled two-sample test under local alternatives with skewness. The power increment equals $L_{N,\gamma_1,\gamma_2,\lambda}$ defined

n_1	skewness	Tests							
		T-Test		Bootstrap-t		Wilcox		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.302	0.093	0.270	0.059	0.194	0.108	0.307	0.082	0.246
25	0.283	0.088	0.324	0.060	0.256	0.138	0.443	0.076	0.289
35	0.273	0.087	0.379	0.057	0.301	0.153	0.555	0.073	0.337
45	0.259	0.085	0.425	0.058	0.354	0.172	0.650	0.070	0.383
55	0.248	0.080	0.469	0.060	0.396	0.188	0.721	0.068	0.425
65	0.240	0.081	0.514	0.057	0.441	0.198	0.780	0.067	0.469
75	0.231	0.080	0.552	0.055	0.481	0.218	0.830	0.065	0.506
85	0.223	0.077	0.580	0.055	0.516	0.224	0.864	0.064	0.536
95	0.219	0.074	0.612	0.053	0.550	0.239	0.894	0.061	0.570
105	0.213	0.072	0.641	0.054	0.580	0.256	0.916	0.061	0.597
115	0.207	0.072	0.659	0.052	0.599	0.266	0.936	0.062	0.617
125	0.203	0.071	0.687	0.052	0.629	0.276	0.947	0.060	0.646
135	0.199	0.071	0.713	0.052	0.657	0.287	0.961	0.060	0.672
145	0.195	0.067	0.736	0.051	0.683	0.298	0.970	0.057	0.695

Table 3.9: *Proportion of rejections of one-sided lower-tailed tests when* 1*st population is Lognormal*

in equation (3.2.18). In practice, majority of settings satisfy the sample ratio condition in Corollary 3.2.6, which leads to higher power for the TCF test.

Furthermore, the results from the example in Section 3.2.3.2 and the simulation study in Section 3.3 support the theoretical results in Section 3.2 well. Under the simulation settings in Section 3.3, the two-sample TCF test provided the highest power and should be recommended over the pooled two-sample t-test, Bootstrap t-test and Wilcoxon Rank-Sum Test for skewed data.

Note that this simulation study focuses on the empirical results based on a fixed sample ratio $\lambda = 0.6$ and population skewness, i.e., $\gamma_1 \approx 6$ and $\gamma_2 = 0$. In Section 4.4, we will conduct a more completed simulation study to figure out the effects of the following factors on the empirical type I error rate and power of the test:

1. Different population distributions;
2. Different levels of λ;
3. Different levels of common population variance σ^2;
4. Different levels of population skewness γ_i.

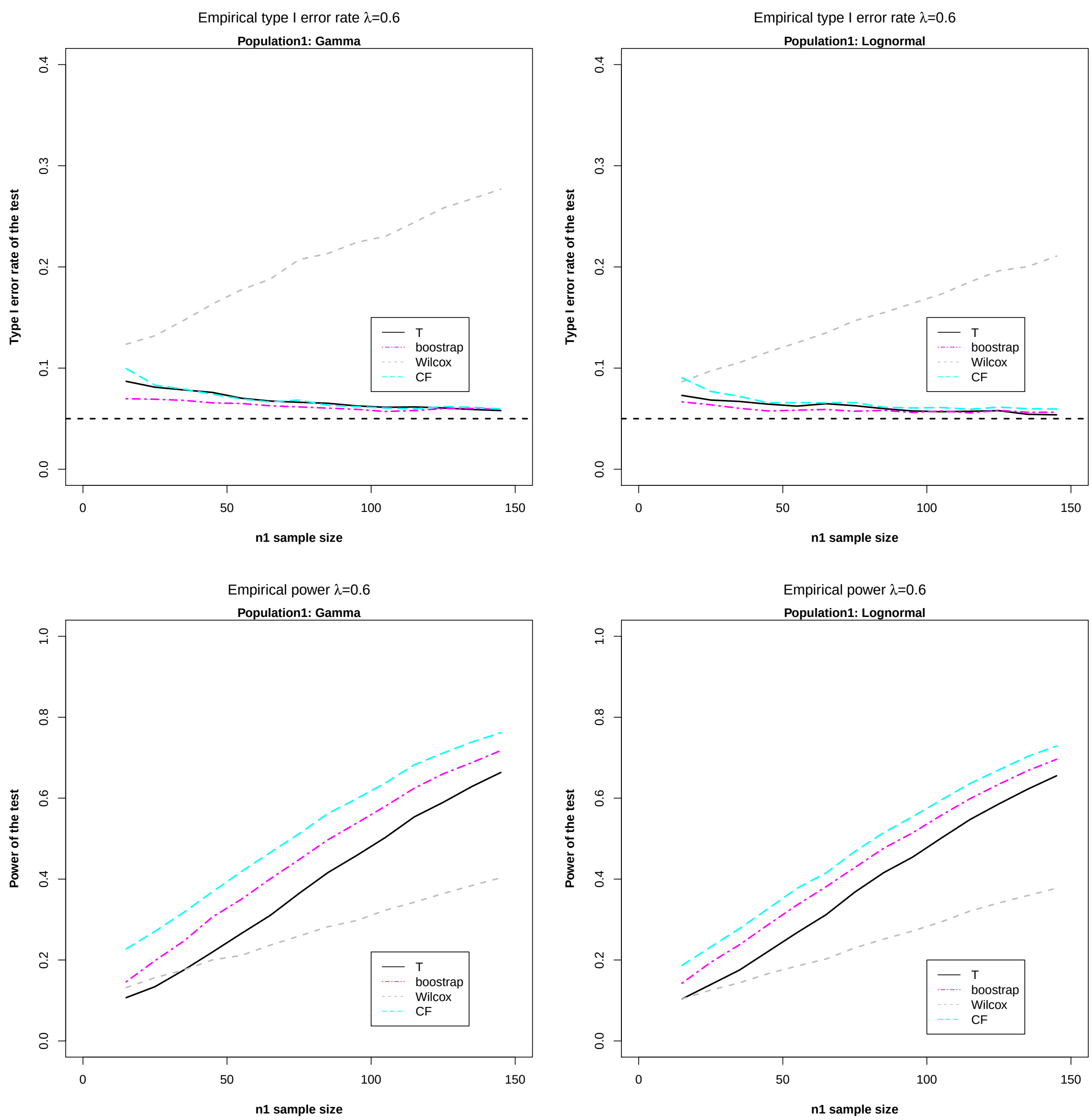

Figure 3.3: *Proportion of rejections of two-sided test. In top 2 panels the data were generated under H_0. In bottom 2 panels, the data were generated under H_a. The left 2 panels correspond to population 1 being Gamma distribution. The right two panels correspond to population 1 being Lognormal distribution.*

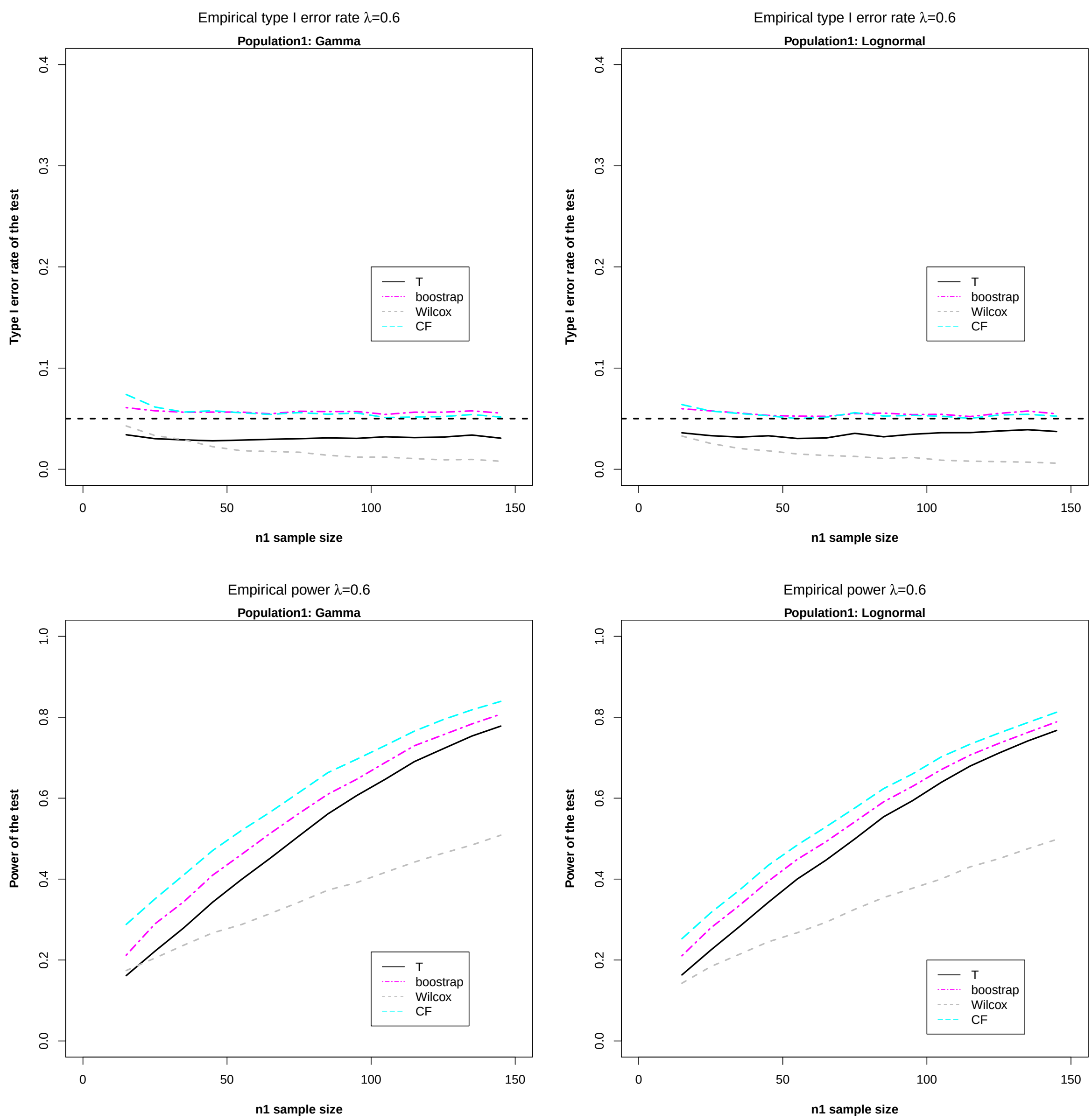

Figure 3.4: *Proportion of rejections of one-sided upper tail test. In top 2 panels the data were generated under H_0. In bottom 2 panels, the data were generated under H_a. The left 2 panels correspond to population 1 being Gamma distribution. The right two panels correspond to population 1 being Lognormal distribution.*

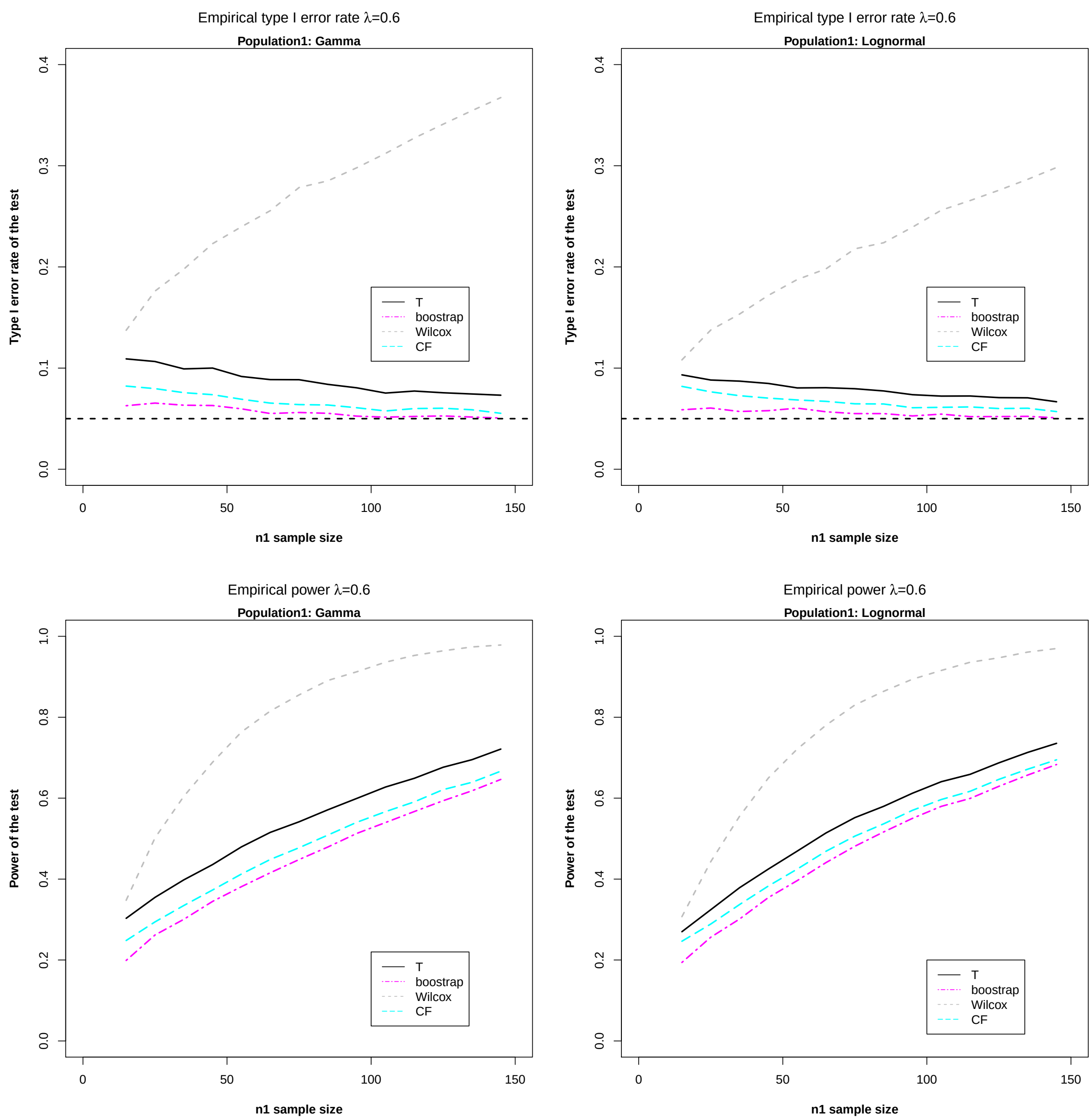

Figure 3.5: *Proportion of rejections of one-sided lower tail test. In top 2 panels the data were generated under H_0. In bottom 2 panels, the data were generated under H_a. The left 2 panels correspond to population 1 being Gamma distribution. The right two panels correspond to population 1 being Lognormal distribution.*

Chapter 4

New tests through transformations based on Edgeworth expansions

4.1 Two-sample tests based on three transformations

Here we introduce three new tests based on transformations proposed by Hall (1992b) and Zhou and Philip (2005). These three transformations can be used to improve the coverage probability of confidence intervals based on studentized statistic under skewness by eliminating the skewness from the corresponding distribution of the studentized statistic. In particular, when the cumulative distribution function of a studentized statistic has an Edgeworth expansion

$$\Phi(x) + \frac{\phi(x)}{\sqrt{n}}\Big[\gamma(ax^2 + b)\Big] + O(n^{-1}), \tag{4.1.1}$$

then the three transformations have the forms

$$T_1 = T_1(U) = U + a\hat{\gamma}U^2 + \frac{1}{3}a^2\hat{\gamma}^2U^3 + n^{-1}b\hat{\gamma},$$

$$T_2 = T_2(U) = (2an^{-1/2}\hat{\gamma})^{-1}\{exp(2an^{-1/2}\hat{\gamma}U) - 1\} + n^{-1}b\hat{\gamma},$$

$$T_3 = T_3(U) = U + U^2 + \frac{1}{3}U^3 + n^{-1}b\hat{\gamma},$$

where a, b and γ are the coefficients of Edgeworth expansion in (4.1.1). In our Edgeworth expansion under the two-sample settings in Section 3.2.1, the cumulative distribution function of the test statistic T in (3.1.1) under H_0 has the form

$$P(T \leq x) = \Phi(x) + \frac{\phi(x)}{\sqrt{n}}[A + B(x^2-1)] + O(N^{-min(1,r+1/2)}),$$

where A and B are defined in equation (3.1.10).

Clearly, our Edgeworth expansion does not have the same format as the expansion in (4.1.1) of Hall (1992b), in that our expansion has an extra term A. In this case, we can not directly follow the standard procedures in Section 2.3.2.1 to get the values of a, b and γ for each transformation and derive the values of percentiles. To solve this problem, we first apply a linear transformation on T as in Phillip and Zhou (2005):

Theorem 4.1.1. *Let $T^{'} = T + N^{-1/2}\hat{A}$, then under H_0,*

$$P_{H_0}(T^{'} \leq x) = \Phi(x) + N^{-1/2}B(x^2-1)\phi(x) + O(N^{-min(1,r+1/2)}). \tag{4.1.2}$$

And under H_a, assume $\delta = O(N^{-1/2})$. Then

$$\begin{aligned} P_{H_a}(T^{'} \leq x) &= \Phi(x - c_N) + N^{-1/2}B((x-c_N)^2-1)\phi(x-c_N) \\ &\quad + Q(x) + O(N^{-min(1,r+1/2)}) \\ &= P_{H_0}(T^{'} \leq x - c_N) + Q(x) + O(N^{-min(1,r+1/2)}), \end{aligned} \tag{4.1.3}$$

where $c_N = \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ and $Q(x)$ is defined in (3.2.12). And $\hat{A}$ is the estimate of A given by

$$\hat{A} = [\lambda_N(1-\lambda_N)]^{1/2}(\hat{\gamma_1} - \hat{\gamma_2})/2,$$

$$B = [\lambda_N(1-\lambda_N)]^{1/2}\left(\frac{8\lambda_N - 2}{\lambda_N}\gamma_1 - \frac{6-8\lambda_N}{1-\lambda_N}\gamma_2\right)/12,$$

where

$$\begin{aligned} \hat{\gamma_i} &= \frac{n_i}{(n_i-1)(n_i-2)}\sum_{j=1}^{n_i}\left\{\frac{X_{ij}-\bar{X}_i}{S_i}\right\}^3, \quad i = 1, 2, \\ \lambda_N &= \frac{n_1}{(n_1+n_2)}. \end{aligned}$$

The proof is given in Appendix A.4. Obviously, $T^{'}$ in (4.1.2) has the same format as the expansion in (4.1.1). Therefore, we can first use the three transformations to find the α^{th} and $(1-\alpha)^{th}$ percentiles of distribution of $T^{'}$, then transfer them back to get the α^{th} and $(1-\alpha)^{th}$ percentiles of distribution of T. Based on the distribution of $T^{'}$ in Theorem 4.1.1, we have $a = 1/3$, $b = -1/3$ and $\gamma = 3N^{-1/2}B$ with its estimate $\hat{\gamma} = 3N^{-1/2}\hat{B}$, where

$$\hat{B} = [\lambda_N(1-\lambda_N)]^{1/2}\left(\frac{8\lambda_N - 2}{\lambda_N}\hat{\gamma}_1 - \frac{6-8\lambda_N}{1-\lambda_N}\hat{\gamma}_2\right)/12.$$

Given a two-sided hypothesis with significance level α and the test statistic T in (3.1.1), the three new rejection regions can be computed through the following steps:

1. Do a T_i transformation on the $U = \frac{1}{\sqrt{N}}T^{'}$ to derive a new statistic. The distribution of the transformed statistic $\sqrt{N}T_i(U)$ is virtually symmetric and approximately standard normal.

2. Use the percentile of standard normal distribution, say $z_{\alpha/2}$, to approximate the percentile of the transformed statistic $\sqrt{N}T_i(U)$.

3. Denote $\eta_{i,\alpha/2} = \frac{z_{\alpha/2}}{\sqrt{N}}$, then $\eta_{i,\alpha/2}$ is the $\frac{\alpha}{2}^{th}$ percentile of $T_i(U)$. Accordingly, $T_i^{-1}(\eta_{i,\alpha/2})$ gives $\frac{\alpha}{2}^{th}$ percentile of U, where $T_i^{-1}(.)$ is the inverse transform of $T_i(.)$. They are given as follows:

$$T_1^{-1}(t) = (a\hat{\gamma})^{-1}\{1 + 3a\hat{\gamma}(t - b\hat{\gamma}/N)\}^{1/3} - (a\hat{\gamma})^{-1}$$
$$T_2^{-1}(t) = (2aN^{-1/2}\hat{\gamma})^{-1}log\{2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma}) + 1\}$$
$$T_3^{-1}(t) = \{1 + 3(t - N^{-1}b\hat{\gamma})\}^{1/3} - 1.$$

4. Since $U = \frac{1}{\sqrt{N}}T^{'}$, the $\frac{\alpha}{2}^{th}$ percentile of the distribution $T^{'}$ is $\sqrt{N}T_i^{-1}\left(\frac{z_{\alpha/2}}{\sqrt{N}}\right)$

5. Report the rejection regions of $T^{'}$ based on each of the three transformations as:

$$T^{'} \le \sqrt{N}T_i^{-1}\left(\frac{z_{\alpha/2}}{\sqrt{N}}\right) \;\; or \;\; T^{'} \ge \sqrt{N}T_i^{-1}\left(\frac{z_{1-\alpha/2}}{\sqrt{N}}\right), \tag{4.1.4}$$

   where $i = 1, 2, 3$.

6. Finally, the new rejection regions of T can be derived from (4.1.4) based on the linear transformation in Theorem 4.1.1,

$$T \leq \frac{-\hat{A}}{\sqrt{N}} + \sqrt{N} T_i^{-1}(\frac{z_{\alpha/2}}{\sqrt{N}}) \;\; or \;\; T \geq \frac{-\hat{A}}{\sqrt{N}} + \sqrt{N} T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}), \tag{4.1.5}$$

where $i = 1, 2, 3$.

We reject the null hypothesis if T falls into the rejection regions in (4.1.5) for the two-sided hypothesis. In the further discussions, we will refer the two sample test based on three transformations as "T_i" test, $i = 1, 2, 3$. Accordingly, the rejection region for one-sided upper-tailed T_i test is

$$T \geq \frac{-\hat{A}}{\sqrt{N}} + \sqrt{N} T_i^{-1}(\frac{z_{1-\alpha}}{\sqrt{N}}), \tag{4.1.6}$$

and the rejection region for one-sided lower-tailed T_i test is

$$T \leq \frac{-\hat{A}}{\sqrt{N}} + \sqrt{N} T_i^{-1}(\frac{z_{\alpha}}{\sqrt{N}}). \tag{4.1.7}$$

4.2 Type I error rate of the test based on three transformations

In this section, we present the theoretical type I error rate for the three tests introduced in Section 4.1. Under the two-sample setting in Section 3.2.1, the distribution of the test statistic T is still $F_T^{(1)}(t)$, and the rejection regions are given by (4.1.5), (4.1.6) and (4.1.7) for two-sided, upper-tailed and lower-tailed T_i test, respectively.

4.2.1 Type I error rate of the two-sided T_i test

Based on the results from Theorem 4.1.1, we have $T^{'} = T + N^{-1/2}\hat{A}$ and the rejection region of $T^{'}$ shows in (4.1.4) as

$$T^{'} \leq \sqrt{N} T_i^{-1}(\frac{z_{\alpha/2}}{\sqrt{N}}) \;\; or \;\; T^{'} \geq \sqrt{N} T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}),$$

where $i = 1, 2, 3$ and $T^{'}$ has the cdf given in (4.1.2). Denote the lower threshold in formula (4.1.4) as $t^{(i)}_{\alpha/2}$ and the upper threshold as $t^{(i)}_{1-\alpha/2}$. That is

$$t^{(i)}_{\alpha/2} = \sqrt{N} T_i^{-1}(\frac{z_{\alpha/2}}{\sqrt{N}})$$
$$t^{(i)}_{1-\alpha/2} = \sqrt{N} T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}).$$

These two values are not symmetric about zero. When H_0 is true, the theoretical type I error rate of the T_i test is obtained by:

$$P(T^{'} \geq t^{(i)}_{1-\alpha/2}) + P(T^{'} \leq t^{(i)}_{\alpha/2}) \tag{4.2.1}$$

In addition, based on the results from Hall (1992b) introduced in equation (2.3.7) and the results from Theorem 3.1.3, in two-sample case we have

$$P(\sqrt{N} T_i(U) \leq x) = \Phi(x) + O(N^{-min(1,r+1/2)}) \tag{4.2.2}$$

where $U = \frac{1}{\sqrt{N}} T^{'}$. Based on the above results, the type I error rate of these tests based on transformations can be derived as:

$$\begin{aligned}
&P(\text{type I error of } T_i \text{ test}) = P(T^{'} \geq t^{(i)}_{1-\alpha/2}) + P(T^{'} \leq t^{(i)}_{\alpha/2}) \\
&= 1 - P(T^{'} \leq t^{(i)}_{1-\alpha/2}) + P(T^{'} \leq t^{(i)}_{\alpha/2}) \\
&= 1 - P(\sqrt{N} T_i(U) \leq \sqrt{N} T_i(t^{(i)}_{1-\alpha/2}/\sqrt{N}) + P(\sqrt{N} T_i(U) \leq \sqrt{N} T_i(t^{(i)}_{\alpha/2}/\sqrt{N}),
\end{aligned}$$

where the last equality holds due to the fact that $T_i(t)$ are monotone functions. Replacing $t^{(i)}_{1-\alpha/2}$ and $t^{(i)}_{\alpha/2}$ with their definitions, we have

$$\begin{aligned}
&P(\text{type I error of } T_i \text{ test}) = 1 - P(\sqrt{N} T_i(U) \leq z_{1-\alpha/2}) + P(\sqrt{N} T_i(U) \leq z_{\alpha/2}) \\
&= 1 - \Phi(z_{1-\alpha/2}) + \Phi(z_{\alpha/2}) + O(N^{-min(1,r+1/2)}) \qquad (4.2.3) \\
&= \alpha + O(N^{-min(1,r+1/2)}).
\end{aligned}$$

Therefore, the approximated type I error rate of the two-sample T_i test has an order of $O(N^{-min(1,r+1/2)})$, which is the same order of the type I error rate of the two-sample TN test.

4.2.2 Type I error rate of the one-sided upper-tailed T_i test

For the one-sided upper-tailed T_i test, the rejection region in (4.1.6) can be expressed by $T^{'}$ based on the linear transformation $T^{'} = T + N^{-1/2}\hat{A}$ from Theorem 4.1.1. The new rejection region for one-sided upper-tailed T_i test becomes:

$$T^{'} \geq \sqrt{N}T_i^{-1}(\frac{z_{1-\alpha}}{\sqrt{N}}), \tag{4.2.4}$$

where $i = 1, 2, 3$. Denote the threshold in (4.2.4) as $t_{1-\alpha}^{(i)}$. That is

$$t_{1-\alpha}^{(i)} = \sqrt{N}T_i^{-1}(\frac{z_{1-\alpha}}{\sqrt{N}}).$$

When H_0 is true, the theoretical type I error rate of one-sided upper-tailed T_i test is obtained by:

$$P(T^{'} \geq t_{1-\alpha}^{(i)}) \tag{4.2.5}$$

Recall that we have $P(\sqrt{N}T_i(U) \leq x) = \Phi(x) + O(N^{-min(1,r+1/2)})$ from (4.2.2), where $U = \frac{1}{\sqrt{N}}T^{'}$. Then the type I error rate of one-sided upper-tailed T_i test can be derived as:

$$\begin{aligned}
&P(\text{type I error of one-sided upper-tailed } T_i \text{ test}) = P(T^{'} \geq t_{1-\alpha}^{(i)}) \\
&= 1 - P(T^{'} \leq t_{1-\alpha}^{(i)}) = 1 - P(\sqrt{N}T_i(U) \leq \sqrt{N}T_i(t_{1-\alpha}^{(i)}/\sqrt{N}),
\end{aligned}$$

where the equality holds due to the fact that $T_i(t)$ are monotone functions. Plugging in $t_{1-\alpha}^{(i)}$ with its definition, we have

$$\begin{aligned}
&P(\text{type I error of one-sided upper-tailed } T_i \text{ test}) = 1 - P(\sqrt{N}T_i(U) \leq z_{1-\alpha}) \\
&= 1 - \Phi(z_{1-\alpha}) + O(N^{-min(1,r+1/2)}) = \alpha + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{4.2.6}$$

4.2.3 Type I error rate of the one-sided lower-tailed T_i test

Following the same procedures in Section 4.2.2, the new rejection region of one-sided lower-tailed T_i test under H_0 can be expressed by $T^{'}$ as

$$T^{'} \leq \sqrt{N}T_i^{-1}(\frac{z_{\alpha}}{\sqrt{N}}), \tag{4.2.7}$$

where $i = 1, 2, 3$. Denote

$$t_\alpha^{(i)} = \sqrt{N} T_i^{-1}(\frac{z_\alpha}{\sqrt{N}}).$$

When H_0 is true, the theoretical type I error rate of one-sided lower-tailed T_i test is:

$$P(T^{'} \leq t_\alpha^{(i)}) \tag{4.2.8}$$

Based on the results from (4.2.2), we have

$$\begin{aligned}&P(\text{type I error of one-sided lower-tailed } T_i \text{ test}) = P(T^{'} \leq t_\alpha^{(i)})\\&= P(T^{'} \leq t_\alpha^{(i)}) = P(\sqrt{N} T_i(U) \leq \sqrt{N} T_i(t_\alpha^{(i)}/\sqrt{N}).\end{aligned}$$

Then plug in $t_\alpha^{(i)} = \sqrt{N} T_i^{-1}(\frac{z_\alpha}{\sqrt{N}})$, we have

$$\begin{aligned}&P(\text{type I error of one-sided lower-tailed } T_i \text{ test}) = P(\sqrt{N} T_i(U) \leq z_\alpha)\\&= \Phi(z_\alpha) + O(N^{-min(1,r+1/2)}) = \alpha + O(N^{-min(1,r+1/2)}).\end{aligned} \tag{4.2.9}$$

From Section 3.2.4.1, we know that the approximated type I error rates of the same one-sided upper-tailed and lower-tailed TN tests are both $\alpha + O(N^{-1/2})$. Then the type I error rate accuracy for the one-sided T_i test has higher approximation accuracy than the one-sided TN test.

4.3 Power of the test based on three transformations

Now consider the data generated under H_a. We derive the power function of the one-sided and two-sided T_i tests under local alternative hypothesis.

4.3.1 Power of the two-sided T_i test under local alternative hypothesis

Based on Theorem 4.1.1, the theoretical power of the two-sided T_i test is given by

$$\begin{aligned}&P_{H_a}(T \leq \frac{-\hat{A}}{\sqrt{N}} + \sqrt{N} T_i^{-1}(\frac{z_{\alpha/2}}{\sqrt{N}})) + P_{H_a}(T \geq \frac{-\hat{A}}{\sqrt{N}} + \sqrt{N} T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}))\\&= P_{H_0}(T^{'} \geq t_{1-\alpha/2}^{(i)} - c_N) + P_{H_0}(T^{'} \leq t_{\alpha/2}^{(i)} - c_N) - Q(t_{1-\alpha/2}^{(i)}) + Q(t_{\alpha/2}^{(i)}),\end{aligned} \tag{4.3.1}$$

where $c_N = \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ and $Q(x) = \frac{q_N}{2}(x-c_N)\phi(x-c_N)$. Denoting $t^i_{1-\alpha/2} - c_N$ as $L_u^{(i)}$ and $t^i_{\alpha/2} - c_N$ as $L_l^{(i)}$, then the power can be expressed as:

$$
\begin{aligned}
&\text{power of two-sided } T_i \text{ test} = 1 - P_{H_0}(T^{'} \le L_u^{(i)}) + P_{H_0}(T^{'} \le L_l^{(i)}) - Q(L_u^{(i)}) + Q(L_l^{(i)}) \\
=&1 - P_{H_0}(\sqrt{N}T_i(U) \le \sqrt{N}T_i(L_u^{(i)}/\sqrt{N}) + P_{H_0}(\sqrt{N}T_i(U) \le \sqrt{N}T_i(L_l^{(i)}/\sqrt{N}) \\
&-Q(L_u^{(i)}) + Q(L_l^{(i)}),
\end{aligned}
$$

where the last equality is due to the fact that $T_i(t)$ are monotone functions. Then replacing $L_u^{(i)}$ and $L_l^{(i)}$ with their definitions, we have

$$
\begin{aligned}
&\text{power of two-sided } T_i \text{ test} \\
=&1 - P_{H_0}\left(\sqrt{N}T_i(U) \le \sqrt{N}T_i\left[(t^{(i)}_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})})/\sqrt{N}\right]\right) \\
&+ P_{H_0}\left(\sqrt{N}T_i(U) \le \sqrt{N}T_i\left[(t^{(i)}_{\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})})/\sqrt{N}\right]\right) \\
&- Q(L_u^{(i)}) + Q(L_l^{(i)}).
\end{aligned}
\tag{4.3.2}
$$

Under the local alternative, $\delta = O(N^{-1/2})$ and $\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} = O(1)$. Therefore we have

$$
\begin{aligned}
&P_{H_0}\left(\sqrt{N}T_i(U) \le \sqrt{N}T_i\left[\left(t^i_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right)/\sqrt{N}\right]\right) \\
=&E\left(E\left(\sqrt{N}T_i(U) \le \sqrt{N}T_i\left[\left(t^i_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right)/\sqrt{N}\right]\Big|\hat{B}\right)\right) \\
=&E\left(\Phi\left\{\sqrt{N}T_i\left[\left(t^i_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right)/\sqrt{N}\right]\right\}\right) + O(N^{-min(1,r+1/2)}) \\
=&E\left(\Phi\left\{\sqrt{N}T_i\left[T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}) - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}/\sqrt{N}\right]\right\}\right) + O(N^{-min(1,r+1/2)})
\end{aligned}
$$

Due to the fact that

$$
\begin{aligned}
&T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}) - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}/\sqrt{N} \\
&= T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}) + O(N^{-1/2}).
\end{aligned}
$$

We can apply Taylor expansion to $T_i(U)$ at $T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}})$ to get

$$
\begin{aligned}
&P_{H_0}\left(\sqrt{N}T_i(U) \le \sqrt{N}T_i\left[\left(t^i_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right)/\sqrt{N}\right]\right) \\
=&\quad E\left(\Phi\left\{z_{1-\alpha/2} - T_i^{'}\left[T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right\}\right) + O(N^{-min(1,r+1/2)}).
\end{aligned}
$$

Now equation (4.3.2) can be simplified as:

$$\text{power of two-sided } T_i \text{ test} = 1 - E\left(\Phi\left\{z_{1-\alpha/2} - T_i^{'}\left[T_i^{-1}(\tfrac{z_{1-\alpha/2}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\tfrac{1}{n_1}+\tfrac{1}{n_2})}\right\}\right)$$
$$+E\left(\Phi\left\{z_{\alpha/2} - T_i^{'}\left[T_i^{-1}(\tfrac{z_{\alpha/2}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\tfrac{1}{n_1}+\tfrac{1}{n_2})}\right\}\right)$$
$$-Q(L_u^{(i)}) + Q(L_l^{(i)}) + O(N^{-min(1,r+1/2)}). \tag{4.3.3}$$

Comparing the power of the two-sided T_i test in (4.3.3) with the power of the two-sample TN test in (3.2.19), they do not have the same functional form even though their approximation order is identical. The approximation in equation (4.3.3) adjusts the term $\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ by a coefficient $T_i^{'}\left[T_i^{-1}(\frac{z_{\alpha/2}}{\sqrt{N}})\right]$, which is a function of standard normal percentile.

Next, we further investigate the power of T_i tests and compare them with the power of TCF test. Recall that the three transformations have the forms:

$$T_1 = T_1(U) = U + a\hat{\gamma}U^2 + \frac{1}{3}a^2\hat{\gamma}^2U^3 + N^{-1}b\hat{\gamma}, \tag{4.3.4}$$

$$T_2 = T_2(U) = (2aN^{-1/2}\hat{\gamma})^{-1}\{exp(2aN^{-1/2}\hat{\gamma}U) - 1\} + N^{-1}b\hat{\gamma}, \tag{4.3.5}$$

$$T_3 = T_3(U) = U + U^2 + \frac{1}{3}U^3 + N^{-1}b\hat{\gamma}, \tag{4.3.6}$$

where $a = 1/3$, $b = -1/3$ and $\hat{\gamma} = 3N^{-1/2}\hat{B}$. Conditional on $\hat{B}$, $\hat{\gamma} = O(N^{-1/2})$. The derivatives of $T_i(U)$ are given by

$$T_1^{'} = T_1^{'}(U) = 1 + 2a\hat{\gamma}U + a^2\hat{\gamma}^2U^2 = (1 + a\hat{\gamma}U)^2, \tag{4.3.7}$$

$$T_2^{'} = T_2^{'}(U) = exp(2aN^{-1/2}\hat{\gamma}U), \tag{4.3.8}$$

$$T_3^{'} = T_3^{'}(U) = 1 + 2U + U^2 = (1+U)^2. \tag{4.3.9}$$

To evaluate $T_i^{'}\left[T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}})\right]$, recall that the inverse transforms are

$$T_1^{-1}(t) = (a\hat{\gamma})^{-1}\{1 + 3a\hat{\gamma}(t - b\hat{\gamma}/N)\}^{1/3} - (a\hat{\gamma})^{-1}, \tag{4.3.10}$$

$$T_2^{-1}(t) = (2aN^{-1/2}\hat{\gamma})^{-1}log\{2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma}) + 1\}, \tag{4.3.11}$$

$$T_3^{-1}(t) = \{1 + 3(t - N^{-1}b\hat{\gamma})\}^{1/3} - 1. \tag{4.3.12}$$

Both $\frac{z_{1-\alpha/2}}{\sqrt{N}}$ and $\frac{z_{\alpha/2}}{\sqrt{N}}$ are $O(N^{-1/2})$. So we can apply Taylor expansion to $T_i^{-1}(t)$ at $t = 0$. Specifically, since conditional on $\hat{B}$, $\hat{\gamma} = O(N^{-1/2})$, then $\{1 + 3a\hat{\gamma}(t - b\hat{\gamma}/N)\}^{1/3}$, $log\{2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma}) + 1\}$ and $\{1 + 3(t - N^{-1}b\hat{\gamma})\}^{1/3}$ can be approximated with Taylor expansion as

$$\{1 + 3a\hat{\gamma}(t - b\hat{\gamma}/N)\}^{1/3} = 1 + 3a\hat{\gamma}(t - b\hat{\gamma}/N)1/3 + O([a\hat{\gamma}(t - b\hat{\gamma}/N)]^2),$$

$$\begin{aligned}&log\{2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma}) + 1\}\\ &= log(1) + 2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma}) + O([2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma})]^2),\end{aligned}$$

$$\{1 + 3(t - N^{-1}b\hat{\gamma})\}^{1/3} = 1 + 3(t - N^{-1}b\hat{\gamma})1/3 + O([t - N^{-1}b\hat{\gamma}]^2).$$

Therefore for $t = O(N^{-1/2})$, conditional on $\hat{B}$,

$$\begin{aligned}T_1^{-1}(t) &= (a\hat{\gamma})^{-1}\{1 + 3a\hat{\gamma}(t - b\hat{\gamma}/N)1/3 + O(N^{-2})\} - (a\hat{\gamma})^{-1}\\ &= t - N^{-1}b\hat{\gamma} + O(N^{-3/2}), \qquad (4.3.13)\\ T_2^{-1}(t) &= (2aN^{-1/2}\hat{\gamma})^{-1}\{log(1) + 2aN^{-1/2}\hat{\gamma}(t - N^{-1}b\hat{\gamma}) + O(N^{-3})\}\\ &= t - N^{-1}b\hat{\gamma} + O(N^{-2}), \qquad (4.3.14)\\ T_3^{-1}(t) &= 1 + 3(t - N^{-1}b\hat{\gamma})1/3 + O(N^{-1}) - 1\\ &= t - N^{-1}b\hat{\gamma} + O(N^{-1}). \qquad (4.3.15)\end{aligned}$$

4.3.1.1 Power of the two-sided T_1 test under local alternative hypothesis

Recall that, conditional on $\hat{B}$

$$\begin{aligned}&T_1^{'} = T_1^{'}(U) = (1 + a\hat{\gamma}U)^2\\ &T_1^{-1}(t) = t - N^{-1}b\hat{\gamma} + O(N^{-3/2}) \text{ for } t = O(N^{-1/2}).\end{aligned}$$

Thus the first two terms of the T_1 power function from (4.3.3) becomes

$$
\begin{aligned}
& 1-E\left(\Phi\left\{z_{1-\alpha/2}-T_1'\left[T_1^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right\}\right)\\
= \quad & 1-E\left(\Phi\left\{z_{1-\alpha/2}-\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\left[\left(1+a\hat{\gamma}\left(\frac{z_{1-\alpha/2}}{\sqrt{N}}-\frac{b\hat{\gamma}}{N}+O_p(N^{-3/2})\right)\right)^2\right]\right\}\right)\\
= \quad & 1-E\left(\Phi\left\{z_{1-\alpha/2}-\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\left[1+\frac{2a\hat{\gamma}z_{1-\alpha/2}}{\sqrt{N}}+O_p(N^{-2})\right]\right\}\right)
\end{aligned}
$$

Apply Taylor expansion to $\Phi(x)$, we know the above term is equal to

$$
= \quad 1-E\left(\Phi(U_{N,1-\alpha/2})-\frac{N^{-1/2}\delta 2a\hat{\gamma}z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}+O(N^{-2})\right), \tag{4.3.16}
$$

where $U_{N,1-\alpha/2}=z_{1-\alpha/2}-\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$. Then the power function of the two-sided T_1 test becomes

$$
\begin{aligned}
& 1-E\left(\Phi(U_{N,1-\alpha/2})-\frac{N^{-1/2}\delta 2a\hat{\gamma}z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right)-Q(L_u^{(1)})+Q(L_l^{(1)})\\
& +E\left(\Phi(U_{N,\alpha/2})-\frac{N^{-1/2}\delta 2a\hat{\gamma}z_{\alpha/2}\phi(U_{N,\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right)+O(N^{-min(1,r+1/2)}) \qquad (4.3.17)\\
& =1-\Phi(U_{N,1-\alpha/2})+\frac{2aN^{-1/2}\delta z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\phi(U_{N,1-\alpha/2})E(\hat{\gamma})-Q(L_u^{(1)})+Q(L_l^{(1)})\\
& +\Phi(U_{N,\alpha/2})-\frac{2aN^{-1/2}\delta z_{\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\phi(U_{N,\alpha/2})E(\hat{\gamma})+O(N^{-min(1,r+1/2)})
\end{aligned}
$$

Note that $\hat{\gamma}=3N^{-1/2}\hat{B}=3N^{-1/2}(B+O_p(N^{-min(r,1/2)}))$ and $\hat{B}$ is a linear combination of $\hat{\gamma}_1$, $\hat{\gamma}_2$, whose distribution is continues, we know $E(\hat{\gamma})=\gamma+O(N^{-min(r+1/2,1)})$. Hence the power of the two-sided T_1 test is

$$\begin{aligned}
&1-\Phi(U_{N,1-\alpha/2})+\frac{N^{-1/2}\delta 2a\gamma z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}-Q(L_u^{(1)})+Q(L_l^{(1)})\\
&+\Phi(U_{N,\alpha/2})-\frac{N^{-1/2}\delta 2a\gamma z_{\alpha/2}\phi(U_{N,\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}+O(N^{-min(1,r+1/2)})\\
&=1-\Phi(U_{N,1-\alpha/2})+\Phi(U_{N,\alpha/2})-Q(L_u^{(1)})+Q(L_l^{(1)})\\
&+N^{-1/2}[\phi(U_{N,1-\alpha/2})+\phi(U_{N,\alpha/2})]\frac{\delta 2a\gamma z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}+O(N^{-min(1,r+1/2)}) \qquad (4.3.18)
\end{aligned}$$

Recall that the power of the two-sided TCF test in (3.2.17) is,

$$\begin{aligned}
&1-\Phi(U_{N,1-\alpha/2})-N^{-1/2}[A+B(U^2_{N,1-\alpha/2}-1)]\phi(U_{N,1-\alpha/2})-Q(L_u^{cf})+Q(L_l^{cf})\\
&+\Phi(U_{N,\alpha/2})+N^{-1/2}[A+B(U^2_{N,\alpha/2}-1)]\phi(U_{N,\alpha/2})+L_{N,\gamma_1,\gamma_2,\lambda}+O(N^{-min(1,r+1/2)}),
\end{aligned}$$

where

$$L_{N,\gamma_1,\gamma_2,\lambda}=\Delta_{N,\alpha/2}[\phi(U_{N,\alpha/2})-\phi(U_{N,1-\alpha/2})]+O(N^{-1}),$$

and $\Delta_{N,\alpha/2}=-N^{-1/2}[A+B(z^2_{\alpha/2}-1)]$. The power difference between the two-sided T_1 test and the two-sided TCF test can be expressed as:

$$\begin{aligned}
&\text{Power of two-sided } T_1 \text{ test - Power of two-sided } TCF \text{ test}\\
&=N^{-1/2}[\phi(U_{N,1-\alpha/2})+\phi(U_{N,\alpha/2})]\frac{\delta 2a\gamma z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}+N^{-1/2}A[\phi(U_{N,1-\alpha/2})-\phi(U_{N,\alpha/2})]\\
&+N^{-1/2}B[\phi(U_{N,1-\alpha/2})(U^2_{N,1-\alpha/2}-1)-\phi(U_{N,\alpha/2})(U^2_{N,\alpha/2}-1)]\\
&-\left\{-N^{-1/2}[A+B(z^2_{\alpha/2}-1)]\right\}[\phi(U_{N,\alpha/2})-\phi(U_{N,1-\alpha/2})]\\
&-Q(L_u^{(1)})+Q(L_l^{(1)})+Q(L_u^{cf})-Q(L_l^{cf})+O(N^{-min(1,r+1/2)})\\
&=N^{-1/2}[\phi(U_{N,1-\alpha/2})+\phi(U_{N,\alpha/2})]\frac{\delta 2a\gamma z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}+\\
&\frac{B}{\sqrt{N}}[\phi(U_{N,1-\alpha/2})(U^2_{N,1-\alpha/2}-z^2_{1-\alpha/2})-\phi(U_{N,\alpha/2})(U^2_{N,\alpha/2}-z^2_{\alpha/2})]\\
&-Q(L_u^{(1)})+Q(L_l^{(1)})+Q(L_u^{cf})-Q(L_l^{cf})+O(N^{-min(1,r+1/2)}).
\end{aligned}$$

After further investigating the power difference, we have

$$-Q(L_u^{(i)}) + Q(L_l^{(i)}) + Q(L_u^{cf}) - Q(L_l^{cf}) = O(N^{-1}). \tag{4.3.19}$$

The proof is given in Appendix A.7. In addition, $a = 1/3$, $\gamma = 3N^{-1/2}B$ and

$$U_{N,\alpha/2}^2 - z_{\alpha/2}^2 = \frac{\delta^2}{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})} - \frac{2\delta z_{\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}},$$

then we have the following result.

Corollary 4.3.1. *The power difference between two-side T_1 test and two-sided TCF test at level α is*

$$\begin{aligned} &\frac{B\delta}{\sqrt{N}\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\left\{\phi(U_{N,1-\alpha/2})\left[-2z_{1-\alpha/2}+\frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right. \\ &\left.+\phi(U_{N,\alpha/2})\left[-2z_{1-\alpha/2}-\frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right\}+O(N^{-min(1,r+1/2)}) \end{aligned} \tag{4.3.20}$$

4.3.1.2 Power of the two-sided T_2 test under local alternative hypothesis

In this Section, we will derive the power function of the two-sided T_2 test and compare it with the power of the two-sided TCF test. Recall that conditional on $\hat{B}$,

$$T_2' = T_2'(U) = exp(2aN^{-1/2}\hat{\gamma}U)$$
$$T_2^{-1}(t) = t - N^{-1}b\hat{\gamma} + O(N^{-2}).$$

Then we can compute the first two terms of the T_2 power function from (4.3.3) as

$$\begin{aligned} & 1 - E\left(\Phi\left\{z_{1-\alpha/2} - T_2'\left[T_2^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right\}\right) \\ = \quad & 1 - E\left(\Phi\left\{z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\left[\exp\left\{\frac{2a\hat{\gamma}}{\sqrt{N}}\left(\frac{z_{1-\alpha/2}}{\sqrt{N}} - \frac{b\hat{\gamma}}{N} + O(N^{-3/2})\right)\right\}\right]\right\}\right) \\ & \text{Apply Taylor expansion to } e^x\text{, we have} \\ = \quad & 1 - E\left(\Phi\left\{z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\left[1 + \frac{2a\hat{\gamma}z_{1-\alpha/2}}{N} + O(N^{-3})\right]\right\}\right) \\ & \text{Apply Taylor expansion to } \Phi(x)\text{, we have} \\ = \quad & 1 - \Phi(U_{N,1-\alpha/2}) + E\left(\frac{\delta 2a\hat{\gamma}z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{N\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-3})\right), \end{aligned} \tag{4.3.21}$$

where $U_{N,1-\alpha/2} = z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$. Now we have the power function of two-sided T_2 test:

$$\begin{aligned}
& 1-\Phi(U_{N,1-\alpha/2}) + \frac{\delta 2aE(\hat{\gamma})z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{N\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} - Q(L_u^{(2)}) + Q(L_l^{(2)}) \\
& +\Phi(U_{N,\alpha/2}) - \frac{\delta 2aE(\hat{\gamma})z_{\alpha/2}\phi(U_{N,\alpha/2})}{N\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-min(1,r+1/2)}) \\
= \quad & 1-\Phi(U_{N,1-\alpha/2}) + \frac{\delta 2a\gamma z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{N\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} - Q(L_u^{(2)}) + Q(L_l^{(2)}) \\
& +\Phi(U_{N,\alpha/2}) - \frac{\delta 2a\gamma z_{\alpha/2}\phi(U_{N,\alpha/2})}{N\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-min(1,r+1/2)}) \\
= \quad & 1-\Phi(U_{N,1-\alpha/2}) + \Phi(U_{N,\alpha/2}) - Q(L_u^{(2)}) + Q(L_l^{(2)}) \\
& +N^{-1/2}[\phi(U_{N,1-\alpha/2}) + \phi(U_{N,\alpha/2})]\frac{N^{-1/2}\delta 2a\gamma z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{4.3.22}$$

Based on the result in (4.3.19), the power difference between two-sided T_2 test and two-sided TCF test can be obtained:

$$\begin{aligned}
& \text{Power of two-sided } T_2 \text{ test - Power of two-sided } TCF \text{ test} \\
& = N^{-1/2}[\phi(U_{N,1-\alpha/2}) + \phi(U_{N,\alpha/2})]\frac{N^{-1/2}\delta 2a\gamma z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + N^{-1/2}A[\phi(U_{N,1-\alpha/2}) - \phi(U_{N,\alpha/2})] \\
& +N^{-1/2}B[\phi(U_{N,1-\alpha/2})(U^2_{N,1-\alpha/2}-1) - \phi(U_{N,\alpha/2})(U^2_{N,\alpha/2}-1)] \\
& - \left\{-N^{-1/2}[A + B(z^2_{\alpha/2}-1)]\right\}[\phi(U_{N,\alpha/2}) - \phi(U_{N,1-\alpha/2})] + O(N^{-min(1,r+1/2)}) \\
& = N^{-1/2}[\phi(U_{N,1-\alpha/2}) + \phi(U_{N,\alpha/2})]\frac{N^{-1/2}\delta 2a\gamma z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + \\
& \frac{B}{\sqrt{N}}[\phi(U_{N,1-\alpha/2})(U^2_{N,1-\alpha/2} - z^2_{1-\alpha/2}) - \phi(U_{N,\alpha/2})(U^2_{N,\alpha/2} - z_{\alpha/2})] + O(N^{-min(1,r+1/2)}).
\end{aligned}$$

Here $a = 1/3$ and $\gamma = 3N^{-1/2}B$. Then Corollary below states this result.

Corollary 4.3.2. *The power difference between two-side T_2 test and two-sided TCF test at level α is*

$$\begin{aligned}
& \frac{B\delta}{\sqrt{N}\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\left\{\phi(U_{N,1-\alpha/2})\left[-2z_{1-\alpha/2} + \frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right. \\
& \left. +\phi(U_{N,\alpha/2})\left[-2z_{1-\alpha/2} - \frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right\} + O(N^{-min(1,r+1/2)})
\end{aligned} \tag{4.3.23}$$

4.3.1.3 Power of the two-sided T_3 test under local alternative hypothesis

In this Section, we will investigate the power difference between the two-sided T_3 test and the two-sided TCF test. Let's first recall that, conditional on $\hat{B}$,

$$T_3' = T_1'(U) = (1+U)^2,$$
$$T_3^{-1}(t) = t - N^{-1}b\hat{\gamma} + O(N^{-1}) \text{ for } t = O(N^{-1/2}).$$

Then the first two terms of the T_3 power function based on (4.3.3) can be derived as:

$$\begin{aligned}
& 1 - E\left(\Phi\left\{z_{1-\alpha/2} - T_3'\left[T_3^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right\}\right) \\
= \quad & 1 - E\left(\Phi\left\{z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\left[(1+\frac{z_{1-\alpha/2}}{\sqrt{N}} - \frac{b\hat{\gamma}}{N} + O(N^{-1}))^2\right]\right\}\right) \\
= \quad & 1 - E\left(\Phi\left\{z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\left[1+\frac{2z_{1-\alpha/2}}{\sqrt{N}} + O(N^{-1})\right]\right\}\right)
\end{aligned}$$

Apply Taylor expansion to $\Phi(x)$, the above term is equal to

$$= \quad 1 - \Phi(U_{N,1-\alpha/2}) + \frac{N^{-1/2}\delta 2z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-1}), \tag{4.3.24}$$

where $U_{N,1-\alpha/2} = z_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$. Then the power function of the two-sided T_3 test becomes

$$\begin{aligned}
& 1-\Phi(U_{N,1-\alpha/2}) + \frac{N^{-1/2}\delta 2z_{1-\alpha/2}\phi(U_{N,1-\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} - Q(L_u^{(3)}) + Q(L_l^{(3)}) \\
& +\Phi(U_{N,\alpha/2}) - \frac{N^{-1/2}\delta 2z_{\alpha/2}\phi(U_{N,\alpha/2})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-min(1,r+1/2)}) \\
= \quad & 1-\Phi(U_{N,1-\alpha/2}) + \Phi(U_{N,\alpha/2}) - Q(L_u^{(3)}) + Q(L_l^{(3)}) \\
& +N^{-1/2}[\phi(U_{N,1-\alpha/2}) + \phi(U_{N,\alpha/2})]\frac{\delta 2z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{4.3.25}$$

Based on the result in (4.3.19), the power difference between two-sided T_3 test and two-sided TCF test can be expressed as:

Power of two-sided T_3 test - Power of two-sided TCF test

$$\begin{aligned}
&= N^{-1/2}[\phi(U_{N,1-\alpha/2}) + \phi(U_{N,\alpha/2})]\frac{\delta 2z_{1-\alpha/2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \\
&+\frac{B}{\sqrt{N}}[\phi(U_{N,1-\alpha/2})(U^2_{N,1-\alpha/2} - z^2_{1-\alpha/2}) - \phi(U_{N,\alpha/2})(U^2_{N,\alpha/2} - z_{\alpha/2})] + O(N^{-min(1,r+1/2)}).
\end{aligned}$$

Then we have the following result.

Corollary 4.3.3. *The power difference between two-side T_3 test and two-sided TCF test at level α is*

$$\frac{\delta}{\sqrt{N}\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\left\{\phi(U_{N,1-\alpha/2})\left[2z_{1-\alpha/2}-2Bz_{1-\alpha/2}+\frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right. \tag{4.3.26}$$
$$\left.+\phi(U_{N,\alpha/2})\left[2z_{1-\alpha/2}-2Bz_{1-\alpha/2}-\frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right\}+O(N^{-min(1,r+1/2)})$$

4.3.1.4 Power comparison of three two-sided T_i tests under local alternative hypothesis

In this section, we will further investigate the power of three two-sided T_i tests under local alternative hypothesis. Recall that the power function of the three two-sided T_i tests are given in equations (4.3.18), (4.3.22) and (4.3.25) for $i = 1, 2, 3$ respectively. Denote the common term of these three equations as

$$\Omega_N = N^{-1/2}z_{1-\alpha/2}\left[\phi(U_{N,1-\alpha/2})+\phi(U_{N,\alpha/2})\right]\frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}. \tag{4.3.27}$$

Denote $H_{two} = -Q(U_{N,1-\alpha/2}) + Q(U_{N,\alpha/2})$. In Section A.7, we showed that

$$-Q(L_u^{(i)}) + Q(L_l^{(i)}) = H_{two} + O(N^{-m}),$$

where $m \geq 1$. Then the three power functions are

- Power of T_1

$$= 1 - \Phi(U_{N,1-\alpha/2}) + \Phi(U_{N,\alpha/2}) + \frac{2B}{\sqrt{N}}\Omega_N + H_{two} + O(N^{-min(1,r+1/2)}), \tag{4.3.28}$$

- Power of T_2

$$= 1 - \Phi(U_{N,1-\alpha/2}) + \Phi(U_{N,\alpha/2}) + \frac{2B}{N}\Omega_N + H_{two} + O(N^{-min(1,r+1/2)}), \tag{4.3.29}$$

- Power of T_3

$$= 1 - \Phi(U_{N,1-\alpha/2}) + \Phi(U_{N,\alpha/2}) + 2\Omega_N + H_{two} + O(N^{-min(1,r+1/2)}). \qquad (4.3.30)$$

It is obvious that the only difference among these three equations (4.3.28), (4.3.29) and (4.3.30) is the coefficient of Ω_N. We can compare the power of three T_i tests by investigating the three coefficients above. We know that B is $O(1)$ under local alternative hypothesis, and Ω_N has order $O(N^{-1/2})$, which sign depends on δ. Besides, we can always arrange the two populations such that B or Ω_N is positive. Without loss of generality, we arrange the two populations and make $\Omega_N > 0$. Then for large N and $B > 0$

$$2\Omega_N > \frac{2B}{\sqrt{N}}\Omega_N > \frac{2B}{N}\Omega_N > 0 \qquad (4.3.31)$$

Therefore, as N gets larger, T_3 test is more powerful than T_1 and T_2 tests. In addition, T_1 test is more powerful than T_2 test. On the other hand, for large N and $B < 0$

$$2\Omega_N > 0 > \frac{2B}{N}\Omega_N > \frac{2B}{\sqrt{N}}\Omega_N \qquad (4.3.32)$$

Therefore, as N gets larger, T_3 test is more powerful than T_1 and T_2 tests. In addition, T_2 test is more powerful than T_1 test.

4.3.1.5 Power comparison between TCF and three two-sided T_i tests under local alternative hypothesis

In this section, we investigate the power difference between the three T_i test and TCF test. Denoting

$$D = B\frac{c_N}{\sqrt{N}}\left[\phi(U_{N,1-\alpha/2})\left(2z_{1-\alpha/2} - c_N\right) + \phi(U_{N,\alpha/2})\left(2z_{1-\alpha/2} + c_N\right)\right], \qquad (4.3.33)$$

where $c_N = \frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}$. Here we have $B = [\lambda(1-\lambda)]^{1/2}\left(\frac{8\lambda-2}{\lambda}\gamma_1 - \frac{6-8\lambda}{1-\lambda}\gamma_2\right)/12$ and $\lambda = \frac{n_1}{n_1+n_2} + O(N^{-r})$ as defined in Theorem (3.1.3). Then the three power differences become

- Power of T_1- power of TCF from (4.3.20):

$$-D + O(N^{-min(1,r+1/2)}), \qquad (4.3.34)$$

- Power of T_2- power of TCF from (4.3.23):

$$-D + O(N^{-min(1,r+1/2)}), \tag{4.3.35}$$

- Power of T_3- power of TCF from (4.3.26):

$$2C_N - D + O(N^{-min(1,r+1/2)}). \tag{4.3.36}$$

Recall that

$$D = B\frac{c_N}{\sqrt{N}}\left[\phi(U_{N,1-\alpha/2})\left(2z_{1-\alpha/2} - c_N\right) + \phi(U_{N,\alpha/2})\left(2z_{1-\alpha/2} + c_N\right)\right]$$

Note that, in Section 3.2.3 we studied the sign of $z_{1-\alpha/2} - c_N$ and presented Figure 3.1 to show the upper bound of $N \leq \frac{1.96^2\sigma^2}{\delta^2\lambda_N(1-\lambda_N)}$, which satisfies the above inequality $z_{1-\alpha/2} - c_N > 0$ when $\alpha = 0.05$. Now, when it comes to the inequality $2z_{1-\alpha/2} - c_N > 0$, the upper bound of N will increase 4 times. In this case, under the local alternative hypothesis and the main focus of this study, without loss of generality we can assume $2z_{1-\alpha/2} - c_N > 0$.

Since $\phi(U_{N,1-\alpha/2}) > \phi(U_{N,\alpha/2}) > 0$, the sign of D depends on the sign of Bc_N. Besides, we can always arrange the two populations such that B or c_N is positive. Without loss of generality, we arrange the two populations and make $c_N > 0$. Then,

- if $B > 0$, we have $D > 0$ and the power of two-sided TCF test is higher than the two-sided T_1 and $T2$ tests;

- if $B < 0$, we have $D < 0$ and the power of two-sided TCF test is smaller than the two-sided T_1 and $T2$ tests.

As shown in (4.3.26), the power difference between two-side T_3 test and two-sided TCF test at level α is:

$$\frac{\delta}{\sqrt{N}\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\left\{\phi(U_{N,1-\alpha/2})\left[2z_{1-\alpha/2} - 2Bz_{1-\alpha/2} + \frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right.$$
$$\left.+\phi(U_{N,\alpha/2})\left[2z_{1-\alpha/2} - 2Bz_{1-\alpha/2} - \frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right]\right\} + O(N^{-min(1,r+1/2)})$$

As we can see, this power difference depends on α, common population standard deviation σ, sample sizes n_1 and n_2, skewness γ_1 and γ_2 and effect size δ. So it is not easy to give an explicit cutoff to pick the one with higher power. However, we can give a rule of thumb cutoff, which is satisfied by the majority of real applications.

Again, without loss of generality, we arrange the two populations and make $\delta > 0$. When the sample size ratio λ is around 0.5 and $\gamma_1 - \gamma_2 \leq 6$, we have $B \leq 1$ and $2z_{1-\alpha/2} - 2Bz_{1-\alpha/2} + \frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} > 0$. Then

$$\begin{aligned}
&\text{Power of two-sided } T_3 \text{ test - Power of two-sided } TCF \text{ test} \\
&= \frac{\delta}{\sqrt{N}\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \left\{ \phi(U_{N,1-\alpha/2}) \left[2z_{1-\alpha/2} - 2Bz_{1-\alpha/2} + \frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \right] \right. \\
&\left. + \phi(U_{N,\alpha/2}) \left[2z_{1-\alpha/2} - 2Bz_{1-\alpha/2} - \frac{B\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \right] \right\} + O(N^{-min(1,r+1/2)}) \\
&\geq \frac{\delta 4 z_{1-\alpha/2}}{\sqrt{N}\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \phi(U_{N,\alpha/2})(1-B) + O_p(N^{-min(1,r+1/2)}) \geq 0.
\end{aligned}$$

Thus the power of the two-sided T_3 test will be larger than the TCF when $B \leq 1$. On the other hand, when population is highly skewed and the sample are very unbalanced, the data can yield a much larger B. Then the power of the two-sided T_3 test will be smaller than the TCF test. This result means the two-sided TCF test can provide more accurate two sample mean comparison when the two populations are highly skewed.

4.3.2 Power of the one-sided upper-tailed T_i test under local alternative hypothesis

Under H_a, the theoretical power of the one-sided upper-tailed T_i test is obtained by:

$$1 - P_{H_0}(T' \leq t^{(i)}_{1-\alpha} - c_N) - Q(t^{(i)}_{1-\alpha}). \tag{4.3.37}$$

Denoting $t^{i}_{1-\alpha} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ as $L^{(i)}_{1-\alpha}$, then the power can be expressed as:

$$\begin{aligned}
&\text{power of one-sided upper-tailed } T_i \text{ test} \\
=&1 - P_{H_0}(\sqrt{N}T_i(U) \leq \sqrt{N}T_i(L^{(i)}_{1-\alpha}/\sqrt{N}) - Q(t^{(i)}_{1-\alpha}),
\end{aligned}$$

where the last equality is due to the fact that $T_i(t)$ are monotone functions. Then replacing $L_{1-\alpha}^{(i)}$ with its definition, we can get

$$\begin{aligned}
&\text{power of one-sided upper-tailed } T_i \text{ test} \\
=&1 - P_{H_0}(\sqrt{N}T_i(U) \leq \sqrt{N}T_i[(t_{1-\alpha}^{(i)} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})})/\sqrt{N}]) - Q(t_{1-\alpha}^{(i)}).
\end{aligned} \tag{4.3.38}$$

Based on the (4.3.3) from Section 4.3.1, we can write

$$\begin{aligned}
&\text{power of one-sided upper-tailed } T_i \text{ test} \\
=&1 - E\left(\Phi\left\{z_{1-\alpha} - T_i^{'}\left[T_i^{-1}(\frac{z_{1-\alpha}}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right\}\right) - Q(t_{1-\alpha}^{(i)}).
\end{aligned} \tag{4.3.39}$$

Next, based on the results from Section 4.3.1.1, 4.3.1.2 and 4.3.1.3, the power functions of the one-sided upper-tailed T_i tests, $i = 1, 2, 3$ can also be obtained. Define

$$\begin{aligned}
C_N^{upper} &= N^{-1/2} z_{1-\alpha}\phi(U_{N,1-\alpha})\frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \\
H_{upper} &= -Q(U_{N,1-\alpha}).
\end{aligned} \tag{4.3.40}$$

Then the power functions are

- Power of the one-sided upper-tailed T_1 tests

$$\begin{aligned}
&1-\Phi(U_{N,1-\alpha}) + \frac{N^{-1/2}\delta 2a\gamma z_{1-\alpha}\phi(U_{N,1-\alpha})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + H_{upper} + O(N^{-min(1,r+1/2)}) \\
=\quad &1-\Phi(U_{N,1-\alpha}) + \tfrac{2B}{\sqrt{N}}C_N^{upper} + H_{upper} + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{4.3.41}$$

- Power of the one-sided upper-tailed T_2 tests

$$\begin{aligned}
&1-\Phi(U_{N,1-\alpha/2}) + \frac{N^{-1}\delta 2a\gamma z_{1-\alpha}\phi(U_{N,1-\alpha})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + H_{upper} + O(N^{-min(1,r+1/2)}) \\
=\quad &1-\Phi(U_{N,1-\alpha}) + \tfrac{2B}{N}C_N^{upper} + H_{upper} + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{4.3.42}$$

- Power of the one-sided upper-tailed T_3 tests

$$\begin{aligned}
&1-\Phi(U_{N,1-\alpha}) + \frac{N^{-1/2}\delta 2z_{1-\alpha}\phi(U_{N,1-\alpha})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + H_{upper} + O(N^{-min(1,r+1/2)}) \\
=\quad &1-\Phi(U_{N,1-\alpha}) + 2C_N^{upper} + H_{upper} + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{4.3.43}$$

For the upper-tailed tests, we always have $\delta > 0$ and $C_N^{upper} > 0$. Then for large N and $B > 0$

$$2C_N^{upper} > \frac{2B}{\sqrt{N}} C_N^{upper} > \frac{2B}{N} C_N^{upper} > 0 \tag{4.3.44}$$

Therefore, as N gets larger, T_3 test is more powerful than T_1 and T_2 tests. In addition, T_1 test is more powerful than T_2 test. On the other hand, for large N and $B < 0$

$$2C_N^{upper} > 0 > \frac{2B}{N} C_N^{upper} > \frac{2B}{\sqrt{N}} C_N^{upper} \tag{4.3.45}$$

Therefore, as N gets larger, T_3 test is more powerful than T_1 and T_2 tests. In addition, T_2 test is more powerful than T_1 test.

Furthermore, the power difference between the one-sided upper-tailed T_i tests and the one-sided upper-tailed TCF test at level α can be obtained as well. Define

$$D^{upper} = B\frac{\Omega}{\sqrt{N}}\phi(U_{N,1-\alpha})\left(2z_{1-\alpha} - c_N\right), \tag{4.3.46}$$

then we have

- Power difference between the one-sided upper-tailed T_1 test and TCF test

$$-D^{upper} + O(N^{-min(1,r+1/2)}); \tag{4.3.47}$$

- Power difference between the one-sided upper-tailed T_2 test and TCF test

$$-D^{upper} + O(N^{-min(1,r+1/2)}); \tag{4.3.48}$$

- Power difference between the one-sided upper-tailed T_3 test and TCF test

$$2C_N^{upper} - D^{upper} + O(N^{-min(1,r+1/2)}). \tag{4.3.49}$$

Following the same discussion for the two-sided test, without loss of generality we can also assume $2z_{1-\alpha} - c_N > 0$ under upper-tailed local alternative hypothesis. Since δ and c_N are both positive under the local alternative of one-sided upper-tailed test, then

- if $B > 0$, we have $D^{upper} > 0$ and the power of two-sided TCF test is higher than the two-sided T_1 and $T2$ tests;
- if $B < 0$, we have $D^{upper} < 0$ and the power of two-sided TCF test is smaller than the two-sided T_1 and $T2$ tests.

Note that the power difference between the one-sided upper-tailed T_3 tests and the one-sided upper-tailed TCF test at level α is:

$$\frac{c_N}{\sqrt{N}}\left\{\phi(U_{N,1-\alpha})\left[2z_{1-\alpha}-2Bz_{1-\alpha}+Bc_N\right]\right\}+O(N^{-min(1,r+1/2)})$$

Based on the same rule of thumb in two-sided case, when $B \leq 1$, we have $2z_{1-\alpha}-2Bz_{1-\alpha}+Bc_N > 0$. Then the power of the one-sided upper-tailed T_3 tests will be larger than the one-sided upper-tailed TCF test.

4.3.3 Power of the one-sided lower-tailed T_i test under local alternative hypothesis

Similarly, the theoretical power of the one-sided lower-tailed T_i test is obtained by:

$$P_{H_0}(T^{'} \leq t_\alpha^{(i)} - c_N) + Q(t_\alpha^{(i)}). \tag{4.3.50}$$

Denoting $t_\alpha^{(i)} - c_N$ as $L_\alpha^{(i)}$, then the power can be expressed as:

$$\begin{aligned}&\text{power of one-sided lower-tailed } T_i \text{ test}\\ =&P_{H_0}(\sqrt{N}T_i(U) \leq \sqrt{N}T_i(L_\alpha^{(i)}/\sqrt{N}) + Q(t_\alpha^{(i)}),\end{aligned}$$

where the last equality is due to the fact that $T_i(t)$ are monotone functions. Then replacing $L_\alpha^{(i)}$ with its definition, we can get

$$\begin{aligned}&\text{power of one-sided lower-tailed } T_i \text{ test}\\ =&P_{H_0}(\sqrt{N}T_i(U) \leq \sqrt{N}T_i[(t_\alpha^{(i)} - \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})})/\sqrt{N}]) + Q(t_\alpha^{(i)}).\end{aligned} \tag{4.3.51}$$

First of all, based on the results (4.3.3) from Section 4.3.1, we have

$$
\begin{aligned}
&\text{power of one-sided lower-tailed } T_i \text{ test} \\
=&E\left(\Phi\left\{z_\alpha + T_i^{'}\left[T_i^{-1}(\frac{z_\alpha}{\sqrt{N}})\right]\delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}\right\}\right) + Q(t_\alpha^{(i)}).
\end{aligned} \tag{4.3.52}
$$

Next, based on the results from Section 4.3.1.1, 4.3.1.2 and 4.3.1.3, the power functions of the one-sided lower-tailed T_i tests, $i = 1, 2, 3$ can also be obtained. Define

$$
\begin{aligned}
C_N^{lower} &= N^{-1/2} z_{1-\alpha}\phi(U_{N,\alpha})\frac{\delta}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} \\
H_{lower} &= Q(U_{N,\alpha}).
\end{aligned} \tag{4.3.53}
$$

Then below are the power functions of the T_i tests:

- Power of the one-sided lower-tailed T_1 test

$$
\begin{aligned}
& \Phi(U_{N,\alpha}) - \frac{N^{-1/2}\delta 2a\gamma z_\alpha \phi(U_{N,\alpha})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + H_{lower} + O(N^{-min(1,r+1/2)}) \\
= \quad & \Phi(U_{N,\alpha}) + \tfrac{2B}{\sqrt{N}} C_N^{lower} + H_{lower} + O(N^{-min(1,r+1/2)})
\end{aligned} \tag{4.3.54}
$$

- Power of the one-sided lower-tailed T_2 test

$$
\begin{aligned}
& \Phi(U_{N,\alpha}) - \frac{N^{-1}\delta 2a\gamma z_\alpha \phi(U_{N,\alpha})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + H_{lower} + O(N^{-min(1,r+1/2)}) \\
= \quad & \Phi(U_{N,\alpha}) + \tfrac{2B}{N} C_N^{lower} + H_{lower} + O(N^{-min(1,r+1/2)})
\end{aligned} \tag{4.3.55}
$$

- Power of the one-sided lower-tailed T_3 test

$$
\begin{aligned}
& \Phi(U_{N,\alpha}) - \frac{N^{-1/2}\delta 2 z_\alpha \phi(U_{N,\alpha})}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} + H_{lower} + O(N^{-min(1,r+1/2)}) \\
= \quad & \Phi(U_{N,\alpha}) + 2C_N^{lower} + H_{lower} + O(N^{-min(1,r+1/2)})
\end{aligned} \tag{4.3.56}
$$

For the lower-tailed tests, we always have $\delta < 0$ and $C_N^{lower} < 0$. Then for large N and $B > 0$

$$
2C_N^{lower} < \frac{2B}{\sqrt{N}} C_N^{lower} < \frac{2B}{N} C_N^{lower} < 0 \tag{4.3.57}
$$

Therefore, as N gets larger, T_2 test is more powerful than T_1 and T_3 tests. In addition, T_1 test is more powerful than T_3 test. On the other hand, for large N and $B < 0$

$$\frac{2B}{\sqrt{N}} C_N^{lower} > \frac{2B}{N} C_N^{lower} > 0 > 2C_N^{lower} \tag{4.3.58}$$

Therefore, as N gets larger, T_1 test is more powerful than T_2 and T_3 tests. In addition, T_2 test is more powerful than T_3 test.

Furthermore, the power difference between the one-sided lower-tailed T_i tests and the one-sided lower-tailed TCF test at level α can be obtained as well. Define

$$D^{lower} = B \frac{\Omega}{\sqrt{N}} \phi(U_{N,\alpha}) \left(2z_{1-\alpha} + \Omega\right), \tag{4.3.59}$$

then we have

- Power difference between the one-sided upper-tailed T_1 test and TCF test

$$-D^{lower} + O(N^{-min(1,r+1/2)}); \tag{4.3.60}$$

- Power difference between the one-sided upper-tailed T_2 test and TCF test

$$-D^{lower} + O(N^{-min(1,r+1/2)}); \tag{4.3.61}$$

- Power difference between the one-sided upper-tailed T_3 test and TCF test

$$2C_N^{lower} - D^{upper} + O(N^{-min(1,r+1/2)}). \tag{4.3.62}$$

Following the same discussion for the two-sided test, without loss of generality we can also assume $2z_{1-\alpha} + \Omega > 0$ under lower-tailed local alternative hypothesis. Since δ and Ω are both negative under the local alternative of one-sided lower-tailed test, then

- if $B > 0$, we have $D^{lower} < 0$ and the power of two-sided TCF test is smaller than the two-sided T_1 and $T2$ tests;
- if $B < 0$, we have $D^{lower} > 0$ and the power of two-sided TCF test is higher than the two-sided T_1 and $T2$ tests.

Note that the power difference between the one-sided lower-tailed T_3 tests and the one-sided lower-tailed TCF test at level α is:

$$\frac{\Omega}{\sqrt{N}}\left\{\phi(U_{N,\alpha})\left[2z_{1-\alpha}-2Bz_{1-\alpha}-B\Omega\right]\right\}+O(N^{-min(1,r+1/2)})$$

When $B \leq 1$, we have $2z_{1-\alpha}-2Bz_{1-\alpha}-B\Omega > 0$. In addition, since $\Omega < 0$, the power of the one-sided lower-tailed T_3 tests will be smaller than the one-sided lower-tailed TCF test.

For given population parameters and fixed sample size, the theoretical type I error rate of the two-sample TCF tests and three T_i test introduced in Section 3.2 and Section 4.1 approach to α by an order of $O(N^{-min(1,r+1/2)})$. The theoretical powers of the four tests depend on terms listed in below:

- How close the $F_T^{(1)}(t)$ to the true distribution of the test statistic T;
- The significance level α;
- The two sample sizes n_1, n_2 and their ratio $\lambda_N = \frac{n_1}{n_1+n_2}$;
- The effect size δ;
- The two population variances σ_1^2 and σ_2^2;
- The two population skewness γ_1 and γ_2.

In next Section 4.4, we will conduct a simulation study aiming to figure out that to what extent, the four new two sample tests can improve the power of the test and maintain the type I error rate under skewness.

4.4 Simulation study

4.4.1 Main purpose of the simulation study

Zhou and Philip (2005) used the coefficient $\frac{A}{\sqrt{N}}$ in equation (2.3.11) in Section 2.3.2.2 to represent the relative skewness in two-sample scenario. Zhou and Philip (2005) found the relative skewness $\frac{A}{\sqrt{N}}$ can affect the coverage accuracy of confidence intervals based on normal approximation. Under our two-sample setting with equal variances, the relative skewness can be calculated as $\gamma = 3N^{-1/2}B$, where B is in equation (3.1.10) in Theorem 3.1.3. So the main purpose of this simulation study is to figure out how the relative skewness affect the test results accuracy and to which extent, the four new tests derived from Edgeworth expansion can improve the test accuracy.

The simulation study compares the type I error rate and the power of seven tests under skewness. The seven tests are pooled two sample t-test, tests by three transformations (T_i), test by Cornish Fisher expansion (TCF), Bootstrap-t test (Davison and Hinkley, 1997; Efron and Tibshirani, 1993) and Bias acceleration bootstrap-t test(BCa) (Efron, 1987). The above seven tests in this simulation study, are testing the population mean difference between two independent populations with equal variances. Furthermore, this simulation study focuses on two-sided hypothesis test, one-sided lower-tailed hypothesis test and one-sided upper-tailed hypothesis test, which assume equal population means under the null hypothesis and unequal population means under the alternative hypothesis.

4.4.2 Detailed settings of simulation study

The two skewed populations are chosen from Gamma family and Log-normal family, which are both right skewed and their relative skewness depends on the value of $\gamma = 3N^{-1/2}B$. Based on the formula of B in Theorem 3.1.3, it is clear that if both samples are skewed, their relative skewness can cancel each other and yield a small value of $\gamma = 3N^{-1/2}B$. To prevent this happen, we let the first population follow Gamma or Log-normal distribution and let the second population follow normal distribution.

Denote normal distribution as $N(\mu_1, \sigma_1^2)$, where μ_1 and σ_1^2 are the mean and variance of the normal distribution. Let $Gamma(\alpha, \beta)$ be the notation for gamma distribution with shape parameter α and rate parameter β. We know, when $\alpha = 1$, the gamma distribution is an exponential distribution and when $\beta = 0.5$, the gamma distribution is a χ^2 distribution. Let $LN(\mu_2, \sigma_2^2)$ be the notation for Log-normal distribution with log-transformed mean μ_2

and log-transformed variance σ_2^2. We summarize these three distributions in Table 4.1

	$N(\mu_1, \sigma_1^2)$	$Gamma(\alpha, \beta)$	$Lognormal(\mu_2, \sigma_2^2)$
parameter	mean=$\mu_1 \in R$	shape=$\alpha > 0$	Location= $\mu_2 \in R$
	variance= $\sigma_1^2 > 0$	rate= $\beta > 0$	scale=$\sigma_2 > 0$
mean	μ_1	α/β	$e^{\mu_2+\sigma_2^2/2}$
variance	σ_1^2	α/β^2	$(e^{\sigma_2^2} - 1)e^{2\mu_2+\sigma_2^2}$
skewness	0	$2/\sqrt{\alpha}$	$(e^{\sigma_2^2} + 2)\sqrt{e^{\sigma_2^2} - 1}$

Table 4.1: *Population Parameters of Three Distribution Families*

There are six pairs of populations used in this simulation study. Under null hypothesis, the setting of each population parameters is listed in Table 4.2 and the corresponding distributions of each pair are presented in Figure 4.1. From Table 4.2, the distributions in pair1, pair3 and pair5 have small common variance, while the distributions from pair2, pair4 and pair6 have big common variance. In addition the second population in pair3, pair4 and pair6 have higher skewness than the second populations in pair1, pair2 and pair5.

	Population2	Population1	γ_1
Pair1	$N(\mu_1 = 1, \sigma_1^2 = 1)$	$Gamma(\alpha = 1, \beta = 1)$	2.0
Pair2	$N(\mu_1 = 8, \sigma_1^2 = 16)$	$Gamma(\alpha = 4, \beta = 0.5)$	1.0
Pair3	$N(\mu_1 = 0.4, \sigma_1^2 = 2)$	$Gamma(\alpha = 0.08, \beta = 0.2)$	7.1
Pair4	$N(\mu_1 = 1.25, \sigma_1^2 = 15.625)$	$Gamma(\alpha = 0.1, \beta = 0.08)$	6.3
Pair5	$N(\mu_1 = e^{0.5/2}, \sigma_1^2 = e^{0.5}(e^{0.5} - 1))$	$Lognormal(\mu_2 = 0, \sigma_2^2 = 0.5)$	2.9
Pair6	$N(\mu_1 = e^{1.5/2}, \sigma_1^2 = e^{1.5}(e^{1.5} - 1))$	$Lognormal(\mu_2 = 0, \sigma_2^2 = 1.5)$	12.1

Table 4.2: *6 Pairs of Population Settings*

Under the alternative hypothesis of one-sided upper-tailed and two-sided tests, a constant in the amount of 0.3σ was added to the first population mean. That is the population mean of the first population is bigger than the second population, and the mean difference is 0.3σ.

Under the alternative hypothesis of one-sided lower-tailed test, a constant in the amount of 0.3σ was subtracted from the first population mean. Again, the reason we choose the value 0.3 is according to Cohen (1988), we are here to check if these seven tests have enough power to detect a small amount of population mean difference. The other population parameter settings are the same with those in Section 3.3.2:

- Significance level $\alpha = 0.05$;
- First population sample size $n_1 = 5, 10, 15, \ldots, 150$;
- $\lambda_N = n_1/N = 0.3, 0.5, 0.8$, here $N = n_1 + n_2$;
- Second population sample size $n_2 = (1 - \lambda_N)N$;
- Effect size: $\delta = (\mu_1 - \mu_2) - Hypothesized(\mu_1 - \mu_2) = 0.3\sigma$, where σ^2 is the common variance.

Note that when $\lambda_N = 0.3$ or $\lambda_N = 0.8$, the two-sample data set becomes highly unbalanced. Same as in Section 3.3.2, there are $10,000$ simulated samples generated for each parameter setting and each sample size. For bootstrap tests, $1,000$ bootstrap samples are resampled from each generated data set.

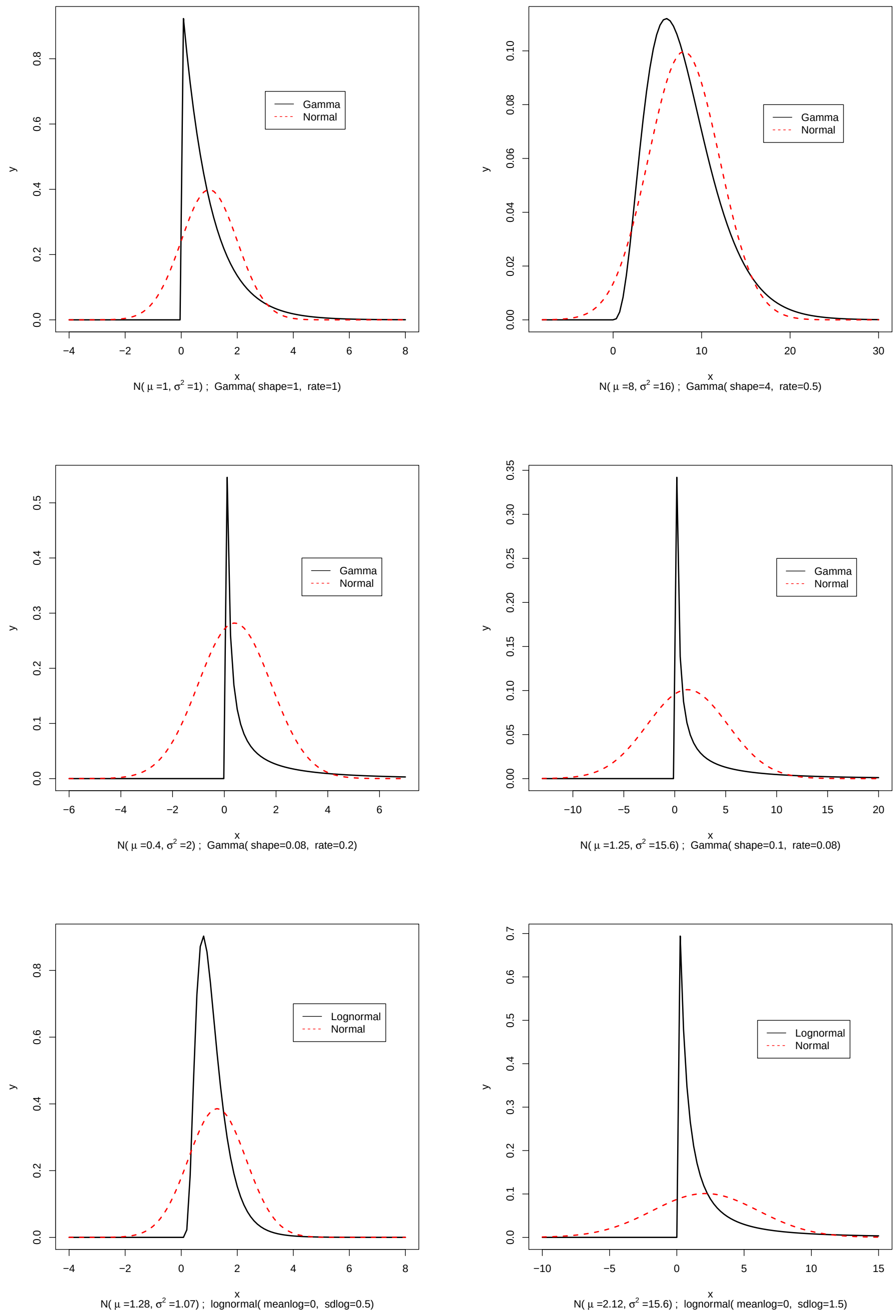

Figure 4.1: *6 pairs of distributions*

4.4.3 Steps of simulation study for each test

For each simulation setting in Subsection 4.4.2, we generate $10,000$ data sets of size $N = n_1 + n_2$ from the two populations respectively, with sample of size n_1 from population 1 and sample of size n_2 from population 2. Then apply the seven tests on each data set. When the alternative hypothesis is true, the two populations having different means, the proportion of times rejecting the null hypothesis out of $10,000$ generated data set is the empirical power of the test. When the null hypothesis is true, the two populations having equal means, the proportion of times rejecting the null hypothesis out of $10,000$ generated data set is the empirical type I error rate.

Below we describe how each test is conducted. The pooled two sample t-test, tests by three transformations and test by Cornish Fisher expansion are calculated as follows:

1. For each generated data set, compute the test statistic T defined in equation (2.1.2) as
$$T = \frac{\bar{Y_1} - \bar{Y_2} - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$
2. For each of the five tests above, reject H_0 if the test statistic T falls into their corresponding rejection regions. Repeat this testing process for each of the $10,000$ generated data set.

Follow the same procedure in Subsubsection 2.2.2.3, the empirical type I error rate and power of the bootstrap-t test can be computed by the following step:

1. Draw $B = 1,000$ bootstrap samples of size $N = n_1 + n_2$ with replacement from each of the $10,000$ generated data set, with sample of size n_1 from sample one and sample of size n_2 from sample two.
2. For each bootstrap sample, compute
$$T_b^* = \frac{\bar{X}_{1b}^* - \bar{X}_{2b}^* - \bar{X}_{1n} + \bar{X}_{2n}}{S_{p_b}^*\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\bar{X}_{1b}^*$ and $\bar{X}_{2b}^*$ are the bth bootstrap sample means of sample one and sample two respectively; $\bar{X}_{1n}$ and $\bar{X}_{2n}$ are the sample means for original sample one and original sample two respectively; $S_{p_b}^* = \sqrt{\frac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}}$ is the pooled two sample standard deviation of bth bootstrap sample; $S_i^* = \sqrt{\frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}^* - \bar{X}_{ib}^*)^2}$, $i = 1, 2$ are the bth bootstrap sample standard deviation for sample i.

3. Estimate the $\alpha/2^{th}$ percentile of test statistic T by the value $\hat{t}_{\alpha/2}$ such that

$$B^{-1}\sum_{b=1}^{B} I(T_b^* \leq \hat{t}_{\alpha/2}) = \alpha/2$$

4. The rejection region of bootstrap-t test with significance level α is then:

$$T \leq \hat{t}_{\alpha/2} \ \ or \ \ T \geq \hat{t}_{1-\alpha/2}.$$

   Reject H_0 is the test statistic T falls into the rejection region above.

Follow the same procedure in Subsubsection 2.2.2.4, the empirical type I error rate and power of the BCa test can be obtained by the following step:

1. Draw $B = 1,000$ bootstrap samples of size $N = n_1 + n_2$ with replacement from each of the $10,000$ generated data set, with sample of size n_1 from sample one and sample of size n_2 from sample two.

2. For each bootstrap sample, compute the BCa confidence interval of $\mu_1 - \mu_2$ introduced in Subsubsection 2.2.2.4, which has a form

$$\left(\hat{G}^{-1}(\Phi(z_{[\alpha/2]})), \hat{G}^{-1}(\Phi(z_{[1-\alpha/2]}))\right).$$

   Equate the lower and upper bounds of this interval with the t-statistic based confidence intervals for $\mu_1 - \mu_2$:

$$\left(\bar{X}_{1n} - \bar{X}_{2n} - t_{1-\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X}_{1n} - \bar{X}_{2n} - t_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right).$$

3. By solving the above the equations, the rejection region of BCa test with significance level α is then:

$$T \leq \frac{\bar{X}_{1n} - \bar{X}_{2n} - \hat{G}^{-1}(\Phi(z_{[1-\alpha/2]}))}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \text{ or } T \geq \frac{\bar{X}_{1n} - \bar{X}_{2n} - \hat{G}^{-1}(\Phi(z_{[\alpha/2]}))}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}. \tag{4.4.1}$$

Reject H_0 if the test statistic T falls into the rejection region above.

So far, we described the steps to compute type I error rate and power of each seven tests in this simulation study. In the following section, we will present the results of the simulation study with differen levels of population skewness, sample size ratio and common variance.

4.4.4 Simulation results for two-sided test

In this simulation study we have six pairs of population settings shown in Table 4.2. The population 1 of the first four pairs follows Gamma distribution and the population 1 of the last two pairs follows Log-normal distribution. The population 2 of all six pairs follows normal distribution.

The simulation results consist of two parts. The first part includes the theoretical results of the type I error and the power of the five two-sample tests including pooled two-sample t-test, TCF test and three T_i tests based on transformations. Recall that these five two-sample tests have the same test statistic T defined in equation (2.1.2). Therefore, our new approximation of T in 3.1.13 based on the first order Edgeworth expansion can be used to calculated the type I error and power of each test. Note that, for the two-sided test, we have already provided the forms of the $(\alpha/2)^{th}$ and $(1-\alpha/2)^{th}$ percentiles of the above five two-sample tests in Chapter 3 and 4. Thus, without generating any data sets, the theoretical results can be obtained on the basis of the two percentiles calculated by the population parameters from each simulation setting:

$$\text{Type I error rate} = 1 - F_T^{(1)}(T_{1-\alpha/2}) + F_T^{(1)}(T_{\alpha/2})$$
$$\text{Power} = 1 - F_T^{(1)}(T_{1-\alpha/2} - \delta/\sqrt{\sigma^2(\tfrac{1}{n_1}+\tfrac{1}{n_2})}) + F_T^{(1)}(T_{\alpha/2} - \delta/\sqrt{\sigma^2(\tfrac{1}{n_1}+\tfrac{1}{n_2})})$$

However, we can not get the theoretical results of the Bootstrap t-test and BCa test, because these two methods need to resample from a data set.

The second part of this simulation study is the empirical results from seven tests mentioned in Section 4.4.1. And the empirical type I error rate and power of these seven tests were obtained by calculating the proportion of rejections in each simulation setting under H_0 and H_a respectively.

The results of the theoretical and empirical type I error rate and power of the seven tests from each pair of population settings were presented with twelve figures in Figures 4.2 through 4.13 and eighteen tables from Table 4.3 to Table 4.20. In the further discussion, we will refer the simulation results for the first pair of population setting as "$Pair1$", the second pair of population setting as "$Pair2$" and so forth. The results of $Pair1$ will be given first then followed by the results from $Pair2$ and so on.

$Pair1$ has a distribution combination with small skewness and small common variance. Figure 4.2 is the theoretical type I error rate and power for $Pair1$ from five tests including pooled two sample t-test denoted as "T", the three transformation-based two sample tests denoted as "T_1", "T_2", "T_3" respectively and Cornish Fisher expansion based two sample test denoted as "CF". There are six panels in Figure 4.2. The top three panels are the theoretical type I error rate of five tests with $\lambda = 0.3, 0.5, 0.8$ respectively. The bottom three panels provide the corresponding theoretical power of each test. From Figure 4.2, under every level of λ, all five tests have a fairly stable theoretical type I error rate, which is close to $\alpha = 0.05$ for $n_1 \geq 20$, and the T_3 test gives a constantly higher theoretical power than the other four tests. When $\lambda > 0.5$, the theoretical power of T_1 and CF becomes higher than T and T_2.

The six panels in Figure 4.3 give the corresponding empirical type I error rate and power for the tests in Figure 4.2. Besides the five tests above, Bootstrap-t test denoted as "$bootstrap$" and BCa test denoted as "BCa" are also included. For the empirical results, the BCa test gives higher type I error rate than the other tests when sample size is small. As the sample size increases, the type I error rate of all seven tests reduced to α. When $\lambda = 0.8$, the type I error rate of BCa test became much larger than the other six tests. The patterns of empirical power in Figure 4.3 support the theoretical power results in Figure 4.2. The pooled two-sample t-test gives the lowest power while T_3 test provides the highest power across all levels of λ. When $\lambda > 0.3$, TCF test and BCa test give the second largest power. The numerical results of $Pair1$ empirical power were presented in Tables 4.3, 4.4 and 4.5.

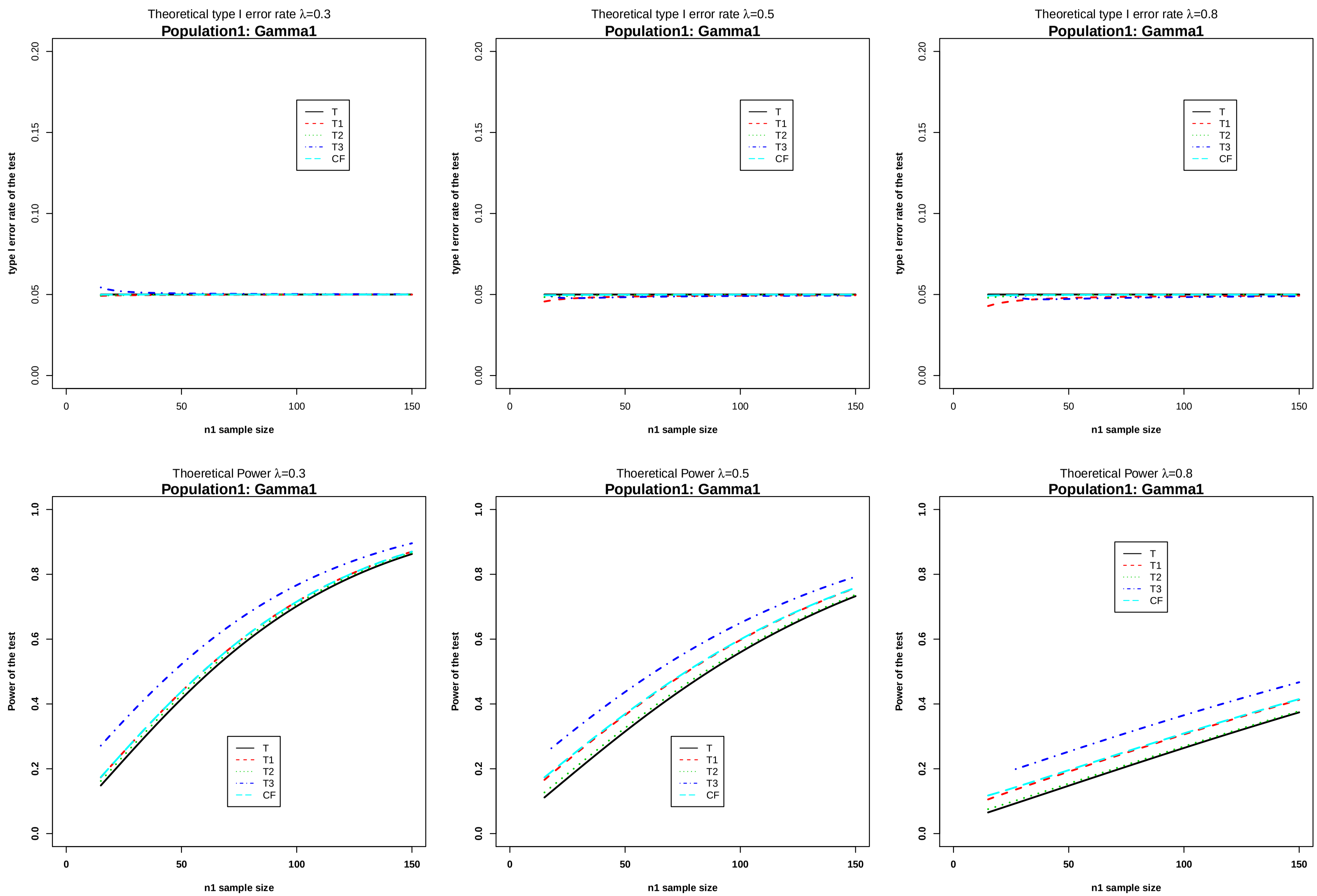

Figure 4.2: *Theoretical Power and Type I Error rate for Pair1*

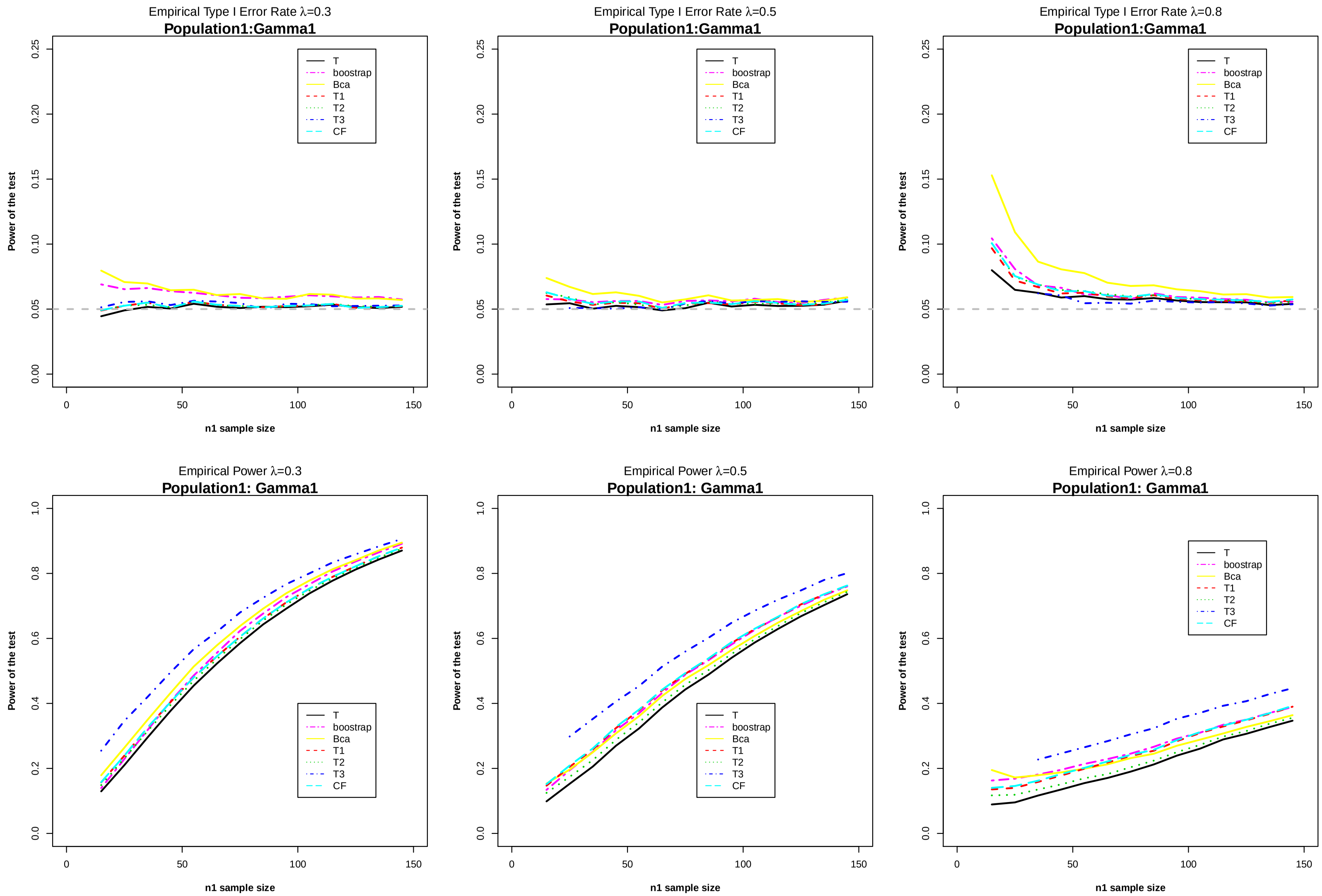

Figure 4.3: *Empirical Power and Type I Error rate for Pair1*

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.044	0.044	0.130	0.069	0.140	0.080	0.179	0.049	0.157	0.049	0.149	0.051	0.255	0.049	0.158
25	0.039	0.049	0.210	0.065	0.229	0.071	0.264	0.052	0.239	0.053	0.231	0.056	0.346	0.053	0.239
35	0.034	0.052	0.295	0.066	0.318	0.070	0.349	0.055	0.325	0.053	0.314	0.056	0.420	0.055	0.325
45	0.031	0.050	0.377	0.064	0.407	0.064	0.432	0.051	0.404	0.051	0.395	0.053	0.496	0.051	0.405
55	0.029	0.054	0.455	0.062	0.486	0.065	0.513	0.056	0.480	0.056	0.470	0.056	0.567	0.056	0.480
65	0.027	0.052	0.522	0.061	0.556	0.061	0.579	0.053	0.545	0.053	0.536	0.056	0.620	0.053	0.545
75	0.026	0.051	0.585	0.059	0.623	0.062	0.638	0.052	0.605	0.052	0.598	0.055	0.680	0.052	0.605
85	0.025	0.052	0.643	0.058	0.676	0.058	0.692	0.051	0.661	0.051	0.655	0.050	0.725	0.051	0.661
95	0.023	0.051	0.692	0.059	0.727	0.058	0.739	0.052	0.711	0.052	0.705	0.054	0.767	0.052	0.711
105	0.022	0.052	0.738	0.061	0.767	0.062	0.778	0.052	0.753	0.053	0.747	0.054	0.800	0.053	0.753
115	0.022	0.053	0.778	0.060	0.806	0.061	0.813	0.054	0.789	0.054	0.785	0.052	0.834	0.054	0.789
125	0.021	0.052	0.812	0.059	0.837	0.058	0.841	0.051	0.822	0.052	0.818	0.052	0.859	0.051	0.822
135	0.020	0.051	0.843	0.059	0.865	0.058	0.871	0.051	0.853	0.052	0.849	0.053	0.883	0.051	0.853
145	0.020	0.052	0.870	0.057	0.891	0.057	0.895	0.052	0.879	0.052	0.876	0.053	0.906	0.052	0.879

Table 4.3: *Pair1 Empirical Power when* $\lambda = 0.3$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.179	0.054	0.099	0.058	0.134	0.074	0.153	0.060	0.147	0.062	0.125	0.988	0.998	0.063	0.152
25	0.158	0.054	0.153	0.057	0.194	0.067	0.197	0.056	0.205	0.059	0.174	0.051	0.298	0.058	0.209
35	0.144	0.050	0.206	0.055	0.251	0.062	0.250	0.053	0.258	0.053	0.226	0.051	0.352	0.054	0.262
45	0.132	0.052	0.270	0.056	0.317	0.063	0.308	0.055	0.324	0.055	0.288	0.050	0.406	0.056	0.327
55	0.123	0.051	0.323	0.056	0.370	0.060	0.361	0.054	0.379	0.053	0.342	0.051	0.454	0.056	0.381
65	0.115	0.049	0.387	0.053	0.432	0.055	0.423	0.050	0.440	0.051	0.403	0.050	0.513	0.051	0.442
75	0.108	0.051	0.443	0.056	0.488	0.058	0.475	0.054	0.491	0.053	0.458	0.054	0.560	0.055	0.493
85	0.103	0.055	0.489	0.057	0.533	0.061	0.518	0.055	0.537	0.056	0.502	0.056	0.602	0.055	0.538
95	0.098	0.052	0.540	0.056	0.581	0.057	0.564	0.054	0.586	0.054	0.552	0.054	0.648	0.054	0.589
105	0.094	0.053	0.587	0.058	0.626	0.057	0.608	0.056	0.629	0.054	0.597	0.056	0.685	0.056	0.630
115	0.091	0.052	0.629	0.055	0.666	0.058	0.648	0.054	0.665	0.054	0.638	0.056	0.719	0.055	0.666
125	0.087	0.052	0.668	0.055	0.701	0.056	0.684	0.053	0.705	0.054	0.678	0.056	0.748	0.053	0.706
135	0.084	0.053	0.703	0.057	0.733	0.056	0.719	0.054	0.735	0.055	0.711	0.056	0.780	0.054	0.736
145	0.082	0.056	0.736	0.059	0.760	0.059	0.748	0.057	0.762	0.058	0.743	0.056	0.801	0.057	0.763

Table 4.4: *Pair1 Empirical Power when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.248	0.080	0.089	0.104	0.163	0.153	0.195	0.097	0.135	0.101	0.117	1.000	1.000	0.101	0.140
25	0.221	0.065	0.096	0.081	0.169	0.109	0.172	0.072	0.140	0.075	0.119	0.741	0.904	0.075	0.146
35	0.197	0.062	0.117	0.068	0.181	0.086	0.179	0.067	0.159	0.069	0.135	0.062	0.227	0.069	0.162
45	0.183	0.059	0.135	0.066	0.196	0.081	0.188	0.062	0.178	0.065	0.151	0.060	0.246	0.064	0.183
55	0.169	0.060	0.155	0.061	0.214	0.078	0.199	0.063	0.199	0.064	0.170	0.054	0.265	0.064	0.202
65	0.160	0.058	0.171	0.061	0.229	0.070	0.213	0.060	0.218	0.061	0.183	0.055	0.284	0.060	0.221
75	0.150	0.057	0.190	0.058	0.245	0.068	0.232	0.058	0.238	0.060	0.204	0.054	0.305	0.059	0.242
85	0.143	0.058	0.212	0.062	0.267	0.068	0.245	0.061	0.254	0.061	0.224	0.056	0.324	0.061	0.257
95	0.136	0.057	0.239	0.059	0.292	0.065	0.270	0.058	0.284	0.059	0.250	0.056	0.353	0.059	0.287
105	0.131	0.055	0.261	0.059	0.310	0.064	0.288	0.057	0.309	0.057	0.272	0.055	0.371	0.058	0.312
115	0.126	0.055	0.289	0.058	0.334	0.061	0.308	0.056	0.329	0.057	0.298	0.056	0.393	0.057	0.332
125	0.122	0.055	0.306	0.057	0.350	0.061	0.328	0.056	0.348	0.056	0.315	0.054	0.407	0.057	0.350
135	0.117	0.053	0.327	0.055	0.371	0.059	0.345	0.055	0.368	0.055	0.337	0.053	0.429	0.056	0.370
145	0.114	0.054	0.347	0.056	0.389	0.059	0.364	0.057	0.390	0.056	0.356	0.054	0.447	0.057	0.393

Table 4.5: *Pair1 Empirical Power when* $\lambda = 0.8$

$Pair2$ has a distribution combination with small skewness and big common variance. Figure 4.4 is the theoretical type I error rate and power for $Pair2$ and Figure 4.5 gives the corresponding empirical results. The numerical results of empirical power were presented in Table 4.6, 4.7 and 4.8. The simulation results of $Pair2$ is very similar to $Pair1$. All the discussions from $Pair1$ still hold for the results of $Pair2$.

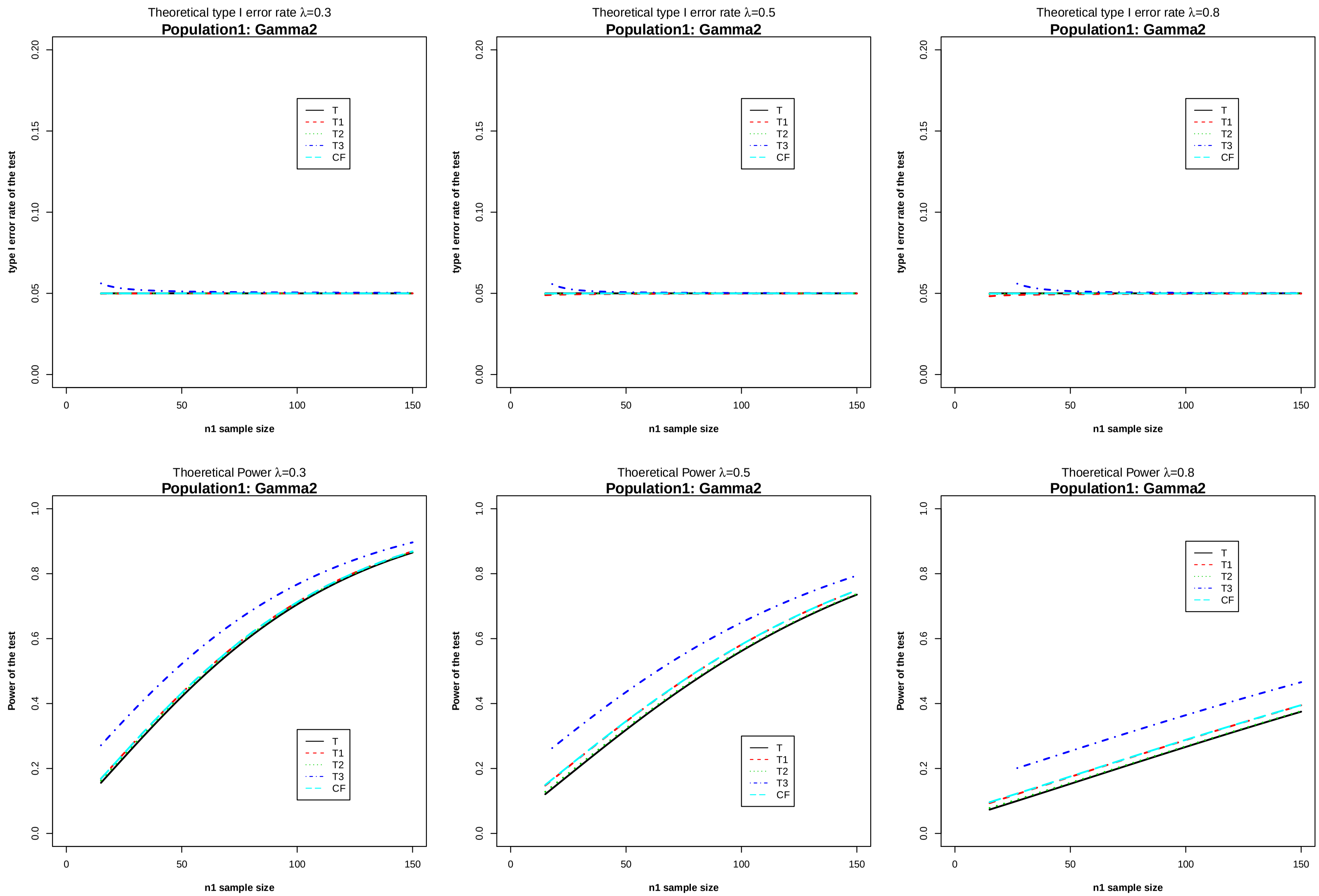

Figure 4.4: *Theoretical Power and Type I Error rate for Pair2*

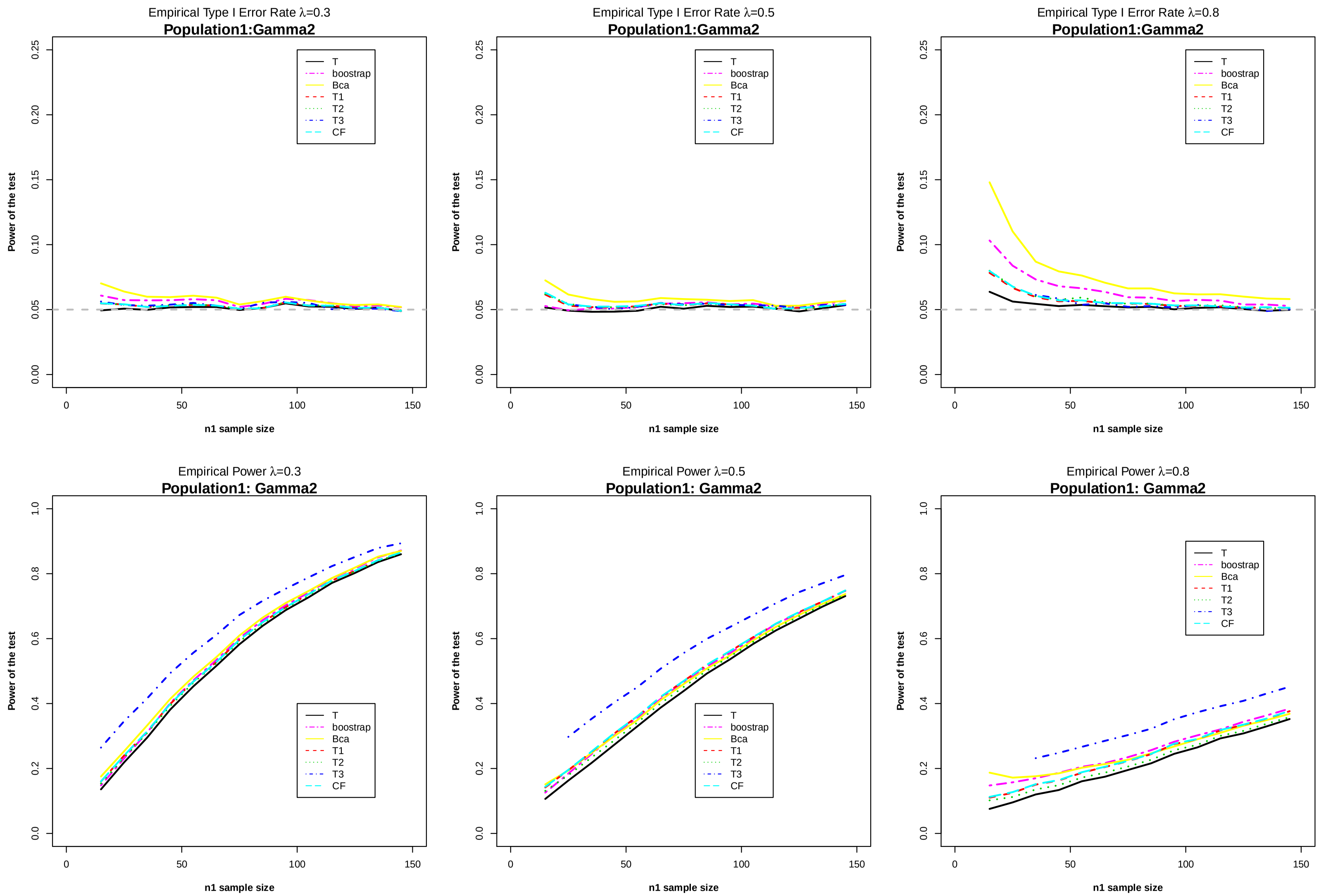

Figure 4.5: *Empirical Power and Type I Error rate for Pair2*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.024	0.049	0.136	0.061	0.148	0.070	0.175	0.055	0.160	0.055	0.154	0.056	0.264	0.055	0.160
25	0.021	0.051	0.219	0.057	0.231	0.064	0.253	0.054	0.240	0.054	0.234	0.053	0.345	0.054	0.240
35	0.018	0.050	0.295	0.057	0.311	0.060	0.333	0.052	0.314	0.052	0.309	0.053	0.416	0.052	0.314
45	0.016	0.052	0.381	0.057	0.401	0.060	0.414	0.053	0.399	0.053	0.393	0.054	0.494	0.053	0.399
55	0.015	0.052	0.453	0.058	0.471	0.061	0.483	0.054	0.468	0.054	0.464	0.055	0.556	0.054	0.468
65	0.014	0.052	0.516	0.057	0.534	0.059	0.543	0.053	0.528	0.053	0.525	0.053	0.612	0.053	0.528
75	0.013	0.050	0.583	0.052	0.601	0.054	0.610	0.050	0.596	0.051	0.591	0.050	0.673	0.051	0.596
85	0.013	0.051	0.639	0.054	0.657	0.056	0.664	0.051	0.650	0.051	0.646	0.055	0.716	0.051	0.650
95	0.012	0.055	0.688	0.058	0.703	0.060	0.710	0.056	0.697	0.056	0.693	0.056	0.754	0.056	0.697
105	0.012	0.052	0.727	0.057	0.742	0.057	0.747	0.053	0.737	0.053	0.734	0.055	0.789	0.053	0.737
115	0.011	0.052	0.771	0.055	0.786	0.055	0.786	0.053	0.779	0.053	0.776	0.050	0.823	0.053	0.779
125	0.011	0.050	0.802	0.052	0.816	0.053	0.819	0.051	0.809	0.051	0.807	0.051	0.851	0.051	0.809
135	0.010	0.051	0.836	0.053	0.850	0.054	0.852	0.051	0.841	0.052	0.839	0.051	0.879	0.051	0.841
145	0.010	0.049	0.860	0.052	0.872	0.052	0.870	0.049	0.865	0.049	0.863	0.049	0.894	0.049	0.865

Table 4.6: *Pair2 Empirical Power when* $\lambda = 0.3$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.096	0.052	0.106	0.053	0.126	0.072	0.151	0.062	0.142	0.062	0.131	1.000	1.000	0.063	0.144
25	0.085	0.049	0.163	0.050	0.184	0.062	0.194	0.053	0.195	0.054	0.179	0.054	0.297	0.054	0.196
35	0.077	0.048	0.217	0.051	0.244	0.058	0.248	0.052	0.251	0.052	0.234	0.052	0.353	0.052	0.252
45	0.070	0.048	0.274	0.051	0.302	0.056	0.300	0.052	0.308	0.052	0.287	0.050	0.404	0.052	0.309
55	0.065	0.049	0.331	0.052	0.357	0.056	0.350	0.052	0.361	0.052	0.342	0.052	0.451	0.053	0.361
65	0.060	0.052	0.388	0.054	0.414	0.059	0.410	0.055	0.420	0.054	0.401	0.055	0.507	0.055	0.421
75	0.057	0.051	0.439	0.055	0.467	0.058	0.460	0.053	0.469	0.053	0.452	0.054	0.556	0.053	0.470
85	0.053	0.053	0.492	0.056	0.514	0.058	0.508	0.055	0.519	0.054	0.501	0.055	0.599	0.055	0.520
95	0.051	0.052	0.536	0.054	0.555	0.057	0.548	0.054	0.561	0.054	0.545	0.053	0.636	0.054	0.562
105	0.049	0.052	0.583	0.055	0.601	0.057	0.595	0.053	0.604	0.053	0.590	0.053	0.673	0.053	0.605
115	0.047	0.051	0.625	0.053	0.645	0.053	0.636	0.050	0.646	0.052	0.633	0.053	0.709	0.050	0.646
125	0.045	0.049	0.662	0.051	0.679	0.053	0.673	0.051	0.682	0.050	0.669	0.051	0.743	0.051	0.682
135	0.044	0.051	0.698	0.054	0.714	0.055	0.707	0.052	0.714	0.052	0.703	0.053	0.770	0.052	0.714
145	0.042	0.053	0.731	0.057	0.747	0.057	0.737	0.054	0.748	0.054	0.736	0.054	0.796	0.054	0.748

Table 4.7: *Pair2 Empirical Power when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.130	0.064	0.076	0.103	0.148	0.148	0.187	0.078	0.110	0.080	0.102	1.000	1.000	0.080	0.112
25	0.118	0.056	0.096	0.084	0.157	0.110	0.172	0.067	0.125	0.068	0.112	0.980	0.993	0.068	0.127
35	0.105	0.054	0.120	0.073	0.170	0.087	0.176	0.060	0.150	0.060	0.135	0.061	0.232	0.061	0.152
45	0.097	0.053	0.134	0.068	0.187	0.079	0.185	0.056	0.163	0.058	0.149	0.058	0.248	0.057	0.164
55	0.089	0.054	0.161	0.066	0.205	0.076	0.202	0.057	0.187	0.059	0.172	0.054	0.267	0.057	0.189
65	0.083	0.053	0.175	0.064	0.217	0.071	0.214	0.055	0.205	0.056	0.187	0.055	0.285	0.055	0.205
75	0.078	0.052	0.196	0.059	0.235	0.066	0.227	0.055	0.222	0.054	0.206	0.052	0.303	0.055	0.223
85	0.075	0.052	0.216	0.059	0.257	0.066	0.247	0.054	0.244	0.055	0.227	0.053	0.323	0.054	0.245
95	0.071	0.050	0.245	0.057	0.282	0.062	0.268	0.053	0.275	0.053	0.256	0.052	0.352	0.053	0.276
105	0.068	0.051	0.265	0.058	0.302	0.062	0.290	0.053	0.290	0.053	0.273	0.054	0.373	0.053	0.291
115	0.065	0.052	0.293	0.057	0.320	0.062	0.310	0.052	0.317	0.053	0.301	0.052	0.392	0.052	0.318
125	0.063	0.050	0.308	0.054	0.344	0.060	0.332	0.051	0.334	0.052	0.316	0.051	0.408	0.052	0.335
135	0.060	0.049	0.330	0.054	0.363	0.058	0.349	0.052	0.353	0.051	0.338	0.049	0.431	0.052	0.354
145	0.059	0.050	0.352	0.053	0.384	0.058	0.368	0.051	0.376	0.052	0.358	0.050	0.452	0.051	0.378

Table 4.8: *Pair2 Empirical Power when* $\lambda = 0.8$

$Pair3$ has a distribution combination with big skewness and small common variance. Figure 4.6 gives the theoretical type I error rate and power for $Pair3$ and Figure 4.7 gives the corresponding empirical results. The numerical results of empirical power were presented in Tables 4.9, 4.10 and 4.11. Since $Pair3$ has bigger skewness than $Pair1$ and $Pair2$, the simulation results are quite different. From Figure 4.6, across all levels of λ the theoretical type I error rate of five tests are all smaller or equal to α. The pooled two-sample t-test still provides the lowest power. When $\lambda = 0.3$, T_3 test gives the highest power followed by the TCF and T_1 test. When $\lambda > 0.5$, TCF test becomes the test with the largest power and T_1 test gives the second largest followed by the third largest from T_3 test.

Note that the theoretical results in Figure 4.6 are consistent with the empirical results in Figure 4.7. When $\lambda = 0.3$, the type I error rates of Bootstrap t-test and BCa test are larger than 0.05 and close to 0.1. While the type I error rates of all the other tests are smaller than $\alpha = 0.05$. BCa test, T_3 test and Bootstrap-t test give higher power than the other tests. When $\lambda = 0.5$ and $n_1 > 15$, the type I error rate of all the tests are close to $\alpha = 0.05$ except BCa test. The power of TCF test and T_3 test are higher than the power of BCa test and Bootstrap-t test. As λ reaches 0.8, i.e., there are 80% of the data coming from the skewed population 1, the type I error rate of all the tests are bigger than 0.10 when n_1 is small. Among seven tests, Bootstrap-t test and BCa test have a smaller type I error rate than the other tests. However, the Bootstrap-t test and BCa test also provide smaller power comparing to the TCF, T_3 and T_1 tests. Among all the seven tests, the pooled two-sample t-test offers smallest power of the test across all levels of λ.

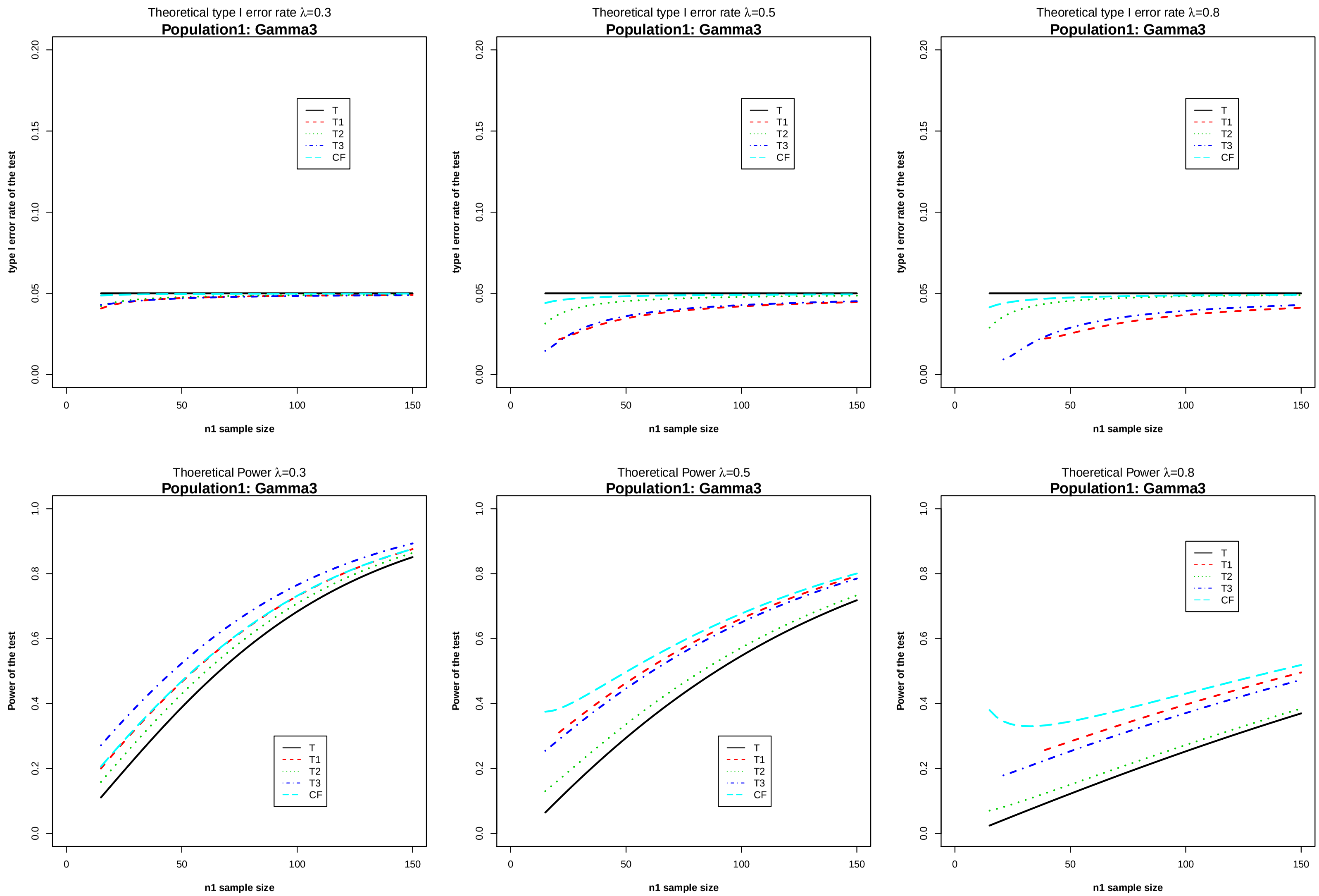

Figure 4.6: *Theoretical Power and Type I Error rate for Pair3*

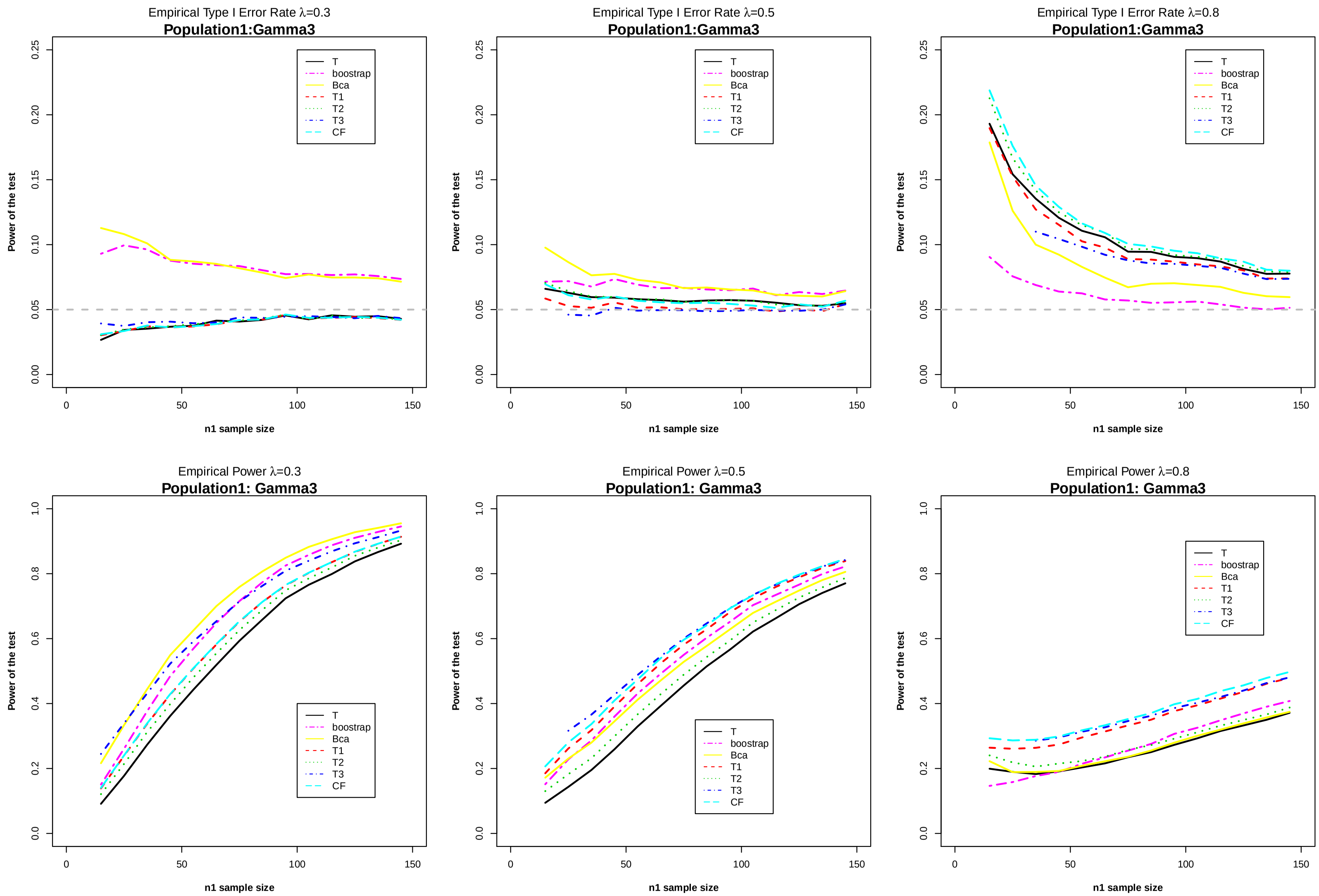

Figure 4.7: *Empirical Power and Type I Error rate for Pair3*

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.096	0.027	0.091	0.093	0.151	0.113	0.216	0.030	0.139	0.030	0.121	0.039	0.245	0.031	0.140
25	0.090	0.034	0.177	0.100	0.261	0.108	0.334	0.034	0.239	0.035	0.214	0.037	0.339	0.034	0.240
35	0.080	0.035	0.273	0.096	0.376	0.101	0.444	0.037	0.338	0.036	0.312	0.040	0.431	0.038	0.339
45	0.077	0.037	0.362	0.088	0.484	0.088	0.549	0.036	0.428	0.037	0.399	0.041	0.523	0.036	0.429
55	0.074	0.037	0.442	0.085	0.569	0.087	0.625	0.037	0.508	0.039	0.480	0.040	0.591	0.037	0.509
65	0.070	0.042	0.519	0.084	0.648	0.085	0.700	0.039	0.582	0.040	0.555	0.040	0.653	0.039	0.583
75	0.068	0.041	0.593	0.084	0.716	0.082	0.759	0.042	0.652	0.042	0.626	0.044	0.715	0.042	0.654
85	0.066	0.042	0.659	0.080	0.774	0.078	0.807	0.042	0.713	0.043	0.689	0.044	0.763	0.042	0.714
95	0.063	0.045	0.724	0.077	0.825	0.074	0.849	0.046	0.764	0.046	0.749	0.044	0.809	0.046	0.765
105	0.061	0.042	0.766	0.077	0.858	0.077	0.882	0.043	0.802	0.043	0.785	0.045	0.840	0.044	0.803
115	0.060	0.046	0.799	0.077	0.887	0.075	0.906	0.044	0.835	0.045	0.819	0.044	0.869	0.044	0.836
125	0.058	0.044	0.838	0.077	0.910	0.075	0.928	0.044	0.867	0.043	0.855	0.043	0.894	0.044	0.868
135	0.057	0.045	0.867	0.076	0.929	0.074	0.941	0.043	0.892	0.044	0.880	0.045	0.912	0.044	0.892
145	0.056	0.043	0.892	0.074	0.945	0.071	0.955	0.042	0.914	0.042	0.903	0.043	0.933	0.042	0.914

Table 4.9: *Pair3 Empirical Power when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.401	0.066	0.094	0.071	0.152	0.098	0.172	0.059	0.185	0.070	0.130	0.536	0.872	0.070	0.207
25	0.369	0.063	0.143	0.072	0.228	0.086	0.231	0.053	0.262	0.064	0.182	0.046	0.316	0.061	0.280
35	0.346	0.060	0.195	0.068	0.286	0.076	0.280	0.051	0.318	0.060	0.232	0.045	0.366	0.058	0.337
45	0.327	0.059	0.259	0.073	0.360	0.077	0.345	0.056	0.391	0.060	0.299	0.051	0.427	0.060	0.410
55	0.310	0.058	0.329	0.069	0.431	0.073	0.412	0.051	0.460	0.058	0.367	0.049	0.488	0.057	0.474
65	0.298	0.057	0.393	0.066	0.492	0.071	0.471	0.052	0.524	0.058	0.429	0.050	0.545	0.056	0.539
75	0.287	0.056	0.456	0.067	0.550	0.066	0.529	0.050	0.581	0.056	0.489	0.050	0.600	0.055	0.596
85	0.276	0.057	0.515	0.066	0.603	0.067	0.578	0.050	0.629	0.056	0.544	0.049	0.647	0.055	0.641
95	0.267	0.057	0.566	0.065	0.652	0.066	0.628	0.050	0.681	0.057	0.594	0.049	0.694	0.054	0.693
105	0.260	0.057	0.622	0.066	0.703	0.065	0.679	0.051	0.724	0.057	0.649	0.050	0.735	0.053	0.734
115	0.253	0.055	0.663	0.061	0.735	0.061	0.714	0.048	0.760	0.054	0.688	0.049	0.766	0.051	0.768
125	0.246	0.053	0.706	0.064	0.766	0.061	0.748	0.050	0.789	0.054	0.727	0.049	0.794	0.054	0.797
135	0.241	0.053	0.741	0.062	0.799	0.060	0.780	0.049	0.817	0.053	0.758	0.050	0.822	0.052	0.823
145	0.235	0.055	0.770	0.065	0.822	0.064	0.806	0.054	0.839	0.056	0.786	0.054	0.842	0.057	0.846

Table 4.10: *Pair3 Empirical Power when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.549	0.193	0.199	0.090	0.146	0.179	0.222	0.190	0.264	0.213	0.240	1.000	1.000	0.219	0.293
25	0.515	0.154	0.190	0.076	0.158	0.126	0.189	0.153	0.261	0.167	0.219	0.167	0.374	0.176	0.286
35	0.476	0.135	0.183	0.069	0.177	0.100	0.189	0.127	0.264	0.142	0.206	0.110	0.286	0.145	0.288
45	0.456	0.121	0.191	0.064	0.190	0.092	0.192	0.115	0.274	0.125	0.215	0.104	0.296	0.129	0.299
55	0.431	0.111	0.203	0.062	0.214	0.083	0.208	0.103	0.295	0.115	0.223	0.098	0.313	0.116	0.318
65	0.415	0.106	0.216	0.058	0.234	0.075	0.222	0.098	0.314	0.109	0.236	0.092	0.327	0.109	0.333
75	0.397	0.095	0.235	0.057	0.255	0.067	0.236	0.089	0.333	0.097	0.258	0.088	0.346	0.101	0.352
85	0.385	0.094	0.250	0.055	0.276	0.070	0.255	0.088	0.350	0.096	0.272	0.086	0.362	0.099	0.371
95	0.371	0.091	0.273	0.056	0.307	0.070	0.280	0.087	0.377	0.092	0.292	0.085	0.386	0.095	0.398
105	0.361	0.090	0.293	0.056	0.326	0.069	0.301	0.085	0.395	0.091	0.311	0.084	0.403	0.093	0.414
115	0.349	0.087	0.315	0.054	0.348	0.068	0.320	0.083	0.415	0.089	0.332	0.082	0.420	0.089	0.438
125	0.342	0.081	0.333	0.051	0.370	0.063	0.339	0.080	0.436	0.084	0.349	0.078	0.440	0.087	0.456
135	0.333	0.077	0.350	0.050	0.390	0.060	0.357	0.074	0.461	0.079	0.366	0.073	0.463	0.081	0.479
145	0.327	0.078	0.371	0.051	0.408	0.060	0.375	0.074	0.479	0.079	0.388	0.074	0.481	0.080	0.497

Table 4.11: *Pair3 Empirical Power when* $\lambda = 0.8$

$Pair4$ has a distribution combination with big skewness and big common variance. Figure 4.8 gives the theoretical results and Figure 4.9 gives the corresponding empirical results. The numerical results of empirical power were shown in Tables 4.12, 4.13 and 4.14. We found that simulation results of $Pair4$ is very similar to $Pair3$, so we omit the discussion of simulation results form $Pair4$.

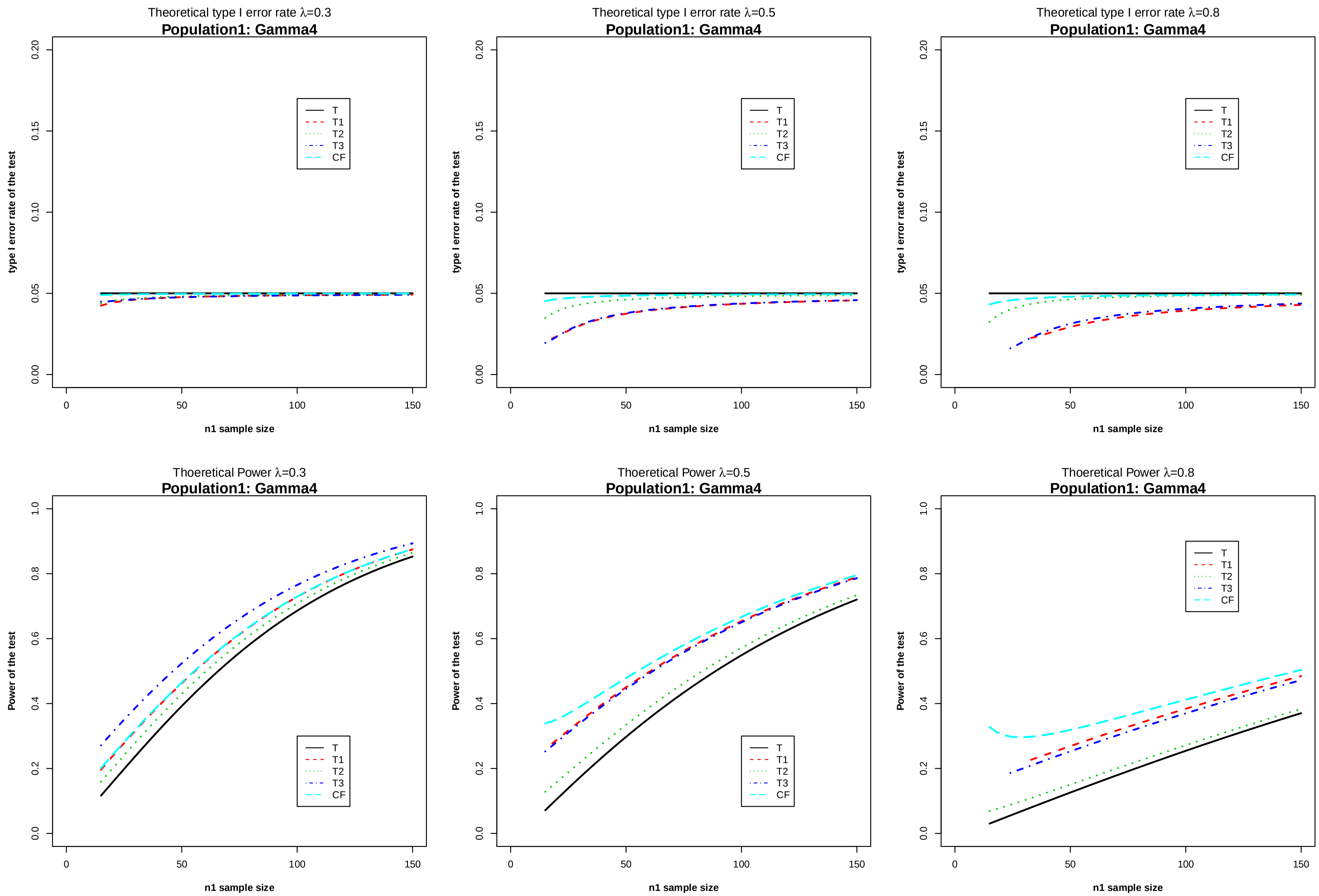

Figure 4.8: *Theoretical Power and Type I Error rate for Pair4*

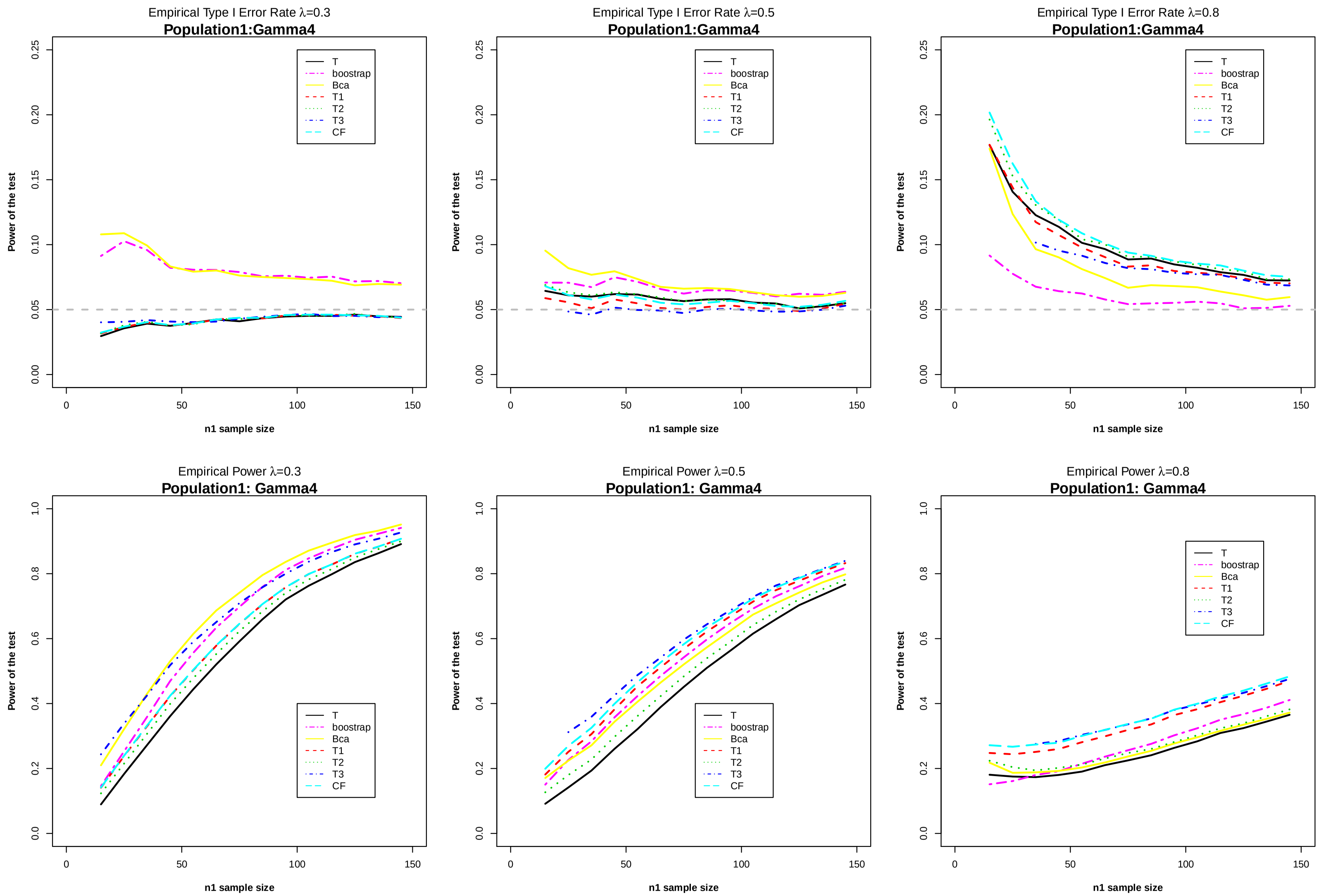

Figure 4.9: *Empirical Power and Type I Error rate for Pair4*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.092	0.030	0.089	0.091	0.146	0.108	0.210	0.032	0.141	0.032	0.123	0.040	0.244	0.032	0.142
25	0.085	0.036	0.183	0.103	0.253	0.109	0.321	0.037	0.238	0.038	0.215	0.041	0.338	0.037	0.239
35	0.076	0.039	0.272	0.096	0.359	0.099	0.431	0.040	0.331	0.041	0.307	0.042	0.425	0.040	0.333
45	0.073	0.037	0.361	0.082	0.470	0.083	0.531	0.038	0.423	0.038	0.399	0.041	0.518	0.038	0.424
55	0.070	0.040	0.444	0.081	0.556	0.079	0.615	0.039	0.503	0.039	0.478	0.040	0.591	0.039	0.504
65	0.065	0.042	0.521	0.080	0.634	0.080	0.687	0.042	0.579	0.042	0.553	0.041	0.651	0.042	0.580
75	0.064	0.041	0.592	0.079	0.698	0.076	0.741	0.044	0.645	0.043	0.623	0.043	0.709	0.044	0.646
85	0.062	0.043	0.660	0.076	0.759	0.075	0.796	0.043	0.706	0.044	0.684	0.044	0.758	0.044	0.707
95	0.058	0.045	0.720	0.076	0.811	0.074	0.836	0.045	0.758	0.046	0.741	0.046	0.800	0.046	0.758
105	0.057	0.045	0.763	0.074	0.847	0.073	0.871	0.046	0.799	0.045	0.782	0.047	0.837	0.046	0.799
115	0.056	0.045	0.798	0.075	0.877	0.072	0.895	0.046	0.828	0.045	0.814	0.045	0.867	0.046	0.829
125	0.054	0.046	0.836	0.072	0.904	0.069	0.919	0.045	0.861	0.047	0.850	0.045	0.890	0.045	0.862
135	0.053	0.045	0.863	0.072	0.923	0.070	0.932	0.045	0.883	0.045	0.875	0.044	0.907	0.045	0.883
145	0.052	0.044	0.891	0.070	0.941	0.069	0.951	0.043	0.907	0.044	0.900	0.044	0.927	0.044	0.907

Table 4.12: *Pair4 Empirical Power when* $\lambda = 0.3$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.383	0.064	0.091	0.071	0.150	0.095	0.171	0.059	0.181	0.069	0.127	0.596	0.896	0.068	0.199
25	0.350	0.061	0.142	0.071	0.226	0.082	0.224	0.056	0.251	0.063	0.180	0.048	0.312	0.061	0.270
35	0.327	0.060	0.194	0.067	0.284	0.077	0.271	0.051	0.306	0.061	0.228	0.046	0.359	0.058	0.325
45	0.307	0.062	0.261	0.075	0.357	0.080	0.344	0.058	0.382	0.063	0.298	0.051	0.425	0.062	0.399
55	0.291	0.062	0.322	0.071	0.423	0.073	0.406	0.055	0.453	0.062	0.362	0.050	0.488	0.059	0.466
65	0.279	0.058	0.389	0.066	0.485	0.068	0.465	0.051	0.512	0.059	0.423	0.049	0.542	0.055	0.527
75	0.268	0.056	0.451	0.062	0.542	0.066	0.520	0.050	0.568	0.056	0.484	0.047	0.597	0.054	0.583
85	0.257	0.058	0.510	0.065	0.597	0.067	0.573	0.052	0.622	0.057	0.539	0.050	0.644	0.055	0.634
95	0.249	0.058	0.562	0.065	0.647	0.066	0.624	0.053	0.668	0.057	0.589	0.051	0.686	0.057	0.678
105	0.241	0.055	0.615	0.063	0.692	0.063	0.674	0.051	0.714	0.055	0.642	0.049	0.728	0.055	0.723
115	0.235	0.055	0.660	0.060	0.731	0.061	0.709	0.051	0.750	0.055	0.683	0.048	0.764	0.053	0.758
125	0.228	0.051	0.703	0.062	0.761	0.060	0.742	0.049	0.778	0.051	0.721	0.049	0.788	0.052	0.786
135	0.223	0.053	0.735	0.061	0.792	0.060	0.773	0.051	0.806	0.053	0.752	0.050	0.815	0.054	0.813
145	0.217	0.055	0.766	0.064	0.818	0.063	0.798	0.054	0.833	0.055	0.781	0.053	0.840	0.057	0.837

Table 4.13: *Pair4 Empirical Power when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.524	0.177	0.181	0.092	0.151	0.174	0.218	0.177	0.248	0.196	0.224	1.000	1.000	0.202	0.272
25	0.489	0.141	0.175	0.078	0.161	0.124	0.187	0.144	0.244	0.153	0.204	0.172	0.386	0.162	0.267
35	0.449	0.123	0.173	0.068	0.180	0.096	0.188	0.117	0.251	0.130	0.194	0.102	0.276	0.133	0.274
45	0.429	0.114	0.180	0.064	0.192	0.090	0.193	0.107	0.260	0.119	0.202	0.095	0.284	0.119	0.279
55	0.404	0.101	0.190	0.062	0.215	0.081	0.203	0.098	0.281	0.104	0.212	0.091	0.304	0.109	0.300
65	0.389	0.097	0.210	0.058	0.236	0.074	0.218	0.090	0.299	0.100	0.231	0.086	0.319	0.101	0.319
75	0.372	0.089	0.225	0.054	0.256	0.067	0.237	0.083	0.319	0.091	0.248	0.082	0.336	0.094	0.337
85	0.359	0.089	0.241	0.055	0.276	0.069	0.254	0.084	0.336	0.091	0.260	0.081	0.354	0.091	0.353
95	0.346	0.085	0.264	0.055	0.302	0.068	0.277	0.080	0.364	0.087	0.282	0.078	0.380	0.088	0.381
105	0.336	0.082	0.284	0.056	0.324	0.067	0.296	0.078	0.383	0.084	0.302	0.077	0.397	0.085	0.399
115	0.324	0.079	0.309	0.055	0.350	0.064	0.316	0.077	0.404	0.081	0.324	0.077	0.416	0.084	0.422
125	0.317	0.077	0.324	0.051	0.367	0.061	0.335	0.074	0.424	0.079	0.339	0.073	0.433	0.080	0.440
135	0.308	0.073	0.344	0.051	0.387	0.058	0.353	0.071	0.445	0.073	0.361	0.069	0.454	0.076	0.462
145	0.302	0.072	0.365	0.053	0.411	0.060	0.372	0.070	0.470	0.073	0.382	0.069	0.476	0.075	0.484

Table 4.14: *Pair4 Empirical Power when* $\lambda = 0.8$

The population 1 of $Pair5$ follows Log-normal distribution and the $Pair5$ has a distribution combination with small skewness and small common variance, which is similar to the setting of $Pair1$. We found that the theoretical results showed in Figure 4.10 for $Pair5$ have the same pattern as the results in Figure 4.2 for $Pair1$. The empirical results of $Pair5$ in Figure 4.11 also resemble that for $Pair1$. The numerical results of empirical power are given in Tables 4.15, 4.16 and 4.17.

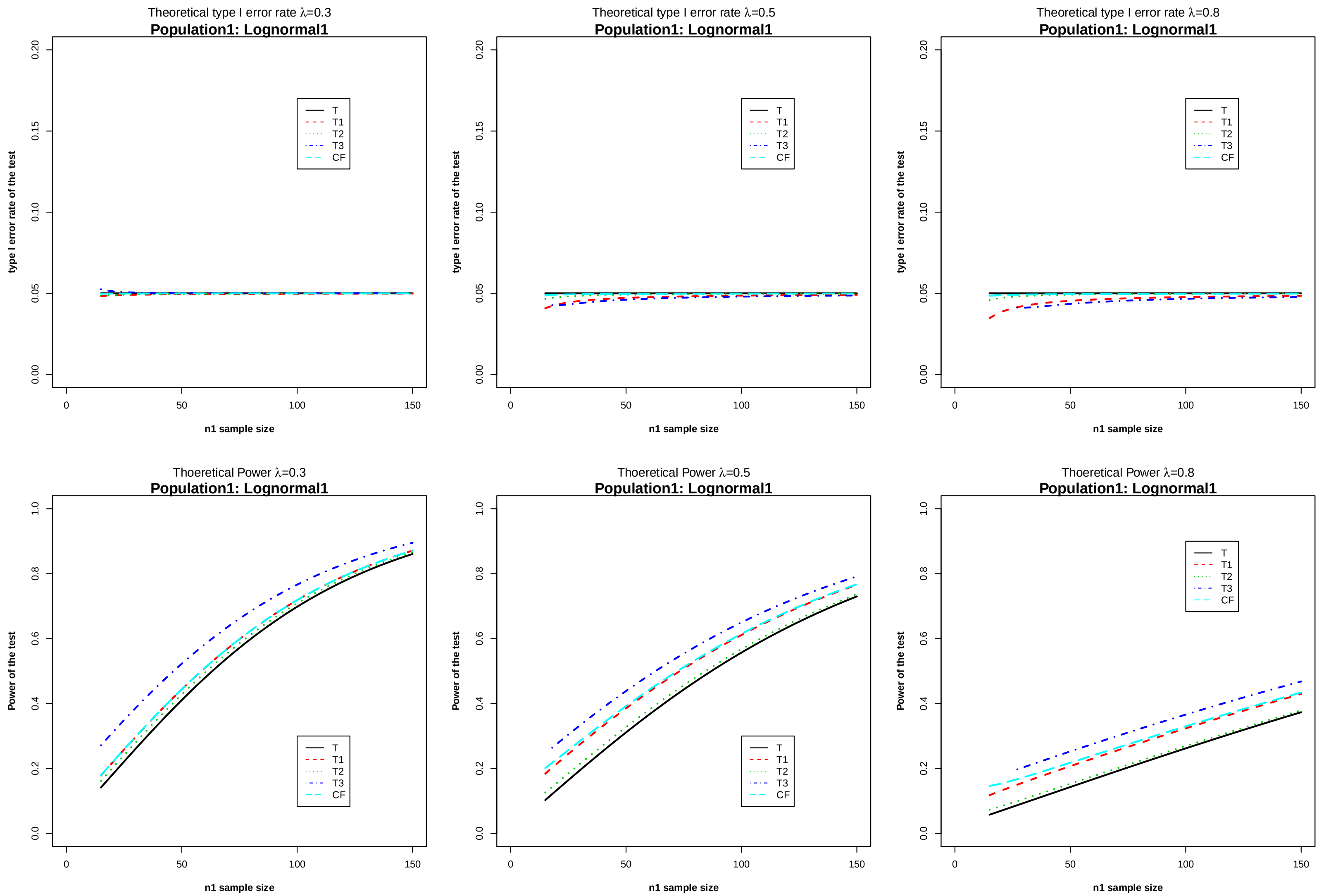

Figure 4.10: *Theoretical Power and Type I Error rate for Pair5*

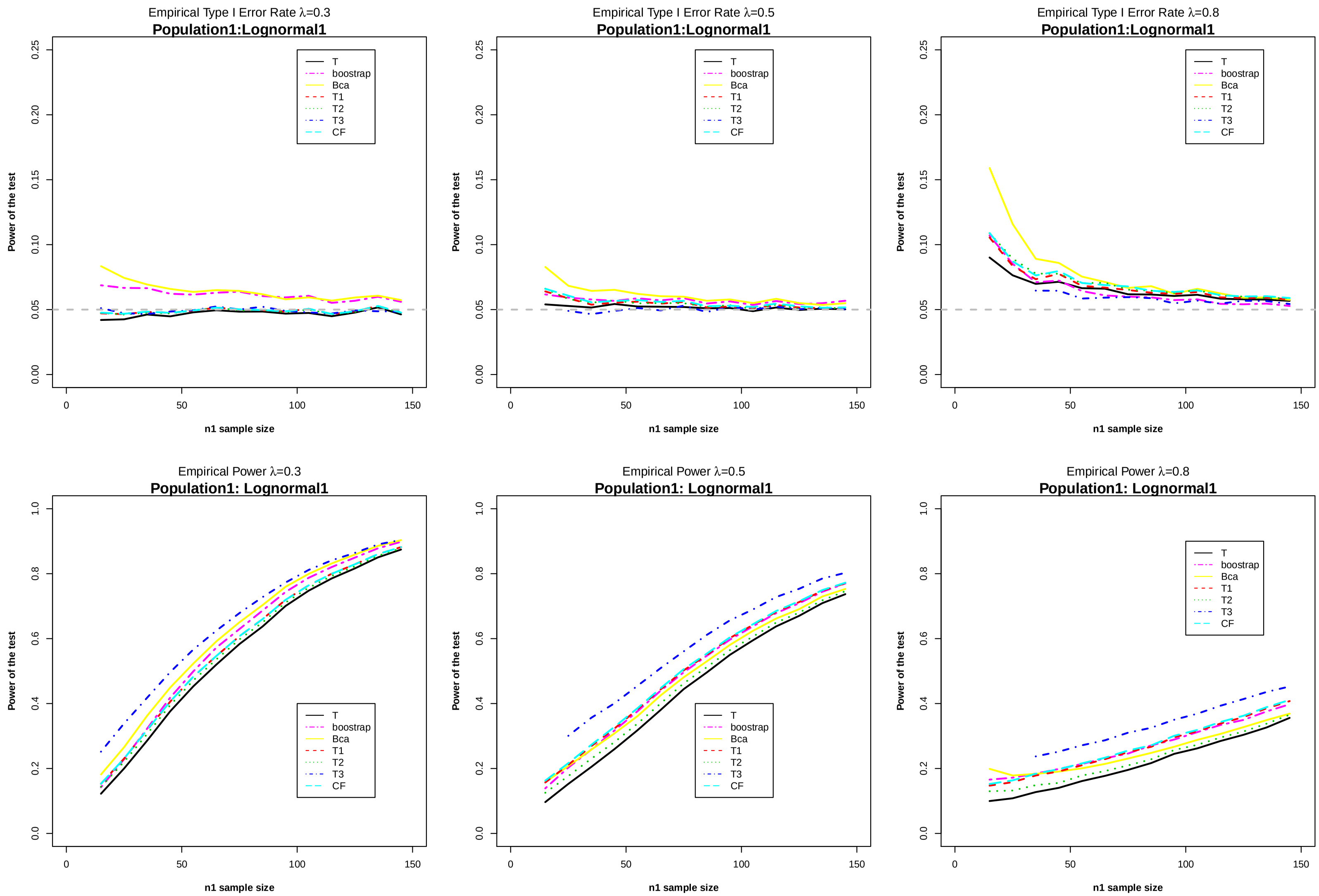

Figure 4.11: *Empirical Power and Type I Error rate for Pair5*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.045	0.042	0.122	0.069	0.142	0.083	0.182	0.047	0.153	0.047	0.144	0.051	0.252	0.048	0.153
25	0.042	0.042	0.200	0.067	0.229	0.074	0.266	0.046	0.230	0.046	0.220	0.047	0.339	0.047	0.231
35	0.037	0.046	0.286	0.066	0.323	0.069	0.363	0.048	0.318	0.048	0.306	0.047	0.418	0.048	0.319
45	0.036	0.045	0.376	0.062	0.418	0.066	0.450	0.048	0.406	0.047	0.395	0.048	0.497	0.048	0.407
55	0.034	0.048	0.453	0.061	0.499	0.064	0.523	0.049	0.483	0.049	0.471	0.050	0.566	0.049	0.483
65	0.032	0.050	0.520	0.063	0.572	0.065	0.592	0.051	0.547	0.051	0.536	0.052	0.625	0.051	0.547
75	0.031	0.048	0.583	0.064	0.629	0.064	0.650	0.050	0.607	0.049	0.598	0.050	0.679	0.050	0.608
85	0.030	0.048	0.637	0.060	0.686	0.062	0.703	0.050	0.660	0.050	0.650	0.052	0.728	0.050	0.660
95	0.028	0.047	0.701	0.059	0.744	0.058	0.760	0.048	0.720	0.048	0.711	0.049	0.773	0.049	0.720
105	0.028	0.047	0.748	0.061	0.787	0.059	0.799	0.051	0.764	0.050	0.757	0.048	0.812	0.051	0.764
115	0.027	0.045	0.785	0.055	0.821	0.057	0.832	0.046	0.800	0.046	0.793	0.047	0.841	0.047	0.800
125	0.026	0.048	0.816	0.057	0.850	0.059	0.859	0.049	0.829	0.048	0.823	0.049	0.864	0.049	0.829
135	0.025	0.052	0.850	0.060	0.878	0.061	0.886	0.053	0.860	0.052	0.855	0.049	0.890	0.053	0.860
145	0.025	0.046	0.874	0.056	0.898	0.057	0.903	0.048	0.882	0.048	0.878	0.048	0.904	0.048	0.882

Table 4.15: *Pair5 Empirical Power when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.186	0.054	0.096	0.062	0.139	0.083	0.158	0.064	0.157	0.066	0.125	0.984	0.997	0.066	0.163
25	0.173	0.053	0.153	0.059	0.203	0.068	0.208	0.059	0.211	0.058	0.177	0.049	0.301	0.061	0.217
35	0.161	0.052	0.205	0.058	0.259	0.064	0.258	0.054	0.269	0.055	0.231	0.046	0.356	0.056	0.273
45	0.151	0.054	0.260	0.057	0.317	0.065	0.308	0.055	0.323	0.057	0.281	0.049	0.401	0.057	0.328
55	0.142	0.052	0.318	0.059	0.377	0.062	0.362	0.056	0.383	0.055	0.339	0.051	0.455	0.057	0.386
65	0.135	0.052	0.380	0.057	0.440	0.060	0.425	0.055	0.444	0.055	0.401	0.049	0.509	0.056	0.448
75	0.130	0.052	0.445	0.059	0.496	0.060	0.481	0.056	0.502	0.055	0.463	0.053	0.561	0.057	0.506
85	0.125	0.051	0.496	0.055	0.547	0.057	0.530	0.052	0.550	0.052	0.512	0.048	0.612	0.053	0.554
95	0.120	0.051	0.550	0.056	0.597	0.058	0.581	0.052	0.602	0.053	0.564	0.052	0.656	0.053	0.604
105	0.116	0.049	0.595	0.054	0.638	0.055	0.624	0.051	0.642	0.050	0.607	0.050	0.690	0.052	0.644
115	0.113	0.052	0.637	0.057	0.678	0.058	0.662	0.053	0.682	0.054	0.649	0.052	0.728	0.055	0.685
125	0.109	0.050	0.670	0.054	0.709	0.055	0.691	0.052	0.713	0.051	0.680	0.050	0.753	0.052	0.715
135	0.106	0.051	0.709	0.055	0.744	0.054	0.730	0.051	0.748	0.052	0.718	0.051	0.785	0.051	0.749
145	0.104	0.050	0.737	0.057	0.770	0.055	0.753	0.051	0.771	0.051	0.746	0.050	0.803	0.052	0.772

Table 4.16: *Pair5 Empirical Power when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.254	0.090	0.100	0.107	0.166	0.159	0.199	0.105	0.147	0.109	0.130	1.000	1.000	0.109	0.152
25	0.240	0.076	0.108	0.085	0.172	0.116	0.178	0.084	0.158	0.089	0.132	0.548	0.802	0.087	0.164
35	0.219	0.070	0.127	0.071	0.184	0.089	0.183	0.073	0.178	0.078	0.149	0.065	0.237	0.076	0.184
45	0.208	0.071	0.140	0.072	0.199	0.086	0.191	0.077	0.191	0.078	0.156	0.064	0.252	0.080	0.198
55	0.197	0.066	0.161	0.064	0.214	0.075	0.201	0.068	0.210	0.070	0.178	0.058	0.271	0.070	0.216
65	0.188	0.066	0.177	0.061	0.231	0.071	0.214	0.067	0.229	0.071	0.192	0.059	0.287	0.069	0.233
75	0.180	0.062	0.196	0.060	0.246	0.067	0.231	0.065	0.250	0.065	0.210	0.060	0.310	0.068	0.255
85	0.173	0.062	0.217	0.059	0.271	0.068	0.248	0.063	0.267	0.065	0.228	0.059	0.325	0.064	0.271
95	0.166	0.060	0.245	0.057	0.289	0.062	0.267	0.062	0.296	0.063	0.257	0.055	0.350	0.064	0.301
105	0.162	0.061	0.262	0.058	0.312	0.066	0.288	0.064	0.315	0.065	0.273	0.057	0.369	0.065	0.319
115	0.156	0.058	0.285	0.054	0.334	0.062	0.306	0.059	0.338	0.061	0.295	0.055	0.393	0.061	0.343
125	0.152	0.058	0.304	0.054	0.350	0.059	0.328	0.059	0.358	0.060	0.314	0.057	0.414	0.060	0.362
135	0.148	0.058	0.326	0.055	0.376	0.059	0.349	0.059	0.385	0.059	0.338	0.057	0.436	0.060	0.389
145	0.144	0.057	0.356	0.053	0.398	0.058	0.369	0.058	0.408	0.059	0.365	0.054	0.453	0.059	0.411

Table 4.17: *Pair5 Simulation Power lamda=0.8*

The simulation setting of $Pair6$ is similar to $Pair4$ with big skewness and big common variance. We found the simulation results of $Pair6$ in Figure 4.12 and Figure 4.13 are similar to the simulation results of $Pair4$. The numerical results of empirical power are in Tables 4.18, 4.19 and 4.20.

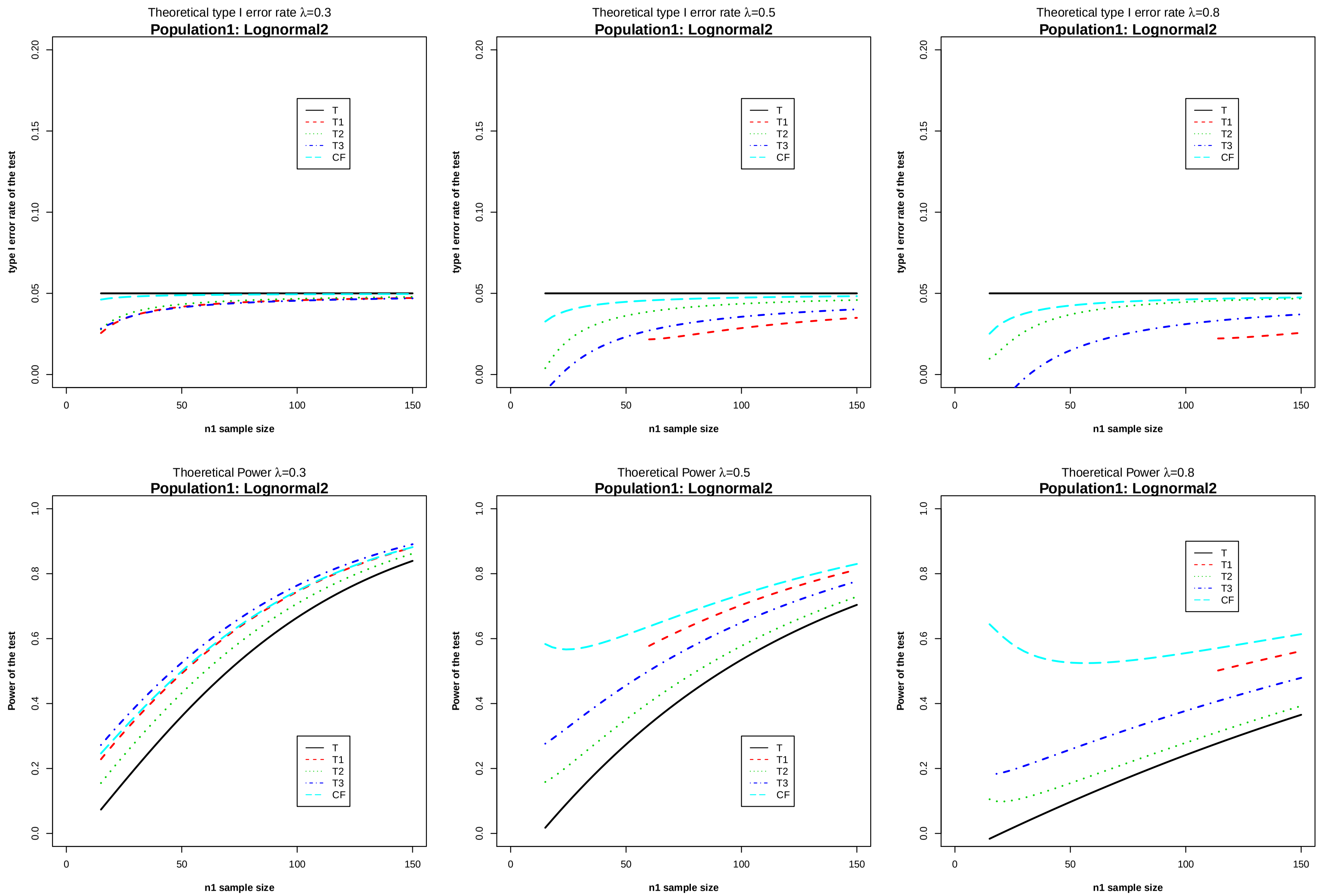

Figure 4.12: *Theoretical Power and Type I Error rate for Pair6*

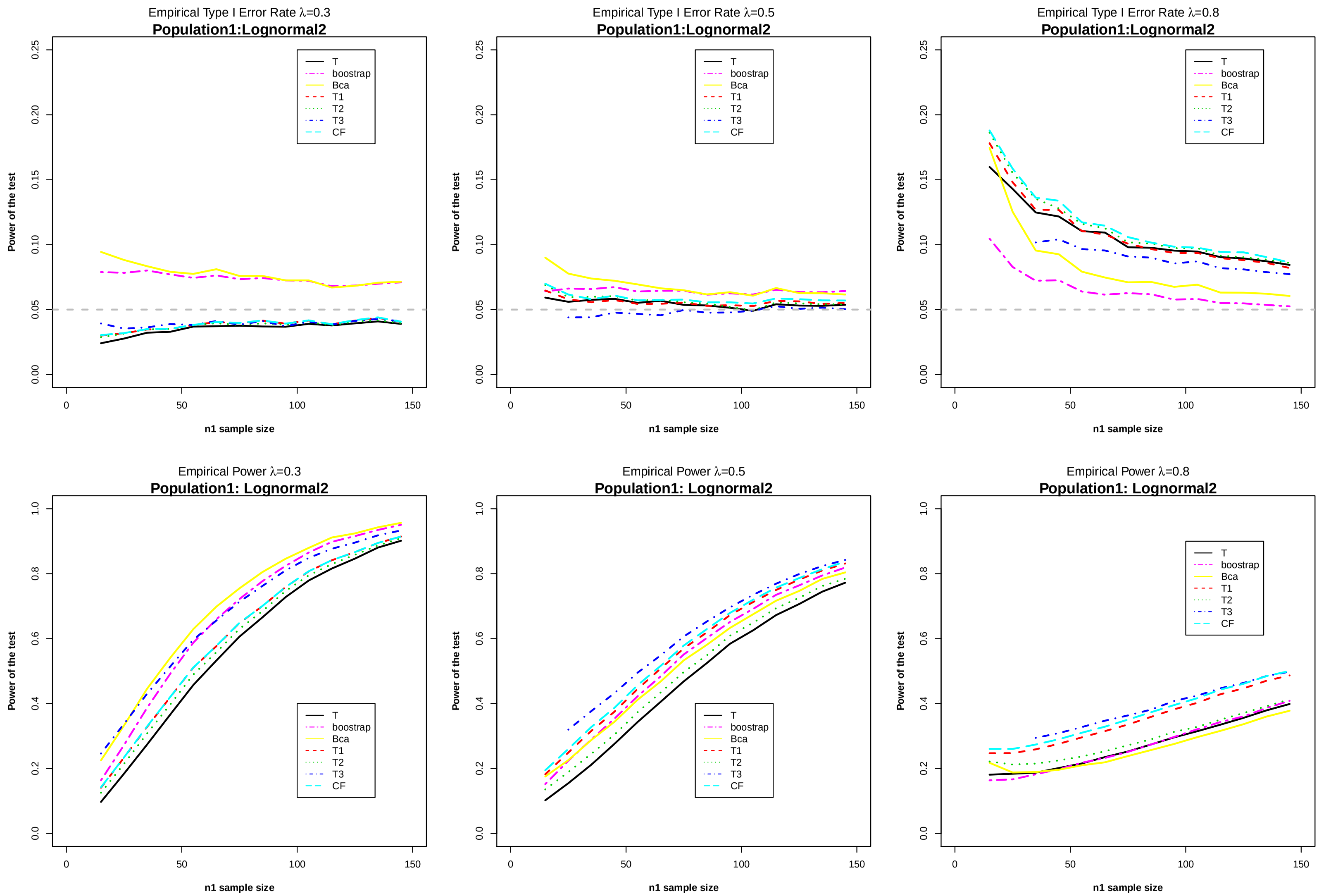

Figure 4.13: *Empirical Power and Type I Error rate for Pair6*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.067	0.024	0.097	0.079	0.164	0.094	0.225	0.030	0.141	0.029	0.125	0.039	0.246	0.030	0.142
25	0.064	0.028	0.184	0.078	0.272	0.088	0.333	0.032	0.234	0.032	0.214	0.035	0.338	0.032	0.234
35	0.059	0.032	0.274	0.080	0.387	0.083	0.446	0.035	0.328	0.035	0.308	0.036	0.430	0.035	0.330
45	0.058	0.033	0.366	0.077	0.491	0.079	0.541	0.035	0.420	0.035	0.397	0.039	0.514	0.035	0.420
55	0.057	0.037	0.457	0.074	0.588	0.077	0.629	0.038	0.511	0.038	0.489	0.038	0.596	0.038	0.511
65	0.054	0.037	0.533	0.076	0.660	0.081	0.699	0.040	0.577	0.040	0.558	0.041	0.656	0.040	0.578
75	0.053	0.038	0.606	0.073	0.722	0.076	0.755	0.040	0.648	0.038	0.630	0.038	0.714	0.040	0.648
85	0.052	0.037	0.666	0.074	0.777	0.076	0.805	0.042	0.700	0.039	0.685	0.041	0.762	0.042	0.701
95	0.050	0.037	0.728	0.072	0.824	0.072	0.846	0.039	0.759	0.038	0.747	0.037	0.810	0.039	0.759
105	0.049	0.039	0.779	0.072	0.865	0.072	0.879	0.042	0.807	0.040	0.795	0.041	0.849	0.042	0.807
115	0.049	0.038	0.816	0.068	0.898	0.067	0.911	0.039	0.841	0.038	0.829	0.038	0.877	0.039	0.842
125	0.047	0.039	0.846	0.068	0.916	0.068	0.924	0.042	0.866	0.041	0.858	0.041	0.896	0.042	0.866
135	0.047	0.041	0.880	0.070	0.935	0.071	0.943	0.044	0.894	0.042	0.887	0.043	0.918	0.044	0.895
145	0.046	0.039	0.901	0.071	0.951	0.071	0.957	0.041	0.914	0.040	0.909	0.041	0.933	0.041	0.914

Table 4.18: *Pair6 Empirical Power when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.277	0.059	0.102	0.064	0.152	0.090	0.175	0.065	0.182	0.069	0.136	0.864	0.974	0.070	0.195
25	0.266	0.056	0.155	0.066	0.224	0.078	0.226	0.058	0.249	0.060	0.189	0.044	0.320	0.062	0.260
35	0.254	0.058	0.212	0.066	0.290	0.074	0.288	0.056	0.316	0.060	0.246	0.044	0.377	0.058	0.328
45	0.244	0.058	0.276	0.067	0.354	0.072	0.344	0.057	0.375	0.060	0.304	0.048	0.432	0.061	0.387
55	0.235	0.055	0.343	0.064	0.422	0.069	0.412	0.054	0.445	0.057	0.373	0.047	0.495	0.057	0.457
65	0.229	0.057	0.405	0.064	0.484	0.066	0.468	0.054	0.508	0.058	0.434	0.046	0.548	0.057	0.516
75	0.223	0.054	0.468	0.064	0.551	0.065	0.533	0.055	0.570	0.056	0.497	0.050	0.607	0.058	0.579
85	0.218	0.053	0.525	0.061	0.602	0.062	0.581	0.053	0.619	0.055	0.549	0.048	0.653	0.056	0.629
95	0.213	0.051	0.584	0.062	0.651	0.063	0.633	0.053	0.672	0.052	0.609	0.048	0.696	0.056	0.679
105	0.209	0.049	0.625	0.062	0.691	0.061	0.675	0.053	0.712	0.050	0.648	0.049	0.735	0.055	0.720
115	0.205	0.054	0.672	0.065	0.735	0.067	0.717	0.057	0.750	0.056	0.694	0.053	0.769	0.059	0.757
125	0.202	0.053	0.707	0.064	0.764	0.063	0.747	0.056	0.782	0.055	0.726	0.051	0.799	0.058	0.786
135	0.198	0.053	0.745	0.063	0.794	0.063	0.784	0.054	0.810	0.054	0.762	0.051	0.824	0.057	0.814
145	0.195	0.054	0.772	0.064	0.819	0.062	0.804	0.055	0.831	0.054	0.784	0.050	0.843	0.057	0.837

Table 4.19: *Pair6 Empirical Power when* $\lambda = 0.5$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.378	0.160	0.181	0.105	0.164	0.174	0.217	0.178	0.247	0.186	0.222	1.000	1.000	0.188	0.260
25	0.370	0.143	0.184	0.083	0.167	0.125	0.188	0.148	0.247	0.155	0.212	0.292	0.565	0.158	0.260
35	0.348	0.125	0.187	0.072	0.182	0.096	0.189	0.127	0.259	0.135	0.215	0.102	0.294	0.136	0.274
45	0.338	0.122	0.201	0.073	0.198	0.093	0.196	0.127	0.276	0.128	0.225	0.104	0.309	0.134	0.290
55	0.328	0.110	0.216	0.064	0.218	0.079	0.210	0.110	0.296	0.116	0.237	0.097	0.328	0.117	0.310
65	0.319	0.109	0.235	0.061	0.233	0.075	0.219	0.108	0.315	0.113	0.253	0.096	0.347	0.115	0.329
75	0.311	0.098	0.252	0.063	0.251	0.071	0.238	0.101	0.336	0.102	0.272	0.091	0.363	0.106	0.351
85	0.304	0.098	0.274	0.062	0.274	0.071	0.257	0.097	0.359	0.101	0.290	0.090	0.381	0.102	0.373
95	0.296	0.095	0.296	0.058	0.297	0.068	0.276	0.094	0.382	0.097	0.313	0.086	0.408	0.098	0.395
105	0.291	0.095	0.315	0.058	0.321	0.069	0.297	0.094	0.402	0.097	0.328	0.087	0.424	0.098	0.416
115	0.285	0.090	0.335	0.055	0.343	0.063	0.316	0.090	0.429	0.092	0.350	0.082	0.447	0.094	0.443
125	0.281	0.089	0.356	0.055	0.360	0.063	0.336	0.088	0.447	0.090	0.371	0.081	0.463	0.094	0.461
135	0.275	0.087	0.379	0.054	0.386	0.062	0.360	0.086	0.470	0.088	0.392	0.079	0.485	0.090	0.485
145	0.272	0.084	0.399	0.052	0.408	0.060	0.378	0.082	0.487	0.086	0.413	0.077	0.498	0.086	0.500

Table 4.20: *Pair6 Empirical Power when* $\lambda = 0.8$

4.4.5 Simulation results for one-sided test

For the simulation study of the one-sided test, we only present the empirical results from the seven tests mentioned in Section 4.4.1. And the empirical type I error rate and power of these seven tests were obtained by calculating the proportion of rejections in each simulation setting under H_0 and H_a respectively. Figure 4.14 and Figure 4.15 give the empirical type I error rate and power for one-sided upper-tailed and lower-tailed tests respectively. The numerical results of Figure 4.14 are presented in Tables 4.21, 4.23 and 4.25; while the numerical results of Figure 4.15 are given in Tables 4.22, 4.24 and 4.26.

Under the simulation settings of $Pair1$, all the seven two-sample one-sided upper-tailed tests keep the empirical type I error rate under 0.1 shown in the upper panel of Figure 4.14. Among all the seven tests, the one-sided upper-tailed T_3 test gives higher empirical type I error rate than the other six tests across all levels of λ from 0.3 to 0.8. The empirical power of the seven tests are close to each other except the empirical power of T_3 test, which is slightly higher than the rest of the six tests.

For the one-sided lower-tailed tests in Figure 4.15, T_3 test gives the smallest empirical type I error rate than the other tests. Across all levels of λ, the type I error rate of all seven tests are below 0.1. The highest empirical power comes from the pooled two-sample t-test, which also keeps its empirical type I error rate close to 0.05. Although the type I error rate of T_3 is consistently smaller than 0.05, its empirical power is much lower than the other tests.

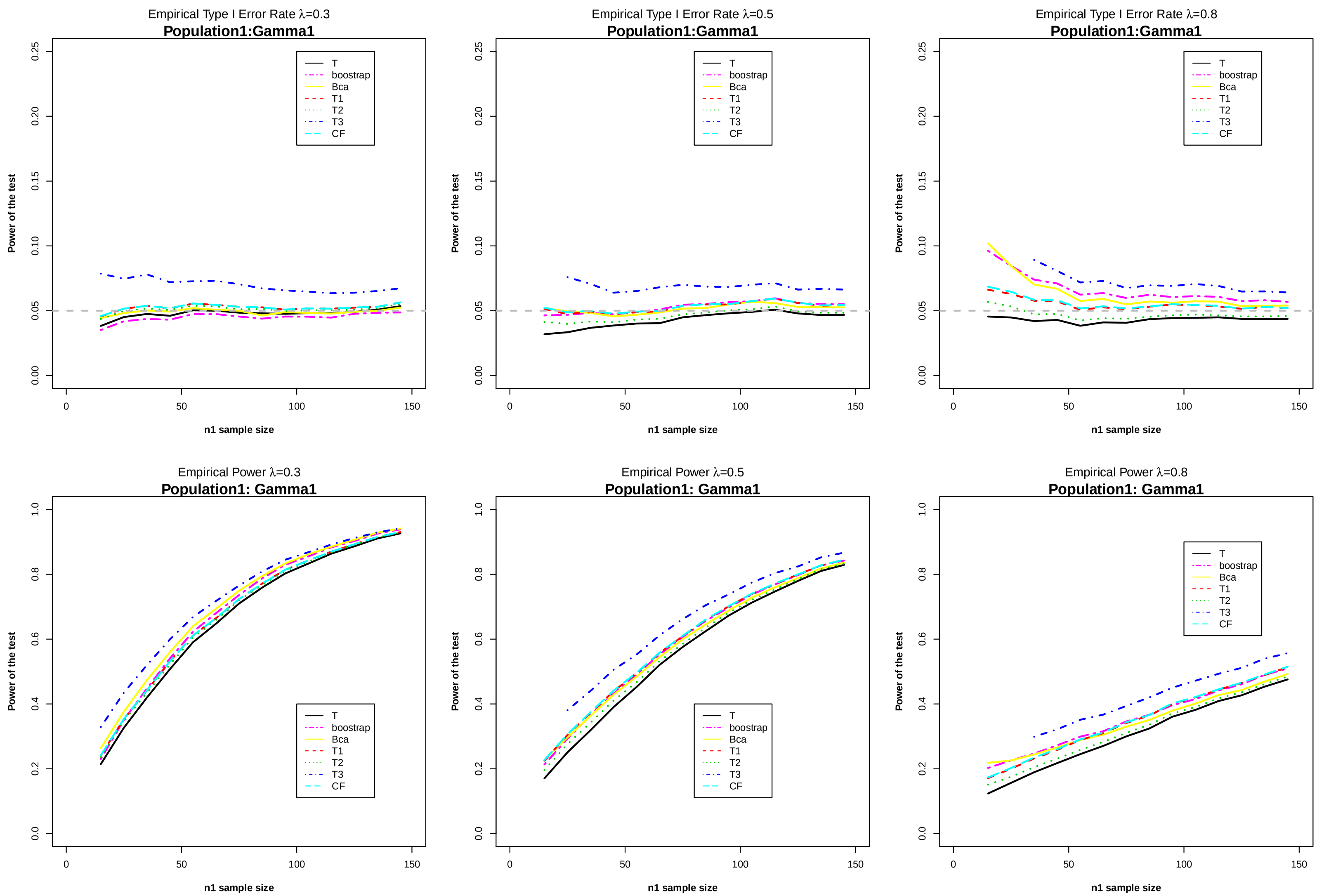

Figure 4.14: *One-sided upper-tailed Empirical Power and Type I Error rate for Pair1*

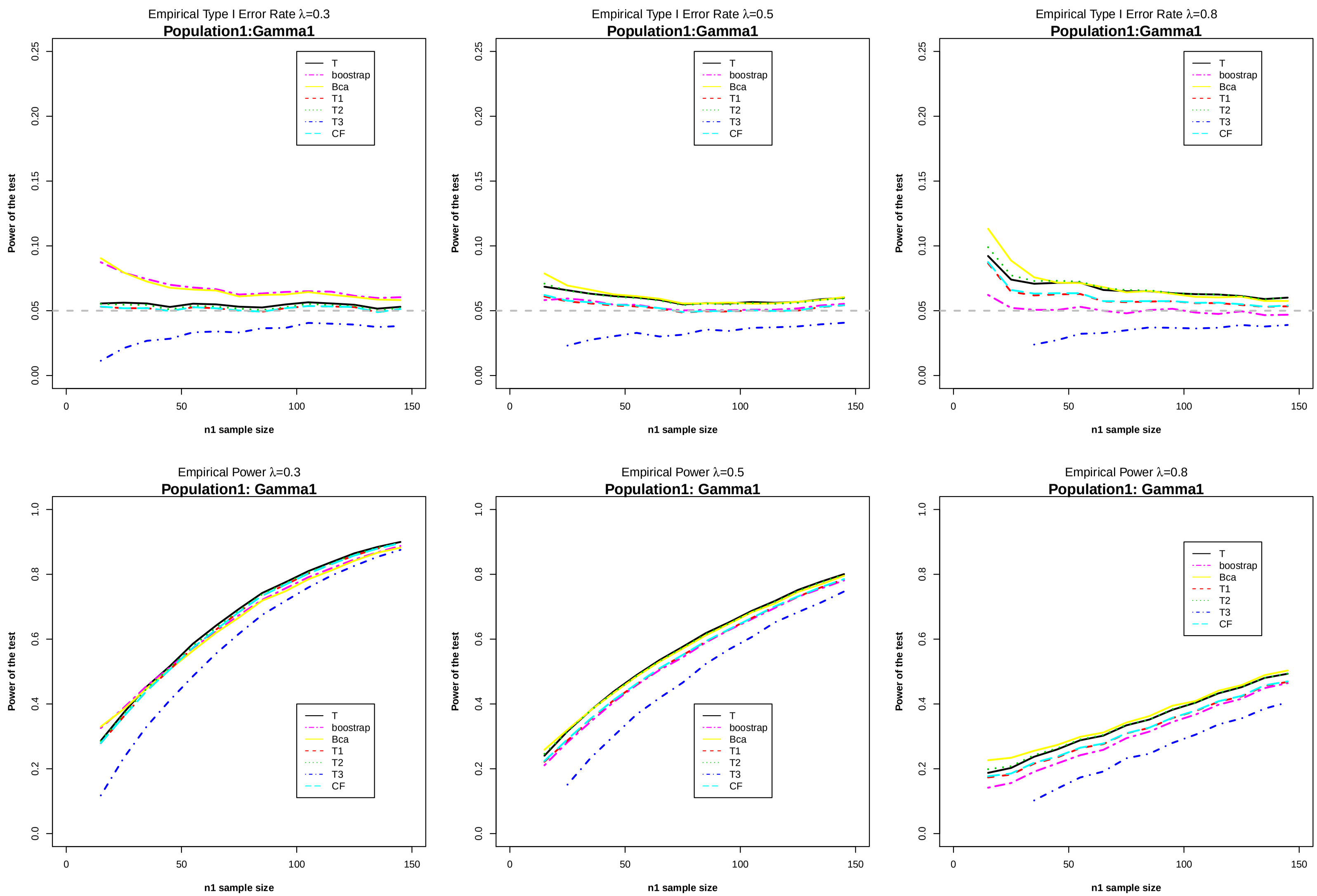

Figure 4.15: *One-sided lower-tailed Empirical Power and Type I Error rate for Pair1*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.044	0.038	0.214	0.035	0.230	0.044	0.264	0.046	0.241	0.044	0.236	0.079	0.329	0.046	0.241
25	0.039	0.045	0.327	0.042	0.351	0.048	0.376	0.052	0.353	0.050	0.346	0.075	0.435	0.052	0.354
35	0.034	0.048	0.420	0.044	0.448	0.050	0.474	0.054	0.442	0.052	0.435	0.078	0.520	0.054	0.442
45	0.031	0.046	0.507	0.043	0.541	0.049	0.560	0.052	0.532	0.050	0.522	0.072	0.599	0.052	0.532
55	0.029	0.050	0.590	0.047	0.623	0.052	0.640	0.056	0.611	0.054	0.605	0.073	0.668	0.056	0.611
65	0.027	0.050	0.648	0.047	0.680	0.051	0.694	0.054	0.665	0.053	0.659	0.073	0.718	0.054	0.665
75	0.026	0.048	0.710	0.045	0.736	0.050	0.750	0.053	0.724	0.051	0.719	0.070	0.765	0.053	0.724
85	0.025	0.048	0.759	0.044	0.787	0.046	0.795	0.053	0.771	0.051	0.767	0.067	0.809	0.053	0.771
95	0.023	0.048	0.803	0.045	0.829	0.049	0.833	0.051	0.814	0.050	0.810	0.066	0.845	0.051	0.814
105	0.022	0.048	0.833	0.045	0.855	0.048	0.862	0.052	0.843	0.050	0.839	0.065	0.868	0.052	0.843
115	0.022	0.048	0.864	0.045	0.883	0.048	0.885	0.052	0.870	0.050	0.868	0.064	0.892	0.052	0.870
125	0.021	0.049	0.886	0.048	0.903	0.049	0.907	0.052	0.893	0.051	0.891	0.064	0.912	0.052	0.893
135	0.020	0.051	0.911	0.048	0.926	0.050	0.928	0.053	0.915	0.052	0.914	0.065	0.929	0.053	0.915
145	0.020	0.053	0.926	0.049	0.939	0.052	0.942	0.056	0.930	0.055	0.929	0.067	0.942	0.056	0.930

Table 4.21: *Proportion of rejections for one-sided upper-tailed tests for Pair1 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.044	0.056	0.287	0.087	0.326	0.090	0.330	0.053	0.279	0.055	0.286	0.011	0.118	0.053	0.279
25	0.039	0.056	0.373	0.079	0.389	0.079	0.384	0.052	0.362	0.055	0.369	0.021	0.234	0.052	0.362
35	0.034	0.056	0.454	0.074	0.456	0.072	0.449	0.052	0.441	0.054	0.448	0.027	0.331	0.052	0.441
45	0.031	0.053	0.517	0.070	0.513	0.068	0.507	0.050	0.506	0.051	0.512	0.028	0.412	0.050	0.506
55	0.029	0.055	0.586	0.068	0.573	0.066	0.564	0.053	0.574	0.054	0.579	0.033	0.486	0.053	0.574
65	0.027	0.055	0.642	0.067	0.623	0.066	0.620	0.052	0.629	0.053	0.636	0.034	0.555	0.052	0.629
75	0.026	0.053	0.694	0.062	0.674	0.061	0.667	0.050	0.684	0.052	0.689	0.033	0.617	0.050	0.684
85	0.025	0.052	0.742	0.063	0.721	0.062	0.719	0.049	0.734	0.050	0.738	0.036	0.675	0.049	0.734
95	0.023	0.055	0.775	0.064	0.756	0.062	0.747	0.052	0.768	0.053	0.772	0.037	0.718	0.052	0.768
105	0.022	0.056	0.810	0.065	0.791	0.064	0.784	0.054	0.802	0.055	0.806	0.041	0.759	0.054	0.802
115	0.022	0.056	0.838	0.065	0.819	0.062	0.812	0.053	0.831	0.054	0.834	0.040	0.796	0.053	0.831
125	0.021	0.055	0.865	0.061	0.846	0.061	0.842	0.053	0.858	0.054	0.862	0.039	0.826	0.053	0.858
135	0.020	0.052	0.884	0.060	0.868	0.059	0.867	0.049	0.879	0.050	0.882	0.037	0.854	0.049	0.879
145	0.020	0.053	0.900	0.060	0.888	0.058	0.882	0.051	0.896	0.052	0.897	0.038	0.875	0.051	0.896

Table 4.22: *Proportion of rejections for one-sided lower-tailed tests for Pair1 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.179	0.032	0.171	0.046	0.213	0.052	0.227	0.051	0.225	0.041	0.196	0.088	0.316	0.052	0.227
25	0.158	0.033	0.251	0.047	0.293	0.050	0.297	0.047	0.304	0.040	0.276	0.076	0.381	0.049	0.306
35	0.144	0.037	0.319	0.048	0.366	0.048	0.363	0.049	0.373	0.042	0.342	0.071	0.440	0.050	0.375
45	0.132	0.039	0.390	0.047	0.432	0.046	0.428	0.047	0.439	0.041	0.410	0.064	0.506	0.048	0.441
55	0.123	0.040	0.453	0.049	0.490	0.047	0.484	0.049	0.495	0.043	0.467	0.065	0.552	0.049	0.497
65	0.115	0.040	0.521	0.051	0.553	0.049	0.545	0.050	0.558	0.044	0.532	0.068	0.612	0.050	0.558
75	0.108	0.045	0.576	0.055	0.607	0.052	0.600	0.054	0.609	0.047	0.585	0.070	0.662	0.054	0.611
85	0.103	0.047	0.625	0.055	0.657	0.052	0.646	0.055	0.660	0.049	0.636	0.069	0.705	0.055	0.661
95	0.098	0.048	0.673	0.057	0.698	0.055	0.689	0.055	0.702	0.050	0.681	0.068	0.738	0.055	0.702
105	0.094	0.049	0.713	0.057	0.738	0.057	0.728	0.057	0.739	0.051	0.721	0.070	0.775	0.057	0.740
115	0.091	0.051	0.747	0.060	0.769	0.056	0.760	0.059	0.770	0.053	0.753	0.071	0.804	0.059	0.771
125	0.087	0.048	0.780	0.056	0.798	0.053	0.789	0.056	0.799	0.049	0.785	0.066	0.824	0.056	0.800
135	0.084	0.047	0.810	0.055	0.827	0.053	0.818	0.054	0.828	0.048	0.815	0.067	0.852	0.054	0.828
145	0.082	0.047	0.829	0.055	0.843	0.053	0.836	0.054	0.845	0.048	0.833	0.066	0.867	0.054	0.845

Table 4.23: *Proportion of rejections for one-sided upper-tailed tests for Pair1 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.179	0.068	0.240	0.058	0.211	0.079	0.259	0.061	0.222	0.071	0.246	0.010	0.056	0.062	0.224
25	0.158	0.066	0.316	0.059	0.282	0.069	0.320	0.057	0.287	0.065	0.315	0.023	0.151	0.058	0.291
35	0.144	0.063	0.380	0.058	0.345	0.066	0.380	0.056	0.352	0.063	0.379	0.028	0.233	0.056	0.354
45	0.132	0.061	0.438	0.055	0.406	0.063	0.434	0.054	0.411	0.061	0.436	0.030	0.301	0.055	0.412
55	0.123	0.060	0.489	0.055	0.458	0.061	0.485	0.053	0.461	0.060	0.487	0.033	0.368	0.054	0.462
65	0.115	0.058	0.536	0.052	0.505	0.059	0.531	0.052	0.509	0.058	0.533	0.030	0.419	0.052	0.510
75	0.108	0.055	0.576	0.050	0.544	0.056	0.571	0.049	0.551	0.054	0.574	0.032	0.466	0.049	0.551
85	0.103	0.056	0.619	0.051	0.590	0.056	0.612	0.050	0.593	0.055	0.616	0.035	0.523	0.050	0.594
95	0.098	0.056	0.652	0.050	0.628	0.056	0.647	0.049	0.629	0.055	0.649	0.034	0.568	0.050	0.630
105	0.094	0.057	0.687	0.051	0.662	0.055	0.683	0.050	0.665	0.056	0.685	0.037	0.607	0.050	0.665
115	0.091	0.056	0.718	0.051	0.696	0.056	0.710	0.050	0.701	0.055	0.716	0.037	0.652	0.050	0.701
125	0.087	0.057	0.752	0.052	0.731	0.057	0.746	0.050	0.731	0.056	0.749	0.038	0.683	0.050	0.732
135	0.084	0.059	0.777	0.054	0.756	0.058	0.769	0.053	0.760	0.058	0.775	0.040	0.713	0.053	0.760
145	0.082	0.060	0.801	0.055	0.781	0.060	0.795	0.054	0.785	0.059	0.799	0.041	0.747	0.054	0.785

Table 4.24: *Proportion of rejections for one-sided lower-tailed tests for Pair1 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.248	0.045	0.124	0.096	0.203	0.102	0.218	0.066	0.171	0.057	0.151	0.114	0.250	0.068	0.173
25	0.221	0.045	0.157	0.085	0.226	0.085	0.226	0.063	0.200	0.053	0.175	0.097	0.270	0.065	0.202
35	0.197	0.042	0.189	0.074	0.247	0.070	0.244	0.058	0.232	0.047	0.205	0.089	0.299	0.058	0.235
45	0.183	0.043	0.217	0.071	0.273	0.067	0.266	0.057	0.258	0.048	0.231	0.081	0.322	0.058	0.260
55	0.169	0.038	0.245	0.062	0.299	0.058	0.289	0.051	0.289	0.042	0.258	0.072	0.351	0.052	0.291
65	0.160	0.041	0.271	0.064	0.316	0.059	0.305	0.053	0.310	0.044	0.283	0.073	0.367	0.053	0.312
75	0.150	0.041	0.300	0.060	0.346	0.055	0.330	0.051	0.341	0.044	0.311	0.068	0.394	0.052	0.344
85	0.143	0.043	0.324	0.062	0.367	0.057	0.351	0.053	0.364	0.045	0.336	0.070	0.420	0.053	0.366
95	0.136	0.044	0.361	0.060	0.396	0.056	0.379	0.055	0.400	0.046	0.370	0.069	0.450	0.055	0.402
105	0.131	0.044	0.382	0.061	0.415	0.057	0.401	0.054	0.420	0.047	0.391	0.071	0.472	0.054	0.421
115	0.126	0.045	0.409	0.061	0.441	0.057	0.427	0.053	0.444	0.046	0.417	0.069	0.493	0.054	0.445
125	0.122	0.044	0.427	0.057	0.461	0.053	0.443	0.051	0.465	0.046	0.436	0.065	0.511	0.052	0.466
135	0.117	0.044	0.453	0.058	0.490	0.054	0.469	0.053	0.489	0.046	0.461	0.065	0.540	0.053	0.491
145	0.114	0.044	0.476	0.057	0.509	0.054	0.492	0.052	0.513	0.046	0.485	0.064	0.557	0.052	0.515

Table 4.25: *Proportion of rejections for one-sided upper-tailed tests for Pair1 when* $\lambda = 0.8$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.248	0.092	0.187	0.062	0.142	0.113	0.226	0.087	0.173	0.099	0.199			0.087	0.177
25	0.221	0.074	0.203	0.052	0.156	0.089	0.234	0.065	0.182	0.077	0.208	0.015	0.056	0.066	0.185
35	0.197	0.071	0.237	0.051	0.191	0.076	0.256	0.062	0.216	0.073	0.242	0.024	0.102	0.063	0.219
45	0.183	0.071	0.260	0.051	0.216	0.072	0.273	0.063	0.235	0.073	0.263	0.027	0.139	0.064	0.238
55	0.169	0.072	0.288	0.053	0.242	0.071	0.298	0.063	0.264	0.072	0.291	0.032	0.173	0.064	0.265
65	0.160	0.066	0.302	0.050	0.258	0.068	0.312	0.057	0.276	0.067	0.304	0.033	0.192	0.057	0.278
75	0.150	0.065	0.334	0.048	0.294	0.064	0.343	0.057	0.309	0.066	0.336	0.035	0.232	0.057	0.310
85	0.143	0.065	0.352	0.050	0.315	0.065	0.362	0.057	0.326	0.066	0.352	0.037	0.246	0.057	0.327
95	0.136	0.064	0.382	0.051	0.345	0.063	0.395	0.057	0.356	0.064	0.384	0.037	0.280	0.057	0.357
105	0.131	0.063	0.404	0.049	0.367	0.061	0.409	0.056	0.378	0.063	0.404	0.036	0.305	0.056	0.379
115	0.126	0.062	0.433	0.048	0.398	0.060	0.441	0.056	0.407	0.062	0.434	0.037	0.337	0.056	0.408
125	0.122	0.061	0.452	0.050	0.416	0.061	0.458	0.054	0.423	0.061	0.453	0.039	0.355	0.055	0.425
135	0.117	0.059	0.480	0.047	0.449	0.058	0.489	0.053	0.456	0.059	0.480	0.038	0.386	0.053	0.458
145	0.114	0.060	0.493	0.047	0.464	0.058	0.503	0.053	0.468	0.060	0.494	0.039	0.405	0.054	0.469

Table 4.26: *Proportion of rejections for one-sided lower-tailed tests for Pair1 when* $\lambda = 0.8$

As mentioned previously, the $Pair2$ has a distribution combination with small skewness and big common variance. The empirical results of the one-sided upper-tailed and lower-tailed tests are given in Figure 4.16 and Figure 4.17 respectively. Their corresponding numerical results are given in Table 4.27, 4.29, 4.31 and Table 4.28, 4.30, 4.32 respectively. The empirical results of seven one-sided tests under the simulation settings of $Pair2$ are very similar to those from $Pair1$. All the discussions from $Pair1$ still hold for the results of $Pair2$.

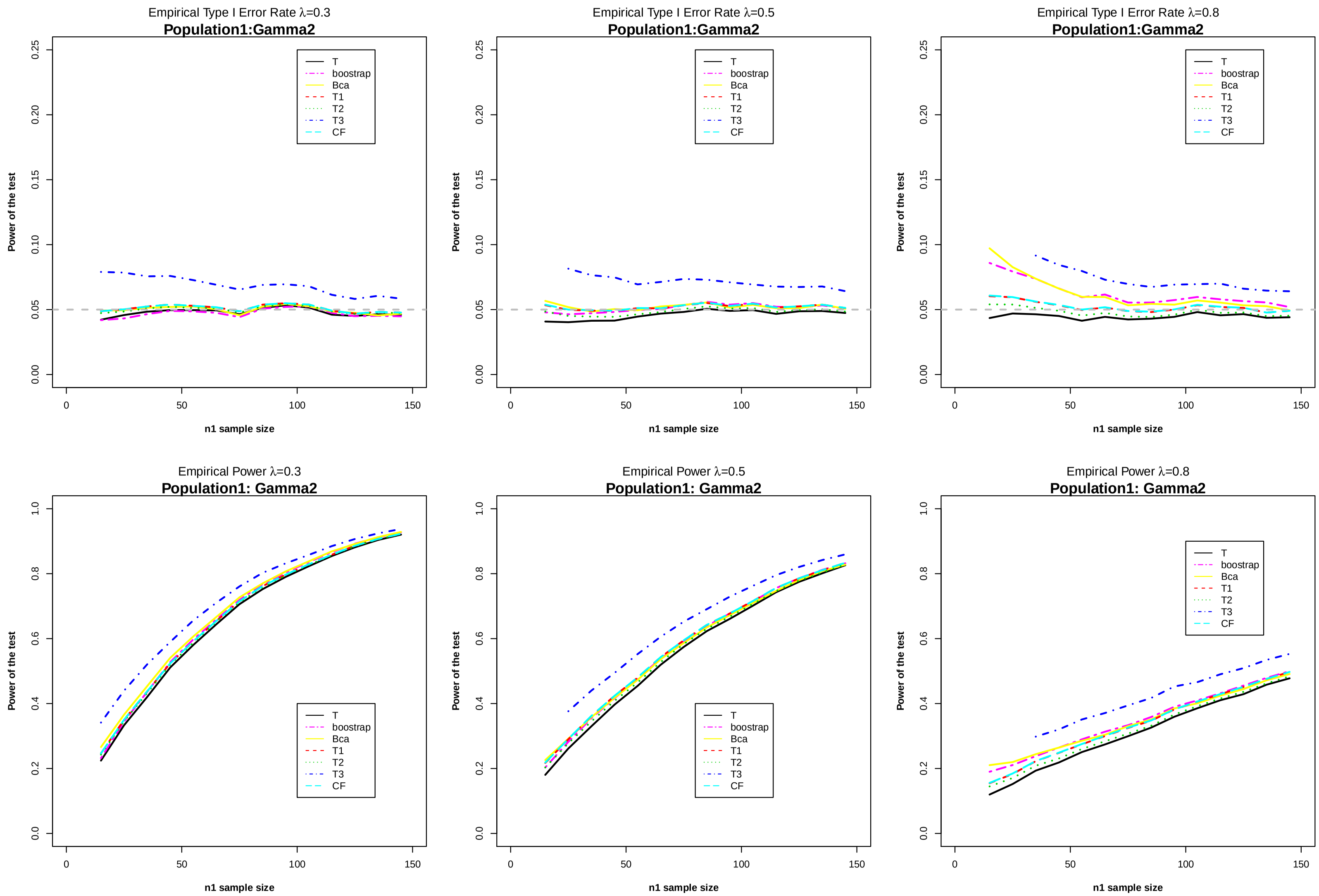

Figure 4.16: *One-sided upper-tailed Empirical Power and Type I Error rate for Pair2*

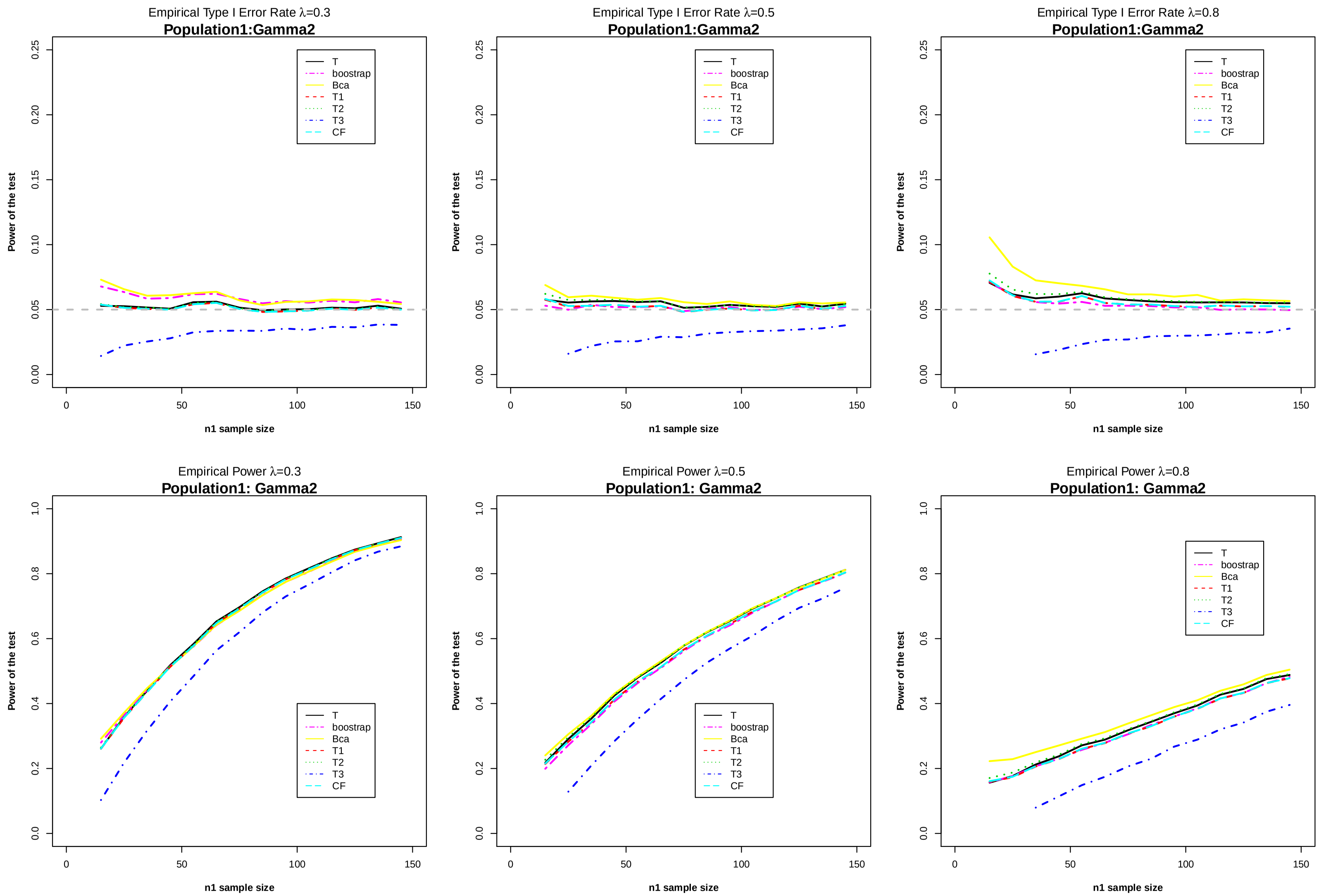

Figure 4.17: *One-sided lower-tailed Empirical Power and Type I Error rate for Pair2*

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.024	0.042	0.224	0.042	0.232	0.049	0.266	0.049	0.246	0.047	0.243	0.079	0.341	0.049	0.247
25	0.021	0.046	0.333	0.043	0.344	0.050	0.366	0.050	0.350	0.048	0.345	0.078	0.440	0.050	0.350
35	0.018	0.048	0.420	0.046	0.436	0.051	0.454	0.052	0.437	0.051	0.433	0.076	0.521	0.052	0.437
45	0.016	0.050	0.511	0.049	0.527	0.052	0.541	0.054	0.524	0.052	0.520	0.076	0.590	0.054	0.524
55	0.015	0.050	0.580	0.049	0.597	0.052	0.607	0.053	0.592	0.051	0.588	0.073	0.656	0.053	0.592
65	0.014	0.049	0.644	0.048	0.659	0.050	0.667	0.052	0.654	0.051	0.652	0.069	0.711	0.052	0.654
75	0.013	0.046	0.706	0.044	0.719	0.046	0.727	0.049	0.715	0.048	0.712	0.065	0.761	0.049	0.715
85	0.013	0.051	0.752	0.051	0.767	0.052	0.770	0.054	0.761	0.052	0.758	0.069	0.802	0.054	0.761
95	0.012	0.053	0.790	0.052	0.804	0.054	0.807	0.055	0.797	0.054	0.795	0.070	0.832	0.055	0.797
105	0.012	0.052	0.823	0.052	0.836	0.053	0.838	0.054	0.829	0.053	0.827	0.068	0.858	0.054	0.829
115	0.011	0.046	0.854	0.048	0.866	0.049	0.869	0.049	0.858	0.049	0.857	0.061	0.885	0.049	0.858
125	0.011	0.045	0.881	0.045	0.893	0.048	0.892	0.047	0.886	0.047	0.884	0.058	0.907	0.047	0.886
135	0.010	0.046	0.904	0.045	0.912	0.045	0.912	0.048	0.907	0.048	0.906	0.061	0.924	0.048	0.907
145	0.010	0.046	0.920	0.045	0.927	0.047	0.928	0.048	0.923	0.047	0.922	0.058	0.938	0.048	0.923

Table 4.27: *Proportion of rejections for one-sided upper-tailed tests for Pair2 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.024	0.053	0.261	0.068	0.280	0.073	0.292	0.054	0.261	0.054	0.265	0.014	0.103	0.054	0.261
25	0.021	0.053	0.359	0.064	0.364	0.066	0.372	0.051	0.356	0.052	0.360	0.022	0.218	0.051	0.356
35	0.018	0.052	0.440	0.058	0.443	0.061	0.448	0.050	0.436	0.052	0.439	0.025	0.317	0.050	0.435
45	0.016	0.051	0.518	0.059	0.515	0.061	0.513	0.050	0.513	0.051	0.517	0.028	0.406	0.050	0.513
55	0.015	0.056	0.583	0.062	0.575	0.063	0.576	0.054	0.577	0.055	0.581	0.032	0.484	0.054	0.577
65	0.014	0.056	0.653	0.062	0.641	0.064	0.642	0.055	0.648	0.056	0.652	0.034	0.564	0.055	0.648
75	0.013	0.051	0.697	0.058	0.686	0.057	0.686	0.050	0.692	0.051	0.695	0.034	0.620	0.050	0.692
85	0.013	0.050	0.746	0.055	0.735	0.053	0.733	0.048	0.742	0.049	0.744	0.034	0.680	0.048	0.742
95	0.012	0.050	0.785	0.056	0.774	0.056	0.774	0.049	0.782	0.050	0.784	0.035	0.730	0.049	0.782
105	0.012	0.050	0.816	0.055	0.805	0.056	0.806	0.049	0.813	0.050	0.815	0.034	0.766	0.049	0.813
115	0.011	0.051	0.847	0.057	0.839	0.058	0.837	0.051	0.844	0.051	0.847	0.037	0.805	0.051	0.844
125	0.011	0.051	0.874	0.056	0.869	0.057	0.867	0.050	0.872	0.050	0.873	0.036	0.841	0.050	0.872
135	0.010	0.053	0.894	0.058	0.887	0.056	0.887	0.051	0.892	0.052	0.893	0.038	0.868	0.051	0.892
145	0.010	0.051	0.913	0.056	0.905	0.054	0.903	0.050	0.910	0.050	0.911	0.038	0.884	0.050	0.910

Table 4.28: *Proportion of rejections for one-sided lower-tailed tests for Pair2 when* $\lambda = 0.3$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.096	0.041	0.180	0.048	0.203	0.057	0.226	0.053	0.217	0.048	0.203	0.092	0.317	0.054	0.218
25	0.085	0.040	0.262	0.046	0.282	0.052	0.292	0.050	0.290	0.045	0.278	0.082	0.376	0.050	0.291
35	0.077	0.041	0.330	0.047	0.355	0.048	0.358	0.049	0.362	0.045	0.345	0.076	0.440	0.050	0.362
45	0.070	0.042	0.397	0.048	0.416	0.050	0.418	0.049	0.424	0.044	0.408	0.075	0.494	0.049	0.424
55	0.065	0.045	0.454	0.050	0.477	0.050	0.473	0.051	0.480	0.046	0.465	0.069	0.552	0.051	0.480
65	0.060	0.047	0.519	0.050	0.538	0.052	0.537	0.051	0.543	0.048	0.528	0.071	0.606	0.051	0.543
75	0.057	0.048	0.574	0.053	0.590	0.054	0.587	0.053	0.595	0.050	0.583	0.074	0.652	0.053	0.595
85	0.053	0.051	0.623	0.056	0.639	0.056	0.634	0.055	0.642	0.052	0.630	0.073	0.691	0.055	0.642
95	0.051	0.049	0.661	0.054	0.678	0.051	0.672	0.053	0.678	0.051	0.668	0.071	0.729	0.053	0.678
105	0.049	0.050	0.702	0.055	0.715	0.054	0.710	0.054	0.716	0.051	0.707	0.069	0.763	0.055	0.716
115	0.047	0.047	0.743	0.052	0.756	0.051	0.747	0.052	0.756	0.048	0.747	0.068	0.795	0.052	0.756
125	0.045	0.049	0.775	0.051	0.786	0.051	0.780	0.052	0.786	0.050	0.778	0.067	0.821	0.052	0.786
135	0.044	0.049	0.801	0.053	0.812	0.054	0.808	0.054	0.810	0.050	0.804	0.068	0.842	0.054	0.811
145	0.042	0.047	0.826	0.051	0.832	0.051	0.828	0.051	0.834	0.048	0.828	0.064	0.860	0.051	0.834

Table 4.29: *Proportion of rejections for one-sided upper-tailed tests for Pair2 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.096	0.058	0.218	0.053	0.199	0.069	0.240	0.058	0.214	0.062	0.227	0.007	0.041	0.058	0.215
25	0.085	0.055	0.291	0.050	0.272	0.059	0.306	0.052	0.283	0.057	0.296	0.016	0.129	0.053	0.283
35	0.077	0.056	0.354	0.054	0.337	0.061	0.363	0.053	0.343	0.057	0.357	0.022	0.209	0.053	0.343
45	0.070	0.057	0.425	0.052	0.407	0.059	0.431	0.054	0.413	0.058	0.427	0.025	0.284	0.054	0.413
55	0.065	0.056	0.480	0.052	0.463	0.058	0.483	0.052	0.468	0.056	0.481	0.026	0.352	0.052	0.469
65	0.060	0.056	0.525	0.053	0.509	0.059	0.530	0.053	0.512	0.056	0.526	0.029	0.414	0.053	0.512
75	0.057	0.051	0.578	0.049	0.562	0.056	0.579	0.048	0.566	0.052	0.579	0.029	0.473	0.048	0.566
85	0.053	0.052	0.619	0.050	0.608	0.054	0.621	0.050	0.609	0.052	0.620	0.032	0.526	0.050	0.609
95	0.051	0.054	0.654	0.052	0.641	0.056	0.656	0.051	0.645	0.053	0.654	0.033	0.570	0.051	0.646
105	0.049	0.053	0.693	0.050	0.680	0.054	0.695	0.049	0.684	0.053	0.693	0.033	0.609	0.049	0.684
115	0.047	0.052	0.725	0.050	0.715	0.053	0.725	0.050	0.715	0.052	0.725	0.034	0.655	0.050	0.715
125	0.045	0.055	0.758	0.052	0.750	0.056	0.757	0.053	0.750	0.054	0.758	0.035	0.695	0.053	0.750
135	0.044	0.052	0.785	0.050	0.775	0.055	0.784	0.050	0.776	0.052	0.784	0.036	0.723	0.050	0.776
145	0.042	0.054	0.812	0.052	0.802	0.056	0.812	0.053	0.803	0.054	0.811	0.038	0.757	0.053	0.804

Table 4.30: *Proportion of rejections for one-sided lower-tailed tests for Pair2 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.130	0.043	0.120	0.086	0.190	0.097	0.210	0.060	0.155	0.054	0.145	0.108	0.243	0.061	0.156
25	0.118	0.047	0.152	0.079	0.211	0.083	0.220	0.059	0.183	0.054	0.171	0.100	0.264	0.060	0.185
35	0.105	0.046	0.193	0.074	0.238	0.074	0.244	0.056	0.223	0.051	0.208	0.092	0.298	0.056	0.223
45	0.097	0.045	0.218	0.066	0.264	0.066	0.264	0.053	0.248	0.049	0.231	0.084	0.320	0.054	0.249
55	0.089	0.041	0.251	0.059	0.290	0.060	0.285	0.050	0.275	0.045	0.260	0.080	0.351	0.050	0.275
65	0.083	0.044	0.274	0.062	0.313	0.060	0.304	0.052	0.300	0.047	0.284	0.073	0.371	0.052	0.301
75	0.078	0.042	0.299	0.055	0.333	0.053	0.330	0.049	0.325	0.045	0.307	0.070	0.394	0.049	0.326
85	0.075	0.043	0.326	0.055	0.358	0.054	0.351	0.048	0.347	0.044	0.332	0.067	0.417	0.048	0.348
95	0.071	0.044	0.359	0.057	0.390	0.054	0.385	0.050	0.382	0.046	0.366	0.069	0.453	0.050	0.382
105	0.068	0.048	0.386	0.060	0.410	0.057	0.401	0.053	0.406	0.050	0.392	0.070	0.466	0.054	0.406
115	0.065	0.046	0.410	0.058	0.432	0.055	0.424	0.052	0.429	0.048	0.416	0.070	0.490	0.052	0.429
125	0.063	0.046	0.429	0.056	0.455	0.053	0.445	0.051	0.451	0.048	0.436	0.066	0.509	0.051	0.451
135	0.060	0.044	0.458	0.056	0.479	0.053	0.472	0.048	0.476	0.045	0.462	0.065	0.534	0.048	0.476
145	0.059	0.044	0.478	0.052	0.500	0.049	0.491	0.049	0.497	0.045	0.484	0.064	0.553	0.049	0.497

Table 4.31: *Proportion of rejections for one-sided upper-tailed tests for Pair2 when* $\lambda = 0.8$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.130	0.071	0.156	0.071	0.158	0.106	0.223	0.071	0.159	0.078	0.171			0.072	0.161
25	0.118	0.062	0.177	0.061	0.177	0.083	0.229	0.060	0.173	0.065	0.188	0.008	0.035	0.061	0.175
35	0.105	0.059	0.212	0.056	0.207	0.072	0.250	0.056	0.206	0.062	0.219	0.016	0.080	0.056	0.207
45	0.097	0.060	0.237	0.055	0.230	0.070	0.271	0.056	0.229	0.062	0.242	0.019	0.114	0.056	0.229
55	0.089	0.063	0.271	0.056	0.258	0.068	0.292	0.060	0.260	0.064	0.275	0.023	0.149	0.060	0.261
65	0.083	0.059	0.289	0.053	0.279	0.066	0.312	0.055	0.278	0.059	0.293	0.027	0.174	0.055	0.279
75	0.078	0.057	0.318	0.053	0.306	0.062	0.339	0.054	0.307	0.058	0.321	0.027	0.206	0.054	0.308
85	0.075	0.056	0.344	0.053	0.334	0.062	0.364	0.054	0.330	0.057	0.346	0.029	0.230	0.054	0.330
95	0.071	0.056	0.370	0.052	0.360	0.060	0.389	0.053	0.360	0.056	0.373	0.030	0.267	0.053	0.360
105	0.068	0.056	0.394	0.052	0.385	0.061	0.410	0.052	0.384	0.056	0.397	0.030	0.289	0.052	0.384
115	0.065	0.056	0.427	0.050	0.416	0.057	0.440	0.053	0.415	0.056	0.430	0.031	0.321	0.053	0.416
125	0.063	0.056	0.445	0.050	0.433	0.058	0.459	0.052	0.433	0.056	0.447	0.032	0.341	0.052	0.433
135	0.060	0.055	0.475	0.050	0.463	0.057	0.488	0.053	0.463	0.055	0.478	0.032	0.375	0.053	0.463
145	0.059	0.055	0.488	0.050	0.482	0.057	0.504	0.052	0.478	0.056	0.491	0.035	0.396	0.052	0.478

Table 4.32: *Proportion of rejections for one-sided lower-tailed tests for Pair2 when* $\lambda = 0.8$

$Pair3$ has a distribution combination with big skewness and small common variance. The empirical results of $Pair3$ are presented in Figure 4.18 and Figure 4.19. The numerical results are in Table 4.33, 4.35, 4.37 and Table 4.34, 4.36, 4.38. Since $Pair3$ has bigger skewness than $Pair1$ and $Pair2$, the simulation results are quite different.

From Figure 4.18, when $\lambda = 0.3$, the bootstrap-t test and BCa test provide an empirical type I error rate below 0.05, but also give the highest power. when $\lambda = 0.5$ with balanced sample, the highest power goes to TCF test and T_3 test, which both keep the type I error rate close to $\alpha = 0.05$. when $\lambda = 0.8$, the majority of the samples are from the first population which is a highly skewed Gamma distribution. The type I error rate of all the seven tests are bigger than α except the pooled two-sample t-test. As the sample size n_1 increasing, the type I error rate of all seven tests decrease and drop below 0.1. The empirical power of TCF, T_1 and T_3 tests are higher than the rest of tests, but these three test also give a bigger type I error rate.

From Figure 4.19, when $\lambda = 0.3$ the empirical type I error rates of bootstrap-t test and BCa test are consistently above 0.1 even with a large sample size $n_1 = 150$. The rest of the tests maintain the type I error rate at or below α, while give the similar power. As λ increases to 0.5, only TCF, T_1 and T_3 tests keep the type I error rate close to α, then give slightly smaller power than the other four tests. Given $\lambda = 0.8$, all the tests have a type I error rate bigger or equal to 0.1 except for bootstrap-t test. Although the bootstrap-t test maintains the type I error rate below 0.05, its empirical power is much smaller than the other six tests.

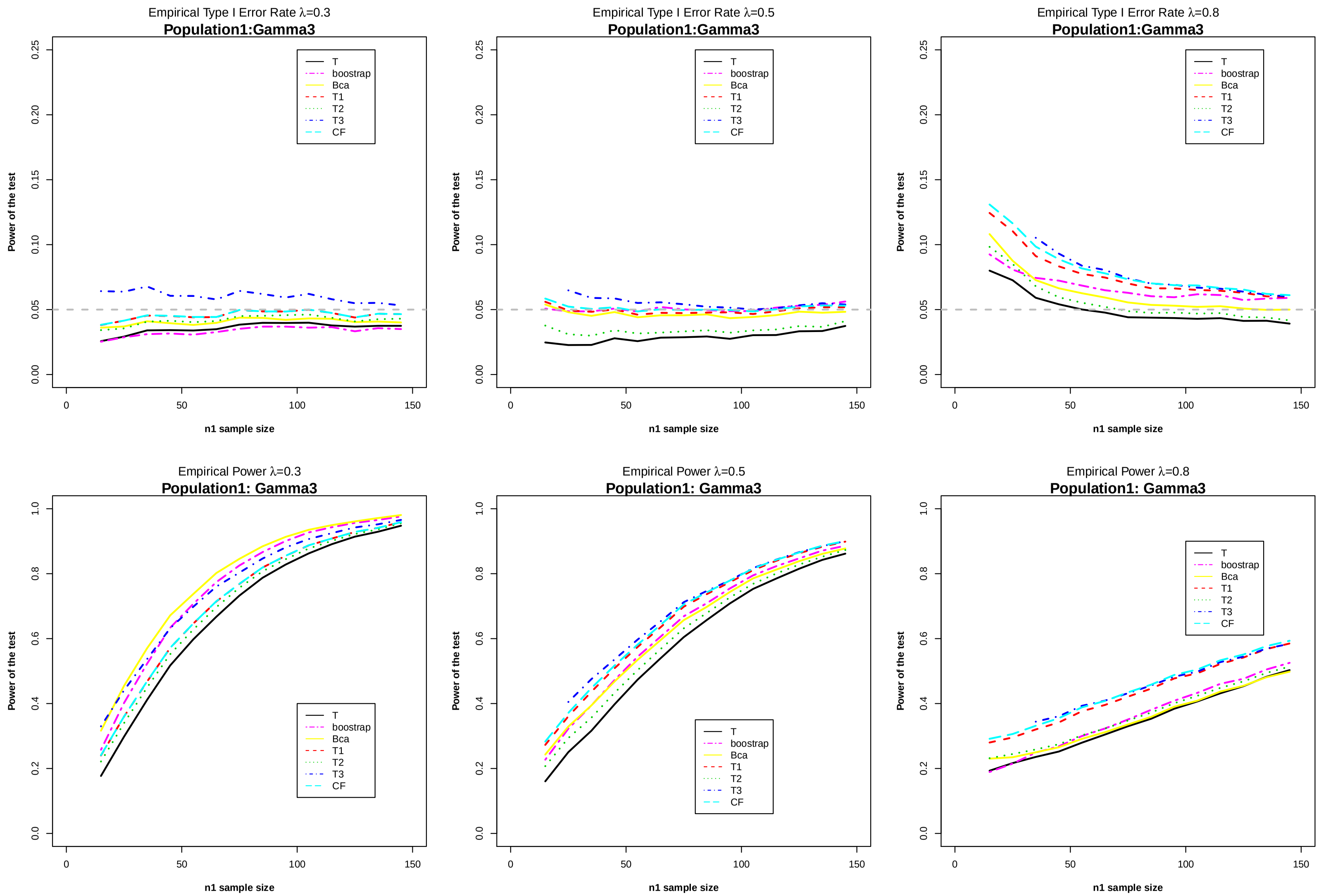

Figure 4.18: *One-sided upper-tailed Empirical Power and Type I Error rate for Pair3*

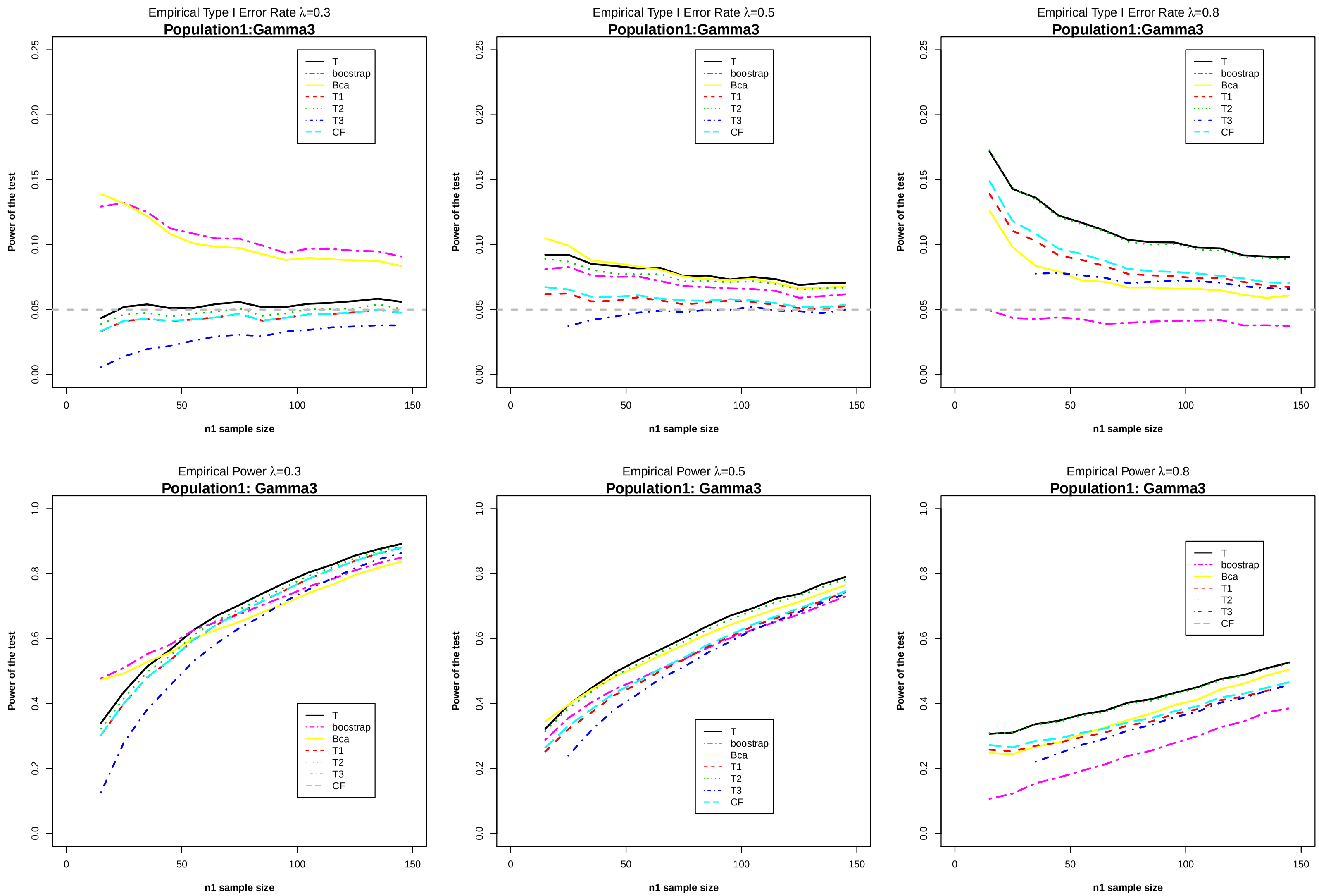

Figure 4.19: *One-sided lower-tailed Empirical Power and Type I Error rate for Pair3*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.096	0.026	0.177	0.025	0.258	0.036	0.316	0.038	0.240	0.034	0.222	0.064	0.330	0.038	0.241
25	0.090	0.029	0.298	0.029	0.402	0.037	0.455	0.042	0.360	0.035	0.340	0.064	0.441	0.042	0.360
35	0.080	0.034	0.412	0.031	0.523	0.041	0.571	0.045	0.468	0.041	0.450	0.068	0.537	0.046	0.469
45	0.077	0.034	0.517	0.032	0.634	0.040	0.672	0.045	0.572	0.041	0.552	0.061	0.632	0.045	0.572
55	0.074	0.034	0.598	0.031	0.707	0.038	0.737	0.044	0.646	0.040	0.628	0.060	0.698	0.044	0.646
65	0.070	0.035	0.667	0.033	0.774	0.040	0.802	0.044	0.715	0.041	0.697	0.058	0.760	0.044	0.715
75	0.068	0.038	0.733	0.035	0.825	0.044	0.846	0.050	0.769	0.045	0.757	0.064	0.803	0.050	0.769
85	0.066	0.040	0.787	0.037	0.866	0.044	0.884	0.049	0.819	0.045	0.807	0.062	0.847	0.049	0.819
95	0.063	0.040	0.828	0.037	0.901	0.042	0.913	0.048	0.855	0.046	0.845	0.059	0.881	0.048	0.855
105	0.061	0.040	0.862	0.036	0.927	0.043	0.935	0.050	0.887	0.046	0.877	0.062	0.907	0.050	0.887
115	0.060	0.038	0.891	0.036	0.943	0.043	0.950	0.047	0.908	0.044	0.901	0.058	0.925	0.047	0.908
125	0.058	0.037	0.914	0.033	0.957	0.040	0.961	0.044	0.928	0.041	0.922	0.055	0.942	0.044	0.929
135	0.057	0.037	0.929	0.036	0.965	0.041	0.971	0.047	0.941	0.043	0.936	0.055	0.952	0.047	0.941
145	0.056	0.037	0.948	0.035	0.976	0.040	0.981	0.046	0.959	0.043	0.956	0.053	0.966	0.046	0.959

Table 4.33: *Proportion of rejections for one-sided upper-tailed tests for Pair3 when* $\lambda = 0.3$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.096	0.043	0.340	0.129	0.477	0.139	0.474	0.033	0.303	0.039	0.323	0.006	0.126	0.034	0.304
25	0.090	0.052	0.436	0.132	0.510	0.132	0.493	0.041	0.400	0.046	0.418	0.014	0.281	0.041	0.401
35	0.080	0.054	0.514	0.125	0.553	0.122	0.528	0.043	0.481	0.048	0.497	0.020	0.381	0.043	0.481
45	0.077	0.051	0.566	0.112	0.582	0.108	0.557	0.041	0.533	0.045	0.548	0.022	0.456	0.041	0.533
55	0.074	0.051	0.626	0.108	0.626	0.101	0.600	0.042	0.595	0.047	0.610	0.026	0.530	0.042	0.596
65	0.070	0.054	0.670	0.105	0.651	0.098	0.628	0.044	0.642	0.048	0.656	0.029	0.585	0.044	0.642
75	0.068	0.056	0.704	0.105	0.675	0.097	0.651	0.047	0.680	0.051	0.690	0.031	0.633	0.047	0.680
85	0.066	0.052	0.739	0.099	0.704	0.093	0.681	0.041	0.716	0.045	0.725	0.029	0.670	0.042	0.716
95	0.063	0.052	0.773	0.093	0.731	0.088	0.709	0.044	0.750	0.047	0.759	0.033	0.716	0.044	0.750
105	0.061	0.054	0.804	0.097	0.761	0.090	0.741	0.046	0.784	0.050	0.793	0.034	0.752	0.046	0.785
115	0.060	0.055	0.827	0.097	0.783	0.089	0.766	0.047	0.812	0.050	0.819	0.036	0.785	0.047	0.813
125	0.058	0.057	0.856	0.095	0.810	0.088	0.796	0.048	0.840	0.051	0.848	0.037	0.817	0.048	0.840
135	0.057	0.058	0.875	0.095	0.832	0.087	0.818	0.050	0.862	0.054	0.869	0.038	0.844	0.050	0.862
145	0.056	0.056	0.892	0.091	0.849	0.084	0.837	0.047	0.880	0.050	0.885	0.038	0.863	0.048	0.880

Table 4.34: *Proportion of rejections for one-sided lower-tailed tests for Pair3 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.401	0.025	0.160	0.051	0.227	0.054	0.243	0.056	0.273	0.038	0.207	0.079	0.335	0.058	0.283
25	0.369	0.023	0.250	0.049	0.323	0.048	0.329	0.049	0.361	0.031	0.294	0.065	0.405	0.052	0.371
35	0.346	0.023	0.316	0.049	0.394	0.045	0.394	0.048	0.436	0.030	0.357	0.059	0.475	0.050	0.448
45	0.327	0.028	0.398	0.051	0.472	0.048	0.468	0.050	0.509	0.034	0.433	0.059	0.536	0.052	0.517
55	0.310	0.026	0.473	0.048	0.544	0.044	0.535	0.046	0.574	0.032	0.503	0.055	0.597	0.049	0.581
65	0.298	0.028	0.540	0.052	0.606	0.046	0.596	0.048	0.635	0.032	0.568	0.056	0.653	0.050	0.643
75	0.287	0.029	0.605	0.050	0.669	0.046	0.657	0.047	0.698	0.033	0.631	0.054	0.712	0.050	0.705
85	0.276	0.029	0.657	0.050	0.709	0.046	0.698	0.048	0.737	0.034	0.679	0.052	0.747	0.050	0.743
95	0.267	0.028	0.708	0.049	0.755	0.043	0.744	0.048	0.774	0.032	0.727	0.051	0.779	0.050	0.779
105	0.260	0.030	0.753	0.049	0.795	0.044	0.785	0.047	0.811	0.034	0.769	0.050	0.817	0.049	0.816
115	0.253	0.030	0.785	0.051	0.823	0.046	0.812	0.049	0.840	0.035	0.800	0.051	0.843	0.050	0.843
125	0.246	0.033	0.815	0.053	0.847	0.048	0.839	0.051	0.863	0.037	0.828	0.053	0.866	0.052	0.865
135	0.241	0.034	0.842	0.054	0.871	0.048	0.861	0.052	0.883	0.037	0.852	0.055	0.885	0.053	0.886
145	0.235	0.037	0.862	0.056	0.886	0.048	0.878	0.052	0.899	0.041	0.873	0.054	0.900	0.053	0.901

Table 4.35: *Proportion of rejections for one-sided upper-tailed tests for Pair3 when* $\lambda = 0.5$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.401	0.092	0.322	0.081	0.288	0.105	0.344	0.062	0.252	0.089	0.315	0.019	0.116	0.067	0.264
25	0.369	0.092	0.395	0.083	0.355	0.099	0.396	0.062	0.321	0.087	0.385	0.037	0.240	0.066	0.331
35	0.346	0.085	0.448	0.076	0.404	0.088	0.443	0.056	0.372	0.081	0.436	0.042	0.318	0.060	0.382
45	0.327	0.084	0.495	0.075	0.444	0.086	0.482	0.057	0.426	0.078	0.485	0.045	0.382	0.060	0.433
55	0.310	0.082	0.533	0.076	0.474	0.083	0.513	0.059	0.460	0.077	0.521	0.048	0.428	0.061	0.468
65	0.298	0.082	0.567	0.072	0.507	0.081	0.548	0.057	0.501	0.077	0.557	0.049	0.478	0.058	0.508
75	0.287	0.076	0.601	0.068	0.537	0.075	0.580	0.054	0.535	0.072	0.591	0.048	0.513	0.057	0.541
85	0.276	0.076	0.638	0.067	0.571	0.073	0.614	0.055	0.574	0.072	0.627	0.050	0.555	0.057	0.580
95	0.267	0.073	0.670	0.066	0.602	0.073	0.642	0.057	0.605	0.071	0.659	0.050	0.590	0.058	0.611
105	0.260	0.075	0.694	0.066	0.628	0.074	0.667	0.056	0.638	0.072	0.686	0.052	0.627	0.057	0.644
115	0.253	0.073	0.723	0.064	0.652	0.070	0.692	0.053	0.663	0.070	0.712	0.049	0.654	0.055	0.669
125	0.246	0.069	0.738	0.059	0.674	0.066	0.714	0.051	0.689	0.065	0.731	0.049	0.682	0.052	0.693
135	0.241	0.070	0.767	0.060	0.703	0.067	0.740	0.050	0.716	0.066	0.759	0.047	0.710	0.052	0.720
145	0.235	0.071	0.789	0.062	0.730	0.068	0.764	0.053	0.743	0.067	0.782	0.050	0.738	0.054	0.746

Table 4.36: *Proportion of rejections for one-sided lower-tailed tests for Pair3 when* $\lambda = 0.5$

n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.549	0.080	0.193	0.092	0.189	0.108	0.230	0.124	0.280	0.098	0.232	0.150	0.322	0.131	0.291
25	0.515	0.073	0.217	0.081	0.215	0.088	0.235	0.110	0.296	0.085	0.244	0.128	0.327	0.116	0.306
35	0.476	0.059	0.236	0.074	0.249	0.073	0.250	0.091	0.320	0.068	0.259	0.105	0.344	0.098	0.332
45	0.456	0.054	0.252	0.072	0.268	0.066	0.265	0.083	0.342	0.059	0.275	0.093	0.361	0.089	0.355
55	0.431	0.050	0.280	0.068	0.300	0.063	0.291	0.077	0.376	0.055	0.302	0.084	0.393	0.082	0.388
65	0.415	0.048	0.304	0.065	0.323	0.059	0.312	0.075	0.396	0.052	0.323	0.080	0.409	0.078	0.408
75	0.397	0.044	0.330	0.063	0.351	0.056	0.336	0.070	0.421	0.049	0.348	0.074	0.432	0.073	0.434
85	0.385	0.044	0.353	0.060	0.381	0.054	0.359	0.066	0.446	0.047	0.372	0.070	0.456	0.070	0.458
95	0.371	0.043	0.384	0.059	0.409	0.053	0.390	0.066	0.477	0.048	0.401	0.069	0.483	0.069	0.488
105	0.361	0.043	0.406	0.062	0.433	0.052	0.409	0.065	0.493	0.047	0.424	0.067	0.498	0.069	0.505
115	0.349	0.043	0.432	0.061	0.461	0.053	0.437	0.064	0.522	0.047	0.448	0.066	0.528	0.067	0.533
125	0.342	0.041	0.454	0.057	0.476	0.051	0.455	0.063	0.542	0.044	0.468	0.064	0.544	0.065	0.551
135	0.333	0.041	0.482	0.058	0.505	0.050	0.481	0.060	0.568	0.044	0.494	0.061	0.570	0.062	0.577
145	0.327	0.039	0.503	0.059	0.525	0.050	0.498	0.059	0.585	0.042	0.514	0.060	0.584	0.061	0.594

Table 4.37: *Proportion of rejections for one-sided upper-tailed tests for Pair3 when* $\lambda = 0.8$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.549	0.172	0.307	0.049	0.107	0.126	0.248	0.139	0.258	0.173	0.310			0.149	0.272
25	0.515	0.143	0.311	0.044	0.123	0.098	0.244	0.111	0.253	0.144	0.310	0.074	0.179	0.118	0.265
35	0.476	0.136	0.337	0.043	0.155	0.084	0.268	0.103	0.270	0.135	0.335	0.078	0.221	0.108	0.285
45	0.456	0.122	0.347	0.044	0.172	0.079	0.280	0.092	0.280	0.121	0.344	0.078	0.247	0.097	0.292
55	0.431	0.117	0.366	0.043	0.193	0.072	0.306	0.088	0.296	0.116	0.362	0.076	0.273	0.093	0.310
65	0.415	0.111	0.378	0.039	0.213	0.071	0.327	0.084	0.311	0.110	0.374	0.074	0.292	0.087	0.323
75	0.397	0.104	0.403	0.040	0.239	0.067	0.349	0.077	0.333	0.102	0.399	0.070	0.317	0.081	0.343
85	0.385	0.102	0.413	0.041	0.255	0.067	0.369	0.076	0.345	0.100	0.408	0.071	0.334	0.080	0.355
95	0.371	0.102	0.433	0.041	0.278	0.066	0.395	0.076	0.368	0.100	0.429	0.072	0.358	0.079	0.376
105	0.361	0.098	0.450	0.042	0.299	0.066	0.412	0.074	0.383	0.096	0.447	0.072	0.375	0.078	0.392
115	0.349	0.097	0.476	0.042	0.327	0.065	0.444	0.074	0.410	0.095	0.472	0.071	0.403	0.076	0.419
125	0.342	0.092	0.488	0.038	0.345	0.061	0.461	0.071	0.421	0.091	0.484	0.068	0.417	0.074	0.430
135	0.333	0.091	0.509	0.038	0.373	0.059	0.486	0.069	0.440	0.089	0.505	0.066	0.439	0.071	0.449
145	0.327	0.090	0.527	0.037	0.386	0.061	0.505	0.067	0.457	0.089	0.524	0.066	0.458	0.070	0.466

Table 4.38: *Proportion of rejections for one-sided lower-tailed tests for Pair3 when* $\lambda = 0.8$

$Pair4$ has a distribution combination with big skewness and big common variance. Figure 4.20 and Figure 4.21 present the empirical results of type I error rate and power of upper-tailed and lower-tailed tests respectively. The numerical results are given in Table 4.39, 4.41, 4.43 and Table 4.40, 4.42, 4.44. Despite that the populations setting of $Pair4$ has a much bigger common variance than the population setting of $Pair3$, their simulation results are quite the same. In this case, we omit the discussion of the simulation results of $Pair4$.

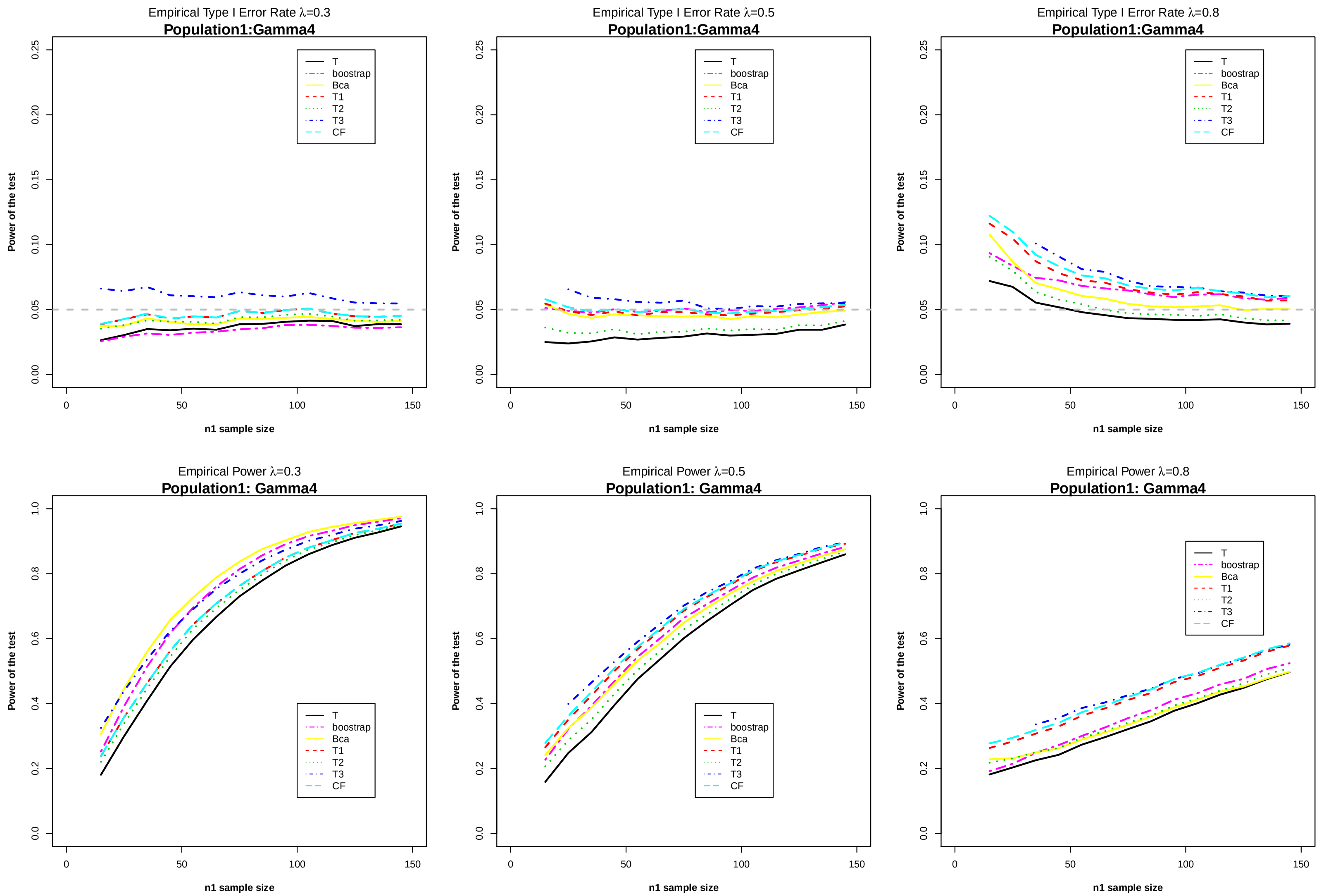

Figure 4.20: *One-sided upper-tailed Empirical Power and Type I Error rate for Pair4*

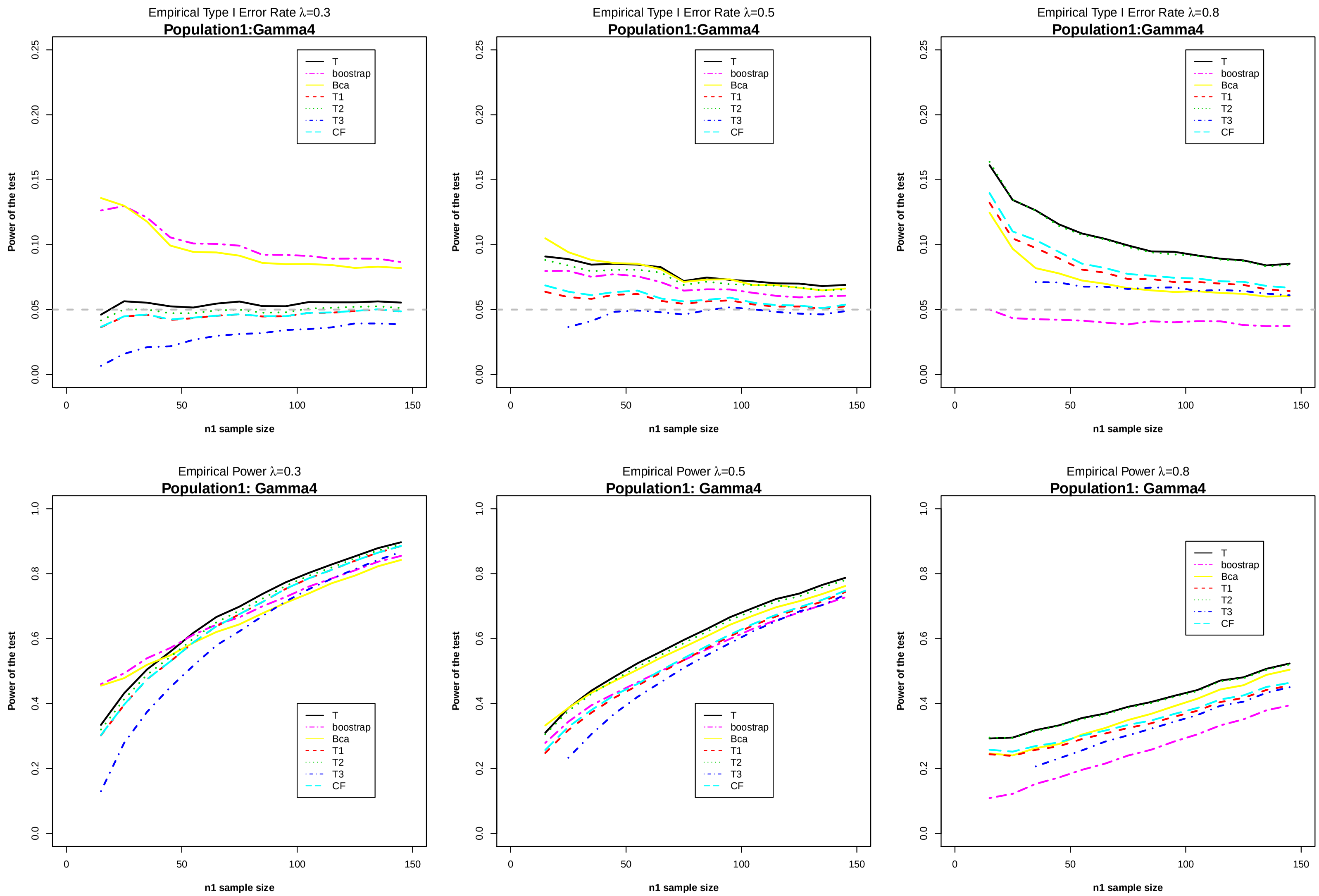

Figure 4.21: *One-sided lower-tailed Empirical Power and Type I Error rate for Pair4*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.092	0.026	0.181	0.025	0.252	0.037	0.307	0.039	0.239	0.035	0.221	0.066	0.325	0.039	0.239
25	0.085	0.030	0.299	0.029	0.392	0.038	0.446	0.043	0.358	0.038	0.339	0.064	0.439	0.043	0.358
35	0.076	0.035	0.409	0.032	0.514	0.043	0.560	0.046	0.463	0.042	0.446	0.067	0.537	0.047	0.464
45	0.073	0.034	0.514	0.030	0.617	0.040	0.658	0.043	0.563	0.041	0.544	0.061	0.623	0.043	0.563
55	0.070	0.035	0.598	0.032	0.696	0.039	0.728	0.045	0.644	0.041	0.631	0.060	0.691	0.045	0.645
65	0.065	0.034	0.667	0.033	0.760	0.039	0.789	0.044	0.708	0.039	0.694	0.059	0.753	0.044	0.709
75	0.064	0.039	0.730	0.035	0.814	0.043	0.837	0.049	0.762	0.044	0.750	0.063	0.800	0.049	0.763
85	0.062	0.039	0.779	0.036	0.857	0.043	0.876	0.047	0.809	0.044	0.798	0.061	0.842	0.048	0.809
95	0.058	0.041	0.825	0.038	0.891	0.044	0.903	0.050	0.850	0.046	0.841	0.060	0.874	0.050	0.850
105	0.057	0.042	0.860	0.038	0.916	0.044	0.928	0.051	0.880	0.047	0.874	0.063	0.901	0.051	0.880
115	0.056	0.041	0.888	0.037	0.932	0.043	0.944	0.047	0.903	0.045	0.896	0.059	0.919	0.047	0.903
125	0.054	0.037	0.911	0.036	0.949	0.042	0.956	0.045	0.925	0.041	0.919	0.055	0.939	0.045	0.925
135	0.053	0.039	0.927	0.036	0.960	0.040	0.966	0.044	0.938	0.042	0.934	0.055	0.949	0.044	0.938
145	0.052	0.039	0.946	0.036	0.971	0.041	0.975	0.045	0.954	0.042	0.952	0.055	0.962	0.045	0.954

Table 4.39: *Proportion of rejections for one-sided upper-tailed tests for Pair4 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.092	0.046	0.334	0.126	0.460	0.136	0.455	0.036	0.302	0.042	0.320	0.007	0.130	0.037	0.303
25	0.085	0.056	0.431	0.130	0.493	0.130	0.478	0.045	0.396	0.050	0.413	0.016	0.278	0.045	0.396
35	0.076	0.055	0.505	0.121	0.540	0.118	0.520	0.046	0.475	0.050	0.489	0.021	0.375	0.046	0.475
45	0.073	0.052	0.560	0.106	0.571	0.099	0.548	0.042	0.529	0.047	0.543	0.022	0.450	0.042	0.530
55	0.070	0.052	0.618	0.101	0.612	0.094	0.587	0.043	0.588	0.047	0.600	0.027	0.516	0.044	0.588
65	0.065	0.055	0.667	0.101	0.643	0.094	0.620	0.045	0.638	0.049	0.650	0.030	0.579	0.045	0.638
75	0.064	0.056	0.699	0.099	0.666	0.091	0.644	0.046	0.675	0.050	0.686	0.031	0.623	0.046	0.675
85	0.062	0.053	0.738	0.092	0.700	0.086	0.678	0.045	0.713	0.048	0.723	0.032	0.669	0.045	0.713
95	0.058	0.053	0.773	0.092	0.728	0.085	0.710	0.045	0.753	0.048	0.762	0.034	0.714	0.045	0.753
105	0.057	0.056	0.802	0.091	0.760	0.085	0.739	0.047	0.785	0.051	0.793	0.035	0.752	0.048	0.785
115	0.056	0.056	0.829	0.089	0.785	0.084	0.771	0.048	0.812	0.051	0.819	0.036	0.785	0.048	0.812
125	0.054	0.056	0.853	0.089	0.810	0.082	0.794	0.049	0.840	0.052	0.847	0.039	0.814	0.049	0.840
135	0.053	0.056	0.879	0.089	0.837	0.083	0.823	0.050	0.865	0.052	0.871	0.039	0.842	0.050	0.865
145	0.052	0.055	0.897	0.087	0.855	0.082	0.842	0.048	0.885	0.051	0.890	0.039	0.867	0.049	0.885

Table 4.40: *Proportion of rejections for one-sided lower-tailed tests for Pair4 when* $\lambda = 0.3$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.383	0.025	0.159	0.051	0.227	0.054	0.241	0.055	0.265	0.036	0.206	0.081	0.333	0.058	0.279
25	0.350	0.024	0.248	0.049	0.322	0.047	0.324	0.049	0.351	0.032	0.287	0.066	0.399	0.052	0.362
35	0.327	0.025	0.312	0.048	0.393	0.044	0.387	0.046	0.425	0.032	0.351	0.059	0.465	0.048	0.436
45	0.307	0.029	0.395	0.050	0.469	0.046	0.462	0.049	0.500	0.035	0.431	0.058	0.529	0.050	0.509
55	0.291	0.027	0.475	0.048	0.544	0.046	0.533	0.045	0.567	0.031	0.503	0.056	0.590	0.048	0.575
65	0.279	0.028	0.538	0.049	0.603	0.045	0.588	0.048	0.626	0.033	0.564	0.055	0.646	0.050	0.632
75	0.268	0.029	0.601	0.051	0.663	0.045	0.649	0.048	0.686	0.033	0.628	0.057	0.702	0.050	0.691
85	0.257	0.032	0.654	0.048	0.706	0.045	0.695	0.046	0.728	0.035	0.674	0.051	0.743	0.048	0.733
95	0.249	0.030	0.703	0.049	0.748	0.043	0.739	0.045	0.765	0.034	0.722	0.050	0.777	0.047	0.770
105	0.241	0.031	0.750	0.049	0.788	0.045	0.777	0.047	0.808	0.035	0.765	0.053	0.816	0.048	0.811
115	0.235	0.031	0.784	0.050	0.818	0.044	0.808	0.048	0.836	0.034	0.799	0.052	0.841	0.049	0.838
125	0.228	0.034	0.810	0.052	0.841	0.046	0.832	0.050	0.856	0.038	0.823	0.054	0.861	0.050	0.858
135	0.223	0.034	0.835	0.053	0.863	0.048	0.855	0.050	0.877	0.038	0.845	0.055	0.882	0.052	0.879
145	0.217	0.038	0.860	0.056	0.883	0.049	0.875	0.053	0.892	0.041	0.869	0.055	0.896	0.054	0.895

Table 4.41: *Proportion of rejections for one-sided upper-tailed tests for Pair4 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.383	0.091	0.310	0.080	0.279	0.105	0.333	0.064	0.248	0.088	0.305	0.019	0.112	0.069	0.257
25	0.350	0.089	0.386	0.080	0.344	0.094	0.386	0.060	0.318	0.084	0.377	0.037	0.233	0.064	0.328
35	0.327	0.085	0.440	0.075	0.395	0.088	0.433	0.058	0.372	0.080	0.430	0.041	0.306	0.061	0.380
45	0.307	0.085	0.483	0.077	0.431	0.086	0.469	0.061	0.418	0.081	0.473	0.048	0.369	0.064	0.426
55	0.291	0.085	0.524	0.076	0.466	0.085	0.504	0.062	0.455	0.081	0.513	0.049	0.421	0.065	0.461
65	0.279	0.083	0.559	0.071	0.500	0.081	0.541	0.057	0.495	0.078	0.549	0.048	0.465	0.059	0.502
75	0.268	0.072	0.596	0.065	0.533	0.071	0.574	0.054	0.534	0.069	0.586	0.046	0.510	0.056	0.540
85	0.257	0.075	0.630	0.066	0.567	0.073	0.608	0.056	0.571	0.071	0.621	0.050	0.549	0.058	0.577
95	0.249	0.073	0.666	0.066	0.599	0.073	0.643	0.057	0.607	0.070	0.657	0.052	0.586	0.059	0.612
105	0.241	0.072	0.694	0.063	0.630	0.069	0.671	0.054	0.641	0.069	0.686	0.050	0.622	0.055	0.645
115	0.235	0.070	0.722	0.061	0.657	0.069	0.697	0.052	0.668	0.068	0.714	0.048	0.654	0.053	0.673
125	0.228	0.070	0.739	0.059	0.680	0.067	0.715	0.052	0.693	0.067	0.730	0.047	0.684	0.053	0.697
135	0.223	0.068	0.765	0.060	0.704	0.065	0.737	0.050	0.714	0.065	0.758	0.046	0.703	0.051	0.720
145	0.217	0.069	0.787	0.061	0.727	0.066	0.762	0.052	0.744	0.066	0.780	0.049	0.735	0.054	0.748

Table 4.42: *Proportion of rejections for one-sided lower-tailed tests for Pair4 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.524	0.072	0.182	0.093	0.192	0.108	0.228	0.116	0.263	0.091	0.218	0.144	0.309	0.122	0.277
25	0.489	0.068	0.203	0.084	0.214	0.087	0.231	0.105	0.283	0.080	0.231	0.123	0.316	0.110	0.295
35	0.449	0.055	0.225	0.074	0.248	0.070	0.249	0.087	0.307	0.064	0.250	0.101	0.336	0.092	0.318
45	0.429	0.052	0.242	0.072	0.272	0.066	0.263	0.078	0.330	0.058	0.263	0.091	0.356	0.083	0.341
55	0.404	0.048	0.274	0.068	0.300	0.060	0.289	0.073	0.363	0.054	0.295	0.081	0.386	0.076	0.373
65	0.389	0.046	0.296	0.066	0.326	0.058	0.311	0.071	0.385	0.050	0.315	0.079	0.403	0.074	0.394
75	0.372	0.043	0.321	0.065	0.355	0.054	0.335	0.066	0.411	0.047	0.341	0.072	0.425	0.069	0.420
85	0.359	0.043	0.346	0.062	0.380	0.052	0.360	0.063	0.433	0.046	0.363	0.068	0.446	0.066	0.444
95	0.346	0.042	0.378	0.060	0.412	0.052	0.388	0.061	0.466	0.046	0.394	0.067	0.476	0.065	0.476
105	0.336	0.042	0.401	0.061	0.432	0.052	0.412	0.063	0.484	0.045	0.416	0.067	0.492	0.067	0.494
115	0.324	0.042	0.428	0.062	0.459	0.053	0.437	0.062	0.511	0.046	0.440	0.064	0.520	0.064	0.520
125	0.317	0.040	0.448	0.059	0.476	0.050	0.452	0.060	0.533	0.043	0.462	0.063	0.540	0.062	0.541
135	0.308	0.039	0.475	0.058	0.506	0.050	0.477	0.057	0.559	0.042	0.490	0.061	0.565	0.059	0.567
145	0.302	0.039	0.497	0.059	0.524	0.050	0.498	0.057	0.579	0.042	0.509	0.060	0.582	0.060	0.586

Table 4.43: *Proportion of rejections for one-sided upper-tailed tests for Pair4 when* $\lambda = 0.8$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.524	0.161	0.292	0.050	0.109	0.125	0.246	0.132	0.244	0.164	0.295			0.140	0.258
25	0.489	0.134	0.295	0.043	0.122	0.097	0.240	0.105	0.239	0.135	0.294	0.068	0.162	0.110	0.252
35	0.449	0.126	0.318	0.043	0.153	0.082	0.263	0.097	0.258	0.126	0.315	0.071	0.206	0.103	0.269
45	0.429	0.116	0.333	0.042	0.172	0.078	0.275	0.089	0.268	0.114	0.331	0.071	0.231	0.094	0.280
55	0.404	0.109	0.355	0.042	0.196	0.072	0.305	0.081	0.291	0.107	0.352	0.068	0.256	0.085	0.301
65	0.389	0.104	0.369	0.040	0.215	0.070	0.325	0.078	0.307	0.104	0.366	0.068	0.283	0.082	0.317
75	0.372	0.099	0.390	0.039	0.240	0.066	0.349	0.073	0.325	0.098	0.387	0.066	0.302	0.077	0.334
85	0.359	0.095	0.405	0.041	0.258	0.065	0.368	0.074	0.339	0.094	0.401	0.067	0.323	0.076	0.348
95	0.346	0.094	0.424	0.040	0.283	0.064	0.392	0.071	0.359	0.092	0.420	0.067	0.344	0.074	0.368
105	0.336	0.092	0.442	0.041	0.305	0.064	0.415	0.071	0.378	0.091	0.438	0.064	0.364	0.074	0.386
115	0.324	0.089	0.471	0.041	0.333	0.063	0.444	0.070	0.405	0.088	0.468	0.065	0.393	0.072	0.413
125	0.317	0.088	0.481	0.038	0.351	0.062	0.456	0.069	0.417	0.087	0.478	0.064	0.406	0.071	0.425
135	0.308	0.084	0.507	0.037	0.379	0.060	0.488	0.066	0.442	0.083	0.504	0.062	0.433	0.068	0.451
145	0.302	0.085	0.523	0.037	0.394	0.060	0.504	0.064	0.455	0.084	0.520	0.061	0.450	0.067	0.464

Table 4.44: *Proportion of rejections for one-sided lower-tailed tests for Pair4 when* $\lambda = 0.8$

As we know, the population 1 of $Pair5$ follows Log-normal distribution and the $Pair5$ has a distribution combination with small skewness and small common variance, which is similar to the setting of $Pair1$. Although the two population settings choose different skewed populations, their simulation results are similar. We found that the empirical results for the one-sided upper-tailed test in Figure 4.22 for $Pair5$ have the same pattern as the results in Figure 4.14 for $Pair1$. Besides, the empirical results of the one-sided lower-tailed test of $Pair5$ in Figure 4.23 also resemble those for $Pair1$. The numerical results of the empirical type I error rate and power of the upper-tailed and lower-tailed test are given in Table 4.45, 4.47, 4.49 and Table 4.46, 4.48, 4.50 respectively.

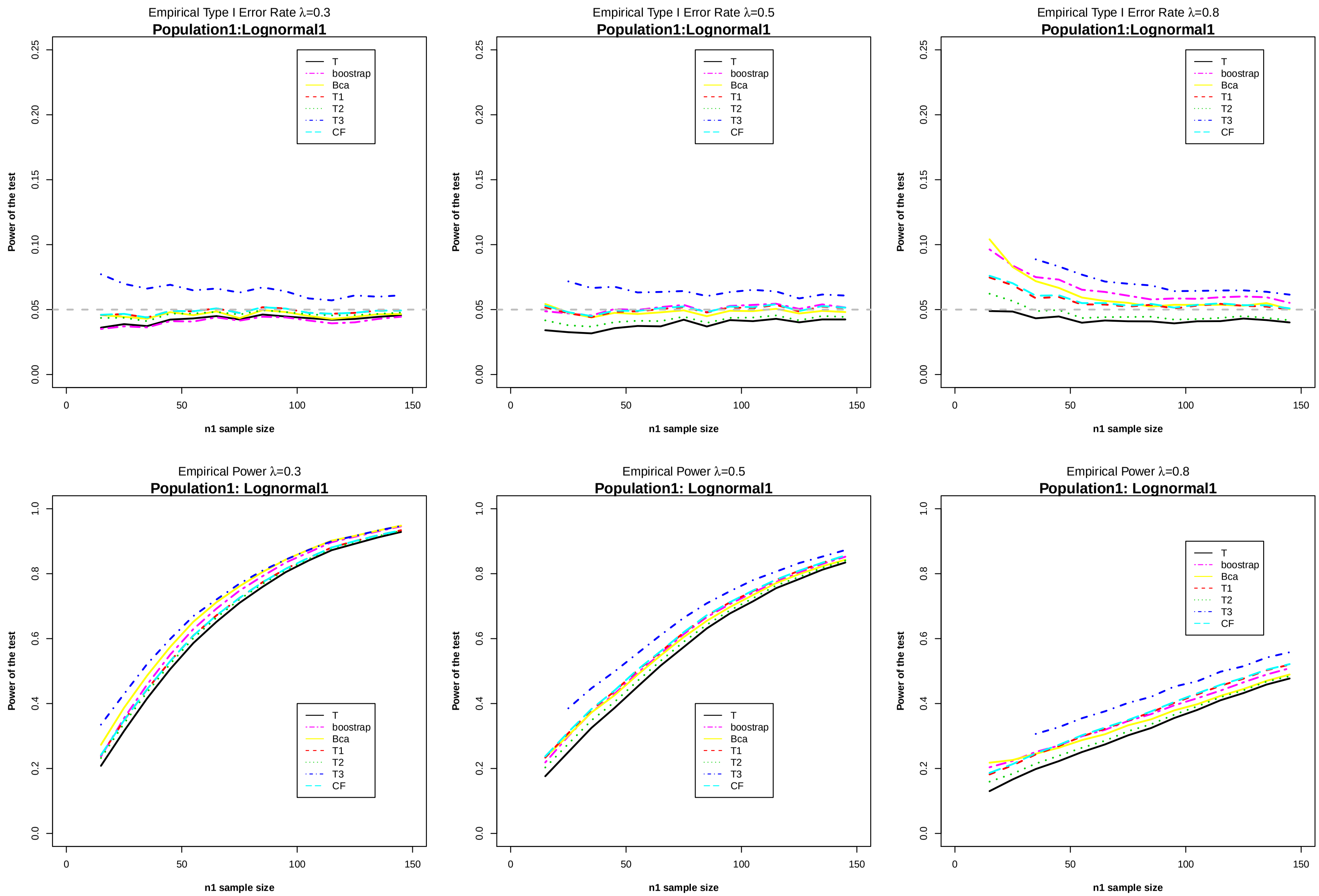

Figure 4.22: *One-sided upper-tailed Empirical Power and Type I Error rate for Pair5*

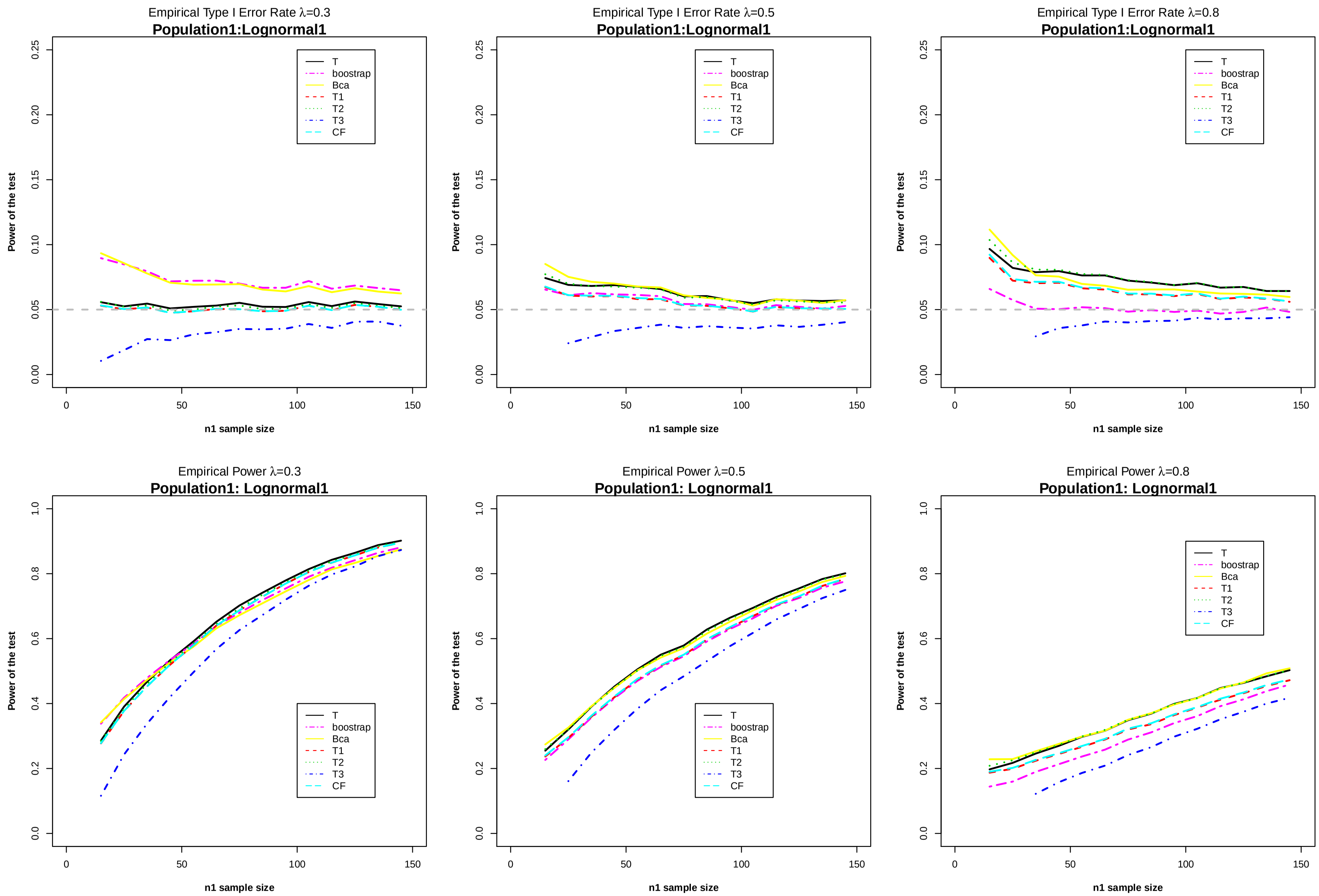

Figure 4.23: *One-sided lower-tailed Empirical Power and Type I Error rate for Pair5*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.045	0.036	0.208	0.035	0.237	0.046	0.273	0.046	0.241	0.044	0.232	0.077	0.335	0.046	0.241
25	0.042	0.039	0.316	0.037	0.355	0.044	0.387	0.046	0.348	0.044	0.339	0.070	0.429	0.047	0.348
35	0.037	0.037	0.416	0.036	0.459	0.043	0.486	0.044	0.444	0.041	0.435	0.066	0.522	0.044	0.444
45	0.036	0.042	0.506	0.041	0.551	0.048	0.575	0.049	0.530	0.047	0.522	0.069	0.599	0.049	0.531
55	0.034	0.043	0.586	0.041	0.629	0.046	0.652	0.049	0.609	0.047	0.601	0.065	0.669	0.049	0.609
65	0.032	0.045	0.651	0.044	0.692	0.049	0.712	0.051	0.671	0.048	0.664	0.066	0.720	0.051	0.671
75	0.031	0.042	0.710	0.042	0.748	0.043	0.761	0.047	0.725	0.046	0.721	0.063	0.769	0.047	0.725
85	0.030	0.046	0.759	0.044	0.792	0.050	0.805	0.052	0.774	0.050	0.768	0.067	0.810	0.052	0.774
95	0.028	0.045	0.805	0.044	0.834	0.048	0.843	0.051	0.815	0.048	0.811	0.064	0.843	0.051	0.815
105	0.028	0.043	0.841	0.042	0.866	0.045	0.874	0.048	0.849	0.046	0.845	0.059	0.873	0.048	0.849
115	0.027	0.042	0.873	0.039	0.897	0.043	0.902	0.047	0.881	0.045	0.877	0.057	0.900	0.047	0.881
125	0.026	0.043	0.892	0.040	0.914	0.044	0.916	0.048	0.900	0.046	0.897	0.061	0.916	0.048	0.900
135	0.025	0.044	0.912	0.043	0.930	0.046	0.933	0.049	0.918	0.047	0.916	0.060	0.933	0.049	0.918
145	0.025	0.046	0.928	0.044	0.946	0.047	0.948	0.049	0.934	0.048	0.932	0.061	0.947	0.049	0.934

Table 4.45: *Proportion of rejections for one-sided upper-tailed tests for Pair5 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.045	0.056	0.286	0.090	0.338	0.093	0.342	0.053	0.277	0.056	0.285	0.010	0.116	0.053	0.278
25	0.042	0.052	0.391	0.085	0.418	0.086	0.414	0.050	0.377	0.052	0.384	0.019	0.243	0.050	0.377
35	0.037	0.055	0.467	0.080	0.479	0.078	0.473	0.052	0.453	0.053	0.461	0.027	0.338	0.052	0.453
45	0.036	0.051	0.534	0.072	0.535	0.071	0.527	0.048	0.520	0.049	0.527	0.026	0.421	0.048	0.520
55	0.034	0.052	0.591	0.072	0.583	0.069	0.575	0.049	0.579	0.050	0.585	0.031	0.496	0.049	0.579
65	0.032	0.053	0.652	0.072	0.639	0.069	0.632	0.050	0.640	0.052	0.645	0.033	0.568	0.051	0.640
75	0.031	0.055	0.702	0.070	0.680	0.070	0.672	0.050	0.687	0.053	0.693	0.035	0.626	0.050	0.687
85	0.030	0.052	0.742	0.067	0.719	0.065	0.709	0.049	0.730	0.050	0.735	0.035	0.673	0.049	0.730
95	0.028	0.052	0.779	0.067	0.754	0.064	0.745	0.049	0.769	0.051	0.774	0.035	0.720	0.049	0.769
105	0.028	0.056	0.814	0.072	0.790	0.068	0.780	0.053	0.805	0.054	0.810	0.039	0.763	0.053	0.805
115	0.027	0.053	0.843	0.066	0.818	0.063	0.813	0.050	0.834	0.051	0.838	0.036	0.797	0.050	0.834
125	0.026	0.056	0.864	0.068	0.842	0.066	0.833	0.054	0.856	0.055	0.859	0.041	0.822	0.054	0.856
135	0.025	0.054	0.888	0.067	0.864	0.064	0.855	0.052	0.880	0.053	0.885	0.041	0.854	0.052	0.880
145	0.025	0.052	0.902	0.065	0.881	0.062	0.875	0.050	0.896	0.051	0.898	0.038	0.873	0.050	0.896

Table 4.46: *Proportion of rejections for one-sided lower-tailed tests for Pair5 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.186	0.034	0.176	0.049	0.219	0.054	0.235	0.052	0.234	0.042	0.203	0.085	0.317	0.053	0.237
25	0.173	0.033	0.251	0.047	0.300	0.048	0.302	0.047	0.309	0.038	0.276	0.072	0.385	0.048	0.312
35	0.161	0.032	0.325	0.045	0.373	0.044	0.372	0.044	0.381	0.037	0.348	0.067	0.447	0.045	0.383
45	0.151	0.036	0.387	0.050	0.430	0.048	0.425	0.048	0.438	0.040	0.405	0.068	0.498	0.049	0.440
55	0.142	0.037	0.452	0.050	0.495	0.047	0.489	0.049	0.503	0.041	0.470	0.063	0.555	0.049	0.505
65	0.135	0.037	0.517	0.052	0.554	0.048	0.548	0.050	0.559	0.041	0.532	0.064	0.611	0.051	0.561
75	0.130	0.042	0.574	0.054	0.614	0.049	0.605	0.052	0.618	0.045	0.590	0.064	0.664	0.052	0.620
85	0.125	0.037	0.631	0.048	0.666	0.045	0.655	0.048	0.671	0.040	0.643	0.060	0.709	0.048	0.672
95	0.120	0.042	0.677	0.053	0.707	0.049	0.697	0.052	0.712	0.044	0.687	0.064	0.746	0.052	0.712
105	0.116	0.041	0.715	0.054	0.742	0.049	0.734	0.051	0.746	0.044	0.724	0.065	0.780	0.052	0.748
115	0.113	0.043	0.755	0.054	0.774	0.051	0.771	0.054	0.781	0.045	0.763	0.064	0.806	0.054	0.782
125	0.109	0.040	0.783	0.050	0.803	0.047	0.799	0.049	0.809	0.042	0.790	0.059	0.833	0.049	0.809
135	0.106	0.042	0.812	0.054	0.830	0.049	0.823	0.052	0.834	0.045	0.818	0.062	0.853	0.053	0.834
145	0.104	0.042	0.834	0.052	0.852	0.048	0.843	0.051	0.855	0.044	0.841	0.061	0.873	0.051	0.856

Table 4.47: *Proportion of rejections for one-sided upper-tailed tests for Pair5 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.186	0.074	0.255	0.065	0.226	0.085	0.274	0.067	0.237	0.077	0.260	0.010	0.063	0.068	0.240
25	0.173	0.069	0.320	0.061	0.290	0.075	0.326	0.061	0.294	0.070	0.320	0.024	0.161	0.061	0.296
35	0.161	0.068	0.391	0.063	0.357	0.071	0.392	0.060	0.359	0.068	0.389	0.029	0.249	0.061	0.362
45	0.151	0.069	0.453	0.062	0.417	0.070	0.449	0.060	0.420	0.068	0.450	0.034	0.320	0.061	0.422
55	0.142	0.068	0.506	0.061	0.470	0.068	0.502	0.058	0.474	0.067	0.502	0.036	0.385	0.059	0.476
65	0.135	0.066	0.550	0.060	0.512	0.067	0.542	0.058	0.516	0.066	0.547	0.038	0.441	0.058	0.518
75	0.130	0.060	0.579	0.054	0.546	0.061	0.571	0.053	0.549	0.059	0.575	0.036	0.484	0.053	0.550
85	0.125	0.060	0.628	0.054	0.591	0.059	0.615	0.053	0.597	0.060	0.624	0.037	0.531	0.053	0.598
95	0.120	0.057	0.664	0.052	0.629	0.058	0.652	0.051	0.634	0.057	0.660	0.036	0.577	0.051	0.635
105	0.116	0.055	0.695	0.050	0.663	0.053	0.687	0.048	0.670	0.054	0.692	0.035	0.618	0.049	0.671
115	0.113	0.058	0.728	0.053	0.701	0.058	0.719	0.052	0.703	0.057	0.725	0.038	0.659	0.052	0.705
125	0.109	0.057	0.755	0.052	0.725	0.057	0.746	0.051	0.729	0.056	0.752	0.037	0.692	0.051	0.731
135	0.106	0.057	0.783	0.051	0.757	0.055	0.774	0.051	0.762	0.056	0.782	0.038	0.725	0.051	0.763
145	0.104	0.057	0.801	0.053	0.776	0.057	0.793	0.050	0.784	0.055	0.798	0.040	0.750	0.050	0.784

Table 4.48: *Proportion of rejections for one-sided lower-tailed tests for Pair5 when* $\lambda = 0.5$

		Tests													
n_1	skewness	T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.524	0.072	0.182	0.093	0.192	0.108	0.228	0.116	0.263	0.091	0.218	0.144	0.309	0.122	0.277
25	0.489	0.068	0.203	0.084	0.214	0.087	0.231	0.105	0.283	0.080	0.231	0.123	0.316	0.110	0.295
35	0.449	0.055	0.225	0.074	0.248	0.070	0.249	0.087	0.307	0.064	0.250	0.101	0.336	0.092	0.318
45	0.429	0.052	0.242	0.072	0.272	0.066	0.263	0.078	0.330	0.058	0.263	0.091	0.356	0.083	0.341
55	0.404	0.048	0.274	0.068	0.300	0.060	0.289	0.073	0.363	0.054	0.295	0.081	0.386	0.076	0.373
65	0.389	0.046	0.296	0.066	0.326	0.058	0.311	0.071	0.385	0.050	0.315	0.079	0.403	0.074	0.394
75	0.372	0.043	0.321	0.065	0.355	0.054	0.335	0.066	0.411	0.047	0.341	0.072	0.425	0.069	0.420
85	0.359	0.043	0.346	0.062	0.380	0.052	0.360	0.063	0.433	0.046	0.363	0.068	0.446	0.066	0.444
95	0.346	0.042	0.378	0.060	0.412	0.052	0.388	0.061	0.466	0.046	0.394	0.067	0.476	0.065	0.476
105	0.336	0.042	0.401	0.061	0.432	0.052	0.412	0.063	0.484	0.045	0.416	0.067	0.492	0.067	0.494
115	0.324	0.042	0.428	0.062	0.459	0.053	0.437	0.062	0.511	0.046	0.440	0.064	0.520	0.064	0.520
125	0.317	0.040	0.448	0.059	0.476	0.050	0.452	0.060	0.533	0.043	0.462	0.063	0.540	0.062	0.541
135	0.308	0.039	0.475	0.058	0.506	0.050	0.477	0.057	0.559	0.042	0.490	0.061	0.565	0.059	0.567
145	0.302	0.039	0.497	0.059	0.524	0.050	0.498	0.057	0.579	0.042	0.509	0.060	0.582	0.060	0.586

Table 4.49: *Proportion of rejections for one-sided upper-tailed tests for Pair5 when* $\lambda = 0.8$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.254	0.097	0.197	0.066	0.144	0.112	0.228	0.090	0.187	0.104	0.208			0.092	0.190
25	0.240	0.082	0.217	0.057	0.160	0.092	0.229	0.072	0.199	0.086	0.225	0.019	0.066	0.073	0.202
35	0.219	0.079	0.246	0.051	0.190	0.076	0.253	0.070	0.224	0.081	0.250	0.029	0.122	0.071	0.226
45	0.208	0.080	0.270	0.050	0.213	0.075	0.276	0.071	0.245	0.081	0.273	0.036	0.158	0.071	0.247
55	0.197	0.076	0.297	0.052	0.237	0.070	0.298	0.066	0.267	0.077	0.299	0.038	0.186	0.067	0.269
65	0.188	0.076	0.316	0.051	0.258	0.068	0.316	0.065	0.289	0.077	0.319	0.041	0.209	0.066	0.291
75	0.180	0.072	0.349	0.048	0.289	0.065	0.350	0.062	0.320	0.073	0.350	0.040	0.241	0.062	0.323
85	0.173	0.071	0.369	0.050	0.311	0.066	0.370	0.062	0.336	0.071	0.369	0.041	0.266	0.062	0.339
95	0.166	0.069	0.399	0.048	0.340	0.066	0.397	0.060	0.364	0.069	0.400	0.041	0.298	0.061	0.367
105	0.162	0.070	0.417	0.049	0.361	0.064	0.417	0.062	0.388	0.071	0.416	0.044	0.322	0.062	0.389
115	0.156	0.067	0.448	0.047	0.392	0.062	0.446	0.058	0.412	0.067	0.448	0.042	0.351	0.058	0.415
125	0.152	0.067	0.464	0.048	0.413	0.062	0.465	0.059	0.432	0.067	0.464	0.043	0.375	0.060	0.434
135	0.148	0.064	0.484	0.052	0.438	0.061	0.493	0.058	0.454	0.064	0.485	0.043	0.400	0.058	0.457
145	0.144	0.064	0.503	0.048	0.459	0.060	0.508	0.056	0.472	0.064	0.503	0.044	0.418	0.056	0.474

Table 4.50: *Proportion of rejections for one-sided lower-tailed tests for Pair5 when* $\lambda = 0.8$

The simulation setting of $Pair6$ is similar to $Pair4$, which has a big skewness and common variance. Although the two settings have difference skewed populations, one from Gamma and one from Log-normal, their simulation results are quite similar. The simulation results of $Pair6$ in Figure 4.24 and Figure 4.25 are similar to the simulation results of $Pair4$. The numerical results of the empirical type I error rate and power are given in Table 4.51, 4.53, 4.55 and Table 4.52, 4.54, 4.56.

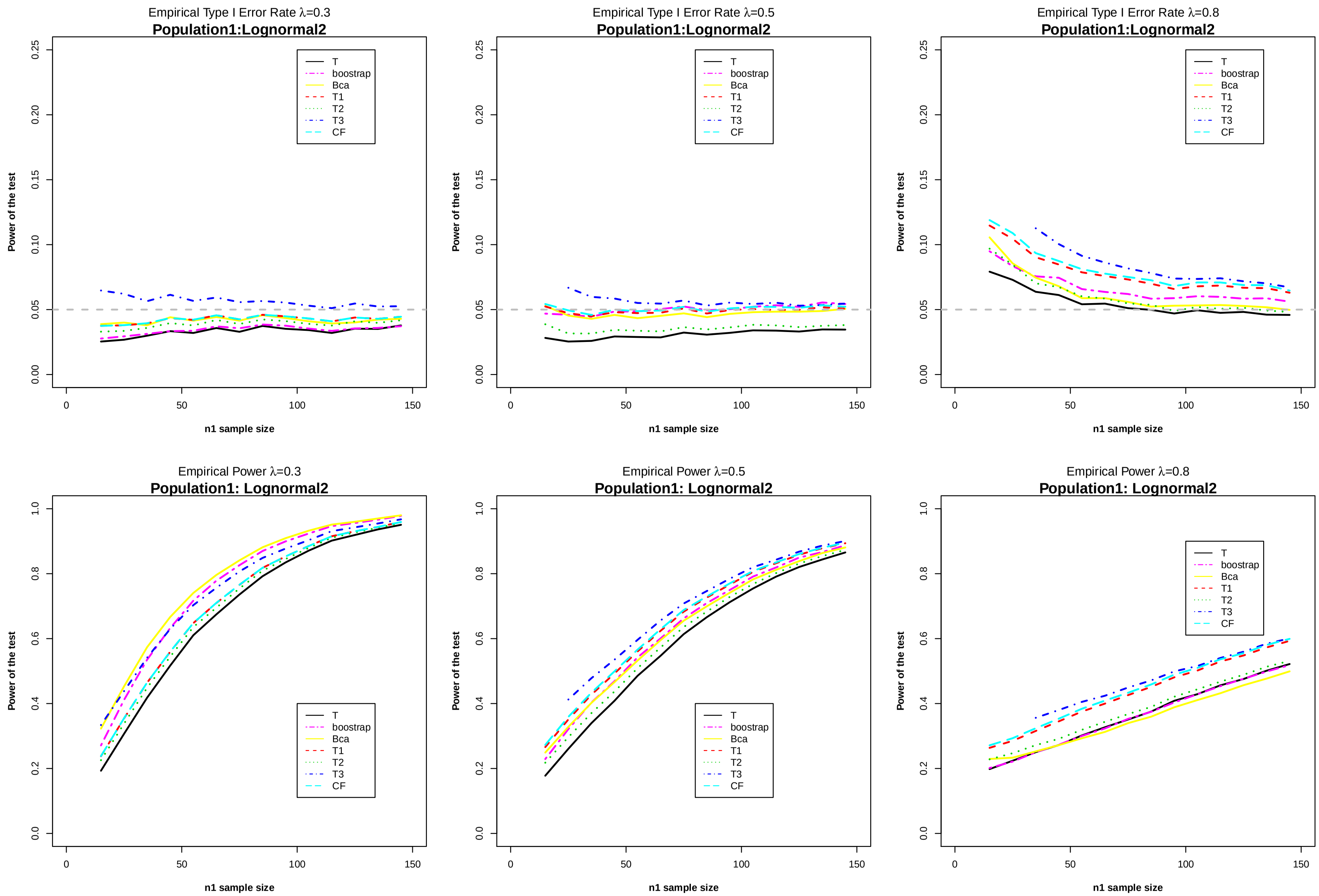

Figure 4.24: *One-sided upper-tailed Empirical Power and Type I Error rate for Pair6*

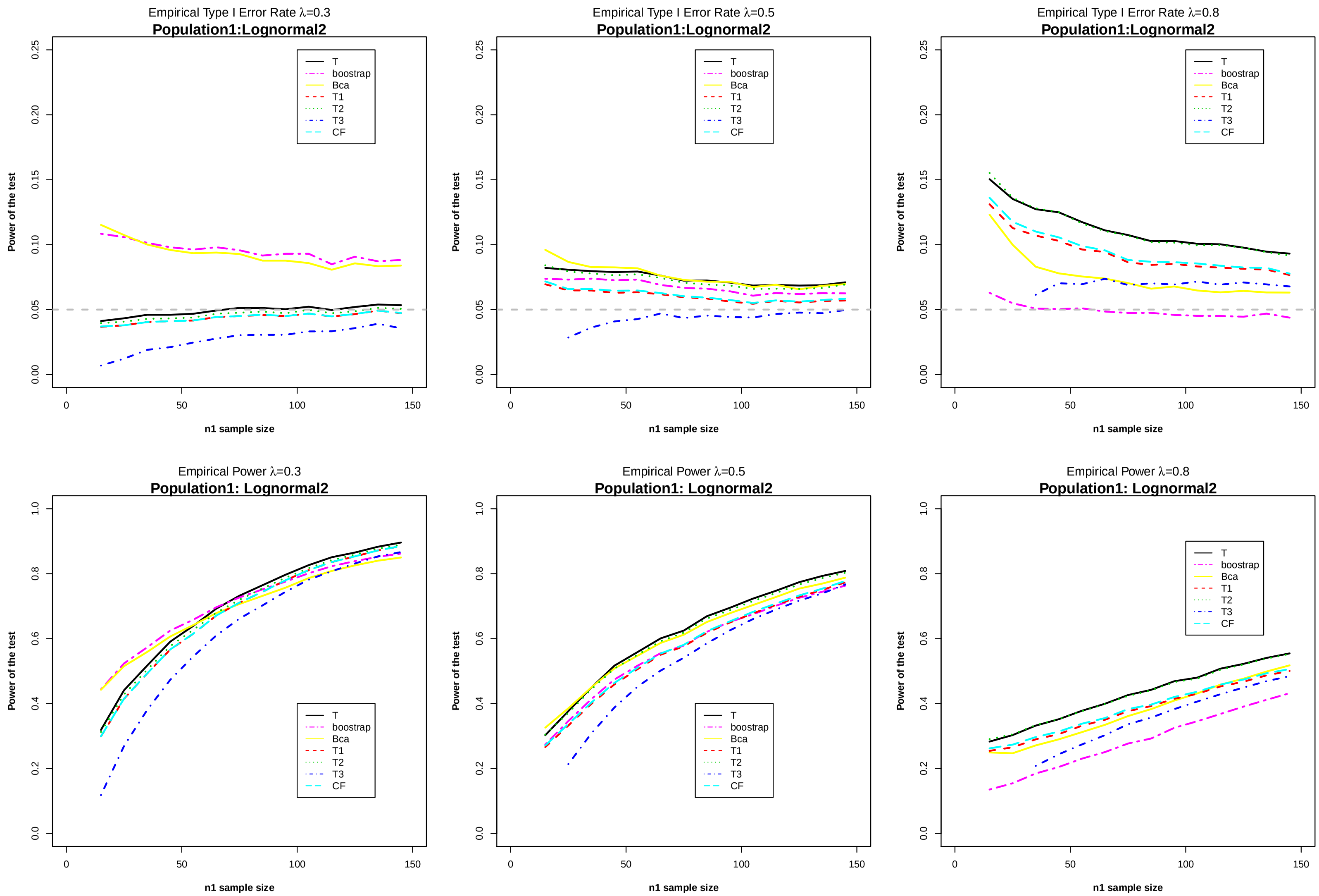

Figure 4.25: *One-sided lower-tailed Empirical Power and Type I Error rate for Pair6*

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.067	0.025	0.193	0.028	0.271	0.039	0.323	0.037	0.238	0.033	0.226	0.065	0.334	0.037	0.238
25	0.064	0.027	0.307	0.029	0.411	0.040	0.453	0.038	0.354	0.034	0.338	0.062	0.438	0.038	0.355
35	0.059	0.030	0.418	0.031	0.535	0.038	0.574	0.039	0.464	0.036	0.449	0.056	0.541	0.040	0.465
45	0.058	0.033	0.517	0.033	0.633	0.044	0.667	0.043	0.559	0.040	0.544	0.061	0.630	0.043	0.560
55	0.057	0.032	0.610	0.034	0.717	0.042	0.741	0.042	0.648	0.038	0.635	0.057	0.703	0.042	0.648
65	0.054	0.036	0.675	0.037	0.778	0.044	0.796	0.046	0.710	0.042	0.695	0.059	0.757	0.046	0.710
75	0.053	0.033	0.736	0.036	0.826	0.042	0.841	0.042	0.766	0.039	0.754	0.056	0.807	0.042	0.766
85	0.052	0.037	0.792	0.038	0.870	0.046	0.881	0.046	0.818	0.042	0.809	0.057	0.849	0.046	0.818
95	0.050	0.035	0.835	0.038	0.899	0.043	0.909	0.045	0.853	0.041	0.845	0.055	0.877	0.045	0.853
105	0.049	0.034	0.871	0.035	0.924	0.041	0.932	0.043	0.885	0.039	0.879	0.053	0.903	0.043	0.885
115	0.049	0.032	0.902	0.034	0.947	0.039	0.951	0.041	0.915	0.038	0.911	0.051	0.931	0.041	0.915
125	0.047	0.035	0.919	0.035	0.956	0.040	0.960	0.044	0.930	0.041	0.927	0.055	0.943	0.044	0.930
135	0.047	0.035	0.936	0.036	0.966	0.042	0.970	0.043	0.944	0.040	0.942	0.052	0.955	0.043	0.944
145	0.046	0.038	0.951	0.037	0.978	0.043	0.980	0.044	0.959	0.042	0.955	0.053	0.967	0.044	0.959

Table 4.51: *Proportion of rejections for one-sided upper-tailed tests for Pair6 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.067	0.041	0.319	0.108	0.444	0.115	0.442	0.037	0.299	0.040	0.313	0.007	0.118	0.037	0.299
25	0.064	0.043	0.441	0.106	0.524	0.107	0.515	0.038	0.415	0.041	0.428	0.012	0.269	0.038	0.415
35	0.059	0.046	0.517	0.101	0.573	0.100	0.559	0.040	0.493	0.043	0.504	0.019	0.380	0.040	0.493
45	0.058	0.046	0.591	0.098	0.625	0.096	0.607	0.041	0.566	0.043	0.577	0.021	0.473	0.041	0.566
55	0.057	0.047	0.639	0.096	0.658	0.093	0.642	0.042	0.615	0.044	0.627	0.025	0.545	0.042	0.615
65	0.054	0.049	0.692	0.098	0.697	0.094	0.679	0.044	0.671	0.047	0.680	0.028	0.610	0.044	0.671
75	0.053	0.051	0.732	0.096	0.727	0.093	0.707	0.045	0.711	0.048	0.720	0.030	0.662	0.045	0.711
85	0.052	0.051	0.764	0.092	0.752	0.088	0.732	0.046	0.744	0.048	0.754	0.031	0.702	0.046	0.744
95	0.050	0.050	0.797	0.093	0.776	0.088	0.757	0.045	0.780	0.047	0.788	0.031	0.744	0.045	0.780
105	0.049	0.052	0.826	0.093	0.801	0.086	0.786	0.047	0.811	0.050	0.817	0.033	0.782	0.047	0.811
115	0.049	0.050	0.851	0.085	0.823	0.081	0.808	0.045	0.836	0.047	0.842	0.033	0.808	0.045	0.836
125	0.047	0.052	0.865	0.091	0.839	0.086	0.826	0.046	0.853	0.049	0.859	0.036	0.831	0.047	0.853
135	0.047	0.054	0.883	0.087	0.852	0.083	0.840	0.049	0.872	0.051	0.877	0.039	0.853	0.049	0.872
145	0.046	0.053	0.896	0.088	0.862	0.084	0.850	0.047	0.885	0.051	0.889	0.036	0.867	0.047	0.885

Table 4.52: *Proportion of rejections for one-sided lower-tailed tests for Pair6 when* $\lambda = 0.3$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.277	0.028	0.178	0.047	0.229	0.053	0.249	0.052	0.266	0.039	0.218	0.080	0.337	0.054	0.273
25	0.266	0.025	0.261	0.046	0.321	0.045	0.328	0.047	0.351	0.032	0.296	0.067	0.412	0.049	0.359
35	0.254	0.026	0.340	0.044	0.403	0.043	0.402	0.045	0.427	0.032	0.371	0.060	0.478	0.046	0.433
45	0.244	0.029	0.409	0.048	0.469	0.046	0.466	0.048	0.493	0.034	0.437	0.058	0.535	0.050	0.499
55	0.235	0.029	0.485	0.049	0.542	0.043	0.532	0.047	0.560	0.034	0.509	0.055	0.596	0.049	0.566
65	0.229	0.029	0.547	0.050	0.601	0.045	0.596	0.048	0.624	0.033	0.573	0.055	0.656	0.049	0.629
75	0.223	0.032	0.614	0.052	0.662	0.047	0.655	0.051	0.683	0.036	0.636	0.057	0.708	0.052	0.688
85	0.218	0.031	0.666	0.050	0.710	0.044	0.700	0.047	0.727	0.035	0.683	0.053	0.747	0.048	0.730
95	0.213	0.032	0.713	0.050	0.750	0.047	0.742	0.050	0.767	0.036	0.729	0.055	0.785	0.051	0.770
105	0.209	0.034	0.754	0.052	0.791	0.048	0.782	0.050	0.805	0.038	0.768	0.054	0.820	0.052	0.808
115	0.205	0.034	0.791	0.053	0.819	0.048	0.812	0.050	0.832	0.038	0.802	0.055	0.843	0.052	0.835
125	0.202	0.033	0.821	0.052	0.850	0.048	0.839	0.050	0.859	0.036	0.831	0.053	0.868	0.051	0.861
135	0.198	0.035	0.844	0.056	0.867	0.049	0.863	0.052	0.878	0.037	0.854	0.053	0.886	0.053	0.880
145	0.195	0.035	0.865	0.054	0.887	0.051	0.881	0.051	0.894	0.038	0.873	0.055	0.901	0.052	0.895

Table 4.53: *Proportion of rejections for one-sided upper-tailed tests for Pair6 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.277	0.082	0.302	0.074	0.274	0.096	0.326	0.070	0.266	0.084	0.304	0.012	0.089	0.072	0.271
25	0.266	0.081	0.377	0.073	0.346	0.087	0.385	0.065	0.332	0.079	0.371	0.028	0.214	0.066	0.337
35	0.254	0.080	0.450	0.074	0.414	0.083	0.450	0.065	0.399	0.078	0.444	0.036	0.308	0.066	0.403
45	0.244	0.079	0.517	0.073	0.474	0.082	0.508	0.063	0.459	0.076	0.508	0.041	0.388	0.064	0.464
55	0.235	0.079	0.559	0.073	0.517	0.082	0.547	0.063	0.506	0.077	0.550	0.043	0.452	0.065	0.510
65	0.229	0.076	0.601	0.069	0.555	0.076	0.587	0.062	0.550	0.074	0.592	0.047	0.502	0.063	0.554
75	0.223	0.072	0.625	0.067	0.579	0.073	0.612	0.059	0.576	0.071	0.618	0.043	0.541	0.060	0.581
85	0.218	0.072	0.669	0.066	0.621	0.071	0.651	0.058	0.618	0.069	0.662	0.045	0.585	0.059	0.623
95	0.213	0.071	0.695	0.064	0.651	0.071	0.678	0.056	0.650	0.068	0.688	0.044	0.625	0.057	0.653
105	0.209	0.068	0.723	0.061	0.675	0.067	0.703	0.054	0.678	0.066	0.716	0.044	0.660	0.055	0.682
115	0.205	0.069	0.748	0.063	0.701	0.069	0.728	0.057	0.704	0.066	0.741	0.047	0.690	0.057	0.708
125	0.202	0.068	0.774	0.062	0.726	0.066	0.754	0.056	0.729	0.066	0.767	0.048	0.717	0.056	0.733
135	0.198	0.069	0.793	0.063	0.744	0.068	0.769	0.057	0.750	0.067	0.786	0.047	0.739	0.057	0.754
145	0.195	0.071	0.808	0.062	0.763	0.070	0.788	0.057	0.773	0.069	0.803	0.050	0.767	0.058	0.777

Table 4.54: *Proportion of rejections for one-sided lower-tailed tests for Pair6 when* $\lambda = 0.5$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.378	0.079	0.198	0.095	0.202	0.106	0.229	0.115	0.264	0.097	0.228	0.151	0.321	0.119	0.271
25	0.370	0.073	0.224	0.083	0.221	0.086	0.233	0.104	0.285	0.085	0.247	0.133	0.332	0.109	0.293
35	0.348	0.064	0.250	0.076	0.250	0.075	0.253	0.090	0.316	0.070	0.272	0.113	0.356	0.093	0.324
45	0.338	0.061	0.272	0.074	0.273	0.068	0.272	0.085	0.345	0.067	0.291	0.100	0.380	0.087	0.353
55	0.328	0.054	0.302	0.066	0.300	0.059	0.293	0.079	0.375	0.060	0.319	0.091	0.405	0.081	0.384
65	0.319	0.055	0.327	0.064	0.324	0.059	0.313	0.076	0.400	0.058	0.344	0.086	0.424	0.078	0.409
75	0.311	0.051	0.351	0.062	0.353	0.056	0.340	0.073	0.426	0.055	0.367	0.082	0.449	0.075	0.433
85	0.304	0.050	0.375	0.058	0.375	0.052	0.360	0.070	0.451	0.054	0.390	0.078	0.471	0.073	0.459
95	0.296	0.047	0.408	0.059	0.403	0.053	0.389	0.066	0.481	0.049	0.421	0.074	0.499	0.068	0.489
105	0.291	0.050	0.429	0.060	0.429	0.053	0.411	0.068	0.501	0.052	0.444	0.074	0.516	0.071	0.510
115	0.285	0.048	0.456	0.060	0.454	0.053	0.432	0.068	0.530	0.051	0.467	0.074	0.540	0.071	0.536
125	0.281	0.048	0.476	0.058	0.476	0.053	0.457	0.067	0.548	0.051	0.489	0.072	0.560	0.069	0.555
135	0.275	0.046	0.501	0.059	0.499	0.052	0.477	0.066	0.573	0.049	0.513	0.070	0.584	0.069	0.581
145	0.272	0.046	0.522	0.056	0.518	0.050	0.499	0.063	0.593	0.048	0.531	0.067	0.601	0.065	0.599

Table 4.55: *Proportion of rejections for one-sided upper-tailed tests for Pair6 when* $\lambda = 0.8$

n_1	skewness	Tests													
		T-Test		Bootstrap-t		BCa		T_1		T_2		T_3		CF	
		H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a	H_0	H_a
15	0.378	0.150	0.283	0.063	0.135	0.123	0.249	0.131	0.254	0.155	0.290			0.136	0.262
25	0.370	0.135	0.303	0.055	0.154	0.100	0.247	0.113	0.265	0.136	0.305	0.051	0.149	0.118	0.273
35	0.348	0.127	0.332	0.051	0.185	0.083	0.271	0.107	0.290	0.128	0.333	0.061	0.208	0.110	0.297
45	0.338	0.125	0.352	0.050	0.205	0.078	0.290	0.103	0.306	0.125	0.352	0.070	0.244	0.106	0.314
55	0.328	0.117	0.378	0.051	0.230	0.075	0.312	0.096	0.332	0.116	0.376	0.070	0.274	0.099	0.338
65	0.319	0.111	0.399	0.048	0.251	0.074	0.335	0.094	0.350	0.110	0.399	0.074	0.303	0.096	0.355
75	0.311	0.107	0.426	0.047	0.277	0.070	0.361	0.086	0.376	0.107	0.423	0.069	0.336	0.088	0.383
85	0.304	0.103	0.442	0.048	0.292	0.066	0.383	0.084	0.392	0.102	0.441	0.070	0.357	0.087	0.397
95	0.296	0.103	0.469	0.046	0.325	0.068	0.409	0.085	0.415	0.102	0.466	0.069	0.383	0.086	0.420
105	0.291	0.101	0.480	0.045	0.345	0.065	0.431	0.083	0.430	0.100	0.477	0.072	0.406	0.086	0.436
115	0.285	0.100	0.507	0.045	0.368	0.063	0.458	0.082	0.453	0.100	0.504	0.069	0.429	0.084	0.458
125	0.281	0.098	0.522	0.044	0.391	0.064	0.477	0.081	0.467	0.098	0.520	0.071	0.449	0.082	0.474
135	0.275	0.095	0.541	0.047	0.411	0.063	0.499	0.081	0.487	0.094	0.538	0.069	0.469	0.082	0.493
145	0.272	0.093	0.554	0.044	0.432	0.063	0.518	0.077	0.500	0.092	0.553	0.068	0.484	0.078	0.506

Table 4.56: *Proportion of rejections for one-sided lower-tailed tests for Pair6 when* $\lambda = 0.8$

4.4.6 Discussion

From the simulation study, we can conclude several findings as follows:

1. Across all levels of λ, as the sample size gets larger, the power of all the seven tests increases and the type I error rate of the seven tests approaches the significant level. When the sample size n_1 remains constant, the type I error rate of the seven tests increases as the relative skewness $\gamma = 3N^{-1/2}B$ gets bigger.

2. The theoretical type I error rate of TCF test and T_i test is smaller than α under each simulation settings. However, the empirical type I error rate of the four new tests is close to α when the relative skewness $\gamma = 3N^{-1/2}B$ is smaller. When the relative skewness is big, the empirical type I error rate of the four new tests is bigger than 0.1 even with a big sample size.

3. Among all the seven tests, the two-sample TCF test and T_3 test not only give consistently bigger power but also control the type I error rate well. As the population relative skewness $\gamma = 3N^{-1/2}B$ increases, the power of these two tests outperform the power of Bootstrap-t test and BCa test.

4. The two-sample Bootstrap-t test and BCa test give better power when the two-sample data are balanced and less skewed; The pooled two-sample t-test gives the smallest power among the seven tests across all the simulation settings. In addition, the pooled two-sample t-test also gives the highest type I error rate when the data set is small and highly skewed.

5. This simulation study can not reflect the effect of the common variance σ^2 on the power of the test. That's because the term of σ^2 was canceled out during the calculation of two sample relative skewness $\gamma = 3N^{-1/2}B$ and $\delta/\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}$, in that δ/σ is considered a constant in our simulation setting.

6. In this simulation study, we investigated two positive skewed families, Gamma distribution and Log-normal distribution. We found that the power of the test was not

significantly affected by the types of the skewed distribution. In both lognormal and gamma families, the amount of their population skewness γ_i affect the performance of the tests.

4.5 Two sample comparisons with population skewness from outliers

In our previous simulation study, the population skewness comes from the population distribution being skewed, i.e., from Gamma distribution or log-normal distribution. In this section, we will study the case that the population skewness is only due to some outliers. In other words, the skewed population becomes symmetric if we delete all its outliers. Thus the main purpose of this section is to check if the TCF test and the three T_i tests have enough power to detect the population mean difference under this situation.

To investigate the above questions, a simulation was conducted to compare the type I error rate and power of the eight tests: pooled two-sample t-test, Wilcoxon Rank-Sum Test, Bootstrap t-test, BCa test, three T_i tests and TCF test. Under H_0, the data were generated from standard normal distribution without outliers. Under H_a, outliers were added to population 1. The outliers were randomly generated integers from 10 to 20, which accounts for 5% of the population 1 original sample size.

	Population2	Population1
Pair1	$N(\mu_1 = 0, \sigma_1^2 = 1)$	$N(\mu_1 = 0, \sigma_1^2 = 1)+$ Outliers
Pair2	$N(\mu_1 = 0, \sigma_1^2 = 1)$	$N(\mu_1 = 0.3, \sigma_1^2 = 1)+$ Outliers

Table 4.57: *Two Pairs of Population Settings*

Since all the outliers are positive integers, the first population mean μ_1 is bigger than the second population mean μ_2. These outliers also make the first population right skewed with a positive population skewness γ_1. The second population is symmetric with $\gamma_2 = 0$. The other simulation settings are the same as in Section 3.3.2. $10,000$ samples were generated for each parameter setting and each sample size. For bootstrap tests, $1,000$ bootstrap samples are resampled from each generated data set.

- Significance level $\alpha = 0.05$;
- First population sample size $n_1 = 15, 25, \ldots, 145$;
- $\lambda_N = n_1/N = 0.6$, here $N = n_1 + n_2$;
- Second population sample size $n_2 = (1 - \lambda_N)N$;
- Effect size: $\delta = (\mu_1 - \mu_2) - Hypothesized(\mu_1 - \mu_2)$, depends totally on the outliers in pair1 setting and is equal to 0.3 plus the effects due to added outliers to popoulation1 for pair2 setting .

The simulation results are presented in Figure 4.26, Table 4.58 and Table 4.59.

Figure 4.26 shows the empirical type I error rate and power from these eight two-sample tests including pooled two-sample t-test denoted as "T", Wilcoxon Rank-Sum Test denoted as "$Wilcox$", three T_i tests denoted as "T_i", $i = 1, 2, 3$ and two sample TCF test denoted as "CF". The two-sample Bootstrap-t test and two-sample BCa test are constructed based on two different resampling methods. The first resampling method generates bootstrap samples from each group separately. In the second resampling method the bootstrap samples were generated from the pooled $X_{ij} - \bar{X}_{i.}$, all i, j. We denote the tests based on the first resampling method as "$boot1$" and "$BCa1$". And denote the tests based on the second resampling method as "$boot2$" and "$BCa2$". The top two panels in Figure 4.26 give the empirical type I error rate of the eight tests with $\lambda = 0.6$. The bottom two panels provide their corresponding empirical powers.

Table 4.58 and Table 4.59 present the numeric results of the empirical power under each pair of population settings in Figure 4.26. The $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{B}$ are the mean estimates of population 1 skewness γ_1, population 2 skewness γ_1 and B respectively. The "skewness" calculated as $\frac{3\hat{B}}{2\sqrt{N}}$ stands for the mean estimate of the relative skewness of two populations defined as $\frac{3B}{2\sqrt{N}}$.

From Figure 4.26, all the eight tests control the type I error rate well. For the empirical power of pair1, the Bootstrap-t test and BCa test based on the first resampling method give the best power, which are consistently higher than the power of the other tests. The TCF test, T_1 and T_3 tests provide the second best power. The power of Wilcoxon Rank-Sum Test is very small, which is almost 0. The empirical power results of pair2 keep the same pattern as the results from pair1, except that the power of the Wilcoxon Rank-Sum Test is higher than the pooled two-sample t-test and T_2 test as long as $n_1 < 65$. The empirical powers of the eight tests for pair2 increase faster than the powers of pair1, because the two population mean difference for pair2 is larger than pair1.

From Table 4.58 and Table 4.59, the mean estimate of $\hat{\gamma}_1$ is around 4 and $\hat{\gamma}_2$ is around 0, which means the first population is heavily skewed and the second population is symmetric. Moreover, the mean estimate of the relative skewness is between 0.23 and 0.65, which means the two pairs of population settings have a high relative skewness.

Under this simulation setting, the bootstrap-t test and BCa test based on the first resampling method have significantly better power than the other tests in testing the two population mean difference when the population skewness comes from the outliers. Comparing with the pooled two-sample t-test, our four new tests are more robust with the existence of outliers. The Wilcoxon Rank-Sum Test gives the worst power in the two simulation settings. Regarding the two resampling methods for the bootstrap-t test and BCa test, we found the empirical powers based on the first resampling method provides much higher power than the second resampling method in this data generating setting.

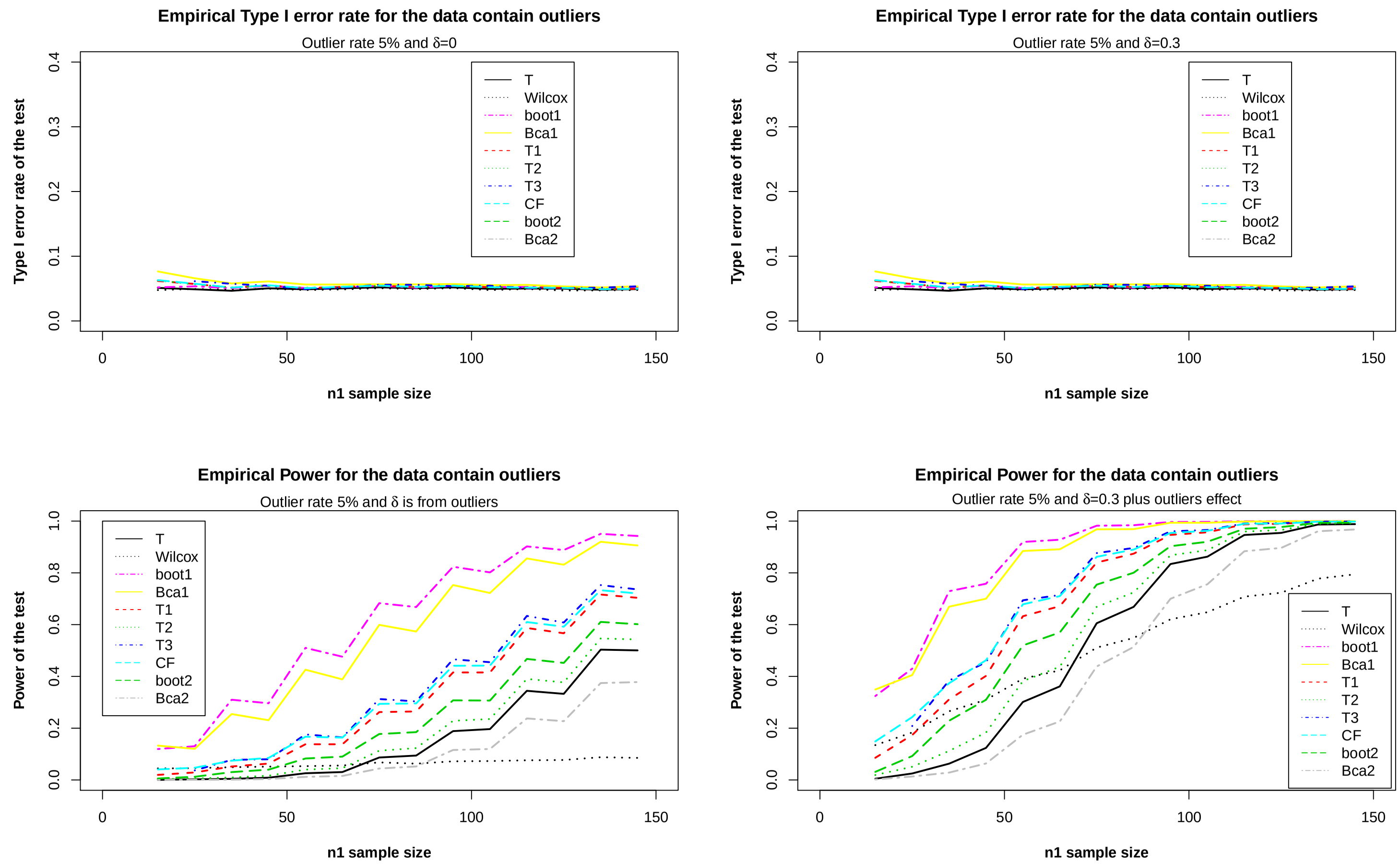

Figure 4.26: *Proportion of rejections of two-sided tests. In top 2 panels the data were generated under H_0. In bottom 2 panels, the data were generated under H_a. The left 2 panels correspond to $Pair1$ distribution setting. The right two panels correspond to $Pair2$ distribution setting.*

n_1	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{B}$	skewness	T-test	Wilcox	Boot1	BCa1	T1	T2	T3	CF	Boot2	BCa2
15	3.825	-0.000	0.729	0.656	0.000	0.044	0.119	0.133	0.020	0.003	0.000	0.041	0.006	0.000
25	4.408	-0.007	0.841	0.580	0.002	0.045	0.130	0.120	0.029	0.006	0.037	0.047	0.012	0.001
35	3.967	-0.006	0.757	0.449	0.004	0.049	0.310	0.255	0.052	0.009	0.078	0.074	0.030	0.001
45	4.318	-0.008	0.824	0.428	0.009	0.051	0.296	0.231	0.063	0.016	0.080	0.085	0.040	0.004
55	4.028	-0.010	0.769	0.360	0.026	0.053	0.510	0.426	0.138	0.040	0.176	0.167	0.083	0.012
65	4.268	-0.008	0.814	0.353	0.030	0.056	0.476	0.389	0.138	0.045	0.164	0.164	0.090	0.015
75	4.063	-0.008	0.775	0.312	0.087	0.068	0.683	0.599	0.262	0.113	0.313	0.294	0.177	0.044
85	4.247	-0.006	0.810	0.305	0.094	0.063	0.668	0.573	0.265	0.123	0.303	0.296	0.185	0.052
95	4.081	-0.004	0.778	0.279	0.189	0.072	0.824	0.753	0.415	0.229	0.466	0.441	0.307	0.116
105	4.228	-0.004	0.806	0.274	0.197	0.073	0.802	0.722	0.415	0.235	0.455	0.442	0.307	0.120
115	4.093	-0.004	0.780	0.253	0.344	0.076	0.902	0.856	0.587	0.390	0.634	0.610	0.468	0.238
125	4.218	-0.003	0.804	0.251	0.333	0.077	0.888	0.832	0.567	0.377	0.608	0.592	0.452	0.227
135	4.102	-0.004	0.782	0.235	0.503	0.088	0.951	0.920	0.716	0.546	0.753	0.733	0.610	0.374
145	4.209	-0.003	0.802	0.232	0.500	0.085	0.942	0.906	0.704	0.543	0.736	0.721	0.602	0.378

Table 4.58: *Proportion of rejections for Pair1 when $\lambda = 0.6$ and δ is from outliers*

n_1	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{B}$	skewness	T-test	Wilcox	Boot1	BCa1	T1	T2	T3	CF	Boot2	BCa2
15	3.804	-0.000	0.725	0.652	0.005	0.135	0.324	0.349	0.086	0.019	0.000	0.149	0.031	0.002
25	4.373	-0.007	0.834	0.576	0.025	0.183	0.430	0.405	0.173	0.051	0.208	0.243	0.092	0.013
35	3.952	-0.006	0.754	0.447	0.063	0.265	0.729	0.669	0.311	0.112	0.384	0.374	0.229	0.028
45	4.296	-0.008	0.819	0.426	0.124	0.310	0.758	0.700	0.402	0.183	0.455	0.462	0.310	0.064
55	4.014	-0.010	0.766	0.358	0.301	0.391	0.919	0.884	0.632	0.382	0.694	0.678	0.520	0.175
65	4.249	-0.008	0.811	0.352	0.361	0.423	0.928	0.891	0.671	0.437	0.715	0.710	0.570	0.226
75	4.050	-0.008	0.773	0.311	0.605	0.511	0.982	0.968	0.840	0.669	0.876	0.862	0.754	0.438
85	4.231	-0.006	0.807	0.304	0.668	0.548	0.984	0.969	0.874	0.724	0.896	0.890	0.800	0.513
95	4.068	-0.004	0.776	0.278	0.834	0.621	0.997	0.994	0.947	0.868	0.960	0.954	0.903	0.700
105	4.212	-0.004	0.803	0.273	0.863	0.648	0.997	0.994	0.957	0.888	0.966	0.963	0.920	0.756
115	4.080	-0.004	0.778	0.252	0.946	0.708	1.000	0.998	0.987	0.960	0.991	0.990	0.970	0.883
125	4.203	-0.003	0.801	0.250	0.954	0.723	1.000	0.999	0.990	0.966	0.992	0.991	0.977	0.896
135	4.089	-0.004	0.780	0.234	0.987	0.778	1.000	1.000	0.998	0.990	0.998	0.998	0.993	0.961
145	4.195	-0.003	0.800	0.231	0.988	0.794	1.000	1.000	0.998	0.992	0.998	0.998	0.994	0.968

Table 4.59: *Proportion of rejections for Pair2 when $\lambda = 0.6$ and $\delta = 0.3$ plus effect of outliers*

4.6 Real data analysis

In this section we report the result of applying our four new two-sample tests, pooled two-sample t-test and bootstrap t-test to analyze two real data sets. These two data sets are both from textbook Ott and Longnecker (2008).

The first data set is about bonus percentage of employees. A personnel officer took samples of 24 female and 36 male managers to see whether there was any difference in bonuses, expressed as a percentage of yearly salary. The data are listed here:

Gender	Bonus Percentage								
F	9.2	7.7	11.9	6.2	9.0	8.4	6.9	7.6	7.4
	8.0	9.9	6.7	8.4	9.3	9.1	8.7	9.2	9.1
	8.4	9.6	7.7	9.0	9.0	8.4			
M	10.4	8.9	11.7	12.0	8.7	9.4	9.8	9.0	9.2
	9.7	9.1	8.8	7.9	9.9	10.0	10.1	9.0	11.4
	8.7	9.6	9.2	9.7	8.9	9.2	9.4	9.7	8.9
	9.3	10.4	11.9	9.0	12.0	9.6	9.2	9.9	9.0

Table 4.60: *Bonus percentage data set. The data set is from exercise* 6.13 *of textbook Ott and Longnecker (2008)*

The data set is unbalanced with $\lambda_N = 0.4$ and total sample size 60. Standard practice is to first conduct an equal variance F-test to decide whether pooled or unpooled t-test should be used. Here the sample standard deviation of the female group is 1.188959 and that for the male group is 1.00385, which leads to non-rejection for the F-test. As a results, the two population variances can be treated as equal. The F-test however, is sensitive to the assumption of normality for the population distribution. In Figure 4.27, we can see that the sample median of Bonus Percentage from the male is higher than that from the female. And the sample from male does not follow normal distribution. The pooled two-sample t statistic value is -4.036748. The cut offs for the rejection regions and the conclusions of the tests are given in Table 4.61. We can see that all the six two-sample tests reject the null hypothesis and conclude that the two populations have significantly different means. In addition the Wilcoxon Rank-sum test also reject the null hypothesis with a p-value of 0.00015. Even though all tests reject H_0, they have quite different rejection regions except for TCF and T_1 tests. Comparing to the rejection region of the two-sample t-test, the cutoff values for all other tests have different amount of shift in both ends to reflect the correction on skewness.

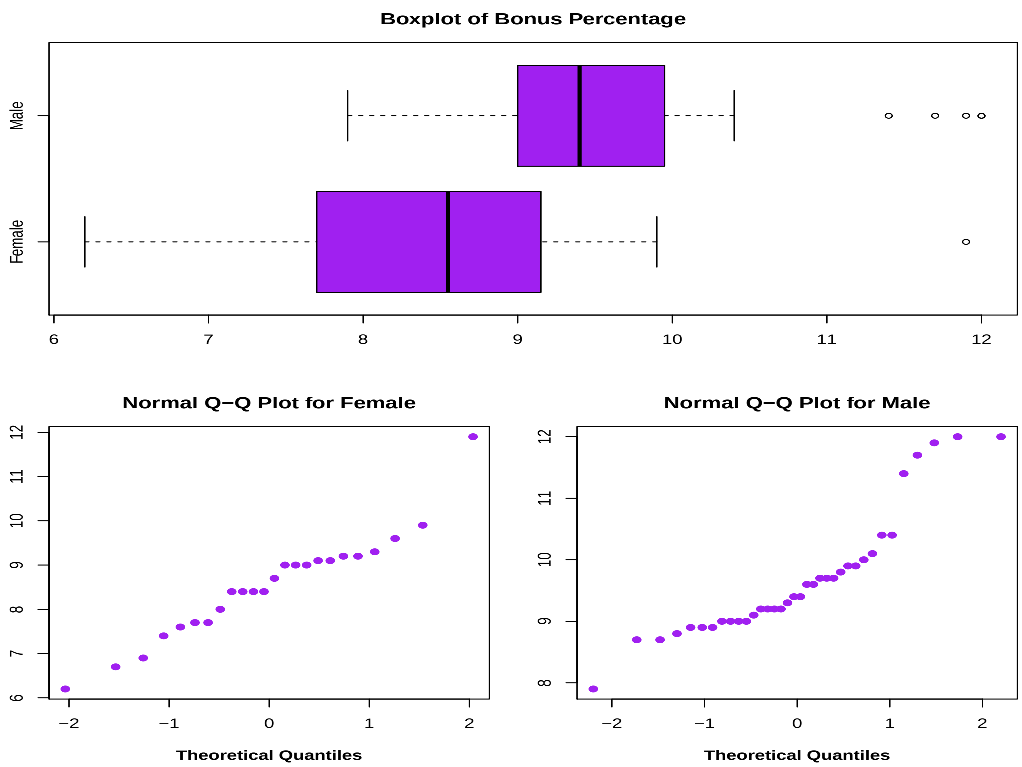

Figure 4.27: *Box-plot and Q-Q plot for bonus percentage data*

The second two-sample data set is from a cable TV company who was interested in making its operation more efficient by cutting down on the distance between service calls while still maintaining at least the same level of service quality. A treatment group of 18 repair-persons was assigned to a dispatcher who monitored all the incoming requests for cable repairs and then provided a service strategy for that day's work orders. A control group of 18 repair-persons was to perform their work in a normal fashion, by providing

	lower cutoff	upper cutoff	Conclusion at level 0.05
T.test	-2.002	2.002	Reject H0
bootstrap	-2.081	2.037	Reject H0
TCF	-1.881	2.039	Reject H0
T1	-1.884	2.043	Reject H0
T2	-1.949	1.971	Reject H0
T3	-2.959	1.612	Reject H0

Table 4.61: *The cutoff for rejection region and conclusion of different tests*

service in roughly a sequential order as requests for repairs were received. The average daily mileage for the 36 repair-persons are in Table 4.62:

Groups	Mileage					
Treatment Group	62.2	79.3	83.2	82.2	84.1	89.3
	95.8	97.9	91.5	96.6	90.1	98.6
	85.2	87.9	86.7	99.7	101.1	88.6
Control Group	97.1	70.2	94.6	182.9	85.6	89.5
	109.5	101.7	99.7	193.2	105.3	92.9
	63.9	88.2	99.1	95.1	92.4	87.3

Table 4.62: *Cable TV Company data set. The data set is from exercise* 6.17 *of textbook Ott and Longnecker (2008)*

The two-sample data set is balanced, with $\lambda_N = 0.5$ and total sample size 36. In Figure 4.28, the two samples have different medians and at least one population appears to be skewed. The pooled two-sample t statistic value is 1.70509. Then comparing to the cut off values of the rejection region listed in Table 4.63, we can see that the two-sample TCF, T_1, T_3 and bootstrap t-tests reject the null hypothesis and conclude that the two populations have significantly different means. But the pooled two-sample t-test, Wilcoxon Rank-Sum Test(p-value $= 0.082$) and T_2 test fail to reject the null hypothesis and conclude the two populations means are not significantly different with p-value all close to 0.1.

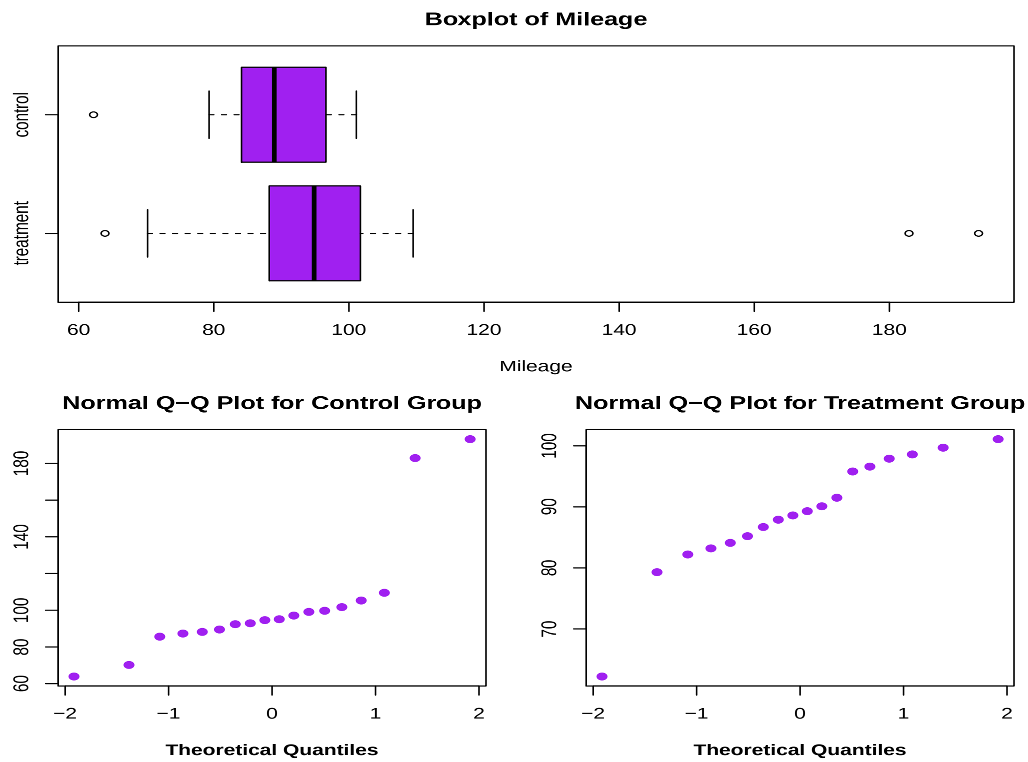

Figure 4.28: *Box-plot and Q-Q plot for mileage data*

In this first example, our four new test gave consistent testing results as the pooled two-sample t-test, bootstrap t-test and Wilcoxon Rank-Sum Test. Furthermore, when it comes to the second example whose data are more skewed, our proposed new two-sample tests reject the null hypothesis and conclude that the two population means are significantly different from each other except for T_2 test, while the commonly used pooled two-sample t-test and Wilcoxon Rank-Sum Test did not reject the null hypothesis and conclude that the two population means are not significantly different.

4.7 Summary

In this chapter, we presented three new two-sample tests based on transformations followed by extensive simulation studies. When testing the two population mean difference, our study shows that the four new two-sample tests, one based on Cornish Fisher expansion

	lower cutoff	upper cutoff	Conclusion at level 0.05
T.test	-2.088	2.088	Do Not Reject H0
bootstrap	-3.760	1.570	Do Not Reject H0
TCF	-2.362	1.558	Reject H0
T1	-2.475	1.614	Reject H0
T2	-2.062	1.851	Do Not Reject H0
T3	-3.710	1.454	Reject H0

Table 4.63: *The cutoff for rejection region and conclusion of different tests*

(TCF) and three based on transformations (T_i), $i = 1, 2, 3$, can provide more accurate tests under skewness. Comparing to the two-sample test based on normal approximation (TN), TCF and T_i tests have the same type I error rate but give higher power for the same sample size.

The Bootstrap-t test and BCa test not only provide elevated type I error rate but also need significantly more computation time due to bootstrap resampling. Among the seven two-sample tests in the simulation study, the two-sample TCF test and T_3 test are recommended for comparing skewed populations over the TN test since they have better power both theoretically and empirically in addition to well maintained type I error control.

Chapter 5

Proposed two-sample test using Cornish Fisher expansion for skewed populations with unequal variances

In this chapter, we will extend this research in previous chapter by considering two-sample comparison for the skewed population with unequal variances. As introduced in Section 2.3.2.2, Zhou and Philip (2005) derived an approximation distribution for the test statistic of unpooled two sample t-test by Edgeworth expansion theory as follows:

Let $\lambda_N = n_1/(n_1+n_2) = n_1/N$. Assume $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$. Under regularity conditions, the distribution of the unpooled two sample t-statistic $T^{'}$ given in (2.1.1) has the following expansion:

$$F_T^{(U)}(x) = P(T^{'} \leq x) = \Phi(x) + \frac{A^{'}}{6\sqrt{N}}(2x^2+1)\phi(x) + O(N^{-min(1,r+1/2)}), \qquad (5.0.1)$$

where $\phi(x)$ is the probability density function of the standard normal distribution, $\Phi(x)$ is the cumulative distribution function of the standard normal distribution and

$$A^{'} = \left\{\frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1-\lambda}\right\}^{-3/2}\left\{\frac{\sigma_1^3\gamma_1}{\lambda^2} - \frac{\sigma_2^3\gamma_2}{(1-\lambda)^2}\right\}. \qquad (5.0.2)$$

Based on the above approximation distribution in (5.0.1), we follow the same procedures in Chapter 3 to derive a new two-sample t-test by the Cornish Fisher expansion theory.

5.1 New test based on Cornish Fisher expansion under unequal variances

5.1.1 A two-sample test based on Cornish Fisher expansion under unequal variances

As introduced in Section 2.3.1, the percentiles of the distribution in (5.0.1) admits a Cornish Fisher expansion, which has the form as follows.

Corollary 5.1.1. *Let τ_α denote the α^{th} percentile of the distribution $F_T^{(U)}(t)$ in (5.0.1). Then based on Cornish Fisher expansion theory, the value of τ_α admits an expansion with the form below:*

$$\tau_\alpha = z_\alpha - \frac{A'}{6\sqrt{N}}(2z_\alpha^2 + 1) + O(N^{-min(1,r+1/2)}), \tag{5.1.1}$$

where z_α is the α^{th} percentile of the standard normal distribution and A' is defined in (5.0.2).

This corollary is a direct result of the theory for Fisher expansion from Hall (1992a). Hence we omit the proof.

Now define $\hat{\tau}_\alpha = z_\alpha - \frac{\hat{A}'}{6\sqrt{N}}(2z_\alpha^2 + 1)$, where

$$\hat{A}' = \left\{\frac{S_1^2}{\lambda_N} + \frac{S_2^2}{1-\lambda_N}\right\}^{-3/2} \left\{\frac{S_1^3\hat{\gamma}_1}{\lambda_N^2} - \frac{S_2^3\hat{\gamma}_2}{(1-\lambda_N)^2}\right\},$$

and

$$\hat{\gamma}_i = \frac{n_i}{(n_i-1)(n_i-2)} \sum_{j=1}^{n_i} \left\{\frac{X_{ij} - \bar{X}_i}{S_i}\right\}^3$$
$$\lambda_N = \frac{n_1}{(n_1+n_2)} \qquad , i = 1, 2.$$

With the test statistic T' defined in equation (2.1.1), we have:

1. The rejection region for two-sided test with hypothesis $H_0 : \mu_1 - \mu_2 = \mu_{10} - \mu_{20}$ vs. $H_a : \mu_1 - \mu_2 \neq \mu_{10} - \mu_{20}$ is

$$T' \leq \hat{\tau}_{\alpha/2} \ \text{ or } \ T' \geq \hat{\tau}_{1-\alpha/2} \tag{5.1.2}$$

2. The rejection region for one-sided upper-tailed test with hypothesis $H_0 : \mu_1 - \mu_2 = \mu_{10} - \mu_{20}$ vs. $H_a : \mu_1 - \mu_2 > \mu_{10} - \mu_{20}$ is

$$T^{'} \geq \hat{\tau}_{1-\alpha}. \tag{5.1.3}$$

3. The rejection region for one-sided lower-tailed test with hypothesis $H_0 : \mu_1 - \mu_2 = \mu_{10} - \mu_{20}$ vs. $H_a : \mu_1 - \mu_2 < \mu_{10} - \mu_{20}$ is

$$T^{'} \leq \hat{\tau}_{\alpha}. \tag{5.1.4}$$

We reject the null hypothesis if $T^{'}$ falls into the rejection regions for corresponding alternative hypothesis. In the further discussions, we will refer this two sample test based on Cornish Fisher expansion under unequal variances as "unpooled-TCF test"

5.1.2 Type I error rate of the two-sided unpooled-TCF test

In this section, we calculate the order of approximation to type I error rate for the two-sided unpooled-TCF test with rejection region in (5.1.2) from the first order Cornish Fisher expansion. Under the two-sample setting in above Section 3.2.1, the distribution of the test statistic $T^{'}$ is $F_T^{(U)}(x)$ defined in (5.0.1). Denote the two cutoffs in formula (5.1.2) as

$$\begin{aligned} \hat{\tau}_{\alpha/2} &= z_{\alpha/2} - \frac{\hat{A}^{'}}{6\sqrt{N}}(2z_{\alpha/2}^2 + 1) \triangleq z_{\alpha/2} + \hat{\Delta}_{N,\alpha/2}^{'} \\ \hat{\tau}_{1-\alpha/2} &= z_{1-\alpha/2} - \frac{\hat{A}^{'}}{6\sqrt{N}}(2z_{1-\alpha/2}^2 + 1) \triangleq z_{1-\alpha/2} + \hat{\Delta}_{N,1-\alpha/2}^{'}, \end{aligned} \tag{5.1.5}$$

where

$$\begin{aligned} \hat{\Delta}_{N,\alpha/2}^{'} &= -\frac{\hat{A}^{'}}{6\sqrt{N}}(2z_{\alpha/2}^2 + 1) \\ \hat{\Delta}_{N,1-\alpha/2}^{'} &= -\frac{\hat{A}^{'}}{6\sqrt{N}}(2z_{1-\alpha/2}^2 + 1). \end{aligned}$$

Under standard regularity conditions and previous results, we have

$$\begin{aligned} \hat{\gamma}_i &= \gamma_i + O_p(N^{-1/2}), \\ S_i &= \sigma_i + O_p(N^{-1/2}), \\ \lambda_N &= \lambda + O(N^{-r}). \end{aligned}$$

Then we can show that

$$
\begin{aligned}
\hat{A}' &= \left\{ \frac{S_1^2}{\lambda_N} + \frac{S_2^2}{1-\lambda_N} \right\}^{-3/2} \left\{ \frac{S_1^3 \hat{\gamma}_1}{\lambda_N^2} - \frac{S_2^3 \hat{\gamma}_2}{(1-\lambda_N)^2} \right\} \\
&= \left\{ \frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1-\lambda} + O_p(N^{-min(r,\frac{1}{2})}) \right\}^{-3/2} \left\{ \frac{\sigma_1^3 \gamma_1}{\lambda^2} - \frac{\sigma_2^3 \gamma_2}{(1-\lambda)^2} + O_p(N^{-min(r,\frac{1}{2})}) \right\} \\
&= A' + O_p(N^{-min(r,\frac{1}{2})}),
\end{aligned}
$$

Based on the above result of $\hat{A}'$, we can obtain

$$
\begin{aligned}
\hat{\Delta}'_{N,\alpha/2} &= -\frac{\hat{A}'}{6\sqrt{N}}(2z_{\alpha/2}^2 + 1) \\
&= -\frac{(A' + O_p(N^{-min(r,\frac{1}{2})}))}{6\sqrt{N}}(2z_{\alpha/2}^2 + 1) \\
&= -\frac{A'}{6\sqrt{N}}(2z_{\alpha/2}^2 + 1) + O_p(N^{-min(1,r+1/2)}) \\
&= \Delta'_{N,\alpha/2} + O_p(N^{-min(1,r+1/2)}),
\end{aligned}
$$

and

$$
\hat{\Delta}'_{N,1-\alpha/2} = \Delta'_{N,1-\alpha/2} + O_p(N^{-min(1,r+1/2)}).
$$

Now we have

$$
\begin{aligned}
\hat{\tau}_{\alpha/2} &= \tau_{\alpha/2} + O_p(N^{-min(1,r+1/2)}) \\
\hat{\tau}_{1-\alpha/2} &= \tau_{1-\alpha/2} + O_p(N^{-min(1,r+1/2)}),
\end{aligned}
$$

and the following Lemma:

Lemma 5.1.2. *Let τ_α denote the α^{th} percentile of distribution $F_T^{(U)}(t)$ in (5.0.1) and $\hat{\tau}_\alpha$ denote the estimate of τ_α given in (5.1.5). Then under standard regularity conditions, the following result holds:*

$$
\begin{aligned}
&P(T \le \hat{\tau}_{\alpha/2}) + P(T \ge \hat{\tau}_{1-\alpha/2}) \\
&= 1 - F_T^{(U)}(\tau_{1-\alpha/2}) + F_T^{(U)}(\tau_{\alpha/2}) + O(N^{-min(1,r+1/2)}).
\end{aligned}
$$

Recall that the distribution of $F_T^{(2)}(x)$ in equation (5.0.1) takes the following form

$$F_T^{(U)}(x) = P(T^{'} \leq x) = \Phi(x) + \frac{A^{'}}{6\sqrt{N}}(2x^2+1)\phi(x) + O(N^{-min(1,r+1/2)}),$$

Hence, based on the result of Lemma 5.1.2, the type I error rate of the two-sample unpooled-TCF test can be obtained as

$$\begin{aligned}
&P(\text{type I error of the two-sided unpooled-TCF test})\\
=&1 - F_T^{(U)}(\tau_{1-\alpha/2}) + F_T^{(U)}(\tau_{\alpha/2}) + O(N^{-min(1,r+1/2)})\\
=&1 - \Phi(\tau_{1-\alpha/2}) + \Phi(\tau_{\alpha/2}) - \frac{A^{'}}{6\sqrt{N}}(2\tau_{1-\alpha/2}^2+1)\phi(\tau_{1-\alpha/2})\\
&+ \frac{A^{'}}{6\sqrt{N}}(2\tau_{\alpha/2}^2+1)\phi(\tau_{\alpha/2}) + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{5.1.6}$$

Due to the fact that $A^{'}$ are finite constant, then we have

$$\Delta_{N,\alpha/2}^{'} = \Delta_{N,1-\alpha/2}^{'} = O(N^{-1/2}) \tag{5.1.7}$$

Note that

$$\begin{aligned}
&-\frac{A^{'}}{6\sqrt{N}}(2\tau_{\alpha/2}^2+1)\\
&= -\frac{A^{'}}{6\sqrt{N}}(2(z_{\alpha/2}+\Delta_{N,\alpha/2}^{'})^2+1)\\
&= -\frac{A^{'}}{6\sqrt{N}}(2z_{\alpha/2}^2+1+O(N^{-1/2}))\\
&= \Delta_{N,\alpha/2}^{'} + O(N^{-1}).
\end{aligned} \tag{5.1.8}$$

Similarly,

$$-\frac{A^{'}}{6\sqrt{N}}(2\tau_{1-\alpha/2}^2+1) = \Delta_{N,1-\alpha/2}^{'} + O(N^{-1}) = \Delta_{N,\alpha/2}^{'} + O(N^{-1}). \tag{5.1.9}$$

Then apply Taylor expansion to $\Phi(\tau_{1-\alpha/2})$, $\phi(\tau_{1-\alpha/2})$ at $z_{1-\alpha/2}$ and to $\Phi(\tau_{\alpha/2})$, $\phi(\tau_{\alpha/2})$ at $z_{\alpha/2}$ correspondingly, we have

$$\begin{aligned}
&\Phi(\tau_{1-\alpha/2}) = \Phi(z_{1-\alpha/2}) + \phi(z_{1-\alpha/2})\Delta_{N,1-\alpha/2}^{'} + O(\Delta_{N,1-\alpha/2}^{'2}),\\
&\Phi(\tau_{\alpha/2}) = \Phi(z_{\alpha/2}) + \phi(z_{\alpha/2})\Delta_{N,\alpha/2}^{'} + O(\Delta_{N,\alpha/2}^{'2}),\\
&\phi(\tau_{1-\alpha/2}) = \phi(z_{1-\alpha/2}) + \phi^{'}(z_{1-\alpha/2})\Delta_{N,1-\alpha/2}^{'} + O(\Delta_{N,1-\alpha/2}^{'2}),\\
&\phi(\tau_{\alpha/2}) = \phi(z_{\alpha/2}) + \phi^{'}(z_{\alpha/2})\Delta_{N,\alpha/2}^{'} + O(\Delta_{N,\alpha/2}^{'2}).
\end{aligned} \tag{5.1.10}$$

Using these four Taylor expansions and (5.1.8), (5.1.9) to replace the terms in (5.1.6), we have:

$$\begin{aligned}
& P(\text{type I error of the two-sided unpooled-TCF test }) \\
= \quad & 1-\Phi(\tau_{1-\alpha/2})+\Phi(\tau_{\alpha/2})-\frac{A^{'}}{6\sqrt{N}}(2\tau^2_{1-\alpha/2}+1)\phi(\tau_{1-\alpha/2}) \\
& +\frac{A^{'}}{6\sqrt{N}}(2\tau^2_{\alpha/2}+1)\phi(\tau_{\alpha/2})+O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha/2-\phi(z_{1-\alpha/2})\Delta^{'}_{N,\alpha/2}+\alpha/2+\phi(z_{\alpha/2})\Delta^{'}_{N,\alpha/2}+O(\Delta^{'2}_{N,\alpha/2}) \\
& +[\Delta^{'}_{N,\alpha/2}+O(N^{-1})][\phi(z_{1-\alpha/2})+\phi^{'}(z_{1-\alpha/2})\Delta^{'}_{N,1-\alpha/2}+O(\Delta^{'2}_{N,1-\alpha/2})] \\
& -[\Delta^{'}_{N,\alpha/2}+O(N^{-1})][\phi(z_{\alpha/2})+\phi^{'}(z_{\alpha/2})\Delta^{'}_{N,\alpha/2}+O(\Delta^{'2}_{N,\alpha/2})] \\
& +O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha+\Delta^{'}_{N,\alpha/2}\phi(z_{1-\alpha/2})-\Delta^{'}_{N,\alpha/2}\phi(z_{\alpha/2})+O(N^{-1})+O(N^{-min(1,r+1/2)}) \\
= \quad & \alpha+O(N^{-min(1,r+1/2)}),
\end{aligned} \tag{5.1.11}$$

Then we have the follow theorem,

Theorem 5.1.3. *Under standard regularity conditions, when H_0 is true, the theoretical type I error rate of the two-sample unpooled-TCF test, with level of significance α is*

$$P(T^{'}\leq\hat{\tau}_{\alpha/2})+P(T^{'}\geq\hat{\tau}_{1-\alpha/2})=\alpha+O(N^{-min(1,r+1/2)}).$$

Note that the approximated type I error rate of unpooled two-sample t-test based on normal approximation is also $\alpha+O(N^{-min(1,r+1/2)})$. This means that the two tests have the same type I error rate accuracy.

5.1.3 Power of the two-sided unpooled-TCF test

Now consider data generated under $H_a:\mu_1-\mu_2\neq\mu_{10}-\mu_{20}$. For power calculation, let $\delta=|(\mu_1-\mu_2)-(\mu_{10}-\mu_{20})|$. When H_a is true, the theoretical power equals the probability of rejecting the null hypothesis with formula:

$$P_{H_a}\left(T^{'}\geq\hat{\tau}_{1-\alpha/2}\right)+P_{H_a}\left(T^{'}\leq\hat{\tau}_{\alpha/2}\right)$$

Following the same Edgeworth expansion procedures in Section 3.2.3, the distribution of the test statistic $T^{'}$ under H_a can be obtained. We state the result in the Theorem below.

Theorem 5.1.4. *Let $\lambda_N = n_1/(n_1+n_2) = n_1/N$. Assume $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$, and $\delta = O(N^{-1/2})$ under H_a. Then under regularity conditions (in Appendix A), the distribution of the unpooled two sample t-statistic $T^{'}$ given in (2.1.1) has the following expansion under H_a:*

$$F^{(H_a)}_{T.unpooled}(x) = F^{(U)}_{T}(x - c_N) + \frac{q_N}{2}(x - c_N)\phi(x - c_N) + O(N^{-min(1,r+1/2)}), \quad (5.1.12)$$

where $\phi(x)$ is the probability density function of the standard normal distribution, $\Phi(x)$ is the cumulative distribution function of the standard normal distribution and $F^{(U)}_{T}(x)$ is the distribution of $T^{'}$ under H_0 defined in (5.0.1). Here, $c_N = \delta/\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}$ and $q_N = \delta\sigma^{-1}[\lambda(1-\lambda)](\gamma_1 - \gamma_2)$.

Denoting

$$\begin{aligned} \hat{L}^{'}_u = \hat{\tau}_{1-\alpha/2} - \delta/\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \\ = \quad & z_{1-\alpha/2} + \hat{\Delta}^{'}_{N,1-\alpha/2} - \delta/\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \\ = \quad & U^{'}_{N,1-\alpha/2} + \hat{\Delta}^{'}_{N,1-\alpha/2} \\ = \quad & U^{'}_{N,1-\alpha/2} + \Delta^{'}_{N,1-\alpha/2} + O_p(N^{-min(1,r+1/2)}) \\ = \quad & L^{'}_u + O_p(N^{-min(1,r+1/2)}) \end{aligned}$$

where $U^{'}_{N,1-\alpha/2} = z_{1-\alpha/2} - \delta/\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ and $L^{'}_u = U^{'}_{N,1-\alpha/2} + \Delta^{'}_{N,1-\alpha/2}$. Similarly

$$\hat{L}^{'}_l = \hat{\tau}_{\alpha/2} - \delta/\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = L^{'}_l + O_p(N^{-min(1,r+1/2)})$$

where $U^{'}_{N,\alpha/2} = z_{\alpha/2} - \delta/\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ and $L^{'}_l = U^{'}_{N,\alpha/2} + \Delta^{'}_{N,\alpha/2}$. We have

$$\hat{L}^{'}_u = L^{'}_u + O_p(N^{-min(1,r+1/2)}),$$
$$\hat{L}^{'}_l = L^{'}_l + O_p(N^{-min(1,r+1/2)}).$$

Then based on the result of Lemma 5.1.2, under standard regularity conditions, the theoretical power of the two-sample unpooled-TCF test can be obtained by

$$\begin{aligned}
& P_{H_a}\left(T \geq \hat{L}_u'\right) + P_{H_a}\left(T \leq \hat{L}_l'\right) \\
= \;& 1 - F_T^{(U)}(L_u') + F_T^{(U)}(L_l') - Q(\hat{L}_u') + Q(\hat{L}_l') + O(N^{-min(1,r+1/2)}) \\
= \;& 1 - \Phi(L_u') + \Phi(L_l') - \tfrac{A'}{6\sqrt{N}}(2L_u'^2 + 1)\phi(L_u') - Q(\hat{L}_u') \\
& + \tfrac{A'}{6\sqrt{N}}(2L_l'^2 + 1)\phi(L_l') + Q(\hat{L}_l') + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{5.1.13}$$

5.1.4 Power of the two-sided unpooled-TCF test under local alternative hypothesis

Recall that, under the local alternative, both $U'_{N,\alpha/2}$ and $U'_{N,1-\alpha/2}$ have the same order of $O(1)$. Therefore we have

$$-\frac{A'}{6\sqrt{N}}(2L_u'^2 + 1) = -\frac{A'}{6\sqrt{N}}(2U_{N,1-\alpha/2}'^2 + 1) + O(N^{-1}). \tag{5.1.14}$$

Similarly, $-\frac{A'}{6\sqrt{N}}(2L_l'^2 + 1) = -\frac{A'}{6\sqrt{N}}(2U_{N,\alpha/2}'^2 + 1) + O(N^{-1})$. Then apply Taylor expansion to $\Phi(L_u')$, $\phi(L_u')$ at $U'_{N,1-\alpha/2}$ and to $\Phi(L_l')$, $\phi(L_l')$ at $U'_{N,\alpha/2}$ correspondingly, we have

$$\begin{aligned}
\Phi(L_u') &= \Phi(U'_{N,1-\alpha/2}) + \phi(U'_{N,1-\alpha/2})\Delta'_{N,1-\alpha/2} + O(\Delta'^2_{N,1-\alpha/2}), \\
\Phi(L_l') &= \Phi(U'_{N,\alpha/2}) + \phi(U'_{N,\alpha/2})\Delta'_{N,\alpha/2} + O(\Delta'^2_{N,\alpha/2}), \\
\phi(L_u') &= \phi(U'_{N,1-\alpha/2}) + \phi'(U'_{N,1-\alpha/2})\Delta'_{N,1-\alpha/2} + O(\Delta'^2_{N,1-\alpha/2}), \\
\phi(L_l') &= \phi(U'_{N,\alpha/2}) + \phi'(U'_{N,\alpha/2})\Delta'_{N,\alpha/2} + O(\Delta'^2_{N,\alpha/2}).
\end{aligned} \tag{5.1.15}$$

Then (5.1.13) can be further calculated as

$$\begin{aligned}
& \text{power of the two-sided unpooled-TCF test} = \\
& 1 - \Phi(U'_{N,1-\alpha/2}) - \frac{A'}{6\sqrt{N}}(2U'^2_{N,1-\alpha/2} + 1)\phi(U'_{N,1-\alpha/2}) - Q(U'_{N,1-\alpha/2}) + Q(U'_{N,\alpha/2}) \quad (5.1.16) \\
& + \Phi(U'_{N,\alpha/2}) + \frac{A'}{6\sqrt{N}}(2U'^2_{N,\alpha/2} + 1)\phi(U'_{N,\alpha/2}) + L'_{N,\gamma_1,\gamma_2,\lambda} + O(N^{-min(1,r+1/2)}),
\end{aligned}$$

where $L'_{N,\gamma_1,\gamma_2,\lambda}$ has the form

$$L'_{(N,\gamma_1,\gamma_2,\lambda)} = \Delta'_{N,\alpha/2}[\phi(U'_{N,\alpha/2}) - \phi(U'_{N,1-\alpha/2})] + O(N^{-1}) \tag{5.1.17}$$

Note that, the approximated power of the unpooled two-sample t-test based on standard normal approximation is:

$$\text{power of the unpooled two-sample t-test} =$$
$$1-\Phi(U'_{N,1-\alpha/2})-\frac{A'}{6\sqrt{N}}(2U'^2_{N,1-\alpha/2}+1)\phi(U'_{N,1-\alpha/2})-Q(U'_{N,1-\alpha/2})+Q(U'_{N,\alpha/2}) \quad (5.1.18)$$
$$+\Phi(U'_{N,\alpha/2})+\frac{A'}{6\sqrt{N}}(2U'^2_{N,\alpha/2}+1)\phi(U'_{N,\alpha/2})+O(N^{-min(1,r+1/2)}).$$

From (5.1.16) and (5.1.18) we have the following Corollary:

Corollary 5.1.5. *The two-sided unpooled-TCF test at level α is more powerful than the unpooled two-sample t-test if and only if the following inequality holds*

$$c'_\alpha(1-\lambda)^2\sigma_1^3\gamma_1 > c'_\alpha\lambda^2\sigma_2^3\gamma_2, \tag{5.1.19}$$

where $c'_\alpha = 2z^2_{\alpha/2}-1$.

In real practice, the sign of $L'_{(N,\gamma_1,\gamma_2,\lambda)}$ can be manipulated by adjusting $\lambda = \lambda_N + O(N^{-r}) = n_1/N + O(N^{-r})$, since γ_1 and γ_2 are determined by the population. With a fixed value of γ_1 and γ_2, we can further rewrite the equation (5.1.19) as

$$(c'_\alpha\sigma_1^3\gamma_1 - c'_\alpha\sigma_2^3\gamma_2)\lambda^2 - 2c'_\alpha\sigma_1^3\gamma_1\lambda + c'_\alpha\sigma_1^3\gamma_1 > 0. \tag{5.1.20}$$

Now we can solve for λ which satisfy the inequality in (5.1.19). The left side of inequality (5.1.20) will have real roots if (5.1.21) is nonnegative.

$$(2c'_\alpha\sigma_1^3\gamma_1)^2 - 4(c'_\alpha\sigma_1^3\gamma_1 - c'_\alpha\sigma_2^3\gamma_2)c'_\alpha\sigma_1^3\gamma_1 > 0.$$
$$\text{That is} \qquad 4c'^2_\alpha\sigma_1^3\sigma_2^3\gamma_1\gamma_2 > 0. \tag{5.1.21}$$

Clearly, (5.1.21) satisfied if $\gamma_1\gamma_2 \geq 0$, since c'_α, σ_1 and σ_2 are all greater than 0 in real practise. When $\gamma_1\gamma_2 < 0$, we can switch the two samples and let $\gamma_1 > 0$ and $\gamma_2 < 0$. Then the inequality (5.1.19) always hold. Suppose ω'_1 and ω'_2 are the two roots of the left side of (5.1.20). Without loss of generality, let $\omega'_1 \leq \omega'_2$. Then under $\gamma_1\gamma_2 \geq 0$, we have the following solutions of λ for $L'_{N,\gamma_1,\gamma_2,\lambda} > 0$:

- if $\sigma_1^3\gamma_1 - \sigma_2^3\gamma_2 \geq 0$ then $\lambda \leq \omega_1^{'}$ or $\lambda \geq \omega_2^{'}$;

- if $\sigma_1^3\gamma_1 - \sigma_2^3\gamma_2 < 0$ then $\omega_1^{'} \leq \lambda \leq \omega_2^{'}$.

To demonstrate the relationship between λ and $L^{'}_{(N,\gamma_1,\gamma_2,\lambda)}$, consider one specific example. Suppose population 1 follows Gamma distribution with population skewness $\gamma_1 = 2$ and population standard deviation $\sigma_1 = 3$; population 2 follows Log-normal distribution with population skewness $\gamma_1 = 6$ and population standard deviation $\sigma_1 = 1$. With $\alpha = 0.05$, the unpooled-TCF and unpooled two-sample t-tests were applied to test the two population mean difference. Based on the given population parameters, we can solve for λ that gives a higher power in the two-sample unpooled-TCF test. Since $\gamma_1\gamma_2 \geq 0$ and $\sigma_1^3\gamma_1 - \sigma_2^3\gamma_2 \geq 0$, by solving the inequality (5.1.20), the solutions can be obtained as $\lambda \leq \omega_1^{'}$ or $\lambda \geq \omega_2^{'}$, where $\omega_1^{'} = 0.75$ and $\omega_2^{'} = 1.5$.

5.1.5 Summary

In this chapter, we derived the order of type I error rate accuracy and the power function of the two-sample unpooled-TCF test. Besides, we also provide the detailed conditions under which the theoretical power of the two-sample unpooled-TCF test is higher than the two-sample TN test. Comparing with the TCF test, the unpooled-TCF test can test the population mean difference under unequal variance assumption.

Dissertation summary

In this dissertation, we applied the Edgeworth expansion on the test statistic of pooled two sample t-test to derive a new approximation of its distribution under skewness. On the basis of this new expansion, a new two-sample test from Cornish Fisher expansion theory (TCF test) was constructed. We proved that the TCF test can maintain the type I error at rate $O(N^{-min(1,r+1/2)})$, but give a higher power than the pooled two-sample t-test when the underlying data are skewed. We quantified the power increment as a function of the sample size ratio and population skewnesses. A sufficient and necessary condition was given in the dissertation to guide users on the range of sample size ratio such that the TCF test has higher power than the two-sample t-test.

We also developed three new tests based on three transformations (T_i test, $i = 1, 2, 3$) for the pooled two-sample case. These three transformations help to eliminate the skewness of the studentized statistic. We proved that the three T_i tests have the same order of type I error rate accuracy as the pooled two-sample t-test based on normal approximation (TN test) and the TCF test. In terms of power, we proved that the two-sided TCF test has higher power than the two-sided T_1 and T_2 tests; the power difference between two-sided TCF test and T_3 test depends on the relative skewness B defined in (3.1.10). Small population skewness difference yield a small B, which leads to a higher power for the T_3 test than the TCF test. On the other hand, a big population skewness difference will yield a large B, which leads to a lower power in T_3 test than TCF test. This result means the two-sided TCF test can provide more accurate two sample comparison when the populations are highly skewed.

Beyond theoretical development, this dissertation also give extensive simulation studies to compare the proposed tests with commonly used two-sample tests. The simulation study in Section 3.3 shows that the two-sample TCF test not only maintained the type I error rate but provided highest power and should be recommended over the pooled two-sample t-test, Bootstrap t-test and Wilcoxon Rank-Sum Test for skewed data. Moreover, the simulation results for two-sided hypothesis or upper-tailed test show that, among all the seven tests, the two-sample TCF test and T_3 test not only give consistently higher power but also control the type I error rate well. As the population relative skewness increases, the power of these two tests outperform the power of Bootstrap-t test and BCa test. For the lower-tailed test, we observed interesting phenomenon. For populations with smaller relative skewness, the TN test performs the best in terms of both type I error and power. When the relative skewness increases, the TN test exhibits higher power but also has the most elevated type I error rate.

Finally, this dissertation extends the study to compare two skewed populations with unequal variances. We derived the unpooled-TCF test based on the Edgeworth expansion of the test statistic of the unpooled two-sample t-test. The theoretical result shows that the unpooled-TCF test gives the same order of type I error rate accuracy as the unpooled two-sample t-test based on normal approximation. We also provided the condition on the sample ratio to yield a higher power for the unpooled-TCF test.

The five new two-sample tests i.e., TCF test, unpooled-TCF test and three T_i tests are especially designed for the two-sample comparison under the skewed populations. Comparing with the commonly used pooled and unpooled two-sample t-tests and Wilcoxon Rank-Sum test, the five new tests have less restricted assumptions and provide better power to detect departure from the null hypothesis. Compared to resampling based bootstrap-t test and BCa test, the five new tests have better type I error control while giving a comparable power.

Bibliography

Bag, A. K. (1966). Binomial theorem in ancient india. *Indian J. History Sci*, 1:68–74.

Bai, J. and NG, S. (2005). Tests for skewness, kurtosis and normality for time series data. *Journal of Business and Economic Statistics*, 23:49–60.

Barrett, J. and Goldsmith, L. (1976). When is n sufficiently large? *American Statistician*, 30:67–70.

Bochner, S. and Chandrasekharan, K. (1949). *Fourier Transforms*. Princeton University Press.

Boos, D. and Hughes, O. J. (2000). How large does n have to be for z and t intervals? *American Statistician*, 54:121–128.

Box, J. F. (1987). Guinness, gosset, fisher, and small samples. *Statistical Science*, 2:45–52.

Chaffin, W. and Rhiel, G. (1993). The effect of skewness and kurtosis on the one-sample t test and the impact of knowledge of the population standard deviation. *Journal of Computation and Simulation*, 46:79–90.

Chen, L. (1995). Testing the mean of skewed distributions. *Journal of American Statistical Association*, 90:767–772.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.

Cornish, E. and Fisher, R. (1938). Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute*, 5:307–320.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82:171–185.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Fedoryuk, M. (2001). *Hermite functions*. Springer, Berlin.

Fisher, N. and Hall, P. (1989). New statistical methods for directional data. *Geophysical Journal International*, 101:305–313.

Fisher, R. (1935). *The Design of Experiments*. New York Hafner.

Fisher, R. A. and Cornish, E. A. (1960). The percentile point of distribution having know cumulants. *Technometrics*, 2:209–225.

Gayen, A. (1949). The distribution of student t in random samples of any size drawn from non-normal universes. *Biometrika*, 36:353–369.

Good, I. (1977). A new formula for k-statistics. *The Annals of Statistics*, 5:224–228.

Hall, P. (1992a). *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics, New York.

Hall, P. (1992b). On the removal of skewness by transformation. *Journal of Royal Statistical Society*, 54:221–228.

Johnson, N. (1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of American Statistical Association*, 73:536–544.

Mankiewicz, R. (2004). *The Story of Mathematics*. Princeton University Press, Princeton.

Ott, R. L. and Longnecker, M. T. (2008). *An Introduction to Statistical Methods and Data Analysis*. Duxbury Press, Michigan.

Phillip, D. and Zhou, X. H. (2005). Nonparametric statistical methods for the cost-effectiveness analysis. *Biostatistics*, 62:576–588.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any population. *Royal Statistical Society Supplement*, 4:119–130.

Reineke, D. M., Baggett, J., and Elfessi, A. (2003). A note on the effect of skewness, kurtosis, and shifting on one-sample t and sign tests. *Journal of Statistics Education*, 11:71–85.

Zhou, X. H. and Philip, D. (2005). Nonparametric confidence intervals for the one- and two-sample problems. *Biostatistics*, 6:187–200.

Appendix A

Proof of Theorems

A.1 Proof of Theorem 3.1.3

Proof of Theorem 3.1.3:

Under the sufficient regularity conditions from Hall (1992a):

$$E(|X|^{j+2}) < \infty \quad and \quad \limsup_{|t|\to\infty} |X(t)| < 1 \tag{A.1.1}$$

holds if the distribution of X is nonsingular.

The pooled two sample t-test is as follows,

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$, $S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$ and $S_i^2 = \frac{1}{n_i-1}\sum_{i=1}^{n_i}(Y_i - \bar{Y}_i)^2$ with $i = 1, 2$.

Let $Y_{ij}^* = \frac{Y_{ij}-\mu_i}{\sigma_i}$, $\bar{Y}_i^* = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}^*$ and ${S_i^*}^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(Y_{ij}^* - \bar{Y}_i^*)^2$, for $i = 1, 2$ and $j = 1, ..., n_i$. Use these new defined variables to replace the original statistics in pooled two sample t-statistic, we have

$$\begin{aligned} T &= \frac{\sigma_1\bar{Y}_1^* - \sigma_2\bar{Y}_2^*}{\sqrt{\frac{(n_1-1)\sigma_1^2{S_1^*}^2+(n_2-1)\sigma_2^2{S_2^*}^2}{n_1+n_2-2}\frac{(n_1+n_2)}{n_1n_2}}} \\ &= \sqrt{N}\frac{\sigma_1\bar{Y}_1^* - \sigma_2\bar{Y}_2^*}{\sqrt{\frac{(n_1-1)\sigma_1^2{S_1^*}^2+(n_2-1)\sigma_2^2{S_2^*}^2}{(N-2)\lambda_N(1-\lambda_N)}}}, \end{aligned} \tag{A.1.2}$$

where $\lambda_N = n_1/N = n_1/(n_1 + n_2)$.

Furthermore, let $X = (X_1, X_2, X_3, X_4)$, where

$$X_1 = \bar{Y}_1^*, \; X_2 = \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j}^{*^2},$$
$$X_3 = \bar{Y}_2^*, \; X_4 = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j}^{*^2}.$$

Now plug X into (A.1.2) to further transform the pooled two sample t-statistic, finally we can write test statistic T as a function of X with $T = \sqrt{N}g(X)$, which has the form as follows:

$$T = \sqrt{N}g(X) = \frac{\sigma_1 X_1 - \sigma_2 X_3}{k(X)^{1/2}}$$
$$k(X) = \frac{(n_1-1)\sigma_1^2 S_1^{*^2} + (n_2-1)\sigma_2^2 S_2^{*^2}}{(N-2)\lambda_N(1-\lambda_N)}$$
$$= \frac{(n_1-1)\sigma_1^2(X_2 - X_1^2) + (n_2-1)\sigma_2^2(X_4 - X_3^2)}{(N-2)\lambda_N(1-\lambda_N)}$$

Next, apply Taylor expansion to $g(X)$ with $EX \equiv U \equiv (U_1, U_2, U_3, U_4) = (0, 1, 0, 1)$, we have

$$g(X) = g(U) + \frac{\partial g(U)}{\partial U}(X-U) + \frac{1}{2}\frac{\partial^2 g(U)}{\partial U^2}(X-U)^2 + \dots$$
$$T = \sqrt{N}\{\frac{\partial g(U)}{\partial U}(X-U) + \frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U) + \dots\},$$

where $g(U) = 0$.

Now, if we let

$$W_N = \sqrt{N}\{\frac{\partial g(U)'}{\partial U}(X-U) + \frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)\}.$$

Under regularity condition in (A.1.1), from Zhou and Philip (2005) we can show that

$$T = W_N + O(N^{-1}).$$

Assuming $EY_{ij}^6 < \infty$, the first three moments of W_N are as follows:

$$E(W_N) = -\frac{1}{2}N^{-1/2}[\lambda(1-\lambda)]^{1/2}(\gamma_1 - \gamma_2) + O(N^{-min(1,r+\frac{1}{2})}), \tag{A.1.3}$$

$$E(W_N^2) = 1 + O(N^{-1}), \tag{A.1.4}$$

$$E(W_N^3) = -\left[\frac{\lambda(1-\lambda)}{4N}\right]^{1/2}(\frac{11\lambda-2}{\lambda}\gamma_1 - \frac{9-11\lambda}{1-\lambda}\gamma_2) + O(N^{-min(1,r+\frac{1}{2})}), \tag{A.1.5}$$

where $\gamma_i = E[(\frac{Y_{ij}-\mu_i}{\sigma_i})^3]$ is the population skewness and $i = 1, 2$.

A.1.1 Proof of Corollary 3.1.1

The proof of Corollary 3.1.1 is shown as follows: First we will find $E(W_N)$.

$$EW_N = \sqrt{N}E\frac{\partial g(U)'}{\partial U}(X-U) + \frac{1}{2}\sqrt{N}E(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U).$$

To calculate $E\{\frac{\partial g(U)'}{\partial U}(X-U)\}$ and $E\{(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)\}$, denote $P = \frac{N^2(n_1-1)\sigma_1^2}{(N-2)n_1n_2}$ and $Q = \frac{N^2(n_2-1)\sigma_2^2}{(N-2)n_1n_2}$. Then,

$$\begin{aligned}
&E\left\{\frac{\partial g(U)'}{\partial U}(X-U)\right\} \\
&= E\left\{[(\frac{\sigma_1^2}{P+Q})^{1/2}\ ,\ 0\ ,\ -(\frac{\sigma_2^2}{P+Q})^{1/2})\ ,\ 0]\begin{bmatrix} X_1 \\ X_2-1 \\ X_3 \\ X_4-1 \end{bmatrix}\right\} \\
&= E\left(X_1(\frac{\sigma_1^2}{P+Q})^{1/2} - X_3(\frac{\sigma_2^2}{P+Q})^{1/2}\right) \\
&= 0
\end{aligned} \tag{A.1.6}$$

Then the expectation of the second term

$$\begin{aligned}
&E\left\{(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)\right\} \\
&= E\left\{[X_1,\ X_2-1,\ X_3,\ X_4-1]\begin{bmatrix} \frac{\partial^2 g(X)}{\partial X_1^2} & \cdots & \frac{\partial^2 g(X)}{\partial X_1 \partial X_4} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g(X)}{\partial X_4 \partial X_1} & \cdots & \frac{\partial^2 g(X)}{\partial X_4^2} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2-1 \\ X_3 \\ X_4-1 \end{bmatrix}\right\}
\end{aligned} \tag{A.1.7}$$

We break (A.1.7) down to calculate the middle term first:

$$\begin{bmatrix} \frac{\partial^2 g(X)}{\partial X_1^2} & \cdots & \frac{\partial^2 g(X)}{\partial X_1 \partial X_4} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g(X)}{\partial X_4 \partial X_1} & \cdots & \frac{\partial^2 g(X)}{\partial X_4^2} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & -\frac{P}{2}\sigma_1(P+Q)^{-3/2} & 0 & -\frac{Q}{2}\sigma_1(P+Q)^{-3/2} \\ -\frac{P}{2}\sigma_1(P+Q)^{-3/2} & 0 & \frac{P}{2}\sigma_2(P+Q)^{-3/2} & 0 \\ 0 & \frac{P}{2}\sigma_2(P+Q)^{-3/2} & 0 & \frac{Q}{2}\sigma_2(P+Q)^{-3/2} \\ -\frac{Q}{2}\sigma_1(P+Q)^{-3/2} & 0 & -\frac{Q}{2}\sigma_2(P+Q)^{-3/2} & 0 \end{bmatrix} \tag{A.1.8}$$

Denote the matrix in (A.1.8) as W, and we put W back to (A.1.7)

$$E\left\{ [X_1,\ X_2-1,\ X_3,\ X_4-1]\mathbf{W} \begin{bmatrix} X_1 \\ X_2-1 \\ X_3 \\ X_4-1 \end{bmatrix} \right\} \tag{A.1.9}$$
$$= E\left[(-P)(P+Q)^{-3/2}\sigma_1 X_1(X_2-1)\right] + E\left[(Q)(P+Q)^{-3/2}\sigma_2 X_3(X_4-1)\right]$$

Here (X_1, X_2) are independent with (X_3, X_4), and P , Q and σ_i are all constants. So the random terms in (A.1.9) are $E\left[(X_2-1)X_1\right]$ and $E\left[(X_4-1)X_3\right]$.

$$\begin{aligned}E\left[(X_2-1)X_1\right]=&E\left[(\frac{1}{n_1}\sum_{j=1}^{n_1}Y_{1j}^{*^2}-1)(\frac{1}{n_1}\sum_{i=1}^{n_1}Y_{1i}^{*})\right]\\=&E\left[\frac{1}{n_1}\sum_{j=1}^{n_1}(Y_{1j}^{*^2}-1)(\frac{1}{n_1}\sum_{i=1}^{n_1}Y_{1i}^{*})\right]\\=&\frac{1}{n_1^2}\sum_{j=1}^{n_1}\sum_{i=1}^{n_1}E[(Y_{1j}^{*^2}-1)Y_{1i}^{*}]\\&\text{if } i\neq j \text{ then } Y_{1j}^{*} \text{ independent with } Y_{1i}^{*}\\=&\frac{1}{n_1^2}\sum_{j=1}^{n_1}E[(Y_{1j}^{*^2}-1)Y_{1i}^{*}]\\=&\frac{1}{n_1}E[(Y_{1j}^{*^3}]\\=&\frac{\gamma_1}{n_1}.\end{aligned}\tag{A.1.10}$$

Based on the similar calculation, we get $E\left[(X_4-1)X_3\right]=\frac{\gamma_2}{n_2}$. Put the results in (A.1.10) back to (A.1.9), we have

$$\begin{aligned}&E\left\{[X_1,\ X_2-1,\ X_3,\ X_4-1]\mathbf{W}\begin{bmatrix}X_1\\X_2-1\\X_3\\X_4-1\end{bmatrix}\right\}\\=&(-P)(P+Q)^{-3/2}\sigma_1\frac{\gamma_1}{n_1}+(Q)(P+Q)^{-3/2}\sigma_2\frac{\gamma_2}{n_2}.\end{aligned}\tag{A.1.11}$$

Combing the results from (A.1.6) and (A.1.11), we have

$$\begin{aligned}E(W_N)=&E\left\{\sqrt{N}[\frac{\partial g(U)'}{\partial U}(X-U)+\frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)]\right\}\\=&\frac{1}{2}\sqrt{N}\left\{Q(P+Q)^{-3/2}\sigma_2\frac{\gamma_2}{n_2}-P(P+Q)^{-3/2}\sigma_1\frac{\gamma_1}{n_1}\right\}\end{aligned}\tag{A.1.12}$$

The first term $\frac{1}{2}\sqrt{N}\left\{Q(P+Q)^{-3/2}\sigma_2\frac{\gamma_2}{n_2}\right\}$ can be computed as,

$$\begin{aligned}&\frac{1}{2}\sqrt{N}\left\{Q(P+Q)^{-3/2}\sigma_2\frac{\gamma_2}{n_2}\right\}\\=&\frac{1}{2}\sqrt{N}\left(\frac{N^2(n_2-1)\sigma_2^2}{(N-2)n_1n_2}\right)\left(\frac{N^2[(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2]}{(N-2)n_1n_2}\right)^{-\frac{3}{2}}\sigma_2\frac{\gamma_2}{n_2}.\end{aligned}\tag{A.1.13}$$

let $B_1 = \frac{N^2\sigma_2^{\frac{3}{2}}\gamma_2}{(N-2)n_1n_2}\left(\frac{N^2[(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2]}{(N-2)n_1n_2}\right)^{-\frac{3}{2}}$ then, above equation is equal to $\frac{1}{2}\sqrt{N}(B_1 - \frac{B_1}{n_2})$. Then we apply Tyler expansion to term $[(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2]^{-\frac{3}{2}}$ in B_1 and have,

$$\begin{aligned}
&[(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2]^{-\frac{3}{2}} \\
=&[(\lambda_N N - 1)\sigma_1^2 + (N - \lambda_N N - 1)\sigma_2^2]^{-\frac{3}{2}} \\
=&[\lambda_N N\sigma_1^2 + N(1-\lambda_N)\sigma_2^2]^{-\frac{3}{2}} + \frac{3}{2}[\lambda_N N\sigma_1^2 + N(1-\lambda_N)\sigma_2^2]^{-\frac{5}{2}}(\sigma_1^2+\sigma_2^2) + O(N^{-\frac{7}{2}}) \\
&\text{(since } \lambda_N = \lambda + O(N^{-r})\text{, we have)} \\
=&[(\lambda + O(N^{-r}))N\sigma_1^2 + N(1-\lambda+O(N^{-r}))\sigma_2^2]^{-\frac{3}{2}} + O(N^{-\frac{5}{2}}) \\
=&N^{-3/2}\left\{[\lambda\sigma_1^2 + (1-\lambda)\sigma_2^2 + O(N^{-r})]^{-\frac{3}{2}})\right\} + O(N^{-\frac{5}{2}}) \\
=&N^{-3/2}\left\{[\lambda\sigma_1^2 + (1-\lambda)\sigma_2^2]^{-\frac{3}{2}} + O(N^{-r})\right\} + O(N^{-\frac{5}{2}}).
\end{aligned} \tag{A.1.14}$$

Then putting the results from (A.1.14) back to B_1, we have,

$$\begin{aligned}
B_1 =&\frac{N^2\sigma_2^{\frac{3}{2}}\gamma_2}{(N-2)n_1n_2}\left(\frac{N^2[(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2]}{(N-2)n_1n_2}\right)^{-\frac{3}{2}} \\
=&\sigma_2^{\frac{3}{2}}\gamma_2(\frac{N-2}{N^2})^{1/2}\frac{[(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2]^{-3/2}}{(n_1n_2)^{-1/2}} \\
=&\sigma_2^{\frac{3}{2}}\gamma_2(\frac{N-2}{N^2})^{1/2}\frac{N^{-3/2}\left\{[\lambda\sigma_1^2+(1-\lambda)\sigma_2^2]^{-\frac{3}{2}}+O(N^{-r})\right\}}{N^{-1}[\lambda(1-\lambda)+O(N^{-r})]^{-\frac{1}{2}}} \\
=&\sigma_2^{\frac{3}{2}}\gamma_2(\frac{N-2}{N^3})^{1/2}\frac{\left\{[\lambda\sigma_1^2+(1-\lambda)\sigma_2^2]^{-\frac{3}{2}}+O(N^{-r})\right\}}{[\lambda(1-\lambda)]^{-\frac{1}{2}}+O(N^{-r})} \\
=&\sigma_2^{\frac{3}{2}}\gamma_2(\frac{N-2}{N^3})^{1/2}\left\{\frac{[\lambda\sigma_1^2+(1-\lambda)\sigma_2^2]^{-\frac{3}{2}}}{[\lambda(1-\lambda)]^{-\frac{1}{2}}}+O(N^{-r})\right\} \\
=&\sigma_2^{\frac{3}{2}}\gamma_2\left\{\frac{[\lambda\sigma_1^2+(1-\lambda)\sigma_2^2]^{-\frac{3}{2}}}{N[\lambda(1-\lambda)]^{-\frac{1}{2}}}+O(N^{-r-1})\right\}.
\end{aligned} \tag{A.1.15}$$

Then plug B_1 back to (A.1.13), we have

$$\begin{aligned}
&\frac{1}{2}\sqrt{N}(B_1 - \frac{B_1}{n_2}) \\
=&\frac{1}{2\sqrt{N}}\sigma_2^{\frac{3}{2}}\gamma_2\left\{\frac{[\lambda\sigma_1^2+(1-\lambda)\sigma_2^2]^{-\frac{3}{2}}}{[\lambda(1-\lambda)]^{-\frac{1}{2}}}\right\} + O(N^{-r-1/2}).
\end{aligned} \tag{A.1.16}$$

Placing (A.1.16) back to (A.1.12), we have

$$\begin{aligned}E(W_N) =&E\left\{\sqrt{N}[\frac{\partial g(U)'}{\partial U}(X-U)+\frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)]\right\}\\ =&-\frac{1}{2\sqrt{N}}\left\{\frac{\sigma_1^3\gamma_1-\sigma_2^3\gamma_2}{\lambda(1-\lambda)}\right\}\left\{\frac{\sigma_1^2}{1-\lambda}+\frac{\sigma_2^2}{\lambda}\right\}^{-3/2}+O(N^{-r-1/2})\\ =&-\frac{1}{2\sqrt{N}}[\lambda(1-\lambda)]^{1/2}(\gamma_1-\gamma_2)+O(N^{-r-1/2}),\end{aligned} \tag{A.1.17}$$

where the last equality is due to $\sigma_1=\sigma_2$.

Next, we will compute $E(W_N^2)$:

$$\begin{aligned}E(W_N^2) &= E\left\{N\{\frac{\partial g(U)'}{\partial U}(X-U)+\frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)\}^2\right\}\\ &= NE\left\{[(G1)+(G2)+(G3)]\right\},\end{aligned} \tag{A.1.18}$$

where

$$\begin{aligned}(G1) &= \left\{\tfrac{\partial g(U)'}{\partial U}(X-U)\right\}^2,\\ (G2) &= \{\tfrac{\partial g(U)'}{\partial U}(X-U)\}\{(X-U)'\tfrac{\partial^2 g(U)}{\partial U^2}(X-U)\},\\ (G3) &= \tfrac{1}{4}\left\{(X-U)'\tfrac{\partial^2 g(U)}{\partial U^2}(X-U)\right\}^2.\end{aligned}$$

The expectations $E(G1)$, $E(G2)$ and $E(G3)$ are

$$\begin{aligned}E(G1) &= \frac{\sigma_1^2}{n_1(P+Q)}+\frac{\sigma_2^2}{n_2(P+Q)},\\ E(G2) &= \frac{-P\sigma_1^2}{n_1^2(P+Q)^2}(\tau_1-1)-\frac{Q\sigma_2^2}{n_2^2(P+Q)^2}(\tau_2-1)=O(N^{-2}),\\ E(G3) &= O(N^{-2}).\end{aligned} \tag{A.1.19}$$

Based on the equations in (A.1.19) and (A.1.18), we can get

$$\begin{aligned}E(W_N^2) &= N\left\{[E(G1)+E(G2)+E(G3)]\right\}\\ &= N\left\{\frac{\sigma_1^2}{n_1(P+Q)}+\frac{\sigma_2^2}{n_2(P+Q)}+O(N^{-2})\right\}\\ &= 1+O(N^{-1}).\end{aligned} \tag{A.1.20}$$

Finally, we will calculate $E(W_N^3)$:

$$\begin{aligned}E(W_N^3) &= E\left\{N^{3/2}\{\frac{\partial g(U)'}{\partial U}(X-U)+\frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)\}^3\right\}\\ &= N^{3/2}E\left\{[(D1)+(D2)+(D3)+(D4)]\right\},\end{aligned} \tag{A.1.21}$$

where

$$(D1) = \left\{ \tfrac{\partial g(U)'}{\partial U}(X-U) \right\}^3,$$
$$(D2) = \tfrac{3}{2}\{ \tfrac{\partial g(U)'}{\partial U}(X-U)\}^2\{(X-U)' \tfrac{\partial^2 g(U)}{\partial U^2}(X-U)\},$$
$$(D3) = \tfrac{3}{4}\{ \tfrac{\partial g(U)'}{\partial U}(X-U)\}\{(X-U)' \tfrac{\partial^2 g(U)}{\partial U^2}(X-U)\}^2,$$
$$(D4) = \tfrac{1}{8}\left\{(X-U)' \tfrac{\partial^2 g(U)}{\partial U^2}(X-U)\right\}^3.$$

Their expectations are

$$\begin{aligned}
E(D1) =& \frac{a_1^3\gamma_1}{n_1^2} + \frac{a_3^3\gamma_2}{n_2^2}, \\
E(D2) =& 3\left\{ a_1^2 b_{12}\frac{3\gamma_1}{n_1^2} + (a_1^2 b34 + 2a_1a_3b_{14})\frac{\gamma_2}{n_1n_2} \right\} \\
& + 3\left\{ (a_3^2 b_{12} + 2a_1a_3b_{23})\frac{\gamma_1}{n_1n_2} + a_3^2 b_{34}\frac{3\gamma_2}{n_2^2} + O(N^{-3}) \right\}, \\
E(D3) =& O(N^{-3}), \\
E(D4) =& O(N^{-3}),
\end{aligned} \tag{A.1.22}$$

where,

$$\begin{aligned}
a_1 = (\frac{\sigma_1^2}{P+Q})^{\frac{1}{2}}, \quad a_2 = 0, \quad a_3 = -(\frac{\sigma_2^2}{P+Q})^{\frac{1}{2}}, \quad a_4 = 0, \\
b_{12} = b_{21} = -\frac{P\sigma_1}{2}(P+Q)^{-\frac{3}{2}}, \quad b_{14} = b_{41} = -\frac{Q\sigma_1}{2}(P+Q)^{-\frac{3}{2}}, \\
b_{23} = b_{32} = \frac{P\sigma_2}{2}(P+Q)^{-\frac{3}{2}}, \quad b_{34} = b_{43} = -\frac{Q\sigma_2}{2}(P+Q)^{-\frac{3}{2}},
\end{aligned} \tag{A.1.23}$$

and all other b_{ij}s are zero, $i = 1,2,3,4$ and $j = 1,2,3,4$. Based on the results from (A.1.22) and (A.1.21), we have

$$\begin{aligned}
E(W_N^3) &= N^{3/2}E\{[(D1)+(D2)+(D3)+(D4)]\} \\
&= N^{3/2}E\{[(D1)+(D2)]\} + O(N^{-3}) \\
&= -\frac{1}{2\sqrt{N}}[\lambda(1-\lambda)]^{1/2}\left[(\frac{11\lambda-2}{\lambda})\gamma_1 - (\frac{9-11\lambda}{1-\lambda})\gamma_2 \right] + O(N^{-r-1/2}).
\end{aligned} \tag{A.1.24}$$

A.1.2 Proof of Corollary 3.1.2

Let K_{1N}, K_{2N} and K_{3N} be the first three cumulants of W_N. Then,

$$
\begin{aligned}
K_{1N} &= EW_N \\
K_{2N} &= EW_N^2 - (EW_N)^2 \\
K_{3N} &= E(W_N - EW_N)^3
\end{aligned} \tag{A.1.25}
$$

Plugging in the results of Corollary 3.1.1, we can get the following equations

$$
\begin{aligned}
K_{1N} &= -\frac{1}{2\sqrt{N}}[\lambda(1-\lambda)]^{1/2}(\gamma_1 - \gamma_2) + O(N^{-min(3/2, r+1/2)}), \\
K_{2N} &= 1 + O(N^{-min(1, r+1/2)}), \\
K_{3N} &= -\frac{1}{2\sqrt{N}}[\lambda(1-\lambda)]^{1/2}\left[(\frac{8\lambda-2}{\lambda})\gamma_1 - (\frac{6-8\lambda}{1-\lambda})\gamma_2\right] + O(N^{-min(3/2, r+1/2)}).
\end{aligned} \tag{A.1.26}
$$

Let $\chi_N(t)$ be the characteristic function of W_N. Then,

$$
\begin{aligned}
\chi_N(t) =& exp\left\{K_{1N}(it) + K_{2N}\frac{(it)^2}{2} + K_{3N}\frac{(it)^3}{6} + \ldots\right\} \\
=& exp\left\{K_{1N}(it) + K_{2N}\frac{(it)^2}{2} + K_{3N}\frac{(it)^3}{6} + O(N^{-min(1, r+1/2)})\right\} \\
=& exp\left(-\frac{t^2}{2}\right) exp\left\{N^{-1/2}\left(-A(it) - B(it)^3\right) + O(N^{-min(1, r+1/2)})\right\}.
\end{aligned} \tag{A.1.27}
$$

By Taylor expansion, we have

$$
\chi_N(t) = exp\left(-\frac{t^2}{2}\right) exp\left\{1 + N^{-1/2}\left(-A(it) - B(it)^3\right) + O(N^{-min(1, r+1/2)})\right\}. \tag{A.1.28}
$$

Based on the results of Hermite polynomials (Fedoryuk, 2001) and Fourier Transformation (Bochner and Chandrasekharan, 1949), the probability density function of W_N is

$$
\begin{aligned}
& f_{W_N}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\chi_N(t)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} exp\left(-\frac{t^2}{2}\right) exp\left\{1 + N^{-1/2}\left(-A(it) - B(it)^3\right)\right\} dt + O(N^{-min(1, r+1/2)}) \\
&= \phi(x) - N^{-1/2}AH_1(x)\phi(x) - N^{-1/2}BH_3(x)\phi(x) + O(N^{-min(1, r+1/2)}) \\
&= \phi(x)[1 + N^{-1/2}(3B - A)x - N^{-1/2}Bx^3] + O(N^{-min(1, r+1/2)}),
\end{aligned} \tag{A.1.29}
$$

where $H_1 = x$ and $H_3 = x^3 - 3x$. Then the cumulative distribution function of W_N is

$$\begin{aligned}
P(W_N \leq x) &= \int_{-\infty}^{\infty} f_{W_N}(x)dx + O(N^{-min(1,r+1/2)}) \\
&= \int_{-\infty}^{\infty} \phi(x) - N^{-1/2}AH_1(x)\phi(x) - N^{-1/2}BH_3(x)\phi(x)dx + O(N^{-min(1,r+1/2)}) \\
&= \Phi(x) - N^{-1/2}AH_0(x)\phi(x) - N^{-1/2}BH_2(x)\phi(x) + O(N^{-min(1,r+1/2)}) \\
&= \Phi(x) + N^{-1/2}[A + B(x^2-1)]\phi(x) + O(N^{-min(1,r+1/2)}),
\end{aligned} \tag{A.1.30}$$

where $H_0 = 1$ and $H_2 = x^2 - 1$. Since $T = W_N + O(N^{-1})$, Theorem 3.1.3 follows.

A.2 Proof of Theorem 3.2.4

The test statistic of the pooled two sample t-test under H_a is

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1^0 - \mu_2^0)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2) + \delta}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\delta = \mu_1 - \mu_2 - (\mu_1^0 - \mu_2^0)$. Similarly, with the same $X = (X_1, X_2, X_3, X_4)$ defined in Section A.1, we can express the T under H_a as

$$T = \sqrt{N}(g(X) + m(X)) = \frac{X_1 - X_3 + \delta}{k(X)^{1/2}},$$

where

$$\begin{aligned}
g(X) &= (X_1 - X_3)k(X)^{-1/2} \\
m(X) &= \delta k(X)^{-1/2},
\end{aligned}$$

and

$$k(X) = \frac{(n_1-1)S_1^{*^2} + (n_2-1)S_2^{*^2}}{(N-2)\lambda_N(1-\lambda_N)} = \frac{(n_1-1)(X_2 - X_1^2) + (n_2-1)(X_4 - X_3^2)}{(N-2)\lambda_N(1-\lambda_N)}.$$

Next, apply Taylor expansion to $g(X)$ and $m(X)$ with $EX \equiv U \equiv (U_1, U_2, U_3, U_4) = (0, 1, 0, 1)$, we have

$$\begin{aligned}
g(X) &= g(U) + \frac{\partial g(U)}{\partial U}(X-U) + \frac{1}{2}\frac{\partial^2 g(U)}{\partial U^2}(X-U)^2 + \dots \\
m(X) &= m(U) + \frac{\partial m(U)}{\partial U}(X-U) + \frac{1}{2}\frac{\partial^2 m(U)}{\partial U^2}(X-U)^2 + \dots,
\end{aligned}$$

where $g(U) = 0$. Now, we let

$$W_N = \sqrt{N}\{\frac{\partial g(U)'}{\partial U}(X-U) + \frac{1}{2}(X-U)'\frac{\partial^2 g(U)}{\partial U^2}(X-U)\}$$
$$W_\delta = \sqrt{N}\{m(U) + \frac{\partial m(U)'}{\partial U}(X-U) + \frac{1}{2}(X-U)'\frac{\partial^2 m(U)}{\partial U^2}(X-U)\}$$

Under regularity conditions in (A.1.1),

$$T = W_N + W_\delta + O(N^{-1}).$$

Assuming $EY_{ij}^6 < \infty$, the first three moments of $W_N + W_\delta$ can be obtained

$$\begin{aligned}
&E(W_N + W_\delta) = E(W_N) + c_N + O(N^{-min(1,r+\frac{1}{2})}),\\
&E(W_N + W_\delta)^2 = E(W_N)^2 + c_N^2 - \delta\sigma[\lambda(1-\lambda)](\gamma_1 - \gamma_2) + O(N^{-min(1,r+\frac{1}{2})}),\\
&E(W_N + W_\delta)^3 = E(W_N)^3 + c_N^3 + 3\delta\sigma\sqrt{N}[\lambda(1-\lambda)]^{1/2}\\
&-\frac{9}{2}\delta^2\sigma^{-2}\sqrt{N}[\lambda(1-\lambda)]^{3/2}(\gamma_1 - \gamma_2) + O(N^{-min(1,r+\frac{1}{2})}),
\end{aligned}$$

where $c_N = \delta/\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}$ and $E(W_N)^i$ are the first three moments of W_N under H_a, $i = 1, 2, 3$. They follow the same formulae (A.1.3) $-$ (A.1.5) as in the case of H_0. Based on the first three moments of $W_N + W_\delta$, the first three cumulants of $W_N + W_\delta$ can be obtained as follows

$$\begin{aligned}
&K_{1N}^{H_a} = K_{1N} + c_N + O(N^{-min(1,r+\frac{1}{2})}),\\
&K_{2N}^{H_a} = E(W_N + W_\delta)^2 - E^2(W_N + W_\delta) = K_{2N} + q + O(N^{-min(1,r+\frac{1}{2})}),\\
&K_{3N}^{H_a} = E(W_N + W_\delta - E(W_N + W_\delta))^3 = K_{3N},
\end{aligned}$$

where $q_N = \delta\sigma^{-1}[\lambda(1-\lambda)](\gamma_1 - \gamma_2)$ and K_{iN} are the first three cumulants of W_N under H_0, $i = 1, 2, 3$. Let $\chi_N(t)^{H_a}$ be the characteristic function of $W_N + W_\delta$. Then,

$$\begin{aligned}
\chi_N(t)^{H_a} &= exp\left\{K_{1N}^{H_a}(it) + K_{2N}^{H_a}\tfrac{(it)^2}{2} + K_{3N}^{H_a}\tfrac{(it)^3}{6} + \ldots\right\}\\
&= exp\left\{(K_{1N} + c_N)(it) + (K_{2N} + q)\tfrac{(it)^2}{2} + K_{3N}\tfrac{(it)^3}{6}\right\} + O(N^{-min(1,r+\frac{1}{2})}).
\end{aligned}$$

Apply Taylor expansion and result in (A.1.28), we have

$$\chi_N(t)^{H_a} = exp\left(itc_N - \frac{t^2}{2}\right)\left\{1 + N^{-1/2}\left(-A(it) - B(it)^3\right) + \frac{t^2 q}{2} + O(N^{-min(1,r+\frac{1}{2})})\right\}.$$

Based on the results of Hermite polynomials (Fedoryuk, 2001) and Fourier Transformation (Bochner and Chandrasekharan, 1949), the probability density function of $W_N + W_\delta$ is

$$f_{W_N+W_\delta}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\chi_N(t)^{H_a}dt =$$
$$\phi(x-c_N)\left[1 - N^{-1/2}AH_1(x-c_N) - H_2(x-c_N)\frac{q_N}{2} - N^{-1/2}BH_3(x-c_N)\right]$$
$$+O(N^{-min(1,r+1/2)}),$$

where $H_1 = x$, $H_2 = x^2 - 1$ and $H_3 = x^3 - 3x$. Then the cumulative distribution function of $W_N + W_\delta$ is

$$P(W_N + W_\delta \leq u) = \int_{-\infty}^{u} f_{W_N+W_\delta}(x)dx + O(N^{-min(1,r+1/2)})$$
$$= \Phi(u-c) + N^{-1/2}[A + B((u-c)^2 - 1)]\phi(x-c) + \tfrac{q}{2}(u-c)\phi(u-c)$$
$$+O(N^{-min(1,r+1/2)}).$$

Since $T = W_N + W_\delta + O(N^{-1})$, Theorem 3.2.4 follows.

A.3 Proof of Corollary 3.2.5

By subtracting the two power functions in (3.2.17) and (3.2.19), we can get

Power of TCF test - Power of TN test

$$= -Q(L_u^{cf}) + Q(L_l^{cf}) + Q(U_{N,1-\alpha/2}) - Q(U_{N,\alpha/2}) + L_{N,\gamma_1,\gamma_2,\lambda} + O(N^{-min(1,r+1/2)}).$$

Thus, if we can prove

$$-Q(L_u^{cf}) + Q(L_l^{cf}) + Q(U_{N,1-\alpha/2}) - Q(U_{N,\alpha/2}) = O(N^{-1}),$$

then we are done. Note that

$$Q(L_u^{cf}) = \frac{q_N}{2}(U_{N,1-\alpha/2} + \Delta_{N,1-\alpha/2})\phi(U_{N,1-\alpha/2} + \Delta_{N,1-\alpha/2}).$$

Since $\Delta_{N,1-\alpha/2} = -N^{-1/2}[A + B((t^{cf}_{1-\alpha/2})^2 - 1)] = O(N^{-1/2})$ and $q_N = O(N^{-1/2})$, we have

$$
\begin{aligned}
Q(L_u^{cf}) &= \frac{q_N}{2}(U_{N,1-\alpha/2} + O(N^{-1/2}))\phi(U_{N,1-\alpha/2} + O(N^{-1/2})) \\
&= \frac{q_N}{2}[(U_{N,1-\alpha/2})\phi(U_{N,1-\alpha/2}) + O(N^{-1/2})] \\
&= \frac{q_N}{2}(U_{N,1-\alpha/2})\phi(U_{N,1-\alpha/2}) + O(N^{-1}) \\
&= Q(U_{N,1-\alpha/2}) + O(N^{-1})
\end{aligned}
$$

Thus, we have

$$Q(L_u^{cf}) - Q(U_{N,1-\alpha/2}) = O(N^{-1}).$$

Similarly, we can also show that

$$Q(L_l^{cf}) - Q(U_{N,\alpha/2}) = O(N^{-1}).$$

Therefore,

$$-Q(L_u^{cf}) + Q(L_l^{cf}) + Q(U_{N,1-\alpha/2}) - Q(U_{N,\alpha/2}) = O(N^{-1}).$$

Then the results in Corollary 3.2.5 holds.

A.4 Proof of Theorem 4.1.1

Proof of Theorem 4.1.1. Under H_0,

$$
\begin{aligned}
P(T^{'} \leq x) &= P(T + N^{-1/2}\hat{A} \leq x) = P(T \leq x - N^{-1/2}\hat{A}) \\
&= \Phi(x - N^{-1/2}\hat{A}) + N^{-1/2}[A + B((x - N^{-1/2}\hat{A})^2 - 1)] \\
&\phi(x - N^{-1/2}\hat{A}) + O(N^{-min(1,r+1/2)}).
\end{aligned} \tag{A.4.1}
$$

Applying Taylor expansion at x, we can calculate the probability as

$$\begin{aligned}
&\Phi(x - N^{-1/2}\hat{A}) + N^{-1/2}[A + B((x - N^{-1/2}\hat{A})^2 - 1)]\phi(x - N^{-1/2}\hat{A}) + O(N^{-min(1,r+1/2)})\\
=&[\Phi(x) + \phi(x)(-N^{-1/2}\hat{A}) + O(N^{-1})] + N^{-1/2}[A + B(x^2 - N^{-1/2}\hat{A}x - 1 + O(N^{-1})]\\
&[\phi(x) + \phi^{'}(x)(-N^{-1/2}\hat{A}) + O(N^{-1})] + O(N^{-min(1,r+1/2)})\\
=&\Phi(x) + N^{-1/2}[-\hat{A} + \hat{A} + \hat{B}(x^2 - 1) + O(N^{-1/2})]\phi(x) + O(N^{-1}) + O(N^{-min(1,r+1/2)})\\
=&\Phi(x) + N^{-1/2}\hat{B}(x^2 - 1)\phi(x) + O(N^{-1}) + O(N^{-min(1,r+1/2)})\\
=&\Phi(x) + N^{-1/2}\hat{B}(x^2 - 1)\phi(x) + O(N^{-min(1,r+1/2)}),
\end{aligned}$$

then the distribution of $P(T^{'} \leq x)$ follows the same form as that in Hall (1992b) and Zhou and Philip (2005). The proof under H_a is similar and is thus omitted.

A.5 Proof of Corollary 3.2.6

Proof of Corollary 3.2.6:

We know that

$$L_N = \Delta_{N,\alpha/2}[\phi(U_{N,\alpha/2}) - \phi(U_{N,1-\alpha/2})] + O(N^{-1}),$$

where $\phi(U_{N,\alpha/2}) < \phi(U_{N,1-\alpha/2})$. So $L_N > 0$ if and only if $\Delta_{N,\alpha/2} < 0$. Note that

$$\Delta_{N,\alpha/2} = -N^{-1/2}[A + B(z^2_{\alpha/2} - 1)].$$

So $\Delta_{N,\alpha/2} < 0$ if and only if $A + B(z^2_{\alpha/2} - 1) > 0$. Recall that

$$\begin{aligned}
A &= [\lambda(1-\lambda)]^{1/2}(\gamma_1 - \gamma_2)/2,\\
B &= [\lambda(1-\lambda)]^{1/2}\left(\frac{8\lambda - 2}{\lambda}\gamma_1 - \frac{6 - 8\lambda}{1-\lambda}\gamma_2\right)/12.
\end{aligned}$$

Plugging A and B into $A + B(z^2_{\alpha/2} - 1) > 0$, we have

$$\begin{aligned}
&\frac{8\lambda - 2}{\lambda}\gamma_1 - \frac{6 - 8\lambda}{1-\lambda}\gamma_2 > (\gamma_2 - \gamma_1)\frac{6}{z^2_{\alpha/2} - 1}\\
\Leftrightarrow&(8 + \frac{6}{z^2_{\alpha/2} - 1} - \frac{2}{\lambda})\gamma_1 > (8 + \frac{6}{z^2_{\alpha/2} - 1} - \frac{2}{1-\lambda})\gamma_2\\
\Leftrightarrow&(8 + c - \frac{2}{\lambda})\gamma_1 > (8 + c - \frac{2}{1-\lambda})\gamma_2,
\end{aligned}$$

where $c = \frac{6}{z_{\alpha/2}^2 - 1}$. Then corollary 3.2.6 holds.

A.6 Proof of Corollary 5.1.5

Proof of Corollary 5.1.5:

We know that

$$L'_{(N,\gamma_1,\gamma_2,\lambda)} = \Delta'_{N,\alpha/2}[\phi(U'_{N,\alpha/2}) - \phi(U'_{N,1-\alpha/2})] + O(N^{-1}),$$

where $\phi(U'_{N,\alpha/2}) < \phi(U'_{N,1-\alpha/2})$. So $L'_{(N,\gamma_1,\gamma_2,\lambda)} > 0$ if and only if $\Delta'_{N,\alpha/2} < 0$. Note that

$$\Delta'_{N,\alpha/2} = -\frac{A'}{6\sqrt{N}}(2z_{\alpha/2}^2 + 1).$$

So $\Delta'_{N,\alpha/2} < 0$ if and only if $A'(2z_{\alpha/2}^2 + 1) > 0$. Recall that

$$A' = \left\{\frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1-\lambda}\right\}^{-3/2}\left\{\frac{\sigma_1^3\gamma_1}{\lambda^2} - \frac{\sigma_2^3\gamma_2}{(1-\lambda)^2}\right\}.$$

Plugging A' into $A'(2z_{\alpha/2}^2 + 1) > 0$, we have

$$\begin{aligned}
&\left\{\frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1-\lambda}\right\}^{-3/2}\left\{\frac{\sigma_1^3\gamma_1}{\lambda^2} - \frac{\sigma_2^3\gamma_2}{(1-\lambda)^2}\right\}(2z_{\alpha/2}^2 + 1) > 0 \\
\Leftrightarrow &\left\{\frac{\sigma_1^3\gamma_1}{\lambda^2} - \frac{\sigma_2^3\gamma_2}{(1-\lambda)^2}\right\}(2z_{\alpha/2}^2 + 1) > 0 \\
\Leftrightarrow &\left\{(1-\lambda)^2\sigma_1^3\gamma_1 - \lambda^2\sigma_2^3\gamma_2\right\}(2z_{\alpha/2}^2 + 1) > 0 \\
\Leftrightarrow &c'_\alpha(1-\lambda)^2\sigma_1^3\gamma_1 > c'_\alpha\lambda^2\sigma_2^3\gamma_2,
\end{aligned}$$

where $c'_\alpha = 2z_{\alpha/2}^2 - 1$. Then corollary 5.1.5 holds.

A.7 Proof of Results 4.3.19

We want to prove

$$-Q(L_u^{(i)}) + Q(L_l^{(i)}) + Q(L_u^{cf}) - Q(L_l^{cf}) = O(N^{-1}).$$

Recall that

$$T_i^{-1}(t) = t - N^{-1}b\hat{\gamma} + O(N^{-m}) \text{ and } \hat{\gamma} = O(N^{-1/2}),$$

where $m \geq 1$ and

$$\begin{aligned} L_u^{(i)} = t^i_{1-\alpha/2} - c_N = \sqrt{N}T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}) - c_N \\ L_l^{(i)} = t^i_{\alpha/2} - c_N = \sqrt{N}T_i^{-1}(\frac{z_{\alpha/2}}{\sqrt{N}}) - c_N. \end{aligned}$$

We have

$$\begin{aligned} Q(L_u^{(i)}) \quad &= \frac{q_N}{2}\phi\left(\sqrt{N}T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}) - c_N\right)\left(\sqrt{N}T_i^{-1}(\frac{z_{1-\alpha/2}}{\sqrt{N}}) - c_N\right) \\ &= \frac{q_N}{2}\phi\left(\sqrt{N}(\frac{z_{1-\alpha/2}}{\sqrt{N}} - N^{-1}b\hat{\gamma}) - c_N\right)\left(\sqrt{N}(\frac{z_{1-\alpha/2}}{\sqrt{N}} - N^{-1}b\hat{\gamma}) - c_N\right) \\ &= \frac{q_N}{2}\phi(z_{1-\alpha/2} - c_N + O(N^{-1/2}))(z_{1-\alpha/2} - c_N + O(N^{-1/2})) \\ &= \frac{q_N}{2}\phi(z_{1-\alpha/2} - c_N)(z_{1-\alpha/2} - c_N) + O(N^{-1}) \\ &= \frac{q_N}{2}\phi(U_{N,1-\alpha/2})(U_{N,1-\alpha/2}) + O(N^{-1}) \\ &= Q(U_{N,1-\alpha/2}) + O(N^{-1}) \end{aligned}$$

Similarly, we can also show that

$$Q(L_l^{(i)}) = Q(U_{N,\alpha/2}) + O(N^{-1}).$$

In Section A.3 we have proved that

$$\begin{aligned} Q(L_u^{cf}) = Q(U_{N,1-\alpha/2}) + O(N^{-1}) \\ Q(L_l^{cf}) = Q(U_{N,\alpha/2}) + O(N^{-1}). \end{aligned}$$

Then the result in equation (4.3.19) holds.