Quantile Estimation based on the Almost Sure Central Limit Theorem

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch–Naturwissenschaftlichen Fakultäten
der Georg–August–Universität zu Göttingen

vorgelegt von
Karthinathan Thangavelu
aus
Erode, Indien

Göttingen 2005

D7

Referent	Prof. Dr. Edgar Brunner
Korreferent	Prof. Dr. Manfred Denker
Tag der mündlichen Prüfung	25 Januar 2006

To the Almighty God

Srimad Bhagavad Gita (Chapter 8, Verse 7):

sri-bhagavan uvaca:

tasmat sarveshu kaleshu
mam anusmara yudhya ca
mayy arpita-mano-buddhir
mam evaishyasy asamsayah.

Translation:

Lord Sri Krishna says to Arjuna, "Therefore, Arjuna, you should always think of Me in the form of Krishna and at the same time carry out your prescribed duty of fighting (for success). With your activities dedicated to Me and your mind and intelligence fixed on Me, you will attain Me without doubt".

Acknowledgements

I would like to take this opportunity to express my thanks to Prof. Dr. Edgar Brunner and Prof. Dr. Manfred Denker (my thesis advisors), both of whom guided me meticulously through the research work. Not only did they guide me on scientific issues, but they were also very understanding and cooperative. I also thank Prof. Dr. Walter Zucchini for his encouraging support and comments.

I acknowledge the financial support from the *Lichtenberg Stipendium*.

I would also like to thank my colleagues in the the Center for Statistics, in general and, Departments of Medical Statistics and Genetic Epidemiology, in particular, for providing me with an excellent and friendly work environment. Special mention is also due to Dr. Aleksey Min, who guided me through some of the important mathematical aspects of the project.

My family members in India, and friends and house mates in Göttingen were also supportive.

I express my deepest gratitude to Sri P. M. Nachimuthu Mudaliyar (my late grandfather) for his great inspiration, encouragement and support, all of which I will cherish throughout my life.

Finally, I thank the Almighty God, for all the grace and blessing that He has showered on me to reach this stage of my academic and personal life!

Göttingen, 2005 **Karthinathan Thangavelu**

Contents

Chapter 1

Introduction

Statistics is the theory of decision making when the probabilistic model is unknown. The theory as it stands today, was developed in the last century and is based on a statistical problem $(E, \mathcal{B}, P_\theta \{\theta \in \Theta\})^\dagger$ and a decision, termed estimation or hypothesis testing. In both cases, the decision is based on quantiles of the unknown distribution, hence estimation of these quantiles is the most important issue in the theory. On the basis of these quantiles one can calculate the error probability for the decision. The main issue of the present dissertation is to introduce a new method for estimating quantiles. It is based on the Almost Sure Central Limit Theorem.

The Almost Sure Central Limit Theorem (ASCLT) was first presented independently by Fisher (1987), Schatte (1988) and Brosamler (1988). The classical central limit theorem says that for an i.i.d. L^2-sequence of random variables X_i with expectation 0 and variance 1, the distribution on $\frac{X_1+\ldots+X_n}{\sqrt{n}}$ converges weakly to the standard normal distribution represented by Φ. The ASCLT states that

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} 1_{\left\{X_1+\ldots+X_n < t\sqrt{n}\right\}} = \Phi(t) \qquad \text{a.s.} \tag{1.1}$$

One motivation to this type of theorem comes from Brownian motion $B(t)$. Note that Brownian motion on $\mathbb{R}_+$ has the property that $\frac{1}{\sqrt{s}} B(st)$, $(t \geq 0)$ is the same Brownian

†Here, $(E, \mathcal{B})$ is a sample space and $P_\theta \{\theta \in \Theta\}$ is a family of probability measures. For basic definitions relating to statistical decision theory, we refer to Strasser (1985).

motion for any $s > 0$, in the sense of distributions. Therefore the maps $g_s : C(\mathbb{R}_+) \to C(\mathbb{R}_+)$ defined by $g_s(f)(t) = \frac{1}{\sqrt{s}} f(st)$ define a flow G_s ($s \in \mathbb{R}$) by $G_s f = g_{e^s}(f)$. This flow has an invariant measure given by the Wiener measure P. It is known that it is ergodic. Hence by the ergodic theorem, for any measurable $h : C(\mathbb{R}_+) \to \mathbb{R}$,

$$\frac{1}{T}\int_0^T h(G_s(\cdot))\,\mathrm{d}\,s \quad\to\quad \int h(f)\,\mathrm{d}\,P(f) \qquad \text{a.s.}$$

Making the variable transform from $\tau = e^s$ and $S = e^T$, we arrive at

$$\frac{1}{\log S}\int_1^S \frac{1}{\tau} h(g_\tau(\cdot))\,\mathrm{d}\,\tau \quad\to\quad \int h(f)\,\mathrm{d}\,P(f).$$

Now take $h(f) = 1_{(-\infty,\,t]} \circ f(1)$, to obtain

$$\begin{aligned} \frac{1}{\log S}\int_1^S \frac{1}{\tau} 1_{(-\infty,\,t]} \circ (g_\tau(\cdot)(1))\,\mathrm{d}\,\tau &\to \int h(f)\,P(\mathrm{d}\,f) \\ &= E(h(B)) = P(B(1) \le t) \\ &= \Phi(t). \end{aligned}$$

The discrete version of this is exactly of the form (1.1).

Another aspect of the ASCLT method we would like to address here is the possibility of making new decision procedures. This may be done like in quality control methods/procedures when constantly observed data forces decision when crossing a given quality level. Note that the classical theories are based on the facts from distribution theory while our proposed approach is using the almost sure concept which permits extension of data even when knowing the past. This is a variant of the sequential testing.

Further, we note that all of the results concerning the theorem are asymptotic in nature and are based on logarithmic averages. Thus the rate of convergence (proposed in the theorem) is very slow. Due to this, the general application of the newly proposed methods of hypothesis testing in data analysis, particularly for data from biological and medical experiments, would be nearly impossible since these data are usually characterized by very small samples sizes. Thus, we intend to also propose small-sample approximations to the corresponding asymptotic results presented.

The proposed hypothesis testing methods have several good properties. These will be discussed in the respective chapters. One of the key properties of these methods would be that estimation or use of variance of the observations will never be done. This has important implications in practical data analysis situations. Through this thesis, we are thus opening the path of research with two aspects - making almost sure decisions and variance(-estimation)-free direct method of estimating the limiting distribution of the statistics.

Due to the nature of the new approach presented, there are several open and unsolved problems arising out of the proposals. Thus we also intend to present such problems and challenges as and when they arise naturally and follow in an intuitive manner.

Also, in the literature, results can be found regarding the ASCLT for several types of statistics. For example, Berkes and Csáki (2001) and later Holzmann et al. (2004) present ASCLT for U-Statistics. In our work, we will state and prove the ASCLT for rank statistics. Rank statistics form the foundation for several nonparametric methods and a short introduction to this important class of statistics is presented. We will also state some results from the literature which aid the proof of the theorem.

In order to evaluate the performance of the proposed tests, we apply them in both parametric as well as nonparametric test situations. Another main aspect of the thesis is that, we discuss in detail about the famous Behrens-Fisher Problem, which was first discussed by German researcher Behrens in 1929 and then pursued by Fisher in later years. We discuss several commonly used solutions for the problem and also present some information on associated software packages available to implement them. We also propose the new solutions for the Behrens-Fisher problem based on the ASCLT from the viewpoint of small-sample approximation.

Chapter 2

Hypothesis Testing based on ASCLT

Introducing a new approach of thinking and handling the statistical inferential methods, particularly testing of hypothesis, is one of the fundamental aims of this thesis. For a review of the underlying theoretical principles, ideas and methods of hypothesis testing, we refer to standard books by Kendall and Stuart (1973) and Lehmann (1986) and, for a more intuitive and applied approach towards hypothesis testing, the recent book by Casella and Berger (2002).

The mathematical foundations of the theory of hypothesis testing based on the Almost Sure Central Limit Theorems will be laid in this chapter. Asymptotic results and the general procedures and proposals for hypothesis testing will also be presented here. These results will be used in the Chapter 4 to develop tests for specific situations and also situation-specific small-sample approximation procedures will be presented there.

2.1 Introduction to ASCLT

The ASCLT was first introduced in literature independently by Fisher (1987), Brosamler (1988) and Schatte (1988). The work by Fisher (in 1987) in Göttingen presented the theorem from the point of view of the ergodic theory, as explained briefly in Chapter 1. The theorem proposed by these authors extended the classical central limit theorem to an *almost sure* (or *pointwise*) version and hence the name *Almost Sure Central Limit Theorem.* For brevity, we will use 'ASCLT' to represent 'Almost Sure Central Limit Theorem'. The version of the theorem as introduced by Fisher (1987), Brosamler (1988) and Schatte (1988) is presented below.

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $S_k = X_1 + \cdots + X_k$, $1 \leq k \leq n$ being the partial sums. If $EX_1 = 0, EX_1^2 = 1$ and $E|X_1|^{2+\delta}$ is finite for some $\delta > 0$ then,*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} 1_{\left\{\frac{S_k}{\sqrt{k}} < x\right\}} = \Phi(x) \qquad a.s. \text{ for any } x, \tag{2.1}$$

where Φ is the standard normal distribution function and $1_{\{A\}}$ is the indicator function of the set A.

In the above theorem, Schatte (1988) assumed $\delta = 1$. It can be noted that a similar result was stated (without proof) in p. 270 of Lévy (1937).

Following the above discovery, during the past decade and half, there were many interesting developments of limit theorems involving log averages and log densities. Several authors have investigated the ASCLT for independent random variables, e.g., Atlagh (1993), Atlagh and Weber (1992) and Berkes and Dehling (1993). Recently, Berkes and Csáki (2001) discuss several examples of applications of ASCLT, for e.g., limit theorems for extrema, distribution of local times, U-Statistics, etc. Holzmann et al. (2004) and Min (2004) also present the ASCLT for U-statistics. For detailed survey and discussion on the papers relating to ASCLT, we refer to Berkes (1998) and Atlagh and Weber (2000).

We will not go into the details and discussion surrounding the literature on ASCLT, as most of them treat the theorem from a pure mathematical perspective. Whereas, our interest lies more in using the standard version of the theorem in order to develop hypothesis testing procedure(s). For this purpose we will use the result of the following form presented by Berkes and Dehling (1993).

Theorem 2 (Berkes and Dehling, 1993). *Let $X_1, X_2, \ldots$ be independent random variables and a_n, $b_n > 0$ numerical sequences such that setting $S_n = X_1 + \cdots + X_n$, we have*

$$Ef\left(\left|\frac{S_n - a_n}{b_n}\right|\right) \leq (\log\log n)^{-1} f(e^{(\log n)^{1-\epsilon}}), \qquad n \geq n_0 \tag{2.2}$$

for some $\epsilon > 0$ *where* $f \geq 0$ *is a Borel measurable function on* $(0, \infty)$ *such that both* $f(x)$ *and* $x/f(x)$ *are eventually nondecreasing and the right-hand side of (2.2) is non-decreasing for* $n \geq n_0$. *Assume also that*

$$b_l/b_k \geq C(l/k)^\gamma, \qquad l \geq k \tag{2.3}$$

for some constants $C > 0, \gamma > 0$. *Then for any distribution function* G, *the following statements are equivalent:*

- *For any Borel set* $A \subset \mathbb{R}$ *with* $G(\delta A) = 0$ *we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} 1_{\left\{\frac{S_k - a_k}{b_k} \in A\right\}} = G(A) \qquad a.s., \tag{2.4}$$

  where the exceptional set of probability zero is independent of A.

- *For any Borel set* $A \subset \mathbb{R}$ *with* $G(\delta A) = 0$ *we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} P\left\{\frac{S_k - a_k}{b_k} \in A\right\} = G(A). \tag{2.5}$$

2.2 Hypothesis Testing, Quantiles and Random Intervals

As briefly explained in the Chapter Introduction, our main proposal towards the ASCLT-based theory of hypothesis testing will be via the estimation of the quantiles of the distribution of the concerned statistic(s). First, the results concerning the estimation of the quantiles will be explained, and then two methods of testing hypothesis based on the estimated quantiles will be described. Most of the developments here will be addressing a setting of an one-sample situation. These results can be generalised to other complex situations, though some care and mathematical thinking would be involved in doing so. Some discussion in that direction would be considered when dealing with the situation of a general two-sample testing problem in Chapter 4.

Notation and Assumptions

For $n \geq 1 (n \in \mathbb{N})$, let T_n be a sequence of real-valued statistics defined on some measurable space $(\Omega, \mathcal{B})$ and $\mathcal{P}$ be a family of probabilities on $\mathcal{B}$. Also let, $E\left(T_n\right) = n\mu\left(P\right)$ for $P \in \mathcal{P}$, where $\mu(P) \in \mathbb{R}$ is unknown. We assume that T_n satisfies the Central Limit Theorem(CLT) and the ASCLT for each $P \in \mathcal{P}$ with constants $b_n = n^{-1/2}$, $a_n(P) = n\mu(P)$ and distribution function G_P, where G_P is unknown, for example, Normal $\mathcal{N}(\mu, \sigma^2)$, where μ and σ as unknown. That is,

$$P\left(\{\omega \in \Omega : \; b_n(T_n(\omega) - a_n(P)) \leq t\}\right) \longrightarrow G_P(t), \qquad \text{for } t \in C_G \tag{2.6}$$

and

$$\frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} 1_{\{b_n(T_n - a_n(P)) \leq t\}} \longrightarrow G_P(t) \qquad P - a.s., \;\; \text{for } t \in C_G, \tag{2.7}$$

where C_G denotes the set of continuity points of G_P. We would like to make the following remark with reference to the above equation (2.7).

Remark: For sufficiently large value of $t \in \mathbb{R}$ in (2.7) such that $1_{\{b_n(T_n - a_n(P)) \leq t\}} \equiv 1$, $\forall n \leq N$, the LHS of the equation will be of the form:

$$\frac{\sum_{n=1}^{N} \frac{1}{n}}{\log N}.$$

This fraction should be (and is expected to be) equal to 1. But even for very large values of N, this is not the case. For examples, for $N = 10^2$, the above ratio is approximately 1.13, and for $N = 10^{10}$, it is approximately 1.025. For any application in statistics, we will seldom come across a situation with sample size of $N = 10^{10}$. Hence,

$$\frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} 1_{\{b_n(T_n - a_n(P)) \leq t\}}$$

will not be a distribution function for even very large values of N. Thus, in the sequel we propose to use directly the averaging term $\sum_{n=1}^{N} \frac{1}{n}$ instead of "$\log N$" in formulae of the form 2.7. Further, for convenience, we denote $C_N = \sum_{n=1}^{N} \frac{1}{n}$.

Consequently, the following two functions are now defined for each $\omega \in \Omega$ and $t \in \mathbb{R}$,

$$\begin{aligned}
\widetilde{G}_N(t,\omega) &= C_N^{-1}\sum_{n=1}^{N}\frac{1}{n}1_{\{b_n(T_n-a_n(P))\le t\}} \\
&= C_N^{-1}\sum_{n=1}^{N}\frac{1}{n}1_{(-\infty,t]}\left(b_n(T_n-a_n(P))\right) \qquad (2.8)
\end{aligned}$$

and

$$\widehat{G}_N(t,\omega) = C_N^{-1}\sum_{n=1}^{N}\frac{1}{n}1_{\{b_nT_n\le t\}} = C_N^{-1}\sum_{n=1}^{N}\frac{1}{n}1_{(-\infty,t]}\left(b_nT_n\right). \qquad (2.9)$$

In the sequel we will be presenting results for fixed $\omega \in \Omega$, though all the results would be applicable to each $\omega \in \Omega$. Thus, for simplicity we will denote the functions defined in (2.8) and (2.9) as $\widetilde{G}_N(t,\omega)$ by $\widetilde{G}_N(t)$ and $\widehat{G}_N(t,\omega)$ by $\widehat{G}_N(t)$, respectively. Similarly, we will also denote $\mu\left(P\right)$ and $a_n\left(P\right)$ simply by μ and a_n, respectively, since the results will hold true for every $P \in \mathcal{P}$.

Now some properties and results establishing the relationship of the two functions defined above in (2.8) and (2.9), with the true distribution G_P are presented below.

Lemma 3. *$\widetilde{G}_N$ and $\widehat{G}_N$ are empirical distribution functions. Moreover, $\widetilde{G}_N\left(t\right)$ converges $G_P\left(t\right)$ a.s. for every $t \in C_G$.*

Proof. Let us first consider $\widetilde{G}_N$.

Now for, $t < s \in \mathbb{R}$, it is clear that,

$$1_{(-\infty,t]}\left(x\right) \le 1_{(-\infty,s]}\left(x\right) \text{ for x} \in \mathbb{R}.$$

This implies that, $\widetilde{G}_N(t) \le \widetilde{G}_N(s)$ for $n \le N$, $N \in \mathbb{N}$ fixed. Thus the function is monotonically increasing in $t \in \mathbb{R}$.

We also observe that,

$$\lim_{t\to-\infty} 1_{(-\infty,t]}\left(b_n(T_n-a_n)\right) = 0 \quad \Longrightarrow \quad \lim_{t\to-\infty}\widetilde{G}_N(t) = 0$$

and

$$\lim_{t\to\infty} 1_{(-\infty,t]}\left(b_n(T_n - a_n)\right) = 1 \quad \Longrightarrow \quad \lim_{t\to\infty} \widetilde{G}_N(t) = 1.$$

Further we note that the function $\widetilde{G}_N$ is a step function in t and it ranges between $[0,1]$. Finally, since the function $\widetilde{G}_N(t)$ has constant values for $t \in (t_{i-1}, t_i]$ for each $i = 2, \ldots, s$, and $\widetilde{G}_N(t) \equiv 0$ for $t \in (-\infty, t_1]$ and $\widetilde{G}_N(t) \equiv 1$, for $t \in (t_s, \infty)$, it is clear that it is left continuous in $t \in \mathbb{R}$. Thus $\widetilde{G}_N$ is an empirical distribution function.

Similarly, observing that $\widehat{G}_N$ is a special case of $\widetilde{G}_N$ with $a_n \equiv 0$, all the above arguments hold true for $\widehat{G}_N$ and thus it is also an empirical distribution function.

The result of $\widetilde{G}_N(t)$ converging $G_P(t)$ a.s. $\forall t \in C_G$, is a special case of the next theorem. So the proof follows from there. $\square$

The next theorem establishes the relation between $\widetilde{G}_N$ and G_P.

Theorem 4 (Glivenko-Cantelli). *We have that,*

$$\lim_{N\to\infty} \sup_{t\in\mathbb{R}} \left|\widetilde{G}_N(t) - G_P(t)\right| = 0 \qquad\qquad a.s.$$

Proof. Let $\epsilon > 0$. Choose t's $\in \mathbb{R}$ such that, $-\infty < t_1 < t_2 < \ldots < t_s < \infty$ so that,

$$\begin{aligned} G_P(t_1) &< \epsilon \\ G_P(t_k) - G_P(t_{k-1}) &< \epsilon,\ k = 2, 3, \ldots, s \\ 1 - G_P(t_s) &< \epsilon. \end{aligned}$$

Now, due to (2.7) and the fact that $\frac{C_N}{\log N} \to 1$, $\exists N_0 \in \mathbb{N}$ such that $\forall N \geq N_0$, $|\widetilde{G}_N(t_i) - G_P(t_i)| \leq \epsilon,\ i = 1, \ldots, s$. We now prove the result for general t, such that $t_{i-1} < t < t_i$ for some $i \geq 2$, or $t < t_1$ or $t > t_s$. Consider,

$$\begin{aligned} \left|\widetilde{G}_N(t) - G_P(t)\right| &= \begin{cases} \widetilde{G}_N(t) - G_P(t), & \widetilde{G}_N(t) > G_P(t) \\ 0, & \widetilde{G}_N(t) = G_P(t) \\ G_P(t) - \widetilde{G}_N(t), & \widetilde{G}_N(t) < G_P(t) \end{cases} \\ &\leq \begin{cases} \widetilde{G}_N(t_i) - G_P(t_i) + G_P(t_i) - G_P(t), & \widetilde{G}_N(t) > G_P(t) \\ 0, & \widetilde{G}_N(t) = G_P(t) \\ G_P(t) - G_P(t_{i-1}) + G_P(t_{i-1}) - \widetilde{G}_N(t_{i-1}), & \widetilde{G}_N(t) < G_P(t) \end{cases} \end{aligned}$$

$$\leq \quad \left\{ \begin{array}{l} \epsilon + G_P(t_i) - G_P(t_{i-1}) \\ 0 \\ G_P(t_i) - G_P(t_{i-1}) + \epsilon \end{array} \right\} \leq \; 2\epsilon$$

□

The above Lemma establishes a version of the Glivenko-Cantelli theorem with respect to the empirical distribution functions under our considerations. Such results have been also presented in literature for several cases in the framework of ASCLT under different settings. For example, Atlagh (1996) shows a version of the Glivenko-Cantelli theorem for independent random variables with normal distribution.

Having shown that the empirical distribution converges to the true distribution, it is now our intention to establish similar results for the quantiles of these distributions. This will lead towards the further idea of hypothesis testing. Before presenting the results relating to the quantiles of the distributions, we need to define certain functions which would be used in the results.

Definition 5 (Inverses of G_P, $\widetilde{G}_N$ and $\widehat{G}_N$). *For fixed $N \in \mathbb{N}$, let the inverse of any distribution function $\widetilde{F}_N$, denoted by function $\widetilde{F}_N^{-1}$, be defined by,*

$$\widetilde{F}_N^{-1}(\alpha) \quad = \quad \begin{cases} \sup\{t \mid \widetilde{F}_N(t) = 0\} \\ \sup\{t \mid \widetilde{F}_N(t) < \alpha\}, & \text{for } \alpha \in (0,1) \\ \inf\{t \mid \widetilde{F}_N(t) = 1\}. \end{cases} \tag{2.10}$$

The inverses of G_P, $\widetilde{G}_N$ and $\widehat{G}_N$ are obtained by substituting these functions appropriately in place of $\widetilde{F}_N$ in the above equation and are represented by G_P^{-1}, $\widetilde{G}_N^{-1}$ and $\widehat{G}_N^{-1}$, respectively.

Definition 6 (Empirical α-Quantiles). *The empirical α-quantiles of statistics T_n, $n \leq N \in \mathbb{N}$, are defined with respect to $\widetilde{G}_N$ and $\widehat{G}_N$, for $\alpha \in [0,1]$, by*

$$\widetilde{t}_\alpha^{(N)} \quad = \quad \widetilde{G}_N^{-1}(\alpha) \tag{2.11}$$

$$\widehat{t}_\alpha^{(N)} \quad = \quad \widehat{G}_N^{-1}(\alpha) \tag{2.12}$$

Lemma 7. *The functions G_P^{-1}, $\widetilde{G}_N^{-1}$ and $\widehat{G}_N^{-1}$ are left continuous for $\alpha \in (0,1)$.*

Proof. For ease of notation, let us denote the inverse functions by $F^{-1}(\alpha)$, for $0 < \alpha < 1$.

Let (α_n) be a sequence such that $\alpha_n \uparrow \alpha$. Further, let $\epsilon > 0$. Now, since

$$F^{-1}(\alpha) = \sup\{t \mid F(t) < \alpha\},$$

there exists a sequence (t_n) such that $F(t_n) < \alpha$, and $t_n \uparrow F^{-1}(\alpha)$. Hence, there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$

$$t_n \geq F^{-1}(\alpha) - \epsilon.$$

Moreover, as $\alpha_n \uparrow \alpha$ and $F(t_{n_0}) < \alpha$, there exists $n_1 \in \mathbb{N}$ such that for $n \geq n_1$

$$\alpha_n > F(t_{n_0}).$$

Consequently, for $n \geq n_1$

$$\begin{aligned} F^{-1}(\alpha_n) &= \sup\{t \mid F(t) < \alpha_n\} \\ &\geq \sup\{t \mid F(t) = F(t_{n_0})\} \\ &\geq t_{n_0} \\ &\geq F^{-1}(\alpha) - \epsilon. \end{aligned}$$

□

Similar to Theorem 4, we have the following result which presents the relation between the empirical and the true quantiles.

Lemma 8. *Let the α-quantile of G_P be denoted and defined by $t_\alpha(P) = G_P^{-1}(\alpha)$ and let $\bar{t}_\alpha(P) = \sup\{t \mid G_P(t) \leq \alpha\}$, then*

$$\bar{t}_\alpha(P) \;\geq\; \limsup_{N\to\infty} \widetilde{t}_\alpha^{(N)} \;\geq\; \liminf_{N\to\infty} \widetilde{t}_\alpha^{(N)} \;\geq\; t_\alpha(P) \qquad\qquad P-a.s.$$

Proof. For $\alpha \in [0,1)$, consider $s \in \mathbb{R}$ such that $s > G_P^{-1}(\alpha)$, which implies, $G_P(s) \geq \alpha$. By Theorem 4, $\exists N_0 \in \mathbb{N}$ such that for $N \geq N_0$,

$$\begin{aligned} & \widetilde{G}_N(s) > \alpha \\ \Rightarrow\quad & s \notin \left\{t \mid \widetilde{G}_N(t) \leq \alpha\right\} \\ \Rightarrow\quad & s > \widetilde{G}_N^{-1}(\alpha) \end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad \inf_s \left\{ s \mid s > G_P^{-1}(\alpha) \right\} &= G_P^{-1}(\alpha) \geq \limsup_{N\to\infty} \widetilde{G}_N^{-1}(\alpha) \\
\Rightarrow \quad t_\alpha &\geq \limsup_{N\to\infty} \widetilde{t}_\alpha^{(N)}
\end{aligned}$$

Let $r \in \mathbb{R}$ such that $r < G_P^{-1}(\alpha) \Rightarrow G_P(r) < \alpha$. Then $\exists N_1 \in \mathbb{N}$ such that for $N \geq N_1$,

$$\begin{aligned}
\widetilde{G}_N(r) &< \alpha \\
r &\in \left\{ t \mid \widetilde{G}_N(t) \leq \alpha \right\} \\
\Rightarrow \quad r &\leq \widetilde{G}_N^{-1}(\alpha) \\
\Rightarrow \quad G_P^{-1}(\alpha) &\leq \liminf_{N\to\infty} \widetilde{G}_N^{-1}(\alpha)
\end{aligned}$$

Thus,

$$G_P^{-1}(\alpha) \leq \liminf_{N\to\infty} \widetilde{G}_N^{-1}(\alpha) \leq \limsup_{N\to\infty} \widetilde{G}_N^{-1}(\alpha) \leq \bar{t}_\alpha(P)$$

For the case when $\alpha = 1$ or $= 0$, using Definition 5 and Theorem 4, the result follows in the same manner. □

Corollary 9. *If G_P is continuous, then*

$$\lim_{N\to\infty} \widetilde{t}_\alpha^{(N)} = t_\alpha(P) \qquad P - a.s.$$

Having shown that the estimated, empirical quantiles converge a.s. to the true ones, we are now interested in constructing intervals using the quantiles, based on which testing of hypothesis concerning the relevant parameter is proposed. It has to be noted here that though most of the above results relate to $\widetilde{G}_N$, these can be applied to $\widehat{G}_N$, which is a special case with sequence $a_n \equiv 0,\ n = 1, \ldots, N$ as defined in (2.9).

There are practical difficulties in using the function $\widetilde{G}_N$ because usually we are interested in some parameter, say μ, and it is unknown. So, we propose our further results on testing of hypothesis based on $\widehat{G}_N$ and finally we will provide a hint on how to develop similar results using $\widetilde{G}_N$.

Before presenting the next result, for simplicity, let us denote the distribution G_P parametrised by μ, by G_μ. Correspondingly, the inverse of G_μ is denoted by G_μ^{-1}. And when $\mu = 0$, we denote G_μ by just G and similarly, G_μ^{-1} by just G^{-1} in order to facilitate easier presentation and distinguish between the special (case of $\mu = 0$) and the general cases.

Lemma 10. *For $\alpha \in [0,1]$ satisfying $G_P^{-1}(\alpha) < 0 < G_P^{-1}(1-\alpha)$, we have the following statements:*

1. *If $a_n(P) = 0$, $\forall n$, then*

$$\lim_{N\to\infty} P\left(0 \in \left[\widehat{t}_{\alpha}^{(N)},\ \widehat{t}_{1-\alpha}^{(N)}\right]\right) = 1 \qquad a.s. \tag{2.13}$$

2. *If $a_n \neq 0$, $\forall n$, and $a_n(P) \to \infty$ or $a_n(P) \to -\infty$,*

$$\lim_{N\to\infty} P\left(0 \in \left[\widehat{t}_{\alpha}^{(N)},\ \widehat{t}_{1-\alpha}^{(N)}\right]\right) = 0 \qquad a.s. \tag{2.14}$$

Proof. Let $\alpha \in \left[0, \frac{1}{2}\right)$. Then we have,

$$\begin{aligned}
\alpha &= \widehat{G}_N\left(\widehat{t}_{\alpha}^{(N)}\right) = \frac{1}{C_N}\sum_{n=1}^{N}\frac{1}{n}1_{\left\{b_nT_n \le \widehat{t}_{\alpha}^{(N)}\right\}} \\
&= \frac{1}{C_N}\sum_{n=1}^{N}\frac{1}{n}1_{\left\{b_n(T_n-a_n) \le \widehat{t}_{\alpha}^{(N)} - a_nb_n\right\}} \\
&= \widetilde{G}_N\left(\widehat{t}_{\alpha}^{(N)} - a_nb_n\right)
\end{aligned}$$

Thus when $a_n = 0$, $\forall n$, using Lemma 8 and Corollary 9,

$$\widehat{t}_{\alpha}^{(N)} - a_nb_n = \widetilde{t}_{\alpha}^{(N)} = \widetilde{G}_N^{-1}(\alpha) \to G_{\mu}^{-1}(\alpha) = t_{\alpha} \qquad \text{a.s.} \tag{2.15}$$

Similarly, $\widehat{t}_{1-\alpha}^{(N)} \to t_{1-\alpha}$, a.s. Now, by assumption that $G_P^{-1}(\alpha) < 0 < G_P^{-1}(1-\alpha)$, the result 1 follows.

Moreover, for $a_nb_n \to \infty$ (or $-\infty$), as $n \to \infty$, also $\widehat{t}_{\alpha}^{(N)} \to \infty$ (or $\widehat{t}_{1-\alpha}^{(N)} \to -\infty$). Therefore, $0 < \widehat{t}_{\alpha}^{(N)}$ (or $0 > \widehat{t}_{1-\alpha}^{(N)}$) $\implies 0 \notin \left[\widehat{t}_{\alpha}^{(N)}, \widehat{t}_{1-\alpha}^{(N)}\right]$ a.s. $\square$

Theorem 11. *Under assumption that $\widetilde{G}_N^{-1}(\alpha) = \widehat{G}_N^{-1}(\alpha) - a_Nb_N$, for $\alpha \in [0,1]$.*

1. *If $a_n(P) = 0\ \forall n$,*

$$\lim_{N\to\infty} P\left(b_NT_N \in A_{\alpha}^{(N)}\right) = 1 - 2\alpha \qquad a.s. \tag{2.16}$$

2. If $a_n(P) \neq 0; a_n(P)b_n \to \infty$ or $-\infty$ as $n \to \infty$,

$$\lim_{N\to\infty} P\left(b_N\left(T_N - a_N\right) \in A_\alpha^{(N)}\right) \quad = \quad 0 \qquad a.s., \tag{2.17}$$

where $A_\alpha^{(N)} = \left[\widehat{t}_\alpha^{(N)},\ \widehat{t}_{1-\alpha}^{(N)}\right]$.

Proof. 1. From (2.15) in the proof of above Lemma 10,

$$\begin{aligned}
\lim_{N\to\infty} P\left(b_N T_N \in A_\alpha^{(N)}\right) &= \lim_{N\to\infty} P\left(b_N T_N \in \left[G^{-1}(\alpha), G^{-1}(1-\alpha),\right]\right) \\
&= 1 - G\left(G^{-1}(\alpha)\right) - \left(1 - G\left(G^{-1}(1-\alpha)\right)\right) \qquad \text{by (2.6)} \\
&= 1 - \alpha - (1 - (1-\alpha)) = 1 - 2\alpha
\end{aligned}$$

2. If $\widehat{t}_\alpha^{(N)} \to \infty$,

$$P\left(b_N\left(T_N - a_N\right) \in A_\alpha^{(N)}\right) \quad \leq \quad \widetilde{G}_N\left(\widehat{G}_N^{-1}(1-\alpha)\right) - \widetilde{G}_N\left(\widehat{G}_N^{-1}(\alpha)\right)$$

Now, by assumption,

$$= \quad \widetilde{G}_N\left(\widetilde{G}_N^{-1}(1-\alpha) + a_N b_N\right) - \widetilde{G}_N\left(\widetilde{G}_N^{-1}(\alpha) + a_N b_N\right)$$

By using the results of Theorem 4, Lemma 8 and Corollary 9, we have,

$$\begin{aligned}
&\overset{N\to\infty}{\longrightarrow} \quad G_P\left(G_P^{-1}(1-\alpha) + \lim_{N\to\infty} a_N b_N\right) - \\
&\qquad\qquad G_P\left(G_P^{-1}(\alpha) + \lim_{N\to\infty} a_N b_N\right) \\
&= \quad 0.
\end{aligned}$$

Similarly the result can be derived for $\widehat{t}_\alpha^{(N)} \to -\infty$.

□

Based on the Lemma 10 and Theorem 11 we can now define the so-called ASCLT-based tests. Before we do so, let the following slightly changed notation be set. Let $\mathbf{X} = (X_1, \ldots, X_N)$ denote a sample of size N with i.i.d. elements $X_i \sim G_\mu$ for parameter $\mu \in \mathbb{R}$. Let also the corresponding observed sample by denoted $\mathbf{x} = (x_1, \ldots, x_N)$. Further, let $T_N(\mathbf{X})$ and $T_N(\mathbf{x})$ be the statistics based on the r.vs and computed from the sample $\mathbf{x}$, respectively, with $E\left(T_N(\mathbf{X})\right) = N\mu$. Finally, let quantiles estimated from the sample $\mathbf{x}$ by denoted by $\widehat{t}_\alpha^{(N)}(\mathbf{x})$ and $\widehat{t}_{1-\alpha}^{(N)}(\mathbf{x})$.

Definition 12 (ASCLT-test Method 1). *For a test of hypothesis of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ at a significance level of 2α, the ASCLT-test method 1 is defined by the decision function,*

$$\delta(\mathbf{x}) = \begin{cases} 1\ (\textit{Accept } H_0), & 0 \in \left[\widehat{t}_{\alpha}^{(N)}(\mathbf{x}),\ \widehat{t}_{1-\alpha}^{(N)}(\mathbf{x})\right] \\ 0\ (\textit{Reject } H_0), & \textit{otherwise.} \end{cases} \tag{2.18}$$

Definition 13 (ASCLT-test Method 2). *For a test of hypothesis of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ at a significance level of 2α, the ASCLT-test method 2 is defined by the decision function,*

$$\delta(\mathbf{x}) = \begin{cases} 1\ (\textit{Accept } H_0), & \frac{T_N(\mathbf{x})}{N} \in A_{\alpha}^{(N)} \\ 0\ (\textit{Reject } H_0), & \textit{otherwise,} \end{cases} \tag{2.19}$$

where $A_{\alpha}^{(N)} = \left[\widehat{\mu} + \frac{\widehat{t}_{\alpha}^{(N)}(\mathbf{x})}{\sqrt{N}}\ ,\ \widehat{\mu} + \frac{\widehat{t}_{1-\alpha}^{(N)}(\mathbf{x})}{\sqrt{N}}\right]$.

We note here that, when the distribution is symmetric around the parameter μ, the above interval $A_{\alpha}^{(N)}$ is equivalent to $\left[\widehat{\mu} - \frac{\widehat{t}_{1-\alpha}^{(N)}(\mathbf{x})}{\sqrt{N}}\ ,\ \widehat{\mu} - \frac{\widehat{t}_{\alpha}^{(N)}(\mathbf{x})}{\sqrt{N}}\right]$.

Further, we also note that the interval $A_{\alpha}^{(N)}$ used in the above definition contains the estimator of parameter μ. In Chapter 4, there is a proposal for such an appropriate estimator.

It can also be noted that the results of Lemma 10 and Theorem 11 can be extended to the situation where the quantiles are estimated based on $\widetilde{G}_N$. But the following issues have to be taken care of, in doing so:

- The estimation of the quantiles using function $\widetilde{G}_N$ is done irrespective of the hypothesis of $\mu = 0$, i.e., the estimated quantile will be centered at mean 0 whether $\mu = 0$ or not. This approach of estimation would be new, but very interesting to be explored in future in a detailed manner. On the contrary, the quantile estimation based on $\widehat{G}_N$ follows the conventional idea, i.e., setting $\mu = 0$ and then estimating the quantiles.
- In practice, to replace the term $a_n = n\mu$ in the equation of $\widetilde{G}_N$ would have to be dealt with carefully, since, μ is unknown and the entire problem revolves around

testing for μ. One of the ways to address this problem is by assuming that T_n satisfies the law of iterated logarithm, i.e.,

$$\limsup_{N\to\infty} \frac{|T_N - N\mu|}{\sqrt{N \log\log N}} \quad < \quad \infty.$$

As a final remark, it can be noticed from the material presented in this chapter that the procedures of hypothesis testing that is proposed are totally free of variance estimation. The methods proposed are thus unique from this perspective too, compared to the other existing tests of hypothesis wherein usually the concerned statistic is standardized with respective variance. Another philosophical perspective of these methods is with respect to making almost sure decisions which was briefly mentioned discussed in Chapter 1. The performance of these methods is evaluated in Chapter 4 via extensive simulation studies in specific situations.

Chapter 3

ASCLT for Rank Statistics

3.1 Introduction

In this chapter we will be concerned with establishing the ASCLT for the two-sample linear rank statistics, which will be defined and discussed in the due course. Let us first set the notation in order to present the material with ease.

Notation:
Notation and definition of terminology that will be considered throughout this chapter is set here. Let $(X_1, \ldots, X_m, X_{m+1}, \ldots, X_n)$ be i.i.d random variables(r.v.s) such that the first m r.v.s correspond to the first sample and are distributed as F and the remaining $n-m$ r.v.s correspond to the second sample and are distributed as G. Also let R_i denote the mid-rank of the X_i, over all n random variables. Further let the weighted average of the two distribution functions be denoted by $H_n(x)$, which is defined by,

$$H_n(x) = \frac{1}{n}\left(mF(x) + (n-m)\,G(x)\right).$$

Let also the empirical distribution of each of the samples be given by

$$\begin{aligned}
\widehat{F}(x) &= \frac{1}{m}\sum_{k=1}^{m} c(x - X_k) \\
\widehat{G}(x) &= \frac{1}{n-m}\sum_{k=m+1}^{n} c(x - X_k),
\end{aligned}$$

where $c(u) = 0, \frac{1}{2}$ or 1 according as $u <, =$ or > 0, is called the *normalized* version of the *count* function $c(\cdot)$. Thus, the corresponding empirical version of $H_n(x)$ is denoted and defined by,

$$\widehat{H}_n(x) = \frac{1}{n}\left(m\widehat{F}(x) + (n-m)\widehat{G}(x)\right) = \frac{1}{n}\sum_{k=1}^{n} c(x - X_k)$$

We refer to Akritas et al. (1997) for some discussion on these notation and terminology, and their practical implications.

Based on the above notation, many (two-sample) nonparametric test statistics have the form of *two-sample linear rank statistics* which is presented as,

$$T_n = \sum_{i=1}^{n} a_i J\left(\widehat{H}_n\left(X_i\right)\right) \tag{3.1}$$

for $n \geq 2; 1 \leq m(n) = m < n$ such that $\frac{m(n)}{n} \overset{n\to\infty}{\longrightarrow} \lambda \in (0,1)$, a constant, and where $a_i = 1$ or 0 according as $1 \leq i \leq m$ or $m < i \leq n$ and $J : (0,1) \to \mathbb{R}$ is absolutely continuous score function. We assume that $m = m(n)$ depends on n, and as $n \to \infty$, both the sample sizes increase. The asymptotic normality of such statistics was first proved by Chernoff and Savage (1958). Subsequent general results were presented by Hájek (1968), Pyke and Shorack (1968), Dupac and Hájek (1969) and Denker and Rösler (1985). For some discussion these developments we refer to the introductory part of Brunner and Denker (1994).

In this work, we are now interested in proving the ASCLT result for the rank statistics given in (3.1). First, two results from literature, which would be used in the proof of the theorem on ASCLT on rank statistics, is presented below. The first result was originally proposed by Berkes and Dehling (1993), and later reported and discussed in Berkes (1998). We present the version which corresponds to Corollary 2.2 of Berkes (1998). The second result is Proposition 1 of Lesigne (1999).

Lemma 14 (Berkes, 1998). *Let $X_1, X_2, \ldots$ be independent random variables with $E\left(X_k\right) = 0$, $E\left(X_k^2\right) = \sigma_k^2$, $(k = 1, 2, \ldots)$ and let $b_n^2 = \sum_{k=1}^{n} \sigma_k^2$. Assume that for some constants $\gamma > 0$, $C > 0$,*

$$\frac{b_l}{b_k} \geq C\left(\frac{l}{k}\right)^{\gamma}, \qquad (1 \leq k \leq l), \tag{3.2}$$

and the sequence (X_n) *satisfies the Lindeberg condition*

$$\lim_{n\to\infty}\frac{1}{b_n^2}\sum_{k=1}^{n}E\left(X_k^2 1_{\{|X_k|\geq \epsilon b_n\}}\right) \quad = \quad 0 \qquad \forall \epsilon > 0. \tag{3.3}$$

Then

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{k\leq N}\frac{1}{k}1_{\left\{\frac{S_k}{b_k}<x\right\}} \quad = \quad \Phi(x) \qquad a.s.\ for\ all\ x,$$

where $S_k = X_1 + \ldots + X_k$, $(k = 1, 2, \ldots)$.

Lemma 15 (Lesigne, 1999). [†] *Let* V_n *and* W_n*, for* $n \in \mathbb{N}$*, be two sequences of random variables such that:*

1. *the sequence* V_n *satisfies the ASCLT, that is to say, almost surely, the sequence of probability measures*

$$\frac{1}{\log n}\sum_{k=1}^{n}\frac{1}{k}1_{\{V_n\}}$$

   converges weakly to the Gaussian law $N(0,1)$*;*

2. *for all* $\epsilon > 0$*, there exists* $\delta > 0$ *such that*

$$P\left(|\, V_n - W_n \,| > \epsilon\right) \quad = \quad O\left(\frac{1}{(\log n)^{\delta}}\right).$$

   Then the sequence W_n *satisfies the ASCLT.*

3.2 ASCLT for Rank Statistics

Along with notation and terminology introduced in the previous section, let $\widetilde{\sigma}_n^2$ denote the variance of the centered rank statistics, $T_n - E(T_n)$. From existing standard results (c.f. Brunner and Denker, 1994), it is also clear that the $E(T_n - E(T_n)) = 0$. We note here that the $\widetilde{\sigma}_n^2$ are strictly positive for all distributions except for which the one-point distributions. So, in the following theorem we exclude such distributions from our consideration.

[†] I would like to specially thank Dr. Aleksey Min for pointing to this result and reference.

Assumptions 3.2.1.

1. *Let the score function J be twice differentiable and J'' be bounded.*
2. *The underlying distribution functions F and G are arbitrary, with the exception of the trivial one-point distribution*
3. $n \to \infty$, $m = m(n) \uparrow$ *and* $\frac{m}{n} \to \lambda$, *such that* $\left|\frac{m}{n} - \lambda\right| = O\left(\frac{1}{(\log n)^{\delta}}\right)$, *for some* $\delta > 0$.
4. *The asymptotic variances of* $J\left(H\left(X_1\right)\right) + h_F\left(X_1\right)$ *and* $h_F\left(X_{m+1}\right)$ *are strictly positive, where* $h_F\left(X\right) = \int J'\left(H\left(t\right)\right) c\left(t - X\right) F\left(\mathrm{d}\, t\right)$.

Theorem 16 (ASCLT for Two-Sample Linear Rank Statistics). *Let the assumptions defined in 3.2, hold. Then the two-sample linear rank statistics satisfies the ASCLT. That is,*

$$\frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} 1_{\left\{k^{-1/2}\left(T_k - m \int J(H_k(t)) dF_{m_k}(t)\right) \leq t\right\}} \to \Phi_{\sigma}(t) \qquad P - a.s.$$

Proof. The basic idea of the proof would be to decompose $n^{-1/2}\left(T_n - m \int J\left(H_n(t)\right) dF_m\left(t\right)\right)$ for $n \in \mathbb{N}\left(n \geq 2\right)$ such that a part of the decomposition satisfies ASCLT and the others go to 0, almost surely in the sense of Lemma 15.

Let F and corresponding empirical counterpart $\widehat{F}$ be denoted by F_m and $\widehat{F}_m$, respectively, in order to emphasize the dependence of these distributions on sample size of first sample, m.

The statistic T_n can be expressed in terms of the empirical distributions via integral as

$$T_n \;=\; \sum_{i=1}^{m} J\left(\widehat{H}_n\left(X_i\right)\right) \;=\; m \int J\left(\widehat{H}_n\left(t\right)\right) d\widehat{F}_m\left(t\right).$$

Now we consider the Taylor expansion of T_n around $H_n(t)$, which is given by,

$$\begin{aligned} T_n = m \Bigg[\int J\left(H_n\left(t\right)\right) \mathrm{d}\,\widehat{F}_m\left(t\right) + \int \left(\widehat{H}_n - H_n\right)\left(t\right) J'\left(H_n\left(t\right)\right) \mathrm{d}\,\widehat{F}_m\left(t\right) \\ + \frac{1}{2} \int \left(\widehat{H}_n - H_n\right)^2 \left(t\right) J''\left(\theta\left(t\right)\right) \mathrm{d}\,\widehat{F}_m\left(t\right) \Bigg], \end{aligned} \qquad (3.4)$$

where $\theta(t) \in \left[\widehat{H}_n(t), H_n(t)\right] \bigcup \left[H_n(t), \widehat{H}_n(t)\right]$.

We use the above expansion (3.4) of T_n, along with subtracting $\frac{m}{\sqrt{n}} \int J(H_n(t)) \,\mathrm{d}\, F_m(t)$ from $n^{-1/2}T_n$ and further decomposing the term $\int \left(\widehat{H}_n - H_n\right)(t)\, J'\left(H_n(t)\right) \mathrm{d}\,\widehat{F}_m(t)$, in order to obtain an expression of the following form:

$$\frac{1}{\sqrt{n}}\left(T_n - m\int J\left(H_n(t)\right) dF_m(t)\right) \quad = \quad \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{B} + \mathcal{C},$$

where the respective terms are given by,

$$\begin{aligned}
\mathcal{A}_1 &= \frac{m}{\sqrt{n}}\left(\int J\left(H_n(t)\right) \mathrm{d}\,\widehat{F}_m(t) - \int J\left(H_n(t)\right) \mathrm{d}\,F_m(t)\right), \\
\mathcal{A}_2 &= \frac{m}{\sqrt{n}}\int J'(H_n(t))\ (\widehat{H}_n - H_n)(t)\,\mathrm{d}\,F_m(t), \\
\mathcal{B} &= \frac{m}{\sqrt{n}}\int J'(H_n(t))\ (\widehat{H}_n - H_n)(t)\,\mathrm{d}\left(\widehat{F}_m(t) - F_m(t)\right), \quad \text{and} \\
\mathcal{C} &= \frac{m}{2\sqrt{n}}\int J''\left(\theta(t)\right)\ (\widehat{H}_n - H_n)^2(t)\,\mathrm{d}\,\widehat{F}_m(t).
\end{aligned}$$

Now, let us consider the first two terms A_1 and A_2 together and then the two other terms are considered individually.

Terms $\mathcal{A}_1$ and $\mathcal{A}_2$: It is straight forward to note that the term $\mathcal{A}_1$ can be expressed of the following way:

$$\mathcal{A}_1 \quad = \quad \sum_{i=1}^{m} \frac{1}{\sqrt{n}}\left[J\left(H_n(X_i)\right) - E\left(J\left(H_n(X_i)\right)\right)\right]. \tag{3.5}$$

Similarly, we get an expression for term $\mathcal{A}_2$.

$$\begin{aligned}
\mathcal{A}_2 &= \frac{m}{\sqrt{n}}\int J'(H_n(t))\ \left(\frac{1}{n}\sum_{i=1}^{n} c(t - X_i) - H_n(t)\right) F(\mathrm{d}\,t) \\
&= \frac{m}{n}\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{m}\int J'(H_n(t))\ \left(c(t - X_i) - F(t)\right) F(\mathrm{d}\,t)\right. \\
&\qquad \left. + \sum_{i=m+1}^{n}\int J'(H_n(t))\ \left(c(t - X_i) - G(t)\right) F(\mathrm{d}\,t)\right]
\end{aligned} \tag{3.6}$$

Our intention is to use the Lemma 14 in order to establish the required result of the ASCLT for rank statistics. But we can not use the above two forms of $\mathcal{A}_1$ and $\mathcal{A}_2$ which presented via the function $H_n(t)$, since it is defined and based on n. We rather need a sequence of i.i.d. r.v.s. Thus, we propose the following modifications of $\mathcal{A}_1$ and $\mathcal{A}_2$, and show that these modifications and the original terms are almost surely the same (in the sense of Lesigne, 1999).

$$\widetilde{\mathcal{A}}_1 = \sum_{i=1}^{m} \frac{1}{\sqrt{n}} \left[J\left(H\left(X_i\right)\right) - E\left(J\left(H\left(X_i\right)\right)\right)\right]. \tag{3.7}$$

$$\begin{aligned} \widetilde{\mathcal{A}}_2 &= \lambda \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{m} \left(h_F\left(X_i\right) - E\left(h_F\left(X_i\right)\right)\right) \right. \\ &\qquad \left. + \sum_{i=m+1}^{n} \left(h_F\left(X_i\right) - E\left(h_F\left(X_i\right)\right)\right) \right], \end{aligned} \tag{3.8}$$

where $H = \lambda F + (1-\lambda)G$ and $h_F\left(X\right) = \int J'\left(H\left(t\right)\right) c\left(t - X\right) F\left(\mathrm{d}\, t\right)$.

We note here that the individual quantities on the R.H.S.'s of the above expressions (3.5) and (3.6), and also expressions (3.7) and (3.8), are not i.i.d. but only independent. Let us introduce random variables Y_i, $i = 1, \ldots, n$ such that,

$$Y_i = \begin{cases} \left[J\left(H\left(X_i\right)\right) - E\left(J\left(H\left(X_i\right)\right)\right)\right] + \lambda\left(h_F\left(X_i\right) - E\left(h_F\left(X_i\right)\right)\right), & i = 1, \ldots, m \\ \lambda\left(h_F\left(X_i\right) - E\left(h_F\left(X_i\right)\right)\right), & i = m+1, \ldots, n. \end{cases}$$

We note here that rank statistics are defined on an array of r.v.s X_i, $i = 1, \ldots, m, m+1, \ldots, n$. But, in order to apply the result of Lemma 14, we need to have a sequence of r.v.s. Since, by assumption, m is increasing, we can rearrange any given array of r.v.s to form a sequence of r.v.s. Let, for $1 \leq k \leq l \in \mathbb{N}$, $I_k \subset \{1, \ldots, k\}$ such that for $i \in I_k$, X_i corresponds to distribution F. Moreover, $I_k \subset I_l \subset \{1, \ldots, l\}$.

By assumption we note that J and J' are bounded. Thus, the r.v.s Y_i, $i = 1, \ldots, n$ are bounded. We also see that $E\left(Y_i\right) = 0$. Further, denoting $b_n = \sum_{i=1}^{n} \sigma_i^2$, where $E\left(Y_i^2\right) = \sigma_i^2$, by Assumption (4) and that the distribution functions F and G are not one-point distributions, we have $b_n \to \infty$.

Thus the sequence of r.v.s Y_i satisfy the Lindeberg condition defined in (3.3) in Lemma 14. That is,

$$\frac{1}{n} \sum_{i=1}^{n} \int_{|Y_i| > b_n \epsilon} \left(Y_i - E\left(Y_i\right)\right)^2 \to 0 \qquad \forall \epsilon > 0.$$

Further let $m_k \leq m_l \in \mathbb{N}$. Then, by assumption, $\left|\frac{m_l}{l} - \lambda\right| \leq \left(\frac{c}{(\log n)^\delta}\right)$, for some $\delta > 0$ and some constant c, and thus also, $\left|\frac{l-m_l}{l} - 1 + \lambda\right| \leq \left(\frac{c'}{(\log n)^{\delta'}}\right)$, for some $\delta' > 0$ and some constant c'.

Let $a_F^2 = \widetilde{\sigma}_{m_l}^2 \operatorname{Var}(h_F(X_1))$ and $a_G^2 = \operatorname{Var}(h_F(X_{m_l+1}))$. Now consider,

$$\begin{aligned}
\sum_{i\in I_l} \sigma_i^2 + \sum_{i\notin I_l} \sigma_i^2 &= m_l a_F^2 + (l - m_l) a_G^2 \\
&\geq \left(l\lambda - \frac{lc}{(\log n)^\delta}\right) a_F^2 + \left(l(1-\lambda) - \frac{lc'}{(\log n)^{\delta'}}\right) a_G^2 \\
&\geq \frac{1}{2} l \left(\lambda a_F^2 + (1-\lambda) a_G^2\right) \qquad \text{for large } l. \qquad (3.9)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{i\in I_k} \sigma_i^2 + \sum_{i\notin I_k} \sigma_i^2 &= m_k a_F^2 + (k - m_k) a_G^2 \\
&\leq 2k \left(\lambda a_F^2 + (1-\lambda) a_G^2\right) \qquad \text{for large } k. \qquad (3.10)
\end{aligned}$$

So we see that the requirement (3.2) of Lemma 14 is established by combining the above two inequalities (3.9) and (3.10), for $1 \leq k \leq l$,

$$\begin{aligned}
\frac{b_l}{b_k} &= \frac{\sum_{i\in I_l} \sigma_i^2 + \sum_{i\notin I_l} \sigma_i^2}{\sum_{i\in I_k} \sigma_i^2 + \sum_{i\notin I_k} \sigma_i^2} \\
&\geq \frac{1}{4} \frac{l}{k} \qquad \text{for large } k, l.
\end{aligned}$$

Thus, having shown that the conditions of Lemma 14 are satisfied, we have $\widetilde{\mathcal{A}}_1 + \widetilde{\mathcal{A}}_2 \xrightarrow{a.s.} \Phi_\sigma$, where Φ_σ is the distribution function of the Normal distribution parameterized by mean 0 and variance σ^2, and σ^2 is the variance of the r.v.s Y_i, $i = 1, \ldots, n$. But we are interested in showing the result for $\mathcal{A}_1 + \mathcal{A}_2$. For this, we establish the following result using Lemma 15,

$$\mathcal{A}_1 + \mathcal{A}_2 - \widetilde{\mathcal{A}}_1 - \widetilde{\mathcal{A}}_2 \quad \to \quad 0 \qquad \text{a.s.}$$

Consider,

$$E\left(\mathcal{A}_1 - \widetilde{\mathcal{A}}_1\right)^2 \quad \leq \quad \frac{m}{n} \operatorname{Var}\left(J\left(H_n\left(X_i\right)\right) - J\left(H\left(X_i\right)\right) + E\left(J\left(H_n\left(X_i\right)\right)\right) - E\left(J\left(H\left(X_i\right)\right)\right)\right)$$

$$\leq 2\lambda \left[E\left(J\left(H_n\left(X_i\right)\right) - J\left(H\left(X_i\right)\right)\right)^2 + E\left(E\left(J\left(H_n\left(X_i\right)\right)\right) - E\left(J\left(H\left(X_i\right)\right)^2\right)\right)\right]$$

By the definitions of H_n and H, and by applying mean value theorem,

$$\leq 2\lambda \left\| J' \right\|_\infty \sup_t \left| H_n\left(t\right) - H\left(t\right) \right|^2$$

Now, by the Assumption 3.2(3),

$$E\left(\mathcal{A}_1 - \widetilde{\mathcal{A}}_1\right)^2 = O\left(\frac{1}{\left(\log n\right)^\gamma}\right) \qquad \text{for some } \gamma > 0. \tag{3.11}$$

Similarly,

$$E\left(\mathcal{A}_2 - \widetilde{\mathcal{A}}_2\right)^2 \leq \left(\frac{m}{n}\right)^2 \operatorname{Var}\left(h_{F_n}\left(X_i\right) - h_F\left(X_i\right) + E\left(h_{F_n}\left(X_i\right)\right) - E\left(h_F\left(X_i\right)\right)\right),$$

where, $h_{F_n}\left(X\right) = \int J'\left(H_n\left(t\right)\right) c\left(t - X\right) F\left(\mathrm{d}\, t\right)$ and h_F introduced earlier. Thus, by similar arguments above, using the mean value theorem, the definitions of H_n and H and the Assumption 3.2(3), we get,

$$\begin{aligned} E\left(\mathcal{A}_2 - \widetilde{\mathcal{A}}_2\right)^2 &\leq 2\lambda^2 \left\| J' \right\|_\infty^2 \sup_t \left| H_n\left(t\right) - H\left(t\right) \right|^2 \\ &= O\left(\frac{1}{\left(\log n\right)^{\gamma'}}\right) \qquad \text{for some } \gamma' > 0. \end{aligned} \tag{3.12}$$

From (3.11) and (3.12), by applying Lemma 15, we have that $\widetilde{\mathcal{A}}_1 + \widetilde{\mathcal{A}}_2 \xrightarrow{a.s.} \Phi_\sigma$ implies $\mathcal{A}_1 + \mathcal{A}_2 \xrightarrow{a.s.} \Phi_\sigma$, for σ defined earlier. We also note here that, by virtue of the same lemma, it is enough to show the following, in order to establish that the rank statistics satisfy the ASCLT:

$$P\left(\left| \frac{1}{\sqrt{n}}\left(T_n - m\int J\left(H_n(t)\right) dF_m\left(t\right)\right) - \mathcal{A}_1 - \mathcal{A}_2 \right| > \epsilon\right) = O\left(\frac{1}{\left(\log n\right)^{\delta'}}\right),$$

for some $\epsilon > 0$ and $\delta' > 0$. Since, we have shown above that $\mathcal{A}_1 + \mathcal{A}_2$ satisfies the ASCLT, we have,

$$\begin{aligned} P\left(\left| \frac{1}{\sqrt{n}}\left(T_n - m\int J\left(H_n(t)\right) dF_m\left(t\right)\right) - \mathcal{A}_1 - \mathcal{A}_2 \right| > \epsilon\right) &= P\left(\left| \mathcal{B} + \mathcal{C} \right| > \epsilon\right) \\ &\leq \frac{E\left(\mathcal{B} + \mathcal{C}\right)^2}{\epsilon^2} \qquad \text{by Chebyshev's inequality,} \end{aligned}$$

(now by using C_r inequality, we get,)

$$\leq \frac{2\left(E\mathcal{B}^2 + E\mathcal{C}^2\right)}{\epsilon^2}. \tag{3.13}$$

Thus it is sufficient to show that,

$$E\mathcal{B}^2,\ E\mathcal{C}^2 \quad = \quad O\left(\frac{1}{(\log n)^{\delta'}}\right), \qquad \text{for some } \delta' > 0,$$

which follows by $E\mathcal{B}^2,\ E\mathcal{C}^2 = O\left(\frac{1}{n}\right)$. So, we consider each of the terms $\mathcal{B}$ and $\mathcal{C}$ and show the required result.

Term $\mathcal{B}$:

Let us first express the term as follows:

$$\mathcal{B} \quad = \quad \frac{m}{\sqrt{n}}\frac{1}{nm}\left[\sum_{r=1}^{n}\sum_{i=1}^{m}\phi_1\left(X_i, X_r\right) - \phi_2\left(X_r\right)\right],$$

where

$$\phi_1\left(X_i, X_r\right) \quad = \quad \begin{cases} J'\left(H\left(X_i\right)\right)\left[c\left(X_i - X_r\right) - F\left(X_i\right)\right], & r = 1, \ldots, m; \\ J'\left(H\left(X_i\right)\right)\left[c\left(X_i - X_r\right) - G\left(X_i\right)\right], & r = m+1, \ldots, n, \end{cases}$$

$$\phi_2\left(X_r\right) \quad = \quad \begin{cases} \int J'\left(H\left(t\right)\right)\left[c\left(t - X_r\right) - F\left(t\right)\right] \mathrm{d}\,F\left(x\right), & r = 1, \ldots, m; \\ \int J'\left(H\left(t\right)\right)\left[c\left(t - X_r\right) - G\left(t\right)\right] \mathrm{d}\,F\left(x\right), & r = m+1, \ldots, n. \end{cases}$$

Taking expectations of the above equation (3.14),

$$E\left(\mathcal{B}^2\right) \quad = \quad \frac{m^2}{n \cdot n^2 m^2}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} E\left(\left[\phi_1\left(X_i, X_r\right) - \phi_2\left(X_r\right)\right] \cdot \left[\phi_1\left(X_j, X_s\right) - \phi_2\left(X_s\right)\right]\right)$$

Now by using the property of independence of the r.v.s, we get

$$\begin{aligned} E\left(\mathcal{B}^2\right) \quad &\leq \quad \frac{1}{n^3}\sum_{r=1}^{n}\sum_{i=1}^{m} E\left(\phi_1\left(X_i, X_r\right) - \phi_2\left(X_r\right)\right)^2 \\ \Longrightarrow E\left(\mathcal{B}^2\right) \quad &= \quad O\left(\frac{n \cdot m \cdot \left\|J'\right\|_\infty^2}{n^3}\right) \quad = \quad O\left(\frac{\lambda\left\|J'\right\|_\infty^2}{n}\right) \end{aligned} \tag{3.14}$$

Term $\mathcal{C}$: The term can be expressed as,

$$\mathcal{C} \quad = \quad \frac{1}{2\sqrt{n}n^2} \sum_{r=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(J'' \left(\theta \left(X_r \right) \right) \ \phi \left(X_i, X_r \right) \cdot \phi \left(X_j, X_r \right) \right),$$

where $\phi(X_i, X_r) = c(X_r - X_i) - H_n(X_r)$. Now squaring and taking expectation of the above equation, we get

$$E\left(\mathcal{C}^2\right) \quad \leq \quad \frac{\|J'\|_\infty^2}{4n^5} \sum_{r=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{s=1}^{m} \sum_{i'=1}^{n} \sum_{j'=1}^{n} \left[\phi \left(X_i, X_r \right) \cdot \phi \left(X_j, X_r \right) \cdot \phi \left(X_{i'}, X_s \right) \cdot \phi \left(X_{j'}, X_s \right) \right].$$

By independence of the r.v.s, we get

$$E\left(\mathcal{C}^2\right) \quad = \quad O\left(\frac{\lambda^2 \left\| J'' \right\|_\infty^2}{n} \right)$$

Finally, using the results of (3.14) and (3.15) in (3.13) and noting that J has bounded second derivative and $\lambda \in (0,1)$, a constant, we see that

$$\begin{aligned} P\left(\left| \frac{1}{\sqrt{n}} \left(T_n - m \int J\left(H_n(t) \right) dF_m\left(t\right) \right) - \mathcal{A}_1 - \mathcal{A}_2 \right| > \epsilon \right) \quad &= \quad O\left(\frac{1}{n} \right) \\ &= \quad O\left(\frac{1}{\left(\log n\right)^{\delta'}} \right), \end{aligned}$$

for some $\delta' > 0$, thus proving the required result. $\square$

Chapter 4

Applications and Numerical Results

4.1 Introduction

The theoretical results presented in Chapters 2 and 3 were set on a general framework. Application-specific tests of hypothesis could be developed using the results from those chapters. One of the main aims of this chapter is to present examples of few such applications. It has to be emphasized that the theoretical results were all asymptotic in nature. Moreover, for hypothesis testing purposes we are interested in the quantiles on the tail of the distribution. Thus, intuitively it is evident that estimation of such tail quantiles, $\widehat{t}_\alpha$ and $\widehat{t}_{1-\alpha}$ (defined in Definition 6, on page 11), based on the ASCLT would require large sample sizes to get close to the true quantiles, t_α and $t_{1-\alpha}$.

For applications in real-time data analysis, particularly for medical data, often one has to deal with situations involving small sample sizes. Thus, there is a need to create a 'bridge' between the asymptotic result and the corresponding proposal of approximations for testing hypothesis based on small sample sizes. In this chapter some proposals for such a link would be suggested based our experience with simulated data. These proposals could be classified into two broad categories:

1. General proposals that can be applied to all tests of hypothesis, and

2. Test-specific proposals.

Before going directly to the proposals to adjust for small-sample cases for the ASCLT-based tests, we note here that, in general, any *small-sample test theory (or) approximation*, should ideally satisfy the following properties and aspects:

- The test, under the concerned null hypothesis, should maintain the pre-assigned level of significance, 2α.
- While moving away from the null hypothesis situation towards the alternative(s), the test should exhibit good power (i.e., *Type-II error*, β, should be made small).
- As the sample sizes increase to become very large, the approximation method proposed for small-samples should lead and coincide with the asymptotic theory, in a natural sense.

It has to be noted here that, in many cases, theoretical evaluation of the above properties of a small-sample test theory is not known. Thus, numerical studies based on computer simulations have to be used to check these properties. Random and quasi-random numbers should be generated under the concerned hypothesis and the type-I error rate of the method should be estimated. Similarly, the power property be evaluated by generating data under the alternative(s). The property of the approximation method converging to the asymptotic result(s) can also be verified numerically, though usually this property is possible to be derived theoretically in a straight forward manner. Finally, the test should work well (in terms of satisfying the above three properties) for cases when the assumed underlying distributions in the asymptotic theory are used in the evaluations via the numerical simulations in the small-sample cases. The performance of the test can also be evaluated for other distributions and the properties should be observed.

The general proposals for such small-sample approximations for the ASCLT-based tests can now be summarized as follows:

- From the statements of the ASCLT (presented in previous chapters), the role of sub-samples (or partial sums) involved in the process of estimating the quantiles can be observed. Though this would not be of concern for very large samples, this would affect the quantile-estimation process for small and even moderately-sized samples (say, sample size $N \leq 1000$). In order to overcome this problem, we propose to use permutations of the full sample in the process of estimation of the quantiles. The number of permutations to perform and how the results are computed based on these permuted samples will be described in the following sections of this chapter.

- As already seen in the previous chapter, the use of quantity $C_N = \sum_{n=1}^{N} \frac{1}{n}$, instead of $\log N$, in the ASCLT is proposed, where N is the sample size.

- The quantiles estimated via the ASCLT-based procedure(s) need to be 'adjusted' or 'transformed' to incorportate the information that these are based on very small samples compared to the large sample sizes that the asmyptotic theory assumes.

- From the Definition 13 of ASCLT-test method 2, we see that there is an unknown part, namely μ, in the proposed interval $A_\alpha^{(N)}$. This has to be obviously substituted with suitable estimator to make this method applicable.

The actual implementation of the proposals made above would be done in the course of this chapter. This would aid the better understanding of what these suggestions could mean in practice. It has to be made clear here that these suggestions and any modifications of the general theory presented earlier, in due course of this chapter, are all based on extensive simulation-based studies which were carried out during the period of the research. There is not enough mathematical justification or foundation for certain suggestions for modifications. This is one of the open problems emulating from this work and has be addressed in future. Thus, there could be comparable or better methods which may be arrived at in the future. By the sheer fact that there is no existing parallel method to the approach presented based on the ASCLT, one has to explore it to the core. So, the basic idea of the proposals made here is an indication towards further building, exploration and fine-tuning of these methods to form more formidable and reliable methods. Further, an aspect of the style of presentation is that, we would suggest considerations for future research and work, as and when such a proposal could follow naturally in an obvious manner. This is also due to numerous open problems and the wide possibilities of to-dos arising out of the new methodologies being proposed. There would, anyway, be a separate short section in the next chapter with a summary of only wider and major proposals for further research.

In this chapter three specific tests of hypothesis will be considered, namely

- the parametric one-sample test for mean, assuming normal distribution of the sample,
- the parametric two-sample test for equality of means, with the assumption of normal distribution of the samples involved, with and without the assumption of equal variances, and
- its nonparametric equivalent, without the assumption of normal distribution of the samples but having equal or unequal variances.

The two-sample testing cases, without the assumption of equal variances, described above, are commonly referred to as (parametric) *Behrens-Fisher problem* and *Nonparametric Behrens-Fisher problem*, respectively, and these will be formally defined in the following, relevant sections. Historical perspectives along with certain controversies surrounding these test problems are also discussed. Detailed review of several existing solutions for the problems are also presented with some associated software mentioned, wherever appropriate. Finally, a summary of the independent evaluations of these methods via monte carlo simulations is presented and discussed. The superiority, or at least a competing, comparative performance, of the ASCLT-based tests to existing methods is verified and established via the simulation results.

It has to be remarked that, being a new approach towards the testing of hypothesis, the ASCLT-based test would need many simulation-based testing and evaluation to establish its performance and/or superiority over other existing methods. Many such simulations, under varied setups with different sample sizes were performed and the results evaluated. Since the results had very similar pattern and due to considerations of space in the thesis, we present only very few of the results to provide a glance of the results. Some more results from similar simulations are planned to be made available via the webpage:

`http://neulat.stochastik.math.uni-goettingen.de/users/tknathan/asclt/`

in order to aid further usage and evaluations by researchers interested in this area.

All simulations were performed via programs written in `C` programming langauge and compiled used the `GNU C compiler 3.3.4` based on interface provided by `Cygwin 1.3`. The scientific computations within the `C` environment were performed using the `GNU Scientific Library (GSL) 1.5`. All the programming was performed using open source software, except for the Operating System, which was `Microsoft Windows XP Professional`.

4.2 One Sample Case

We consider the parametric one-sample situation assuming that the sample is normally distributed (with finite variance) and we will be interested in test of hypothesis regarding the mean. Independent comparisons of simulation-based analysis and results from two existing, standard procedures are made with the ASCLT-based tests (methods 1 and 2).

The notation is set here and it will be followed in the rest of this section on one-sample case. Let the vector of the sample of size N be denoted by $\mathbf{x} = (x_1, \ldots, x_N)$, the elements of which are assumed to be independent realisations of $\mathcal{N}(\mu, \sigma)$. We are then interested in testing a hypothesis,

$$H_0 : \mu = \mu_0 \quad \text{Vs} \quad H_1 : \mu \neq \mu_0; \qquad \text{w.l.o.g, we take } \mu_0 = 0. \tag{4.1}$$

We fix a significance level of 2α, which, for our simulations, would be either 5% or 10%.

The optimal test with such a set-up is the one-sample student's t-test, the description of which could be found in any basic, introductory book on statistics. So, the results of t-test via the simulations are computed and would be considered bench-mark result for comparing the results from other methods presented below. The t-test is available as a standard function in almost all available statistical software packages.

4.2.1 Bootstrap BC$_a$ Method

A short introduction to bootstrap methods is presented in subsection 4.3.2 (on page 45). Since the proposal with ASCLT-based tests would involve a random-interval based testing approach, a bootstrap method based on a similar idealogy is considered here. Construction of several confidence intervals (CIs) following the bootstrapping technique is available from the literature, one among which is the popular book by Efron and Tibshirani (1993). The reader is referred to Efron and Tibshirani (1993) and Carpenter and Bithell (2000) for a discussion on which CIs are applicable/appropriate under what situations. For our consideration, we will present the one which they recommend for wide usage, namely the BC_a (it is the abbreviation standing for *Bias-Corrected and Accelerated*) interval (c.f., Efron and Tibshirani, 1993, Chapter 14). The algorithm for construction of the BC_a interval intended for a $1-2\alpha$ coverage, can be summarized as follows.

`Begin Algorithm - BC`$_a$ `interval`

- Form B bootstrap[†] data sets $\mathbf{x}^{(*b)}$, $b=1,\ldots,B$, where $\mathbf{x}^{(*b)}$, are sampled with replacement from the original sample $\mathbf{x}=(x_1,\ldots,x_N)$.
- Evaluate the mean $\overline{x}_{\cdot}^{(*b)}, b=1,\ldots,B$ for each bootstrap data set. Let $\overline{x}_{\cdot}$ denote the mean of the original sample $\mathbf{x}$.
- The *bias*, say denoted by z_0, is then estimated using the following formula:

$$\widehat{z}_0 \;=\; \Phi^{-1}\left(\frac{\#\left\{\overline{x}_{\cdot}^{(*b)} < \overline{x}_{\cdot}\right\}}{B}\right),$$

  where Φ^{-1} denotes the inverse function of the standard normal cumulative distribution funtion.
- Let the notation $\mathbf{x}_{(i)}$ denote the original sample with the i$^{\text{th}}$ sample point x_i removed ($\Longrightarrow$ the size of $\mathbf{x}_{(i)}$ is $N-1$.) Let then $\overline{x}_{(i)}$ denote mean of the sample $\mathbf{x}_{(i)}$ and let $\overline{x}_{(\cdot)} = \sum_{i=1}^{N}\overline{x}_{(i)}/N$, the mean of quantities $(\overline{x}_{(1)},\ldots,\overline{x}_{(N)})$. The *acceleration* is now estimated with the following expression:

$$\widehat{a} = \frac{\sum_{i=1}^{N}\left(\overline{x}_{(\cdot)}-\overline{x}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{N}\left(\overline{x}_{(\cdot)}-\overline{x}_{(i)}\right)^2\right\}^{3/2}}.$$

- The computed $\overline{x}_{\cdot}^{(*b)}$'s are sorted in ascending order and vector $(\overline{x}_{\cdot}^{(*,(1))},\ldots,x_{\cdot}^{(*,(B))})$ is formed.

[†]Bootstrap data sets are those which are formed by re-sampling (with replacement) from the original samples. These samples are used in the computation of relevant bootstrap estimates.

- Finally, the BC_a interval is given by

$$\text{BC}_a = \left(\overline{x}_{\cdot}^{(*,(B\alpha_1))}, \overline{x}_{\cdot}^{(*,(B\alpha_2))}\right),$$

where,

$$\alpha_1 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z^{(\alpha)}}{1 - \widehat{a}\left(\widehat{z}_0 + z^{(\alpha)}\right)}\right)$$
$$\alpha_2 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z^{(1-\alpha)}}{1 - \widehat{a}\left(\widehat{z}_0 + z^{(1-\alpha)}\right)}\right),$$

and where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $z^{(\alpha)}$ is the $100\alpha^{\text{th}}$ percentile point of a standard normal distribution. Moreover, if $B\alpha_1$ and $B\alpha_2$ are non-integers, then they are approximated to the closest interger.

`End Algorithm - BC`$_a$ `interval`

This algorithm is implemented in SPlus 7/R 2.2.0 in the library `bootstrap` with the function `bcanon`. There is also a macro written in SAS and made available in Tibshirani (1985).

Using the above algorithm, once the BC_a interval is computed, we consider that there is evidence in support of the hypothesis H_0 defined in (4.1), if $\mu_0 \in \left(\overline{x}_{\cdot}^{(*,(B\alpha_1))}, \overline{x}_{\cdot}^{(*,(B\alpha_2))}\right)$; otherwise the hypothesis can be rejected.

4.2.2 ASCLT Tests

Here the modified, small-sample approximation for ASCLT-based tests will be presented for the parametric one-sample situation under consideration.

As said in the Introduction of this chapter, we will propose certain modifications and transformations to the theoretical proposal of ASCLT-based tests in order to make them applicable in situations of small sample sizes.

`Begin Algorithm - ASCLT-based parametric one-sample test for mean`

Let the i$^{\text{th}}$ permuted sample vector be denoted and given by

$$\mathbf{x}^{*i} = \text{permute}(\mathbf{x}), \qquad i = 1, \ldots, \text{nper},$$

where nper represents the total number of permutations of the sample we are interested in.

Now for each of the permuted sample, we compute weighted partial means as

$$SS_n^{*i} \; = \; n \cdot \frac{\overline{x}_n^{*i}}{\sqrt{n}} \; = \; \sqrt{n} \cdot \overline{x}_n^{*i}\,, \qquad n = 1, \ldots, N,$$

where $\overline{x}_n^{*i}$ denotes the mean of the partial, i$^{\text{th}}$ permuted sample $(x_1^{*i}, \ldots, x_n^{*i})$, $n = 1, \ldots, N$.

The average of these SS_n^{*i} is given by,

$$\overline{SS}^{\,*i} = \frac{1}{N} \sum_{n=1}^{N} SS_n^{*i}.$$

For $\alpha \in (0,1)$, the quantiles $\widehat{q}_\alpha^{*i,(N)}$ and $\widehat{q}_{1-\alpha}^{*i,(N)}$ are now estimated via the relationship (2.12) defined by the ASCLT in chapter 2:

$$\begin{aligned}
\widehat{q}_\alpha^{*i,(N)} &= \max\left\{ q \;\middle|\; C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} 1_{\{SS_n^{*i} < q\}} \leq \alpha \right\} \\
\widehat{q}_{1-\alpha}^{*i,(N)} &= \max\left\{ q \;\middle|\; C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} 1_{\{SS_n^{*i} < q\}} \leq 1-\alpha \right\}
\end{aligned}$$

where $C_N = \sum_{n=1}^{N} \frac{1}{n}$.

We split the algorithm into sub parts to implement the two methods proposed via Definitions 12 and 13 (on page 16).

`Begin Sub-Algorithm: ASCLT-test method 1`

$$\overline{q}_\alpha = \frac{\sum_{i=1}^{\text{nper}} \widehat{q}_\alpha^{*i,(N)}}{\text{nper}} \quad \text{and} \quad \overline{q}_{1-\alpha} = \frac{\sum_{i=1}^{\text{nper}} \widehat{q}_{1-\alpha}^{*i,(N)}}{\text{nper}}$$

Reject H_0 if

$$0 \notin \left[\overline{q}_{\alpha},\ \overline{q}_{1-\alpha}\right]. \tag{4.2}$$

`End Sub-Algorithm: ASCLT-test method 1`

`Begin Sub-Algorithm: ASCLT-test method 2`

In an attempt to implement the result of Theorem 11, we propose to transform the estimated quantiles as follows and call them *transformed quantiles*,

$$\widehat{q}^{\,*i,(N)}_{\alpha,\text{trans.}} = \frac{\left(\overline{SS}^{\,*i} - \widehat{q}^{\,*i,(N)}_{1-\alpha}\right)}{\sqrt{N}} \quad \text{and} \quad \widehat{q}^{\,*i,(N)}_{1-\alpha,\text{trans.}} = \frac{\left(\overline{SS}^{\,*i} - \widehat{q}^{\,*i,(N)}_{\alpha}\right)}{\sqrt{N}}.$$

Further we denote the averages of the estimated, transformed quantiles over the different permuted samples as,

$$\overline{q}_{\alpha,\text{trans.}} = \frac{\sum_{i=1}^{\text{nper}} \widehat{q}^{\,*i,(N)}_{\alpha,\text{trans.}}}{\text{nper}} \quad \text{and} \quad \overline{q}_{1-\alpha,\text{trans.}} = \frac{\sum_{i=1}^{\text{nper}} \widehat{q}^{\,*i,(N)}_{1-\alpha,\text{trans.}}}{\text{nper}}.$$

Finally we propose an adjustment of these quantiles above, with the following expressions:

$$\begin{aligned}
\overline{q}_{\alpha,\text{fin}} &= \frac{\overline{q}_{\alpha,\text{trans.}} - 2\alpha \cdot \lambda_{N,2\alpha} \cdot \left(\overline{q}_{\alpha,\text{trans.}} + \overline{q}_{1-\alpha,\text{trans.}}\right)}{\kappa_{N,2\alpha}} \\
\overline{q}_{1-\alpha,\text{fin}} &= \frac{\overline{q}_{1-\alpha,\text{trans.}} - 2\alpha \cdot \lambda_{N,2\alpha} \cdot \left(\overline{q}_{\alpha,\text{trans.}} + \overline{q}_{1-\alpha,\text{trans.}}\right)}{\kappa_{N,2\alpha}},
\end{aligned}$$

where $\kappa_{N,2\alpha}$ and $\lambda_{N,2\alpha}$ are constant coefficients for a given N and 2α, and they have to be numerically determined for each sample size $N \in \mathbb{N}$ and significance level 2α.

Now, reject H_0 if

$$\overline{x}_N \notin \left[\overline{q}_{\alpha,\text{fin}},\ \overline{q}_{1-\alpha,\text{fin}}\right],$$

where $\overline{x}_N$ denotes the mean of the original sample $\mathbf{x}$.

`End Sub-Algorithm: ASCLT-test method 1`

`End Algorithm - ASCLT-based parametric one-sample test for mean`

Table 4.1: Determined coefficients, $\kappa_{n,\alpha}$ and $\lambda_{n,\alpha}$, for different sample sizes and significance levels 2α

	$2\alpha = 5\%$		$2\alpha = 10\%$	
N	$\kappa_{N,2\alpha}$	$\lambda_{N,2\alpha}$	$\kappa_{N,2\alpha}$	$\lambda_{N,2\alpha}$
10	0.52	2.9	0.51	3.9
15	0.59	2.65	0.58	3.6
20	0.645	2.45	0.624	3.3
25	0.68	2.1	0.67	3.1
30	0.72	2.0	0.70	3.0

Through simulation studies, the constant coefficients $\kappa_{N,2\alpha}$ and $\lambda_{N,2\alpha}$ have been found for some select values of sample sizes and significance levels. The values for these coefficients were basically determined by more of an trial-and-error approach, using the required significance level to be achieved and also sometimes trying to match the results from the t-test. Such numerically determined values are presented in Table 4.1. As can be seen, though the finding of the values of these coefficients were not done from a mathematical perspective, they have some pattern (monotonously increasing or decreasing), implying the fact that they are not random. One of the main, immediate further research topic of this thesis is to determine the mathematical foundation of such values of these coefficients and to derive a general formula to compute these for any N and 2α.

4.2.3 Simulation results

Monte carlo simulations were performed to observe the properties of the t-test, BC_a (interval) method and the ASCLT-based methods 1 and 2, in maintaining the pre-assigned significance level 2α and regarding power. Following are the common setting for these simulations:

1. Total number of simulation runs for each result presented, say N_{SIM}, is 10000.
2. For BC_a method, $B = 2000$.
3. For both methods of ASCLT-based tests, nper $= 2000$.
4. All tests were two-sided and performed at a significance level of $2\alpha = 5\%$ or 10%.

Table 4.2: Results for simulated level for 2α (in %) under $H_0 : \mu = 0$ for samples from $\mathcal{N}(\mu = 0, \sigma)$, for varying σ given in the first column. Columns labelled 'A-M1' and 'A-M2' represent the results from ASCLT-test methods 1 and 2, respectively.

σ	t-test	BC_a	A-M1	A-M2	t-test	BC_a	A-M1	A-M2
	$n = 10$ and $2\alpha = 5\%$				$n = 10$ and $2\alpha = 10\%$			
1.0	5.05	10.24	*19.70*	4.98	9.99	15.37	*20.45*	9.92
1.5	5.46	10.75	*20.68*	5.31	10.00	15.29	*21.03*	9.92
2.0	5.22	10.18	*19.51*	5.22	9.76	14.78	*20.20*	9.73
3.0	5.24	10.68	*20.70*	5.26	10.07	15.53	*20.81*	10.20
4.0	5.22	10.00	*19.93*	4.99	9.92	15.25	*20.76*	10.00
5.0	5.03	9.88	*19.92*	4.89	10.06	15.96	*21.56*	10.26
	$n = 15$ and $2\alpha = 5\%$				$n = 15$ and $2\alpha = 10\%$			
1.0	4.71	7.93	*10.08*	4.68	10.13	13.22	*11.06*	10.02
1.5	5.01	8.27	*10.42*	4.99	9.91	13.56	*11.17*	9.99
2.0	4.86	8.73	*10.90*	5.06	10.31	13.75	*11.57*	10.38
3.0	5.14	8.25	*10.76*	5.23	10.40	14.05	*11.82*	10.61
4.0	4.90	7.94	*10.37*	4.96	9.81	13.29	*11.27*	10.05
5.0	5.20	8.32	*10.95*	5.41	9.55	12.86	*10.89*	9.73

As mentioned in the beginning of this chapter, only few parts of all the simulations that we performed are reported here since the general trend of the results remain the same over several settings. Simulations evaluating the property of maintaining significance level are presented in Table 4.2 and those relating to the study of power are plotted in Figure 4.1. Note that, in the table, the results corresponding to ASCLT-test method 1 are presented in italized text. By this we intend to imply that, we do not recommend using this method for real-data analysis purposes. But the results are presented here only to observe and show the general performance of this method and for possible further work.

It is clear from the table and the figure that the t-test maintains the pre-assigned level 2α and at the same time exhibiting good power. On the other hand, the BC_a method displays very liberal performance with respect to maintaining the pre-assigned level 2α. This result comes in as quite a surprise given the support it has been provided independently by Efron and Tibshirani (1993) and Carpenter and Bithell (2000). But as we observed the pattern, the liberality of this method reduces as the sample size increases. But even for sample size $n = 30$ for $2\alpha = 5\%$, the method had, on the

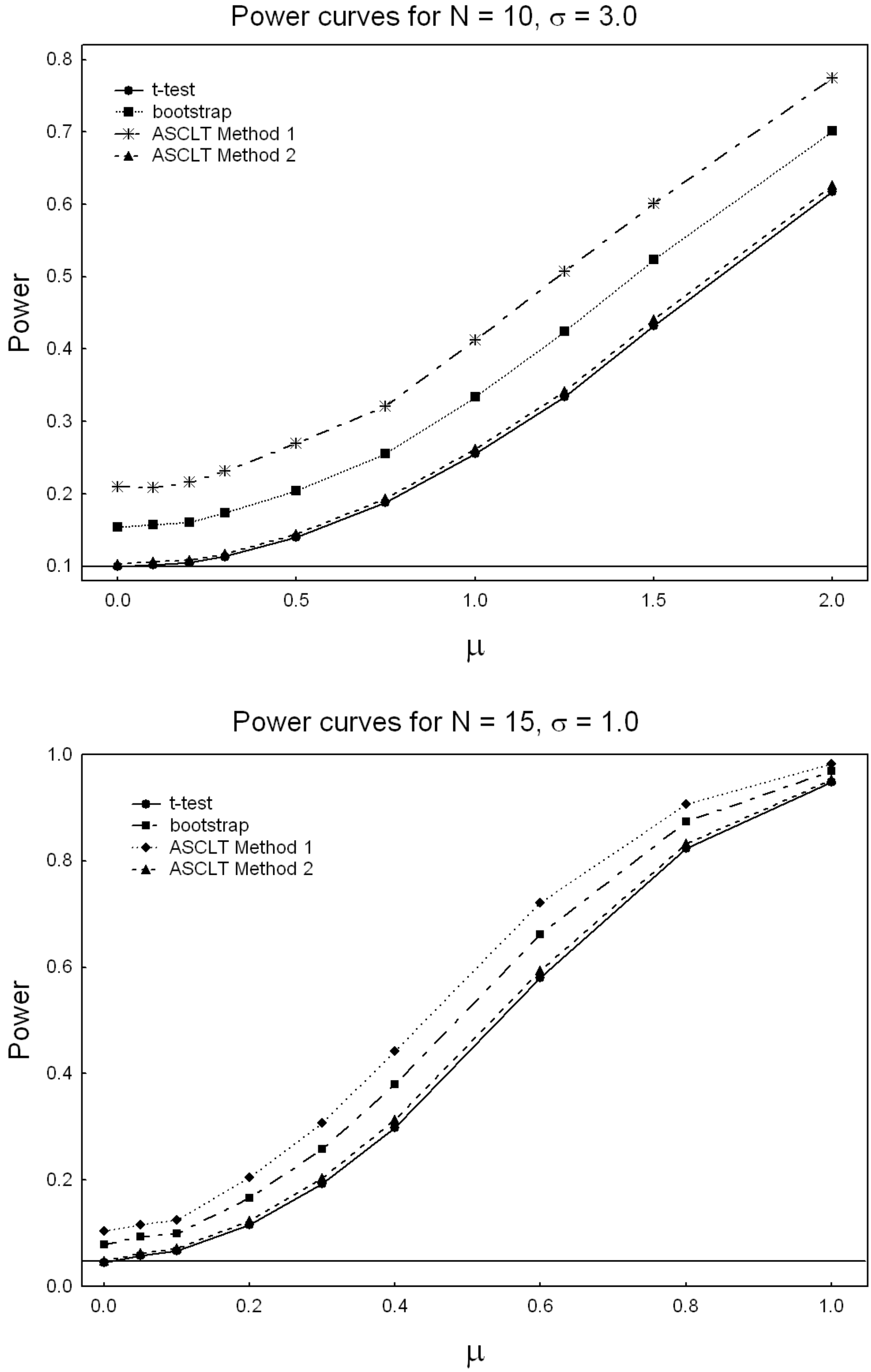


Figure 4.1: Power curves of different methods for $N = 10$ at significance level $2\alpha = 10\%$, and $N = 15$ at significance level $2\alpha = 5\%$, based on samples generated from $\mathcal{N}(\mu, \sigma = 3)$ and $N(\mu, \sigma = 1)$, respectively. (μ's are varying and it is plotted in the x-axis.)

average, an estimated type-I error rate of 6.5%. The results of ASCLT-test method 1 are provided only for observational purpose and not for real comparison with other methods. Finally, the ASCLT-test method 2 displays comparable results to the t-test, both in terms of maintaining level and power property. In fact, in many cases relating to the power, the ASCLT-test method 2 has, approximately, 1% better power than t-test in spite of maintaining level 2α. Thus, a completely satisfactory performance of the ASCLT-test method 2 is shown via these simulations concerning level and power.

4.3 Two-sample case - Behrens Fisher Problem

Often in medical data, the problem of unequal variances of observations between the groups is encountered. Considerable literature has been developed in this area and even in the simple two-sample situation, this problem is still actively tackled in order to obtain accurate results. In spite of the vast literature, a common drawback of many such sophisticatedly developed methodology is that many of them are based on strong underlying assumptions and strict model conditions. In this section, an overview of one such frequently encountered and widely addressed situation, namely Behrens-Fisher Problem, will be presented, which will be followed by the application of ASCLT-test methods to the problem. Numerical as well as graphical illustrations of simulation-based analysis will be used to demonstrate the performance of the ASCLT-test, along with frequently used competing methods from literature. Subsequent comparison with results from some existing standard methods are also presented. Let us first set the formal definition of the Behrens-Fisher Problem.

Definition 17 (Behrens-Fisher Problem). *Suppose that*

$$\mathbb{E} = \left(\mathbb{R}^{n_1 \times n_2}, \mathcal{B}, \left\{F^{n}_{\mu_1,\sigma_1} \times F^{m}_{\mu_2,\sigma_2} \,|\, \mu_1, \mu_2 \in \mathbb{R}, \sigma_1^2, \sigma_2^2 \geq 0\right\}\right)$$

is a statistical experiment[†]*, where* $n_1, n_2 \in \mathbb{N}$, $(\mathbb{R}^{n_1 \times n_2}, \mathcal{B})$ *represents the (measurable) sample space and* $\left\{F^{n_1}_{\mu_1,\sigma_1} \times F^{n_2}_{\mu_2,\sigma_2} \,|\, \mu_1, \mu_2 \in \mathbb{R}, \sigma_1^2, \sigma_2^2 \geq 0\right\}$ *is the product of family of probability measures. Further, let* $\mathcal{T} = (H_0, H_1)$ *be a partition of* $\left\{\mu_1, \mu_2 \in \mathbb{R}, \sigma_1^2, \sigma_2^2 \geq 0\right\}$*. Then, the experiment* $\mathbb{E}$ *along with partition* $\mathcal{T}$*, is called a Behrens-Fisher Problem if* $H_0 = \{(\mu_1, \mu_2, \sigma_1, \sigma_2) \,|\, \mu_1 = \mu_2\}$*.*

[†]For basic definitions (from a mathematical perspective) relating to hypothesis testing, we refer to Strasser (1985).

For the parametric Behrens-Fisher problem, in the above definition, the F's are taken to be Normal distributions. Thus, from an applied point of view, for i.i.d $\mathcal{N}\left(\mu_i, \sigma_i^2\right)$ random variables $(X_{i1}, \ldots, X_{in_i})$, $i = 1, 2$, the parametric Behrens-Fisher problem is that of testing

$$H_0 : \mu_1 = \mu_2 \quad \text{Vs.} \quad H_0 : \mu_1 \neq \mu_2, \tag{4.3}$$

without the assumption of $\sigma_1^2 = \sigma_2^2$.

4.3.1 Behrens Fisher Problem - Overview

Behrens (1929) proposed a method to test the hypothesis of comparing means of two samples from normal distributions with unequal variances. Fisher (1935, 1941) developed a solution (for the test proposed by Behrens, 1929) using *fiducial* inference-arguments. Since then the test problem has been popularly called as *Behrens-Fisher Problem* (and sometimes also referred to as Fisher-Behrens Problem or, simply, Behrens Problem). For brevity, we will use abbreviation 'BFP' to read as 'Behrens-Fisher Problem'. Moreover, considering the setup addressed by Behrens (1929) as a stituation of *parametric* case of BFP, later a generalized version of BFP without the assumption of normally distributed samples was introduced and has also been extensively studied in the literature. This case has been called as the 'Generalized BFP' or also as 'Nonparametric BFP', which will be discussed in the next section.

From the time of the article by Behrens (1929), BFP has been an active topic of consideration in statistical literature. It was so common and widely addressed that it prompted the well-known statistician Professor Scheffé to write (as the first sentence of one of his papers),
"*The most frequently occuring problem in applied statistics is, in my opinion ... the Behrens-Fisher Problem ...*" (Scheffé, 1970).

The problem has being approached via different techniques and with wide range of mathematical aspects and treatment. To mention a few, there are exact tests (e.g. Behrens, 1929, or Bayesian approaches) and approximate tests (e.g. Smith, 1936; Satterthwaite, 1946; Welch, 1947; Aspin, 1948; Cochran and Cox, 1957; Howe, 1974); Bayesian solutions (e.g. Jeffreys, 1939; Patil, 1964; Duong and Shorrock, 1992; Lee, 1997), Bootstrapping and Re-sampling methods (e.g. Beran, 1988; Hall and Martin, 1988; Compagnone and

Denker, 1996; Babu and Padmanabhan, 2002; Reiczigel et al., 2005), Confidence-Interval approach (e.g. Banerjee, 1961; Howe, 1974; Lee, 1997; Wang and Chow, 2002), Likelihood Ratio test (e.g. Bozdogan and Ramirez, 1986; Troendle, 2002; Dong, 2004), Permutation methods (e.g. Pitman, 1937; Janssen, 1997), Generalized p-value approach (e.g. Weeranhandi, 1995) and Rank-based approaches for Generalized BFP (e.g. Sen, 1962; Fligner and Policello II, 1981; Brunner and Neumann, 1982, 1986; Brunner and Munzel, 2000). (This list also includes solutions for nonparametric BFP). It can be noted that the literature spans from early Twentieth century to very recent times.

Furthermore, historical views, reviews and controversies (particularly between the approaches proposed by Behrens-Fisher and those by Satterthwaite-Welch-Smith) surrounding the developments of the solutions are available from Robinson (1976); Wallace (1980); Barnard (1984); Moser and Stevens (1992); Miller (1997) and Lee (1997). Thomasse (1974) gives summarized practical recipes of several popular, interesting and even controversial methods that were available during the time of his writing.

4.3.2 Solutions for BFP

In this subsection, a brief overview and description of the algorithm of certain commonly used methods for the (parametric) BFP is given. It should be emphasized that there are many authors who consider one or the other method as more, or even most, favorable compared to others. Here, as a general rule, the methods suggested most in our experience from applications and literature survey are presented and discussed. First, a common set of notation and terminology for all the procedures/solutions for BFP described below, is presented here. Let the vector of i.i.d random variables corresponding to the two samples be denoted by $\mathbf{X}_{(i)} = (X_{i1}, \ldots, X_{in_i})'$, $i = 1, 2$, and the corresponding observed sample values be $\mathbf{x}_{(i)} = (x_{i1}, \ldots, x_{in_i})'$, $i = 1, 2$, where n_1 and n_2 are the respective sample sizes. Now, let $\overline{x}_{1\cdot}$ and $\overline{x}_{2\cdot}$ denote the respective sample means. Also, let s_1^2 and s_2^2 denote the corresponding empirical variances. Finally, let the notation $t_{a,b}$ denote the a^{th} quantile of a (central) t-distribution with b degrees of freedom.

SWS approximation

This approximation is due to the independent results from Smith (1936), Satterthwaite (1946) and Welch (1947). The idea here is to reject H_0, defined in Definition 17, at a given level of significance 2α if,

$$\frac{|\overline{x}_{1\cdot} - \overline{x}_{2\cdot}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad > \quad t_{2\alpha,r} \quad ,$$

where the approximate degrees of freedom r is estimated by,

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

This test has been widely recommended in literature since it maintains the pre-assigned significance level, 2α, and also exhibits good power properties. For example, Mehta and Srinivasan (1970); Lee and Gurland (1975) and, more recently, Wang and Chow (2002) have conducted extensive study of many solutions available for the BFP. These and several other authors concluded, or at least remarked, that the SWS approximation works the best for the BFP among the several methods that they compared. Because of its good performance, this approximation has been widely implemented in several statistical software including Microsoft MS-Office XP Professional Excel (via `TTEST`), SAS 9.1 (via procedure `PROC TTEST`), SPlus 7/R 2.2.0 (via `t.test` function), SPSS 13.0 (via `T-TEST`) and STATA 9.0 (via `.ttest`). We will thus consider the SWS approximation as a gold-standard test for comparisons (in the situation of parametric BFP) in our simulation studies presented in the later part of this section.

Software STATISTICA 7 also offers `T-test` procedure providing for BFP, but the method used for this purpose is not explicitly stated in the documentation. But from what we have observed with few example data sets, it became clear that STATISTICA 7 also implements the SWS approximation.

Cochran-Cox's approximation

According to Cochran and Cox (1957), for a given level of significance 2α, the null hypothesis of H_0, given in Definition 17, can be rejected if,

$$\frac{|\overline{x}_{1\cdot} - \overline{x}_{2\cdot}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad > \quad \frac{t_{1-\alpha,n_1-1}\frac{s_1^2}{n_1} + t_{1-\alpha,n_2-1}\frac{s_2^2}{n_2}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \tag{4.4}$$

Note that for $n_1 = n_2 = n$ (say), RHS of the above inequality is just $t_{1-\alpha,n-1}$. It is well known that this test is quite conservative in maintaining pre-assigned significance level, 2α (c.f. Lee and Gurland, 1975). In their original proposal, Cochran and Cox (1957) make it clear that during the time of publishing their work, there was extensive dependence on the printed Fisher and Yates statistical tables for the BFP. Thus, Cochran and Cox (1957) suggested the above approximation (4.4), in cases when these tables were "not readily accessible". They also agree that the proposed approximation (4.4) "probably errs slightly on the conservative side". According to our knowledge, particularly with the advent of extensive computing systems, there has not been any account of when the test is recommended in preference to other competing methods. This approximation has been implemented in SAS's `PROC TTEST`.

Bootstrap Method

Bootstrap methods (and related idea of *jackknife* and *resampling* procedures) were introduced during mid-Twentieth century. But they have been extensively explored only during the last two decades or so. One of the main reasons for such a recent popularity is owing to the availability of stronger and powerful computational capacity and efficiency. For a technical introduction of Bootstrap methods we refer to Davison and Hinkley (1997), and for applied overview, Efron and Tibshirani (1993).

There are several proposals of bootstrap two-sample t-test, particularly addressing the BFP, available from the literature. We will be using the one which is widely used in practice due to Efron and Tibshirani (1993).

One of the general ideas of bootstrap test of hypothesis (i.e., not restricted to the BFP) is to estimate the so-called *Achieved Significance Level* (ASL) based on the observed dataset(s) and the concerned (null) hypothesis is rejected or accepted depending on whether $ASL \leq 2\alpha$ or $ASL > 2\alpha$, respectively, where 2α is the pre-assigned significance level. In their book, Efron and Tibshirani (1993) propose a method to address the problem of testing equality of means without making any assumptions on the equality (or unequality) of the variances. Moreover, they recommend and perform only one-sided test of hypothesis. The reason for them to do so, according to them, is that "The two-sided ASL is always larger than the one-sided ASL, giving less reason for rejecting H_0. The two-sided test is inherently more conservative"(pp. 212). This argument may suit a general-purpose application. But for biostatistical investigations two-sided tests are widely preferred and, sometimes, also expected to be performed by regulatory authorities. So, their proposed algorithm is slightly modified (according to their suggestion) to implement a two-sided test (refer to Efron and Tibshirani, 1993, page 212 and Algorithm 16.2 on page 224). This modified algorithm can be summarized as follows:

`Begin Algorithm - A Bootstrap solution for BFP`

- Transform the observed variables to $\tilde{x}_{ij} = x_{ij} - \overline{x}_{i\cdot} + \overline{z},\ i = 1, 2;\ j = 1, \ldots, n_i$, where $\overline{z}$ is the mean of the combined sample $(x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{1n_2})$.
- Form B bootstrap† data sets $\left(\mathbf{x}_1^{(*b)}, \mathbf{x}_2^{(*b)}\right)$, $b = 1, \ldots, B$, where $\mathbf{x}_i^{(*b)},\ i = 1, 2$, are sampled with replacement from $(\tilde{x}_{i1}, \ldots, \tilde{x}_{in_i})$.
- Evaluate for each data set,

$$t^{(*b)} = \frac{\overline{x}_{1\cdot}^{(*b)} - \overline{x}_{2\cdot}^{(*b)}}{\sqrt{\frac{s_1^{2,(*b)}}{n_1} + \frac{s_2^{2,(*b)}}{n_2}}},\ b = 1, \ldots, B, \tag{4.5}$$

  where $\overline{x}_{1\cdot}^{(*b)}$ and $\overline{x}_{2\cdot}^{(*b)}$ are respective sample means and $s_1^{2,(*b)}$ and $s_2^{2,(*b)}$ are respective variances of each of the bootstrap data sets $\mathbf{x}_1^{(*b)}$ and $\mathbf{x}_2^{(*b)}$, $b = 1, \ldots, B$.
- Calculate the approximate *Achieved Significance Level*, $\widehat{\mathrm{ASL}}$, given by,

$$\widehat{\mathrm{ASL}} = \frac{\#\left\{|t^{(*b)}| \geq |t_{\mathrm{obs}}|\right\}}{B},$$

  where, t_{obs} is the t-value computed by using the original samples (instead of the bootstrap data sets) in (4.5).
- Reject H_0 if $\widehat{\mathrm{ASL}} \leq 2\alpha$.

`End Algorithm - A Bootstrap solution for BFP`

According to our knowledge, this algorithm has not been implemented directly in any of the existing standard statistical software packages. This is the case even in SPlus or R, where the authors have their own ‘library’ of functions. But it is quite straight-forward to implement the above algorithm using SAS’s IML (Interactive Matrix Language) or SPlus’s/R’s standard commands and functions.

†Bootstrap data sets are those which are formed by re-sampling (with replacement) from the original samples. These samples are used in the computation of relevant bootstrap estimates.

Bayesian t-test

Bayesian approach towards hypothesis testing is based on a completely different philosophy compared to the conventional, *frequentist* approach of hypothesis testing. But, since the approach is gaining popularity among more and more users, we include a Bayesian test here in order to facilitate comparison to other competing methods. We will be using the recently proposed Bayesian version of two-sample t-test by Gönen et al. (2005). Since the Bayesian methodology being a whole area of itself, we intend to present only the minimum explanation here, so as to make the implementation of the method as clear as possible than to going into the theoretical aspects of them. For an introduction to general Bayesian theory we refer to Lee (1997); Bolstad (2004) and for a more practical view, Woodworth (2004).

The general, fundamental Bayesian idea is to combine the sample information along with *prior* knowledge in terms of probabilities in deriving the final *posterior* probabilities (or probability distributions) and thus basing the inferences on them. The Bayesian formulation of hypothesis testing for two sample t-test involves placing prior probabilities π_0 and π_1 ($\pi_0 + \pi_1 = 1$) on the respective hypotheses $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$. Then these probabilities are 'updated' via the Bayes' theorem to obtain the posterior probabilities

$$P\left(H_j|\,\mathrm{data}\right) \quad = \quad \frac{\pi_j P\left(\mathrm{data}\,|H_j\right)}{\pi_0 P\left(\mathrm{data}\,|H_0\right) + \pi_1 P\left(\mathrm{data}\,|H_1\right)}, \quad j = 0, 1,$$

where $P(H_j|\,\mathrm{data})$ denotes the marginal density of the data under H_j. It is suggested often to use the *Bayes Factor*(BayFac), as defined below, instead of the formulation above in order to overcome partially the sensitivity of the posteriors on the priors π_0 and π_1:

$$\mathrm{BayFac} \quad = \quad \frac{P\left(\mathrm{data}\,|H_0\right)}{P\left(\mathrm{data}\,|H_1\right)}.$$

It is interpreted as the data providing evidence in support of H_0 when BayFac ≥ 1, and against H_0 if BayFac < 1.

Gönen et al. (2005) consider the set up the two-sample problem such that the random variables $X_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)\,, i = 1, 2;\ j = 1, \ldots, n_i$. Then the pooled-variance, two-sample

t-statistic is defined and denoted by

$$t = \frac{\overline{X}_{1\cdot} - \overline{X}_{2\cdot}}{s_p/\sqrt{n_\delta}}, \tag{4.6}$$

where $s_p^2 = \left((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right)/(n_1 + n_2 - 2)$ is the pooled variance estimate and $n_\delta = 1/(n_1^{-1} + n_2^{-1})$. For testing

$$H_0 : \mu_1 = \mu_2 \quad \text{Vs.} \quad H_1 : \mu_1 \neq \mu_2, \tag{4.7}$$

Gönen et al. (2005) propose the following form of the Bayesian Factor:

$$\text{BayFac} \quad = \quad \frac{T_\nu\,(t|0,1)}{T_\nu\left(t\,|\,\lambda\sqrt{n_\delta}, 1 + n_\delta\sigma^2\right)}, \tag{4.8}$$

where t is defined in (4.6), λ and σ^2 are the prior mean and variance of the standardised effect size $(\mu_1 - \mu_2)/\sigma$ under H_1 (defined in equation 4.7), $\nu = n_1 + n_2 - 2$, and notation $T_a(\cdot|b, c)$ denotes the probability density function of non-central t-distribution having location b, scale $\sqrt{c}$ and degrees of freedom a.

Instead of going into the merits and discussion surrounding the choice of priors to be implemented in (4.8), we directly provide here the general rule-of-thumb choice proposed by Gönen et al. (2005). They suggest the use of sample size-power relationship to estimate these priors to be used in (4.8) as follows (assuming a power of 80%):

- for $2\alpha = 5\%$: $\lambda = \frac{2.80}{\sqrt{n_\delta}}$ and $\sigma_\delta = \frac{2.19}{\sqrt{n_\delta}}$, and
- for $2\alpha = 10\%$: $\lambda = \frac{2.49}{\sqrt{n_\delta}}$ and $\sigma_\delta = \frac{1.94}{\sqrt{n_\delta}}$,

where n_δ is defined in (4.6).

The following important points have to be remarked with respect to this procedure.

- It is well known that a careful selection of the prior distributions is an important step while using Bayesian methods. This usually requires expert assistance. Gönen et al. (2005) also gives clear indication to this effect, though they recommend a general-purpose prior based sample size-power relationship assuming underlying normal distribution of the samples. We will use these priors in our simulations.

It will be evident from the simulation results as to how crucial and important, careful selection of priors is when using Bayesian methods.

- The Bayesian procedure by Gönen et al. (2005) assumes homoscedastic variances, thus can not be applied to the BFP. However, as stated above, our intention is to merely compare the results among competing methods, at least in situations where they are validly applicable. For sake of completeness, we will also present the results of simulations with heteroscedastic variances, but in such cases we will denote the results for Bayesian test with italized text in order to represent that the method should not have been used 'in principle'. This is explained more in Section 4.3.4, where the simulation results are presented and discussed.

4.3.3 ASCLT-test for BFP

As mentioned earlier in Chapter 2, one of the major advantages of the ASCLT-based tests of hypothesis is that the procedures do not involve variance estimation in them. Though the origination of BFP is based on the concept of unequal variances, the study by both Moser and Stevens (1992) and Sprott and Farewell (1993) give detailed accounts of how important and crucial the assumption of equal or unequal variances (or regarding the ratio of the variances) between the two concerned samples are, in order to arrive at inferences regarding the population mean difference[†]. Thus, BFP would be one of the ideal situations to demonstrate the ASCLT-tests' effectiveness.

Further more, it has to be mentioned here that the general hypothesis-testing theory developed in Chapter 2 considers an one-sample setup, i.e., only one sequence of identically and independently distributed variables was considered. Formally one has to extend this theory to the two or more variables' case. But in view of Lindeberg-Levy version of ASCLT, and noting that the variables here are independent, but not necessarily identically distributed, the theory can be derived for such extensions. We assume the result for a two-sample case under discussion, and proceed with the two-sample approximate tests based on ASCLT.

With extensive computations and programming, we have identified modifications to the ASCLT-based general theory of hypothesis-testing from Chapter 2 to make it applicable

[†]We refer to the unpublished work by Dickson available online at *http://uts.cc.utexas.edu/∼ccbv001/stats/cond.html* for an interesting, general discussion and summary on assumptions and pre-testing of equality of variance.

to the BFP in discussion. The proposed modifications are presented here as the following algorithm. We will follow the same notation and terminology for the samples defined in the beginning of Section 4.3.2, and additionally let $n = \min(n_1, n_2)$.

`Begin Algorithm - ASCLT tests for parametric two-sample testing problem`

The two samples are <u>independently</u> permuted nper times (nper is chosen by user - as a general rule, the larger the value of nper, the better is the approximation), with p^{th} permuted sample vector denoted and given by

$$\mathbf{x}_i^{*p} = \text{permute}(\mathbf{x}_i), \qquad i = 1, 2;\ p = 1, \ldots, \text{nper}\,.$$

Note that this are not "bootstrapping", but pure permutations, that is, only 'shuffling' the data points with each vector of the sample $\mathbf{x}_i$, $i = 1, 2$. Note also that independent permutations imply that the shuffling scheme for the first sample in the p^{th} permutation is highly likely to be different than the shuffling scheme for the second sample.

Now for each of the permuted sample, we compute quantities:

$$SS_k^{*p} \;=\; k \cdot \frac{\overline{x}_{1,k}^{\,*p} - \overline{x}_{2,k}^{\,*p}}{\sqrt{k}} \;=\; \sqrt{k} \cdot \left(\overline{x}_{1,k}^{\,*p} - \overline{x}_{2,k}^{\,*p}\right), \qquad k = 1, \ldots, n, \tag{4.9}$$

where $\overline{x}_{i,k}^{\,*p}$, $i = 1, 2$, denotes the mean of the partial sample $\left(x_{i1}^{*p}, \ldots, x_{ik}^{*p}\right)$ of size k. And let, $\overline{SS}^{\,*p} = \frac{\sum_{k=1}^{n} SS_k^{*p}}{n}$ for each permutation $p = 1, \ldots, \text{nper}$.

The quantiles $\widehat{t}_{\alpha}^{*p,(n)}$ and $\widehat{t}_{1-\alpha}^{*p,(n)}$ are now estimated by

$$\begin{aligned}
\widehat{t}_{\alpha}^{*p,(n)} &= \max\left\{ t \;\middle|\; C_n^{-1} \sum_{k=1}^{n} \frac{1}{k} 1_{\{SS_k^{*p} < t\}} \le \alpha \right\} \\
\widehat{t}_{1-\alpha}^{*p,(n)} &= \max\left\{ t \;\middle|\; C_n^{-1} \sum_{k=1}^{n} \frac{1}{k} 1_{\{SS_k^{*p} < t\}} \le 1 - \alpha \right\}
\end{aligned}$$

where $C_n = \sum_{i=1}^{n} \frac{1}{i}$. From here, we propose the two methods of arriving at a decision to accept or reject the hypothesis H_0 defined in (4.3).

`Begin Sub-Algorithm: ASCLT-test Method 1`

$$\bar{t}_{\alpha} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{\alpha}^{*p,(n)}}{\text{nper}} \quad \text{and} \quad \bar{t}_{1-\alpha} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{1-\alpha}^{*p,(n)}}{\text{nper}}$$

Now, we propose to reject H_0 if

$$0 \notin \left[\bar{t}_{\alpha},\ \bar{t}_{1-\alpha}\right]. \tag{4.10}$$

`End Sub-Algorithm: ASCLT-test Method 1`

`Begin Sub-Algorithm: ASCLT-test Method 2`

Subsequently, define *transformed* quantiles,

$$\widehat{t}_{\alpha,\text{trans.}}^{*p,(n)} = \frac{\left(\overline{SS}^{\,*p} - \widehat{t}_{1-\alpha}^{*p,(n)}\right)}{\sqrt{n}} \quad \text{and} \quad \widehat{t}_{1-\alpha,\text{trans.}}^{*p,(n)} = \frac{\left(\overline{SS}^{\,*p} - \widehat{t}_{\alpha}^{*p,(n)}\right)}{\sqrt{n}}.$$

Further define,

$$\bar{t}_{\alpha,\text{trans.}} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{\alpha,\text{trans.}}^{*p,(n)}}{\text{nper}} \quad \text{and} \quad \bar{t}_{1-\alpha,\text{trans.}} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{1-\alpha,\text{trans.}}^{*p,(n)}}{\text{nper}}$$

Finally, a transformation of the quantiles with the following expression is performed:

$$\bar{t}_{\alpha,\text{fin}} = \frac{\bar{t}_{\alpha,\text{trans.}} - 2\alpha \cdot \lambda_{n_1,n_2,2\alpha} \cdot \left(\bar{t}_{\alpha,\text{trans.}} + \bar{t}_{1-\alpha,\text{trans.}}\right)}{\kappa_{n_1,n_2,2\alpha}} \tag{4.11}$$

$$\bar{t}_{1-\alpha,\text{fin}} = \frac{\bar{t}_{1-\alpha,\text{trans.}} - 2\alpha \cdot \lambda_{n_1,n_2,2\alpha} \cdot \left(\bar{t}_{\alpha,\text{trans.}} + \bar{t}_{1-\alpha,\text{trans.}}\right)}{\kappa_{n_1,n_2,2\alpha}}, \tag{4.12}$$

where $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$ are constant coefficients for given values of n_1, n_2 and 2α, and they have to be numerically determined for each sample size n_1, $n_2 \in \mathbb{N}$ and significance level 2α.

Now, reject H_0 if

$$\overline{x}_{n_1} - \overline{x}_{n_2} \notin \left[\bar{t}_{\alpha,\text{fin}},\ \bar{t}_{1-\alpha,\text{fin}}\right],$$

where $\overline{x}_{n_1}$ and $\overline{x}_{n_2}$ denote the respective means of the original sample $\mathbf{x}_1$ and $\mathbf{x}_2$.

`End Sub-Algorithm: ASCLT-test Method 2`

`End Algorithm - ASCLT tests for parametric two-sample testing problem1`

The numerically found values of $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$, introduced in the above algorithm, for certain cases are enlisted in Table 4.3. These values for the coefficients were determined by more of an trial-and-error approach, using the required significance level to be achieved and also sometimes trying to match the results from SWS appraoximation (which we set as the gold standard test in the situation of the BFP). As can be observed from the Table 4.3 that, though the finding of these values were not done following a pure mathematical rule, the values do have some regular pattern (monotonously increasing or decreasing). This implies that they are not ad-hoc, random quantities, but they could very well have a mathematical structure underlying them.

Moreover, before proposing transformation $\bar{t}_{\alpha,\text{fin}}$ and $\bar{t}_{1-\alpha,\text{fin}}$ in equations (4.11) and (4.12), we did perform extensive computations to have simpler form of the transformation. For example, we tried methods without the additive term corresponding to $\lambda_{n_1,n_2,2\alpha}$, and having only the multiplicative coefficient similar to $\kappa_{n_1,n_2,2\alpha}$. In such a setup, for the normally distributed samples we were able to produce results similar to the one produced by using such complex transformation. Whereas, in the case of non-normal samples, the results were very close to that of SWS approximation which does not perform well in such situations (as will be observed from the next section on simulation results). With the complex transformation proposed above in (4.11) and (4.12), the results are better than the SWS approximation. Therefore, we propose and use these transformations.

The ASCLT procedure(s) described here forms the first step towards developing a general method for the situation two samples with unequal sample sizes. The ASCLT-test method 2 (on which we would mainly focus) presented above performs very good for

Table 4.3: Numerically determined coefficients $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$ for the ASCLT-test method 2, for different sample sizes and significance levels

	$2\alpha = 5\%$		$2\alpha = 10\%$	
$n_1 = n_2$	$\kappa_{n_1,n_2,2\alpha}$	$\lambda_{n_1,n_2,2\alpha}$	$\kappa_{n_1,n_2,2\alpha}$	$\lambda_{n_1,n_2,2\alpha}$
10	0.523	3.00	0.512	3.98
15	0.595	2.70	0.585	3.705
20	0.65	2.50	0.63	3.40
25	0.685	2.15	0.67	3.14
30	0.72	2.00	0.70	3.00

equal sample sizes $n_1 = n_2 = n$, as will be seen from the next subsection with simulation results. Though the procedure can be applied for situations with unequal sample sizes, our studies reveal that the method does not consistently maintain the pre-assigned significance level, 2α, for unequal sample sizes. That is, for any given combination of n_1 and n_2, ASLCT-test method 2 maintains the pre-assigned level for certain patterns of equal and unequal variances (increasing or decreasing) but breaks down for some others. But, the solution for equal sample sizes-case, presented here, would be the first step towards developing a general method for the situation two samples with unequal sample sizes. According to the current understanding and exploration, following are the very likely indications towards modifying the procedure to properly address the situation of unequal sample sizes:

- In the current approach, the minimum of n_1 and n_2 is taken and considered in the procedure, mainly through the equation (4.9). There could be a modification in this, by implementing the knowledge of unequal sample sizes in this part. This aspect may get clearer when the specific mathematical theory surrounding the two-sample ASCLT is explicitly derived.

- There are currently two additive components in the numerator on the RHS of equations (4.11) and (4.12). There could be a third (or more) component involved and/or even the current two components could be split to form more components. There could also be some multiplicative term or constant. All these have to be explored from either mathematical viewpoint or simulation-based studies, or both.

Parts of the proposals above are planned to be worked on, subsequent to the formal completion of this thesis.

Along with all of the explained methods above, the two-sample student's t-test was also implemented to study the performance of the test, mainly in comparison to the independent results from the SWS approximation. Moreover, we simulated situations of equal and unequal variances of the parent distribution. So, t-test would be the optimal test to use under homoscedastic samples.

4.3.4 Simulation Results and Discussion

The methods detailed in the Subsection 4.3.2 will be implemented along with the ASCLT-tests described in Subsection 4.3.3, via monte-carlo simulations in order to compare the results among these procedures and also observe the pattern, if any, of the individual methods. It has to be noted that the simulation results for each method is independent of the others and by 'comparison of results of these methods' we imply that the finally arrived-at results of these methods are reviewed independently and with reference to each other with respect to maintaining the pre-assigned level of significance and power properties. Numerical results will be presented via tables and figures, and also summarized through discussion in this section.

Note that not all methods described in the two Subsections 4.3.2 and 4.3.3 above, are, in principle, applicable to all situations which are simulated. So, in order to make comparison and interpretation easier we follow the scheme that the numerical results of methods which are validly applicable (i.e., simulation setting matching with those of the theoretical assumptions) are presented in straight, normal-face font. If a method is not applicable due to conflict between the simulation setting and the theoretical assumptions, then we present the results of those in italized text. Such italized text would then indicate that the concerned method should not have been used 'in principle'. In spite of this remark, it should also be realised that in real-life situations it is quite difficult to recognize such differences (w.r.t. the theoretical assumptions) with accuracy. It can be noted that Moser and Stevens (1992) and, later, Sprott and Farewell (1993) show results concerning pre-testing for the assumptions before moving on to perform the actual test of hypothesis of interest. They advice that this should not be performed and it could have serious consequences on the final results. Further, one of the intentions of presenting the results of even such "non-applicable" methods is to observe how 'good' or 'bad' the methods could perform when the underlying assumptions with respect to variances, are not fulfilled, without formally testing for such assumptions.

The following are the common setting of parameters for all simulations:

- Total number of simulation runs for each result presented, say N_{SIM}, is 10000.
- For Bootstrap test, $B = 2000$
- For ASCLT tests, $P = 2000$
- All tests were two-sided and performed at a significance level of $2\alpha = 5\%$ or 10%.

Simulations based on Normal Distribution

The idea here is to evaluate the performance of the two-sample tests when the samples come from normal distributions. Within this framework, we consider two scenarios - one, where the two samples originate from populations with equal variances, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, say, and the other where the variances are different in the parent populations, i.e., the situation of the BFP.

The results of the simulations corresponding to sample sizes $n_1 = n_2 = 10$, 15 and 30 for significance levels $2\alpha = 5\%$ and $2\alpha = 10\%$ are presented in the Tables 4.4 and 4.5, respectively, and the power curves for sample sizes $n_1 = n_2 = 10$ and 15, respectively, covering cases of unequal and equal variances are presented in Figure 4.2. The power curves are shown only for select methods since some of the curves are very close to each other and would be difficult to distinguish between themselves. The main findings from the results presented in the tables and figures can be summarized as follows.

- The SWS approximation is slightly conservative compared to the t-test in the case of $\sigma_1 = \sigma_2$. The performance of this approximation improves as sample sizes increases. Moreover, this pattern continues in the situation of varying variances, but here the t-test tends to be liberal, while the SWS keeps the level. So assuming normal distribution of the sample, SWS could very well be used, in general, instead of t-test, particularly for larger sample sizes without much cause of concern about the assumption on variances.

- The Cochran-Cox approximation is very conservative for the situation with $\sigma_1 = \sigma_2$. It tends to get slightly better with increasing difference between the variances - still being on the conservative side. This approximation is better avoided in practice.

- The bootstrap method tends to be very conservative too and displays similar results as the Cochran-Cox approximation.

- A surprising part is the results from Bayesian method at $2\alpha = 5\%$ level. The method seems to work 'better' at $2\alpha = 10\%$ compared to the performance at $2\alpha = 5\%$. It can also be noted that the pattern over equal and unequal variances situations do not vary too much. Moreover, from the two power curves in Figure 4.2, it can be seen that there is some severe problem with the shape of the curves

- the main difference between the two plots being the variation of mean of the first or the second sample fixing the other mean as constant. As mentioned earlier, the priors suggested by the authors were used here. Thus, issues with both the achieved significance level and power, indicate a strong signal for the need of Bayesian expert(s) to advice on appropriate priors for every situation independently and on a case-by-case basis.

- The ASCLT-test method 1 test results show a pattern of moving from being extremely liberal for small sample sizes to very conservative as the sample sizes increase. This numerical observation is in agreement with the Lemma 10. This is an important point to note, since this could be another further topic for detailed research - considering and constructing the theory of hypothesis-testing from an 'almost sure' point of view with debate over the conventional methods used in practice, at present. The reason why it can not discussed straight-away from the results presented here is that, the power of such a test gets to become worse. So, one of the main directions for further work would be to have a procedure which can lead to an almost sure decision. More of this will be discussed in the next Chapter.

- Finally, the performance of the ASCLT-test method 2 test is sometimes little liberal and sometimes little conservative. On an average, the method maintains the level and also has good power. Indeed the power of this method is approximately 1-2% better than that of SWS approximation, in spite of maintaining level on par with the SWS approximation. In summary, the results of this method are very much comparable and similar to the SWS approximation, which we set as the gold-standard for comparisons.

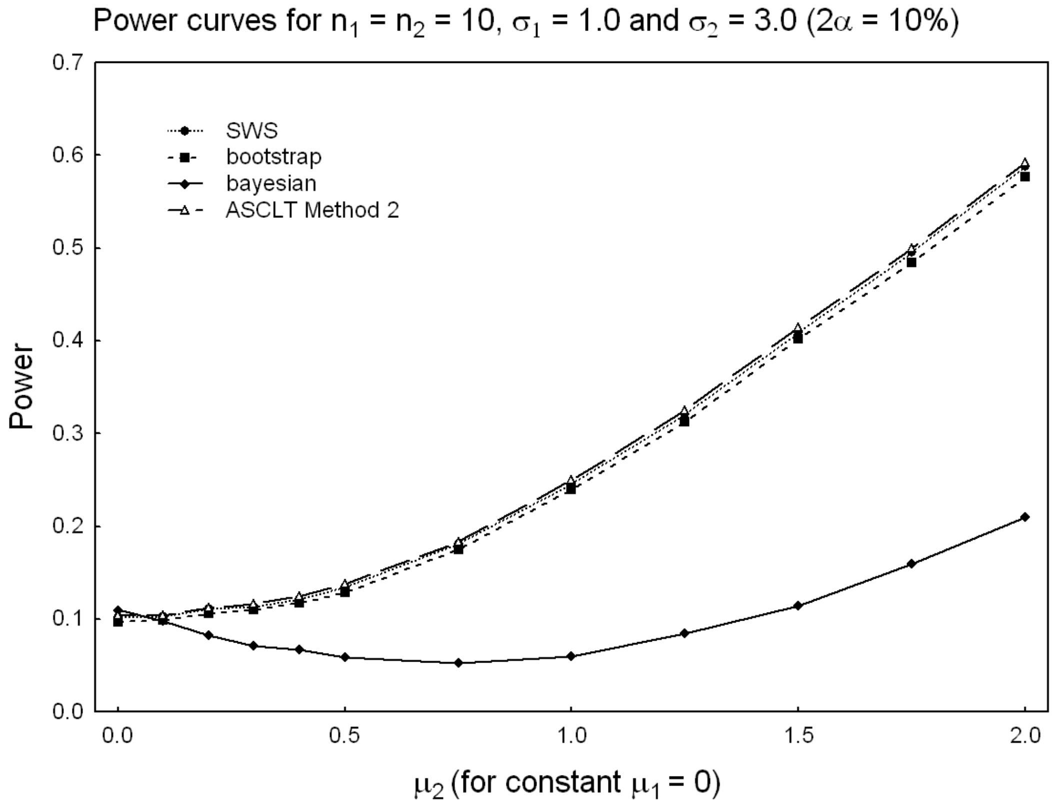
It has to be noted that these summaries are based on the equal sample sizes that we have considered. The pattern of these findings may change for unequal sample sizes.

Table 4.4: Results for simulated level at $2\alpha = \mathbf{5\%}$ under $H_0 : \mu_1 = \mu_2$ based on samples from $\mathcal{N}(\mu_i, \sigma_i)$. Column labelled 'Boot' gives the result for Bootstrap method, and 'A-M1' and 'A-M2' represent the results from ASCLT-test methods 1 and 2, respectively.

σ_1	σ_2	t-test	SWS	CocCox	Boot	Bayes	A-M1	A-M2
				$n_1 = n_2 = 10$				
1.0	1.0	5.22	*5.05*	*3.61*	3.39	8.80	*18.59*	4.85
2.0	2.0	5.23	*5.08*	*3.71*	3.31	8.95	*19.14*	4.92
3.0	3.0	5.32	*5.21*	*3.93*	3.53	8.85	*18.77*	5.08
4.0	4.0	4.99	*4.88*	*3.59*	3.22	8.53	*18.09*	4.79
2.0	1.0	*5.26*	4.75	3.97	3.23	*8.62*	*18.59*	4.99
3.0	1.0	*5.96*	5.14	4.60	3.83	*10.08*	*19.74*	5.58
4.0	1.0	*6.25*	5.33	4.90	3.94	*9.50*	*19.54*	5.92
1.0	2.0	*5.68*	5.21	4.31	4.22	*9.41*	*19.52*	5.34
1.0	3.0	*5.96*	5.28	4.69	4.71	*9.70*	*19.84*	5.79
1.0	4.0	*6.20*	5.22	4.74	5.04	*10.04*	*19.81*	5.86
				$n_1 = n_2 = 15$				
1.0	1.0	4.72	*4.65*	*3.86*	3.63	8.46	*9.01*	4.46
2.0	2.0	5.01	*4.94*	*4.00*	3.96	8.37	*9.54*	4.77
3.0	3.0	4.65	*4.61*	*3.83*	3.75	8.24	*9.62*	4.50
4.0	4.0	5.05	*4.97*	*4.14*	3.94	8.86	*9.88*	4.79
2.0	1.0	*5.06*	4.84	4.26	3.94	*8.92*	*9.70*	4.88
3.0	1.0	*5.21*	4.74	4.31	3.78	*8.55*	*9.73*	4.85
4.0	1.0	*5.96*	5.36	5.05	4.42	*9.43*	*10.73*	5.61
1.0	2.0	*5.43*	5.10	4.34	4.38	*8.76*	*10.07*	5.04
1.0	3.0	*5.82*	5.23	4.72	4.85	*9.11*	*10.40*	5.44
1.0	4.0	*5.20*	4.58	4.34	4.49	*9.07*	*9.94*	4.84
				$n_1 = n_2 = 30$				
σ_1	σ_2	t-test	SWS	CocCox	Boot	Bayes	A-M1	A-M2
1.0	1.0	4.85	*4.85*	*4.34*	4.22	7.93	*1.45*	4.61
2.0	2.0	5.08	*5.07*	*4.67*	4.62	8.24	*1.61*	4.90
3.0	3.0	4.95	*4.93*	*4.61*	4.57	8.16	*1.63*	4.82
4.0	4.0	5.00	*4.99*	*4.65*	4.56	8.72	*1.80*	4.86
5.0	5.0	4.85	*4.83*	*4.47*	4.38	7.80	*1.37*	4.73
2.0	1.0	*5.15*	4.96	4.60	4.43	*8.81*	*1.64*	4.89
3.0	1.0	*4.84*	4.42	4.29	4.10	*8.57*	*1.56*	4.59
4.0	1.0	*5.16*	4.75	4.65	4.40	*8.41*	*1.79*	4.87
5.0	1.0	*5.28*	4.91	4.83	4.62	*8.03*	*2.05*	5.05
1.0	2.0	*4.90*	4.69	4.41	4.45	*8.67*	*1.66*	4.66
1.0	3.0	*5.52*	5.17	4.99	5.04	*8.57*	*1.89*	5.31
1.0	4.0	*5.44*	4.89	4.78	4.93	*8.33*	*1.92*	5.09
1.0	5.0	*5.53*	5.11	5.06	5.00	*8.84*	*2.08*	5.30

Table 4.5: Results for simulated level at $2\alpha = \mathbf{10\%}$ under $H_0 : \mu_1 = \mu_2$ based on samples from $\mathcal{N}(\mu_i, \sigma_i)$. Column labelled 'Boot' gives the result for Bootstrap method, and 'A-M1' and 'A-M2' represent the results from ASCLT-test methods 1 and 2, respectively.

σ_1	σ_2	t-test	SWS	CocCox	Boot	Bayes	A-M1	A-M2
				$n_1 = n_2 = 10$				
1.0	1.0	9.84	*9.65*	*8.13*	8.02	10.33	*19.36*	9.08
2.0	2.0	10.05	*9.90*	*8.33*	8.21	9.85	*19.54*	9.30
3.0	3.0	9.90	*9.76*	*8.38*	8.22	9.62	*19.32*	9.26
4.0	4.0	10.44	*10.26*	*8.73*	8.55	10.37	*20.26*	9.73
5.0	5.0	9.85	*9.66*	*8.01*	7.76	10.20	*19.96*	9.13
2.0	1.0	*10.59*	9.96	8.81	8.25	*10.44*	*20.21*	9.75
3.0	1.0	*10.50*	9.66	9.01	8.10	*10.26*	*19.97*	9.75
4.0	1.0	*11.58*	10.38	9.77	8.90	*10.97*	*20.72*	10.69
5.0	1.0	*11.30*	9.96	9.68	8.64	*11.23*	*20.52*	10.31
1.0	2.0	*10.46*	9.91	8.70	8.79	*10.72*	*20.48*	9.60
1.0	3.0	*11.70*	10.70	9.90	10.29	*10.93*	*21.00*	10.86
1.0	4.0	*10.77*	9.73	9.21	9.26	*10.75*	*20.25*	10.00
1.0	5.0	*11.50*	10.10	9.73	10.09	*10.90*	*20.91*	10.59
				$n_1 = n_2 = 15$				
1.0	1.0	9.75	*9.63*	*8.53*	8.57	9.41	*9.91*	9.33
2.0	2.0	9.82	*9.79*	*8.87*	8.77	9.39	*10.35*	9.67
3.0	3.0	9.77	*9.73*	*8.59*	8.60	9.38	*10.21*	9.59
4.0	4.0	10.12	*10.03*	*8.88*	8.93	9.96	*10.58*	9.93
5.0	5.0	10.15	*10.09*	*8.89*	8.88	10.24	*10.64*	9.96
2.0	1.0	*10.08*	9.70	8.94	8.70	*9.39*	*10.63*	9.87
3.0	1.0	*11.11*	10.44	10.07	9.76	*10.65*	*11.43*	10.80
4.0	1.0	*10.28*	9.54	9.33	8.92	*9.66*	*10.59*	9.98
5.0	1.0	*10.61*	9.76	9.66	9.25	*10.36*	*10.96*	10.33
1.0	2.0	*10.41*	10.19	9.42	9.49	*9.77*	*10.79*	10.15
1.0	3.0	*10.27*	9.68	9.26	9.62	*9.63*	*10.59*	10.00
1.0	4.0	*10.55*	9.77	9.48	9.69	*10.76*	*11.07*	10.28
1.0	5.0	*10.34*	9.46	9.22	9.46	*9.89*	*10.73*	10.01
				$n_1 = n_2 = 30$				
1.0	1.0	9.74	*9.70*	*9.11*	9.24	9.60	*2.18*	9.83
2.0	2.0	10.02	*10.01*	*9.49*	9.59	9.40	*1.77*	10.17
3.0	3.0	10.19	*10.18*	*9.67*	9.60	9.90	*1.92*	10.32
4.0	4.0	10.40	*10.36*	*9.90*	9.94	9.86	*2.16*	10.58
2.0	1.0	*9.96*	9.70	9.40	9.44	*9.28*	*2.00*	10.03
3.0	1.0	*9.94*	9.63	9.44	9.27	*9.35*	*1.91*	10.07
4.0	1.0	*10.53*	10.28	10.19	9.97	*10.29*	*2.28*	10.65
5.0	1.0	*9.86*	9.42	9.34	9.21	*9.64*	*2.31*	9.95
1.0	2.0	*10.18*	10.10	9.76	9.81	*9.51*	*2.23*	10.43
1.0	3.0	*10.65*	10.32	10.06	10.24	*9.54*	*2.16*	10.76
1.0	4.0	*10.24*	9.80	9.71	9.84	*9.41*	*2.25*	10.35
1.0	5.0	*10.70*	10.18	10.09	10.28	*10.47*	*2.59*	10.81



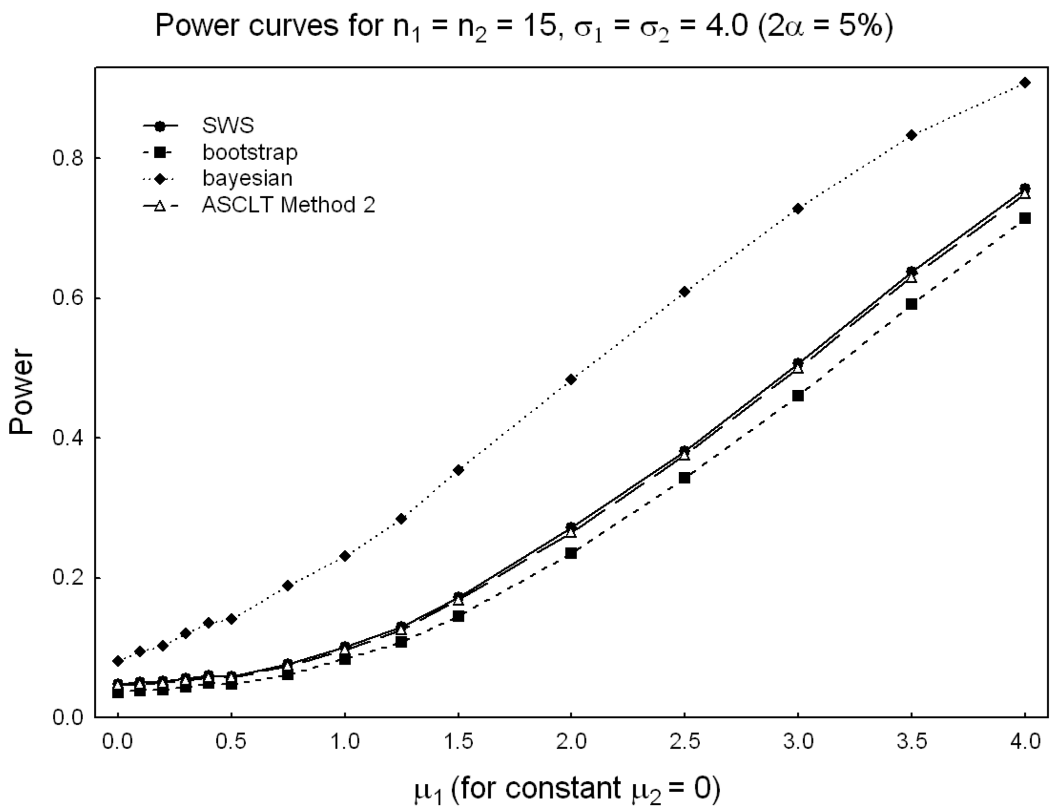


Figure 4.2: Power curves of different methods for cases with $n_1 = n_2 = 10$ and $n_1 = n_2 = 15$ based on samples from $\mathcal{N}(\mu_i, \sigma_i)$.

Simulations based on lognormal and exponential Distributions

In reality, it is usually not possible to determine the form of the parent distribution from which the data comes. In medical applications there are situations where the source of the data could have been lognormal or exponential distribution. So, we wanted to see how the results for the methods presented above, change if the assumption of normality is not fulfilled. Tables 4.6 presents the results concerning the simulated significance level of the methods in situations where the two samples come from a lognormal distributions having equal means, with equal and unequal variances. Similarly, Table 4.7 presents results concerning samples from exponential distributions with equal means. Note that for two exponential distributions with same parameter, say, λ, then mean $= \lambda$ and variance $= \lambda^2$. Thus, having equal means for two exponential distributions, the situation of BFP can not be generated. So here, we are interested in studying only the performance of the different methods under the influence of exponential distribution. As usual, wherever certain method is not applicable, the results are presented in italics.

From the results presented in Tables 4.6 and 4.7, the following observations can be made:

- The three tests, namely, t-test, SWS and Cochran and Cox approximations tend to be quite conservative in exponential distribution situation and both conservative and liberal under lognormally distributed samples. Surprisingly, in most situations t-test seems to perform better than the SWS approximation and the worst performance (among these three tests) is by Cochran and Cox approximation.
- Bootstrap method tends to be very conservative.
- With exponentially distributed samples, bayesian test exhibits similar result as it did in the case of normally distributed samples and it shows conservativeness under lognormally distributed samples for $2\alpha = 10\%$. The $2\alpha = 5\%$ case seems to have similar problems as observed in earlier the simulations with normally distributed samples.
- The ASCLT-test method 2 exhibits both conservativeness and liberality under lognormal distribution case. But still it shows results closer to the pre-designated level in majority of the cases, compared with the performance of other methods. And in the exponentially distributed samples, in spite of being slightly conservative, it performs the best.

Table 4.6: Results for simulated level under $H_0 : \mu_1 = \mu_2$, for samples from lognormal distributions* with parameters λ_{i1} and λ_{i2}, $i = 1, 2$. Columns 'Boot', 'A-M1' and 'A-M2' give the results from Bootstrap and ASCLT-test methods 1 and 2, respectively

$\mu_1 = \mu_2$	σ_1^2	σ_2^2	λ_{11}	λ_{21}	λ_{12}	λ_{22}	t-test	SWS	CocCox	Boot	Bayes	A-M1	A-M2
$n_1 = n_2 = 10$ and $2\alpha = 5\%$													
3.5	42.47	4.26	0.50	1.10	1.225	0.548	*13.74*	13.45	11.95	8.07	*8.02*	*31.13*	14.03
3.5	42.47	12.37	0.50	0.90	1.225	0.837	*6.20*	5.61	4.58	3.18	*4.98*	*27.77*	7.24
3.5	42.47	24.44	0.50	0.70	1.225	1.049	*3.66*	2.87	2.47	1.56	*5.42*	*26.20*	4.83
3.5	42.47	42.47	0.50	0.50	1.225	1.225	3.10	*2.29*	*1.96*	1.27	7.74	*26.95*	4.22
3.5	4.26	12.37	1.10	0.90	0.548	0.837	*6.05*	5.59	4.65	3.26	*13.90*	*23.10*	6.37
2.0	4.95	3.34	0.30	0.40	0.894	0.775	*4.43*	3.76	3.15	2.44	*7.01*	*23.62*	5.05
2.5	23.92	23.92	0.10	0.10	1.265	1.265	2.87	*2.07*	*1.62*	1.05	7.70	*28.05*	4.13
2.5	23.92	14.00	0.10	0.30	1.265	1.095	*3.40*	2.60	2.18	1.62	*5.79*	*27.08*	4.74
2.5	23.92	7.39	0.10	0.50	1.265	0.894	*5.99*	5.20	4.44	2.98	*4.86*	*28.49*	7.18
$n_1 = n_2 = 10$ and $2\alpha = 10\%$													
2.01	6.96	3.34	0.20	0.40	1.000	0.775	*10.20*	9.52	8.67	7.11	*7.39*	*25.41*	10.29
2.01	9.38	3.34	0.10	0.40	1.095	0.775	*11.54*	10.68	9.77	7.93	*6.12*	*26.44*	11.32
2.5	18.46	10.39	0.20	0.40	1.183	1.000	*9.52*	8.48	7.61	6.07	*7.10*	*28.31*	9.51
2.71	37.39	29.22	0.10	0.20	1.342	1.265	*7.90*	6.85	6.31	4.35	*8.73*	*29.78*	8.48
2.71	22.54	29.22	0.30	0.20	1.183	1.265	*8.02*	6.96	6.29	4.37	*10.97*	*29.32*	8.50
2.71	22.54	12.69	0.30	0.50	1.183	1.000	*8.56*	7.61	6.88	5.38	*7.16*	*27.17*	8.79
3.0	35.69	20.89	0.30	0.50	1.265	1.095	*8.48*	7.55	6.88	5.32	*6.91*	*27.86*	8.70
3.0	11.03	20.89	0.70	0.50	0.894	1.095	*9.78*	8.81	7.79	5.90	*14.51*	*26.70*	9.87
3.5	24.44	42.47	0.70	0.50	1.049	1.225	*9.12*	7.91	7.23	4.91	*13.64*	*28.34*	9.52
3.5	24.44	12.37	0.70	0.90	1.049	0.837	*9.79*	9.07	7.93	6.56	*6.72*	*26.60*	9.87

* The density of lognormal distribution that we consider is of the following form:

$$f(x) \;=\; \frac{1}{x\lambda_{i2}\sqrt{2\pi}}\, e^{-\frac{(\ln(x)-\lambda_{i1})^2}{2\lambda_{i2}^2}},$$

for $x \geq 0$, $\lambda_{i2} > 0$ and $i = 1, 2$.

Table 4.7: Results for simulated level $H_0 : \mu_1 = \mu_2$, for samples generated from exponential distributions* with parameters λ_i, $i = 1, 2$. Columns 'Boot', 'A-M1' and 'A-M2' give the results from Bootstrap and ASCLT-test methods 1 and 2, respectively

$\lambda_1 = \lambda_2$	t-test	SWS	CocCox	Boot	Bayes	A-M1	A-M2
			$n_1 = n_2 = 10$ and $2\alpha = 5\%$				
1	3.83	*3.45*	*2.79*	2.07	8.57	*20.55*	4.13
2	4.57	*3.87*	*2.96*	2.25	9.01	*21.20*	4.87
3	4.20	*3.71*	*2.90*	2.14	9.13	*21.77*	4.57
4	4.37	*3.85*	*3.06*	2.25	8.63	*20.75*	4.78
5	4.17	*3.54*	*2.93*	2.18	8.63	*20.98*	4.51
			$n_1 = n_2 = 10$ and $2\alpha = 10\%$				
1	9.30	*8.73*	*7.68*	6.62	10.22	*22.42*	9.31
2	9.11	*8.49*	*7.45*	6.32	9.44	*21.91*	9.15
3	10.15	*9.37*	*8.07*	6.86	10.43	*22.71*	10.20
4	9.68	*9.09*	*7.89*	6.81	10.23	*22.44*	9.70
5	9.63	*9.05*	*7.90*	6.63	10.44	*22.40*	9.91

* The density of exponential distribution that we consider is of the form $f(x) = \frac{1}{\lambda_i}\, e^{-\frac{x}{\lambda_i}}$, for $x \geq 0$, $\lambda_i > 0$ and $i = 1, 2$.

4.4 Nonparametric Behrens-Fisher Problem

The nonparametric BFP is also referred by some authors as the generalized BFP. For sake of simplicity, we abbreviate 'Nonparametric Behrens-Fisher Problem' by 'NP-BFP'.

Several authors propose their own way of interpreting and defining their version of the hypothesis for NP-BFP. We will look at the individual definitions of the problem when presenting the authors' solution corresponding to their definition of the NP-BFP. The main reason for such differences is that prior to the past decade, the so-called 'nonparametric tests' were defined via hypothesis in a semi-parametric setup (cf. Lehmann, 1986, pp. 323). Only recently the fully nonparametric approach has been advocated by experts like Akritas, Brunner, Thompson, etc,. So, for our consideration, we propose the following definition from literature which is based on the fully nonparametric approach from an applied perspective.

Definition 18 (Nonparametric Behrens-Fisher Problem). *Let* $X_{i1}, \ldots, X_{in_i}$, $i = 1, 2$, *be i.i.d random variables such that* $X_{i1} \sim F_i$, $i = 1, 2$. *Then the Nonparametric*

Behrens-Fisher Problem is that of testing

$$H_0^p : p = \frac{1}{2} \quad Vs. \quad H_1 : p \neq \frac{1}{2}, \tag{4.13}$$

where $p = P\{X_{11} < X_{21}\} + \frac{1}{2}\{X_{11} = X_{21}\}$*, the so-called Relative Treatment Effect (RTE).*

For the standard nonparametric two-sample location problem, when $\sigma_1 = \sigma_2$, the Wilcoxon-Mann-Whitney(WMW) test is robust in maintaining level and has good power (Lehmann, 1975, pages 76-81). In this situation, it is also distribution-free and its critical values have been tabulated extensively. However, if the scales are unequal, then the WMW statistic is not even asymptotically distribution-free, since its asymptotic variance would depend on the unknown distribution. Consequently, using the critical values for equal scales in this case, leads to grossly inflated levels (c.f. Fligner and Policello, 1981, Table 2). Thus the WMW test can not be considered as a solution to the NP-BFP. So, below we present and discuss some solutions proposed in the literature for the NP-BFP. We then present the solution via the ASCLT-test methods based on rank statistics. This will be followed by the presentation of simulations results comparing the independent results of the performance of the methods.

We first fix a common set of notation which we will use throughout this section in the description of the methods which follow. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})$ be the vectors of i.i.d random variables for the i$^{\text{th}}$ sample for $i = 1, 2$. Further, let the overall mid-rank of X_{ij} be denoted and defined by $R_{ij} = \frac{1}{2} + \sum_{k=1}^{2} \sum_{l=1}^{n_k} c(X_{ij} - X_{kl})$, where $c(u) = 0, \frac{1}{2}$ or 1 according as $u <, =$ or > 0 is called the *normalized* version of the *count* function $c(\cdot)$. Let also the corresponding observed sample be denoted by $\mathbf{x}_i = (x_{i1}, \ldots, x_{in_i})$, $i = 1, 2$.

4.4.1 Babu and Padmanabhan (2002) Resampling Method

Babu and Padmanabhan (2002) propose two resampling-based procedures for the NP-BFP of testing for the equality of the medians of two continuous distributions having the same shape, but possibly unequal variances. Thus the hypothesis they are interested in testing is

$$H_0^{BP} : \widetilde{\mu}_1 = \widetilde{\mu}_2 \quad \text{Vs.} \quad H_1^{BP} : \widetilde{\mu}_1 \neq \widetilde{\mu}_2, \tag{4.14}$$

where $\widetilde{\mu}_i$, $i = 1, 2$, denotes the median of the respective parent distribution of first and second sample.

For our application here, of the two procedures suggested by them, we will consider only the one that they report to perform better for small sample sizes.

We follow the same notation of variables introduced in the beginning of this section. Additionally, let $\widetilde{X}_1$ and $\widetilde{X}_2$ denote the sample medians of first and second samples. Let also s_1 and s_2 denote the respective sample standard deviations.

`Begin Algorithm - Babu and Padmanabhan Method`

Firs, define transformed observations z_i as

$$z_i = \begin{cases} \left(X_{1i} - \widetilde{X}_1\right)/s_1 & \text{if } 1 \leq i \leq n_1 \\ \left(X_{2(i-n_1)} - \widetilde{X}_2\right)/s_2 & \text{if } n_1 < i \leq n_1 + n_2 \end{cases}$$

Denoting $Q = n_1 + n_2$, the following terms are computed,

$$\begin{aligned} U &= \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} 1_{\{X_{1i} \leq X_{2j}\}}, \\ \tilde{p} &= \frac{1}{Q^2} \sum_{i=1}^{Q} \sum_{j=1}^{Q} 1_{\{z_i s_1 \leq z_j s_2\}} \text{ and} \\ T &= \sqrt{n}\left(U - \tilde{p}\right). \end{aligned}$$

Now, let $(z_1^{*,b}, \ldots, z_Q^{*,b})$ be sample values of bootstrap sample b, $b = 1, \ldots, B$, from the original sample of $(z_1, \ldots, z_Q)$. Also let $y_{1i}^b = z_i^{*,b} s_1,\ i = 1, \ldots, n_1$ and $y_{2i}^b = z_{i+n_1}^{*,b} s_2,\ i = 1, \ldots, n_2$. Based on these values, the following quantities are computed:

$$\begin{aligned} U^{*,b} &= \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} 1_{\{y_{1i}^b \leq y_{2j}^b\}}, \\ p^{*,b} &= \frac{1}{Q^2} \sum_{i=1}^{Q} \sum_{j=1}^{Q} 1_{\left\{z_i^{*,b} s_1^{*,b} \leq z_j^{*,b} s_2^{*,b}\right\}} \text{ and} \\ T^{*,b} &= \sqrt{n}\left(U^{*,b} - p^{*,b}\right), \end{aligned}$$

where $s_1^{*,b}$ and $s_1^{*,b}$ are the respective standard deviations of samples $\left(y_{11}^b,\ldots,y_{1n_1}^b\right)$ and $\left(y_{21}^b,\ldots,y_{2n_2}^b\right)$. The computed $\left(T^{*,1},\ldots,T^{*,B}\right)$ are arranged in ascending order to form vector $\left(T^{*,(1)},\ldots,T^{*,(B)}\right)$. Finally the $\alpha\cdot B^{\text{th}}$ and $(1-\alpha)\cdot B^{\text{th}}$ percentile are determined as $T^{*,(\alpha\cdot B)}$ and $T^{*,((1-\alpha)\cdot B)}$ (if $\alpha\cdot B$ and $(1-\alpha)\cdot B$ are not integers, then they are rounded to the closest integer.) The decision for evidence to accept the hypothesis H_0^{BP} defined in (4.14), is reached if $T\in\left(T^{*,(\alpha\cdot B)},T^{*,((1-\alpha)\cdot B)}\right)$.

`End Algorithm - Babu and Padmanabhan Method`

The authors present several simulation results regarding the achieved significance level using their method. They show that their method exhibits a mixed result - being conservative or liberal sometimes, and some others when the test maintains the pre-assigned level.

4.4.2 Reiczigel el al. (2005) Bootstrap Method

Unlike complicated-looking the Babu and Padmanabhan (2002) resampling method, a more straight-forward bootstrap method is proposed by Reiczigel et al. (2005). The technical difference between the two methods is that, Babu and Padmanabhan (2002) propose a overall resampling from the transformed, pooled samples; whereas, Reiczigel et al. (2005) suggest bootstrapping within each of the respective samples. The authors also discuss extensively the necessity and importance of the two sample problem in biometry, in particular they discuss problems in parasitology and psychology, where non-parametric methods are recommended due to the nature of the data arising out of such fields.

In their paper, the authors suggest four different transformations of the original sample before proceeding with the bootstrapping and further computations based on it. They also recommend a particular transformation to be the "most appropriate". Thus, we will use this transformation in the following algorithm of their method.

`Begin Algorithm - A Bootstrap solution for NP-BFP`

- Transform the observations of the second data set $\mathbf{x}_2$ to, say, $\mathbf{x}_3$, where each element of is $\mathbf{x}_3$ given by $x_{3i}=x_{2i}-c_1$, where the constant c_1 is the median of the values $x_{1i} - x_{2j}$ over all $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$. Note that $\mathbf{x}_3$ is just the notation for transformed $\mathbf{x}_2$. Thus, $n_2 = n_3$, the size of vector $\mathbf{x}_3$.

- Form B bootstrap data sets $\left(\mathbf{x}_1^{(*b)}, \mathbf{x}_3^{(*b)}\right)$, $b = 1, \ldots, B$, where $\mathbf{x}_i^{(*b)}$, $i = 1, 3$, are sampled with replacement from $(x_{i1}, \ldots, x_{in_i})$. Let ranks of corresponding observations the combined bootstrap sample $\left(\mathbf{x}_1^{(*b)}, \mathbf{x}_3^{(*b)}\right)$ be denoted by $R_{i,j}^{(*b)}$ for $i = 1, 3$ and $b = 1, \ldots, B$.

- Evaluate for each bootstrap data set,

$$t_{\mathrm{RW}}^{(*b)} = \frac{\overline{R}_{1\cdot}^{(*b)} - \overline{R}_{3\cdot}^{(*b)}}{\sqrt{\frac{s_1^{2,(*b)}}{n_1} + \frac{s_3^{2,(*b)}}{n_3}}}, \; b = 1, \ldots, B, \tag{4.15}$$

where $\overline{R}_{1\cdot}^{(*b)}$ and $\overline{R}_{3\cdot}^{(*b)}$ are respective means of the ranks corresponding to each sample and $s_1^{2,(*b)}$ and $s_2^{2,(*b)}$ are respective variances of ranks for each of the bootstrap data sets $\mathbf{x}_1^{(*b)}$ and $\mathbf{x}_3^{(*b)}$, $b = 1, \ldots, B$. Let also the $t_{\mathrm{RW,\,obs}}$ be the t-value computed by using the ranks of the original samples in equation (4.15) above.

- The bootstrap p value (or what is commonly known as "Achieved Significance level") is then calculated as $\mathrm{boot}_p = 2\min(p_1, p_2)$, where,

$$p_1 = \frac{1}{B}\sum_{b=1}^{B} 1_{\left\{t_{\mathrm{RW}}^{(*b)} \geq t_{\mathrm{RW,\,obs}}\right\}} \quad \text{and} \quad p_2 = \frac{1}{B}\sum_{b=1}^{B} 1_{\left\{t_{\mathrm{RW}}^{(*b)} \leq t_{\mathrm{RW,\,obs}}\right\}}.$$

- Reject H_0^p if $\mathrm{boot}_p \leq 2\alpha$.

`End Algorithm - A Bootstrap solution for NP-BFP`

In simple terminology, the authors suggest a bootstrapped, rank-version of the SWS approximation for t-test. The method is not directly implemented in and statistical packages known. But the authors have a SPlus/R function and also an Microsoft windows-based executable file to perform the computations for given samples. This can be downloaded from `http://www.univet.hu/users/jreiczig/brw/`

4.4.3 Brunner and Munzel (2000) Method

Along with the notation set in the beginning of this section, let $R_{ij}^{(i)} = \frac{1}{2} + \sum_{l=1}^{n_i} c(X_{ij} - X_{il})$, the internal ranks of within sample i, where $c(\cdot)$ is the normalized count function.

Brunner and Munzel (2000) consider a test of hypothesis of the form defined in Definition 18. An unbiased and consistent estimator of the RTE, defined in the (4.13), is given by,

$$\widehat{p} \;=\; \frac{1}{n_1}\left(\overline{R}_{2\cdot} - \frac{n_2+1}{2}\right)$$

Based on the above estimator of the RTE, following statistic is proposed,

$$W_N^{\mathrm{BF}} = \frac{1}{\sqrt{N}} \frac{\overline{R}_{2\cdot} - \overline{R}_{1\cdot}}{\widehat{\sigma}_N},$$

which has, asymptotically, a standard normal distribution under the hypothesis H_0^p : $p = \frac{1}{2}$, where

$$\widehat{\sigma}_N^2 \;=\; N \cdot \left[\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}\right],$$

where in turn,

$$\widehat{\sigma}_i^2 \;=\; \frac{1}{(n_i-1)\,(N-n_i)^2} \sum_{k=1}^{n_i} \left(R_{ik} - R_{ik}^{(i)} - \overline{R}_{i\cdot} + \frac{n_i+1}{2}\right)^2.$$

4.4.4 ASCLT Methods for NP-BFP

Based on the Theorem on ASCLT for Rank Statistics (Chapter 3, page 22, Theorem 16), here we propose a small-sample approximation for the two-sample NP-BFP. The only difference here would be that, we will not be using the standard normal distribution as proposed in the Theorem, but allowing the estimation of the quantiles of general distribution, say G, like in the parametric situation (Section 4.3.3). For our consideration, we use the hypothesis defined in the (4.13).

Thus, the small-sample approximation is similar to the one proposed for the parametric BFP, though here overall mid-ranks of the observations are appropriately used. Also the coefficients used in the transformations of the quantiles would be slightly different in terms of their numerical values. Along with the standard notation set previously for NP-BFP, let $n = \min(n_1, n_2)$.

`Begin Algorithm - ASCLT Methods for NP-BFP`

The two samples are independently permuted nper times (nper is chosen by user - as a general rule, the larger the value of nper, the better is the approximation), with p^{th} permuted sample vector be denoted and given by

$$\mathbf{x}_i^{*p} \;=\; \text{permute}(\mathbf{x}_i), \qquad i = 1, 2;\; p = 1, \ldots, \text{nper}\,.$$

Now for each of the permuted sample, we compute quantities:

$$SS_k^{*p} \;=\; k \cdot \frac{\overline{R}_{1\cdot,k}^{*p} - 0.5}{\sqrt{k}} \;=\; \sqrt{k} \cdot \left(\overline{R}_{1\cdot,k}^{*p} - 0.5\right) \qquad k = 1, \ldots, n, \tag{4.16}$$

where $\overline{R}_{1\cdot,k}^{*p} = \frac{\sum_{j=1}^{k} R_{1j,k}^{*p}}{k}$, where $R_{ij,k}^{*p}$, $i = 1, 2$; $j = 1, \ldots, k$ denote the overall mid-rank of observations of the combined, partial samples $(X_{11}^{*p}, \ldots, X_{1k}^{*p}, X_{21}^{*p}, \ldots, X_{2k}^{*p})$ for each $k = 1, \ldots, n$. That is,

$$R_{ij,k}^{*p} \;=\; \frac{1}{2} + \sum_{l=1}^{2} \sum_{m=1}^{k} c(X_{ij}^{*p} - X_{lm}^{*p}), \quad k = 1, \ldots, n.$$

Now, let $\overline{SS}^{\,*p} = \frac{\sum_{k=1}^{n} SS_k^{*p}}{n}$ for each permutation $p = 1, \ldots, \text{nper}$.

The quantiles $\widehat{t}_{\alpha}^{*p,(n)}$ and $\widehat{t}_{1-\alpha}^{*p,(n)}$ are now estimated by

$$\begin{aligned}
\widehat{t}_{\alpha}^{*p,(n)} &= \max\left\{ t \;\middle|\; C_n^{-1} \sum_{k=1}^{n} \frac{1}{k} 1_{\{SS_k^{*p} < t\}} \le \alpha \right\} \\
\widehat{t}_{1-\alpha}^{*p,(n)} &= \max\left\{ t \;\middle|\; C_n^{-1} \sum_{k=1}^{n} \frac{1}{k} 1_{\{SS_k^{*p} < t\}} \le 1 - \alpha \right\}
\end{aligned}$$

where $C_n = \sum_{i=1}^{n} \frac{1}{i}$. From here, we propose the two *methods* of arriving at a decision to accept or reject the hypothesis H_0 defined in (4.13).

`Begin Sub-Algorithm: ASCLT-test Method 1`

$$\bar{t}_\alpha = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_\alpha^{*p,(n)}}{\text{nper}} \quad \text{and} \quad \bar{t}_{1-\alpha} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{1-\alpha}^{*p,(n)}}{\text{nper}}$$

Now, we propose to reject H_0 if

$$0 \notin \left[\bar{t}_\alpha,\ \bar{t}_{1-\alpha}\right]. \tag{4.17}$$

`End Sub-Algorithm: ASCLT-test Method 1`

`Begin Sub-Algorithm: ASCLT-test Method 2`

Subsequently, define *transformed* quantiles,

$$\begin{aligned}
\widehat{t}_{\alpha,\text{trans.}}^{*p,(n)} &= \frac{\sqrt{n}\left(\overline{SS}^{\,*p} - \widehat{t}_{1-\alpha}^{*p,(n)}\right)}{n} = \frac{\left(\overline{SS}^{\,*p} - \widehat{t}_{1-\alpha}^{*p,(n)}\right)}{\sqrt{n}} \\
\widehat{t}_{1-\alpha,\text{trans.}}^{*p,(n)} &= \frac{\sqrt{n}\left(\overline{SS}^{\,*p} - \widehat{t}_{\alpha}^{*p,(n)}\right)}{n} = \frac{\left(\overline{SS}^{\,*p} - \widehat{t}_{\alpha}^{*p,(n)}\right)}{\sqrt{n}}.
\end{aligned}$$

Further define,

$$\bar{t}_{\alpha,\text{trans.}} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{\alpha,\text{trans.}}^{*p,(n)}}{\text{nper}} \quad \text{and} \quad \bar{t}_{1-\alpha,\text{trans.}} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{1-\alpha,\text{trans.}}^{*p,(n)}}{\text{nper}}$$

Finally, a transformation of the quantiles with the following expression is performed:

$$\bar{t}_{\alpha,\text{fin}} = \frac{\bar{t}_{\alpha,\text{trans.}} - 2\alpha \cdot \lambda_{n_1,n_2,2\alpha} \cdot \left(\bar{t}_{\alpha,\text{trans.}} + \bar{t}_{1-\alpha,\text{trans.}}\right)}{\kappa_{n_1,n_2,2\alpha}} \tag{4.18}$$

$$\bar{t}_{1-\alpha,\text{fin}} = \frac{\bar{t}_{1-\alpha,\text{trans.}} - 2\alpha \cdot \lambda_{n_1,n_2,2\alpha} \cdot \left(\bar{t}_{\alpha,\text{trans.}} + \bar{t}_{1-\alpha,\text{trans.}}\right)}{\kappa_{n_1,n_2,2\alpha}}, \tag{4.19}$$

where $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$ are constant coefficients for a given n_1, n_2 and 2α, and they have to be numerically determined for each sample size n_1, $n_2 \in \mathbb{N}$ and significance level 2α.

Now, reject H_0 defined in (4.13) if

$$\overline{R}_{1\cdot} - 0.5 \notin \left[\bar{t}_{\alpha,\text{fin}},\ \bar{t}_{1-\alpha,\text{fin}}\right],$$

where $\overline{R}_{1\cdot} = \frac{\sum_{j=1}^{n_1} R_{1j}}{n_1}$, where the R_{ij} denotes the overall mid-rank of observation X_{ij}.

```
End Sub-Algorithm: ASCLT-test Method 2
End Algorithm - ASCLT Methods for NP-BFP
```

As mentioned earlier, the numerical values of the coefficients $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$ differ from those estimated for the parametric BFP. But the concept and underlying idea of transforming the quantiles remain the same. The numerically found values of these coefficients for some combinations of sample sizes $n_1 = n_2 = n$ and $2\alpha = 5\%$ or 10% are presented in Table 4.8. The same discussion regarding such transformations of quantiles in page 52, holds here.

4.4.5 Simulation Results

Similar to the simulations for the parametric BFP, monte carlo simulations were also performed to evaluate the performance of the above solutions for the NP-BFP. The same common settings as for the parametric BFP (set on page 54) were also used for the simulations here. As usual, we present only a small portion of the output and the remaining will be made available online since the pattern of the results were similar.

The results for two simulation setting are presented in Tables 4.9 and 4.10. The first table corresponds to the results of normally distributed samples , while the second one presents the results for samples generated from lognormal distribution. The parameters used for the distribution under each simulations are also tabulated in the tables. Moreover, we

Table 4.8: Numerically determined coefficients $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$, for different sample sizes and significance levels for the ASCLT-test method 2 for NP-BFP

	$2\alpha = 5\%$		$2\alpha = 10\%$	
$n_1 = n_2$	$\kappa_{n_1,n_2,2\alpha}$	$\lambda_{n_1,n_2,2\alpha}$	$\kappa_{n_1,n_2,2\alpha}$	$\lambda_{n_1,n_2,2\alpha}$
10	0.6	3.00	0.58	3.95
15	0.65	2.60	0.63	3.7
20	0.696	2.45	0.67	3.35
25	0.727	2.10	0.705	3.05
30	0.76	1.90	0.73	2.90

Table 4.9: Results for simulated level based on samples from $\mathcal{N}(\mu = 0, \sigma)$ of 2α, under H_0 defined in (4.13). Columns 'A-M1' and 'A-M2' represent the ASCLT-test method 1 and 2, respectively.

σ_1	σ_2	**Reiczigel**	**Bab-Pad**	**Br-Mun**	**A-M1**	**A-M2**
		$n_1 = n_2 = 15$ and $2\alpha = 5\%$				
1.0	1.0	4.15	4.41	5.14	*4.15*	3.83
2.0	1.0	4.90	5.24	5.44	*5.02*	4.72
3.0	1.0	4.94	6.06	5.19	*5.33*	4.80
4.0	1.0	5.29	6.44	5.26	*5.78*	5.23
5.0	1.0	5.61	7.33	5.56	*6.50*	5.77
		$n_1 = n_2 = 15$ and $2\alpha = 10\%$				
1.0	1.0	9.05	9.10	9.94	*4.75*	8.02
1.0	2.0	9.34	10.33	9.79	*5.26*	8.50
1.0	3.0	9.98	11.28	9.79	*5.63*	9.22
1.0	4.0	10.65	12.92	10.43	*6.43*	10.33
1.0	5.0	10.16	12.74	9.74	*6.31*	9.92

have followed the same convention as in the previous sections, by presenting results of procedure not applicable 'in principle' in italized text.

Based on the results seen in Tables 4.9 and 4.10, we summarize the findings here:

- It can be noticed that none of the methods is consistently conservative or liberal in maintaining the pre-assigned level 2α (of course, except for the ASCLT-test method 1, which we do not discuss).

- In the normal sample situation, Reiczigel et al. (2005) method seems to be a bit conservative, but as the difference in the variance between the two groups increases, this method seems to get liberal. The same pattern also seems to hold for Babu and Padmanabhan (2002) method. On the other hand, Brunner and Munzel (2000) method seems to be consistently bit liberal at $2\alpha = 5\%$ and bit conservative at $2\alpha = 10\%$. Finally, the newly proposed ASCLT-test method 2, at $2\alpha = 5\%$ seems to move from being conservative to liberal as the difference in the variance increases and is consistently conservative for the case when $2\alpha = 10\%$.

- Under the lognormal distribution of the samples, there is mixed performance among all methods, though the ASCLT-test method 2 seems to be performing the 'best' among the lot, exhibiting results closest to the pre-assigned level.

Table 4.10: Results for simulated level under the H_0 defined in (4.13), based on samples from lognormal distributions* with parameters λ_{i1} and λ_{i2}, $i = 1, 2$. Columns 'A-M1' and 'A-M2' represent the results from ASCLT-test methods 1 and 2, respectively)

$\mu_1 = \mu_2$	σ_1^2	σ_2^2	λ_{11}	λ_{21}	λ_{12}	λ_{22}	**Reiczigel**	**Bab-Pad**	**Br-Mun**	**A-M1**	**A-M2**
$n_1 = n_2 = 15$ and $2\alpha = 5\%$											
3.5	69.10	109.63	0.30	0.10	1.38	1.52	4.36	6.67	6.88	*5.57*	5.14
3.5	69.10	42.47	0.30	0.50	1.38	1.23	5.01	4.73	7.27	*6.05*	5.54
3.5	42.47	42.47	0.50	0.50	1.23	1.23	3.79	4.53	5.41	*4.60*	4.30
3.5	24.44	42.47	0.70	0.50	1.05	1.23	5.69	7.88	8.02	*6.72*	6.15
$n_1 = n_2 = 15$ and $2\alpha = 10\%$											
3.5	109.63	69.10	0.10	0.30	1.52	1.38	9.50	9.50	12.05	*5.97*	9.82
3.5	42.47	69.10	0.50	0.30	1.23	1.38	10.17	13.51	12.59	*6.34*	10.27
3.5	42.47	42.47	0.50	0.50	1.23	1.23	8.13	10.25	10.24	*4.87*	8.29
3.5	42.47	24.44	0.50	0.70	1.23	1.05	11.16	10.28	13.04	*6.71*	11.09

* The density of lognormal distribution that we consider is of the following form:

$$f(x) = \frac{1}{x\lambda_{i2}\sqrt{2\pi}} e^{-\frac{(\ln(x)-\lambda_{i1})^2}{2\lambda_{i2}^2}},$$

for $x \geq 0$, $\lambda_{i2} > 0$ and $i = 1, 2$.

It is clear from the results that the performance of ASCLT-test method 2 which was nearly the best in the parametric situation is only of a satisfactory nature here. One possibility to fine-tune the method to produce better results is by exploring and modifying the coefficients $\kappa_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$ carefully. But it has to be noted that we arrived at the numerical values provided in Table 4.8 after extensive simulation-based studies. Moreover, the procedure directly implementing the Theorem on ASCLT for Rank Statistics (Theorem 16), should be performed to observe the pattern of results from them.

4.5 Conclusion

We discussed several analysis methods for the parametric one-sample and two-sample t-test situations, the BFP and the NP-BFP. We also had some detailed discussion surrounding the BFP. Along with the other existing methods, small-sample approximation for the ASCLT-test methods were also presented. Extensive simulations were performed to evaluate these methods and also to compare the independently produced results in terms of their power and maintenance of pre-assigned level. From a general overview, we see that the results of ASCLT-test method 2, in most of the cases, performs either comparably to the existing standard method(s) or even better than all competing methods. The simulation results from ASCLT-test method 1 were always presented, but not discussed. This is because, we are interested in leaving the results for future explorations of this method from the point of view of almost sure decisions regarding the concerned hypothesis.

Chapter 5

Discussion and Conclusion

5.1 Further Plans of Research and Open Problems

By virtue of being a new approach of testing based on ASCLT, the field throws open a lot of interesting and challenging problems - both from pure mathematical viewpoint, as well as from the perspective of applications. Though the open problems could be addressed by taking solely one or the other viewpoint, according to our experience, the best solutions are possible with the mixture of mathematical theory along with supporting simulation-based, application-oriented evaluation. To some extent this was done in this thesis, wherein we maintained a balance between the asymptotic theory and real-life, small sample approximations.

Throughout the thesis, wherever appropriate, there have been suggestions for modification, improvement and new proposals for further research. There are also many more suggestions possible. Here we highlight only the major open questions to be addressed and immediate consequences of work to be done.

- The mathematics of using the function $\widetilde{G}_N$ proposed in Chapter 2, for applications should be explored in the similar manner as was done using function $\widehat{G}_N$. This could lead to results similar to the one using ASCLT-test method 2, since in ASCLT-test method 2 the estimated quantiles are adjusted for μ after estimation and when using $\widetilde{G}_N$ a similar adjustment would already happen at the stage of estimation of the quantiles.

- As mentioned in Chapter 4, one of the immediate, important further development of the methods proposed in the situation with two-sample testing is regarding the unequal sample sizes. Though the proposed procedures in Chapter 4 can be implemented in situations with unequal sample sizes, their properties with respect to maintaining level and power are not satisfactory. Since many real-life situations involve samples of different sizes, this is important to be carried out in order to facilitate wider usage of the new ASCLT-test method.

- Mathematical formulation of the coefficients $\kappa_{N,2\alpha}$ and $\lambda_{N,2\alpha}$ proposed in Chapter 4, Section 4.2.2 (on page 35) and other such coefficients in other subsequent sections of that chapter have to be worked out. This would be very helpful for general applications and there would be no need to look up tables like 4.1, 4.3 and 4.8.

- The approach we have taken here is to consider a interval-based testing - more specifically, without the conventional '*p-value*' concept. Thus, another possible approach using the ASCLT is the one similar to the concepts of bootstrap or permutation tests, namely the '*Achieved Significance Levels*'. With this approach, a p-value is constructed based on the ASCLT directly and the decision could be based on whether this p-value is greater or lesser than or equal to 2α.

- The concept of Group Sequential Trials/Experiments is getting more common in clinical trials, as there are several advantageous of following such an approach. One of the main problems in implementing such trials is due to inflation of level α because of the multiple tests performed during the course of the trial. Whereas, this problem of α-inflation could be solved by using the ASCLT procedure in such multi-stage trials. This idea has to be explored in detail as well.

- We performed some further simulation studies for multi-sample, one-way layout design, with a similar approach that we adopted in the two-sample testing problems done in Chapter 4. Though the results are not presented here, we observed that the method performs better also in the one-way layout and shows clear and positive signs of extending to complex designs. Two main problems in such extensions are the issues with unequal sample sizes and the mathematical justification with respect to the coefficients. These have to addressed adequately and the methods developed further.

5.2 Conclusions

This thesis presented some existing theoretical results on the ASCLT. This was followed by proposals for estimating quantiles of the distribution of the concerned statistics. Two methods of hypothesis testing based on these estimated quantiles were proposed. The ASCLT for rank statistics was also proposed and proved.

From the asymptotic, theoretical results presented, modifications for small-sample-based applications were proposed in Chapter 4. We considered the parametric one sample test, the BFP and NP-BFP. Of these, the BFP was discussed in detail with several indications to the past literature and reviews. The results of ASCLT-test method 2 was very satisfactory in nearly all the situations presented and compared to existing, standard methods. For the BFP, this method was having similar and comparable results to the SWS approximation, which is widely accepted as the best approximate solution for the BFP. Similarly, ASLCT-test method 2 for NP-BFP performed quite satisfactorily. On the other hand, though having a similar resampling-based approach, the bootstrap methods, seem to completely breakdown in several situations discussed.

5.3 Future Outlook

The fundamental statistical theories of hypothesis testing were laid in early Twentieth century. Some of the remarkable developments concerning hypothesis testing were due to series of papers by Neyman and Pearson starting from 1928 (Neyman and Pearson, 1928). These theories were further developed and fine-tuned by mathematical statisticians until the concepts were universally set and well accepted among the scientific community. It is evident that such theories were developed on pure mathematical and hypothetical setting and there were not enough computing facilities to simulate artificial settings in order to sufficiently evaluate or modify the theory.

On the other hand, with the advent of rapid and efficient computing, in recent times, several methods based on the ideas of Bootstrapping, Jack-knifing, Re-sampling, etc,. are being proposed. Such procedures are built on the conventional foundations for the theory of hypothesis testing, and thus, they improve only the methods of arriving at inference - not directly the inference itself. There has not been much thinking into reinventing the basics of statistical decision-making process, as such, in order to overhaul the theory in an effort to make better and more reliable decisions.

In this thesis, we get a step ahead in this direction of starting a new way of thinking and looking at statistical decision problems. We proposed the usage the ASCLTs to construct conventional-type hypothesis-testing procedure. At the same time, we also proposed and left open (for further development) the new concept of, what we call, almost sure decision theory (via the ASCLT-test method 1). This procedure has to be explored very carefully, from mathematical, philosophical and application-oriented approaches.

Bibliography

Akritas, M. G., S. F. Arnold, and E. Brunner (1997). Nonparametric hypothesis and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association 92*, 258 – 265.

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika 35*, 88 – 96.

Atlagh, M. and M. Weber (1992). Un théorème central limite presque sûr relatif à des sous-suites. *C. R. Acad. Sci. Paris Sér. I. 315*, 203 – 206.

Atlagh, M. and M. Weber (2000). Le théoreme central limite presque sûr. *Expo. Math. 18*, 97 – 126.

Atlagh, M. (1993). Théorème central limite presque sûr et loi du logarithme itéré pour des sommes de variables aléeatoires indéependantes. *C. R. Acad. Sci. Paris Sér. I. 316*, 929 – 933.

Atlagh, M. (1996). Almost sure central limit theorem and associated invariance principle for sums of independent random variables. *Pub. Inst. Stat. Paris XXXX*, 3 – 20.

Babu, G. J. and A. R. Padmanabhan (2002). Re-sampling methods for the non-parametric Behrens-Fisher problem. *Sankhyā : The Indian Journal of Statistics (Series A) 64*, 678 – 692.

Banerjee, S. (1961). On confidence interval for two-means problem based on separate estimates of variances and tabulated values of t-table. *Sankhya 23*, 359 – 378.

Barnard, G. A. (1984). Comparing the means of two independent samples. *Applied Statistics 33*, 266 – 271.

Behrens (1929). Ein beitrag zur fehlerberechnung bei wenigen beobachtungen. *Landwirtschaftliche Jahrbücher 68*, 807 – 837.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of American Statistical Association 83*, 687 – 697.

Berkes, I. and E. Csáki (2001). A universal result in almost sure central limit theory. *Stochastic Processes and their Applications 94*, 105 – 134.

Berkes, I. and H. Dehling (1993). Some limit theorems in log density. *The Annals of Probability 21*, 1640 – 1670.

Berkes, I. (1998). Results and problems related to the pointwise central limit theorem. In B. S. (ed.) (Ed.), *Asymptotic Methods in Probability and Statistics - A volume in Honour of Miklos Csorgo*, pp. 59 – 96. Elsevier, Amsterdam.

Bolstad, W. M. (2004). *Introduction to Bayesian Statistics*. John Wiley and Sons, New York.

Bozdogan, H. and D. E. Ramirez (1986). An adjusted likelihood-ratio approach to the Behrens-Fisher problem. *Communications in Statistics - Theory and Methods 15*, 2405 – 2433.

Brosamler, G. (1988). An almost everywhere central limit theorem. *Mathematical Proceedings of the Cambridge Philosophical Society 104*, 561 – 574.

Brunner, E. and M. Denker (1994). Rank statistics under dependent observations and applications to factorial designs. *Journal of Statistical Planning and Inference 42*, 353 – 378.

Brunner, E. and U. Munzel (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal 42*, 17 – 23.

Brunner, E. and N. Neumann (1982). Rank tests for correlated random variables. *Biometrical Journal 24*, 373 – 389.

Brunner, E. and N. Neumann (1986). Two-sample rank tests in general models. *Biometrical Journal 28*, 395 – 402.

Carpenter, J. and J. Bithell (2000). Bootstrap confidence intervals when, which, what? a practical guide for medical statisticians. *Statistics in Medicine 19*, 1141 – 1164.

Casella, G. and R. L. Berger (2002). *Statistical Inference* (Second ed.). Duxbury - Thomson Learning, USA.

Chernoff, H. and I. R. Savage (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *The Annals of Mathematical Statistics 29*, 972 – 994.

Cochran, W. G. and G. M. Cox (1957). *Experimental Designs* (Second ed.). John Wiley & Sons, Inc., New York.

Compagnone, D. and M. Denker (1996). Nonparametric tests for scale and location. *Journal of Nonparametric Statistics 7*, 123 – 154.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge.

Denker, M. and U. Rösler (1985). Some contributions to the chernoff-savage theorems. *Statistical Decisions 3*, 49 – 75.

Dong, L. B. (2004). Econometrics working paper ewp0404: The Behrens-Fisher problem: An empirical likelihood approach. Technical report, Department of Economics, University of Victoria, Canada.

Duong, Q. P. and R. W. Shorrock (1992). An empirical bayes approach to the Behrens-Fisher problem. *Environmentrics 3*, 183 – 192.

Dupac, V. and J. Hájek (1969). Asymptotic normality of simple linear rank statistics under alternatives II. *The Annals of Mathematical Statistics 40*, 1992 – 2017.

Efron, B. and R. J. Tibshirani (1993). *An introduction ot the Bootstrap.* Chapman & Hall, New York.

Fisher, A. (1987). Convey invariant measure and pathwise central limit theorem. *Advances in Mathematics 63*, 213 – 246.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics 6*, 391 – 198.

Fisher, R. A. (1941). The asymptotic appraoch to behrens's integral, with further tables for the d test of significance. *Annals of Eugenics 11*, 141 – 172.

Fligner, M. A. and G. E. Policello II (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of American Statistical Association 76*, 162 – 168.

Gönen, M., W. O. Johnson, Y. Lu, and P. H. Westfall (2005). The bayesian two-sample t test. *American Statistician 59*, 252 – 257.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *The Annals of Mathematical Statistics 39*, 325 – 346.

Hall, P. and M. Martin (1988). On the bootstrap and two-sample problems. *Australian Journal of Statistics A30*, 179 – 192.

Holzmann, H., S. Koch, and A. Min (2004). Almost sure limit theorems for u-statistics. *Statistics & Probability Letters 69*, 261 – 269.

Howe, W. G. (1974). Approximate con dence limits on the mean of x + y where x and y are two tabled independent random variables. *Journal of the American Statistical Association 69*, 789 – 794.

Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hy potheses and the generalized Behrens-Fisher problem. *Statistics & Probability Letters 36*, 9 – 21.

Jeffreys, H. (1939). *Theory of Probability.* Oxford University Press.

Kendall, M. G. and A. Stuart (1973). *The Advanced Theory of Statistics* (Third ed.), Volume II. Griffin, London.

Lee, A. F. S. and J. Gurland (1975). Size and power of tests for equality of means of two normal populations. *Journal of the American Statistical Association 70*, 933 – 944.

Lee, P. M. (1997). *Bayesian Statistics - An Introduction* (Second ed.). Arnold Publishers, London.

Lehmann, E. L. (1986). *Testing Statistical Hypothesis* (Second ed.). John Wiley & Sons, New York.

Lesigne, E. (1999). Almost central limit theorem for strictly stationary processes. *Proceedings of the American Mathematical Society 128*, 1751 – 1759.

Lévy, P. (1937). *Théorie de l'addition des variables aléatories.* Gauthier-Villars, Paris.

Mehta, J. S. and R. Srinivasan (1970). On the Behrens-Fisher problem. *Biometrika 57*, 649 – 655.

Miller, R. G. (1997). *Beyond ANOVA, Basics of Applied Statistics* (Second ed.). Springer-Verlag, New York.

Min, A. (2004). *Limit theorems for statistical functionals with applications to dimension estimation*. PhD Thesis, Institute of Mathematical Stochastics, University of Göttingen, Germany.

Moser, B. K. and G. R. Stevens (1992). Homogeneity of variance in the two sample mean test. *American Statistician 46*, 19 – 21.

Neyman, J. and E. S. Pearson (1928). On the use and interpretation of certain test criteria for the purposes of statistical inference - parts 1 and 2. *Biometrika 20A*, 175 – 240, and 263 – 294.

Patil, V. H. (1964). The Behrens Fisher problem and its bayesian solution. *Journal of Indian Statistical Association 2*, 21 – 31.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society 4*, 119 – 130.

Pyke, R. and G. R. Shorack (1968). Weak convergence of a two-sample empirical process and a new approach to chernoff-savage theorems. *The Annals of Mathematical Statistics 39*, 755 – 771.

Reiczigel, J., I. Zakariás, and L. Rózsa (2005). A bootstrap test of stochastic equality of two populations. *The American Statistician 59*, 156 – 161.

Robinson, G. K. (1976). Properties of students t and of the Behrens-Fisher solution to the two means problem. *The Annals of Statistics 4*, 963 – 971.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin 2*, 110 – 114.

Schatte, P. (1988). On strong version of central limit theorem. *Mathematische Nachricten 137*, 249 – 256.

Scheffé, H. (1970). Practical solutions of the Behrens-Fisher problem. *Journal of American Statistical Association 65*, 1501 – 1508.

Sen, P. K. (1962). On studentized non-parametric multi-sample location test. *Annals of Mathematical Statistics 14*, 119 – 131.

Smith, H. F. (1936). The problem of comparing the results of two experiments with unequal errors. *Journal of the Council of Scientific and Industrial Research 9*, 211 – 212.

Sprott, D. A. and V. T. Farewell (1993). The difference between two normal means. *The American Statistician 47*, 126 – 128.

Strasser, H. (1985). *Mathematical Theory of Statistics: de Gruyter Studies in Mathematics 7.* Walter de Gruyter, Berlin.

Thomasse, A. H. (1974). Practical recipes to solve the Behrens-Fisher problem. *Statistica Neerlandica 28*, 127 – 138.

Tibshirani, R. (1985). Bootstrapping computations. *Proceedings of the SAS Users group conference, Reno, Neveda.*

Troendle, J. F. (2002). A likelihood ratio test for the nonparametric Behrens-Fisher problem. *Biometrical Journal 44*, 813 – 824.

Wallace, D. L. (1980). The Behrens-Fisher and fieller-creasy problems. In S. E. Fienberg and D. V. H. (eds.) (Eds.), *In R. A. Fisher: An Appreciation*, pp. 119 – 147. Springer, New York.

Wang, H. and S. C. Chow (2002). A practical approach for comparing means of two groups without equal variance assumption. *Statistics in Medicine 21*, 3137 – 3151.

Weeranhandi, S. (1995). *Exact Statistical Methods for Data Analysis.* Springer Verlag, New York.

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika 34*, 28 – 35.

Woodworth, G. G. (2004). *Biostatistics: A Bayesian Introduction.* John Wiley and Sons, New York.

Curriculum Vitae

28 Nov. 1980	Born in Chennimalai, Erode, India to Valliyammal Thangavelu and Thangavelu Laxmanan
Jun. 1984 – Apr. 1998	Primary and Secondary education at the *Vanavani Matriculation Higher Secondary School*, Chennai, India
Jun. 1998 – May 2001	Bachelor of Science in Statistics with Computer Science and Mathematics from *Loyola College, University of Madras*, Chennai, India
Oct. 2001 – Jul. 2002	Master of Philosophy in Statistical Science at the *University of Cambridge*, United Kingdom Thesis Title: Bayesian Analysis of Electronic Essay Marking Data, for Chalkface Project Supervisor: Dr. Steve Brooks
Sep. 2002 – Jan. 2006	Doctor of Philosophy in Applied Statistics and Empirical Methods (Biostatistics) at the *Department of Medical Statistics, Center for Statistics, Georg-August University of Göttingen*, Germany, under the supervision of Prof. Dr. Edgar Brunner and Prof. Dr. Manfred Denker