UNIVERSITY OF Southampton

Faculty of Engineering, Science and Mathematics

School of Mathematics

Thesis

In

Robustness of Triple Sampling Inference Procedures to Underlying Distributions

By

Ali Saleh Ali Yousef

Thesis for the degree of Doctor of Philosophy

February 2010

Acknowledgement

By the Name of GOD, Most Gracious, Most Merciful

"Facts *are many, but the truth is one*". **Rabindranath Tagore**

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Alan C. Kimber for his supervision, guidance, and continuous support for my PhD study and research at the University of Southampton.

Besides my supervisor, I gratefully acknowledge, Prof. Hosny I. Hamdy, for his motivation, encouragement, immense knowledge and insightful comments. Actually, he is the one who seriously motivated and encouraged me towards the area of sequential analysis.

My sincere thanks are due to the official referees, Prof. Christopher Jennison and Prof. Wei Liu for their constructive and excellent advice that improved the mathematical proofs of this thesis.

I would like to dedicate this thesis to my mother's soul, who taught me the meaning of passions and patience and also to my lovely boys.

Many thanks to all friends and colleagues for their advice help and support.

Last but not least; I would like to thank my family for their support and patience.

The University of Southampton, U.K.

Ali Yousef

Abstract

In this study, the sensitivity of the sequential normal-based triple sampling procedure for estimating the population mean to departures from normality is discussed. We assume only that the underlying population has finite but unknown first six moments. Two main inferential methodologies are considered. First point estimation of the unknown population mean is investigated where a squared error loss function with linear sampling cost is assumed to control the risk of estimating the unknown population mean by the corresponding sample measure. We find that the behaviour of the estimators and of the sample size depends asymptotically on both the skewness and kurtosis of the underlying distribution and we quantify this dependence. Moreover, the asymptotic regret of using the triple sampling inference instead of the fixed sample size approach, had the nuisance parameter been known, is a finite but non-vanishing quantity that depends on the kurtosis of the underlying distribution. We also supplement our findings with a simulation experiment to study the performance of the estimators and the sample size in a range of conditions and compare the asymptotic and finite sample results. The second part of the thesis deals with constructing a triple sampling fixed width confidence interval for the unknown population mean with a prescribed width and coverage while protecting the interval against Type II error. An account is given of the sensitivity of the normal-based triple sampling sequential confidence interval for the population when the first six moments are assumed to exist but are unknown. First, triple sampling sequential confidence intervals for the mean are constructed using Hall's (1981) methodology. Hence asymptotic characteristics of the constructed interval are discussed and justified. Then an asymptotic second order approximation of a continuously differentiable and bounded function of the stopping time is given to calculate both asymptotic coverage based on a second order Edgeworth asymptotic expansion and the Type II error probability. The impact of several parameters on the Type II error probability is explored for various continuous distributions. Finally, a simulation experiment is performed to investigate the methods in finite sample cases and to compare the finite sample and asymptotic results.

Table of Contents

Simulation Tables

List of Figures

List of Tables

DECLARATION OF AUTHORSHIP

I …Ali Saleh Ali Yousef…………………………………………………………………………….

declare that the thesis entitled

Robustness of triple sampling inference procedures to underlying distribution …………………….………………………………………………………………………

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission, **or** [delete as appropriate] parts of this work have been published as: [please list references]

Signed: …Ali Saleh Yousef………………………………………………………………………..

Date:……February 2010…………………………………………………………………….

Chapter I

Introduction and Review

1.1 Historical background

1.1.1 Historical developments of sequential procedures for inference

Introduction

In statistics, the term *sequential analysis* refers to a statistical analysis in which the sample size is not fixed in advance. Instead inference is made within the course of sampling. Further, sampling is terminated in accordance with a predefined stopping rule upon satisfying certain optimality criteria. Inferences may be made much earlier by using sequential procedures than would be possible with classical fixed sample size inference procedures at much lower cost.

Regardless of the type of inference we seek to make, sequential sampling procedures are derived basically under some optimality criteria. These optimality criteria could involve minimizing a given loss (cost) function while estimating the unknown parameter(s) by the corresponding sample measures, or constructing a fixed width confidence interval of a targeted parameter with a predetermined coverage, or justifying a given claim regarding the unknown parameter(s) while controlling the Type I and/or Type II error probabilities.

Let n^* be the optimal sample size required to satisfy some given optimality criterion. Usually n^* is a function of the underlying population parameters (nuisance parameters). Had n^* been known, the fixed sample size procedure would be an appropriate sampling technique to implement in order to accomplish the required inference. However n^* is unknown because the nuisance parameters are often unknown. Therefore, it would be impracticable to apply classical inference techniques (fixed sample size sampling procedures) to make inferences about the targeted parameter(s) and at the same time satisfy the predetermined optimality criteria. Alternatively, however, we may resort to sequential sampling procedures by mimicking the form of n^* by a sequential stopping rule which will terminate only if the predetermined optimality criteria are satisfied. Let N be the sequential sample size required to satisfy certain given criteria. Then clearly N is an integer-valued discrete random variable. Regarding the performance of sequential sampling procedures, several measures may need to be satisfied for a certain sequential sampling design to be declared efficient. For example, for point estimation, a sequential sampling procedure may be deemed to be efficient if it satisfies the following criteria.

$i)\ \lim_{n^*\to\infty} E\left(N/n^*\right)=1;$ that is, on the average, the ratio of the sequential sample size and the optimal sample size should be asymptotically one. Property *(i)* is referred to as first order asymptotic efficiency.

$ii)\ \lim_{n^* \to \infty} E\left(N - n^*\right) = K$, where K is a finite constant unrelated to n^*. Property (ii) reveals that on the average, the difference between the optimal and sequential sample size is asymptotically bounded by a finite constant. This property is known as second order asymptotic efficiency.

Let us define the optimal risk of the fixed sample size procedure, had n^* been known, as $E\left(L\left(n^*\right)\right)$ where $L(\cdot)$ is the loss function incurred in estimating the targeted parameters and let $E\left(L\left(N\right)\right)$ be the risk associated with a multistage sequential procedure as an alternative procedure to estimate the targeted parameters. Then the regret of using a multistage sequential sampling procedure instead of the fixed sampling procedure is defined as $\omega\left(n^*\right)$ where $\omega\left(n^*\right) = E\left(L\left(N\right)\right) - E\left(L\left(n^*\right)\right)$. Therefore, for an efficient multistage sequential procedure, it is also required to have

$iii)\ \lim_{n^* \to \infty} \omega\left(n^*\right) < \infty$. Property *(iii)* ensures that the asymptotic regret is bounded by a finite constant unrelated to n^* and also means that the multistage sequential sampling procedure is as risky as the corresponding fixed sample size procedure had n^* been known.

On the other hand, the efficiency of a fixed width confidence interval sequential sampling procedure is granted by $i)$ and $ii)$ above in addition to the measure of consistency, which affirms that the probability that the fixed width confidence interval covers the unknown parameter(s) is at least the nominal value of $100(1-\alpha)$ percent, where $1-\alpha$ is the confidence coefficient of the desired confidence interval. The above requirements are still valid to obtain a confidence region for the mean vector of a q - variate normal distribution, see Khan 1968.

Moreover, for a sequential hypothesis test to be efficient, both Type I and Type II error probabilities, α and β_t, should be controlled through the sample size required to perform the sequential procedure of the form $n \geq n^* = (a+b)^2 \theta / d^2$, where under the normal distribution, a and b are the upper $\alpha/2$ and β_t critical points of the standard normal distribution, θ is the population variance, and $d\ (>0)$ is a fixed prescribed constant. Usually the operating characteristic function (OC) is used to measure the performance of a certain test procedure. For more details, see Mukhopadhyay and de Silva (2009).

1.1.2 Historical developments of sequential sampling schemes

The idea of sequential sampling was first developed during World War II as a tool to establish more efficient quality control in equipment inspection. From Wallis' view, see Govindarajulu (1987), the story began with Captain Schyler's advice to Wallis in 1943 that one should be able to achieve some economy savings in sampling by applying the single sample test sequentially (one sample at a time). Moreover, Captain Schyler attributed this suggestion to Gwathmay, who, however, seemed to have no recollection of making such a suggestion. Wallis and Friedman had several discussions about this sequential setup, both with each other and with Wolfowitz and Paulson. Later, when Wallis and Friedman realized that the sequential method might involve a higher level of mathematical statistics, they brought the problem to Wald who actually put forward the formal theory in which optimal tests are derived for simple statistical hypotheses. Such a sequential test is known as a sequential probability ratio test (SPRT). Since then the theory and methods have been extended to a wide variety of statistical problems.

In the following section, we will demonstrate the main sampling procedures and their properties: the two stage sampling procedure proposed by Stein (1945, 1949) and Cox (1952) (also referred to as the double sampling procedure) and the one-by-one purely sequential sampling procedure proposed by Anscombe (1953), Robbins (1959) and Chow and Robbins (1965). Moreover, we will illustrate briefly Hall's (1981) three stage procedure and its properties, also referred to as the triple sampling procedure, as mentioned in his seminal paper of 1981.

1.1.2.1 Multistage sampling procedures

Multistage sampling procedures are used in statistical inference when no suitable fixed sample size procedure is available, especially when the optimal fixed sample size needed to meet certain specific goals depends on unknown nuisance parameters. The word "optimal" refers to the minimum fixed sample size needed to satisfy certain criteria had the nuisance parameters been known.

Problem

Let $\left(X_1, X_2, ..., X_n\right)$ be a random sample drawn from $N\left(\mu, \theta\right)$, where both parameters are unknown but finite with $\mu \in \left(-\infty, \infty\right)$ and $\theta \in \left(0, \infty\right)$. Given $d\left(>0\right)$ and $\alpha \in \left(0,1\right)$, we wish to construct a $100\left(1-\alpha\right)$ percent confidence interval $I = \left\{-d \le \bar{X}_n - \mu \le d\right\}$ for μ such that the length of the interval is $2d$ and $P\left(\mu \in I\right) \ge 1-\alpha$ uniformly for all μ and θ.

Solution

If θ is known, then it can be shown that $n^* = a^2\theta/d^2$ solves the above problem uniformly over all θ and n^* is referred to as the *optimal fixed sample size* required to solve the problem. So n^* is the optimal fixed sample size required to construct such a confidence interval for μ had θ been known. Here a is the upper $\alpha/2$ point of the standard normal distribution. If θ is unknown then multistage sequential procedures should be used to solve the above problem; see Mukhopadhyay and de Silva (2009) for details.

Stein's double sampling procedure

Stein (1945, 1949) and Cox (1952) proposed sequential sampling in two stages, known as double sampling. The procedure was developed mainly for constructing a fixed width confidence interval for a normal mean with unknown variance θ and with a prescribed coverage probability $1-\alpha$.

Let $\left(X_1, X_2, ..., X_m\right)$ be a random sample of size $m\left(\ge 2\right)$, and let $\bar{X}_m$ and S_m^2 be estimators of the normal mean μ and variance θ respectively. Then Stein's stopping rule is

$$N = \max\left\{m, \left[\left(t\left(m-1\right)S_m\right)^2 / d^2\right] + 1\right\},$$

where $\left[x\right]$ denotes the integer part of x, $t\left(m-1\right)$ is the upper $\alpha/2$ point of the t distribution with $\left(m-1\right)$ degrees of freedom, and $d\left(>0\right)$ is a fixed prescribed constant.

The procedure goes like this: if $N = m$, that is when m is larger than the estimator of n^*, then we do not need to undertake any more sampling at the second stage, while if $N > m$, this indicates that we have started with too few observations at the pilot stage. Hence, we take new observations $\left(X_{m+1}, X_{m+2}, ..., X_N\right)$ at the second stage. Finally we have $\left(X_1, X_2, ..., X_N\right)$, and then the interval $I_N = \left(\bar{X}_N \pm d\right)$ is the proposed interval estimator of μ, while $\bar{X}_N$ is the sequential point estimator of μ.

The properties of Stein's double sampling procedure are

$i)$ $P\left(\mu \in I_N\right) \geq 1-\alpha$, for all μ and θ

$ii)$ $\lim_{d\to 0} E\left(N/n^*\right) = \left(t\left(m-1\right)/a\right)^2 > 1$

$iii)$ $\lim_{d\to 0} E\left(N - n^*\right) = \infty$

$iv)$ $\lim_{d\to 0} P\left(\mu \in I_N\right) = 1-\alpha$, for all μ and θ.

Property *(i)* is known as consistency or exact consistency in the sense of Chow and Robbins (1965). Property *(ii)* shows that the procedure is first order asymptotically inefficient, also in the sense of Chow and Robbins (1965). Property *(iii)* shows that the procedure is second order asymptotically inefficient, while property *(iv)* is known as asymptotic consistency or first order consistency, in the sense of Chow and Robbins (1965) and Ghosh and Mukhopadhyay (1981).

Stein's procedure has the advantage of reducing the number of sampling operations, and hence reduces the operational cost. Generally, however, a double sampling procedure will lead to over sampling (i.e. the procedure is second order asymptotically inefficient). This means that the expectation of N is much greater than n^* as $n^* \to \infty$, i.e. on average, the difference between the optimal and the second stage sample size is asymptotically unbounded, especially when the initial sample size is chosen much smaller than the optimal sample size n^*.

Remark

The major success of Stein's double sampling procedure is its exact consistency property.

One-by-one purely sequential procedure

In order to reduce the problem of over sampling in the double sampling procedure, one can estimate the variance successively in a sequential manner.

Anscombe (1953), Ray (1957) and Chow and Robbins (1965) proposed the core of one-by-one purely sequential sampling for estimation, where the stopping rule is

$N \equiv N\left(d\right)$: is the smallest integer $n\left(\geq m\right)$ for which we observe $n \geq a^2 S_n^2 / d^2$.

The one-by-one purely sequential procedure is implemented as follows:

1. In the initial stage we obtain S_m^2 based on the pilot sample $\left(X_1, X_2, ..., X_m\right)$, $m(\geq 2)$ and check whether $m \geq a^2 S_m^2 / d^2$.

2. If $m \geq a^2 S_m^2 / d^2$, stop sampling at this stage and record m as the final sample size. If $m < a^2 S_m^2 / d^2$, take one additional observation X_{m+1} and update the sample variance estimator to S_{m+1}^2 based on the new sample of size $m+1$. Next check whether $m+1 \geq a^2 S_{m+1}^2 / d^2$. If so, sampling terminates and the final sample size is $m+1$, otherwise continue sampling by taking another one additional observation X_{m+2} and update the sample. This process continues until for the first time we arrive at a sample of size n which is at least as large as $a^2 S_n^2 / d^2$, constructed from $\left(X_1, X_2, ..., X_n\right)$. In other words, stop sampling as soon as the sample size exceeds the estimate of n^* and hence the confidence interval is constructed.

The asymptotic characteristics of the one-by-one purely sequential procedure based on the above stopping rule is

$i)$ $\lim_{d \to 0} E\left(N/n^*\right) = 1.$ Asymptotic first order efficiency.

$ii)$ $\lim_{d \to 0} E\left(N - n^*\right) < \infty.$ Asymptotic second order efficiency.

$iii)$ $\lim_{d \to 0} P\left\{\mu \in I_N\right\} = 1 - \alpha.$ Asymptotic consistency.

Remarks

1. The asymptotic second order efficiency is stronger than the asymptotic first order efficiency property since *(ii)* leads to *(i)* but the converse is not true.

2. The one-by-one purely sequential procedure does not over-sample as Stein's double sampling procedure does, but it fails to have the exact consistency property, unlike Stein's procedure.

Although the one-by-one purely sequential sampling procedure is more efficient than the Stein (1945, 1949) and Cox (1952) two stage procedures, it is inexact (i.e. it only attains the prescribed coverage probability asymptotically). The one-by-one purely sequential sampling procedure of Anscombe (1952, 1953), Robbins (1959) and Chow and Robbins (1965) was developed to tackle both point and confidence interval estimation problems. Specifically, in point estimation generally a loss (cost) function is assumed and hence the risk incurred in estimating the unknown population parameter by the corresponding sample measures is calculated. Then the optimal sample size, n^*, required to minimize the associated risk is obtained. Mimicking the structure of the optimal sample size, a one-by-one purely sequential procedure for point estimation is then developed. It has been shown (see, for example, Chow and Robbins, 1965 and Woodroofe, 1977) that the one-by-one purely sequential procedure is asymptotically efficient (first and second order), while the Stein (1945) and Cox (1952) two stage procedures suffer from lack of efficiency. On the other hand, the one-by-one purely sequential sampling procedure of Anscombe and of Chow and

Robbins is not cost effective since it takes time to terminate and becomes impractical. We would also emphasize that most of these sampling techniques pertain to making inferences about the normal mean.

During the period from 1965 until the early 1980's, no substantial progress was made towards devising sequential sampling procedures which satisfy both the operational savings made possible by the two stage procedure and the asymptotic efficiency of the one-by-one purely sequential procedure. Rather, during this period the research in sequential inference was devoted mainly to applications of these well established sequential sampling procedures to other distributions, such as: the exponential distribution (Basu (1971), Starr and Woodroofe (1972) and Mukhopadhyay (1974)), the negative exponential distribution (Mukhopadhyay *et. al.* (1986) and Mukhopadhyay (1988)), the difference of the means of two negative exponential populations (Mukhopadhyay and Darmanto (1988)), the difference of location parameters of two negative exponential distributions (Mukhopadhyay and Hamdy (1984)), the uniform distribution (Graybill and Connell (1964) and Ghosh and Mukhopadhyay (1975b)), the pareto distribution (Wang (1973) and Mukhopadhyay and Ekwo (1987a)), the Bernoulli distribution (Robbins and Siegmund (1974) and Cabilio (1977)), the binomial distribution (Wolfowitz (1946), Degroot (1959) and Wasan (1964)), the negative binomial distribution (Mukhopadhyay and Diaz (1985)), the Poisson distribution (McCabe (1970)), the hypergeometric distribution (Ifram (1965)), the bivariate normal distribution (Sinha and Mukhopadhyay (1976)) and the lognormal distribution (Zacks (1966)), the range in a power family distribution (Mukopadhyay *et. al.* (1983)) and the multivariate normal distribution (Khan (1968), Wang (1981), Callahan (1969) and Ghosh *et al.* (1976)). Other applications involved classes of distributions such as the one-parameter exponential family; see Lorden (1978), McCabe (1974) and Mukhopadhyay (1974).

Hall's (1981) triple sampling sequential procedure

In the early 1980's a three stage procedure was introduced by Hall (1981, 1983) to construct a fixed width confidence interval for the normal mean μ with unknown finite variance θ and with a predetermined coverage probability $1-\alpha$. His procedure is based on the following stopping rule

$$N_1 = \max\left\{m, \left[\delta\left(a^2 S_m^2 / d^2\right)\right] + 1\right\},$$

and

$$N = \max\left\{N_1, \left[\left(a^2 S_{N_1}^2 / d^2\right)\right] + 1\right\}.$$

A modified version of his procedure is based on modifying the last stage to

$$N = \max\left\{N_1, \left[\left(a^2 S_{N_1}^2 / d^2\right) + \left(5 + a^2 - \delta\right) / 2\delta\right] + v + 1\right\},$$

where a is the $\alpha/2$ critical point of $N(0,1)$, $v \geq 0$ is an arbitrary integer added to improve the convergence of the procedure, $d(>0)$ is a fixed prescribed constant and δ is a fixed number between $\delta \in (0,1)$.

We will illustrate Hall's triple sampling procedure in detail in Chapter II section 2.2.

The asymptotic characteristics of the triple sampling sequential procedure based on the above stopping rules are

$i)$ $\lim_{d \to 0} E\left(N/n^*\right) = 1.$

$ii)$ $\lim_{d \to 0} E\left(N - n^*\right) < \infty.$

$iii)$ $\lim_{d \to 0} P\left\{\mu \in I_N\right\} = 1 - \alpha.$

Hall's asymptotic theory begins with the assumption that for $m \geq 2$ and $s \geq 1$

$\lim_{m \to \infty} sup\left(m/n^*\right) < \delta$ and $n^* = O\left(m^s\right)$.

Since Hall's triple sampling procedure is the main focus of this thesis, so it would be better to show the effect of increasing the arbitrary number v on improving the coverage probability.

Table 1.1 as seen in Hall (1981, p. 1232) shows the effect of increasing the number v on improving the performance of the coverage probability based on 1000 replicate samples from the standard normal distribution with $\delta = 0.5$, $1 - \alpha = 0.95$ and $m = 10$. For more details see Hall (1981, p. 1232).

n^*	$v = 0$	$v = 3$	$v = 5$	$v = 8$
24	0.950	0.964	0.973	0.978
43	0.956	0.949	0.958	0.963
61	0.949	0.948	0.955	0.962
76	0.953	0.959	0.945	0.952
96	0.930	0.942	0.951	0.953
125	0.948	0.955	0.949	0.951
171	0.936	0.970	0.954	0.954
246	0.956	0.952	0.959	0.958
384	0.958	0.958	0.954	0.954

Table 1.1: The effect of increasing v on the performance of the coverage probability

Remark

The two-stage procedure, the one-by-one purely sequential procedure and the triple sampling procedure enjoy the property that $P\left(N < \infty\right) = 1.$

Hall's three stage procedure combines the efficiency of the Anscombe, Chow and Robbins one-by-one purely sequential procedure and the operational saving made possible by sampling in bulks by applying Stein's group sampling techniques.

Since the introduction of multistage sequential sampling by Hall (1981, 1983), applications to other distributions and inferences for other parameters have been of interest; see Hamdy and Pallotta (1987), AlMahmeed *et al.* (1998, 1990), Hamdy *et al.* (1995), Mukhopadhyay (1985, 1988, 1990), Mukhopadhyay and Mauromoustakos (1987) and Mukhopadhyay and Padmanabhan (1993).

Mukhopadhyay (1990) made further developments to triple sampling by focusing on higher order moments of the stopping variable N. Hamdy (1988) extended Hall's (1983) triple sampling results and proposed a triple sampling point estimation procedure to estimate the normal mean. The extension of Hall's results to tackle hypothesis testing problems of the normal mean was developed by Liu (1995). Son *et al.* (1997) proposed a triple sampling sequential procedure which yields both a fixed width confidence interval and a hypothesis test for the normal mean while controlling the Type II error probability. Their procedure also provided second order approximations to the operating characteristic curves of the inference.

Closely related applications of sequential procedures in clinical trials are found in Whitehead (1983, 1991, 1992, 1994, 1997, 2001 and 2005), Brunier and Whitehead (1993) and Jennison *et al.* (1999).

Accelerated sequential procedure

We have seen that the one-by-one purely sequential procedure is asymptotically second order efficient, but it has a disadvantage that we need to record observations one by one until the process terminates. To accelerate the one-by-one purely sequential procedure and to preserve second order asymptotic efficiency, we combine the one-by-one purely sampling procedure with an additional stopping rule. Such a procedure is called an improved accelerated sequential procedure. The procedure as proposed by Mukhopadhyay (1996) is: start with a pilot sample $X_1, \ldots, X_m$ of size $m(\geq 2)$ and let δ be a fixed number between 0 and 1. Then the stopping rule is

$$N_1 = \min\left\{n \geq m,\ n \geq \delta a^2 S_m^2 / d^2\right\},$$

and the final sample size is

$$N = \left\{\left(\delta^{-1} N_1 + q\right) + 1\right\} \ , \ q = (2\delta)^{-1}\left(5 + a^2\right),$$

where a is the upper $\alpha/2$ point of the standard normal distribution and $d\left(>0\right)$ is a fixed prescribed constant. The final data set is $\left\{X_1, \ldots, X_{N_1}, X_{N_1+1}, \ldots, X_N\right\}$. Then the fixed width confidence interval for the normal mean μ is $I_N = \left(\bar{X}_N \pm d\right)$. For more details, see Mukhopadhyay (1996).

The asymptotic characteristics of the accelerated sequential procedure are summarized in Mukhopadhyay (1996) as follows:

(i) $\lim_{d \to 0} E\left(N/n^*\right) = 1$, asymptotic first order efficiency.

$(ii)\ \lim_{d \to 0} P\{\mu \in I_N\} = 1 - \alpha$, asymptotic consistency.

$(iii)\ \lim_{d \to 0} E\left(N - n^* \right) < \infty$, asymptotic second order efficiency.

Remark

If δ is chosen near zero, the accelerated sequential procedure would clearly behave more like Stein's two stage procedure, but if chosen near one it would behave more like a one-by-one purely sequential procedure. Therefore, an accelerated sequential procedure is often implemented with $\delta = 0.4$, 0.5 or 0.6. In numerous problems, one tends to use $\delta = 0.5$; see Mukhopadhyay and de Silva (2009).

The objective in this thesis is to study the robustness of normal-based triple sampling inference procedures to departures from normality of the underlying distribution. Therefore, it is useful to give some brief statements about the meaning of robustness in statistics and its role in the area of sequential sampling.

1.2 Robustness of sequential procedures

Since many statistical methods make specific assumptions about the nature of the underlying distribution, it has long been a concern of statisticians to determine how far conclusions might be affected if these assumptions were false. In particular a considerable literature exists on the effect of non-normality on analysis of variance tests to compare means and on the test to compare the variances of independent samples; see Geary (1936), Gayen (1950), and Box and Anderson (1955).

The word *robustness* in statistics was first coined by Box in 1953. Box *et al.* (1964) studied the behaviour under non-normality of the probability distribution of a specific criterion. Such specific property is referred as criterion robustness to non-normality. Moreover, they distinguished between two types of sensitivity to non-normality: *criterion robustness* and *inference robustness*. Criterion robustness means that the distribution of the statistic used to estimate parameters or test hypotheses about the parameters under the original model is not substantially affected by changing the model. Inference robustness means that inferences made about parameters on the basis of the data do not change substantially with a change in the model. For example, changes in the significance level when appropriate changes were made in the nature of the criterion to correspond with the changes in the underlying distribution. An excellent discussion of this distinction between the two types of robustness was given by Box *et al.* (1973) where he illustrated this distinction for the t and F distributions of normal theory statistical inference.

Huber (1964) used the word robustness to mean the insensitivity of a statistical procedure to small deviations from the assumed assumptions. He was concerned with *distributional robustness* where the shape of the true underlying distribution deviates slightly from the assumed model, usually the normal distribution.

In real life problems, underlying distributions are not perfectly known and can be a member of larger class of distributions. For example, the distribution of errors which are assumed to be normal could belong to some larger symmetric class of distributions not necessarily normally distributed; or the true underlying distribution might be a mixture of several normal distributions or a contaminated normal distributions. Also in linear models, least square estimators are sensitive to deviations from an assumed normal distribution of the errors, thus least square estimators are non-robust with respect to deviations from the assumed normal distribution.

Pearson (1931) showed the sensitivity of classical ANOVA procedures to departures from the assumed normal model, mostly in terms of the skewness and kurtosis of the underlying distribution. Many articles examine the effect of such deviations on the size and power of the ANOVA tests; see Tukey (1960). The robustness of the t test against mixtures of normal populations that differ in location parameters has been studied by Tukey and McLaughlin (1963). Subrahmaniam *et al.* (1975) studied the robustness of the one-sample t procedure against slight contamination of the population with another normal population having a different mean.

As we know the t test, based on $t = \sqrt{n}\left(\bar{X} - \mu\right)/S$, is used to test the hypothesis about the population mean of a normal distribution when the variance is unknown. The power of this test is a function of the unknown variance. It was shown by Dantzig (1940) that for a fixed sample size there does not exist a test whose power is independent of the variance. The idea of studying the effect of non-normality on the t test has been investigated by, for example, Pearson and Adyanthaya (1929), Bartlett (1935), Geary (1936), Gayen (1949), Ghurye (1949) and Srivastava (1958). Pearson and Adyanthaya (1929) have shown that the effect of skewness and kurtosis of the underlying distribution on t may be considerable. Bartlett (1935) confirmed Pearson's results theoretically by obtaining an approximate distribution of t in non-normal samples, assuming the underlying distribution can be represented by the first two terms of Edgeworth series. Geary (1936), obtained the approximate of the t distribution. Gayen (1949) considered the effect of skewness and kurtosis by using the first four terms of the Edgeworth series as the density function of the population to derive the distribution of t. A theoretical study on the effect of non-normality on the power of the t test was first made by Ghurye (1949) by considering the first two terms of the Edgeworth series and later Srivastava (1958) extended this work by considering the effects of the skewness and kurtosis of the underlying distribution.

Stein (1945) gave a two sample test for a linear hypothesis whose power is independent of the unknown variance. He used it to test the hypothesis about the mean of a normal population and to estimate the mean by a confidence interval of a prescribed width with a given confidence coefficient. As in other tests of significance, the basic assumption in Stein's test is the normality of the underlying distribution. Since this assumption may not hold in practice, Bhattacharjee (1965) was the first to study the robustness of Stein's two stage procedure against departures from normality by deriving the distribution of Stein's t test for non-normal populations represented by the first four terms of an Edgeworth series. He considered the power function of Stein's test and the confidence level of the fixed width confidence interval. He concluded that the procedure is sensitive to Edgeworth type of expansion. On the contrary, Blumenthal and Govindarajulu (1977) investigated the departure of Stein's two stage results from normality under the assumption that the underlying distribution is a mixture of two normal populations with common unknown variance but with different means. Both of these studies are concerned with criterion robustness of the procedure (Box and Tiao, 1973). Their results indicated that the procedure is remarkably robust. The controversy between these two different conclusions was settled by Ramkaran (1983) who investigated the same problem and found that the Stein's two stage sampling procedure is quite robust even under Edgeworth series model.

Applications of robust statistical procedures can be found in Hampel (1968, 1971), Hampel *et al.* (1986), Huber (1964, 1981), Huber and Dutter (1974), Govindarajulu and Leslie (1972) and Jureckova and Sen (1996).

Although it is vital, as one can see, the quantity of research in the area of robustness in sequential sampling is limited. However, Jureckova and Sen (1996) devoted several chapters to discuss robustness of sequential statistical inference (point and interval estimation).

In this current study the robustness of triple sampling procedures to non-normality of the underlying distribution will be of greatest importance. The problem arises when the underlying population is misspecified (normality is assumed when in fact it is not normal). Also for some distributions it is almost impossible to express the optimal sample size explicitly and therefore, if the normal-based triple sampling procedure were robust to departures from normality, it would be convenient to use the triple sampling procedure with normal stopping rule. Several measures are devised to assess the robustness of sampling procedure. However, in this study we rely on the skewness and kurtosis of the underlying distribution to measure the extent of departures from normality.

1.3 Skewness and kurtosis

In this section we shall define what we mean by skewness and kurtosis.

It is commonly noted that distributions can be characterized in terms of location or central tendency, variation and shape. With respect to shape, two important measures are the skewness γ, which is a measure of asymmetry, and the kurtosis β, which is a measure of peakedness and/or tail behaviour. Higher kurtosis indicates more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly sized deviations. The formulae that we will use are given by Abramowitz and Stegun (1972) who define the skewness and kurtosis as

$\gamma = E\left(X-\mu\right)^3 \big/ \theta^{3/2}$ and $\beta = E\left(X-\mu\right)^4 \big/ \theta^2$, for all $\mu \in \mathbb{R}, \theta \in \mathbb{R}^+$.

Note that the kurtosis is always greater than one, $(\beta > 1)$ if it exists. When appropriate we will also use the excess kurtosis, $\beta^* = \beta - 3$; see Kenney and Keeping (1951).

In the literature, there are different perspectives for the meaning of kurtosis. Bickel and Lehman (1975a, 1975b) noticed that there is no agreement on precisely what kurtosis measures, while Balanda *et al.* (1988) showed that it is better to define kurtosis obscurely as the location and scale-free movement of a probability density function from the shoulders of a distribution into its centre and tails, and to recognize that it can be formalized in many different ways. Pearson (1905) defined kurtosis as a measure of how flat the top of a symmetric distribution is when compared to a normal distribution of the same variance, while others like Johnson *et al.* (1994) illustrated that kurtosis measures the amount of deviation from normality depending on the relative frequency of values either near the mean or far from it to values an intermediate distance from the mean. Wilcox (1990) used the skewness and kurtosis in studies of robustness of normal theory procedures. The Pearson family of distributions is characterized by the first four moments and skewness and kurtosis may be used to help to select an appropriate member of this family. All such differences occur because kurtosis is not well understood and because the role of kurtosis in various statistical analyses is not widely recognized. For more details about the role of kurtosis, see Hampel (1968, 1974).

The terms mesokurtic, leptokurtic and platykurtic mean as follows:

(i) Mesokurtic: Such distributions have a moderate degree of peakedness and represented by a normal distribution, symmetric around the mean and the three central measures, the mean, the median and the mode are equal. Note that all normal distributions are mesokurtic and the weight/thickness of the tails of a normal distribution is in between the weight/thickness of the tails of distributions that are leptokurtic or platykurtic.

(ii) Leptokurtic: Such distributions have a high degree of peakedness. The tails are heavier\ thicker than the tails of a mesokurtic distribution. An example of this case is the exponential distribution.

(iii) Platykurtic: Such distributions have a low degree of peakedness. The tails are lighter\thinner than the tails of a mesokurtic distribution. An example of this case is the uniform distribution. See Sheskin (2004) for more details.

Table 1.2 gives the values of the skewness γ and kurtosis β for some distributions that we will use in later chapters: the normal distribution $N(\mu,\theta)$, the uniform distribution $U(a,b)$, the t distribution with r degrees of freedom $t(r)$, the beta distribution $beta(a,b)$, the exponential distribution with mean μ, $Exp(\mu)$ and the chi-squared distribution with r degrees of freedom $\chi^2(r)$.

Distribution	γ	β
$N(\mu,\theta)$	0	3.0
$U(a,b)$	0	1.8
$t(r)$	$0,\ \forall\ r>3$	$3(r-2)(r-4)^{-1},\ \forall\ r>4$
$beta(a,b)$	$\frac{2(b-a)}{(a+b+2)}\sqrt{\frac{a+b+1}{ab}}$	$\frac{3(a+b+1)\left(2a^2+2b^2-2ab+a^2b+ab^2\right)}{ab\ (a+b+2)(a+b+3)}$
$Exp(\mu)$	2	9.0
$\chi^2(r)$	$2\sqrt{2/r}$	$12/r+3$

Table 1.2: Skewness and kurtosis for $N(\mu,\theta), U(a,b), t(r)$, beta $(a,b), Exp(\mu)$ and $\chi^2(r)$ distributions

Chapter II

Statement of the Problem

2.1 Problem setting

Let $X_1, X_2, X_3, \ldots$ be a sequence of $i.i.d.$ random variables from a continuous distribution function $F(\,.;\mu,\theta)$, where the two parameters, the mean $\mu \in \mathbb{R}$ and the variance $\theta \in \mathbb{R}^+$ are assumed unknown but finite. We also assume that the skewness γ and the kurtosis β are both unknown but finite $\gamma < \infty$ and $\beta < \infty$. The main interest in this study is estimation of μ in the presence of the unknown variance θ.

Having observed a random sample $X_1, X_2, \ldots, X_n$, $\forall\, n \geq 2$ from the distribution function $F(\,.;\mu,\theta)$, we propose to use the sample mean and the sample variance as point estimates of μ and θ respectively: $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$, $\forall n \geq 2.$

It is well-known that these estimators are unbiased for their respective parameters and minimal sufficient statistics in the case of the normal distribution.

In the literature on sequential sampling for inference of the mean for most distributions it is assumed that the sample size required to satisfy the conditions *(i)*, *(ii)* and *(iii)* in section 1.1.1 can take the general form (2.1) below; see Sen (1985) and Ghosh *et al.*(1997) for details.

$$(2.1) \qquad n^* = \lambda\, g(\,\theta\,),$$

where λ depends on some predetermined constants (which may, for example, appear in a loss (cost) function incurred in point estimation of μ or arise from consideration of a fixed width confidence interval for μ with a prescribed coverage probability). Further, λ is permitted to tend to infinity if the optimal sample size $n^* \to \infty$. Note that $g(\,\cdot\,)$ is a positive real valued twice continuously differentiable function such that g, g' and g'' are bounded. We shall use the representation (2.1) in this thesis to develop theory for point and interval estimation and for hypothesis testing.

2.2 Triple sampling procedure for inference for the population mean

Since n^* in (2.1) is numerically unknown because θ is unknown, then no fixed sample size procedure provides the above point estimation for μ uniformly for $\forall\,\theta > 0$. Therefore, we use sequential sampling procedures to estimate μ via estimation of the optimal sample size n^*. We now give a rigorous account of the triple sampling procedure as described by Hall (1981). As the name suggests, triple sampling can be described by the three phases.

Pilot Phase: Here an initial sample $X_1, X_2, ..., X_m$ of size $m \geq 2$ is taken at random from the distribution $F(.;\mu,\theta)$, from which $\bar{X}_m$ and S_m^2 are calculated as our initial estimates for μ and θ respectively.

The Main Study Phase: In the main study phase a fraction of n^* is estimated during that phase, say, δn^*, where δ is a fixed number between 0 and 1. The sample size required to complete the main study phase is defined by the following stopping rule as

$$(2.2) \qquad N_1^* = \left[\delta \lambda g\left(S_m^2\right)\right] + 1, \ N_1 = \max\left\{m, N_1^*\right\}.$$

where $0 < \delta < 1$ and $0 < \lambda < \infty$ are known constants and $[x]$ denotes the integer part of x. Observe that N_1 estimates δn^* in this phase.

If $m \geq N_1^*$, stop sampling at this stage; otherwise continue to observe an extra random sample of size $N_1 - m$ from the distribution function $F(.;\mu,\theta)$, say $X_{m+1}, X_{m+2}, ..., X_{N_1}$. Hence, we augment the $N_1 - m$ observations by the previous m observations and calculate $\bar{X}_{N_1}$ and $S_{N_1}^2$ as new estimates of μ and θ respectively.

The fine tuning phase: This is defined according to the following final stage stopping rule

$$(2.3) \qquad N^* = \lambda g\left(S_{N_1}^2\right) + 1, \ N = \max\left\{N_1, N^*\right\}.$$

If $N_1 \geq N^*$, then stop at this stage, otherwise continue to sample $N - N_1$ more observations randomly from the distribution $F(.;\mu,\theta)$, say $X_{N_1+1}, X_{N_1+2}, \ldots, X_N$. Whenever sampling is terminated and N is realized, then $\bar{X}_N$ is a natural point estimate for μ, and hence $\bar{X}_N$ is a sequential point estimator of μ. Observe that N estimates n^* in this phase.

Throughout the thesis, the asymptotic characteristics of triple sampling are developed under the assumption made by Hall (1981) that for $0 < \delta < 1$ and λ positive, we assume that

$$(2.4) \qquad n^* \to \infty, m = m\left(n^*\right) \to \infty, \limsup\left(m/n^*\right) < \delta \ \text{ and } \ \lambda = O\left(m^s\right), \ s \geq 1.$$

where $s \geq 1$ is a fixed constant. Moreover we assume that $E|X_1|^6 < \infty$. This moment condition is the same as that used by Chow and Martinsek (1982) to estimate the mean of an unknown distribution using the one-by-one purely sequential procedure proposed by Robbins (1959). Although the assumption $E|X_1|^6 < \infty$ may seem restrictive, but we shall show in the next chapter that second order approximations of a continuously differentiable function of the stopping sample sizes N_1 and

N depend on the first four moments of the distribution in order to capture both the skewness and kurtosis of the underlying distribution. Therefore it is clear that one needs more than the first four moments to be finite to obtain such approximations and ensure that the corresponding error terms tend to zero. Also, it is necessary to put lower and upper bounds on θ in order to ensure that $g(\theta)$ and its first two derivatives are bounded in the inferential situations we shall investigate later in the thesis.

Lemma 2.1 (Hall 1981)

For the triple sampling rule (2.2) – (2.3) as $m \to \infty$ we have

$$P\left(N_1 \neq N_1^*\right) = O\left(\exp\left(-km\right)\right),$$
$$P\left(N \neq N^*\right) = o\left(m^{-q/2+1}\right), \qquad k > 0,\ q \geq 10.$$

See Honda (1992) for the proof.

Lemma 2.1 shows essentially that the probability of not completing all three stages is small for large values of m.

Remarks

1. Mukhopadhyay (1990) noted that if the design factor δ is chosen near zero or one, then a three stage procedure would clearly be rather like Stein's two stage procedure. Therefore a three stage procedure is better implemented with $\delta = 0.4$, 0.5 or 0.6. Hall (1981) mentioned that in practice it seems a reasonable compromise to choose $\delta = 0.5$.

2. In the context of two stage sampling Seelbinder (1953) and Moshman (1958) developed some criteria based on prior information about the variance in order to suggest a reasonable choice of the pilot sample size m, while Mukhopadhyay (2005a) developed an information-based approach to suggest a reasonable choice of m without any prior information about the variance.

Notes

1. The objective of Hall (1981) was to construct a fixed width confidence interval for the normal mean with a prescribed width $d\,(>0)$ and coverage $1-\alpha$ without any concern about the estimate of the optimal sample size n^*. Moreover, he used only the first order approximation of the stopping sample size N.

2. Mukhopadhyay *et al.* (1987) treated the same situation as Hall (1981) but they considered point estimation of the normal mean to achieve the minimum bounded risk.

In our thesis we deal with inference (point and interval estimation and hypothesis testing) about the mean of any continuous underlying distribution whose analytical form is unknown and for any positive twice continuously differentiable and bounded function of θ. We use the second order approximations of a suitable continuously differentiable function of the stopping sample sizes N_1 and N in order to evaluate the asymptotic regret in terms of the first four moments of the underlying distribution.

Chapter III

Developing Mathematical Results for Triple Sampling

3.1 Preliminary results

Lemma 3.1.1

Let X be a random variable drawn from the continuous distribution function $F\left(\,.;\mu,\theta\right)$, where μ is the population mean $\mu \in \mathbb{R}$, θ is the population variance, $\theta \in \mathbb{R}^{+}$, and with finite skewness γ and kurtosis β . Then, it is easy to show that

$$i)\, E\left(X^{3}\right) = \gamma\,\theta^{3/2} + 3\mu\theta + \mu^{3}.$$

$$ii)\, E\left(X^{4}\right) = \beta\,\theta^{2} + 4\mu\,\gamma\,\theta^{3/2} + 6\mu^{2}\theta + \mu^{4}.$$

$$iii)\, Var\left(X^{2}\right) = \theta^{2}\left(\beta - 1\right) + 4\mu\gamma\theta^{3/2} + 4\mu^{2}\theta.$$

Proof: *(i)* and *(ii)* follow immediately from the definition of skewness and kurtosis while *(iii)* follows from *(ii)* and the fact that $E\left(X^{2}\right) = \mu^{2} + \theta.$

Lemma 3.1.2

Let $X_1, X_2, ..., X_m$ be a random sample of size $m \geq 2$ drawn from the continuous distribution function $F\left(\,.\,;\mu,\theta\right)$ with finite mean μ , variance θ, skewness γ and kurtosis β . Let $\bar{X}_m$ and S_m^{2} be the sample mean and sample variance respectively. Then it follows that

Part (I)

$$i)\, E\left(\bar{X}_m^{3}\right) = m^{-2}\left(\gamma\,\theta^{3/2} + 3m\mu\theta + m^{2}\mu^{3}\right).$$

$$ii)\, E\left(\bar{X}_m^{4}\right) = m^{-3}\left(\left(\beta + 3m - 3\right)\theta^{2} + 4m\,\mu\gamma\,\theta^{3/2} + 6m^{2}\mu^{2}\theta + m^{3}\mu^{4}\right).$$

$$iii)\, E\left(S_m^{2} - \theta\right)^{2} = \left(m\left(m-1\right)\right)^{-1}\left(\left(\,m-1\,\right)\beta - \left(\,m-3\,\right)\right)\theta^{2}.$$

$$iv)\, E\left(\bar{X}_m\, S_m^{2}\right) = m^{-1}\gamma\,\theta^{3/2} + \mu\theta.$$

$$v)\, E\left(\bar{X}_m^{2}\, S_m^{2}\right) = m^{-2}\theta^{2}\left(\beta + m - 3\right) + 2m^{-1}\mu\,\gamma\,\theta^{3/2} + \mu^{2}\theta.$$

Part (II)

Define the location shift $Z = X - \mu$. Then

$i)\ E\left(Z^3 \right) = \gamma\, \theta^{3/2}.$

$ii)\ E\left(Z^4 \right) = \beta\, \theta^2 .$

Also for a random sample $\left(Z_1, Z_2, ..., Z_m\right)$ from the random variable Z above,

$$iii)\, E\left(\overline{Z}_m^{\,3}\right) = m^{-2}\gamma\, \theta^{3/2}.$$

$$iv)\, E\left(\overline{Z}_m^{\,4}\right) = m^{-3}\theta^2\, (\beta + 3m - 3).$$

$$v)\, E\left(\overline{Z}_m\, S_m^2\right) = m^{-1}\gamma\, \theta^{3/2}.$$

$$vi)\, E\left(\sum_{i=1}^{m} Z_i^2\, S_m^2\right) = \theta^2\left(\beta + m - 1\right).$$

$$vii)\, E\left(\sum_{i \neq j}^{m}\sum Z_i Z_j\, S_m^2\right) = -2\theta^2.$$

$$viii)\, E\left(\overline{Z}_m^{\,2}\, S_m^2\right) = m^{-2}\theta^2 (\beta + m - 3).$$

Proof:

It suffices to prove part (II) of Lemma 3.1.2. *(i)* and (*ii)* are immediate from Lemma 3.1.1., while *(iii)* and *(iv)* follow from Rohatgi (1976, p 303).

Note $S_m^2 = \left(m-1\right)^{-1}\left(\sum_{i=1}^{m} Z_i^2 - m\overline{Z}_m^{\,2}\right)$ and $\left(\sum_{i=1}^{m} Z_i\right)^2 = \sum_{i=1}^{m} Z_i^2 + \sum_{i \neq j}^{m}\sum Z_i\, Z_j$.

The proof of (*v)*, (*vi)*, *(vii)* and (*viii*) follows by taking the expectation over the identities:

$$\overline{Z}_m\, S_m^2 = \left(m\,(m-1)\right)^{-1}\left\{\sum_{i=1}^{m} Z_i^3 + \sum_{i \neq j}^{m}\sum Z_i\, Z_j^2 - m^2\overline{Z}_m^{\,3}\right\},$$

$$\sum_{i=1}^{m} Z_i^2\, S_m^2 = \left(m-1\right)^{-1}\sum_{i=1}^{m} Z_i^2\left(\sum_{i=1}^{m} Z_i^2 - m\overline{Z}_m^{\,2}\right)$$

$$= m^{-1}\left(\sum_{i=1}^{m} Z_i^2\right)^2 - \left(m\left(m-1\right)\right)^{-1}\sum_{i=1}^{m} Z_i^2 \sum_{i \neq j}\sum Z_i Z_j$$

$$= m^{-1}\left(\sum_{i=1}^{m} Z_i^4 + \sum_{i \neq j}\sum Z_i^2 Z_j^2\right) - \left(m\left(m-1\right)\right)^{-1}\sum_{i=1}^{m} Z_i^2 \sum_{i \neq j}\sum Z_i\, Z_j ,$$

$$\sum_{i\neq j}\sum Z_i Z_j S_m^2 = m^{-1}\left[2\sum_{i\neq j}\sum Z_i^3 Z_j + \sum_{i\neq j\neq k}\sum\sum Z_i^2 Z_j Z_k\right]$$

$$-(m(m-1))^{-1}\left[2\sum_{i\neq j}\sum Z_i^2 Z_j^2 + 4\sum_{i\neq j\neq k}\sum\sum Z_i^2 Z_j Z_k + \sum_{i\neq j\neq k\neq l}\sum\sum\sum Z_i Z_j Z_k Z_l\right]$$

and

$$\overline{Z}_m^2 S_m^2 = m^{-2}\left(\left(\sum_{i=1}^{m} Z_i^2 + \sum_{i\neq j}^{m}\sum Z_i Z_j\right)S_m^2\right)$$

$$= m^{-2}\left(\sum_{i=1}^{m} Z_i^2 S_m^2 + \sum_{i\neq j}^{m}\sum Z_i Z_j S_m^2\right).$$

The proof of part (I) follows from part (II) by using the above location shift transformation.

Lemma 3.1.3

Let $X_1, X_2, ..., X_m$ be a random sample of size $m \geq 2$ drawn from the continuous distribution function $F(\,.\,;\mu,\theta)$ with finite mean μ, variance θ, skewness γ and kurtosis β. Let S_m^2 be the sample variance. Then as $m \to \infty$, $\sqrt{m}\left(S_m^2 - \theta\right) \xrightarrow{L} N\left(0, (\beta - 1)\theta^2\right)$.

Proof:

The proof follows immediately from the central limit theorem and Lemma 3.1.2 *(iii)*. See also Serfling (1980) section 2.2 for more results.

3.2 Asymptotic characteristics of the triple sampling sequential procedure for inference

In the following sections 3.2.1 and 3.2.2, we will study the asymptotic characteristics of both the main study phase given by (2.2) and the fine tuning phase given by (2.3) under the assumption set forward by Hall (1981) given by (2.4) and our assumption that $E|X_1|^6 < \infty$.

Definition 3.2.1

A sequence of random variables $\{Y_n, n \geq 1\}$ is defined to be uniformly integrable if $\limsup E\left\{|Y_n| I\left(|Y_n| > c\right)\right\} = 0$ as $c \to \infty$, where $I(\,\cdot\,)$ is an indicator function of $(\,\cdot\,)$; see Serfling (1980), page 13 for more details. In other words, a sequence of random variables is defined to be uniformly integrable if it is dominated by some integrable random variable.

Lemma 3.2.1 is found in Chow and Yu (1981) while Lemma 3.2.2 is a uniform integrability result obtained from Lemmas 2, 4 and 5 of Chow and Yu (1981) and are necessary to establish our proofs regarding the asymptotic characteristics of N_1 and N.

Lemma 3.2.1

Let $\{T_a; 0<a<1\}$ be a family of random variables such that $P(T_a \geq 1)=1$. If for some $\delta_1 \in (0,1)$, $p>0,\ q>0$ we have $P(a^q T_a < \delta_1) = o(a^{pq})$ as $a \to 0$, then the family of random variables $\left\{(a^q T_a)^{-p}\right\}$ is uniformly integrable.

Proof: See Lemma 1 in Chow and Yu (1981).

Lemma 3.2.2

Let $Z_1, Z_2, \ldots$ be independent and identically distributed random variables, each with mean zero. Let G_n be $\sigma-$field generated by $\{Z_1, Z_2, \ldots, Z_n\}$. Let $\{T_\lambda; \lambda \in B\}$ be G_n stopping times with $B \subset (0,\infty)$ Then

(i) $E|Z_1|^{2p} < \infty$, $p \geq 1 \Rightarrow \left\{(T_\lambda/\lambda)^p\right\}$ is uniformly integrable.

(ii) $E(Z_1^2) < \infty \Rightarrow \left\{(T_\lambda/\lambda)^{-p}\right\}$ is uniformly integrable for all $p>0$.

(iii) $E|Z_1|^{2p} < \infty$, $p \geq 1 \Rightarrow \left\{\left(\lambda^{-1}\sum_{i=1}^{T_\lambda} Z_i\right)^{2p}\right\}$ is uniformly integrable.

(iv) $E|Z_1|^{2p} < \infty$, $p \geq 1 \Rightarrow \left\{\left(\lambda^{-1/2}\sum_{i=1}^{T_\lambda} Z_i\right)^{2p}\right\}$ is uniformly integrable.

(iv) $E|Z_1|^{2p} < \infty$, $p \geq 2 \Rightarrow \left\{\left|\lambda^{-1/2}\left(\sum_{i=1}^{T_\lambda} Z_i^2 - T_\lambda\right)\right|^p\right\}$ is uniformly integrable.

Proof: See Lemmas 2, 4 and 5 of Chow and Yu (1981).

Theorem 3.2.1 (Anscombe's Theorem)

Let $Z_1, Z_2, \ldots$ be independent and identically distributed random variables with mean zero and finite variance θ. Let $\nu(t)$ denote a positive integer-valued random variable for any $t>0$ such that $\nu(t)/t \to c > 0$ in probability as $t \to +\infty$. Then $(\theta\,\nu(t))^{-1/2}\sum_{i=1}^{\nu(t)} Z_i$ converges in distribution to a standard normal distribution as $t \to +\infty$. See Renyi (1957) for the proof of the Theorem.

3.2.1 Asymptotic characteristics of the main study phase

The asymptotic characteristics of the main study phase of the triple sampling procedure are discussed through the following Theorems. Theorem 3.2.1.1 provides results regarding the asymptotic characteristics of the main study phase. Specifically, second order approximations of the expectation and the variance of the second stage sample mean are given as the initial sample size m gets large. First we introduce some Lemmas that help us to construct Theorems about N_1. Lemma 3.2.1.1 shows the uniform integrability for positive and negative powers of N_1, Lemma 3.2.1.2 shows the asymptotic distribution of N_1, Lemma 3.2.1.3 shows the uniform integrability and the asymptotic distribution of $S_{N_1}^2$ and Lemma 3.2.1.4 gives some useful results that simplify our proof of Theorem 3.2.1.1.

Lemma 3.2.1.1

For the triple sampling rule given by (2.2) we have

(i) The set $\left\{ \left(\lambda / N_1 \right)^p \right\}$ is uniformly integrable for every $p > 0$.

(ii) The set $\left\{ \left(N_1 / \lambda \right)^p \right\}$ is uniformly integrable for every $p > 0$.

Proof: Part *(i)* follows directly from Lemma 3.2.2 *(ii)*. It also follows from the fact that $\delta\lambda g\left(S_m^2\right) \le N_1$, by making use of Lemmas 3 and 4 of Chow and Yu (1981) and Lemma 3.2.1we obtain the result. Part *(ii)* follows directly from Lemma 3.2.2 *(i)*. It also follows from (2.2) and the fact that $\left(N_1/\lambda\right)^p \approx \left(\delta g\left(S_m^2\right)\right)^p < \infty$ where g is a bounded function.

Lemma 3.2.1.2

For the triple sampling rule given by (2.2) and if condition (2.4) holds, then as $\lambda \to \infty$,

$$N_1 / n^* \xrightarrow{a.s} \delta.$$

Proof:

The proof follows directly from (2.4) and the strong law of large numbers.

Lemma 3.2.1.3

Let $X_1, X_2, \dots$ be independent and identically distributed random variables from the continuous distribution function $F\left(. ; \mu, \theta\right)$ such that $E\left|X_1\right|^6 < \infty$. Then

(i) The set $\left\{ \left| \lambda^{1/2} \left(S_{N_1}^2 - \theta\right)\right|^p \right\}$ is uniformly integrable for $0 < p \le 3$.

(ii) $\sqrt{N_1}\left(S_{N_1}^2-\theta\right)\xrightarrow{L} N\left(0,\theta^2\left(\beta-1\right)\right)$ as $\lambda\to\infty$.

Proof:

Part *(i)* can be proved as follows

Recall $S_{N_1}^2=\left(N_1-1\right)^{-1}\sum_{i=1}^{N_1}\left(X_i-\bar{X}_{N_1}\right)^2$. Using the transformation $Z_i=X_i-\mu$, we have $S_{N_1}^2=\left(N_1-1\right)^{-1}\left(\sum_{i=1}^{N_1}Z_i^2-N_1\bar{Z}_{N_1}^2\right)$. Hence we can write $\lambda^{1/2}\left(S_{N_1}^2-\theta\right)$ as follows

$$\lambda^{1/2}\left(S_{N_1}^2-\theta\right)=\frac{\lambda}{N_1-1}\left\{\sum_{i=1}^{N_1}\left(Z_i^2-\theta\right)+\theta-N_1^{-1}\left(\sum_{i=1}^{N_1}Z_i\right)^2\right\}\frac{1}{\sqrt{\lambda}}.$$

By making use of Lemmas 3.2.1.1and 3.2.2 parts *(iii)*,*(iv)* and *(v)* we obtain the result. Hence the proof is complete.

Part *(ii)* Recall that $N_1/\delta n^*\to 1$ almost surely as $\lambda\to\infty$ (Lemma 3.2.1.2 *(i)*). Then the result follows by Anscombe's Theorem.

Lemma 3.2.1.4

Let $Z_1,Z_2,...,Z_m$ be a random sample of size $m\geq 2$ drawn from the continuous distribution function $F\left(\,.\,;\mu,\theta\right)$ with mean zero and $E\left|Z_1\right|^6<\infty$. Let g be a positive, twice continuously differentiable function such that $g,g^{'}$ and $g^{''}$ are bounded. Then for the triple sampling rule $\left(2.2\right)$ and condition $\left(2.4\right)$ as $\lambda\to\infty$ we have

$$(i)\ E\left(\frac{1}{N_1}\sum_{i=1}^{m}Z_i\right)=-\frac{\gamma\theta^{3/2}}{\delta n^*}\frac{d}{d\theta}\ln\left(g\left(\theta\right)\right)+o\left(\lambda^{-1}\right).$$

$$(ii)\ E\left(\frac{1}{N_1^2}\sum_{i=1}^{m}Z_i^2\right)=\frac{\theta}{\delta n^*}-\frac{2\theta^2\left(\beta-1\right)}{\left(\delta n^*\right)^2}\frac{d}{d\theta}\ln\left(g\left(\theta\right)\right)+o\left(\lambda^{-2}\right).$$

$$(iii)\ E\left(\frac{1}{N_1^2}\sum_{i\neq j}^{m}\sum Z_i\,Z_j\right)=\frac{4\theta^2}{\left(\delta n^*\right)^2}\frac{d}{d\theta}\ln\left(g\left(\theta\right)\right)+o\left(\lambda^{-2}\right).$$

Proof:

The proof of part *(i)* follows by expanding the function $f\left(N_1\right)=N_1^{-1}$ around δn^* where $f\left(\,\cdot\,\right)$ is a continuously differentiable function. Note that $f^{'}$, $f^{''}$ and $f^{'''}$ are continuous over $\left[N_1,\delta n^*\right]$. Then

$$N_1^{-1}=\left(\delta n^*\right)^{-1}-\left(\delta n^*\right)^{-2}\left(N_1-\delta n^*\right)+\left(\delta n^*\right)^{-3}\left(N_1-\delta n^*\right)^2+\frac{1}{6}f'''(\eta)\left(N_1-\delta n^*\right)^3,$$

where η is a positive random variable that lies between N_1 and δn^*. By collecting the above terms we have

$$(3.1)\quad N_1^{-1}=3\left(\delta n^*\right)^{-1}-3\left(\delta n^*\right)^{-2}\left(N_1\right)+\left(\delta n^*\right)^{-3}\left(N_1\right)^2+\frac{1}{6}f'''(\eta)\left(N_1-\delta n^*\right)^3.$$

Multiplying (3.1) by $\sum_{i=1}^{m}Z_i$ and substituting for $N_1=\delta\,\lambda\,g\left(S_m^2\right)$ a.s. as $\lambda\to\infty,$ we have

$$(3.2)\quad N_1^{-1}\sum_{i=1}^{m}Z_i=3\sum_{i=1}^{m}Z_i\left(\delta n^*\right)^{-1}-3\sum_{i=1}^{m}Z_i\left(\delta n^*\right)^{-2}\left(\delta\lambda g\left(S_m^2\right)\right)+\sum_{i=1}^{m}Z_i\left(\delta n^*\right)^{-3}\left(\delta\lambda g\left(S_m^2\right)\right)^2$$
$$+\frac{1}{6}\sum_{i=1}^{m}Z_i f'''(\eta)\left(N_1-\delta n^*\right)^3.$$

Since g is a bounded twice differentiable function, then g^2 is also a bounded function. Hence by expanding $g\left(S_m^2\right)$ and $g^2\left(S_m^2\right)$ around θ we have

$$(3.3)\quad N_1^{-1}\sum_{i=1}^{m}Z_i=3\sum_{i=1}^{m}Z_i\left(\delta n^*\right)^{-1}$$
$$-3\sum_{i=1}^{m}Z_i\left(\delta n^*\right)^{-2}\left(\delta\lambda\left\{g(\theta)+g'(\theta)\left(S_m^2-\theta\right)+\frac{1}{2}g''(\eta_1)\left(S_m^2-\theta\right)^2\right\}\right)$$
$$+\sum_{i=1}^{m}Z_i\left(\delta n^*\right)^{-3}(\delta\lambda)^2\left\{g^2(\theta)+g^{2'}(\theta)\left(S_m^2-\theta\right)+\frac{1}{2}g^{2''}(\eta_2)\left(S_m^2-\theta\right)^2\right\}$$
$$+\frac{1}{6}\sum_{i=1}^{m}Z_i f'''(\eta)\left(N_1-\delta n^*\right)^3,$$

where η_1 and η_2 are random variables that lie between S_m^2 and θ. Taking the expectation over (3.3) and making use of the identity $E\left\{\left(S_m^2-\theta\right)\sum_{i=1}^{m}Z_i\right\}=\gamma\,\theta^{3/2}$ we have

$$(3.4)\quad E\left(N_1^{-1}\sum_{i=1}^{m}Z_i\right)=-3\left(\delta n^*\right)^{-2}\delta\,\lambda g'(\theta)\gamma\,\theta^{3/2}+\left(\delta n^*\right)^{-3}(\delta\lambda)^2 g^{2'}(\theta)\gamma\,\theta^{3/2}$$
$$-(3/2)\delta\lambda\left(\delta n^*\right)^{-2}E\left\{g''(\eta_1)\sum_{i=1}^{m}Z_i\left(S_m^2-\theta\right)^2\right\}+(1/2)(\delta\,\lambda)^2\left(\delta n^*\right)^{-3}E\left\{g^{2''}(\eta_2)\sum_{i=1}^{m}Z_i\left(S_m^2-\theta\right)^2\right\}$$
$$+E\left\{\frac{1}{6}\sum_{i=1}^{m}Z_i f'''(\eta)\left(N_1-\delta n^*\right)^3\right\}.$$

To show that the errors vanish as $\lambda\to\infty$ we proceed as follows. From Jensen's inequality we have

$$(3.5)\quad \left|E\left(R_1\right)\right| = \left|E\left\{\frac{1}{6}\sum_{i=1}^{m} Z_i f'''(\eta)\left(N_1-\delta n^*\right)^3\right\}\right| \le E\left\{\eta^{-4}\left|\sum_{i=1}^{m} Z_i\right|\left|N_1-\delta n^*\right|^3\right\},$$

where η is a random variable between N_1 and δn^*. Consider the case $\eta > \delta n^*$. Then

$$\left|E\left\{\frac{1}{6}\sum_{i=1}^{m} Z_i f'''(\eta)\left(N_1-\delta n^*\right)^3\right\}\right| \le E\left\{\eta^{-4}\left|\sum_{i=1}^{m} Z_i\right|\left|N_1-\delta n^*\right|^3\right\} \le \left(\delta n^*\right)^{-4} E\left\{\left|\sum_{i=1}^{m} Z_i\right|\left|N_1-\delta n^*\right|^3\right\}$$
$$\le \left(\delta n^*\right)^{-4} E\left\{\left|\sum_{i=1}^{m} Z_i\right| N_1^3\right\} = \left(\delta n^*\right)^{-4}\left(\delta\lambda\right)^3 E\left\{g^3\left(S_m^2\right)\left|\sum_{i=1}^{m} Z_i\right|\right\}$$

The last identity follows from the fact that $\frac{N_1}{\delta n^*}-1 \le \frac{N_1}{\delta n^*}$ for large values N_1, while as $\lambda \to \infty$ we have $\left|\frac{N_1}{\delta n^*}-1\right| < \left|\frac{N_1}{\delta n^*}\right|$, such inequality can be justified from Lemma 3.2.1.2 *(i)* where $N_1/\delta n^* \to 1$ almost surely as $\lambda \to \infty$. Since g is bounded there exists a generic constant M independent of $N_1, \delta n^*, \eta$ and m such that $\left| g^3 \right| \le M$. Hence multiplying by $g^3(\theta)$ in the numerator and denominator we obtain

$$\left|E\left(R_1\right)\right| \le M\left(\delta n^*\right)^{-1} g^{-3}(\theta) E\left\{\left|\sum_{i=1}^{m} Z_i\right|\right\},$$

but from central limit theorem $\left(m\theta\right)^{-1/2}\sum_{i=1}^{m} Z_i \to N(0,1)$ in distribution as $m \to \infty$ and since the quantity $\left\{\left|m^{-1/2}\sum_{i=1}^{m} Z_i\right|\right\}$ is uniformly integrable, then $E\left\{\left(m\theta\right)^{-1/2}\left|\sum_{i=1}^{m} Z_i\right|\right\} = \sqrt{2/\pi},\ m \to \infty$ which implies that $E\left\{m^{-1/2}\left|\sum_{i=1}^{m} Z_i\right|\right\} = \sqrt{2\theta/\pi},\ m \to \infty.$

Hence $\left|E\left(R_1\right)\right| \le M\left(\delta n^*\right)^{-1} g^{-3}(\theta) m^{1/2}\sqrt{2\theta/\pi} < M_1\left(\delta n^*\right)^{-1/2} \to 0,\ \lambda \to \infty.$

Note that from condition (2.4) we have $m/n^* \approx \delta$ as $m \to \infty$.

Consider the case $\eta > N_1$ and the fact that $N_1 \ge m$ then

$$\left|E\left\{\frac{1}{6}\sum_{i=1}^{m} Z_i f'''(\eta)\left(N_1-\delta n^*\right)^3\right\}\right| \le E\left\{\eta^{-4}\left|\sum_{i=1}^{m} Z_i\right|\left|N_1-\delta n^*\right|^3\right\} \le E\left\{N_1^{-4}\left|\sum_{i=1}^{m} Z_i\right|\left|N_1-\delta n^*\right|^3\right\}$$
$$\le E\left\{N_1^{-1}\left|\sum_{i=1}^{m} Z_i\right|\right\} \le E\left\{m^{-1}\left|\sum_{i=1}^{m} Z_i\right|\right\} = E\left\{\left|\overline{Z}_m\right|\right\}$$

but from the strong law of large numbers and using Theorem A.3.4 in Mukhopadhyay and de Silva (2009) page 442 and the fact that $\left\{\left|\overline{Z}_m\right|\right\}$ is uniformly integrable, we have $E\left|\overline{Z}_m\right| \to 0$ as $m \to \infty$. Also

$$(3.6) \quad \left|E\left(R_2\right)\right| = \delta\lambda\left(\delta n^*\right)^{-2}\left|E\left\{g''\left(\eta_1\right)\sum_{i=1}^{m} Z_i\left(S_m^2-\theta\right)^2\right\}\right|,$$

where η_1 is a random variable between S_m^2 and θ. Since $g''\left(\eta_1\right)$ is bounded there exists a generic constant K_1 independent of S_m^2, θ, m and η_1 such that $\left|g''\left(\eta_1\right)\right| \le K_1$. Then

$$\begin{aligned}\left|E\left(R_2\right)\right| &= \delta\lambda\left(\delta n^*\right)^{-2}\left|E\left\{g''\left(\eta_1\right)\sum_{i=1}^{m} Z_i\left(S_m^2-\theta\right)^2\right\}\right| \\ &\le K_1\left(\delta n^*\right)^{-1} g^{-1}\left(\theta\right) E\left\{\left|\sum_{i=1}^{m} Z_i\right|\left(S_m^2-\theta\right)^2\right\}.\end{aligned}$$

By using lemma 3.2.1.1 *(ii)* the quantity $\left\{\theta^2\left(\beta-1\right)\right\}^{-1} m\left(S_m^2-\theta\right)^2 \to \chi^2\left(1\right)$ in distribution as $\lambda \to \infty$. Moreover, from the strong law of large numbers and Theorem A.3.4 (as mentioned above), we have $\left|m^{-1}\sum_{i=1}^{m} Z_i\right| \to 0$ a.s. as $m \to \infty$. Hence it follows from Slutsky's Theorem that $\left(\theta^2\left(\beta-1\right)\right)^{-1}\left|\sum_{i=1}^{m} Z_i\right|\left(S_m^2-\theta\right)^2 \to 0$ in distribution as $m \to \infty$, but since $\left\{\left|\sum_{i=1}^{m} Z_i\left(S_m^2-\theta\right)^2\right|\right\}$ is uniformly integrable, then $E\left\{\left(\theta^2\left(\beta-1\right)\right)^{-1}\left|\sum_{i=1}^{m} Z_i\right|\left(S_m^2-\theta\right)^2\right\} \to 0$ as $m \to \infty$.

A similar procedure can be used to prove that

$$\begin{aligned}(3.7) \quad \left|E\left(R_3\right)\right| &= \left(\delta\lambda\right)^2\left(\delta n^*\right)^{-3}\left|E\left\{g^{2''}\left(\eta_2\right)\sum_{i=1}^{m} Z_i\left(S_m^2-\theta\right)^2\right\}\right| \\ &\le K_2\left(\delta n^*\right)^{-1} g^{-2}\left(\theta\right) E\left\{\sum_{i=1}^{m} Z_i\left(S_m^2-\theta\right)^2\right\} \to 0,\ \lambda \to \infty,\end{aligned}$$

where K_2 is a generic constant independent of S_m^2, θ, m and η_2 such that $\left|g^{2''}\left(\eta_2\right)\right| \le K_2$. Substituting (3.5), (3.6) and (3.7) in (3.4) we have

$$(3.8)\quad E\left(N_1^{-1}\sum_{i=1}^{m} Z_i\right) = -3\left(\delta n^*\right)^{-2}\delta\lambda g^{'}(\theta)\gamma\theta^{3/2} + \left(\delta n^*\right)^{-3}(\delta\lambda)^2 g^{2'}(\theta)\gamma\theta^{3/2} + o\left(\lambda^{-1}\right)$$

$$= -3\left(\delta n^*\right)^{-1}\frac{g^{'}(\theta)}{g(\theta)}\gamma\theta^{3/2} + \left(\delta n^*\right)^{-1}\frac{g^{2'}(\theta)}{g^2(\theta)}\gamma\theta^{3/2} + o\left(\lambda^{-1}\right)$$

$$= -\gamma\theta^{3/2}\left(\delta n^*\right)^{-1}\frac{d}{d\theta}\ln g(\theta) + o\left(\lambda^{-1}\right).$$

The proof is now complete.

Part *(ii)* follows by expanding the function $f(N_1) = N_1^{-2}$ around δn^*. Then by collecting the expansion terms we have

$$(3.9)\quad N_1^{-2} = 6\left(\delta n^*\right)^{-2} - 8\left(\delta n^*\right)^{-3}(N_1) + 3\left(\delta n^*\right)^{-4}(N_1)^2 + \frac{1}{6}f^{'''}(\eta)\left(N_1 - \delta n^*\right)^3,$$

where η is a random variable that lies between N_1 and δn^*.

Substituting for N_1 and expanding $g\left(S_m^2\right)$ and $g^2\left(S_m^2\right)$ around θ and entering the expectation over the terms we obtain

$$(3.10)\quad E\left(N_1^{-2}\sum_{i=1}^{m} Z_i^2\right) = \frac{\theta}{\delta n^*} - \frac{2\theta^2(\beta-1)}{\left(\delta n^*\right)^2}\frac{d}{d\theta}\ln\left(g(\theta)\right)$$

$$+ E\left\{\frac{1}{6}\sum_{i=1}^{m} Z_i^2\, f^{'''}(\eta)\left(N_1 - \delta n^*\right)^3\right\} - 4\delta\lambda\left(\delta n^*\right)^{-3} E\left\{g^{''}(\eta_1)\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\}$$

$$+ (3/2)(\delta\lambda)^2\left(\delta n^*\right)^{-4} E\left\{g^{2''}(\eta_2)\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\}.$$

But from Jensen's inequality and the fact $\left|N_1 - \delta n^*\right| \le N_1$ we have

$$\left|E(R_1)\right| = \left|E\left\{\frac{1}{6}\sum_{i=1}^{m} Z_i^2\, f^{'''}(\eta)\left(N_1 - \delta n^*\right)^3\right\}\right| \le E\left\{\eta^{-5}\sum_{i=1}^{m} Z_i^2\,\left|N_1 - \delta n^*\right|^3\right\} \le E\left\{\eta^{-5}\sum_{i=1}^{m} Z_i^2\, N_1^3\right\}.$$

Since g is bounded, there exists a generic constant M independent of $N_1, \delta n^*, \eta$ and m such that $\left|g^3\right| \le M$. Hence $\left|E(R_1)\right| \le E\left\{\eta^{-5}\sum_{i=1}^{m} Z_i^2\, N_1^3\right\} \le M(\delta\lambda)^3 E\left\{\eta^{-5}\sum_{i=1}^{m} Z_i^2\right\}.$

Consider the case $\eta > \delta n^*$. Then multiplying by $g^3(\theta)$ in the numerator and denominator and using the fact that as $m \to \infty,\ m \approx \delta n^*$ we obtain

$$\left|E\left(R_1\right)\right| \le M\left(\delta\lambda\right)^3\left(\delta n^*\right)^{-5} E\left\{\sum_{i=1}^{m} Z_i^2\right\},$$

$$= M\left(\delta n^*\right)^{-2} g^{-3}\left(\theta\right) m\theta < M\theta\left(\delta n^*\right)^{-1} g^{-3}\left(\theta\right) \to 0, \quad \lambda \to \infty.$$

Consider the case $\eta > N_1$ and the fact $N_1 \ge m$. Then we have

$$\left|E\left(R_1\right)\right| \le E\left\{N_1^{-5}\left(\sum_{i=1}^{m} Z_i^2 N_1^3\right)\right\} = E\left\{\sum_{i=1}^{m} Z_i^2 N_1^{-2}\right\} \le E\left\{m^{-2}\sum_{i=1}^{m} Z_i^2\right\} = m^{-1}\theta \to 0, \ m \to \infty.$$

Also

$$\left|E\left(R_2\right)\right| = 4\delta\lambda\left(\delta n^*\right)^{-3}\left|E\left\{g''\left(\eta_1\right)\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\}\right|$$

$$\le K_1\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-2} E\left\{\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\} = o\left(\lambda^{-2}\right), \quad \lambda \to \infty$$

as illustrated before as $\lambda \to \infty$ the quantity $\left\{\theta^2\left(\beta - 1\right)\right\}^{-1} m\left(S_m^2 - \theta\right)^2 \to \chi^2\left(1\right)$

and $m^{-1}\sum_{i=1}^{m} Z_i^2 \to \theta$ in probability as $m \to \infty$. Thus, from Slutsky's Theorem

$\left(S_m^2 - \theta\right)^2 \sum_{i=1}^{m} Z_i^2 \to \theta^3\left(\beta - 1\right)\chi_1^2$ in distributions as $m \to \infty.$ since the quantity $\left\{\left(S_m^2 - \theta\right)^2 \sum_{i=1}^{m} Z_i^2\right\}$

is uniformly integrable, then $E\left\{\left(S_m^2 - \theta\right)^2 \sum_{i=1}^{m} Z_i^2\right\} = \theta^3\left(\beta - 1\right), m \to \infty.$

Hence it follows that

$$\left|E\left(R_2\right)\right| \le K_1\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-2} E\left\{\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\}$$

$$= K_1\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-2} \theta^3\left(\beta - 1\right) \to 0, \quad \lambda \to \infty.$$

A similar procedure can be used to show that

$$\left|E\left(R_3\right)\right| = (3/2)\left(\delta\lambda\right)^2\left(\delta n^*\right)^{-4}\left|E\left\{g^{2''}\left(\eta_2\right)\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\}\right|$$

$$\le K_2\left(\delta n^*\right)^{-2} E\left\{\sum_{i=1}^{m} Z_i^2\left(S_m^2 - \theta\right)^2\right\} \to 0, \quad \lambda \to \infty.$$

Hence the proof is complete.

Part *(iii)*

The proof of part *(iii)* follows as for part *(ii)* except putting $\sum_{i\neq j}^{m}\sum Z_iZ_j$ instead of $\sum_{i=1}^{m} Z_i^2$. Then we have the following

$$(3.11)\quad E\left(N_1^{-2}\sum_{i=1}^{m}\sum Z_iZ_j\right)=\frac{4\theta^2}{\left(\delta n^*\right)^2}\frac{d}{d\theta}\ln\left(g\left(\theta\right)\right)+E\left\{\frac{1}{6}\sum_{i=1}^{m}\sum Z_iZ_j f'''\left(\eta\right)\left(N_1-\delta n^*\right)^3\right\}$$

$$-4\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-2}E\left\{g''\left(\eta_1\right)\sum_{i=1}^{m}\sum Z_iZ_j\left(S_m^2-\theta\right)^2\right\}$$

$$+\left(3/2\right)\left(g\left(\theta\right)\right)^{-2}\left(\delta n^*\right)^{-2}E\left\{g^{2''}\left(\eta_2\right)\sum_{i=1}^{m}\sum Z_iZ_j\left(S_m^2-\theta\right)^2\right\}.$$

But

$$\left|E\left(R_1\right)\right|=\left|E\left\{\eta^{-5}\sum_{i\neq j=1}^{m}\sum Z_iZ_j\left(N_1-\delta n^*\right)^3\right\}\right|\leq E\left\{\eta^{-5}\left|\sum_{i\neq j=1}^{m}\sum Z_iZ_j\right|\left|N_1-\delta n^*\right|^3\right\},$$

$$\eta>\delta n^*\Rightarrow E\left\{\eta^{-5}\left|\sum_{i\neq j=1}^{m}\sum Z_iZ_j\right|\left|N_1-\delta n^*\right|^3\right\}\leq E\left\{\eta^{-5}\left|\sum_{i\neq j=1}^{m}\sum Z_iZ_j\right|\left|N_1\right|^3\right\}$$

$$\leq M\left(\delta n^*\right)^{-5}\left(\delta\lambda\right)^3E\left\{\left|\sum_{i\neq j=1}^{m}\sum Z_iZ_j\right|\right\}=M\left(\delta n^*\right)^{-2}g^{-3}\left(\theta\right)E\left\{\left|\sum_{i\neq j=1}^{m}\sum Z_iZ_j\right|\right\}.$$

To find $E\left\{\left|\sum_{i\neq j=1}^{m}\sum Z_iZ_j\right|\right\}$ we proceed as follows. Let $U=\left(m\left(m-1\right)\right)^{-1}\sum_{i\neq j=1}^{m}\sum Z_iZ_j$. Then it can be shown that $\frac{mU}{\theta}\to\chi_1^2-1$ in distribution as $m\to\infty$; see Theorem 5.5.1A in Serfling (1980) page 192. This implies that $\left(\theta\left(m-1\right)\right)^{-1}\sum_{i\neq j=1}^{m}\sum Z_iZ_j\to\chi_1^2-1$ as $m\to\infty$. If we set $V\sim\chi_1^2$, then by making a linear transformation between U and V , the probability density function of the random variable U , $h\left(u\right)=\frac{\exp\left(-\left(u+1\right)/2\right)}{\sqrt{2\pi\left(u+1\right)}}$, $-1<u<\infty$.

By integration we have $E\left|U\right|=E\left|\left(\theta\left(m-1\right)\right)^{-1}\sum_{i\neq j}^{m}\sum Z_iZ_j\right|=\frac{4}{\sqrt{2\pi e}}$, $m\to\infty$ which implies that

$$E\left|\left(m-1\right)^{-1/2}\sum_{i\neq j}^{m}\sum Z_iZ_j\right|=\frac{4\theta}{\sqrt{2\pi e}}\text{ as }m\to\infty.$$

Thus

$$E\left|R_1\right| \le M\left(\delta n^*\right)^{-2} g^{-3}(\theta)\frac{4\theta}{\sqrt{2\pi e}}(m-1)$$
$$< M\left(\delta n^*\right)^{-1} g^{-3}(\theta)\frac{4\theta}{\sqrt{2\pi e}} \to 0,\ \lambda \to \infty.$$

Also

$$\eta > N_1 \Rightarrow E\left\{N_1^{-5}\sum_{i\ne j}^{m}\sum Z_i Z_j \left|N_1-\delta n^*\right|^3\right\} \le E\left\{N_1^{-5}\sum_{i\ne j}^{m}\sum Z_i Z_j \left|N_1\right|^3\right\}$$
$$= E\left\{N_1^{-2}\sum_{i\ne j}^{m}\sum Z_i Z_j\right\} \le E\left\{m^{-2}\sum_{i\ne j}^{m}\sum Z_i Z_j\right\} = \frac{(m-1)}{m^2}\frac{4\theta}{\sqrt{2\pi e}} \to 0,\ m \to \infty.$$

For the other error term, we have

$$\left|E\left(R_2\right)\right| = \left(g(\theta)\right)^{-1}\left(\delta n^*\right)^{-2}\left|E\left\{g''\left(\eta_1\right)\sum_{i=1}^{m}\sum Z_i Z_j\left(S_m^2-\theta\right)^2\right\}\right|,$$
$$\le K_1\left(g(\theta)\right)^{-1}\left(\delta n^*\right)^{-2} E\left\{\left|\sum_{i=1}^{m}\sum Z_i Z_j\right|\left(S_m^2-\theta\right)^2\right\},$$

but $\left(m(m-1)\right)^{-1}\sum_{i\ne j}^{m}\sum Z_i Z_j \to 0$ in probability and $\frac{m\left(S_m^2-\theta\right)^2}{\theta^2(\beta-1)} \to \chi^2(1)$ in distribution, as $\lambda \to \infty$. Thus $E\left\{(m-1)^{-1}\left|\sum_{i=1}^{m}\sum Z_i Z_j\right|\left(S_m^2-\theta\right)^2\right\} \to 0$ in distribution, as $\lambda \to \infty$ by Slutsky's Theorem, provided that $\left\{(m-1)^{-1}\left|\sum_{i=1}^{m}\sum Z_i Z_j\left(S_m^2-\theta\right)^2\right|\right\}$ is uniformly integrable. Hence $E\left|R_2\right| \to 0$ as $\lambda \to \infty$. A similar procedure can be used to show that

$$\left|E\left(R_3\right)\right| = (3/2)\,g^{-2}(\theta)\left(\delta n^*\right)^{-2}\left|E\left\{g^{2''}\left(\eta_2\right)\sum_{i=1}^{m}\sum Z_i Z_j\left(S_m^2-\theta\right)^2\right\}\right|$$
$$\le K_2\ g^{-2}(\theta)\left(\delta n^*\right)^{-2} E\left\{\left|\sum_{i=1}^{m}\sum Z_i Z_j\right|\left(S_m^2-\theta\right)^2\right\} \to 0,\ \lambda \to \infty.$$

The proof is now complete.

Theorem 3.2.1.1

Let g be a positive, twice continuously differentiable function such that g, g' and g'' are bounded. Then for the triple sampling rule (2.2) and condition (2.4) as $\lambda \to \infty$, we have

$$i)\, E\left(\bar{X}_{N_1}\right) = \mu - \gamma\,\theta^{3/2}\left(d\ln g(\theta)/d\theta\right)\left(\delta n^*\right)^{-1} + o\left(\lambda^{-1}\right).$$

$$ii)\, Var\left(\bar{X}_{N_1}\right) = \theta\left(\delta n^*\right)^{-1} - 2\theta^2(\beta-3)\left(d\ln g(\theta)/d\theta\right)\left(\delta n^*\right)^{-2} + o\left(\lambda^{-2}\right).$$

Proof:

To prove *(i)*, consider the transformation $Z = X - \mu$, and we may write

$$(3.12)\quad E\left(\bar{X}_{N_1} - \mu\right) = E\left(N_1^{-1}\sum_{i=1}^{N_1} Z_i\right) = E\left\{E\left(N_1^{-1}\sum_{i=1}^{N_1} Z_i \,|N_1\right)\right\}.$$

Then, conditioning on the $\sigma-$field generated by the random variables $Z_1, Z_2, ..., Z_m$, we have

$$E\left(\bar{X}_{N_1} - \mu\right) = E\left\{N_1^{-1}\, E\left(\sum_{i=1}^{m} Z_i + \sum_{i=m+1}^{N_1} Z_i \,|\, Z_1, Z_2, ..., Z_m\right)\right\}.$$

Given $Z_1, Z_2, ..., Z_m$, the first sum is non-random and the second has expectation zero. Hence

$$(3.13)\quad E\left(\bar{X}_{N_1} - \mu\right) = E\left\{N_1^{-1}\sum_{i=1}^{m} Z_i\right\}.$$

But from Lemma 3.2.1.3 part *(i)* $E\left(\frac{1}{N_1}\sum_{i=1}^{m} Z_i\right) = -\frac{\gamma\,\theta^{3/2}}{\delta n^*}\frac{d}{d\theta}\ln\left(g(\theta)\right) + o\left(\lambda^{-1}\right).$

The remainder term is of order $o\left(\lambda^{-1}\right)$. By substituting this in (3.13), *(i)* follows and hence the proof of *(i)* is complete.

To prove *(ii)*, we also write

$$(3.14)\quad E\left(\bar{X}_{N_1} - \mu\right)^2 = E\left(E\left[N_1^{-2}\left(\sum_{i=1}^{N_1} Z_i^2 + \sum_{i\neq j}^{N_1}\sum Z_i Z_j\right)\middle|Z_1, Z_2, ..., Z_m\right]\right).$$

The first term in (3.14), conditional on the $\sigma-$field generated by the random variables $Z_1, Z_2, ..., Z_m$, can be written as

$$(3.15)\quad E\left(N_1^{-2}\sum_{i=1}^{N_1} Z_i^2\right) = E\left\{N_1^{-2}\, E\left(\sum_{i=1}^{m} Z_i^2 + \sum_{i=m+1}^{N_1} Z_i^2 \,\middle|\, Z_1, Z_2, ..., Z_m\right)\right\}.$$

Therefore,

$$(3.16)\quad E\left(N_1^{-2}\sum_{i=1}^{N_1} Z_i^2\right) = E\left(N_1^{-2}\sum_{i=1}^{m} Z_i^2\right) + \theta E\left(N_1^{-1} - m\, N_1^{-2}\right).$$

The first term on the right hand side of (3.16) is in Lemma 3.2.1.3 part *(ii)*. The second term of (3.16) can be written as

$$E\left\{N_1^{-2}\left(N_1 - m\right)\right\} \le E\left(N_1^{-1}\right) = o\left(\lambda^{-1}\right), \text{ as } \lambda \to \infty.$$

Finally,

$$(3.17)\quad E\left(N_1^{-2}\sum_{i=1}^{N_1} Z_i^2\right) = \theta\left(\delta n^*\right)^{-1} - 2\theta^2\left(\beta - 1\right)\left(\delta n^*\right)^{-2}\left(d \ln g\left(\theta\right)/d\theta\right) + o\left(\lambda^{-1}\right).$$

Similarly, the second term of (3.14) conditioned on the $\sigma-$ field generated by the random variables $Z_1, Z_2, ..., Z_m$, can be written as

$$(3.18)\quad E\left(N_1^{-2}\sum_{i\ne j}^{N_1}\sum Z_i Z_j\right) = E\left\{N_1^{-2}\, E\left(\sum_{i\ne j}^{m}\sum Z_i Z_j + \sum_{i\ne j}^{N_1}\sum Z_i Z_j \,\middle|\, Z_1, \ldots, Z_m\right)\right\}.$$

The first term of (3.18) is in Lemma 3.2.1.3 part *(iii)* while the second term tends to zero.

Hence (3.18) leads to

$$(3.19)\quad E\left(N_1^{-2}\sum_{i\ne j}^{N_1}\sum Z_i Z_j\right) = 4\theta^2\left(\delta n^*\right)^{-2}\left(d \ln g\left(\theta\right)/d\theta\right) + o\left(\lambda^{-2}\right),$$

where we have used Lemma 3.1.2 and the assumption that $g\left(\cdot\right)$ and its derivatives are bounded. By adding (3.17) and (3.19) the proof of part *(ii)* is complete.

It is obvious from (i) of Theorem 3.2.1.1 that $\bar{X}_N$ may be a biased estimator of μ. The bias depends on the variance θ and the skewness γ of the underlying distribution, together with the form of the function $g\left(\cdot\right)$, the optimal sample size n^* and the design factor δ. Clearly, if γ and $g^{'}\left(\theta\right)$ have the same/different signs, then the bias is negative/positive. Also, the magnitude of the bias increases as $|\gamma|$ increases. However, as n^* increases the magnitude of the bias decreases and approaches zero as $n^* \to \infty$.

From *(ii)* of Theorem 3.2.1.1, $Var\left(\bar{X}_{N_1}\right)$ depends on the form of $g(\cdot)$, the kurtosis β, the variance of the underlying distribution θ, and the optimal sample size n^*, as well as the design factor δ. If $g'(\theta)$ and $\beta^*(=\beta-3)$ have the same/different signs, then $Var\left(\bar{X}_{N_1}\right)$ is less than/greater than $\theta\left(\delta n^*\right)^{-1}$. Note that if the underlying distribution is normal, then $\bar{X}_{N_1}$ is an unbiased estimator for μ with variance $\theta\left(\delta n^*\right)^{-1}$. So from Theorem 3.2.1.1, the skewness appears in the expression for the mean of the estimator $\bar{X}_{N_1}$, while the kurtosis appears in the expression for the variance of the estimator $\bar{X}_{N_1}$.

The following results, presented in Lemma 3.2.1.5 and Theorem 3.2.1.2, which involve $S^2_{N_1}$, will be useful in the derivation of asymptotic results for the fine tuning phase.

Lemma 3.2.1.5

For the triple sampling rule (2.2), if condition (2.4) holds and $E|Z_1|^6<\infty$, then conditioning on the σ-field generated by $Z_1, Z_2, \ldots, Z_m$, we have

$$(i)\, E\left[\left(S^2_{N_1}-\theta\right)\sum_{i=1}^{N_1}\sum_{i\neq j}^{N_1} Z_i Z_j \,|Z_1, Z_2, \ldots, Z_m\right] = -2N_1^{-2}\left(\sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i^2 Z_j^2 + \theta^2\left(N_1-m\right)\left(N_1-m-1\right)\right).$$

$$(ii)\, E\left[\left(S^2_{N_1}-\theta\right)\sum_{i=1}^{N_1} Z_i^2 \,|Z_1, Z_2, \ldots, Z_m\right] = N_1^{-1}\left(\sum_{i=1}^{m} Z_i^4 + \beta\theta^2\left(N_1-m\right) + \sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i^2 Z_j^2\right)$$
$$+\theta^2 N_1^{-1}\left(N_1^2-2mN_1-N_1+m^2+m\right)-\theta^2 N_1.$$

$$(iii)\, E\left[\left(S^2_{N_1}-\theta\right)\sum_{i=1}^{N_1} Z_i \,|Z_1, Z_2, \ldots, Z_m\right] = N_1^{-1}\left(\sum_{i=1}^{m} Z_i^3 + \sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i^2 Z_j - \mu\sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i Z_j\right)$$
$$+N_1^{-1}\left(\gamma\,\theta^{3/2}\left(N_1-m\right)\right)+\mu\left(\sum_{i=1}^{m} Z_i^2+\theta\left(N_1-m\right)\right).$$

$$(iv)\, E\left[N_1\left(S^2_{N_1}-\theta\right)|Z_1, Z_2, \ldots, Z_m\right] = -N_1^{-1}\sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i Z_j.$$

$$(v)\, E\left[N_1\left(S_{N_1}^2-\theta\right)^2 \middle| Z_1, Z_2, ..., Z_m \right] = \theta^2 N_1 + N_1^{-1}\left(\sum_{i=1}^{m} Z_i^4 + \beta\theta^2\left(N_1-m\right)\right)$$

$$+ N_1^{-1}\left(\sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i^2 Z_j^2 + \theta^2\left(N_1^2-2mN_1-N_1+m^2+m\right)\right)$$

$$+ N_1^{-3}\left(\sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i^2 Z_j^2 + \theta^2\left(N_1-m\right)\left(N_1-m-1\right)\right).$$

Proof:

To prove *(i)*, we use the fact that $S_{N_1}^2 = \left(N_1-1\right)^{-1}\left\{\sum_{i=1}^{N_1} Z_i^2 - N_1 \overline{Z}_{N_1}^2\right\}$.

Then

$$S_{N_1}^2 \sum_{i\neq j}^{N_1}\sum Z_i Z_j = N_1^{-1}\sum_{i=1}^{N_1} Z_i^2 \sum_{i\neq j}^{N_1} Z_i Z_j - \left(N_1\left(N_1-1\right)\right)^{-1}\left(\sum_{i\neq j}^{N_1}\sum Z_i Z_j\right)^2.$$

But $\left(\sum_{i\neq j}^{N_1} Z_i Z_j\right)^2 = 4\left(\sum_{i\neq j}^{N_1}\sum Z_i^2 Z_j^2 + 2\sum_{i\neq j\neq k}^{N_1}\sum\sum Z_i^2 Z_j Z_k\right).$

By conditioning on the σ -field generated by $Z_1, Z_2, ..., Z_m$, we have

$$E\left\{\left(S_{N_1}^2-\theta\right)\sum_{i\neq j}^{N_1}\sum Z_i Z_j \middle| Z_1, Z_2, ..., Z_m\right\} = -2N_1^{-2}\, E\left\{\sum_{i\neq j}^{N_1}\sum Z_i^2 Z_j^2 \middle| Z_1, Z_2, ..., Z_m\right\}$$

$$= -2N_1^{-2}\, E\left\{\sum_{i\neq j}^{m}\sum Z_i^2 Z_j^2 + \sum_{i\neq j=m+1}^{N_1}\sum Z_i^2 Z_j^2 \middle| Z_1, Z_2, ..., Z_m\right\}$$

$$= -2N_1^{-2}\, E\left\{\sum_{i\neq j}^{m}\sum Z_i^2 Z_j^2 + \theta^2\left(N_1-m\right)\left(N_1-m-1\right)\right\}.$$

Similar arguments can be used to prove *(ii)* and *(iii)* of Lemma 3.2.1.5.

To prove *(iv)* we use the identity

$$N_1 S_{N_1}^2 = \sum_{i=1}^{N_1} Z_i^2 - \left(N_1-1\right)^{-1}\sum_{i\neq j}^{N_1}\sum Z_i Z_j\ .$$

Thus

$$E\left\{N_1\left(S_{N_1}^2-\theta\right)|Z_1,Z_2,...,Z_m\right\}=-E\left\{\left(N_1-1\right)^{-1}\left(\sum_{i\neq j=1}^{m}\sum Z_iZ_j+\sum_{i\neq j=m+1}^{N_1}\sum Z_iZ_j\right)|\,Z_1,Z_2,...,Z_m\right\}$$

$$=-\left(N_1-1\right)^{-1}\sum_{i\neq j=1}^{m}\sum Z_iZ_j.$$

Similar arguments and the identity

$$N_1S_{N_1}^4=N_1^{-1}\left(\sum_{i=1}^{N_1}Z_i^4+\sum_{i\neq j}^{N_1}\sum Z_i^2Z_j^2\right)+2N_1^{-3}\sum_{i\neq j}^{N_1}\sum Z_i^2Z_j^2$$

can be used to prove *(v)*.

Theorem 3.2.1.2

For the triple sampling rule (2.2), if condition (2.4) holds and $E|Z_1|^6<\infty$ as $\lambda\to\infty$, we have

$$(i)\,E\left[\left(S_{N_1}^2-\theta\right)\sum_{i=1}^{N_1}\sum_{i\neq j}^{N_1}Z_iZ_j\right]=-2\theta^2+2\theta^2\left(\delta n^*\right)^{-1}+o\left(\lambda^{-1}\right).$$

$$(ii)\,E\left[\left(S_{N_1}^2-\theta\right)\sum_{i=1}^{N_1}Z_i^2\right]=\theta^2\left(\beta-1\right)+o\left(\lambda^{-1}\right).$$

$$(iii)\,E\left[\left(S_{N_1}^2-\theta\right)\sum_{i=1}^{N_1}Z_i\right]=\gamma\,\theta^{3/2}+o\left(\lambda^{-1}\right).$$

$$(iv)\,E\left[N_1\left(S_{N_1}^2-\theta\right)\right]=-2\theta^2\left(\delta n^*\right)^{-1}\left(\frac{d}{d\theta}\ln g\left(\theta\right)\right)+o\left(\lambda^{-1}\right).$$

$$(v)\,E\left[N_1\left(S_{N_1}^2-\theta\right)^2\right]=\theta^2\left(\beta-1\right)+2\theta^2\left(\delta n^*\right)^{-1}+o\left(\lambda^{-1}\right).$$

Proof:

To prove *(i)*, we write

$$E\left[\left(S_{N_1}^2-\theta\right)\sum_{i=1}^{N_1}\sum_{i\neq j}^{N_1}Z_iZ_j\right]=E\left\{E\left[\left(S_{N_1}^2-\theta\right)\sum_{i=1}^{N_1}\sum_{i\neq j}^{N_1}Z_iZ_j\,|Z_1,Z_2,...Z_m\right]\right\}.$$

Hence, from *(i)* of Lemma 3.2.1.5, we have

$$E\left[\left(S_{N_1}^2-\theta\right)\sum_{i=1}^{N_1}\sum_{i\neq j}^{N_1}Z_iZ_j\right]=E\left\{-2N_1^{-2}\left(\sum_{i=1}^{m}\sum_{i\neq j}^{m}Z_i^2Z_j^2+\theta^2\left(N_1-m\right)\left(N_1-m-1\right)\right)\right\}.$$

Consider the expansion of N_1^{-1} and N_1^{-2} around δn^*. The first term leads to

$$E\left\{-2N_1^{-2}\left(\sum_{i=1}^{m}\sum_{i\neq j}^{m} Z_i^2 Z_j^2\right)\right\} = -2\theta^2 + 2\theta^2\left(\delta n^*\right)^{-1} + o\left(\lambda^{-1}\right),$$

and the second term leads to

$$E\left\{-2N_1^{-2}\left(\theta^2\left(N_1 - m\right)\left(N_1 - m - 1\right)\right)\right\} = o\left(\lambda^{-1}\right),$$

where we have used the assumptions in (2.4) and the fact that $g(\cdot)$ and its derivatives are bounded, from which *(i)* follows.

Similar arguments can be used to verify *(ii)* and *(v)* using $E\left[N_1^{-1}\sum_{i=1}^{N_1} Z_i^4\right] = \beta\theta^2 + o\left(\lambda^{-1}\right)$ and

$$E\left[N_1^{-3}\sum_{i\neq j}^{N_1}\sum Z_i^2 Z_j^2\right] = o\left(\lambda^{-1}\right)\cdot$$

Part *(iii)* follows along similar lines and the fact that $E\left[N_1^{-1}\sum_{i=1}^{m} Z_i^3\right] = \gamma\theta^{3/2} + o\left(\lambda^{-1}\right)$.

To prove *(iv)*, recall *(iv)* of Lemma 3.2.1.5 and *(v)*, *(vi)* and *(vii)* of Lemma 3.1.2.

The proof of Theorem 3.2.1.2 is thus complete. We delete details for brevity.

The following Theorem 3.2.1.3 gives asymptotic results for the estimator of θ after the main study phase.

Theorem 3.2.1.3

Let g be a positive, twice continuously differentiable function such that g, g' and g'' are bounded. Then for the triple sampling rule (2.2), if condition (2.4) holds and $E\left|Z_1\right|^6 < \infty$ as $\lambda \to \infty$, we have:

$i)\, E\left(S_{N_1}^2\right) = \theta - \theta^2(\beta - 1)\left(d\ln g(\theta)/d\theta\right)\left(\delta n^*\right)^{-1} + o\left(\lambda^{-1}\right).$

$ii)\, E\left(S_{N_1}^4\right) = \theta^2 + \theta^2(\beta - 1)\left(\delta n^*\right)^{-1} - 2\theta^3(\beta - 1)\left(d\ln g(\theta)/d\theta\right)\left(\delta n^*\right)^{-1} + o\left(\lambda^{-1}\right).$

$iii)\, Var\left(S_{N_1}^2\right) = \theta^2(\beta - 1)\left(\delta n^*\right)^{-1} + o\left(\lambda^{-1}\right).$

$iv)\, E\left(g\left(S_{N_1}^2\right)\right) = g(\theta) - \theta^2(\beta - 1)\left(\lambda\delta n^*\right)^{-1}\left[\left(n^*\right)^{-1}\left(\frac{dn^*}{d\theta}\right)^2 - (1/2)\frac{d^2 n^*}{d\theta^2}\right] + o\left(\lambda^{-1}\right).$

Proof:

To prove *(i)*, we write

$$E\left(S_{N_1}^2\right)=E\left(N_1^{-1}E\left(\sum_{i=1}^{N_1}Z_i^2\,|Z_1,Z_2,...,Z_m\right)\right)$$

$$-E\left(\left(N_1\left(N_1-1\right)\right)^{-1}E\left(\sum_{i\neq j}^{N_1}\sum Z_iZ_j\,|Z_1,Z_2,...,Z_m\right)\right).$$

Consequently,

$$E\left(N_1^{-1}E\left(\sum_{i=1}^{N_1}Z_i^2\,|Z_1,Z_2,...,Z_m\right)\right)=\theta-\theta^2\left(\beta-1\right)\left(\delta n^*\right)^{-1}\left[\frac{d}{d\theta}\ln g\left(\theta\right)\right]+o\left(\lambda^{-1}\right),$$

while

$$(3.20)\quad E\left(\left(N_1\left(N_1-1\right)\right)^{-1}E\left(\sum_{i\neq j}^{N_1}\sum Z_iZ_j\,|Z_1,Z_2,...,Z_m\right)\right)$$

$$=E\left(\left(N_1\left(N_1-1\right)\right)^{-1}\sum_{i=1}^{m}\sum_{i\neq j}^{m}Z_iZ_j\right)$$

$$\leq E\left(\sum_{i=1}^{m}\sum_{i\neq j}^{m}Z_iZ_j\right)=0.$$

Also from Honda (1992) equation (3.7) we have

$$\lambda E\left(S_{N_1}^2-\theta\right)=\lambda E\left\{\frac{m\left(S_m^2-\theta\right)}{N_1}\right\}+o\left(1\right).$$

By expanding $f\left(N_1\right)=N_1^{-1}$ around δn^* and collecting the terms, we have

$$N_1^{-1}m\left(S_m^2-\theta\right)=m\left(S_m^2-\theta\right)\left(\delta n^*\right)^{-1}-m\left(\delta n^*\right)^{-1}\left(g'/g\right)\left(S_m^2-\theta\right)^2$$

$$-3/2\,m\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-1}g''\left(\eta_1\right)\left(S_m^2-\theta\right)^2+1/2\,m\left(g\left(\theta\right)\right)^{-2}\left(\delta n^*\right)^{-1}g^{2''}\left(\eta_2\right)\left(S_m^2-\theta\right)^2,$$

where η_1 and η_2 are random variables that lie between S_m^2 and θ. By taking the expectation over the above expression, we have

$$E\left\{N_1^{-1}m\left(S_m^2-\theta\right)\right\}=-m\left(\delta n^*\right)^{-1}\left(g'/g\right)E\left(S_m^2-\theta\right)^2-3/2\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-1}E\left\{mg''\left(\eta_1\right)\left(S_m^2-\theta\right)^2\right\}$$

$$+1/2\left(g\left(\theta\right)\right)^{-2}\left(\delta n^*\right)^{-1}E\left\{mg^{2''}\left(\eta_2\right)\left(S_m^2-\theta\right)^2\right\}$$

$$=-\left(\delta n^*\right)^{-1}\left(g'/g\right)\theta^2\left(\beta-1\right)-3/2\left(g\left(\theta\right)\right)^{-1}\left(\delta n^*\right)^{-1}E\left\{mg''\left(\eta_1\right)\left(S_m^2-\theta\right)^2\right\}$$

$$+1/2\left(g\left(\theta\right)\right)^{-2}\left(\delta n^*\right)^{-1}E\left\{mg^{2''}\left(\eta_2\right)\left(S_m^2-\theta\right)^2\right\}.$$

But

$$\left(\delta n^*\right)^{-1}E\left\{mg''\left(\eta_1\right)\left(S_m^2-\theta\right)^2\right\}\leq\left(\delta n^*\right)^{-1}k_1E\left(S_m^2-\theta\right)^2$$
$$=o\left(\lambda^{-1}\right).$$

Also

$$\left(\delta n^*\right)^{-1}E\left\{mg^{2''}\left(\eta_2\right)\left(S_m^2-\theta\right)^2\right\}\leq\left(\delta n^*\right)^{-1}k_2E\left\{m\left(S_m^2-\theta\right)^2\right\},$$
$$=o\left(\lambda^{-1}\right).$$

Thus $E\left\{N_1^{-1}m\left(S_m^2-\theta\right)\right\}=\left(\delta n^*\right)^{-1}\theta^2\left(\beta-1\right)\frac{d}{d\theta}\ln\left(g\left(\theta\right)\right)+o\left(\lambda^{-1}\right).$

The proof is now complete.

To prove *(ii)*, we write

$$E\left(S_{N_1}^4\right)=E\left(N_1^{-2}\left(E\left(\sum_{i=1}^{N_1}Z_i^4\right)\middle|Z_1,Z_2,...,Z_m\right)\right)$$
$$+E\left(N_1^{-2}E\left(\sum_{i\neq j}^{N_1}\sum Z_i^2Z_j^2\,\middle|Z_1,Z_2,...,Z_m\right)\right)$$
$$+E\left(\left(N_1\left(N_1-1\right)\right)^{-2}E\left(\sum_{i\neq j}^{N_1}\sum Z_i^2Z_j\,\middle|Z_1,Z_2,...,Z_m\right)^2\right).$$

Arguments similar to those used to verify (3.17) can be used to prove

$$E\left(\left(N_1\left(N_1-1\right)\right)^{-2}E\left(\sum_{i\neq j}^{N_1}\sum Z_i^2Z_j^2\,\middle|Z_1,Z_2,...,Z_m\right)\right)=0,$$

while

$$E\left(N_1^{-2}\left(E\left(\sum_{i=1}^{N_1} Z_i^{\,4}\right)\middle|Z_1,Z_2,...,Z_m\right)\right)=\beta\theta^2\left(\delta n^*\right)^{-1}+o\left(\lambda^{-1}\right),$$

and

$$E\left(N_1^{-2}E\left(\sum_{i\neq j}^{N_1}\sum Z_i^2 Z_j^2\,\middle|Z_1,Z_2,...,Z_m\right)\right)=\theta^2-\theta^2\left(\delta n^*\right)^{-1}+o\left(\lambda^{-1}\right).$$

Part *(iii)* follows immediately from parts *(i)* and *(ii)* which implies that $Var\left(S_{N_1}^2\right)=\left(\delta n^*\right)^{-1}\theta^2\left(\beta-1\right)+\left(\delta n^*\right)^{-2}\theta^4\left(\beta-1\right)\left(g^{-1}g'\right)^2+o\left(\lambda^{-2}\right),$

$$=\left(\delta n^*\right)^{-1}\theta^2\left(\beta-1\right)+o\left(\lambda^{-1}\right).$$

Also from Lemma 3.2.1.3 *(ii)* we have $\dfrac{\sqrt{N_1}\left(S_{N_1}^2-\theta\right)}{\theta\sqrt{\beta-1}}\rightarrow N\left(0,1\right)$ in distribution, as $\lambda\rightarrow\infty$. This implies that $\dfrac{N_1\left(S_{N_1}^2-\theta\right)^2}{\theta^2\left(\beta-1\right)}\rightarrow\chi^2\left(1\right)$ in distribution, as $\lambda\rightarrow\infty$. Using Lemma 3.2.1.3 *(i)* , we have

$$E\left\{\frac{N_1\left(S_{N_1}^2-\theta\right)^2}{\theta^2\left(\beta-1\right)}\right\}\rightarrow 1 \text{ as, } \lambda\rightarrow\infty .$$

Using Lemma 3.2.1.2 part *(i)*

$$E\left\{\frac{\delta n^* N_1\left(S_{N_1}^2-\theta\right)^2}{\delta n^*\theta^2\left(\beta-1\right)}\right\}\rightarrow 1\Rightarrow E\left\{\frac{\delta n^* N_1\left(S_{N_1}^2-\theta\right)^2}{\delta n^*\theta^2\left(\beta-1\right)}\right\}=\frac{\theta^2\left(\beta-1\right)}{\delta n^*}+o\left(\lambda^{-1}\right).$$

The proof is now complete.

The proof of *(iv)* follows by expanding the function $g\left(S_{N_1}^2\right)$ around $g\left(\theta\right)$ and using *(i), (ii)* and *(iii)* of Theorem 3.2.1.3.

From Theorem 3.2.1.3*(i),* $S_{N_1}^2$ may be a biased estimator of θ. The bias depends on θ, the form of the function $g\left(\cdot\right)$ and its derivative $g'\left(\cdot\right)$, the kurtosis of the underlying distribution, which is always positive, the design factor δ and the optimal sample size n^*. However, as $n^*\rightarrow\infty$ the bias of $S_{N_1}^2$ tends to zero. Similar arguments can be made regarding (ii) and (iii) above.

Moreover a second order asymptotic expansion of the expectation of the function $g\left(\cdot\right)$ of $S_{N_1}^2$ as $m\rightarrow\infty$ can be given in part *(iv).*

3.2.2 Asymptotic characteristics of the fine tuning phase

It is of interest to see whether the fine tuning stage (third stage) reduces the magnitude of the bias noticed in Theorem 3.2.1.1.

Theorem 3.2.2.1 presents some asymptotic characteristics of the third stage sample, but first we give some useful lemmas.

Lemma 3.2.2.1

For the triple sampling rule given by (2.3), if condition (2.4) holds then as $\lambda \to \infty$, we have

(i) $N/n^* \xrightarrow{a.s} 1.$

(ii) $\left(N - n^*\right)/\sqrt{n^*} \xrightarrow{L} N\left(0, (4\delta)^{-1}\theta^2(\beta-1)\right).$

Proof:

Part *(i)* follows directly from (2.4) and the strong law of large numbers. Part *(ii)* follows from *(i)* and Anscombe's Theorem.

Lemma 3.2.2.2

For the triple sampling rule given by (2.3) we have

(i) $\left\{\left(N/\lambda\right)^p\right\}, \left\{\left(\lambda/N\right)^p\right\}$ are uniformly integrable for $p > 0$.

(ii) $E\left\{\left(\lambda/N\right)^p\right\} \to 1, \quad p > 0.$

(iii) $\left\{\left|\lambda^{1/2}\left(S_N^2 - \theta\right)\right|^p\right\}$ is uniformly integrable for $0 < p \le 3.$

(iv) $\left\{\left|\lambda^{-1/2}\left(N - n^*\right)\right|^p\right\}$ is uniformly integrable for $p > 0$.

Proof: The first part of *(i)* follows directly from (2.3), while the second part can be proved as follows: $n^*/N \to 1$ a.s. as $\lambda \to \infty$, let $I\{\cdot\}$ be an indicator function of $\{\cdot\}$. Then $\left(n^*/N\right)^p I\left\{N > \delta n^*\right\} < \delta^{-p}$ for $p > 0$ and thus $\left(n^*/N\right)^p I\left\{N > \delta n^*\right\}$ is uniformly integrable. To prove Part *(ii)* we need first to find $P\left\{N \le \delta n^*\right\}$. Let $\delta \in (0,1)$. Then

$$P\left\{N \le \delta n^*\right\} \le P\left\{\lambda g\left(S_{N_1}^2\right) \le \delta\lambda g(\theta)\right\},$$

$$\le P\left\{\left|g\left(S_{N_1}^2\right) - g(\theta)\right| \ge (1-\delta)\, g(\theta)\right\}$$

$$\le P\left\{\max_{m \le n \le [\delta n^*]} \left|g^{-1}(\theta)\left(g\left(S_n^2\right) - g(\theta)\right)\right| \ge (1-\delta)\right\} = O\left(\lambda^{-r^*/s}\right),$$

where r^* is an arbitrary fixed positive integer and the last expression valid by using Hajek-Renyi inequality; see Sung (2008) for more details about the inequality.

Now $E\left\{\left(n^*/N\right)^p I\left\{N \le \delta n^*\right\}\right\} \le \left(n^*/m\right)^p P\left\{N \le \delta n^*\right\} = O\left(\lambda^{p\left(1-\frac{1}{s}\right)-\frac{r^*}{s}}\right)$, which is $o(1)$ if we choose r^* to be larger than $p(s-1)$. The proof is now complete. Also part *(i)* follows from Lemmas 3.2.1.2 and 3.2.3. Parts *(iii)* and (*iv)* follows from Honda (1992) and Liu (2002).

Lemma 3.2.2.3

If g is a positive twice continuously differentiable function such that g, g' and g'' are bounded, then for the triple sampling rule $(2.2)-(2.3)$ and under condition (2.4) and $E|Z_1|^6 < \infty$ as $\lambda \to \infty$ we have

$$(i) \quad E\left(N^{-1}\sum_{i=1}^{N_1} Z_i\right) = -\gamma\theta^{3/2}\frac{d}{d\theta}\ln g(\theta) + o\left(\lambda^{-1}\right).$$

$$(ii) \quad E\left(N^{-2}\sum_{i=1}^{N_1} Z_i^2\right) = \theta\delta n^{*-1} - 2\theta^2(\beta-1)n^{*-2}\frac{d}{d\theta}\ln g(\theta) + o\left(\lambda^{-2}\right).$$

$$(iii) \quad E\left(N^{-2}\sum_{i=1}^{N_1} Z_i Z_j\right) = 4\theta^2 n^{*-2}\frac{d}{d\theta}\ln g(\theta) - 4\theta^2\delta n^{*-3}\frac{d}{d\theta}\ln g(\theta) + o\left(\lambda^{-3}\right).$$

$$(iv) \quad E\left(N^{-1} - N^{-2}N_1\right) \le o\left(\lambda^{-1}\right).$$

Proof:

Part *(i)* follows by expanding $f(N) = N^{-1}$ around n^*, then after collecting the terms we have

$$N^{-1} = 3n^{*-1} - 3Nn^{*-2} + N^2 n^{*-3} + 6^{-1} f'''(\eta)\left(N - n^*\right)^3.$$

By substituting for $N \approx \lambda g\left(S_{N_1}^2\right)$ and expanding g and g^2 around $g(\theta)$ we obtain

$$(3.20)\quad N^{-1}=3n^{*-1}-3\lambda g\left(S_{N_1}^2\right)n^{*-2}+\lambda^2 g^2\left(S_{N_1}^2\right)n^{*-3}+6^{-1}f'''(\eta)\left(N-n^*\right)^3$$

$$=3n^{*-1}-3\lambda n^{*-2}\left\{g(\theta)+g'(\theta)\left(S_{N_1}^2-\theta\right)+1/2\,g''(\eta_1)\left(S_{N_1}^2-\theta\right)^2\right\}$$

$$+\lambda^2 n^{*-3}\left\{g^2(\theta)+g^{2'}(\theta)\left(S_{N_1}^2-\theta\right)+1/2\,g^{2''}(\eta_2)\left(S_{N_1}^2-\theta\right)^2\right\}+6^{-1}f'''(\eta)\left(N-n^*\right)^3.$$

Multiplying (3.20) by $\sum_{i=1}^{N_1} Z_i$ we obtain

$$(3.21)\quad N^{-1}\sum_{i=1}^{N_1}Z_i=3n^{*-1}\sum_{i=1}^{N_1}Z_i-3\lambda n^{*-2}\left\{g(\theta)+g'(\theta)\left(S_{N_1}^2-\theta\right)+1/2\,g''(\eta_1)\left(S_{N_1}^2-\theta\right)^2\right\}\sum_{i=1}^{N_1}Z_i$$

$$+\lambda^2 n^{*-3}\left\{g^2(\theta)+g^{2'}(\theta)\left(S_{N_1}^2-\theta\right)+1/2\,g^{2''}(\eta_2)\left(S_{N_1}^2-\theta\right)^2\right\}\sum_{i=1}^{N_1}Z_i+6^{-1}f'''(\eta)\sum_{i=1}^{N_1}Z_i\left(N-n^*\right)^3.$$

Taking the expectation over (3.21) and recalling Theorem 3.2.1.2 *(iii)* we have

$$(3.23)\quad E\left(N^{-1}\sum_{i=1}^{N_1}Z_i\right)=-3\lambda n^{*-2}\left\{g'(\theta)\gamma\,\theta^{3/2}+1/2\,Eg''(\eta_1)\sum_{i=1}^{N_1}Z_i\left(S_{N_1}^2-\theta\right)^2\right\}$$

$$+\lambda^2 n^{*-3}\left\{g^{2'}(\theta)\gamma\,\theta^{3/2}+1/2\,Eg^{2''}(\eta_2)\sum_{i=1}^{N_1}Z_i\left(S_{N_1}^2-\theta\right)^2\right\}+6^{-1}E\left\{f'''(\eta)\sum_{i=1}^{N_1}Z_i\left(N-n^*\right)^3\right\},$$

but

$$\left|E\left(R_1\right)\right|=6^{-1}\left|Ef'''(\eta)\sum_{i=1}^{N_1}Z_i\left(N-n^*\right)^3\right|\le E\left\{\eta^{-4}\left|\sum_{i=1}^{N_1}Z_i\right|\left|N-n^*\right|^3\right\}.$$

Consider the case

$$\eta>n^*\Rightarrow\left|E\left(R_1\right)\right|\le E\left\{\eta^{-4}\left|\sum_{i=1}^{N_1}Z_i\right|\left|N-n^*\right|^3\right\}$$

$$\le n^{*-4}E\left\{\left|\sum_{i=1}^{N_1}Z_i\right|N^3\right\}\le Mg^{-3}(\theta)n^{*-1}E\left\{\left|\sum_{i=1}^{N_1}Z_i\right|\right\}.$$

From Anscombe's Theorem $\left(\theta N_1\right)^{-1/2}\sum_{i=1}^{N_1}Z_i\to N(0,1)$ as $\lambda\to\infty$. But since $\left\{\left|\left(N_1\right)^{-1/2}\sum_{i=1}^{N_1}Z_i\right|\right\}$ is uniformly integrable, then $E\left|\left(N_1\right)^{-1/2}\sum_{i=1}^{N_1}Z_i\right|=\sqrt{\frac{2\theta}{\pi}}$ as $\lambda\to\infty.$ Using Lemma 3.2.1.2 *(i)*, we have $E\left|\sum_{i=1}^{N_1}Z_i\right|=\sqrt{\frac{2\theta\delta n^*}{\pi}}$ as $\lambda\to\infty.$ Thus

$$\left|E\left(R_1\right)\right| \le Mg^{-3}\left(\theta\right)n^{*-1/2}\sqrt{\frac{2\theta\delta}{\pi}} \to 0,\ \lambda \to \infty.$$

Whilst

$$\eta > N \Rightarrow \left|E\left(R_1\right)\right| \le E\left\{\eta^{-4}\sum_{i=1}^{N_1} Z_i \left|N-n^*\right|^3\right\}$$

$$\le E\left(N^{-1}\sum_{i=1}^{N_1} Z_i\right) \le E\left(N_1^{-1}\sum_{i=1}^{N_1} Z_i\right) \le E\left(m^{-1}\sum_{i=1}^{N_1} Z_i\right) < \left(\delta n^*\right)^{-1/2}\sqrt{\frac{2\theta}{\pi}} \to 0,\quad \lambda \to \infty.$$

Also,

$$\left|E\left(R_2\right)\right| = 3/2\,\lambda n^{*-2}\left|E\left\{g^{''}\left(\eta_1\right)\sum_{i=1}^{N_1} Z_i\left(S_{N_1}^2-\theta\right)^2\right\}\right| \le 3/2\,g^{-1}\left(\theta\right)n^{*-1}K_1 E\left\{\left|\sum_{i=1}^{N_1} Z_i\right|\left(S_{N_1}^2-\theta\right)^2\right\}.$$

But from Lemma 3.2.1.3 *(ii)*, we have $\frac{N_1\left(S_{N_1}^2-\theta\right)^2}{\theta^2\left(\beta-1\right)} \to \chi_1^2$ in distribution, as $\lambda \to \infty$. Moreover from Govindarajulu (1987), $N_1^{-1}\sum_{i=1}^{N_1} Z_i \to 0$ in probability, as $\lambda \to \infty$. Hence from Slutsky's Theorem we have $\left(\theta^2\left(\beta-1\right)\right)^{-1}\left|\sum_{i=1}^{N_1} Z_i\right|\left(S_{N_1}^2-\theta\right)^2 \to 0$ in distribution, as $\lambda \to \infty$. Since from Lemma 3.2.2 the quantity $\left\{\left|\sum_{i=1}^{N_1} Z_i\left(S_{N_1}^2-\theta\right)^2\right|\right\}$ is uniformly integrable, then it follows that

$E\left\{\left|\sum_{i=1}^{N_1} Z_i\right|\left(S_{N_1}^2-\theta\right)^2\right\} \to 0$ as $\lambda \to \infty$. Hence

$$\left|E\left(R_2\right)\right| = 3/2\,g^{-1}\left(\theta\right)n^{*-1}K_1 E\left\{\left|\sum_{i=1}^{N_1} Z_i\right|\left(S_{N_1}^2-\theta\right)^2\right\} \to 0,\quad \lambda \to \infty.$$

It can be shown also that $\left|E\left(R_3\right)\right| \le K_2 g^{-2}\left(\theta\right)n^{*-1}\left\{E\left|\sum_{i=1}^{N_1} Z_i\left(S_{N_1}^2-\theta\right)^2\right|\right\} \to 0,\quad \lambda \to \infty.$

Hence $\left(3.23\right)$ becomes

$$(3.24)\quad E\left(N^{-1}\sum_{i=1}^{N_1} Z_i\right) = -3\lambda n^{*-2}\left\{g^{'}(\theta)\gamma\,\theta^{3/2}\right\} + \lambda^2 n^{*-3}\left\{g^{2'}(\theta)\gamma\,\theta^{3/2}\right\} + o\left(\lambda^{-1}\right)$$

$$= -3n^{*-1}\left\{\frac{g^{'}(\theta)}{g(\theta)}\gamma\,\theta^{3/2}\right\} + n^{*-1}\left\{\frac{g^{2'}(\theta)}{g^2(\theta)}\gamma\,\theta^{3/2}\right\} + o\left(\lambda^{-1}\right)$$

$$= -n^{*-1}\gamma\,\theta^{3/2}\frac{d}{d\theta}\ln g(\theta) + o\left(\lambda^{-1}\right).$$

The proof is now complete.

Part *(ii)* follows by expanding N^{-2} around n^*, then substituting for N and expanding g and g^2 around θ. We have

$$(3.25)\quad N^{-2} = 6n^{*-2} - 8n^{*-3}N + 3n^{*-4}N^2 + 1/6\, f^{'''}(\eta)\left(N - n^*\right)^3$$

$$= 6n^{*-2} - 8n^{*-3}\lambda g\left(S_{N_1}^2\right) + 3n^{*-4}\lambda^2 g^2\left(S_{N_1}^2\right) + 1/6\, f^{'''}(\eta)\left(N - n^*\right)^3$$

$$= 6n^{*-2} - 8n^{*-3}\lambda\left\{g(\theta) + g^{'}(\theta)\left(S_{N_1}^2 - \theta\right) + 1/2\, g^{''}(\eta_1)\left(S_{N_1}^2 - \theta\right)^2\right\}$$

$$+ 3n^{*-4}\lambda^2\left\{g^2(\theta) + g^{2'}(\theta)\left(S_{N_1}^2 - \theta\right) + 1/2\, g^{2''}(\eta_1)\left(S_{N_1}^2 - \theta\right)^2\right\} + 1/6\, f^{'''}(\eta)\left(N - n^*\right)^3.$$

Multiplying (3.25) by $\sum_{i=1}^{N_1} Z_i^2$ gives

$$(3.26)\quad N^{-2}\sum_{i=1}^{N_1} Z_i^2 = 6n^{*-2}\sum_{i=1}^{N_1} Z_i^2 - 8n^{*-3}\lambda\left\{g(\theta) + g^{'}(\theta)\left(S_{N_1}^2 - \theta\right) + 1/2\, g^{''}(\eta_1)\left(S_{N_1}^2 - \theta\right)^2\right\}\sum_{i=1}^{N_1} Z_i^2$$

$$+ 3n^{*-4}\lambda^2\left\{g^2(\theta) + g^{2'}(\theta)\left(S_{N_1}^2 - \theta\right) + 1/2\, g^{2''}(\eta_1)\left(S_{N_1}^2 - \theta\right)^2\right\}\sum_{i=1}^{N_1} Z_i^2 + 1/6\, f^{'''}(\eta)\sum_{i=1}^{N_1} Z_i^2\left(N - n^*\right)^3.$$

Taking the expectation of (3.26), we have

$$(3.27)\quad E\left(N^{-2}\sum_{i=1}^{N_1} Z_i^2\right) = n^{*-2}\theta E\left(N_1\right) - 8n^{*-2}\frac{g^{'}(\theta)}{g(\theta)}\theta^2(\beta - 1) - 4n^{*-3}\lambda E g^{''}(\eta_1)\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2 - \theta\right)^2$$

$$+ 3n^{*-2}\frac{g^{2'}(\theta)}{g^2(\theta)}\theta^2(\beta - 1) + (3/2)\, n^{*-4}\lambda^2 E g^{2''}(\eta_1)\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2 - \theta\right)^2 + 1/6\, E f^{'''}(\eta)\sum_{i=1}^{N_1} Z_i^2\left(N - n^*\right)^3.$$

However,

$$\left|E\left(R_1\right)\right| = \left|E\left\{\eta^{-5}\sum_{i=1}^{N_1} Z_i^2\left(N - n^*\right)^3\right\}\right| \le E\left\{\eta^{-5}\sum_{i=1}^{N_1} Z_i^2\left|N - n^*\right|^3\right\}$$

Consider the case

$$\eta > n^* \Rightarrow \left|E\left(R_1\right)\right| \le E\left\{\eta^{-5}\sum_{i=1}^{N_1} Z_i^2 \left|N-n^*\right|^3\right\} \le E\left\{n^{*-5}\sum_{i=1}^{N_1} Z_i^2 \left|N-n^*\right|^3\right\}$$

$$\le E\left\{n^{*-5}\sum_{i=1}^{N_1} Z_i^2 N^3\right\} \le M\lambda^3 n^{*-5} E\left\{\sum_{i=1}^{N_1} Z_i^2\right\} \le Mn^{*-2} g^{-3}\left(\theta\right)\theta E\left(N_1\right),$$

by making use of Lemma 3.2.1.1 *(ii)* we have $E\left(\dfrac{N_1}{\delta n^*}\right) \to 1$ in probability as $\lambda \to \infty$. Hence

$$E\left|R_1\right| \le M\, n^{*-1} g^{-3}\left(\theta\right)\theta\delta \to 0,\ \lambda \to \infty.$$

However,

$$\eta > N \Rightarrow \left|E\left(R_1\right)\right| \le E\left\{\eta^{-5}\sum_{i=1}^{N_1} Z_i^2 \left|N-n^*\right|^3\right\} \le E\left\{N^{-5}\sum_{i=1}^{N_1} Z_i^2 \left|N-n^*\right|^3\right\}$$

$$\le E\left(N^{-2}\sum_{i=1}^{N_1} Z_i^2\right) \le E\left(m^{-2}\sum_{i=1}^{N_1} Z_i^2\right) = m^{-2}\theta\delta n^* < \theta\left(\delta n^*\right)^{-1} \to 0,\ \ \lambda \to \infty.$$

Also

$$\left|E\left(R_2\right)\right| = n^{*-3}\lambda\left|Eg^{''}\left(\eta_1\right)\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2-\theta\right)^2\right| \le n^{*-2} g^{-1} K_1 E\left\{\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2-\theta\right)^2\right\},$$

$\dfrac{N_1\left(S_{N_1}^2-\theta\right)^2}{\theta^2\left(\beta-1\right)} \to \chi_1^2$ and $N_1^{-1}\sum_{i=1}^{N_1} Z_i^2 \to \theta$ in probability, as $\lambda \to \infty$. Hence

$\dfrac{\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2-\theta\right)^2}{\theta^2\left(\beta-1\right)} \to \theta\chi_1^2$, as $\lambda \to \infty$. Since $\left\{\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2-\theta\right)^2\right\}$ is uniformly integrable from Chow and Yu (1981), then $E\left\{\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2-\theta\right)^2\right\} \to \theta^3\left(\beta-1\right)$ as $\lambda \to \infty$. This indicates that

$$\left|E\left(R_2\right)\right| \le n^{*-2} g^{-1} K_1 \theta^3\left(\beta-1\right) \to 0,\ \lambda \to \infty.$$

A similar procedure can be used to show that the last error term

$$\left|E\left(R_3\right)\right| \le K_2 n^{*-2} g^{-2} E\left\{\sum_{i=1}^{N_1} Z_i^2\left(S_{N_1}^2-\theta\right)^2\right\} \to 0,\ \ \lambda \to \infty.$$

By substituting the results of the error expectations in $\left(3.26\right)$ the proof is complete.

Part *(iii)* follows by multiplying $\sum_{i\neq j}^{N_1}\sum Z_i Z_j$ in $\left(3.25\right)$ and making use of Theorem 3.2.2.1.. Note here that

$$\left|E\left(R_1\right)\right|=\left|E\left(f'''(\eta)\left(N-n^*\right)^3\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right)\right|\leq E\left\{\eta^{-5}\left|N-n^*\right|^3\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\right\}.$$

Consider the case $\eta > n^* \Rightarrow \left|E\left(R_1\right)\right|\leq n^{*-5}E\left(\left|N-n^*\right|^3\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\right)\leq M\, g^{-3}n^{*-2}E\left(\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\right)$

Note that it can be shown easily that $E\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|=\sqrt{\frac{2}{\pi}}\,\theta\left(\delta n^*\right)$. Thus

$$\left|E\left(R_1\right)\right|\leq M\, g^{-3}n^{*-2}\sqrt{2/\pi}\,\theta\left(\delta n^*\right)\rightarrow 0,\ \lambda\rightarrow\infty.$$

However, the case

$$\eta > N \Rightarrow \left|E\left(R_1\right)\right|\leq E\left(N^{-5}\left|N-n^*\right|^3\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right)\leq E\left(N^{-2}\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right)$$

$$\leq E\left(m^{-2}\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right)<\left(\delta n^*\right)^{-2}\sqrt{2/\pi}\,\theta\left(\delta n^*\right)\rightarrow 0,\ \lambda\rightarrow\infty.$$

Also

$$\left|E\left(R_2\right)\right|=n^{*-2}\left|E\left(g''\left(\eta_1\right)\left(S_{N_1}^2-\theta\right)^2\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right)\right|\leq n^{*-2}K_1E\left(\left(S_{N_1}^2-\theta\right)^2\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\right)$$

To show that $E\left\{\left(N_1-1\right)^{-1}\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\left(S_{N_1}^2-\theta\right)^2\right\}\rightarrow 0$, as $\lambda\rightarrow\infty$ we proceed as follows.

Using Lemma 3.2.1.3 *(i)* and *(ii)*, we have $E\left\{\frac{N_1\left(S_{N_1}^2-\theta\right)^2}{\theta^2\left(\beta-1\right)}\right\}\rightarrow\chi_1^2$, as $\lambda\rightarrow\infty$. Moreover $\left\{\left(N_1\left(N_1-1\right)\right)^{-1}\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\right\}\rightarrow 0$ in probability as $\lambda\rightarrow\infty$. Using Lemma 3.2.1.3 *(i)* and the fact that $\left\{\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right\}$ can be expressed as a linear combination of uniformly integrable terms, that is $\left\{\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right\}=\left\{\sum_{i=1}^{N_1}Z_i\right\}^2-\sum_{i=1}^{N_1}Z_i^2$, then it follows from Chow and Yu (1981) that $\left\{\left(\left(N_1-1\right)\right)^{-1}\left(S_{N_1}^2-\theta\right)^2\left|\sum_{i\neq j}^{N_1}\sum Z_iZ_j\right|\right\}$ is uniformly integrable. Also from Lemma 3.2.1.2 *(i)* we have

$N_1/\delta n^* \to 1$ in probability as $\lambda \to \infty$. Then from Slutsky's Theorem it follows that $E\left\{\left|\sum_{i\neq j}^{N_1}\sum Z_i Z_j\right|\left(S_{N_1}^2-\theta\right)^2\right\} \to 0$ in distribution, as $\lambda \to \infty$ and therefore $E\left|R_2\right| \to 0$ as $\lambda \to \infty$.

A similar procedure can be used for the remaining error. The proof is then complete.

The proof of part *(iv)* follows easily from the fact that if $N_1 \leq N$, then $E\left(N_1 N^{-2}\right) \leq E\left(N^{-1}\right) = o\left(\lambda^{-1}\right)$.

The proof is now complete.

Theorem 3.2.2.1

If g is a positive twice continuously differentiable function, such that g, g' and g'' are bounded, then for the triple sampling rule (2.3), if condition (2.4) holds and $E\left|Z_1\right|^6 < \infty$ as $\lambda \to \infty$, we have

$$i)\, E\left(\bar{X}_N\right) = \mu - \gamma\,\theta^{3/2}\left(n^*\right)^{-1}\left(d\ln g(\theta)/d\theta\right) + o\left(\lambda^{-1}\right).$$

$$ii)\, Var\left(\bar{X}_N\right) = \theta\left(n^*\right)^{-1} - 2\theta^2(\beta-3)\left(n^*\right)^{-2}\left(d\ln g(\theta)/d\theta\right)$$
$$+\theta^3(\beta-1)\delta^{-1}\left(n^*\right)^{-2}\left[2\left(\frac{d}{d\theta}\ln g(\theta)\right)^2 - \left(2n^*\right)^{-1}\left(\frac{d^2 n^*}{d\theta^2}\right)\right] + o\left(\lambda^{-2}\right).$$

iii) $\bar{X}_N$ is asymptotically normally distributed with mean and variance given in *(i)* and *(ii)* respectively.

Proof:

To prove *(i)*, conditioning on the σ-field generated by the random variables $Z_1, Z_2, \ldots, Z_{N_1}$, we write

$$(3.28)\quad E\left(\bar{X}_N - \mu\right) = E\left\{N^{-1} E\left(\sum_{i=1}^{N_1} Z_i + \sum_{i=N_1+1}^{N} Z_i \,\middle|\, Z_1, Z_2, \ldots, Z_{N_1}\right)\right\}.$$

Again the first sum $\sum_{i=1}^{N_1} Z_i$ in (3.28) is non-random. Thus, (3.28) reduces to

$$(3.29)\quad E\left(\bar{X}_N - \mu\right) = E\left(N^{-1}\sum_{i=1}^{N_1} Z_i\right).$$

Applying Lemma 3.2.2.3 part *(i)* gives the result and the proof is complete**.**

The proof of (*ii*) can be obtained directly from the following

$$(3.30)\quad Var\left(\bar{X}_N\right)=E\left(N^{-2}E\left(\sum_{i=1}^{N_1}Z_i^2+\sum_{i\neq j}^{N_1}\sum Z_iZ_j\,\middle|Z_1,Z_2,...,Z_{N_1}\right)\right)$$
$$+\theta E\left(E\left(N^{-1}-N_1N^{-2}\,\middle|Z_1,Z_2,...,Z_{N_1}\right)\right).$$

Using Lemma 3.2.2.3 *(ii)*, *(iii)* and *(iv)* and substituting them in (3.30) gives the result. The proof is now complete.

The proof of *(iii)* can be obtained either by using Anscombe's Theorem or by using the moment generating function technique . We proceed first by using moment generating function technique.

Let $u(\cdot)$ be a continuously differentiable function around the population mean μ such that $u^{'}, u^{''}$ and $u^{'''}$ are continuous over the interval $\left[\bar{X}_N,\mu\right]$. Then a Taylor expansion gives

$$u\left(\bar{X}_N\right)=u(\mu)+u^{'}(\mu)\left(\bar{X}_N-\mu\right)+\frac{1}{2}u^{''}(\mu)\left(\bar{X}_N-\mu\right)^2+\frac{1}{6}u^{'''}(\eta)\left(\bar{X}_N-\mu\right)^3,$$

where η lies between $\bar{X}_N$ and μ . Thus

$$E\left(u\left(\bar{X}_N\right)\right)=u(\mu)+u^{'}(\mu)E\left(\bar{X}_N-\mu\right)+\frac{1}{2}u^{''}(\mu)E\left(\bar{X}_N-\mu\right)^2+\frac{1}{6}E\left\{u^{'''}(\eta)\left(\bar{X}_N-\mu\right)^3\right\}.$$

If we set $u\left(\bar{X}_N\right)=exp\left(t\left(\bar{X}_N-\mu\right)/\sigma_{\bar{X}_N}\right)$, and use part s *(i)* and *(ii)* of Theorem 3.2.2.1, we obtain

$$M_{\bar{X}_N}(t)\approx 1+\frac{1}{2}t^2+o\left(\lambda^{-1}\right),\ \text{as}\ \lambda\to\infty\ ,$$

which is the limiting moment generating function of the standard normal distribution.

Now we need to show that the error term converges to zero as $\lambda\to\infty$.

$$\left|E(R)\right|=\left|E\left\{u^{'''}(\eta)\left(\bar{X}_N-\mu\right)^3\right\}\right|\le t^3\sigma_{\bar{X}_N}^{-3}\ E\left\{\exp\left(t(\eta-\mu)/\sigma_{\bar{X}_N}\right)\left|\bar{X}_N-\mu\right|^3\right\}.$$

Consider the case $\eta<\mu, t>0\Rightarrow \exp\left\{t(\eta-\mu)/\sigma_{\bar{X}_N}\right\}<1$. Thus

$$\left|E(R)\right|\le t^3\sigma_{\bar{X}_N}^{-3}\ E\left\{\left|\bar{X}_N-\mu\right|^3\right\}\le t^3\sigma_{\bar{X}_N}^{-3}E\left\{\left|\bar{Z}_N\right|^3\right\}=t^3\sigma_{\bar{X}_N}^{-3}E\left(\left|N^{-1}\sum_{i=1}^{N}Z_i\right|^3\right)$$
$$\le t^3\sigma_{\bar{X}_N}^{-3}m^{-3}E\left(\left|\sum_{i=1}^{N}Z_i\right|^3\right)$$

By Anscombe's Theorem $(N\theta)^{-1/2}\sum_{i=1}^{N} Z_i \to N(0,1)$ as $\lambda \to \infty$ and since $\left\{\left|(N)^{-1/2}\sum_{i=1}^{N} Z_i\right|^3\right\}$ is uniformly integrable then it follows

$E\left\{\left|(\theta N)^{-1/2}\sum_{i=1}^{N} Z_i\right|^3\right\} = 2\sqrt{2/\pi},\ \lambda \to \infty \Rightarrow E\left\{\left|(N)^{-1/2}\sum_{i=1}^{N} Z_i\right|^3\right\} = 2\theta\sqrt{2\theta/\pi},\ \lambda \to \infty$. Using Lemma 3.2.2.1, it follows that $E\left\{\left|\sum_{i=1}^{N} Z_i\right|^3\right\} = 2\theta\left(n^*\right)^{3/2}\sqrt{2\theta/\pi}$ as $\lambda \to \infty$. Using condition (2.4), we have

$$\left|E(R)\right| \le t^3 \sigma_{\bar{X}_N}^{-3}\left(\delta n^*\right)^{-3} 2\theta\left(n^*\right)^{3/2}\sqrt{2\theta/\pi} \to 0,\ \lambda \to \infty.$$

While if

$\eta < \bar{X}_N,\ t < 0 \Rightarrow 0 < \eta - \mu < \bar{X}_N - \mu \Rightarrow 0 > t(\eta - \mu) > t\left(\bar{X}_N - \mu\right)$. This implies the above case.

The proof is complete.

For more details about the convergence of moment generating functions, see Mukherjea *et al.* (2006).

Now we proceed to prove *(iii)* using Theorem 3.2.1. By choosing $\nu(t) = N$ and $t = n^*$ in

Lemma 3.2.2.1, we have $(\theta N)^{-1/2}\sum_{i=1}^{N} Z_i \to N(0,1)$ as $n^* \to \infty$.

But note that $\sum_{i=1}^{N} Z_i = \sum_{i=1}^{N} X_i - N\mu$, which implies that $\sqrt{N\theta^{-1}}\left(\bar{X}_N - \mu\right) \to N(0,1)$ as $n^* \to \infty$. The proof is now complete.

In view of $i)$ and $ii)$ of Theorem 3.2.2.1, it is worth mentioning that the third stage has indeed reduced the magnitude of the bias noticed in $i)$ and $ii)$ of Theorem 3.2.1.1.

The expectation of the final stage sample size N and other asymptotic characteristics can be easily obtained from (iv) of Theorem 3.2.1.3 above, as given in the following Theorem 3.2.2.2.

Theorem 3.2.2.2

Let g be a positive twice continuously differentiable function, such that g, g' and g'' are bounded and let N be defined as in (2.3) and assume that condition (2.4) holds with $E\left|Z_1\right|^6 < \infty$ then as $\lambda \to \infty$, we have

$$i) E(N) = n^* - \theta^2(\beta-1)\left(\delta n^*\right)^{-1}\left(n^*\left(\frac{d}{d\theta}\ln\left(n^*\right)\right)^2 - (1/2)\frac{d^2n^*}{d\theta^2}\right) + E\left(\varepsilon_{N_1}\right) + o(1).$$

$$ii) Var\left(N\right) = \theta^2(\beta-1)\left(\delta n^*\right)^{-1}\left(dn^*/d\theta\right)^2 + o(\lambda).$$

$$iii) E\,|\,N - E\left(N\right)|^3 = o\left(\lambda^2\right),$$

where the random variable $\varepsilon_{N_1} = 1 - \left(\lambda g\left(S^2_{N_1}\right) - \left[\lambda g\left(S^2_{N_1}\right)\right]\right)$, is defined over the interval $(0,1)$.

Proof:

To prove part *(i)*, note that $N = \left[\lambda g\left(S^2_{N_1}\right)\right] + 1$, $a.s.$ except possibly on a set $\xi = \left\{\left(N_1 > \left[\lambda g\left(S^2_{N_1}\right)\right] + 1\right) \cup \left(m > \left[\delta\,\lambda\,g\left(S^2_m\right)\right] + 1\right)\right\}$ of measure zero, such that $\int_\xi N\,dP = o(1)$; see, for example, Hall (1981) for details. Hence,

$$(3.31)\quad N = \left[\lambda g\left(S^2_{N_1}\right)\right] + 1$$
$$= \lambda g\left(S^2_{N_1}\right) - \left\{\lambda g\left(S^2_{N_1}\right) - \left[\lambda g\left(S^2_{N_1}\right)\right]\right\} + 1$$
$$= \lambda g\left(S^2_{N_1}\right) + \varepsilon_{N_1}.$$

Thus, $E(N) = \lambda E\left(g\left(S^2_{N_1}\right)\right) + E\left(\varepsilon_{N_1}\right) + o(1)$, as $m \to \infty$.

Using Theorem 3.2.1.3 part *(iv)*, we obtain the result.

Proof of *(ii):*

$Var(N) = Var\left(\lambda g\left(S^2_{N_1}\right)\right) = \lambda^2\,Var\left(g\left(S^2_{N_1}\right)\right)$, as $m \to \infty$

but

$$Var\left(g\left(S^2_{N_1}\right)\right) = Var\left(g(\theta) + g'(\theta)\left(S^2_{N_1} - \theta\right) + \frac{1}{2}g''(\tau)\left(S^2_{N_1} - \theta\right)^2\right)$$
$$= \left(g'(\theta)\right)^2 Var\left(S^2_{N_1}\right) + o\left(\lambda^{-1}\right),$$

where τ is a random variable lies between $S^2_{N_1}$ and θ. To show that the error vanishes as $\lambda \to \infty$ we proceed as follows:

$$\left|E\left(R\right)\right| = \left|E\left(g''(\tau)\left(S^2_{N_1} - \theta\right)\right)\right| \le E\left(\left|g''(\tau)\left(S^2_{N_1} - \theta\right)^2\right|\right) \le E\left\{\left|g''(\tau)\right|\left|\left(S^2_{N_1} - \theta\right)^2\right|\right\}$$
$$\le K_1\,E\left(\left(S^2_{N_1} - \theta\right)^2\right) = K_1\theta^2(\beta-1)\left(\delta n^*\right)^{-1} \to 0,\ \lambda \to \infty.$$

Here K_1 is a generic constant independent of $S_{N_1}^2$, θ and τ such that $\left|g^{''}(\tau)\right| \le K_1$. By using Theorem 3.2.1.3 part *(iii)*, and the assumption that $g(\cdot)$ and its derivatives are bounded, part (*ii*) of Theorem 3.2.2.2 is complete.

Proof of *(iii)* can be illustrated as follows

$$
\begin{aligned}
E\left|N - E(N)\right|^3 &= \lambda^3 E\left|g\left(S_{N_1}^2\right) - E\left(g\left(S_{N_1}^2\right)\right)\right|^3 \\
&\le \lambda^3 E\left|g\left(S_{N_1}^2\right)\right|^3 \\
&= \lambda^3 o\left(\lambda^{-1}\right) \\
&= o\left(\lambda^2\right).
\end{aligned}
$$

By using Theorem 3.2.1.3 part *(iv)*, we obtain the required result.

Remark

Hall (1981) proved that when the underlying distribution is normal, ε_{N_1} is asymptotically uniformly distributed over the interval $(0,1)$, i.e. $\varepsilon_{N_1} \xrightarrow{L} U(0,1)$, as $\lambda \to \infty$. We show in chapter V section 5.7, using simulation, that this result appears to hold more generally than simply for the case in which the underlying distribution is normal, see Yousef *et al.* (2009) for more details. It is also evident that under the normal distribution, Theorem 1 in Hall's (1981) paper is a special case of Theorem 3.2.2.2 above when $\beta = 3$ and the optimal sample size $n^* = \lambda\sqrt{\theta}$ provided that $\theta < \theta_0 \in (0, \infty)$. We also emphasise that both the expectation and variance of N depend on the kurtosis of the underlying distribution and accordingly will reflect the amount of departure from normality. It is also of interest to give a general form of the expectation of a real valued continuously differentiable function of the fine tuning phase sample size N to be able to derive asymptotic results of all moments of N. We also want to stress that we have not assumed independence of $\bar{X}_{N_1}$ and S_m^2 or of $\bar{X}_N$ and $S_{N_1}^2$, and therefore the above results are more general in that sense (i.e., the estimate of the nuisance parameter and the estimate of the targeted parameter could be correlated).

Problems which yield independence, such as those of the normal and exponential distributions, may be treated as special cases of our findings above.

Theorem 3.2.2.3

Let g be a twice continuously differentiable function, such that $g, g^{'}$ and $g^{''}$ are bounded and let N be defined as in (2.3) and let $h(>0)$ be a continuously differentiable real valued function in a neighborhood around n^*, such that $\underset{n \ge m}{Sup}\, h(n) = O\left(h'''\left(n^*\right)\right)$. Then

$$E\left(h(N)\right)=h\left(n^*\right)-\theta^2\left(\beta-1\right)\left(\delta n^*\right)^{-1}\left\{\left(\frac{dn^*}{d\theta}\right)^2\left[\left(n^*\right)^{-1}h'\left(n^*\right)-(1/2)h''\left(n^*\right)\right]-(1/2)h'\left(n^*\right)\frac{d^2n^*}{d\theta^2}\right\}$$

$$+h'\left(n^*\right)E\left(\varepsilon_{N_1}\right)+o\left(\lambda^{-1}\left(|h'''|\right)\right).$$

Proof:

The proof follows by expanding the differentiable function $h(N)$ around $h\left(n^*\right)$ using Taylor series and applying the results of Theorem 3.2.2.2. The general form of the second order asymptotic expansion of the expectation of a real valued continuously differentiable function $h(>0)$ enables one to obtain the expectations of positive and negative moments of N in subsequent analysis.

Moreover, it also helps to provide a second order asymptotic expansion of the coverage probability while constructing a fixed width confidence interval of the unknown mean μ (see Chapter VIII, section 8.2).

3.3 Asymptotic normality of the stopping variable N

Theorem 3.3.1

Let g be a positive twice continuously differentiable and bounded function of θ and let N be defined as in (2.3) such that $E(N)<\infty$ and $Var(N)<\infty$ also assume that condition (2.4) holds. Then, as $\lambda\to\infty$, N is asymptotically normal with mean $E(N)$ and variance $\sigma_N^2=Var(N)$, $\sigma_N^2=\theta^2(\beta-1)\left(\delta n^*\right)^{-1}\left(dn^*/d\theta\right)^2+o(\lambda)$.

Proof: (using the moment generating function technique)

The proof of Theorem 3.3.1 is a straightforward application of Theorem 3.2.2.3 above by setting $h(v)=e^{tv}$, where $v=\left(N-n^*\right)/\sqrt{Var(N)}$. This yields

$$E\left[Exp\left(t\left(\frac{N-n^*}{\sigma_N}\right)\right)\right]=1+\left(t^2/2\right)-\left(t\sigma_N/n^*\right)+(t/2)\theta\sqrt{\frac{\beta-1}{\delta n^*}}\ \ln\left(g'(\theta)\right)+o\left(\lambda^{-1}\right).$$

By letting $n^*\to\infty$, we have

$$E\left[Exp\left(t\left(\frac{N-n^*}{\sigma_N}\right)\right)\right]\approx 1+\left(t^2/2\right)+o\left(\lambda^{-1}\right),$$

which is the moment generating function of the standard normal distribution. The proof is complete.

Proof: (By using Anscombe's Theorem)

The proof follows immediately from Lemma 3.2.2.1 part *(i)* and from Lemma 3.2.2.2 parts *(i)* and *(ii)*. Also see Lemma 3.10 in Honda (1992).

Use of Theorem 3.3.1 will enable us to find all the moments of N, or a bounded differentiable function of N, in addition to the distribution function of $g\left(S_{N_1}^2\right)$, which will facilitate us in proving theorems about N and $g\left(S_{N_1}^2\right)$. Bhattacharya *et al.* (1973) discussed the asymptotic normality of the stopping times of some sequential procedures and proved that the asymptotic distribution of the stopping time of a procedure due to Robbins (1959) is normal. Further investigations about Theorem 3.3.1, using simulation, will be discussed in Chapter V section 5.4.

Theorem 3.3.2

Let $\bar{X}_N$ be as in Theorem 3.2.2.1, where N is a triple sampling rule given by (2.3). Then $\bar{X}_N$ and N are asymptotically uncorrelated, as $\lambda \to \infty$.

Proof:

$$Cov\left(\bar{X}_N, N\right) = Cov\left(N^{-1}\sum_{i=1}^{N} Z_i + \mu, N\right) = Cov\left(N^{-1}\sum_{i=1}^{N} Z_i, N\right)$$
$$= E\left(\sum_{i=1}^{N} Z_i\right) - E\left(N^{-1}\sum_{i=1}^{N} Z_i\right) E(N).$$

The first part tends to zero from Wald's (1947) first equation, while the second part vanishes because $N^{-1} \le 1, E\left(\sum_{i=1}^{N} Z_i\right) = E(Z)\,E(N) = 0$ and $E\left(\bar{Z}_N\right) = E\left(N^{-1}\sum_{i=1}^{N} Z_i\right) \le E\left(\sum_{i=1}^{N} Z_i\right) = 0$. Hence $\left(\bar{X}_N, N\right)$ are asymptotically uncorrelated. The proof is complete.

It would be fairly difficult in this stage to prove that $\left(\bar{X}_N, N\right)$ are asymptotically independent since the underlying distribution is unknown. Therefore as a conjecture we need to assume that $\left(\bar{X}_N, N\right)$ are asymptotically independent as $\lambda \to \infty$. Moreover we will support our conjecture from the side of the simulation results regarding the asymptotic coverage probability and the asymptotic Type II error probability when the underlying distribution is normal. We will show that our simulation results under the normal distribution agree with the results of Hall (1981), Mukhopadhyay *et al.* (1987), Hamdy (1988) , Costanza *et al.* (1995) and Son *et al.* (1997) . Note here that under the normal distribution, $\left(\bar{X}_N, N\right)$ are independent.

Having constructed the theory of triple sampling given by $(2.2)-(2.3)$, we consider in chapter IV the effect of departures from normality of the underlying distribution on Hall's (1981) triple sampling scheme under the squared error loss function. Moreover, we will compute the asymptotic regret in this case. This will be compared with the corresponding asymptotic regret when the underlying distribution is known to belong to the one-parameter exponential family.

Chapter IV

Point Estimation of the Population Mean using the Triple Sampling Procedure

4.1 Squared error loss function to estimate the mean

In this section our main objective is to develop a triple sampling point estimation procedure to estimate the mean μ of the unknown distribution. Specifically, if a point estimate of the unknown μ is required, we assume that the incurred cost of estimating the mean μ by the corresponding sample mean $\bar{X}_n$ can be approximated by the following squared error loss function in (4.1) with linear sampling cost. The literature in sequential sampling has considered several forms of higher order loss (cost) functions to model estimation cost. However, squared error loss functions are recommended and commonly used in sequential point estimation problems (see, for example, Degroot, 1970). Therefore, we write the loss (cost) function as

$$(4.1) \qquad L_n(A) = A\left(\bar{X}_n - \mu\right)^2 + Cn, \quad A > 0, C > 0,$$

where C is the known cost per unit sampled. The constant A is permitted to approach infinity and represents the monetary amount that needs to be paid to achieve the minimum risk, while $A\left(\bar{X}_n - \mu\right)^2$ is the estimation cost. We shall elaborate further on determination of A in subsequent developments. The risk associated with (4.1) is

$$(4.2) \qquad \begin{aligned} R_n(A) &= E\left(L_n(A)\right) = AE\left(\bar{X}_n - \mu\right)^2 + Cn \\ &= A(\theta/n) + Cn. \end{aligned}$$

Treating n as a continuous variable in (4.2), we differentiate (4.2) with respect to n and equate the results to zero to obtain the optimal sample size as

$$(4.3) \qquad n^* = \sqrt{A\theta/C}.$$

The numerical value of n^* in (4.3) is unknown because the population variance θ is unknown in the context of this thesis. It has been shown by Dantzig (1940), Stein (1945) and Seelbinder (1953) that no fixed sample size procedure exists to achieve the above optimal requirement uniformly over $\theta > 0$. In other words, since the optimal sample size in (4.3) depends on the unknown variance θ, no fixed sample size procedure can be used to estimate μ optimally over all θ. Therefore, the triple sampling procedure in $(2.2)-(2.3)$ may be used to estimate μ with $\lambda = \sqrt{A/C}$ and $g(\theta) = \sqrt{\theta}$.

From $\left(4.3\right)$, the value $A=\left(Cn^{*}\right)\left(n^{*}/\theta\right)$, where $\left(C\,n^{*}\right)$ is the cost of optimal sampling, while $\left(n^{*}/\theta\right)$ is the total information contained in the sample : the amount of information required to explore a unit of variance in order to achieve the minimum risk. Hence, A is the cost of perfect information and contrarily to what has been said that it represents the cost of estimation.

The question arises: how efficient is the triple sampling procedure estimator of μ relative to other estimators?

4.2 The asymptotic regret of triple sampling point estimation under a squared error loss function

In the literature in sequential point estimation several measures have been developed of the efficiency of sequential (triple sampling, accelerated sequential schemes) procedures relative to the fixed sample size counterpart had the form of $g(\theta)$ in $\left(2.1\right)$ been completely specified; see Ghosh *et al.* (1977) and Starr (1966) for details. The regret reflects the expected cost of missed opportunity; it measures the expected loss in using triple sampling to estimate μ rather than using the corresponding fixed sample size procedure had the nuisance parameter(s) been known. One of the measures that can be used to assess the efficiency of a sequential procedure is the difference between the sequential risk and the optimal risk, but such a measure is useless in the case of Stein's two stage procedure since the measure goes to infinity. Other weaker measures, like the asymptotic relative efficiency (risk efficiency) $\eta(A)=E(L_N(A))\big/E(L_{n^*}(A))$ which is the ratio of the triple sampling risk compared to the optimal risk, may also be used to assess the efficiency of the triple sampling procedure relative to fixed sample size procedure. For an efficient sampling procedure we expect $\eta(A)\to 1$ and $\omega(A)\to 0$ as $A\to\infty$ where

$$\omega\left(A\right)=E\left(L\left(N\right)\right)-E\left(L\left(n^{*}\right)\right).$$

Note

A procedure is called asymptotically risk-efficient or asymptotically first order risk efficient if $\lim_{A\to\infty}\eta\left(A\right)=1$, while it is called asymptotically second order risk efficient if $\lim_{c\to 0}c^{-1}\,\omega\left(A\right)<\infty$.

Recall the squared error loss function in $\left(4.1\right)$, the risk function in $\left(4.2\right)$ and the optimal sample size in $\left(4.3\right)$; the asymptotic characteristics of the efficiency of the estimator of μ using triple sampling are discussed in the following Theorem.

Theorem 4.2.1

For the triple sampling rule $\left(2.2\right)-\left(2.3\right)$, the asymptotic risk for squared error loss $\left(4.1\right)$ and under $g\left(\theta\right)=\sqrt{\theta}$ as $m\to\infty$ is given by

$$R_N\left(A\right)=E\left[L_N\left(A\right)\right]=2Cn^{*}-C\left(\beta-3\right)+(1/4)\left(\beta-1\right)\left(C/\delta\right)+C\,E\left(\varepsilon_{N_1}\right)+o\left(1\right).$$

Moreover, the asymptotic relative efficiency of the triple sampling procedure and the asymptotic regret as $A \to \infty$ are given by

$i)\ \eta(A) = 1 + o\left(\lambda^{-1}\right)$

$ii)\ \omega(A) = -C(\beta - 3) + (1/4)(\beta - 1)(C/\delta) + C\,E\left(\varepsilon_{N_1}\right) + o(1).$

Proof:

Recall Theorem 3.2.2.3, evaluate the terms $E(N)$ and $E\left(N^{-1}\right)$ and consider the optimal risk $R_{n^*}(A) = 2Cn^*$. Then Theorem 4.2.1 is immediate.

It is worth making some comments about Theorem 4.2.1. Firstly, the results for the normal distribution treated by Mukhopadhyay *et al.* (1987), Hamdy (1988) and Hamdy *et al.* (1988) are special cases. Secondly, for distributions with $\beta < 3$ a non-vanishing positive regret is expected. In addition, for distributions with $\beta > 3$ (fatter tailed than the normal) we expect either positive or negative non-vanishing regret, depending on the values of β and δ. Specifically, for distributions with $\beta > 6$, negative regret is expected when $\delta = 1/2$.

Martinsek (1988) argued that for the one-by-one purely sequential procedure negative regret is expected when $\beta > 3$. It is also worth mentioning that the regret of one-by-one purely sequential procedures involves both kurtosis and skewness of the underlying distribution, as indicated by Martinsek (1988), while our findings in Theorem 4.2.1 emphasise that the triple sampling procedure involves only the kurtosis of the underlying distribution and effectively treats underlying distributions as if they are symmetric. This could be due to the nature of one-by-one purely sequential procedures, which filter data. This filtration may cause either accelerating or delaying termination of the procedure (to cross over the boundaries). On the other hand, triple sampling uses bulks (batches) to decide whether to stop or to continue sampling. Therefore, if an extreme observation presents, it will not affect the decision compared to the influence of the rest of the bulk at that stage. This may cause the triple sampling procedure to be less sensitive to extreme observations compared to the one-by-one purely sequential procedures. Consequently, the skewness does not play a role in determining the regret of triple sampling.

A general formula for the asymptotic regret incurred in estimating the unknown mean μ under squared error loss (4.1) is

$$(4.4)\quad \omega(A) = -2C\theta(\beta - 3)\left(d\ln\left(n^*\right)/d\theta\right) + C\theta^2(\beta - 1)\delta^{-1}\left(n^*\right)^{-2}\left(dn^*/d\theta\right)^2 + C\,E\left(\varepsilon_{N_1}\right) + o(1).$$

Obviously, the non-vanishing regret in (4.4) above depends on the kurtosis β, the design factor δ, the cost of unit sampling C, the variance of the underlying distribution θ and the form of the function $g(\theta)$.

Consider now the loss function in Martinsek (1988) of the form

$$L_n(A) = A\theta^{b-1}\left(\bar{X}_n - \mu\right)^2 + n, \quad A > 0, b > 0.$$

Under the triple sampling scheme and this loss function, it can be shown that the asymptotic regret is

$$(4.5) \quad \omega(A) = -(\beta - 3)b + b^2\left((\beta - 1)/4\delta\right) + E\left(\varepsilon_{N_1}\right) + o(1).$$

The proof of (4.5) follows from Theorem 3.2.2.3 by evaluating the terms $E(N)$ and $E\left(N^{-1}\right)$ and considering the optimal risk $R_{n^*}(A) = 2\sqrt{A\theta^b}$, where $n^* = \sqrt{A\theta^b}$.

The asymptotic regret of the triple sampling procedure in (4.5) is the same as equation (7) of Martinsek (1988) for symmetric underlying distributions but under the one-by-one purely sequential procedure proposed by Robbins (1959) and the above loss function and without the design factor δ and $E\left(\varepsilon_{N_1}\right)$. The asymptotic regret of Martinsek (1988), equation (7) is

$$\omega(A) = -(\beta - 3)b + b^2\left((\beta - 1)/4\right) + o(1), \quad \text{as } A \to \infty.$$

4.3 The case of the one parameter exponential family

In previous sections, we have considered the case of an underlying distribution that is completely unspecified except that the first six moments are finite. We have developed asymptotic results for triple sampling to estimate μ in the presence of the unknown nuisance variance θ. The results for triple sampling depend on the skewness and kurtosis of the underlying distribution. A natural question is: if extra information is known about the structure of the underlying distribution, for example we know the class to which the distribution belongs, would it improve the results we have obtained under the assumption that the underlying distribution is unspecified? In this section we consider the case in which the underlying distribution is known to be in the natural one-parameter exponential family, defined by

$$dF_\nu(x) = \int_R e^{\nu x - \psi(\nu)} dP(x), \qquad x \in R, \ \nu \in \Omega$$

with respect to a σ - finite measure *P*. The natural parameter space Ω is an open interval on the real line *R* over which:

$$\int_R e^{\nu x} dP(x) < \infty, \qquad x \in R, \ \nu \in \Omega.$$

The function $\psi(\cdot)$ is convex on the sample space Ω (see, Lehman, 1986, p. 57) satisfying the moment generating function $M_X(t) = E(e^{tX}) = e^{\psi(t+\nu) - \psi(\nu)}$ (see AlMahmeed *et al.*, 1998). Hence the first four moments are

$$E(X)=\psi'(\nu),\ Var(X)=\psi''(\nu)$$
$$E\left(X-\psi'(\nu)\right)^3=\psi'''(\nu),\ E\left(X-\psi'(\nu)\right)^4=\psi''''(\nu)+3\left(\psi''(\nu)\right)^2,$$

and hence

$$\gamma=\psi'''(\nu)/\left(\psi''(\nu)\right)^{3/2} \text{ and } \beta=\psi''''(\nu)/\left(\psi''(\nu)\right)^2+3.$$

AlMahmeed *et al.* (1998) consider estimating the mean $\mu=\psi'(\nu)$ of the general form of a one-parameter exponential family using triple sampling procedure under the squared error loss function with linear sampling cost similar to (4.1), but with $C=1$ and A replaced by A^2. However, they checked the results for the gamma distribution and for the normal distribution with known mean but unknown variance. They considrerd the following loss function,

$$(4.6)\qquad L_n(A)=A^2\left(\bar{X}_n-\psi'(\nu)\right)^2+n, \text{ where } A>0.$$

The following Lemmas and Theorems are given in AlMahmeed *et al.* (1998).

The three stage exponential family (triple sampling procedure applied to the one parameter exponential family) with optimal stopping rule $n^*=Ag(\mu)$, where $\mu=\psi'(\nu)$ was defined in AlMahmeed *et al*. (1998) as follows:

Let $X_1,\ldots,X_m$ be a random sample of size $m(\geq 2)$ from the distribution function $F_\nu(\cdot)$ to compute the estimate $g(\bar{X}_m)$ of $g(\psi'(\nu))$. Then a fraction $\delta\in(0,1)$ is selected to determine the percentage of n^* to be estimated in the second stage. Accordingly, the second stage sample size is determined by the following stopping rule

$$N_1=\max\left\{m,\left[\delta Ag(\bar{X}_m)\right]+1\right\},$$

If the decision is to continue sampling, the initial sample is augmented by a second randomly selected sample of size N_1-m to determine the final sample size from the stopping rule

$$N=\max\left\{N_1,\left[Ag(\bar{X}_{N_1})\right]+1\right\}.$$

If necessary a third batch of size $N-N_1$ is randomly selected and combined with the previous N_1 observations to compute the sequential estimator $\bar{X}_N$ for the unknown parameter $\psi'(\nu)$.

Lemma 4.3.1

For the three stage exponential family rule, see AlMahmeed *et al.* (1998), if $g(\cdot)$ and its derivative are bounded, then as $m\to\infty$, we have

$i)\, AE\left(\bar{X}_{N_1} - \psi^{'}(\nu)\right) = -(1/2)\gamma\,\delta^{-1} + o(1).$

$ii)\, AE\left(\bar{X}_{N_1} - \psi^{'}(\nu)\right)^2 = \sqrt{\psi^{''}(\nu)}\,\delta^{-1} + o(1).$

$iii)\, AE\left(g\left(\bar{X}_{N_1}\right) - g\left(\psi^{'}(\nu)\right)\right) = (1/4)\left(\beta - (5/2)\gamma^2 - 3\right)\delta^{-1} + o(1).$

Lemma 4.3.1(i) shows directly that $\bar{X}_{N_1}$ may be a biased estimator of $\psi^{'}(\nu)$. The bias depends crucially on the skewness of the underlying distribution. If $\gamma = 0$, then $\bar{X}_{N_1}$ is an unbiased estimator of $\psi^{'}(\nu)$. Part(ii) shows that $Var\left(\bar{X}_{N_1}\right) = \psi^{''}(\nu)\left(\delta n^*\right)^{-1} + o\left(A^{-1}\right)$, which is obtained by setting $A = n^* / \sqrt{\theta}$ in *(ii)*, exhibits the same pattern as in Theorem 3.2.1.1*(i).*

Part *(iii)*, shows that if $\gamma = 0$ and $\beta = 3$, then $g\left(\bar{X}_{N_1}\right)$ is an unbiased estimator of $g\left(\psi^{'}(\nu)\right)$. Therefore, the results in Lemma 4.3.1 depend on the structure of the underlying distribution and are sensitive to the departure from normal theory for moderate sample sizes. However, robustness is expected asymptotically.

Lemma 4.3.2

Under the conditions in Lemma 4.3.1, we have, as $m \to \infty$, the following

$i)\, E(N) = n^* + (1/2)\left(1 + (1/2)\left(\beta - (5/2)\gamma^2 - 3\right)\delta^{-1}\right) + o(1)$

$ii)\, A^{-1} Var(N) = (1/4)\sqrt{\psi^{''}(\nu)}\,\gamma^2\,\delta^{-1} + o(1)$

$iii)\, A^{-2} E\left(\left|N - n^*\right|^3\right) = o(1).$

Lemma 4.3.2(i) shows directly that the mean of the actual sample size N depends on both the skewness and the kurtosis of the underlying distribution beside the values of the design factor δ and the optimal sample size n^*. Part *(ii)* shows that the variance of the actual sample size N depends on the skewness and the variance of the underlying distribution as well as δ . Part *(iii)* shows that the absolute third moment of the actual sample size N around n^* is of $o\left(\lambda^2\right)$.

Theorem 4.3.1

Under the condition of Lemma 4.3.1, let $h(.)$ be a continuously differentiable function in a neighbourhood of n^*, such that $\underset{n \geq m}{Sup}\, h^{'''}\left(n^*\right) = O\left(\left|h'''(n^*)\right|\right)$. Then as $m \to \infty$,

$$E\left(h(N)\right) = h\left(n^*\right) + (1/2)\left\{\left(\delta + (1/2)\left(\beta - (5/2)\gamma^2 - 3\right)\right)h^{'}\left(n^*\right) + (1/4)n^*\gamma^2 h^{''}\left(n^*\right)\right\}\delta^{-1} + o\left(A^2\left(|h'''(A)|\right)\right).$$

Proof:

The proof follows by expanding $h(N)$ around n^*; that is, by using a Taylor expansion we have

$$h(N)=h(n^*)+h^{'}(n^*)(N-n^*)+\frac{1}{2}h^{''}(n^*)(N-n^*)+\frac{1}{6}h^{'''}(\xi)(N-n^*),$$

where ξ is a number between N and n^*. By taking the expectation and using Lemma 4.3.2, we get the result. The proof is then complete.

Similarly, the representation above illustrates that the procedure is sensitive to departures from normality, since it depends on the skewness and the kurtosis of the underlying distribution.

Lemma 4.3.3

Under the condition of Lemma 4.3.1, and as $m \to \infty$, we have

$$(i)\, E(\bar{X}_N)=\psi^{'}(\nu)-(1/2)\gamma A^{-1}+o(A^{-1})$$
$$(ii)\, A^2 E(\bar{X}_N-\psi^{'}(\nu))^2=n^*+(1/4)\{(1/2)(7+16\delta)\gamma^2-(1+4\delta)\beta+(3+10\delta)\}\delta^{-1}+o(1).$$

Although the triple sampling fine tuning stage reduces the magnitude of the bias noticed in Lemma 4.3.1, the results remain sensitive to departures from normality since they depend on the kurtosis and skewness of the underlying distribution. However, robustness is attained asymptotically.

Theorem 4.3.2

Under the quadratic loss function given by (4.6), the asymptotic regret of the three stage exponential family rule is given by

$$(4.7) \qquad \omega(A)=(1/4)(8\delta+1)\gamma^2\delta^{-1}-(\beta-3)+o(1), \text{ as } m\to\infty.$$

If $\delta\to 1$ in (4.7) then we obtain the asymptotic regret of the one-by-one purely sequential procedure, which is $\omega=(9/4)\gamma^2-(\beta-3)+o(1)$, as $m\to\infty$. Moreover, the asymptotic regret in (4.7) is a non-vanishing quantity that is independent of m and A and which takes negative values when $\beta>(1/4)(8\delta+1)\gamma^2\delta^{-1}+3$.

Obviously the asymptotic regret for estimating the mean of the one parameter exponential family (4.7) depends on the skewness and the kurtosis of the underlying distribution and tends to zero in the case of the normal distribution. In contrast the asymptotic regret when the underlying distribution is unspecified (Theorem 4.2.1*(ii)*) depends only on the kurtosis of the underlying distribution and tends to a non-vanishing but finite quantity in the case of the normal distribution.

Finally, extra knowledge regarding the structure of the underlying distribution undoubtedly will enhance our knowledge regarding the performance of the triple sampling as justified by the class of the one parameter exponential family. In other words, knowing the family of the underlying distribution will determine the exact values of the skewness and kurtosis and consequently will provide a precise measure of the risk, while if the underlying distribution is not known then we have to estimate the skewness and kurtosis and this will ensure more cost (cost of ignorance).

To illustrate the performance of the triple sampling procedure when the underlying distribution is analytically known, we give the following example.

Example

Let X be an exponential random variable with mean μ , then the probability density function of X is

$$f(x,\mu)=\frac{1}{\mu}\exp(-x/\mu),\ \ x>0 \text{ and } 0<\mu<\infty .$$

It can be shown easily from Lemma 4.3.2 and Lemma 4.3.3 that the triple sampling asymptotic characteristics under the one-parameter exponential family with normal stopping rule are as follows

$$E(N)=n^{*}+(1/2)-\delta^{-1}+o(1),\ Var(N)=n^{*}\delta^{-1}+o(\lambda),$$

$$E(\bar{X}_{N})=\mu-(\sqrt{\theta}/n^{*})+o(\lambda^{-1}),$$

and

$$Var(\bar{X}_{N})=(\theta/n^{*})+\theta(2+1.5\delta)\delta^{-1}(n^{*})^{-2}+o(\lambda^{-2}).$$

However, from Theorem 3.2.5 and Theorem 3.2.3 the asymptotic characteristics under the normal stopping rule $n^{*}=\lambda\sqrt{\theta}$ are

$$E(N)=n^{*}-3\delta^{-1}+E(\varepsilon_{N_1})+o(1),Var(N)=2n^{*}\delta^{-1}+o(\lambda),$$

$$E(\bar{X}_{N})=\mu-(\sqrt{\theta}/n^{*})+o(\lambda^{-1}),$$

and

$$Var(\bar{X}_{N})=(\theta/n^{*})+\theta(5-6\delta)\delta^{-1}(n^{*})^{-2}+o(\lambda^{-2}).$$

To illustrate this comparison, consider the case $\mu=2$. It follows that $\theta=4,\gamma=2$ and $\beta=9$.

By direct substitution in the above formulae with $\delta=0.5$ we obtain the results in Table 4.1 below.

Asymptotic characteristics of the triple sampling under normal stopping rule	Asymptotic characteristics of the triple sampling under the one parameter exponential family
$E(N) = n^* - 6 + E(\varepsilon_{N_1}) + o(1)$	$E(N) = n^* - 1.5 + o(1)$
$Var(N) = 4n^* + o(\lambda)$	$Var(N) = 2n^* + o(\lambda)$
$E(\bar{X}_N) = 2 - 2/n^* + o(\lambda^{-1})$	$E(\bar{X}_N) = 2 - 2/n^* + o(\lambda^{-1})$
$Var(\bar{X}_N) = 4/n^* + 16/n^{*2} + o(\lambda^{-2})$	$Var(\bar{X}_N) = 4/n^* + 22/n^{*2} + o(\lambda^{-2})$
$\omega = -2 + E(\varepsilon_{N_1}) + o(1)$	$\omega = 4 + o(1)$

Table 4.1: Asymptotic characteristics of the triple sampling procedure in two cases: the underlying distribution is unknown and the underlying distribution known to be exponential distribution with mean two

Table 4.1 shows the difference between the asymptotic characterstics of the triple sampling procedure under two cases: firstly, when the underlying distribution is analytically unknown and secondly when the underlying distribution is known to be a member of the class of a one-parameter exponential family. It is clear from Table 4.1 that $E(N)$ in the second case is less biased than in the first case, moreover the variance of the stopping sample size N in the second case is less than that in the first case. Thus the performance of the actual sample size N is better in the case where the family of the underlying distribution is analytically known. For the sequential estimator $\bar{X}_N$ of μ, the bias is the same in both cases, while the variance of $\bar{X}_N$ is slightly less than in the first.

The regrets in both cases are bounded by a non vanishing quantity. Under the first case we have a negative regret while under the second case the regret is positive. The reason behind this is as follows: under the first case the stopping rule depends on the nuisance parameter and the estimate of the nuisance parameter depends on the estimate of the mean, but since their distributions are dependent, then this creates asymptotically negative regret (see Martinseck, 1988 and Takada, 1992). Under the second case the stopping rule depends only on the estimate of the mean and is independent of the nuisance parameter (variance) and this causes a bounded positive regret (see Hamdy *et al.*, 1989).

Collectively, if we derive a triple sampling point estimation procedure under the exponential distribution, the results would be consistent with the asymptotic theory in terms of $E(N) \to n^*$ as $d \to 0$, $E(\bar{X}_N) \to \mu$ as $n^* \to \infty$ and ω is bounded as $A \to \infty$. However, if we are sampling from the exponential distribution under the normal theory, noticeable deviations from the true values are present for small values of n^*. This indicates that the triple sampling procedure is sensitive to the underlying distribution.

4.4 Other continuous classes of distributions

In section 4.3 we considered the effect of restricting the class of underlying distributions on the triple sampling point estimation. We now return to the case where no assumptions are made about the underlying distribution except that the first six moments are finite.

In practice we might use Hall's triple sampling scheme if we believed that the underlying distribution was in some sense close to the normal. For example, perhaps a central limit theorem argument may lead us to believe that our data will be approximately normally distributed. In the classical robustness literature various authors have argued that approximately normal data are often somewhat fatter tailed than normal in practice. For example, the location-scale version of the t distribution (Hampel, 1968), the symmetric contaminated normal distribution (Tukey, 1960) and (Huber, 1964) have been used to model this situation. Moreover, an asymmetric contaminated normal distribution is often used to model basically normal data with outliers (Barnett and Lewis, 1994).

In this section we give the results for Hall's triple sampling method when the unknown underlying distribution is in fact $t(r)$, Huber's least favourable or contaminated normal (symmetric and asymmetric).

4.4.1. The t distribution

Let T be a $t(r)$ random variable with r degrees of freedom, where $r > 4$. Then the first four moments of T are:

$E(T)=0$, $Var(T)=r/(r-2)$, $\gamma=0$ and $\beta=3(r-2)/(r-4)$.

By direct substitution in Theorem 3.2.2.1, Theorem 3.2.2.2 and Theorem 4.2.1, we have

$$E(N)=n^{*}-0.75\delta^{-1}\frac{(r-1)}{(r-4)}+E\left(\varepsilon_{N_1}\right)+o(1),$$

$$Var(N)=n^{*}(2\delta)^{-1}(r-1)/(r-4)+o(\lambda),$$

$$E\left(\bar{X}_N\right)=o\left(\lambda^{-1}\right),$$

$$Var\left(\bar{X}_N\right)=\frac{r}{n^{*}(r-2)}-\frac{6r}{\left(n^{*}\right)^{2}(r-2)(r-4)}+\frac{5r(r-1)\delta^{-1}}{4\left(n^{*}\right)^{2}(r-2)(r-4)}+o\left(\lambda^{-2}\right),$$

$$\omega=\frac{-6C}{(r-4)}+\frac{1}{2}\left(\frac{r-1}{r-4}\right)C\delta^{-1}+C\,E\left(\varepsilon_{N_1}\right)+o(1).$$

Clearly, the asymptotic behaviour of the triple sampling procedure under the t distribution depends mainly on the degrees of freedom r and, as $r\to\infty$, the corresponding normal results are obtained:

$E(N) = n^* - 0.75\delta^{-1} + 0.5 + o(1)$,

$Var(N) = n^*(2\delta)^{-1} + o(\lambda)$,

$E(\bar{X}_N) = o(\lambda^{-1})$,

$Var(\bar{X}_N) = \theta(n^*)^{-1} + 1.25\theta\delta^{-1}(n^*)^{-2} + o(\lambda^{-2})$,

$\omega = C(2\delta)^{-1} + 0.5C + o(1)$.

Let $E_T(\cdot)$ and $Var_T(\cdot)$ denotes the mean and variance under the t distribution and let $E_N(\cdot)$ and $Var_N(\cdot)$ be the mean and variance under the normal distribution. Then from the above characterstics we have the following:

1. $E_T(N) < E_N(N)$, for all $r > 4$, because $((r-1)/(r-4)) > 1, \forall r > 4$ with equality attained only as $r \to \infty$. This indicates that the estimators of the actual sample size N under the t distribution will on average be less than under the normal distribution (earlier stopping than normal), and as r increases they nearly attain the same behaviour on average.

2. $Var_T(N) > Var_N(N)$. As r increases the ratio tends to one as $r \to \infty$.

3. $E_T(\bar{X}_N) = E_N(\bar{X}_N)$.

4. To compare $Var_T(\bar{X}_N)$ and $Var_N(\bar{X}_N)$ directly, let $\theta = r/(r-2)$ in order to match the variances of the underlying distributions. If $\delta = 5/8$, then clearly $Var_T(\bar{X}_N) = Var_N(\bar{X}_N)$, while if $\delta < 5/8$ we have that $Var_T(\bar{X}_N) > Var_N(\bar{X}_N)$ and if $\delta > 5/8$, then $Var_T(\bar{X}_N) < Var_N(\bar{X}_N)$. Of course, in the limiting case, as $r \to \infty$, then $\theta \to 1$ and in this case $Var_T(\bar{X}_N) = Var_N(\bar{X}_N)$.

Thus, the triple sampling procedure is sensitive to the underlying distribution and hence is not robust to departures from normality.

4.4.2. Contaminated normal distribution

Let X be a random variable that has a contaminated normal distribution with distribution function

$$F = p F_N(\mu_1, \sigma_1^2) + (1-p) F_N(\mu_2, \sigma_2^2),$$

where $F_N(\mu_1, \sigma_1^2)$ and $F_N(\mu_2, \sigma_2^2)$ are the normal distribution functions at μ_1, σ_1^2 and μ_2, σ_2^2 respectively and $p \in (0,1)$.

In the classical robustness context p is taken to be relatively large, so that the first component represents “good” observations, whereas the second component represents an outlier generating mechanism.

We will consider two cases to illustrate the idea of robustness.

Case *(i)*: $\mu_1 = \mu_2 = \mu$ and $\sigma_1^2 \neq \sigma_2^2$

The associated variance and kurtosis are respectively,

$$\sigma_F^2 = \sigma_2^2 + p\left(\sigma_1^2 - \sigma_2^2\right) \text{ and } \beta_F = \left(3p\left(\sigma_1^4 - \sigma_2^4\right) + 3\sigma_2^4\right)\left(\sigma_2^2 + p\left(\sigma_1^2 - \sigma_2^2\right)\right)^{-2}$$

For example, if $\mu = 0$, $\sigma_1 = 1$ and $\sigma_2 = 3$, then the values of the variance and the kurtosis for selected values of p are shown in Table 4.2.

p	σ_F^2	β_F
0.900	1.800	8.33333
0.950	1.400	7.65306
0.990	1.080	4.62963
0.999	1.008	3.18877

Table 4.2: The variance and kurtosis of the contaminated normal distribution with equal means but different variances at selected values of p

By substituting in Theorem 3.2.2.1, Theorem 3.2.2.2 and Theorem 4.2.1, we have

$$E(N) = n^* - (3/8)(\beta_F - 1)\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1),$$

$$Var(N) = n^*\left(\frac{3p\left(\sigma_1^4 - \sigma_2^4\right) + 3\sigma_2^4}{\left(p\left(\sigma_1^2 - \sigma_2^2\right) + \sigma_2^2\right)^2} - 1\right)(4\delta)^{-1} + o(\lambda),$$

$$E\left(\bar{X}_N\right) = \mu_F = \mu + o\left(\lambda^{-1}\right),$$

and

$$Var\left(\bar{X}_N\right) = (8\delta)^{-1}\left(n^*\right)^{-2}\left(\xi_1/\sigma_F^2\right) + o\left(\lambda^{-2}\right),$$

where

$$\xi_1 = \begin{Bmatrix} 8\left(\sigma_1^2-\sigma_2^2\right)^2\left(-5/8+\left(n^*+3\right)\delta\right)p^2+16\left(\sigma_1^2-\sigma_2^2\right)p\left(5/16+\left(n^*+3/2\right)\delta\sigma_2^2-\left(3/2\right)\left(\delta-5/8\right)\sigma_1^2\right) \\ +\sigma_2^4\left(8n^*\delta+10\right) \end{Bmatrix}.$$

Also, the asymptotic regret will be completely specified by the kurtosis of the underlying distribution, $\omega(A) = -C\left(\beta_F-3\right)+(1/4)\left(\beta_F-1\right)\left(C/\delta\right)+C\,E\left(\varepsilon_{N_1}\right)+o(1)$.

Obviously, $Var\left(\bar{X}_N\right)\to 0$, as $n^*\to\infty$ and if $\sigma_1=\sigma_2=\sigma$, then we obtain the case of the normal distribution, where

$$E(N) = n^* - 0.75\delta^{-1} + 0.5 + o(1).$$

$$Var(N) = n^*\left(2\delta\right)^{-1} + o(\lambda).$$

$$E\left(\bar{X}_N\right) = \mu + o\left(\lambda^{-1}\right).$$

$$Var\left(\bar{X}_N\right) = \sigma^2\left(n^*\right)^{-1} + 1.25\sigma^2\delta^{-1}\left(n^*\right)^{-2} + o\left(\lambda^{-2}\right).$$

The asymptotic regret is $\omega = \left(2\delta\right)^{-1}C + 2^{-1}C + o(1)$.

To illustrate the above equations and show the effect of increasing p on the performance of the actual sample size N, the estimator $\bar{X}_N$ and the regret ω see Tables 4.3, 4.4 and 4.5, which show the asymptotic characterstics of the triple sampling scheme under the contaminated normal distribution with equal means but different variances, $\mu=0$, $\sigma_1=1, \sigma_2=3$ for $p=0.9, 0.99$ and 0.999. We see that for a specific value of p, $\left|E(N)-n^*\right|$ is fixed for all values of n^*. At $p=0.9$ the absolute difference is 5, while at $p=0.99$ and 0.999 the absolute differences are 2.22 and 1.14 respectively. The standard deviation of N increases with n^*. It is clear that the mean of $\bar{X}_N$ is asymptotically zero while its standard deviation decreases as n^* increases. The regret values for $p=0.9, 0.99$ and 0.999 are -1.166665, 0.685185 and 1.405615 respectively.

n^*	$E(N)$	$s.d.(N)$	$E(\bar{X}_N)$	$s.d.(\bar{X}_N)$	ω
24	19	9.3808315	0	0.395504	-1.166665
43	38	12.556539	0	0.259279	-1.166665
61	56	14.955490	0	0.205213	-1.166665
76	71	16.693312	0	0.178337	-1.166665
96	91	18.761663	0	0.154399	-1.166665
125	120	21.408721	0	0.131915	-1.166665
171	166	25.039968	0	0.110137	-1.166665
246	241	30.033315	0	0.089956	-1.166665
500	495	42.817442	0	0.061543	-1.166665

Table 4.3: Asymptotic characteristics of the triple sampling scheme with underlying contaminated normal distribution with, $\mu = 0$, $\sigma_1 = 1$ and $\sigma_2 = 3$; $p = 0.9$, $\delta = 0.5$ and $E(\varepsilon_{N_1}) \cong 0.5$.

n^*	$E(N)$	$s.d.(N)$	$E(\bar{X}_N)$	$s.d.(\bar{X}_N)$	ω
24	21.78	6.5996633	0	0.421623	0.685185
43	40.78	8.8338574	0	0.257829	0.685185
61	58.78	10.521583	0	0.195593	0.685185
76	73.78	11.744187	0	0.165682	0.685185
96	93.78	13.199327	0	0.139814	0.685185
125	122.78	15.061602	0	0.116337	0.685185
171	168.78	17.616280	0	0.094504	0.685185
246	243.78	21.129232	0	0.075193	0.685185
500	497.78	30.123204	0	0.049658	0.685185

Table 4.4: Asymptotic characteristics of the triple sampling scheme with underlying contaminated normal distribution with $\mu = 0$, $\sigma_1 = 1$ and $\sigma_2 = 3$; $p = 0.99$, $\delta = 0.5$ and $E(\varepsilon_{N_1}) \cong 0.5$.

n^*	$E(N)$	$s.d.(N)$	$E(\bar{X}_N)$	$s.d.(\bar{X}_N)$	ω
24	22.86	5.1249689	0	0.429076	1.405615
43	41.86	6.8599325	0	0.260213	1.405615
61	59.86	8.1705357	0	0.196271	1.405615
76	74.86	9.1199490	0	0.165633	1.405615
96	94.86	10.249938	0	0.139218	1.405615
125	123.86	11.696088	0	0.115337	1.405615
171	169.86	13.679924	0	0.093242	1.405615
246	244.86	16.407906	0	0.073825	1.405615
500	498.86	23.392176	0	0.048409	1.405615

Table 4.5: Asymptotic characteristics of the triple sampling scheme with underlying contaminated normal distribution with $\mu = 0$, $\sigma_1 = 1$ and $\sigma_2 = 3$; $p = 0.999$, $\delta = 0.5$ and $E(\varepsilon_{N_1}) \cong 0.5$.

Conclusions

1. As p increases $E(N)$ increases and the absolute value between $E(N)$ and n^* decreases while $s.d.(N)$ decreases from $n^* \geq 61$.

2. As p increases, the asymptotic mean is zero while the $s.d.(\bar{X}_N)$ decreases from $n^* \geq 76$.

3 As p increases, so does the regret.

Case *(ii)*: $\mu_1 \neq \mu_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$

It can be shown that the values of the skewness and kurtosis are respectively,

$$\gamma_F = \left(2p(p-1)(p-0.5)(\mu_1-\mu_2)^3\right)\left(p(\mu_1-\mu_2)^2 - p^2(\mu_1-\mu_2)^2 + \sigma^2\right)^{-3/2},$$

and

$$\beta_F = \frac{\left(-3p^4+6p^3-4p^2+p\right)(\mu_1-\mu_2)^4 - 6\sigma^2 p(p-1)(\mu_1-\mu_2)^2 + 3\sigma^4}{\left(p(\mu_1-\mu_2)^2 - p^2(\mu_1-\mu_2)^2 + \sigma^2\right)^2}.$$

For example, if $\mu_1 = 0$, $\mu_2 = 3$ and $\sigma = 1$, then the values of the skewness and kurtosis for different values of p are shown in Table 4.6.

p	σ_F^2	γ_F	β_F
0.900	1.81	0.798323	4.023595
0.950	1.4275	0.676762	4.349996
0.990	1.0891	0.230475	3.635901
0.999	1.008991	0.026560	3.079007

Table 4.6: The variance, skewness and kurtosis of the contaminated normal ($\mu_1 = 0, \mu_2 = 3, \sigma = 1$) distribution at selected values of p.

The forms of $E(N)$, $Var(N)$ and the asymptotic regret may be found as before. For simplicity,

Let $M_1 = (\mu_1-\mu_2)^2$. Then

$$E(\bar{X}_N) = (p\mu_1 + (1-p)\mu_2) + \frac{p(p-0.5)(p-1)M_1^{3/2}}{\left(p^2M_1 - pM_1 - \sigma^2\right)n^*} + o\left(\lambda^{-1}\right),$$

$$Var(\bar{X}_N) = (8\delta)^{-1}(n^*)^{-2}\xi_2/\sigma_F^2 + o\left(\lambda^{-2}\right),$$

where

$$\sigma_F^2 = p\left(\mu_1 - \mu_2\right)^2 - p^2\left(\mu_1 - \mu_2\right)^2 + \sigma^2 ,$$

and

$$\xi_2 = 8p\left(p-1\right)\left(\left(-5/2+\left(n^*+6\right)\delta\right)p^2+\left(5/2-\left(n^*+6\right)\delta\right)p+\delta-5/8\right)M_1^2$$
$$-16p\sigma^2\left(p-1\right)\left(\delta n^*+5/4\right)M_1+2\sigma^4\left(4\delta n^*+5\right).$$

Moreover,

$$E\left(N\right)=\left(8\delta\right)^{-1}\xi_3/\sigma_F^4+E\left(\varepsilon_{N_1}\right)+o\left(1\right),$$

where

$$\xi_3 = 8p\left(p-1\right)\left(3/8+\left(3/2+\delta n^*\right)p^2-\left(\left(3/2+\delta n^*\right)\right)p\right)M_1^2-16p\left(p-1\right)\sigma^2\left(\delta n^*-3/4\right)M_1$$
$$-2\sigma^4\left(3-4\delta n^*\right).$$

$$Var\left(N\right)=n^*\delta^{-1}\xi_4/\sigma_F^4+o\left(\lambda\right),$$

where

$$\xi_4 = -p\left(p-1/2\right)^2\left(p-1\right)M_1^2-p\left(p-1\right)\sigma^2 M_1+\left(1/2\right)\sigma^4$$

and the asymptotic regret is $\omega = -C\left(\beta_F-3\right)+(1/4)\left(\beta_F-1\right)\left(C/\delta\right)+C\,E\left(\varepsilon_{N_1}\right)+o\left(1\right)$.

To illustrate the above equations and show the effect of increasing p on the behaviour of N, $\bar{X}_N$ and the regret, see Tables 4.7, 4.8 and 4.9, which show the asymptotic characteristics of the triple sampling scheme under the contaminated normal distribution with different means but equal variances, $\mu_1 = 0$, $\mu_2 = 3$, $\sigma = 1$ while $p = 0.9$, 0.99 and 0.999.

We see that at $p = 0.9$, $\left|E\left(N\right)-n^*\right| = 2.5177$ while for $p = 0.99$ and 0.999 the absolute value between $E\left(N\right)$ and n^* are 1.4769 and 1.0593 respectively and for all values of n^*. The standard deviation of N increases with n^*. For all values of p, $\bar{X}_N$ is a biased estimator of μ and the amount of bias decreases as n^* increases. Moreover, the standard deviation of $\bar{X}_N$ decreases as n^* increases. The regret values are 0.9882025, 1.1820495 and 1.4604965 respectively.

n^*	$E(N)$	$s.d.(N)$	$E(\bar{X}_N)$	$s.d.(\bar{X}_N)$	ω
24	21.4823	6.02354894	0.32237570	0.2293707	0.9882025
43	40.4823	8.06271017	0.31248875	0.1768155	0.9882025
61	58.4823	9.60310634	0.30880355	0.1501248	0.9882025
76	73.4823	10.71898385	0.30706601	0.1351967	0.9882025
96	93.4823	12.04709783	0.30559392	0.1208176	0.9882025
125	122.4823	13.74680674	0.30429613	0.1062835	0.9882025
171	168.4823	16.07847578	0.30314045	0.0911772	0.9882025
246	243.4823	19.28476603	0.30218299	0.0762292	0.9882025
500	497.4823	27.49361348	0.301074033	0.0536403	0.9882025

Table 4.7: Asymptotic characteristics of the triple sampling scheme with underlying contaminated normal distribution with $\mu_1 = 0$, $\mu_2 = 3$ and $\sigma = 1$; $p = 0.9$, $\delta = 0.5$ and $E(\varepsilon_{N_1}) \cong 0.5$.

n^*	$E(N)$	$s.d.(N)$	$E(\bar{X}_N)$	$s.d.(\bar{X}_N)$	ω
24	22.5231	5.62412792	0.03501090	0.2172883	1.1820495
43	41.5231	7.52807257	0.03279678	0.1607005	1.1820495
61	59.5231	8.96632521	0.03197150	0.1344086	1.1820495
76	74.5231	10.0082090	0.03158239	0.1201983	1.1820495
96	94.5231	11.2482558	0.03125273	0.1067827	1.1820495
125	123.5231	12.8352572	0.03096209	0.0934523	1.1820495
171	169.5231	15.0123135	0.03070330	0.0798029	1.1820495
246	244.5231	18.0059949	0.03044889	0.0664678	1.1820495
500	498.5231	25.6705144	0.030240523	0.0465677	1.1820495

Table 4.8: Asymptotic characteristics of the triple sampling scheme with underlying contaminated normal distribution with $\mu_1 = 0$, $\mu_2 = 3$ and $\sigma = 1$; $p = 0.99$, $\delta = 0.5$ and $E(\varepsilon_{N_1}) \cong 0.5$.

n^*	$E(N)$	$s.d.(N)$	$E(\bar{X}_N)$	$s.d.(\bar{X}_N)$	ω
24	22.9407	4.994805561	0.00355582	0.2147986	1.4604965
43	41.9407	6.685704748	0.00331022	0.1572909	1.4604965
61	59.9407	7.963021407	0.00321868	0.1310506	1.4604965
76	74.9407	8.888321639	0.00317552	0.1169789	1.4604965
96	94.9407	9.989611122	0.00313895	0.1037582	1.4604965
125	123.9407	11.39903198	0.00310672	0.0906777	1.4604965
171	169.9407	13.33248246	0.00307801	0.0773359	1.4604965
246	244.9407	15.99118028	0.00305423	0.0643451	1.4604965
500	498.9407	22.79806397	0.00302668	0.0450252	1.4604965

Table 4.9: Asymptotic characteristics of the triple sampling scheme with underlying contaminated normal distribution with $\mu_1 = 0$, $\mu_2 = 3$ and $\sigma = 1$; $p = 0.999$, $\delta = 0.5$ and $E(\varepsilon_{N_1}) \cong 0.5$.

Conclusion

1. As p increases, the difference between $E(N)$ and n^* decreases and the standard deviation of N also decreases.
2. For all values of p, $\bar{X}_N$ is a biased estimator of μ but the bias decreases as n^* increases. The standard deviation also decreases as n^* increases.
3. As p increases the regret also increases.

In the case with $\mu_1 = \mu_2$, we return to the normal distribution results.

4.4.3. Huber's least favourable distribution

Huber's least favourable distribution has density function in the form

$$f(t) = \begin{cases} \dfrac{a}{\sqrt{2\pi}} exp\left(-t^2/2\right) & ,|t| \le k \\ \dfrac{a}{\sqrt{2\pi}} exp\left(k^2/2 - k|t|\right) & ,|t| \ge k \end{cases},$$

where $a = \dfrac{\sqrt{2\pi}k}{2\,Exp\left(-k^2/2\right) + \sqrt{2\pi}k\left(2\Phi(k)-1\right)}$.

It is easy to show that the variance and kurtosis are respectively,

$$\sigma_t^2 = \frac{w_1 + k^2 w_2}{k^2 w_3} \text{ and } \beta_h = \frac{w_3\left(3k^4 w_2 + w_4\right)}{\left(w_1 + k^2 w_2\right)^2},$$

where

$$w_1 = 2\left(1+k^2\right)exp\left(-k^2/2\right),\ w_2 = \sqrt{2\pi}\,k\left(\Phi(k)-1\right),\ w_3 = exp\left(-k^2/2\right) + w_2 \text{ and}$$
$$w_4 = \left(k^6 + 12k^4 + 24k^2 + 24\right)exp\left(-k^2/2\right).$$

The triple sampling asymptotic results are as follows:

$$E(N) = n^* + (8\delta)^{-1}\xi_1^*\left(w_1 + k^2 w_2\right)^{-2} + E\left(\varepsilon_{N_1}\right) + o(1),$$

where

$$\xi_1^* = 3k^4 w_2^2 + 3k^2\left(-3w_3k^2 + 2w_1\right)w_2 - 3w_3 w_4 + 3w_1^2.$$

$$Var\left(N\right)=(4\delta)^{-1}n^{*}\xi_{2}^{*}\left(w_{1}+k^{2}w_{2}\right)^{-2}+o\left(\lambda\right),$$

where

$$\xi_{2}^{*}=3w_{2}w_{3}k^{4}+w_{3}w_{4}-w_{1}^{2}-2w_{1}w_{2}k^{2}-w_{2}^{2}k^{4}.$$

$$E\left(\bar{X}_{N}\right)=o\left(\lambda^{-1}\right),$$

$$Var\left(\bar{X}_{N}\right)=\left(8\delta\right)^{-1}\left(n^{*}\right)^{-2}\left(\xi_{3}^{*}/\left(\left(w_{1}+k^{2}w_{2}\right)k^{2}w_{3}\right)\right)+o\left(\lambda^{-2}\right),$$

where

$$\xi_{3}^{*}=8w_{2}\left(\left(\left(3+n\right)w_{2}-3w_{3}\right)\delta+(15/8)w_{3}-(5/8)w_{2}\right)k^{4}$$
$$+16w_{1}w_{2}\left(-5/8+\left(3+n\right)\delta\right)k^{2}+\left(\left(8n+24\right)w_{1}^{2}-8w_{3}w_{4}\right)\delta$$
$$+5w_{3}w_{4}-5w_{1}^{2}$$

and the asymptotic regret is $\omega=-C\left(\beta_{h}-3\right)+(1/4)\left(\beta_{h}-1\right)\left(C/\delta\right)+C\,E\left(\varepsilon_{N_{1}}\right)+o(1)$.

At $k=0$ the distribution is the double exponential (or Laplace), which has $\beta=6$, while as $k\rightarrow\infty$ the normal distribution is obtained. Commonly used values of k in robustness studies $k=1.0,1.5$ and 2 for which the kurtosis values are respectively $\beta_{h}=5.371428,4.303949$ and 3.511003.

k	a	β_h
0	0	6.0
1.0	0.857176	5.371428
1.5	0.962393	4.303949
2.0	0.991581	3.511003

Table 4.10: The value of a and the kurtosis of Huber's least favourable distribution for selected values of k.

Notes

1. From the above we deduce that the triple sampling procedure is sensitive to the underlying distribution. However, the effect reduces as the skewness and kurtosis of the underlying distribution approach zero and three respectively.

2. We also see that if the stopping rule depends on the estimates of the parameter of interest, then an early stopping is expected. This may result in a negative regret (see Martinsek, 1988 and AlMahmeed *et al.*, 1998). Also, if the stopping rule depends on an estimate of a certain nuisance parameter, which is independent of the targeted estimate, the procedure naturally terminates with bounded regret (see Hamdy, 1989).

Chapter V

Simulation Results for Point Estimation of the Population Mean

In this chapter we use simulation to investigate the finite sample properties of the triple sampling procedure $(2.2)-(2.3)$ and to compare them to the asymptotic results obtained in the previous chapters.

5.1 Experimental setup

A series of Monte Carlo studies was carried out in order to investigate the finite sample size performance of the triple sampling point estimation sampling procedure $(2.2)-(2.3)$ under the squared error loss function (4.1) and to compare these results with the asymptotic results.

First we allowed various aspects of the triple sampling scheme to vary: $m = 5, 15, 20$; $\delta = 0.3, 0.5, 0.8$ and $n^* = 24$, 43, 61, 76, 96, 125, 171, 246 and 500. These values of n^* are the same as those used by Hall (1981) and represent small, moderate and large optimal sample sizes.

In addition various underlying distributions were used to cover symmetric and skewed distributions and light and heavy tailed distributions: normal, uniform, $t(r)$, beta and exponential. For each experimental situation $50,000$ replicate samples were used (see Costanza *et al*., 1995).

The following steps explain how we obtain the simulation results. For the i th sample generated for a particular combination of m, δ and distribution:

1. Take an initial sample of size m (say, $X_{1,i}, \ldots, X_{m,i}$).
2. Compute the sample mean and sample variance for the pilot sample.
3. Apply the triple sampling procedure, as presented in $(2.2)-(2.3)$ to determine the stopping sample size at this iteration whether in the first or second stage (say, N_i^*).
4. Record the resultant sample size and the sample mean $\left(N_i^*, \bar{X}_i^*\right)$.

Hence, for each experimental combination we have two vectors of size 50,000 as follows:

Vector I contains all the stopping sizes, say $N_1^*, N_2^*, ..., N_{50,000}^*$,

and

Vector II contains all the sample means, $\overline{X}_1^*, \overline{X}_2^*, \ldots, \overline{X}_{50,000}^*$.

Here N_i^* may be thought of as the estimate of n^* and $\overline{X}_i^*$ is the estimate of μ at sample *i*. Let

$$\overline{N} = \sum_{i=1}^{50000} \left(N_i^* / 50,000 \right) \text{ and } \widehat{\mu} = \overline{X} = \sum_{i=1}^{50000} \left(\overline{X}_i^* / 50,000 \right),$$

where $\overline{N}$ and $\overline{X}$ are respectively the estimated mean sample size and the estimated mean of the estimator of the population mean across replicates. Thus $\widehat{\mu} = \overline{X}$ may be regarded as an estimate of the expected value of the estimator of the population mean μ.

The standard errors are

$$S.E.\left(\widehat{\mu}\right) = s_1 \Big/ \sqrt{50,000}\text{, where } s_1 = \sqrt{\sum_{i=1}^{50000} \left(\overline{X}_i - \widehat{\mu} \right)^2 \Big/ 49,999}.$$

and

$$S.E.\left(\overline{N}\right) = s_2 \Big/ \sqrt{50,000}\text{, where } s_2 = \sqrt{\sum_{i=1}^{50000} (\overline{N}_i - \overline{N})^2 \Big/ 49,999}.$$

To calculate the estimated regret, we proceed as follows. First for simplicity take $C = 1$ and from (4.3), $A = \left(n^*\right)^2 \big/ \theta$. Then $\widehat{\omega}$ is the point estimate for the asymptotic regret, which is the difference between the sequential risk and the optimal risk. The optimal risk is $2Cn^*$. Secondly, we calculate the estimated variance of $\overline{X}_N$ which can be obtained easily from the simulation, $\widehat{\text{var}}\left(\overline{X}_N\right) = s_1^2$. Finally, the estimated regret is obtained as $\widehat{\omega} = As_1^2 - \overline{N}$.

Although our Theorems are valid for all $\delta \in (0,1)$ we shall concentrate mostly on the case $\delta = 0.5$ for two reasons. First, this has been recommended previously for practical reasons (see Hall, 1981 and Mukhopadhyay and de Silva, 2009). Secondly, we shall see later that this is indeed a good choice.

The featured underlying distributions are: standard normal, standard uniform, the t distribution with $r = 5, 25, 50$ and 100 degrees of freedom, beta$(2,3)$ and exponential with mean one. These cover a variety of distributional shapes. Results for other distributions have been omitted but are available from the author.

We proceed in this chapter by studying the behaviour of $\overline{X}_N$, the actual sample size N and the regret at $\delta = 0.5$ and $m = 5$, 15 and 20 from two sides; one from the simulation view and the other from the asymptotic view obtained from Theorems 3.2.2.1, 3.2.2.2 and 4.2.1 for the above underlying distributions. Moreover, we show the effect of increasing δ on the performance of the above estimators at $m = 15$.

5.2 The behaviour of the estimator of the population mean

From Theorem 3.2.2.1, the mean and variance of the estimator $\bar{X}_N$ under the normal optimal stopping rule $n^* = \lambda\sqrt{\theta}$ are

$$(5.1)\ E\left(\bar{X}_N\right) = \mu - \gamma\sqrt{\theta}\left(2n^*\right)^{-1} + o\left(\lambda^{-1}\right),$$

and

$$(5.2)\ Var\left(\bar{X}_N\right) = \theta\left(n^*\right)^{-1} - \theta\left(n^*\right)^{-2}\left\{(\beta-3)-(5/8)(\beta-1)\delta^{-1}\right\} + o\left(\lambda^{-2}\right).$$

The estimator $\bar{X}_N$ is a biased estimator of μ when the underlying distribution is skewed. The variance of the estimator depends on the variance and the kurtosis of the underlying distribution as well as the design factor δ. Clearly, as $n^* \to \infty$ the asymptotic mean and variance of the estimator $\bar{X}_N$ are respectively μ and zero.

In the following subsections we discuss the simulation results in comparison with the asymptotic results obtained from equations $(5.1)-(5.2)$ for the underlying distributions mentioned above. Each table includes $\mu_s(m)$ and $ssd(m)$ (the simulated estimate of $E\left(\bar{X}_N\right)$ and its standard deviation at a specific value of m, where $m=$5, 15 and 20) and also $E\left(\bar{X}_N\right)$ and $sd\left(\bar{X}_N\right)$, the asymptotic mean and standard deviation of $\bar{X}_N$. Moreover, we discuss the effect of increasing δ on the performance of the estimator $\bar{X}_N$ through the simulation results arranged in separate tables. Each table includes $\mu_s(\delta)$ and $ssd(\delta)$: the simulated estimates of $E\left(\bar{X}_N\right)$ and its standard deviation at a specific value of δ, where $\delta=$0.3, 0.5 and 0.8.

5.2.1 Standard normal distribution

From $(5.1)-(5.2)$ the mean and variance of $\bar{X}_N$ under the above optimal fixed sample size are

$$E\left(\bar{X}_N\right) = o\left(\lambda^{-1}\right) \text{ and } Var\left(\bar{X}_N\right) = \left(1/n^*\right) + (5/4)\delta^{-1}\left(n^*\right)^{-2} + o\left(\lambda^{-2}\right).$$

The asymptotic mean and variance of $\bar{X}_N$ under the standard normal distribution are both zero as $n^* \to \infty$.

To illustrate the above equations and compare them with the simulation results, Table 5.1 shows the behaviour of the estimator $\bar{X}_N$ from two sides: one from the simulation view at $\delta = 0.5$ while m increases from $m=$5, 15 to 20 and the second from the asymptotic view. From the simulation view, we see that for all values of m, $\bar{X}_N$ is essentially unbiased. Moreover, the standard deviation decreases as n^* increases. Asymptotically, the mean of $\bar{X}_N$ is zero and its standard deviation decreases as n^* increases. Clearly there is good agreement between the simulated and asymptotic means. The asymptotic standard deviations are underestimates, but are in good agreement with the simulated values when n^* is 61 or more.

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E(\bar{X}_N)$	$sd(\bar{X}_N)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	0.00115	0.00082	0.00176	0	0.214492	0.231433	0.252899	0.223383
43	0.00084	0.00069	-0.00095	0	0.156869	0.162339	0.162115	0.180003
61	0.00068	-0.00047	-0.00026	0	0.130634	0.132375	0.130586	0.131928
76	0.00031	-0.00083	0.00021	0	0.116579	0.117394	0.116276	0.117170
96	0.00003	-0.00043	-0.00031	0	0.103382	0.103306	0.103306	0.103530
125	0.00033	0.00011	-0.00003	0	0.090333	0.090561	0.090114	0.090337
171	-0.00027	0.00052	0.00013	0	0.077029	0.077368	0.077144	0.077144
246	0.00021	0.00035	-0.00007	0	0.064081	0.064175	0.063952	0.064175
500	-0.00005	-0.00020	0.00009	0	0.044833	0.044945	0.044721	0.044721

Table 5.1: Comparison between the simulated estimates of $E(\bar{X}_N)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under $N(0,1)$ as n^* increases and at $m = 5,15,20$ and $\delta = 0.5$

To show the impact of increasing δ on the performance of the estimator $\bar{X}_N$ at $m = 15$ Table 5.2 shows the simulated estimates the mean and standard deviations of $\bar{X}_N$ at $\delta = 0.3$, 0.5 and 0.8. We see that the effect of δ is small, except at small values of n^* where the performance at is $\delta = 0.3$ is slightly inferior.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	0.00117	0.256701	0.00082	0.252899	0.00002	0.212650
43	0.00053	0.242166	0.00069	0.162115	-0.00004	0.155183
61	-0.00014	0.161891	-0.00047	0.130586	0.00024	0.128798
76	-0.00121	0.124549	-0.00083	0.116276	-0.00036	0.115381
96	-0.00073	0.105319	-0.00043	0.103306	-0.00031	0.103083
125	-0.00039	0.091008	0.00011	0.090114	-0.00017	0.089443
171	0.00003	0.076921	0.00052	0.077144	0.00022	0.076250
246	0.00000	0.064622	0.00035	0.063952	-0.00014	0.063728
500	-0.00005	0.044945	-0.00020	0.044721	0.00011	0.044721

Table 5.2 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with $N(0,1)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

5.2.2 Standard uniform distribution

From $(5.1)-(5.2)$ the mean and variance of the estimator $\bar{X}_N$ under the normal optimal fixed sample size are

$$E\left(\bar{X}_N\right)=(1/2)+o\left(\lambda^{-1}\right),$$

and

$$Var\left(\bar{X}_N\right)=\left(1/12n^*\right)-(1/12)\left(n^*\right)^{-2}\left\{-1.2-0.5\delta^{-1}\right\}+o\left(\lambda^{-2}\right).$$

The asymptotic mean and variance of $\bar{X}_N$ are $(1/2)$ and zero respectively as $n^* \to \infty$.

Table 5.3 shows the behaviour of $\bar{X}_N$ at $\delta = 0.5$ and $m=5$, 15, 20. We see that $\bar{X}_N$ is essentially unbiased, in agreement with the asymptotic result. The asymptotic standard deviation of the estimator tends to be an underestimate but the effect reduces as n^* increases.

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E\left(\bar{X}_N\right)$	$sd\left(\bar{X}_N\right)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	0.49974	0.50030	0.49955	0.5	0.0615671	0.0724486	0.0744611	0.0641752
43	0.50012	0.49999	0.49937	0.5	0.0451347	0.0476282	0.0469574	0.0534420
61	0.50008	0.49988	0.50011	0.5	0.0376217	0.0386840	0.0380132	0.0377895
76	0.50008	0.50026	0.50004	0.5	0.0335892	0.0342118	0.0335410	0.0337646
96	0.50003	0.50006	0.50016	0.5	0.0297985	0.0299633	0.0297397	0.0299633
125	0.49990	0.50015	0.49987	0.5	0.0260461	0.0261620	0.0259384	0.0259384
171	0.50000	0.50004	0.50012	0.5	0.0222171	0.0221371	0.0221371	0.0221371
246	0.49995	0.49988	0.49998	0.5	0.0184874	0.0185594	0.0185594	0.0185594
500	0.49991	0.50000	0.50007	0.5	0.0129383	0.0129692	0.0129692	0.0129692

Table 5.3: Comparison between the simulated estimates of $E\left(\bar{X}_N\right)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under $U(0,1)$ as n^* increases and at $m=5,15,20$ and $\delta = 0.5$

To show the effect of increasing δ on the performance of the estimator $\bar{X}_N$ at $m=15$, Table 5.4 shows the simulation estimates of $E\left(\bar{X}_N\right)$ and its standard deviation at $\delta=0.3$, 0.5 and 0.8. We see that the effect of δ is small, except at small values of n^* where the performance at is $\delta=0.3$ is slightly inferior.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	0.50076	0.0742375	0.50030	0.0744611	0.50026	0.0621627
43	0.50045	0.0733430	0.49999	0.0469574	0.50011	0.0453922
61	0.49978	0.0462866	0.49988	0.0380132	0.50036	0.0373423
76	0.49980	0.0351063	0.50026	0.0335410	0.49983	0.0335410
96	0.50005	0.0301869	0.50006	0.0297397	0.50007	0.0297397
125	0.49981	0.0261620	0.50015	0.0259384	0.49995	0.0257148
171	0.49995	0.0223607	0.50004	0.0221371	0.49995	0.0221371
246	0.49993	0.0185594	0.49988	0.0185594	0.49996	0.0183358
500	0.50006	0.0129692	0.50000	0.0129692	0.50002	0.0129692

Table 5.4 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with $U(0,1)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

5.2.3 The t distribution

From $(5.1)-(5.2)$ the mean of the estimator $\bar{X}_N$ under the normal optimal fixed sample size is

$$E(\bar{X}_N) = o(\lambda^{-1}),$$

while the variances at $r = 5, 25$ and 50 degrees of freedom are respectively,

$$Var(\bar{X}_N) = (5/3n^*) - (5/3)(n^*)^{-2}\{6 - 5\delta^{-1}\} + o(\lambda^{-2}),$$

$$Var(\bar{X}_N) = (25/23n^*) - (50/161)(n^*)^{-2}\{1 - 5\delta^{-1}\} + o(\lambda^{-2}),$$

and

$$Var(\bar{X}_N) = (25/24n^*) - (25/552)(n^*)^{-2}\{3 - 30.625\delta^{-1}\} + o(\lambda^{-2}).$$

Clearly, as r increases, the variance of $\bar{X}_N$ decreases and the asymptotic mean and variance of the estimator $\bar{X}_N$ are both zero as $n^* \to \infty$.

Tables 5.5, 5.6 and 5.7 show the effect of r and illustrate the comparison between the simulation results and the asymptotic results for $t(5), t(25)$ and $t(50)$ respectively. Broadly the simulation results and the corresponding asymptotic results are in good agreement for all three values of r.

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E(\bar{X}_N)$	$sd(\bar{X}_N)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	-0.00092	-0.00090	0.00125	0	0.284638	0.292701	0.307907	0.282415
43	-0.00190	0.00010	-0.00068	0	0.205828	0.208849	0.210190	0.227408
61	0.00081	-0.00133	-0.00018	0	0.170628	0.171954	0.170836	0.172624
76	0.00004	-0.00003	-0.00042	0	0.151934	0.152947	0.151829	0.151382
96	-0.00083	-0.00024	-0.00045	0	0.134479	0.134835	0.133940	0.134164
125	-0.00061	-0.00071	-0.00037	0	0.117303	0.117170	0.117170	0.116723
171	-0.00061	-0.00036	0.00020	0	0.099873	0.099505	0.099281	0.099505
246	0.00019	0.00026	-0.00052	0	0.082977	0.082735	0.082511	0.082511
500	-0.00047	-0.00029	-0.00005	0	0.057966	0.057691	0.058138	0.057914

Table 5.5: Comparison between the simulated estimates of $E(\bar{X}_N)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under $t(5)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E(\bar{X}_N)$	$sd(\bar{X}_N)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	-0.00116	-0.00101	0.00025	0	0.223925	0.240601	0.262514	0.233445
43	0.00055	-0.00037	0.00003	0	0.163676	0.169047	0.169718	0.186041
61	0.00072	-0.00012	-0.00045	0	0.136272	0.138413	0.136400	0.137965
76	0.00046	-0.00076	-0.00042	0	0.121598	0.122313	0.121418	0.121866
96	-0.00084	0.00033	-0.00010	0	0.107823	0.108673	0.108226	0.107555
125	-0.00074	0.00059	-0.00039	0	0.094205	0.094586	0.094362	0.094362
171	0.00009	-0.00007	0.00056	0	0.080325	0.080051	0.080051	0.080498
246	-0.00076	-0.00043	-0.00041	0	0.066818	0.066635	0.066858	0.066635
500	-0.00011	-0.00018	-0.00043	0	0.046745	0.046734	0.046734	0.046957

Table 5.6: Comparison between the simulated estimates of $E(\bar{X}_N)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under $t(25)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E(\bar{X}_N)$	$sd(\bar{X}_N)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	0.00062	-0.00022	0.00010	0	0.219050	0.235682	0.256477	0.228303
43	-0.00070	0.00029	-0.00024	0	0.160161	0.165693	0.165469	0.182016
61	-0.00017	0.00004	-0.00044	0	0.133362	0.135506	0.133270	0.135059
76	-0.00024	-0.00070	-0.00116	0	0.119008	0.120748	0.118735	0.118959
96	-0.00114	0.00052	-0.00113	0	0.105532	0.105990	0.105766	0.105095
125	0.00000	-0.00080	0.00026	0	0.092207	0.092573	0.092573	0.091902
171	-0.00064	-0.00043	-0.00041	0	0.078625	0.078710	0.078486	0.078710
246	0.00004	-0.00014	0.00011	0	0.065407	0.065517	0.065740	0.065740
500	-0.00070	-0.00027	-0.00018	0	0.045759	0.045839	0.045839	0.045839

Table 5.7: Comparison between the simulated estimates of $E(\bar{X}_N)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under $t(50)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$.

The impact of δ on the performance of the estimator $\bar{X}_N$ is shown in Tables 5.8, 5.9 and 5.10 for $t(5)$, $t(25)$ and $t(50)$ respectively. We see that for a each value of r the higher the value of δ the better the behaviour. However, this effect is small for lager values of n^*.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	-0.00441	0.329820	-0.00090	0.307907	-0.00113	0.266092
43	-0.00298	0.290242	0.00010	0.210190	0.00133	0.198339
61	0.00063	0.210414	-0.00133	0.170836	-0.00073	0.164798
76	0.00083	0.167034	-0.00003	0.151829	-0.00055	0.147580
96	0.00083	0.138860	-0.00024	0.133940	0.00022	0.130363
125	0.00068	0.118959	-0.00071	0.117170	-0.00003	0.115158
171	-0.00028	0.100623	-0.00036	0.099281	0.00000	0.097940
246	-0.00027	0.083405	0.00026	0.082511	0.00024	0.081393
500	-0.00042	0.058138	-0.00029	0.058138	-0.00003	0.057467

Table 5.8 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with $t(5)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	0.00133	0.269893	-0.00101	0.262514	-0.00117	0.220253
43	0.00117	0.251110	-0.00037	0.169718	-0.00046	0.160773
61	-0.00023	0.168823	-0.00012	0.136400	0.00000	0.135282
76	-0.00021	0.131481	-0.00076	0.121418	-0.00025	0.120077
96	-0.00102	0.110238	0.00033	0.108226	-0.00091	0.106884
125	-0.00001	0.095256	0.00059	0.094362	-0.00020	0.093244
171	-0.00088	0.081169	-0.00007	0.080051	0.00000	0.079828
246	0.00024	0.067306	-0.00043	0.066858	-0.00003	0.066635
500	-0.00057	0.046510	-0.00018	0.046734	-0.00023	0.046734

Table 5.9 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with $t(25)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	0.00034	0.262514	-0.00022	0.256477	0.00012	0.216899
43	0.00047	0.245967	0.00029	0.165469	-0.00015	0.157643
61	0.00063	0.165022	0.00004	0.133270	-0.00040	0.131257
76	-0.00090	0.126785	-0.00070	0.118735	-0.00026	0.118064
96	0.00010	0.107778	0.00052	0.105766	0.00046	0.104201
125	-0.00047	0.092797	-0.00080	0.092573	-0.00018	0.091902
171	-0.00018	0.079157	-0.00043	0.078486	-0.00094	0.078262
246	0.00009	0.065740	-0.00014	0.065740	-0.00019	0.065070
500	-0.00042	0.045839	-0.00027	0.045839	-0.00032	0.045392

Table 5.10 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with $t(50)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

5.2.4 Beta $(2,3)$ distribution

From $(5.1)-(5.2)$ the mean and variance of the estimator $\bar{X}_N$ under the normal optimal fixed sample size are respectively,

$$E(\bar{X}_N) = 0.4 - (1/35)(n^*)^{-1} + o(\lambda^{-1}),$$

and

$$Var(\bar{X}_N) = (1/25n^*) + (1/350)(n^*)^{-2}\{9 + 11.875\delta^{-1}\} + o(\lambda^{-2}).$$

Thus as $n^* \to \infty$, the asymptotic mean and variance converge to 0.4 and zero respectively.

Table 5.11 shows the simulation results for beta$(2,3)$ and the corresponding asymptotic results. Clearly the bias of the estimator is small and is in line with the asymptotic result. Again the asymptotic results for the standard deviation of the estimator are also in reasonable agreement with the simulated values.

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E(\bar{X}_N)$	$sd(\bar{X}_N)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	0.39700	0.39929	0.40022	0.398810	0.0427682	0.0478519	0.0509823	0.0447214
43	0.39887	0.39862	0.39762	0.399336	0.0313184	0.0326466	0.0328702	0.0366715
61	0.39936	0.39927	0.39925	0.399532	0.0260938	0.0266092	0.0263856	0.0263856
76	0.39976	0.39957	0.39942	0.399624	0.0232920	0.0237023	0.0232551	0.0234787
96	0.39957	0.39964	0.39953	0.399702	0.0206596	0.0207954	0.0205718	0.0207954
125	0.39991	0.39967	0.39988	0.399771	0.0180552	0.0181122	0.0181122	0.0181122
171	0.39987	0.39981	0.39994	0.399833	0.0153986	0.0154289	0.0154289	0.0154289
246	0.39974	0.39996	0.39986	0.399884	0.0128120	0.0127456	0.0127456	0.0129692
500	0.39991	0.39996	0.39994	0.399943	0.0089652	0.0089443	0.0089443	0.0089443

Table 5.11: Comparison between the simulated estimates of $E(\bar{X}_N)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under *beta* $(2,3)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$.

Table 5.12 shows the behaviour of the estimator $\bar{X}_N$ for different values of δ. Once again the performance at $\delta = 0.3$ is the least good.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	0.40019	0.0516532	0.39929	0.0509823	0.39842	0.0431561
43	0.39823	0.0491935	0.39862	0.0328702	0.39949	0.0313050
61	0.39714	0.0328702	0.39927	0.0263856	0.39967	0.0259384
76	0.39902	0.0248204	0.39957	0.0232551	0.39950	0.0232551
96	0.39972	0.0207954	0.39964	0.0205718	0.39977	0.0205718
125	0.39982	0.0181122	0.39967	0.0181122	0.39988	0.0178885
171	0.39984	0.0154289	0.39981	0.0154289	0.39985	0.0152053
246	0.39981	0.0127456	0.39996	0.0127456	0.39980	0.0127456
500	0.39996	0.0089443	0.39996	0.0089443	0.39998	0.0089443

Table 5.12 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with beta $(2,3)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

5.2.5 Exponential distribution with mean one

The mean and variance of the estimator $\bar{X}_N$ of μ under the normal optimal fixed sample size are respectively,

$$E(\bar{X}_N) = 1 - (n^*)^{-1} + o(\lambda^{-1}),$$

and

$$Var(\bar{X}_N) = (1/n^*) - (n^*)^{-2}\{6 - 5\delta^{-1}\} + o(\lambda^{-2}).$$

The asymptotic mean and variance of $\bar{X}_N$ are one and zero respectively as $n^* \to \infty$.

Table 5.13 shows that the estimator is somewhat more biased and more variable than indicated by the asymptotic results. However, this effect reduces as the optimal sample size increases.

n^*	$\mu_s(5)$	$\mu_s(15)$	$\mu_s(20)$	$E(\bar{X}_N)$	$sd(\bar{X}_N)$	$ssd(5)$	$ssd(15)$	$ssd(20)$
24	0.91062	0.96999	0.99486	0.958333	0.220479	0.270564	0.219358	0.215781
43	0.95096	0.95095	0.95662	0.976744	0.159434	0.203482	0.183358	0.172624
61	0.97055	0.97141	0.96587	0.983607	0.132168	0.160326	0.152053	0.154289
76	0.97799	0.98051	0.97832	0.986842	0.117688	0.138636	0.130139	0.132822
96	0.98547	0.98597	0.98632	0.989583	0.104167	0.116723	0.110909	0.112027
125	0.98934	0.99106	0.99068	0.992000	0.090863	0.098834	0.095033	0.094362
171	0.99284	0.99283	0.99342	0.994152	0.077361	0.080722	0.079157	0.079380
246	0.99513	0.99610	0.99518	0.995935	0.064274	0.065517	0.065517	0.065517
500	0.99761	0.99772	0.99795	0.998000	0.044900	0.045392	0.045169	0.045169

Table 5.13: Comparison between the simulated estimates of $E(\bar{X}_N)$ and standard deviation of $\bar{X}_N$ with the asymptotic results under $Exp(1)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$

Table 5.14 shows the performance of the estimator for different values of δ. Here the standard deviations tend to decrease as δ increases, but the effect reduces with increasing n^*.

n^*	$\mu_s(0.3)$	$ssd(0.3)$	$\mu_s(0.5)$	$ssd(0.5)$	$\mu_s(0.8)$	$ssd(0.8)$
24	0.99728	0.255135	0.96999	0.219358	0.95964	0.208402
43	0.95447	0.203035	0.95095	0.183358	0.96937	0.167705
61	0.93888	0.175531	0.97141	0.152053	0.98034	0.136847
76	0.95264	0.162562	0.98051	0.130139	0.98446	0.120971
96	0.97270	0.137295	0.98597	0.110909	0.98774	0.105319
125	0.98560	0.108002	0.99106	0.095033	0.99072	0.091455
171	0.99151	0.083182	0.99283	0.079157	0.99376	0.077368
246	0.99505	0.066411	0.99610	0.065517	0.99567	0.063728
500	0.99785	0.045839	0.99772	0.045169	0.99757	0.044721

Table 5.14 The simulated estimates of the expected value and standard deviation of $\bar{X}_N$ with $Exp(1)$ underlying distribution at $m = 15$ and $\delta = 0.3, 0.5, 0.8$ and selected values of n^*.

5.3 The behaviour of the actual sample size N

In this section we investigate the mean and variance of the stopping variable N. By using Theorem 3.2.2.2, the mean and variance of the stopping variable N under the normal optimal fixed sample size are respectively,

$$(5.3)\quad E(N) = n^* - (3/8)(\beta - 1)\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1),$$

and

$$(5.4)\quad Var(N) = (4\delta)^{-1} n^* (\beta - 1) + o(\lambda).$$

We investigate the behaviour of ε_{N_1} in section 5.5 but to a good approximation it has mean 0.5.

In the following subsections we compare the simulation results with the asymptotic results obtained from equations $(5.3)-(5.4)$ for the underlying distributions mentioned above. Each table includes $\overline{N}_s(m)$ and $ssd\overline{N}(m)$, the simulated mean and the simulated standard deviation of the actual sample size N at a specific value of m, where $m = 5,15$ and 20, and also $E(N)$, $sd(N)$, the asymptotic mean and standard deviation of N.

5.3.1 Standard normal distribution

From $(5.3)-(5.4)$ the mean and variance of N when the underlying distribution is standard normal are respectively

$$E(N) = n^* - 0.75\delta^{-1} + 0.5 + o(1),$$

and

$$Var(N) = (2\delta)^{-1} n^* + o(\lambda).$$

Table 5.15 below presents the simulation and asymptotic results for N at $\delta = 0.5$ and $m = 5, 15, 20$. We see that $\overline{N} < n^*$ (early stopping on average) for all values of n^* and the quantity $\left|\overline{N} - n^*\right|$ decreases as n^* increases. Table 5.15 also shows the performance of the estimator $\overline{N}$ as the pilot sample size and the optimal sample size increases in comparison with the asymptotic results. Problems arise when *m* is close to *n**, but in other cases the asymptotic results are optimistic in that they tend to underestimate the early stopping and underestimate the variability, particularly for $m = 5$.

n^*	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\overline{N}(5)$	$ssd\overline{N}(15)$	$ssd\overline{N}(20)$
24	22.18	16.36	20.00	23	4.8990	6.4806	4.6356	0.0000
43	41.25	41.11	36.39	42	6.5575	8.2068	8.8318	13.0754
61	59.37	59.78	59.52	60	7.8101	9.5813	8.5404	9.2989
76	74.44	74.90	74.83	75	8.7178	10.4874	9.3295	9.3009
96	94.51	94.88	94.88	95	9.7980	11.6886	10.4245	10.2917
125	123.59	123.90	123.92	124	11.1803	12.9920	11.7378	11.6647
171	169.56	169.83	169.87	170	13.0767	15.1040	13.7169	13.4962
246	244.62	244.95	244.92	245	15.6845	18.0156	16.3383	16.2531
500	498.70	498.82	498.92	499	22.3607	25.3201	23.1905	23.0038

Table 5.15: Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results under $N(0,1)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$.

To show the impact of δ on the performance N, we consider Figure 5.1 at $m = 15$. We see that for $\delta = 0.3$ and 0.5, $\overline{N} < n^*$ (early stopping on average) for all values of n^*, but the amount $\left|\overline{N} - n^*\right|$, is larger at $\delta = 0.3$, which means that the procedure tends to terminate much early than in the case of $\delta = 0.5$, this consequently causes bad estimates at $\delta = 0.3$. While for $\delta = 0.8$, $\overline{N} > n^*$ (over sampling in average) at $n^* = 246$ and 500. But such behaviour tends to vanish as m increases. This indicate that the choice of $\delta = 0.5$ is much better than other values of δ and this supports Hall (1981) recommendation.

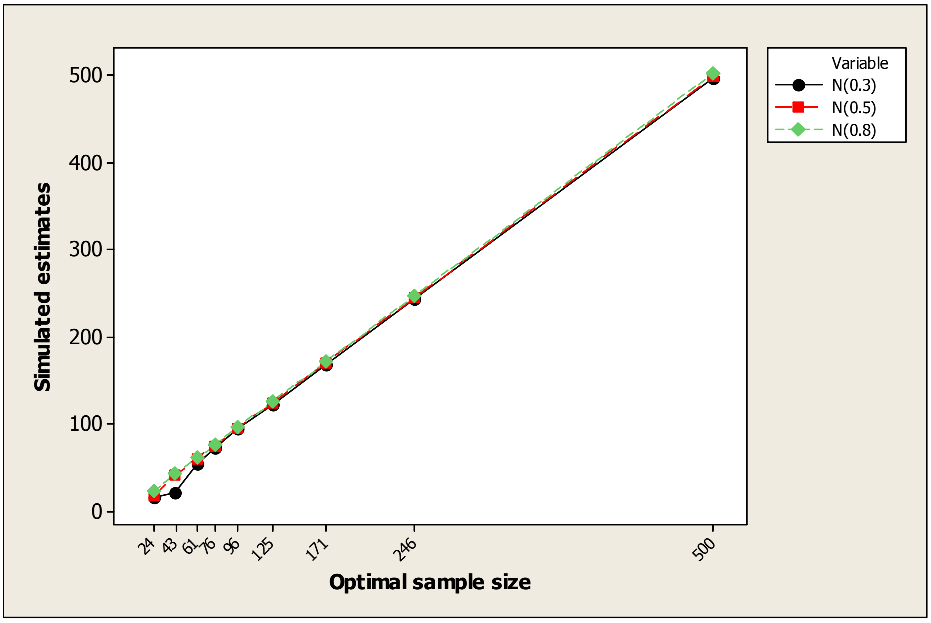

Figure 5.1: The simulated estimates of $E(N)$ for underlying $N(0,1)$ as the optimal sample size increases at $m = 15$ and $\delta = 0.3, 0.5, 0.8$.

5.3.2 Standard uniform distribution

From $(5.3)-(5.4)$ the mean and variance for the actual sample size N when the underlying distribution is standard uniform are respectively,

$$E(N)=n^{*}-0.3\delta^{-1}+E\left(\varepsilon_{N_1}\right)+o(1),$$

and

$$Var(N)=0.2\delta^{-1}n^{*}+o(\lambda).$$

Table 5.16 below present the simulation and asymptotic results for N at $\delta=0.5$ and $m=5,\ 15,\ 20$. We see the same pattern as in the previous case. Provided *m* and n^{*} are not too close, we see that N has a small negative bias but this is well captured by the asymptotic result. However, the asymptotic standard deviation underestimates the true variability in *N*, especially when *m* = 5.

n^{*}	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\overline{N}(5)$	$ssd\overline{N}(15)$	$ssd\overline{N}(20)$
24	23.05	15.24	20.00	23.9	3.0984	4.9817	1.9566	0.0000
43	42.35	42.52	38.15	42.9	4.1473	5.7319	5.5457	11.3579
61	60.52	60.77	60.73	60.9	4.9396	6.2733	5.4940	5.5958
76	75.52	75.82	75.77	75.9	5.5136	6.8513	5.9743	5.9549
96	95.60	95.76	95.80	95.9	6.1968	7.3638	6.6440	6.5669
125	124.62	124.86	124.80	124.9	7.0711	8.2198	7.4689	7.4130
171	170.65	170.86	170.82	170.9	8.2704	9.4713	8.6022	8.6216
246	245.64	245.78	245.83	245.9	9.9197	11.0600	10.2582	10.1725
500	499.69	499.86	499.89	499.9	14.1421	15.4454	14.5423	14.3878

Table 5.16: Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results for underlying $U(0,1)$ as n^{*} increases and at $m=5,15,20$ and $\delta=0.5$

To illustrate the effect of increasing δ on the performance of the actual sample size N we consider Figure 5.2 at $m=15$. We noticed $\overline{N}<n^{*}$ (early stopping on average) occurs at $\delta=0.3$ and 0.5 and the absolute bias $\left|\overline{N}-n^{*}\right|$ is larger at $\delta=0.3$. While at $\delta=0.8$ we noticed that $\overline{N}>n^{*}$ for all $n^{*}\geq 43$.

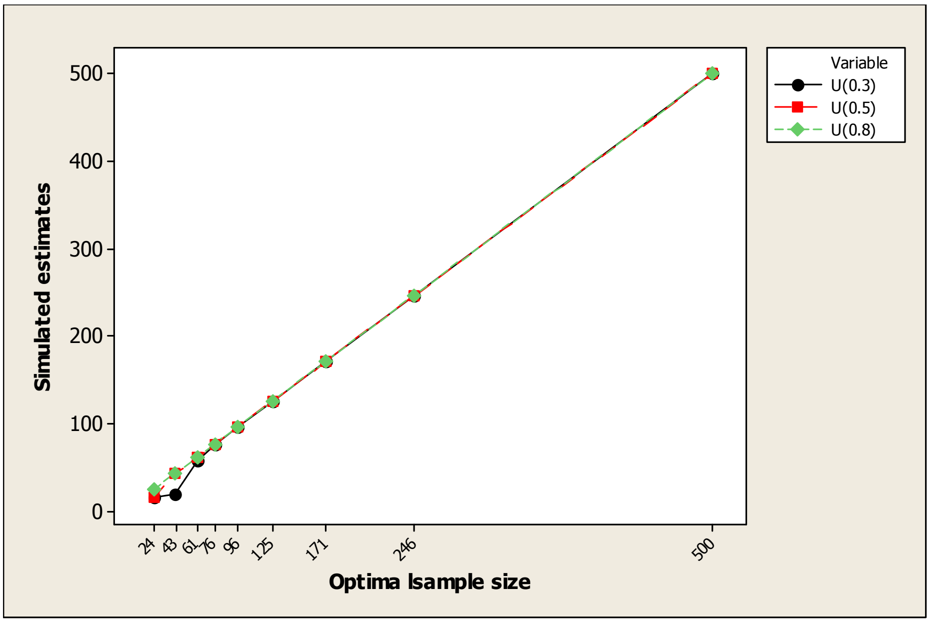

Figure 5.2: The simulated estimates of $E(N)$ for underlying $U(0,1)$ as the optimal sample size increases at $m=15$ and $\delta = 0.3, 0.5, 0.8$

5.3.3 The t distribution

From $(5.3)-(5.4)$ the mean and variance of N when the underlying distribution is t with $r = 5$ degrees of freedom are

$$E(N) = n^* - 3\delta^{-1} + E(\varepsilon_{N_1}) + o(1),$$

and

$$Var(N) = 2(\delta)^{-1} n^* + o(\lambda).$$

At $r = 25$, the mean and variance are

$$E(N) = n^* - (6/7)\delta^{-1} + E(\varepsilon_{N_1}) + o(1),$$

$$Var(N) = (4/7)(\delta)^{-1} n^* + o(\lambda).$$

At $r = 100$, the mean and variance are

$$E(N) = n^* - (99/128)\delta^{-1} + E(\varepsilon_{N_1}) + o(1),$$

$$Var(N) = (33/64)(\delta)^{-1} n^* + o(\lambda).$$

Clearly, for a fixed δ and as r increases, the variance of N decreases. Tables 5.17, 5.18 and 5.19 show the simulation results for $r=5$, 25 and 50 respectively. From Table 5.10 we observe early stopping behaviour: $\overline{N}<n^*$ for all values of n^*, while the magnitude of the bias decreases as the m increases. Similar behaviour occurs at $r=25$ and 50. Also, note that $\left|\overline{N}-n^*\right|$ decreases as r increases. We see from these tables that the differences between the simulation estimates of N and the corresponding asymptotic values are: for $r=5, m=15$, it is 2.35 and for $r=5, m=20$ it is 1.97, while for $r=25, m=15$ and 20 it is 0.156 and finally for $r=50, m=15$ it is 0.012 and at $r=50$, $m=20$ it is 0.108. This indicates that better estimates occur as r increases and at $m=15$. Similar arguments can be said regarding the ratio between the simulation standard deviation and the asymptotic standard deviation.

n^*	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\,\overline{N}(5)$	$ssd\,\overline{N}(15)$	$ssd\,\overline{N}(20)$
24	20.88	17.33	20.32	18.5	9.7980	8.1961	6.3670	2.8087
43	39.66	38.65	34.20	37.5	13.1150	11.7470	12.2199	14.6013
61	57.66	57.65	57.04	55.5	15.6205	14.8871	12.7521	14.1375
76	72.68	72.70	72.73	70.5	17.4355	16.8409	14.3761	14.2860
96	92.72	92.81	92.69	90.5	19.5960	18.7890	16.6569	16.2441
125	121.86	121.68	121.75	119.5	22.3607	22.2898	19.6488	18.6857
171	167.80	167.73	167.60	165.5	26.1535	26.9987	23.2992	23.4483
246	243.19	242.64	242.50	240.5	31.3687	33.3313	28.6809	28.3670
500	498.39	496.85	496.47	494.5	44.7214	54.2577	43.3122	42.6409

Table 5.17: Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results for underlying $t(5)$ as n^* increases and at $m=5,15,20$ and $\delta=0.5$

n^*	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\,\overline{N}(5)$	$ssd\,\overline{N}(15)$	$ssd\,\overline{N}(20)$
24	22.02	16.53	20.00	22.786	5.2373	6.6946	4.9066	0.2985
43	41.09	40.94	36.14	41.786	7.0103	8.6156	9.2562	13.3169
61	59.18	59.60	59.34	59.786	8.3495	10.0753	9.0494	9.8421
76	74.27	74.65	74.61	74.786	9.3197	11.2199	9.9038	9.9304
96	94.26	94.66	94.70	94.786	10.4744	12.4200	11.0795	10.9013
125	123.27	123.67	123.67	123.786	11.9522	13.8983	12.4741	12.4428
171	169.49	169.72	169.73	169.786	13.9797	16.0288	14.4569	14.3819
246	244.54	244.71	244.72	244.786	16.7674	19.1620	17.5697	17.3608
500	498.77	498.63	498.63	498.786	23.9045	27.3455	24.7271	24.5377

Table 5.18: Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results for underlying $t(25)$ as n^* increases and at $m=5,15,20$ and $\delta=0.5$.

n^*	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\,\overline{N}(5)$	$ssd\,\overline{N}(15)$	$ssd\,\overline{N}(20)$
24	22.14	16.48	20.00	22.902	5.0562	6.5499	4.8250	0.2625
43	41.20	41.05	36.31	41.902	6.7679	8.4957	8.9946	13.1400
61	59.31	59.64	59.46	59.902	8.0608	9.8128	8.7839	9.4170
76	74.40	74.70	74.74	74.902	8.9975	10.7159	9.6140	9.5822
96	94.39	94.76	94.88	94.902	10.1124	11.8632	10.7253	10.5618
12	123.56	123.69	123.83	123.902	11.5392	13.3943	12.1143	11.9873
17	169.43	169.79	169.84	169.902	13.4965	15.6862	14.1161	13.9649
24	244.47	244.84	244.68	244.902	16.1878	18.6126	16.8774	16.6869
50	498.62	498.89	499.01	498.902	23.0782	26.2615	23.7602	23.5959

Table 5.19: Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results for underlying $t(50)$ as n^* increases and at $m = 5$, 15, 20 and $\delta = 0.5$.

To show the effect of increasing δ on the performance of the actual sample size N, we consider Figures 5.3, 5.4 and 5.5.

We noticed that for $r = 5, 25$ and 50 we have the following: at $\delta = 0.3$ and $\delta = 0.5$ an early stopping occurs, while at $\delta = 0.8$, $\overline{N}$ is over estimating n^*, and the amount of $\left|\overline{N} - n^*\right|$ decreases as n^* increases.

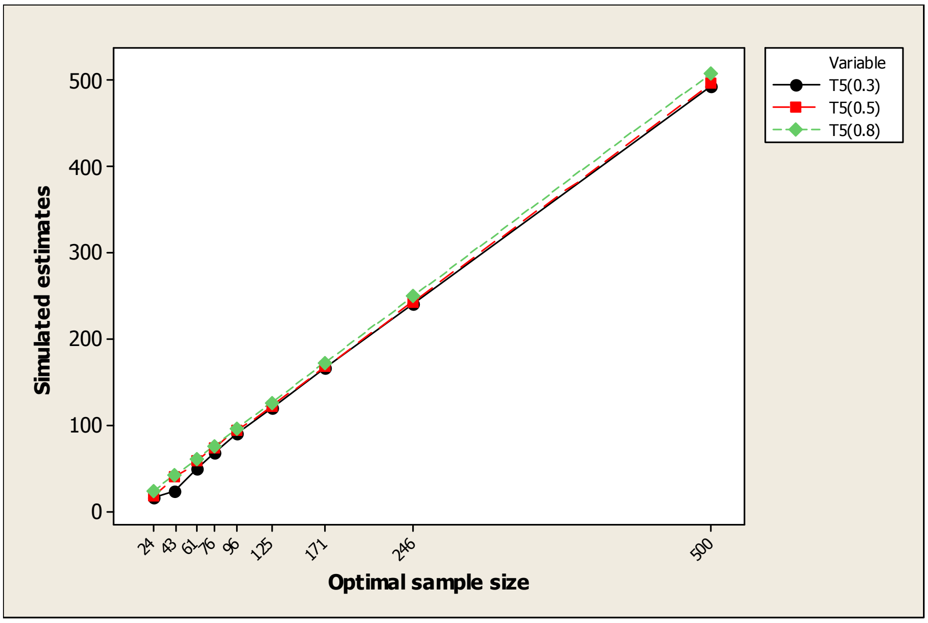

Figure 5.3: The simulated estimates of $E(N)$ for underlying $t(5)$ as the optimal sample size increases at $m = 15$ and $\delta = 0.3, 0.5, 0.8$.

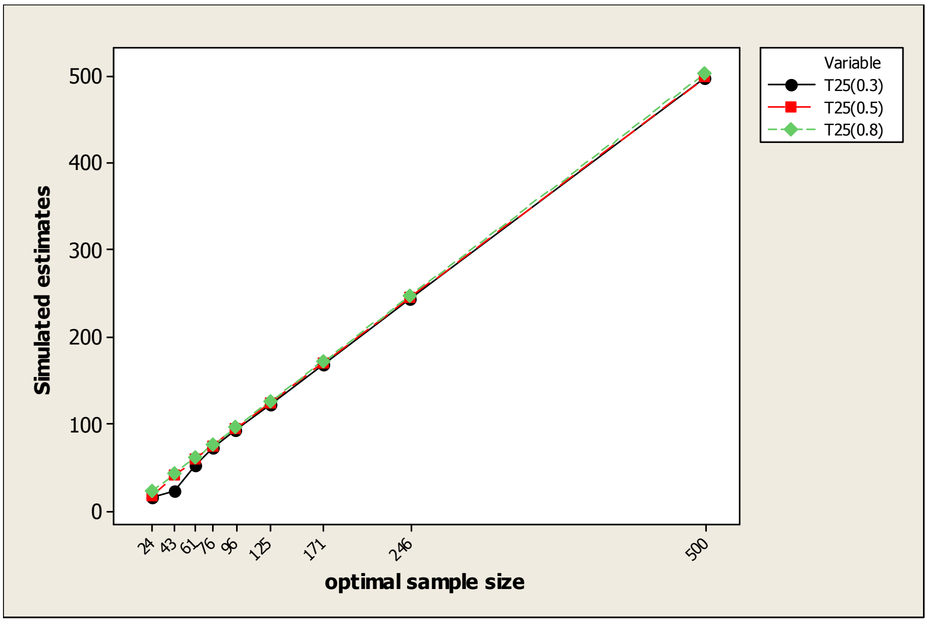

Figure 5.4: The simulated estimates of $E(N)$ for underlying $t(25)$ as the optimal sample size increases at $m = 15$ and $\delta = 0.3, 0.5, 0.8$.

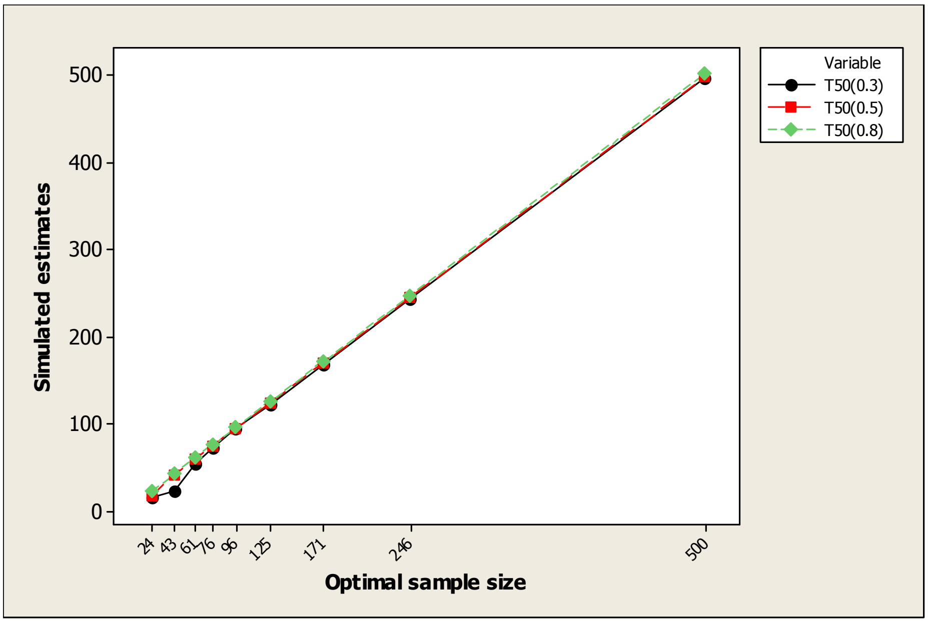

Figure 5.5: The simulated estimates of $E(N)$ for underlying $t(50)$ as the optimal sample size increases at $m = 15$ and $\delta = 0.3, 0.5, 0.8$.

5.3.4 Beta $(2,3)$ distribution

From $(5.3)-(5.4)$ the mean and variance of N when the underlying distribution is beta $(2,3)$ are respectively

$$E(N)=n^{*}-(57/112)\delta^{-1}+E\left(\varepsilon_{N_1}\right)+o(1),$$

and

$$Var(N)=(19/56)(\delta)^{-1}n^{*}+o(\lambda).$$

Table 5.20 below presents the simulation and asymptotic results for N at $\delta=0.5$ and $m=5$, 15, 20. Regarding the simulation results we have seen same pattern as in the previous cases. Also Table 5.20 shows the comparison between the simulation results and the asymptotic results and similar arguments can as stated as in the previous cases, but note here that the difference between the simulation estimates of $E(N)$ and its asymptotic value is 0.072 at $m=15$ and 0.122 at $m=20$. The asymptotic standard deviation underestimates the simulated standard deviation for all values of m, especially at $m=5$.

n^{*}	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\,\overline{N}(5)$	$ssd\,\overline{N}(15)$	$ssd\,\overline{N}(20)$
24	22.63	15.78	20.00	23.482	4.0356	5.8212	3.5015	0.0000
43	41.78	41.80	37.19	42.482	5.4017	7.1659	7.3627	12.3679
61	59.94	60.28	60.15	60.482	6.4337	8.0981	7.1384	7.5738
76	75.04	75.28	75.37	75.482	7.1813	8.7097	7.8482	7.7167
96	95.06	95.29	95.34	95.482	8.0711	9.5590	8.6209	8.5836
12	124.09	124.36	124.41	124.482	9.2099	10.7197	9.7016	9.7204
17	170.06	170.32	170.45	170.482	10.7720	12.2948	11.3156	11.2078
24	245.20	245.30	245.41	245.482	12.9201	14.6145	13.4678	13.3914
50	499.12	499.41	499.36	499.482	18.4197	20.5264	18.9885	18.8657

Table 5.20: Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results under *beta* $(2,3)$ as n^{*} increases and at $m=5,15,20$ and $\delta=0.5$.

Figure 5.6 shows the effect of increasing δ on the performance of N. Similar arguments as before can be said here.

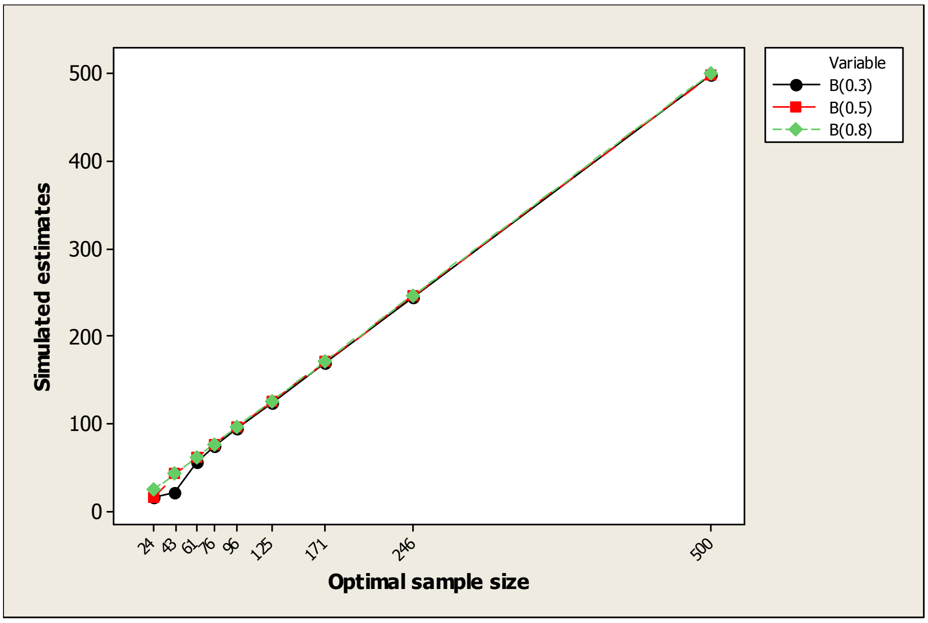

Figure 5.6: The simulated estimates of $E(N)$ for underlying beta $(2,3)$ as the optimal sample size increases at $m=15$ and $\delta=0.3, 0.5, 0.8$.

5.3.5 Exponential distribution with mean one

From $(5.3),(5.4)$ the mean and variance for N when the underlying distribution is exponential with mean one are respectively,

$$E(N)=n^{*}-3\delta^{-1}+E\left(\varepsilon_{N_1}\right)+o(1),$$

and

$$Var(N)=2(\delta)^{-1}n^{*}+o(\lambda).$$

To show the performance of the simulated estimates of the actual sample size N at $\delta=0.5$ while $m=5$, 15 and 20 and compare the results with the asymptotic values we consider Table 5.21 From Table 5.21 we see that $\overline{N}<n^{*}$ for all values of n^{*} and m. However, the case of the underlying exponential distribution is different from the previous distributions because here we see that as m increases, so $\left|\overline{N}-n^{*}\right|$ increases which is opposite to the previous distributions.

We see that the differences between the simulated and asymptotic values of $E(N)$ are relatively small. However, the asymptotic standard deviation tends to underestimate the variability, especially when m = 5.

n^*	$\overline{N}_s(5)$	$\overline{N}_s(15)$	$\overline{N}_s(20)$	$E(N)$	$s.d(N)$	$ssd\overline{N}(5)$	$ssd\overline{N}(15)$	$ssd\overline{N}(20)$
24	19.05	18.00	20.46	18.5	9.7980	9.4548	7.1087	3.2374
43	37.36	36.70	33.97	37.5	13.1149	14.9365	14.2833	15.2229
61	55.27	55.42	54.51	55.5	15.6205	18.5862	16.2180	17.5464
76	70.43	70.67	70.40	70.5	17.4356	21.2979	18.1126	18.1727
96	90.82	90.71	90.50	90.5	19.5959	24.3286	20.4415	19.9214
125	119.63	119.59	119.70	119.5	22.3607	28.1069	23.3593	22.9483
171	166.33	165.67	165.69	165.5	26.1534	33.7398	27.6396	27.0486
246	241.72	241.02	240.62	240.5	31.3688	41.0723	33.5526	32.7117
500	497.95	495.23	495.16	494.5	44.7214	62.3749	48.7127	47.2864

Table 5.21: : Comparison between the simulated estimates of $E(N)$ and standard deviation of N with the asymptotic results under $Exp(1)$ as n^* increases and at $m = 5, 15, 20$ and $\delta = 0.5$.

Figure 5.7 shows the performance of the simulated extimates of N for selected values of n^* and δ. We see that at $\delta = 0.3$, $\overline{N} < n^*$ for all values of n^* and the bias of $\overline{N}$ decreases as n^* increases. However, at $\delta = 0.8$ we see that $\overline{N} \geq n^*$ for $n^* = 61$ and the bias tends to decrease as n^* increases. But, the procedure takes more time to overcome such behaviour $\overline{N} > n^*$ for large values of n^*, and this delay goes to the sharp skewness and high kurtosis of the exponential distribution.

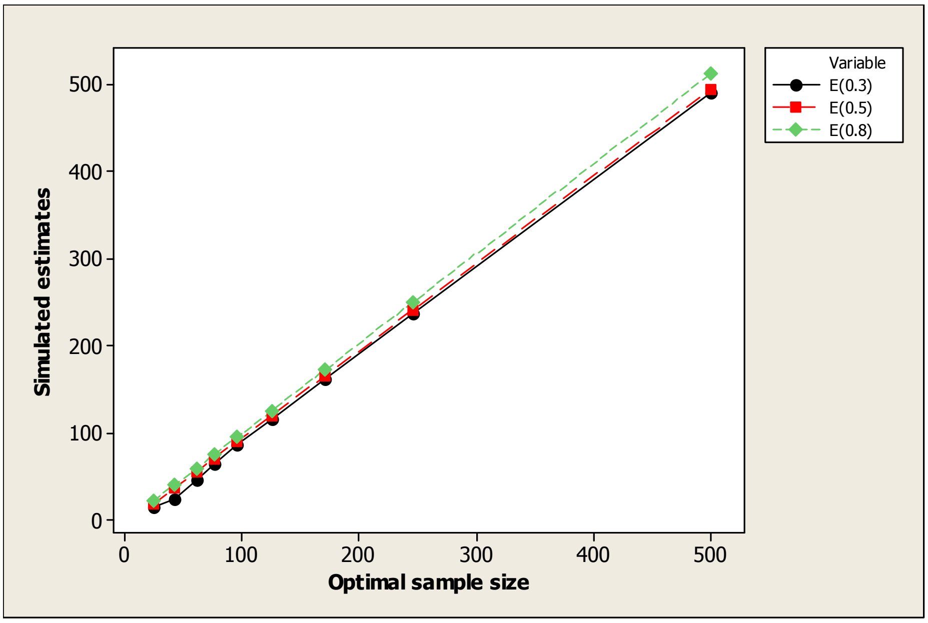

Figure 5.7: The simulated estimates of $E(N)$ for underlying $Exp(1)$ as the optimal sample size increases at $m = 15$ and $\delta = 0.3, 0.5, 0.8$.

5.4 The asymptotic distribution of the actual sample size N

In Theorem 3.3.4 we proved that the asymptotic distribution of N is a standard normal distribution. To investigate the rate of convergence to normality, we ran a FORTRAN program using the IMSL library at $\delta = 0.5$, $m = 15$ and all values of n^* as mentioned before. For each of 1000 replicates and for a specific value of n^* the program runs the triple sampling procedure and stores the stopping sample size in a vector. At the end for each value of n^* we will have 1000 replicate samples from the actual sample size N. To check the normality of these samples we used two standard goodness of fit tests for normality: the Anderson-Darling (AD) test and the Kolomogorov-Smirnov (KS) test. The sample mean, sample standard deviation, sample skewness and sample kurtosis were also recorded for each n^*.

5.4.1. Standard normal distribution

Table 5.22 shows the p-values for tests of normality of N when the underlying distribution is standard normal and with $m = 15$, $\delta = 0.5$ and $n^* = 76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	75.005	9.214	-0.24	3.39	<0.005	>0.150
96	94.707	10.404	-0.33	3.15	<0.005	<0.01
125	124.170	11.68	-0.15	2.91	0.005	0.071
171	169.900	13.60	-0.20	3.37	0.031	>0.150
246	243.840	17.24	-0.20	3.66	0.061	>0.150
500	498.260	23.07	-0.05	2.99	0.255	> 0.150

Table 5.22: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying $N(0,1)$ as the optimal sample size increases.

Figure 5.8 shows the normal probability plot of the simulated values of N when the underlying distribution is standard normal at $n^* = 500$. We proceed our conclusions by using AD test rather than KS test, since KS test is not sensitive to the presence of outliers in the tails. Now by using AD test the p – value is 0.225 and hence we have a good evidence that our data follows the normal distribution, though the evidence for normality is much weaker for smaller values of n^*, so convergence to normality appears to be slow.

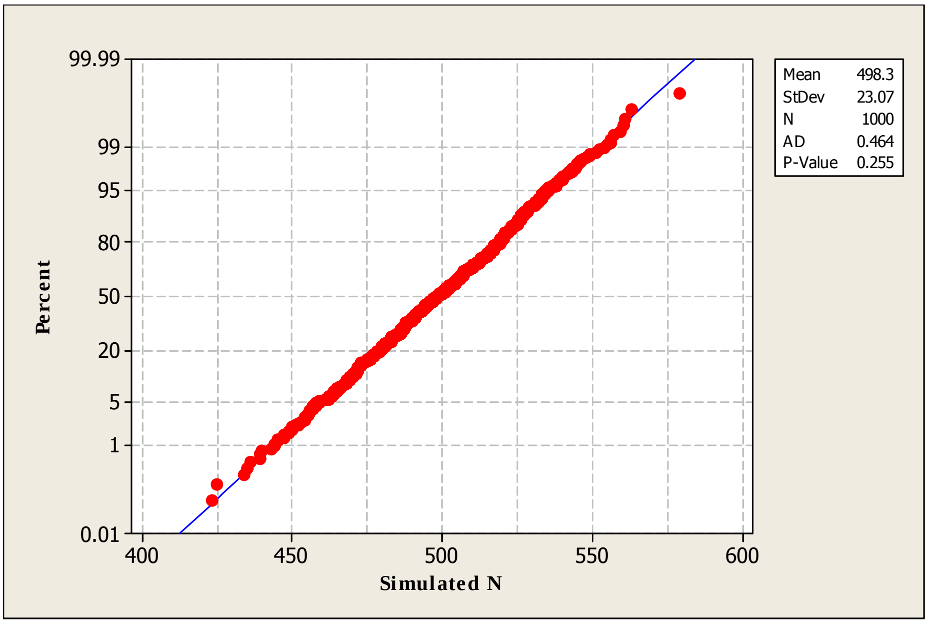

Figure 5.8. Normal probability plot of the actual sample size N for underlying distribution $N(0,1)$ using AD test at n^*=500, $\delta = 0.5$ and $m = 15$; p – value = 0.255.

5.4.2. Standard uniform distribution

Similarily, Table 5.23 shows the p-values for tests of normality of N when the underlying distribution is standard uniform and with $m = 15$, $\delta = 0.5$ and $n^* = 76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	75.859	5.893	-0.310	3.070	< 0.005	>0.15
96	95.676	6.681	-0.330	3.310	< 0.005	0.132
125	124.95	7.47	-0.180	3.340	0.011	>0.15
171	171.28	8.49	-0.360	3.630	< 0.005	0.029
246	246.02	10.10	-0.260	3.310	< 0.005	>0.15
500	500.10	14.69	-0.150	3.160	0.042	>0.15

Table 5.23: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying $U(0,1)$ as the optimal sample size increases.

Figure 5.9 below shows the normal probability plot of the simulated estimates of N when the underlying distribution is standard uniform at $n^* = 500$. By using AD test the p – value is 0.042, which means that normality is rejected. However, the plot shows that the extent of the departure from normality is perhaps rather small.

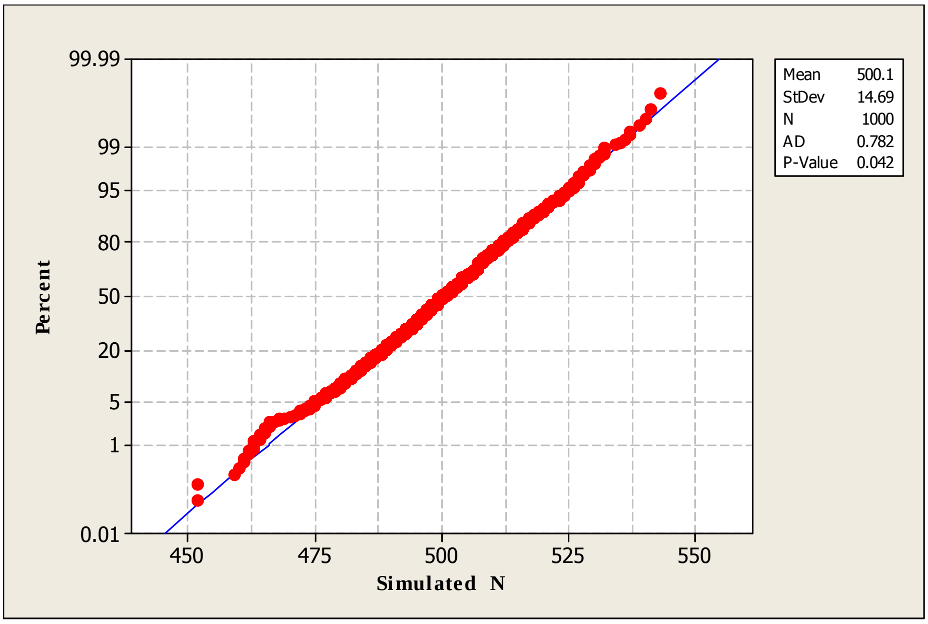

Figure 5.9. Normal probability plot of the actual sample size N for underlying distribution $U(0,1)$ using AD test at n^*=500, $\delta = 0.5$ and $m = 15$; p – value = 0.042.

5.4.3. The t distribution

Similarily, Table 5.24 shows the p-values for tests of normality of N when the underlying distribution is $t(5)$ and with $m = 15$, $\delta = 0.5$ and $n^* = 76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	72.952	13.797	0.53	5.11	<0.005	<0.01
96	92.600	17.419	2.53	33.75	<0.005	<0.01
125	121.36	19.310	0.89	7.93	<0.005	<0.01
171	168.64	22.940	0.77	5.67	<0.005	<0.01
246	243.10	29.200	1.82	15.53	<0.005	<0.01
500	499.01	48.670	3.07	30.64	<0.005	<0.01

Table 5.24: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying $t(5)$ as the optimal sample size increases.

Figure 5.10 below shows the normal probability plot of the simulated estimates of N when the underlying distribution is $t(5)$ at $n^* = 500$. By using AD test the p- value is less than 0.005. The extreme outliers in the upper tail are clear. Thus, convergence to normality is very slow.

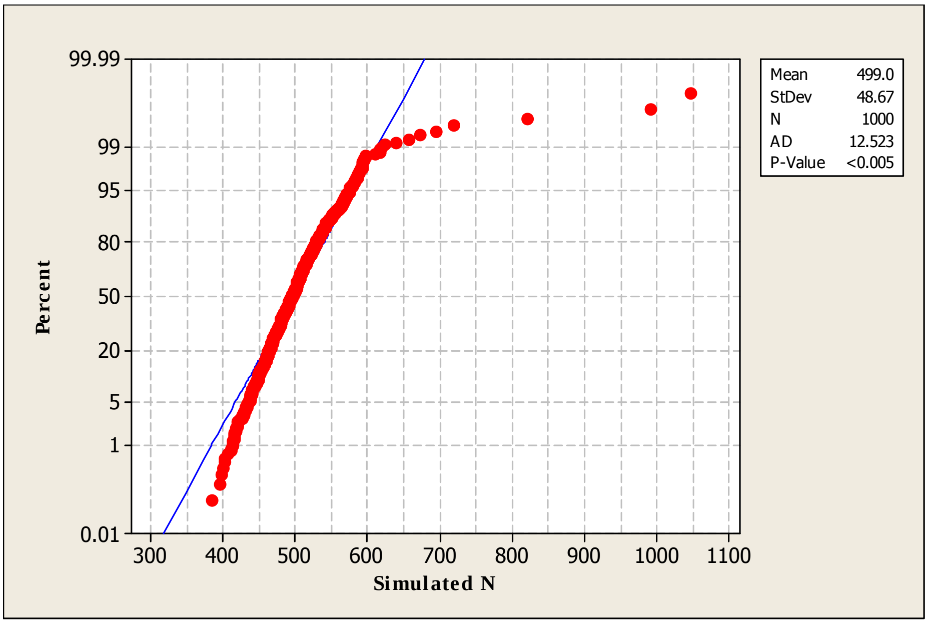

Figure 5.10. Normal probability plot of the actual sample size N for underlying distribution $t(5)$ using AD test at n^*=500, $\delta = 0.5$ and $m = 15$; p – value < 0.005.

Similarily, Table 5.25 shows the p-values for tests of normality of N when the underlying distribution is $t(25)$ and with $m = 15$, $\delta = 0.5$ and $n^* = 76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	75.123	9.642	-0.26	3.16	<0.005	0.039
96	94.035	11.201	-0.46	3.78	<0.005	<0.01
125	124.09	12.470	-0.14	3.27	0.245	>0.150
171	169.80	14.280	-0.10	3.43	0.123	>0.150
246	244.70	18.410	-0.13	3.36	0.029	0.107
500	499.49	25.760	-0.11	3.04	0.159	0.118

Table 5.25: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying $t(25)$ as the optimal sample size increases.

Figure 5.11 shows the normal probability plot of the simulated estimates of N at n*=500 when the underlying distribution is $t(25)$ at $n^* = 500$. Clearly the normality of N seems plausible here.

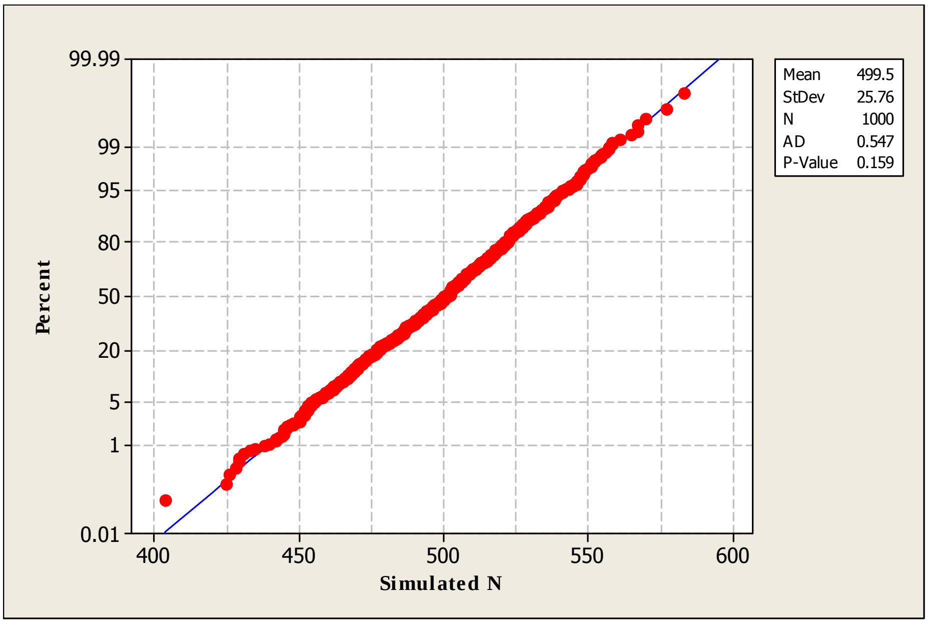

Figure 5.11. Normal probability plot of the actual sample size N for underlying distribution $t(\,25\,)$ using AD test at n^*=500, $\delta = 0.5$ and $m = 15$; p – value = 0.159.

Similarily, Table 5.26 shows the p-values for tests of normality of N when the underlying distribution is $t(\,50\,)$ and with $m = 15$, $\delta = 0.5$ and $n^* = 76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	75.279	9.329	-0.25	3.20	<0.005	<0.01
96	94.822	10.802	-0.20	3.22	0.009	0.097
125	124.00	12.36	-0.20	3.34	0.01	<0.01
171	170.39	14.01	-0.25	3.43	<0.005	0.033
246	245.63	16.63	-0.14	3.18	0.044	>0.150
500	499.27	23.16	-0.10	3.08	0.577	>0.150

Table 5.26: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying $t(50)$ as the optimal sample size increases.

As expected, the higher value of r, the nearest the kurtosis to the normal distribution, and thus more acceleration to normality. Figure 5.12 shows the normal probability plot for the simulated estimates of N at n^* = 500. By using AD test the p – value is 0.577. Clearly the normality is strong here.

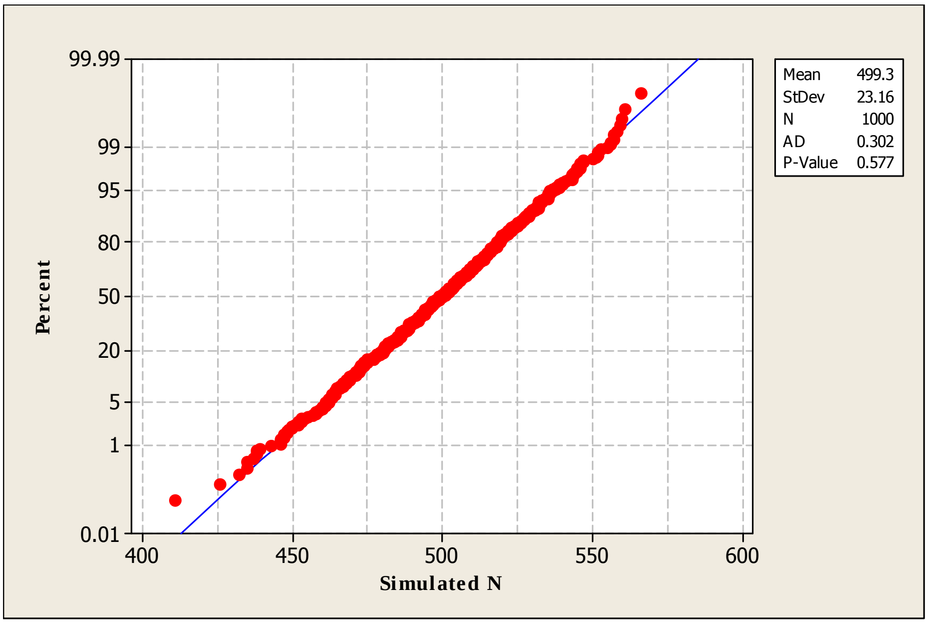

Figure 5.12. Normal probability plot of the actual sample size N for underlying distribution $t(50)$ using AD test at n^*=500, $\delta=0.5$ and $m=15$; p – value = 0.577.

5.4.4. Beta $(2,3)$ distribution

Similarily, Table 5.27 shows the p-values for tests of normality of N when the underlying distribution is beta $(2,3)$ and with $m=15$, $\delta=0.5$ and $n^*=76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	75.216	7.893	-0.6	3.95	<0.005	<0.01
96	95.202	8.680	-0.36	3.53	<0.005	0.148
125	124.66	9.72	-0.23	3.04	<0.005	0.037
171	169.98	11.22	-0.27	3.90	<0.005	0.039
246	245.22	13.68	-0.28	3.20	<0.005	0.064
500	499.20	20.33	-0.10	2.91	0.195	>0.150

Table 5.27: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying beta$(2,3)$ as the optimal sample size increases.

Figure 5.13 shows the normal probability plot for the simulated estimates of N at $n^*=500$. By using AD test the p – value is 0.195, so we have a satisfactory evidence for normality here.

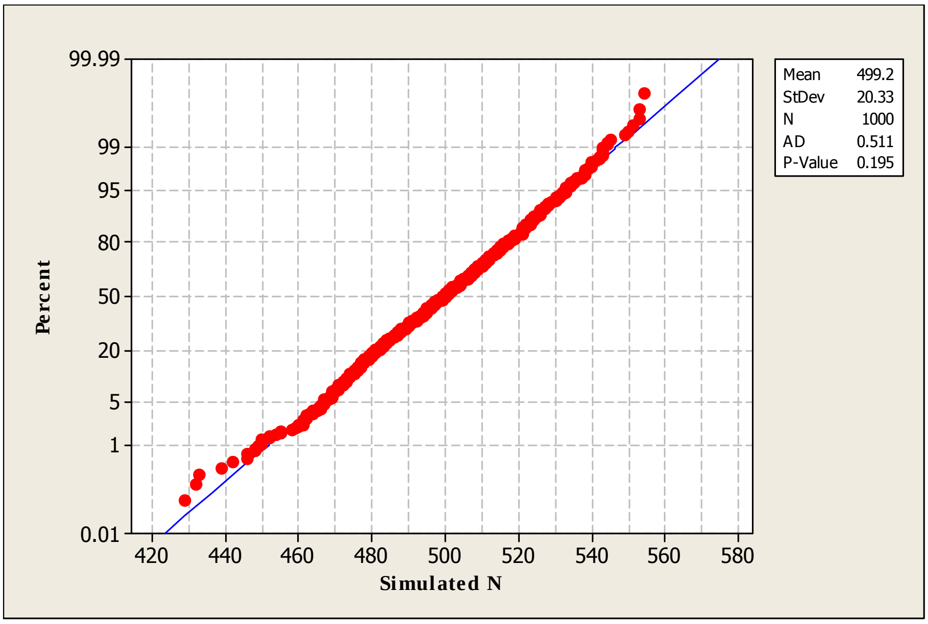

Figure 5.13. Normal probability plot of the actual sample size N for underlying distribution beta $(2,3)$ using AD test at n^*=500, $\delta = 0.5$ and $m = 15$; p – value = 0.195.

5.4.5. Exponential distribution with mean one

Similarily, Table 5.28 shows the p-values for tests of normality of N when the underlying distribution is $Exp(1)$ and with $m = 15$, $\delta = 0.5$ and $n^* = 76$, 96, 125, 246 and 500 using the AD and KS tests. It also shows the basic descriptive measures of the distribution of N at each value of n^*: sample mean, sample standard deviation, sample skewness and sample kurtosis.

n^*	$\overline{N}$	$s.d.(\overline{N})$	$\hat{\gamma}$	$\hat{\beta}$	P_{AD}	P_{KS}
76	70.205	17.853	-0.20	3.58	<0.005	>0.150
96	91.135	20.012	0.27	3.35	<0.005	0.062
125	120.11	23.280	-0.15	3.51	<0.005	0.033
171	167.00	28.440	0.02	3.48	0.321	>0.150
246	240.77	33.990	0.56	6.61	0.028	>0.150
500	493.90	48.360	0.19	3.81	<0.005	0.06

Table 5.28: The Descriptives: mean, standard deviation, skewness and kurtosis and the p –values for the AD and KS statistic for testing the asymptotic normality of N for underlying $Exp(1)$ as the optimal sample size increases.

Figure 5.14 below shows the normal probability plot of the simulated estimates of N for underlying exponential distribution at $n^* = 500$. By using AD test the p – value is less than 0.005. Note the non-normal tail behaviour. Clearly convergence to normality is slow.

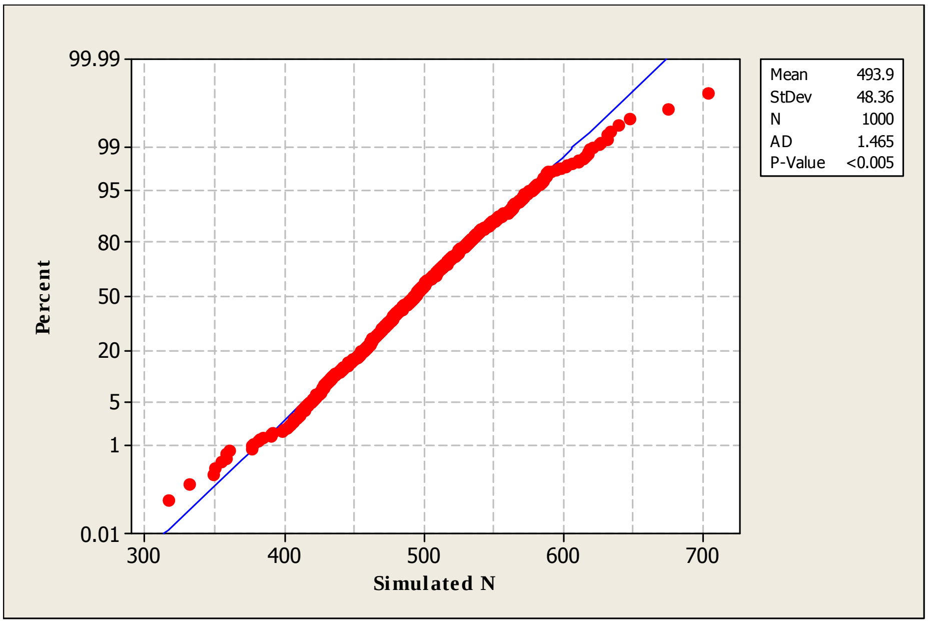

Figure 5.14. Normal probability plot of the actual sample size N for underlying distribution $Exp(1)$ using AD test at n^*=500, $\delta = 0.5$ and $m = 15$; p – value < 0.005.

5.5 The behaviour of the asymptotic regret

From Theorem 4.2.1 the asymptotic regret of the triple sampling procedure $(2.2)-(2.3)$ under the squared error loss function (4.1) as $A \to \infty$ is

$$(5.5)\quad \omega(A) = -C(\beta-3) + C(\beta-1)(4\delta)^{-1} + C\,E(\varepsilon_{N_1}) + o(1).$$

As stated in Chapter IV, the asymptotic regret depends mainly on the kurtosis of the underlying distribution. Moreover $\lim_{c\to 0} C^{-1}\omega(A) < \infty$, which means that the triple sampling procedure with (5.5) have asymptotically second order risk efficient. Equation (5.5) means that the regret due to using the triple sampling procedure presented in $(2.1)-(2.2)$ in ignorance of the population variance θ is bounded above by the quantity $-C(\beta-3)+C(\beta-1)(4\delta)^{-1}+CE(\varepsilon_{N_1})+o(1)$ as $m \to \infty$. That is in the limit, one loses at most the cost of $-C(\beta-3)+C(\beta-1)(4\delta)^{-1}+CE(\varepsilon_{N_1})$ observations when using the stopping rule (2.2) instead of the optimal fixed sample size.

Woodroofe (1977) showed that the regret in estimating the normal mean by using the one-by-one purely sequential procedure under the loss function given in (4.1) is $\omega = 1/2 + o(1)$ as $A \to \infty$. In other words, for normal data the asymptotic regret is half the cost of a single observation. Martinsek (1983) extended Woodroofe's result to the non-parametric case (distribution free), showing that the asymptotic regret of using the one-by-one purely sequential procedure instead of the optimal fixed sample size procedure is $\omega = 2.75 - 0.75\beta + 2\gamma^2 + o(1)$, as $A \to \infty$. He further extended his results to a more general loss function given by

$$L_n(A) = A\theta^{b-1}\left(\bar{X}_n - \mu\right)^2 + n,\ A > 0, b > 0.$$

He showed that under this loss function, the asymptotic regret in using the one-by-one purely sequential procedure instead of the optimal fixed sample size procedure is

$$\omega = \left(b^2/4\right)(\beta - 1) - b(\beta - 3) + b(b+1)\gamma^2 + o(1), \text{ as } A \to \infty.$$

See Martinsek (1988) for details.

Martinsek (1988) argued that the early stopping phenomenon may cause negative regret, and in this case the triple sampling sequential procedure could perform better estimation than the optimal. Negative regret occurs when the triple sampling sequential risk is less than the optimal risk, and may be due to the dependency between N and $\bar{X}_N$. For more details about negative regret, see Martinsek (1988) and Takada (1992).

In our case we can expect negative regret when $\beta > (14\delta - 1)/(4\delta - 1)$, provided that $\delta \neq 1/4$.

We now compare the regret from the main simulation experiment and the asymptotic regret at $\delta = 0.5$ and $m =$ 5, 15 and 20 for the same class of distributions as before. Each table given below includes $\omega_s(m)$ and ω, where $\omega_s(m)$ is the simulated regret at the initial sample size m, and ω is the asymptotic regret.

5.5.1 Standard normal distribution

From (5.5) the asymptotic regret under the normal optimal fixed sample size and taking $C = 1$ is

$$\omega = (2\delta)^{-1} + 1/2 + o(1).$$

Thus the asymptotic regret is bounded by a finite non-vanishing positive quantity $(2\delta)^{-1} + 1/2$, which depends on δ. In particular at $\delta = 0.5$ the regret is at most 1.5 times the cost of one observation.

Table 5.29 illustrates the effect of m on the regret as the optimal sample size increases.

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	5.0388	5.2289	0.7149	1.5
43	4.0344	3.6876	10.3735	1.5
61	2.5455	1.1441	2.3210	1.5
76	2.0155	1.0326	2.2575	1.5
96	0.7541	1.1022	1.6784	1.5
125	1.9341	0.9575	1.5963	1.5
171	2.6415	2.1620	2.0388	1.5
246	2.2415	0.7743	2.5890	1.5
500	3.6976	-0.0303	0.3383	1.5

Table 5.29: Comparison between the simulated estimates of the regret and the asymptotic regret for underlying $N(0,1)$ at $m = 5,15,20$ and $\delta = 0.5$

The asymptotic regret values at $\delta = 0.3$, 0.5 and 0.8 are respectively $\omega = 2.166667$, 1.5 and 1.125. The reason of having simulated regret far from the asymptotic value goes to the fact that our chosen optimal sample sizes are not large enough to ensure the limiting regret. But, still we have a non-vanishing regret with small quantities and not in a disordered manner.

5.5.2 Standard uniform distribution

From (5.5) the asymptotic regret for underlying standard uniform distribution and under the normal optimal fixed sample size and taking $C = 1$ is

$$\omega = 1.2 + 0.2\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1).$$

The above equation states that the asymptotic regret is bounded by a finite non-vanishing positive quantity that depends on δ and $E\left(\varepsilon_{N_1}\right)$. At $\delta = 0.5$ the asymptotic regret is $\omega = 2.1$.

Table 5.30 shows the behaviour of the regret at $\delta = 0.5$ and as m increases and the asymptotic regret.

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	11.3524	5.4707	0.5572	2.1
43	6.6890	5.2190	15.6292	2.1
61	4.9614	2.9704	2.5278	2.1
76	4.6738	2.1537	2.3136	2.1
96	2.4396	1.7100	2.6052	2.1
125	2.9169	1.2865	0.9192	2.1
171	0.7481	1.4656	2.1333	2.1
246	2.3900	5.9230	3.8632	2.1
500	4.1204	7.6406	1.4336	2.1

Table 5.30: Comparison between the simulation estimates of the regret and the asymptotic regret for underlying $U(0,1)$ at $m = 5,15,20$ and $\delta = 0.5$

The asymptotic regret at $\delta = 0.3$ and 0.8 are respectively $\omega = 2.366667$ and 1.95.

5.5.3 The t distribution

From (5.5) the asymptotic regret for underlying t distribution and under the normal optimal fixed sample size and taking $C = 1$ at $r = 5, 25$ and 50 are respectively,

$$\omega = -6 + 2\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1),$$

$$\omega = -2/7 + (4/7)\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1),$$

and

$$\omega = -3/23 + (49/92)\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1).$$

Clearly the regret depends on both the design factor δ and on $E\left(\varepsilon_{N_1}\right)$. The asymptotic regret values at $\delta = 0.5$ and $r = 5, 25$ and 50 are -1.5 , 1.35714285 and 1.4347826 respectively. We noticed that as r increases the regret values also increases, and this agree to the fact that as r increases we get closer to the normal distribution which have a positive regret.

Tables 5.31, 5.32 and 5.33 show the behaviour of the regret as m increases and also as r increases.

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	2.5043	2.1192	-0.1202	-1.5
43	2.0134	1.6830	5.5412	-1.5
61	1.6291	0.7942	1.4883	-1.5
76	1.8168	0.5190	0.2072	-1.5
96	1.4098	0.0544	0.1309	-1.5
125	0.7889	0.3386	-0.4909	-1.5
171	-0.1928	-1.4392	-0.4245	-1.5
246	0.2370	-1.8762	-1.9978	-1.5
500	-1.7375	4.4777	-0.7673	-1.5

Table 5.31: Comparison between the simulated estimates of the regret and the asymptotic regret for underlying distribution $t(5)$ at $m = 5, 15, 20$ and $\delta = 0.5$

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	4.6844	5.0694	0.9021	1.3571
43	3.6912	3.9828	9.0250	1.3571
61	2.7728	1.3662	2.5277	1.3571
76	1.7548	1.0648	1.4274	1.3571
96	2.2458	2.1484	0.8447	1.3571
125	1.9584	1.4517	1.3883	1.3571
171	0.2642	0.0084	2.1650	1.3571
246	-0.9581	1.4579	-0.2633	1.3571
500	-0.0232	1.5594	6.2678	1.3571

Table 5.32: : Comparison between the simulated estimates of the regret and the asymptotic regret for underlying distribution $t(25)$ at $m = 5, 15, 20$ and $\delta = 0.5$

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	4.8482	4.8474	0.8401	1.4348
43	3.9371	3.6657	9.0564	1.4348
61	2.9058	1.0596	2.6648	1.4348
76	3.2878	0.7468	1.2739	1.4348
96	1.8918	1.6086	0.5378	1.4348
125	2.0669	2.1794	0.3645	1.4348
171	0.9446	0.9590	1.2998	1.4348
246	1.2752	4.1042	3.1244	1.4348
500	4.7263	1.0304	5.6271	1.4348

Table 5.33: : Comparison between the simulated estimates of the regret and the asymptotic regret for underlying distribution $t(50)$ at $m = 5, 15, 20$ and $\delta = 0.5$

5.5.4 Beta $(2,3)$ distribution

From (5.5) the asymptotic regret under the normal optimal fixed sample size and taking $C = 1$

$$\omega = 9/14 + (19/56)\delta^{-1} + E(\varepsilon_{N_1}) + o(1).$$

Similarily the asymptotic regret depends on δ and on the value of $E(\varepsilon_{N_1})$. In particular at $\delta = 0.5$ the asymptotic regret is 1.821428572.

Table 5.34 shows the effect of increasing the initial sample size m on the performance of the regret as the optimal sample size increases. Clearly, the simulated estimates of the regret are all positive.

The regret values at $\delta = 0.3$ and 0.8 are respectively 2.2738095 and 1.5669643.

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	7.8013	5.1932	0.6887	1.8214
43	5.2428	5.6713	13.3263	1.8214
61	4.0645	2.5505	3.2058	1.8214
76	3.8069	1.8228	3.0726	1.8214
96	3.5709	1.3700	2.8478	1.8214
125	3.6107	1.2696	1.4720	1.8214
171	2.8993	2.1447	1.3234	1.8214
246	2.4580	0.2243	4.0616	1.8214
500	1.1577	-0.3856	3.6948	1.8214

Table 5.34: Comparison between the simulated estimates of the regret and the asymptotic regret for underlying beta $(2,3)$ at $m = 5,15,20$ and $\delta = 0.5$

5.5.5 Exponential distribution with mean one

From (5.5) the asymptotic regret under the normal optimal fixed sample size and taking $C = 1$

$$\omega = -6 + 2\delta^{-1} + E\left(\varepsilon_{N_1}\right) + o(1).$$

The above equation states that the limiting regret is bounded by a non vanishing negative quantity regardless the value of δ and the value of $E\left(\varepsilon_{N_1}\right)$. In particular at $\delta = 0.5$ the asymptotic regret is -1.5. Therefore, one would expect negative regret values.

From Table 5.35 we noticed mostly positive values and little negative values of the regret. The reason behind this is due to the large value of the kurtosis $\beta = 9$ and the optimal sample sizes are not large enough. We may also add that the convergence is from above (i.e. through positive values of regret). This indicates that the sequential procedures are risky as the optimal fixed sample size procedure.

n^*	$\omega_s(5)$	$\omega_s(15)$	$\omega_s(20)$	ω
24	17.7940	-1.7872	-0.7321	-1.5
43	32.3492	17.3048	6.4769	-1.5
61	32.0484	22.4790	25.3587	-1.5
76	32.2888	18.6754	23.0115	-1.5
96	26.3778	14.0712	15.7395	-1.5
125	23.7438	11.8602	10.3143	-1.5
171	16.7938	8.7925	8.7394	-1.5
246	10.2735	9.2764	9.0376	-1.5
500	13.6584	7.3364	4.0084	-1.5

Table 5.35: : Comparison between the simulated estimates of the regret and the asymptotic regret for underlying $Exp(1)$ at $m = 5,15,20$ and $\delta = 0.5$

Finally, the asymptotic regret at $\delta = 0.3$ and 0.8 are respectively 1.16667 and -3.

It is now clear that the triple sampling procedures with normal stopping rule can result extremely accurate estimates of the targeted parameters with non normal underlying distributions.

5.6 The probability of early stopping

We also estimated the probability of early stopping in our simulation. During the simulations we calculated the number of times the triple sampling procedure stops at the first stage, second stage and third stage. Then we calculated the relative frequency of terminating the procedure at each stage. In the tables in Appendix B we denote the estimated probabilities of stopping at the first, second and third stages as $p(m)$, $p(N_1)$ and $p(N)$ respectively. Note that the sum of these probabilities is equal to one.

When the underlying distribution is uniform, if the design factor δ is less than 0.5 and for small starting sample sizes m, the procedure tends to terminate at the first or third stage. However, for fixed δ, as m increases, the procedure terminates at the third stage almost surely for large values of n^* ($\lambda \to \infty$). As δ increases, the probability of stopping at the second stage increases with m. Specifically, when the initial sample size increases to 20 and at $\delta = 0.5$ and 0.8 the procedure terminates at the third stage almost surely.

When the underlying distribution is exponential, the triple sampling exhibits almost the same pattern except that the convergence to one of $p(N)$ is slower than in the case of the uniform distribution. This could be due to the skewness of the exponential distribution, see Appendix B for all distributions.

The Tables in Appendix B support our conjecture regarding the almost sure termination of the triple sampling procedure at the third stage, see, chapter III, Theorem 3.2.2.2.

5.7 Investigation of the asymptotic distribution of ε_{N_1}

In the main simulation experiment we also investigated the distribution of the rounded off portion of the final stage sample size N. Hall (1981) proved that the random variable ε_{N_1} that appears in our Theorem 3.2.2.2 is asymptotically uniform over $(0,1)$ as $n^* \to \infty$. His proof only works if the underlying distribution is normal. A generalization of Hall's result to other underlying distributions is not currently available. Meanwhile, we have tested the distribution of the continuous part of the final stage during simulation for a particular combination of δ, m, α and n^*. We saved the values of ε_{N_1} for each replicate. The Kolmogorov-Smirnov test (KS) was used to test the simulated ε_{N_1} values for uniformity.

Tables C1, C2 and C3 in Appendix C, give summary statistics regarding the distribution of the random variable ε_{N_1} under three different classes of distributions: normal, uniform and exponential at $\delta = 0.5$ and $m = 5$, 15 and 20.

From Tables C1, C2 and C3 with $m = 15$ and $n^* = 500$, the p – values of the KS statistic are 0.154, 0.303 and 0.439 respectively for the three underlying distributions. Thus, asymptotic uniformity over a range of underlying distributions seems plausible. Further, in Theorem 3.2.2.2 we have $E(\varepsilon_{N_1})$ is equal to $E\left\{1-\left(\lambda g\left(S_{N_1}^2\right)-\left[\lambda g\left(S_{N_1}^2\right)\right]\right)\right\}$. In all the simulations this value was close to 0.5. More details concerning the behaviour of the distribution of ε_{N_1} at $m=5$ and $m=20$ can be found in Appendix C.

Before we end up this chapter we would like to show the performance of but the method for larger values of n*. We consider here $n^*=500$, 1000, 2000, 2500, 3000, 3500, 4000, 4500 and 5000 and for brevity only three underlying distributions: normal, uniform and exponential. Tables 5.29, 5.30 and 5.31 show the effect of increasing the optimal sample size on the performance at $m=15$ and $\delta=0.5$. In particular they show the behaviour of N, $\overline{X}_N$ and the regret at the above values of n^*. We see excellent performance for all estimates in comparison with the previous selection of n^*. Moreover the regret here is mostly negative for all the above underlying distributions. This indicates that the triple sampling procedure performs better than the fixed sample size procedure for large values of n^*.

n^*	$\overline{N}_s(15)$	$ssd\overline{N}(15)$	$\widehat{\mu}_s(15)$	$ssd(0.5)$	$\omega_s(15)$
500	498.92	23.2551	0.000109	0.0447214	0.486
1000	999.04	32.6466	0.000122	0.0315286	-6.474
2000	1998.94	46.2866	-0.000040	0.0223607	-14.264
2500	2499.21	51.2060	-0.000103	0.0199010	-34.768
3000	2998.97	56.3489	-0.000061	0.0181122	-41.039
3500	3498.81	61.2683	-0.000002	0.0167705	-62.163
4000	3999.19	65.2932	-0.000078	0.0156525	-112.354
4500	4499.31	69.5417	-0.000076	0.0147580	-124.769
5000	4999.23	73.1194	-0.000039	0.0138636	-143.308

Table 5.36: Large sample performance for all estimates when the underlying distribution is $N(0,1)$ and $m=15, \delta=0.5$.

n^*	$\overline{N}_s(15)$	$ssd\overline{N}(15)$	$\widehat{\mu}_s(15)$	$ssd(0.5)$	$\omega_s(15)$
500	499.84	14.5344	0.500138	0.0129692	-1.6559
1000	999.80	20.3482	0.499971	0.0091679	5.7754
2000	1999.99	28.8453	0.499989	0.0064846	-31.9243
2500	2499.80	31.9758	0.499978	0.0058138	-13.7676
3000	2999.81	35.5535	0.499989	0.0051430	-60.8281
3500	3499.81	38.0132	0.500016	0.0049193	-56.5469
4000	4000.00	40.6964	0.499992	0.0044721	-18.3623
4500	4499.98	42.9325	0.500011	0.0042485	-49.5195
5000	4999.69	45.3922	0.500009	0.0040249	-72.9023

Table 5.37: Large sample performance for all estimates when the underlying distribution is $U(0,1)$ and $m=15, \delta=0.5$.

n^*	$\overline{N}_s(15)$	$ssd\overline{N}(15)$	$\hat{\mu}_s(15)$	$ssd(0.5)$	$\omega_s(15)$
500	494.94	48.299	0.997648	0.0451686	6.024
1000	995.59	69.989	0.998916	0.0317522	2.384
2000	1996.58	100.847	0.999557	0.0223607	-18.247
2500	2496.27	113.369	0.999623	0.0199010	-22.274
3000	2997.85	127.009	0.999733	0.0181122	-38.203
3500	3497.97	135.953	0.999742	0.0167705	-70.302
4000	3997.72	146.239	0.999720	0.0156525	-41.623
4500	4498.20	154.512	0.999901	0.0147580	-130.177
5000	4998.80	168.152	0.999914	0.0140872	-76.789

Table 5.38: Large sample performance for all estimates when the underlying distribution is $Exp(1)$ and $m = 15, \delta = 0.5$.

We prefer the asymptotically negative regret values if they provide better estimates because this indicate that the triple sampling procedure does better than the fixed sample size . Otherwise they may delay the procedure and cause early stopping and consequently obtain bad estimates.

Chapter VI

Triple Sampling Fixed Width Confidence Intervals for the Population Mean

6.1 Introduction

As before we assume nothing about the underlying distribution except that the first six moments are finite. The objective of this chapter is to construct a fixed width confidence interval for the population mean using the triple sampling procedure as presented in $(2.2)-(2.3)$. The approach used here involves using second order Edgeworth approximations. Moreover, we will find the coverage asymmetric and symmetric confidence intervals. Then we will discuss the sensitivity of triple sampling fixed width confidence intervals to shifts in the population mean.

In the next section we shed light on some characteristics of the Edgeworth expansion and its limitations.

6.2 Edgeworth asymptotic expansion

Let $X_1, \ldots, X_n$ be a collection of $i.i.d.$ random variables drawn from the distribution function $F(.;\mu,\theta)$, with finite population mean μ and variance $\theta < \infty$. Let $F_n(x) = P(Z_n \le x)$, where $Z_n = \sqrt{n}\left(\bar{X}_n - \mu\right)\big/\sqrt{\theta}$. Then by the central limit theorem $\lim_{n\to\infty} F_n(x) = \Phi(x)$, for every fixed x and $\Phi(x)$ is the standard normal cumulative distribution function.

Cramer's Condition

A cumulative distribution function $F(.;\mu,\theta)$ defined over $\Re$ satisfies Cramer's condition if $\lim_{t\to\infty} Sup \left|E_F\left(e^{itX}\right)\right| < 1$.

Remark

All absolutely continuous distribution functions satisfy Cramer's condition.

The Two-term Edgeworth Expansion Theorem (see, DasGupta 2008) is

Theorem 6.1.1 (Two-term Edgeworth Expansion)

Suppose the distribution function, $F(.;\mu,\theta)$ satisfies Cramer's condition and $E(X^4)<\infty$. Then

$$F_n(x)=\Phi(x)+c_1 p_1\phi(x)/\sqrt{n}+(c_2 p_2+c_3 p_3)\phi(x)/n+O(n^{-3/2})$$

uniformly in x and as $n\to\infty$, where $\phi(x)$ and $\Phi(x)$ are the standard normal density and distribution functions, $c_1=\frac{\gamma}{6}$, $c_2=\frac{(\beta-3)}{24}$, $c_3=\frac{\gamma^2}{72}$, $p_1=1-x^2$, $p_2=3x-x^3$ and $p_3=-15x+10x^3-x^5$.

Proof: See Kokic *et al.* (1990) and Hall (1992) for the proof.

Note that c_1 and c_2 are called the skewness and kurtosis corrections of the underlying distribution function. The Edgeworth expansion is an improvement over the central limit theorem, which fails to take account of the skewness in the distribution of the sample mean of a given finite sample of size n. Expanding successive terms will capture both the skewness and the kurtosis of the underlying distribution. Using further terms in the expansion may cause it to become unstable because of the presence of higher order polynomials. We therefore restrict our expansion to the first two terms. The error of the leading term in the expansion, the standard normal density, is of $O(n^{-1/2})$ provided that $\gamma\neq 0$. This suggests that convergence to normality is relatively slow, especially in the tails of the underlying distribution (see Barndorff-Nielsen and Cox, 1989, chapter 4). Bhattacharya *et al.* (1978) mentioned that the error $O(n^{-3/2})$ can be improved if one makes more stringent moment assumptions.

We must stress that the Edgeworth expansion in Theorem 6.1.1 is only an asymptotic expansion and not a convergent series. This means that if the expansion is stopped after a specific number of terms, then the remainder will be smaller than the last term that has been included; see Hall (1992).

From Theorem 6.1.1, by direct substitutions for c_1,c_2,c_3,p_1,p_2 and p_3 and by taking $n=1$, we have the heuristic result

$$(6.1)\quad F(x)=\Phi(x)-\phi(x)x(x^4-10x^2+15)\gamma^2/72-\phi(x)(x^2-1)\gamma/6$$
$$-\phi(x)x(x^2-3)(\beta-3)/24+O(1).$$

If the distribution function $F(\cdot)$ of an absolutely continuous random variable admits an Edgeworth expansion, then we can obtain an expansion of the density function heuristically by differentiating (6.1) with respect to x. Hence the probability density function of the Edgeworth expansion is

$$(6.2)\quad f(x)=\phi(x)\begin{Bmatrix}1+\gamma x\left(x^2-3\right)/6+(\beta-3)\left(x^4-6x^2+3\right)/24\\+\gamma^2\left(x^6-15x^4+45x^2-15\right)/72\end{Bmatrix}+O(1).$$

Equation (6.2) shows how to represent a continuous probability density function $f(x)$ in terms of the standardized normal probability density function and which is an approximating standardized density with the desired γ and β. The density function $f(x)$ in (6.2) is called a standardized Edgeworth asymptotic expansion; see Johnson *et al.* (1994).

The Edgeworth expansion is more useful in many applications than asymptotic series such as the Gauss-Hermite and Gram-Charlier series. This so because first, it is directly connected to the moments and cumulants of a probability density functions, a property that is lost in the Gauss-Hermite series. Secondly, it is a true asymptotic expansion since the error of the approximation is controlled by estimating the error of the expansion until the order $O\left(n^{-3/2}\right)$. It is worth mentioning the disadvantage of the Edgeworth series as an approximation to the standardized density function. It was shown by Barton and Dennis (1952) that the Edgeworth series can give negative values for some values of x. They found the region in the plane of values of skewness and kurtosis where the density is positive. This region was further studied by Draper and Tierney (1972) in detail using numerical methods. They found that the validity region that ensures the Edgeworth series to represent a positive definite and unimodal probability density function is $R_e=\{|\gamma|\le 0.45, 3.0\le\beta\le 5.35\}$. If the parameters are lie outside the validity region (as we shall see in the forthcoming examples), the results may be misleading. Furthur analytical investigations about the validity region were undertaken by Balitskaya and Zolotuhina (1988).

The first few terms of the series (6.2) are given in standard references (such as Cramer, 1957, Abramowitz and Stegun, 1972, and Juszkiewicz *et al.*, 1995). References for the main results on Edgeworth expansions are Bhattacharya and Ghosh (1978), Barndorff-Nielsen and Cox (1989), Hall (1992) and Lahiri (2003).

We now give some examples of the Edgeworth asymptotic expansion involving the uniform, $t(r)$ and chi-squared distributions.

6.2.1. Uniform distribution $U(a,b)$

From (6.2), the Edgeworth asymptotic expansion for the standardized uniform distribution $U(a,b)$ is

$$f(x)=\phi(x)\left\{0.85+0.3x^2-0.05x^4\right\}+O(1),$$

while the standardized uniform density function is

$$g(x)=\frac{1}{2\sqrt{3}},\ -\sqrt{3}<x<\sqrt{3}.$$

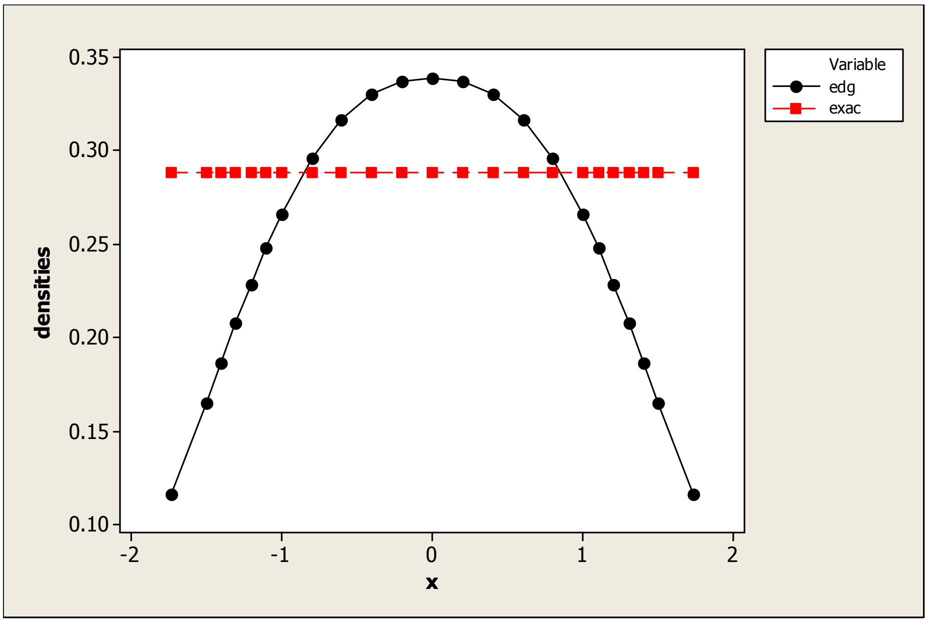

Figure 6.1. The standardized uniform density and its Edgeworth approximation

Figure 6.1 shows that Edgeworth approximation to the standardized uniform density is poor. Note that the kurtosis of the uniform distribution, $\beta = 1.8$ lies outside the validity region.

6.2.2. The t distribution

From (6.2), the Edgeworth asymptotic expansion for the standardized $t(r)$ distribution is

$$f(x) = \phi(x)\left\{1 + \frac{1}{4(r-4)}\left(x^4 - 6x^2 + 3\right)\right\} + O(1), \text{ defined for all } r > 4.$$

The standardized density for $t(r)$ is

$$g(x) = \frac{\Gamma((r+1)/2)}{\Gamma(r/2)\sqrt{\pi(r-2)}}\left(1 + \frac{x^2}{r-2}\right)^{-\left(\frac{r+1}{2}\right)}, -\infty < x < \infty,\ r > 2.$$

The Edgeworth expansion for $t(r)$ depends on r and clearly $\lim_{r\to\infty} f(x) = \phi(x) + O(1)$.

To illustrate further the role of r, consider $r = 5,10$ and 20.

Case (I): $r = 5$

Figure 6.2 shows the poor performance of the Edgeworth series for the standardized density of the t distribution with 5 degrees of freedom. Here the kurtosis lies outside the validity region.

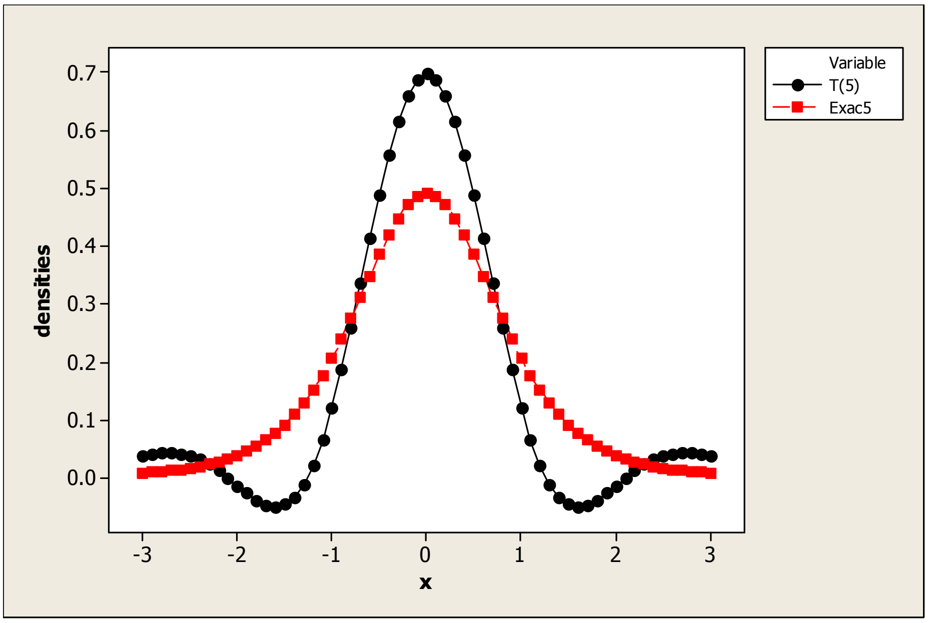

Figure 6.2. The standardized $t(5)$ density and its Edgeworth approximation

Case (II): $r = 10$

Figure 6.3 shows the good performance of the Edgeworth approximation for the standardized density of the $t(10)$. Here the skewness and the kurtosis lie inside the validity region.

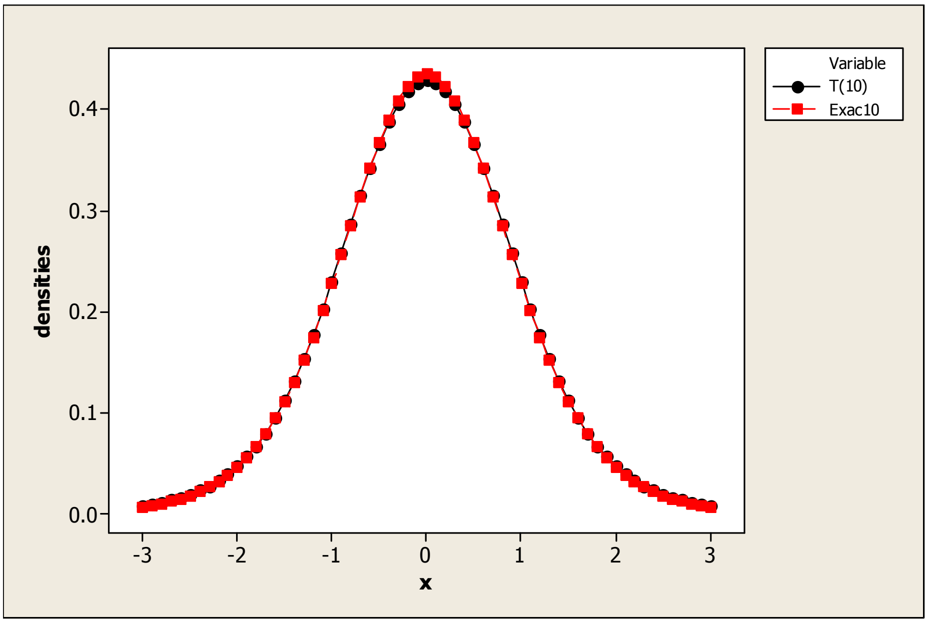

Figure 6.3. The standardized $t(10)$ density and its Edgeworth approximation

Case (III): $r = 20$

Similarly, Figure 6.4 shows the good performance of the Edgeworth approximation to the standardized density of the t distribution at $r = 20$. Both the skewness and the kurtosis lie inside the validity region.

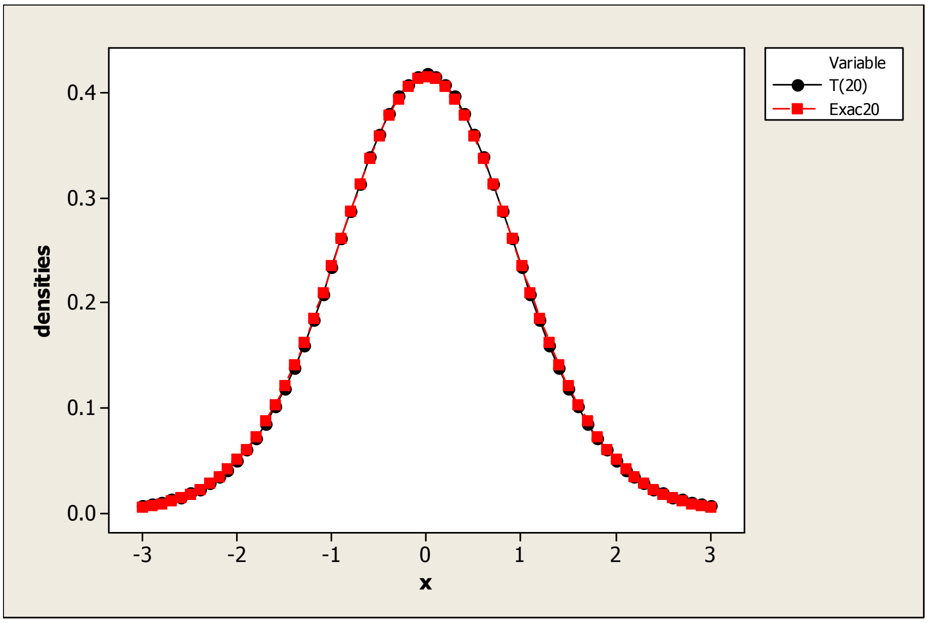

Figure 6.4. The standardized $t(\ 20\ )$ density and its Edgeworth approximation

6.2.3. Chi-Squared distribution

The Edgeworth asymptotic expansion for the standardized chi-square distribution with r degrees of freedom is

$$f(x) = \phi(x)\left\{\frac{6\sqrt{2r}\,x\left(x^2-3\right)+3(6r-1)+x^2\left(2x^4-21x^2+36\right)}{18r}\right\}+O(1).$$

The standardized chi-squared density is

$$g(x) = \frac{\sqrt{r\,2^{1-r}}}{\Gamma(r/2)}\left(r+\sqrt{2r}\,x\right)^{\frac{r}{2}-1}\exp\left(\frac{-\left(r+\sqrt{2r}\,x\right)}{2}\right),\ x > -\sqrt{r/2}\,.$$

The Edgeworth expansion of the chi-square distribution depends on the degrees of freedom. Note also that $\lim_{r\to\infty} f(x) = \phi(x) + O(1)$.

To illustrate the effect of the increase of the degrees of freedom r on the accuracy of the Edgeworth approximation, we take $r = 2, 5$ and 10.

Case (I): $r = 2$

Clearly, $r = 2$ yields the case of an exponential distribution with mean two, and the Edgeworth expansion for this case is

$$f(x) = \phi(x)\left\{\frac{1}{18}x^6 - \frac{7}{12}x^4 + \frac{1}{3}x^3 + x^2 - x + \frac{11}{12}\right\} + O(1).$$

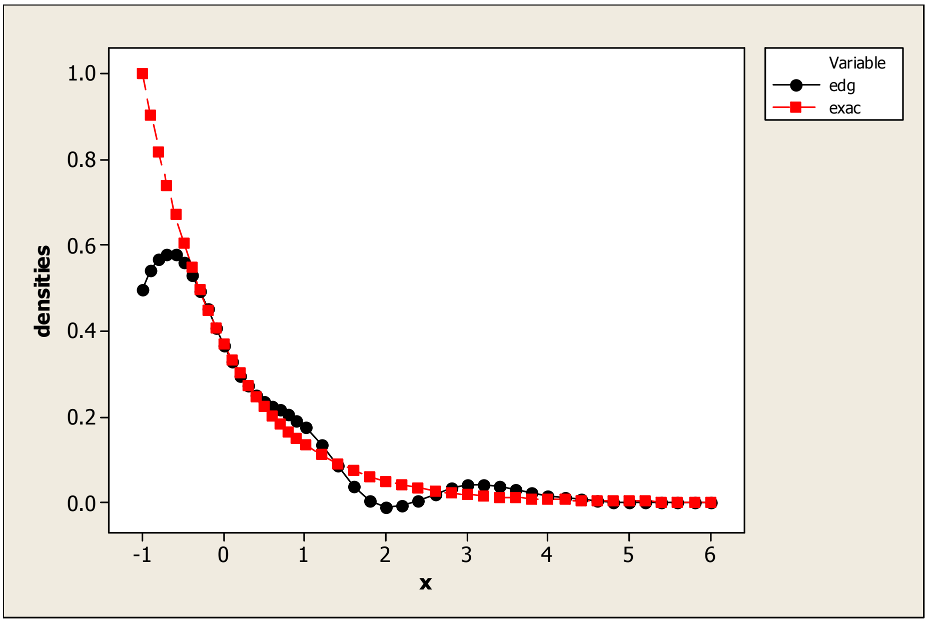

Figure 6.5. The standardized exponential density and its Edgeworth approximation

Figure 6.5 shows the poor performance of the Edgeworth expansion in the case of an exponential distribution, and the highest order approximation behaves poorly in the upper tail and the oscillations in the tail are more severe. Here the skewness and kurtosis, $\gamma = 2$, $\beta = 9$, are both outside the validity region.

Case (II): $r = 5$

This case leads to a better performance than in Case (I). Note that $\gamma = \sqrt{8/5}$, $\beta = 5.4$, both of which are outside the validity region. However, the value of the kurtosis is close to the upper bound specified in the validity region.

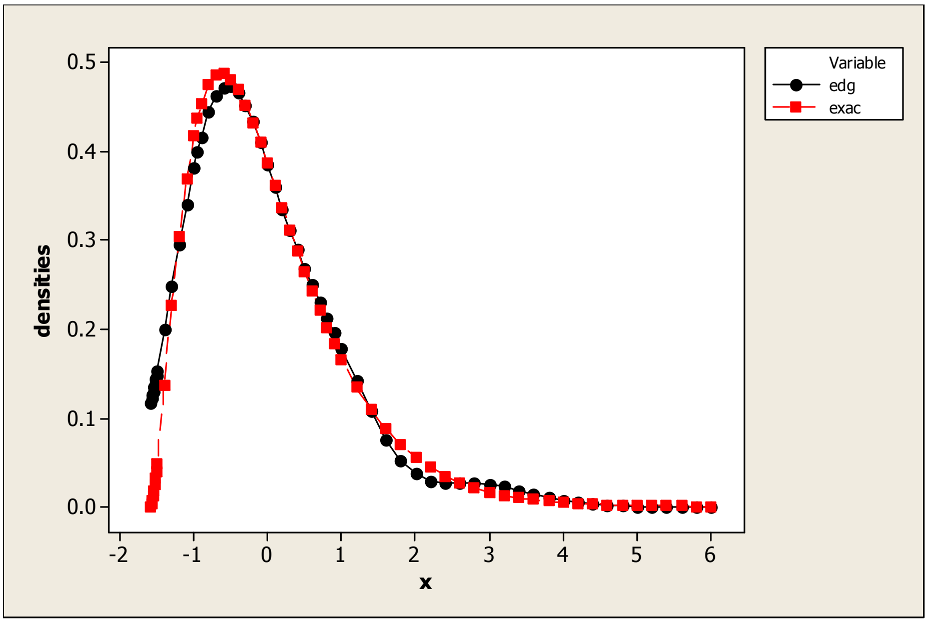

Figure 6.6. The standardized chi-squared density with 5 degrees of freedom and its Edgeworth approximation

Case (III): $r = 10$

Figure 6.7 shows the good performance of the Edgeworth approximation. Note that $\gamma = \sqrt{0.8}$ and $\beta = 4.2$. This is a nice case where the skewness is outside the validity region but the kurtosis is inside the validity region. Nevertheless the skewness is not large.

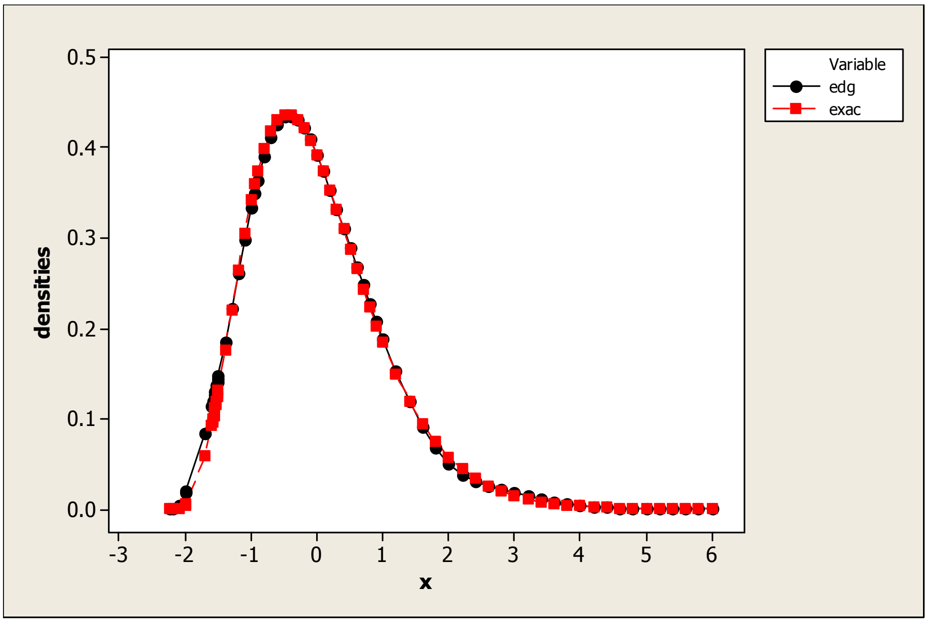

Figure 6.7. The standardized chi-squared density with 10 degrees of freedom and its Edgeworth approximation

Clearly, the Edgeworth approximations improve as *r* increases.

The above results illustrate how the values of the skewness and the kurtosis of the underlying distribution determine the behaviour of the Edgeworth series.

In the following section, the Edgeworth expansion is used to describe the coverage accuracy of triple fixed width confidence intervals for the unknown population mean. Specifically, an asymptotic second order expansion of the coverage probability of the proposed interval is given. Then to guarantee better coverage a triple sampling fixed width confidence interval with a controlled Type II error is proposed, and its characteristics are assessed. Operating characteristic curves are developed to study the effect of skewness, kurtosis and the design factor on the probability of committing a Type II error.

6.3 The coverage probability of the triple sampling sequential procedure

In this section the main objective is to construct a fixed $d_2 - d_1(>0)$ width confidence interval $I_n = \left(\bar{X}_n - d_2, \bar{X}_n - d_1\right)$ for μ, where $d_1 < d_2$ are predetermined constants, such that the confidence coefficient is at least the nominal value $100(1-\alpha)$ percent. Assume further that a random sample $X_1, \ldots, X_n$ for $n \geq 2$ has been observed with sample mean $\bar{X}_n$. Hence we construct the required interval for the unknown mean μ so that the coverage probability is at least the nominal value $100(1-\alpha)\%$. This implies that

$$(6.3) \quad \begin{aligned} P(\mu \in I_n) &= P\left(d_1 \leq \bar{X}_n - \mu \leq d_2\right) \\ &= P\left(d_1\sqrt{n/\theta} \leq Z_n \leq d_2\sqrt{n/\theta}\right) \\ &= H\left(d_2\sqrt{n/\theta}\right) - H\left(d_1\sqrt{n/\theta}\right) \geq (1-\alpha), \end{aligned}$$

where $H(\cdot)$ is the cumulative distribution function of Z_n and from which we may obtain the necessary sample size required to satisfy the above requirements. Generally, however, investigation of (6.3) demonstrates that obtaining an explicit general form of the optimal sample size n^* will be troublesome for several reasons. Firstly, determination of the cutoff point requires complete knowledge of the form of the cumulative distribution function and its inverse probability. Secondly, explicit determination of the optimal sample size in a simple form (as required to define stopping rules in the case of sequential sampling) may be impossible because of the difficulty of solving (6.3) for n, as with the beta or gamma distributions, for example.

Let the optimal sample size n^* takes the general form $n^* = \lambda g(\theta)$ that satisfies (6.3) and assume further that the triple sampling sequential procedure is applied to construct a fixed $d_2 - d_1(>0)$ width confidence interval for the population mean. Then the coverage probability of the required confidence interval is

$$(6.4)\quad P(\mu \in I_N) = P\left(d_1 \le \bar{X}_N - \mu \le d_2\right)$$
$$= \sum_{n=m}^{\infty} P\left(d_1 \le \bar{X}_N - \mu \le d_2, N = n\right)$$
$$= \sum_{n=m}^{\infty} P\left(d_1 \le \bar{X}_N - \mu \le d_2 \mid N = n\right) P(N = n).$$

From Theorem 3.2.2.1 *(iii)*, Theorem 3.3.1 and Theorem 3.3.2 we proved that $\bar{X}_N$ and N are asymptotically normal as $\lambda \to \infty$, moreover they are asymptotically uncorrelated as $\lambda \to \infty$. Recalling the conjecture we made in Chapter 3, that is, $\bar{X}_N$ and N are asymptotically independent as $\lambda \to \infty$ then this implies that the events $\left\{d_1 \le \bar{X}_N - \mu \le d_2\right\}$ and $\{N = n\}$ are asymptotically independent as $\lambda \to \infty$, where N is the third stage sample size of the triple sampling sequential procedure. Moreover, Theorems 3.2.2.2, 3.2.2.3 and 3.3.1 still apply in this case. For $n^* \to \infty$, equation (6.4) gives:

$$(6.5)\quad P(\mu \in I_N) = \sum_{n=m}^{\infty} P\left(d_1\sqrt{n/\theta} \le Z_n \le d_2\sqrt{n/\theta}\right) P(N = n)$$
$$= E_N\left(H\left(d_2\sqrt{N/\theta}\right)\right) - E_N\left(H\left(d_1\sqrt{N/\theta}\right)\right) \ge (1-\alpha)$$

and therefore we can use equation (6.1) to approximate $H\left(d_1\sqrt{N/\theta}\right)$ and $H\left(d_2\sqrt{N/\theta}\right)$.

For simplicity, let u_N be a random variable defined by $u_N = \sqrt{N/\theta}$. Thus equation (6.5) yields the following coverage

$$(6.6)\quad P(\mu \in I_N) = E_N\left\{\Phi(d_2 u_N) - \Phi(d_1 u_N) + (1/72)\left(\phi(d_1 u_N)\, f_1 + \phi(d_2 u_N)\, f_2\right)\right\} + O(1),$$

where

$$f_1 = d_1^5\, u_N^5\, \gamma^2 + 3 d_1^3\, u_N^3 \left(\beta - 3 - (10/3)\gamma^2\right) + 12\gamma\, d_1^2\, u_N^2 - 9\, d_1\, u_N \left(\beta - 3 - (5/3)\gamma^2\right) - 12\gamma,$$

and

$$f_2 = -d_2^5\, u_N^5\, \gamma^2 - 3 d_2^3\, u_N^3 \left(\beta - 3 - (10/3)\gamma^2\right) - 12\gamma\, d_2^2\, u_N^2 + 9\, d_2\, u_N \left(\beta - 3 - (5/3)\gamma^2\right) + 12\gamma.$$

Equation (6.6) shows that the coverage probability depends mainly on the values of the skewness and kurtosis of the underlying distribution.

Hence, using Theorem 3.2.2.3 and for simplicity $E\left(\varepsilon_{N_1}\right) \approx 1/2$ asymptotically, we obtain an expression involving the skewness and kurtosis of the underlying distribution, δ and n^*.

Theorem 6.3.1

For the triple sampling procedure given by $(2.2)-(2.3)$ and the optimal fixed sample size $n^* = a^2\theta/d^2$, the coverage probability of I_N , as $d \to 0$ is

$$P(\mu \in I_N) = \Phi(2d_2 a/d) - \Phi(2d_1 a/d) - t_0\,\gamma/6 + \Upsilon(a,\gamma,\beta,\delta,d) + o(d^2),$$

where,

$$t_n = a^n\left(\psi_1\, d_1^{\,n} - \psi_2\, d_2^{\,n}\right),\ d = (d_2 - d_1),\ \psi_1 = \phi(2d_1 a/d),\ \psi_2 = \phi(2d_2 a/d),$$

$$\Upsilon(a,\gamma,\beta,\delta,d) = (\delta n^*)^{-1}\begin{pmatrix} (-1/4)t_1v_1d^{-1} + (2/3)\gamma\, t_2v_2d^{-2} + (1/3)t_3v_3d^{-3} \\ -(1/3)\gamma\, t_4v_4d^{-4} + (4/9)t_5v_5d^{-5} + (4/3)\gamma\, t_6v_6d^{-6} \\ +(2/3)t_7v_7d^{-7} + (8/9)\gamma^2 t_9v_9d^{-9} \end{pmatrix},$$

and

$$v_1 = \left[(-5\delta/3)\left(n^* + 1/4\right) + (25/24)(\beta - 1)\right]\gamma^2 + \left[\left(n^* + 1/4\right)\beta + 5/4 - 3n^*\right]\delta - (5/8)(\beta + 5)(\beta - 1)$$

$$v_2 = \left(n^* + 3/4\right)\delta - (3/2)(\beta - 1)$$

$$v_3 = \left[(-5/12)\left(8n^* + 9\right)\delta + 5(\beta - 1)\right]\gamma^2 + \left(n^* + 3/2\right)(\beta - 3)\delta - (15/8)(\beta - 1)(\beta - 23/5)$$

$$v_4 = (2\delta + \beta - 1)$$

$$v_5 = \left[\left(n^* + 15/4\right)\delta + (15/4)(\beta - 1)\right]\gamma^2 - (15/8)(\beta - 3)(\beta + 2\delta/5 - 1)$$

$$v_6 = (\beta - 1)$$

$$v_7 = \left[(-16/3)(\beta - 1) - 2\delta/3\right]\gamma^2 + (\beta - 1)(\beta - 3)$$

$$v_9 = (\beta - 1)$$

Proof:

The proof follows from equations (6.4),(6.5) and (6.6) and then making use of Theorem 3.2.2.3. This completes the proof.
Theorem 6.3.1 shows that the coverage is completely determined by the values of the skewness, kurtosis, design factor, optimal sample size and width of the interval.

In the special case where $d_1 = -d_2$ (symmetric intervals) Theorem 6.3.1 reduces to

$$(6.7)\quad P(\mu \in I_N) = (1-\alpha) - \frac{a\,\phi(a)}{288\,\delta n^*}\left(k_0 + k_1 a^2 + k_2 a^4 + k_3 a^6 + k_4 a^8\right) + o\left(d^2\right),$$

where

$k_4 = \gamma^2(\beta - 1).$

$k_3 = 3(\beta-1)\left(\beta - 3 - (16/3)\gamma^2\right) - 2\gamma^2\delta.$

$k_2 = -15(\beta-1)\left(\beta-3-2\gamma^2\right) + \left\{-6(\beta-3) + 2\left(4n^* + 15\right)\gamma^2\right\}\delta.$

$k_1 = -45(\beta-1)\left(\beta - 23/5 - (8/3)\gamma^2\right) + \left\{12\left(2n^*+3\right)\beta - 10\left(9+8n^*\right)\gamma^2 - 72n^* - 108\right\}\delta.$

$k_0 = 45(\beta-1)\left(\beta+5-(5/3)\gamma^2\right) + \left\{-18\left(1+4n^*\right)\beta + 30\left(1+4n^*\right)\gamma^2 + 216n^* - 90\right\}\delta.$

As $d \to 0$, $n^* \to \infty$ the coverage will be asymptotically insensitive (robust). However, for small n^* and also for considering symmetric intervals equation (6.7) will be in the following form

$$(6.8) \quad P\left(\mu \in I_N\right) = (1-\alpha) - \frac{a\phi(a)}{288n^*\delta}\left(k_0 + k_1 a^2 + k_2 a^4 + k_3 a^6 + k_4 a^8\right)$$

where

$k_4 = \gamma^2(\beta - 1).$

$k_3 = (\beta-1)\left(3\beta - 9 - 16\gamma^2\right) - 4\gamma^2\delta E\left(\varepsilon_{N_1}\right).$

$k_2 = -15(\beta-1)\left(\beta-3-2\gamma^2\right) + \left\{-12\beta E\left(\varepsilon_{N_1}\right) + 4\left(2n^* + 15E\left(\varepsilon_{N_1}\right)\right)\gamma^2 + 36E\left(\varepsilon_{N_1}\right)\right\}\delta.$

$$k_1 = -3(\beta-1)\left(15\beta - 69 - 40\gamma^2\right) + \begin{Bmatrix} 24\left(n^* + 3E\left(\varepsilon_{N_1}\right)\right)\beta - 20\left(9E\left(\varepsilon_{N_1}\right) + 4n^*\right)\gamma^2 \\ -72n^* - 216E\left(\varepsilon_{N_1}\right) \end{Bmatrix}\delta.$$

$$k_0 = \quad 15(\beta-1)\left(3\beta + 15 - 5\gamma^2\right) + \begin{Bmatrix} -36\left(E\left(\varepsilon_{N_1}\right) + 2n^*\right)\beta + 60\left(E\left(\varepsilon_{N_1}\right) + 2n^*\right)\gamma^2 \\ +216n^* - 180E\left(\varepsilon_{N_1}\right) \end{Bmatrix}\delta.$$

and the coverage will be sensitive to values of the skewness and kurtosis of the underlying distribution and also to the choice of δ. The coverage is asymptotically robust when the width of the interval approaches zero, which implies that the optimal sample size n^* goes to infinity.

As a special case, consider the normal distribution, where $\gamma = 0$, $\beta = 3$ and $E\left(\varepsilon_{N_1}\right) \simeq 1/2$. Equation (6.8) gives

$$(6.9) \qquad P\left(\mu \in I_N\right) = (1-\alpha) - \frac{a\phi(a)}{2n^*\delta}\left(a^2 - \delta + 5\right) + o(1),$$

which agrees with the corresponding result in Hall (1981). The asymptotic coverage in (6.9) is less than the nominal value $1-\alpha$ and approaches it from below as n^* increases. To maintain a coverage probability of at least the nominal value, Hall (1981) suggested taking an extra sample of size

$\left[\left(a^2-\delta+5\right)/2\delta\right]$ after termination of the triple sampling procedure to improve the coverage, as mentioned in Chapter I.

We now obtain the asymptotic coverage for the uniform, $t(r)$, chi – squared and exponential distributions.

From (6.7), the coverage when the underlying distribution is uniform, where $\gamma=0$ and $\beta=1.8$ is

$$(6.10)\quad P(\mu\in I_N)=(1-\alpha)-\frac{a\phi(a)}{288n^*\delta}\left(k_0+k_1a^2+k_2a^4+k_3a^6\right)+o\left(d^2\right),$$

where

$k_3=-2.88.$
$k_2=14.4+7.2\delta.$
$k_1=100.8-\left\{28.8n^*+43.2\right\}\delta.$
$k_0=244.8+\left\{86.4n^*-122.4\right\}\delta.$

Similarly, the 100 $(1-\alpha)\%$ coverage probability when the underlying distribution is $t(r)$ is

$$(6.11)\quad P(\mu\in I_N)=(1-\alpha)-\frac{a\phi(a)}{8n^*\delta}\left\{k_6a^6+k_4a^4+k_2a^2+k_0\right\}+o\left(d^2\right),$$

where

$k_6=(r-1).$
$k_4=(-5-\delta)r+(5+4\delta).$
$k_2=4r^2+\left(-35+(4n+6)\delta\right)r+31+(-16n-24)\delta.$
$k_0=(-4\delta+20)r^2+\left[-85+(-12n+29)\delta\right]r+65+(-52+48n)\delta.$

while from (6.7) the $100(1-\alpha)\%$ coverage probability when the underlying distribution is chi-squared with r degrees of freedom is

$$(6.12)\quad P(\mu\in I_N)=(1-\alpha)-\frac{a\phi(a)}{18r^2\delta n^*}\left\{k_2r^2+k_1r+k_0\right\}+o\left(d^2\right),$$

where

$k_2=\left(9a^2-9\delta+45\right).$
$k_1=a^8+\left(-\frac{23}{2}-\delta\right)a^6+\left(\frac{15}{2}+\left(4n+\frac{21}{2}\right)\delta\right)a^4+\left(\frac{213}{2}+(-22n-18)\delta\right)a^2+\left(\frac{525}{2}+\left(6n+\frac{3}{2}\right)\delta\right).$
$k_0=6a^8-69a^6+45a^4+315a^2-45.$

A special case of (6.12) is the exponential distribution where $r=2$. Here, taking $E\left(\varepsilon_{N_1}\right)\simeq 1/2$ the asymptotic coverage is

$$(6.13)\quad P\left(\mu\in I_N\right)=\left(1-\alpha\right)-T+o\left(d^2\right),$$

where

$$T=\frac{a\phi(a)}{72\delta n^*}\left\{\begin{array}{l}\left(8a^8-92a^6+60a^4+564a^2+660\right)+\delta\left(-2a^6+21a^4-36a^2-33\right)\\+\delta n^*\left(8a^4-44a^2+12\right)\end{array}\right\}.$$

To illustrate the coverage probabilities based on the Edgeworth asymptotic expansion, we take the nominal values to be $1-\alpha=0.9, 0.95$ and 0.99 with $\delta=0.3, 0.5$ and 0.8 and optimal sample sizes $n^*=24, 43, 61, 76, 96, 125, 171, 246$ and 500 respectively. We denote $Cg(\delta)$ as the asymptotic coverage probability at a specific value of δ. We consider the following underlying distributions: standard normal, standard uniform, t, beta and chi-squared.

6.3.1. Standard normal distribution

To illustrate the asymptotic coverage probabilities obtained by Edgeworth approximation at the above values δ and $n*$ and at $1-\alpha=0.95$ consider Table 6.1. The coverage probabilities based on the Edgeworth approximation approach the nominal coverage probability from below.

n^*	Cg(0.3)	Cg(0.5)	Cg(0.8)
24	0.8821	0.9102	0.9260
43	0.9121	0.9278	0.9366
61	0.9233	0.9343	0.9406
76	0.9285	0.9374	0.9424
96	0.9330	0.9401	0.9440
125	0.9370	0.9424	0.9454
171	0.9405	0.9444	0.9466
246	0.9434	0.9461	0.9477
500	0.9467	0.9481	0.9489

Table 6.1: The asymptotic coverage probability when the underlying distribution is standard normal for various values of δ and n^*; $1-\alpha=0.95$.

6.3.2. Standard uniform distribution

Similarly, to illustrate the asymptotic coverage probability based on an Edgeworth approximation for different values of δ and n^* with $1-\alpha = 0.95$ see Table 6.2. We see that the probabilities tend to increase with δ and with n^*, actually exceeding the nominal value in some cases. However, these results should be treated with caution since the Edgeworth approximation is poor.

n^*	Cg(0.3)	Cg(0.5)	Cg(0.8)
24	0.925025	0.940078	0.948545
43	0.940320	0.948722	0.953448
61	0.946021	0.951944	0.955275
76	0.948709	0.953463	0.956136
96	0.950986	0.954750	0.956866
125	0.952994	0.955884	0.957510
171	0.954782	0.956895	0.958083
246	0.956263	0.957732	0.958558
500	0.957979	0.958701	0.959108

Table 6.2: The asymptotic coverage probability when the underlying distribution is standardized uniform for various values of δ and n^*; $1-\alpha = 0.95$.

Table 6.3 shows the corresponding asymptotic coverage probabilities when the nominal coverage is 0.99. The effect of the Edgeworth approximation being poor is clear!

n^*	Cg(0.3)	Cg(0.5)	Cg(0.8)
24	0.99135	0.99642	0.99928
43	0.99673	0.99957	1.00116
61	0.99874	1.00074	1.00186
76	0.99969	1.00129	1.00219
96	1.00049	1.00176	1.00247
125	1.00120	1.00217	1.00272
171	1.00183	1.00254	1.00294
246	1.00235	1.00284	1.00312
500	1.00295	1.00320	1.00333

Table 6.3: The asymptotic coverage probability when the underlying distribution is standardized uniform for various values of δ and n^*; $1-\alpha = 0.99$

6.3.3. The t distribution

To study the impact of increasing the degrees of freedom on the performance of the coverage probability at $1-\alpha=0.95$, we take $r=5,10,20,50$ and 100 while δ is allowed to vary from 0.3, 0.5 and 0.8.

Case (I): $r=5$

The asymptotic coverage probability at $\delta=0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N)=0.95+\left(4.630692279/n^*\right)-0.04820022046+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95+\left(2.771017473/n^*\right)-0.04820022046+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95+\left(1.724950396/n^*\right)-0.04820022046+\mathrm{o}\left(d^2\right).$$

Case (II) $r=10$

The asymptotic coverage at $\delta=0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N)=0.95-\left(1.7746865840/n^*\right)-0.008033370077+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95-\left(1.0469542250/n^*\right)-0.008033370077+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95-\left(0.6376047727/n^*\right)-0.008033370077+\mathrm{o}\left(d^2\right).$$

Case (III): $r=20$

The asymptotic coverage at $\delta=0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N)=0.95-\left(1.7403333110/n^*\right)-0.003012513779+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95-\left(1.0231853080/n^*\right)-0.003012513779+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95-\left(0.6197895568/n^*\right)-0.003012513779+\mathrm{o}\left(d^2\right).$$

Case (IV): $r=50$

The asymptotic coverage at $\delta=0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N)=0.95-\left(1.6603389800/n^*\right)-0.0009940518632+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95-\left(0.9740748628/n^*\right)-0.0009940518632+\mathrm{o}\left(d^2\right).$$
$$P(\mu \in I_N)=0.95-\left(0.5880512975/n^*\right)-0.0009940518632+\mathrm{o}\left(d^2\right).$$

Case (V): $r=100$

The asymptotic coverage at $\delta=0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N) = 0.95 - (1.6535712050/n^*) - 0.00050208563 + o(d^2).$$
$$P(\mu \in I_N) = 0.95 - (0.9695495681/n^*) - 0.00050208563 + o(d^2).$$
$$P(\mu \in I_N) = 0.95 - (0.5847873977/n^*) - 0.00050208563 + o(d^2).$$

The idea of listing the equations in this way is to show the following

1. The leading coefficients of the term $1/n^*$ decreases as δ increases. Moreover, the coefficient at $\delta = 0.5$ lies between that of the other values of δ .

2. For a specific value of r, the last term in the equation is fixed for all values of δ and decreases as r increases.

3. At $r = 5$ the sign of the leading coefficients of the term $1/n^*$ is positive, while for the remaining value of r the sign is negative. The reason of this nice picture is due to the kurtosis. Note here that at $r = 5$ the kurtosis is 9, which is outside the validity region, while for the other values of r the kurtosis values are inside the validity region.

Table 6.4 shows the asymptotic coverage probabilities for various values of r with $\delta = 0.5$ and $n^* = 500$ for nominal coverage probabilities 90%, 95% and 99%. Clearly as r increases the asymptotic results approach the nominal values from above at 90% and from below at 95% and 99%.

r	90%	95%	99%
5	0.924900	0.901800	0.922310
10	0.904155	0.941970	0.978720
20	0.901558	0.946987	0.985770
50	0.900542	0.949006	0.988530
100	0.900260	0.949498	0.989295

Table 6.4: The asymptotic coverage probabilities for underlying t distribution with $r = 5, 10, 20, 50$ and 100 and $1-\alpha = 0.9, 0.95$ and 0.99 at $\delta = 0.5$ and $n^* = 500$.

6.3.4. Beta $(4,4)$ distribution

For the beta $(4,4)$ distribution the standardized density is

$$g(x) = 0.000500114(9 - x^2)^3, \ -3 < x < 3.$$

The corresponding Edgeworth asymptotic expansion is

$$f(x) = \phi(x)\{1 - 0.022727272(x^4 - 6x^2 + 3)\} + O(1).$$

For nominal coverage 95% the asymptotic coverages at $\delta = 0.3, 0.5$ and 0.8 are

$$P(\mu \in I_N) = 0.95 - (1.3518654720/n^*) + 0.00438183823 + o(d^2).$$
$$P(\mu \in I_N) = 0.95 - (0.7854552756/n^*) + 0.00438183823 + o(d^2).$$
$$P(\mu \in I_N) = 0.95 - (0.4668495391/n^*) + 0.00438183823 + o(d^2).$$

For $n^* = 500$ and $\delta = 0.5$, the $90\%, 95\%$ and 99% coverage probabilities are 0.897734, 0.954382 and 0.990662 respectively. Clearly, the coverage probability exceeds the nominal value at 95% and 99% while it is less than the nominal value at 90%.

Note that the skewness of the Beta(4, 4) distribution is inside the validity region while the kurtosis (2.45455) is outside.

6.3.5. Chi-squared distribution with r degrees of freedom

Case (I): $r = 2$

Table 6.5 shows the asymptotic coverage probabilities for an underlying exponential distribution at $\delta = 0.5$ for selected n^* values at nominal coverage 90%. As with the uniform case, the Edgeworth-based approach leads to poor results.

n^*	Cg(0.3)	Cg(0.5)	Cg(0.8)
24	0.95759	0.97868	0.99055
43	0.98163	0.99340	1.00002
61	0.99059	0.99889	1.00355
76	0.99481	1.00147	1.00522
96	0.99839	1.00366	1.00663
125	1.00155	1.00560	1.00787
171	1.00436	1.00732	1.00898
246	1.00668	1.00874	1.00990
500	1.00938	1.01039	1.01096

Table 6.5: The asymptotic coverage probability, showing the effect of the poor behaviour of the Edgeworth approximation under the chi-squared distribution with $r = 2$, $1-\alpha = 0.9$.

Case (II) $r = 5$

The asymptotic coverage probability at $\delta = 0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N) = 0.95 - (2.488184025/n^*) + 0.02479709643 + o(d^2).$$
$$P(\mu \in I_N) = 0.95 - (1.485587290/n^*) + 0.02479709643 + o(d^2).$$
$$P(\mu \in I_N) = 0.95 - (0.921626627/n^*) + 0.02479709643 + o(d^2).$$

Case (III): $r = 10$

The asymptotic coverage probability at $\delta = 0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N) = 0.95 - \left(2.224000864/n^*\right) + 0.01239854821 + o\left(d^2\right).$$
$$P(\mu \in I_N) = 0.95 - \left(1.319284531/n^*\right) + 0.01239854821 + o\left(d^2\right).$$
$$P(\mu \in I_N) = 0.95 - \left(0.810381594/n^*\right) + 0.01239854821 + o\left(d^2\right).$$

Case (IV): $r = 30$

The asymptotic coverage probability at $\delta = 0.3, 0.5$ and 0.8 is respectively

$$P(\mu \in I_N) = 0.95 - \left(1.865008892/n^*\right) + 0.004132849406 + o\left(d^2\right).$$
$$P(\mu \in I_N) = 0.95 - \left(1.098694106/n^*\right) + 0.004132849406 + o\left(d^2\right).$$
$$P(\mu \in I_N) = 0.95 - \left(0.667642039/n^*\right) + 0.004132849406 + o\left(d^2\right).$$

In Table 6.6 the asymptotic coverage probabilities at $\delta = 0.5$, $n^* = 500$ and *r*=5, 10 and 30 are shown for nominal coverage probabilities 90%, 95% and 99%. Clearly the asymptotic value approaches the nominal value as *r* increases, from above when the nominal coverages are 90% and 95%, and from below at 99%. Note that when *r*=5 the skewness and kurtosis are both outside the validity range, while at *r*=10 and 30 only the skewness is outside.

r	90%	95%	99%
5	0.945688	0.974790	0.975053
10	0.922844	0.962399	0.982527
30	0.907615	0.954133	0.987509

Table 6.6: The asymptotic coverage probability for underlying chi-squared distribution with $r = 5, 10, 30$ at $n^* = 500$ and $\delta = 0.5$.

We shall investigate the accuracy of the asymptotic coverage probabilities in the next chapter using simulation.

6.4 Sensitivity of triple sampling fixed width confidence intervals to a shift in the population mean

We now introduce the idea of controlling the probability of committing a Type II error and at the same time improving the coverage probability along the lines of Costanza *et. al*. (1995), Son *et. al*. (1997) and Hamdy (1997).

Confidence intervals in general provide satisfactory information regarding the quality and the reliability of inference (see, for example, Nelson, 1990 and 1994). It is also evident that a confidence interval shows by its width the precision of estimation as well as which parameter values would not be rejected if they were hypothesized as point null values. This view of the duality between the union of all non-rejectable point null hypotheses and a single confidence interval is referred to by Tukey (1991) who said

"Many *of us are familiar with deriving a confidence interval from an infinite array of tests of significance, one for each potential null hypothesis*".

His point was made in the two following sentences

"*Fewer of us perhaps have thought of the use of a confidence interval as the reverse process. This is the most important reason for a confidence interval…*"

Based on his view, the intersection of an infinite array of confidence intervals is used to locate a single set of plausible point null hypotheses. Lehmann (1986, pp. 89-96) used the concept of confidence intervals to test hypotheses. He studied the relation between uniformly most powerful one-sided tests and the corresponding lower or upper confidence bounds.

In order to design and perform tests of hypotheses based on confidence intervals using the idea of Tukey (1991), we need to consider the idea of controlling the Type II error probabilities. This is essentially the approach employed when statistical quality control charts are designed to detect shifts in a process mean; see Montgomery (1982) and Rahim (1993) for details. Such work on the relationship between confidence intervals and the power of tests has received little attention in the literature. To the best of our knowledge, no-one has studied the use of triple sampling fixed width confidence intervals to test hypotheses in the manner of Tukey for the class of continuous distributions with finite first six moments.

Costanza *et. al*. (1995) evaluated the sensitivity of fixed width confidence intervals for detecting shifts in the normal mean based on Hall's (1981) triple and modified triple sampling against the corresponding fixed sample size sampling procedure. They found that the usual triple sampling fixed width confidence intervals were more sensitive to shifts occurring within the intervals than their fixed sample size counterparts. However, the corresponding Type II error probabilities were still large.

Hall's triple sampling attains the nominal value asymptotically and his modified triple sampling improved the coverage probability when the underlying distribution is normal, but it has the disadvantage of increasing the Type II error probabilities for shifts that occur, both inside and outside the confidence intervals. The reason is that the usual optimal sample sizes used to establish the triple sampling estimation procedures do not reflect any requirements regarding the control of Type II error probabilities. The use of another optimal fixed sample size that actually reflects some form of Type II error will improve the coverage probability.

In the following section, we will describe the hypotheses to be tested when controlling type II error probabilities, and derive the coverage and operating characteristic function for symmetrical intervals based on the new approximate optimal fixed sample size

6.4.1 Triple sampling fixed width confidence intervals with controlled Type II error probability

In this context we formulate the following two hypotheses in order to signify such a shift if it takes place, where $I_N\left(\bar{X}_N - d, \bar{X}_N + d\right)$.

Hypotheses

$$(6.14)\quad H_0 : \mu = \mu_0, \quad \mu_0 \in I_N$$

vs.

$$H_1 : \mu = \mu_1 = \mu_0 \pm d\left(1+k\right), \quad \mu_1 \notin I_N, \forall k \geq 0$$

The null hypothesis asserts that no shift in the μ_0 has occurred against the alternative that the parameter value differs from μ_0 by a "distance" $1+k$ measured in units of d. To clarify this point, assume that the null parameter value is μ_0. If the parameter value has shifted to $\mu_1 > (<) \mu_0$, assessment of the amount of risk associated with the departure from μ_0 provides an indication of the ability of I_N to detect such a departure.

In section 6.3, the fixed width confidence interval based on the optimal fixed sample size without controlling the Type II error probability is $n^* = a^2\theta/d^2$. Now we find the fixed width confidence interval based on controlling the Type II error probability.

It was shown by Brownlee (1965, pp. 117-118) that the optimal fixed sample size required to control the Type II error probabilities of detecting shifts in μ of magnitude $\pm d(1+k)$ units away from $\mu = \mu_0$ outside the interval for a prespecified value of k and β_t is given by

$$(6.15) \quad n^* = \frac{(a+b)^2\theta}{d^2(1+k)^2},$$

where a is the $\alpha/2$ point and b is the upper β_t point of the standard normal distribution. Note that we assume that the value of $\mu = \mu_0$ is located in the centre of the interval in order to provide equal Type II error probabilities for equidistant shifts to $\mu = \mu_1$ outside the interval in either direction. Clearly, as k approaches zero, the coverage probability based on (6.15) will be greater than the nominal value than when using $n^* = a^2\theta/d^2$, while for larger values of k, the coverage will be less than when $n^* = a^2\theta/d^2$ since the effect of increasing the shift k will dominate the effect of b. Moreover, we need this modification in the optimal sample size in order to ensure full protection of the triple sampling sequential fixed width confidence interval against type II error.

It was shown that the optimal fixed sample size for testing (6.14) is

$$(6.16) \quad n^* = (a+b)^2\theta/d^2,$$

see Son *et al*. (1997) for details. The reason behind this choice may be illustrated as follows:

1. It ensures the coverage to be at least the nominal value $100(1-\alpha)\%$.
2. It has Type II error probability less than the prescribed 100 $\beta_t\%$ at the particular shift of interest indexed by the value of k.
3. It has Type II error probabilities that are uniformly less than those corresponding to (6.16) for all $k \geq 0$.
4. It is independent of k, which facilitates the algebra in computing the coverage of the triple sampling procedure with controlled optimal sample size as well as the Type II error probability.

Thus the triple sampling procedure based on the new optimal fixed sample size (6.16) is

$$(6.17)\quad N_1 = \max\left\{m, \left[\ \delta\left(d^{-1}(a+b)\right)^2 g\left(S_m^2\right)\right]+1\right\},$$

and

$$(6.18)\quad N = \max\left\{N_1, \left[\ \left(d^{-1}(a+b)\right)^2 g\left(S_{N_1}^2\right)\right]+1\right\}.$$

Moreover the same arguments and asymptotic characteristics apply as described in chapter III, and we propose the confidence interval $I_N = \left(\bar{X}_N - d, \bar{X}_N + d\right)$ for μ.

Theorem 6.4.1

The coverage probability of I_N based on the triple sampling procedure given by $(6.17)-(6.18)$ and the controlled optimal fixed sample size (6.16) as $d \to 0$ is given by

$$P\left(\mu \in I_N\right) = \left(2\Phi(a+b)-1\right) - \frac{(a+b)\phi(a+b)}{288 n^* \delta} S(a,b) + O\left(d^2\right),$$

where

$$S(a,b) = \left(k_0 + k_1(a+b)^2 + k_2(a+b)^4 + k_3(a+b)^6 + k_4(a+b)^8\right),$$

and

$$k_4 = \gamma^2(\beta-1)$$

$$k_3 = 3(\beta-1)\left(\beta-3-(16/3)\gamma^2\right) - 2\gamma^2\delta$$

$$k_2 = -15(\beta-1)\left(\beta-3-2\gamma^2\right) + \left\{-6(\beta-3) + 2\left(4n^*+15\right)\gamma^2\right\}\delta$$

$$k_1 = -45(\beta-1)\left(\beta-23/5-(8/3)\gamma^2\right) + \left\{12\left(2n^*+3\right)\beta - 10\left(9+8n^*\right)\gamma^2 - 72n^* - 108\right\}\delta$$

$$k_0 = 45(\beta-1)\left(\beta+5-(5/3)\gamma^2\right) + \left\{-18\left(1+4n^*\right)\beta + 30\left(1+4n^*\right)\gamma^2 + 216n^* - 90\right\}\delta$$

Proof:

The proof follows exactly as for the coverage probability in Chapter V, after replacing the usual optimal fixed sample size $n^* = a^2\theta/d^2$ by (6.16).

Clearly, the modified coverage is greater than the nominal value for any value of n^* since

$$\begin{aligned} P\left(\mu \in I_N\right) &= \left(2\Phi(a+b)-1\right) - \frac{(a+b)\phi(a+b)}{288 n^* \delta} S(a,b) + O\left(d^2\right) \\ &> (1-\alpha) - \frac{(a+b)\phi(a+b)}{288 n^* \delta} S(a,b) + O\left(d^2\right), \end{aligned}$$

The second term will be negligible for most reasonable choices of α and β_t and even for small values of n^*. The coverage probability under (6.16) is greater than under $n^* = a^2\theta/d^2$. Moreover, the controlled coverage depends mainly on the skewness and kurtosis of the underlying distribution and n^*. The effect of the design factor δ is small since it appears only in the negligible term.

6.4.2 Operating characteristic function of Type II error controlled confidence intervals

Our last assertion in this study is to investigate the sensitivity of the constructed confidence interval to shifts in the population mean μ.

The probability (β_{tc}-risk) of not detecting such a shift in the true parameter μ_0 when, in fact, a shift has actually occurred is given by

$$\beta_{tc} = P\left(\mu \in I_N \,|H_1\right),$$

see, for example, Costanza *et al*. (1995), Son *et al*. (1997) and Hamdy (1997, 1999) for details.

Therefore, β_{tc} may be written as

$$(6.19) \quad \beta_{tc} = \sum_{n=m}^{\infty} P\left(\left|\bar{X}_N - \mu_1\right| \le d, N = n\right)$$
$$= \sum_{n=m}^{\infty} P\left(\left|\bar{X}_N - \mu_1\right| \le d \,|N = n\right) P\left(N = n\right).$$

Since $\bar{X}_N$ is asymptotically normally distributed, independent of N (Theorem 3.2.2.1, (*iii*)), then $\bar{X}_N$ and the event $N = n, n+1, n+2, \ldots$ are asymptotically independent. Therefore, β_{tc} may be expressed as follows

$$(6.20) \quad \beta_{tc} = \sum_{n=m}^{\infty} P\left(\left|\bar{X}_n - \mu_1\right| \le d\right) P(\, N = n\,)$$
$$= \sum_{n=m}^{\infty} P\left(-(\,2+k\,)d \le \bar{X}_n - \mu_0 \le -k\,d\right) P(\, N = n\,)$$
$$= E_N\left(H\left(-k\,d\sqrt{N/\theta}\right)\right) - E_N\left(H\left(-(\,2+k\,)\,d\sqrt{N/\theta}\right)\right).$$

Theorem 6.4.2

The operating characteristic function based on the triple sampling procedure given by $(6.17)-(6.18)$ and controlled optimal fixed sample size (6.16) as $d \to 0$ is

$$\beta_{tc} = \Phi\left(-k\left(a+b\right)\right) - \Phi\left(-(2+k)\left(a+b\right)\right) + Q\left(a,b,\gamma,\beta,\delta\right) + o\left(d^2\right),$$

where

$$Q(a,b,\gamma,\beta,\delta) = (1/72)(\phi_1 e_1 - \phi_2 e_2),$$

with

$$e_1 = \gamma^2 (a+b)^5 k^5 + 3\left(\beta - 3 - (10/3)\gamma^2\right)(a+b)^3 k^3 + \left(12\delta^{-1}(\beta - 1 - 3\delta/2)\right)(a+b)^2 k^2 \gamma$$
$$-9(a+b)k\left(\beta - 3 - (5/3)\gamma^2\right) + 12\gamma$$

$$e_2 = \gamma^2 (a+b)^5 (k+2)^5 + 3(a+b)^3 (k+2)^3 \left(\beta - 3 - (10/3)\gamma^2\right) - 12\gamma (a+b)^2 (k+2)^2$$
$$-9(a+b)(k+2)\left(\beta - 3 - (5/3)\gamma^2\right) + 12\gamma,$$

and ϕ_1 and ϕ_2 are the standard normal densities at $-k(a+b)$ and $-(2+k)(a+b)$.

The proof follows immediately from (6.1) and Theorem 3.2.2.3.

Theorem 6.4.2 shows that the probability of Type II error β_{tc} is completely determined by the shift k and the skewness and kurtosis of the underlying distribution. The influence of δ is small, which is different from the case of the coverage as shown in Theorem 6.3.1.

Remark

It was shown by Son *et al*. (1997) that the operating characterstic function of Hall's (1981) triple sampling procedure is greater than the operating characterstic function with controlled optimal sample size uniformly in k.

As a special case of Theorem 6.4.2, consider the case in which the underlying distribution is normal so that $e_1 = e_2 = 0$, and thus the operating characteristic function reduces to

$$\beta_{tc} = \Phi\left(-k(a+b)\right) - \Phi\left(-(2+k)(a+b)\right) + o\left(d^2\right) \text{ as } d \to 0.$$

At $k = 0$ the value of $\beta_{tc} = 1/2$.

For a uniform underlying distribution, the probability of Type II error is

$$\beta_{tc} = \Phi\left(-k(a+b)\right) - \Phi\left(-(2+k)(a+b)\right) + Q(a,b,\gamma=0,\beta=1.8,\delta) + o\left(d^2\right).$$

So, it is clear that the probability of Type II error for any continuous distribution satisfying the existence of the first six moments and as $m \to \infty$ is an additively modified version of the Type II error probability when the underlying distribution is normal.

Chapter VII

Simulation Results for Triple Sampling Fixed Width Confidence Intervals for the Population Mean

In this chapter, we use simulation to study the performance of normal based triple sampling fixed width confidence intervals for small, moderate and large sample sizes. Moreover, we compare the simulation results with the corresponding asymptotic results found in Chapter VI.

7.1 Experimental setup

As in Chapter V, a series of Monte Carlo studies was carried out in order to study the performance of normal based triple sampling fixed width confidence intervals and compare them with the confidence intervals based on the second order Edgeworth asymptotic expansion.

First we allowed aspects of the triple sampling scheme to vary: $m = 5$, 15, 20; $\delta = 0.3$, 0.5, 0.8; $n^* =$ 24, 43, 61, 76, 96, 125, 171, 246, 500 and $1-\alpha = 0.9$, 0.95 and 0.99. In addition, we consider the same class of underlying distributions as in Chapter V in order to enable comparison with the point estimation results.

For each experimental situation the same 50,000 replicate samples were used as in Chapter V and for each experimental situation we estimated the coverage probability $1-\widehat{\alpha}$. The standard error of the estimated nominal coverage probabilities for the above nominal values are respectively, 0.001341, 0.000974 and 0.000445.

7.2 The coverage probabilities of the triple sampling procedure

In the following subsections, we investigate in detail the coverage probabilities of the triple sampling fixed width confidence intervals for $m = 5, 15, 20$, $\delta = 0.5$ and $\alpha = 0.05$. Results for other situations are tabulated in Appendix B.

7.2.1 The underlying distribution is standard normal

From Table B1 in Appendix B we see that the estimated coverage probabilities do not attain the nominal coverage values for any of our values of m.

Remark

1. The coverage probability improves (attains the nominal value) as n^* increases and also as m increases.

2. Hall (1981) pointed out that for moderate n^* $P\left\{\left|\overline{X}_N - \mu\right| \le d\right\} < 1-\alpha$, while for large values of n^*, $P\left\{\left|\overline{X}_N - \mu\right| \le d\right\} \to (1-\alpha)$ as $n^* \to \infty$.

To show the effect of increasing n^* on the performance of the coverage probability at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$ see Figure 7.1, which shows the simulated estimates of the coverage probability as $n*$ increases. It is clear that the coverage probabilities do not attain the nominal value under these conditions.

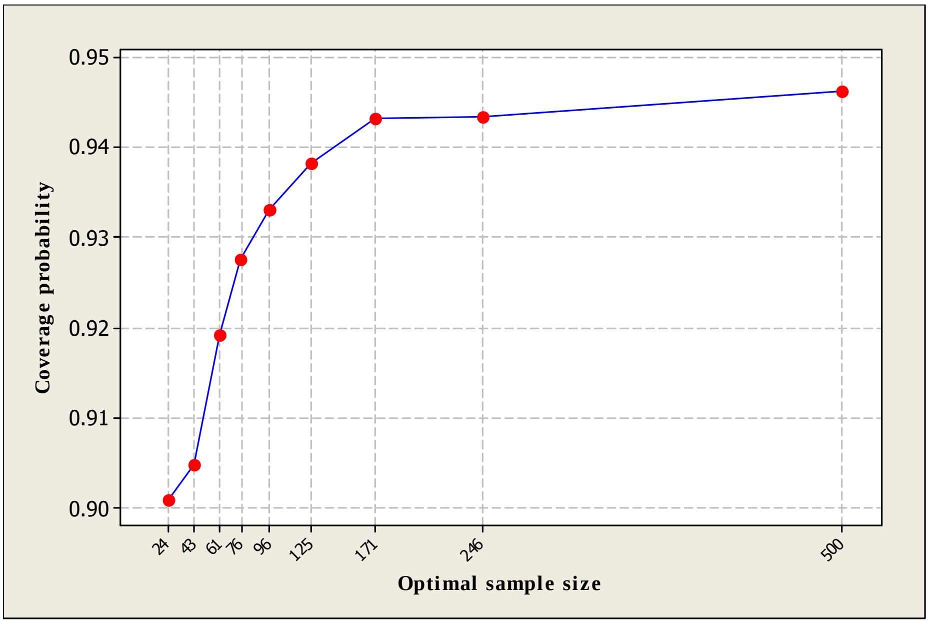

Figure 7.1: The simulated estimates of the coverage probability by optimal sample size for underlying distribution $N(0,1)$ at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

To realize the effect of δ on the coverage probability and compare this with the asymptotic results obtained in Table 6.1 see Table 7.1, which shows the coverage probability as the optimal sample size increases. We see that the coverage probability never attains the targeted nominal values at $\delta = 0.3$. However, at $\delta = 0.8$, the coverage probability exceeds the nominal value but only at large values of n^* ($n^* = 246$ and 500).

By comparing the simulation results in Table 7.1 with the asymptotic results in Table 6.1 we see that there are consistencies (nearly same values and less than the nominal value) at $\delta = 0.3$ and 0.5 while at $\delta = 0.8$ the simulation estimates of the coverage probability are greater than the asymptotic results for large n^*.

n^*	$1-\hat{\alpha}, \delta=0.3$	$1-\hat{\alpha}, \delta=0.5$	$1-\hat{\alpha}, \delta=0.8$
24	0.8801	0.9009	0.9331
43	0.8211	0.9048	0.9308
61	0.8538	0.9192	0.9383
76	0.8833	0.9276	0.9421
96	0.9066	0.9331	0.9444
125	0.9248	0.9383	0.9475
171	0.9337	0.9433	0.9486
246	0.9404	0.9435	0.9508
500	0.9449	0.9463	0.9523

Table 7.1: The simulated estimates of the coverage probability for underlying $N(0,1)$ as the optimal sample size increases, $\delta = 0.3$, 0.5, 0.8 and at $\alpha = 0.05$, $m = 15$.

Figure 7.2. shows the effect of changing δ on the coverage probability at $m = 15$. We see from the graph that the coverage exceeds the nominal value at $\delta = 0.8$ as n^* increases. The graph supports our discussion above and support the choice of taking δ as a compromise choice.

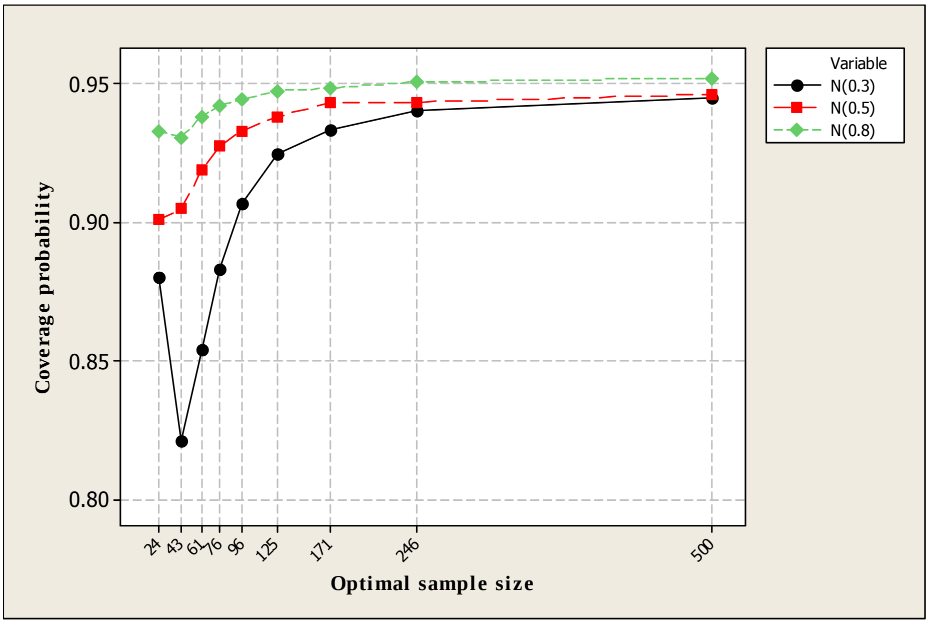

Figure 7.2: The effect of changing δ on the simulated estimates of the coverage probability when the underlying distribution is $N(0,1)$, $\alpha = 0.05$, $m = 15$ and $\delta = 0.5$.

7.2.2 The underlying distribution is standard uniform

From Table B2 in Appendix B we see that the estimated coverage probabilities do not attain the nominal coverage values for any of our values of m.

Figure 7.3 below shows the effect of increasing the optimal sample size on the performance of the coverage probability at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$. It is obvious again that the coverage probabilities do not attain the nominal value. To show the performance of the coverage probability as n^* increases at $m = 15, \delta = 0.5$ and $\alpha = 0.05$ see Figure 7.3, which shows the simulated coverage probability as the optimal sample size increases. Clearly the coverage never attains the nominal value.

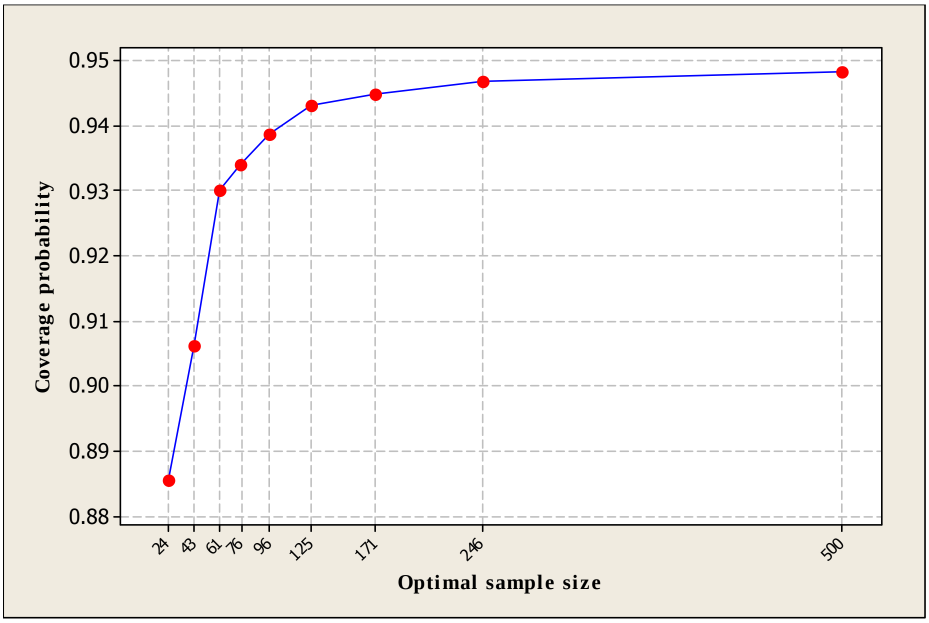

Figure 7.3: The simulated estimates of the coverage probability by optimal sample size for underlying distribution $U(0,1)$ at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

To show the effect of δ on the coverage probability at $m = 15$ see Table 7.2. At $\delta = 0.3$ and 0.5 the simulated estimates of the coverage probability is always less than the nominal value, while at $\delta = 0.8$ the coverage exceeds the nominal value only at large values of n^* ($n^* = 500$) but this behaviour tends to be adjusted and turn back to be less than the nominal value as m increases.

In Chapter VI we saw that the asymptotic coverage probabilities in this case are unreliable, particularly at high nominal coverage values. By comparing the simulation results with the asymptotic results that are presented in Table 6.2, we see the differences between the asymptotic and simulated coverage values. Collectively, the coverage probability improves as n^* increases and as m increases (see Table B2 in Appendix B).

n^*	$1-\hat{\alpha}, \delta = 0.3$	$1-\hat{\alpha}, \delta = 0.5$	$1-\hat{\alpha}, \delta = 0.8$
24	0.8803	0.8854	0.9209
43	0.7844	0.9062	0.9285
61	0.8572	0.9301	0.9399
76	0.9009	0.9341	0.9411
96	0.9269	0.9387	0.9447
125	0.9361	0.9432	0.9468
171	0.9431	0.9449	0.9480
246	0.9461	0.9470	0.9483
500	0.9484	0.9483	0.9510

Table 7.2: The effect of changing the design factor δ on the performance of the coverage probability when the underlying distribution is $U(0,1), \alpha = 0.05, m = 15$ and $\delta = 0.5$. The coverage probability

Figure 7.4 shows the effect of δ on the coverage probability at $\alpha = 0.05, m = 15$. Clearly the graph support our discussion above.

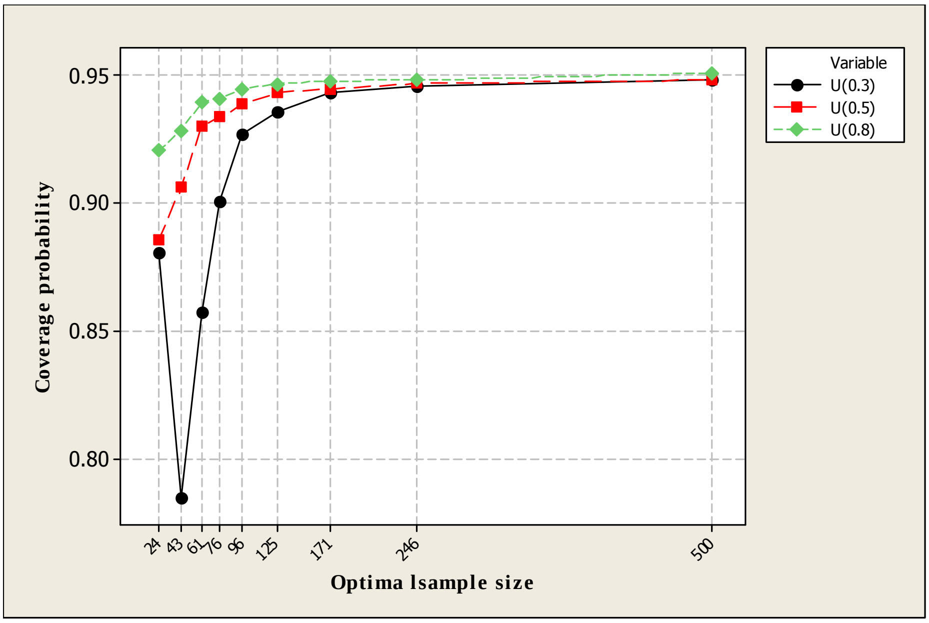

Figure 7.4: The effect of changing δ on the simulated estimates of the coverage probability for underlying distribution $U(0,1), \alpha = 0.05, m = 15$ and $\delta = 0.5$.

7.2.3 The underlying distribution is t

From Tables B3, B4, B5 and B6, we have the following

Regarding the coverage probability, the procedure starts with bad estimates at $r = 5$, and tends to improve as n^* and m increase. While, as r increases we noticed improve performance of the coverage probability as expected.

To show the effect of increasing the optimal sample size on the performance of the coverage probability under the t distribution we consider Figures 7.5 and 7.6. They show the coverage probability when the underlying distribution is t as the degrees of freedom increases. It is clear from the graphs that as r increases, we attain better coverage probability. This support the theory that the limiting distribution of the $t(r)$ is normal as $r \to \infty$.

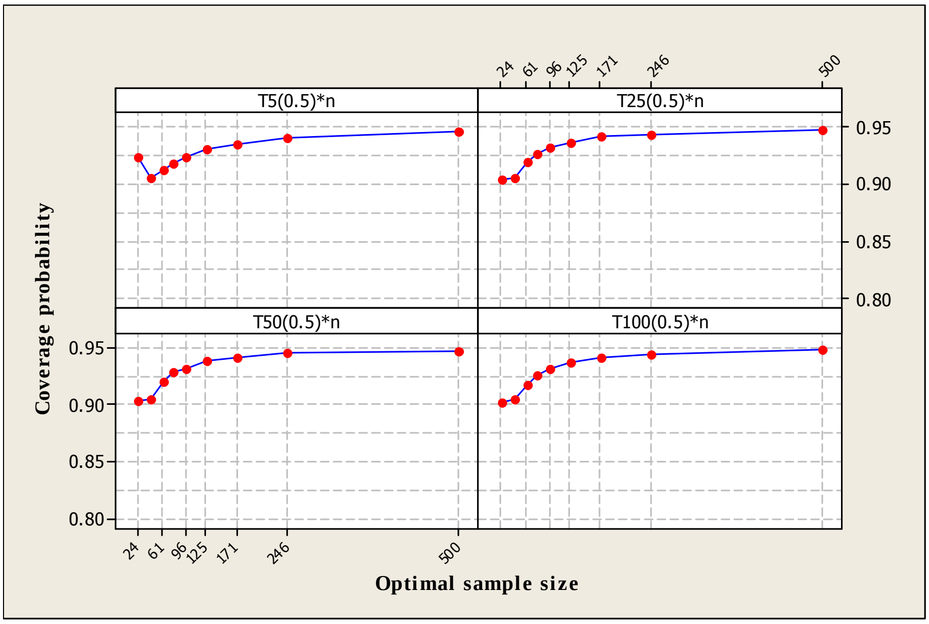

Figure 7.5: The simulated estimates of the coverage probability for underlying t distribution with $r = 5, 25, 50$ and 100 as the optimal sample size increases at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

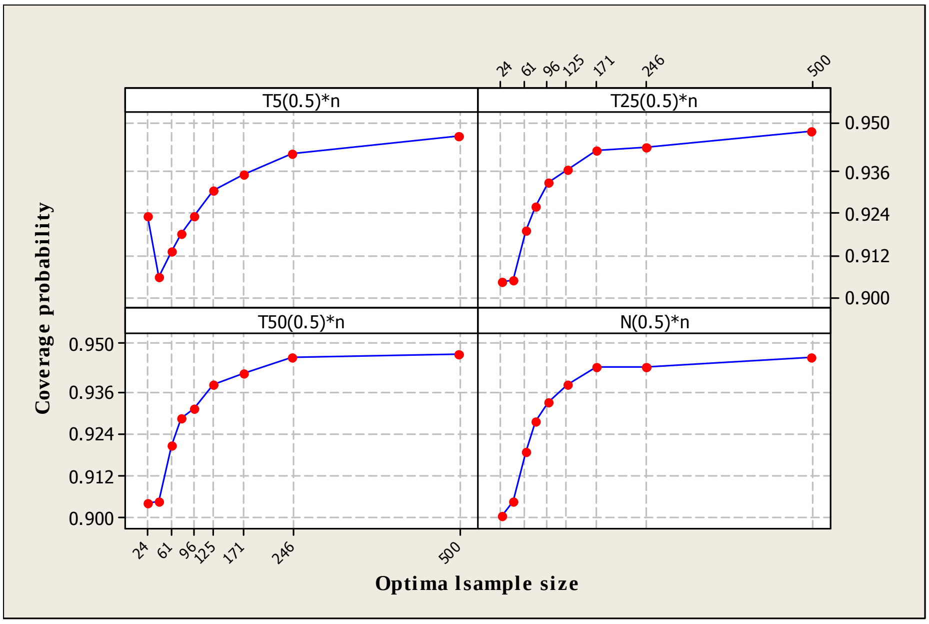

Figure 7.6: Comparison between the simulated estimates of the coverage probability for underlying $N(0,1)$ and t distributions with $r = 25, 50$ and 100; $\alpha = 0.05, m = 15$ and $\delta = 0.5$.

To show the impact of δ on the coverage probability at $m = 15$ and as r increases we consider Tables 7.3, 7.4 and 7.5, which illustrate that the coverage probability under the t distribution improves as r increases. By comparing the simulation results with the asymptotic results that are presented in Table 6.4, we found that for underlying $t(5)$ distribution the simulation estimates of the coverage probability is larger than the asymptotic value for all values of n^*, and this due to the bad behaviour of the Edgeworth series in approximating the coverage probability as shown in the previous chapter. While for underlying $t(25)$ and $t(50)$ distributions both the simulation estimates of the coverage probability and the asymptotic coverage are consistent as n^* increases.

n^*	$1-\widehat{\alpha}, \delta = 0.3$	$1-\widehat{\alpha}, \delta = 0.5$	$1-\widehat{\alpha}, \delta = 0.8$
24	0.8980	0.9232	0.9456
43	0.8477	0.9058	0.9332
61	0.8511	0.9129	0.9374
76	0.8669	0.9181	0.9401
96	0.8886	0.9230	0.9443
125	0.9011	0.9304	0.9475
171	0.9196	0.9353	0.9473
246	0.9292	0.9411	0.9505
500	0.9387	0.9462	0.9536

Table 7.3: The simulated coverage probability for underlying $t(5)$ as the optimal sample size increases; $1-\alpha = 0.95$ and $m = 15$.

n^*	$1-\hat{\alpha}, \delta=0.3$	$1-\hat{\alpha}, \delta=0.5$	$1-\hat{\alpha}, \delta=0.8$
24	0.8780	0.9042	0.9336
43	0.8242	0.9048	0.9334
61	0.8538	0.9192	0.9383
76	0.8818	0.9257	0.9406
96	0.9045	0.9327	0.9435
125	0.9223	0.9365	0.9471
171	0.9313	0.9422	0.9479
246	0.9379	0.9431	0.9508
500	0.9445	0.9474	0.9511

Table 7.4: The simulated coverage probability for underlying $t(25)$ as the optimal sample size increases; $1-\alpha=0.95$ and $m=15$.

n^*	$1-\hat{\alpha}, \delta=0.3$	$1-\hat{\alpha}, \delta=0.5$	$1-\hat{\alpha}, \delta=0.8$
24	0.8806	0.9043	0.9349
43	0.8242	0.9049	0.9316
61	0.8531	0.9209	0.9378
76	0.8804	0.9289	0.9425
96	0.9068	0.9316	0.9451
125	0.9213	0.9385	0.9458
171	0.9339	0.9417	0.9486
246	0.9400	0.9460	0.9510
500	0.9447	0.9469	0.9515

Table 7.5: The simulated coverage probability for underlying $t(50)$ as the optimal sample size increases; $1-\alpha=0.95$ and $m=15$.

7.2.4 The underlying distribution is beta$(2,3)$

From Table B7 we have the following

Regarding the coverage probability we noticed similar behaviour as the case of normal and uniform.

To show the effect of increasing n^* on the coverage probability under the beta distribution see Figure 7.7, which shows the coverage probability when the underlying distribution is beta as n^* increases and $m=15$. As before the confidence interval do not attain the coverage probability, and this supports Hall (1981) that we only attain the coverage asymptotically.

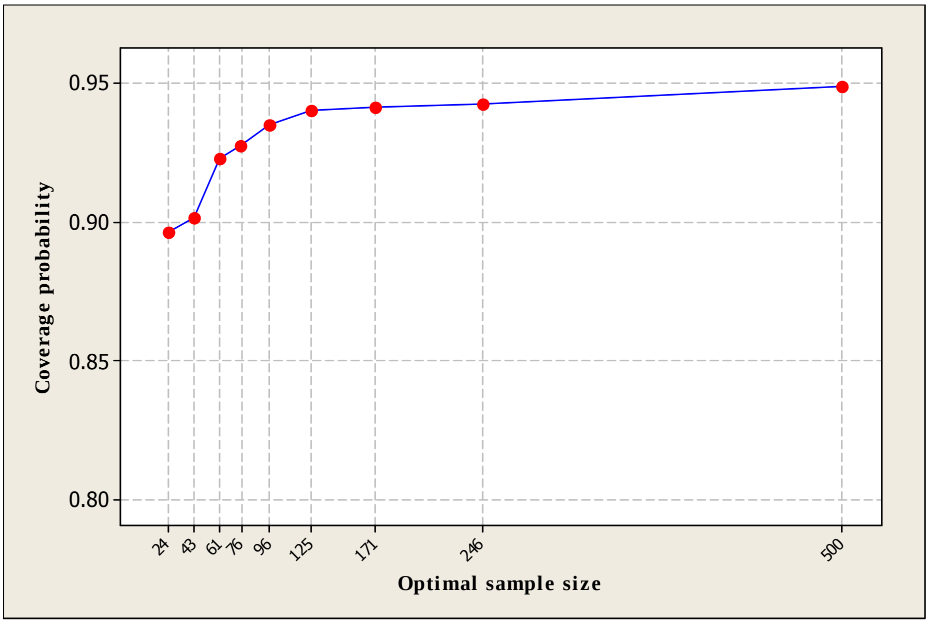

Figure 7.7: The simulated estimates of the coverage probability for underlying beta $(2,3)$ as the optimal sample size increases at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

To show the effect of δ on the coverage see Table 7.6 and Figure 7.8, which show the coverage probability as n^* increases at $m = 15$. We have same arguments as the previous cases. Similar arguments can be made regarding the comparison between the simulation results and the asymptotic results as in the case of the normal distribution.

n^*	$1-\widehat{\alpha}, \delta = 0.3$	$1-\widehat{\alpha}, \delta = 0.5$	$1-\widehat{\alpha}, \delta = 0.8$
24	0.8792	0.8962	0.9282
43	0.8088	0.9017	0.9276
61	0.8544	0.9228	0.9365
76	0.8887	0.9278	0.9421
96	0.9135	0.9353	0.9435
125	0.9276	0.9405	0.9460
171	0.9362	0.9415	0.9498
246	0.9420	0.9429	0.9487
500	0.9467	0.9490	0.9516

Table 7.6: The simulated coverage probability for underlying beta $(2,3)$ as the optimal sample size increases; $1-\alpha = 0.95$ and $m = 15$.

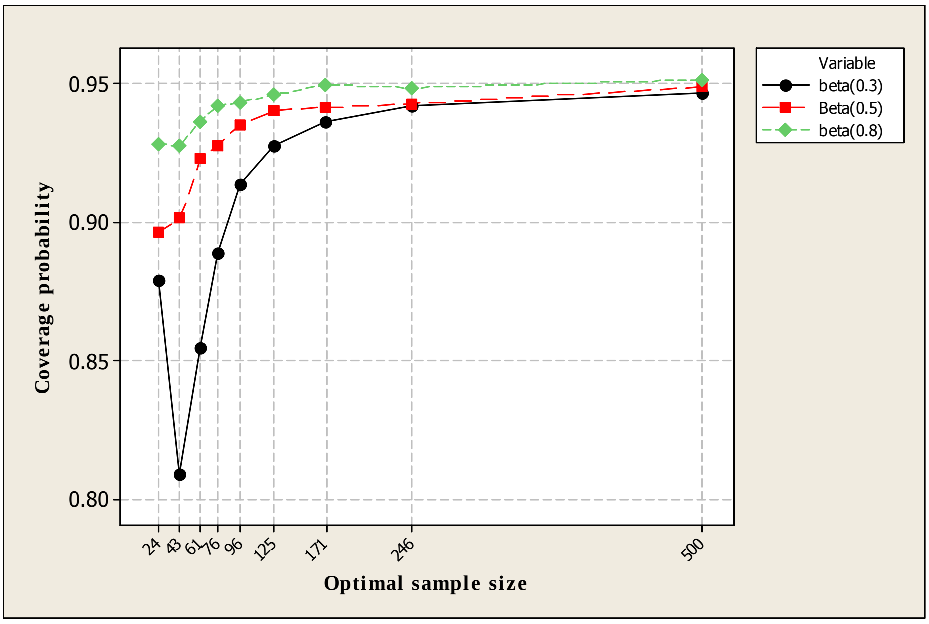

Figure 7.8: The effect of changing δ on the simulated estimates of the coverage probability for underlying beta$(2,3)$ distribution at $\alpha=0.05, m=15$ and $\delta=0.5$.

7.2.5 The underlying distribution is exponential with mean one

From Table B8 and for brevity we have the following

Regarding the coverage probability, we noticed bad estimates for the coverage, and the reason goes due to the sharp value of the skewness and high value of the kurtosis, which delay the convergence of the coverage probability. Better performance expected as both n^* and m increase.

Figure 7.9 shows the coverage probability when the underlying distribution is exponential with mean one, $Exp(\mu=1)$. Clearly the convergence to the nominal coverage is slow..

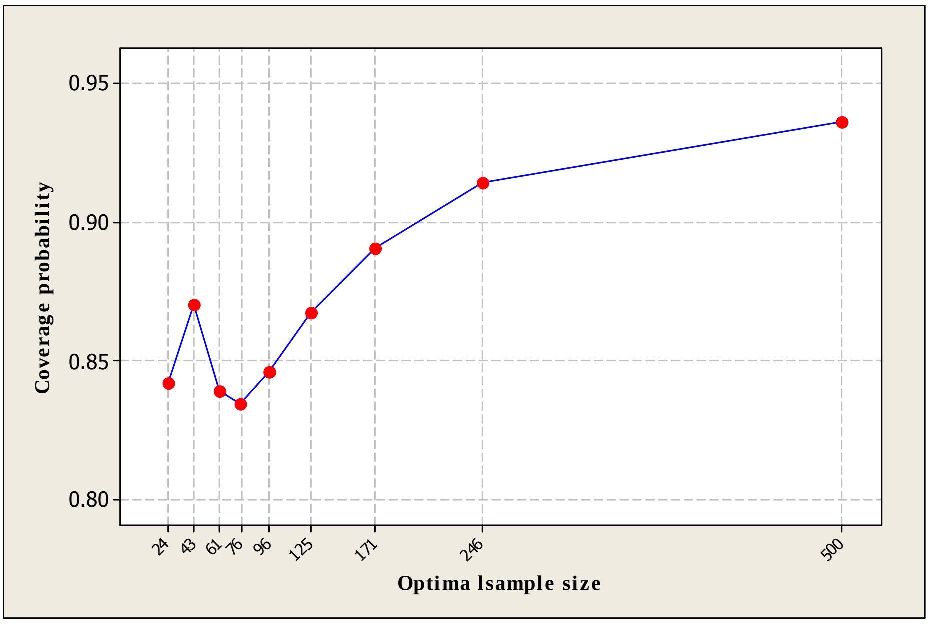

Figure 7.9: The simulated estimates of the coverage probability for underlying distribution $Exp(1)$ as the optimal sample size increases at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

Similarly to show the effect of increasing δ on the performance of the coverage probability at $1-\alpha = 0.95$, $m = 15$ we consider Table 7.7 and Figure 7.10. Both Table 7.9 and Figure 7.10 support our discussion above and reflect the poor coverage for all values of δ. Recall that the asymptotic coverage probabilities are unsatisfactory because of the poor approximation provided by the Edgeworth series.

n^*	$1-\hat{\alpha}, \delta = 0.3$	$1-\hat{\alpha}, \delta = 0.5$	$1-\hat{\alpha}, \delta = 0.8$
24	0.9173	0.8420	0.9496
43	0.8522	0.8706	0.8818
61	0.8128	0.8389	0.8671
76	0.7943	0.8346	0.8754
96	0.7873	0.8463	0.8876
125	0.8085	0.8676	0.9053
171	0.8442	0.8908	0.9207
246	0.8806	0.9141	0.9323
500	0.9245	0.9361	0.9478

Table 7.7: The simulated coverage probability for underlying $Exp(1)$ as the optimal sample size increases; $1-\alpha = 0.95$ and $m = 15$.

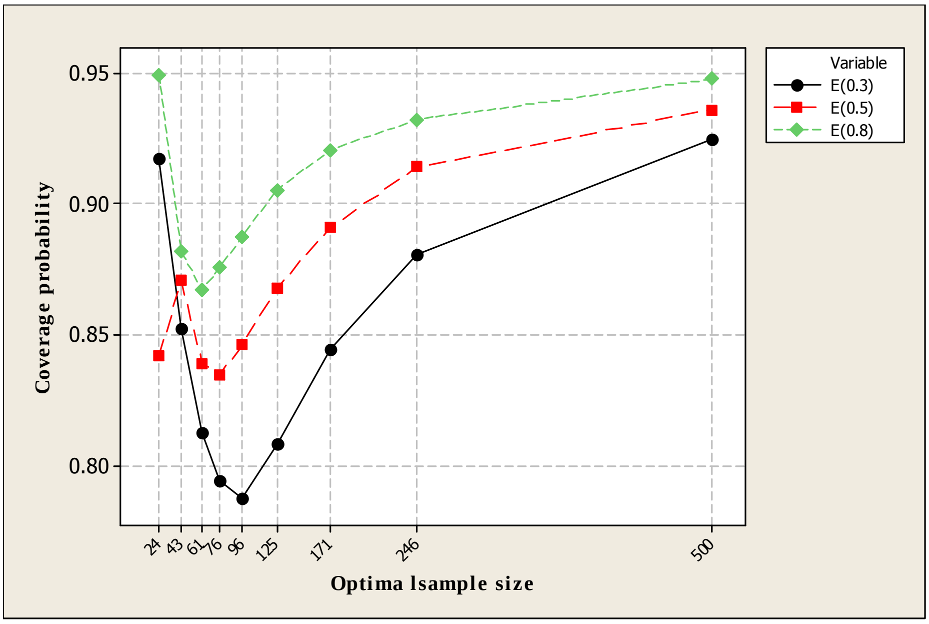

Figure 7.10: The effect of changing δ on the simulated estimates of the coverage probability for underlying distribution $Exp(1)$ as the optimal sample size increases at $\alpha = 0.05, m = 15$ and $\delta = 0.5$.

Before finishing the chapter we show the behaviour of the coverage for a collection of underlying distributions together. Figure 7.11 shows the behaviour of the coverage probability for the three distributions: normal, uniform and exponential all treated at $1-\alpha = 0.95$, $\delta = 0.5$ and $m = 15$. It is obvious that the coverage probability under the normal distribution is between the coverage probabilities under the uniform and exponential distributions. As n^* increases all the coverage probabilities approach the nominal coverage.

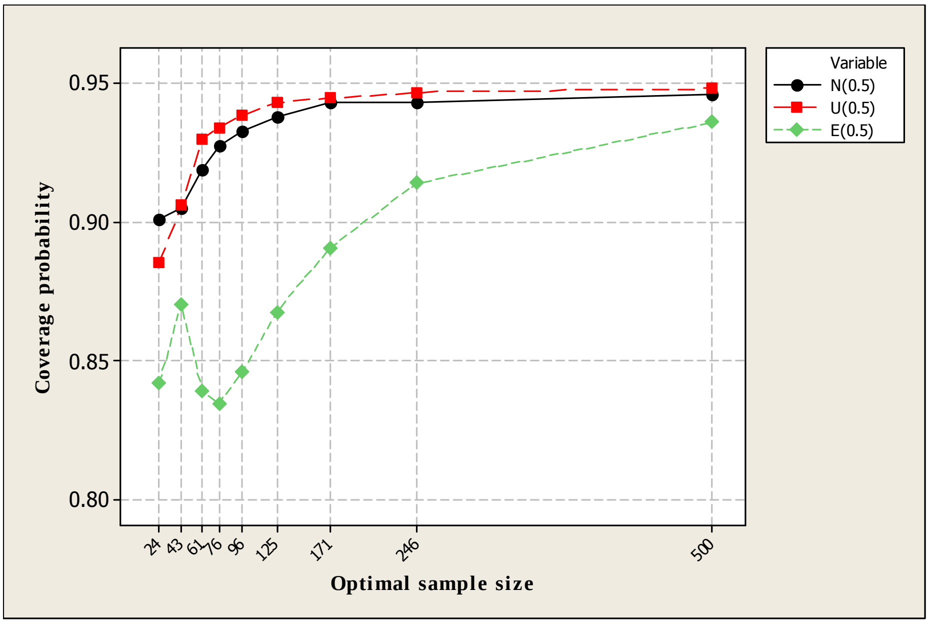

Figure 7.11: The simulated estimates of the coverage probability for underlying $N(0,1), U(0,1)$ and $Exp(1)$ distributions as the optimal sample size increases at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

Figure 7.12 shows the behaviour of the coverage probability for the three distributions for different values of δ and as the optimal sample size increases. The graph shows the following: first, the bad behaviour of the coverage probability under the exponential distribution and secondly, the improved estimate of the coverage probability occurs as the design factor increases.

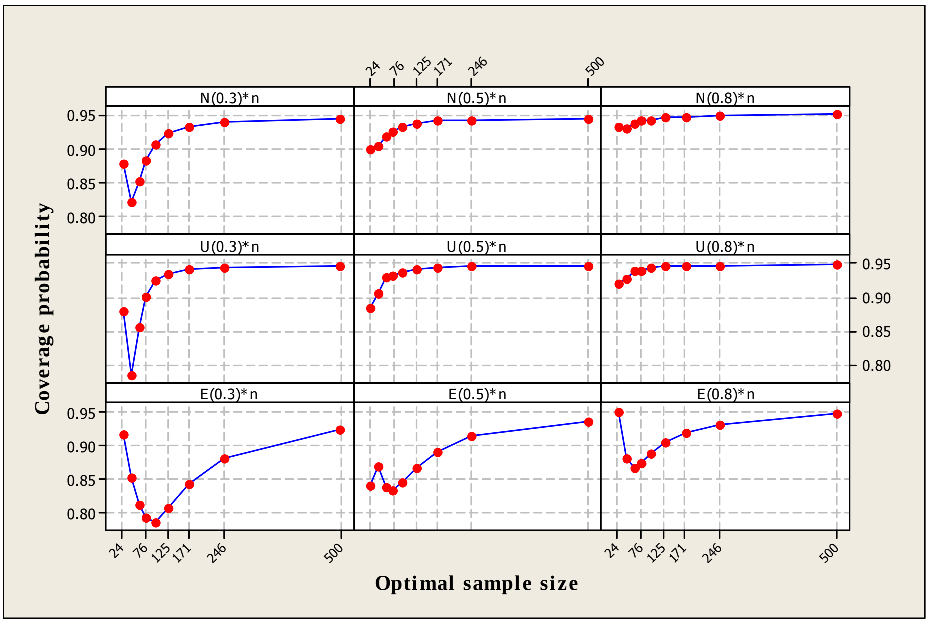

Figure 7.12: The simulated estimates of the coverage probability for underlying $N(0,1), U(0,1)$ and $Exp(1)$ at $\alpha = 0.05$, $\delta = 0.3, 0.5, 0.8$ and $m = 15$.

Figure 7.13 below discuss the same point as before but for the following underlying distributions: normal, uniform, beta and exponential. Obviously, the coverage probability of the normal, uniform and beta have same behaviour while the case of the exponential is completely different.

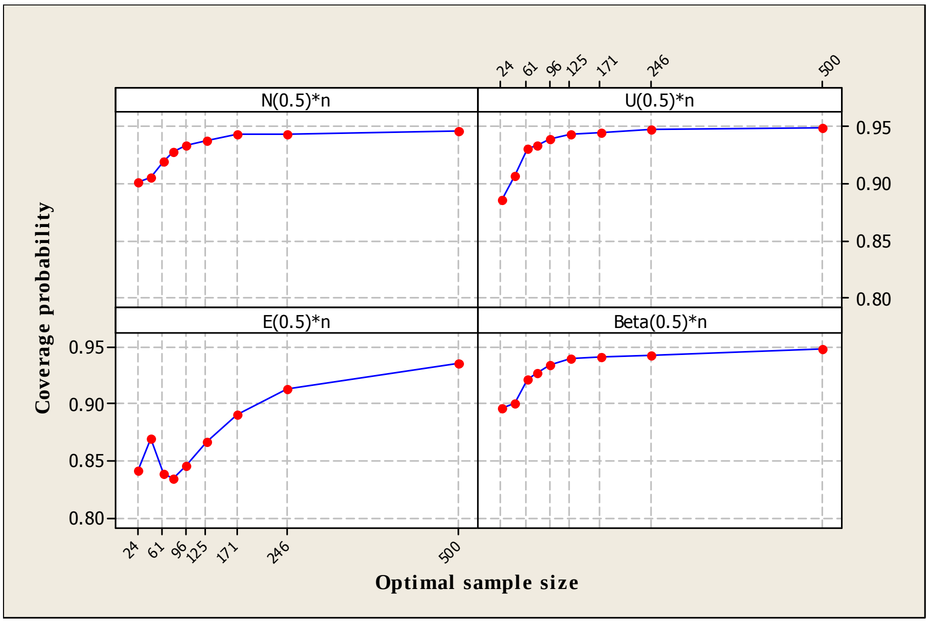

Figure 7.13: The simulated estimates of the coverage probability for underlying $N(0,1), U(0,1)$, beta $(2,3)$ and $Exp(1)$ as the optimal sample size increases at $\alpha = 0.05$, $\delta = 0.5$ and $m = 15$.

Finally, the good or bad behaviour of the asymptotic coverage probability is controlled mainly by the behaviour of the Edgeworth approximation to the underlying distribution and therefore to the values of the skewness and kurtosis of the underlying distribution. So in the cases where the underlying distribution is poorly approximated, the asymptotic coverages naturally differ considerably from the simulated coverage values.

Chapter VIII

Simulation Results of Triple Sampling Fixed Width Confidence Intervals with Controlled Type II Error Probabilities

In Chapter VI, we obtained an asymptotic mathematical representations for the triple sampling fixed width confidence intervals with controlled Type II error (Theorem 6.4.1), and also the corresponding Type II error probability (Theorem 6.4.2).

8.1. Simulation results regarding the triple sampling fixed width confidence interval with controlled Type II error probability

In order to investigate the performance of the triple sampling fixed width confidence intervals with controlled Type II error under$(6.17),(6.18)$, a series of simulation studies were performed with small, moderate and large optimal sample sizes. The same class of underlying distributions, design factors and pilot sample sizes were used as in previous simulation studies for this thesis. Moreover same number of replicate samples as in the previous simulations.

Tables D1 to D9 in Appendix D represent all the simulation results for the underlying distributions: standard normal, standard uniform, t with $r = 5,10,25,50$ and 100 degrees of freedom, beta$(2,3)$ and exponential with mean one at $\alpha = 0.05, \delta = 0.5$ and $m = 5,15,20$. In all cases the coverage probability under the controlled optimal fixed sample size exceeds the nominal value even for small values of n^* and for all values of m and δ.

8.2 Simulation results to estimate the Type II error probability

Although the Type II error probability is asymptotically and mathematically expressed in Theorem 6.4.2, it is of interest to estimate the Type II error probability for small, moderate to large optimal sample sizes and to investigate the effect of underlying distribution, δ, *m*, shift k and n^*.

For brevity, we consider here only three underlying distributions: standard normal, standard uniform and exponential with mean one with $m = 5,15,20$, $\delta = 0.3, 0.5$, 0.8 ,$\alpha = 0.05$ and $\beta_t = 0.05$.

Tables 8.1, 8.2 and 8.3 show the estimated Type II error probability under the normal, uniform and exponential distribution respectively at $\delta = 0.5$, $\alpha = 0.05$, $\beta_t = 0.05$, $m = 5,15,20$ and k = 0 (0.01) 0.1 (0.1) 0.5. We see from the tables that cases with uniform and normal underlying distributions exhibit a similar pattern of Type II error probabilities across values of m, n^*, α, β_t and δ. We also see that for small shift there is a high probability of committing a Type II error, as anticipated. These probabilities decrease as k increases for both distributions. The exponential case behaves similarly except with changing *m*. In this case for fixed k and n^* the probability of committing a Type II error increases with m. One explanation is that symmetric distributions give similar Type II error results whatever the kurtosis, especially with large values of n^*.

	$n^*, m = 5,15,20$																	
	24			**76**			**96**			**125**			**246**			**500**		
k	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**
.00	.499	.501	.499	.499	.500	.502	.504	.502	.501	.499	.501	.501	.500	.499	.502	.499	.498	.499
.01	.487	.488	.485	.487	.486	.487	.490	.489	.486	.484	.487	.487	.485	.486	.487	.485	.484	.485
.02	.475	.476	.471	.473	.473	.474	.476	.474	.472	.470	.473	.472	.471	.471	.473	.470	.470	.470
.03	.462	.464	.458	.460	.459	.460	.463	.461	.459	.456	.460	.458	.457	.457	.460	.457	.455	.457
.04	.449	.451	.445	.447	.446	.447	.450	.446	.445	.444	.445	.445	.443	.442	.445	.442	.442	.443
.05	.437	.438	.430	.433	.432	.433	.436	.432	.431	.430	.431	.431	.430	.428	.431	.428	.427	.430
.06	.425	.426	.417	.419	.419	.419	.422	.419	.419	.417	.418	.417	.416	.414	.416	.414	.414	.415
.07	.412	.413	.404	.406	.406	.407	.408	.405	.405	.404	.405	.403	.402	.400	.402	.400	.400	.401
.08	.401	.402	.390	.393	.393	.394	.394	.391	.392	.392	.391	.389	.388	.386	.389	.386	.385	.388
.09	.388	.390	.378	.380	.380	.381	.381	.378	.379	.377	.378	.377	.374	.373	.375	.373	.371	.374
.10	.375	.378	.365	.367	.367	.368	.368	.366	.365	.363	.365	.363	.361	.359	.362	.359	.358	.361
.20	.266	.265	.250	.252	.245	.249	.247	.245	.244	.244	.242	.243	.240	.238	.238	.234	.237	.240
.30	.183	.175	.158	.162	.151	.153	.155	.151	.151	.152	.147	.148	.145	.141	.143	.139	.142	.143
.40	.121	.107	.091	.100	.087	.087	.092	.087	.086	.088	.082	.082	.079	.077	.077	.075	.077	.078
.50	.081	.062	.048	.060	.048	.047	.052	.047	.045	.048	.043	.043	.040	.037	.039	.037	.038	.038

Table 8.1: The simulated estimates of Type II error probabilities for underlying $N(0,1)$ distribution as the optimal sample size increases and k increases with $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 5,15,20$.

	$n^*, m = 5,15,20$																	
	24			76			96			125			246			500		
K	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20
.00	.502	.500	.501	.497	.497	.498	.494	.504	.500	.495	.503	.499	.498	.495	.499	.503	.500	.501
.01	.489	.488	.487	.483	.484	.484	.480	.490	.486	.481	.489	.485	.484	.481	.484	.487	.485	.487
.02	.476	.475	.473	.470	.470	.470	.466	.476	.471	.467	.475	.471	.469	.467	.471	.473	.471	.472
.03	.463	.464	.460	.457	.456	.455	.452	.462	.457	.454	.460	.458	.455	.453	.456	.458	.455	.458
.04	.451	.452	.447	.443	.442	.442	.438	.447	.443	.440	.445	.444	.441	.440	.442	.444	.440	.443
.05	.439	.440	.434	.430	.428	.428	.425	.433	.429	.426	.432	.431	.427	.425	.427	.429	.426	.429
.06	.426	.429	.421	.417	.415	.413	.412	.419	.416	.412	.418	.417	.412	.411	.414	.415	.412	.415
.07	.414	.417	.409	.404	.401	.399	.398	.405	.402	.399	.404	.403	.398	.397	.399	.402	.398	.401
.08	.401	.405	.396	.391	.388	.384	.385	.391	.389	.386	.390	.389	.384	.383	.386	.388	.384	.387
.09	.388	.394	.383	.377	.375	.371	.371	.377	.375	.373	.377	.376	.372	.370	.373	.375	.371	.374
.10	.377	.382	.370	.364	.362	.358	.360	.364	.361	.360	.363	.363	.359	.357	.359	.362	.358	.359
.20	.269	.276	.254	.247	.241	.240	.241	.242	.241	.238	.239	.239	.236	.235	.236	.236	.235	.237
.30	.185	.188	.161	.155	.147	.146	.149	.147	.147	.145	.144	.144	.140	.140	.141	.141	.140	.141
.40	.126	.121	.094	.091	.083	.083	.085	.081	.082	.081	.080	.079	.078	.076	.077	.075	.074	.076
.50	.088	.073	.050	.052	.044	.043	.048	.041	.041	.042	.040	.040	.038	.037	.038	.037	.036	.036

Table 8.2: The simulated estimates of Type II error probabilities for underlying $U(0,1)$ distribution as the optimal sample size increases and k increases with $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 5,15,20$.

	$n^*, m = 5,15,20$																	
	24			**76**			**96**			**125**			**246**			**500**		
k	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**	**5**	**15**	**20**
.00	.355	.421	.457	.390	.411	.412	.406	.420	.424	.415	.435	.438	.449	.460	.463	.468	.470	.474
.01	.345	.408	.443	.378	.398	.400	.393	.407	.412	.403	.422	.424	.436	.447	.449	.453	.457	.460
.02	.333	.393	.428	.366	.387	.387	.381	.395	.399	.390	.409	.412	.421	.433	.436	.438	.443	.446
.03	.322	.380	.414	.354	.374	.374	.369	.381	.387	.377	.396	.398	.409	.418	.422	.424	.430	.431
.04	.311	.367	.400	.343	.362	.362	.357	.368	.374	.365	.384	.385	.396	.405	.408	.410	.415	.417
.05	.301	.354	.386	.332	.350	.350	.345	.356	.362	.352	.370	.372	.382	.392	.395	.396	.401	.403
.06	.290	.341	.373	.320	.337	.338	.333	.343	.350	.340	.357	.360	.368	.378	.382	.382	.387	.390
.07	.280	.328	.359	.309	.325	.326	.321	.331	.338	.327	.344	.347	.356	.365	.368	.367	.373	.377
.08	.268	.315	.346	.297	.315	.314	.310	.320	.325	.316	.332	.334	.343	.352	.354	.354	.359	.363
.09	.258	.302	.333	.285	.303	.303	.298	.308	.312	.304	.319	.321	.331	.339	.341	.340	.345	.349
.10	.249	.290	.320	.274	.291	.291	.287	.297	.300	.292	.306	.308	.317	.327	.328	.326	.332	.335
.20	.157	.181	.201	.174	.185	.183	.182	.188	.192	.187	.194	.196	.204	.206	.210	.207	.211	.213
.30	.090	.100	.113	.103	.105	.104	.105	.107	.111	.108	.110	.112	.119	.116	.119	.119	.120	.123
.40	.049	.050	.057	.057	.053	.053	.057	.055	.056	.058	.055	.056	.063	.059	.059	.064	.061	.061
.50	.027	.024	.026	.030	.023	.024	.030	.025	.026	.030	.025	.025	.031	.027	.027	.031	.028	.027

Table 8.3: The simulated estimates of Type II error probabilities for underlying $Exp(1)$ distribution as the optimal sample size increases and k increase with $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 5,15,20$.

Figures 8.1 to 8.3 show the effect of increasing n^* on the estimated Type II error probability for the three underlying distributions at $m = 15$, $\delta = 0.5$, $\alpha = 0.05$ and $\beta_t = 0.05$. We see that for small values of n^* and for small values of k (less than 0.1), the probability of committing Type II error gets as high as 0.5 for both the uniform and the normal, while it gets as high as 0.47 for the exponential. The difference between the uniform and the normal cases becomes clear for larger values of k, while for the exponential the gap is significant over the entire range of k. As n^* increases the difference between the uniform and the normal results becomes small and the gap between the results of the exponential and the other two distributions gets narrower (see Table E1 for more details).

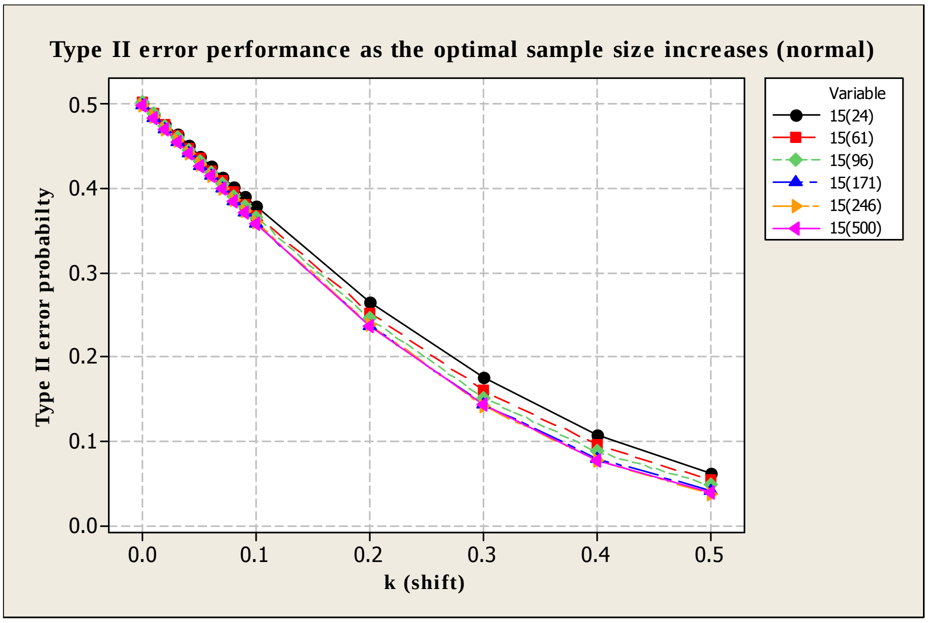

Figure 8.1: Performance of the Type II error probability for underlying $N(0,1)$ as the optimal sample size and k increase at $\delta = 0.5, \alpha = 0.05, \beta_t = 0.05$ and $m = 15$.

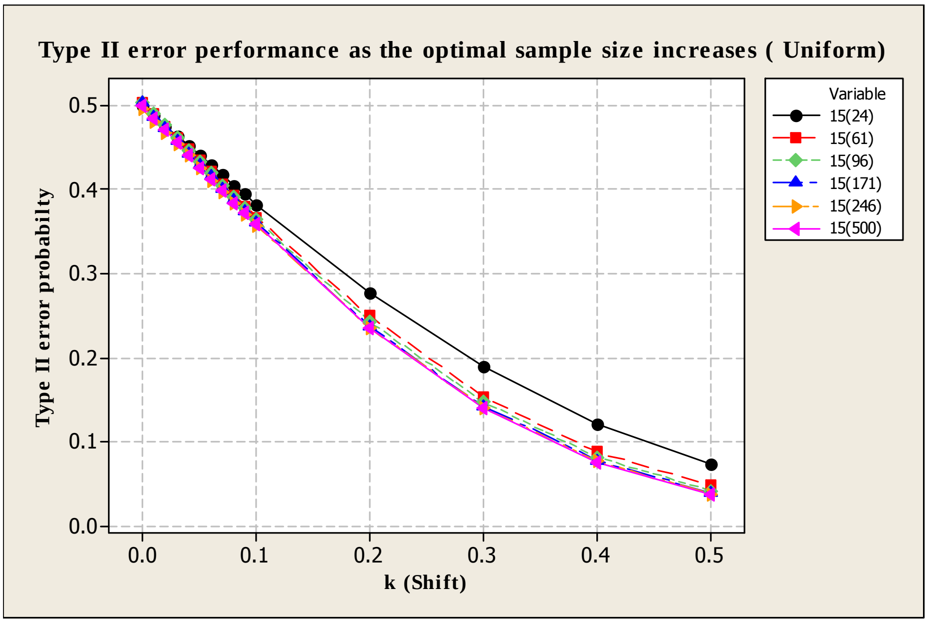

Figure 8.2: Performance of the Type II error probability for underlying $U(0,1)$ as the optimal sample size and k increase at $\delta = 0.5, \alpha = 0.05, \beta_t = 0.05$ and $m = 15$.

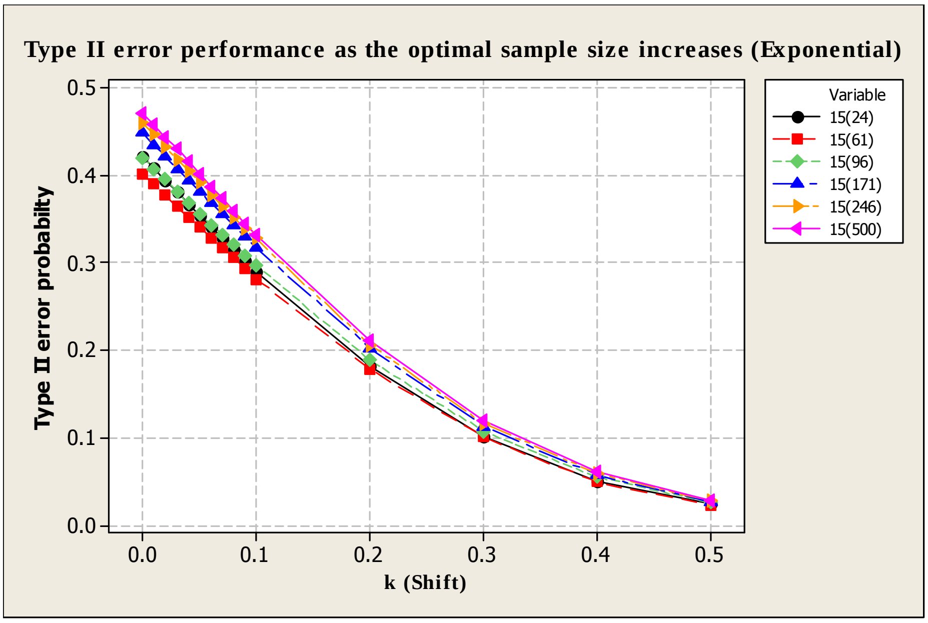

Figure 8.3: Performance of the Type II error probability for underlying $Exp(1)$ as the optimal sample size and k increase at $\delta = 0.5$, $\alpha = 0.05$, $\beta_t = 0.05$ and $m = 15$.

To illustrate the difference between the simulation results and the asymptotic results obtained in chapter VI we consider Tables 8.4, 8.5 and 8.6 for the above underlying distributions respectively. Tables 8.4, 8.5 and 8.6 compare the simulation results and the asymptotic results at $\alpha = 0.05$, $\beta_t = 0.05$, $\delta = 0.5$ and $m = 15$. The asymptotic values β_{asy} and the corresponding estimated Type II error probabilities for the normal and the uniform underlying distributions are in broad agreement for all values of n^* and k. Moreover, as n^* increases the difference between the asymptotic and estimated values decreases. From Table 8.6 we see the bad behaviour of the asymptotic Type II error probability in comparison with the simulation results in the exponential distribution case, as expected given the poor performance of the Edgeworth approximation (see Chapter IV, section 6.2).

k	β_{asy}	24	43	61	76	96	125	171	246	500
.00	0.500	0.501	0.500	0.503	0.500	0.502	0.501	0.499	0.499	0.498
.01	0.486	0.488	0.486	0.489	0.486	0.489	0.487	0.484	0.486	0.484
.02	0.471	0.476	0.473	0.475	0.473	0.474	0.473	0.470	0.471	0.470
.03	0.457	0.464	0.460	0.462	0.459	0.461	0.46	0.455	0.457	0.455
.04	0.443	0.451	0.448	0.448	0.446	0.446	0.445	0.441	0.442	0.442
.05	0.429	0.438	0.434	0.435	0.432	0.432	0.431	0.427	0.428	0.427
.06	0.414	0.426	0.421	0.421	0.419	0.419	0.418	0.414	0.414	0.414
.07	0.400	0.413	0.408	0.409	0.406	0.405	0.405	0.400	0.400	0.400
.08	0.387	0.402	0.396	0.395	0.393	0.391	0.391	0.385	0.386	0.385
.09	0.373	0.390	0.383	0.381	0.380	0.378	0.378	0.372	0.373	0.371
.10	0.359	0.378	0.371	0.368	0.367	0.366	0.365	0.358	0.359	0.358
.20	0.236	0.265	0.253	0.252	0.245	0.245	0.242	0.237	0.238	0.237
.30	0.140	0.175	0.167	0.160	0.151	0.151	0.147	0.143	0.141	0.142
.40	0.075	0.107	0.100	0.095	0.087	0.087	0.082	0.079	0.077	0.077
.50	0.036	0.062	0.059	0.053	0.048	0.047	0.043	0.040	0.037	0.038

Table 8.4: Comparison between the asymptotic and simulation results for the Type II error probability for underlying distribution $N(0,1)$ as the optimal sample size and k increase; $\alpha = 0.05$, $\beta_t = 0.05$, $\delta = 0.5$ and $m = 15$.

k	β_{asy}	24	43	61	76	96	125	171	246	500
.00	0.500	0.500	0.501	0.503	0.497	0.504	0.503	0.503	0.495	0.500
.01	0.488	0.488	0.489	0.490	0.484	0.490	0.489	0.487	0.481	0.485
.02	0.476	0.475	0.474	0.475	0.470	0.476	0.475	0.473	0.467	0.471
.03	0.463	0.464	0.460	0.462	0.456	0.462	0.460	0.458	0.453	0.455
.04	0.451	0.452	0.447	0.448	0.442	0.447	0.445	0.443	0.440	0.440
.05	0.439	0.440	0.434	0.434	0.428	0.433	0.432	0.429	0.425	0.426
.06	0.427	0.429	0.421	0.421	0.415	0.419	0.418	0.416	0.411	0.412
.07	0.415	0.417	0.408	0.407	0.401	0.405	0.404	0.401	0.397	0.398
.08	0.403	0.405	0.395	0.393	0.388	0.391	0.390	0.387	0.383	0.384
.09	0.391	0.394	0.383	0.380	0.375	0.377	0.377	0.374	0.37	0.371
.10	0.379	0.382	0.371	0.367	0.362	0.364	0.363	0.361	0.357	0.358
.20	0.263	0.276	0.253	0.249	0.241	0.242	0.239	0.236	0.235	0.235
.30	0.162	0.188	0.163	0.153	0.147	0.147	0.144	0.142	0.140	0.140
.40	0.084	0.121	0.100	0.087	0.083	0.081	0.080	0.077	0.076	0.074
.50	0.034	0.073	0.059	0.047	0.044	0.041	0.040	0.039	0.037	0.036

Table 8.5: Comparison between the asymptotic and simulation results for the Type II error probability for underlying distribution $U(0,1)$ as the optimal sample size and k increase; $\alpha = 0.05$, $\beta_t = 0.05$, $\delta = 0.5$ and $m = 15$.

k	β_{asy}	24	43	61	76	96	125	171	246	500
.00	0.633	0.421	0.391	0.401	0.411	0.420	0.435	0.448	0.460	0.470
.01	0.622	0.408	0.379	0.390	0.398	0.407	0.422	0.434	0.447	0.457
.02	0.616	0.393	0.367	0.377	0.387	0.395	0.409	0.421	0.433	0.443
.03	0.615	0.380	0.354	0.364	0.374	0.381	0.396	0.407	0.418	0.430
.04	0.618	0.367	0.342	0.352	0.362	0.368	0.384	0.394	0.405	0.415
.05	0.626	0.354	0.331	0.340	0.350	0.356	0.370	0.381	0.392	0.401
.06	0.638	0.341	0.319	0.328	0.337	0.343	0.357	0.368	0.378	0.387
.07	0.654	0.328	0.308	0.317	0.325	0.331	0.344	0.356	0.365	0.373
.08	0.674	0.315	0.297	0.305	0.315	0.320	0.332	0.343	0.352	0.359
.09	0.698	0.302	0.285	0.293	0.303	0.308	0.319	0.329	0.339	0.345
.10	0.725	0.290	0.274	0.281	0.291	0.297	0.306	0.316	0.327	0.332

Table 8.6: Comparison between the asymptotic and simulation results for the Type II error probability for underlying distribution $Exp(1)$ as the optimal sample size and k increase; $\alpha = 0.05$, $\beta_t = 0.05$, $\delta = 0.5$ and $m = 15$.

To compare the simulation results and the asymptotic results for the normal and uniform underlying distributions at $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 15$, Figures 8.4 and 8.5 show the estimated Type II error probabilities in comparison with the corresponding asymptotic values as n^* increases. For brevity we consider only specific values of n^*; $n^* = 24$, 96, 246 and 500. Clearly as n^* increases we attain consistent behaviour between the simulation and the asymptotic results.

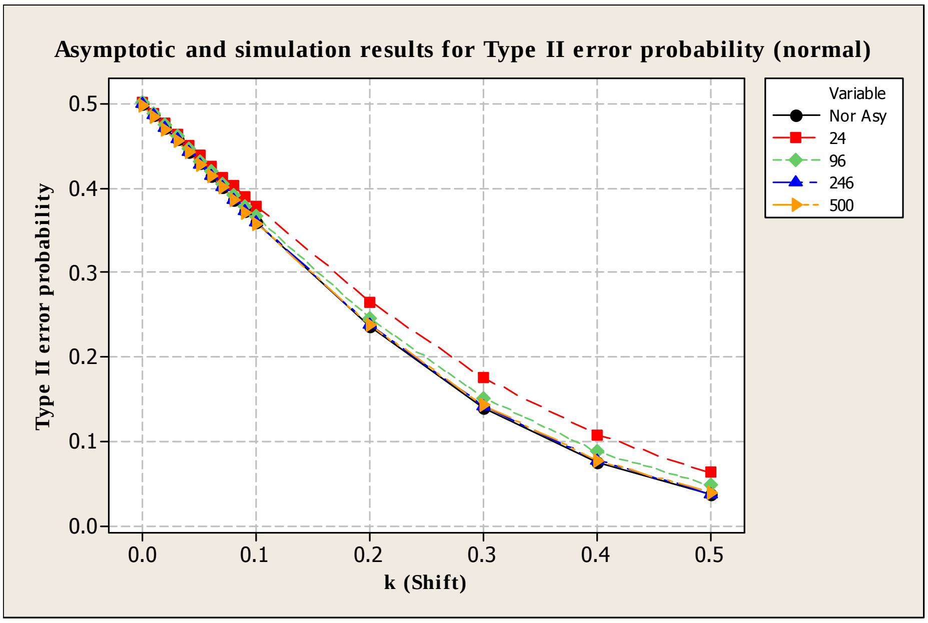

Figure 8.4. The difference between the asymptotic and the simulation Type II error probability for the underlying $N(0,1)$ as the optimal sample size and k increase; $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 15$.

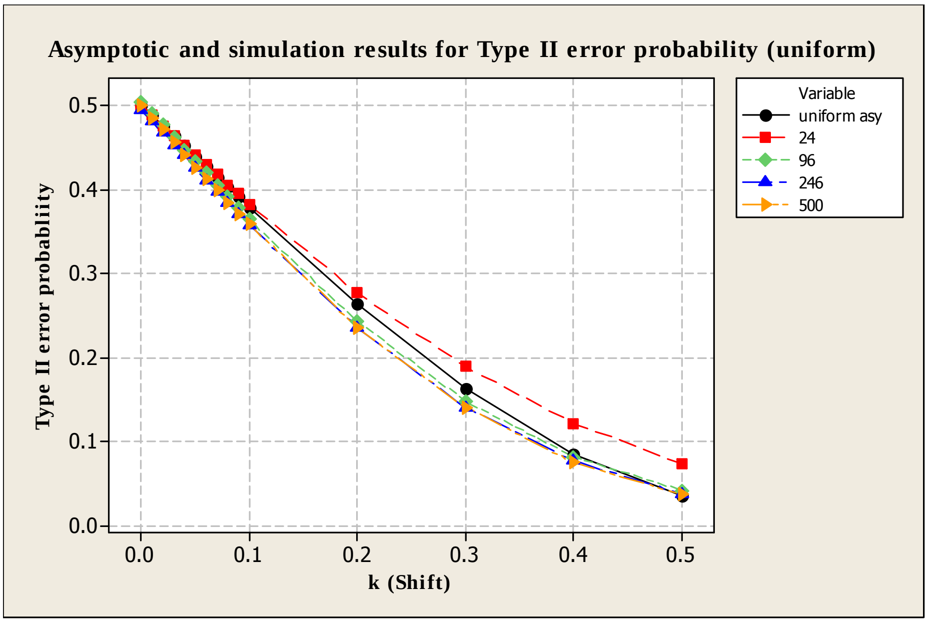

Figure 8.5. The difference between the asymptotic and the simulation Type II error probability for the underlying $U(0,1)$ as the optimal sample size and k increase; $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 15$.

To illustrate the asymptotic Type II error probabilities for other underlying distributions, we consider the case of the t distribution with $r = 5, 25, 50$ and 100 degrees of freedom in Table 8.7 and of the chi-squared distribution with $r = 5, 10$ and 50 degrees of freedom in Table 8.8.

Tables 8.7 and 8.8 show the asymptotic Type II error probabilities as r increases and for different values of k for the t and chi-squared underlying distributions respectively. Note that $\beta_{asy}(r)$ denotes the asymptotic Type II error probability at a specific value of r.

We see in Table 8.7 that at $k = 0$ (no shift occurs) the asymptotic Type II error probability is 0.5 for all r, while as k increases for a fixed value of r, the asymptotic Type II error probability decreases. A s r increases and for a fixed value of $k > 0.5$ the asymptotic Type II error probability increases.

k	β_{asy} (5)	β_{asy} (25)	β_{asy} (50)	β_{asy} (100)
0.00	0.5000	0.5000	0.5000	0.5000
0.01	0.4748	0.4851	0.4854	0.4855
0.02	0.4498	0.4702	0.4708	0.4710
0.03	0.4249	0.4554	0.4562	0.4566
0.04	0.4003	0.4407	0.4417	0.4422
0.05	0.3760	0.4260	0.4273	0.4279
0.06	0.3521	0.4114	0.4130	0.4137
0.07	0.3288	0.3970	0.3988	0.3996
0.08	0.3060	0.3827	0.3848	0.3857
0.09	0.2839	0.3686	0.3709	0.3719
0.10	0.2625	0.3546	0.3571	0.3582
0.20	0.0979	0.2289	0.2325	0.2340
0.30	0.0297	0.1345	0.1373	0.1386
0.40	0.0278	0.0724	0.0736	0.0742
0.50	0.0446	0.0362	0.0359	0.0358

Table 8.7. The asymptotic Type II error probabilities for underlying t distribution with $r = 5, 25, 50$ and 100 as k increases.

In Table 8.8 we see that for a fixed r and as k increases and for a fixed k and as r increases, the asymptotic Type II error probability decreases. At $k = 0$, the asymptotic Type II error probability is 0.584 and not 0.5 as in the case of symmetric distributions.

k	β_{asy} (5)	β_{asy} (10)	β_{asy} (50)
0.00	0.5841	0.5595	0.5266
0.01	0.5709	0.5457	0.5124
0.02	0.5593	0.5326	0.4983
0.03	0.5491	0.5201	0.4844
0.04	0.5403	0.5084	0.4708
0.05	0.5329	0.4973	0.4574
0.06	0.5269	0.4870	0.4442
0.07	0.5222	0.4772	0.4313
0.08	0.5188	0.4681	0.4186
0.09	0.5165	0.4596	0.4062
0.10	0.5152	0.4517	0.3941
0.20	0.5411	0.3954	0.2872
0.30	0.5677	0.3533	0.2043
0.40	0.5304	0.2972	0.1398
0.50	0.4270	0.2248	0.0899

Table 8.8. The asymptotic Type II error probabilities for underlying chi- squared distribution with $r = 5$, 10 and 50 as k increases.

We end this section by plotting the simulated estimates of Type II error probabilities for the three distributions together at $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5$ and $m = 15$. Figures 8.6 to 8.9 show this for n* = 24, 76, 125, 246 and 500 respectively. Clearly the estimated Type II error probabilities in the normal and uniform cases are very similar, whereas the corresponding probabilities for the exponential case tend to be lower. However, the difference declines as n* increases.

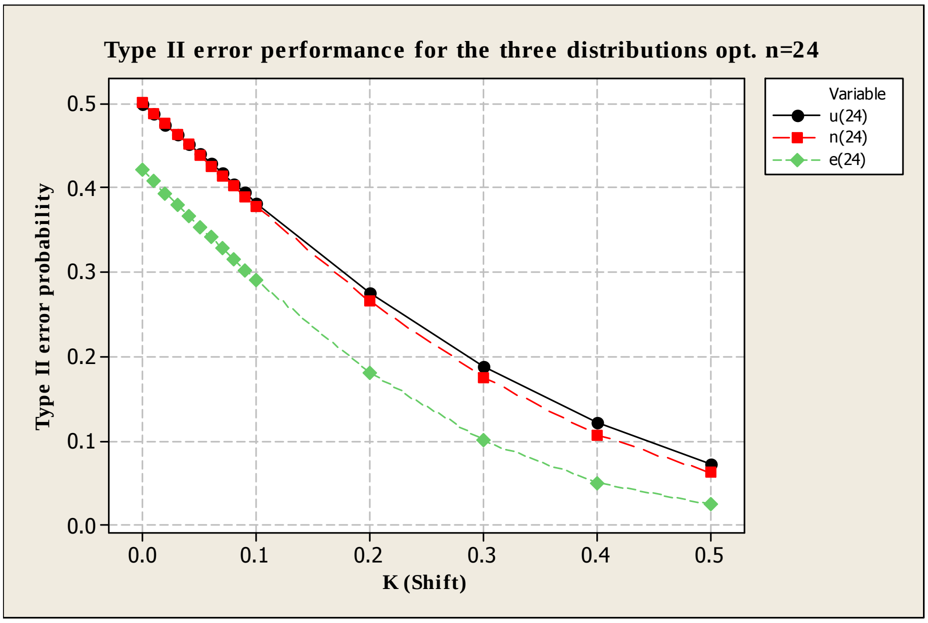

Figure 8.6: The simulated estimates of Type II error probability for underlying normal, uniform and exponential distributions as k increases at $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5, m = 15$ and $n^* = 24$.

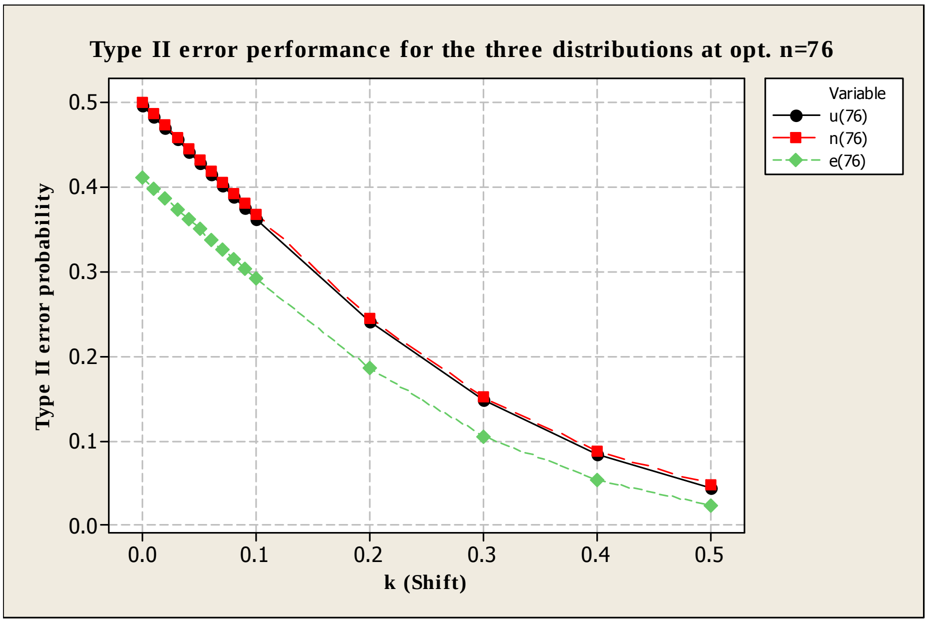

Figure 8.7: The simulated estimates of Type II error probability for underlying normal, uniform and exponential distributions as k increases at $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5, m = 15$ and $n^* = 76$.

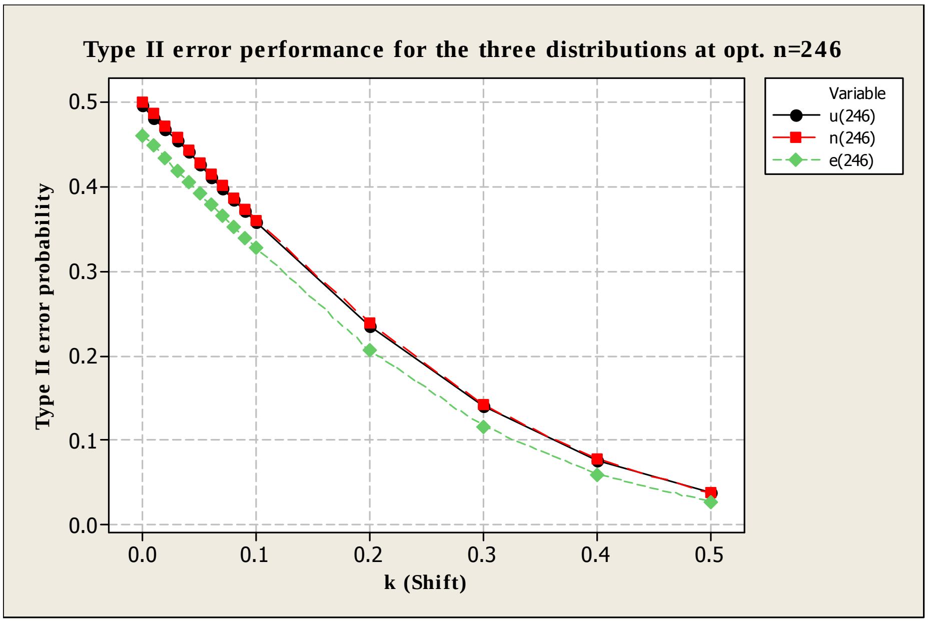

Figure 8.8: The simulated estimates of Type II error probability for underlying normal, uniform and exponential distributions as k increases at $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5, m = 15$ and $n^* = 246$.

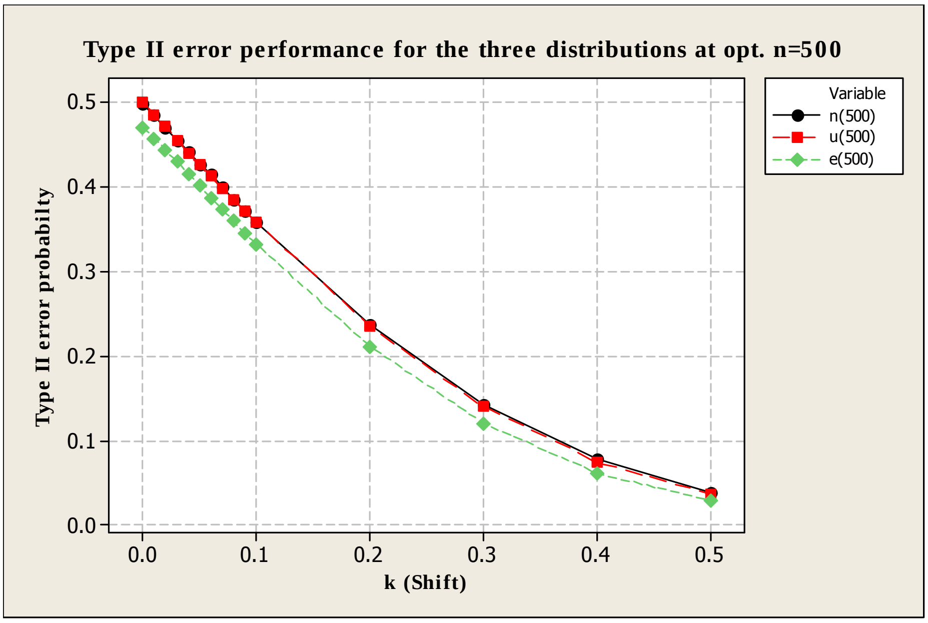

Figure 8.9: The simulated estimates of Type II error probability for underlying normal, uniform and exponential distributions as k increases at $\alpha = 0.05, \beta_t = 0.05, \delta = 0.5, m = 15$ and $n^* = 500$.

8.3 The effect of δ on the Type II error probability

The effect of δ on the estimated Type II probability at $\alpha = 0.05, \beta_t = 0.05$ and $m = 15$ is illustrated in Tables 8.9, 8.10 and 8.11 for the normal, uniform and exponential underlying distributions respectively. We see that for small values of n^* and by holding other parameters constant the Type II error probabilities at $\delta = 0.5$ lie between those at $\delta = 0.3$ and $\delta = 0.8$. As n^* increases the estimated Type II error probabilities tend the coincide for all values of δ . This new findings support the choice of $\delta = 0.5$ suggested by Hall (1981) even our justification came from different perspective.

n^*	24			76			96			125			246			500		
k	$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$		
.00	0.501	0.501	0.499	0.500	0.500	0.502	0.496	0.502	0.504	0.500	0.501	0.499	0.496	0.499	0.499	0.501	0.498	0.502
.01	0.489	0.488	0.485	0.487	0.486	0.487	0.482	0.489	0.490	0.486	0.487	0.485	0.481	0.486	0.484	0.487	0.484	0.487
.02	0.478	0.476	0.470	0.475	0.473	0.472	0.469	0.474	0.475	0.473	0.473	0.470	0.468	0.471	0.470	0.474	0.470	0.473
.03	0.466	0.464	0.457	0.462	0.459	0.458	0.455	0.461	0.461	0.460	0.460	0.456	0.454	0.457	0.456	0.460	0.455	0.458
.04	0.455	0.451	0.443	0.450	0.446	0.444	0.441	0.446	0.447	0.446	0.445	0.441	0.441	0.442	0.442	0.445	0.442	0.444
.05	0.444	0.438	0.428	0.437	0.432	0.430	0.429	0.432	0.433	0.432	0.431	0.426	0.428	0.428	0.427	0.432	0.427	0.429
.06	0.433	0.426	0.415	0.424	0.419	0.416	0.416	0.419	0.419	0.419	0.418	0.411	0.414	0.414	0.414	0.419	0.414	0.413
.07	0.422	0.413	0.402	0.412	0.406	0.402	0.404	0.405	0.405	0.405	0.405	0.398	0.401	0.400	0.399	0.406	0.400	0.399
.08	0.411	0.402	0.388	0.400	0.393	0.388	0.392	0.391	0.390	0.392	0.391	0.383	0.387	0.386	0.385	0.392	0.385	0.385
.09	0.399	0.390	0.375	0.387	0.380	0.373	0.379	0.378	0.376	0.379	0.378	0.369	0.374	0.373	0.370	0.378	0.371	0.372
.10	0.388	0.378	0.360	0.374	0.367	0.360	0.366	0.366	0.362	0.366	0.365	0.355	0.361	0.359	0.356	0.364	0.358	0.358
.20	0.286	0.265	0.242	0.261	0.245	0.237	0.251	0.245	0.238	0.247	0.242	0.231	0.240	0.238	0.232	0.239	0.237	0.232
.30	0.198	0.175	0.149	0.173	0.151	0.144	0.162	0.151	0.145	0.153	0.147	0.136	0.147	0.141	0.135	0.145	0.142	0.137
.40	0.129	0.107	0.087	0.110	0.087	0.078	0.098	0.087	0.078	0.091	0.082	0.073	0.081	0.077	0.072	0.078	0.077	0.072
.50	0.078	0.062	0.047	0.069	0.048	0.039	0.056	0.047	0.039	0.050	0.043	0.035	0.041	0.037	0.034	0.038	0.038	0.034

Table 8.9: The effect of increasing δ on the simulated estimates of Type II error probability for underlying $N(0,1)$ as the optimal sample size and k increase; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

n^*	24			76			96			125			246			500		
k	$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$			$\delta = 0.3, 0.5, 0.8$		
.00	0.500	0.500	0.502	0.497	0.497	0.501	0.501	0.504	0.498	0.501	0.503	0.498	0.503	0.495	0.498	0.498	0.500	0.501
.01	0.489	0.488	0.487	0.483	0.484	0.486	0.487	0.490	0.485	0.487	0.489	0.483	0.489	0.481	0.484	0.485	0.485	0.486
.02	0.477	0.475	0.473	0.470	0.470	0.473	0.474	0.476	0.470	0.473	0.475	0.469	0.474	0.467	0.469	0.471	0.471	0.471
.03	0.467	0.464	0.461	0.456	0.456	0.459	0.460	0.462	0.456	0.459	0.460	0.454	0.459	0.453	0.456	0.456	0.455	0.457
.04	0.455	0.452	0.447	0.442	0.442	0.444	0.446	0.447	0.442	0.445	0.445	0.438	0.445	0.440	0.441	0.441	0.440	0.442
.05	0.444	0.440	0.435	0.429	0.428	0.431	0.432	0.433	0.428	0.432	0.432	0.423	0.432	0.425	0.427	0.428	0.426	0.428
.06	0.434	0.429	0.421	0.416	0.415	0.415	0.420	0.419	0.414	0.418	0.418	0.409	0.417	0.411	0.413	0.414	0.412	0.414
.07	0.422	0.417	0.406	0.403	0.401	0.401	0.406	0.405	0.400	0.405	0.404	0.395	0.403	0.397	0.399	0.400	0.398	0.400
.08	0.411	0.405	0.392	0.391	0.388	0.388	0.392	0.391	0.387	0.391	0.390	0.382	0.389	0.383	0.386	0.386	0.384	0.387
.09	0.400	0.394	0.379	0.378	0.375	0.374	0.378	0.377	0.372	0.377	0.377	0.368	0.377	0.370	0.372	0.372	0.371	0.372
.10	0.388	0.382	0.367	0.365	0.362	0.360	0.366	0.364	0.359	0.364	0.363	0.356	0.363	0.357	0.359	0.358	0.358	0.359
.20	0.286	0.276	0.247	0.251	0.241	0.236	0.246	0.242	0.237	0.241	0.239	0.234	0.239	0.235	0.234	0.234	0.235	0.233
.30	0.198	0.188	0.153	0.160	0.147	0.141	0.154	0.147	0.142	0.148	0.144	0.140	0.143	0.140	0.139	0.140	0.140	0.139
.40	0.127	0.121	0.090	0.098	0.083	0.078	0.090	0.081	0.078	0.084	0.080	0.076	0.080	0.076	0.076	0.076	0.074	0.073
.50	0.076	0.073	0.050	0.059	0.044	0.039	0.050	0.041	0.039	0.044	0.040	0.037	0.040	0.037	0.036	0.037	0.036	0.035

Table 8.10: The effect of increasing δ on the simulated estimates of Type II error probability for underlying $U\left(0,1\right)$ as the optimal sample size and k increase; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

n^*	24			76			96			125			246			500		
k	$\delta=0.3,0.5,0.8$			$\delta=0.3,0.5,0.8$			$\delta=0.3,0.5,0.8$			$\delta=0.3,0.5,0.8$			$\delta=0.3,0.5,0.8$			$\delta=0.3,0.5,0.8$		
.00	0.453	0.421	0.408	0.387	0.411	0.423	0.399	0.420	0.430	0.417	0.435	0.448	0.450	0.460	0.462	0.472	0.470	0.475
.01	0.440	0.408	0.394	0.376	0.398	0.410	0.387	0.407	0.416	0.406	0.422	0.433	0.438	0.447	0.449	0.459	0.457	0.462
.02	0.427	0.393	0.380	0.365	0.387	0.396	0.375	0.395	0.402	0.394	0.409	0.419	0.425	0.433	0.434	0.445	0.443	0.446
.03	0.414	0.380	0.366	0.354	0.374	0.381	0.363	0.381	0.389	0.382	0.396	0.405	0.411	0.418	0.420	0.432	0.430	0.432
.04	0.402	0.367	0.352	0.342	0.362	0.368	0.351	0.368	0.375	0.369	0.384	0.391	0.399	0.405	0.405	0.419	0.415	0.417
.05	0.389	0.354	0.339	0.331	0.350	0.355	0.340	0.356	0.362	0.358	0.370	0.378	0.387	0.392	0.390	0.405	0.401	0.402
.06	0.378	0.341	0.326	0.320	0.337	0.341	0.329	0.343	0.348	0.346	0.357	0.363	0.374	0.378	0.376	0.390	0.387	0.388
.07	0.366	0.328	0.312	0.309	0.325	0.328	0.318	0.331	0.335	0.334	0.344	0.348	0.361	0.365	0.363	0.377	0.373	0.373
.08	0.355	0.315	0.299	0.298	0.315	0.315	0.307	0.320	0.321	0.323	0.332	0.335	0.348	0.352	0.349	0.364	0.359	0.358
.09	0.344	0.302	0.286	0.286	0.303	0.302	0.296	0.308	0.307	0.312	0.319	0.321	0.336	0.339	0.333	0.351	0.345	0.344
.10	0.332	0.290	0.273	0.276	0.291	0.289	0.285	0.297	0.294	0.300	0.306	0.308	0.324	0.327	0.320	0.338	0.332	0.329
.20	0.227	0.181	0.161	0.179	0.185	0.177	0.184	0.188	0.179	0.196	0.194	0.187	0.210	0.206	0.197	0.219	0.211	0.205
.30	0.146	0.100	0.084	0.105	0.105	0.094	0.109	0.107	0.097	0.117	0.110	0.099	0.123	0.116	0.109	0.127	0.120	0.112
.40	0.086	0.050	0.037	0.055	0.053	0.045	0.059	0.055	0.045	0.064	0.055	0.048	0.066	0.059	0.052	0.068	0.061	0.056
.50	0.050	0.024	0.014	0.027	0.023	0.018	0.029	0.025	0.019	0.030	0.025	0.019	0.032	0.027	0.022	0.031	0.028	0.024

Table 8.11: The effect of increasing δ on the simulated estimates of Type II error probability for underlying $Exp(1)$ as the optimal sample size and k increase; $\alpha = 0.05, \beta_t = 0.05, m = 15.$

To illustrate the above discussion graphically we consider Figures 8.10 till 8.21. The below graphs are graphical representations that illustrate the above idea carefully as the optimal sample size increases. Figure 8.10 till 8.13 represent the effect of changing the design factor δ on the performance of the Type II error under the normal distribution while Figures from 8.14 till 8.17 illustrate the same idea under the uniform distribution while Figures 8.18 till 8.21 represent the case under the exponential distribution. We noticed from Figures 8.11 till 8.15 how fast the convergence of the Type II error probabilities occur at $\delta = 0.5$ while from Figures 8.15 till 8.18 we have same pattern as the normal distribution while from Figures 8.10 till 8.21 where we can realize the slow convergence of the Type II error probabilities towards the case $\delta = 0.5$ in comparative with the normal and uniform.

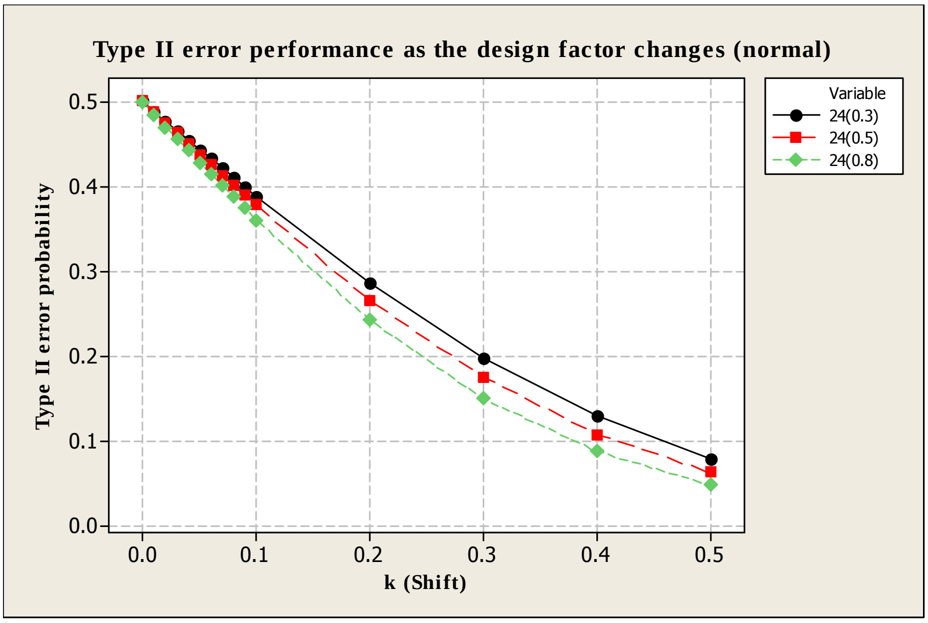

Figure 8.10: The effect of increasing δ on the estimated Type II error probability for underlying $N(0,1)$ as k increases, $n^* = 24$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

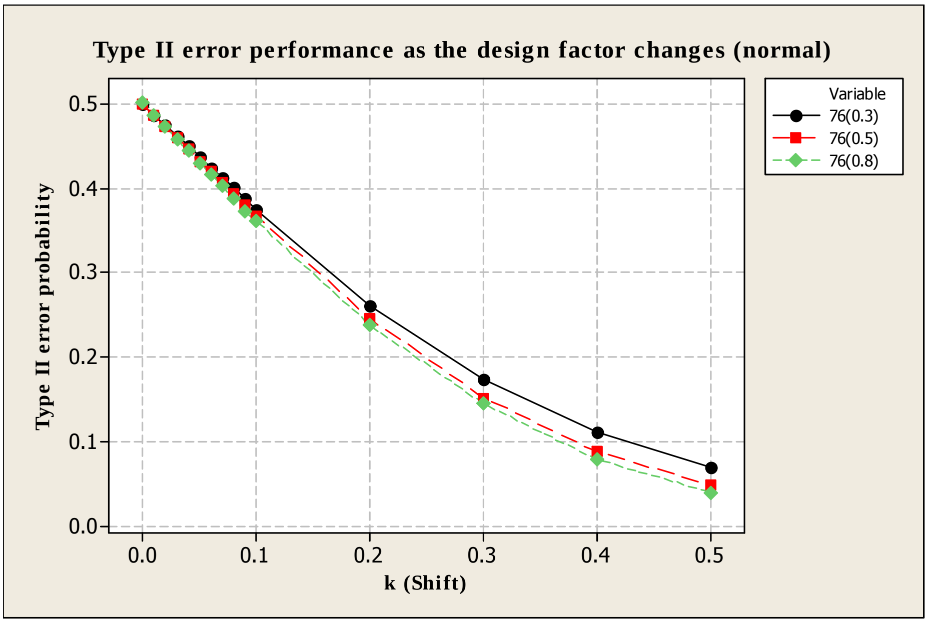

Figure 8.11: The effect of increasing δ on the estimated Type II error probability for underlying $N(0,1)$ as k increases at $n^* = 76$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

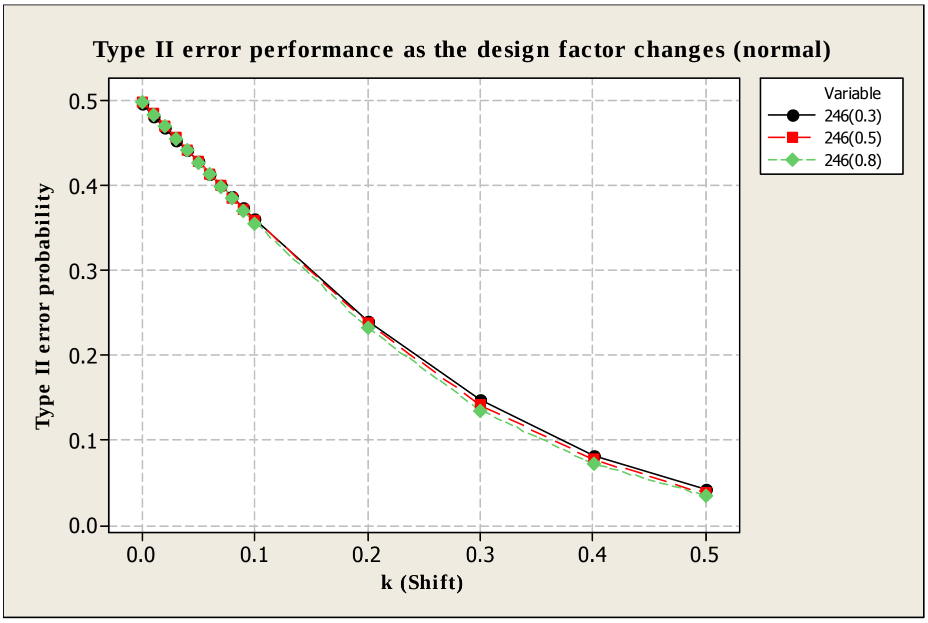

Figure 8.12: The effect of increasing δ on the estimated Type II error probability for underlying $N(0,1)$ as k increases at $n^* = 246$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

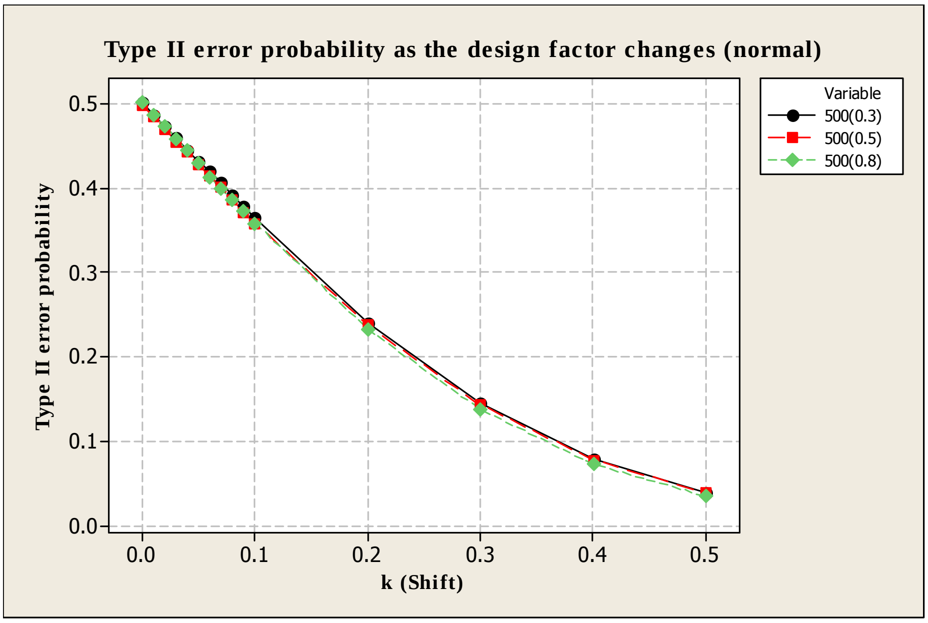

Figure 8.13: The effect of increasing δ on the estimated Type II error probability for underlying $N(0,1)$ as k increases at $n^* = 500$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

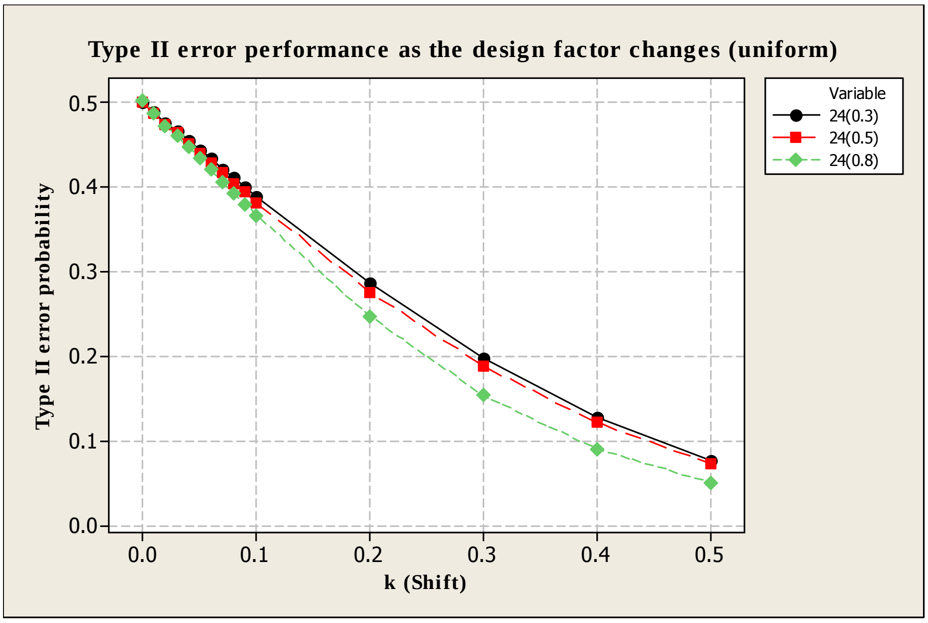

Figure 8.14: The effect of increasing δ on the estimated Type II error probability for underlying $U(0,1)$ as k increases at $n^* = 24$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

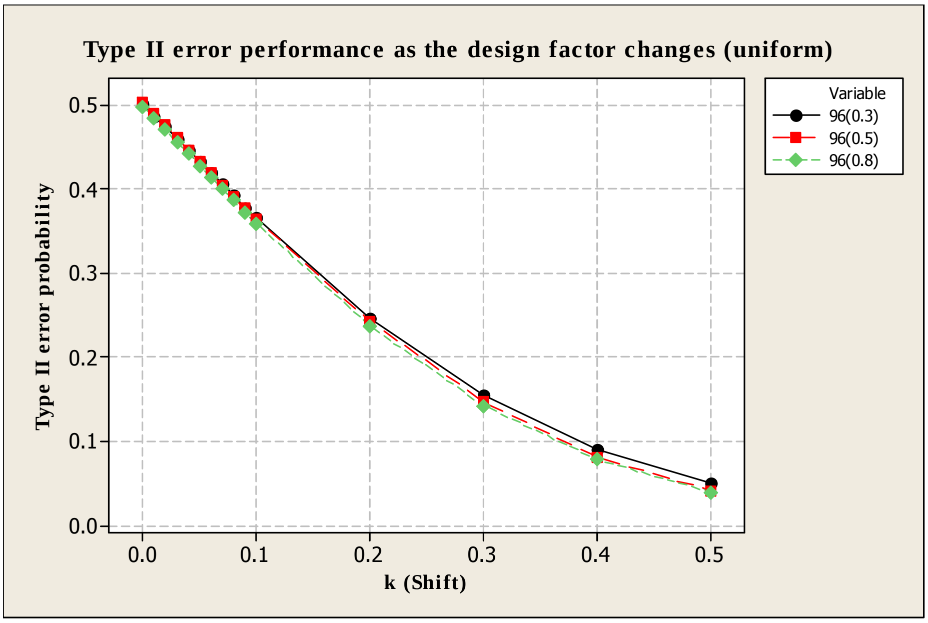

Figure 8.15: The effect of increasing δ on the estimated Type II error probability for underlying $U(0,1)$ as k increases at $n^* = 96$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

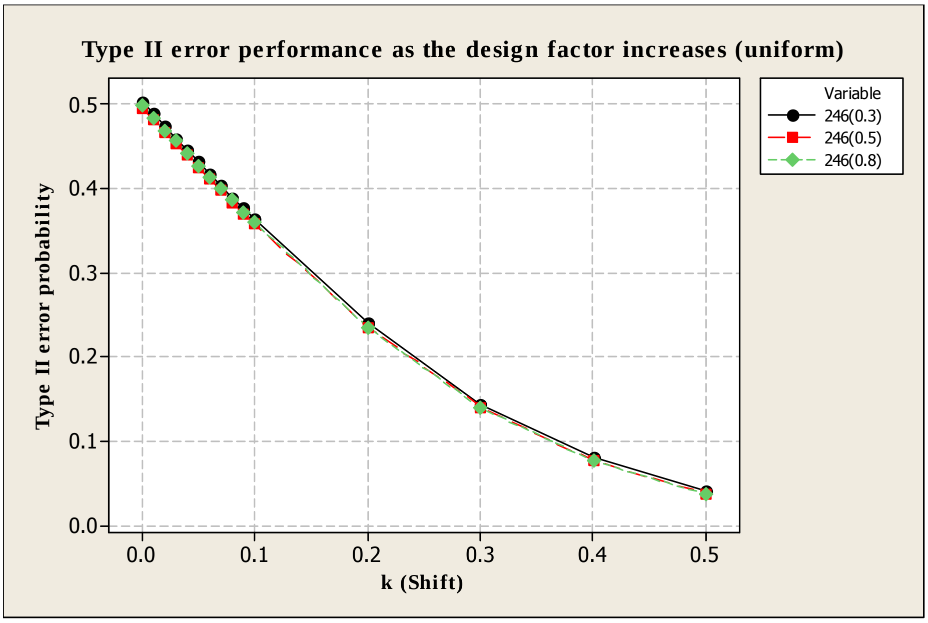

Figure 8.16: The effect of increasing δ on the estimated Type II error probability for underlying $U(0,1)$ as k increases at $n^* = 246$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

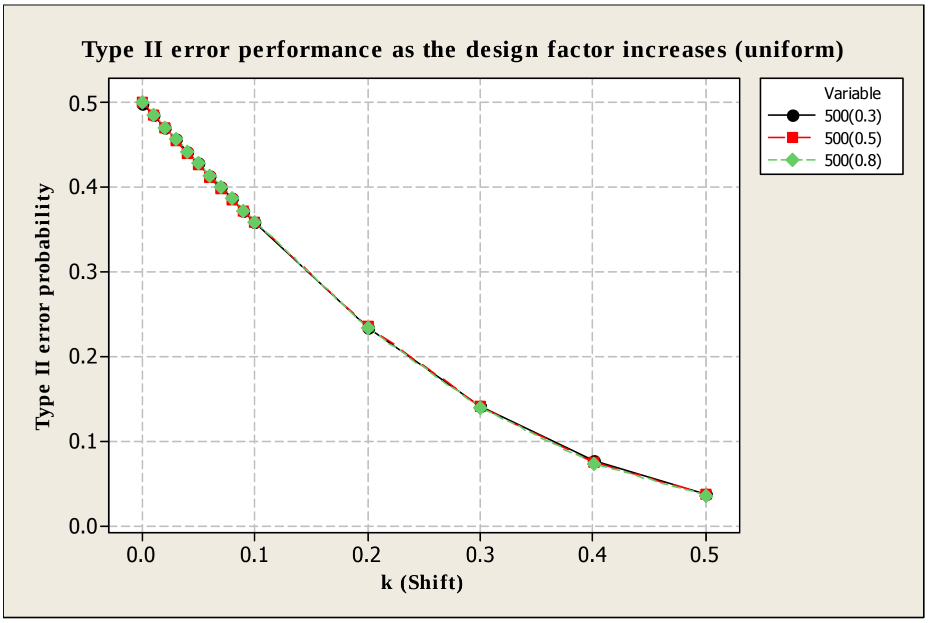

Figure 8.17: The effect of increasing δ on the estimated Type II error probability for underlying $U(0,1)$ as k increases at $n^* = 500$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

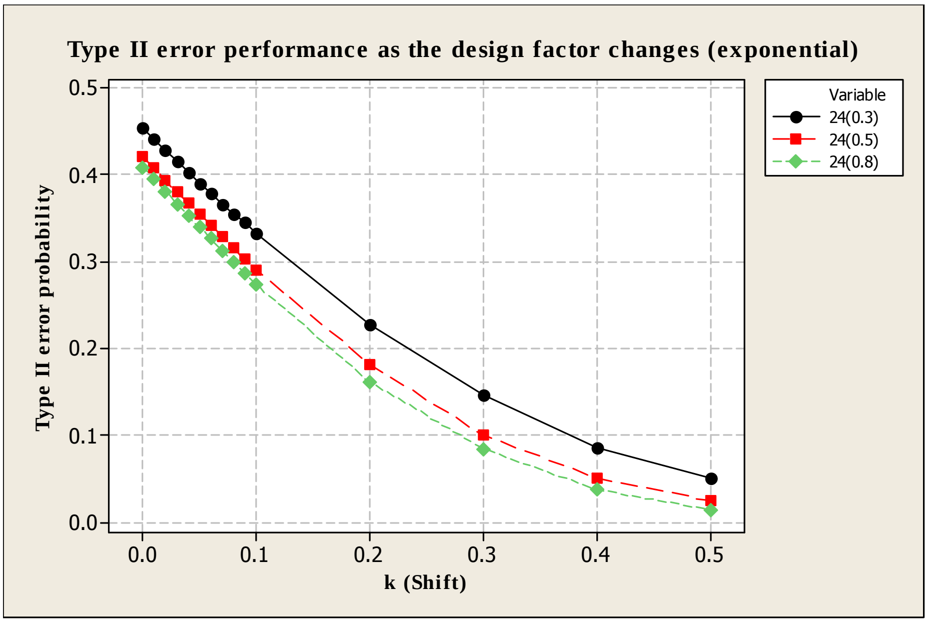

Figure 8.18: The effect of increasing δ on the estimated Type II error probability for underlying $Exp(1)$ as k increases at $n^* = 24$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

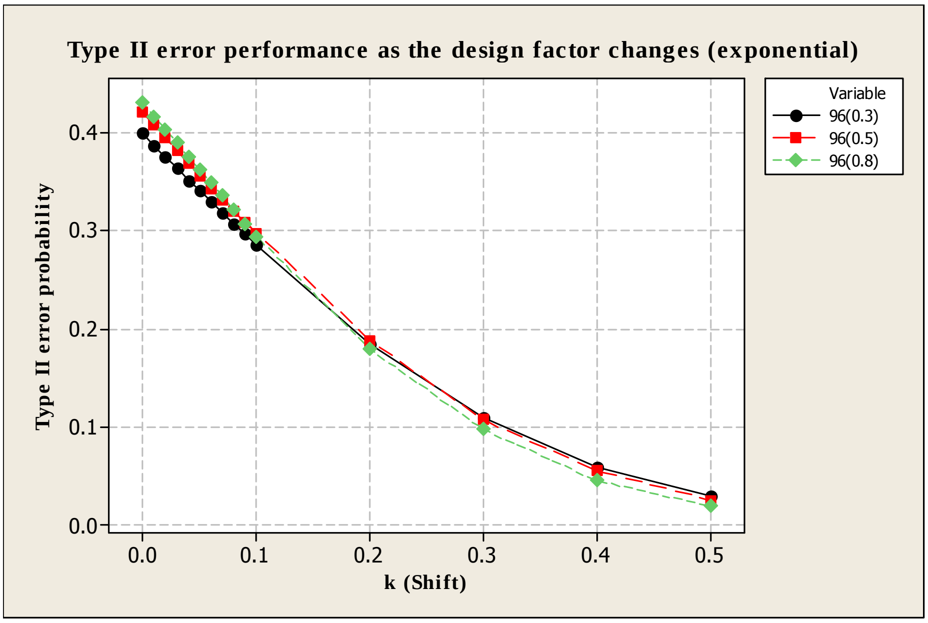

Figure 8.19: The effect of increasing δ on the estimated Type II error probability for underlying $Exp(1)$ as k increases at $n^* = 96$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

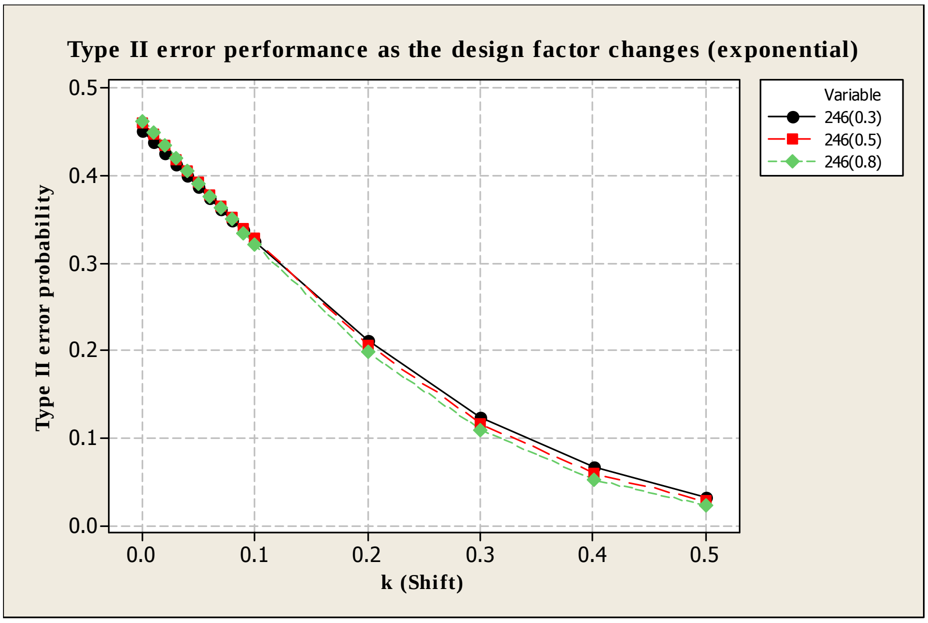

Figure 8.20: The effect of increasing δ on the estimated Type II error probability for underlying $Exp(1)$ as k increases at $n^* = 246$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

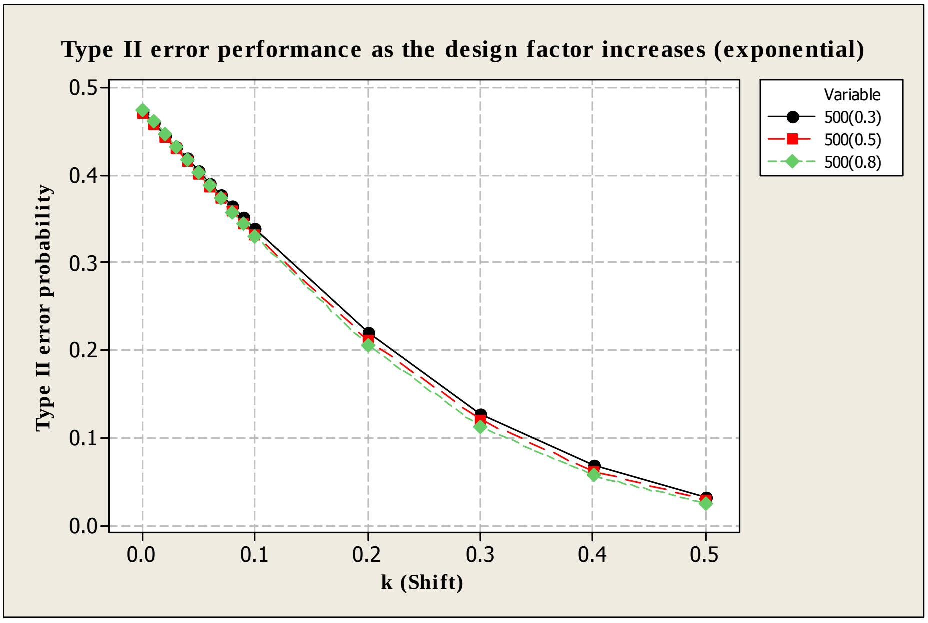

Figure 8.21: The effect of increasing δ on the estimated Type II error probability for underlying $Exp(1)$ as k increases at $n^* = 500$; $\alpha = 0.05, \beta_t = 0.05, m = 15$.

To conclude the thesis, we now summarize our .

We have seen that the results of the sequential triple sampling procedure depend mainly on the characteristics of the underlying distribution. In particular, the skewness and the kurtosis play a major role in determining the properties of the procedure.

In point estimation the asymptotic regret is bounded by non vanishing quantities and depends basically on the kurtosis. On the other hand the asymptotic coverage probability of the fixed width confidence interval depends on both the skewness and the kurtosis of the underlying distribution. Moreover, it is common with Hall's triple sampling for fixed width confidence interval estimation that we attain the nominal value asymptotically. Hall (1981) recommended a modified sample size N^* by increasing the sample size N in (4.3) to $N^* = N + \left[\left(a^2 - \delta + 5 \right) / 2\delta \right]$ to modify the coverage up to the nominal value. We have found that controlling Type II errors during the course of estimation while building the confidence interval provides coverage with at least the nominal value. Therefore, we may say that controlling Type II error will act in two ways. First, improving the coverage and second, signifying any shifts in the targeted mean.

In addition the design factor δ was recommended by many sequential scientists to be 0.5 in practical situations. Intensive simulation results supported this fact from the prospective of controlling Type II error. Moreover, the rounded off random error ε_{N_1}, was found by intensive simulations, that their asymptotic distribution is uniformly distributed over that interval $(0,1)$.

Appendix A

Triple Sampling Simulation to Estimate the Optimal Sample Size, the Population Mean and the Regret at $m = 5, 15, 20$ and $\delta = 0.5$

Note: Each table is divided into three sub attached tables, the first one at $m = 5$, the second at $m = 15$ and the last at $m = 20$.

TABLE A1

The underlying distribution is a *standard normal*

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	22.18	0.028982	0.00115	0.001035	5.0388
43	41.25	0.036702	0.00084	0.000726	4.0344
61	59.37	0.042849	0.00068	0.000592	2.5455
76	74.44	0.046901	0.00031	0.000525	2.0155
96	94.51	0.052273	0.00003	0.000462	0.7541
125	123.59	0.058102	0.00033	0.000405	1.9341
171	169.56	0.067547	-0.00027	0.000346	2.6415
246	244.62	0.080568	0.00021	0.000287	2.2415
500	498.70	0.113235	-0.00005	0.000201	3.6976
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	16.36	0.020731	0.00082	0.001131	5.2289
43	41.11	0.039497	0.00069	0.000725	3.6876
61	59.78	0.038194	-0.00047	0.000584	1.1441
76	74.90	0.041723	-0.00083	0.000520	1.0326
96	94.88	0.046620	-0.00043	0.000462	1.1022
125	123.90	0.052493	0.00011	0.000403	0.9575
171	169.83	0.061344	0.00052	0.000345	2.1620
246	244.95	0.073067	0.00035	0.000286	0.7743
500	498.82	0.103711	-0.00020	0.000200	-0.0303
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.00	0.000000	0.00176	0.000999	0.7149
43	36.39	0.058475	-0.00095	0.000805	10.3735
61	59.52	0.041586	-0.00026	0.000590	2.3210
76	74.83	0.041595	0.00021	0.000524	2.2575
96	94.88	0.046026	-0.00031	0.000463	1.6784
125	123.92	0.052166	-0.00003	0.000404	1.5963
171	169.87	0.060357	0.00013	0.000345	2.0388
246	244.92	0.072686	-0.00007	0.000287	2.5890
500	498.92	0.102876	0.00009	0.000200	0.3383

TABLE A2

The underlying distribution is a *standard uniform*

n^*	$\overline{N}$	$s.e.\overline{N}$	$\widehat{\mu}$	$s.e.\widehat{\mu}$	$\widehat{\omega}$
24	23.05	0.022279	0.49974	0.000324	11.3524
43	42.35	0.025634	0.50012	0.000213	6.6890
61	60.52	0.028055	0.50008	0.000173	4.9614
76	75.52	0.030640	0.50008	0.000153	4.6738
96	95.60	0.032932	0.50003	0.000134	2.4396
125	124.62	0.036760	0.49990	0.000117	2.9169
171	170.65	0.042357	0.50000	0.000099	0.7481
246	245.64	0.049462	0.49995	0.000083	2.3900
500	499.69	0.069074	0.49991	0.000058	4.1204
n^*	$\overline{N}$	$s.e.\overline{N}$	$\widehat{\mu}$	$s.e.\widehat{\mu}$	$\widehat{\omega}$
24	15.24	0.008750	0.50030	0.000333	5.4707
43	42.52	0.024801	0.49999	0.000210	5.2190
61	60.77	0.024570	0.49988	0.000170	2.9704
76	75.82	0.026718	0.50026	0.000150	2.1537
96	95.76	0.029713	0.50006	0.000133	1.7100
125	124.86	0.033402	0.50015	0.000116	1.2865
171	170.86	0.038470	0.50004	0.000099	1.4656
246	245.78	0.045876	0.49988	0.000083	5.9230
500	499.86	0.065035	0.50000	0.000058	7.6406
n^*	$\overline{N}$	$s.e.\overline{N}$	$\widehat{\mu}$	$s.e.\widehat{\mu}$	$\widehat{\omega}$
24	20.00	0.000000	0.49955	0.000287	0.5572
43	38.15	0.050794	0.49937	0.000239	15.6292
61	60.73	0.025025	0.50011	0.000169	2.5278
76	75.77	0.026631	0.50004	0.000151	2.3136
96	95.80	0.029368	0.50016	0.000134	2.6052
125	124.80	0.033152	0.49987	0.000116	0.9192
171	170.82	0.038557	0.50012	0.000099	2.1333
246	245.83	0.045493	0.49998	0.000083	3.8632
500	499.89	0.064344	0.50007	0.000058	1.4336

TABLE A3

The underlying distribution is $t(5)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.88	0.036654	-0.00092	0.001309	2.5043
43	39.66	0.052534	-0.00190	0.000934	2.0134
61	57.66	0.066577	0.00081	0.000769	1.6291
76	72.68	0.075315	0.00004	0.000684	1.8168
96	92.72	0.084027	-0.00083	0.000603	1.4098
125	121.86	0.099683	-0.00061	0.000524	0.7889
171	167.80	0.120742	-0.00061	0.000445	-0.1928
246	243.19	0.149062	0.00019	0.000370	0.2370
500	498.39	0.242648	-0.00047	0.000258	-1.7375
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	17.33	0.028474	-0.00090	0.001377	2.1192
43	38.65	0.054649	0.00010	0.000940	1.6830
61	57.65	0.057029	-0.00133	0.000764	0.7942
76	72.70	0.064292	-0.00003	0.000679	0.5190
96	92.81	0.074492	-0.00024	0.000599	0.0544
125	121.68	0.087872	-0.00071	0.000524	0.3386
171	167.73	0.104197	-0.00036	0.000444	-1.4392
246	242.64	0.128265	0.00026	0.000369	-1.8762
500	496.85	0.193698	-0.00029	0.000260	4.4777
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.32	0.012561	0.00125	0.001263	-0.1202
43	34.20	0.065299	-0.00068	0.001017	5.5412
61	57.04	0.063225	-0.00018	0.000772	1.4883
76	72.73	0.063889	-0.00042	0.000677	0.2072
96	92.69	0.072646	-0.00045	0.000600	0.1309
125	121.75	0.083565	-0.00037	0.000522	-0.4909
171	167.60	0.104864	0.00020	0.000445	-0.4245
246	242.50	0.126861	-0.00052	0.000369	-1.9978
500	496.47	0.190696	-0.00005	0.000259	-0.7673

TABLE A4

The underlying distribution is $t(25)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	22.02	0.029939	-0.00116	0.001076	4.6844
43	41.09	0.038530	0.00055	0.000756	3.6912
61	59.18	0.045058	0.00072	0.000619	2.7728
76	74.27	0.050177	0.00046	0.000547	1.7548
96	94.26	0.055544	-0.00084	0.000486	2.2458
125	123.27	0.062155	-0.00074	0.000423	1.9584
171	169.49	0.071683	0.00009	0.000358	0.2642
246	244.54	0.085695	-0.00076	0.000298	-0.9581
500	498.77	0.122293	-0.00011	0.000209	-0.0232
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	16.53	0.021943	-0.00101	0.001174	5.0694
43	40.94	0.041395	-0.00037	0.000759	3.9828
61	59.60	0.040470	-0.00012	0.000610	1.3662
76	74.65	0.044291	-0.00076	0.000543	1.0648
96	94.66	0.049549	0.00033	0.000484	2.1484
125	123.67	0.055786	0.00059	0.000422	1.4517
171	169.72	0.064653	-0.00007	0.000358	0.0084
246	244.71	0.078574	-0.00043	0.000299	1.4579
500	498.63	0.110583	-0.00018	0.000209	1.5594
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.00	0.001335	0.00025	0.001044	0.9021
43	36.14	0.059555	0.00003	0.000832	9.0250
61	59.34	0.044015	-0.00045	0.000617	2.5277
76	74.61	0.044410	-0.00042	0.000545	1.4274
96	94.70	0.048752	-0.00010	0.000481	0.8447
125	123.67	0.055646	-0.00039	0.000422	1.3883
171	169.73	0.064318	0.00056	0.000360	2.1650
246	244.72	0.077640	-0.00041	0.000298	-0.2633
500	498.63	0.109736	-0.00043	0.000210	6.2678

TABLE A5

The underlying distribution is $t(50)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	22.14	0.029292	0.00062	0.001054	4.8482
43	41.20	0.037994	-0.00070	0.000741	3.9371
61	59.31	0.043884	-0.00017	0.000606	2.9058
76	74.40	0.047923	-0.00024	0.000540	3.2878
96	94.39	0.053054	-0.00114	0.000474	1.8918
125	123.56	0.059901	0.00000	0.000414	2.0669
171	169.43	0.070151	-0.00064	0.000352	0.9446
246	244.47	0.083238	0.00004	0.000293	1.2752
500	498.62	0.117445	-0.00070	0.000205	4.7263
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	16.48	0.021578	-0.00022	0.001147	4.8474
43	41.05	0.040225	0.00029	0.000740	3.6657
61	59.64	0.039283	0.00004	0.000596	1.0596
76	74.70	0.042995	-0.00070	0.000531	0.7468
96	94.76	0.047965	0.00052	0.000473	1.6086
125	123.69	0.054177	-0.00080	0.000414	2.1794
171	169.79	0.063129	-0.00043	0.000351	0.9590
246	244.84	0.075478	-0.00014	0.000294	4.1042
500	498.89	0.106259	-0.00027	0.000205	1.0304
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.00	0.001174	0.00010	0.001021	0.8401
43	36.31	0.058764	-0.00024	0.000814	9.0564
61	59.46	0.042114	-0.00044	0.000604	2.6648
76	74.74	0.042853	-0.00116	0.000532	1.2739
96	94.88	0.047234	-0.00113	0.000470	0.5378
125	123.83	0.053609	0.00026	0.000411	0.3645
171	169.84	0.062453	-0.00041	0.000352	1.2998
246	244.68	0.074626	0.00011	0.000294	3.1244
500	499.01	0.105524	-0.00018	0.000205	5.6271

TABLE A6

The underlying distribution is $t(100)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	22.11	0.029307	-0.00054	0.001051	5.2683
43	41.29	0.037196	-0.00057	0.000728	3.3283
61	59.29	0.043189	-0.00041	0.000599	2.8020
76	74.44	0.047089	-0.00068	0.000532	2.4863
96	94.46	0.052633	-0.00054	0.000469	2.0045
125	123.50	0.059256	-0.00032	0.000412	3.4297
171	169.50	0.068577	0.00004	0.000349	2.0338
246	244.58	0.082121	-0.00013	0.000289	0.9453
500	498.77	0.116592	-0.00028	0.000202	-3.2164
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	16.41	0.021003	-0.00023	0.001142	5.2045
43	41.09	0.039855	0.00043	0.000734	3.9395
61	59.81	0.038443	-0.00002	0.000593	1.9960
76	74.73	0.042378	-0.00002	0.000525	0.8356
96	94.92	0.046767	-0.00060	0.000465	0.4081
125	123.90	0.053644	-0.00048	0.000408	1.1102
171	169.80	0.061847	-0.00007	0.000346	-1.0170
246	244.81	0.074448	-0.00044	0.000290	1.8997
500	498.81	0.105261	-0.00030	0.000204	6.3207
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.00	0.000000	0.00164	0.001009	0.7379
43	36.32	0.058804	0.00128	0.000813	10.1842
61	59.55	0.041776	0.00058	0.000599	2.9800
76	74.82	0.041891	0.00046	0.000526	1.0016
96	94.79	0.046664	-0.00064	0.000466	1.0195
125	123.86	0.052835	-0.00029	0.000409	1.7239
171	169.71	0.061537	-0.00085	0.000347	-0.0869
246	244.87	0.073605	-0.00004	0.000288	-0.4451
500	498.92	0.104293	-0.00039	0.000202	-3.5525

TABLE A7

The underlying distribution is *beta* $(2,3)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	22.63	0.026033	0.39700	0.000214	7.8013
43	41.78	0.032047	0.39887	0.000146	5.2428
61	59.94	0.036216	0.39936	0.000119	4.0645
76	75.04	0.038951	0.39976	0.000106	3.8069
96	95.06	0.042749	0.39957	0.000093	3.5709
125	124.09	0.047940	0.39991	0.000081	3.6107
171	170.06	0.054984	0.39987	0.000069	2.8993
246	245.20	0.065358	0.39974	0.000057	2.4580
500	499.12	0.091797	0.39991	0.000040	1.1577
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	15.78	0.015659	0.39929	0.000228	5.1932
43	41.80	0.032927	0.39862	0.000147	5.6713
61	60.28	0.031924	0.39927	0.000118	2.5505
76	75.28	0.035098	0.39957	0.000104	1.8228
96	95.29	0.038554	0.39964	0.000092	1.3700
125	124.36	0.043387	0.39967	0.000081	1.2696
171	170.32	0.050605	0.39981	0.000069	2.1447
246	245.30	0.060230	0.39996	0.000057	0.2243
500	499.41	0.084919	0.39996	0.000040	-0.3856
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.00	0.000000	0.40022	0.000200	0.6887
43	37.19	0.055311	0.39762	0.000164	13.3263
61	60.15	0.033871	0.39925	0.000118	3.2058
76	75.37	0.034510	0.39942	0.000105	3.0726
96	95.34	0.038387	0.39953	0.000093	2.8478
125	124.41	0.043471	0.39988	0.000081	1.4720
171	170.45	0.050123	0.39994	0.000069	1.3234
246	245.41	0.059888	0.39986	0.000058	4.0616
500	499.36	0.084370	0.39994	0.000040	3.6948

TABLE A8

The underlying distribution is *exponential* with mean one

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	19.05	0.042283	0.91062	0.001210	17.7940
43	37.36	0.066798	0.95096	0.000910	32.3492
61	55.27	0.083120	0.97055	0.000717	32.0484
76	70.43	0.095247	0.97799	0.000620	32.2888
96	90.82	0.108801	0.98547	0.000522	26.3778
125	119.63	0.125698	0.98934	0.000442	23.7438
171	166.33	0.150889	0.99284	0.000361	16.7938
246	241.72	0.183681	0.99513	0.000293	10.2735
500	497.95	0.278949	0.99761	0.000203	13.6584
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	18.00	0.031791	0.96999	0.000981	-1.7872
43	36.70	0.063877	0.95095	0.000820	17.3048
61	55.42	0.072529	0.97141	0.000680	22.4790
76	70.67	0.081002	0.98051	0.000582	18.6754
96	90.71	0.091417	0.98597	0.000496	14.0712
125	119.59	0.104466	0.99106	0.000425	11.8602
171	165.67	0.123608	0.99283	0.000354	8.7925
246	241.02	0.150052	0.99610	0.000293	9.2764
500	495.23	0.217850	0.99772	0.000202	7.3364
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$\hat{\omega}$
24	20.46	0.014478	0.99486	0.000965	-0.7321
43	33.97	0.068079	0.95662	0.000772	6.4769
61	54.51	0.078470	0.96587	0.000690	25.3587
76	70.40	0.081271	0.97832	0.000594	23.0115
96	90.50	0.089091	0.98632	0.000501	15.7395
125	119.70	0.102628	0.99068	0.000422	10.3143
171	165.69	0.120965	0.99342	0.000355	8.7394
246	240.62	0.146291	0.99518	0.000293	9.0376
500	495.16	0.211471	0.99795	0.000202	4.0084

Appendix B

Triple Sampling Simulation to Estimate the Optimal Sample Size, the Population Mean, the Probability of Stopping and the Coverage Probability at $m = 5, 15, 20$, $\delta = 0.5$ and $\alpha = 0.05$

Note: Each table is divided into three sub attached tables, the first one at $m = 5$, the second at $m = 15$ and the last at $m = 20$.

TABLE B1

The underlying distribution is a *standard normal*

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.03	0.051	-0.0011	0.0013	0.202	0.076	0.722	0.8651
43	37.92	0.081	0.0007	0.0009	0.079	0.091	0.830	0.8831
61	55.72	0.104	0.0003	0.0008	0.044	0.086	0.870	0.8985
76	70.87	0.120	-0.0001	0.0007	0.031	0.092	0.877	0.9063
96	91.21	0.139	0.0000	0.0006	0.018	0.091	0.891	0.9179
125	120.52	0.164	0.0004	0.0005	0.012	0.092	0.897	0.9254
171	167.48	0.200	-0.0008	0.0004	0.007	0.091	0.902	0.9315
246	244.49	0.249	0.0002	0.0003	0.003	0.089	0.908	0.9394
500	505.39	0.405	-0.0001	0.0002	0.001	0.089	0.910	0.9462

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.56	0.038	0.0011	0.0011	0.769	0.000	0.231	0.9009
43	38.12	0.068	-0.0006	0.0008	0.221	0.001	0.779	0.9048
61	56.32	0.081	0.0004	0.0007	0.062	0.004	0.935	0.9192
76	71.55	0.089	0.0002	0.0006	0.022	0.005	0.973	0.9276
96	91.48	0.101	0.0004	0.0005	0.008	0.008	0.985	0.9331
125	120.70	0.113	-0.0003	0.0004	0.002	0.009	0.989	0.9383
171	166.85	0.131	0.0006	0.0004	0.000	0.010	0.990	0.9433
246	242.08	0.156	0.0000	0.0003	0.000	0.010	0.990	0.9435
500	496.69	0.223	0.0001	0.0002	0.000	0.013	0.987	0.9463

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.77	0.019	-0.0016	0.0010	0.966	0.000	0.034	0.9297
43	36.25	0.071	-0.0017	0.0008	0.452	0.000	0.548	0.8974
61	55.85	0.084	-0.0001	0.0007	0.133	0.000	0.867	0.9161
76	71.55	0.089	-0.0006	0.0006	0.046	0.001	0.953	0.9269
96	91.71	0.098	0.0001	0.0005	0.012	0.002	0.987	0.9335
125	120.72	0.111	0.0000	0.0004	0.002	0.002	0.996	0.9398
171	166.88	0.127	0.0006	0.0004	0.000	0.003	0.997	0.9411
246	241.63	0.151	0.0001	0.0003	0.000	0.004	0.996	0.9457
500	496.14	0.214	-0.0002	0.0002	0.000	0.005	0.995	0.9487

TABLE B2

The underlying distribution is a *standard uniform*

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.92	0.04	0.5	0.0	0.136	0.022	0.842	0.8487
43	39.27	0.06	0.5	0.0	0.047	0.028	0.925	0.8854
61	57.12	0.07	0.5	0.0	0.026	0.029	0.945	0.9040
76	72.20	0.08	0.5	0.0	0.016	0.030	0.953	0.9151
96	92.15	0.09	0.5	0.0	0.011	0.031	0.958	0.9220
125	121.37	0.10	0.5	0.0	0.007	0.032	0.961	0.9319
171	167.75	0.12	0.5	0.0	0.004	0.032	0.965	0.9385
246	243.38	0.14	0.5	0.0	0.001	0.031	0.968	0.9447
500	498.77	0.19	0.5	0.0	0.001	0.032	0.967	0.9481
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	17.89	0.03	0.5	0.0	0.839	0.000	0.161	0.8854
43	40.47	0.05	0.5	0.0	0.112	0.000	0.887	0.9062
61	59.07	0.05	0.5	0.0	0.016	0.000	0.984	0.9301
76	74.21	0.06	0.5	0.0	0.005	0.000	0.995	0.9341
96	94.27	0.06	0.5	0.0	0.001	0.000	0.999	0.9387
125	123.36	0.07	0.5	0.0	0.000	0.000	1.000	0.9432
171	169.32	0.08	0.5	0.0	0.000	0.000	1.000	0.9449
246	244.34	0.10	0.5	0.0	0.000	0.000	1.000	0.9470
500	498.42	0.13	0.5	0.0	0.000	0.000	1.000	0.9483
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.03	0.00	0.5	0.0	0.999	0.000	0.001	0.9258
43	37.11	0.06	0.5	0.0	0.381	0.000	0.619	0.8842
61	58.82	0.06	0.5	0.0	0.050	0.000	0.950	0.9230
76	74.25	0.06	0.5	0.0	0.010	0.000	0.990	0.9329
96	94.45	0.06	0.5	0.0	0.001	0.000	0.999	0.9412
125	123.45	0.07	0.5	0.0	0.000	0.000	1.000	0.9447
171	169.47	0.08	0.5	0.0	0.000	0.000	1.000	0.9430
246	244.49	0.09	0.5	0.0	0.000	0.000	1.000	0.9467
500	498.77	0.13	0.5	0.0	0.000	0.000	1.000	0.9470

TABLE B3

The underlying distribution is $t(5)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.03	0.08	0	0	0.288	0.117	0.595	0.8741
43	36.68	0.21	0	0	0.130	0.128	0.742	0.8773
61	55.27	0.23	0	0	0.075	0.125	0.800	0.8848
76	70.00	0.24	0	0	0.053	0.123	0.824	0.8927
96	90.62	0.28	0	0	0.035	0.118	0.847	0.8993
125	121.49	0.37	0	0	0.023	0.115	0.863	0.9074
171	171.01	0.44	0	0	0.012	0.115	0.873	0.9216
246	250.70	0.61	0	0	0.006	0.111	0.883	0.9322
500	521.82	1.19	0	0	0.002	0.112	0.887	0.9417

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.06	0.05	0	0	0.770	0.007	0.222	0.9232
43	35.45	0.10	0	0	0.348	0.021	0.631	0.9058
61	52.88	0.13	0	0	0.144	0.034	0.821	0.9129
76	67.92	0.16	0	0	0.070	0.040	0.890	0.9181
96	88.02	0.19	0	0	0.027	0.047	0.925	0.9230
125	117.76	0.24	0	0	0.010	0.050	0.941	0.9304
171	164.97	0.29	0	0	0.002	0.053	0.945	0.9353
246	241.67	0.37	0	0	0.000	0.055	0.944	0.9411
500	503.30	0.82	0	0	0.000	0.055	0.945	0.9462

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	22.35	0.04	0	0	0.914	0.003	0.084	0.9417
43	34.76	0.09	0	0	0.553	0.009	0.438	0.9088
61	52.10	0.13	0	0	0.256	0.016	0.728	0.9138
76	66.85	0.14	0	0	0.126	0.022	0.851	0.9191
96	87.19	0.17	0	0	0.050	0.026	0.923	0.9259
125	116.76	0.21	0	0	0.014	0.033	0.954	0.9333
171	163.54	0.26	0	0	0.002	0.037	0.961	0.9354
246	239.37	0.33	0	0	0.000	0.040	0.960	0.9395
500	498.27	0.54	0	0	0.000	0.044	0.956	0.9479

TABLE B4

The underlying distribution is $t(10)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.73	0.06	0	0	0.239	0.089	0.671	0.8709
43	37.28	0.10	0	0	0.096	0.103	0.801	0.8802
61	55.23	0.13	0	0	0.053	0.103	0.844	0.8936
76	70.20	0.15	0	0	0.038	0.105	0.857	0.9004
96	91.01	0.17	0	0	0.025	0.101	0.874	0.9104
125	120.92	0.21	0	0	0.016	0.104	0.880	0.9199
171	168.47	0.26	0	0	0.009	0.103	0.889	0.9282
246	246.80	0.34	0	0	0.004	0.105	0.890	0.9363
500	512.31	0.59	0	0	0.001	0.104	0.895	0.9450
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.88	0.04	0	0	0.764	0.001	0.235	0.9120
43	37.13	0.08	0	0	0.271	0.006	0.723	0.9040
61	54.98	0.09	0	0	0.090	0.013	0.897	0.9180
76	69.79	0.11	0	0	0.037	0.017	0.947	0.9233
96	89.98	0.12	0	0	0.013	0.020	0.967	0.9283
125	119.38	0.14	0	0	0.003	0.023	0.974	0.9352
171	165.89	0.17	0	0	0.001	0.026	0.973	0.9399
246	241.44	0.20	0	0	0.000	0.029	0.971	0.9440
500	496.88	0.30	0	0	0.000	0.030	0.970	0.9469
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	21.43	0.03	0	0	0.941	0.000	0.059	0.9334
43	35.70	0.07	0	0	0.489	0.001	0.510	0.9020
61	54.35	0.09	0	0	0.180	0.003	0.817	0.9156
76	69.80	0.10	0	0	0.072	0.006	0.923	0.9230
96	89.83	0.12	0	0	0.024	0.008	0.968	0.9298
125	119.01	0.13	0	0	0.005	0.011	0.984	0.9377
171	165.33	0.16	0	0	0.000	0.013	0.987	0.9405
246	240.64	0.19	0	0	0.000	0.015	0.985	0.9439
500	495.78	0.28	0	0	0.000	0.018	0.982	0.9472

TABLE B5

The underlying distribution is $t(25)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.85	0.05	0	0	0.215	0.081	0.704	0.8682
43	37.71	0.09	0	0	0.089	0.095	0.817	0.8828
61	55.35	0.11	0	0	0.047	0.096	0.857	0.8940
76	70.92	0.13	0	0	0.033	0.097	0.870	0.9070
96	91.12	0.15	0	0	0.020	0.099	0.881	0.9163
125	120.67	0.18	0	0	0.013	0.094	0.893	0.9236
171	168.38	0.22	0	0	0.007	0.096	0.897	0.9305
246	245.30	0.28	0	0	0.004	0.096	0.900	0.9364
500	508.23	0.46	0	0	0.001	0.095	0.904	0.9482
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.68	0.04	0	0	0.769	0.000	0.231	0.9042
43	37.70	0.07	0	0	0.237	0.002	0.761	0.9048
61	55.83	0.08	0	0	0.072	0.006	0.922	0.9192
76	70.95	0.09	0	0	0.027	0.008	0.965	0.9257
96	91.18	0.11	0	0	0.009	0.011	0.980	0.9327
125	119.97	0.12	0	0	0.002	0.013	0.985	0.9365
171	166.24	0.14	0	0	0.000	0.014	0.985	0.9422
246	241.76	0.17	0	0	0.000	0.017	0.983	0.9431
500	496.31	0.24	0	0	0.000	0.019	0.981	0.9474
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	21.01	0.02	0	0	0.957	0.000	0.043	0.9309
43	36.19	0.07	0	0	0.468	0.000	0.531	0.8979
61	55.32	0.09	0	0	0.150	0.001	0.849	0.9160
76	70.89	0.09	0	0	0.055	0.002	0.944	0.9265
96	90.96	0.10	0	0	0.016	0.003	0.981	0.9332
125	120.32	0.12	0	0	0.003	0.004	0.993	0.9375
171	166.48	0.14	0	0	0.000	0.006	0.994	0.9407
246	241.10	0.16	0	0	0.000	0.007	0.993	0.9441
500	495.96	0.23	0	0	0.000	0.009	0.991	0.9480

TABLE B6

The underlying distribution is $t(50)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.89	0.05	0	0	0.207	0.078	0.715	0.8697
43	37.70	0.08	0	0	0.085	0.091	0.824	0.8806
61	55.58	0.11	0	0	0.047	0.093	0.861	0.8983
76	70.73	0.12	0	0	0.030	0.091	0.879	0.9080
96	90.92	0.15	0	0	0.020	0.095	0.885	0.9148
125	120.61	0.17	0	0	0.012	0.091	0.897	0.9222
171	168.09	0.21	0	0	0.007	0.091	0.902	0.9328
246	245.03	0.26	0	0	0.003	0.093	0.904	0.9392
500	506.02	0.42	0	0	0.001	0.096	0.903	0.9467

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.65	0.04	0	0	0.769	0.000	0.231	0.9043
43	37.95	0.07	0	0	0.226	0.002	0.772	0.9049
61	56.23	0.08	0	0	0.063	0.005	0.932	0.9209
76	71.25	0.09	0	0	0.024	0.006	0.970	0.9289
96	91.27	0.10	0	0	0.008	0.008	0.984	0.9316
125	120.37	0.12	0	0	0.002	0.011	0.987	0.9385
171	166.45	0.14	0	0	0.000	0.012	0.988	0.9417
246	241.69	0.16	0	0	0.000	0.014	0.986	0.9460
500	496.56	0.23	0	0	0.000	0.016	0.984	0.9469

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.86	0.02	0	0	0.963	0.000	0.037	0.9294
43	36.16	0.07	0	0	0.463	0.000	0.537	0.8974
61	55.64	0.09	0	0	0.142	0.000	0.858	0.9156
76	71.24	0.09	0	0	0.050	0.001	0.949	0.9276
96	91.26	0.10	0	0	0.014	0.002	0.984	0.9347
125	120.43	0.11	0	0	0.002	0.003	0.995	0.9368
171	166.48	0.13	0	0	0.000	0.004	0.996	0.9429
246	241.59	0.16	0	0	0.000	0.005	0.995	0.9442
500	496.03	0.22	0	0	0.000	0.006	0.994	0.9476

TABLE B7

The underlying distribution is $t(100)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.93	0.05	0	0	0.204	0.076	0.721	0.8654
43	37.72	0.08	0	0	0.079	0.090	0.831	0.8834
61	55.67	0.11	0	0	0.044	0.091	0.865	0.8964
76	70.75	0.12	0	0	0.029	0.093	0.878	0.9060
96	90.75	0.14	0	0	0.019	0.092	0.889	0.9151
125	120.49	0.17	0	0	0.012	0.091	0.897	0.9240
171	168.12	0.21	0	0	0.006	0.091	0.903	0.9343
246	245.19	0.26	0	0	0.003	0.093	0.903	0.9407
500	506.01	0.42	0	0	0.001	0.093	0.907	0.9476
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	19.63	0.04	0	0	0.768	0.000	0.232	0.9030
43	38.15	0.07	0	0	0.224	0.001	0.775	0.9047
61	56.40	0.08	0	0	0.066	0.004	0.931	0.9184
76	71.39	0.09	0	0	0.024	0.007	0.969	0.9268
96	91.50	0.10	0	0	0.007	0.008	0.985	0.9322
125	120.47	0.11	0	0	0.002	0.010	0.988	0.9380
171	166.73	0.13	0	0	0.000	0.011	0.989	0.9420
246	241.93	0.16	0	0	0.000	0.012	0.987	0.9450
500	496.91	0.23	0	0	0.000	0.014	0.986	0.9493
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.86	0.02	0	0	0.965	0.000	0.035	0.9301
43	36.14	0.07	0	0	0.460	0.000	0.540	0.8974
61	55.71	0.08	0	0	0.137	0.000	0.863	0.9161
76	71.27	0.09	0	0	0.049	0.001	0.950	0.9244
96	91.45	0.10	0	0	0.013	0.001	0.986	0.9335
125	120.58	0.11	0	0	0.003	0.003	0.994	0.9382
171	166.72	0.13	0	0	0.000	0.004	0.996	0.9423
246	242.11	0.15	0	0	0.000	0.005	0.995	0.9446
500	496.06	0.22	0	0	0.000	0.005	0.995	0.9478

TABLE B8

The underlying distribution is *beta* $(3,2)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.24	0.05	0.61	0	0.173	0.050	0.777	0.8535
43	38.53	0.07	0.60	0	0.064	0.064	0.872	0.8823
61	56.28	0.09	0.60	0	0.033	0.067	0.900	0.8995
76	71.28	0.10	0.60	0	0.023	0.066	0.911	0.9090
96	91.52	0.11	0.60	0	0.014	0.067	0.919	0.9187
125	120.95	0.13	0.60	0	0.009	0.069	0.921	0.9288
171	167.32	0.16	0.60	0	0.004	0.069	0.926	0.9349
246	243.33	0.19	0.60	0	0.003	0.072	0.926	0.9409
500	501.04	0.28	0.60	0	0.001	0.072	0.927	0.9490
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	18.91	0.03	0.60	0	0.788	0.000	0.212	0.8962
43	39.20	0.06	0.60	0	0.172	0.000	0.828	0.9017
61	57.76	0.07	0.60	0	0.037	0.000	0.963	0.9228
76	72.68	0.08	0.60	0	0.012	0.001	0.987	0.9278
96	92.95	0.08	0.60	0	0.004	0.002	0.994	0.9353
125	121.96	0.09	0.60	0	0.001	0.001	0.998	0.9405
171	168.10	0.11	0.60	0	0.000	0.002	0.998	0.9415
246	243.14	0.13	0.60	0	0.000	0.002	0.998	0.9429
500	497.11	0.18	0.60	0	0.000	0.003	0.997	0.9490
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.28	0.01	0.60	0	0.986	0.000	0.014	0.9274
43	36.61	0.07	0.60	0	0422	0.000	0.578	0.8909
61	57.33	0.07	0.60	0	0.094	0.000	0.906	0.9167
76	72.69	0.08	0.60	0	0.026	0.000	0.974	0.9310
96	93.04	0.08	0.60	0	0.006	0.000	0.994	0.9341
125	122.16	0.09	0.60	0	0.001	0.000	0.999	0.9401
171	168.26	0.11	0.60	0	0.000	0.000	1.000	0.9435
246	243.34	0.12	0.60	0	0.000	0.000	1.000	0.9458
500	497.13	0.17	0.60	0	0.000	0.000	1.000	0.9477

TABLE B9

The underlying distribution is *exponential* with mean one

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	17.71	0.083	0.8701	0.0012	0.388	0.129	0.483	0.7935
43	33.82	0.151	0.8755	0.0011	0.227	0.153	0.620	0.7418
61	50.80	0.213	0.8915	0.0011	0.154	0.152	0.694	0.7544
76	65.89	0.258	0.9061	0.0010	0.117	0.152	0.731	0.7736
96	85.76	0.312	0.9188	0.0009	0.088	0.148	0.765	0.7932
125	116.15	0.399	0.9357	0.0008	0.060	0.141	0.799	0.8209
171	165.41	0.516	0.9517	0.0007	0.039	0.139	0.823	0.8458
246	245.54	0.693	0.9673	0.0006	0.021	0.134	0.845	0.8771
500	525.70	1.277	0.9879	0.0004	0.007	0.133	0.860	0.9169
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	20.49	0.049	0.9621	0.0009	0.745	0.010	0.245	0.8420
43	33.89	0.097	0.9368	0.0008	0.415	0.035	0.550	0.8706
61	49.56	0.136	0.9377	0.0008	0.233	0.055	0.712	0.8389
76	63.51	0.166	0.9440	0.0007	0.150	0.065	0.785	0.8346
96	82.86	0.201	0.9520	0.0007	0.084	0.074	0.842	0.8463
125	112.37	0.249	0.9649	0.0006	0.041	0.078	0.881	0.8676
171	159.19	0.311	0.9753	0.0005	0.015	0.082	0.902	0.8908
246	236.66	0.394	0.9856	0.0004	0.004	0.084	0.912	0.9141
500	500.20	0.635	0.9949	0.0002	0.000	0.083	0.917	0.9361
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$P(M)$	$P(N_1)$	$P(N)$	$1-\hat{\alpha}$
24	22.90	0.039	0.9808	0.0009	0.880	0.002	0.118	0.9608
43	34.35	0.089	0.9535	0.0008	0.563	0.013	0.424	0.9062
61	49.56	0.128	0.9480	0.0007	0.330	0.026	0.644	0.8666
76	63.10	0.154	0.9502	0.0007	0.213	0.037	0.751	0.8508
96	82.32	0.188	0.9571	0.0006	0.118	0.046	0.835	0.8529
125	111.23	0.227	0.9672	0.0006	0.052	0.056	0.892	0.8731
171	157.35	0.280	0.9777	0.0005	0.017	0.061	0.922	0.8977
246	235.12	0.353	0.9875	0.0003	0.004	0.062	0.934	0.9186
500	493.58	0.553	0.9953	0.0002	0.000	0.069	0.931	0.9363

Appendix C

Testing the Uniformity of the Round Off Errors Using the One-Sample Kolmogorov-Smirnov Test

TABLE C1: The underlying distribution is *standard normal*

Sample mean and sample variance

n^*	24	43	61	76	96	125	171	246	500
$m = 5$	0.49364	0.49834	0.49812	0.50048	0.49845	0.49906	0.50137	0.50071	0.49853
$m = 15$	0.49650	0.49538	0.49810	0.49667	0.49977	0.50109	0.49910	0.49743	0.49820
$m = 20$	0.49818	0.49541	0.49534	0.49732	0.49847	0.49827	0.50134	0.50023	0.50016
n^*	24	43	61	76	96	125	171	246	500
$m = 5$	0.289296	0.288961	0.288435	0.288331	0.288164	0.288462	0.289102	0.289191	0.288560
$m = 15$	0.287997	0.288547	0.288268	0.288428	0.288969	0.288467	0.287896	0.288917	0.288368
$m = 20$	0.288265	0.289189	0.288286	0.289392	0.288015	0.287741	0.288949	0.289133	0.287682
Kolmogorov-Smirnov Z and asymptotic significance (2-tailed)									
n^*	24	43	61	76	96	125	171	246	500
$m = 5$	2.849	1.024	0.984	0.738	1.194	0.765	0.885	0.693	0.957
p - value	0.000	0.245	0.288	0.648	0.115	0.602	0.413	0.723	0.319
$m = 15$	1.677	1.923	1.440	1.798	0.631	0.823	0.823	1.561	1.131
p - value	0.007	0.001	0.032	0.003	0.821	0.507	0.507	0.015	0.154
$m = 20$	1.190	1.990	2.375	1.440	1.328	1.087	0.962	0.626	0.818
p - value	0.118	0.001	0.000	0.032	0.059	0.188	0.314	0.828	0.515

TABLE C2: The underlying distribution is *standard uniform*

Sample mean and sample variance

n^*	24	43	61	76	96	125	171	246	500
$m=5$	0.49590	0.49988	0.50084	0.49959	0.49868	0.49820	0.49737	0.49838	0.49902
$m=15$	0.49613	0.49710	0.50148	0.50055	0.49843	0.49992	0.49840	0.50111	0.49988
$m=20$	0.49954	0.49365	0.49778	0.49908	0.49792	0.50043	0.49915	0.49922	0.50037
n^*	24	43	61	76	96	125	171	246	500
$m=5$	0.288914	0.288738	0.288710	0.287476	0.287966	0.289064	0.288151	0.289227	0.289059
$m=15$	0.288364	0.289391	0.288393	0.288332	0.289708	0.288167	0.289492	0.288656	0.287506
$m=20$	0.288685	0.288945	0.289072	0.288661	0.288475	0.289036	0.289519	0.288739	0.289506

Kolmogorov-Smirnov Z and asymptotic significance (2-tailed)

n^*	24	43	61	76	96	125	171	246	500
$m=5$	1.816	0.631	0.769	0.948	1.091	0.863	1.216	0.935	0.912
p - value	0.003	0.821	0.595	0.330	0.185	0.446	0.104	0.347	0.376
$m=15$	1.699	1.695	1.020	0.671	1.199	0.894	1.136	1.060	0.970
p - value	0.006	0.006	0.250	0.759	0.113	0.400	0.151	0.211	0.303
$m=20$	0.648	2.379	1.337	0.769	1.140	0.778	0.970	0.859	0.738
p - value	0.794	0.000	0.056	0.595	0.148	0.580	0.303	0.452	0.648

TABLE C3: The underlying distribution is *exponential* with mean one

Sample mean and sample variance

n^*	24	43	61	76	96	125	171	246	500
$m=5$	0.49415	0.49557	0.49810	0.49972	0.49618	0.49862	0.49799	0.49751	0.49877
$m=15$	0.49514	0.49895	0.49812	0.49643	0.49854	0.49855	0.49884	0.49765	0.50121
$m=20$	0.49887	0.49990	0.49724	0.49735	0.49892	0.50013	0.50151	0.49999	0.49979
n^*	24	43	61	76	96	125	171	246	500
$m=5$	0.288686	0.288570	0.288959	0.287910	0.289384	0.288984	0.289011	0.288865	0.288169
$m=15$	0.289140	0.289373	0.288103	0.288940	0.288650	0.288360	0.289308	0.288476	0.288297
$m=20$	0.288723	0.287310	0.288421	0.288647	0.288829	0.288311	0.288754	0.288124	0.289311

Kolmogorov-Smirnov Z and asymptotic significance (2-tailed)

n^*	24	43	61	76	96	125	171	246	500
$m=5$	2.312	2.312	1.123	0.760	1.641	1.261	1.127	1.440	0.823
p - value	0.000	0.000	0.161	0.610	0.009	0.083	0.158	0.032	0.507
$m=15$	2.446	1.167	1.279	1.588	1.118	1.145	0.953	1.525	0.868
p - value	0.000	0.131	0.076	0.013	0.164	0.145	0.324	0.019	0.439
$m=20$	0.814	0.944	1.476	1.377	0.769	0.792	1.297	0.684	0.868
p - value	0.522	0.335	0.026	0.045	0.595	0.558	0.069	0.737	0.439

Appendix D

Controlled Triple Sampling Simulation to Estimate the Optimal Sample Size, the Population Mean and the Coverage Probability at $m = 5, 15, 20,\ \alpha = 0.05$ **and** $\delta = 0.5$

Note: Each table is divided into three sub attached tables, the first one at $m = 5$, the second at $m = 15$ and the last at $m = 20$.

TABLE D1

The underlying distribution is a *standard normal*

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.94	0.051	0.0007	0.0013	0.9768
43	37.80	0.081	0.0005	0.0009	0.9780
61	55.67	0.104	-0.0002	0.0008	0.9800
76	70.88	0.120	0.0001	0.0007	0.9841
96	91.16	0.140	-0.0003	0.0006	0.9871
125	120.85	0.164	-0.0005	0.0005	0.9908
171	167.48	0.201	0.0000	0.0004	0.9926
246	244.92	0.250	0.0002	0.0003	0.9962
500	505.86	0.410	-0.0001	0.0002	0.9982
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.55	0.038	0.0007	0.0011	0.9966
43	38.16	0.068	-0.0009	0.0008	0.9925
61	56.47	0.081	0.0021	0.0006	0.9946
76	71.51	0.090	0.0006	0.0006	0.9957
96	91.80	0.100	0.0004	0.0005	0.9976
125	120.58	0.114	-0.0002	0.0004	0.9984
171	166.70	0.132	-0.0001	0.0004	0.9989
246	241.86	0.156	0.0002	0.0003	0.9991
500	496.49	0.223	-0.0004	0.0002	0.9994
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.79	0.019	0.0007	0.0010	0.9991
43	36.38	0.071	0.0005	0.0008	0.9937
61	55.90	0.083	-0.0003	0.0007	0.9944
76	71.50	0.090	-0.0001	0.0006	0.9963
96	91.71	0.098	-0.0003	0.0005	0.9976
125	120.82	0.111	-0.0001	0.0004	0.9988
171	166.83	0.127	0.0001	0.0004	0.9991
246	242.05	0.151	0.0005	0.0003	0.9995
500	496.43	0.215	-0.0002	0.0002	0.9996

TABLE D2

The underlying distribution is *standard uniform*

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.88	0.04	0.4993	0.0004	0.9528
43	39.25	0.06	0.4997	0.0003	0.9657
61	57.08	0.07	0.4999	0.0002	0.9754
76	72.10	0.08	0.4996	0.0002	0.9813
96	92.45	0.09	0.5001	0.0002	0.9867
125	121.43	0.10	0.5001	0.0002	0.9910
171	167.73	0.12	0.4999	0.0001	0.9948
246	243.32	0.14	0.5001	0.0001	0.9969
500	498.53	0.19	0.4999	0.0001	0.9988
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	17.93	0.03	0.5000	0.0003	0.9962
43	40.32	0.05	0.4999	0.0002	0.9876
61	58.97	0.05	0.4999	0.0002	0.9933
76	74.20	0.06	0.5000	0.0002	0.9964
96	94.38	0.06	0.4999	0.0001	0.9979
125	123.46	0.07	0.5001	0.0001	0.9986
171	169.55	0.08	0.4998	0.0001	0.9993
246	244.45	0.09	0.5000	0.0001	0.9995
500	498.38	0.13	0.4999	0.0001	0.9996
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.03	0.00	0.4997	0.0003	0.9991
43	37.02	0.06	0.4999	0.0002	0.9907
61	58.82	0.06	0.5000	0.0002	0.9931
76	74.24	0.06	0.5000	0.0002	0.9968
96	94.45	0.06	0.5000	0.0001	0.9980
125	123.41	0.07	0.5000	0.0001	0.9991
171	169.59	0.08	0.5000	0.0001	0.9994
246	244.59	0.09	0.5000	0.0001	0.9994
500	498.56	0.13	0.5000	0.0001	0.9996

TABLE D3

The underlying distribution is $t(5)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.13	0.08	-0.0014	0.0015	0.9851
43	36.93	0.19	-0.0019	0.0012	0.9788
61	54.63	0.19	-0.0006	0.0010	0.9796
76	70.04	0.25	-0.0005	0.0009	0.9801
96	91.49	0.35	0.0000	0.0008	0.9830
125	121.40	0.36	0.0001	0.0006	0.9853
171	169.18	0.50	-0.0018	0.0005	0.9896
246	250.39	0.62	0.0002	0.0004	0.9920
500	525.01	1.15	-0.0005	0.0003	0.9971
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.13	0.05	0.0014	0.0013	0.9985
43	35.55	0.09	-0.0004	0.0010	0.9944
61	53.01	0.14	0.0005	0.0009	0.9941
76	67.74	0.16	-0.0003	0.0008	0.9953
96	88.13	0.19	0.0004	0.0007	0.9954
125	117.74	0.24	-0.0008	0.0006	0.9967
171	164.90	0.30	0.0004	0.0005	0.9979
246	242.06	0.41	-0.0008	0.0004	0.9985
500	502.28	0.70	-0.0003	0.0003	0.9993
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	22.32	0.04	0.0001	0.0012	0.9992
43	34.71	0.09	0.0005	0.0010	0.9964
61	51.91	0.12	-0.0005	0.0009	0.9951
76	66.64	0.14	-0.0009	0.0008	0.9955
96	87.39	0.19	-0.0003	0.0007	0.9961
125	116.88	0.23	0.0008	0.0006	0.9976
171	163.58	0.35	0.0000	0.0005	0.9983
246	239.62	0.36	-0.0001	0.0004	0.9988
500	498.40	0.72	-0.0002	0.0003	0.9995

TABLE D4

The underlying distribution is $t(10)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\widehat{\mu}$	$s.e.\widehat{\mu}$	$1-\widehat{\alpha}$
24	19.59	0.06	-0.0021	0.0014	0.9813
43	37.44	0.10	-0.0011	0.0010	0.9785
61	55.42	0.13	-0.0012	0.0009	0.9811
76	70.41	0.15	-0.0002	0.0007	0.9824
96	90.76	0.18	0.0003	0.0006	0.9854
125	120.55	0.21	-0.0004	0.0005	0.9881
171	168.65	0.27	0.0001	0.0005	0.9914
246	246.56	0.34	-0.0004	0.0004	0.9949
500	511.53	0.60	-0.0004	0.0002	0.9980
n^*	$\overline{N}$	$s.e.\overline{N}$	$\widehat{\mu}$	$s.e.\widehat{\mu}$	$1-\widehat{\alpha}$
24	19.94	0.04	-0.0014	0.0012	0.9971
43	37.24	0.08	-0.0008	0.0009	0.9930
61	55.15	0.09	-0.0015	0.0007	0.9949
76	69.95	0.11	-0.0004	0.0006	0.9957
96	90.07	0.12	-0.0004	0.0006	0.9972
125	119.57	0.14	-0.0007	0.0005	0.9980
171	165.62	0.16	-0.0004	0.0004	0.9988
246	240.94	0.20	-0.0002	0.0003	0.9990
500	497.18	0.30	-0.0004	0.0002	0.9995
n^*	$\overline{N}$	$s.e.\overline{N}$	$\widehat{\mu}$	$s.e.\widehat{\mu}$	$1-\widehat{\alpha}$
24	21.42	0.03	0.0003	0.0011	0.9990
43	35.68	0.08	0.0005	0.0009	0.9949
61	54.47	0.09	0.0001	0.0007	0.9945
76	70.01	0.10	-0.0002	0.0006	0.9959
96	90.02	0.12	-0.0005	0.0006	0.9972
125	119.14	0.13	-0.0004	0.0005	0.9981
171	165.27	0.16	-0.0004	0.0004	0.9990
246	240.84	0.19	0.0000	0.0003	0.9993
500	495.73	0.28	0.0001	0.0002	0.9996

TABLE D5

The underlying distribution is $t(25)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.73	0.05	0.0014	0.0013	0.9782
43	37.59	0.09	0.0011	0.0010	0.9774
61	55.50	0.11	-0.0015	0.0008	0.9806
76	70.63	0.13	-0.0004	0.0007	0.9831
96	91.07	0.15	-0.0005	0.0006	0.9865
125	120.74	0.18	-0.0007	0.0005	0.9895
171	167.78	0.22	-0.0001	0.0004	0.9926
246	245.52	0.28	-0.0002	0.0003	0.9953
500	508.14	0.46	-0.0006	0.0002	0.9985
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.66	0.04	-0.0002	0.0011	0.9968
43	37.91	0.07	-0.0008	0.0008	0.9927
61	55.98	0.08	0.0006	0.0007	0.9941
76	70.89	0.10	-0.0004	0.0006	0.9964
96	90.99	0.11	-0.0013	0.0005	0.9971
125	120.28	0.12	0.0000	0.0004	0.9978
171	166.09	0.14	-0.0005	0.0004	0.9987
246	241.34	0.17	0.0000	0.0003	0.9991
500	496.62	0.24	-0.0002	0.0002	0.9995
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	21.02	0.02	-0.0001	0.0010	0.9992
43	36.09	0.07	-0.0009	0.0008	0.9940
61	55.45	0.09	0.0000	0.0007	0.9947
76	70.87	0.09	0.0000	0.0006	0.9954
96	91.13	0.10	-0.0004	0.0005	0.9976
125	120.19	0.12	-0.0004	0.0004	0.9986
171	166.38	0.14	0.0003	0.0004	0.9992
246	241.88	0.16	-0.0007	0.0003	0.9995
500	495.73	0.23	-0.0007	0.0002	0.9995

TABLE D6

The underlying distribution is $t(50)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.88	0.05	0.0003	0.0013	0.9787
43	37.78	0.08	0.0002	0.0010	0.9772
61	55.81	0.11	-0.0004	0.0008	0.9802
76	70.65	0.12	-0.0008	0.0007	0.9843
96	90.94	0.14	-0.0008	0.0006	0.9865
125	120.46	0.17	0.0001	0.0005	0.9902
171	167.91	0.21	-0.0011	0.0004	0.9927
246	244.88	0.26	-0.0006	0.0003	0.9961
500	507.06	0.43	-0.0003	0.0002	0.9983
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.54	0.04	-0.0009	0.0011	0.9967
43	38.10	0.07	-0.0004	0.0008	0.9922
61	56.02	0.08	-0.0003	0.0007	0.9940
76	71.21	0.09	0.0006	0.0006	0.9954
96	91.23	0.10	0.0000	0.0005	0.9970
125	120.52	0.12	-0.0002	0.0004	0.9979
171	166.39	0.14	0.0004	0.0004	0.9990
246	241.67	0.16	-0.0008	0.0003	0.9994
500	496.49	0.23	-0.0005	0.0002	0.9994
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.86	0.02	-0.0012	0.0010	0.9990
43	36.25	0.07	-0.0008	0.0008	0.9936
61	55.73	0.09	0.0008	0.0007	0.9948
76	71.15	0.09	-0.0001	0.0006	0.9960
96	91.46	0.10	-0.0006	0.0005	0.9975
125	120.45	0.11	-0.0001	0.0004	0.9984
171	166.78	0.13	-0.0001	0.0004	0.9989
246	241.71	0.16	0.0001	0.0003	0.9995
500	495.86	0.22	-0.0001	0.0002	0.9997

TABLE D7

The underlying distribution is $t(100)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.91	0.05	0.0033	0.0013	0.9780
43	37.77	0.08	0.0003	0.0009	0.9773
61	55.55	0.10	-0.0006	0.0008	0.9810
76	70.84	0.12	0.0000	0.0007	0.9832
96	91.00	0.14	0.0000	0.0006	0.9866
125	120.81	0.17	0.0000	0.0005	0.9901
171	167.67	0.20	0.0002	0.0004	0.9936
246	244.61	0.26	-0.0007	0.0003	0.9960
500	506.48	0.42	-0.0002	0.0002	0.9984
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	19.52	0.04	-0.0005	0.0011	0.9968
43	38.09	0.07	0.0001	0.0008	0.9919
61	56.29	0.08	-0.0005	0.0007	0.9945
76	71.38	0.09	0.0006	0.0006	0.9961
96	91.46	0.10	-0.0005	0.0005	0.9970
125	120.57	0.12	0.0003	0.0004	0.9981
171	166.64	0.13	-0.0003	0.0004	0.9989
246	242.02	0.16	0.0002	0.0003	0.9995
500	496.39	0.23	0.0001	0.0002	0.9997
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.82	0.02	0.0000	0.0010	0.9990
43	36.25	0.07	-0.0006	0.0008	0.9940
61	55.85	0.08	-0.0008	0.0007	0.9946
76	71.28	0.09	-0.0011	0.0006	0.9961
96	91.65	0.10	-0.0001	0.0005	0.9978
125	120.74	0.11	-0.0007	0.0004	0.9987
171	166.83	0.13	-0.0008	0.0004	0.9991
246	242.03	0.15	-0.0007	0.0003	0.9993
500	496.04	0.22	-0.0003	0.0002	0.9997

TABLE D8

The underlying distribution is *beta* $(3,2)$

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.29	0.05	0.61	0	0.9651
43	38.40	0.07	0.60	0	0.9709
61	56.25	0.09	0.60	0	0.9779
76	71.24	0.10	0.60	0	0.9824
96	91.39	0.11	0.60	0	0.9862
125	120.78	0.13	0.60	0	0.9891
171	167.72	0.16	0.60	0	0.9938
246	243.44	0.19	0.60	0	0.9964
500	501.39	0.28	0.60	0	0.9985
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	18.85	0.03	0.6	0	0.9967
43	39.22	0.06	0.6	0	0.9902
61	57.74	0.07	0.6	0	0.9915
76	72.74	0.08	0.6	0	0.9948
96	93.04	0.08	0.6	0	0.9970
125	122.04	0.09	0.6	0	0.9983
171	168.08	0.11	0.6	0	0.9991
246	243.12	0.13	0.6	0	0.9994
500	497.35	0.18	0.6	0	0.9996
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.30	0.01	0.6	0	0.9993
43	36.62	0.07	0.6	0	0.9929
61	57.28	0.07	0.6	0	0.9917
76	72.91	0.08	0.6	0	0.9951
96	93.09	0.08	0.6	0	0.9975
125	122.00	0.09	0.6	0	0.9984
171	168.32	0.10	0.6	0	0.9991
246	243.19	0.12	0.6	0	0.9994
500	497.58	0.17	0.6	0	0.9996

TABLE D9

The underlying distribution is *exponential* with mean one

n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	18.03	0.08	0.8730	0.0012	0.9873
43	34.09	0.15	0.8822	0.0011	0.9231
61	50.94	0.21	0.8987	0.0010	0.8975
76	65.64	0.25	0.9110	0.0010	0.8958
96	86.45	0.31	0.9237	0.0009	0.9010
125	116.01	0.39	0.9395	0.0008	0.9178
171	164.76	0.52	0.9548	0.0007	0.9340
246	247.48	0.71	0.9714	0.0006	0.9563
500	526.05	1.27	0.9884	0.0004	0.9826
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	20.88	0.05	0.9576	0.0009	0.9998
43	34.62	0.10	0.9380	0.0008	0.9959
61	50.04	0.14	0.9397	0.0008	0.9809
76	64.02	0.17	0.9461	0.0007	0.9675
96	83.38	0.20	0.9550	0.0007	0.9599
125	112.43	0.25	0.9650	0.0006	0.9621
171	159.31	0.31	0.9772	0.0005	0.9759
246	237.08	0.39	0.9868	0.0004	0.9876
500	498.73	0.63	0.9948	0.0002	0.9967
n^*	$\overline{N}$	$s.e.\overline{N}$	$\hat{\mu}$	$s.e.\hat{\mu}$	$1-\hat{\alpha}$
24	23.27	0.04	0.9791	0.0009	0.9998
43	35.10	0.09	0.9536	0.0008	0.9986
61	50.13	0.13	0.9490	0.0007	0.9918
76	63.53	0.15	0.9504	0.0007	0.9823
96	82.65	0.19	0.9589	0.0006	0.9720
125	111.11	0.23	0.9672	0.0005	0.9678
171	157.53	0.28	0.9782	0.0005	0.9777
246	234.52	0.36	0.9874	0.0003	0.9894
500	494.99	0.55	0.9952	0.0002	0.9972

Appendix E

Simulation Results for the Type II Error Probability at $\alpha = 0.05, \beta_t = 0.05$

Note: The symbols P, M and L stands for platykurtic (standard uniform), mesokurtic (standard normal) and leptokurtic (exponential with mean one) respectively.

TABLE E1: The underlying distributions are *standard uniform*, *standard normal* and *exponential* respectively at $\delta = 0.5$ and $m = 15$

Selected values of n^*

	24			76			96			125			246			500		
k	P	M	L	P	M	L	P	M	L	P	M	L	P	M	L	P	M	L
.00	.500	.501	.421	.497	.500	.411	.504	.502	.420	.503	.501	.435	.495	.499	.460	.500	.498	.470
.01	.488	.488	.408	.484	.486	.398	.490	.489	.407	.489	.487	.422	.481	.486	.447	.485	.484	.457
.02	.475	.476	.393	.470	.473	.387	.476	.474	.395	.475	.473	.409	.467	.471	.433	.471	.470	.443
.03	.464	.464	.380	.456	.459	.374	.462	.461	.381	.460	.460	.396	.453	.457	.418	.455	.455	.430
.04	.452	.451	.367	.442	.446	.362	.447	.446	.368	.445	.445	.384	.440	.442	.405	.440	.442	.415
.05	.440	.438	.354	.428	.432	.350	.433	.432	.356	.432	.431	.370	.425	.428	.392	.426	.427	.401
.06	.429	.426	.341	.415	.419	.337	.419	.419	.343	.418	.418	.357	.411	.414	.378	.412	.414	.387
.07	.417	.413	.328	.401	.406	.325	.405	.405	.331	.404	.405	.344	.397	.400	.365	.398	.400	.373
.08	.405	.402	.315	.388	.393	.315	.391	.391	.320	.390	.391	.332	.383	.386	.352	.384	.385	.359
.09	.394	.390	.302	.375	.380	.303	.377	.378	.308	.377	.378	.319	.370	.373	.339	.371	.371	.345
.10	.382	.378	.290	.362	.367	.291	.364	.366	.297	.363	.365	.306	.357	.359	.327	.358	.358	.332
.20	.276	.265	.181	.241	.245	.185	.242	.245	.188	.239	.242	.194	.235	.238	.206	.235	.237	.211
.30	.188	.175	.100	.147	.151	.105	.147	.151	.107	.144	.147	.110	.140	.141	.116	.140	.142	.120
.40	.121	.107	.050	.083	.087	.053	.081	.087	.055	.080	.082	.055	.076	.077	.059	.074	.077	.061
.50	.073	.062	.024	.044	.048	.023	.041	.047	.025	.040	.043	.025	.037	.037	.027	.036	.038	.028

TABLE E2: The underlying distribution is a *standard normal* at $\delta = 0.5$ and $m = 5, 15, 20$

	24			76			96			125			246			500		
k	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20
.00	.499	.501	.499	.499	.500	.502	.504	.502	.501	.499	.501	.501	.500	.499	.502	.499	.498	.499
.01	.487	.488	.485	.487	.486	.487	.490	.489	.486	.484	.487	.487	.485	.486	.487	.485	.484	.485
.02	.475	.476	.471	.473	.473	.474	.476	.474	.472	.470	.473	.472	.471	.471	.473	.470	.470	.470
.03	.462	.464	.458	.460	.459	.460	.463	.461	.459	.456	.460	.458	.457	.457	.460	.457	.455	.457
.04	.449	.451	.445	.447	.446	.447	.450	.446	.445	.444	.445	.445	.443	.442	.445	.442	.442	.443
.05	.437	.438	.430	.433	.432	.433	.436	.432	.431	.430	.431	.431	.430	.428	.431	.428	.427	.430
.06	.425	.426	.417	.419	.419	.419	.422	.419	.419	.417	.418	.417	.416	.414	.416	.414	.414	.415
.07	.412	.413	.404	.406	.406	.407	.408	.405	.405	.404	.405	.403	.402	.400	.402	.400	.400	.401
.08	.401	.402	.390	.393	.393	.394	.394	.391	.392	.392	.391	.389	.388	.386	.389	.386	.385	.388
.09	.388	.390	.378	.380	.380	.381	.381	.378	.379	.377	.378	.377	.374	.373	.375	.373	.371	.374
.10	.375	.378	.365	.367	.367	.368	.368	.366	.365	.363	.365	.363	.361	.359	.362	.359	.358	.361
.20	.266	.265	.250	.252	.245	.249	.247	.245	.244	.244	.242	.243	.240	.238	.238	.234	.237	.240
.30	.183	.175	.158	.162	.151	.153	.155	.151	.151	.152	.147	.148	.145	.141	.143	.139	.142	.143
.40	.121	.107	.091	.100	.087	.087	.092	.087	.086	.088	.082	.082	.079	.077	.077	.075	.077	.078
.50	.081	.062	.048	.060	.048	.047	.052	.047	.045	.048	.043	.043	.040	.037	.039	.037	.038	.038

TABLE E3: The underlying distribution is a *standard uniform* at $\delta = 0.5$ and $m = 5, 15, 20$

	24			76			96			125			246			500		
k	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20
.00	.502	.500	.501	.497	.497	.498	.494	.504	.500	.495	.503	.499	.498	.495	.499	.503	.500	.501
.01	.489	.488	.487	.483	.484	.484	.480	.490	.486	.481	.489	.485	.484	.481	.484	.487	.485	.487
.02	.476	.475	.473	.470	.470	.470	.466	.476	.471	.467	.475	.471	.469	.467	.471	.473	.471	.472
.03	.463	.464	.460	.457	.456	.455	.452	.462	.457	.454	.460	.458	.455	.453	.456	.458	.455	.458
.04	.451	.452	.447	.443	.442	.442	.438	.447	.443	.440	.445	.444	.441	.440	.442	.444	.440	.443
.05	.439	.440	.434	.430	.428	.428	.425	.433	.429	.426	.432	.431	.427	.425	.427	.429	.426	.429
.06	.426	.429	.421	.417	.415	.413	.412	.419	.416	.412	.418	.417	.412	.411	.414	.415	.412	.415
.07	.414	.417	.409	.404	.401	.399	.398	.405	.402	.399	.404	.403	.398	.397	.399	.402	.398	.401
.08	.401	.405	.396	.391	.388	.384	.385	.391	.389	.386	.390	.389	.384	.383	.386	.388	.384	.387
.09	.388	.394	.383	.377	.375	.371	.371	.377	.375	.373	.377	.376	.372	.370	.373	.375	.371	.374
.10	.377	.382	.370	.364	.362	.358	.360	.364	.361	.360	.363	.363	.359	.357	.359	.362	.358	.359
.20	.269	.276	.254	.247	.241	.240	.241	.242	.241	.238	.239	.239	.236	.235	.236	.236	.235	.237
.30	.185	.188	.161	.155	.147	.146	.149	.147	.147	.145	.144	.144	.140	.140	.141	.141	.140	.141
.40	.126	.121	.094	.091	.083	.083	.085	.081	.082	.081	.080	.079	.078	.076	.077	.075	.074	.076
.50	.088	.073	.050	.052	.044	.043	.048	.041	.041	.042	.040	.040	.038	.037	.038	.037	.036	.036

TABLE E4: The underlying distribution is *exponential* with mean one at $\delta = 0.5$ and $m = 5, 15, 20$

	24			76			96			125			246			500		
k	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20	5	15	20
.00	.355	.421	.457	.390	.411	.412	.406	.420	.424	.415	.435	.438	.449	.460	.463	.468	.470	.474
.01	.345	.408	.443	.378	.398	.400	.393	.407	.412	.403	.422	.424	.436	.447	.449	.453	.457	.460
.02	.333	.393	.428	.366	.387	.387	.381	.395	.399	.390	.409	.412	.421	.433	.436	.438	.443	.446
.03	.322	.380	.414	.354	.374	.374	.369	.381	.387	.377	.396	.398	.409	.418	.422	.424	.430	.431
.04	.311	.367	.400	.343	.362	0.362	.357	.368	.374	.365	.384	.385	.396	.405	.408	.410	.415	.417
.05	.301	.354	.386	.332	.350	.350	.345	.356	.362	.352	.370	.372	.382	.392	.395	.396	.401	.403
.06	.290	.341	.373	.320	.337	.338	.333	.343	.350	.340	.357	.360	.368	.378	.382	.382	.387	.390
.07	.280	.328	.359	.309	.325	.326	.321	.331	.338	.327	.344	.347	.356	.365	.368	.367	.373	.377
.08	.268	.315	.346	.297	.315	.314	.310	.320	.325	.316	.332	.334	.343	.352	.354	.354	.359	.363
.09	.258	.302	.333	.285	.303	.303	.298	.308	.312	.304	.319	.321	.331	.339	.341	.340	.345	.349
.10	.249	.290	.320	.274	.291	.291	.287	.297	.300	.292	.306	.308	.317	.327	.328	.326	.332	.335
.20	.157	.181	.201	.174	.185	.183	.182	.188	.192	.187	.194	.196	.204	.206	.210	.207	.211	.213
.30	.090	.100	.113	.103	.105	.104	.105	.107	.111	.108	.110	.112	.119	.116	.119	.119	.120	.123
.40	.049	.050	.057	.057	.053	.053	.057	.055	.056	.058	.055	.056	.063	.059	.059	.064	.061	.061
.50	.027	.024	.026	.030	.023	.024	.030	.025	.026	.030	.025	.025	.031	.027	.027	.031	.028	.027

TABLE E5: Comparison between the asymptotic and simulation results for Type II error probability for underlying distribution standard normal at $\delta = 0.5$ and $m = 15$. β_{asy} is the asymptotic value .

k	β_{asy}	24	43	61	76	96	125	171	246	500
.00	0.500	0.501	0.500	0.503	0.500	0.502	0.501	0.499	0.499	0.498
.01	0.486	0.488	0.486	0.489	0.486	0.489	0.487	0.484	0.486	0.484
.02	0.471	0.476	0.473	0.475	0.473	0.474	0.473	0.470	0.471	0.470
.03	0.457	0.464	0.460	0.462	0.459	0.461	0.46	0.455	0.457	0.455
.04	0.443	0.451	0.448	0.448	0.446	0.446	0.445	0.441	0.442	0.442
.05	0.429	0.438	0.434	0.435	0.432	0.432	0.431	0.427	0.428	0.427
.06	0.414	0.426	0.421	0.421	0.419	0.419	0.418	0.414	0.414	0.414
.07	0.400	0.413	0.408	0.409	0.406	0.405	0.405	0.400	0.400	0.400
.08	0.387	0.402	0.396	0.395	0.393	0.391	0.391	0.385	0.386	0.385
.09	0.373	0.390	0.383	0.381	0.380	0.378	0.378	0.372	0.373	0.371
.10	0.359	0.378	0.371	0.368	0.367	0.366	0.365	0.358	0.359	0.358
.20	0.236	0.265	0.253	0.252	0.245	0.245	0.242	0.237	0.238	0.237
.30	0.140	0.175	0.167	0.160	0.151	0.151	0.147	0.143	0.141	0.142
.40	0.075	0.107	0.100	0.095	0.087	0.087	0.082	0.079	0.077	0.077
.50	0.036	0.062	0.059	0.053	0.048	0.047	0.043	0.040	0.037	0.038

TABLE E6: Comparison between the asymptotic and simulation results for Type II error probability for underlying distribution standard uniform at $\delta = 0.5$ and $m = 15$. β_{asy} is the asymptotic value.

k	β_{asy}	24	43	61	76	96	125	171	246	500
.00	0.500	0.500	0.501	0.503	0.497	0.504	0.503	0.503	0.495	0.500
.01	0.488	0.488	0.489	0.490	0.484	0.490	0.489	0.487	0.481	0.485
.02	0.476	0.475	0.474	0.475	0.470	0.476	0.475	0.473	0.467	0.471
.03	0.463	0.464	0.460	0.462	0.456	0.462	0.460	0.458	0.453	0.455
.04	0.451	0.452	0.447	0.448	0.442	0.447	0.445	0.443	0.440	0.440
.05	0.439	0.440	0.434	0.434	0.428	0.433	0.432	0.429	0.425	0.426
.06	0.427	0.429	0.421	0.421	0.415	0.419	0.418	0.416	0.411	0.412
.07	0.415	0.417	0.408	0.407	0.401	0.405	0.404	0.401	0.397	0.398
.08	0.403	0.405	0.395	0.393	0.388	0.391	0.390	0.387	0.383	0.384
.09	0.391	0.394	0.383	0.380	0.375	0.377	0.377	0.374	0.37	0.371
.10	0.379	0.382	0.371	0.367	0.362	0.364	0.363	0.361	0.357	0.358
.20	0.263	0.276	0.253	0.249	0.241	0.242	0.239	0.236	0.235	0.235
.30	0.162	0.188	0.163	0.153	0.147	0.147	0.144	0.142	0.140	0.140
.40	0.084	0.121	0.100	0.087	0.083	0.081	0.080	0.077	0.076	0.074
.50	0.034	0.073	0.059	0.047	0.044	0.041	0.040	0.039	0.037	0.036

TABLE E7: Comparison between the asymptotic and simulation results for Type II error probability for underlying distribution exponential with mean one at $\delta = 0.5$ and $m = 15$. β_{asy} is the asymptotic value.

k	β_{asy}	24	43	61	76	96	125	171	246	500
.00	0.633	0.421	0.391	0.401	0.411	0.420	0.435	0.448	0.460	0.470
.01	0.622	0.408	0.379	0.390	0.398	0.407	0.422	0.434	0.447	0.457
.02	0.616	0.393	0.367	0.377	0.387	0.395	0.409	0.421	0.433	0.443
.03	0.615	0.380	0.354	0.364	0.374	0.381	0.396	0.407	0.418	0.430
.04	0.618	0.367	0.342	0.352	0.362	0.368	0.384	0.394	0.405	0.415
.05	0.626	0.354	0.331	0.340	0.350	0.356	0.370	0.381	0.392	0.401
.06	0.638	0.341	0.319	0.328	0.337	0.343	0.357	0.368	0.378	0.387
.07	0.654	0.328	0.308	0.317	0.325	0.331	0.344	0.356	0.365	0.373
.08	0.674	0.315	0.297	0.305	0.315	0.320	0.332	0.343	0.352	0.359
.09	0.698	0.302	0.285	0.293	0.303	0.308	0.319	0.329	0.339	0.345
.10	0.725	0.290	0.274	0.281	0.291	0.297	0.306	0.316	0.327	0.332

TABLE E8: The effect of δ on the Type II error probability for underlying distribution standard normal at $\alpha = 0.05$, $\beta_t = 0.05$ and $m = 15$.

k	24			76			96			125			246			500		
.00	0.501	0.501	0.499	0.500	0.500	0.502	0.496	0.502	0.504	0.500	0.501	0.499	0.496	0.499	0.499	0.501	0.498	0.502
.01	0.489	0.488	0.485	0.487	0.486	0.487	0.482	0.489	0.490	0.486	0.487	0.485	0.481	0.486	0.484	0.487	0.484	0.487
.02	0.478	0.476	0.470	0.475	0.473	0.472	0.469	0.474	0.475	0.473	0.473	0.470	0.468	0.471	0.470	0.474	0.470	0.473
.03	0.466	0.464	0.457	0.462	0.459	0.458	0.455	0.461	0.461	0.460	0.460	0.456	0.454	0.457	0.456	0.460	0.455	0.458
.04	0.455	0.451	0.443	0.450	0.446	0.444	0.441	0.446	0.447	0.446	0.445	0.441	0.441	0.442	0.442	0.445	0.442	0.444
.05	0.444	0.438	0.428	0.437	0.432	0.430	0.429	0.432	0.433	0.432	0.431	0.426	0.428	0.428	0.427	0.432	0.427	0.429
.06	0.433	0.426	0.415	0.424	0.419	0.416	0.416	0.419	0.419	0.419	0.418	0.411	0.414	0.414	0.414	0.419	0.414	0.413
.07	0.422	0.413	0.402	0.412	0.406	0.402	0.404	0.405	0.405	0.405	0.405	0.398	0.401	0.400	0.399	0.406	0.400	0.399
.08	0.411	0.402	0.388	0.400	0.393	0.388	0.392	0.391	0.390	0.392	0.391	0.383	0.387	0.386	0.385	0.392	0.385	0.385
.09	0.399	0.390	0.375	0.387	0.380	0.373	0.379	0.378	0.376	0.379	0.378	0.369	0.374	0.373	0.370	0.378	0.371	0.372
.10	0.388	0.378	0.360	0.374	0.367	0.360	0.366	0.366	0.362	0.366	0.365	0.355	0.361	0.359	0.356	0.364	0.358	0.358
.20	0.286	0.265	0.242	0.261	0.245	0.237	0.251	0.245	0.238	0.247	0.242	0.231	0.240	0.238	0.232	0.239	0.237	0.232
.30	0.198	0.175	0.149	0.173	0.151	0.144	0.162	0.151	0.145	0.153	0.147	0.136	0.147	0.141	0.135	0.145	0.142	0.137
.40	0.129	0.107	0.087	0.110	0.087	0.078	0.098	0.087	0.078	0.091	0.082	0.073	0.081	0.077	0.072	0.078	0.077	0.072
.50	0.078	0.062	0.047	0.069	0.048	0.039	0.056	0.047	0.039	0.050	0.043	0.035	0.041	0.037	0.034	0.038	0.038	0.034

TABLE E9: The effect of δ on the Type II error probability for underlying distribution standard uniform at $\alpha = 0.05$, $\beta_t = 0.05$ and $m = 15$.

k	24			76			96			125			246			500		
.00	0.500	0.500	0.502	0.497	0.497	0.501	0.501	0.504	0.498	0.501	0.503	0.498	0.503	0.495	0.498	0.498	0.500	0.501
.01	0.489	0.488	0.487	0.483	0.484	0.486	0.487	0.490	0.485	0.487	0.489	0.483	0.489	0.481	0.484	0.485	0.485	0.486
.02	0.477	0.475	0.473	0.470	0.470	0.473	0.474	0.476	0.470	0.473	0.475	0.469	0.474	0.467	0.469	0.471	0.471	0.471
.03	0.467	0.464	0.461	0.456	0.456	0.459	0.460	0.462	0.456	0.459	0.460	0.454	0.459	0.453	0.456	0.456	0.455	0.457
.04	0.455	0.452	0.447	0.442	0.442	0.444	0.446	0.447	0.442	0.445	0.445	0.438	0.445	0.440	0.441	0.441	0.440	0.442
.05	0.444	0.440	0.435	0.429	0.428	0.431	0.432	0.433	0.428	0.432	0.432	0.423	0.432	0.425	0.427	0.428	0.426	0.428
.06	0.434	0.429	0.421	0.416	0.415	0.415	0.420	0.419	0.414	0.418	0.418	0.409	0.417	0.411	0.413	0.414	0.412	0.414
.07	0.422	0.417	0.406	0.403	0.401	0.401	0.406	0.405	0.400	0.405	0.404	0.395	0.403	0.397	0.399	0.400	0.398	0.400
.08	0.411	0.405	0.392	0.391	0.388	0.388	0.392	0.391	0.387	0.391	0.390	0.382	0.389	0.383	0.386	0.386	0.384	0.387
.09	0.400	0.394	0.379	0.378	0.375	0.374	0.378	0.377	0.372	0.377	0.377	0.368	0.377	0.370	0.372	0.372	0.371	0.372
.10	0.388	0.382	0.367	0.365	0.362	0.360	0.366	0.364	0.359	0.364	0.363	0.356	0.363	0.357	0.359	0.358	0.358	0.359
.20	0.286	0.276	0.247	0.251	0.241	0.236	0.246	0.242	0.237	0.241	0.239	0.234	0.239	0.235	0.234	0.234	0.235	0.233
.30	0.198	0.188	0.153	0.160	0.147	0.141	0.154	0.147	0.142	0.148	0.144	0.140	0.143	0.140	0.139	0.140	0.140	0.139
.40	0.127	0.121	0.090	0.098	0.083	0.078	0.090	0.081	0.078	0.084	0.080	0.076	0.080	0.076	0.076	0.076	0.074	0.073
.50	0.076	0.073	0.050	0.059	0.044	0.039	0.050	0.041	0.039	0.044	0.040	0.037	0.040	0.037	0.036	0.037	0.036	0.035

TABLE E10: The effect of δ on the Type II error probability for underlying distribution exponent tial with mean one at $\alpha = 0.05$, $\beta_t = 0.05$ and $m = 15$.

k	24			76			96			125			246			500		
.00	0.453	0.421	0.408	0.387	0.411	0.423	0.399	0.420	0.430	0.417	0.435	0.448	0.450	0.460	0.462	0.472	0.470	0.475
.01	0.440	0.408	0.394	0.376	0.398	0.410	0.387	0.407	0.416	0.406	0.422	0.433	0.438	0.447	0.449	0.459	0.457	0.462
.02	0.427	0.393	0.380	0.365	0.387	0.396	0.375	0.395	0.402	0.394	0.409	0.419	0.425	0.433	0.434	0.445	0.443	0.446
.03	0.414	0.380	0.366	0.354	0.374	0.381	0.363	0.381	0.389	0.382	0.396	0.405	0.411	0.418	0.420	0.432	0.430	0.432
.04	0.402	0.367	0.352	0.342	0.362	0.368	0.351	0.368	0.375	0.369	0.384	0.391	0.399	0.405	0.405	0.419	0.415	0.417
.05	0.389	0.354	0.339	0.331	0.350	0.355	0.340	0.356	0.362	0.358	0.370	0.378	0.387	0.392	0.390	0.405	0.401	0.402
.06	0.378	0.341	0.326	0.320	0.337	0.341	0.329	0.343	0.348	0.346	0.357	0.363	0.374	0.378	0.376	0.390	0.387	0.388
.07	0.366	0.328	0.312	0.309	0.325	0.328	0.318	0.331	0.335	0.334	0.344	0.348	0.361	0.365	0.363	0.377	0.373	0.373
.08	0.355	0.315	0.299	0.298	0.315	0.315	0.307	0.320	0.321	0.323	0.332	0.335	0.348	0.352	0.349	0.364	0.359	0.358
.09	0.344	0.302	0.286	0.286	0.303	0.302	0.296	0.308	0.307	0.312	0.319	0.321	0.336	0.339	0.333	0.351	0.345	0.344
.10	0.332	0.290	0.273	0.276	0.291	0.289	0.285	0.297	0.294	0.300	0.306	0.308	0.324	0.327	0.320	0.338	0.332	0.329
.20	0.227	0.181	0.161	0.179	0.185	0.177	0.184	0.188	0.179	0.196	0.194	0.187	0.210	0.206	0.197	0.219	0.211	0.205
.30	0.146	0.100	0.084	0.105	0.105	0.094	0.109	0.107	0.097	0.117	0.110	0.099	0.123	0.116	0.109	0.127	0.120	0.112
.40	0.086	0.050	0.037	0.055	0.053	0.045	0.059	0.055	0.045	0.064	0.055	0.048	0.066	0.059	0.052	0.068	0.061	0.056
.50	0.050	0.024	0.014	0.027	0.023	0.018	0.029	0.025	0.019	0.030	0.025	0.019	0.032	0.027	0.022	0.031	0.028	0.024

References

Abramowitz, M., and Stegun, I. A. 1972. *Handbook of Mathematical Functions with Formulas, and Graphs, and Mathematical Tables*. New York: Dover, pp. 928.

AlMahmeed, M., AlHessainan, A., Son, M. S., and Hamdy, H. I., 1998. Three stage estimation for the mean of a one parameter exponential family. *Korean Commun. Statist.*, *5*, pp. 539-557.

AlMahmeed, M., and Hamdy, H. I., 1990. Sequential estimation of linear models in three stages. *Metrika,* **37**, pp. 19-36.

Anderson, T, W., and Darling, D. A. 1954. A test of goodness of fit. *J. Amer. Statist. Assoc.*, **49**, 367, pp. 765-769.

Anscombe, F. J. 1952. Large sample theory of sequential estimation. *Proc. Cambridge. Phil. Soc.*, **45**, pp. 600-607.

Anscombe, F. J. 1953. Sequential estimation. *J. Roy. Statist. Soc.,* **15** pp. 1-21.

Balanda, K. P., and MacGillivray, H. L. 1988. Kurtosis: a critical review. *J. Amer. Statist. Assoc.*, **42**, pp. 111-119.

Balitskaya, E. O., and Zolotuhina, L. A. 1988. On the representation of a density by an Edqeworth series. *Biometrika*, **75**, pp. 185-187.

Barnett, V., Lewis, T. (1994). *Outliers in Statistical Data. 3rd Edition.* Wiley, Chichester.

Bartlett, M. S. (1935). The effect of non normality on the t distribution. *Proc. Camb. Phil.* Soc. **31**, pp. 223-231.

Barndorff-Nielsen, O. E., and Cox, D. R. 1989. *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.

Basu, A. P. 1971. On a sequential rule for estimating the location parameter of an exponential distribution. *Naval Res. Logist. Quart.*, **18**, pp. 329-337.

Bhattacharjee, G. P. 1965. Effect of non-normality on Stein's two sample test. *Ann. Math. Statist.,* **36**, pp. 651-663.

Bhattacharjee, G. P., and Malik, A. 1973. Asymptotic normality of the stopping times of some sequential procedures. *Ann. Statist.*, **1**, pp. 1203-1211.

Bhattacharjee, G. P., and Nagendra, Y. 1964. Effect of non-normality on a sequential test for mean. *Biometrika,* **51**, pp. 281-287.

Bhattacharya, R. N., and Ghosh, J. K. 1978. On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, pp. 434-451.

Bickel, P. J., and Lehmann, E. L. (1975a). Descriptive statistics for non parametric model. I Introduction. *Ann. Statist.* **3**, pp. 1038-1044.

Bickel, P. J., and Lehmann, E. L. (1975b). Descriptive statistics for non parametric model. II Location. *Ann. Statist.* 3, pp. 1045-1069.

Blinnikov, S., and Moessner, R. 1998. Expansions for nearly Gaussian distributions. *Astron. Astrophys. Suppl.* Ser. 130, pp. 193-205.

Blumenthal, S., and Govindarajulu, Z. 1977. Robustness of Stein's two stage procedure for mixtures of normal populations. *J. Amer. Statist. Assoc.*, **72** , pp. 192-196.

Box, G.E.P. 1953. Non-normality and tests of variance. *Biometrika* **40**, pp. 318-335.

Box, G.E.P., Anderson, S. L. 1955. Permutation theory in the derivation of robust criteria and the study of departures from assumption. *J. Roy. Statist. Soc.* Ser. B **17**, pp. 1-34.

Box, G. E. P., and Tiao, G. C. 1964. A note on criterion robustness and inference robustness. *Biometrika*, **51**, pp. 169-173.

Brownlee, K. A. 1965. *Statistical Theory and Methodology in Science and Engineering*. Wiley, NewYork.

Brunier, H. and Whitehead, J. 1993. PEST-Planning and Evaluation of Sequential Trials-operating manual..Reading, England: PEST Project, University of Reading, 1993 (and software).

Cabilio, P. 1977. Sequential estimation in Bernoulli trials. *Ann. Statist.* **5,** pp. 342-356.

Callahan, J. D. 1969. One sequential estimation procedures for the mean vector of a multivariate normal distribution. Tech. Rep. No. 132, Dept. of Statistics, *John Hopkins University* pp. 518-521.

Chow, Y. S. and Robbins, H. 1965. On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36,** pp. 1203-1212.

Chow, Y. S. and Martinsek, A. T. 1982. Bounded regret of a sequential procedure for estimation of the mean. *Ann. Statist.*, **10**, pp. 909-914.

Chow, Y. S. and Yu, K. F. 1981. The performance of a sequential procedure for the estimation of the mean. *Ann. Statist.*, **9**, pp. 184-189.

Costanza, M. C., Hamdy, H. I., Haugh, L. D., and Son, M. S. 1995. Type II error performance of triple sampling fixed precision confidence intervals for the normal mean. Metron, LIII, pp. 69-82.

Cox, D. R. 1952(a). A note on sequential estimation of means. *Proc. Cam. Phil. Soc.* **48**, pp. 449-450.

Cox, D. R. 1952(b). Estimation by double sampling. *Biometrika*, **39** pp. 217-227.

Cramer, H. 1957. *Mathematical Methods of Statistics.* Princeton Univ. Press, Princeton.

Dantzig, G. B. 1940. On the non existence of tests of 'students' hypothesis having power functions independent of σ . *Ann. Math. Statist.* **11**, pp. 186-192.

DasGupta, A. 2008. *Asymptotic Theory of Statistics and Probability*. Springer Sciences and Business Media, Berlin.

DeCarlo, L. T. 1997 . On the meaning and use of kurtosis. *Psychological Methods*, *2*, pp. 292-307.

Degroot, M. H. 1959. Unbiased binomial sequential estimation. *Ann. Math. Statist.* **30**, pp. 80-101.

Degroot, M. H. 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York.

Draper N. R., and Tierney, D. E. 1972. Regions of positive and unimodal series expansion of the Edgeworth and Gram-Charlier approximations. *Biometrika*, **59**, pp. 463-465.

Gayen, A. K. (1949). The distribution of Student's t in random samples of any size drawn from non normal universes. *Biometrika*, 36, pp. 353-369.

Geary, R. C. (1936). The distribution of Student's ratio for non normal samples. *J. Roy. Statist. Soc.* , **3**, pp. 178-184.

Geary, R. C. (1947). Testing for normality. *Biometrika*, 34, pp. 209-241.

Ghosh, M., and Mukhopadhyay, N. 1975b. Sequential point estimation of the parameter of a rectangular distribution. *Calcutta Statist. Assoc. Bull.*, **24**, pp. 117-122.

Ghosh, M., and Mukhopadhyay, N. 1981. Consistency and asymptotic efficiency of two stage and sequential procedures. *Sankhya*, Series A, **43**, pp. 220-227.

Ghosh, M., Mukhopadhyay, N., and Sen, P. 1997. *Sequential Estimation*. Wiley, New York.

Ghosh, M., Sinha, B. K., and Mukhopadhyay, N. 1976. Multvariate sequential point estimation. *J. Multivariate Analysis*, **6,** pp. 281-294.

Ghurye, S. G. (1949). On the use of Student's t test in an asymmetrical population. *Biometrika*, **36**, pp. 426-430.

Govindarajulu, Z. 1987 . *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation and Decision Theory. Amer. Series in Math. Manag. Sc.*, American Sciences Press, New York.

Govindarajulu, Z. and Leslie, R. T. 1972. Annotated bibliography on robustness studies of statistical procedures. U.S. Dept. of Health, *Education and Welfare Publications*. No. (HSM) pp. 72-1051.

Graybill, F. A., and Connell, T. L. 1964. Sample size required to estimate the parameter in the uniform density within d units. *J. Amer. Statist. Assoc.*, **59**, pp. 550-556.

Hall, P. 1981. Asymptotic theory and triple sampling of sequential estimation of a mean. *Ann. Statist.* **,** **9**, pp. 1229-1238.

Hall, P. 1983. Sequential estimation saving sampling operations. *J. Roy. Statist. Soc.*, B., **45**, pp. 1229-1238.

Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. Springer, New York.

Hamdy, H. I. 1988. Remarks on the asymptotic theory of triple stage estimation of the normal mean. *Scand. J. Statist.*, **15**, pp. 303-310.

Hamdy, H. I., AlMahmeed, M., Nigm, A., and Son, M. S. 1989. Three stage estimation for the exponential location parameters. *Metron*, XLVII, pp. 279-294.

Hamdy, H. I. 1997. Performance of fixed width confidence intervals under Type II errors: The exponential case. *South African Statist. J.*, 31, 259-269.

Hamdy, H. I., Son, M. S., and Al-Ibrahim, A. H. 2005. Multistage inference of the unknown normal mean. *AMS*, Advances in Modelling, Series B, Computer and Statistics, 2, pp. 1-18.

Hamdy, H. I., Costanza, M. C., and Ashikaga, T. 1995. On the Behrens-Fisher problem: an integrated triple sampling approach. (Unpublished Manuscript).

Hamdy, H. I., Mukhopadhyay N., Costanza MC., and Son MS. 1988. Triple stage point estimation for the exponential location parameter. *Ann. Instit. Statist. Math.*, **40,** pp. 785-797.

Hamdy, H. I., and Pallotta, W. J. 1987. Triple sampling procedure for estimating the scale parameter of Pareto distribution. *Commun. Statist.*, *Theory and Methods*, **16**, pp. 2155-2164.

Hampel, F. R. 1968. Contributions to the theory of robust estimators. Ph.D. dissertation, Univ. California, Berkeley.

Hampel, F. R. 1971. A general quantitative definition of robustness. *Ann. Math. Statist.* **42**, pp. 1887-1896.

Hampel, F. R. 1973. Robust estimation: A condensed partial survey. Z. *Wahrscheinlichkeitstheorie und Verw. Gebiete*, **27**, pp. 87-172.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. and Stahel, W. 1986. *Robust Statistics – The Approach Based on Influence Functions*. Wiley, New York.

Honda, T. 1992. Esquentian of the mean by three stage procedure. *Seq. Anal*, **11**, pp. 73-89.

Huber, P. J. 1964. Robust estimation of a location parameter. *Ann. Math. Statist.* , **35**, pp. 73-101.

Huber, P. J. 1981. *Robust Statistics*. Wiley, New York.

Huber, P. J., and Dutter, R. 1974. Numerical solutions of robust regression problems. COMPSTAT 1974 (eds. G. Bruckmann et al.). Physica Verlag, Vienna, pp. 165-172.

Ifram, A. F. 1965. Hypergeometric functions in sequential analysis. *Ann. Math. Statist.* **36**, pp. 1870-1872.

Jennison, C., and Turnbull, B. 1999. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, UK.

Johnson, N. L., Kotz, S., and Balakrishnan, N. 1994. *Continuous Univariate Distributions*. Vol. 1 Wiley, New York.

Jureckova, J. and Sen, P. K. 1996. *Robust Statistical Procedures, Asymptotic and Interrelations*. Wiley, New York.

Kenney, J. F., and Keeping, E. S. 1951. *Mathematics of Statistics*. NJ: Van Nostrand , Princeton.

Khan, R. A. 1968. Sequential estimation of the mean vector of a multivariate normal distribution. *Indian Statist. Instit,* 30, pp. 331-334.

Kokic P. N., and Weber N. C. 1990. An Edgeworth expansion for U-statistics based on samples from finite populations. *Ann. Probab.*, 18, pp. 390-404.

Lahiri, P. 2003. On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science*, 18, pp. 199-210.

Law, A. M. 2007. *Simulation Modelling and Analysis*. McGraw-Hill, NewYork.

Lax, D. A. 1985. Robust estimators of scale: finite-sample performance in long tailed symmetric distributions. *J. Amer. Statist. Assoc.*, 80, pp. 736-741.

Lehmann, E. L., and Casella G. 1998. *Theory of Point Estimation*. Springer, NewYork.

Liu, J-F. 2002. Three stage point estimation of the mean in poststratification. *Seq. Anal*, 21, pp. 43-57.

Liu, W. 1995. A k-stage sequential sampling procedure for estimation of a normal mean. *J. Statist. Planng. Inf.,* 65, pp. 109-127.

Liu, W. 1997. On some sample size formulae for controlling both size and power in clinical trials. *The Statistician*, 46, pp. 239-251.

Lohr, S. L. 1990. Accurate multivariate estimation using triple sampling. *Ann. Statist.*, 18, pp. 1615-1633.

Lorden, G. 1978. Nearly optimal sequential tests for exponential families. Unpublished manuscript. [Chapter 3]

Martinsek, A. T. 1983. Second order approximation to the risk of a sequential procedure. *Ann. Statist.*, 11, pp. 827-836.

Martinsek, A. T. 1988. Negative regret, optimal stopping, and the elimination of outliers. *J. Amer. Statist. Assoc.*, 83, pp. 160-163.

McCabe, G. P. 1972. Sequential estimation of a Poisson integer mean. *Ann. Math. Statist.*, 43, pp. 803-813.

McCabe, G. P. Sequential estimation of a restricted mean parameter of an exponential family. *Ann. Inst. Statist. Math.* 26, pp. 103-115.

Moshman, J. 1958. A method for selecting the size of the initial sample in Stein's two sample procedure. *Ann. Math. Statist.*, 29, pp. 667-671.

Mukopadhyay, N. 1974. Sequential estimation of a location parameter in exponential distributions. *Calcutta Statist. Assoc.* Bull., 23, pp. 85-95.

Mukopadhyay, N. 1985. A note on three-stage and sequential point estimation procedures for a normal mean. *Sequential Anal.*, 4, pp. 311-319.

Mukopadhyay, N. 1988. Sequential estimation problems for negative exponential populations. *Commun. Statist. ,Theory Methods A.* , **17**, pp. 2471-2506.

Mukopadhyay, N. 1990. Some properties of a three-stage procedure with applications in sequential analysis. *Indian J. Statist.*, **52**, Series A, pp. 218-231.

Mukopadhyay, N. 2005a. A new approach to determine the pilot sample size in two stage sampling. *Commun. Statist.*, Theory & Methods, **34**, pp. 1275-1295.

Mukopadhyay, N., and Darmanto, S.1988. Sequential estimation of the difference of means of two negative exponential populations. *Seq. Anal.*, **7**, pp. 165-190.

Mukopadhyay, N., and de Silva, B. M. 1950. *Sequential Methods and Their Applications.* Chapman and Hall/CRC, UK.

Mukopadhyay, N. and Diaz, J. 1985. Two stage sampling for estimating the mean of a negative binomial distribution. *Seq. Anal.*, **4**, pp. 1-18.

Mukopadhyay, N., and Ekwo, M. E. 1987a. Sequential estimation problems for the scale parameter of a Pareto distribution. *Scand. Actuar. J.*, pp. 83-103.

Mukopadhyay, N., and Hamdy, H. I. 1984. On estimating the difference of location parameters of two negative exponential distributions. *Canad. J. Statist.*, **12**, pp. 67-76.

Mukopadhyay, N., Hamdy H. I., AlMahmeed M., and Costanza M. C., 1987. Three stage point estimation procedures for a normal mean. *Seq. Anal.*, **6**, pp. 21-36.

Mukopadhyay, N., Hamdy H. I., Ghosh, M., and Wackerly, D. 1983. Sequential and two stage point estimation for the range in a power family distribution. *Seq. Anal.*, **2**, pp. 259-288.

Mukopadhyay, N., and Hilton, G. F. 1986. Two stage and sequential procedures for estimating the location parameter of a negative exponential distribution. *S. Afr. Statist. J.*, **20**, pp. 117-136.

Mukopadhyay, N., and Mauromoustakos, A. 1987. Three stage estimation procedures of the negative exponential distribution. *Metrika*, **34**, pp. 83-93.

Mukopadhyay, N., and Padmanabhan, A. R. 1993. A note on three-stage confidence intervals for the difference of locations: the exponential case. *Metrika*, **40**, pp. 121-128.

Pearson, K. 1905. Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder. *Biometrika, 4*, pp. 169-212.

Pearson, E. S. 1931. The analysis of variance in cases of non normal variation. *Biometrika*, **23**, pp. 114-133.

Pearson, E. S., and Adyanthaya, N. K. 1929. The distribution of frequency constants in small samples from non normal symmetrical and skew populations. *Biometrika*, **21**, pp. 259-286.

Pollock, K. H. 1978. Inference robustness versus criterion robustness: An example. *Amer. Statist.*, 32, 4, pp. 133-136.

Rahim, M. A. 1993. Economic design of control charts assuming Weibull distribution in control times. *J. Qual. Tech.*, 25, pp. 296-305.

Ramkaran, 1983. The robustness of Stein's two stage procedure. *J. Sequential Anal., 5*, pp. 139-168.

Ray, W. D. 1957. Sequential confidence intervals for the mean of a normal population with unknown variance. *J. Roy. Statist. Soc.*, B, 19, pp. 133-143.

Renyi, A. 1957. On the asymptotic distribution of the sum of a random number of independent random variables. *Acta Math. Acad. Sci. Hung.*, 8, pp. 193-199.

Robbins, H. 1948. On the asymptotic distribution of the sum of a random number of random variables. *Bull. Amer. Math. Soc.*, 54, pp. 1151-1161.

Robbins, H. 1959. Sequential estimation of the mean of a normal population. Probability and Statistics (Harald Cramer Volume). Almguist and Wiksell, Uppsala, Sweden, pp. 235-245.

Robbins, H. and Siegmund, D. O. 1974. Sequential estimation of p in Bernoulli trials. (Ed. E. J. Williams) *Jerusalem Academic Press*, pp. 103-107.

Rohatgi, V. K. 1976. *An Introduction to Probability Theory and Mathematical Statistics*. Wiley, NewYork.

Rousseeuw, P.J. and Leroy, A.M. 1987. *Robust Regression and Outlier Detection.* Wiley, New York.

Seelbinder, B. M. 1953. On Stein's two stage sampling scheme. *Ann. Math. Statist.* , 24, pp. 640-900.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Sen, P. K. 1985. *Theory and Applications of Sequential Non parametric.* Society for Industrial and Applied Mathematics, Philadelphia.

Sheskin, D. J. 2004. *Handbook of Parametric and Non-parametric Statistical Procedures.* Chapman & Hall/CRC, UK.

Sinha, B. K. and Mukopadhyay, N. 1976. Sequential estimation of a bivariate normal mean vector. Sankhya Ser. B 38, pp. 219-230.

Son, M. S., Haugh, H. I., Hamdy, H. I., and Costanza, M. C. 1997. Controlling type II error while constructing triple sampling fixed precision confidence intervals for the normal mean. *Ann. Instit. Statist. Math.* 49, pp. 681-692.

Srivistava, M. S. 1984. A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics and Probability Letters*, 2, pp. 263-267.

Starr, N. 1966. On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.* 37, pp. 1173-1185.

Starr, N., and Woodroofe, M. 1972. Further remarks on sequential point estimation: The exponential case. *Ann. Math. Statist.* 43, pp. 1147-1154.

Stein, C. 1945. A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16**, pp. 243-258.

Stein, C. 1949. Some problems in sequential estimation. (Abstract), *Econometrica*, **17**, pp. 77-78.

Stephens, M. A. 1974. Statistics for goodness of fit and some comparisons. *J. Amer. Statist. Assoc.*, 69, 363, pp. 730-737.

Sung, S. H. 2008. A note on the Hajek-Renyi inequality for associated random variables. *Statistics and Probability Letters,* **78**, pp. 885-889.

Takada, Y. 1992. A sequential procedure with asymptotically negative regret for estimating a normal mean. *Ann. Statist.*, **20**, pp. 562-569.

Takada, Y. 1993. Three stage estimation procedure of the multivariate normal mean. *Indian J. Statist.*, **55**, Series B, pp. 124-129.

Tukey, J. W. 1960. A Survey of Sampling From Contaminated Distributions. In: *Contributions to Probability and Statistics*. Olkin, I. (Ed). Standford University Press: Stanford, California.

Wald, A. 1947. *Sequential Analysis*. Wiley, NewYork.

Wang, Y. H. 1973. Sequential estimation of the scale parameter of the Pareto distribution. *Commun. Statist.* **A2**, pp. 145-154.

Wang, Y. H. 1981. Sequential estimation of the mean of the multi-normal population. *J. Amer. Statist. Assoc.* **75**, pp. 977-983.

Wasan, M. T. 1964. Sequential optimum procedures for unbiased estimation of a binomial parameter. *Technometrics,* **6,** pp. 259-272.

Whitehead, J. 1983. *The Design and Analysis of Sequential Clinical Trials.* Halsted Press, New York.

Whitehead, J. 1991. *Handbook of Sequential Analysis*. Chapter 26, Marcel Dekker, New York.

Whitehead, J. 1992. *The Design and Analysis of Sequential Clinical Trials*. Ellis Horwood, London.

Whitehead, J. 1994. *Sequential methods based on the boundaries approach for the clinical comparison of survival times. Statistics in Medicine* **13,** pp. 1357-1368.

Whitehead, J. 1997. *The Design and Analysis of Sequential Clinical Trials*, Revised second edition, *Statistics in Practice*. Wiley, New York.

Whitehead, J.2001. *Handbook in Clinical Oncology*. Chapter 12, Marcel Dekker, New York.

Whitehead, J. 2005. Analysis of clinical trials using SAS (R): A Practical Guide. Wiley, New York.

Wilcox, R. R. (2001). *Fundamentals of Modern Statistical Methods*. Springer-Verlag, New York.

Wilcox, R. R. (1998). *How many discoveries have been lost by ignoring modern statistical methods? American Psychologist*, **53**, pp. 300-314.

Wolfowitz, J. 1946. On sequential binomial estimation. *Ann. Math. Statist.* 17, pp. 489-493.

Woodroofe, M. 1977. Second order approximations for sequential point and interval estimation. *Ann. Statist.* 5, pp. 984-995.

Yousef, A. S., Kimber, A. C., Hamdy, H. I. (2009). Sensitivity of Normal-based triple sampling sequential point estimation to the normality assumption. Southampton, UK, Southampton *Statistical Sciences Research Institute*, 23pp. (S3RI Methodology Working Papers, M09/06).

Zacks, S. 1966. Sequential estimation of the mean of a lognormal distribution having a prescribed proportional closeness. *Ann. Math. Statist.* 37, pp. 1688-1696.