7-20-2018

Multi-Stage Estimation Methods with Termination Defined Via Gini's Mean Difference or Mean Absolute Deviation

Jun Hu
University of Connecticut - Storrs, jun.hu@uconn.edu

Multi-Stage Estimation Methods with Termination Defined Via Gini's Mean Difference or Mean Absolute Deviation

Jun Hu, Ph.D.

University of Connecticut, 2018

ABSTRACT

This thesis consists of multi-stage methodologies handling two fundamental estimation problems. These are (i) *fixed-width confidence interval estimation* (FCIE), and (ii) *minimum risk point estimation* (MRPE) problems for the unknown mean μ of a normal distribution whose variance σ^2 is also assumed unknown.

We first develop purely sequential estimation methodologies for both FCIE and MRPE problems. New stopping rules are constructed by replacing the sample variance with appropriate multiples of *Gini's mean difference* (GMD) and *mean absolute deviation* (MAD) in defining the conditions for boundary crossing. A number of asymptotic first-order consistency, efficiency, and risk efficiency properties associated with these new estimation strategies has been investigated. These are followed by summaries obtained from extensive sets of simulations by drawing samples from (i) normal universes or (ii) mixture-normal universes where samples may be reasonably treated as observations from a normal universe in a large majority of simulations. We also include illustrations using sales data and horticulture data.

Jun Hu, University of Connecticut, 2018

By revisiting Stein (1945,1949) as well as Mukhopadhyay and Duggan (1997), we then move on to propose new two-stage estimation methodologies under both MRPE and FCIE configurations for a normal mean μ when a lower bound of variance $\sigma_L^2\,(0 < \sigma_L < \sigma)$ is known to us. Again, unbiased estimators based on sample standard deviation, GMD and MAD are constructed to define the final sample sizes. The new methodologies prove to enjoy both asymptotic first-order and second-order properties, followed by simulated performances. Real data illustrations of horticulture data are also included.

Next, we further explore the asymptotic second-order approximations for the regret function associated with the purely sequential MRPE methodologies, providing a general structure.

Overall, we empirically feel confident that our newly developed GMD-based or MAD-based multi-stage estimation methodologies are more robust for practical purposes when we compare them with the sample variance-based methodologies respectively, especially when suspect outliers may be expected. We conclude with some interesting directions of future research work that we can follow to make our multi-stage estimation methodologies more widely applicable for a lot of inference problems.

Multi-Stage Estimation Methods with Termination Defined Via Gini's Mean Difference or Mean Absolute Deviation

Jun Hu

B.Sc., Wuhan University, China, 2014

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2018

APPROVAL PAGE

Doctor of Philosophy Dissertation

Multi-Stage Estimation Methods with Termination Defined Via Gini's Mean Difference or Mean Absolute Deviation

Presented by

Jun Hu, B.Sc.

Major Advisor ____

Nitis Mukhopadhyay

Associate Advisor ____

Joseph Glaz

Associate Advisor ____

Vladimir Pozdnyakov

University of Connecticut

2018

ACKNOWLEDGEMENTS

In the blink of an eye, it has been four years since I joined the Department of Statistics at the University of Connecticut in Fall 2014. Standing here in retrospect, I would like to express my sincere and foremost gratitude to my mother Xuefang Zhong and my father Lifeng Hu for their infinite moral and emotional support in every stage of my life. I am also grateful to a list of my other family members and friends throughout the four years of my Ph.D. program, without whose support I could have never ever gone this far, with a special mention to Aiqing Zhong, Yuqing Zhong and Xiangyi Li.

I would also like to thank my advisor Professor Nitis Mukhopadhyay from the bottom of my heart. It was so fortunate for me to have taken Prof. Mukhopadhyay's math stat courses when I came to UConn, where I was very impressed by his passion for Statistics. His comprehensive knowledge in Statistics as well as wide-ranged collection of anecdotes always kept me spellbound. I believe it was the best decision, which was not difficult to make at all, to select Prof. Mukhopadhyay as my academic advisor. Over these past four years, his abundant experience has kept me from getting detoured along the way; and his warm encouragements have eliminated most frustrations in my research. I also highly appreciated Prof. Mukhopadhyay's dedication in his teaching and research as well as his commitment to students. This has firmed my desire to plunge into academia further upon my graduation. I cherished the opportunity to work with Prof. Mukhopadhyay under his guidance, and hope to continue working with him in the future.

My sincere gratitude also goes to Prof. Ming-Hui Chen, who played the role as my program advisor during my initial year at UConn. I benefited a lot from his advise about how to get acclimated to the program. I would like to extend my gratitude to Prof. Joseph Glaz and Prof. Vladimir Pozdnyakov, my associate advisors for their continuous support in research. They were always ready to provide help whenever I had any questions. I would also like to especially thank Prof. Richard Vitale and Prof. Zhiyi Chi for their time to take charge in my general exam. I am also grateful to all the other professors who have taught me during the four years of my Ph.D. program.

Many thanks are owed to the following department staff: Megan Petsa, Tracy Burke and Anthony Luis for their dedicated service over these past few years, which have made our lives much easier and let us feel at home.

And finally, last but not least, I would like to thank a lot of my fellow graduate students, without whom my time at UConn would have never been full of joy and memories. I will cherish our friendship forever.

Table of Contents

Chapter 1

Introduction

In this thesis, we revisit both of the *fixed-width confidence interval estimation* (FCIE) and *minimum risk point estimation* (MRPE) problems under multi-stage stopping rules for an unknown normal mean μ when the population variance σ^2 also remains unknown. We propose to replace the sample standard deviation with appropriate multiples of either *Gini's mean difference* (GMD) or *mean absolute deviation* (MAD) in defining the stopping boundaries. We do in order to accommodate possibilities of encountering outlying observations even though we consider a normal population distribution under investigation.

1.1. Literature Review

The sampling designs of multi-stage methodologies have been growing rapidly over the past decades. Stein (1945,1949) had put forward his ingenious two-stage strategy for gathering data. His fundamental contribution solved the fixed-width confidence interval estimation (FCIE) problem for a normal mean μ with preassigned coverage probability when the variance σ^2 is unknown. Later, Anscombe (1952,1953) developed breakthrough large-sample theories of sequential estimation. In his 1952 paper, Anscombe gave the fundamental formulation of the random central limit theorem. His 1953 paper proposed a purely sequential sampling strategy for the same problem. Ray (1957) broadened Anscombe's (1952,1953) ideas by modifying the stopping boundary.

Chow and Robbins (1965) developed pioneering asymptotic theories of FCIE problems in the distribution-free case and laid the foundation for ground-breaking results. These eventually led to a broad class of crucial nonlinear renewal theoretic tools for arriving at asymptotic second-order approximations developed by Woodroofe (1977,1982) and Lai and Siegmund (1977,1979).

In order to reduce the oversampling problem associated with Stein's (1945,1949) two-stage methodology, Mukhopadhyay (1980) designed a modification giving rise to a surprising two-stage estimation strategy with the asymptotic efficiency property. This direction led Ghosh and Mukhopadhyay (1981) to give the foundation of the notion of *asymptotic second-order efficiency* property. Mukhopadhyay and Duggan (1997,1999) revisited this problem when a positive lower bound of the population variance σ^2 is known and proved that their two-stage methodology enjoyed both asymptotic first-order and second-order efficiency properties.

In a parallel path, Robbins (1959) introduced the pioneering formulation for minimum risk point estimation (MRPE) problems for a normal mean μ. He put forward a purely sequential stopping rule to estimate μ. Starr (1966) and Starr and Woodroofe (1969) subsequently broadened Robbins' (1959) methodology in more details and concluded a number of interesting asymptotic properties enjoyed by that procedure. Second-order approximations were summarized by Woodroofe (1977,1982), Siegmund (1985), Ghosh et al. (1997) among other sources. Mukhopadhyay and Duggan (1997,1999) also incorporated their new Stein-type (1945,1949) methodology in this area.

In traditional normal problems, the boundary crossing criteria often utilized the sample variance unbiasedly estimating the population variance σ^2 at every step of the way. In a different direction, in lieu of using the sample variance, some researchers had looked into other estimators such as those based on GMD and MAD, both explicitly defined in (2.3.1).

GMD was originally proposed by Gini (1914,1921).

Later, Nair (1936) developed the standard error of GMD so that one could specify an unbiased estimator for the population variance based on GMD and its standard error. This was way before Hoeffding (1948,1961) came up with his pathbreaking theory of U-statistics. The literature on sequential U-statistics is very broad. One may review from Lee (1990), Mukhopadhyay and Vik (1985), Sen (1981,1985), Jurečkovā and Sen (1996) and other sources.

Revisiting Downton (1966) and Barnett et al. (1967), David (1968) used a linear function of order statistics as an estimator of the standard deviation of a normal distribution which turned out to be essentially GMD. See also Sen (1986). Barnett et al. (1967) established an unbiased estimator of the standard deviation based on GMD (G) and investigated the role of a sample mean standardized with G as a competitor of the usual Student's t test when the sample size is fixed. Mukhopadhyay and Chattopadhyay (2013) recently revisited this problem charting interesting new directions.

Yitzhaki (2003) studied properties of GMD as a superior measure of variability than the usual sample variance in some non-normal cases. Meanwhile, MAD had been adopted as an estimator of the standard deviation, too. Herrey (1965) derived the expression of the variance of MAD using results from Helmert (1876). Herrey (1965) focused on MAD and constructed confidence intervals based on MAD instead of the sample standard deviation. Babu and Rao (1992) worked on expansions for statistics involving MAD. We may note that MAD is widely used in portfolio optimization models. See Konno and Yamazaki (1991) where MAD was preferred. One may also refer to Markowitz (1959).

In Chattopadhyay and Mukhopadhyay (2013), both GMD and MAD were used to construct two-stage FCIE methodologies for a normal mean μ in the presence of suspect outliers and these were compared with the customary Stein-type (1945,1949) two-stage methodology based on the sample variance. We have found interesting recent trends in sequential estimation problems for Gini's index itself from Chattopadhyay and De (2016) and De and Chattopadhyay (2017). Mukhopadhyay and Hu (2017,2018) and Hu and Mukhopadhyay (2018) have extended the application of GMD and MAD into more multi-stage sampling strategies for FCIE and MRPE problems.

We briefly mention that there is a large volume of literature available on multi-stage estimation problems and associated sampling strategies. More generally, one may accomplish a broad ranging review in this field by combining selected parts of interest from many resources including the following monographs and references therein: Sen (1981,1985), Woodroofe (1982), Siegmund (1985), Ghosh and Sen (1991), Mukhopadhyay and Solanky (1994), Jurečkovā and Sen (1996), Ghosh et al. (1997), Govindarajulu (2004), Mukhopadhyay et al. (2004a), Mukhopadhyay and de Silva (2009) and Zacks (2009,2017).

1.2. Thesis Outline

In this thesis we introduce multi-stage methodologies associated with two fundamental estimation problems, namely

(i) the fixed-width confidence interval estimation (FCIE), and

(ii) the minimum risk point estimation (MRPE).

We mainly focus on the purely sequential and two-stage sampling designs with the purpose to estimate an unknown normal mean.

Chapter 2 begins with basic formulations of FCIE and MRPE for an unknown mean μ in a $N(\mu, \sigma^2)$ population where σ^2 is also assumed unknown. Alternative robust estimators of σ^2 other than the sample variance are introduced, including appropriate functions of GMD and MAD.

In Chapter 3, we develop purely sequential FCIE and MRPE methodologies with estimators of σ^2 based on GMD and MAD as well as the customary sample variance in defining the stopping boundary conditions. We discuss a series of interesting asymptotic first-order properties as well as the limited robustness, followed by extensive sets of simulations as a reasonable validation. Real data analyses are also provided to address the possible application of our newly proposed methodologies in the fields of economics and agriculture with illustrations using (i) sales data and (ii) horticulture data. The chapter largely comes from the publication, Mukhopadhyay and Hu (2017).

In the contexts of both FCIE and MRPE problems, Chapter 4 introduces newly proposed GMD-based and MAD-based two-stage estimation methodologies, assuming that a positive lower bound σ_L^2 for the variance σ^2 is known to us in the spirit of Mukhopadhyay and Duggan (1997,1999). We lay out a number of desirable asymptotic first-order and second-order properties. And simulated performances as well as illustrations with real datasets are also presented. Chapter 4 is based on the publication, Mukhopadhyay and Hu (2018).

In Chapter 5, we again focus on the purely sequential MRPE methodologies based on robust estimators of σ. However, we provide a general structure with appropriate sufficient conditions on such estimators, which will allow us in general to claim that the associated methodologies would enjoy both asymptotic first-order and second-order asymptotic properties. The chapter is based on Hu and Mukhopadhyay (2018), which has been submitted for publication.

Chapter 6 gives a brief summary of our work. In Chapter 7, we provide a lot of possible directions that we can follow for future research to make our methodologies more applicable for practical purposes.

Chapter 2

Formulations for Two Fundamental Estimation Problems: Fixed-Width Confidence Intervals and Minimum Risk Point Estimation

More than forty years ago, Ghosh and Mukhopadhyay (1976) gave a broad review of both sequential and purely sequential methodologies for constructing fixed-width confidence interval and minimum risk point estimators for unknown μ in a $N(\mu, \sigma^2)$ population having σ^2 also unknown. In the context of our present investigation, that paper remains relevant even today.

We begin with a sequence of independent observations $X_1, X_2, ...$ from a $N(\mu, \sigma^2)$ population with $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$, both parameters unknown. Having recorded $X_1, ..., X_n,\ n \geq 2$, we denote the customarily used unbiased estimators for μ and σ^2 as follows:

$$\text{Sample mean:} \quad \overline{X}_n = n^{-1}\Sigma_{i=1}^n X_i$$

$$\text{Sample variance:} \quad S_n^2 = (n-1)^{-1}\Sigma_{i=1}^n (X_i - \overline{X}_n)^2.$$

In Sections 2.1-2.2, we respectively lay out the fixed-width confidence interval estimation methodology and the minimum risk point estimation methodology.

2.1. Fixed-Width Confidence Interval Estimation (FCIE)

In order to estimate the mean μ, first we aim to construct a confidence interval with *preassigned* (i) fixed-width $2d(> 0)$ and (ii) the confidence coefficient at least $1 - \alpha$, or approximately at least $1-\alpha$, $0<\alpha<1$. The customary fixed-width confidence interval

$$J_n = [\overline{X}_n \pm d] \tag{2.1.1}$$

is considered, which additionally must satisfy the requirement that $P_{\mu,\sigma}\{\mu \in J_n\}$ is at least $1-\alpha$, or approximately $1-\alpha$, with $0<\alpha<1$ preassigned. Dantzig (1940) proved that there did not exist any fixed-sample-size procedure to solve the problem. Pathbreaking papers that provided a solution to this problem were due to Stein (1945,1949) by invoking the ingenious two-stage sampling methodology.

From the problem specifications, one has:

$$P_{\mu,\sigma}\{\mu \in J_n\} = P_{\mu,\sigma}\{\left|\overline{X}_n - \mu\right| \le d\} = 2\Phi(\sqrt{n}d/\sigma) - 1,$$

so that

$$P_{\mu,\sigma}\{\mu \in J_n\} \ge 1-\alpha \tag{2.1.2}$$

provided that n is the smallest integer satisfying:

$$\sqrt{n}d/\sigma \ge a \Leftrightarrow n \ge a^2\sigma^2/d^2 = C, \text{ say.}$$

Here, we write $a \equiv z_{\alpha/2}$, the upper $100(\alpha/2)\%$ point of a standard normal distribution. Thus, we define the optimal fixed sample size, had σ^2 been known, as follows:

$$C \equiv C(d) = a^2\sigma^2/d^2, \tag{2.1.3}$$

by tacitly disregarding the fact that C may not be an integer.

2.2. Minimum Risk Point Estimation (MRPE)

On the other hand, Robbins (1959) proposed his original and fundamental formulation of a MRPE problem for the normal mean μ. We will consider the loss function given by the squared error loss (SEL) due to estimation of μ with $\overline{X}_n$ plus linear cost of sampling:

$$L_n \equiv L_n\left(\mu, \overline{X}_n\right) = A\left(\overline{X}_n - \mu\right)^2 + cn, \text{ where } A \text{ and } c \text{ are both known.} \tag{2.2.1}$$

Here, $A(> 0)$ is an appropriate weight function and $c(> 0)$ is the cost of each observation and "n" obviously indicates a sample size.

The risk function associated with (2.2.1) is then given by:

$$R_n(c) \equiv E_{\mu,\sigma}\left[L_n\left(\mu, \overline{X}_n\right)\right] = A\sigma^2 n^{-1} + cn. \tag{2.2.2}$$

We can thus obtain the optimal fixed sample size, $n^* = n^*(c)$, had σ^2 been known, which minimizes the risk function $R_n(c)$ from (2.2.2). We have:

$$n^* \equiv n^*(c) = \sigma\sqrt{A/c}, \tag{2.2.3}$$

again by tacitly disregarding the fact that n^* may not be an integer.

2.3. Introducing Gini's Mean Difference (GMD) and Mean Absolute Deviation (MAD)

One should note that in (2.1.3) and (2.2.3), σ^2 is actually assumed unknown. Therefore, it is essential for us to estimate σ^2 by updating its estimator at various stages of the methodologies as needed. A customarily used unbiased estimator of σ^2 is the sample variance, which, however, in certain scenarios, has its own drawback in estimating the population variance. For example, when skewness appears pronounced due to the existence of possible outliers even though the parent population is assumed normal, sample variance may not estimate σ^2 well. In nonsequential portfolio theory, Markowitz (1959) suggested using semivariance as a substitute of the sample variance.

Back to multi-stage estimation problems, Mukhopadhyay (1982) opened the possibilities of using estimators of σ^2 other than the sample variance. Recall that Chattopadhyay and Mukhopadhyay (2013), estimators based on GMD and MAD were used for constructing two-stage fixed-width confidence interval methodologies for a normal mean μ in the presence of suspect outliers and these were compared with the customary Stein-type (1945,1949) two-stage methodology based on the sample variance.

In this thesis, we therefore propose to push similar ideas lot further and utilize alternative more robust estimators for σ^2 based on (i) GMD and (ii) MAD, as needed, formally defined in (2.3.1).

Having recorded $X_1, ..., X_n$ from a $N(\mu, \sigma^2)$ population, we define GMD and MAD as follows:

$$\begin{aligned} &\text{(i)} \quad \text{GMD: } G_n = \binom{n}{2}^{-1} \Sigma\Sigma_{1\le i<j\le n} \left|X_i - X_j\right|; \\ &\text{(ii)} \quad \text{MAD: } M_n = n^{-1}\Sigma_{i=1}^{n} \left|X_i - \overline{X}_n\right|. \end{aligned} \tag{2.3.1}$$

Unbiased estimators for the population variance σ^2 based on GMD or MAD were proposed in Nair (1936) and Herrey (1965) respectively. In the spirit of notation used in Chattopadhyay and Mukhopadhyay (2013), we denote two unbiased estimators as well as the sample variance by $U_{n;i}^2$, $i = 0, 1, 2$, for σ^2 as follows:

$$(0)\ \widehat{\sigma}^2 \equiv U_{n;0}^2 = S_n^2;\ (1)\ \widehat{\sigma}^2 \equiv U_{n;1}^2 = c_{n;1}^{-2} G_n^2;\ (2)\ \widehat{\sigma}^2 \equiv U_{n;2}^2 = c_{n;2}^{-2} M_n^2, \tag{2.3.2}$$

where we denote

$$\begin{aligned} c_{n;1} &= \left\{ \tfrac{4}{\pi n(n-1)} \left(\tfrac{\pi}{3}(n+1) + 2(n-2)\sqrt{3} + n^2 - 5n + 6 \right) \right\}^{1/2}, \\ c_{n;2} &= \left\{ \left(\tfrac{\pi}{2} + \arcsin(\tfrac{1}{n-1}) - n + \sqrt{n(n-2)} \right) \tfrac{2(n-1)}{\pi n^2} + \tfrac{2(n-1)}{\pi n} \right\}^{1/2}. \end{aligned} \tag{2.3.3}$$

We also construct unbiased estimators for σ based on GMD or MAD as well as the sample standard deviation as follows:

$$\begin{aligned} &(0) \quad \widehat{\sigma} \equiv V_{n;0} = \frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)} S_n; \\ &(1) \quad \widehat{\sigma} \equiv V_{n;1} = \tfrac{\sqrt{\pi}}{2} G_n; \\ &(2) \quad \widehat{\sigma} \equiv V_{n;2} = \sqrt{\tfrac{\pi n}{2(n-1)}} M_n. \end{aligned} \tag{2.3.4}$$

One will find many more specific details in Chapters 3 and 4, respectively, dealing with the two fundamental multi-stage estimation problems, namely, (i) FCIE and (ii) MRPE for the normal mean. Throughout the text, we write $\phi(z) = (\sqrt{2\pi})^{-1} \exp(-z^2/2)$ and $\Phi(z) = \int_{-\infty}^{z} \phi(y) dy$, $z \in R$.

Chapter 3

Purely Sequential Methodologies Involving GMD or MAD

In this chapter, we propose the GMD-based and MAD-based purely sequential methodologies for both fixed-width confidence interval estimation (FCIE) and minimum risk point estimation (MRPE) problems. In comparison, the customary sample variance-based purely sequential methodology is also included. The chapter is mainly based on Mukhopadhyay and Hu (2017).

Section 3.1 develops purely sequential estimation strategies with stopping boundaries based on the sample variance, GMD or MAD (defined in 2.3.1-2.3.3) for the FCIE problems. Three main results have been summarized in Theorems 3.1.1-3.1.3, which show that our newly proposed purely sequential FCIE strategies enjoy asymptotic first-order efficiency and asymptotic consistency properties.

Section 3.2 develops purely sequential estimation strategies with stopping boundaries based on the sample standard deviation, GMD or MAD (defined in 2.3.1 and 2.3.4) for the MRPE problems. Three main results are summarized in Theorems 3.2.1-3.2.3, which show that our newly proposed purely sequential MRPE strategies enjoy asymptotic first-order efficiency and first-order risk efficiency.

In both Sections 3.1 and 3.2, we have presented summaries of data analysis obtained via extensive sets of simulations with random samples drawn from a normal as well as a mixture normal distribution (3.1.7). The mixture normal scenarios were so constructed that outlying observations could be expected, but the gathered data could overwhelmingly be treated as random samples from a normal universe for practical purposes. We are reassured that the new GMD-based and MAD-based purely sequential methodologies are more robust (Theorem 3.1.4) than the existing sample variance-based purely sequential methodologies. This happens in both problems when we may expect to encounter some outliers as we sample from a normal universe.

Section 3.3 includes illustrations of both FCIE and MRPE methodologies for the mean (μ) using two real datasets: (a) net profit or loss percentages (sales data) from department stores (McNair 1930; http://www.stat.ufl.edu/~winner/), and (b) number of days that seeds of marigold varieties 2 and 3 needed to flower (horticulture data) from Mukhopadhyay et al. (2004b).

The chapter ends with some brief overall thoughts in Section 3.4.

3.1. Fixed-Width Confidence Intervals

Under the fixed-width confidence interval formulation (2.1.1)-(2.1.3), we propose the following sampling procedure to determine the final sample size. Beginning with pilot data $X_1, ..., X_m$ of size m, $m \geq 2$, we record one additional observation at-a-time successively as needed until we stop according to the following stopping rule:

$$N_i \equiv N_i(d) = \inf\{n \geq m : n \geq a^2 U_{n;i}^2 / d^2\}, i = 0, 1, 2. \tag{3.1.1}$$

Observe that the index $i = 0, 1, 2$ respectively corresponds to the customary sample variance-based, the GMD-based or the MAD-based methodology with $U_{n;i}^2, i = 0, 1, 2$, defined via (2.3.2)-(2.3.3).

That is, for $i = 0, 1, 2$, if $m \geq a^2 U_{m;i}^2 / d^2$ is satisfied, we do not take any additional observation and the sample size is $N_i = m$. Otherwise, we record one more observation and obtain updated $U_{m+1;i}^2$ in order to check with the stopping rule (3.1.1). We terminate sampling at the first time $N_i = n(\geq m)$ such that $n \geq a^2 U_{n;i}^2/d^2$ occurs. Finally, having observed the full dataset $\{N_i, X_1, ..., X_m, X_{m+1}, ..., X_{N_i}\}$, we construct the fixed-width confidence interval for μ as follows:

$$J_{N_i} = [\overline{X}_{N_i} \pm d]. \tag{3.1.2}$$

We should add that the customary sample variance-based methodology was first developed by Anscombe (1952,1953), Ray (1957), and Chow and Robbins (1965). In Anscombe (1953), he used "a", the upper $100(\alpha/2)\%$ point of a standard normal distribution and his stopping rule was similar to (3.1.1) with $i = 0$. Ray (1957) used the upper $100(\alpha/2)\%$ point of the Student's t_{n-1} distribution instead of "a" in defining the stopping boundary $(= t_{n-1,\alpha/2}^2 S_n^2/d^2)$ in (3.1.1). Obviously, $\lim_{n\to\infty} t_{n-1,\alpha/2} = a$. One may refer to Mukhopadhyay (2010), Gut and Mukhopadhyay (2010), and also look at the references included therein. Chow and Robbins (1965) used a general sequence of positive constants $\{a_n; n \geq 2\}$ instead of "a" in defining their stopping boundary $(= a_n^2 S_n^2/d^2)$ in (3.1.1) where $\lim_{n\to\infty} a_n = a$.

For the purely sequential methodologies $\{N_i, J_{N_i}\}$ from (3.1.1)-(3.1.2), we can clearly claim that $P_{\mu,\sigma}\{N_i < \infty\} = 1$ and $N_i \uparrow \infty$ w.p.1 as $d \downarrow 0$, $i = 0, 1, 2$.

3.1.1. Some Useful Lemmas

We begin with a number of technical lemmas (Lemmas 3.1.1-3.1.4) which will be helpful in justifying some of the more substantial properties.

Lemma 3.1.1. *The statistics* $\{U^2_{k;i}, k \leq n\}, i = 0, 1, 2,$ *are distributed independently of* $\{\overline{X}_n, n \geq m\}$.

Proof: We fix $\sigma = \sigma_0(> 0)$ and $n(\geq m)$, but otherwise they remain arbitrary. Now, one can show easily that $\{U^2_{k;i}, k \leq n\}$ consists of ancillary statistics for μ and $\overline{X}_n$ is a complete and sufficient statistic for μ in the model $N(\mu, \sigma^2_0)$. Thus, by Basu's (1955) theorem, we immediately claim that $\{U^2_{k;i}, k \leq n\}$ and $\overline{X}_n$ are independent under the model $N(\mu, \sigma^2_0)$ for every fixed σ^2_0. But, since σ_0 is fixed but arbitrary, it means that $\{U^2_{k;i}, k \leq n\}$ and $\overline{X}_n$ are independently distributed. ■

Lemma 3.1.2. $U^2_{n;i} \stackrel{P_{\mu,\sigma}}{\to} \sigma^2$ *as* $n \to \infty$ *for* $i = 0, 1, 2$.

Proof: With $i = 0$, it is obvious that $S^2_n \stackrel{P_{\mu,\sigma}}{\to} \sigma^2$ as $n \to \infty$.

With $i = 1$, we see that G_n from (2.3.1) is a U-statistic so that $G_n \stackrel{P_{\mu,\sigma}}{\to} E_{\mu,\sigma}\left[|X_1 - X_2|\right]$ as $n \to \infty$ in view of Hoeffding (1948,1961). See also Lee (1990). Since $X_1 - X_2 \sim N(0, 2\sigma^2)$, we have $E_{\mu,\sigma}\left[|X_1 - X_2|\right] = 2\sigma/\sqrt{\pi}$ and thus $G^2_n \stackrel{P_{\mu,\sigma}}{\to} 4\sigma^2/\pi$ as $n \to \infty$. Also, $c^{-2}_{n;1} \to \pi/4$ as $n \to \infty$. Thus, $U^2_{n;1} \stackrel{P_{\mu,\sigma}}{\to} \sigma^2$ as $n \to \infty$.

With $i = 2$, we work with M_n from (2.3.1) and observe that with probability 1 (w.p.1):

$$n^{-1}\Sigma^n_{i=1}\left|X_i - \mu\right| - \left|\overline{X}_n - \mu\right| \leq M_n \leq n^{-1}\Sigma^n_{i=1}\left|X_i - \mu\right| + \left|\overline{X}_n - \mu\right|.$$

Obviously,

$$\left|\overline{X}_n - \mu\right| \stackrel{P_{\mu,\sigma}}{\to} 0,\ n^{-1}\Sigma^n_{i=1}\left|X_i - \mu\right| \stackrel{P_{\mu,\sigma}}{\to} E_{\mu,\sigma}\left[|X_1 - \mu|\right] = \sigma\sqrt{2/\pi},$$

as $n \to \infty$. Therefore, $M^2_n \stackrel{P_{\mu,\sigma}}{\to} 2\sigma^2/\pi$ as $n \to \infty$. Combining this with the fact that $c^{-2}_{n;2} \to \pi/2$ as $n \to \infty$, one immediately concludes that $U^2_{n;2} \stackrel{P_{\mu,\sigma}}{\to} \sigma^2$ as $n \to \infty$. ■

Lemma 3.1.3. *For any arbitrary* $p > 1$, *we have* $E_{\mu,\sigma}\left[\sup_n U^p_{n;i}\right] < \infty$, $i = 0, 1, 2$.

Proof. With $i = 1$, we have $U^p_{n;1} = c^{-p}_{n;1} G^p_n$. Clearly, $c^{-p}_{n;1} \to (\pi/4)^{p/2} < 1$ as $n \to \infty$, so that there exists large enough n_1 and $c^{-p}_{n;1} < 1$ for all $n > n_1$. Let us denote

$$k = \max\left\{c^{-p}_{1;1}, ..., c^{-p}_{n_1;1}, 1\right\}$$

and we can express:

$$\begin{aligned} &\sup_n U^p_{n;1} \\ &\quad \le k \sup_n \left\{\binom{n}{2}^{-1} \Sigma_{1\le i<j\le n} \left|X_i - X_j\right|\right\}^p \\ &\quad \le 4^p k \left\{\sup_n n^{-1} \Sigma_{1\le i\le n} \left|X_i - \mu\right|\right\}^p . \end{aligned} \tag{3.1.3}$$

We know that $\left|X_i - \mu\right|$, $i = 1, 2, ...$ are independent and identically distributed (i.i.d.) and belong to L^p, $p > 1$, and so does $\sup_n \left\{n^{-1}\Sigma_{1\le i\le n} \left|X_i - \mu\right|\right\}$ by Wiener's (1939) ergodic theorem. Hence, the result follows from (3.1.3).

With $i = 2$, we have $U^p_{n;2} = c^{-p}_{n;2} M^p_n$. Again, clearly, $c^{-p}_{n;2} \to (\pi/2)^{p/2}$ as $n \to \infty$, so that there exists large enough n_2 and $c^{-p}_{n;2} < 2^p$ for all $n > n_2$. Let us denote

$$l = \max\left\{c^{-p}_{1;2}, ..., c^{-p}_{n_2;2}, 2^p\right\}$$

and we can express:

$$\begin{aligned} &\sup_n U^p_{n;2} \\ &\quad \le l \sup_n \left[n^{-1}\Sigma_{1\le i\le n} \left|X_i - \overline{X}_n\right|\right]^p \\ &\quad \le l \left\{\sup_n \left[n^{-1}\Sigma_{1\le i\le n} \left|X_i - \mu\right| + \left|\overline{X}_n - \mu\right|\right]\right\}^p , \end{aligned}$$

in the spirit of (3.1.3). Again, $\sup_n \left\{ n^{-1} \Sigma_{1 \le i \le n} |X_i - \mu| + \left| \overline{X}_n - \mu \right| \right\}$ belongs to L^p by invoking Wiener's (1939) ergodic theorem so that we have $E_{\mu,\sigma} \left[\sup_n U_{n;2}^p \right] < \infty$.

With $i = 0$, one may refer to Chow and Robbins (1965) or look at another source. ■

Lemma 3.1.4. *Under the purely sequential estimation strategy* $(N_i, \overline{X}_{N_i})$ *from (3.1.1)-(3.1.2), for all fixed* $\mu, \sigma, d,$ *and* α, *we conclude the following properties*:

(i) $E_{\mu,\sigma}[N_i] < \infty, V_{\mu,\sigma}[N_i] < \infty$;

(ii) $E_{\mu,\sigma}[\overline{X}_{N_i}] = \mu, V_{\mu,\sigma}[\overline{X}_{N_i}] = \sigma^2 E_{\mu,\sigma}[N_i^{-1}]$;

for $i = 0, 1, 2$.

Proof:

P*art* (i): We surely have $P_{\mu,\sigma}\{N_i < \infty\} = 1$. The stopping rule (3.1.1) provides the inequality (w.p.1):

$$N_i - 1 \le m - 1 + a^2 U_{N_i - 1;i}^2 / d^2,$$

which gives (w.p.1):

$$N_i^2 \le m^2 + 2m \tfrac{a^2}{d^2} \sup_n U_{n;i}^2 + \tfrac{a^4}{d^4} \sup_n U_{n;i}^4, \tag{3.1.4}$$

$i = 0, 1, 2$. Thus, Lemma 3.1.3 and (3.1.4) show that $E_{\mu,\sigma}[N_i^2] < \infty$ with $p = 4$. Part (i) follows.

P*art* (ii): The event $\{N_i = n\}$ depends only on the vector $U_i^* = \left(U_{m+1;i}^2, U_{m+2;i}^2, ..., U_{n;i}^2 \right)$ whereas U_i^* is independent of $\overline{X}_n$ in view of Lemma 3.1.1. As a result, the event $\{N_i = n\}$ and $\overline{X}_n$ are independent for every fixed $n \ge m$.

Thus, we have:

$$E_{\mu,\sigma}[\overline{X}_{N_i}]$$
$$= \Sigma_{n=m}^{\infty} E_{\mu,\sigma}[\overline{X}_{N_i}|N_i = n] P_{\mu,\sigma}\{N_i = n\}$$
$$= \Sigma_{n=m}^{\infty} E_{\mu,\sigma}[\overline{X}_n] P_{\mu,\sigma}\{N_i = n\}$$
$$= \mu.$$

Similarly, we can express:

$$E_{\mu,\sigma}[\overline{X}_{N_i}^2] = \mu^2 + \sigma^2 E_{\mu,\sigma}\left[N_i^{-1}\right],$$

$i = 0, 1, 2$. Part (ii) follows. ■

3.1.2. Asymptotic First-Order Properties of the Purely Sequential Methodology (3.1.1)-(3.1.2)

In this section, we lay down a number of interesting first-order asymptotic properties in the form of Theorems 3.1.1-3.1.3 associated with the estimation strategies (N_i, J_{N_i}) proposed in (3.1.1)-(3.1.2), $i = 0, 1, 2$.

Theorem 3.1.1. *For the stopping time N_i defined by the purely sequential estimation strategy (3.1.1)-(3.1.2), for all fixed μ, σ, d, and α, we have $Q_i \equiv \sqrt{N_i}(\overline{X}_{N_i} - \mu)/\sigma \sim N(0, 1)$, $i = 0, 1, 2$.*

Proof: For all $z \in R$, we have:

$$P_{\mu,\sigma}\{Q_i \leq z\}$$
$$= \Sigma_{n=m}^{\infty} P_{\mu,\sigma}\{Q_i \leq z \mid N_i = n\} P_{\mu,\sigma}\{N_i = n\}$$

$$= \Sigma_{n=m}^{\infty} P_{\mu,\sigma}\{\sqrt{n}(\overline{X}_n - \mu)/\sigma \le z\} P_{\mu,\sigma}\{N_i = n\}$$

$$= \Sigma_{n=m}^{\infty} \Phi(z) P_{\mu,\sigma}\{N_i = n\},$$

which simplifies to $\Phi(z)$, $i = 0, 1, 2$. ■

Theorem 3.1.2. *For the stopping time N_i defined by the purely sequential estimation strategy (3.1.1)-(3.1.2), for all fixed μ, σ, and α, we have:*

$$\lim_{d \to 0} E_{\mu,\sigma}[N_i/C] = 1,\ i = 0, 1, 2\ [\textit{Asymptotic First-Order Efficiency}], \tag{3.1.5}$$

with C defined by (2.1.3).

Proof. We prove the theorem along the lines of Chow and Robbins (1965). Let us denote $y_{n;i} = U_{n;i}^2/\sigma^2$, and we know from Lemma 3.1.2 that $\lim_{n\to\infty} y_{n;i} = 1$ w.p.1. Denote $f(n) = n$ so that $\lim_{n\to\infty} f(n) = \infty$ and $\lim_{n\to\infty} f(n)/f(n-1) = 1$. Then, the stopping rule (3.1.1) would match the stopping rule of Chow and Robbins (1965). Using Lemma 3.1.3, we claim: $E_{\mu,\sigma}[\sup_n y_{n;i}] < \infty$ which immediately shows (3.1.5). ■

Theorem 3.1.3. *For the stopping time N_i and the proposed fixed-width confidence interval J_{N_i} defined by the purely sequential estimation strategy (3.1.1)-(3.1.2), for all fixed μ, σ, and α, we have:*

$$\lim_{d \to 0} P_{\mu,\sigma}\{\mu \in J_{N_i}\} = 1 - \alpha,\ i = 0, 1, 2\ [\textit{Asymptotic Consistency}]. \tag{3.1.6}$$

Proof. First, it is obvious that $\lim_{d\to 0} N_i/C = 1$ with C defined by (2.1.3). Next, we express:

$$P_{\mu,\sigma}\{\mu \in J_{N_i}\} = E_{\mu,\sigma}\left[2\Phi\left(N_i^{1/2} d/\sigma\right) - 1\right] \to 1 - \alpha,$$

as $d \to 0$, which follows from the dominated convergence theorem. ■

Theorems 3.1.2-3.1.3 give us reasonable assurances that the terminal sample sizes on an average and the confidence coefficients associated with the constructed confidence intervals will hover around C and $1-\alpha$ respectively when C (or d) is large (small). Small and moderate sample size performances of the purely sequential estimation strategies (N_i, J_{N_i}) from (3.1.1)-(3.1.2) will be summarized shortly via computer simulations.

3.1.3. Limited Robustness Properties of the Purely Sequential Methodology (3.1.1)-(3.1.2)

Here, we briefly address some of the issues surrounding potentially limited robustness properties associated with the GMD-based or MAD-based purely sequential estimation strategies (3.1.1)-(3.1.2) with $i = 1, 2$ when compared with the existing and customary strategy (3.1.1)-(3.1.2) with $i = 0$.

We deal with a situation when the normality of the parent population is satisfied, but one may encounter some outliers during the sequential process of gathering observations. Since an outlier is expected to be either extremely large or extremely small, it is reasonable to presume that the true population variance, say, σ'^2 would be pulled upward compared with σ^2, the assumed variance of the population without outliers.

As a consequence, if we fix the significance level $1-\alpha$ and the confidence interval's width $2d$, the true behind-the-scene optimal fixed sample size C will go up. That is, if C'(or C) corresponds to the optimal fixed sample size when the population produces some (or no) outliers, then it may be reasonable to expect $C' > C$.

For convenience, we simulate data from a population that will be expected to produce some outliers by generating observations from a *mixture normal* distribution specified by:

$$(1-\epsilon)N(\mu,\sigma_1^2)+\epsilon N(\mu,\sigma_2^2) \text{ where } \sigma_2>\sigma_1 \text{ and } 0<\epsilon<1 \text{ is small,} \tag{3.1.7}$$

that is, we view the normal population $N(\mu,\sigma_1^2)$ as mildly contaminated allowing it to possibly generate some outliers. Next, we summarize some interesting conclusions.

Theorem 3.1.4. *Under the model (3.1.7), with* $U_{n;i}^2$ *defined in (2.3.2)-(2.3.3),* $i=0,1,2$, *and for all fixed* $(\mu,\sigma_1,\sigma_2)\in R\times R^+\times R^+$, *we have:*

$$\sigma_1^2<E_{\mu,\sigma_1,\sigma_2}[U_{n;2}^2]<E_{\mu,\sigma_1,\sigma_2}[U_{n;1}^2]<E_{\mu,\sigma_1,\sigma_2}[U_{n;0}^2]<\sigma_2^2. \tag{3.1.8}$$

as $n\to\infty$.

Proof. Note that as $n\to\infty$, we first have:

$$E_{\mu,\sigma_1,\sigma_2}[U_{n;0}^2]=(1-\epsilon)\sigma_1^2+\epsilon\sigma_2^2. \tag{3.1.9}$$

Also, we obtain:

$$\begin{aligned} U_{n;1}^2 &\overset{P_{\mu,\sigma_1,\sigma_2}}{\rightarrow} \tfrac{\pi}{4}E_{\mu,\sigma_1,\sigma_2}^2[|X_1-X_2|]=\left[(1-\epsilon)^2\sigma_1+\epsilon(1-\epsilon)\sqrt{2(\sigma_1^2+\sigma_2^2)}+\epsilon^2\sigma_2\right]^2; \\ U_{n;2}^2 &\overset{P_{\mu,\sigma_1,\sigma_2}}{\rightarrow} \tfrac{\pi}{2}E_{\mu,\sigma_1,\sigma_2}^2[|X_1-\mu|]=\left[(1-\epsilon)\sigma_1+\epsilon\sigma_2\right]^2. \end{aligned} \tag{3.1.10}$$

Next, it follows that

$$\begin{aligned} &\left[(1-\epsilon)^2\sigma_1+\epsilon(1-\epsilon)\sqrt{2(\sigma_1^2+\sigma_2^2)}+\epsilon^2\sigma_2\right]^2 \\ &<(1-\epsilon)^2\sigma_1^2+\epsilon(1-\epsilon)(\sigma_1^2+\sigma_2^2)+\epsilon^2\sigma_2^2=(1-\epsilon)\sigma_1^2+\epsilon\sigma_2^2, \end{aligned} \tag{3.1.11}$$

and also,

$$\begin{aligned}&\left[(1-\epsilon)^2\sigma_1+\epsilon(1-\epsilon)\sqrt{2(\sigma_1^2+\sigma_2^2)}+\epsilon^2\sigma_2\right]^2\\ >&\left[(1-\epsilon)^2\sigma_1+\epsilon(1-\epsilon)(\sigma_1+\sigma_2)+\epsilon^2\sigma_2\right]^2=\left[(1-\epsilon)\sigma_1+\epsilon\sigma_2\right]^2.\end{aligned} \tag{3.1.12}$$

Based on the proof of Lemma 3.1.3, the dominated convergence theorem could be applied on $U_{n;1}^2$ and $U_{n;2}^2$. Thus, combining (3.1.9)-(3.1.12), the desired result follows. ■

Now, we may summarize by writing that mild robustness of the procedure (3.1.1)-(3.1.2) is felt through the spirit of taking into account what our Theorem 3.1.4 has to say. It indicates that when a population along the line of (3.1.7) is more prone to produce outliers, but yet we continue to assume that the population behaves like normal with variance σ_1^2, the estimators $U_{n;1}^2, U_{n;2}^2$, and $U_{n;0}^2$ tend to overestimate σ_1^2. As a result, they are no longer unbiased or even consistent for σ_1^2. Additionally, Theorem 3.1.4 also indicates that under procedures (3.1.1)-(3.1.2) the final sample sizes N_1, N_2 based on $U_{n;1}^2, U_{n;2}^2$ on an average could asymptotically fall below the average N_0 based on $U_{n;0}^2$. In other words, N_1, N_2 could be expected to stay nearer to the correspondingly presumed optimal fixed sample size, $C^* = a^2\sigma_1^2/d^2$, on an average than N_0 would.

3.1.4. Simulations and Data Analysis

In this section, we summarize performances of the purely sequential FCIE strategies (N_i, J_{N_i}), $i = 0, 1, 2$ from (3.1.1)-(3.1.2) respectively when sample sizes were small (≤ 50), moderate (100) or large (200). Throughout this section, we use the following system codes in (3.1.13) for identifying each specific methodology under implementation. We present summary performances in both normal cases and mixture normal cases (see 3.1.7) obtained via simulations.

$$
\begin{aligned}
i = 0 &\quad : \text{Sample variance-based methodology (3.1.1)-(3.1.2)} \\
i = 1 &\quad : \text{GMD-based methodology (3.1.1)-(3.1.2)} \\
i = 2 &\quad : \text{MAD-based methodology (3.1.1)-(3.1.2)}
\end{aligned}
\tag{3.1.13}
$$

3.1.4.1. Samples from a Normal Population

We generated pseudo random samples from a $N(5, 4)$ population, that is, we had fixed $\mu = 5$ and $\sigma = 2$. We also fixed $\alpha = 0.01, 0.05, 0.10$, $m = 10, 15, 20$, the optimal fixed sample size C over a wide range of values including $50, 100, 200$, and then determined the values of d accordingly by referring to (2.1.3). Table 3.1 represents this situation. One may however critique that we should have alternatively first fixed $d (= 0.2, 0.4, 0.6)$ and determined C accordingly. Table 3.2 represents that situation.

In Tables 3.1-3.2, we respectively present selected summaries in the two situations that we have mentioned when $\alpha = 0.05$ and $m = 10, 15$ for brevity alone even though we ran simulations with other values of C, α, and m.

Using $T\,(= 1000$, say) independent replications or runs, we estimated the average final sample size $\overline{n}$ with its estimated standard error $s(\overline{n})$, the minimum $\min(n)$, the lower quartile $Q_L(n)$, the median $\text{med}(n)$, the upper quartile $Q_U(n)$, and the maximum $\max(n)$ obtained from 1000 observed sample sizes. We also looked at the estimated coverage probability $\overline{p}$ with its estimated standard error $s(\overline{p})$ along with the observed values of both $\overline{n}/C$ and $\overline{n} - C$.

More specifically, in Tables 3.1-3.2, as well as in other tables which will follow, we have used the following notation:

$$\begin{aligned}
n_t &: \text{sample size in } t^{\text{th}} \text{ run;} \\
\overline{n} = T^{-1}\Sigma_{t=1}^{T} n_t &: \text{should estimate } C; \\
s(\overline{n}) = \left\{ \left(T^2 - T\right)^{-1} \Sigma_{t=1}^{T} \left(n_t - \overline{n}\right)^2 \right\}^{1/2} &: \text{estimated standard error of } \overline{n}; \\
\overline{x}_{n_t} &: \text{sample mean in } t^{\text{th}} \text{ run;} \\
p_t &: \text{1 if interval } J_{n_t} \text{ covers } \mu \text{ (or 0 otherwise)} \\
& \quad \text{in } t^{\text{th}} \text{ run;} \\
\overline{p} = T^{-1}\Sigma_{t=1}^{T} p_t &: \text{should compare with } 1 - \alpha; \\
s(\overline{p}) = \left\{ T^{-1} \overline{p} \left(1 - \overline{p}\right) \right\}^{1/2} &: \text{estimated standard error of } \overline{p}.
\end{aligned} \tag{3.1.14}$$

As a typical case, for example, we prefixed the optimal sample size $C = 50$ (small), 100 (moderate), and 200 (large) with different pilot sample sizes $m = 10$ and 15 respectively when $\alpha = 0.05$. Thus, with small, moderate and large optimal fixed sample sizes, the corresponding fixed-width of the confidence intervals was $d = 0.5544$, 0.3920, and 0.2772. Overall performances did not differ over other choices of C, α or m. In summary, we discover that the procedures (3.1.1)-(3.1.2) based on either GMD or MAD as the estimator of σ^2 tend to perform as well as the procedure based on the sample variance.

Looking at column 4, it is fair to say that the average sample sizes ($\overline{n}$) derived were close to each other, all staying a little smaller than C no matter whether C was small, moderate, or large. From column 9, we can see that the first-order term $\overline{n}/C$ is close to 1, whereas larger the sample sizes were, the closer was $\overline{n}/C$ to 1.

A similar conclusion could be in order while looking at the estimated second-order values $\overline{n} - C$ which look rather close to the theoretical asymptotic value -1.18 (see Mukhopadhyay and de silva 2009, eq. 6.2.34, p. 119) when $i = 0$. In some cases, the average sample sizes under procedure (3.1.1)-(3.1.2) with $i = 0$ turned out to be smaller than those of other two procedures ($i = 1, 2$), however the average sample sizes of GMD-based or MAD-based methodologies were often bit smaller.

Next, we focus on columns 5–7, the five number summary (min(n), $Q_L(n)$, med(n), $Q_U(n)$, and max(n)) from 1000 observed n values in each row. From Tables 3.1-3.2, one may sense that these quantities reflect a right-skewness property of the stopping times, while at the same time, the difference between $\overline{n}$ and med(n) was largely within one observation. For small sample sizes, the MAD-based procedure seemed to be more right-skewed when m became larger. For moderate and large sample sizes, while the values of $Q_L(n)$, med(n), and $Q_U(n)$ look close to each other, we noticed some apparent differences in the magnitude of $\max(n) - \min(n)$ when different procedures were implemented.

Now let us attend to column 8, that is $\overline{p}$, the estimated coverage probability, and $s(\overline{p})$. In a large majority of cases, the interval of $[\overline{p} \pm 2s(\overline{p})]$ did cover our nominal target $1 - \alpha$, however sometimes $[\overline{p} \pm 2s(\overline{p})]$ did not cover $1 - \alpha$. This feature should not be surprising, since Theorem 3.1.3 shows only asymptotic consistency property.

In summary, the purely sequential procedure (3.1.1)-(3.1.2) has performed remarkably well under the normality assumption for the parent population. We found very little to no significant difference in the overall performances between the existing purely sequential procedure (3.1.1)-(3.1.2) with $i = 0$ and our newly developed purely sequential procedures (3.1.1)-(3.1.2) with $i = 1, 2$ based on GMD or MAD across the large range of choices of m, C, and α. Recall that Tables 3.1-3.2 highlight selected summaries.

3.1.4.2. Samples from an ϵ-Contaminated Normal Population (3.1.7)

Next, we drew pseudo random samples from a mixture normal population (3.1.7), namely, a population distribution given by $(1-\epsilon)N(5,4)+\epsilon N(5,9)$ to investigate performances of the purely sequential fixed-width confidence interval estimation strategies (N_i, J_{N_i}) from (3.1.1)-(3.1.2) with $i=0,1,2$ respectively when suspect outliers may occur. We fixed $\alpha=0.01,0.05,0.10$, $m=10,15,20$, $C=50,100,200$, and considered $\epsilon=0.1$ (10% contamination), 0.2 (20% contamination).

We implemented the purely sequential procedures similarly to what we had explained in connection with Table 3.1. Tables 3.3-3.4 show findings from 1000 runs when $m=10,15$, $C=50,100,200$, $\alpha=0.05$, and $\epsilon=0.10,0.20$ in the spirit of Table 3.1. In our larger simulation runs, we had additionally included the choices (i) $m=20$, $C=50,100,200$, $\alpha=0.10,0.01$, $\epsilon=0.10,0.20$ and (ii) $m=10,15,20$, $d=0.6,0.4,0.2$, $\alpha=0.10,0.01$, $\epsilon=0.1,0.2$. However, we exhibit only a brief summary especially since we found very little to no significant difference in the overall performances. In particular, we include no tables here in the spirit of Table 3.2.

At the same time, one will find a number of additional entries in Tables 3.3-3.4 compared with those shown in Tables 3.1-3.2. Let us explain the extra notation.

x_{tj} : j^{th} observation in t^{th} run;

γ_{tj} : $=1$ if x_{tj} came from $N(5,4)$ or 0 if x_{tj} was from $N(5,9)$;

$\gamma_t=\Sigma_{j=1}^{n_t}\gamma_{tj}$: proportion of observations sampled from $N(5,4)$ in t^{th} run;

$\overline{\gamma}=\Sigma_{t=1}^{T}\gamma_t$: should estimate $\gamma(=1-\epsilon)$;

$s(\overline{\gamma})=\left\{T^{-1}\overline{\gamma}\,(1-\overline{\gamma})\right\}^{1/2}$: estimated standard error of $\overline{\gamma}$;

$$
\begin{aligned}
KS_t \quad &: = 1 \text{ if K-S test's } p\text{-value} \geq 0.05 \text{ in } t^{\text{th}} \\
&\quad \text{run or 0 otherwise;} \\
KS = \Sigma_{t=1}^{T} KS_t \quad &: \text{proportion of data from } T \text{ runs that} \\
&\quad \text{passed 5\% level K-S normality test.}
\end{aligned} \tag{3.1.15}
$$

We denoted $\gamma = 1 - \epsilon$ and estimated the average sampling from $N(5, 4)$ population proportion $\overline{\gamma}$ and its standard error $s(\overline{\gamma})$. We observe column 10 that $\overline{\gamma}$ is close to the target $\gamma = 90\%(80\%)$ in Table 3.3(3.4) while its standard error $s(\overline{\gamma})$ hovers around 0.01. We also used Kolmogorov-Smirnov test to check the normality of our samples within each run and we recorded in column 11 the proportion KS, that passed the Kolmogorov-Smirnov test at 0.05 significance level among all 1000 runs. From column 11 in Tables 3.3-3.4, we get an impression that our observed data looked like normal between 89.7% to 95% of the time among 1000 runs even though in the background the data came from (3.1.7)!

Again, we gather data purely sequentially from a population described by (3.1.7), but **pretend** that the population is indeed $N(\mu = 5, \sigma_1^2 = 4)$ for all practical purpose. In column 1 we show the C-values (in the top) computed using (2.1.3) replacing $\sigma^2 \equiv \sigma_1^2 = 4$. Obviously, this C is a rather "fake" optimal fixed sample size, and this is clear from the fact that the $\overline{n}$ values appear off from such C-values shown in the top in column 1. But, in column 1, we also show the bold $\mathbf{C}$-values computed from (2.1.3) using the correct σ^2 value from the population (3.1.7). Clearly, the $\overline{n}$ values appear much closer to the bold $\mathbf{C}$-values shown in column 1.

3.1.4.3. Some Overall Sentiments

From Tables 3.3-3.4 and many other simulations that we had run, we may summarize our sentiments as follows when a population may be reasonably treated as a normal universe for practical purposes (as validated by Kolmogorov-Smirnov test), but it is mildly contaminated behind the scene:

(**a**) All three methodologies oversampled which is clear by comparing the $\overline{n}$ values with C-values or $\mathbf{C}$-values.

(**b**) Pilot size m has not impacted overall performances in a significant way.

(**c**) Average sample sizes of GMD-based and MAD-based purely sequential procedures (3.1.1)-(3.1.2) with $i = 1, 2$ were generally smaller than those associated with the sample variance-based procedure (3.1.1)-(3.1.2) with $i = 0$. We interpret this feature as a reasonable validation of mild robustness of the procedures (3.1.1)-(3.1.2) with $i = 1, 2$ over the procedure (3.1.1)-(3.1.2) with $i = 0$.

(**d**) MAD-based purely sequential procedure generally came up with a smaller average sample size than that associated with GMD-based procedure.

(**e**) Average sample sizes from the customary procedure (3.1.1)-(3.1.2) with $i = 0$ came bit closer to the prefixed C-values. This is also reflected in the feature that first-order terms $(\overline{n}/C)$ for the purely sequential procedures (3.1.1)-(3.1.2) with $i = 1, 2$ were bit lower than those of the procedure (3.1.1)-(3.1.2) with $i = 0$.

(**f**) From columns 5–7, we feel that the distributions of all three stopping times were skewed to the right which is validated by the feature that $\text{med}(n)$ turned out to be a little larger than $\overline{n}$, but they were largely within one observation of each other. Also, $\text{med}(n)$ seen from procedures (3.1.1)-(3.1.2) with $i = 1, 2$ kept being smaller than those derived from procedure (3.1.1)-(3.1.2) with $i = 0$.

(**g**) MAD-based procedure gave the smallest $\text{med}(n)$ and $\overline{n}$ values. Similar conclusions generally hold for the other two quartiles $(Q_L(n), Q_U(n))$.

(**h**) Nearly all coverage probability estimates ($\overline{p}$) fell slightly below the target confidence level of 0.95. We suspect that this may be partly due to contamination.

By combining our findings from Tables 3.1-3.4 with those obtained from numerous other sets of extensive simulations, we empirically feel confident that our newly developed GMD-based or MAD-based methodologies (3.1.1)-(3.1.2) with $i = 1, 2$ are more robust for practical purposes when we compare them with the sample variance-based methodology (3.1.1)-(3.1.2) with $i = 0$. At the same time, however, the GMD-based or MAD-based methodologies from (3.1.1)-(3.1.2) with $i = 1, 2$ are not very sharply divided. Hence, we suggest implementing either GMD-based or MAD-based methodologies from (3.1.1)-(3.1.2) with $i = 1, 2$ with confidence when sampling from a normal universe, especially when suspect outliers may be expected.

3.2. Minimum Risk Point Estimation

In this section, we estimate the unknown σ unbiasedly by the estimators from (2.3.4), respectively, based on the sample standard deviation, GMD, and MAD. The corresponding purely sequential procedures will be proposed to construct minimum risk point estimators for the unknown mean μ under the formulation (2.2.1)-(2.2.3). Again, we begin with pilot data $X_1, ..., X_m$ of size m, $m \geq 2$, and record one additional observation at-a-time successively as needed until we stop according to the following stopping rule:

$$N_i \equiv N_i(c) = \inf\{n \geq m : n \geq \sqrt{A/c}\left(V_{n;i} + n^{-\lambda}\right)\}, \qquad (3.2.1)$$

where $\lambda(> \frac{1}{2})$ is held fixed and σ is replaced with $V_{n;i}$, $i = 0, 1, 2$, in defining the boundary condition as we estimate $n^* \equiv n^*(c)$ from (2.2.3). Observe that the index i corresponds to the sample standard deviation-based, GMD-based or the MAD-based methodology with $V_{n;i}, i = 0, 1, 2$, defined via (2.3.1) and (2.3.4).

The stopping rule (3.2.1) looks much like the representations in Chow and Robbins (1965) and Ghosh et al. (1997, equation 9.2.30, p. 274). But, in comparison with (3.1.1), the present boundary condition in (3.2.1) looks slightly different because it includes an additional term, $n^{-\lambda}$. This is done in order to (i) avoid very early stopping and (ii) facilitate derivations of various convergences with "desired" rates.

With regard to implementation of (3.2.1), for $i = 0, 1, 2$, if $m \geq \sqrt{A/c}(V_{m;i} + m^{-\lambda})$ is satisfied, we do not take any additional observation and the sample size is $N_i = m$. Otherwise, we record one more observation and obtain updated $V_{m+1;i}$ to check with the stopping rule (3.2.1). We terminate sampling at the first time $N_i = n(\geq m)$ such that $n \geq \sqrt{A/c}\left(V_{n;i} + n^{-\lambda}\right)$ occurs. Finally, having observed the dataset

$$\{N_i, X_1, ..., X_m, X_{m+1}, ..., X_{N_i}\},$$

we construct the minimum risk point estimator for μ as follows:

$$\overline{X}_{N_i} \equiv N_i^{-1}\Sigma_{j=1}^{N_i} X_j, i = 0, 1, 2. \tag{3.2.2}$$

Clearly, $P_{\mu,\sigma}\{N_i < \infty\} = 1$ and $N_i \uparrow \infty$ w.p.1 as $c \downarrow 0$, $i = 0, 1, 2$.

3.2.1. Asymptotic First-Order Properties of the Purely Sequential Methodology (3.2.1)-(3.2.2)

In the spirit of our previous Lemma 3.1.4 and Theorem 3.1.1, modified very mildly, we first summarize the following properties for the purely sequential estimation strategy

(3.2.1)-(3.2.2):

$$
\begin{aligned}
&\text{(i)} \quad E_{\mu,\sigma}[N_i] < \infty, V_{\mu,\sigma}[N_i] < \infty; \\
&\text{(ii)} \quad E_{\mu,\sigma}[\overline{X}_{N_i}] = \mu, V_{\mu,\sigma}[\overline{X}_{N_i}] = \sigma^2 E_{\mu,\sigma}(N_i^{-1}); \\
&\text{(iii)} \quad \sqrt{N_i}(\overline{X}_{N_i} - \mu)/\sigma \sim N(0,1);
\end{aligned}
\tag{3.2.3}
$$

for all fixed μ, σ, A, c, and $i = 0, 1, 2$.

In the present situation, upon termination according to (3.2.1), we can express the risk associated with the purely sequential estimation strategy $\left(N_i, \overline{X}_{N_i}\right)$ defined via (3.2.1)-(3.2.2) as follows:

$$
R_{N_i}(c) \equiv E_{\mu,\sigma}[L_{N_i}\left(\mu, \overline{X}_{N_i}\right)] = A\sigma^2 E_{\mu,\sigma}[N_i^{-1}] + cE_{\mu,\sigma}[N_i], i = 0, 1, 2, \tag{3.2.4}
$$

since $\overline{X}_n$ and $I\{N_i = n\}$ are independent for each fixed $n(\geq m)$. The minimum fixed-sample-size risk can be expressed as:

$$
R_{n^*}(c) \equiv A\sigma^2 n^{*-1} + cn^* = 2cn^*, \tag{3.2.5}
$$

with $n^* \equiv n^*(c)$ from (2.2.3).

Robbins (1959) and Starr (1966) formulated two crucial notions, namely, the *risk efficiency* and *regret* originally developed in the context of comparing $R_{N_0}(c)$ associated with (3.2.4) and $R_{n^*}(c)$. We implement those notions for comparing $R_{N_i}(c)$ and $R_{n^*}(c)$:

$$
\begin{aligned}
\textit{Risk Efficiency}: \quad & \xi_i(c) \equiv \frac{R_{N_i}(c)}{R_{n^*}(c)} = \tfrac{1}{2}E_{\mu,\sigma}[N_i/n^*] + \tfrac{1}{2}E_{\mu,\sigma}[n^*/N_i]; \\
\textit{Regret}: \quad & \omega_i(c) \equiv R_{N_i}(c) - R_{n^*}(c) = cE_{\mu,\sigma}[(N_i - n^*)^2/N_i],
\end{aligned}
\tag{3.2.6}
$$

for $i = 0, 1, 2$.

Clearly, the following property will also hold:

$$N_i/n^* \stackrel{P_{\mu,\sigma}}{\rightarrow} 1 \text{ as } \; c \rightarrow 0, i = 0, 1, 2, \tag{3.2.7}$$

which results from the customary basic inequality and the facts that $N_i \rightarrow \infty$ and $\overline{X}_{N_i} \rightarrow \mu$ w.p.1($P_{\mu,\sigma}$) as $c \rightarrow 0$.

Next, in view of the first part of (3.2.7) and then slightly improvising the proof shown in the case of our earlier Theorem 3.1.2, we can immediately claim the following property:

Theorem 3.2.1. *For the stopping time N_i defined by the purely sequential estimation strategy (3.2.1)-(3.2.2), for all fixed μ, σ, and A, we have*:

$$\lim_{c \rightarrow 0} E_{\mu,\sigma}[N_i/n^*] = 1, i = 0, 1, 2 \text{ [\textit{Asymptotic First-Order Efficiency}]}, \tag{3.2.8}$$

with $n^* \equiv n^*(c)$ from (2.2.3).

Next, we show the following lemma, which will lead to the asymptotic first-order risk efficiency of the methodology (3.2.1)-(3.2.2).

Lemma 3.2.1. *For the stopping time N_i defined by the purely sequential estimation strategy (3.2.1)-(3.2.2), for all fixed μ, σ, and A, we have*: with any arbitrary $0 < \eta < 1$ and $r \geq 2$,

$$P_{\mu,\sigma}\{N_i \leq \eta n^*\} = O\left(n^{*-r/(2(1+\lambda))}\right), i = 0, 1, 2. \tag{3.2.9}$$

Proof: Let $\lfloor w \rfloor$ denote the largest integer that is smaller ($<$) than w and we define:

$$n_{1c} = \left\lfloor (A/c)^{\frac{1}{2(1+\lambda)}} \right\rfloor = O(c^{-\frac{1}{2(1+\lambda)}}) \text{ and } n_{2c} = \eta n^* = \eta\sigma\sqrt{A/c}. \tag{3.2.10}$$

From the definition of N_i in (3.2.1), it should be clear that $N_i \geq n_{1c}$ w.p.1$(P_{\mu,\sigma})$, $i = 0, 1, 2$.

Next, in view of (3.2.7), it is obvious that $P_{\mu,\sigma}\{N_i \leq \eta n^*\}$ ought to converge to zero as $c \to 0$. Hence, we set out to obtain the rate at which $P_{\mu,\sigma}\{N_i \leq \eta n^*\}$ may converge to zero for small c.

Case $i = 1$:

We write for small c, with $k_1 = \frac{2(1-\eta)}{\sqrt{\pi}}$ (> 0):

$$\begin{aligned} & P_{\mu,\sigma}\{N_1 \leq \eta n^*\} \\ & \leq P_{\mu,\sigma}\left\{\tfrac{\sqrt{\pi}}{2} G_n \leq \eta\sigma \text{ for some } n \text{ such that } n_{1c} \leq n \leq n_{2c}\right\} \\ & \leq P_{\mu,\sigma}\left\{\left|G_n - \tfrac{2}{\sqrt{\pi}}\sigma\right| \geq k_1\sigma \text{ for some } n \text{ such that } n_{1c} \leq n \leq n_{2c}\right\} \\ & \leq P_{\mu,\sigma}\left\{\max_{n_{1c} \leq n \leq n_{2c}} \left|G_n - \tfrac{2}{\sqrt{\pi}}\sigma\right| \geq k_1\sigma\right\}. \end{aligned} \tag{3.2.11}$$

We know that G_n is a U-statistic with kernel $|X_{i_1} - X_{i_2}|$ and $E_{\mu,\sigma}[G_n] = \frac{2}{\sqrt{\pi}}\sigma$. As a result, $\{G_n; n \geq m\}$ is a reverse martingale adapted to the σ-fields $\mathcal{F}_n = \sigma\{G_n, G_{n+1}, ...\}$, and so is also $\left\{G_n - \frac{2}{\sqrt{\pi}}\sigma; n \geq m\right\}$. Note the fact that all positive moments of $|X_{i_1} - X_{i_2}|$ are finite. Hence, by Kolmogorov's inequality for reverse martingales, the lemma from Sen and Ghosh (1981), in view of (3.2.10) we obtain from (3.2.11) with $r \geq 2$:

$$P_{\mu,\sigma}\{N_1 \leq \eta n^*\} \leq (k_1\sigma)^{-r} E_{\mu,\sigma}\left[\left|G_{n_{1c}} - \tfrac{2}{\sqrt{\pi}}\sigma\right|^r\right] = O\left(n_{1c}^{-r/2}\right) = O\left(n^{*-r/(2(1+\lambda))}\right). \tag{3.2.12}$$

See also Ghosh et al. (1997, Lemma 9.2.3, pp. 275-276).

Case $i = 0$:

Observing that in (2.3.4) $\frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)} \geq 1$ as $E_{\mu,\sigma}[S_n] \leq \sqrt{E_{\mu,\sigma}[S_n^2]} = \sigma$, we again write for small c, with $k_0 = 1 - \eta\,(> 0)$:

$$\begin{aligned}
&P_{\mu,\sigma}\{N_0 \leq \eta n^*\} \\
&\quad \leq P_{\mu,\sigma}\left\{S_n \leq \eta\sigma \text{ for some } n \text{ such that } n_{1c} \leq n \leq n_{2c}\right\} \\
&\quad \leq P_{\mu,\sigma}\left\{|S_n - \sigma| \geq k_0\sigma, \text{ for some } n \text{ such that } n_{1c} \leq n \leq n_{2c}\right\} \\
&\quad \leq P_{\mu,\sigma}\left\{\max_{n_{1c} \leq n \leq n_{2c}} \left|S_n^2 - \sigma^2\right| \geq k_0\sigma^2\right\}.
\end{aligned} \tag{3.2.13}$$

Note that S_n^2 is a common U-statistic with all positive moments finite, with $r \geq 2$, we claim:

$$P_{\mu,\sigma}\{N_0 \leq \eta n^*\} = O\left(n_{1c}^{-r/2}\right) = O\left(n^{*-r/(2(1+\lambda))}\right), \tag{3.2.14}$$

in the same way we had concluded (3.2.12) from (3.2.11).

Case $i = 2$:

We note that

$$M_n \geq n^{-1}\Sigma_{i=1}^n |X_i - \mu| - \left|\overline{X}_n - \mu\right| = M_{1n} - M_{2n}, \text{ say,}$$

w.p.1($P_{\mu,\sigma}$). Similarly, observing that in (2.3.4) $\sqrt{\frac{\pi n}{2(n-1)}} \geq \sqrt{\frac{\pi}{2}}$, we again write for small c, with $k_2 = \frac{2(1-\eta)}{\sqrt{2\pi}}$:

$$\begin{aligned}
&P_{\mu,\sigma}\{N_2 \leq \eta n^*\} \\
&\quad \leq P_{\mu,\sigma}\left\{\sqrt{\tfrac{\pi}{2}}M_n \leq \eta\sigma \text{ for some } n \text{ such that } n_{1c} \leq n \leq n_{2c}\right\} \\
&\quad \leq P_{\mu,\sigma}\left\{\left|M_{1n} - \sqrt{\tfrac{2}{\pi}}\sigma\right| + |M_{2n}| \geq k_2\sigma, \text{ for some } n \text{ such that } n_{1c} \leq n \leq n_{2c}\right\}
\end{aligned}$$

$$\leq P_{\mu,\sigma}\left\{\max_{n_{1c}\leq n\leq n_{2c}}\left|M_{1n}-\sqrt{\tfrac{2}{\pi}}\sigma\right|\geq \tfrac{1}{2}k_2\sigma\right\}+P_{\mu,\sigma}\left\{\max_{n_{1c}\leq n\leq n_{2c}}|M_{2n}|\geq \tfrac{1}{2}k_2\sigma\right\}. \tag{3.2.15}$$

Since M_{1n} and $\overline{X}_n$ are both U-statistics with all positive moments finite, with $r \geq 2$, we claim:

$$P_{\mu,\sigma}\{N_2 \leq \eta n^*\} = O\left(n_{1c}^{-r/2}\right) = O\left(n^{*-r/(2(1+\lambda))}\right). \tag{3.2.16}$$

The proof of the lemma is complete that $P_{\mu,\sigma}\{N_i \leq \eta n^*\} = O\left(n^{*-r/(2(1+\lambda))}\right)$ holds for $i = 0, 1, 2$. ■

Now we are in a position to state and prove the main result in this section.

Theorem 3.2.2. *For the stopping time N_i defined by the purely sequential estimation strategy (3.2.1)-(3.2.2), for all fixed μ, σ, and A, we have:*

$$\lim_{c\to 0}\xi_i(c) = 1, i = 0, 1, 2 \text{ [\textit{Asymptotic First-Order Risk Efficiency}]}, \tag{3.2.17}$$

where the term $\xi_i(c)$ comes from (3.2.6).

Proof: In view of (3.2.6) and (3.2.8), it will suffice to prove the following result:

$$\lim_{c\to 0} E_{\mu,\sigma}[n^*/N_i] = 1,\ i = 0, 1, 2. \tag{3.2.18}$$

With $\eta = \frac{1}{2}$ in Lemma 3.2.1, we express:

$$\begin{aligned} E_{\mu,\sigma}[n^*/N_i] &= E_{\mu,\sigma}\left[\tfrac{n^*}{N_i} I\left(N_i \leq \tfrac{1}{2}n^*\right)\right] + E_{\mu,\sigma}\left[\tfrac{n^*}{N_i} I\left(N_i > \tfrac{1}{2}n^*\right)\right] \\ &= E_{\mu,\sigma}[W_1] + E_{\mu,\sigma}[W_2], \text{ say, } i = 0, 1, 2. \end{aligned} \tag{3.2.19}$$

But, W_2 is bounded and hence W_2 is uniformly integrable. Also, $W_2 \stackrel{P_{\mu,\sigma}}{\to} 1$ as $c \to 0$ so that $\lim_{c\to 0} E_{\mu,\sigma}[W_2] = 1$. In view of (3.2.9) and (3.2.10), we can write:

$$E_{\mu,\sigma}[W_1] \le E_{\mu,\sigma}\left[\frac{n^*}{n_{1c}} I\left(N_i \le \tfrac{1}{2}n^*\right)\right] = O\left(n^*\right) O\left(n^{*-1/(1+\lambda)}\right) O\left(n^{*-r/(2(1+\lambda))}\right)$$
$$= O\left(n^{*1-\frac{1}{1+\lambda}-\frac{r}{2(1+\lambda)}}\right) = O\left(n^{*\frac{2\lambda-r}{2(1+\lambda)}}\right) \to 0,$$

as long as we pick some $r > \max\{2, 2\lambda\}$. Thus, from (3.2.18), it follows that (3.2.17) is justified. ■

3.2.2. Simulations and Data Analysis

Along the lines of what we did in Section 3.1.4, we had analogously implemented purely sequential minimum risk point estimation strategies $\left(N_i, \overline{X}_{N_i}\right), i = 0, 1, 2$ from (3.2.1)-(3.2.2) under both normal cases and mixture-normal cases (3.1.7). In all tables to follow, we continue to use the index i along the lines of (3.1.13), but define more precisely as follows:

$i = 0$: Sample standard deviation-based methodology (3.2.1)-(3.2.2) with $\lambda = 2$

$i = 1$: GMD-based methodology (3.2.1)-(3.2.2) with $\lambda = 2$

$i = 2$: MAD-based methodology (3.2.1)-(3.2.2) with $\lambda = 2$

(3.2.20)

as well as other notation used in the spirit of (3.1.14) obtained on the basis of T (= 1000, say) replications:

n_t : sample size in t^{th} run;

$\overline{n} = \frac{1}{T}\Sigma_{t=1}^{T} n_t$: should estimate n^*;

$s(\overline{n})$: estimated standard error (s.e.) of $\overline{n}$;

$R_{n_t} = A s_{n_t}^2 n_t^{-1} + c n_t$: the method of calculation of risk using sample variance, $s_{n_t}^2$, in t^{th} run;

$$
\begin{aligned}
\overline{R}_n = \tfrac{1}{T}\Sigma_{t=1}^{T} R_{n_t} \quad &: \text{ should estimate } R_{n^*}(c); \\
s\left(\overline{R}_n\right) \quad &: \text{ estimated s.e. of } \overline{R}_n; \\
\widehat{\xi} = \overline{R}_n / R_{n^*}(c) \quad &: \text{ should estimate } \xi_i(c),\ i = 0,1,2; \\
\widehat{\omega} = c\tfrac{1}{T}\Sigma_{t=1}^{T} \tfrac{(n_t - n^*)^2}{n_t} \quad &: \text{ should estimate } \omega_i(c),\ i = 0,1,2; \\
\delta_i^2 c \quad &: \text{ a theoretical approximation of } \omega_i(c), i = 0,1,2.
\end{aligned}
\tag{3.2.21}
$$

3.2.2.1. Samples from a Normal Population

We generated pseudo random samples from a $N(5,4)$ population, that is, we had fixed $\mu = 5$ and $\sigma = 2$. We also fixed $A = 100$ and $m = 10, 15, 20$. In addition, we varied the optimal fixed sample size $n^*(c)$ over a wide range of values including $50, 100, 200$, and then determined the values of $c(= 0.16, 0.04, 0.01)$ accordingly by referring to (2.2.3). Table 3.5 represents this situation.

Given the set of notation defined in (3.2.21), we present selected summaries based on Table 3.5 when $m = 10, 15$ and $n^*(c) = 50, 100, 200$ for brevity alone even though we ran simulations with other values of $n^*(c), A, c, m$. Using $R = 1000$ independent replications or runs, we estimated the average final sample size $\overline{n}$ with its estimated standard error $s(\overline{n})$, the minimum $\min(n)$, the lower quartile $Q_L(n)$, the median $\text{med}(n)$, the upper quartile $Q_U(n)$, and the maximum $\max(n)$ obtained from 1000 observed sample sizes.

We also looked at the rather natural estimated value, $\overline{R}_n$, the terminal risk along with its estimated standard error $s\left(\overline{R}_n\right)$ in column 8, $\widehat{\xi}$, the risk efficiency and $\widehat{\omega}$, the regret both in column 9. Considering the the values of $\overline{R}_n$ along with the estimated standard error, we feel tempted to suggest that the number is generally consistent with Remark 3.1 given at the end of Section 3.2.

3.2.2.2. Samples from an ϵ-Contaminated Normal Population (3.1.7)

This section is similar in principle with what we had reported in Section 3.1.4.2. We generated pseudo random samples from a mixture-normal population distribution given by $(1-\epsilon)N(5,4)+\epsilon N(5,9)$ with $\epsilon = 0.1, \epsilon = 0.2$, and $\lambda = 2$. Our Tables 3.6-3.7 show brief summaries of our findings, parallel to those shown in Tables 3.3-3.4. A majority of the entries in Tables 3.6-3.7 ought to be interpreted as those in Tables 3.3-3.4, respectively. The columns showing $\overline{R}_n$, the terminal risk, $\widehat{\xi}$, the risk efficiency, and $\widehat{\omega}$, the regret, should be interpreted just like those in Table 3.5.

3.2.2.3. Some Overall Sentiments

From Tables 3.5-3.7 and many other simulations that we had run, we may summarize our overall sentiments as follows when a population may be reasonably treated as a normal universe for practical purposes (as validated by Kolmogorov-Smirnov test), but it is mildly contaminated behind the scene:

We empirically feel confident that our newly developed GMD-based or MAD-based methodologies (3.2.1)-(3.2.2) with $i = 1, 2$ are more robust for practical purposes when we compare them with the sample standard deviation-based methodology (3.2.1)-(3.2.2) with $i = 0$. At the same time, however, the GMD-based or MAD-based methodologies from (3.2.1)-(3.2.2) with $i = 1, 2$ are not very sharply divided. Hence, we suggest implementing either GMD-based or MAD-based methodologies from (3.2.1)-(3.2.2) with $i = 1, 2$ with confidence when sampling from a normal universe, especially when suspect outliers may be expected.

Remark 3.1. For all fixed μ, σ, c and A, we can check easily that $\xi_i(c) > 1$, $i = 0, 1, 2$. This is consistent with the observations noted generally by Ghosh and Mukhopadhyay (1976).

Remark 3.2. A more original customary purely sequential minimum risk point estimation methodology excludes the term $n^{-\lambda}$ in the stopping rule from (3.2.1), which is given by

$$N_0' \equiv N_0'(c) = \inf\{n \geq m : n \geq S_n \sqrt{A/c}\},$$

instead. The conclusion from Theorems 3.2.3-3.2.3 also holds for the associated strategy $\left(N_0', \overline{X}_{N_0'}\right)$ when $m \geq 3$. One may refer to Ghosh et al. (1997, pp. 174-175), Mukhopadhyay and de Silva (2009, pp. 146-147) or look at another source.

Remark 3.3. Besides the asymptotic first-order properties given in Section 3.2.1, the purely sequential MRPE methodology (3.2.1)-(3.2.2) further enjoys asymptotic second-order risk efficiency. Stronger results will be provided in Chapter 5.

3.3. Illustrations with Real Datasets

In this section, we include illustrations of both FCIE and MRPE methodologies for the mean (μ) using two separate real datasets.

First, in Section 3.3.1, we will include data analysis associated with our proposed methodologies using net profit or loss percentages from 352 department stores with sales less than one million dollars in 1925 referred to as the "sales data" in the sequel. The original data can be found from McNair (1930) or it can be downloaded from the Miscellaneous Datasets page of Winner's personal website (http://www.stat.ufl.edu/~winner/).

Next, in Section 3.3.2, we will include illustrations using the real data associated with the number of days that the seeds of marigold varieties 2 and 3 needed to flower referred to as the "horticulture data" in the sequel. The data were collected and recorded by Mukhopadhyay et al. (2004b).

3.3.1. Illustrations Using Sales Data

The sales data on net profit or loss percentages for 352 department stores came with min $= -14.21$ to max $= 10.31$. The whole dataset looked bit skewed to the left with 5 outlying observations, all on the left, namely $-14.21, -12.61, -11.86, -11.04$, and -9.97. These were clear from the histogram and boxplot which are not included here for brevity. The whole data (with $n = 352$) gave the following descriptive statistics:

n	$\overline{x}$	s	min	Q_L	med	Q_U	max
352	0.261	3.854	−14.21	−2.1325	0.360	2.9300	10.31

Normality of the data was violated with the p-value 0.008611 associated with the Shapiro-Wilk test on the whole data of size 352. We decided to remove the most extreme outlier, -14.21, and then the normality of the data of size 351 gave the p-value 0.09033 associated with the Shapiro-Wilk test on the data of size 351. The revised descriptive statistics from the data (with $n = 351$) after removing the most extreme outlier, -14.21, were:

n	$\overline{x}$	s	min	Q_L	med	Q_U	max
351	0.302	3.781	−12.61	−2.1250	0.360	2.9400	10.31

For purposes of illustrations, we treated the dataset of size 351 as our population after removing the most extreme outlier, -14.21. We drew random samples from this population separately under simple random sampling with replacement (SRSWR) as well as under simple random sampling without replacement (SRSWOR). They did not give appreciably different results and hence we summarize our findings obtained under SRSWOR only.

We began with a pilot sample of size $m = 20$ on each occasion and implemented the fixed-width 95% confidence interval methodologies using purely sequential strategies (3.1.1)-(3.1.2) based on the sample variance ($i = 0$), GMD ($i = 1$), and MAD ($i = 2$), respectively. Analogously, we implemented the minimum risk point estimation methodologies using purely sequential strategies (3.2.1)-(3.2.2).

After we began each strategy with m pilot observations, we checked with all three stopping rules simultaneously based whether to terminate or not based on the same sequence of observations. That is, as one strategy terminated, we continued with the other two based on data accrued thus far. Then, eventually, all three strategies based on rather comparable sets of observations from the same population terminated at some stage depending on the sample path. Thus, one will find descriptive statistics presented in Tables 3.8-3.9 for the observed data (x) are comparable for $i = 0, 1, 2$ within a fixed set of configuration. These are summarized in Tables 3.8-3.9 where the choices of respective design constants are made explicit as needed.

One should reiterate the point that each row in Tables 3.8-3.9 is obtained upon termination from a single run of an appropriate estimation methodology under consideration. We emphasize that we had repeated similar illustrations with all 352 observations included in the population, but we saw no appreciably different results compared with those shown in Tables 3.8-3.9. Hence, we leave them out for brevity.

3.3.2. Illustrations Using Horticulture Data

Here, we utilized the horticulture data collected and recorded by Mukhopadhyay et al. (2004b) on the number of days that the seeds of marigold varieties 2 and 3 needed to flower. We treated the real datasets on variety 2 and variety 3 which seemed not to contradict normal distributions confirmed via Shapiro-Wilk test with associated p-values 0.1473 and

0.4101 for variety 2 and variety 3 respectively.

However, we saw suspect outliers, for example, the observation 22.0 from variety 2 and observations $22.5, 23.0, 33.5, 34.5$ from variety 3. The simple descriptive statistics from full datasets on varieties 2 and 3 are summarized as follows:

Variety	n	$\overline{x}$	s	min	Q_L	med	Q_U	max
2	460	35.13	4.03	22.0	32.0	35.0	38.0	46.5
3	162	28.20	2.17	22.5	27.0	28.5	29.5	34.5

For purposes of illustrations, we treated these datasets of sizes 460 and 162 respectively as our populations. We drew random samples from these populations in the spirit of Section 3.3.1 under SRSWR as well as SRSWOR. But, SRSWOR and SRSWR did not produce appreciably different results and hence we summarize our findings obtained under SRSWOR only.

As in Section 3.1.1, again we began with a pilot sample of size $m = 20$ on each occasion and implemented the fixed-width 95% confidence interval methodologies using purely sequential strategies (3.1.1)-(3.1.2) based on the sample variance ($i = 0$), GMD ($i = 1$), and MAD ($i = 2$), respectively. Analogously, we implemented the minimum risk point estimation methodologies using purely sequential strategies (3.2.1)-(3.2.2).

After we began each strategy with m pilot observations, we checked with all three stopping rules simultaneously based whether to terminate or not based on the same sequence of observations. That is, as one strategy terminated, we continued with the other two based on data accrued thus far. Then, eventually, all three strategies based on rather comparable sets of observations from the same population terminated at some stage depending on the sample path.

Thus, one will find descriptive statistics presented in Tables 3.10-3.11 for the observed data (x) are comparable for $i = 0, 1, 2$ within a fixed set of configuration. These are summarized in Tables 3.10-3.11 where the choices of respective design constants are made explicit as needed. One should reiterate the point that each row in Tables 3.10-3.11 is obtained upon termination from a single run of an appropriate estimation methodology under consideration. We leave out many details for brevity.

3.4. Overall Concluding Thoughts

In summary, the purely sequential procedures (3.1.1)-(3.1.2) and (3.2.1)-(3.2.2) have performed remarkably well, whether based on GMD or MAD, under the normality assumption for the parent population. We saw very little to no significant differences in the overall performances between the three purely sequential procedures, that is (3.1.1)-(3.1.2) and (3.2.1)-(3.2.2), across the board when the population can be reasonably assumed normal.

We empirically feel strongly confident that our newly developed GMD-based or MAD-based methodologies are more robust for practical purposes when we compare them with the customary methodologies based on the sample variance or sample standard deviation, respectively. Hence, we enthusiastically suggest implementing either GMD-based or MAD-based methodologies with confidence when sampling from a normal universe, especially when suspect outliers may be expected.

Table 3.1. Simulations under $\alpha = 0.05$ with 1000 runs implementing methods (3.1.1)-(3.1.2) with i from (3.1.13)

C / d	m	i	$\overline{n}$ / $s(\overline{n})$	$\min(n)$ / $\max(n)$	$Q_L(n)$ / $Q_U(n)$	$\text{med}(n)$	$\overline{p}$ / $s(\overline{p})$	$\overline{n}/C$ / $\overline{n}-C$
50	10	0	47.823	10	40.00	48.0	**0.925**	0.956
0.5544			0.374	82	56.00		0.008	−2.177
		1	47.917	10	41.00	49.0	0.939	0.958
			0.370	84	56.00		0.008	−2.083
		2	47.307	10	40.00	48.0	0.934	0.946
			0.394	78	56.00		0.008	−2.693
	15	0	47.926	15	41.00	49.0	**0.926**	0.959
			0.352	78	55.00		0.008	−2.074
		1	48.269	15	42.00	49.0	**0.932**	0.965
			0.355	79	56.00		0.008	−1.731
		2	47.831	15	40.00	48.0	0.936	0.957
			0.371	81	56.00		0.008	−2.169
100	10	0	98.226	10	88.00	99.0	0.941	0.982
0.3920			0.513	145	109.00		0.007	−1.774
		1	98.411	13	88.00	99.0	**0.929**	0.984
			0.509	145	109.00		0.008	−1.589
		2	98.368	10	88.00	99.0	0.948	0.984
			0.519	149	110.00		0.007	−1.632
	15	0	98.565	24	89.00	99.0	0.951	0.986
			0.497	141	109.00		0.007	−1.435
		1	98.446	38	89.00	99.0	0.952	0.984
			0.492	149	109.00		0.007	−1.554
		2	98.780	30	88.00	99.0	0.945	0.988
			0.515	158	110.00		0.007	−1.220
200	10	0	198.998	138	185.00	199.0	0.944	0.995
0.2772			0.652	264	212.25		0.007	−1.002
		1	198.681	123	185.00	199.0	**0.931**	0.993
			0.652	272	213.00		0.008	−1.319
		2	198.573	10	183.00	199.0	0.942	0.993
			0.712	267	213.00		0.007	−1.427
	15	0	198.662	120	185.00	199.0	0.944	0.993
			0.651	266	212.00		0.007	−1.338
		1	198.644	92	185.00	199.0	0.945	0.993
			0.677	276	213.00		0.007	−1.356
		2	198.649	121	183.00	198.0	0.937	0.993
			0.707	277	214.00		0.008	−1.351

Table 3.2. Simulations under $\alpha = 0.05$ with 1000 runs implementing methods (3.1.1)-(3.1.2) picking d first with i from (3.1.13)

C / d	m	i	$\overline{n}$ / $s(\overline{n})$	$\min(n)$ / $\max(n)$	$Q_L(n)$ / $Q_U(n)$	$\text{med}(n)$	$\overline{p}$ / $s(\overline{p})$	$\overline{n}/C$ / $\overline{n}-C$
42.683	10	0	41.262	10	35.00	42.0	0.940	0.967
0.6			0.338	71	49.00		0.008	−1.421
		1	41.281	10	35.00	42.0	**0.929**	0.967
			0.341	68	49.00		0.008	−1.402
		2	41.221	10	34.00	42.0	0.935	0.966
			0.357	72	49.00		0.008	−1.462
	15	0	41.257	15	34.75	42.0	**0.921**	0.967
			0.324	73	49.00		0.009	−1.426
		1	41.469	15	35.00	42.0	**0.927**	0.972
			0.338	74	49.00		0.008	−1.214
		2	41.266	15	34.00	42.0	**0.927**	0.967
			0.347	71	49.00		0.008	−1.417
96.036	10	0	94.854	10	85.00	96.0	0.936	0.988
0.4			0.486	140	105.00		0.008	−1.182
		1	94.837	45	85.00	96.0	0.940	0.988
			0.481	141	105.00		0.008	−1.199
		2	95.035	11	85.00	96.0	**0.928**	0.990
			0.509	138	106.00		0.008	−1.001
	15	0	95.093	46	86.00	96.0	**0.931**	0.990
			0.475	138	105.00		0.008	−0.943
		1	94.947	19	85.00	96.0	**0.926**	0.989
			0.496	141	106.00		0.008	−1.089
		2	94.753	23	84.00	95.0	0.937	0.987
			0.513	137	106.00		0.008	−1.283
384.146	10	0	382.746	277	364.00	384.0	0.935	0.996
0.2			0.928	470	403.00		0.008	−1.400
		1	383.010	267	363.00	384.0	0.950	0.997
			0.904	470	402.00		0.007	−1.136
		2	382.866	287	361.00	383.0	**0.931**	0.997
			0.961	467	404.00		0.008	−1.280
	15	0	382.887	274	363.00	385.0	0.940	0.997
			0.910	470	403.00		0.008	−1.259
		1	383.104	259	365.00	384.0	0.945	0.997
			0.924	470	402.00		0.007	−1.042
		2	382.751	269	363.00	384.0	**0.932**	0.996
			0.977	474	405.00		0.008	−1.395

Table 3.3. Simulations under $\alpha = 0.05$ and $\epsilon = 0.10$ in (3.1.7) with 1000 runs implementing methods (3.1.1)-(3.1.2) with i from (3.1.13)

C d	m	i	$\overline{n}$ $s(\overline{n})$	min max	Q_L Q_U	med	$\overline{p}$ $s(\overline{p})$	$\overline{n}/C$ $\overline{n}-C$	$\overline{\gamma}$ $s(\overline{\gamma})$	KS
50	10	0	54.127	10	47.00	55.0	0.934	1.083	0.901	0.926
56.25			0.391	95	62.00		0.008	4.127	0.009	
0.5544		1	53.367	10	46.00	54.0	0.922	1.067	0.901	0.930
			0.394	86	62.00		0.008	3.367	0.009	
		2	52.595	10	44.00	53.0	0.927	1.052	0.902	0.930
			0.420	95	62.00		0.008	2.595	0.009	
	15	0	54.176	15	46.00	55.0	0.952	1.084	0.900	0.946
			0.395	92	62.25		0.007	4.176	0.009	
		1	53.636	15	46.00	54.0	0.946	1.073	0.900	0.937
			0.378	90	61.00		0.007	3.636	0.009	
		2	53.028	15	45.00	54.0	0.939	1.061	0.899	0.936
			0.402	87	61.25		0.008	3.028	0.010	
100	10	0	109.924	11	99.00	110.5	0.947	1.099	0.901	0.940
112.5			0.541	165	121.00		0.007	9.924	0.009	
0.3920		1	109.314	48	99.00	109.0	0.930	1.093	0.901	0.935
			0.511	158	121.00		0.008	9.314	0.009	
		2	107.829	56	97.00	108.0	0.930	1.078	0.901	0.933
			0.539	159	120.00		0.008	7.829	0.009	
	15	0	110.357	53	100.00	111.0	0.944	1.104	0.901	0.940
			0.520	155	121.00		0.007	10.357	0.009	
		1	109.284	60	98.00	109.5	0.950	1.093	0.902	0.950
			0.504	153	120.00		0.007	9.284	0.009	
		2	107.985	44	97.00	108.5	0.933	1.080	0.901	0.939
			0.541	158	120.00		0.008	7.985	0.009	
200	10	0	222.790	150	208.00	224.0	0.929	1.114	0.901	0.938
225			0.713	293	238.00		0.008	22.790	0.009	
0.2772		1	219.633	143	204.00	220.0	0.932	1.098	0.901	0.941
			0.725	292	235.00		0.008	19.633	0.009	
		2	218.486	135	202.00	219.0	0.940	1.092	0.900	0.939
			0.727	289	234.00		0.008	18.486	0.009	
	15	0	222.277	153	207.00	223.0	0.929	1.111	0.901	0.937
			0.721	294	238.00		0.008	22.277	0.009	
		1	219.796	153	205.00	220.0	0.925	1.099	0.901	0.927
			0.695	276	235.00		0.008	19.796	0.009	
		2	218.059	144	202.00	218.0	0.937	1.090	0.900	0.936
			0.744	289	234.00		0.008	18.059	0.009	

Table 3.4. Simulations under $\alpha = 0.05$ and $\epsilon = 0.20$ in (3.1.7) with 1000 runs implementing methods (3.1.1)-(3.1.2) with i from (3.1.13)

C d	m	i	$\overline{n}$ $s(\overline{n})$	min max	Q_L Q_U	med	$\overline{p}$ $s(\overline{p})$	$\overline{n}/C$ $\overline{n}-C$	$\overline{\gamma}$ $s(\overline{\gamma})$	KS
50	10	0	59.114	10	51.75	60.0	**0.931**	1.182	0.805	0.921
62.5			0.443	101	68.00		0.008	9.114	0.013	
0.5544		1	58.879	10	51.00	59.0	0.938	1.178	0.803	0.937
			0.409	96	68.00		0.008	8.879	0.013	
		2	58.155	10	50.00	59.0	0.935	1.163	0.803	0.938
			0.426	96	67.00		0.008	8.155	0.013	
	15	0	60.021	15	51.00	61.0	0.940	1.200	0.803	0.923
			0.431	96	69.00		0.008	10.021	0.013	
		1	59.358	15	51.00	60.0	**0.930**	1.187	0.802	0.930
			0.395	93	68.00		0.008	9.358	0.013	
		2	58.139	15	50.00	59.0	**0.933**	1.163	0.804	0.932
			0.431	104	67.00		0.008	8.139	0.013	
100	10	0	122.534	13	111.00	123.0	0.934	1.225	0.801	0.925
125			0.570	171	135.00		0.008	22.534	0.013	
0.3920		1	120.809	30	109.75	122.0	0.935	1.208	0.801	0.930
			0.538	180	132.00		0.008	20.809	0.013	
		2	118.579	10	107.00	119.0	0.945	1.186	0.801	0.941
			0.568	172	130.00		0.007	18.579	0.013	
	15	0	122.738	68	110.00	123.0	0.939	1.227	0.801	0.932
			0.553	181	135.00		0.008	22.738	0.013	
		1	121.306	53	111.00	122.0	0.944	1.213	0.801	0.929
			0.539	172	133.00		0.007	21.306	0.013	
		2	118.898	66	107.00	120.0	0.937	1.189	0.801	0.933
			0.573	173	131.00		0.008	18.898	0.013	
200	10	0	247.049	166	230.00	248.0	0.936	1.235	0.802	0.897
250			0.790	318	264.00		0.008	47.049	0.013	
0.2772		1	242.648	165	226.00	242.5	0.935	1.213	0.801	0.914
			0.756	311	259.00		0.008	42.648	0.013	
		2	239.725	150	224.00	240.0	**0.932**	1.199	0.801	0.917
			0.750	318	255.00		0.008	39.725	0.013	
	15	0	246.803	168	231.00	248.0	0.936	1.234	0.801	0.912
			0.752	303	263.00		0.008	46.803	0.013	
		1	243.253	173	228.00	244.0	0.938	1.216	0.801	0.909
			0.743	308	259.00		0.008	43.253	0.013	
		2	239.931	149	224.00	240.0	0.935	1.200	0.801	0.907
			0.765	310	256.00		0.008	39.931	0.013	

Table 3.5. Simulations under $A = 100$ with 1000 runs implementing methods (3.2.1)-(3.2.2) with i from (3.2.20) and $\lambda = 2$

n^* c	m	i	$\overline{n}$ $s(\overline{n})$	$\min(n)$ $\max(n)$	$Q_L(n)$ $Q_U(n)$	$\text{med}(n)$	$\overline{R}_n$ $s(\overline{R}_n)$	$\widehat{\xi}$ $\widehat{\omega}$
50	10	0	50.199	22	47.00	50.0	15.777	0.986
0.16			0.173	65	54.00		0.055	0.104
		1	50.313	30	47.00	51.0	15.806	0.988
			0.170	65	54.00		0.054	0.098
		2	50.259	29	46.00	51.0	15.795	0.987
			0.178	66	54.00		0.053	0.107
	15	0	50.247	28	47.00	50.0	15.801	0.988
			0.167	67	54.00		0.053	0.093
		1	50.460	33	47.00	51.0	15.839	0.990
			0.168	65	54.00		0.053	0.095
		2	50.457	31	47.00	51.0	15.843	0.990
			0.171	66	54.00		0.053	0.098
100	10	0	100.107	72	95.00	100.0	7.937	0.992
0.04			0.238	123	106.00		0.019	2.33×10^{-2}
		1	100.335	70	95.00	100.0	7.954	0.994
			0.235	121	106.00		0.018	2.25×10^{-2}
		2	100.332	76	95.00	101.0	7.951	0.994
			0.249	123	106.00		0.019	2.55×10^{-2}
	15	0	100.159	76	95.00	101.0	7.941	0.993
			0.229	124	105.00		0.018	2.15×10^{-2}
		1	100.239	73	95.00	101.0	7.947	0.993
			0.233	121	105.25		0.018	2.23×10^{-2}
		2	100.245	69	95.00	100.0	7.944	0.993
			0.250	122	106.00		0.019	2.55×10^{-2}
200	10	0	200.132	165	194.00	201.0	3.985	0.996
0.01			0.333	231	207.00		0.007	5.64×10^{-3}
		1	200.255	167	193.00	201.0	3.988	0.997
			0.338	230	207.00		0.007	5.80×10^{-3}
		2	200.026	162	193.00	200.0	3.985	0.996
			0.356	230	208.00		0.007	6.44×10^{-3}
	15	0	200.199	163	193.00	201.0	3.986	0.997
			0.328	229	207.25		0.007	5.46×10^{-3}
		1	200.201	160	193.00	200.0	3.987	0.997
			0.338	233	208.00		0.007	5.80×10^{-3}
		2	200.103	151	193.00	201.0	3.989	0.997
			0.346	229	208.00		0.007	6.16×10^{-3}

Table 3.6. Simulations under $A = 100$ and $\epsilon = 0.10$ in (3.1.7) with 1000 runs implementing methods (3.2.1)-(3.2.2) with i from (3.2.20) and $\lambda = 2$

n^* c	m	i	$\overline{n}$ $s(\overline{n})$	min max	Q_L Q_U	med	$\overline{R}_n$ $s(\overline{R}_n)$	$\widehat{\xi}$ $100\widehat{\omega}$	$\overline{\gamma}$ $s(\overline{\gamma})$	KS
50	10	0	53.191	26	50.00	53.0	16.744	1.046	0.900	0.938
53.033			0.183	70	57.00		0.058	12.7	0.009	
0.16		1	52.973	30	49.00	53.0	16.749	1.047	0.901	0.931
			0.178	69	57.00		0.057	11.8	0.009	
		2	52.756	33	49.00	53.0	16.751	1.047	0.901	0.948
			0.189	71	57.00		0.059	12.5	0.009	
	15	0	53.059	28	49.00	53.0	16.699	1.044	0.902	0.948
			0.182	70	57.00		0.058	12.2	0.009	
		1	53.152	29	50.00	53.0	16.809	1.051	0.900	0.949
			0.171	71	57.00		0.056	11.3	0.010	
		2	52.917	35	49.00	53.0	16.787	1.049	0.901	0.941
			0.177	70	56.00		0.055	11.5	0.009	
100	10	0	106.393	84	101.00	106.0	8.439	1.055	0.901	0.943
106.066			0.250	130	112.00		0.020	3.64	0.009	
0.04		1	105.719	72	100.00	105.0	8.427	1.053	0.901	0.959
			0.249	130	111.00		0.021	3.41	0.009	
		2	105.254	75	100.00	105.0	8.424	1.053	0.900	0.939
			0.260	129	111.00		0.021	3.42	0.009	
	15	0	106.611	84	102.00	107.0	8.456	1.057	0.900	0.947
			0.242	131	112.00		0.019	3.62	0.009	
		1	105.974	79	101.00	106.0	8.446	1.056	0.901	0.961
			0.239	128	111.00		0.019	3.31	0.009	
		2	105.353	79	100.00	105.0	8.434	1.054	0.901	0.945
			0.253	128	110.00		0.020	3.34	0.009	
200	10	0	211.699	175	204.00	211.0	4.217	1.054	0.901	0.949
212.132			0.348	241	220.00		0.007	0.116	0.009	
0.01		1	210.866	178	204.00	211.0	4.221	1.055	0.901	0.953
			0.350	242	218.00		0.007	0.109	0.009	
		2	210.113	164	203.00	210.5	4.224	1.056	0.901	0.947
			0.358	244	217.00		0.007	0.105	0.009	
	15	0	212.014	176	205.00	212.0	4.223	1.056	0.901	0.944
			0.339	248	219.00		0.007	0.117	0.009	
		1	210.620	175	204.00	211.0	4.217	1.054	0.901	0.959
			0.347	246	218.00		0.007	0.106	0.009	
		2	210.045	177	203.00	210.0	4.224	1.056	0.901	0.956
			0.349	244	217.00		0.007	0.101	0.009	

Table 3.7. Simulations under $A = 100$ and $\epsilon = 0.20$ in (3.1.7) with 1000 runs implementing methods (3.2.1)-(3.2.2) with i from (3.2.20) and $\lambda = 2$

n^* / c	m	i	$\overline{n}$ / $s(\overline{n})$	min / max	Q_L / Q_U	med	$\overline{R}_n$ / $s(\overline{R}_n)$	$\widehat{\xi}$ / $100\widehat{\omega}$	$\overline{\gamma}$ / $s(\overline{\gamma})$	KS
50	10	0	55.759	35	52.00	56.0	17.552	1.097	0.803	0.920
55.902			0.197	75	60.00		0.063	0.190	0.013	
0.16		1	55.339	33	51.00	56.0	17.554	1.097	0.801	0.929
			0.194	74	59.00		0.063	0.176	0.013	
		2	55.075	27	51.00	55.0	17.589	1.099	0.801	0.938
			0.196	74	59.00		0.062	0.173	0.013	
	15	0	55.934	26	52.00	56.0	17.609	1.101	0.801	0.927
			0.197	75	60.00		0.063	0.197	0.013	
		1	55.423	36	52.00	56.0	17.570	1.098	0.802	0.923
			0.190	72	59.00		0.062	0.174	0.013	
		2	55.085	32	51.00	55.0	17.591	1.099	0.802	0.918
			0.191	72	59.00		0.061	0.166	0.013	
100	10	0	111.491	87	106.00	111.0	8.847	1.106	0.803	0.927
111.803			0.270	136	117.00		0.022	6.87	0.013	
0.04		1	110.892	84	105.00	111.0	8.869	1.109	0.802	0.927
			0.267	139	117.00		0.022	6.40	0.013	
		2	110.008	76	104.00	110.0	8.862	1.108	0.802	0.924
			0.264	136	116.00		0.022	5.78	0.013	
	15	0	111.654	81	106.00	111.0	8.861	1.108	0.802	0.926
			0.267	137	117.00		0.021	6.93	0.013	
		1	110.990	79	106.00	111.0	8.875	1.109	0.802	0.936
			0.252	138	116.00		0.021	6.25	0.013	
		2	110.267	84	105.00	110.0	8.877	1.110	0.803	0.926
			0.267	138	116.00		0.022	5.99	0.013	
200	10	0	223.196	186	216.00	223.0	4.446	1.111	0.800	0.935
223.607			0.371	262	231.00		0.007	0.291	0.013	
0.01		1	221.380	162	214.00	221.0	4.447	1.112	0.801	0.920
			0.366	255	229.00		0.007	0.257	0.013	
		2	219.880	164	212.00	220.0	4.446	1.112	0.801	0.922
			0.378	272	227.00		0.008	0.234	0.013	
	15	0	223.241	175	216.00	223.0	4.447	1.112	0.801	0.924
			0.369	265	231.00		0.007	0.292	0.013	
		1	221.250	179	214.00	221.0	4.445	1.111	0.801	0.930
			0.369	260	229.00		0.008	0.255	0.013	
		2	219.774	180	213.00	220.0	4.446	1.112	0.801	0.924
			0.368	253	227.00		0.007	0.230	0.013	

Table 3.8. Purely sequential fixed-width confidence intervals from sales data with −14.21 removed from it under estimation strategies (3.1.1)-(3.1.2) with $m = 20, \alpha = 0.05$: Sampling without replacement (SRSWOR) from 351 observations

(3.1.13)	Descriptive stats from n_i obs. at Termination								Conf Int	
i	n_i	$\overline{x}_{n_i}$	s_{n_i}	min	Q_L	med	Q_U	max	$\overline{x}_{n_i} - d$	$\overline{x}_{n_i} + d$
					$d = 0.50$					
0	222	0.41	3.79	−12.61	−2.02	0.76	2.97	8.56	−0.09	0.91
1	204	0.37	3.69	−12.61	−1.89	0.61	2.85	8.56	−0.13	0.87
2	205	0.36	3.69	−12.61	−1.96	0.57	2.84	8.56	−0.14	0.86
					$d = 0.75$					
0	73	0.24	3.25	−9.97	−1.65	0.21	2.44	7.17	−0.51	0.99
1	72	0.27	3.26	−9.97	−1.55	0.54	2.48	7.17	−0.48	1.02
2	73	0.24	3.25	−9.97	−1.65	0.21	2.44	7.17	−0.51	0.99

Table 3.9. Purely sequential minimum risk point estimators from sales data with −14.21 removed from it under estimation strategies (3.2.1)-(3.2.2) with $m = 20, A = 100, c = 0.1, \lambda = 2$: Sampling without replacement (SRSWOR) from 351 observations

(3.2.29)	Descriptive stats from n_i obs. at termination							
i	n_i	$\overline{x}_{n_i}$	s_{n_i}	min	Q_L	med	Q_U	max
0	117	−0.09	3.69	−12.61	−2.120	0.10	2.520	7.17
1	114	−0.22	3.65	−12.61	−2.135	0.07	2.148	7.17
2	113	−0.24	3.66	−12.61	−2.140	0.07	2.140	7.17

Table 3.10. Purely sequential fixed-width confidence intervals from horticulture data under estimation strategies (3.1.1)-(3.1.2) with $m = 20, \alpha = 0.05$: Sampling without replacement (SRSWOR)

(3.1.13)	Descriptive stats from n_i obs. at Termination								Conf Int	
i	n_i	$\overline{x}_{n_i}$	s_{n_i}	min	Q_L	med	Q_U	max	$\overline{x}_{n_i} - d$	$\overline{x}_{n_i} + d$
Variety 2 population size 460: $d = 0.75$										
0	98	35.55	3.78	27.5	33.5	35.5	37.9	46.5	34.80	36.30
1	94	35.45	3.73	27.5	33.1	35.5	37.5	46.5	34.70	36.20
2	90	35.41	3.73	27.5	33.1	35.5	37.5	46.5	34.66	36.16
Variety 3 population size 162: $d = 0.50$										
0	75	27.73	2.21	23.0	26.5	27.5	29.0	34.5	27.23	28.23
1	74	27.73	2.22	23.0	26.5	27.5	29.0	34.5	27.23	28.23
2	75	27.73	2.21	23.0	26.5	27.5	29.0	34.5	27.23	28.23

Table 3.11. Purely sequential minimum risk point estimators from horticulture data under estimation strategies (3.2.1)-(3.2.2) with $m = 20, A = 100, c = 0.1, \lambda = 2$: Sampling without replacement (SRSWOR)

(3.2.29)	Descriptive stats from n_i obs. at Termination							
i	n_i	$\overline{x}_{n_i}$	s_{n_i}	min	Q_L	med	Q_U	max
			Variety 2 population size 460					
0	129	35.17	4.06	24.5	32.0	35.0	38.0	46.5
1	130	35.14	4.06	24.5	31.6	35.0	38.0	46.5
2	132	35.17	4.04	24.5	31.9	35.0	38.0	46.5
			Variety 3 population size 162					
0	75	28.07	2.35	22.5	26.5	28.5	29.5	34.5
1	74	28.09	2.35	22.5	26.5	28.5	29.5	34.5
2	72	28.08	2.36	22.5	26.5	28.5	29.5	34.5

Chapter 4

Two-Stage Methodologies Involving GMD or MAD

In this chapter, we propose the sample variance-based, GMD-based and MAD-based Stein-type (1945,1949) two-stage estimation strategies when a positive lower bound, denoted σ_L^2, for the underlying unknown variance σ^2 is known to us. Both minimum risk point estimation (MRPE) and fixed-width confidence interval estimation (FCIE) problems are investigated. The chapter is based on Mukhopadhyay and Hu (2018).

The structure of this chapter looks similar with that of Chapter 3, but we develop the new Stein-type two-stage estimation strategies for the MRPE problems first in Section 4.1. Section 4.1.1 summarizes the existing two-stage strategies which are relevant in our context, followed by newly developed strategies based on sample standard deviation (indicated by: $i = 0$), GMD (indicated by: $i = 1$), and MAD (indicated by: $i = 2$) in Section 4.1.2. The case "$i = 0$" resembles the well-known methodologies due to Robbins (1959) and Starr (1966).

The main results have been summarized in Theorems 4.1.1-4.1.3. These show that our newly proposed estimation strategies with unbiased estimators of σ based on sample standard deviation, GMD, or MAD in defining the stopping boundaries enjoy both asymptotic first-order and second-order efficiency as well as first-order and second-order risk efficiency properties. In Section 4.1.4, simulated performances are briefly discussed.

Section 4.2 develops new two-stage strategies in the context of the FCIE problems.

The main results regarding asymptotic first-order and second-order efficiency and asymptotic consistency properties have been summarized in Theorems 4.2.1-4.2.3, followed with summaries obtained from simulations in Section 4.2.3.

Section 4.3 includes illustrations with real data under both two-stage MRPE and FCIE strategies. In doing so, we again appropriately use the horticulture data recorded and presented by Mukhopadhyay et al. (2004b), which has appeared in Section 3.3.2.

Section 4.4 summarizes some final overall thoughts.

4.1. Minimum Risk Point Estimation

Recall that Robbins (1959) put forward the MRPE problem for a normal mean μ having an unknown population variance σ^2. He developed a purely sequential stopping rule and established the original MRPE formulation. Starr (1966) extended his idea by appropriately modifying the loss function.

Here, we revisit Stein (1945,1949) and apply his breakthrough two-stage procedure to address the MRPE problem.

4.1.1. Existing Two-Stage MRPE Methodologies

Under the MRPE formulation (2.2.1)-(2.2.3), σ remains unknown even though an expression of n^* is known. A straightforward idea is to use the sample standard deviation as an estimator of σ in the expression of n^* in (2.2.3). With the pilot data of size $m(\geq m_0 \geq 2)$, in the spirit of Ghosh and Mukhopadhyay (1976), a Stein-type two-stage MRPE procedure is proposed as follows:

$$N \equiv N(c) = \max\left\{m, \left\lfloor S_m\sqrt{A/c}\right\rfloor + 1\right\}, \tag{4.1.1}$$

where $\lfloor w \rfloor$ continues to denote the largest integer that is smaller ($<$) than w as appeared in (3.2.10).

Here, m and m_0 are fixed positive integers chosen by an experimenter. Note, however, that while S_m^2 is unbiased for σ^2, the sample standard deviation S_m used in (4.1.1) is biased for σ. Another sticky point is this: How to determine m, the size of the pilot data? We have more to say shortly.

But, whatever may be a fixed choice of m, if we observe $N = m$, we do not gather any additional data. Otherwise, that is if $N > m$, we record additional $(N - m)$ observations in the second stage. Then, having recorded the data $\{N, X_1, ..., X_N\}$ from both stages combined, the sample mean $\overline{X}_N$ is used to estimate the unknown normal mean μ.

Although the Stein-type two-stage strategy (4.1.1) is convenient to implement, the amount of oversampling relative to n^* becomes bothersome. It does not have the efficiency property, let alone the first-order efficiency or first-order risk efficiency properties. One may refer to Ghosh and Mukhopadhyay (1976).

Mukhopadhyay and Duggan (1997) made the observation that if a *positive lower bound* for the normal variance σ^2 is known to us, say $0 < \sigma_L^2 \leq \sigma^2$, then an appropriately updated stein-type two-stage strategy would enjoy both asymptotic first-order and second-order efficiency and risk efficiency properties. Their modification of (4.1.1) is summarized as follows: Let

$$m \equiv m\left(c\right) = \max\left\{m_0, \left\lfloor \sigma_L \sqrt{A/c} \right\rfloor + 1\right\}, \tag{4.1.2}$$

be our determined pilot sample size instead of an experimenter's arbitrarily chosen fixed $m (\geq m_0)$. Observe that the pilot size $m \equiv m(c) \to \infty$ as $c \to 0$, and yet $m \equiv m(c)$ remains small compared with $n^*\left(c\right)$ as $c \to 0$ in the sense that $m(c)/n^*\left(c\right) < 1$ as $c \to 0$.

Let the final sample size be given by:

$$N_0 \equiv N_0(c) = \max\left\{m, \left\lfloor S_m \sqrt{A/c} \right\rfloor + 1\right\}, \tag{4.1.3}$$

obtained from the pilot data $\{m, X_1, ..., X_m\}$ and then we implement the procedure as we did in the case of (4.1.1). Having obtained the combined data

$$\{N_0, X_1, ..., X_{N_0}\},$$

the minimum risk point estimator for μ is proposed to be $\overline{X}_{N_0} = N_0^{-1}\Sigma_{j=1}^{N_0} X_j$.

4.1.2. Substituting Sample Standard Deviation with Alternative Estimators

One should note that in (4.1.3), while S_m is a consistent estimator for σ, it is not unbiased and tends to underestimate σ. In view of the inequality: $\sigma = \sqrt{E_{\mu,\sigma}[S_m^2]} > E_{\mu,\sigma}[S_m]$, it is understood that S_m tends to underestimate σ. Therefore, we should consider an appropriate multiple of S_m, denoted by $V_{m;0}$, to make it unbiased in the normal case by explicitly defining in the spirit of (2.3.4) such that

$$V_{m;0} = \Gamma\left(\tfrac{1}{2}(m-1)\right)\sqrt{m-1}\left\{\sqrt{2}\Gamma\left(\tfrac{1}{2}m\right)\right\}^{-1} S_m.$$

We also use $V_{m;0}$ in the sequel to replace S_m in (4.1.3).

Following (2.3.4), we also propose unbiased estimators for σ based GMD and MAD denoted by $V_{m;i}, i = 1, 2$, respectively. The associated estimators are correspondingly expressed as:

(i) $V_{m;1} = \frac{\sqrt{\pi}}{2} G_m$;

(ii) $V_{m;2} = \sqrt{\frac{\pi m}{2(m-1)}} M_m$, where G_m and M_m come from (2.3.1).

Now we are in a position to propose our new two-stage MRPE strategies and the final sample sizes are precisely given by

$$N_i \equiv N_i(c) = \max\left\{m, \left\lfloor V_{m;i}\sqrt{A/c}\right\rfloor + 1\right\},\ i = 0, 1, 2, \tag{4.1.4}$$

where m was defined in (4.1.2). Having obtained the final dataset $\{N_i, X_1, ..., X_{N_i}\}$, the associated minimum risk point estimators are then defined as follows:

$$\overline{X}_{N_i} = N_i^{-1}\Sigma_{j=1}^{N_i} X_j,\ i = 0, 1, 2. \tag{4.1.5}$$

4.1.3. Properties of the Two-Stage Methodology (4.1.4)-(4.1.5) with m from (4.1.2)

In this section, under the newly developed Stein-type two-stage strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), we first enumerate a number of interesting results summarized by a series of lemmas and theorems.

Lemma 4.1.1. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have for $r \geq 2$:*

$$P_{\mu,\sigma}\{N_i = m\} = O\left(c^{r/4}\right), i = 0, 1, 2, \tag{4.1.6}$$

for sufficiently small c.

Proof. We start with $i = 1$. For sufficiently small c, we have:

$$
\begin{aligned}
&P_{\mu,\sigma}\{N_1 = m\} \\
&\quad = P_{\mu,\sigma}\left\{V_{m;1}\sqrt{A/c} \le m\right\} \\
&\quad = P_{\mu,\sigma}\left\{|V_{m;1} - \sigma| \ge \sigma - \sigma_L\right\} \\
&\quad \le \left(\tfrac{2/\sqrt{\pi}}{\sigma-\sigma_L}\right)^r E_{\mu,\sigma}\left[|G_m - 2\sigma/\sqrt{\pi}|^r\right].
\end{aligned}
\tag{4.1.7}
$$

Thus, we can claim in the spirit of the proof of Lemma 3.2.1: by the lemma of Sen and Ghosh (1981),

$$
E_{\mu,\sigma}\left[|G_m - 2\sigma/\sqrt{\pi}|^r\right] = O\left(m^{-r/2}\right) = O\left(c^{r/4}\right), \text{ for sufficiently small } c. \tag{4.1.8}
$$

Therefore, $P_{\mu,\sigma}\{N_1 = m\} = O\left(c^{r/4}\right)$.

Next, for $i = 2$, we note (w.p.1):

$$
M_m \ge m^{-1}\Sigma_{i=1}^{m}|X_i - \mu| - \left|\overline{X}_m - \mu\right| = M_{1m} - M_{2m}, \text{ say.}
$$

Then, with $\varepsilon = \sigma - \sigma_L$, for sufficiently small c, we can express:

$$
\begin{aligned}
&P_{\mu,\sigma}\{N_2 = m\} \\
&\quad = P_{\mu,\sigma}\{V_{m;2} \le \sigma_L\} \\
&\quad \le P_{\mu,\sigma}\{M_{1m} - M_{2m} \le 2\sigma_L/\sqrt{\pi}\} \\
&\quad \le P_{\mu,\sigma}\left\{|M_{1m} - 2\sigma/\sqrt{\pi}| + |M_{2m}| \ge 2\varepsilon/\sqrt{\pi}\right\} \\
&\quad \le P_{\mu,\sigma}\left\{|M_{1m} - 2\sigma/\sqrt{\pi}| \ge \varepsilon/\sqrt{\pi}\right\} + P_{\mu,\sigma}\left\{\left|\overline{X}_m - \mu\right| \ge \varepsilon/\sqrt{\pi}\right\}.
\end{aligned}
\tag{4.1.9}
$$

Thus we can conclude in a similar way that

$$
P_{\mu,\sigma}\{N_2 = m\} = O\left(c^{r/4}\right). \tag{4.1.10}
$$

Finally, when $i = 0$, for sufficiently small c, we express:

$$\begin{aligned} P_{\mu,\sigma}\{N_0 = m\} &\leq P_{\mu,\sigma}\{S_m \leq \sigma_L\} \leq P_{\mu,\sigma}\{|S_m - \sigma| \geq \sigma - \sigma_L\} \\ &\leq O\left(m^{-r/2}\right) = O\left(c^{r/4}\right), \end{aligned} \tag{4.1.11}$$

by Corollary 1.2 in Mukhopadhyay and Vik (1985). The proof is now complete. ■

Lemma 4.1.2. *For the final sample size* N_i *defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with* m *coming from (4.1.2), for all fixed* μ, σ, *and* A, *we have*:

$$N_i/n^* \overset{P_{\mu,\sigma}}{\to} 1 \text{ as } c \to 0, \tag{4.1.12}$$

for $i = 0, 1, 2$ *where* n^* *was defined in (2.2.3).*

Proof. From (4.1.4), for $i = 0, 1, 2$, we have (w.p.1):

$$V_{m;i}\sqrt{A/c} \leq N_i \leq mI\left(N_i = m\right) + V_{m;i}\sqrt{A/c} + 1.$$

Thus, we can rewrite (w.p.1):

$$\frac{V_{m;i}}{\sigma} \leq \frac{N_i}{n^*} \leq \frac{mI\left(N_i = m\right)\sqrt{c/A} + V_{m;i} + \sqrt{c/A}}{\sigma}. \tag{4.1.13}$$

Next by slightly modifying the proof of Lemma 3.1.2, we can show that $V_{m;i}/\sigma \overset{P_{\mu,\sigma}}{\to} 1$ as $c \to 0$. Also, from Lemma 4.1.1 we have:

$$mI\left(N_i = m\right)\sigma^{-1}\sqrt{c/A} = O_{P_{\mu,\sigma}}\left(c^{r/4}\right),$$

so that we can immediately conclude: $N_i/n^* \overset{P_{\mu,\sigma}}{\to} 1$ as $c \to 0, .i = 0, 1, 2.$ ■

Theorem 4.1.1. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have:*

$$\lim_{c \to 0} E_{\mu,\sigma}[N_i/n^*] = 1, i = 0, 1, 2 \text{ [*Asymptotic First-Order Efficiency*]}, \tag{4.1.14}$$

where n^* was defined in (2.2.3).

Proof. For sufficiently small c, we observe (w.p.1): for $i = 0, 1, 2$,

$$\frac{mI(N_i = m)}{\sigma\sqrt{A/c}} \leq 1, \ \frac{V_{m;i}}{\sigma} \leq 2,$$

so that $N_i/n^* \leq 4$. Now, using Lemma 4.1.2, we obtain (4.1.14). ■

Theorem 4.1.2. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have:*

$$\lim_{c \to 0} \xi_i(c) = 1, i = 0, 1, 2 \text{ [*Asymptotic First-Order Risk Efficiency*]} \tag{4.1.15}$$

with $\xi_i(c)$ defined in (3.2.6).

Proof. In view of Theorem 4.1.1, it will suffice to prove that $\lim_{c \to 0} E_{\mu,\sigma}[n^*/N_i] = 1$ for $i = 0, 1, 2$. Now, we show outlines of the proof in the case of three different choices of i. Observe that

(i) $i = 0$: We have

$$P_{\mu,\sigma}\left\{N_0 \leq \tfrac{1}{2}n^*\right\} \leq P_{\mu,\sigma}\left\{V_{m;0}\sqrt{A/c} \leq \tfrac{1}{2}\sigma\sqrt{A/c}\right\} \leq P_{\mu,\sigma}\left\{|S_m - \sigma| \geq \tfrac{1}{2}\sigma\right\}; \tag{4.1.16}$$

(ii) $i = 1$: We have

$$P_{\mu,\sigma}\left\{N_1 \le \tfrac{1}{2}n^*\right\} \le P_{\mu,\sigma}\left\{V_{m;1}\sqrt{A/c} \le \tfrac{1}{2}\sigma\sqrt{A/c}\right\} \le P_{\mu,\sigma}\left(|V_{m;1}-\sigma| \ge \tfrac{1}{2}\sigma\right); \tag{4.1.17}$$

(iii) $i = 2$: Recall the expressions of M_{1m} and M_{2m} from (4.1.9) and write

$$\begin{aligned}
&P_{\mu,\sigma}\left\{N_2 \le \tfrac{1}{2}n^*\right\} \\
&\quad\le P_{\mu,\sigma}\left\{V_{k;2} \le \tfrac{1}{2}\sigma\right\} \\
&\quad\le P_{\mu,\sigma}\left\{M_{1m} - 2\sigma/\sqrt{\pi} - M_{2m} \le \tfrac{\sqrt{2}-4}{2\sqrt{\pi}}\sigma\right\} \\
&\quad\le P_{\mu,\sigma}\left\{|M_{1m} - 2\sigma/\sqrt{\pi}| \ge \tfrac{4-\sqrt{2}}{4\sqrt{\pi}}\sigma\right\} + P_{\mu,\sigma}\left\{|M_{2m}| \ge \tfrac{4-\sqrt{2}}{4\sqrt{\pi}}\sigma\right\}.
\end{aligned} \tag{4.1.18}$$

Thus, we can claim $P_{\mu,\sigma}\left\{N_i \le \frac{1}{2}n^*\right\} = O\left(c^{r/4}\right)$ for all $r > 2, i = 0, 1, 2$.

At this point, we can conclude that $\lim_{c\to 0}\xi_i(c) = 1, i = 0, 1, 2$ in the same way as we did in the proof of Theorem 3.2.2. ■

Theorem 4.1.3. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have:*

$$\lim_{c\to 0} E_{\mu,\sigma}[N_i - n^*] \text{ is bounded [\textit{Asymptotic Second-Order Efficiency}]}, \tag{4.1.19}$$

for $i = 0, 1, 2$ where n^* was defined in (2.2.3).

Proof. Note that for $i = 0, 1, 2$,

$$(V_{m;i} - \sigma)\sqrt{A/c} \le N_i - n^* \le mI(N_i = m) + (V_{m;i} - \sigma)\sqrt{A/c} + 1, \text{ w.p.1.} \tag{4.1.20}$$

We immediately have:

$$0 \le E_{\mu,\sigma}[N_i - n^*] \le 1 + O\left(c^{\frac{r-2}{4}}\right), \tag{4.1.21}$$

provided that $r > 2$. Thus, (4.1.19) holds. ∎

Lemma 4.1.3. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have:*

$$n^{*-1/2}\left(N_i - n^*\right) \xrightarrow{\pounds} N(0, \delta_i^2 \sigma \sigma_L^{-1}) \text{ as } c \to 0, i = 0, 1, 2, \tag{4.1.22}$$

where $\delta_0^2 = \frac{1}{2}$, $\delta_1^2 = \frac{1}{3}(\pi + 6\sqrt{3} - 12)$, $\delta_2^2 = \frac{1}{2}(\pi - 2)$ *and n^* was defined in (2.2.3).*

Proof. From (4.1.20), we can express (w.p.1): for $i = 0, 1, 2$,

$$\begin{aligned} \sigma^{-1/2}\left(A/c\right)^{1/4}\left(V_{m;i} - \sigma\right) \le n^{*-1/2}\left(N_i - n^*\right) \le \sigma^{-1/2}\left(A/c\right)^{1/4}\left(V_{m;i} - \sigma\right) \\ + \sigma^{-1/2}\left(A/c\right)^{-1/4}\left\{mI\left(N_i = m\right) + 1\right\}. \end{aligned} \tag{4.1.23}$$

Now, $\sigma^{-1/2}\left(A/c\right)^{-1/4}\left\{mI\left(N_i = m\right) + 1\right\} \overset{P_{\mu,\sigma}}{\to} 0$ as $c \to 0$, so that by Slutsky's theorem, for sufficiently small c, the asymptotic distribution of $n^{*-1/2}\left(N_i - n^*\right)$ is the same as that of $\left(\sigma\sigma_L\right)^{-1/2} m^{1/2}\left(V_{m;i} - \sigma\right)$.

(i) $i = 0$: With S_m^2, we have as $m \to \infty$:

$$n^{1/2}\left(S_m^2 - \sigma^2\right) \xrightarrow{\pounds} N(0, 2\sigma^4),$$

so that by the delta method, one gets:

$$n^{1/2}\left(S_m - \sigma\right) \xrightarrow{\pounds} N(0, \tfrac{1}{2}\sigma^2).$$

Thus,

$$(\sigma\sigma_L)^{-1/2}\, m^{1/2}\, (V_{m;0} - \sigma) \xrightarrow{\pounds} N(0, \delta_0^2 \sigma\sigma_L^{-1}),$$

as $c \to 0$, where $\delta_0^2 = \frac{1}{2}$.

(ii) $i = 1$: Since G_m is a U-statistic, from Hoeffding (1948,1961), we have as $m \to \infty$,

$$m^{1/2}\, (G_m - 2\sigma/\sqrt{\pi}) \xrightarrow{\pounds} N(0, 4\sigma_1^2),$$

where $\sigma_1^2 = \sigma^2 \left(\frac{1}{3} + \frac{2\sqrt{3}-4}{\pi} \right)$. Thus, we have:

$$(\sigma\sigma_L)^{-1/2}\, m^{1/2}\, (V_{m;1} - \sigma) \xrightarrow{\pounds} N(0, \delta_1^2 \sigma\sigma_L^{-1}),$$

as $c \to 0$, where $\delta_1^2 = \frac{1}{3}(\pi + 6\sqrt{3} - 12)$.

(iii) $i = 2$: By appealing to Babu and Rao (1992), we have as $m \to \infty$:

$$n^{1/2} \left(M_m - \sigma\sqrt{2/\pi} \right) \xrightarrow{\pounds} N(0, \tfrac{\pi-2}{\pi}\sigma^2).$$

Thus, we can claim:

$$(\sigma\sigma_L)^{-1/2}\, m^{1/2}\, (V_{m;2} - \sigma) \xrightarrow{\pounds} N(0, \delta_2^2 \sigma\sigma_L^{-1}),$$

as $c \to 0$, where $\delta_2^2 = \frac{1}{2}(\pi - 2)$. ■

Lemma 4.1.4. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have*:

$$n^{*-1}\, (N_i - n^*)^2\,, i = 0, 1, 2 \textit{ is uniformly integrable,}$$

where n^* was defined in (2.2.3).

Proof. The necessary steps can be developed by utilizing (4.1.23) repeatedly in the spirits of Mukhopadhyay and Duggan (1997,1999). Details are left out for brevity. ∎

Theorem 4.1.4. *For the final sample size N_i defined by the two-stage estimation strategy (4.1.4)-(4.1.5) with m coming from (4.1.2), for all fixed μ, σ, and A, we have:*

$$\omega_i(c) = \delta_i^2 \sigma \sigma_L^{-1} c + o(c), i = 0, 1, 2 \text{ [\textit{Asymptotic Second-Order Risk Efficiency}]}, \tag{4.1.24}$$

where $\omega_i(c)$ was defined in (3.2.6) and $\delta_0^2 = \frac{1}{2}$, $\delta_1^2 = \frac{1}{3}(\pi + 6\sqrt{3} - 12)$, $\delta_2^2 = \frac{1}{2}(\pi - 2)$ remain the same as in Lemma 4.1.3.

Proof. We express: for $i = 0, 1, 2$,

$$\begin{aligned} \omega_i(c) &= cE_{\mu,\sigma}\left\{N_i^{-1}(N_i - n^*)^2 I(N_i = m)\right\} \\ &+ cE_{\mu,\sigma}\left\{N_i^{-1}(N_i - n^*)^2 I(N_i > m)\right\}. \end{aligned} \tag{4.1.25}$$

On the set $\{N_i > m\}$, we observe (w.p.1):

$$N_i^{-1}(N_i - n^*)^2 = \frac{n^*}{\left\lfloor V_{m;i}\sqrt{A/c} \right\rfloor + 1} \frac{(N_i - n^*)^2}{n^*}.$$

Next, we note (w.p.1):

$$\frac{\sigma}{V_{m;i} + \sqrt{c/A}} \leq \frac{n^*}{\left\lfloor V_{m;i}\sqrt{A/c} \right\rfloor + 1} \leq \frac{\sigma}{V_{m;i}}.$$

Since both $\frac{\sigma}{V_{m;i} + \sqrt{c/A}}$ and $\frac{\sigma}{V_{m;i}}$ converge to 1 in probability($P_{\mu,\sigma}$) as $c \to 0$, we can claim that for sufficiently small c, $\frac{n^*}{\lfloor V_{m;i}\sqrt{A/c}\rfloor+1} \le 2$. Thus, $N_i^{-1}(N_i-n^*)^2 I\,(N_i > m)$ is uniformly integrable and then for sufficiently small c, we get:

$$E_{\mu,\sigma}\left\{N_i^{-1}(N_i - n^*)^2 I\,(N_i > m)\right\} = \delta_i^2\sigma\sigma_L^{-1}. \tag{4.1.26}$$

Next, on the set $\{N_i = m\}$, we can express:

$$E_{\mu,\sigma}\left\{N_i^{-1}(N_i - n^*)^2 I\,(N_i = m)\right\} \le O\,(m)\, P_{\mu,\sigma}\{N_i = m\} = o\,(1)\,. \tag{4.1.27}$$

Combining (4.1.26) and (4.1.27), the proof of (4.1.24) is complete. ■

4.1.4. Simulations and Data Analysis

In this section, we summarize simulated performances of the two-stage MRPE methodology $(N_i, \overline{X}_{N_i}), i = 0, 1, 2$ from (4.1.4)-(4.1.5) with m coming from (4.1.2), where the index i identifies each strategy under implementation:

$$\begin{array}{ll} i=0: & \text{Sample standard deviation-based methodology;} \\ i=1: & \text{GMD-based methodology;} \\ i=2: & \text{MAD-based methodology.} \end{array} \tag{4.1.28}$$

We set $\mu = 5$ and $\sigma = 4$, and thus generated pseudorandom samples from a $N(5, 4^2)$ population. For convenience, we fixed $m_0 = 10, A = 100$, and $c = 1.00, 0.16, 0.04, 0.01$ so that the optimal fixed sample size n^* from (2.2.3) would cover a range of values including $40, 100, 200, 400$, that is from small to medium to large.

Further, we had assumed that a prior lower bound of the variance was known to us: We considered $\sigma_L = 1, 2, 3$. Table 4.1 summarizes the findings in these situations obtained under $T\,(=10000$, say) independent replications. We report the average final sample size $\overline{n}$ along with its estimated standard error $s(\overline{n})$, the average risk $\overline{R}_n$ along with its estimated standard error $s(\overline{R}_n)$, estimated risk efficiency $\left(\widehat{\xi}\right)$ and regret $(\widehat{\omega})$ explained in (3.2.40) as well as other entities explained in (4.1.29).

$$\eta_i c \text{ where } \eta_i = \delta_i^2 \sigma \sigma_L^{-1}, \text{ for } i = 0,1,2: \quad \text{a theoretical approximation of the } \omega_i\left(c\right). \tag{4.1.29}$$

In Table 4.1, we show 3 blocks corresponding to assigned values of σ_L. Within each block, there is little or no significant difference among three procedures $(i = 0,1,2)$, no matter whether the sample sizes were small (40), moderate (100), or large $(200, 400)$. We note from the fifth column that all $\overline{n}$ values appear a bit over the corresponding optimal fixed sample sizes n^*, but within one observation. These empirically validate Theorem 4.1.3 about the asymptotic second-order efficiency property from equation (4.1.21).

$\widehat{\xi}$ from column 9 obtained using column 7 for $\overline{R}_n$ reflects the asymptotic first-order risk efficiency property which turns out to remain tightly around 1. Upon comparing the different blocks, it becomes obvious that for larger lower bound σ_L for σ, the corresponding $\widehat{\omega}$ values in the last column become smaller and get closer to 0. The estimated regret values stay close to the theoretical ones shown in Theorem 4.1.4 and they compare well with the asymptotic second-order terms, namely, $\eta_i c = \delta_i^2 \sigma \sigma_L^{-1} c, i = 0,1,2$. All three procedures $(i = 0,1,2)$ appear to perform remarkably well.

4.2. Fixed-Width Confidence Intervals

The fundamental breakthrough ideas of Stein (1945,1949) have been pushed by many researchers to new heights. Mukhopadhyay and Duggan (1997) and Chattopadhyay and Mukhopadhyay (2013) proceeded to propose new FCIE strategies with practical modifications. We continue to suppose that we want to estimate the normal mean μ using a confidence interval having a fixed-width $2d$, $d > 0$, with its associated preassigned confidence level at least $(1-\alpha)$, $0 < \alpha < 1$ under the framework of (2.1.1)-(2.1.3).

4.2.1. A Formulation

While σ^2 remains unknown in (2.1.3), we use sample variance and GMD-based or MAD-based *unbiased* estimators of σ^2, defined in (2.3.1)-(2.3.3) such that with $m(\geq m_0 \geq 2)$:

$$U_{m;i}^2 = \begin{cases} S_m^2, & \text{corresponding to } i = 0 \\ c_{m;1}^{-2} G_m^2, & \text{corresponding to } i = 1 \\ c_{m;2}^{-2} M_m^2, & \text{corresponding to } i = 2 \end{cases} \cdot \tag{4.2.1}$$

Thus far, m and m_0 have been fixed positive integers chosen arbitrarily by an experimenter. But, how to determine m, the size of the pilot data? Under the assumption that $\sigma > \sigma_L > 0$ where σ_L is known, we propose the following choice of m and the associated Stein-type two-stage strategies. Let us define:

$$\begin{aligned} m \equiv m(d) &= \max\{m_0, \lfloor a^2 \sigma_L^2 / d^2 \rfloor + 1\}, \\ N_i \equiv N_i(d) &= \max\{m, \left\lfloor a^2 U_{m;i}^2 / d^2 \right\rfloor + 1\}, \end{aligned} \tag{4.2.2}$$

and accordingly construct the fixed-width confidence intervals:

$$J_{N_i} = \left[\overline{X}_{N_i} \pm d\right], \tag{4.2.3}$$

for the unknown mean μ based on the combined data $\{N_i, X_1, ..., X_{N_i}\}$, for $i = 0, 1, 2$.

We should mention that for $i = 0$, Mukhopadhyay and Duggan (1997) used $b^2_{m;0}$, the upper $100\,(\alpha/2)\,\%$ point of the Student's t distribution with degrees of freedom $(m-1)$ instead of a^2 to define N_0 in (4.2.3). Also, for $i = 1, 2$, Chattopadhyay and Mukhopadhyay (2013) used respectively $b^2_{m;1}$ and $b^2_{m;2}$, the upper $100\,(\alpha/2)\,\%$ point of the corresponding pivotal distribution of the sample mean standardized appropriately by $U_{m;1}$ and $U_{m;2}$, instead.

While our present substitution of a^2 in the place of $b^2_{m;0}, b^2_{m;1}$, and $b^2_{m;2}$ may lead to minor loss in the coverage probability, the newly proposed strategies (4.2.2)-(4.2.3) will continue to enjoy attractive asymptotic first-order and second-order properties. The oversampling problem would also be less pronounced. Such performances are summarized in the following sections.

4.2.2. Properties of the Two-Stage Methodology (4.2.2)-(4.2.3)

Under the strategies (4.2.2)-(4.2.3), we have the following result in the spirit of Lemma 4.1.1. We state it without giving its proof.

Lemma 4.2.1. *For the final sample size N_i defined by the two-stage estimation strategy (4.2.2)-(4.2.3), for all fixed μ, σ, α and sufficiently small d, we have*:

$$P_{\mu,\sigma}\{N_i = m\} = O\left(d^r\right), i = 0, 1, 2, \tag{4.2.4}$$

where $r \geq 2$.

Now, we move forward to formally summarize the following asymptotic first-order and second-order properties. Some steps in the proofs are omitted for brevity since they may be constructed similarly along the lines of those which were shown in the case of the MRPE problem in Section 4.1.

Theorem 4.2.1. *For the final sample size N_i defined by the two-stage estimation strategy (4.2.2)-(4.2.3), for all fixed μ, σ, α and sufficiently small d, we have:*

$$\lim_{d \to 0} E_{\mu,\sigma} \left[N_i / C \right] = 1, i = 0, 1, 2 \text{ [Asymptotic First-Order Efficiency]}, \tag{4.2.5}$$

where C was defined in (2.1.3).

Theorem 4.2.2. *For the final sample size N_i defined by the two-stage estimation strategy (4.2.2)-(4.2.3), for all fixed μ, σ, α and sufficiently small d, we have:*

$$\lim_{d \to 0} P_{\mu,\sigma} \left\{ \mu \in J_{N_i} \right\} = 1 - \alpha, i = 0, 1, 2 \text{ [Asymptotic Consistency]}, \tag{4.2.6}$$

where J_{N_i} was defined in (4.2.3).

Proof. For $i = 0, 1, 2$, one should note that for all $n \geq m$, $I(N_i = n)$ depends only on $U_{m;i}^2$ which are independent of $\overline{X}_n$. Thus, we can write:

$$P_{\mu,\sigma} \left\{ \mu \in J_{N_i} \right\} = E_{\mu,\sigma} \left[2\Phi \left(N_i^{1/2} d/\sigma \right) - 1 \right], \tag{4.2.7}$$

which converges to $1 - \alpha$ as $d \to 0$ in view of the dominated convergence theorem. ∎

Theorem 4.2.3. *For the final sample size N_i defined by the two-stage estimation strategy (4.2.2)-(4.2.3), for all fixed μ, σ, α and sufficiently small d, we have:*

(*i*) $\lim_{d \to 0} E_{\mu,\sigma} \left[N_i - C \right]$ *is bounded [Asymptotic Second-Order Efficiency]*;

(*ii*) $C^{-1/2} \left(N_i - C \right) \xrightarrow{\mathcal{L}} N(0, 4\delta_i^2 \sigma^2 \sigma_L^{-2})$ as $d \to 0$, *where δ_i^2 are defined in (3.2.22)*;

(*iii*) $C^{-1} \left(N_i - C \right)^2$ *is uniformly integrable*;

(4.2.8)

for $i = 0, 1, 2$ *where* C *was defined in (2.1.3).*

Proof. For $i = 0, 1, 2$, in part (i), for small enough d, we note (w.p.1):

$$a^2 \left(U_{m;i}^2 - \sigma^2 \right) / d^2 \leq N_i - C \leq mI \left(N_i = m \right) + a^2 \left(U_{m;i}^2 - \sigma^2 \right) / d^2 + 1, \qquad (4.2.9)$$

where $mI\left(N_i = m\right) = O\left(d^{r-2}\right)$ provided that $r > 2$. Thus, as $d \to 0$, we have:

$$0 \leq E_{\mu,\sigma} \left[N_i - C \right] \leq 1 + O\left(d^{r-2}\right). \qquad (4.2.10)$$

For part (ii), we observe that as $d \to 0$, $C^{-1/2}\left(N_i - C\right)$ has the same asymptotic distribution as that of $\left(\sigma\sigma_L\right)^{-1} m^{1/2}\left(U_{m;i}^2 - \sigma^2\right)$. With delta method in place, we can claim Part (ii). Part (iii) can be derived similarly as in Lemma 4.1.4. ■

4.2.3. Simulations and Data Analysis

In this section, we summarize performances of the FCIE strategies $(N_i, J_{N_i}), i = 0, 1, 2$ from (4.2.2)-(4.2.3), where the index i is used to identify each procedure under implementation as before. We use the previous settings described in Section 4.1.4 with $\mu = 5$ and $\sigma = 4$, and assume that the known lower bound σ_L varies through $\sigma_L = 1, 2, 3$. We also fixed $m_0 = 10$ as what we did in Section 4.1.4.

For brevity, we consider $\alpha = 0.05$ only, and set the optimal fixed sample size C across a range of small, moderate and large values, including $50, 100, 200, 500$, and 1000 so that we are able to compare simulated performances under a variety of sample sizes.

The values of d can be determined by referring to (2.1.3). Table 4.2 summarizes the findings in these situations under $T\,(= 10000$, say) independent replications. We obtained the average final sample size $\overline{n}$ along with its estimated standard error $s\left(\overline{n}\right)$, the average coverage probability $\overline{p}$ along with its estimated standard error $s(\overline{p})$.

The observed values of $\overline{n}/C$ and $\overline{n}-C$, associated with first-order and second-order terms, are both included. We use the basic notation similar in spirit with those reported in (3.1.14).

In Table 4.2, within each block corresponding to different values of σ_L, there is little to no significant difference among three procedures, when the sample sizes increased from small (50) to medium (100, 200) to large (500, 1000). We note that our new two-stage strategies do not suffer from noticeable oversampling problem. This is clear since the $\overline{n}$ values from column 5 appear close to the corresponding optimal fixed sample sizes shown in column 2.

The simulations demonstrate that the consistency indeed holds as an asymptotic feature instead of an exact property seen from column 7 shown for $\overline{p}$. The average coverage probabilities appear close to the target level $0.95\,(=1-\alpha)$ when sample sizes increased. We also note that as σ_L increased from 1 to 3, the standard error of $\overline{n}$ dropped and the average coverage probability increased.

4.3. Illustrations with Real Datasets

In this section, we conduct further data analyses and illustrate both MRPE and FCIE methodologies for the normal mean μ based on methodologies (4.1.4)-(4.1.5) with m from (4.1.2) and (4.2.2)-(4.2.3). Here, we use the horticulture data that comes from Mukhopadhyay et al. (2004b) again, referred to as "horticulture data" recording the number of days that the marigold variety 2 seeds needed to flower. It is assumed that the data followed a normal distribution with a lower bound of standard deviation at least 2.0.

For purposes of illustrations, we consider this data set of size 460 with $\mu = 35.13$ and $\sigma = 4.03$ as a representative of the marigold variety 2 seeds population. In the spirit of bootstrap, we draw random samples from this "population" under simple random sampling with replacement (SRSWR). Since the population is rather large, simple random sampling without replacement (SRSWOR) would produce comparable results. Fixing $m_0 = 10$, we include both MRPE and FCIE methodologies from Sections 4.3.1 and 4.3.2, respectively.

4.3.1. MRPE Methodologies

Here, we take advantage of the following bootstrap technique to help us improve the quality of estimators from strategies (4.1.4)-(4.1.5) with m from (4.1.2). We first fix $A = 100$ and $c = 0.1$. With the known lower bound of standard deviation $\sigma_L = 2$, we are able to determine the pilot sample size m. Then, we applied all three two-stage methodologies (4.1.4)-(4.1.5), namely sample standard deviation-based ($i = 0$), GMD-based ($i = 1$), and MAD-based ($i = 2$), to obtain the final sample size $N = n$.

Then, we selected extra $(N - m)$ observations if needed (that is, if $N > m$) from the population to record our observed sample data to be bootstrapped with $b = 200$ replications. With each bootstrap sample, we can calculate the mean and variance. The detailed notation is provided below:

$$
\begin{aligned}
B_r &: \text{ the } r^{\text{th}} \text{ bootstrap sample;} \\
n &: \text{ the bootstrap sample size;} \\
x_{rj} &: \text{ the } j^{\text{th}} \text{ observation in the } r^{\text{th}} \text{ bootstrap sample} \\
\overline{x}_r = \tfrac{1}{n}\Sigma_{j=1}^{n} x_{rj} &: \text{ the } r^{\text{th}} \text{ bootstrap mean estimate;} \\
\widehat{\sigma}_r^2 = \tfrac{1}{n}\Sigma_{j=1}^{n} \left(x_{rj} - \overline{x}_r\right)^2 &: \text{ the } r^{\text{th}} \text{ bootstrap variance estimate;} \\
\overline{x}_{BS} = \tfrac{1}{b}\Sigma_{r=1}^{b} \overline{x}_r &: \text{ the bootstrap estimate of mean;} \\
\widehat{\sigma}_{BS}^2 = \tfrac{1}{b}\Sigma_{r=1}^{b} \widehat{\sigma}_r^2 &: \text{ the bootstrap estimate of variance;} \\
\widehat{\sigma}_{BS} &: \text{ the bootstrap estimate of standard deviation.}
\end{aligned} \tag{4.3.1}
$$

Table 4.3 summarizes the descriptive statistics for $i = 0, 1, 2$, associated with the strategies (4.1.4)-(4.1.5) with m from (4.1.2). The three different procedures yielded results with little to no difference.

4.3.2. FCIE Methodologies

Having fixed $\alpha = 0.05$, $d = 0.5$, and $\sigma_L = 2$, we were able to determine the pilot sample size m. As in Section 4.3.1, we again applied the bootstrap algorithm. Within each bootstrap, in addition to (4.3.1), we also included the following entities:

$$\begin{aligned} J_r = [\overline{x}_r \pm d] &\quad : \text{the } r^{\text{th}} \text{ bootstrap fixed-width confidence interval;} \\ p_r &\quad : 1\,(0) \text{ if } \mu \in (\notin)\ J_r; \\ \widehat{p}_{BS} = \tfrac{1}{b}\Sigma_{r=1}^{b} p_r &\quad : \text{should be comparable with } 1-\alpha; \\ s\left(\widehat{p}_{BS}\right) = \left\{b^{-1}\widehat{p}_{BS}\left(1-\widehat{p}_{BS}\right)\right\}^{1/2} &\quad : \text{estimated standard error of } \widehat{p}_{BS}. \end{aligned} \tag{4.3.2}$$

Table 4.4 summarizes the descriptive statistics for $i = 0, 1, 2$, associated with the strategies (4.2.2)-(4.2.3). Again, the three different procedures yielded comparable results with little to no difference.

4.4. Overall Concluding Thoughts

In summary, the Stein-type two-stage MRPE and FCIE methodologies from (4.1.4)-(4.1.5) with m from (4.1.2) and (4.2.2)-(4.2.3) have performed remarkably well, when we used unbiased estimators of σ $\left(\text{or } \sigma^2\right)$ involving the sample standard deviation (or sample variance), GMD or MAD in defining the stopping boundaries, under the assumption normality. We did not notice appreciable differences in the overall performances of the proposed procedures in the context of either point estimation problem or the interval estimation problem.

Table 4.1. Simulations from $N(5, 4^2)$ under $A = 100$, $m_0 = 10$ with $T \equiv 10000$ runs implementing methods (4.1.4)-(4.1.5) with m from (4.1.2) and i from (4.1.28)

σ_L	n^*	c	i	$\overline{n}$	$s(\overline{n})$	$\overline{R}_n$	$s(\overline{R}_n)$	$\widehat{\xi}$	$\widehat{\omega}$
				$\eta_0 = 2, \eta_1 = 2.045, \eta_2 = 2.283$					
1	40	1	0	40.452	0.096	80.091	0.101	1.001	2.456
			1	40.547	0.096	80.028	0.102	1.000	2.488
			2	40.539	0.101	80.275	0.104	1.003	2.761
	100	0.16	0	100.499	0.145	32.013	0.024	1.000	0.348
			1	100.332	0.146	32.037	0.023	1.001	0.350
			2	100.509	0.153	32.071	0.024	1.002	0.384
	200	0.04	0	200.216	0.201	15.991	0.008	0.999	0.082
			1	200.043	0.202	15.998	0.008	1.000	0.083
			2	200.266	0.215	16.011	0.008	1.001	0.093
	400	0.01	0	400.691	0.283	8.004	0.003	1.001	0.020
			1	400.836	0.288	8.002	0.003	1.000	0.021
			2	400.161	0.302	8.004	0.003	1.000	0.023
				$\eta_0 = 1, \eta_1 = 1.023, \eta_2 = 1.142$					
2	40	1	0	40.529	0.065	78.953	0.095	0.987	1.095
			1	40.538	0.066	78.900	0.095	0.986	1.115
			2	40.551	0.070	78.960	0.097	0.987	1.235
	100	0.16	0	100.442	0.100	31.828	0.023	0.995	0.163
			1	100.360	0.101	31.822	0.023	0.994	0.167
			2	100.543	0.107	31.881	0.023	0.996	0.184
	200	0.04	0	200.390	0.143	15.954	0.008	0.997	0.041
			1	200.448	0.145	15.959	0.008	0.997	0.042
			2	200.418	0.152	15.956	0.008	0.997	0.047
	400	0.01	0	400.598	0.201	7.993	0.003	0.999	0.010
			1	400.316	0.202	7.990	0.003	0.999	0.010
			2	400.472	0.213	7.991	0.003	0.999	0.011
				$\eta_0 = 0.667, \eta_1 = 0.682, \eta_2 = 0.761$					
3	40	1	0	40.545	0.052	78.646	0.093	0.983	0.672
			1	40.582	0.053	78.686	0.093	0.984	0.687
			2	40.596	0.055	78.849	0.092	0.986	0.739
	100	0.16	0	100.562	0.083	31.813	0.023	0.994	0.110
			1	100.561	0.083	31.819	0.023	0.994	0.111
			2	100.595	0.086	31.839	0.023	0.995	0.119
	200	0.04	0	200.490	0.116	15.943	0.008	0.996	0.027
			1	200.451	0.116	15.943	0.008	0.996	0.027
			2	200.438	0.123	15.948	0.008	0.997	0.030
	400	0.01	0	400.371	0.163	7.987	0.003	0.998	0.007
			1	400.366	0.165	7.985	0.003	0.998	0.007
			2	400.413	0.174	7.985	0.003	0.998	0.008

Table 4.2. Simulations from $N(5, 4^2)$ under $\alpha = 0.05, m_0 = 10$ with $T \equiv 10000$ runs implementing methods (4.2.2)-(4.2.3)

σ_L	C	d	i	$\overline{n}$	$s(\overline{n})$	$\overline{p}$	$s(\overline{p})$	$\overline{n}/C$	$\overline{n}-C$
1	50	1.1087	0	50.550	0.236	0.919	0.003	1.011	0.550
			1	50.297	0.235	0.916	0.003	1.006	0.297
			2	50.778	0.253	0.912	0.003	1.016	0.778
	100	0.7840	0	100.733	0.473	0.914	0.003	1.007	0.733
			1	100.501	0.483	0.914	0.003	1.005	0.501
			2	101.031	0.506	0.912	0.003	1.010	1.031
	200	0.5544	0	201.590	0.817	0.928	0.003	1.008	1.590
			1	200.736	0.821	0.924	0.003	1.004	0.736
			2	200.218	0.860	0.923	0.003	1.001	0.218
	500	0.3506	0	501.255	1.273	0.943	0.002	1.003	1.255
			1	502.996	1.287	0.942	0.002	1.006	2.996
			2	503.503	1.354	0.944	0.002	1.007	3.503
	1000	0.2479	0	1000.787	1.785	0.946	0.002	1.001	0.787
			1	1002.184	1.811	0.948	0.002	1.002	2.184
			2	1000.102	1.944	0.945	0.002	1.000	0.102
2	50	1.1087	0	50.643	0.204	0.930	0.003	1.013	0.643
			1	50.495	0.208	0.931	0.003	1.010	0.495
			2	50.370	0.217	0.927	0.003	1.007	0.370
	100	0.7840	0	100.165	0.286	0.938	0.002	1.002	0.165
			1	100.559	0.292	0.936	0.002	1.006	0.559
			2	100.994	0.309	0.933	0.003	1.010	0.994
	200	0.5544	0	199.635	0.399	0.943	0.002	0.998	−0.366
			1	200.346	0.404	0.942	0.002	1.002	0.346
			2	200.367	0.430	0.943	0.002	1.002	0.367
	500	0.3506	0	501.199	0.635	0.947	0.002	1.002	1.199
			1	499.808	0.641	0.952	0.002	1.000	−0.192
			2	500.253	0.676	0.950	0.002	1.001	0.253
	1000	0.2479	0	1002.160	0.904	0.952	0.002	1.002	2.160
			1	1001.465	0.917	0.951	0.002	1.001	1.465
			2	1001.411	0.958	0.947	0.002	1.001	1.411

Table 4.2 Continued. Simulations from $N(5, 4^2)$ under $\alpha = 0.05$, $m_0 = 10$ with $T \equiv 10000$ runs implementing methods (4.2.2)-(4.2.3)

σ_L	C	d	i	$\overline{n}$	$s(\overline{n})$	$\overline{p}$	$s(\overline{p})$	$\overline{n}/C$	$\overline{n} - C$
3	50	1.1087	0	50.562	0.133	0.939	0.002	1.011	0.562
			1	50.549	0.133	0.940	0.002	1.011	0.549
			2	50.580	0.139	0.941	0.002	1.012	0.580
	100	0.7840	0	100.675	0.189	0.946	0.002	1.007	0.675
			1	100.543	0.192	0.945	0.002	1.005	0.543
			2	100.476	0.204	0.941	0.002	1.005	0.476
	200	0.5544	0	200.632	0.266	0.951	0.002	1.003	0.632
			1	200.242	0.270	0.948	0.002	1.001	0.242
			2	200.512	0.286	0.947	0.002	1.003	0.512
	500	0.3506	0	501.100	0.425	0.950	0.002	1.002	1.100
			1	500.648	0.428	0.949	0.002	1.001	0.648
			2	500.598	0.452	0.949	0.002	1.003	0.512
	1000	0.2479	0	1000.530	0.594	0.948	0.002	1.001	0.530
			1	1000.842	0.607	0.948	0.002	1.001	0.842
			2	1000.426	0.650	0.949	0.002	1.000	0.426

Table 4.3. Point estimators from horticulture data under MRPE methods (4.1.4)-(4.1.5) with m from (4.1.2) with $\sigma_L = 2$

i	N	$\overline{x}_{BS}$	$\widehat{\sigma}_{BS}$
	$A = 100, c = 0.1,$ $m_0 = 10, b = 200$		
0	112	34.811	4.146
1	113	34.785	3.884
2	113	34.785	3.884

Table 4.4. Confidence intervals from horticulture data under FCIE methods (4.2.2)-(4.2.3) with $\sigma_L = 2$

i	N	$\overline{x}_{BS}$	$\widehat{\sigma}_{BS}$	$\widehat{p}_{BS}$	$s(\widehat{p}_{BS})$
$\alpha = 0.05, d = 0.5, m_0 = 10, b = 200$					
0	196	35.027	3.818	0.940	0.0168
1	198	35.073	3.908	0.925	0.0186
2	199	35.087	3.979	0.930	0.0180

Chapter 5

Second-Order Asymptotics in a Class of Purely Sequential MRPE Methodologies

As Mukhopadhyay (1982) had suggested using estimators of σ^2 other than the sample variance, for instance, Chattopadhyay and Mukhopadhyay (2013) incorporated the GMD, MAD and range to construct their two-stage FCIE methodologies. Following these ideas, we may consider other alternative estimators of σ^2(or σ), propose new stopping boundary conditions, and construct even more robust multi-stage methodologies. Chapter 5 is based on the work Hu and Mukhopadhyay (2018).

In Section 5.1, we propose a new class of purely sequential MRPE methodologies, which involve general robust estimators of σ in defining the stopping boundaries. Some appropriate sufficient conditions on such estimators are assumed.

In Section 5.2, we provide both asymptotic first-order and second-order properties enjoyed by the associated methodologies as our main results, summarized in Theorems 5.1.1-5.1.3.

In Section 5.3, we revisit the GMD-based and MAD-based purely sequential MRPE methodologies described earlier in Section 3.2 and develop asymptotic second-order results which are considerably stronger than those reported in Mukhopadhyay and Hu (2017). More specifically, for the purely sequential MRPE methodologies established in (3.2.1)-(3.2.2), we propose to derive second-order approximations of the regret $\omega_i(c), i = 0, 1, 2$. Corresponding simulated performances are provided to validate the second-order approximations of the regret and highlight the usefulness of the methodologies.

Section 5.4 summarizes the chapter with a few overall concluding thoughts.

5.1. A Formulation

Here, we focus on the purely sequential MRPE methodologies alone and take into consideration a class of general robust estimators of σ, say denoted by V_n based on the observations given by $X_1, X_2, ..., X_n, n \geq 2$ independently from a normal distribution. We aim to find a set of sufficient conditions on $\{V_n, n \geq 2\}$ such that the associated methodologies involving V_n will have a number of asymptotic first-order and second-order properties.

5.1.1. Sufficient Conditions on Robust Estimators

Assuming we have a sequence of independent observations $X_1, X_2, ...$ from a $N(\mu, \sigma^2)$ population with $-\infty < \mu < \infty$ and $0 < \sigma < \infty$, both unknown. Having recorded $X_1, X_2, ..., X_n, n \geq 2$, we use $V_n \equiv V_n(X_1, X_2, ..., X_n)$, assumed positive w.p.1, to denote an appropriate robust estimator of σ. The sequence of $\{V_n, n \geq 2\}$ satisfies the following set of conditions:

(C1) *Independence.*

$\overline{X}_n$ and $\{V_k, 2 \leq k \leq n\}$ are distributed independently for all $n \geq 2$.

(C2) *Convergence in probability.*

$V_n \stackrel{P_{\mu,\sigma}}{\to} \sigma$, as $n \to \infty$.

(C3) *Asymptotic normality.*

$\sqrt{n}\left(\sigma^{-1}V_n - 1\right) \stackrel{\pounds}{\to} N(0, \delta^2)$ for some $\delta^2(>0)$ as $n \to \infty$.

(C4) *Uniform continuity in probability (u.c.i.p).* For every $\varepsilon > 0$, there exists a large $\nu \equiv \nu(\varepsilon)$ and small $\gamma > 0$ for which

$P_{\mu,\sigma}\left(\max_{0 \leq k \leq n\gamma} |V_{n+k} - V_n| \geq \varepsilon\right) < \varepsilon$ holds for any $n \geq \nu$.

(C5) *Kolmogorov's inequality.* For every $\varepsilon > 0$, and some $n_2 \geq n_1 \geq 2$,

$$P_{\mu,\sigma}\left(\max_{n_1 \leq n \leq n_2} |V_n - \sigma| \geq \varepsilon\right) \leq \varepsilon^{-r} E_{\mu,\sigma} |V_{n_1} - \sigma|^r, \text{ with } r \geq 2.$$

(C6) *Order of central absolute moments.* For $n \geq 2$, and $r \geq 2$,

$$E_{\mu,\sigma}\left[|V_n - \sigma|^r\right] = O(n^{-r/2}).$$

(C7) *Wiener's condition.*

$$E_{\mu,\sigma}\left[\sup_{n\geq 2} V_n\right] < \infty.$$

Remark 5.1. (C4) can be deduced from (C5) and (C6). Note that for any fixed $n_0 \geq \nu$,

$$\begin{aligned}
&P_{\mu,\sigma}\left(\max_{0\leq k\leq n_0\gamma} |V_{n_0+k} - V_{n_0}| \geq \varepsilon\right) \\
&\leq P_{\mu,\sigma}\left(\max_{0\leq k\leq n_0\gamma} \{|V_{n_0+k} - \sigma| + |V_{n_0} - \sigma|\} \geq \varepsilon\right) \\
&\leq P_{\mu,\sigma}\left(\max_{\nu\leq n\leq n_0(1+\gamma)} |V_n - \sigma| \geq \varepsilon/2\right) + P_{\mu,\sigma}\left(\max_{\nu\leq n\leq n_0} |V_n - \sigma| \geq \varepsilon/2\right) \\
&\leq 2\,(\varepsilon/2)^{-r}\, E_{\mu,\sigma}\, |V_\nu - \sigma|^r = O\left(\left(\nu\varepsilon^2\right)^{-r/2}\right).
\end{aligned}$$

By choosing some large enough ν and appropriate r such that $O\left(\left(\nu\varepsilon^2\right)^{-r/2}\right) < \varepsilon$, we can claim that the u.c.i.p property for $\{V_n, n \geq 2\}$ holds.

5.1.2. Purely Sequential MRPE Methodologies Involving V_n

Now we are in a position to propose a class of purely sequential methodologies based on V_n for an unknown normal mean μ under the MRPE formulation (2.2.1)-(2.2.3) in the spirit of the methodology (3.2.1)-(3.2.2) in Section 3.2. Beginning with the pilot data $X_1, X_2, ..., X_m$ of size $m\,(\geq 2)$, we sample one additional observation at a time sequentially as needed until we terminate according to the following stopping rule:

$$N \equiv N(c) = \inf\left\{n \geq m : n \geq \sqrt{A/c}\left(V_n + n^{-\lambda}\right)\right\}, \tag{5.1.1}$$

where $\lambda\,(>1/2)$ is held fixed. That is, if we have a pilot sample of size m such that $m \geq \sqrt{A/c}\left(V_m + m^{-\lambda}\right)$ already holds, no additional observations will be recorded and the final sample size is $N = m$. Otherwise, we record one more observation at a time and update n in the stopping rule (5.1.1). We terminate the sampling procedure at the first time that $N = n\,(\geq m)$ is observed such that $n \geq \sqrt{A/c}\left(V_n + n^{-\lambda}\right)$. Finally with $\{N, X_1, ..., X_m, ..., X_N\}$, we establish the minimum risk point estimator for μ as follows:

$$\overline{X}_N \equiv N^{-1}\Sigma_{i=1}^{N} X_i. \tag{5.1.2}$$

For the purely sequential MRPE methodology (5.1.1)-(5.1.2), clearly $P_{\mu,\sigma}\left(N < \infty\right) = 1$ and $N \uparrow \infty$ w.p.1 as $c \downarrow 0$.

5.2. Main Results

In this section, we lay down a number of main lemmas and theorems associated with the purely sequential MRPE methodology given by (5.1.1)-(5.1.2) under conditions (C1)-(C7).

Theorem 5.1. *For the stopping time N defined by the purely sequential estimation strategy (5.1.1)-(5.1.2), for all fixed μ, σ, and A, we have:*

$$\lim_{c\to 0} E_{\mu,\sigma}\left[N/n^*\right] = 1 \text{ [\textit{Asymptotic First-Order Efficiency}]}. \tag{5.2.1}$$

Proof. By the stopping rule defined in (5.1.1), we have the following two inequalities (w.p.1):

$$N \geq \sqrt{A/c}\left(V_N + N^{-\lambda}\right),$$

and

$$N - 1 < m - 1 + \sqrt{A/c}\left(V_{N-1} + (N-1)^{-\lambda}\right),$$

from which we conclude

$$\frac{V_N + N^{-\lambda}}{\sigma} \le \frac{N}{n^*} < \frac{V_{N-1} + (N-1)^{-\lambda}}{\sigma} + \frac{m}{n^*}. \tag{5.2.2}$$

Then, it is clear that as $c \to 0$, $N/n^* \overset{P_{\mu,\sigma}}{\to} 1$. Next, note that for some sufficiently small c, the right-hand side of (5.2.2) can be bounded as follows:

$$\frac{N}{n^*} < \sigma^{-1}(\sup_{n\ge 2} V_n + 1) + 1.$$

Thus, under (C7) and by the dominated convergence theorem, $\lim_{c\to 0} E_{\mu,\sigma}\left[N/n^*\right] = 1$ holds immediately. ■

Lemma 5.1. *For the stopping time N defined by the purely sequential estimation strategy (5.1.1)-(5.1.2), for all fixed μ, σ, and A, we have: for any arbitrary $0 < \eta < 1$,* with $r \ge 2$,

$$P_{\mu,\sigma}\left(N \le \eta n^*\right) = O\left(n^{*^{-\frac{r}{2(1+\lambda)}}}\right). \tag{5.2.3}$$

Proof. Along the lines in the proof of Lemma 3.2.1, let $\lfloor w \rfloor$ denote the largest integer that is smaller ($<$) than w and we define:

$$n_{1c} = \left\lfloor (A/c)^{\frac{1}{2(1+\lambda)}} \right\rfloor = O(c^{-\frac{1}{2(1+\lambda)}}) \text{ and } n_{2c} = \eta n^* = \eta\sigma\sqrt{A/c}, \tag{5.2.4}$$

again. It should be obvious that $N \ge n_{1c}$ w.p.1 from the definition of N in (5.1.1). Next, we set out to obtain the rate at which $P_{\mu,\sigma}\{N \le \eta n^*\}$ may converge to zero for small c:

$$
\begin{aligned}
&P_{\mu,\sigma}\{N \le \eta n^*\} \\
&\quad \le P_{\mu,\sigma}\left\{V_n \le \eta\sigma \text{ for some } n \text{ such that } n_{1c} \le n \le n_{2c}\right\} \\
&\quad \le P_{\mu,\sigma}\left\{\max_{n_{1c}\le n\le n_{2c}} |V_n - \sigma| \ge (1-\eta)\,\sigma\right\} \\
&\quad \le \{(1-\eta)\,\sigma\}^{-r} E_{\mu,\sigma}\left|V_{n_{1c}} - \sigma\right|^r = O\left(n_{1c}^{-r/2}\right) = O\left(n^{*-r/(2(1+\lambda))}\right),
\end{aligned}
$$

by conditions (C5) and (C6). Thus, (5.2.3) holds.■

Theorem 5.2. *For the stopping time N defined by the purely sequential estimation strategy (5.1.1)-(5.1.2), for all fixed $\mu, \sigma,$ and A, we have:*

$$\lim_{c\to 0} \xi\,(c) = 1 \text{ [Asymptotic First-Order Risk Efficiency]}, \tag{5.2.5}$$

where the term $\xi\,(c)$ was defined in (3.2.6).

Proof. Note that under (C1), $\xi\,(c) = \frac{1}{2}E_{\mu,\sigma}[N/n^*] + \frac{1}{2}E_{\mu,\sigma}[n^*/N]$. Picking $\eta = \frac{1}{2}$ in Lemma 5.1, (5.2.5) can be justified in the same way as we proved Theorem 3.2.2. ■

Lemma 5.2. *For the stopping time N defined by the purely sequential estimation strategy* (*5.1.1*)-(*5.1.2*), *for all fixed $\mu, \sigma,$ and A, we have*:

$$\left(N - n^*\right)/N^{1/2} \xrightarrow{\pounds} N(0, \delta^2), \text{ as } c \to 0.$$

Proof. Since $N/n^* \overset{P_{\mu,\sigma}}{\to} 1$ as $c \to 0$, we have $\sqrt{n^*}/\sqrt{N} \overset{P_{\mu,\sigma}}{\to} 1$ as $c \to 0$. Then by Slutsky's theorem, it suffices to show

$$\frac{N-n^*}{\sqrt{n^*}} \xrightarrow{\pounds} N(0,\delta^2) \text{ as } c \to 0. \tag{5.2.6}$$

Note that for sufficiently small c, the following holds w.p.1:

$$\sqrt{n^*}\left(\frac{V_N}{\sigma}-1\right) \le \frac{N-n^*}{\sqrt{n^*}} \le \frac{m}{\sqrt{n^*}} + \sqrt{n^*}\left(\frac{V_{N-1}}{\sigma}-1\right). \tag{5.2.7}$$

By the conditions (C3) and (C4), as well as Anscombe's (1952) Random Central Limit Theorem, we have

$$\sqrt{n^*}\left(\frac{V_N}{\sigma}-1\right) \xrightarrow{\pounds} N(0,\delta^2) \text{ and } \frac{m}{\sqrt{n^*}} + \sqrt{n^*}\left(\frac{V_{N-1}}{\sigma}-1\right) \xrightarrow{\pounds} N(0,\delta^2),$$

as $c \to 0$. Hence, the lemma follows from (5.2.7). ■

Lemma 5.3. *For the stopping time N defined by the purely sequential estimation strategy (5.1.1)-(5.1.2), for all fixed μ, σ, and A, we have: $(N-n^*)^2/N$ is uniformly integrable in $c \le c_0$, for some c_0.*

Proof. We prove this lemma in the spirit of Ghosh and Mukhopadhyay (1980) and Ghosh et al. (1997, Lemma 7.2.3, pp. 217-219). First we show $(N-n^*)^2/n^*$ is uniformly integrable in $c \le c_0$. Write for any $b > b_0 + 1$, $b_0 = \left(\sigma\sqrt{A/c_1}\right)^{-1/2}$, where c_1 is some appropriate constant such that $c \le c_1$. We have

$$E_{\mu,\sigma}\left\{\frac{(N-n^*)^2}{n^*}\mathbf{1}\left(\frac{(N-n^*)^2}{n^*} > b^2\right)\right\} = 2\int_b^\infty x P_{\mu,\sigma}\left(|N-n^*| > x\sqrt{n^*}\right)\mathrm{d}x. \tag{5.2.8}$$

Write $k_1 = \lfloor n^* + x\sqrt{n^*} \rfloor + 1$, for $x \geq b$. Then,

$$\begin{aligned} & P_{\mu,\sigma}\left(N > n^* + x\sqrt{n^*}\right) \\ & \leq P_{\mu,\sigma}\left(k_1 - 1 \leq V_{k_1-1}\sqrt{A/c}\right) \leq P_{\mu,\sigma}\left(\frac{V_{k_1-1}}{\sigma} \geq \frac{n^*+x\sqrt{n^*}-1}{n^*}\right) \\ & \leq P_{\mu,\sigma}\left(\left|\frac{V_{k_1-1}}{\sigma} - 1\right| \geq \frac{x\sqrt{n^*}-1}{n^*}\right) \leq \frac{E_{\mu,\sigma}\left|\frac{V_{k_1-1}}{\sigma}-1\right|^{2r_1}}{\left(\frac{x}{\sqrt{n^*}}-\frac{1}{n^*}\right)^{2r_1}}, \end{aligned} \tag{5.2.9}$$

for some r_1. With (C5), we claim that there exists a $\lambda_1\,(>0)$ depending only on r_1 such that

$$P_{\mu,\sigma}\left(N > n^* + x\sqrt{n^*}\right) \leq \lambda_1\,(k_1-1)^{-r_1}\,n^{*r_1}\left(x - 1/\sqrt{n^*}\right)^{-2r_1}. \tag{5.2.10}$$

Note that $(k_1-1)^{-r_1}\,n^{*r_1} < 1$, for $x \geq b > b_0 + 1 = \left(\sigma\sqrt{A/c_1}\right)^{-1/2} + 1$ and $n^* \geq \sigma\sqrt{A/c_1}$. It follows that

$$\begin{aligned} & \textstyle\int_b^\infty x P_{\mu,\sigma}\left(|N - n^*| > x\sqrt{n^*}\right)\mathrm{d}x \leq \lambda_1 \textstyle\int_b^\infty x\,(x-b_0)^{-2r_1}\,\mathrm{d}x \\ & \qquad = \lambda_1 \frac{(b-b_0)^{1-2r_1}((3-2r_1)b-b_0)}{(1-2r_1)(2-2r_1)} \to 0 \end{aligned} \tag{5.2.11}$$

as $b \to \infty$ by choosing $r_1 > 1$ appropriately.

Next, note that if $\sqrt{n^*} \leq b$,

$$\textstyle\int_b^\infty x P_{\mu,\sigma}\left(N - n^* < -x\sqrt{n^*}\right)\mathrm{d}x = 0. \tag{5.2.12}$$

If $\sqrt{n^*} > b$, there exists some $0 < \gamma < 1$ such that $(1-\gamma)\sqrt{n^*} > b$, when $c \leq c_2$, for some technically picked c_2. Then,

$$\begin{aligned} & \textstyle\int_b^\infty x P_{\mu,\sigma}\left(N - n^* < -x\sqrt{n^*}\right)\mathrm{d}x \\ & \leq \textstyle\int_b^{\sqrt{n^*}} x P_{\mu,\sigma}\left(N \leq \gamma n^*\right)\mathrm{d}x + \textstyle\int_b^{(1-\gamma)\sqrt{n^*}} x P_{\mu,\sigma}\left(\gamma n^* < N < n^* - x\sqrt{n^*}\right)\mathrm{d}x. \end{aligned} \tag{5.2.13}$$

By Lemma 5.1, $P_{\mu,\sigma}(N \le \gamma n^*) \le \lambda_2 n^{*-\frac{r_2}{2(1+\lambda)}}$, for some appropriate $r_2\,(> 2+2\lambda)$ and $\lambda_2\,(>0)$ depending on r_2 alone. Hence,

$$\int_b^{\sqrt{n^*}} x P_{\mu,\sigma}(N \le \gamma n^*)\,\mathrm{d}x \le \lambda_2 b^{2-\frac{r_2}{(1+\lambda)}} \to 0 \text{ as } b \to \infty. \tag{5.2.14}$$

As for $b \le x \le (1-\gamma)\sqrt{n^*}$, write

$$k_2 = \lfloor \gamma n^* \rfloor + 1 \text{ and } k_3 = \left\lfloor n^* - x\sqrt{n^*} \right\rfloor.$$

We obtain that

$$\begin{aligned} &P_{\mu,\sigma}\left(\gamma n^* < N < n^* - x\sqrt{n^*}\right) \\ &= P_{\mu,\sigma}\left(\cup_{n=k_2}^{k_3}\{N=n\}\right) \\ &\le P_{\mu,\sigma}\left(\cup_{n=k_2}^{k_3}\left\{\tfrac{V_n}{\sigma} < \tfrac{n}{n^*}\right\}\right). \end{aligned} \tag{5.2.15}$$

But for a small c, say $c \le c_3$, for some c_3,

$$\frac{n}{n^*} \le \frac{k_3}{n^*} \le \frac{n^* - x\sqrt{n^*}}{n^*} = 1 - \frac{x}{\sqrt{n^*}}.$$

Hence, it follows from (5.2.15) that

$$\begin{aligned} &P_{\mu,\sigma}\left(\gamma n^* < N < n^* - x\sqrt{n^*}\right) \\ &\le P_{\mu,\sigma}\left(\tfrac{V_n}{\sigma} - 1 < -\tfrac{x}{\sqrt{n^*}}, \text{ for some } k_2 \le n \le k_3\right) \\ &\le P_{\mu,\sigma}\left(\left|\tfrac{V_n}{\sigma} - 1\right| > \tfrac{x}{\sqrt{n^*}}, \text{ for some } k_2 \le n \le k_3\right) \\ &\le P_{\mu,\sigma}\left(\max_{k_2 \le n \le k_3}\left|\tfrac{V_n}{\sigma} - 1\right| > \tfrac{x}{\sqrt{n^*}}\right) \\ &\le \frac{E_{\mu,\sigma}\left|\frac{V_{k_2-1}}{\sigma} - 1\right|^{2r_3}}{\left(x/\sqrt{n^*}\right)^{2r_3}} \le \lambda_3 k_2^{-r_3} x^{-2r_3} n^{*r_3} \le \lambda_4 x^{-2r_3}, \end{aligned} \tag{5.2.16}$$

for some appropriate $\lambda_3\,(>0)$ and $\lambda_4\,(>0)$, both depending only on r_3. Choosing $r_3 > 1$, we get from (5.2.16) that

$$\begin{aligned}&\int_b^{(1-\gamma)\sqrt{n^*}} xP_{\mu,\sigma}\left(\gamma n^* < N < n^* - x\sqrt{n^*}\right)\mathrm{d}x \\ &\quad\le \lambda_4 \int_b^{(1-\gamma)\sqrt{n^*}} x^{1-2^{r_3}}\mathrm{d}x \le \tfrac{\lambda_4}{2-2r_3} b^{2-2r_3} \to 0 \text{ as } b\to\infty.\end{aligned} \tag{5.2.17}$$

Now choosing $c_0 = \min\{c_1, c_2, c_3\}$, with (5.2.11), (5.2.14) and (5.2.17), we can prove the uniform integrability of $(N-n^*)^2/n^*$ in $c \le c_0$. To complete the proof of the lemma, observe that

$$\begin{aligned}&E_{\mu,\sigma}\left\{\tfrac{(N-n^*)^2}{N}\mathbf{1}\left(\tfrac{(N-n^*)^2}{N} > b^2\right)\mathbf{1}\left(N > \tfrac{1}{2}n^*\right)\right\} \\ &\quad\le 2E_{\mu,\sigma}\left\{\tfrac{(N-n^*)^2}{n^*}\mathbf{1}\left((N-n^*)^2 > \tfrac{1}{2}b^2 n^*\right)\right\} \to 0\end{aligned} \tag{5.2.18}$$

as $b\to\infty$ uniformly in $c \le c_0$. Furthermore, choosing $b > n^*$, it follows that for $c \le c_0$,

$$\begin{aligned}&E_{\mu,\sigma}\left\{\tfrac{(N-n^*)^2}{N}\mathbf{1}\left(\tfrac{(N-n^*)^2}{N} > b^2\right)\mathbf{1}\left(N \le \tfrac{1}{2}n^*\right)\right\} \\ &\quad\le m^{-1}n^{*2}P_{\mu,\sigma}\left(N \le \tfrac{1}{2}n^*\right) \le \lambda_5 n^{*2-r_4} \le \lambda_5 b^{2-r_4} \to 0\end{aligned} \tag{5.2.19}$$

as $b\to\infty$, with $r_4\,(>2)$ appropriately chosen and some $\lambda_5\,(>0)$ depending only on r_4. In view of (5.2.18) and (5.2.19), the lemma holds. ■

Note that under the condition (C1), $\omega(c) = cE_{\mu,\sigma}[(N-n^*)^2/N]$. As an immediate result from Lemmas 5.2 and 5.3, the following asymptotic second-order property holds, summarized in Theorem 5.3.

Theorem 5.3. *For the stopping time N defined by the purely sequential estimation strategy (5.1.1)-(5.1.2), for all fixed μ, σ, and A, we have:*

$$\omega(c) = \delta^2 c + o(c) \text{ [Asymptotic Second-Order Risk Efficiency]}, \tag{5.2.20}$$

as $c\to 0$, where the term $\omega(c)$ was defined in (3.2.6).

5.3. An Application: Purely Sequential Methodology (3.2.1)-(3.2.2)

In this section, we review the purely sequential methodology (3.2.1)-(3.2.2) as a possible application of the methodology (5.1.1)-(5.1.2) newly proposed in this chapter. By verifying the conditions (C1)-(C7), we further claim the asymptotic second-order risk efficiency of the methodology (3.2.1)-(3.2.2) beyond the asymptotic first-order properties concluded in Section 3.2.1. Simulations are provided as a reasonable validation, as well.

5.3.1. Second-Order Approximations of $\omega_i(c), i = 0, 1, 2$

We give the second-order approximations of the regret $\omega_i(c), i = 0, 1, 2$ from the purely sequential methodology (3.2.1)-(3.2.2) in Corollary 5.1.

Corollary 5.1. *For the stopping time N_i defined by the purely sequential estimation strategy (3.2.1)-(3.2.2), for all fixed μ, σ, and A, we have*:

$$\omega_i(c) = \delta_i^2 c + o(c), i = 0, 1, 2 \text{ [\textit{Asymptotic Second-Order Risk Efficiency}]} \tag{5.3.1}$$

as $c \to 0$, where the term $\omega_i(c)$ was defined in (3.2.6), and $\delta_0^2 = \frac{1}{2}, \delta_1^2 = \frac{1}{3}(\pi + 6\sqrt{3} - 12)$, $\delta_2^2 = \frac{1}{2}(\pi - 2)$ remain the same as in Lemma 4.1.3.

Proof. It suffices to show that (C1)-(C7) are satisfied in terms of $V_{n;i}, i = 0, 1, 2$ defined in (2.3.4), respectively. Among the conditions, (C1) and (C7) follow from Mukhopadhyay and Hu (2017, Lemmas 3.1 and 3.3) with minor modifications. (C2) can be simply verified by slightly modifying the proof of Lemma 3.1.2. In the spirit of the proof of Lemma 4.1.3, we have (C3). Both (C5) and (C6) have been indicated in the proof of Lemma 3.2.1. And (C4) comes from combining (C5) and (C6).

Thus, the corollary follows. One may refer to Mukhopadhyay and de Silva (2009), Sen and Ghosh (1981), Mukhopadhyay and Hu (2017,2018) and Hu and Mukhopadhyay (2018) for more details. ■

5.3.2. Simulations and Data Analysis

Based on what we did in Section 3.2.2, we had considered $m = 15$ alone for brevity as we found little to no difference in the overall performances whether the pilot sample size was fixed to be 10 or 15. Besides the notation explained in (3.2.21), we included the following entities in (5.3.2) as well.

$$
\begin{aligned}
s(\widehat{\xi}) &: \text{ estimated standard error of } \widehat{\xi}; \\
s(\widehat{\omega}) &: \text{ estimated standard error of } \widehat{\omega}; \\
\delta_i^2 c &: \text{ a theoretical approximation of } \omega_i(c), i = 0, 1, 2.
\end{aligned}
\tag{5.3.2}
$$

Table 5.1 summarizes selected findings ($m = 15$) in the spirit of Table 4.1. One should note that within each block of Table 5.1, the theoretical second-order approximations of $\omega_i(c), i = 0, 1, 2$ turned out close to the estimated regret values. These empirically validate Corollary 5.1 about the asymptotic second-order risk efficiency property from (5.3.1) of the purely sequential MRPE methodology (3.2.1)-(3.2.2).

5.4. Overall Concluding Thoughts

In summary, the purely sequential MRPE methodology (5.1.1)-(5.1.2) opens the possibility to incorporate a class of robust estimators of σ in defining the stopping boundary conditions without loss of the second-order properties. We have shown that the associated methodology is second-order risk efficient, as long as the robust estimators satisfy the conditions given by (C1)-(C7). This indicates the purely sequential MRPE methodology (5.1.1)-(5.1.2) will have a much wider application.

Table 5.1. Simulations under $A = 100$, $m = 15$ with 1000 runs implementing methods (3.2.1)-(3.2.2) with i from (3.2.20) with $\lambda = 2$

n^* c	i	$\overline{n}$ $s(\overline{n})$	$\overline{R}_n$ $s(\overline{R}_n)$	$\widehat{\xi}$ $s(\widehat{\xi})$	$\widehat{\omega}$ $s(\widehat{\omega})$
$\delta_0^2 c = 8 \times 10^{-2}, \delta_1^2 c = 8.18 \times 10^{-2}, \delta_2^2 c = 9.13 \times 10^{-2}$					
50	0	50.247	15.801	0.988	9.34×10^{-2}
0.16		0.167	0.053	0.033	5.07×10^{-3}
	1	50.460	15.839	0.990	9.49×10^{-2}
		0.168	0.053	0.033	4.65×10^{-3}
	2	50.457	15.843	0.990	9.79×10^{-2}
		0.171	0.053	0.033	5.00×10^{-3}
$\delta_0^2 c = 2 \times 10^{-2}, \delta_1^2 c = 2.05 \times 10^{-2}, \delta_2^2 c = 2.28 \times 10^{-2}$					
100	0	100.159	7.941	0.993	2.15×10^{-2}
0.04		0.229	0.018	0.023	1.05×10^{-3}
	1	100.239	7.947	0.993	2.23×10^{-2}
		0.233	0.018	0.023	1.08×10^{-3}
	2	100.245	7.944	0.993	2.55×10^{-2}
		0.250	0.019	0.024	1.27×10^{-3}
$\delta_0^2 c = 5 \times 10^{-3}, \delta_1^2 c = 5.11 \times 10^{-3}, \delta_2^2 c = 5.71 \times 10^{-3}$					
200	0	200.199	3.986	0.997	5.46×10^{-3}
0.01		0.328	0.007	0.016	2.54×10^{-4}
	1	200.201	3.987	0.997	5.80×10^{-3}
		0.338	0.007	0.017	2.81×10^{-4}
	2	200.103	3.989	0.997	6.16×10^{-3}
		0.346	0.007	0.016	3.27×10^{-4}

Chapter 6

A Brief Summary

Purely sequential and Stein-type two-stage methodologies involving robust estimators such as GMD and MAD have been developed for both fixed-width confidence interval and minimum risk point estimation problems. With the underlying normal population, we assumed the condition that there exists an known positive lower bound of the variance so that the Stein-type methodologies will then enjoy desirable second-order properties, the purely sequential methodologies seem to automatically satisfy the higher order properties. The first-order properties have been proved, and the second-order properties are in progress.

Compared with the customary multi-stage methodologies based on the sample variance as an unbiased estimator of the population variance, our newly proposed multi-stage methodologies are more robust, when the normality of the underlying population is contaminated mildly. Therefore, our methodologies will have a wider applicability for practical purposes.

Chapter 7

Future Work

We have seen the significance of GMD or MAD, as a more robust alternative estimator than the sample variance. There are a lot of directions that we can follow in the future to make our research more broadly applicable in a number of other substantial inference problems. The potential future research may include the following areas.

7.1. Other Multi-Stage Methodologies

We may develop other multi-stage methodologies based on GMD and MAD, for example three-stage and accelerated sequential sampling, and compare them with the existing multi-stage methodologies and highlight the benefits of one approach over another.

While the Stein's (1945,1949) two-stage methodology is quick to terminate, it suffers from the serious oversampling problem, especially when the final sample size is large. Researchers have looked into the sampling designs where an additional stage is imposed so that the methodology will preserve both the first-order and second-order properties. In this spirit of combining procedure efficiency as well as sampling operational savings, the three-stage estimation methodologies have been established.

Mukhopadhyay (1976) first introduced the idea of triple sampling. Later on, Hall (1981) proposed a systematic three-stage fixed-width confidence interval estimation methodology for a normal mean when the population variance is unknown, where sample variance was utilized as an unbiased estimator for the population variance in defining stopping boundaries in each stage. He gave a rigorous account of the first-order and second-order properties. Mukhopadhyay (1990) considered more general cases and put forward a unified three-stage methodology. He discussed both the FCIE and MRPE problems with applications. We should mention there is a wide range of literature on three-stage estimation methodologies that one may refer to, including Mukhopadhyay (1985), Mukhopadhyay et al. (1987), Woodroofe (1987), Hamdy (1988), Hamdy et al. (1988), and Liu and Wang (2007).

On the other hand, researchers have been modifying the purely sequential estimation methodology and considering accelerated sequential stopping rules as well. Originally proposed in Hall (1983), the accelerated sequential methodology was further discussed in Mukhopadhyay and Solanky (1991) and Mukhopadhyay (1996).

Here, we may therefore follow this direction and establish both the three-stage and accelerated sequential estimation methodologies based on GMD or MAD correspondingly. We hope to show these methodologies are potentially more robust without loss of the first-order and second-order properties.

7.2. Distribution-Free Methodologies

Recall that Section 3.1.3 only provided the limited robustness of our newly proposed purely sequential estimation methodologies based on GMD or MAD, as we had assumed the population should follow a mixture normal distribution (3.1.7) instead of a pure normal distribution. Mukhopadhyay (1978), Ghosh and Mukhopadhyay (1979), Chow and Yu (1981) and other work set out to study the performances of multi-stage methodologies when the population distribution was unspecified. See Sen (1981,1985), Ghosh et al. (1997), and Mukhopadhyay and de Silva (2009) for a thorough review.

We may continue to discuss the GMD-based or MAD-based multi-stage methodologies in the distribution-free cases, where stronger robustness is highly likely to be addressed.

7.3. Multivariate Multi-Stage Estimation

Schezhtman and Yitzhaki (1987) proposed a new measure of association between two random variables based on GMD, which opened the possibility of extending our new GMD-based multi-stage methodologies (i) from univariate estimation to multivariate estimation or (ii) from one-sample estimation to multi-sample estimation. We may also investigate the MAD case so that our methodologies will have a significantly broader appeal.

List of Tables

Bibliography

Anscombe, F. J. (1952). Large-Sample Theory of Sequential Estimation, *Proceedings of Cambridge Philosophical Society* 48: 600-607.

Anscombe, F. J. (1953). Sequential Estimation, *Journal of Royal Statistical Society, Series B* 15: 1-29.

Babu, G. J. and Rao, C. R. (1992). Expansions for Statistics Involving the Mean Absolute Deviations, *Annals of Institute of Statistical Mathematics* 44: 387-403.

Barnett, F. C., Mullen, K., and Saw, J. G. (1967). Linear Estimates of a Population Scale Parameter, *Biometrika* 54: 551-554.

Basu, D. (1955). On Statistics Independent of a Complete Sufficient Statistic, *Sankhyā* 15: 377-380.

Carroll, R. J. (1977). On the Asymptotic Normality of Stopping Times Based on Robust Estimators, *Sankhyā, Series A* 39: 355–377.

Chattopadhyay, B. and De, S. K. (2016). Estimation of Gini Index within Pre-Specified Error Bound, *Econometrics* 4: 1-30.

Chattopadhyay, B. and Mukhopadhyay, N. (2013). Two-Stage Fixed-Width Confidence Intervals for a Normal Mean in the Presence of Suspect Outliers, *Sequential Analysis* 32: 134-157.

Chow, Y. S. and Robbins, H. (1965). On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean, *Annals of Mathematical Statistics* 36: 457-462.

Chow, Y. S. and Yu, K. F. (1981). The Performance of a Sequential Procedure for the Estimation of the Mean, *Annals of Statistics* 9: 184-188.

Dantzig, G. B. (1940). On the Non-Existence of Tests of Student's Hypothesis Having Power Functions Independent of σ, *Annals of Mathematical Statistics,* 11: 186-192.

David, H. A. (1968). Miscellanea: Gini's Mean Difference Rediscovered, *Biometrika* 55: 573-575.

De, S. K. and Chattopadhyay, B. (2017). Minimum Risk Point Estimation of Gini Index, *Sankhyā, Series B* 79: 247-277.

Downton, F. (1966). Linear Estimates with Polynomial Coefficients, *Biometrika* 53: 129-141.

Ghosh, B. K. and Sen, P. K. (1991). *Handbook of Sequential Analysis*, edited volume, New York: Dekker.

Ghosh, M. and Mukhopadhyay, N. (1975). Asymptotic Normality of Stopping Times in Sequential Analysis, Unpublished manuscript, Indian Statistical Institute, Calcutta, India.

Ghosh, M. and Mukhopadhyay, N. (1976). On Two Fundamental Problems of Sequential Estimation, *Sankhyā, Series B* 38: 203-218.

Ghosh, M. and Mukhopadhyay, N. (1979). Sequential Point Estimation of the Mean When the Distribution is Unspecified, *Communications in Statistics - Theory and Methods, Series A* 8: 637-652.

Ghosh, M. and Mukhopadhyay, N. (1980). Sequential Point Estimation of the Difference of Two Normal Means, *Annals of Statistics* 8: 221-225.

Ghosh, M. and Mukhopadhyay, N. (1981). Consistency and Asymptotic Efficiency of Two-Stage and Sequential Procedures, *Sankhyā, Series A* 43: 220-227.

Ghosh, M., Mukhopadhyay, N., and Sen, P. K. (1997). *Sequential Estimation,* New York: Wiley.

Gini, C. (1914). Sulla Misura della Concentrazione e della Variabilità dei Caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti* 73: 1203–1248.

Gini, C. (1921). Measurement of Inequality of Incomes, *Economic Journal* 31: 124-126.

Govindarajulu, Z. (2004). *Sequential Statistics*, Singapore: World Scientific.

Gut, A. and Mukhopadhyay, N. (2010). On Asymptotic and Strict Monotonicity of a Sharper Lower Bound for Student's t Percentiles, *Methodology and Computing in Applied Probability* 12: 647-657.

Hall, P. (1981). Asymptotic Theory of Triple Sampling for Estimation of a Mean, *Annals of Statistics* 9: 1229-1238.

Hall, P. (1983). Sequential Estimation Saving Sampling Operations, *Journal of Royal Statistical Society, Series B* 45: 219-223.

Hamdy, H. I. (1988). Remarks on the Asymptotic Theory of Triple Stage Estimation of the Normal Mean, *Scandinavian Journal of Statistics* 15: 303–310.

Hamdy, H. I., Mukhopadhyay, N., Costanza, M. C., and Son, M. S. (1988). Triple Stage Point Estimation for the Exponential Location Parameter, *Annals of Institute of Statistical Mathematics* 40: 785-797.

Helmert, F. R. (1876). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers director Beobachtungen gleicher Genauigkeit, *Astronomische Nachrichten* 88: 113.

Herrey, E. M. (1965). Confidence Intervals Based on the Mean Absolute Deviation of a Normal Sample, *Journal of American Statistical Association* 60: 257-269.

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution, *Annals of Mathematical Statistics* 19: 293-325.

Hoeffding, W. (1961). The Strong Law of Large Numbers for U-Statistics, Institute of Statistics Mimeo Series #302, University of North Carolina, Chapel Hill.

Hu, J. and Mukhopadhyay, N. (2018). Second-Order Asymptotics in a Class of Purely Sequential Minimum Risk Point Estimation (MRPE) Methodologies, *Submitted*.

Jurečkovā, J. and Sen, P. K. (1996). *Robust Statistical Procedures*, New York: Wiley.

Konno, H. and Yamazaki, H. (1991). Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market, *Management Science* 37: 519-531.

Lai, T. L. and Siegmund, D. (1977). A Nonlinear Renewal Theory with Applications to Sequential Analysis I, *Annals of Statistics* 5: 946-954.

Lai, T. L. and Siegmund, D. (1979). A Nonlinear Renewal Theory with Applications to Sequential Analysis II, *Annals of Statistics* 7: 60-76.

Lee, A. J. (1990). *U-statistics. Theory and Practice*, New York: Dekker.

Liu, W. and Wang, N. (2007). A Note on Hall's Triple Sampling Procedure: A Multiple Sample Second Order Sequential Analogue of the Behrens–Fisher Problem, *Journal of Statistical Planning and Inference* 137: 331-343.

Markowitz, H. (1959). *Portfolio Selection, Efficient Diversification of Investments*, New York: Wiley.

McNair, M. P. (1930). Data from "Margins, Expenses and Profits in Retail Trade in the United States As Studied by the Harvard University Bureau of Business Research", *Economic Journal* 40: 599-632.

Mukhopadhyay, N. (1976). Fixed-width Confidence Intervals for the Mean Using a Three-Stage Procedure, Unpublished report.

Mukhopadhyay, N. (1978). Sequential Point Estimation of the Mean When the Distribution is Unspecified, *Technical Report No. 312*, University of Minnesota, Minneapolis.

Mukhopadhyay, N. (1980). A Consistent and Asymptotically Efficient Two-Stage Procedure to Construct Fixed-Width Confidence Intervals for the Mean, *Metrika* 27: 281-284.

Mukhopadhyay, N. (1982). Stein's Two-Stage Procedure and Exact Consistency, *Scandinavian Actuarial Journal* 110-122.

Mukhopadhyay, N. (1985). A Note on Three-Stage and Sequential Point Estimation Procedures for a Normal Mean, *Sequential Analysis* 4: 311-320.

Mukhopadhyay, N. (1990). Some Properties of a Three-Stage Procedure with Applications in Sequential Analysis, *Sankhyā, Series A* 52: 218-231.

Mukhopadhyay, N. (1996). An Alternative Formulation of Accelerated Sequential Procedures with Applications to Parametric and Nonparametric Estimation, *Sequential Analysis* 15: 253-269.

Mukhopadhyay, N. (2010). On a New Sharper Lower Bound for a Percentile of a Student's t Distribution with an Application, *Methodology and Computing in Applied Probability* 12: 609-622.

Mukhopadhyay, N. and Chattopadhyay, B. (2013). Asymptotic Expansion of Percentiles for a Sample Mean Standardized by GMD: An Application in the Normal Case, *Journal of Japan Statistical Society* 42: 165-184.

Mukhopadhyay, N., Datta, S., and Chattopadhyay, S. (2004). *Applied Sequential Methodologies*, edited volume, New York: Dekker.

Mukhopadhyay, N. and de Silva, B. M. (2009). *Sequential Methods and Their Applications*, Boca Raton: CRC.

Mukhopadhyay, N., and Duggan, W. T. (1997). Can a Two-Stage Procedure Enjoy Second-Order Properties? *Sankhyā, Series A* 59: 435-448.

Mukhopadhyay, N., and Duggan, W. T. (1999). On a Two-Stage Procedure Having Second-Order Properties with Applications, *Annals of Institute of Statistical Mathematics* 51: 621-636.

Mukhopadhyay, N., Hamdy, H. I., Al-Mahmeed, M., and Costanza, M. C. (1987). Three-stage Point Estimation Procedures for a Normal Mean, *Sequential Analysis* 6: 21-36.

Mukhopadhyay, N., and Hu, J. (2017). Confidence Intervals and Point Estimators for a Normal Mean Under Purely Sequential Strategies Involving Gini's Mean Difference and Mean Absolute Deviation, *Sequential Analysis* 36: 210-239.

Mukhopadhyay, N., and Hu, J. (2018). Gini's Mean Difference and Mean Absolute Deviation Based Two-Stage Estimation for a Normal Mean with Known Lower Bound of Variance, *Sequential Analysis*, 37: in press.

Mukhopadhyay, N. and Solanky, T. K. S. (1991). Second Order Properties of Accelerated Stopping Times with Applications in Sequential Estimation, *Sequential Analysis* 10: 99-123.

Mukhopadhyay, N. and Solanky, T. K. S. (1994). *Multistage Selection and Ranking Procedures*, New York: Dekker.

Mukhopadhyay, N., Son, M. S., and Ko, Y. C. (2004). A New Two-Stage Sampling Design for Estimating the Maximum Average Time to Flower, *Journal of Agricultural, Biological, and Environmental Statistics* 9: 479-499.

Mukhopadhyay, N. and Vik, G. (1985). Asymptotic Results for Stopping Times Based on U-Statistics, *Sequential Analysis* 4: 83-109.

Nair, U. S. (1936). The Standard Error of Gini's Mean Difference, *Biometrika* 28: 428-436.

Ray, W. D. (1957). Sequential Confidence Intervals for the Mean of a Normal Population with Unknown Variance, *Journal of Royal Statistical Society, Series B* 19: 133-143.

Robbins, H. (1959). Sequential Estimation of the Mean of a Normal Population, in *Probability and Statistics*, H. Cramér volume, Ulf Grenander, ed., pp. 235-245, Uppsala: Almquist & Wiksell.

Schezhtman, E. and Yitzhaki, S. (1987). A Measure of Association Based on Gini's Mean Difference, *Communications in Statistics - Theory and Methods* 16: 207-231.

Sen, P. K. (1981). *Sequential Nonparametrics*, New York: Wiley.

Sen, P. K. (1985). *Theory and Applications of Sequential Nonparametrics*, Philadelphia: SIAM.

Sen, P. K. (1986). The Gini Coefficient and Poverty Indexes: Some Reconciliations, *Journal of American Statistical Association* 81: 1050-1057.

Sen, P. K. and Ghosh, M. (1981). Sequential Point Estimation of Estimable Parameters Based on U-Statistics, *Sankhyā, Series A* 43: 331-344.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer.

Starr, N. (1966). On the Asymptotic Efficiency of a Sequential Procedure for Estimating the Mean, *Annals of Mathematical Statistics* 37: 1173-1185.

Starr, N. and Woodroofe, M. (1969). Remarks on Sequential Point Estimation, *Proceedings of National Academy of Sciences* 63: 285-288.

Stein, C. (1945). A Two Sample Test for a Linear Hypothesis Whose Power Is Independent of the Variance, *Annals of Mathematical Statistics* 16: 243-258.

Stein, C. (1949). Some Problems in Sequential Estimation (Abstract), *Econometrica* 17: 77-78.

Wiener, N. (1939). The Ergodic Theorem, *Duke Mathematical Journal* 5: 1–18.

Winner, L. Personal Website http://www.stat.ufl.edu/~winner/data/ret_earn.dat.

Woodroofe, M. (1977). Second Order Approximations for Sequential Point and Interval Estimation, *Annals of Statistics* 5: 984-995.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, CBMS Lecture Notes #39, Philadelphia: SIAM.

Woodroofe, M. (1987). Asymptotically Optimal Sequential Point Estimation in Three Stages, in *New Perspectives in Theoretical and Applied Statistics*, M. L. Puri, ed., pp. 397-411, New York: Wiley.

Yitzhaki, S. (2003). Gini's Mean Difference: A Superior Measure of Variability for Non-Normal Distributions, *Metron* 61: 285-316.

Zacks, S. (2009). *Stage-Wise Adaptive Designs*, New York: Wiley.

Zacks, S. (2017). *Sample Path Analysis and Distributions of Boundary Crossing Times*, New York: Springer.