7-17-2017

Multistage Sampling Strategies and Inference in Health Studies Under Appropriate Linex Loss Functions

Sudeep R. Bapat
University of Connecticut - Storrs, sudeep.bapat@uconn.edu

Multistage Sampling Strategies and Inference in Health Studies Under Appropriate Linex Loss Functions

Sudeep R. Bapat, Ph.D.

University of Connecticut, 2017

ABSTRACT

A sequential sampling methodology provides concrete results and proves to be beneficial in many scenarios, where a fixed sampling technique fails to deliver. This dissertation introduces several multistage sampling methodologies to estimate the unknown parameters depending on the model in hand. We construct both two-stage and purely sequential sampling rules under different situations. The estimation is carried under a loss function which in our case is either a usual squared error loss or a Linex loss. We adopt a technique known as bounded risk estimation strategy, where we bound the appropriate risk function from above by a fixed and known constant $\omega(> 0)$. At first we draw attention to a negative exponential distribution and applications from health studies. We propose appropriate stopping rules to estimate the location parameter or the threshold μ of a negative exponential distribution under a Linex loss function. This model proves to be relevant to depict failure times of complex equipment or survival times in cancer research. We include some real data applications such as to estimate the minimum threshold of infant mortality rates for different countries.

We then move on to extend this investigation to a two-sample situation, where we estimate the difference in locations $\mu_1 - \mu_2$ of two independent negative exponential populations having scales σ and $b\sigma$. An interesting aspect here is that $b(> 0)$ should be known a'priori. The estimation is again carried out under Linex loss. We introduce some applications from cancer studies and reliability analysis.

The third fold of this dissertation involves the bounded risk multistage point estimation of a negative binomial (NB) mean $\mu(> 0)$, under different loss functions. We assume that the thatch

Sudeep R. Bapat, University of Connecticut, 2017

parameter $\tau(>0)$ is either known or unknown. We use a parameterization of the NB model which was first introduced by Anscombe in (1949). This is on slightly different lines since it involves a discrete population. A negative binomial model finds its use in entomological or ecological studies involving count data. We propose two-stage and purely sequential rules under squared error and Linex loss functions. We include real data applications involving weed count and bird count data.

We next move on to extend this work for a multi-sample situation where we 1) simultaneously estimate a k-vector of NB means and 2) estimate the difference in means $\mu_1 - \mu_2$ of two independent NB populations. We again assume that the thatch parameters are either known or unknown. In the case when the thatch parameters are unknown, we have designed an interesting allocation scheme along with suitable set of stopping rules. The work is supported using interesting real world applications.

We should mention that all of our proposed methodologies enjoy exciting asymptotic efficiency and consistency properties depending on the scenario. Finally, we conclude by discussing some attractive areas of future research that may be of practical significance.

Multistage Sampling Strategies and Inference in Health Studies Under Appropriate Linex Loss Functions

Sudeep R. Bapat

B.Sc., University of Delhi, India, 2010

M.Sc., Indian Institute of Technology, Bombay, India, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Multistage Sampling Strategies and Inference in Health Studies Under Appropriate Linex Loss Functions

Presented by

Sudeep R. Bapat, B.Sc., M.Sc.

Major Advisor ____

Nitis Mukhopadhyay

Associate Advisor ____

Joseph Glaz

Associate Advisor ____

Vladimir Pozdnyakov

University of Connecticut

2017

ACKNOWLEDGEMENTS

The first and very foremost I would like to thank my parents. Their immense love and support proved to be beneficial towards where I stand today. My father Dr. Ravindra B. Bapat, being a professor himself, is a true role model who set an amazing example through his life for me to emulate. My mother Ragini R. Bapat, always believed in the path I took and stood besides me at all times. I would also like to wholeheartedly thank the rest of my family. I would specially like to extend my gratitude to my grandparents, who showed their constant support through their appreciation and positivity.

After I joined the Department of Statistics at the University of Connecticut in Fall 2013, I was soon exposed to courses taught by Professor Nitis Mukhopadhyay. These were so intriguing and riveting that I wished they continued longer. It did not take long for me to admire Professor Mukhopadhyay's immense knowledge of the subject and his thought provoking process of teaching. I feel that I am truly blessed to have an advisor like Professor Mukhopadhyay. Two of his most remarkable qualities are his passion for statistics and his commitment to students. Being a great researcher, an educator, an administrator among many others, his passion for statistics remains admirable. As a teacher and a mentor, his commitment to students has always been remarkable. I consider myself fortunate enough to have cherished these qualities over these four years. Another of his positive aspects is that he is always loaded with interesting ideas and lends a helping hand which makes it easier to grasp any topic. I would like to thank Professor Mukhopadhyay for all the guidance, advise and help extended to me over these years.

I would also like to thank his lovely wife, Mahua Mukhopadhyay who is cheerful, gracious and immensely cordial. I thoroughly enjoyed all the interactions I had with her.

I would also like to extend my sincere thanks to Professor Joseph Glaz and Professor Vladimir Pozdnyakov, my associate advisors for their continuous guidance and support. Their valuable inputs and friendly presence made my dissertation journey a lot comfortable. I would also like to thank all the other professors in the department most of whom have taught me in this process. I would like to extend my thanks to Professor Zhiyi Chi, who provided me with summer teaching opportunities and solo-instructorship in my last year at UConn. This experience would definitely help in my future career. I would also like to specially thank Professor Dipak Dey and Professor Nalini Ravishanker who were always there extending their warm support especially during my initial years which made my transition into the department much smoother.

My sincere thanks also go to Tracy and Megan who were phenomenal with their support through the four years of my Ph.D.

My time at UConn was extremely memorable because of the friendship and camaraderie of my fellow graduate students, both inside and outside of the department. They have helped me live through tough times and made me feel at-home as well. I will always cherish their friendship deeply.

Table of Contents

Chapter 1
Introduction

1.1. THE NEED FOR SEQUENTIAL

Sequential analysis is concerned with gathering information about an unknown parameter θ, by taking random samples in batches or in a sequence. This is in direct contrast with the usual inference methodologies where one takes a fixed number of sample points. A sequential rule or a sampling scheme consists of a positive integer valued random variable N, which denotes the number of sample points collected at termination. We then propose an estimator of some unknown parameter θ, by a randomly stopped statistic T_N which is dependent on N.

A sequential sampling scheme proves to have an edge over the usual fixed-sample strategy, as shown in the following scenarios.

1.1.1. Hypothesis Tests

Consider a random sample $X_1, X_2, ..., X_n$ of size n from a Normal population with mean μ and variance σ^2. Also consider the following hypothesis $H_0 : \mu = \mu_0$ against $H_a : \mu = \mu_1 (> \mu_0)$. We know that a MP test under a known σ^2 at a significance level $\alpha (0 < \alpha < 1)$ is given by:

$$\text{Reject } H_0 \text{ if } \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma} > z_\alpha, \tag{1.1.1}$$

where

$$\begin{aligned} z_\alpha &\equiv \text{upper } 100\alpha^{th} \text{ percentile of a } N(0,1) \text{ distribution} \\ \overline{X}_n &= n^{-1} \textstyle\sum_{i=1}^n X_i, \text{ the sample mean.} \end{aligned} \tag{1.1.2}$$

The type I error probability is clearly α. However one also might be concerned with the type II error probability $\beta (0 < \beta < 1)$. If we define the sample size as follows:

$$n \geq \left(\tfrac{z_\alpha + z_\beta}{\mu_1 - \mu_0}\right)^2 \sigma^2 = n^* \text{ with } z_\gamma \equiv \text{upper } 100\gamma^{th} \text{ percentile of a } N(0,1) \text{ distribution,} \tag{1.1.3}$$

we can also restrict the type II error probability below a predefined β. However in most practical cases since σ^2 is unknown, controlling both α and β is a concern which needs to be addressed.

1.1.2. Estimating a Normal Mean

Again consider a Normal distribution with mean μ and variance σ^2. Now consider a problem of estimating the unknown mean μ. The customary estimator based on a sample of size n is the sample mean $\overline{X}_n$. Now further consider a loss function defined as:

$$L(\overline{X}_n, \mu) = A(\overline{X}_n - \mu)^2 + cn \text{ with known } A(>0) \text{ and } c(>0). \tag{1.1.4}$$

Here, c is taken to be the cost of per unit observation. The risk function R_n is given by:

$$R_n = A\sigma^2 n^{-1} + cn. \tag{1.1.5}$$

Assuming σ^2 known, R_n can be minimized by defining the sample size as:

$$n = (A/c)^{1/2}\sigma : \text{ assuming it is an integer.} \tag{1.1.6}$$

However the issue again arises that σ^2 is often unknown and one then needs to deal with the situation of minimizing the risk function in a proper manner.

1.2. MULTISTAGE AND SEQUENTIAL SAMPLING DESIGNS: A HISTORICAL OVERVIEW

In the previous section we presented a couple of situations where a sequential route seems to be advantageous over a fixed-sample technique. The problems of concern were to control both type I and type II errors in Section 1.1.1 and to minimize the risk function in Section 1.1.2. The important interest is that will these techniques work if σ^2 is unknown. It so turns out that none of these will work under a fixed-sample strategy. We hence need to resort to multistage or sequential sampling schemes.

1.2.1. Evolution of Multistage and Sequential Sampling Methods

The importance of sampling and development of sampling designs was carried out by Mahalanobis in 1940. He was the pioneer of using a pilot sample before large sample surveys. Abraham Wald was the forerunner of sequential analysis who along with his collaborators systematically developed theory and methodology around sequential tests in 1940 to reduce the number of sampling inspections. The problems discussed in Section 1.1 were tackled by Lehmann (1951) who proved the non-existence of fixed-sample size solutions. One may refer to Ghosh et al. (1997) or Mukhopadhyay and de Silva (2009). A similar problem where a fixed-sample solution is impossible is under the framework of finding a confidence interval. Dantzig (1940) proved the non-existence of such a solution under a fixed-sample scenario. A large amount of literature has been developed starting in the 1940's surrounding the hypothesis testing and minimum risk point estimation problems. Some of the notable works include Robbins and Starr (1965) and Robbins (1959). Starr (1966b) extended the ideas developed by Robbins (1959) in detail. Ghosh and Mukhopadhyay (1976) investigated a two-stage procedure to tackle with the minimum risk estimation problem. Mukhopadhyay (1976,1990) gave a three-stage procedure followed by an accelerated sequential procedure in Mukhopadhyay (1987,1996). It was Stein (1945) who developed the confidence interval problem and presented the famous two-stage procedure, followed by Chow and Robbins (1965) who gave the purely sequential procedure. Some of the other notable works in this relation comprise of Hall (1981,1983) and Mukhopadhyay (1980,1990,1996).

1.2.2. Few Other Applications

Sequential probability ratio test (SPRT) is a technique developed by Wald (1947). It addresses the hypothesis testing problem as seen in Section 1.1.1. The sample size in SPRT turns out to be smaller on an average than the one seen in (1.1.3). Wald and Wolfowitz (1948) proved the optimality of their SPRT rule which proves to be efficient as compared to other known methods.

Another widespread area comprises of selection and ranking procedures which handle multiple comparison problems. An illustration of the problem could be to select the sample corresponding to the largest population mean (best treatment) among several samples of distinct population means, with a preassigned probability of correct selection. There does not exist any fixed-sample methodologies to tackle such a problem and one has to resort to multistage/sequential strategies. If the variance σ^2 is known, then an optimal sample size can be determined using the *indifference-zone* approach according to Bechhofer (1954). However if σ^2 is unknown, one still has to resort to multistage/sequential sampling designs. Further literature on selection and ranking problems can be found in Bechhofer (1968,1995), Gibbons et al. (1977), Gupta and Panchapakesan (1979), Mukhopadhyay (1993), Mukhopadhyay and Solanky (1994), Mukhopadhyay and de Silva (2009) among other sources.

Further, vast literature surrounding multistage and or sequential designs can be found in Wald (1947), Chernoff (1972), Ghosh and Sen (1991), Siegmund (1985), Mukhopahyay et al. (2004) and Mukhopadhyay and de Silva (2009).

1.3. THESIS OUTLINE

In this dissertation we introduce multistage and or purely sequential sampling designs for the purposes of point estimation. We mainly concentrate on two distributions namely negative exponential and negative binomial. The problems are to estimate the appropriate parameters under both these distributions under a Linex loss which was introduced by Varian in 1975.

In chapter 2 we develop modified two-stage and purely sequential strategies to estimate the location parameter of a negative exponential distribution under a modified Linex loss. All the methodologies enjoy asymptotic efficiency and consistency properties. Towards the end of chapter 2 we support our methods using simulations and real datasets from health studies namely the infant mortality and bone marrow data. This part comes from the publication, Mukhopadhyay and Bapat (2016a).

In chapter 3 we extend our ideas to a two-sample problem. Here we construct appropriate stopping rules to estimate the difference in means of two independent negative exponential distributions under appropriate Linex loss functions. We assume that the scale parameters are σ and $b\sigma$ respectively where b is known a-priori. Interesting applications are provided using real datasets from cancer research and reliability analysis to support our methodologies. This part comes from the publication, Mukhopadhyay and Bapat (2016b).

Chapter 4 deals with construction of purely sequential stopping rules to estimate the mean of a negative binomial distribution. We cover both one-sample and multi-sample problems. the estimation is again carried out under an appropriate Linex loss and is also supported using a usual squared error loss function. We present some exciting applications using real datasets from ecology namely, the weed count, bird count and raptor count data. This part comes from the submitted manuscripts, Mukhopadhyay and Bapat (2017a,b).

We summarize our work in chapter 5 and then provide some interesting ventures of future research in chapter 6.

Chapter 2
Multistage Point Estimation Methodologies for a Negative Exponential Location Under a Modified Linex Loss Function: Illustrations with Infant Mortality and Bone Marrow Data

2.1. INTRODUCTION

In this chapter we develop estimation methodologies for a negative exponential location under a modified Linex loss function. This chapter is based on Mukhopadhyay and Bapat (2016a). We consider a negative exponential distribution having the following *probability density function* (p.d.f.):

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right) I(x>\mu), \tag{2.1.1}$$

where $\mu \in (-\infty,\infty)$ and $\sigma \in (0,\infty)$ are *unknown parameters.* $I(.)$ denotes an indicator function of (.) which takes the value 1 (or 0) when $x >$ (or $\leq$)μ. This distribution is also known as a two-parameter exponential distribution. The parameter μ, if positive, may be interpreted as the minimum guarantee time or the threshold of the distribution in the sense that no failure will occur before μ. The parameter σ is called a scale.

The p.d.f. (2.1.1) has found its use in many reliability problems such as to depict failure times of electrical component and complex equipment. One may refer to Johnson and Kotz (1970), Bain (1978), and Balakrishnan and Basu (1995) for illustrations. Another area where it has been used is clinical trials such as in studying the behavior of tumor systems in animals and analysis of survival data in cancer research. One may refer to Zelen (1966). We provide illustrations with real data analysis in Section 8.

Before we go any further, we explain our notation clearly. Since both μ,σ are unknown, the parameter vector $\boldsymbol{\theta} = (\mu,\sigma)$ remains unknown. When we write $P(.)$ or $E(.)$, they should be interpreted as $P_{\boldsymbol{\theta}}(.)$ or $E_{\boldsymbol{\theta}}(.)$ respectively. In the same spirit, when we write $\xrightarrow{P}$ (convergence in probability) or w.p.1 (with probability one) or $\xrightarrow{\pounds}$ (convergence in law or distribution), they are all with respect to $P_{\boldsymbol{\theta}}$. We drop subscript $\boldsymbol{\theta}$ for simplicity.

We address methodologies for estimating μ under a *variant* of a customary Linex loss function defined as follows:

$$L_n \equiv L_n(\hat{\mu}_n, \mu) = \exp\left(\frac{a(\hat{\mu}_n - \mu)}{\sigma}\right) - \frac{a(\hat{\mu}_n - \mu)}{\sigma} - 1, \tag{2.1.2}$$

where $\hat{\mu}_n$ is meant to be a generic estimator of μ based on a random sample of size n. This modified Linex loss function (2.1.2) is different from the one that was first proposed by Varian (1975), namely,

$$\exp\left(a(\hat{\mu}_n - \mu)\right) - a(\hat{\mu}_n - \mu) - 1. \tag{2.1.3}$$

The Linex loss (2.1.3) was an appropriate function to be considered in cases of an asymmetric penalty due to bias. It was supposed to address estimation error by penalizing over-estimation and under-estimation unequally where over-estimation is deemed more (less) serious than under-estimation when $a > (<) 0$.

Varian (1975) and Zellner (1986) popularized Linex loss (2.1.3) and brought it to the forefront of statistical science with interesting applications. We will explain in Section 2.2 why we have replaced (2.1.3) with (2.1.2) and then propose to work under this modified Linex loss (2.1.2). We may add that under a squared error loss function (or its variant), sequential inference problems devoted to negative exponential distributions can be found in Mukhopadhyay (1974,1984,1988,1995) and other sources.

A sequential point estimation problem under a Linex loss function (2.1.3) was first developed by Chattopadhyay (1998) utilizing nonlinear renewal theory from Woodroofe (1977,1982) and Lai and Siegmund (1977,1979). Methodologies pertaining to Linex loss (2.1.3) were developed by Chattopadhyay (2000), Chattopadhyay et al. (2000), Takada (2000,2006), Takada and Nagao (2004), Chattopadhyay et al. (2005), Chattopadhyay and Sengupta (2006), Zacks and Mukhopadhyay (2009), and others. Many researchers obtained sequential point estimation methods under the Linex loss (2.1.3) largely under a normal distribution. There is a substantial literature available on Bayes sequential estimation problems under Linex loss (1.3) which may be reviewed from Takada (2001), Jokiel-Rokita (2008), Hwang and Lee (2012), and other sources.

Second- and higher-order approximations were developed by Mukhopadhyay and Duggan (1997) and Mukhopadhyay (1999) under a variety of loss functions in the context of appropriately modified two-stage estimation strategies. For a review of two-stage, purely sequential, and other kinds of stopping rules, one may refer to Sen (1981), Woodroofe (1982), Siegmund (1985), Mukhopadhyay and Solanky (1994), Ghosh et al. (1997), Mukhopadhyay et al. (2004), Mukhopadhyay and de Silva (2009), Zacks (2009) and other sources.

2.1.1. The Layout of This Chapter

In Section 2.2, we include some preliminaries plus an explanation for the modified Linex loss (2.1.2), formulation of the risk function, and its optimization by bounding it from above. In Section 2.3, we consider a Stein-type (1945,1949) two-stage procedure where one first gathers pilot data followed by the remaining requisite data determined by the final sample size. We derive some asymptotic properties for this two-stage estimation strategy (Theorem 2.3.1). Some technical details follow from Chow and Robbins (1965) and Mukhopadhyay (1984,1988,1995).

Section 2.4 introduces a modified two-stage procedure in which we assume a lower bound $\sigma_L(>0)$ for the standard deviation σ along the lines of Mukhopadhyay and Duggan (1997,1999). We develop both first-order (Theorem 2.4.1) as well as second-order (Theorems 2.4.2-2.4.4) properties in the present scenario of a negative exponential setting under the modified Linex loss (2.1.2). Higher-order approximations may be reviewed from Isogai et al. (2011) and seen as natural extensions of Mukhopadhyay and Duggan (1997) and Mukhopadhyay (1999).

Section 2.5 develops a purely sequential methodology followed by its asymptotic first-order properties (Theorem 2.5.1). Then, we exploit nonlinear renewal theory from Woodroofe (1977,1982) and Lai and Siegmund (1977,1979) to obtain requisite second-order approximations (Theorems 2.5.2-2.5.3). For some of the technicalities and second-order properties, we have referred to Mukhopadhyay (1974,1984,1988), Lombard and Swanepoel (1978), Swanepoel and van Wyk (1982), and others.

In Section 2.6, we outline the proofs of selected major results. Section 2.7 highlights performances of the proposed estimation methodologies obtained with the help of computer simulations for a wide range of values of the sample sizes and a large variety of parameter configurations. Our presented data analysis are both extensive and thorough. Selected conclusions from the theorems studied in Sections 2.3-2.6 are critically examined and validated with data analysis in Section 2.7.

These are supplemented with illustrations (Section 2.8) using two real datasets from health studies: The first illustration (Section 2.8.1) uses infant mortality data available from Leinhardt and Wasserman (1979). Our second illustration (Section 2.8.2) uses bone marrow transplant data that came from a multicenter clinical trial with patients prepared for transplantation with a radiation-free conditioning regimen (Klein and Moeschberger, 2003). We end with a brief set of concluding thoughts (Section 2.9).

2.2. MODIFIED LINEX LOSS AND SOME PRELIMINARIES

In this section, we first introduce an appropriately modified Linex loss function (2.1.2) and then calculate the associated risk function. But, why it is that we must modify the customary Linex loss function (2.1.3)? Let us explain.

Having recorded a random sample $X_1, \ldots, X_n$ of *independent and identically distributed* (i.i.d.) observations from a negative exponential distribution (2.1.1), one may customarily estimate μ by the maximum likelihood estimator $X_{n:1}$, the smallest order statistic. Suppose that one begins with the original Linex loss function (2.1.3) from Varian (1975) and Zellner (1986):

$$\exp\left(a(X_{n:1}-\mu)\right) - a(X_{n:1}-\mu) - 1, n \geq 1.$$

Now, the associated risk function requires evaluation of $E[\exp\left(a(X_{n:1}-\mu)\right)]$ which is alternatively expressed as $E\left[\exp\left(\frac{a\sigma}{n}Y\right)\right]$ where $Y \sim Exp(1)$, a standard exponential distribution. But, this particular term, $E\left[\exp\left(\frac{a\sigma}{n}Y\right)\right]$, will be finite provided that $n > a\sigma$. However, since the scale parameter σ remains unknown, there is no way for one to guarantee that a sample size n will certainly exceed $a\sigma$ when $a > 0$. Even if we have some reasonable estimator $\widehat{\sigma}$ of σ available and n exceeds $a\widehat{\sigma}$, it will not guarantee that our sample size n will exceed $a\sigma$ when $a > 0$. Moreover, under two-stage and purely sequential sampling, one will require $n > a\sigma$ for all fixed n which will be hard to guarantee if a is positive.

Remark 2.2.1. If "a" is assumed negative, however, then one may continue to work under the customary Linex loss:

$$\exp\left(a(X_{n:1}-\mu)\right)-a(X_{n:1}-\mu)-1, n\geq 1,$$

without any further modification. We leave out associated details for brevity. In what follows, however, we chart a new direction by providing a unified treatment under (2.2.1) whether a is positive or negative.

2.2.1. Modified Linex Loss Function

We focus on working with fixed a without distinguishing whether a is exclusively positive or negative. Hence, we develop sampling methodologies for estimating μ by $X_{n:1}$ under the *modified Linex loss* function formulated as follows. Instead of (2.1.2), we rewrite it:

$$L_n \equiv L_n(X_{n:1},\mu)=\exp\left(\frac{a(X_{n:1}-\mu)}{\sigma}\right)-\frac{a(X_{n:1}-\mu)}{\sigma}-1, \tag{2.2.1}$$

where a is a constant. We obtain the corresponding risk function by taking expectations across (2.2.1) as follows:

$$\begin{aligned}\text{Risk}_n \equiv E[L_n] &= E\left[\exp\left(\tfrac{a(X_{n:1}-\mu)}{\sigma}\right)-\tfrac{a(X_{n:1}-\mu)}{\sigma}-1\right]\\ &=\left(1-\tfrac{a}{n}\right)^{-1}-\tfrac{a}{n}-1, \text{ for } n>a.\end{aligned} \tag{2.2.2}$$

Upon expanding (2.2.2), we clearly obtain:

$$\text{Risk}_n=\tfrac{a^2}{n^2}+o\left(\tfrac{1}{n^2}\right). \tag{2.2.3}$$

2.2.2. Cost Per Unit Sampling

Next, what we consider is the cost function, $\text{Cost}_n(>0)$, the exact nature of which ought to depend upon the problem on hand. But, it is reasonable to assume that the cost for each observation should go up (or down) as σ goes down (or up). With this understanding, we propose a cost function of the following form:

$$\text{Cost}_n = cn\sigma^{-k} \text{ with fixed and known } c(>0), k(>0). \tag{2.2.4}$$

2.2.3. Proposed Criterion: Bound the Risk Per Unit Cost from Above

We wish to bound the associated "risk" from above where we interpret "risk" as the *risk per unit cost* (RPUC), namely,

$$\text{RPUC}_n \equiv \frac{\text{Risk}_n}{\text{Cost}_n} = \frac{a^2}{n^2}\frac{\sigma^k}{cn} + o\left(n^{-3}\right). \tag{2.2.5}$$

Thus, we fix a suitable number $\omega(>0)$ and require that $\text{RPUC}_n \leq \omega$ for all μ, σ which leads us to determine the required optimal fixed sample size, had σ been known as follows:

$$n \geq \left(\frac{a^2}{c\omega}\right)^{1/3} \sigma^{k/3} = n^*, \text{ say.} \tag{2.2.6}$$

But, the magnitude of n^* is unknown even though its expression is known. Hence, we proceed to develop two-stage, modified two-stage, and purely sequential bounded risk estimation strategies in Sections 2.3-2.5.

Now, having recorded data $X_1, \ldots, X_n$ of fixed size n, we denote $\hat{\sigma}_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{n:1})$ as our customary minimum variance unbiased estimator of σ with $n(>\max\{1,a\})$.

2.2.4. Evaluating the Risk Per Unit Cost When Sample Size Is Random

Clearly, the optimal fixed sample size n^* given by (2.2.6) needs to be estimated. We must begin with pilot data of appropriate size $m(>\max\{1,a\})$ and then move forward step by step with the help of implementing a two-stage, modified two-stage, or purely sequential sampling strategy to record more data subsequently beyond pilot data as needed.

Suppose that a final sample size, denoted by a random variable Q, is determined by an adaptive multistage sampling strategy. Then, the next theorem shows an exact analytical expression for the risk per unit cost associated with the terminal estimator $X_{Q:1} = \min\{X_1, ..., X_Q\}$ of μ once sampling has stopped. Its proof is outlined in Section 2.6.1.

Theorem 2.2.1. *Under a multistage estimation strategy, suppose that* (i) *the final sample size* Q *is an observable random variable that is finite w.p.1, and* (ii) Q *is determined in such a way that the event* $Q = q$ *is measurable with respect to* $\{\hat{\sigma}_j; m \leq j \leq q\}$, for all fixed $q \geq m(>\max\{1,a\})$.

Then, the expression for the risk per unit cost associated with the terminal estimator $X_{Q:1}$ is given by:

$$E\left[\mathrm{RPUC}_Q\right] = \omega \frac{n^{*3}}{a^2}\left\{E\left[\frac{1}{Q}\left(1-\frac{a}{Q}\right)^{-1}\right] - E\left(\frac{a}{Q^2}\right) - E\left(\frac{1}{Q}\right)\right\}. \qquad (2.2.7)$$

2.3. STEIN-TYPE TWO-STAGE PROCEDURE

A Stein-type two-stage procedure along the lines of Stein (1945,1949) may be logistically convenient to implement in certain situations since we may collect data in two batches. At the first stage, we collect pilot data $X_1, \ldots, X_m$ of size $m(> \max\{1, a\})$. Recall that $\hat{\sigma}_m = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - X_{m:1})$ is an estimator of σ obtained from pilot data. We define the stopping rule as:

$$N = \max\left\{m, \left\lfloor d_\omega \hat{\sigma}_m^{k/3}\right\rfloor + 1\right\} \text{ with } d_\omega = \left(a^2(c\omega)^{-1}\right)^{1/3}, \qquad (2.3.1)$$

as an estimator of n^* found in (2.2.6) where

$\lfloor u \rfloor$ denotes the largest integer less than $u(> 0)$.

Now, if $N = m$, then we will not require any more data at the second stage. However, if $N > m$, we sample the difference $N - m$ at the second stage by recording an additional set of observations $X_{m+1}, \ldots, X_N$. From full data obtained by combining both stages, we propose to estimate μ by the smallest order statistic:

$$X_{N:1} = \min\{X_1, ..., X_N\}.$$

Along the lines of (2.2.1), the associated loss function will be:

$$L_N = \exp\left(\frac{a(X_{N:1} - \mu)}{\sigma}\right) - \frac{a(X_{N:1} - \mu)}{\sigma} - 1. \qquad (2.3.2)$$

A major difference between (2.2.1) and (2.3.2) is that the sample size N used in (2.3.2) is a random variable unlike n.

Now, since $X_{n:1}$ and $I(N = n)$ are independent for all fixed $n \geq m$, using (2.2.7) with Q replaced by N from (2.3.1), we get:

$$\omega^{-1}E\left[\mathrm{RPUC}_N\right]=\frac{n^{*3}}{a^2}\left\{E\left[\frac{1}{N}\left(1-\frac{a}{N}\right)^{-1}\right]-E\left(\frac{a}{N^2}\right)-E\left(\frac{1}{N}\right)\right\}. \tag{2.3.3}$$

Theorem 2.3.1. *For the two-stage methodology (2.3.1), for fixed values of μ,σ,c,k,a,m we have as $\omega\to 0$:*

(*i*) $N/n^* \xrightarrow{P} \left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3}$; $n^*/N \xrightarrow{P} \left(\frac{\sigma}{\hat{\sigma}_m}\right)^{k/3}$ *if* $m>\max\{1,a\}$;

(*ii*) $E\left(N/n^*\right)\to E\left[\left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3}\right]$ (>1) *if* $m>\max\{a,1,1-\frac{1}{3}k\}$;

where n^* comes from (2.2.6).

Theorem 2.3.1(ii) shows that the two-stage methodology (2.3.1) oversamples on an average compared with n^* even asymptotically. Such a feature has been observed in the past under numerous Stein-type two-stage estimation strategies. A proof of Theorem 2.3.1 will be sketched in Section 2.6.2.

2.4. MODIFIED TWO-STAGE PROCEDURE

From Theorem 2.3.1, we note that the ratios $E\left(N/n^*\right)$ and $E\left(n^*/N\right)$ did not converge to 1 under the Stein-type two-stage methodology (2.3.1). We thus resort to a modified two-stage procedure along the lines of Mukhopadhyay and Duggan (1997).

A key idea is to introduce a lower bound σ_L such that $0<\sigma_L<\sigma$ with σ_L known. Given this additional input, from the expression of n^* found in (2.2.6), we note that $n^*>\left(a^2(c\omega)^{-1}\right)^{1/3}\sigma_L^{k/3}$. Thus, the pilot size m may be chosen in such a way that $m\approx\left(a^2(c\omega)^{-1}\right)^{1/3}\sigma_L^{k/3}$.

With that spirit, Mukhopadhyay and Duggan (1997) introduced a modification for estimating a normal mean which allowed them to study second-order properties of the associated two-stage estimation strategy. Such a modification has been widely adopted in the literature. Thus, for the problem on hand, we follow along, fix an integer $m_0(>\max\{1,a\})$, and gather pilot data $X_1,...,X_m$ of size m defined as follows:

$$m\equiv m(\omega)=\max\left\{m_0,\left\lfloor d_\omega\sigma_L^{k/3}\right\rfloor+1\right\}\text{ with } d_\omega=\left(a^2(c\omega)^{-1}\right)^{1/3}, \tag{2.4.1}$$

where all the constants remain as defined in Section 2.3.

Based on pilot data, we evaluate the statistic $\hat{\sigma}_m = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - X_{m:1})$, and then determine N as follows:

$$N = \max\left\{m, \left\lfloor d_\omega \hat{\sigma}_m^{k/3} \right\rfloor + 1\right\}, \tag{2.4.2}$$

as an estimator of n^* defined in (2.2.6). Recall that $\lfloor u \rfloor$ denotes the largest integer less than $u(>0)$ as before.

If $N = m$, we would not require any additional observations at the second stage. However, if $N > m$, we sample the difference $N - m$ at the second stage by recording an additional set of observations $X_{m+1}, \ldots, X_N$. From full data $X_1, ..., X_N$ obtained by combining both stages, we propose to estimate μ by the smallest order statistic:

$$X_{N:1} = \min\{X_1, ..., X_N\}.$$

2.4.1. First-Order Asymptotics

We begin with some first-order results to contrast with those stated in Section 2.3. One will surely note that there is no stated sufficient condition involving m in Theorem 2.4.1. This is so because in the present setup, we have $m \equiv m(\omega) \to \infty$ as $\omega \to 0$. A proof will be sketched in Section 2.6.3.

Theorem 2.4.1. *With m and N respectively defined in (2.4.1) and (2.4.2), for each fixed value of μ, σ, c, k, a we have as $\omega \to 0$:*

(*i*) $N/n^* \xrightarrow{P} 1$; $n^*/N \xrightarrow{P} 1$;

(*ii*) $E\left[(N/n^*)^t\right] \to 1, t = -1, 1$ [*asymptotic first-order efficiency*];

where n^* comes from (2.2.6).

We note that all expressions shown in Theorem 2.4.1 converge to 1 which is in direct contrast with those from Theorem 2.3.1. That is, the modified two-stage procedure (2.4.1)-(2.4.2) has more attractive first-order properties than those under customary Stein-type two-stage procedure (2.3.1). One may claim convergence of higher positive and negative moments of N/n^* in part (ii) by referring to Mukhopadhyay and Duggan (1997,1999) and Mukhopadhyay (1999), but we leave them out for brevity.

2.4.2. Second-Order Asymptotics

Again we avoid giving any sufficient condition involving m in Theorems 2.4.2-2.4.4, stated in this section, since $m \equiv m(\omega) \to \infty$ as $\omega \to 0$. We now supplement with second-order properties which become readily accessible along the lines of Mukhopadhyay and Duggan (1997,1999).

The following expressions provide second-order expansions as bounds of $E\left[(N/n^*)^t\right]$:

$$\begin{aligned}
&1 + \left\{t\psi + \tfrac{1}{2}t(t-1)\sigma_0^2\right\}\tfrac{1}{n^*} + o\left(\tfrac{1}{n^*}\right) \le E\left(\tfrac{N}{n^*}\right)^t \\
&\le 1 + \left\{t\psi + t + \tfrac{1}{2}t(t-1)\sigma_0^2\right\}\tfrac{1}{n^*} + o\left(\tfrac{1}{n^*}\right), \text{ if } t > 0; \text{ and} \\
&1 + \left\{t\psi + t + \tfrac{1}{2}t(t-1)\sigma_0^2\right\}\tfrac{1}{n^*} + o\left(\tfrac{1}{n^*}\right) \le E\left(\tfrac{N}{n^*}\right)^t \\
&\le 1 + \left\{t\psi + \tfrac{1}{2}t(t-1)\sigma_0^2\right\}\tfrac{1}{n^*} + o\left(\tfrac{1}{n^*}\right), \text{ if } t < 0,
\end{aligned} \tag{2.4.3}$$

where

$$\psi = \left(\tfrac{c}{a^2}\right)^{1/3}\left\{\tfrac{k}{6}\left(\tfrac{k}{3}-1\right)\left(\tfrac{a^2}{c}\right)^{1/3}\right\}\left(\tfrac{\sigma}{\sigma_L}\right)^{k/3}, \text{ and } \sigma_0^2 = \tfrac{k^2}{9}\left(\tfrac{\sigma}{\sigma_L}\right)^{k/3}. \tag{2.4.4}$$

A proof of (2.4.3) follows directly from a more generally stated Theorem 2.1 in Mukhopadhyay and Duggan (1999). Indeed, Theorem 2.4.1, part (ii) holds for all non-zero t which directly follows from (2.4.3). For completeness, however, we first show a bound for $E(N - n^*)$.

Theorem 2.4.2. *With m and N respectively defined in (2.4.1) and (2.4.2), for each fixed value of μ, σ, c, k, a we have as $\omega \to 0$:*

$$\psi + O(\omega^{1/2}) \le E(N - n^*) \le \psi + 1 + O(\omega^{1/2})[\textit{asymptotic second-order efficiency}]; \tag{2.4.5}$$

where ψ is defined in (2.4.4) and n^* comes from (2.2.6).

A proof can be constructed using (2.6.2) and the rest is omitted for brevity. Theorem 2.4.2 shows the second-order efficiency property of the modified two-stage procedure (2.4.1)-(2.4.2) in the sense of Ghosh and Mukhopadhyay (1981). It is possible, however, to show a sharper result, namely $E(N - n^*) = \psi + O(\omega^{1/2})$, but we leave it out for brevity.

Next, we provide a result which obtains the asymptotic distribution of a standardized version of N along the lines of Ghosh and Mukhopadhyay (1975) and Mukhopadhyay and Duggan (1997,1999). Its proof follows from Lemma 2.1, part (i) in Mukhopadhyay and Duggan (1999) and hence it is omitted.

Theorem 2.4.3. *With m and N respectively defined in (2.4.1) and (2.4.2), for each fixed value of μ, σ, c, k, a we have as $\omega \to 0$:*

$$U \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_0^2), \tag{2.4.6}$$

where σ_0^2 is defined in (2.4.4) and n^* comes from (2.2.6).

Again, since $X_{n:1}$ and $I(N = n)$ are independent for all fixed $n \geq m$, the associated expression for $E\left[\text{RPUC}_N\right]$ will resemble (2.2.7) with Q replaced by N from (2.4.2). The following theorem evaluates the risk per unit cost up to second-order approximation. We outline its proof in Section 2.6.4.

Theorem 2.4.4. *With m and N respectively defined in (2.4.1) and (2.4.2), for each fixed value of μ, σ, c, k, a we have as $\omega \to 0$:*

$$\begin{aligned} 1 + \tfrac{1}{n^*}(6\sigma_0^2 + a - 3\psi - 3) + o(\tfrac{1}{n^*}) &\leq \omega^{-1} E\left[\text{RPUC}_N\right] \\ &\leq 1 + \tfrac{1}{n^*}(6\sigma_0^2 + a - 3\psi) + o(\tfrac{1}{n^*}) \end{aligned} \tag{2.4.7}$$

[*asymptotic second-order risk efficiency*];

where ψ is as defined in (2.4.4) and n^* comes from (2.2.6).

2.5. A PURELY SEQUENTIAL PROCEDURE

Unlike a modified two-stage procedure where we record observations in two batches, a purely sequential procedure is more involved operationally, but it also provides more accurate inferences. We only take as many observations step-by-step as required depending on the rule of termination.

In this section, we develop a purely sequential methodology and make use of nonlinear renewal theory to provide second-order approximations for the average sample size and RPUC, the risk per unit cost.

We recall the expressions of n^* from (2.2.6). We again fix an integer $m(> \max\{1, a\})$ and obtain pilot data $X_1, \ldots, X_m$ of size m. We then proceed by recording one additional observation X at every stage as needed according to the following rule:

$$N = \inf\left\{n \geq m : n \geq d_\omega \hat{\sigma}_n^{k/3}\right\} \text{ with } d_\omega = \left(a^2(c\omega)^{-1}\right)^{1/3}. \tag{2.5.1}$$

As before, this stopping variable N again estimates n^* from (2.2.6). From full data $X_1, ..., X_N$ so obtained upon termination, we propose to estimate μ by the smallest order statistic:

$$X_{N:1} = \min\{X_1, ..., X_N\}.$$

Again, since $X_{n:1}$ and $I(N = n)$ are independent for all fixed $n \geq m$, the associated expression for $E\left[\text{RPUC}_N\right]$ will resemble (2.2.7) with Q replaced by N from (2.5.1). In what follows, if the condition $m(> \max\{1, a\})$ will suffice for a result to hold, then we will not mention it. We will, however, mention a more stringent condition on m if that is what is required for a particular result to hold.

2.5.1. First-Order Asymptotics

We begin with some useful first-order asymptotic properties summarized in the next theorem. A proof of this theorem will be outlined in Section 2.6.5.

Theorem 2.5.1. *For N defined in (2.5.1), for each fixed value of μ, σ, a, k, c we have as $\omega \to 0$:*

- *(i)* $N/n^* \xrightarrow{P} 1$;
- *(ii)* $E\left[(N/n^*)^t\right] \to 1$ *for* $t > 0$ *if* $m(> \max\{2, a\})$ [*asymptotic first-order efficiency*];
- *(iii)* $E\left[(n^*/N)^t\right] \to 1$ *for* $t > 0$ *if* $m > \max\{1 + \frac{1}{3}kt, a\}$;
- *(iv)* $\omega^{-1} E\left[\text{RPUC}_N\right] \to 1$ *if* $m > \max\{1 + \frac{4}{3}k, a\}$ [*asymptotic first-order risk efficiency*];
- *(v)* $V \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2)$ *with* $\sigma_1^2 = \frac{k^2}{9}$;

where n^* comes from (2.2.6).

We note that the expressions shown in Theorem 2.5.1, parts (i)-(iii) converge to 1 which is again in direct contrast with those from Theorem 2.3.1. That is, parallel to the results from Theorem 2.4.1 under the modified two-stage procedure (2.4.1)-(2.4.2), the purely sequential procedure (2.5.1) also has attractive asymptotic first-order properties than those under customary Stein-type two-stage procedure (2.3.1).

However, one major difference between the properties of the modified two-stage procedure (2.4.1)-(2.4.2) and the purely sequential procedure (2.5.1) already stands out. We recall that Theorem 2.4.3 concluded that

$$U \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_0^2) \text{ with } \sigma_0^2 = \tfrac{k^2}{9}(\tfrac{\sigma}{\sigma_L})^{k/3}.$$

In contrast, Theorem 2.5.1, part (v) shows that

$$V \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2) \text{ with } \sigma_1^2 = \tfrac{k^2}{9}.$$

In other words, asymptotic normal distribution for the standardized stopping variable V certainly has a smaller or tighter variance, namely, σ_1^2 under the purely sequential strategy which gives it an edge over the modified two-stage strategy.

2.5.2. Second-Order Asymptotics

Observe that N from (2.5.1) can be rewritten as $J + 1$ w.p.1 where

$$\begin{aligned} J &= \inf\left\{n \geq m - 1 : \Sigma_{i=1}^n Z_i \leq n^{*-3/k} n^{\frac{3}{k}+1}(1 + \tfrac{1}{n})^{\frac{3}{k}}\right\} \\ &= \inf\left\{n \geq m - 1 : \Sigma_{i=1}^n Z_i \leq h^* n^\delta L(n)\right\}, \end{aligned} \tag{2.5.2}$$

where the Z_i's are i.i.d. $Exp(1)$ random variables. This has been established inside the proof of Theorem 2.5.1, part (iii) that is laid out in Section 2.6.5.

Next, we refer to nonlinear renewal theory, originally developed by Woodroofe (1977) and Lai and Siegmund (1977,1979). We match (2.5.2) with the representation laid out in Mukhopadhyay (1988) and Mukhopadhyay and Solanky (1994, Section 2.4.2) as follows:

$$\delta = \tfrac{3}{k} + 1,\ h^* = n^{*-3/k},\ L(n) = 1 + \tfrac{3}{kn} + o\left(\tfrac{1}{n}\right), \text{so that } L_0 = \tfrac{3}{k}, \tag{2.5.3}$$

and also note:

$$\theta = 1\,, \tau^2 = 1, \beta^* = \tfrac{k}{3}, n_0^* = n^* \text{ and } p = \tfrac{k^2}{9}. \tag{2.5.4}$$

Condition (2.5) from Mukhopadhyay (1988), or equivalently (2.4.8) from Mukhopadhyay and

Solanky (1994), is satisfied with $B = 1$ and $b = 1$. We define two special entities:

$$\begin{aligned} \nu \equiv \nu_k &= \tfrac{k}{6}\left(\tfrac{9}{k^2}+1\right) - \tfrac{1}{2}\Sigma_{n=1}^{\infty} n^{-1} E\left\{\max\left(0, \chi^2_{2n} - 2n\left(\tfrac{3}{k}+1\right)\right)\right\}, \\ \eta \equiv \eta_k &= \tfrac{1}{3}k\nu - \left(\tfrac{1}{18}k^2 + \tfrac{1}{6}k + 1\right), \end{aligned} \tag{2.5.5}$$

along the lines of (2.4.9)-(2.4.10) in Mukhopadhyay and Solanky (1994). Table 2.1 shows a few values of $\nu \equiv \nu_k$ and $\eta \equiv \eta_k$ for a couple of combinations of input for k.

To conclude this section, we now specify asymptotic second-order expansions of both positive and negative moments of $\frac{N}{n^*}$ (Theorem 2.5.2) as well as an asymptotic second-order expansion of the risk per unit cost (Theorem 2.5.3) under the purely sequential setting (2.5.1).

Theorem 2.5.2. *For N defined in (2.5.1), for each fixed value of μ, σ, a, k, c and every non-zero real number t, we have as $\omega \to 0$:*

$$E\left[(N/n^*)^t\right] = 1 + \left\{t\eta_k + t + \tfrac{1}{2}t(t-1)p\right\} n^{*-1} + o\left(n^{*-1}\right) \tag{2.5.6}$$

[asymptotic second-order efficiency];

when (i) $m > \frac{(3-t)k}{3} + 1$ for $t \in (-\infty, 2) - \{-1, 1\}$; (ii) $m > \frac{k}{3} + 1$ for $t = 1$ and $t \geq 2$; and (iii) $m > \frac{2k}{3} + 1$ for $t = -1$; where n^*, p, and η_k come from (2.2.6), (2.5.4), and (2.5.5) respectively.

Theorem 2.5.3. *For N defined in (2.5.1), for each fixed value of μ, σ, a, k, c, we have the following second-order expansion of the risk per unit cost as $\omega \to 0$:*

$$\omega^{-1} E\left[\mathrm{RPUC}_N\right] = 1 + (6p + a - 3\eta_k - 3)n^{*-1} + o\left(n^{*-1}\right) \tag{2.5.7}$$

[asymptotic second-order risk efficiency];

when $m > \frac{7k}{3} + 1$ with n^*, p, and η_k coming from (2.2.6), (2.5.4), and (2.5.5) respectively.

Very brief outlines of proofs of Theorems 2.5.2-2.5.3 are sketched in Section 2.6.6. We have simply remarked how to connect what we want to prove here with established details from Mukhopadhyay (1988) and Mukhopadhyay and Solanky (1994).

2.6. TECHNICAL DETAILS AND PROOFS OF THEOREMS

In this section, we sketch some of the proofs of our main results from Sections 2.2-2.5. Often some intermediate steps are left out for brevity.

2.6.1. Proof of Theorem 2.2.1

From (2.2.1), the associated loss function is given by:

$$L_Q = \exp\left(\frac{a(X_{Q:1}-\mu)}{\sigma}\right) - \frac{a(X_{Q:1}-\mu)}{\sigma} - 1.$$

Now, we recall risk per unit cost from (2.2.5) and proceed to evaluate $E\left[\mathrm{RPUC}_Q\right]$ as follows:

$$\begin{aligned}
&E\left[\tfrac{L_Q}{\mathrm{Cost}_Q}\right] \\
&\quad = \sum_{m\le q<\infty} E\left[\tfrac{L_Q}{\mathrm{Cost}_Q} \mid Q=q\right] P(Q=q) \\
&\quad = \sum_{m\le q<\infty} E\left[\tfrac{L_q}{\mathrm{Cost}_q} \mid Q=q\right] P(Q=q).
\end{aligned}$$

But, under the conditions (i) and (ii) stated in Theorem 2.2.1, we may rewrite:

$$\begin{aligned}
&E\left[\mathrm{RPUC}_Q\right] \\
&\quad = \sum_{m\le q<\infty} E\left[\tfrac{L_q}{\mathrm{Cost}_q}\right] P(Q=q) \\
&\quad = \sum_{m\le q<\infty} \tfrac{E[L_q]}{\mathrm{Cost}_q} P(Q=q) \\
&\quad = \sum_{m\le q<\infty} \tfrac{\mathrm{Risk}_q}{\mathrm{Cost}_q} P(Q=q) \\
&\quad = \sum_{m\le q<\infty} \left\{\left(1-\tfrac{a}{q}\right)^{-1} - \tfrac{a}{q} - 1\right\}(cq\sigma^{-k})^{-1},
\end{aligned}$$

utilizing previous expressions of Risk_q and Cost_q from (2.2.2) and (2.2.4) respectively while substituting q for n.

From (2.2.6), we see that $\frac{\sigma^k}{c} = \frac{n^{*3}\omega}{a^2}$, and thus we obtain:

$$E\left[\mathrm{RPUC}_Q\right] = \omega\frac{n^{*3}}{a^2}E\left[\tfrac{1}{Q}\left(1-\tfrac{a}{Q}\right)^{-1} - \tfrac{a}{Q^2} - \tfrac{1}{Q}\right].$$

This is (2.2.7). ■

2.6.2. Stein Type Two-Stage Procedure: Proof of Theorem 2.3.1

From (2.3.1) we have the following basic inequality:

$$d_\omega \hat{\sigma}_m^{k/3} \le N \le m + d_\omega \hat{\sigma}_m^{k/3} \text{ w.p.1},$$

which implies:

$$n^{*-1} d_\omega \hat{\sigma}_m^{k/3} \le n^{*-1} N \le n^{*-1} m + n^{*-1} d_\omega \hat{\sigma}_m^{k/3} \text{ w.p.1.} \tag{2.6.1}$$

Now, as $\omega \to 0$, we have $n^* \to \infty$, $\frac{m}{n^*} \to 0$, so that $\frac{N}{n^*} \to \left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3}$ w.p.1. The rest of part (i) follows immediately.

Part (ii) follows by taking expectations throughout (2.6.1) and then taking limits as $\omega \to 0$. Observe that $\frac{2(m-1)}{\sigma}\hat{\sigma}_m \sim \chi^2_{2(m-1)}$ which implies:

$$E\left[\hat{\sigma}_m^{k/3}\right] = \left(\frac{\sigma}{m-1}\right)^{k/3} \Gamma(m-1+k/3)\{\Gamma(m-1)\}^{-1}.$$

The proof that $E\left[\left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3}\right] > 1$ follows from Mukhopadhyay and Hilton (1986). ■

2.6.3. Modified Two-Stage Procedure: Proof of Theorem 2.4.1

Part (i): Utilizing (2.6.1), we have the following basic inequality:

$$d_\omega \hat{\sigma}_m^{k/3} \le N \le mI(N=m) + d_\omega \hat{\sigma}_m^{k/3} + 1 \text{ w.p.1.} \tag{2.6.2}$$

Along the lines of Mukhopadhyay and Duggan (1999, Lemma 2.1), we know that $P(N=m) = O(\kappa^{m-1})$ where $\kappa = \frac{\sigma_L}{\sigma} \exp\left\{1 - \frac{\sigma_L}{\sigma}\right\}$ so that $\kappa \in (0,1)$. Thus, we can claim that $I(N=m) \xrightarrow{P} 0$ as $\omega \to 0$. Now, after dividing (2.6.2) throughout by n^*, we get:

$$\left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3} \le \frac{N}{n^*} \le \frac{m}{n^*} I(N=m) + \left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3} + \frac{1}{n^*} \text{ w.p.1.} \tag{2.6.3}$$

But, $\hat{\sigma}_m \xrightarrow{P} \sigma$, $\frac{m}{n^*} = O(1)$, and $n^* \longrightarrow \infty$ as $\omega \to 0$. Part (i) is immediate from (2.6.3).

Part (ii): Taking expectations throughout (2.6.3), we have:

$$E\left[\left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3}\right] \le E\left(\frac{N}{n^*}\right) \le \frac{m}{n^*}P(N=m) + E\left[\left(\frac{\hat{\sigma}_m}{\sigma}\right)^{k/3}\right] + \frac{1}{n^*} \tag{2.6.4}$$

Clearly,

$$E\left[\hat{\sigma}_m^{k/3}\right] = \sigma^{k/3}\left\{1 + \tfrac{k}{3}\left(\tfrac{k}{3}-1\right)(2m)^{-1} + O(m^{-2})\right\}.$$

Thus, a proof in the case of $E\left(\frac{N}{n^*}\right)$ is complete in view of (2.6.4). Next, in order to handle $E\left(\frac{n^*}{N}\right)$, we first note that $\liminf_{\omega\to 0} E\left(\frac{n^*}{N}\right) \ge 1$ by Fatou's lemma. Also, we have

$$\limsup_{\omega\to 0} E\left(\frac{n^*}{N}\right) \le \limsup_{\omega\to 0} E\left(\frac{\sigma}{\hat{\sigma}_m}\right)^{k/3} = 1.$$

Now, the proof is complete. ■

2.6.4. Modified Two-Stage Procedure: Proof of Theorem 2.4.4

We may recall (2.2.7) and then rewrite:

$$\begin{aligned} &\omega^{-1}E\left[\mathrm{RPUC}_N\right] \\ &= \tfrac{n^{*3}}{a^2}\left\{E\left[\tfrac{1}{N}\left(1-\tfrac{a}{N}\right)^{-1}\right] - E\left(\tfrac{a}{N^2}\right) - E\left(\tfrac{1}{N}\right)\right\} \\ &= E\left[\left(\tfrac{N}{n^*}\right)^{-3}\right] + an^{*-1}E\left[\left(\tfrac{N}{n^*}\right)^{-4}\right] + a^2 n^{*-2}E\left[U_N\right], \end{aligned} \tag{2.6.5}$$

where U_N is a remainder term which is $O_P\left(\left(\frac{N}{n^*}\right)^{-5}\right)$.

One can show that $E\left[U_N\right] = O(1)$. Next, we may combine (2.6.5) and (2.4.3)-(2.4.4) with $t = -3, -4$ to express:

$$\omega^{-1}E\left[\mathrm{RPUC}_N\right] \ge 1 + \tfrac{1}{n^*}(6\sigma_0^2 + a - 3\psi - 3) + o(\tfrac{1}{n^*}),$$

and

$$\omega^{-1}E\left[\mathrm{RPUC}_N\right] \le 1 + \tfrac{1}{n^*}(6\sigma_0^2 + a - 3\psi) + o(\tfrac{1}{n^*}).$$

This complete the proof. ■

2.6.5. Purely Sequential Procedure: Proof of Theorem 2.5.1

Part (i): Follows from using Lemma 1 of Chow and Robbins (1965).

Part (ii): With $m \geq 3$ we can claim that $\frac{N-1}{N-2} \leq 2$ w.p.1, $U^* = \sup_{n \geq 2} \left\{ \frac{1}{n} \Sigma_{i=1}^{n} (X_i - \mu) \right\}$, for sufficiently small $\omega (> 0)$ so that $n^* > m$, observe the following inequality (w.p.1):

$$N \leq m + d_\omega \hat{\sigma}_{N-1}^{k/3} \leq m + 2^{k/3} d_\omega \left\{ \tfrac{1}{N-2} \Sigma_{i=1}^{N-1} (X_i - \mu) \right\}^{k/3} \leq m + 2^{k/3} d_\omega U^*,$$

which implies:

$$\tfrac{N}{n^*} \leq 1 + 2^{k/3} \sigma^{-k/3} U^*. \tag{2.6.6}$$

Now, by Wiener's (1939) ergodic theorem, it follows that $E[U^{*t}]$ is finite for all fixed positive number t. The right-hand side of (2.6.6) is also free from ω so that we can claim uniform integrability of all positive powers of $\frac{N}{n^*}$. Then, appealing to the dominated convergence theorem and part (i) complete a proof of part (ii).

Part (iii): Let us denote:

$$S_n^* = \Sigma_{i=2}^{n} (n - i + 1)(X_{n:i} - X_{n:i-1}),$$

so that we have $\hat{\sigma}_n = S_n^*/(n-1)$. Let $Y_1, Y_2, \ldots$ be i.i.d. $Exp(\sigma)$ random variables and let $S_n = \Sigma_{i=1}^{n-1} Y_i$. Now, using the embedding ideas from Lombard and Swanepoel (1978) and Swanepoel and van Wyk (1982), we can claim that the distribution of $\{S_n^* : n \leq n_0\}$ is identical to that of $\{S_n : n \leq n_0\}$ for all n_0. Thus, N given by (2.5.1) is equivalently expressed as:

$$\begin{aligned} N &\equiv \inf \left\{ n \geq m : n \geq d_\omega \left(\tfrac{1}{n-1} \Sigma_{i=1}^{n-1} Y_i \right)^{k/3} \right\} \\ &= \inf \left\{ n \geq m : \left(\tfrac{n}{n^*} \right)^{3/k} (n-1) \geq \Sigma_{i=1}^{n-1} Z_i \right\}, \end{aligned} \tag{2.6.7}$$

where Z_i's are i.i.d. $Exp(1)$, that is, the standard exponential random variables.

Then, N from (2.6.7) can be written as $J + 1$ w.p.1 with J defined in (2.5.2). Using Lemma 2.3 from Woodroofe (1977) or Theorem 2.4.8, part (i) of Mukhopadhyay and Solanky (1994) with $b = 1$, we can claim:

$$P(J \leq \tfrac{1}{2} n^*) = O(n^{*-\frac{3}{k}(m-1)}). \tag{2.6.8}$$

Next, with fixed $t > 0$, since $N = J + 1$ w.p.1, we have $0 < \left(\frac{n^*}{N} \right)^t \leq \left(\frac{n^*}{J} \right)^t$ so that $\left(\frac{n^*}{N} \right)^t$ will be uniformly integrable if we show:

$$\left(\frac{n^*}{J}\right)^t \text{ is uniformly integrable.} \tag{2.6.9}$$

Now, we may write:

$$\left(\frac{n^*}{J}\right)^t I(J > \tfrac{1}{2}n^*) < 2^t,$$

so that $\left(\frac{n^*}{J}\right)^t I(J > \frac{1}{2}n^*)$ must be uniformly integrable. But, $\left(\frac{n^*}{J}\right)^t I(J > \frac{1}{2}n^*) \xrightarrow{P} 1$ and hence, we must have:

$$E\left[\left(\frac{n^*}{J}\right)^t I(J > \tfrac{1}{2}n^*)\right] = 1 + o(1). \tag{2.6.10}$$

Additionally, in view of (2.6.8), we also note the following:

$$E\left[\left(\frac{n^*}{J}\right)^t I(J \le \tfrac{1}{2}n^*)\right] \le \left(\frac{n^*}{m-1}\right)^t P(J \le \tfrac{1}{2}n^*) = O(n^{*-\frac{3}{k}(m-1)+t}), \tag{2.6.11}$$

which is $o(1)$ if $m > 1 + \frac{1}{3}kt$.

Combining (2.6.10)-(2.6.11), we note that (2.6.9) will follow so that we can claim:

$$E\left[\left(\frac{n^*}{J}\right)^t\right] = 1 + o(1) \text{ if } m > 1 + \tfrac{1}{3}kt, \text{ with } t > 0, \tag{2.6.12}$$

which is part (iii).

Part (iv): We may go back to (2.6.5) and express:

$$\omega^{-1}E\left[\text{RPUC}_N\right] = E\left[\left(\frac{N}{n^*}\right)^{-3}\right] + an^{*-1}E\left[V_N\right], \tag{2.6.13}$$

where V_N is a remainder term which is $O_P\left(\left(\frac{N}{n^*}\right)^{-4}\right)$.

Clearly, in view of (2.6.12), $E\left[\left(\frac{n^*}{J}\right)^3\right] = 1 + o(1)$ if $m > 1 + k$ and one can show that $E\left[V_N\right] = O(1)$ if $m > 1 + \frac{4}{3}k$. Then, part (iv) follows from (2.6.13).

Part (v): This result follows directly from an application of Ghosh and Mukhopadhyay's (1975) theorem. One may also refer to Theorem 2.4.3 or Theorem 2.4.8, part (ii) in Mukhopadhyay and Solanky (1994). ■

2.6.6. Purely Sequential Procedure: Outlines of Proofs of Theorems 2.5.2-2.5.3

Theorem 2.5.2 follows along the lines of Theorem 2.4.8, part (iv) and from its established applications found in Mukhopadhyay and Solanky (1994).

For a proof of Theorem 2.5.3, we recall that the associated expression for $E\,[\mathrm{RPUC}_N]$ will resemble (2.2.7) with Q replaced by N from (2.5.1). Then, one will proceed with an expansion of $\omega^{-1}E\,[\mathrm{RPUC}_N]$ similar to that in (2.6.5) and exploit Theorem 2.5.1, part (iii) with $t = 5$ and Theorem 2.5.2 with $t = -3, -4$. Additional details are left out for brevity. ■

2.7. DATA ANALYSIS: SIMULATIONS

Thus far we have developed theoretical properties for three proposed estimation strategies in Sections 2.3-2.5 respectively. Section 2.6 gave outlines of some selected derivations. Now, it is time to implement the methodologies and investigate how those estimation strategies may perform when sample sizes are small $(20, 30)$ to moderate $(50, 100, 150)$ to large $(300, 400, 500)$. Computer simulations help in this investigation. All simulations are carried out with R based on $10,000 (= H$, say) replications run under each configuration and each methodology.

Under each procedure, we generated pseudorandom observations from the distribution (2.1.1) with $\mu = 5, \sigma = 10$. Then, we fixed certain values of a, c, k and appropriate rounded values for n^*, thereby solving for a corresponding value of the risk-bound, ω. In other words, a set of preassigned values for a, c, k, ω will have the associated n^* values as shown in our tables (column 1).

We fix a pilot sample size, namely, m in the contexts of Stein-type two-stage methodology (2.3.1) and purely sequential methodology (2.5.1). In the context of modified two-stage methodology (2.4.1)-(2.4.2), we fix a positive lower bound σ_L for σ and a number m_0, thereby determining m from (2.4.1). While implementing a methodology to determine the final sample size (N) and a terminal estimator $(X_{N:1})$ of μ, we pretended that we did not know μ, σ, and n^* values.

Now, let us set the notations that we will use in the tables to follow. We focus on implementing a particular estimation methodology under a fixed configuration of all necessary input (e.g., $a, c, k, \omega, m, m_0, \sigma_L$ as appropriate). We ran the i^{th} replication by beginning with m pilot observations and then eventually ending sampling by recording the final sample size $N = n_i$, terminal estimator $X_{N:1} = x_{n_i:1}$, and the achieved risk per unit cost:

$$\text{RPUC}_{n_i} = \omega \frac{n^{*3}}{a^2} \left\{ \frac{1}{n_i} \left(1 - \frac{a}{n_i}\right)^{-1} - \frac{a}{n_i^2} - \frac{1}{n_i} \right\} = r_i, \text{ say,} \tag{2.7.1}$$

$i = 1, ..., H$. The basic notations are itemized next where H is kept fixed at $10,000$:

$\overline{n} = H^{-1}\Sigma_{i=1}^{H} n_i$	Estimate of $E(N)$ or n^*;
$\text{s}_{\overline{n}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(n_i - \overline{n})^2}$	Estimated standard error of $\overline{n}$;
$\overline{x}_{\min} = H^{-1}\Sigma_{i=1}^{H} x_{n_i:1}$	Estimate of μ;
$\text{s}_{\overline{x}_{\min}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(x_{n_i:1} - \overline{x}_{\min})^2}$	Estimated standard error of $\overline{x}_{\min}$;
r_i	RPUC_{n_i} from (2.7.1)
$\overline{r} = H^{-1}\Sigma_{i=1}^{H} r_i$ with r_i from (7.1)	Risk estimator
$\text{s}_{\overline{r}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(r_i - \overline{r})^2}$	Estimated standard error of $\overline{r}$;
$\overline{z} = \overline{r}/\omega$	Estimated risk efficiency to be; compared with 1;
$\text{s}_{\overline{z}} = \text{s}_{\overline{r}}/\omega$	Estimated standard error of $\overline{z}$;

Now, we are in a position to summarize observed performances of the proposed estimation methodologies laid down in Sections 2.3-2.5. We have many sets of tables and results obtained from simulations under additional configurations. For brevity, however, we outline only a small subset of our findings.

2.7.1. Stein-Type Two-Stage Procedure (2.3.1)

First we present performances of Stein-type two-stage estimation methodology (2.3.1) in Table 2.2 for

$$n^* = 30, 100, 300, 500 \text{ and}$$

$$(k, m) = (1, 3), (2, 4), (3, 5), (4, 6), (5, 7).$$

Table 2.2 specifies μ, σ, a, c. Each block shows (k, m), n^* (column 1), ω (column 2), the estimated (from $10,000$ simulations) values $\overline{x}_{\min}$,$\text{s}_{\overline{x}_{\min}}$ (column 3), values $\overline{n}$,$\text{s}_{\overline{n}}$ (column 4), the ratio $\overline{n}/n^*$ (column 5), and values $\overline{z}$,$\text{s}_{\overline{z}}$ (column 6).

All $\overline{x}_{\min}$ values appear rather close to $5(=\mu)$ with very small estimated standard error values $s_{\overline{x}_{\min}}$, for sample sizes over 300. For smaller (k,m) values, namely, $(1,3)$, $(2,4)$, it appears that $\overline{n}$ slightly underestimates n^*, but these performances reverse for larger choices of (k,m). This is consistent with Theorem 2.3.1, part (ii). The last column shows that the two-stage estimation methodology (2.3.1) is not successful in delivering a risk-bound approximately under (or close to) our preset goal ω.

2.7.2. Modified Two-Stage Procedure (2.4.1)-(2.4.2)

Now, we move to summarize performances for the modified two-stage estimation methodology (2.4.1)-(2.4.2) in Table 2.3 for

$$n^* = 20, 30, 50, 100, 150, 300, 400, 500 \text{ and}$$
$$(k, m_0) = (1,3), (2,4), (3,5), (4,6), (5,7).$$

In this methodology, we need a positive and known lower bound $\sigma_L(=3)$ for true σ, but σ remains unknown. The pilot size m was determined from (2.4.1) but that m is not shown in Table 2.3. The estimation methodology (2.4.2) was implemented as described. Table 2.3 specifies $\mu, \sigma, \sigma_L, a, c$, and each block shows (k, m_0), n^* (column 1), ω (column 2), the estimated (from $10,000$ simulations) values $\overline{x}_{\min}$,$s_{\overline{x}_{\min}}$ (column 3), values $\overline{n}$,$s_{\overline{n}}$ (column 4), the ratio $\overline{n}/n^*$ (column 5), and values $\overline{z}$,$s_{\overline{z}}$ (column 6).

All $\overline{x}_{\min}$ values appear closer to $5(=\mu)$ with very small estimated standard error values $s_{\overline{x}_{\min}}$, for sample size 150 or over. For all (k,m) values, it appears that $\overline{n}$ estimates n^* very accurately across the board. These features are consistent with Theorem 2.4.1, parts (i)-(ii). The last column shows that the modified two-stage estimation methodology (2.4.1)-(2.4.2) is very successful for $(k,m) = (1,3)$ for all n^* values under consideration in delivering a risk-bound approximately our preset goal ω. However, in the case of $k = 2,3,4,5$, a similar level of success in delivering a risk-bound approximately our preset goal ω is observed as n^* successively exceeded $50, 100, 150, 300$ respectively.

The entries in Table 2.4 were obtained as those in Table 2.3 with one major difference. Table 2.4 used another positive and known lower bound $\sigma_L(=5)$ for true σ, but σ continued to remain unknown. Again, the pilot size m was determined from (2.4.1) but m is not shown in Table 2.4.

Comparing columns 5 from Tables 2.3 and 2.4, it is clear $\overline{n}/n^*$ is nearer to 1 in Table 2.4. Also, entries found in the last column of Table 2.4 look more favorable to those in Table 2.3. This feature should be expected since the specified positive and known lower bound $\sigma_L(=5)$ is nearer to $\sigma(=10)$ than $\sigma_L(=3)$ is.

Table 2.5 corresponds to $\sigma_L = 11$ whereas true σ was 10. That is, Table 2.5 shows performances of the modified two-stage estimation methodology (2.4.1)-(2.4.2) if σ_L is misspecified in that it just barely goes over true σ. It is clear that the modified two-stage estimation methodology holds up rather well under mild misspecification of σ_L.

In Tables 2.6-2.8, we provide the values of ψ found in (2.4.4) corresponding to the configurations highlighted in Tables 2.3-2.5 respectively. Importance of ψ comes from the fact that $E(N - n^*)$ values could be expected to lie inside the corresponding interval $[\psi, \psi + 1]$ for large n^* values in view of Theorem 2.4.2. Thus, one could expect $\bar{n} - n^*$ values to lie inside the corresponding interval $[\psi, \psi + 1]$ for large n^* values. In Tables 2.6-2.7, we find that nearly all $\bar{n} - n^*$ values lie very close to (or inside) the corresponding interval $[\psi, \psi + 1]$ along with the bold entries missing the boat by a wider margin. But, the entries in Table 2.8 correspond to the case of misspecifying σ_L considered by Table 2.5, and hence those entries in Table 2.8 are supposed to be completely out of line with regard to any sense of practicality of second-order approximation under misspecification of σ_L.

We now provide Figures 2.1-2.2 to validate empirically the normality result described in (2.4.6). We considered four scenarios, namely $\sigma_L = 3$, $k = 1$, $n^* = 30$ (Figure 2.1a); $\sigma_L = 3$, $k = 4$, $n^* = 500$ (Figure 2.1b); $\sigma_L = 5$, $k = 2$, $n^* = 100$ (Figure 2.2a); and $\sigma_L = 5$, $k = 5$, $n^* = 500$ (Figure 2.2b). Under a specific configuration, we recorded observed values:

$$N = n_i, i = 1, ..., H(= 10,000),$$

and thus calculated $10,000$ associated standardized $(n_i - n^*)/\sqrt{n^*}$ values. Under each specific configuration, such $10,000$ observed $u_i \equiv (n_i - n^*)/\sqrt{n^*}$ values provide the empirical distribution of the standardized sample size (dashed curve in red). We superimpose on it the appropriate theoretical $N(0, \sigma_0^2)$ distributions (solid curve in blue) where $\sigma_0^2 = \frac{1}{9}(\frac{\sigma}{\sigma_L})^{k/3}k^2$ coming from (2.4.4).

The two curves as shown appear slightly off from each other in Figure 2.1a which is meant for small $n^*(= 30)$, however we recall that a visual asymptotic match should be expected for larger n^*

(Theorem 2.4.3). Indeed, Figures 2.1b and 2.2a-2.2b show clearly that the empirical distribution curve and the theoretical distribution curve nearly lie on each other when $n^*(= 100, 500)$ is large.

On $10,000$ observed values of u, we also performed the customary Kolmogorov-Smirnov (K-S) test for normality in the case of each dataset that generated Figures 2.1a-2.1b and Figures 2.2a-2.2b. We summarize the observed values of K-S test statistic (D) under the null hypothesis of normality with associated p-values.

Case	Parameter configuration	K-S stat D	p-value
2.1a	$\sigma_L = 3,\ k = 1,\ n^* = 30$	0.56149	0.8770
2.1b	$\sigma_L = 3,\ k = 4,\ n^* = 500$	0.51959	0.9608
2.2a	$\sigma_L = 5,\ k = 2,\ n^* = 100$	0.51191	0.9762
2.2b	$\sigma_L = 5,\ k = 5,\ n^* = 500$	0.51170	0.9766

The p-values shown under cases 2.1b, 2.2a, 2.2b are sizably larger than the p-value shown under case 2.1a and all four p-values are much larger than 0.05. Our earlier sentiments supported by visual examinations of Figures 2.1a-2.1b and Figures 2.2a-2.2b are clearly validated by K-S test of normality under each scenario. That is, we are reasonably assured of a good fit between the observed values of u and a normal curve with a high level of confidence for all practical purposes.

2.7.3. Purely Sequential Procedure (2.5.1)

In this section, we summarize performances for the purely sequential estimation methodology (2.5.1) in Table 2.9 for

$$n^* = 20, 30, 50, 100, 150, 300, 400, 500 \text{ and}$$
$$(k, m) = (1, 5), (2, 8), (3, 9), (4, 12), (5, 15).$$

The estimation methodology (2.5.1) was implemented as described. Table 2.9 specifies μ, σ, a, c and each block shows (k, m), n^* (column 1), ω (column 2), the estimated (from $10,000$ simulations) values $\overline{x}_{\min}$,$s_{\overline{x}_{\min}}$ (column 3), values $\overline{n}$,$s_{\overline{n}}$ (column 4), the ratio $\overline{n}/n^*$ (column 5), and values $\overline{z}$,$s_{\overline{z}}$ (column 6). An explanation of column 7 comes later.

All $\overline{x}_{\min}$ values appear closer to $5(= \mu)$ with very small estimated standard error values $s_{\overline{x}_{\min}}$, for sample size 100 or over. For all (k, m) values, it appears that $\overline{n}$ estimates n^* very accurately across the board with slight occasional over- or under-estimation. These features are consistent with Theorem 2.5.1, parts (i)-(ii). Column 6 shows that the purely sequential estimation methodology (2.5.1) is overall very successful for $(k, m) = (1, 5)$ for all n^* values under consideration in delivering a risk-bound approximately our preset goal ω. However, in the case of $k = 2, 3, 4, 5$, a similar level of success in delivering a risk-bound approximately our preset goal ω is observed as n^* successively exceeded $50, 100, 150, 300$ respectively. It is our feeling that the purely sequential procedure provides estimates for the risk per unit cost which are generally closer to 1 than the two-stage methodologies (2.3.1) or (2.4.1)-(2.4.2).

Theorem 5.3 showed that $\omega^{-1}E\left[\text{RPUC}_N\right]$ should be close to $1 + \varepsilon$ where we have:

$$\varepsilon = (6p + a - 3\eta_k - 3)n^{*-1}, \tag{2.7.2}$$

from (2.5.7) when n^* is large. η_k was defined by (2.5.5) and it was tabulated in Table 2.1. Column 7 in Table 2.9 shows these ε values under each configuration. Upon comparing $\overline{z}$ values from column 6 with ε values from column 7, it appears that the second-order approximation is little slow to take hold under some configurations. It appears, however, that for all practical purposes, the $\overline{z}$ values are described fairly well by $1 + \varepsilon$, more or less all across the board.

We now provide Figures 2.3-2.4 to validate empirically the normality result described in Theorem 2.5.1, part (v). We considered four scenarios, namely $m = 5,\ k = 1,\ n^* = 30$ (Figure 2.3a); $m = 12$, $k = 4$, $n^* = 500$ (Figure 2.3b); $m = 8$, $k = 2$, $n^* = 100$ (Figure 2.4a); and $m = 15$, $k = 5$, $n^* = 500$ (Figure 2.4b). Under a specific configuration, we again recorded observed values:

$$N = n_i, i = 1, ..., H(= 10,000)$$

and thus calculated $10,000$ associated standardized $(n_i - n^*)/\sqrt{n^*}$ values. Under each specific configuration, such $10,000$ observed $v_i \equiv (n_i - n^*)/\sqrt{n^*}$ values provided the empirical distribution of the standardized sample size (dashed curve in red). We superimpose on it the appropriate theoretical $N(0, \sigma_1^2)$ distributions (solid curve in blue) where $\sigma_1^2 = \frac{1}{9}k^2$ coming from Theorem 2.5.1, part (v).

The two curves as shown appear slightly off from each other in Figure 2.3a which is meant for small $n^*(= 30)$, however we recall that a visual asymptotic match should be expected for larger n^* (Theorem 2.5.1, part (v)). Indeed, Figures 2.3b and 2.4a-2.4b show clearly that the empirical distribution curve and the theoretical distribution curve nearly lie on each other when $n^*(= 100, 500)$ is large.

On $10,000$ observed values of v, we also performed the customary Kolmogorov-Smirnov (K-S) test for normality in the case of each dataset that generated Figures 2.3a-2.3b and Figures 2.4a-2.4b. We summarize the observed values of K-S test statistic (D) under the null hypothesis of normality with associated p-values.

Case	Parameter configuration	K-S stat D	p-value
2.3a	$m = 5,\ k = 1,\ n^* = 30$	0.56679	0.8664
2.3b	$m = 12,\ k = 4,\ n^* = 500$	0.51066	0.9787
2.4a	$m = 8,\ k = 2,\ n^* = 100$	0.50022	0.9996
2.4b	$m = 15,\ k = 5,\ n^* = 500$	0.51380	0.9724

The p-values shown under cases 2.3b, 2.4a, 2.4b are sizably larger than the p-value shown under case 2.3a and all four p-values are much larger than 0.05. Our earlier sentiments supported by visual examinations of Figures 2.3a-2.3b and Figures 2.4a-2.4b are clearly validated by K-S test of normality under each scenario. That is, we are reasonably assured of a good fit between the observed values of v and a normal curve with a high level of confidence for all practical purposes.

2.8. DATA ANALYSIS: ILLUSTRATIONS USING REAL DATA

The modified two-stage estimation methodology (2.4.1)-(2.4.2) and the purely sequential estimation methodology (2.5.1) will now be illustrated using two real datasets from health studies.

The first illustration (Section 2.8.1) uses infant mortality data from Leinhardt and Wasserman (1979). R documentation for this data may be seen from the website:

$$\boxed{\textit{https://vincentarelbundock.github.io/Rdatasets/doc/car/Leinhardt.html.}} \tag{2.8.1}$$

Our second illustration (Section 2.8.2) uses bone marrow transplant data from the text:

$\boxed{\textit{Survival Analysis,}}$

by Klein and Moeschberger (2003). The data came from a multicenter clinical trial with patients prepared for transplantation with a radiation-free conditioning regimen that consisted of allogeneic marrow transplants for patients with AML.

2.8.1. Infant Mortality Data

The data on infant mortality rates were published in the *New York Times* in 1975. This data, named after Leinhardt, were adapted in Leinhardt and Wasserman (1979). Equation (2.8.1) shows a link for R documentation and a description of data pertaining to 105 nations around 1970.

Since infant mortality rates for Iran, Nepal, Laos and Haiti were not available, we omitted them from our illustration. The variable of interest (X) was the infant mortality rate per 1000 live births. The data consisted of $101(= 105 - 4)$ rows and we checked that the negative exponential model (2.1.1) fitted well.

Treating this dataset as the universe, we first found $\hat{\mu} = 9.6$ and $\hat{\sigma} = 80.24$ from full data. Next, we implemented both modified two-stage and purely sequential estimation procedures drawing observations from the full set of data as needed. It is emphasized, however, that implementation of sampling strategies did not exploit the numbers $\hat{\mu} = 9.6$ and $\hat{\sigma} = 80.24$.

We carried out a single run under both procedures for estimating μ. Tables 2.10-2.11 provide the results from implementing the stopping rules from (2.4.1)-(2.4.2) and (2.5.1) respectively corresponding to certain fixed values of c, k, a and the preset risk-bound ω, chosen arbitrarily. We also fixed m or m_0 as needed. These numbers are shown in the tables.

Table 2.10 summarizes the results from the modified two-stage procedure. We assumed a lower bound, $\sigma_L = 40$, for the otherwise unknown scale parameter σ. Table 2.11 summarizes the results from the purely sequential procedure.

Under both methodologies, we notice that the terminal estimated values of μ are not too far away from corresponding $\hat{\mu} = 9.6$ that was obtained from full data. One other comment is in order: The numbers shown in the first column under n^* are computed using (2.2.6) after replacing σ with the number $\hat{\sigma} = 80.24$ obtained from full data. Again, in running the estimation methodologies, we did not exploit the number $\hat{\sigma} = 80.24$.

We have provided n^* values just so that one is able to gauge whether the observed n-values look reasonable. The ratio n/n^* appears reasonably close to 1 which should be desired. The value z, that is, the ratio of achieved risk per unit cost and preset goal ω, appears reasonably under (or close) to 1.

2.8.2. Bone Marrow Data

We looked at the data from a multicenter clinical trial involving 137 bone marrow patients prepared for transplant with a radiation-free conditioning regimen that consisted of allogeneic marrow transplants for patients with AML. The dataset is explained in Section 1.3 (chapter 1), and enumerated in Table D.1 (appendix D), of Klein and Moeschberger's (2003) textbook. For illustrative purposes, we were interested in estimating the minimum waiting time (in days) to death (or time on study time) for these patients. We got rid of two lowest waiting times (1 or 2 days) since they were suspect possible outliers. Other aspects of implementation of our modified two-stage and purely sequential estimation procedures and analysis remained similar to what we had explained in Section 2.8.1.

We checked that the negative exponential model (2.1.1) fitted well to this data. Treating this dataset as our universe, we first found $\hat{\mu} = 10.0$ days and $\hat{\sigma} = 841.57$ days. Next, we implemented both modified two-stage and purely sequential estimation procedures on full data. It is emphasized, however, that implementation of the sampling strategies did not exploit the numbers $\hat{\mu}$ or $\hat{\sigma}$.

We carried out a single run under both procedures for estimating μ. Tables 2.12-2.13 provide the results of implementing the stopping rules from (2.4.1)-(2.4.2) and (2.5.1) respectively corresponding to certain fixed values of c, k, a and the preset risk-bound ω chosen arbitrarily. We also fixed m or m_0 as needed. These numbers are shown in the tables.

Table 2.12 summarizes the results from the modified two-stage procedure. We assumed a lower bound, $\sigma_L = 500$, for the otherwise unknown scale parameter σ. Table 2.13 summarizes the results for the purely sequential procedure.

Here again, we have provided n^* values just so that one is able to gauge whether the observed n-values look reasonable. The ratio n/n^* appears reasonably close to 1 which should be desired. The value z, that is, the ratio of achieved risk per unit cost and preset goal ω, appears reasonably under (or close) to 1.

2.9. A BRIEF SUMMARY OF CHAPTER 2

A negative exponential distribution is widely used to model survival/failure times of electrical equipment and is also used widely as a model in health industries. In this chapter, we have developed methodologies that are operationally convenient and possess interesting efficiency and consistency properties. Our direction of research came from a thorough literature review and we proposed purely sequential, two-stage and modified two-stage methods to estimate the location parameter of a negative exponential distribution.

In comparison with the usual Stein-type two-stage methodology, we came up with a modified two-stage strategy which was shown to perform better than the former. For each of these cases, we validated our theoretical results with simulations and implemented them on real data-sets from health studies.

A quick glance at comparing the entries within Tables 2.10-2.11 or within Tables 2.12-2.13 reveals that the purely sequential estimation strategy appears to have an edge over the modified two-stage estimation strategy. At another level, however, one should realize that a modified two-stage strategy is logistically simpler to implement than a purely sequential estimation strategy. Indeed, both procedures are fully expected to perform very well under comparable experimental circumstances. A practitioner, however, may consider employing one of the two procedures (2.4.1)-(2.4.2) or (2.5.1) that will provide an acceptable level of comfort in running an experiment as one balances it with additional factors deemed locally important, namely, feasibility, efficiency, accuracy, operational convenience, and cost.

Table 2.1. Selected values of $\nu \equiv \nu_k$ and $\eta \equiv \eta_k$ from (2.5.5), $k = 1(1)6$

	k					
	1	2	3	4	5	6
ν_k	1.646	0.963	0.745	0.638	0.577	0.542
η_k	−0.673	−0.913	−1.254	−1.703	−2.259	−2.915

Table 2.2. Simulation results from 10,000 replications for Stein-type two-stage procedure (2.3.1) with $\mu = 5$, $\sigma = 10$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (s_{\overline{z}})$
			$\mathbf{k = 1, m = 3}$		
30	3.7×10^{-3}	5.3720, (0.0040)	28.84, (0.0704)	0.9615	1.9767, (0.0454)
100	1×10^{-4}	5.1120, (0.0012)	94.95, (0.2319)	0.9495	1.9568, (0.0341)
300	3.7×10^{-6}	5.0385, (0.0004)	284.88, (0.7064)	0.9496	1.9635, (0.0320)
500	8×10^{-7}	5.0227, (0.0002)	473.89, (1.1778)	0.9477	1.9872, (0.0367)
			$\mathbf{k = 2, m = 4}$		
30	3.7×10^{-2}	5.3978, (0.0050)	29.51, (0.1130)	0.9837	4.2397, (0.2016)
100	1×10^{-3}	5.1227, (0.0015)	97.74, (0.3790)	0.9774	3.9148, (0.1488)
300	3.7×10^{-5}	5.0413, (0.0004)	289.75, (1.1285)	0.9659	4.5254, (0.5000)
500	8×10^{-6}	5.0249, (0.0003)	482.53, (1.8796)	0.9650	3.9985, (0.1588)
			$\mathbf{k = 3, m = 5}$		
30	3.7×10^{-1}	5.4293, (0.0057)	30.38, (0.1497)	1.0127	7.0348, (0.2574)
100	1×10^{-2}	5.1341, (0.0018)	100.55, (0.5064)	1.0055	9.3451, (0.9627)
300	3.7×10^{-4}	5.0440, (0.0006)	302.75, (1.5099)	1.0091	9.7991, (1.0965)
500	8×10^{-5}	5.0265, (0.0003)	501.48, (2.4959)	1.0029	15.2665, (6.1989)

Table 2.2 contd.. Simulation results from 10,000 replications for Stein-type two-stage procedure (2.3.1) with $\mu = 5$, $\sigma = 10$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (s_{\overline{z}})$
			$\mathbf{k = 4, m = 6}$		
30	3.7	5.4622, (0.0064)	31.53, (0.1874)	1.0510	8.8788, (0.2425)
100	0.1	5.3720, (0.0040)	105.02, (0.6233)	1.0502	16.2954, (1.5576)
300	3.7×10^{-3}	5.0480, (0.0008)	311.15, (1.8387)	1.0371	21.7228, (4.6001)
500	8×10^{-4}	5.0282, (0.0004)	521.66, (3.1057)	1.0433	14.8328(1.4949)
			$\mathbf{k = 5, m = 7}$		
30	37.03	5.4663, (0.0065)	32.64, (0.2190)	1.0880	9.3771, (0.2085)
100	1	5.1456, (0.0023)	110.94, (0.7588)	1.1094	21.6822, (1.5331)
300	3.7×10^{-2}	5.0510, (0.0008)	324.97, (2.2472)	1.0832	41.4347, (9.9591)
500	8×10^{-3}	5.0306, (0.0005)	555.25, (3.8969)	1.1105	30.0233, (4.5692)

Table 2.3. Simulation results from 10,000 replications for the modified two-stage procedure (2.4.1)-(2.4.2) with $\mu = 5$, $\sigma = 10$, $\sigma_L = 3$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (s_{\overline{z}})$
			$\mathbf{k = 1, m_0 = 3}$		
30	3.7×10^{-3}	5.3308, (0.0032)	30.32, (0.0225)	1.0109	1.0364, (0.0024)
100	1×10^{-4}	5.0992, (0.0009)	100.33, (0.0414)	1.0033	1.0103, (0.0012)
150	2.96×10^{-5}	5.0657, (0.0006)	150.35, (0.0499)	1.0023	1.0062, (0.0010)
300	3.7×10^{-6}	5.0333, (0.0003)	300.31, (0.0707)	1.0010	1.0035, (0.0007)
400	1.56×10^{-6}	5.0252, (0.0002)	400.47, (0.0813)	1.0011	1.0014, (0.0006)
500	8×10^{-7}	5.0199, (0.0001)	500.20, (0.0922)	1.0004	1.0028, (0.0005)
			$\mathbf{k = 2, m_0 = 4}$		
30	3.7×10^{-2}	5.3410, (0.0035)	30.35, (0.0554)	1.0119	1.2423, (0.0082)
100	1×10^{-3}	5.1003, (0.0010)	100.22, (0.0995)	1.0022	1.0661, (0.0033)
150	2×10^{-4}	5.0673, (0.0006)	150.25, (0.1203)	1.0016	1.0416, (0.0025)
300	3.7×10^{-5}	5.0330, (0.0003)	300.26, (0.1731)	1.0008	1.0210, (0.0017)
400	1.56×10^{-5}	5.0248, (0.0002)	400.17, (0.1978)	1.0004	1.0160, (0.0015)
500	8×10^{-6}	5.0198, (0.0002)	500.69, (0.2250)	1.0013	1.0100, (0.0013)
			$\mathbf{k = 3, m_0 = 5}$		
30	3.7×10^{-1}	5.3713, (0.0041)	30.65, (0.1069)	1.0217	1.3024, (0.0387)
100	1×10^{-2}	5.1021, (0.0010)	100.27, (0.1862)	1.0027	1.1249, (0.0081)
150	2×10^{-3}	5.0672, (0.0006)	150.69, (0.2258)	1.0046	1.1048, (0.0055)
300	3.7×10^{-4}	5.0336, (0.0003)	300.54, (0.3150)	1.0018	1.0825, (0.0035)
400	1.56×10^{-4}	5.0247, (0.0002)	400.47, (0.3628)	1.0011	1.0274, (0.0029)
500	8×10^{-5}	5.0204, (0.0002)	500.37, (0.4111)	1.0007	1.0087, (0.0018)

Table 2.3 contd.. Simulation results from 10,000 replications for the modified two-stage procedure (2.4.1)-(2.4.2) with $\mu = 5$, $\sigma = 10$, $\sigma_L = 3$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (\mathrm{s}_{\bar{x}_{\min}})$	$\bar{n}, (\mathrm{s}_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (\mathrm{s}_{\overline{z}})$
			$\mathbf{k = 4, m_0 = 6}$		
30	3.7	5.5033, (0.0050)	31.63, (0.1684)	1.0546	1.6420, (0.0908)
100	0.1	5.1084, (0.0010)	101.97, (0.3037)	1.0197	1.1544, (0.0208)
150	0.02	5.0693, (0.0007)	151.53, (0.3659)	1.0102	1.1124, (0.0128)
300	3.7×10^{-3}	5.0336, (0.0003)	302.45, (0.5159)	1.0081	1.0987, (0.0066)
400	1.56×10^{-3}	5.0247, (0.0002)	400.40, (0.5934)	1.0010	1.0487, (0.0055)
500	8×10^{-4}	5.0200, (0.0002)	500.16, (0.6661)	1.0003	1.0215, (0.0048)
			$\mathbf{k = 5, m_0 = 7}$		
30	37.03	5.4653, (0.0064)	33.22, (0.2242)	1.1075	1.7487, (0.0758)
100	1	5.1201, (0.0015)	104.29, (0.4803)	1.0429	1.3241, (0.0597)
150	0.2	5.0334, (0.0008)	155.43, (0.5801)	1.0362	1.2015, (0.0514)
300	0.037	5.0358, (0.0003)	305.21, (0.8044)	1.0173	1.1198, (0.0425)
400	0.015	5.0260, (0.0002)	405.39, (0.9324)	1.0134	1.0489, (0.0398)
500	0.008	5.0204, (0.0002)	507.11, (1.0365)	1.0142	1.0341, (0.0304)

Table 2.4. Simulation results from 10,000 replications for the modified two-stage procedure (2.4.1)-(2.4.2) with $\mu = 5$, $\sigma = 10$, $\sigma_L = 5$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (\mathrm{s}_{\bar{x}_{\min}})$	$\bar{n}, (\mathrm{s}_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (\mathrm{s}_{\overline{z}})$
			$\mathbf{k = 1, m_0 = 3}$		
30	3.7×10^{-3}	5.3290, (0.0033)	30.34, (0.0210)	1.0116	1.0292, (0.0022)
100	1×10^{-4}	5.1008, (0.0010)	100.36, (0.0373)	1.0036	1.0073, (0.0011)
300	3.7×10^{-6}	5.0336, (0.0003)	300.45, (0.0656)	1.0015	1.0016, (0.0006)
500	8×10^{-7}	5.0198, (0.0001)	500.35, (0.0824)	1.0007	1.0015, (0.0004)
			$\mathbf{k = 2, m_0 = 4}$		
30	3.7×10^{-2}	5.3312, (0.0033)	30.32, (0.0469)	1.0109	1.1678, (0.0061)
100	1×10^{-3}	5.1010, (0.0010)	100.33, (0.0841)	1.0033	1.0445, (0.0027)
300	3.7×10^{-5}	5.0340, (0.0003)	300.46, (0.1468	1.0015	1.0132, (0.0015)
500	8×10^{-6}	5.0199, (0.0002)	500.10, (0.1886)	1.0002	1.0099, (0.0011)
			$\mathbf{k = 3, m_0 = 5}$		
30	3.7×10^{-1}	5.3546, (0.0037)	30.51, (0.0805)	1.0173	1.5370, (0.0148)
100	1×10^{-2}	5.1009, (0.0010)	100.54, (0.1421)	1.0054	1.1252, (0.0052)
300	3.7×10^{-4}	5.0340, (0.0003)	300.37, (0.2447)	1.0012	1.0405, (0.0026)
500	8×10^{-5}	5.0195, (0.0001)	500.86, (0.3158)	1.0017	1.0210, (0.0019)

Table 2.4 contd. Simulation results from $10,000$ replications for the modified two-stage procedure (2.4.1)-(2.4.2) with $\mu = 5$, $\sigma = 10$, $\sigma_L = 5$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (s_{\overline{z}})$
			$\mathbf{k = 4, m_0 = 6}$		
30	3.7	5.3681, (0.0041)	31.20, (0.1222)	1.0401	1.5429, (0.0343)
100	0.1	5.1019, (0.0010)	101.35, (0.2167)	1.0135	1.2921, (0.0099)
300	3.7×10^{-3}	5.0338, (0.0003)	301.47, (0.3687)	1.0049	1.0837, (0.0042)
500	8×10^{-4}	5.0201, (0.0002)	501.40, (0.4701)	1.0022	1.0496, (0.0030)
			$\mathbf{k = 5, m_0 = 7}$		
30	37.03	5.4061, (0.0050)	32.52, (0.1771)	1.0841	1.7654, (0.0710)
100	1	5.1074, (0.0011)	102.38, (0.3005)	1.0238	1.5306, (0.0202)
300	3.7×10^{-2}	5.0341, (0.0003)	302.29, (0.5197)	1.0076	1.1724, (0.0065)
500	8×10^{-3}	5.0200, (0.0002)	501.74, (0.6724)	1.0034	1.1062, (0.0047)

Table 2.5. Simulation results from 10,000 replications for the modified two-stage procedure (2.4.1)-(2.4.2) with $\mu = 5$, $\sigma = 10$, $\sigma_L = 11$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (s_{\overline{z}})$
			$\mathbf{k = 1, m_0 = 3}$		
30	3.7×10^{-3}	5.3199, (0.0031)	31.43, (0.0085)	1.0478	0.9014, (0.0015)
100	1×10^{-4}	5.0976, (0.0009)	104.22, (0.0076)	1.0422	0.8921, (0.0011)
300	3.7×10^{-6}	5.0320, (0.0031)	310.10, (0.0067)	1.0336	0.9083, (0.0005)
500	8×10^{-7}	5.0192, (0.0001)	517.03, (0.0040)	1.0340	0.9061, (0.0004)
			$\mathbf{k = 2, m_0 = 4}$		
30	3.7×10^{-2}	5.3078, (0.0030)	32.78, (0.0162)	1.0926	0.8005, (0.0009)
100	1×10^{-3}	5.0935, (0.0009)	107.52, (0.0169)	1.0752	0.8130, (0.0005)
300	3.7×10^{-5}	5.0311, (0.0003)	320.20, (0.0127)	1.0673	0.8250, (0.0002)
500	8×10^{-6}	5.0190, (0.0001)	533.07, (0.0083)	1.0664	0.8267, (0.0001)
			$\mathbf{k = 3, m_0 = 5}$		
30	3.7×10^{-1}	5.2935, (0.0029)	34.11, (0.0249)	1.1372	0.7179, (0.0011)
100	1×10^{-2}	5.0903, (0.0009)	110.90, (0.0278)	1.1090	0.7421, (0.0008)
300	3.7×10^{-4}	5.0301, (0.0002)	330.26, (0.0180)	1.1008	0.7519, (0.0004)
500	8×10^{-5}	5.0179, (0.0001)	550.11, (0.0133)	1.1002	0.7522, (0.0001)

Table 2.5 contd. Simulation results from $10,000$ replications for the modified two-stage procedure (2.4.1)-(2.4.2) with $\mu = 5$, $\sigma = 10$, $\sigma_L = 11$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\overline{n}/n^*$	$\overline{z}, (s_{\overline{z}})$
			$\mathbf{k = 4, m_0 = 6}$		
30	3.7	5.2792, (0.0028)	36.21, (0.0305)	1.2070	0.6026, (0.0010)
100	0.1	5.0876, (0.0008)	115.09, (0.0367)	1.1509	0.6648, (0.0007)
300	3.7×10^{-3}	5.0291, (0.0002)	341.34, (0.0240)	1.1378	0.6810, (0.0005)
500	8×10^{-4}	5.0175, (0.0001)	568.08, (0.0120)	1.1361	0.6830, (0.0001)
			$\mathbf{k = 5, m_0 = 7}$		
30	37.03	5.2649, (0.0026)	37.70, (0.0411)	1.2566	0.5426, (0.0014)
100	1	5.0829, (0.0008)	119.24, (0.0430)	1.1924	0.5985, (0.0008)
300	3.7×10^{-2}	5.0288, (0.0002)	352.42, (0.0300)	1.1747	0.6188, (0.0004)
500	8×10^{-3}	5.0168, (0.0001)	587.12, (0.0166)	1.1742	0.6187, (0.0002)

Table 2.6. Values of $\bar{n} - n^*$, ψ and $\psi + 1$ for each k used in Table 2.3

$n^*\backslash k$	1	2	3	4	5
20	0.34	0.24	0.54	1.42	**2.52**
30	0.32	0.35	0.65	1.63	**3.22**
50	0.36	0.31	0.28	1.85	4.77
100	0.33	0.22	0.27	1.97	4.29
150	0.35	0.25	0.69	1.53	5.06
300	0.31	0.26	0.22	**2.45**	**5.43**
400	0.47	0.17	0.47	**0.40**	**5.21**
500	0.20	0.69	0.37	**0.16**	**5.39**
ψ	−0.1659	−0.2479	0	1.1065	4.1323
$\psi + 1$	0.8341	0.7521	1	2.1065	5.1323

Table 2.7. Values of $\bar{n} - n^*$, ψ and $\psi + 1$ for each k used in Table 2.4

$n^*\backslash k$	1	2	3	4	5
30	0.34	0.32	0.51	1.20	2.52
100	0.36	0.33	0.54	1.35	2.38
300	0.45	0.46	0.37	1.47	2.29
500	0.35	0.10	0.86	1.10	**1.74**
ψ	−0.1399	−0.1763	0	0.5599	1.7637
$\psi + 1$	0.8601	0.8237	1	1.5599	2.7637

Table 2.8. Values of $\bar{n} - n^*$, ψ and $\psi + 1$ for each k used in Table 2.5

$n^*\backslash k$	1	2	3	4	5
30	**1.43**	**2.78**	**4.11**	**6.21**	**7.70**
100	**4.22**	**7.52**	**10.90**	**15.09**	**19.24**
300	**10.10**	**20.20**	**30.26**	**41.34**	**52.42**
500	**17.03**	**33.07**	**50.11**	**68.08**	**87.12**
ψ	-0.1076	-0.1042	0	0.1957	0.4739
$\psi + 1$	0.8924	0.8958	1	1.1957	1.4739

Table 2.9. Simulation results from 10,000 replications of the purely sequential procedure (2.5.1) with $\mu = 5$, $\sigma = 10$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\bar{n}/n^*$	$\overline{z}, (s_{\overline{z}})$	ε in (7.2)
			$\mathbf{k = 1, m = 5}$			
30	3.70×10^{-3}	5.3282, (0.0032)	30.31, (0.0188)	1.0106	1.0265, (0.0243)	0.0228
100	1.00×10^{-4}	5.0996, (0.0009)	100.29, (0.0334)	1.0029	1.0079, (0.0238)	0.0068
150	2.96×10^{-5}	5.0662, (0.0006)	150.38, (0.0410)	1.0025	1.0044, (0.0226)	0.0045
300	3.70×10^{-6}	5.0334, (0.0003)	300.22, (0.0582)	1.0007	1.0033, (0.0216)	0.0022
400	1.56×10^{-6}	5.0251, (0.0002)	400.34, (0.0670)	1.0008	1.0015, (0.0240)	0.0017
500	8.00×10^{-7}	5.0199, (0.0002)	500.42, (0.0743)	1.0008	1.0007, (0.0228)	0.0013
			$\mathbf{k = 2, m = 8}$			
30	3.70×10^{-2}	5.3443, (0.0034)	29.99, (0.0386)	0.9997	1.1714, (0.0323)	0.1135
100	1.00×10^{-3}	5.0993, (0.0009)	100.03, (0.0669)	1.0003	1.0377, (0.0227)	0.0340
150	2.00×10^{-4}	5.0671, (0.0006)	149.90, (0.0842)	0.9993	1.0284, (0.0224)	0.0227
300	3.70×10^{-5}	5.0332, (0.0003)	300.03, (0.1159)	1.0001	1.0120, (0.0235)	0.0113
400	1.56×10^{-5}	5.0249, (0.0002)	399.96, (0.1344)	0.9999	1.0096, (0.0220)	0.0085
500	8.00×10^{-6}	5.0201, (0.0002)	500.31, (0.1483)	1.0006	1.0054, (0.0212)	0.0068
			$\mathbf{k = 3, m = 9}$			
30	3.70×10^{-1}	5.3574, (0.0037)	29.46, (0.0585)	0.9822	1.2464, (0.0478)	0.2588
100	1.00×10^{-2}	5.1014, (0.0010)	99.78, (0.1018)	0.9941	1.0988, (0.0515)	0.0776
150	2.00×10^{-3}	5.0663, (0.0006)	149.81, (0.1240)	0.9987	1.0548, (0.0400)	0.0517
300	3.70×10^{-4}	5.0338, (0.0003)	299.75, (0.1745)	0.9991	1.0270, (0.0489)	0.0258
400	1.56×10^{-4}	5.0249, (0.0002)	399.82, (0.2013)	0.9995	1.0195, (0.0375)	0.0194
500	8.00×10^{-5}	5.0203, (0.0002)	499.85, (0.2252)	0.9997	1.0153, (0.0236)	0.0155

Table 2.9 contd. Simulation results from $10,000$ replications of the purely sequential procedure (2.5.1) with $\mu = 5$, $\sigma = 10$, $a = 1$, $c = 0.1$

n^*	ω	$\bar{x}_{\min}, (s_{\bar{x}_{\min}})$	$\bar{n}, (s_{\bar{n}})$	$\bar{n}/n^*$	$\overline{z}, (s_{\overline{z}})$	ε in (2.7.2)
			$\mathbf{k = 4, m = 12}$			
30	3.70	5.3823, (0.0042)	29.10, (0.0782)	0.9700	1.4793, (0.1318)	0.4592
100	0.10	5.1018, (0.0010)	99.46, (0.1357)	0.9946	1.1498, (0.0283)	0.1377
150	0.02	5.0665, (0.0006)	149.13, (0.1649)	0.9942	1.1108, (0.0253)	0.0918
300	3.70×10^{-3}	5.0339, (0.0003)	299.19, (0.2305)	0.9973	1.0496, (0.0236)	0.0459
400	1.56×10^{-3}	5.0252, (0.0002)	399.10, (0.2681)	0.9977	1.0378, (0.0241)	0.0344
500	8.00×10^{-4}	5.0201, (0.0002)	499.68, (0.2984)	0.9993	1.0298, (0.0234)	0.0275
			$\mathbf{k = 5, m = 15}$			
30	37.03	5.3872, (0.0042)	28.78, (0.0890)	0.9596	1.7024, (0.2270)	0.7148
100	1.00	5.1061, (0.0011)	98.38, (0.1738)	0.9838	1.2409, (0.3578)	0.2144
150	0.20	5.0687, (0.0007)	148.77, (0.2078)	0.9918	1.1531, (0.0529)	0.1429
300	0.037	5.0336, (0.0003)	298.65, (0.2936)	0.9955	1.0815, (0.0422)	0.0714
400	0.015	5.0256, (0.0002)	398.38, (0.3362)	0.9959	1.0613, (0.0410)	0.0536
500	0.008	5.0204, (0.0002)	499.83, (0.3733)	0.9996	1.0442, (0.0394)	0.0428

Table 2.10. Analysis of infant mortality rate data using modified two-stage procedure (2.4.1)-(2.4.2) with $a = 1$, $c = 0.1$, $\sigma_L = 40$

n^*	m_0	k	ω	$\hat{\mu}$: $x_{n:1}$	n	n/n^*	z
40	4	1	0.01253	9.8	41	1.025	0.9518
50	4		0.00642	9.6	50	1.000	1.0204
50	5	2	0.51507	10.1	52	1.040	0.9064
60	5		0.29807	9.6	57	0.950	1.2523

Table 2.11. Analysis of infant mortality rate data using purely sequential procedure (2.5.1) with $a = 1$, $c = 0.1$

n^*	m	k	ω	$\hat{\mu}$: $x_{n:1}$	n	n/n^*	z	ε in (2.7.2)
40	5	1	0.01253	9.6	41	1.025	1.0245	0.0171
50	5		0.00641	9.7	48	0.960	1.0188	0.0137
50	7	2	0.51507	10.1	54	1.080	1.1073	0.0681
60	7		0.29807	9.8	63	1.050	1.1547	0.0567

Table 2.12. Analysis of time to death (or time on study) in bone marrow data using modified two-stage procedure (2.4.1)-(2.4.2) with $a = 1$, $\sigma_L = 500$

n^*	m_0	c	k	ω	$\hat{\mu}$: $x_{n:1}$	n	n/n^*	z
60	4	0.1	1	0.03896	14	57	0.950	1.1871
70	4			0.02453	10	75	1.071	0.9240
80	5	0.4	2	3.45820	15	81	1.013	0.9754
90	5			2.42880	10	89	0.989	1.0458

Table 2.13. Analysis of time to death (or time on study) in bone marrow data using purely sequential procedure (2.5.1) with $a = 1$

n^*	m	c	k	ω	$\hat{\mu}$: $x_{n:1}$	n	n/n^*	z	ε in (2.7.2)
60	4	0.1	1	0.03896	16	61	1.017	1.0256	0.0114
70	4			0.02453	10	70	1.000	1.0187	0.0098
80	5	0.4	2	3.45820	11	77	0.963	1.1362	0.0425
90	5			2.42880	10	88	0.978	1.0820	0.0378

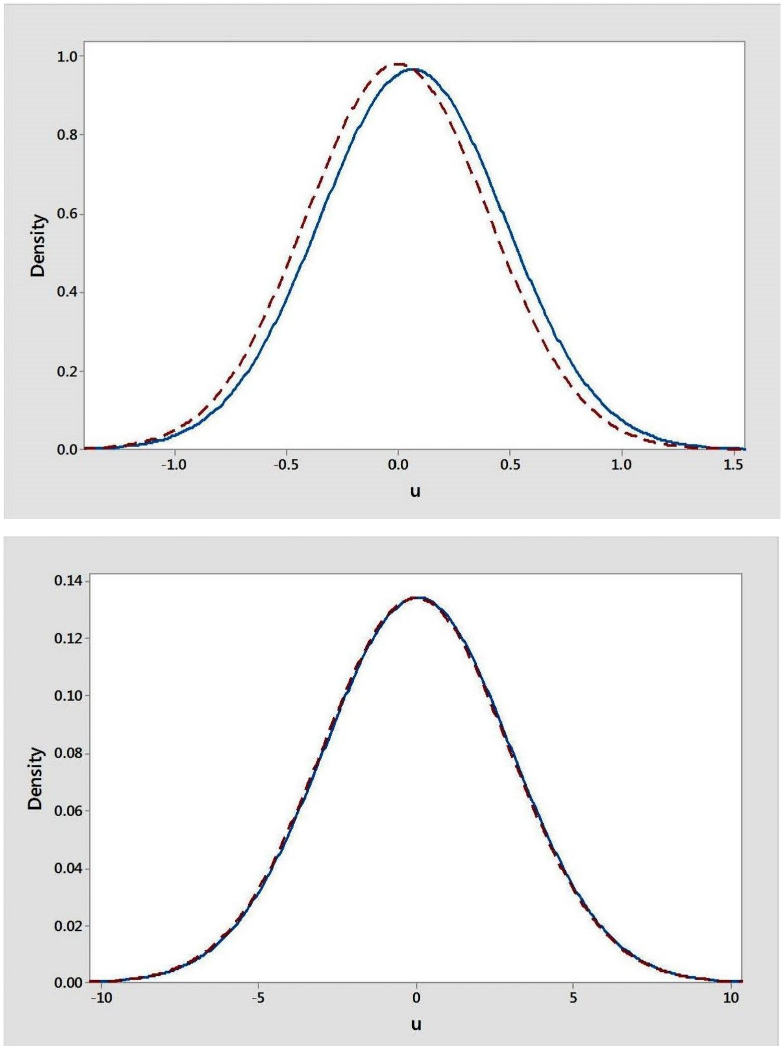


Figure 2.1. Plots of normality curves for the modified two-stage procedure (2.4.1)-(2.4.2) as validation for (2.4.6). The dashed curve (red) and the solid curve (blue) respectively correspond to the empirical distribution of the standardized sample size and the $N(0, \sigma_0^2)$ distribution with $\sigma_0^2 = \frac{1}{9}(\frac{\sigma}{\sigma_L})^{k/3}k^2$ coming from (2.4.4): **(a)** $\sigma_L = 3$, $k = 1$, $n^* = 30$; **(b)** $\sigma_L = 3$, $k = 4$, $n^* = 500$

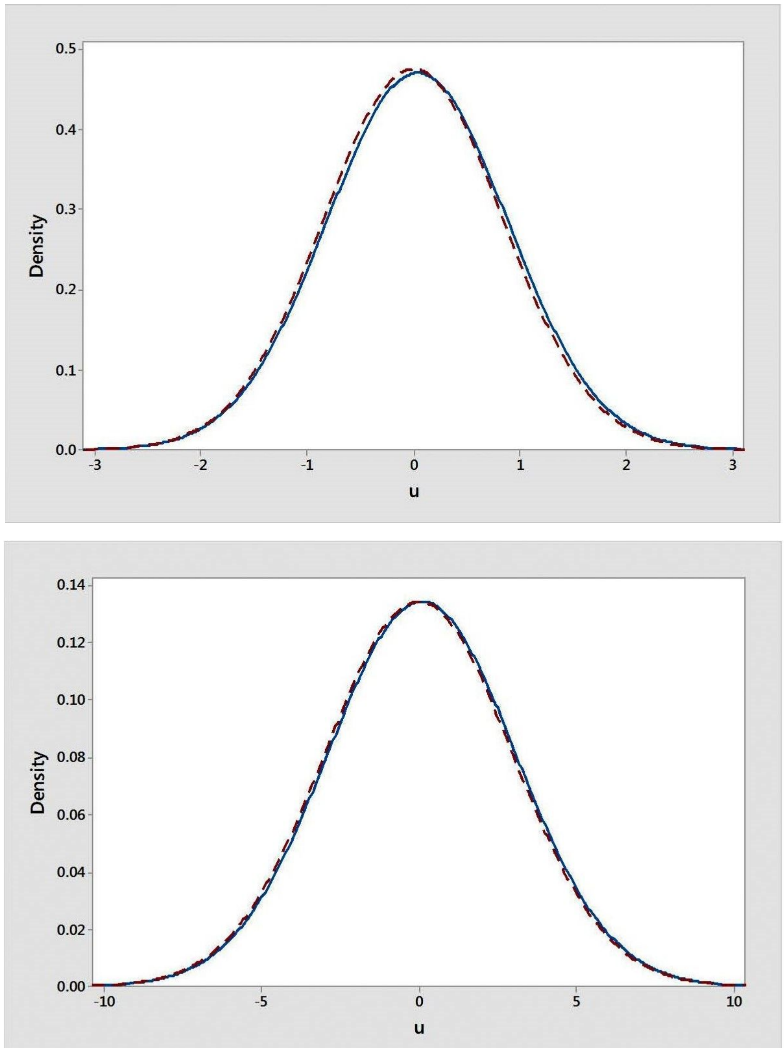
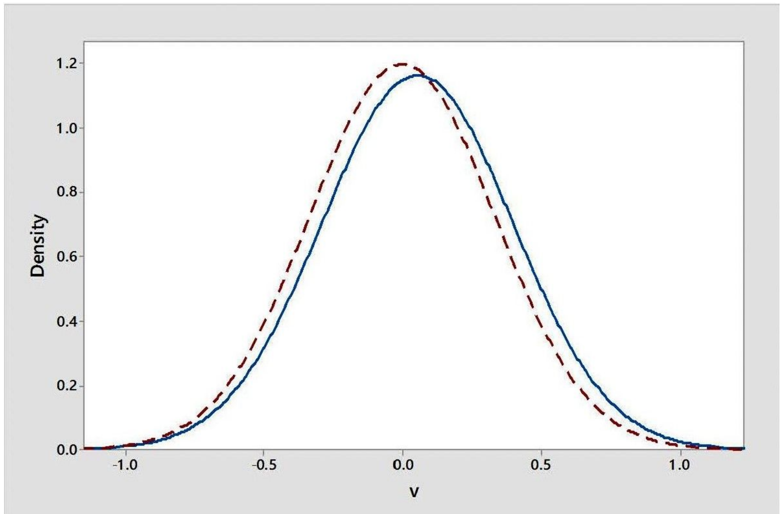


Figure 2.2. Plots of normality curves for the modified two-stage procedure (2.4.1)-(2.4.2) as validation for (2.4.6). The dashed curve (red) and the solid curve (blue) respectively correspond to the empirical distribution of the standardized sample size and the $N(0, \sigma_0^2)$ distribution with $\sigma_0^2 = \frac{1}{9}(\frac{\sigma}{\sigma_L})^{k/3}k^2$ coming from (2.4.4): **(a)** $\sigma_L = 5$, $k = 2$, $n^* = 100$; **(b)** $\sigma_L = 5$, $k = 5$, $n^* = 500$



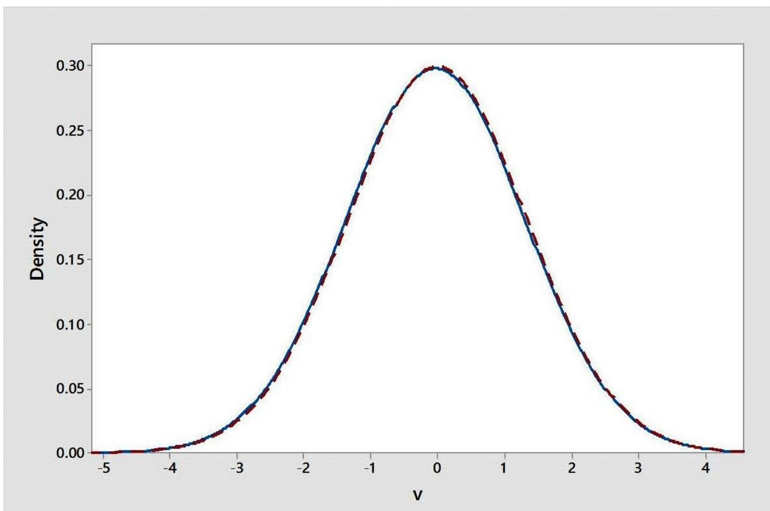


Figure 2.3. Plots of normality curves for the purely sequential procedure (2.5.1) as validation of Theorem 2.5.1, part (v). The dashed curve (red) and the solid curve (blue) respectively correspond to the empirical distribution of the standardized sample size and the $N(0, \sigma_1^2)$ distribution with $\sigma_1^2 = \frac{1}{9}k^2$ coming from Theorem 2.5.1, part (v): **(a)** $k = 1,\ m = 5,\ n^* = 30$; **(b)** $k = 4,\ m = 12,\ n^* = 500$

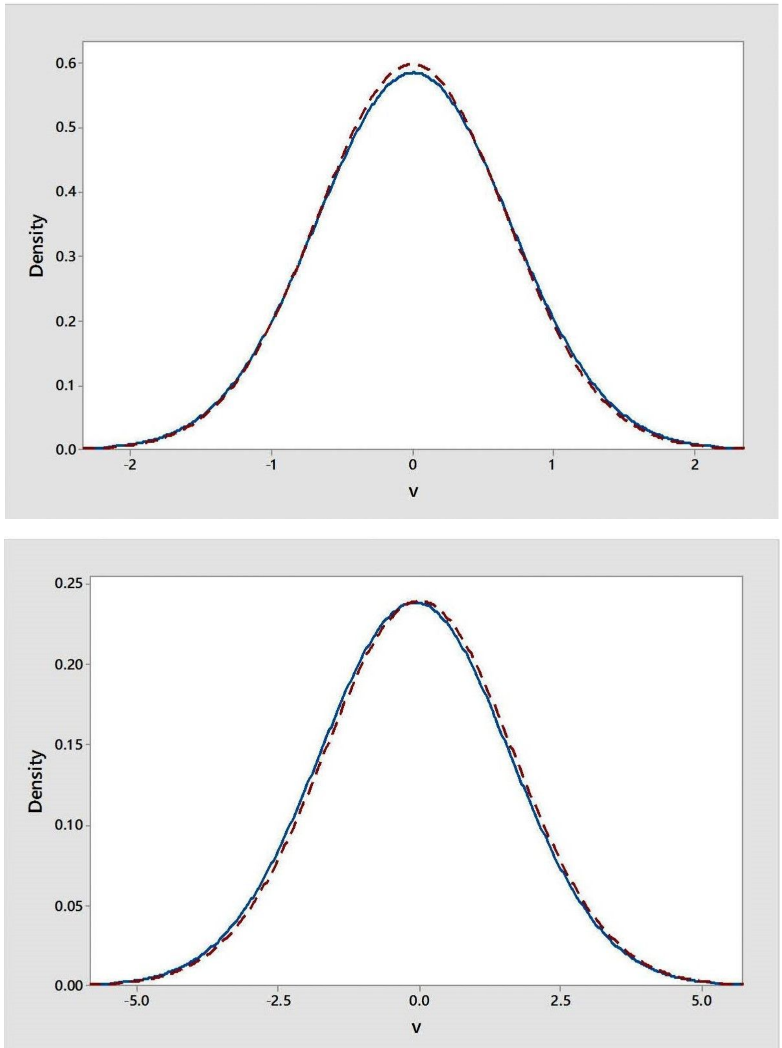


Figure 2.4. Plots of normality curves for the purely sequential procedure (2.5.1) as validation of Theorem 2.5.1, part (v). The dashed curve (red) and the solid curve (blue) respectively correspond to the empirical distribution of the standardized sample size and the $N(0, \sigma_1^2)$ distribution with $\sigma_1^2 = \frac{1}{9}k^2$ coming from Theorem 2.5.1, part (v): **(a)** $k = 2$, $m = 8$, $n^* = 100$; **(b)** $k = 5$, $m = 15$, $n^* = 500$

Chapter 3
Multistage Estimation of the Difference of Locations of Two Negative Exponential Populations Under a Modified Linex Loss Function: Real Data Illustrations from Cancer Studies and Reliability Analysis

3.1. INTRODUCTION

In this chapter we develop sequential and two-stage procedures to estimate the difference of two independent negative exponential locations under a modified Linex loss function. The material of this chapter is based on Mukhopadhyay and Bapat (2016b). There is a wide literature devoted to normal populations, largely under a different form of the loss function.

Point estimation of a negative exponential location parameter under a variant of the Linex loss function has been introduced in Chapter 2. The present investigation expands those ideas by developing appropriate methodologies and associated properties in the context of two-sample comparisons.

3.1.1. Two-Sample Scenario and the Goal

Let $X_1, \ldots, X_n, \ldots$ and $Y_1, \ldots, Y_n, \ldots$ be two independent and identically distributed (i.i.d.) sequences of random variables with the probability density functions (p.d.f.s) $f(x; \mu_1, \sigma)$ and $f(y; \mu_2, b\sigma)$ with $b(>0)$ respectively where we denote:

$$f(t; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{t-\mu}{\sigma}\right) I(t > \mu), \tag{3.1.1}$$

where $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ are both *unknown parameters.* $I(.)$ denotes an indicator function of (.) which takes the value 1(or 0) when $t >$ (or $\leq$)μ.

Additionally, we assume that the X's are independent of the Y's and b is *known.* The location parameter μ, if positive, is viewed as the minimum guarantee time or the threshold of the distribution. The parameter σ is called the scale. We record a pair (X_i, Y_i) at-a-time or a group of pairs of observations (X_i, Y_i) at-a-time as needed, $i = 1, ..., n, ...$. Our goal is to estimate the "difference" parameter, namely:

$$\Delta = \mu_1 - \mu_2. \tag{3.1.2}$$

3.1.2. Motivation for a Scenario Where b Is Known A'priori

In Figure 3.1, we show columns from Parthenon, Greece. In Figure 3.1(a), a single column is shown whereas in Figure 3.1(b) a cluster of eight columns is shown. These columns are identically constructed. In large structures (for example, bridges), one will often see a column or a cluster of columns strategically placed to withstand stress. This is done to raise a structure that may not fall apart easily with a high probability under reasonable application of stress levels.

Now, visualize Figure 3.1(a), expose the column to a stress level (L), and observe the time (U) it takes the column to show up cracks. We may reasonably assume that U is governed by the p.d.f. $f(u; \mu_1, \sigma)$ defined by (3.1.1), that is we will not observe a failure before time μ_1.

Next, we may visualize Figure 3.1(b), expose eight pillars to a stress level (L) each, and observe the time (V) it takes the cluster of columns to show fatal cracks. Consider independent $U_1, ..., U_8$ distributed identically as U. Then, the integrity of the cluster in Figure 3.1(b) may be described by $V = \min\{U_1, ..., U_8\} \sim f(v; \mu_2, \frac{1}{8}\sigma)$, the time of failure of the weakest column among all eight columns. We will not observe a failure of the cluster before time μ_2.

Under a stress test conducted in a laboratory, we may want to estimate $\Delta = \mu_1 - \mu_2$ by observing independent pairs of observations $\{(X_i, Y_i); i = 1, ..., n, ...\}$ where we begin by first identifying X with U and Y with V respectively. That is, X_i $(Y_i) = i^{\text{th}}$ observation on U (V), $i = 1, ..., n, ...$ This setup clearly agrees with a two-sample scenario described in Section 3.1.1 involving a known multiplier, $b = \frac{1}{8}$. We emphasize that b needs to be specified a'priori before data collection begins. Often b will be specified by the subject-matter expert(s) based on personal knowledge from previously run similar studies.

In a situation where it may be more reasonable to postulate that $X \sim f(x; \mu_1, \sigma)$ and $Y \sim f(y; g\mu_2, h\sigma)$ with both g, h positive and known, then obviously $Y^* \equiv g^{-1}Y \sim f(y^*; \mu_2, b\sigma)$ with $b = g^{-1}h$ positive and known. By exploiting one-to-one coding from Y to Y^* and then working with X, Y^* instead of X, Y would bring one back to our proposed two-sample estimation of $\Delta = \mu_1 - \mu_2$.

3.1.3. A Brief Review and Layout of the Chapter

Some of the literature handling to a normal distribution under different loss functions include Mukhopadhyay et al. (2010) and Aoshima et al. (2011). Some notable papers pertaining to two-sample estimation problems under negative exponential settings include Mukhopadhyay and Hamdy (1984) and Mukhopadhyay and Darmanto (1988).

Negative exponential distributions often play a central role in reliability, life testing and clinical experiments. For an elaborate review, one may refer to Zelen (1966), Johnson and Kotz (1970), Lawless and Singhal (1980), Balakrishnan and Basu (1995), and other sources. We provide illustrations with real data in Section 7.

Sequential and two-stage procedures for estimating a negative exponential location have been discussed extensively in Basu (1971), Ghurye (1958), Mukhopadhyay (1974,1980,1984,1988,1995), Swanepoel and van Wyk (1982), and other sources. A sequential point estimation analog under a Linex loss function (1.3) was first developed by Chattopadhyay (1998) utilizing nonlinear renewal theory from Woodroofe (1977,1982) and Lai and Siegmund (1977,1979).

Before we go any further, we explain our notation clearly. Since μ_1, μ_2, σ are all unknown, the parameter vector $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma)$ remains unknown. When we write $P(.)$ or $E(.)$, they should be interpreted as $P_{\boldsymbol{\theta}}(.)$ or $E_{\boldsymbol{\theta}}(.)$ respectively. In the same spirit, when we write $\xrightarrow{P}$ (convergence in probability) or w.p.1 (with probability one) or $\xrightarrow{\pounds}$ (convergence in law or distribution), they are all with respect to $P_{\boldsymbol{\theta}}$. We drop subscript $\boldsymbol{\theta}$ in the sequel for simplicity.

In Section 3.2, we include some preliminaries of a modified Linex loss function in the spirit of Chapter 2, formulation of the risk function, and its optimization by bounding it from above. Section 3.3 introduces a modified two-stage procedure in which we assume a *known* lower bound $\sigma_L(> 0)$ for the unknown standard deviation σ along the lines of Mukhopadhyay and Duggan (1997,1999). Both first- and second-order properties are developed for estimating Δ.

Section 3.4 introduces a purely sequential methodology in the present context for estimating Δ. For some of the technicalities and second-order properties, we have relied upon Mukhopadhyay (1974,1984,1988), Lombard and Swanepoel (1978), and Swanepoel and van Wyk (1982). Generally speaking, the broad literature on sequential estimation as well as many associated tools of trade may be reviewed from Mukhopadhyay and Solanky (1994), Ghosh et al. (1997) and Mukhopadhyay and de Silva (2009).

In Section 3.5, we outline proofs for some selected results. Section 3.6 presents extensive data analysis based upon computer simulations for a large variety of parameter configurations highlighting the performance of the proposed methodologies for a wide range of small, moderate, and large sample sizes. Selected conclusions from the theorems studied in Sections 3.3-3.4 are critically examined and validated with data analysis.

Section 3.7 shows analysis on two real datasets, one from cancer studies and the other from reliability analysis, both supporting the proposed methodologies. First illustration (Section 3.7.1) uses parallel datasets related to survival times of cancer patients from Shanker et al. (2016). Second illustration (Section 3.7.2) uses data on lifetimes of steel components from the text of Lawless (1982). Section 3.8 gives brief conclusions.

3.2. MODIFIED LINEX LOSS AND SOME PRELIMINARIES

In this section, we work under an appropriately modified Linex loss function in the spirit of Chapter 2 and then calculate the associated risk function. For motivation, one may refer back to Chapter 2.

3.2.1. Equal Sample Size and the Estimators

Having recorded pairs of random samples $\{X_i, Y_i; i = 1, ..., n\}, n > 1$, we estimate the difference parameter $\Delta \equiv \mu_1 - \mu_2$ from (3.1.2) with the maximum likelihood estimator

$$\widehat{\Delta}_n \equiv X_{n:1} - Y_{n:1},$$

where $X_{n:1} = \min\{X_1, ..., X_n\}$ and $Y_{n:1} = \min\{Y_1, ..., Y_n\}$.

The minimum variance unbiased estimator of σ is given by

$$\hat{\sigma}_n \equiv W_n = \tfrac{1}{2}\left(U_n + b^{-1}V_n\right),$$

where we let

$$U_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{n:1}) \text{ and } V_n = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - Y_{n:1}). \tag{3.2.1}$$

3.2.2. Modified Linex Loss Function

Let us begin with some fixed a, positive or negative. In the spirit of Chapter 2, a *modified Linex loss* function in estimating Δ from (3.1.2) with $\widehat{\Delta}_n \equiv X_{n:1} - Y_{n:1}$ is formulated as follows:

$$L_n = \exp\left(a\left[\sigma^{-1}(\widehat{\Delta}_n - \Delta)\right]\right) - a\left[\sigma^{-1}(\widehat{\Delta}_n - \Delta)\right] - 1. \tag{3.2.2}$$

The corresponding risk function is obtained by taking expectations across (3.2.2), namely,

$$\text{Risk}_n \equiv E[L_n] = \left(1 - \tfrac{a}{n}\right)^{-1}\left(1 + \tfrac{ab}{n}\right)^{-1} - \tfrac{a}{n} + \tfrac{ab}{n} - 1. \tag{3.2.3}$$

Upon expanding (3.2.3), we clearly obtain:

$$\text{Risk}_n = \frac{a^2(1+b^2-b)}{n^2} + o\left(\frac{1}{n^2}\right), \tag{3.2.4}$$

for $n > \max\{1, a, -ab\}$.

3.2.3. Cost Per Unit Sampling

Next, we consider a cost function, $\text{Cost}_n(> 0)$. The exact form of this function must depend upon the underlying practical scenario. It may be acceptable to assume that the cost for each observation should go up (or down) as σ goes down (or up). With this in mind, we propose a cost function of the following form:

$$\text{Cost}_n = cn\sigma^{-k} \text{ with fixed and known } c(> 0), k(> 0). \tag{3.2.5}$$

One may alternatively think that Cost_n should have been $c(2n)\sigma^{-k}$ instead of $cn\sigma^{-k}$ because afterall we record $2n$ observations. Also, the cost per unit observation on X-data could be surely different from the cost per unit observation on Y-data giving rise to an expression such as $(c_1 n +$

$c_2n)\sigma^{-k}$ instead of $cn\sigma^{-k}$. However, Cost_n from (3.2.5) is general enough because in our formulation, the choice of "c" remains generic, covering easily those other possibilities.

3.2.4. Proposed Criterion: Bounded Risk Per Unit Cost

Adapting the basic formulation from Chapter 2, we bound the associated "risk" where we interpret "risk" as the *risk per unit cost* (RPUC), namely,

$$\text{RPUC}_n \equiv \frac{\text{Risk}_n}{\text{Cost}_n} = \frac{a^2(1+b^2-b)}{n^2}\frac{\sigma^k}{cn} + o\left(n^{-3}\right). \tag{3.2.6}$$

We require that $\text{RPUC}_n \leq \omega$ for all parameter vector $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma)$ where $\omega(>0)$ is a fixed constant which leads us to determine the required optimal fixed sample size, had σ been known as follows:

$$n \geq \left(\frac{a^2(1+b^2-b)}{c\omega}\right)^{1/3} \sigma^{k/3} = n^*, \text{ say.} \tag{3.2.7}$$

But, the magnitude of n^* is unknown even though its expression is known. Hence, we proceed to develop modified two-stage and purely sequential bounded risk estimation strategies in Sections 3.3-3.4.

3.2.5. Evaluating the Risk Per Unit Cost When Sample Size Is Random

The idea is to estimate the optimal fixed sample size n^* given by (3.2.7). We begin with pilot data of appropriate size $m(>\max\{1, a, -ab\})$ from both X, Y and then move forward step by step with the help of implementing a modified two-stage or a purely sequential sampling strategy to record more data subsequently as needed beyond the pilot stage.

Suppose that a final sample size, denoted by a random variable Q, is determined by an adaptive multistage sampling strategy. Then, the next result shows an exact analytical expression for the risk per unit cost associated with the terminal estimator $\widehat{\Delta}_Q = X_{Q:1} - Y_{Q:1}$ for Δ once sampling is terminated.

Theorem 3.2.1. *Under a multistage estimation strategy, suppose that* (i) *the final sample size* Q *is an observable random variable that is finite w.p.1, and* (ii) Q *is determined in such a way that the event* $Q = q$ *is measurable with respect to* $\{\hat{\sigma}_j; m \leq j \leq q\}$, for all fixed $q \geq m(>\max\{1, a, -ab\})$.

Then, the expression for the risk per unit cost associated with the terminal estimator $\widehat{\Delta}_Q$ *is given by*:

$$E\left[\text{RPUC}_Q\right] = \omega \tfrac{n^{*3}}{a^2(1+b^2-b)} \left\{ E\left[\tfrac{1}{Q}\left(1-\tfrac{a}{Q}\right)^{-1}\left(1+\tfrac{ab}{Q}\right)^{-1}\right] - E\left(\tfrac{a-ab}{Q^2}\right) - E\left(\tfrac{1}{Q}\right)\right\}. \qquad (3.2.8)$$

Its proof is similar along the lines of Section 2.6.1 of Chapter 2 and hence it is omitted for brevity.

3.3. A MODIFIED TWO-STAGE PROCEDURE

The customary two-stage procedure developed by Stein (1945,1949) has some advantages and some disadvantages which were elaborated in Chapter 2. We thus resort to a modified two-stage procedure along the lines of Mukhopadhyay and Duggan (1997) to achieve attractive second-order properties.

The key idea is to introduce a lower bound σ_L such that $0 < \sigma_L < \sigma$ with σ_L known. Given this additional input, from the expression of n^* found in (3.2.7), we note that

$$n^* > \left(a^2(1+b^2-b)(c\omega)^{-1}\right)^{1/3} \sigma_L^{k/3}.$$

Thus, we fix an integer $m_0(> \max\{1, a, -ab\})$ and choose the pilot size m in such a way that $m \approx \left(a^2(1+b^2-b)(c\omega)^{-1}\right)^{1/3} \sigma_L^{k/3}$. We formally define m as follows:

$$m \equiv m(\omega) = \max\left\{m_0, \left\lfloor d_\omega \sigma_L^{k/3}\right\rfloor + 1\right\} \text{ with } d_\omega = \left(a^2(1+b^2-b)(c\omega)^{-1}\right)^{1/3}, \qquad (3.3.1)$$

and gather pilot data $\{X_i, Y_i; i = 1, ..., m\}$. Here and elsewhere, $\lfloor s \rfloor$ denotes the largest integer less than $s(> 0)$.

From pilot data, we obtain the statistic $\hat{\sigma}_m \equiv W_m = \frac{1}{2}\left(U_m + b^{-1}V_m\right)$ where U_m and V_m are defined via (3.2.1), and then determine the terminal sample size N as follows:

$$N \equiv N(\omega) = \max\left\{m, \left\lfloor d_\omega W_m^{k/3}\right\rfloor + 1\right\}, \qquad (3.3.2)$$

which is an estimator of n^* defined in (3.2.7).

If $N = m$, we would not require additional observations at the second stage. However, if $N > m$, then we sample the difference $N - m$ at the second stage by recording an additional set of pairs of observations $\{(X_i, Y_i), i = m+1, ..., N\}$. From full data $\{(X_i, Y_i), i = 1, ..., N\}$ obtained by combining both stages, we propose to estimate Δ by the difference of the smallest order statistics, namely:

$$\widehat{\Delta}_N \equiv X_{N:1} - Y_{N:1} = \min\{X_1, ..., X_N\} - \min\{Y_1, ..., Y_N\}.$$

Along the lines of (3.2.2), the associated loss function will be:

$$L_N = \exp\left(a\left[\sigma^{-1}(\widehat{\Delta}_N - \Delta)\right]\right) - a\left[\sigma^{-1}(\widehat{\Delta}_N - \Delta)\right] - 1, \tag{3.3.3}$$

A major difference between (3.2.2) and (3.3.3) is that the sample size N used in (3.3.3) is a random variable unlike n.

Now, since $\widehat{\Delta}_n$ and $I(N = n)$ are independent for all fixed $n \geq m$, using (3.2.8) with Q replaced by N from (3.3.2), we get:

$$\omega^{-1} E\left[\mathrm{RPUC}_N\right] = \tfrac{n^{*3}}{a^2(1+b^2-b)}\left\{E\left[\tfrac{1}{N}\left(1-\tfrac{a}{N}\right)^{-1}\left(1+\tfrac{ab}{N}\right)^{-1}\right] - E\left(\tfrac{a-ab}{N^2}\right) - E\left(\tfrac{1}{N}\right)\right\}. \tag{3.3.4}$$

3.3.1. First-Order Asymptotics

We begin with some attractive first-order results involving N. One will surely note that there is no stated sufficient condition involving m in Theorem 3.3.1. This is so because in the present setup, we have $m \equiv m(\omega) \to \infty$ as $\omega \to 0$.

Theorem 3.3.1. *With m and N respectively defined in (3.3.1) and (3.3.2), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, b, a$ we have as $\omega \to 0$:*

(*i*) $N/n^* \xrightarrow{P} 1$; $n^*/N \xrightarrow{P} 1$;

(*ii*) $E\left[(N/n^*)^t\right] \to 1, t = -1, 1$ [*asymptotic first-order efficiency*];

where n^* comes from (3.2.7).

Its proof follows in similar lines as in Section 2.6.3 of Chapter 2 and we thus leave it out for brevity. One clearly sees that all the expressions in Theorem 3.3.1 converge to 1 which looks very encouraging because we can expect N and n^* to be in the same ball park. Indeed one may claim convergence of higher positive and negative moments of N/n^* in part (ii) by referring to Mukhopadhyay and Duggan (1997,1999) and Mukhopadhyay (1999), but we leave them out for brevity.

3.3.2. Second-Order Asymptotics

The second-order asymptotics are similar to what is presented in equation (2.4.3) of Chapter 2, with N replaced from (3.3.2) and we leave them out for brevity. One should note that all such asymptotics are readily accessible along the lines of Mukhopadhyay and Duggan (1997,1999).

Next, let us define the expressions for ψ and σ_0^2 as follows:

$$\psi = \left(\frac{c}{a^2(1+b^2-b)}\right)^{1/3} \left\{\frac{k}{12}\left(\frac{k}{3}-1\right)\left(\frac{a^2(1+b^2-b)}{c}\right)^{1/3}\right\}\left(\frac{\sigma}{\sigma_L}\right)^{k/3} \text{ and } \sigma_0^2 = \frac{k^2}{18}\left(\frac{\sigma}{\sigma_L}\right)^{k/3}. \quad (3.3.5)$$

For completeness, now we exhibit both lower and upper bounds for $E(N-n^*)$.

Theorem 3.3.2. *With m and N respectively defined in (3.3.1) and (3.3.2), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, b, a$ we have as $\omega \to 0$:*

$$\psi + O(\omega^{1/2}) \le E(N-n^*) \le \psi + 1 + O(\omega^{1/2}) \text{ [\textit{asymptotic second-order efficiency}]}; \quad (3.3.6)$$

where ψ is defined in (3.3.5) and n^* comes from (3.2.7).

Theorem 3.3.2 shows the second-order efficiency property of the modified two-stage procedure (3.3.1)-(3.3.2) in the sense of Ghosh and Mukhopadhyay (1981). Next, we look at a result which obtains the asymptotic distribution of a standardized version of N along the lines of Ghosh and Mukhopadhyay (1975) and Mukhopadhyay and Duggan (1997,1999).

Theorem 3.3.3. *With m and N respectively defined in (3.3.1) and (3.3.2), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, b, a$ we have as $\omega \to 0$:*

$$U^* \equiv n^{*-1/2}(N-n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_0^2), \quad (3.3.7)$$

where σ_0^2 is defined in (3.3.5) and n^* comes from (3.2.7).

Its proof follows from Lemma 2.1, part (i) in Mukhopadhyay and Duggan (1999) and hence it is omitted. The following theorem evaluates the risk per unit cost up to second-order approximation. The proof follows along the lines of section 2.6.4 in Chapter 2 and is thus left out for brevity.

Theorem 3.3.4. *With m and N respectively defined in (3.3.1) and (3.3.2), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, b, a$ we have as $\omega \to 0$:*

$$\begin{aligned} 1 + \tfrac{1}{n^*}(6\sigma_0^2 - 3\psi - 3 + a\left(\tfrac{b^2 - b^3 - b + 1}{1 + b^2 - b}\right)) + o(\tfrac{1}{n^*}) &\leq \omega^{-1} E\left[\mathrm{RPUC}_N\right] \\ &\leq 1 + \tfrac{1}{n^*}(6\sigma_0^2 - 3\psi + a\left(\tfrac{b^2 - b^3 - b + 1}{1 + b^2 - b}\right)) + o(\tfrac{1}{n^*}) \end{aligned} \quad (3.3.8)$$

[*asymptotic second-order risk efficiency*];

where ψ, σ_0^2 are as defined in (3.3.5) and n^* comes from (3.2.7).

3.4. A PURELY SEQUENTIAL PROCEDURE

A purely sequential procedure provides tighter results compared with those (Theorem 3.3.4) for the modified two-stage procedure. This happens perhaps because this methodology allows us to take as many pairs of observations sequentially step-by-step as required depending on the rule of termination.

In this section, we will pursue a purely sequential sampling strategy along the lines of Chapter 2 to address our two-sample problem. We will make use of nonlinear renewal theory to provide second-order approximations for the average sample size and RPUC, the risk per unit cost.

We recall the expressions of n^* from (3.2.7). We again fix an integer $m(> \max\{1, a, -ab\})$ and obtain pilot data $\{X_i, Y_i; i = 1, ..., m\}$. We then proceed by recording one additional pair (X, Y) at every step as needed determined by the following rule:

$$N \equiv N(\omega) = \inf\left\{n \geq m : n \geq d_\omega W_n^{k/3}\right\} \text{ with } d_\omega = \left(a^2(1 + b^2 - b)(c\omega)^{-1}\right)^{1/3}. \quad (3.4.1)$$

Again, this stopping variable N estimates n^*. From full data $\{(X_i, Y_i), i = 1, ..., N\}$ gathered upon termination of sampling, we propose to estimate Δ by the difference of the smallest order statistics, namely:

$$\widehat{\Delta}_N \equiv X_{N:1} - Y_{N:1} = \min\{X_1, ..., X_N\} - \min\{Y_1, ..., Y_N\}.$$

Along the lines of (3.2.2), the associated loss function will be:

$$L_N = \exp\left(a\left[\sigma^{-1}(\widehat{\Delta}_N - \Delta)\right]\right) - a\left[\sigma^{-1}(\widehat{\Delta}_N - \Delta)\right] - 1.$$

Since $\widehat{\Delta}_n$ and $I(N = n)$ are independent for all fixed $n \geq m$, the associated expression for $E\left[\mathrm{RPUC}_N\right]$ will again resemble (3.2.8) with Q replaced by N from (3.4.1).

3.4.1. First-Order Asymptotics

The first order asymptotics that we are now presenting resemble Theorem 2.5.1 in Chapter 2. The proof of this theorem follows along same lines as in Section 2.6.5 and hence we show a partial derivation in Section 3.5.1.

Theorem 3.4.1. *For N defined in (3.4.1), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, b, a$ we have as $\omega \to 0$:*

(*i*) $N/n^* \xrightarrow{P} 1$; $n^*/N \xrightarrow{P} 1$;

(*ii*) $E\left[(N/n^*)^t\right] \to 1$ *for* $t > 0$ *if* $m(> \max\{2, a, -ab\})$ [*asymptotic first-order efficiency*];

(*iii*) $E\left[(n^*/N)^t\right] \to 1$ *for* $t > 0$ *if* $m > \max\{1 + \frac{1}{3}kt, a, -ab\}$;

(*iv*) $\omega^{-1}E\left[\mathrm{RPUC}_N\right] \to 1$ *if* $m > \max\{1 + \frac{4}{3}k, a, -ab\}$ [*asymptotic first-order risk efficiency*];

(*v*) $V^* \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2)$ *with* $\sigma_1^2 = \frac{k^2}{18}$;

where n^* comes from (3.2.7).

The expressions shown in Theorem 3.4.1, parts (i)-(iv) converge to 1 which are parallel to the results from Theorem 3.3.1 under the modified two-stage procedure (3.3.1)-(3.3.2). In other words, the purely sequential procedure (3.4.1) also has attractive asymptotic first-order properties.

Now, in order to contrast the asymptotic normality properties of standardized N of the modified two-stage procedure (3.3.1)-(3.3.2) and the purely sequential procedure (3.4.1), first recall from Theorem 3.3.3 that

$$U^* \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_0^2) \text{ with } \sigma_0^2 = \tfrac{k^2}{18}(\tfrac{\sigma}{\sigma_L})^{k/3}.$$

As opposed to that, Theorem 3.4.1, part (v) shows that

$$V^* \equiv n^{*-1/2}(N - n^*) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2) \text{ with } \sigma_1^2 = \tfrac{k^2}{18}.$$

Thus, we clearly observe that the asymptotic normal distribution for the standardized stopping variable V^* definitely has a smaller asymptotic variance (σ_1^2) compared with that (σ_0^2) for U^*. This gives an edge of the purely sequential strategy over the modified two-stage strategy.

3.4.2. Second-Order Asymptotics

Going back to applications of nonlinear renewal theoretic results provided in Chapter 2 for a one-sample problem, we express N from (3.4.1) as $R+1$ w.p.1 where

$$\begin{aligned} R &= \inf\left\{n \geq m-1 : \Sigma_{i=1}^{n} Z_i \leq n^{*-3/k} n^{\frac{3}{k}+1} (1+\tfrac{1}{n})^{\frac{3}{k}}\right\} \\ &= \inf\left\{n \geq m-1 : \Sigma_{i=1}^{n} Z_i \leq h^* n^{\delta} L(n)\right\}, \end{aligned} \tag{3.4.2}$$

with the Z_i's being i.i.d. Gamma$(2, \frac{1}{2})$ random variables. A brief explanation is laid out inside the proof of Theorem 3.4.1, part (iii) in Section 3.5.1.

Using Woodroofe (1977), Lai and Siegmund (1977,1979), and especially the representations laid out in Mukhopadhyay (1988) and Mukhopadhyay and Solanky (1994, Section 2.4.2), clearly R from (3.4.2) matches with (2.4.7) in Mukhopadhyay and Solanky (1994) where

$$\delta = \tfrac{3}{k} + 1,\; h^* = n^{*-3/k},\; L(n) = 1 + \tfrac{3}{kn} + o\left(\tfrac{1}{n}\right), \text{so that } L_0 = \tfrac{3}{k}, \tag{3.4.3}$$

and also note:

$$\theta = 1\,, \tau^2 = \tfrac{1}{2}, \beta^* = \tfrac{k}{3}, n_0^* = n^* \text{ and } p = \tfrac{k^2}{18}. \tag{3.4.4}$$

Condition (2.5) from Mukhopadhyay (1988) or equivalently (2.4.8) from Mukhopadhyay and Solanky (1994) is satisfied with $B = 2$ and $b = 1$. We define two special entities:

$$\begin{aligned} &\nu \equiv \nu_k = \tfrac{k}{6}\left(\tfrac{9}{k^2}+1\right) - \tfrac{1}{4}\Sigma_{n=1}^{\infty} n^{-1} E\left\{\max\left(0, \chi^2_{4n} - 4n\left(\tfrac{3}{k}+1\right)\right)\right\}, \text{ and} \\ &\eta \equiv \eta_k = \tfrac{1}{3}k\nu - \left(\tfrac{1}{36}k^2 + \tfrac{1}{12}k + 1\right), \end{aligned} \tag{3.4.5}$$

along the lines of (2.4.9)-(2.4.10) in Mukhopadhyay and Solanky (1994) and (2.5.5) in Chapter 2.

Table 3.1 illustrates a few values of $\nu \equiv \nu_k$ and $\eta \equiv \eta_k$ with $k = 1, 2, 3, 4, 5, 6$.

To conclude this section, we now specify asymptotic second-order expansions for both positive and negative moments of $\frac{N}{n^*}$ (Theorem 3.4.2) as well as an asymptotic second-order expansion of the risk per unit cost (Theorem 3.4.3) under the purely sequential setting (3.4.1). Recall that we must also have $m(> \max\{1, a, -ab\})$ satisfied in the background for Theorems 3.4.2-3.4.3 to hold. Brief outlines of proofs of Theorems 3.4.2-3.4.3 are shown in Section 3.5.2.

Theorem 3.4.2. *For N defined in (3.4.1), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, b, a$ and every non-zero real number t, we have as $\omega \to 0$:*

$$E\left[(N/n^*)^t\right] = 1 + \left\{t\eta_k + t + \tfrac{1}{2}t(t-1)p\right\} n^{*-1} + o\left(n^{*-1}\right) \quad \text{[asymptotic second-order efficiency]}; \tag{3.4.6}$$

when (i) $m > \frac{(3-t)k}{3} + 1$ for $t \in (-\infty, 2) - \{-1, 1\}$; (ii) $m > \frac{k}{3} + 1$ for $t = 1$ and $t \geq 2$; and (iii) $m > \frac{2k}{3} + 1$ for $t = -1$; with n^*, p, and η_k coming from (3.2.7), (3.4.4), and (3.4.5) respectively.

Theorem 3.4.3. *For N defined in (3.4.1), for each fixed value of $\mu_1, \mu_2, \sigma, c, k, a$, we have the following second-order expansion of the risk per unit cost as $\omega \to 0$:*

$$\omega^{-1} E\left[\mathrm{RPUC}_N\right] = 1 + (6p - 3\eta_k - 3 + a\left(\tfrac{b^2 - b^3 - b + 1}{1 + b^2 - b}\right))n^{*-1} + o\left(n^{*-1}\right) \quad \text{[asymptotic second-order risk efficiency]}; \tag{3.4.7}$$

when $m > \max\left\{\frac{7k}{3} + 1, a, -ab\right\}$ with n^*, p, and η_k coming from (3.2.7), (3.4.4), and (3.4.5) respectively.

3.5. TECHNICAL DETAILS AND PROOFS OF THEOREMS

Since the technicalities are largely similar to those in Chapter 2, we show some steps selectively. Intermediate steps are kept out for brevity.

3.5.1. Purely Sequential Procedure: Proof of Theorem 3.4.1

Part (i): Follows from Lemma 1 of Chow and Robbins (1965).

Part (ii): With $m \geq 3$ we can claim that $\frac{N-1}{N-2} \leq 2$ w.p.1, and let

$$H^* = \sup_{n \geq 2} \left\{ \tfrac{1}{n} \Sigma_{i=1}^{n} (X_i - \mu_1) + \tfrac{1}{bn} \Sigma_{i=1}^{n} (Y_i - \mu_2) \right\},$$

for sufficiently small $\omega(> 0)$ so that $n^* > m$, observe the following inequality (w.p.1):

$$\begin{aligned} N \leq m + d_\omega W_{N-1}^{k/3} &\leq m + 2^{k/3} d_\omega \left\{ \tfrac{1}{N-2} \Sigma_{i=1}^{N-1} (X_i - \mu_1) + \tfrac{1}{b(N-2)} \Sigma_{i=1}^{N-1} (Y_i - \mu_2) \right\}^{k/3} \\ &\leq m + 2^{k/3} d_\omega H^*, \end{aligned}$$

which implies (w.p.1):

$$\tfrac{N}{n^*} \leq 1 + 2^{k/3} \sigma^{-k/3} H^*. \tag{3.5.1}$$

Now, by Wiener's (1939) ergodic theorem, it follows that $E[H^{*t}]$ is finite for all fixed positive number t. The right-hand side of (3.5.1) is also free from ω so that we can claim uniform integrability of all positive powers of $\frac{N}{n^*}$. Then, by appealing to the dominated convergence theorem and part (i), we claim part (ii).

Part (iii): Let us denote:

$$S_{x,n}^* = \Sigma_{i=2}^{n} (n - i + 1)(X_{n:i} - X_{n:i-1}) \text{ and } S_{y,n}^* = \Sigma_{i=2}^{n} (n - i + 1)(Y_{n:i} - Y_{n:i-1}),$$

so that we have $\hat{\sigma}_n = \frac{1}{2(n-1)} \left(S_{x,n}^* + \frac{1}{b} S_{y,n}^* \right)$. Let $L_1, L_2, \ldots$ and $M_1, M_2, \ldots$ be i.i.d. $Exp(\sigma)$ and $Exp(b\sigma)$ random variables respectively, L's and M's being independent, and denote $S_{x,n} = \Sigma_{i=1}^{n-1} L_i$,. $S_{y,n} = \Sigma_{i=1}^{n-1} M_i$.

Then, utilizing the embedding ideas from Lombard and Swanepoel (1978) and Swanepoel and van Wyk (1982), we can claim that the distribution of $\{S_{x,n}^*, S_{y,n}^*; n \leq n_0\}$ is identical to that of $\{S_{x,n}, S_{y,n}; n \leq n_0\}$ for all n_0. Thus, N given by (3.4.1) can be equivalently expressed as:

$$\begin{aligned} N &\equiv \inf \left\{ n \geq m : n \geq d_\omega \left(\tfrac{1}{2(n-1)} \left(\Sigma_{i=1}^{n-1} L_i + \tfrac{\Sigma_{i=1}^{n-1} M_i}{b} \right) \right)^{k/3} \right\} \\ &= \inf \left\{ n \geq m : \left(\tfrac{n}{n^*} \right)^{3/k} (n-1) \geq \Sigma_{i=1}^{n-1} Z_i \right\}, \end{aligned} \tag{3.5.2}$$

where Z_i's are i.i.d. Gamma$(2, \frac{1}{2})$ random variables.

Now, N from (3.5.2) can be written as $R + 1$ w.p.1 with R defined in (3.4.2). Using Lemma

2.3 from Woodroofe (1977) or Theorem 2.4.8, part (i) of Mukhopadhyay and Solanky (1994) with $b = 1$, we can claim:

$$P(R \le \tfrac{1}{2}n^*) = O(n^{*-\frac{3}{k}(m-1)}). \tag{3.5.3}$$

Next, with fixed $t > 0$, since $N = R + 1$ w.p.1, we have $0 < \left(\frac{n^*}{N}\right)^t \le \left(\frac{n^*}{R}\right)^t$ so that $\left(\frac{n^*}{N}\right)^t$ will be uniformly integrable if we show:

$$\left(\tfrac{n^*}{R}\right)^t \text{ is uniformly integrable.} \tag{3.5.4}$$

Now, we may write (w.p.1):

$$\left(\tfrac{n^*}{R}\right)^t I(R > \tfrac{1}{2}n^*) < 2^t,$$

so that $\left(\frac{n^*}{R}\right)^t I(R > \frac{1}{2}n^*)$ must be uniformly integrable. But, $\left(\frac{n^*}{R}\right)^t I(R > \frac{1}{2}n^*) \xrightarrow{P} 1$ and hence, we must have:

$$E\left[\left(\tfrac{n^*}{R}\right)^t I(R > \tfrac{1}{2}n^*)\right] = 1 + o(1). \tag{3.5.5}$$

Additionally, in view of (3.5.3), we also note the following:

$$E\left[\left(\tfrac{n^*}{R}\right)^t I(R \le \tfrac{1}{2}n^*)\right] \le \left(\frac{n^*}{m-1}\right)^t P(R \le \tfrac{1}{2}n^*) = O(n^{*-\frac{3}{k}(m-1)+t}), \tag{3.5.6}$$

which is $o(1)$ if $m > 1 + \frac{1}{3}kt$.

Combining (3.5.5)-(3.5.6), clearly (3.5.4) follows so that we can claim:

$$E\left[\left(\tfrac{n^*}{R}\right)^t\right] = 1 + o(1) \text{ if } m > 1 + \tfrac{1}{3}kt, \text{ with } t > 0, \tag{3.5.7}$$

which is part (iii).

Part (iv): The proof is similar to that in Section 2.6.5 of Chapter 2.

Part (v): This result follows directly from an application of Ghosh and Mukhopadhyay's (1975) theorem. One may also refer to Theorem 2.4.3 or Theorem 2.4.8, part (ii) in Mukhopadhyay and Solanky (1994). ■

3.5.2. Outlines of Proofs of Theorems 3.4.2-3.4.3

Theorem 3.4.2 follows along the lines of Theorem 2.4.8, part (iv) and from its established applications found in Mukhopadhyay and Solanky (1994).

For a proof of Theorem 3.4.3, we recall that the associated expression for $E[\mathrm{RPUC}_N]$ will resemble (3.2.8) with Q replaced by N from (3.4.1). Then, one will proceed with an expansion of $\omega^{-1}E[\mathrm{RPUC}_N]$ similar to that in Section 2.6.4 of Chapter 2 and exploit Theorem 3.4.2 with $t = -3, -4$. Additional details are kept out for brevity. ■

3.6. DATA ANALYSIS: SIMULATIONS

After developing some interesting theoretical results associated with the two proposed estimation methodologies in Sections 3.3-3.4, we now move on to implement these strategies via computer simulations. We examine how these estimation strategies may perform when sample sizes are small (20) to moderate $(50, 100, 150)$ to large $(300, 500)$. All simulations are carried out with R codes based on $10,000 (= H$, say) replications under each configuration and methodology.

Under each procedure, we generated pseudo-random observations from the distribution (3.1.1) with $\mu_1 = 8, \sigma_1 = \sigma = 10$ and $\mu_2 = 5, \sigma_2 = b\sigma = 10b$ where we fixed b so that σ_2 became an appropriate multiple of σ. Moreover, we also fixed certain values of a, c, k and n^*, thereby solving for a corresponding value of the risk-bound, ω. Thus, a set of preassigned values for $b, a, c, k, \omega, \sigma$ will have the associated n^* values as shown in our tables (column 1) to come.

In the context of the modified two-stage methodology (3.3.1)-(3.3.2), we fixed a positive lower bound σ_L for σ and a number m_0, thereby determining m from (3.3.1). On the other hand, we fixed a pilot sample size m in the context of the purely sequential methodology (3.4.1). While implementing a methodology to determine the final sample size (N) and a terminal estimator $(X_{N:1} - Y_{N:1})$ of $\mu_1 - \mu_2$, we pretended that we did not know μ_1, μ_2, σ, and n^* values.

Now, we specify a set of notation used in the tables. Under a fixed configuration with all necessary input (e.g., $a, c, k, \omega, m, m_0, \sigma_L$ as appropriate), we focus on implementing a particular estimation methodology. We ran the i^{th} replication by beginning with m pilot observations and then eventually ended sampling by recording a final sample size $N = n_i$, terminal estimator $x_{n_i:1} - y_{n_i:1}$, and the achieved risk per unit cost:

$$\text{RPUC}_{n_i} = \omega \frac{n^{*3}}{a^2(1+b^2-b)} \left\{ \frac{1}{n_i} \left(1 - \frac{a}{n_i}\right)^{-1} \left(1 + \frac{ab}{n_i}\right)^{-1} - \left(\frac{a-ab}{n_i^2}\right) - \left(\frac{1}{n_i}\right) \right\} = r_i, \text{ say,} \tag{3.6.1}$$

$i = 1, ..., H(= 10,000)$.

$\overline{n} = H^{-1}\Sigma_{i=1}^{H} n_i$	Estimate of $E(N)$ or n^*;
$\mathrm{s}_{\overline{n}} = \sqrt{(H^2-H)^{-1}\Sigma_{i=1}^{H}(n_i - \overline{n})^2}$	Estimated standard error of $\overline{n}$;
$\overline{x}_{\min} = H^{-1}\Sigma_{i=1}^{H} x_{n_i:1}$, $\overline{y}_{\min} = H^{-1}\Sigma_{i=1}^{H} y_{n_i:1}$	Estimates of μ_1 and μ_2 respectively;
$\overline{\Delta} = H^{-1}\Sigma_{i=1}^{H}(x_{n_i:1} - y_{n_i:1})$	Estimate of Δ, that is $\mu_1 - \mu_2$;
$\mathrm{s}_{\overline{x}_{\min}} = \sqrt{(H^2-H)^{-1}\Sigma_{i=1}^{H}(x_{n_i:1} - \overline{x}_{\min})^2}$	Estimated standard error of $\overline{x}_{\min}$;
$\mathrm{s}_{\overline{y}_{\min}} = \sqrt{(H^2-H)^{-1}\Sigma_{i=1}^{H}(y_{n_i:1} - \overline{y}_{\min})^2}$	Estimated standard error of $\overline{y}_{\min}$;
$\mathrm{s}_{\overline{\Delta}} = \sqrt{(H^2-H)^{-1}\Sigma_{i=1}^{H}(x_{n_i:1} - y_{n_i:1} - \overline{\Delta})^2}$	Estimated standard error of $\overline{\Delta}$;
r_i	RPUC_{n_i} from (3.6.1);
$\overline{r} = H^{-1}\Sigma_{i=1}^{H} r_i$ with r_i from (6.1)	Risk estimator;
$\mathrm{s}_{\overline{r}} = \sqrt{(H^2-H)^{-1}\Sigma_{i=1}^{H}(r_i - \overline{r})^2}$	Estimated standard error of $\overline{r}$;
$\overline{z} = \overline{r}/\omega$	Estimated risk efficiency to be compared with 1;
$\mathrm{s}_{\overline{z}} = \mathrm{s}_{\overline{r}}/\omega$	Estimated standard error of $\overline{z}$;

We now summarize the observed performances of the proposed estimation methodologies laid down in Sections 3.3-3.4. We obtained a large set of tables and results summarizing extensive simulations run under a variety of configurations. For brevity, we outline a small subset of our findings.

3.6.1. Modified Two-Stage Procedure (3.3.1)-(3.3.2)

We now summarize performances for the modified two-stage estimation methodology (3.3.1)-(3.3.2) in Table 3.2 for

$$n^* = 20, 50, 100, 300, 500 \text{ and}$$

$$(k, m_0) = (1,5), (2,6), (3,7).$$

In this methodology, we need a positive and known lower bound $\sigma_L(=3)$ for true σ, but σ remains unknown in practice. The pilot size m was determined from (3.3.1) but it is not shown in Table 3.2. The estimation methodology (3.3.2) was implemented as described. Table 3.2 specifies $\mu_1, \mu_2, \sigma, \sigma_L, b, a, c$, and each block shows (k, m_0), n^* (column 1), ω (column 2), the estimated (from 10,000 simulations) values $\overline{x}_{\min}$,$s_{\overline{x}_{\min}}$ (column 3), $\overline{y}_{\min}$,$s_{\overline{y}_{\min}}$ (column 4), $\overline{\Delta}(=\overline{x}_{\min}-\overline{y}_{\min})$,$s_{\overline{\Delta}}$ (column 5), values $\overline{n}$,$s_{\overline{n}}$ (column 6), the ratio $\overline{n}/n^*$ (column 7), and values $\overline{z}$,$s_{\overline{z}}$ (column 8).

All $\overline{x}_{\min}$ and $\overline{y}_{\min}$values appear close to $8(=\mu_1)$ and $5(=\mu_2)$ respectively with very small estimated standard error values $s_{\overline{x}_{\min}}$, $s_{\overline{y}_{\min}}$for sample size 100 or over. We observe that $\overline{\Delta}$ accurately estimates the true difference Δ in locations which is 3. For all (k, m_0) values, we see that $\overline{n}$ estimates n^* very accurately across the board. These features are consistent with Theorem 3.3.1, parts (i)-(ii).

The last column shows that the modified two-stage estimation methodology (3.3.1)-(3.3.2) is very successful in the case of all selected values of k for all n^* under consideration in delivering a risk-bound approximately preset ω. More specifically, in the case $k=1$, we note that $\overline{z}<1$, that is $\overline{r}<\omega$, for all chosen n^* values which is remarkable. On the other hand, $\overline{z}$ values hang around 1 for all chosen n^* and k values which is certainly encouraging.

The entries in Table 3.3 are similar to those in Table 3.2. Table 3.3 used a different lower bound $\sigma_L(=5)$ for true σ. We highlight performances for $n^*=20, 150, 500$. Again, the pilot size m was determined from (3.3.1) but m is not shown in Table 3.3. There are two sets of two blocks corresponding to different values of $b(=2,3)$ and $k(=1,2)$. We find that $\overline{n}/n^*$ is nearer to 1 in Table 3.3 as compared to those in Table 3.2. Also, entries found in the last column of Table 3.3 look more attractive to those in Table 3.2. This feature should be expected since the specified positive and known lower bound $\sigma_L=5$ is closer to $\sigma=10$ than $\sigma_L=3$ is.

In Table 3.4, we provide the values of ψ found in (3.3.5) corresponding to the configurations highlighted in Table 3.2. We expect that $E(N-n^*)$ values should lie inside the interval $[\psi, \psi+1]$ for large n^* values in view of Theorem 3.3.2. This is indeed true as we find that nearly all $\bar{n}-n^*$ values lie inside the corresponding interval $[\psi, \psi+1]$. Chapter 2 briefly mentioned the case when σ_L may be slightly misspecified.

We provide Figure 3.2 showing empirical validation of asymptotic normality result described in (3.3.7). We considered two scenarios, namely $\sigma_L=3$, $b=1$, $k=1$, $n^*=100$ (Figure 3.2a); and $\sigma_L=3$, $b=2$, $k=3$, $n^*=500$ (Figure 3.2b). Under each configuration, we recorded observed values:

$$N = n_i, i = 1, ..., H(= 10,000),$$

and thus calculated $10,000$ associated standardized $(n_i - n^*)/\sqrt{n^*}$ values. Under each specific configuration, such $10,000$ observed $u_i^* \equiv (n_i - n^*)/\sqrt{n^*}$ values provide the empirical distribution of the standardized sample size (dashed curve in red). We superimpose on it the expected theoretical $N(0, \sigma_0^2)$ distributions (solid curve in blue) where $\sigma_0^2 = \frac{1}{18}(\frac{\sigma}{\sigma_L})^{k/3}k^2$ found from (3.3.5).

Both sets of plots in Figure 3.2(a) and 3.2(b) show very good fit. However, Figure 3.2(a) shows that the curves are slightly off from one another, but this corresponds to a moderate value of $n^*(= 100)$, Indeed, Figure 3.2(b) clearly shows that the empirical and theoretical distribution curves nearly lie on each other when $n^*(= 500)$ is large.

To supplement the graphical presentations in Figure 3.2, we also performed the customary Kolmogorov-Smirnov (K-S) test for normality in the case of each dataset that generated Figures 3.2(a)-3.2(b). Table 3.5 shows associated K-S test statistic (D) values under the null hypothesis of normality with associated p-values.

Both p-values (Table 3.5) are much larger than 0.05. Thus, our earlier thoughts supported by simple visual examinations of Figures 3.2(a)-3.2(b) are clearly validated by K-S test of normality under each scenario. That is, we are reasonably assured of a good fit between the observed values of u^* and a normal curve with a high level of confidence for all practical purposes.

3.6.2. Purely Sequential Procedure (3.4.1)

We now summarize performances for the purely sequential estimation methodology (3.4.1) in Table 3.6 for

$$n^* = 20, 50, 100, 300, 500 \text{ and}$$

$$(k, m) = (1, 4), (2, 6), (3, 9).$$

The estimation methodology (3.4.1) was implemented as described. Table 3.6 specifies $\mu_1, \mu_2, \sigma, b, a, c$ and each block shows (k, m), n^* (column 1), ω (column 2), the estimated (from $10,000$ simulations) values $\overline{x}_{\min}$,$s_{\overline{x}_{\min}}$ (column 3), $\overline{x}_{\min}$,$s_{\overline{x}_{\min}}$ (column 4), $\overline{\Delta}(= \overline{x}_{\min} - \overline{y}_{\min})$,$s_{\overline{\Delta}}$ (column 5) values $\overline{n}$,$s_{\overline{n}}$ (column 6), the ratio $\overline{n}/n^*$ (column 7), and values $\overline{z}$,$s_{\overline{z}}$ (column 8). An explanation of column 9 comes later.

All $\overline{x}_{\min}$ and $\overline{y}_{\min}$values appear closer to $8(=\mu_1)$ and $5(=\mu_2)$ respectively with very small estimated standard error values $s_{\overline{x}_{\min}}$, $s_{\overline{y}_{\min}}$for sample size 100 or over. $\overline{\Delta}$ also accurately estimates the actual difference in locations which is 3. For all (k,m) values, it appears that $\overline{n}$ estimates n^* very accurately throughout. These features are consistent with Theorem 3.4.1, parts (i)-(ii). Column 8 shows that the purely sequential estimation methodology (3.4.1) is very successful for all values of k and n^* under consideration in delivering a risk-bound approximately ω. However, when $k=1,2$, we see an under-estimation in the risk which improves when $k=3$. We feel that the purely sequential procedure provides estimates for the risk per unit cost which are generally closer to 1 than the two-stage methodology (3.3.1)-(3.3.2).

Theorem 3.4.3 showed that $\omega^{-1}E\left[\mathrm{RPUC}_N\right]$ should be close to $1+\varepsilon$ where we have:

$$\varepsilon = (6p - 3\eta_k - 3 + a\left(\frac{b^2-b^3-b+1}{1+b^2-b}\right))n^{*-1}, \tag{3.6.2}$$

from (3.4.7) when n^* is large. η_k was defined by (3.4.5) and it was tabulated in Table 3.1. Column 9 in Table 3.6 shows these ε values under each configuration. Upon comparing $\overline{z}$ values from column 8 with ε values from column 9, we see that for all practical purposes, the $\overline{z}$ values are explained fairly well by $1+\varepsilon$ for practical purposes.

We now address Figure 3.3 consisting of two side-by-side plots in our attempt to validate empirically the normality result described in Theorem 3.4.1, part (v). We considered two scenarios, $b=3,\ k=1,\ m=5,\ n^*=100$ in Figure 3.3(a) and $b=1,\ k=3,\ m=13,\ n^*=500$ in Figure 3.3(b). Under each configuration, we recorded observed values:

$$N = n_i, i = 1, ..., H(=10,000),$$

and thus calculated $10,000$ associated standardized $(n_i-n^*)/\sqrt{n^*}$ values. Under each specific configuration, such $10,000$ observed $v_i^* \equiv (n_i-n^*)/\sqrt{n^*}$ values provided the empirical distribution of the standardized sample size (dashed curve in red). We superimpose on it the appropriate theoretical $N(0,\sigma_1^2)$ distributions (solid curve in blue) where $\sigma_1^2 = \frac{1}{18}k^2$ coming from Theorem 3.4.1, part (v).

Both sets of plots in Figure 3.3(a) and 3.3(b) show very good fit. However, Figure 3.3(a) shows that the curves are slightly off from one another, but this corresponds to a moderate value of $n^*(=100)$, Indeed, Figure 3.3(b) clearly shows that the empirical and theoretical distribution curves nearly lie on each other when $n^*(=500)$ is large.

To supplement the graphical presentations in Figure 3.3, we also performed the customary Kolmogorov-Smirnov (K-S) test for normality in the case of each dataset that generated Figures 3.3(a)-3.3(b). Table 3.7 shows associated K-S test statistic (D) values under the null hypothesis of normality with associated p-values.

Both p-values (Table 3.7) are much larger than 0.05. and our earlier thoughts supported by visual examinations of Figures 3.3a-3.3b are clearly validated by K-S test of normality under each scenario. That is, we are reasonably assured of a good fit between the observed values of v^* and a normal curve with a high level of confidence for all practical purposes.

3.7. DATA ANALYSIS: ILLUSTRATIONS USING REAL DATA

In this section, we illustrate applications of the modified two-stage estimation methodology (3.3.1)-(3.3.2) and the purely sequential estimation methodology (3.4.1) using two real datasets, one from cancer research and the other from reliability. The first example (Section 3.7.1) uses parallel datasets on survival times of cancer patients from Shanker et al. (2016). The data were utilized in Efron's (1988) paper.

A second example (Section 3.7.2) uses lifetimes of steel components data that is available from the textbook of Lawless (2003, *Statistical Models and Methods for Lifetime Data*). This data were presented earlier by Crowder (2000).

3.7.1. Cancer Studies Data

This data consist of two independent datasets representing the survival times of a group of patients suffering from head and neck cancer. The first group was treated using a combination of radiotherapy and chemotherapy whereas the second group was treated with radiotherapy alone. These data on survival times, X from first group and Y from second group, have been presented in Shanker et al. (2016) and were earlier reported in Efron (1988).

The full data consisted of survival times of 44 and 51 patients respectively. As a check for a negative exponential model's fit, in Figure 3.4 we provide the exponential Q-Q plots for both the datasets. Indeed the data points seem to lie fairly well within or near the bands, indicating a good fit.

Treating these two datasets as the universe, we first found $\hat{\mu}_1 = 12.2$, $\hat{\sigma}_1 = 216.19$ and $\hat{\mu}_2 = 6.53$, $\hat{\sigma}_2 = 233.88$ from full data. We note that the scale parameters appear nearly the same. Thus, we assumed $b = 1$. Utilizing (3.2.1), we found the pooled estimator of the scale, $\hat{\sigma} = 225.04$. We then implemented both modified two-stage and purely sequential estimation procedures drawing observations (X, Y) from the full set of data as needed. It is emphasized, however, that implementation of sampling strategies did not exploit the observed numbers $\hat{\mu}_1$, $\hat{\mu}_2$, or $\hat{\sigma}$.

For estimating $\Delta \equiv \mu_1 - \mu_2$, we carried out a single run under both procedures. Tables 3.8-3.9 provide the results from implementing the stopping rules from (3.3.1)-(3.3.2) and (3.4.1) respectively corresponding to certain fixed values of c, k, a and the preset risk-bound ω, chosen arbitrarily. We fixed m or m_0 as needed.

The outcome of these methodologies are summarized in Tables 3.8-3.9. Table 3.8 summarizes results from the modified two-stage procedure with $\sigma_L = 80$, but assuming otherwise that the scale parameter σ remains unknown. Table 3.9 summarizes the results from the purely sequential procedure.

Under both methodologies, we notice that the terminal estimated values of μ_1, μ_2 are not too far away from corresponding $\hat{\mu}_1 = 12.2$ and $\hat{\mu}_2 = 6.53$ that was obtained from full data. Also, the estimate of the difference in locations given by $\widehat{\Delta}$ is also close to the actual value $\Delta = 5.67$. One other comment is in order: The n^* values shown in the first column are computed using (3.2.7) after replacing σ with $\hat{\sigma} = 225.04$ obtained from full data. Again, in running the estimation methodologies, we did not exploit the number $\hat{\sigma} = 225.04$.

We have provided n^* values just so that one is able to gauge whether the observed n-values look reasonable. The ratio n/n^* appears reasonably close to 1 which is nice to see. But, these correspond to a single run each, and hence observed n may not always be very close to n^*. The values of z, that is, the ratio of achieved risk per unit cost and preset goal ω, also appear reasonably under (or close) to 1.

3.7.2. Lifetimes of Steel Specimen Data

We considered Data G4 presented in Appendix G of the textbook, *Statistical Models and Methods for Lifetime Data*, of Lawless (2003). This data, earlier presented in Crowder (2000), consisted of lifetimes of steel specimens tested at 14 stress levels. We divided the whole data into two parts, one corresponding to stress levels 35 or lower (low stress) and the other corresponding to stress levels 35.5 or higher (high stress).

We then considered these two groups to be our two independent groups, low (high) stress group giving rise to failure time X (Y). The two datasets consisted of 140 observations each. As a check for a negative exponential (3.1.1) fit, Figure 3.5 provides the exponential Q-Q plots for both groups. Indeed the data seem to lie largely within or near the bands, indicating a reasonable fit.

The scale parameters are taken to be σ and $b\sigma$ for X, Y respectively, whereas the formulation demands that one should specify a value of b before obtaining the data, perhaps from previous analogous studies or prior experiences. We have thus resorted to the following practical approach.

We randomly picked 10 data points X (Y) from each group under low (high) stress levels and treat them as our data, say, from previous analogous study. In the population from where this prior data supposedly became available through a similar previous stress test, suppose that the low (high) stress level group had the scale parameter value σ_{Low} (σ_{High}). Following were those observed data values:

$$\boxed{\begin{array}{c} \textbf{Low } (X)\text{: } 206, 273, 312, 385, 415, 568, 669, 714, 767, 1056 \\ \Rightarrow \widehat{\sigma}_{\text{Low}} = \frac{1}{9}\sum_{i=1}^{n}(X_i - X_{10:1}) = \frac{1}{9}(5365 - 2060) = 367.22; \\ \textbf{High}(Y) : 66, 90, 105, 108, 121, 122, 127, 164, 255, 318 \\ \Rightarrow \widehat{\sigma}_{\text{High}} = \frac{1}{9}\sum_{i=1}^{n}(Y_i - Y_{10:1}) = \frac{1}{9}(1476 - 660) = 90.67. \end{array}} \quad (3.7.1)$$

We have

$$F_{18,18;0.025} = 0.38527 \text{ and } F_{18,18;0.975} = 2.5956$$

$$\Rightarrow 95\% \text{ confidence interval for } \sigma_{\text{High}}/\sigma_{\text{Low}} \text{ is given by}$$

$$\left(\tfrac{90.67}{367.22} \times \tfrac{1}{2.5959}, \tfrac{90.67}{367.22} \times \tfrac{1}{0.38527}\right), \text{ that is } (0.0951, 0.6409).$$

Thus, a null hypothesis postulating that $\sigma_{\text{High}}/\sigma_{\text{Low}} = 1$ will be rejected in favor of a two-sided test at 5% level. Indeed any number lying between 0.0951 and 0.6409 would seem reasonable for the ratio $\sigma_{\text{High}}/\sigma_{\text{Low}}$.

Thus, we decided to implement our proposed methodologies on the remainder of the stress test data after removing the observations shown in (3.7.1) from full dataset with two possible choices for the number b, namely, Case 1: $b = 0.33$ and Case 2: $b = 0.25$.

We treat the remaining 130 values from each stress groups as our universe and carry out our proposed methodologies. From full datasets, we found $\hat{\mu}_1 = 115$, $\hat{\mu}_2 = 51$ and $\widehat{\sigma} = 507.22$ when $b = 0.33$ whereas $\widehat{\sigma} = 570.78$ when $b = 0.25$. Table 3.10 shows findings from implementing the stopping rule (3.3.1)-(3.3.2) with $\sigma_L = 150$ whether $b = 0.33$ or $b = 0.25$. Table 3.11 shows findings from implementing the stopping rule when $b = 0.33, 0.25$. Both tables correspond to certain fixed choices of c, k, a and the preset risk-bound ω and m or m_0 as needed. The implementation of sampling strategies did not exploit the numbers $\hat{\mu}_1$, $\hat{\mu}_2$ or $\hat{\sigma}$ obtained from full datasets. For estimating $\Delta = \mu_1 - \mu_2$, we carried out a single run under both methodologies.

Here again, we show n^* values just so that one is able to gauge whether the observed n-values look reasonable. The estimate of difference Δ in locations given by $\overline{\Delta}$ appears fairly close to the actual value of Δ under either choices of b. The ratio n/n^* appears reasonably close to 1 which should be desirable. The value z, that is, the ratio of achieved risk per unit cost and preset goal ω, appears reasonably under (or close) to 1.

3.8. A BRIEF SUMMARY OF CHAPTER 3

In this chapter, we have developed methodologies that are operationally convenient and possess interesting efficiency and consistency properties. Our direction of research came from a thorough literature review and we proposed purely sequential and modified two-stage methods to estimate the difference Δ in location parameters of two independent negative exponential distributions.

One may notice that the purely sequential strategy appears to perform better overall than the modified two-stage strategy. However, it is also true that a modified two-stage strategy is logistically simpler to implement than a purely sequential estimation strategy. Indeed, both procedures are fully expected to perform very well. A practitioner, however, may consider employing one of the two procedures (3.3.1)-(3.3.2) or (3.4.1) that will provide an acceptable level of logistical comfort in running an experiment as one balances benefit under the presence of additional factors, namely, feasibility, efficiency, accuracy, operational convenience, and cost.

Table 3.1. Selected values of $\nu \equiv \nu_k$ and $\eta \equiv \eta_k$ from (3.4.5), $k = 1(1)6$

	k					
	1	2	3	4	5	6
ν_k	3.082	1.638	1.171	0.942	0.806	0.716
η_k	-0.084	-0.185	-0.328	-0.521	-0.768	-1.068

Table 3.2. Simulation results from 10,000 replications for the modified two-stage procedure (3.3.1)-(3.3.2) with $\mu_1 = 8$, $\mu_2 = 5$, $\sigma = 10$, $\sigma_L = 3$, $b = 1$, $a = 1$, $c = 0.1$

n^*	ω	$\overline{x}_{\min}$ $s_{\overline{x}_{\min}}$	$\overline{y}_{\min}$ $s_{\overline{y}_{\min}}$	$\overline{\Delta}$ $s_{\overline{\Delta}}$	$\bar{n}$ $s_{\bar{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$
				$\mathbf{k = 1, m_0 = 5}$			
20	1.25×10^{-2}	8.4929 0.0049	5.4901 0.0049	3.0027 0.0069	20.38 0.0132	1.0194	0.9710 0.0019
100	1×10^{-4}	8.0987 0.0010	5.1012 0.0010	2.9974 0.0014	100.45 0.0297	1.0045	0.9917 0.0008
300	3.7×10^{-6}	8.0334 0.0003	5.0336 0.0003	2.9998 0.0004	300.45 0.0496	1.0015	0.9971 0.0004
500	8×10^{-7}	8.0197 0.0001	5.0198 0.0001	2.9998 0.0002	500.43 0.0645	1.0008	0.9983 0.0003
				$\mathbf{k = 2, m_0 = 6}$			
20	1.25×10^{-1}	8.5065 0.0050	5.5084 0.0051	2.9980 0.0071	20.40 0.0334	1.0202	1.1188 0.0063
100	1×10^{-3}	8.1017 0.0010	5.0998 0.0010	3.0018 0.0014	100.42 0.0719	1.0042	1.0187 0.0022
300	3.7×10^{-5}	8.0335 0.0003	5.0336 0.0003	2.9998 0.0004	300.65 0.1213	1.0021	1.0033 0.0012
500	8×10^{-6}	8.0202 0.0002	5.0195 0.0001	3.0006 0.0002	500.20 0.1560	1.0004	1.0046 0.0009

Table 3.2 contd. Simulation results from $10,000$ replications for the modified two-stage procedure (3.3.1)-(3.3.2) with $\mu_1 = 8$, $\mu_2 = 5$, $\sigma = 10$, $\sigma_L = 3$, $b = 1$, $a = 1$, $c = 0.1$

n^*	ω	$\overline{x}_{\min}$ $s_{\overline{x}_{\min}}$	$\overline{y}_{\min}$ $s_{\overline{y}_{\min}}$	$\overline{\Delta}$ $s_{\overline{\Delta}}$	$\bar{n}$ $s_{\bar{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$
		$\mathbf{k = 3, m_0 = 7}$					
20	1.25	8.5184 0.0055	5.5333 0.0058	2.9850 0.0078	20.48 0.0571	1.0242	1.5479 0.0181
100	0.01	8.1045 0.0010	5.1024 0.0010	3.0020 0.0014	100.40 0.1308	1.0024	1.1020 0.0046
300	3.7×10^{-4}	8.0332 0.0003	5.0340 0.0003	2.9992 0.0004	300.46 0.2243	1.0015	1.0292 0.0023
500	8×10^{-5}	8.0201 0.0002	5.0198 0.0001	3.0002 0.0002	500.46 0.2896	1.0009	1.0175 0.0017

Table 3.3. Simulation results from 10,000 replications for the modified two-stage procedure (3.3.1)-(3.3.2) with $\mu_1 = 8,\ \mu_2 = 5,\ \sigma = 10,\ \sigma_L = 5,\ a = 1,\ c = 0.1$

n^*	ω	$\overline{x}_{\min}$ / $s_{\overline{x}_{\min}}$	$\overline{y}_{\min}$ / $s_{\overline{y}_{\min}}$	$\overline{\Delta}$ / $s_{\overline{\Delta}}$	$\bar{n}$ / $s_{\bar{n}}$	$\overline{n}/n^*$	$\overline{z}$ / $s_{\overline{z}}$
		$\mathbf{k=1, m_0=5, b=2}$					
20	3.75×10^{-2}	8.9922	5.4933	3.4989	20.39	1.0197	0.8992
		0.0049	0.0100	0.0112	0.0143		0.0019
150	8.88×10^{-5}	8.1316	5.0656	3.0659	150.46	1.0031	0.9838
		0.0006	0.0013	0.0014	0.0388		0.0007
500	2.4×10^{-6}	8.0402	5.0199	3.0202	500.50	1.0010	0.9948
		0.0001	0.0003	0.0004	0.0713		0.0004
		$\mathbf{k=2, m_0=6, b=2}$					
20	3.75×10^{-1}	9.0348	5.5075	3.5273	20.31	1.0158	1.0370
		0.0052	0.0105	0.0015	0.0332		0.0057
150	8.88×10^{-4}	8.1332	5.0666	3.0665	150.41	1.0027	1.0014
		0.0006	0.0013	0.0014	0.0883		0.0017
500	2.4×10^{-5}	8.0403	5.0195	3.0207	500.44	1.0009	1.0001
		0.0001	0.0002	0.0004	0.1601		0.0009

Table 3.3 contd. Simulation results from 10,000 replications for the modified two-stage procedure (3.3.1)-(3.3.2) with $\mu_1 = 8,\ \mu_2 = 5,\ \sigma = 10,\ \sigma_L = 5,\ a = 1,\ c = 0.1$

n^*	ω	$\overline{x}_{\min}$ / $s_{\overline{x}_{\min}}$	$\overline{y}_{\min}$ / $s_{\overline{y}_{\min}}$	$\overline{\Delta}$ / $s_{\overline{\Delta}}$	$\bar{n}$ / $s_{\bar{n}}$	$\overline{n}/n^*$	$\overline{z}$ / $s_{\overline{z}}$
		$\mathbf{k = 1, m_0 = 5, b = 3}$					
20	8.75×10^{-2}	9.1832	5.5031	3.6801	20.39	1.0195	0.9617
		0.0050	0.0149	0.0156	0.0168		0.0021
150	2.07×10^{-4}	8.1988	5.0673	3.1315	150.32	1.0021	0.9801
		0.0006	0.0019	0.0020	0.0451		0.0008
500	8×10^{-5}	8.0594	5.0198	3.0396	500.31	1.0006	0.9940
		0.0002	0.0005	0.0005	0.0816		0.0004
		$\mathbf{k = 2, m_0 = 6, b = 3}$					
20	8.75×10^{-1}	9.1187	5.5093	3.6094	20.33	1.0168	1.0309
		0.0052	0.0161	0.0167	0.0385		0.0067
150	2.07×10^{-3}	8.2026	5.0661	3.1364	150.60	1.0040	0.9957
		0.0006	0.0020	0.0021	0.1004		0.0020
500	5.6×10^{-5}	8.0597	5.0198	3.0399	500.23	1.0004	1.0011
		0.0001	0.0004	0.0004	0.1852		0.0011

Table 3.4. Values of $\bar{n} - n^*$, ψ from (3.3.5)

and $\psi + 1$ for each k used in Table 3.2

$n^* \backslash k$	1	2	3
20	0.38	0.40	0.48
50	0.40	0.32	0.65
100	0.45	0.42	0.24
300	0.45	0.65	0.46
500	0.43	0.20	0.46
ψ	-0.0829	-0.1239	0
$\psi + 1$	0.9170	0.8760	1

Table 3.5. Kolmogorov-Smirnov test results

corresponding to Figures 3.2(a)-3.2(b)

Case	Parameter configuration	K-S stat D	p-value
2a	$\sigma_L = 3,\ b = 1,\ k = 1,\ n^* = 100$	0.46252	0.8212
2b	$\sigma_L = 3,\ b = 2,\ k = 4,\ n^* = 500$	0.49874	0.9025

Table 3.6. Simulation results from 10,000 replications of the purely sequential procedure (3.4.1) with $\mu_1 = 8,\ \mu_2 = 5,\ \sigma = 10,\ b = 1, a = 1,\ c = 0.1$

n^*	ω	$\overline{x}_{\min}$ $s_{\overline{x}_{\min}}$	$\overline{y}_{\min}$ $s_{\overline{y}_{\min}}$	$\overline{\Delta}$ $s_{\overline{\Delta}}$	$\bar{n}$ $s_{\bar{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$	ε (3.6.2)
				$\mathbf{k = 1, m = 4}$				
20	1.25×10^{-2}	8.4902 0.0049	5.4975 0.0049	2.9926 0.0069	20.40 0.0110	1.0201	0.8612 0.0016	-0.1203
100	1×10^{-4}	8.0990 0.0010	5.1005 0.0009	2.9986 0.0014	100.40 0.0238	1.0040	0.9914 0.0007	-0.0241
300	3.7×10^{-6}	8.0335 0.0003	5.0336 0.0003	2.9999 0.0004	300.31 0.0414	1.0015	0.9979 0.0004	-0.0080
500	8×10^{-7}	8.0197 0.0001	5.0202 0.0002	2.9994 0.0002	500.45 0.0527	1.0009	0.9979 0.0003	-0.0048
				$\mathbf{k = 2, m = 6}$				
20	1.25×10^{-1}	8.4937 0.0050	5.5018 0.0051	2.9918 0.0071	20.30 0.0218	1.0150	0.9610 0.0040	-0.0555
100	1×10^{-3}	8.0990 0.0009	5.0994 0.0010	2.9995 0.0014	100.37 0.0472	1.0037	0.9965 0.0014	-0.0111
300	3.7×10^{-5}	8.0332 0.0003	5.0333 0.0003	2.9998 0.0004	300.38 0.0822	1.0012	0.9980 0.0008	-0.0037
500	8×10^{-6}	8.0197 0.0001	5.0203 0.0001	2.9994 0.0002	500.44 0.1046	1.0008	0.9986 0.0005	-0.0022

Table 3.6 contd. Simulation results from $10,000$ replications of the purely sequential procedure (3.4.1) with $\mu_1 = 8$, $\mu_2 = 5$, $\sigma = 10$, $b = 1, a = 1$, $c = 0.1$

n^*	ω	$\overline{x}_{\min}$ $s_{\overline{x}_{\min}}$	$\overline{y}_{\min}$ $s_{\overline{y}_{\min}}$	$\overline{\Delta}$ $s_{\overline{\Delta}}$	$\bar{n}$ $s_{\bar{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$	ε (3.6.2)
				$\mathbf{k = 3, m = 9}$				
20	1.25	8.5199 0.0054	5.5157 0.0053	3.0041 0.0075	20.14 0.0329	1.0074	1.2011 0.0091	0.0492
100	1×10^{-2}	8.1014 0.0010	5.1012 0.0010	3.0002 0.0014	100.23 0.0706	1.0023	1.0242 0.0022	0.0098
300	3.7×10^{-4}	8.0335 0.0003	5.0333 0.0003	3.0002 0.0004	300.19 0.1236	1.0006	1.0083 0.0012	0.0032
500	8×10^{-5}	8.0200 0.0001	5.0199 0.0001	3.0000 0.0002	500.04 0.1594	1.0000	1.0058 0.0009	0.0019

Table 3.7. Kolmogorov-Smirnov test results corresponding to Figures 3.3(a)-3.3(b)

Case	Parameter configuration	K-S stat D	p-value
3a	$b = 3, m = 5,\ k = 1,\ n^* = 100$	0.55487	0.8978
3b	$b = 1, m = 13,\ k = 3,\ n^* = 500$	0.52314	0.9487

Table 3.8. Analysis of cancer survival data using modified two-stage procedure (3.3.1)-(3.3.2) with $b = 1$, $a = 1$, $c = 0.1$, $\sigma_L = 80$

n^*	m_0	k	ω	$\hat{\mu}_1$: $x_{n:1}$	$\hat{\mu}_2$: $y_{n:1}$	$\widehat{\Delta}$	n	n/n^*	z
15	3	1	0.6667	11.64	6.84	4.8	16	1.067	0.9274
25	3		0.1440	12.52	6.21	6.31	24	0.96	0.9526
15	4	2	150.04	12.03	6.47	5.56	15	1.00	1.4201
25	4		32.410	12.38	6.64	5.74	26	1.04	1.2317

Table 3.9. Analysis of cancer survival data using purely sequential procedure (3.4.1) with $b = 1$, $a = 1$, $c = 0.1$

n^*	m	k	ω	$\hat{\mu}_1$: $x_{n:1}$	$\hat{\mu}_2$: $y_{n:1}$	$\widehat{\Delta}$	n	n/n^*	z	ε in (3.6.2)
15	5	1	0.66677	12.40	6.52	5.88	17	1.13	0.8629	-0.1609
25	5		0.14402	12.27	6.19	6.08	26	1.04	0.8897	-0.0965
15	7	2	150.046	13.02	6.49	6.53	15	1.00	0.9248	-0.0740
25	7		32.4101	13.11	6.62	6.49	28	1.12	0.9589	-0.0444

Table 3.10. Analysis of lifetimes of steel components in stress data using modified two-stage procedure (3.3.1)-(3.3.2) with $a = 1$, $c = 0.1$

n^*	m_0	k	ω	$\hat{\mu}_1$: $x_{n:1}$	$\hat{\mu}_2$: $y_{n:1}$	$\widehat{\Delta}$	n	n/n^*	z
				$b = 0.33$, $\sigma_L = 150$					
50	4	1	0.0316	129	51	78	52	1.04	0.9056
80	4		0.0077	115	57	58	79	0.98	1.0511
50	5	2	16.0311	129	65	64	47	0.94	1.2289
80	5		3.9138	146	51	95	73	0.91	1.3335
				$b = 0.25$, $\sigma_L = 150$					
50	4	1	0.0371	115	57	58	48	0.96	0.9608
80	4		0.0090	115	57	58	73	0.91	1.1816
50	5	2	21.1763	129	51	78	48	0.96	1.1538
80	5		5.1700	129	57	50	72	0.90	1.3340

Table 3.11. Analysis of lifetimes of steel components in stress data using purely sequential procedure (3.4.1) with $a = 1,\ c = 0.1$

n^*	m	k	ω	$\hat{\mu}_1$: $x_{n:1}$	$\hat{\mu}_2$: $y_{n:1}$	$\widehat{\Delta}$	n	n/n^*	z	ε in (3.6.2)
					$b = 0.33$					
50	5	1	0.0316	115	57	58	51	1.02	0.9603	-0.0101
80	5		0.0077	140	51	89	78	0.97	0.9796	-0.0063
50	7	2	16.0311	129	51	78	49	0.98	1.0568	0.0159
80	7		3.9138	143	57	86	79	0.98	1.0233	0.0099
					$b = 0.25$					
50	5	1	0.0371	129	51	78	51	1.02	0.9824	-0.0090
80	5		0.0090	115	59	56	79	0.98	0.9912	-0.0056
50	7	2	21.1763	146	51	95	49	0.98	1.0478	0.0170
80	7		5.1700	129	51	78	81	1.01	1.0321	0.0106

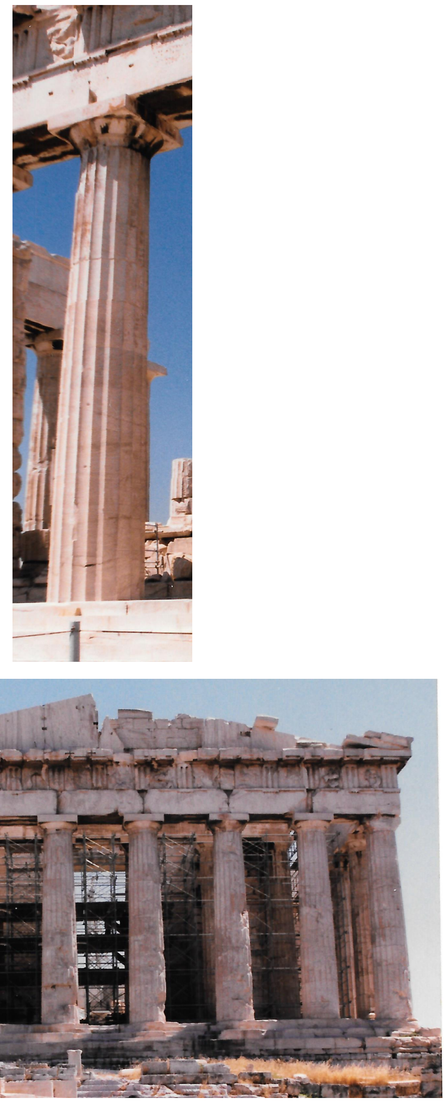
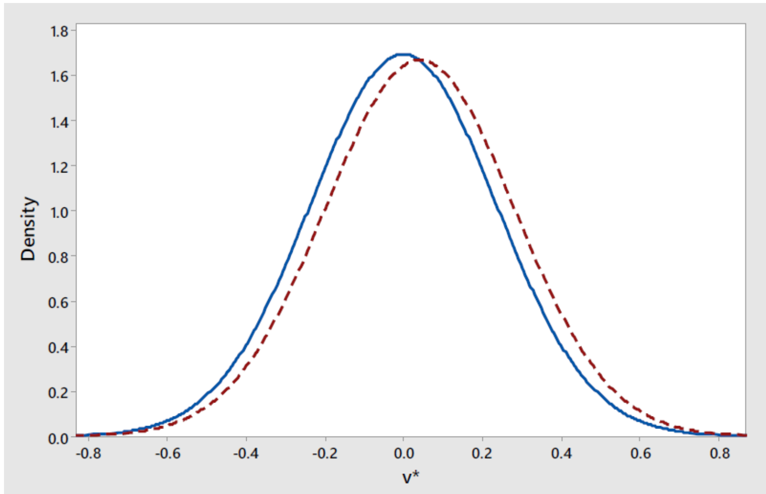
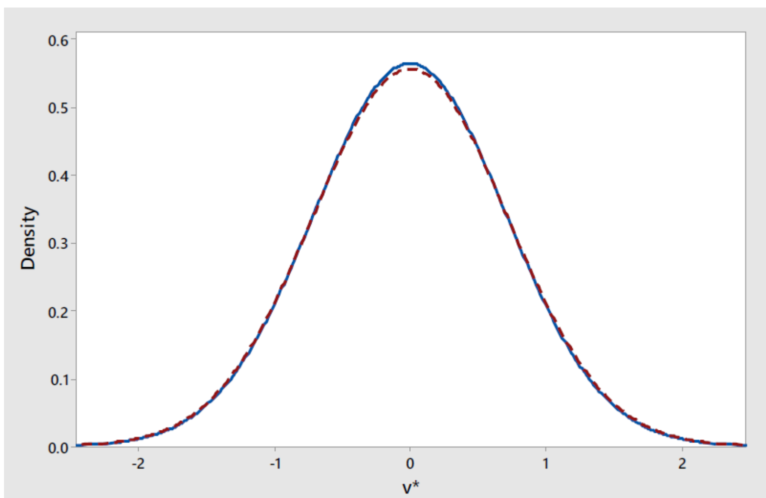


Figure 3.1. Columns from Parthenon, Greece: (a) One column by itself (U), (b) A cluster of eight identical columns (V).

Photo courtesy: Nitis Mukhopadhyay

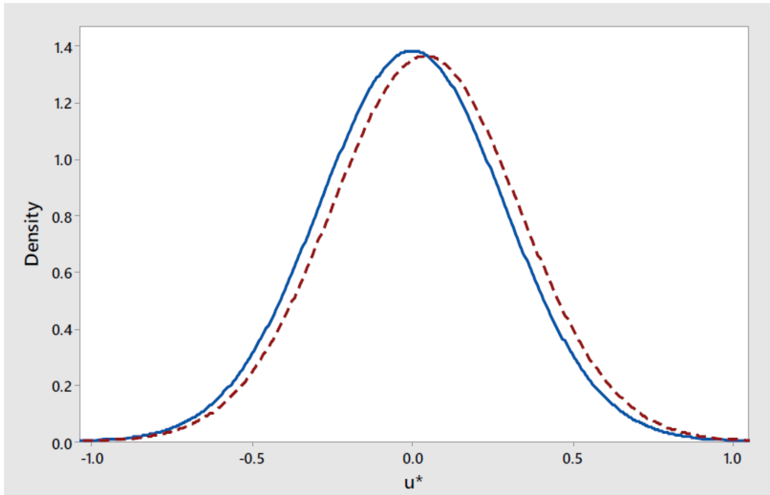


(a)

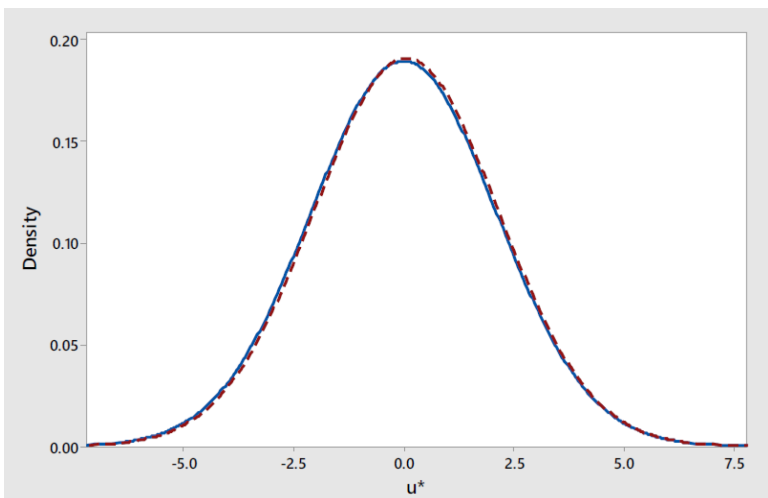


(b)

Figure 3.2. Plots of normality curves for the modified two-stage procedure (3.3.1)-(3.3.2) as validations for (3.3.7). The dashed curve (red) and the solid curve (blue) respectively correspond to the empirical distribution of the standardized sample size and the $N(0, \sigma_0^2)$ distribution with $\sigma_0^2 = \frac{1}{18}(\frac{\sigma}{\sigma_L})^{k/3}k^2$ coming from (3.3.5): **(a)** $\sigma_L = 3$, $b = 1$, $k = 1$, $n^* = 100$; **(b)** $\sigma_L = 3$, $b = 2$, $k = 3$, $n^* = 500$.

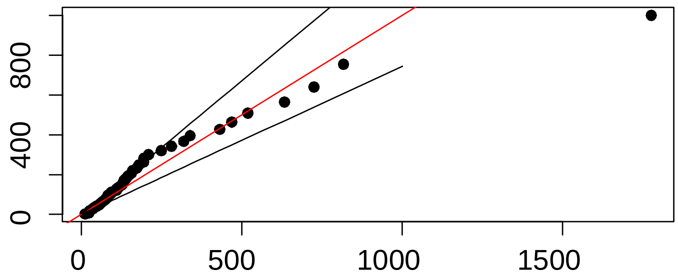


(a)

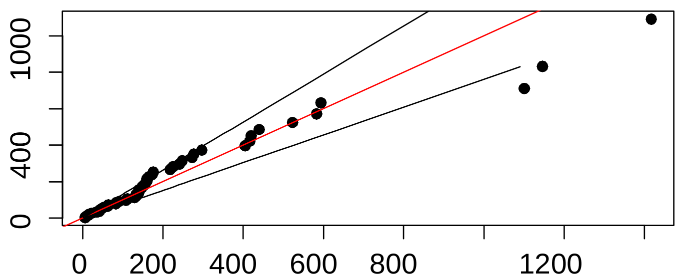


(b)

Figure 3.3. Plots of normality curves for the purely sequential procedure (3.4.1) as validation of Theorem 3.4.1, part (v). The dashed curve (red) and the solid curve (blue) respectively correspond to the empirical distribution of the standardized sample size and the $N(0, \sigma_1^2)$ distribution with $\sigma_1^2 = \frac{1}{18}k^2$ coming from Theorem 3.4.1, part (v): **(a)** $b = 3,\ k = 1,\ m = 5,\ n^* = 100$; **(b)** $b = 1,\ k = 3,\ m = 13,\ n^* = 500$.



(a)

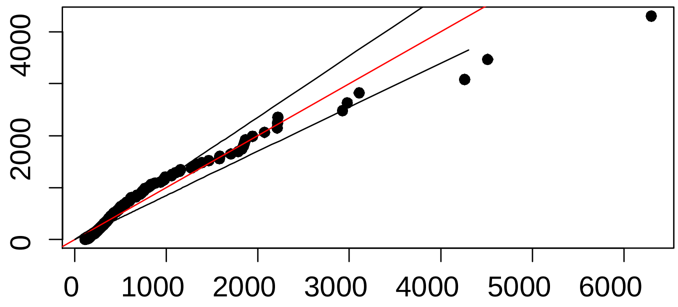


(b)

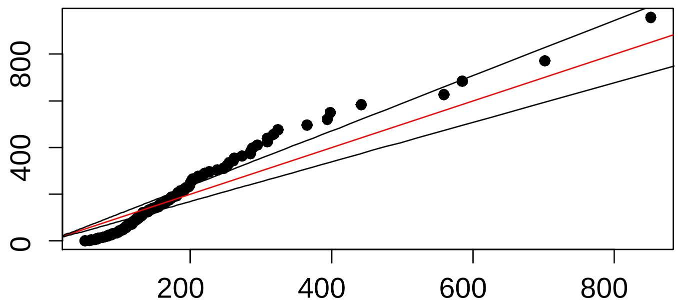
Figure 3.4. Exponential Q-Q plots for cancer data:

(a) Radiotherapy and chemo (X)

(b) Radiotherapy alone (Y)



(a)



(b)

Figure 3.5. Exponential Q-Q plots for stress data:

(a) Low stress (X)

(b) High stress (Y).

Chapter 4
Purely Sequential Bounded-Risk Point Estimation of the Negative Binomial Mean Under Various Loss Functions: One-Sample and Multi-Sample Problems

4.1. INTRODUCTION

In this chapter, we develop only purely sequential methodologies for estimating the mean of a *negative binomial* (NB) distribution under various forms of the loss functions that we have included. We will cover both one-sample and multi-sample situations. The contents of this chapter are based on Mukhopadhyay and Bapat (2017a) and (2017b).

A NB distribution has been used widely to model over-dispersed count data arising from studies in ecology and agriculture. Anscombe (1949,1950) emphasized the role of NB modeling in the case of insect count data often found in entomological research.

We will work with a NB distribution parametrized by Anscombe (1949) where one assumes the following *probability mass function* (p.m.f.):

$$f(x;\mu,\tau) \equiv P_{\mu,\tau}(X=x) = \left(1+\tfrac{\mu}{\tau}\right)^{-\tau} \tfrac{\Gamma(\tau+x)}{x!\Gamma(\tau)} \left(\tfrac{\mu}{\mu+\tau}\right)^{x}, x=0,1,2,... \tag{4.1.1}$$

This probability model is referred to as a NB distribution involving two parameters abbreviated as NB(μ,τ) where $0<\mu,\tau<\infty$. This notation pretends that μ,τ are both unknown. If τ is known, then we will interpret $f(x;\mu,\tau)$ as $f(x;\mu) \equiv P_\mu(X=x)$. In what follows, we use the notation I_A or $I(A)$ interchangeably to denote the indicator function of an event or a set A.

Both parameters μ,τ are assumed finite and positive with μ unknown, but τ may or may not be known. Here, for example, the measured response variable X may stand for the count of insects on plants or count of a particular variety of weed in an agricultural plot.

In such examples, the parameter μ used in (4.1.1) is interpreted as the average insect count or the average number of weed per sampling unit, whereas τ indicates the degree of clumping or thatching of infestation per sampling unit. The mean and variance for the distribution (4.1.1) are given by:

$$E_{\mu,\tau}[X] \equiv \mu \quad \text{and} \quad V_{\mu,\tau}[X] \equiv \sigma^2 = \mu + \tau^{-1}\mu^2. \tag{4.1.2}$$

Again, this notation pretends that μ, τ are both unknown. If τ is known, then we will interpret $E_{\mu,\tau}[.]$ and $V_{\mu,\tau}[.]$ as $E_{\mu}[.]$ and $V_{\mu}[.]$ respectively.

4.1.1. Brief Review: One-Sample and Multi-Sample Problems

Some selected references interfacing a NB model and sequential and/or multistage sampling strategies in agriculture and biology include: Bliss and Owen (1958), Kuno (1969,1972), Barrigossi (1997), Nyrop and Binns (1991), Mulekar and Young (1993,2004), Mukhopadhyay (2002), Mukhopadhyay and de Silva (2005), Mukhopadhyay and Banerjee (2014,2015), and Banerjee and Mukhopadhyay (2016). For a general overview in the broad area of sequential and multi-stage inference and methodologies, one may refer to Sen (1981), Woodroofe (1982), Siegmund (1985), Mukhopadhyay and Solanky (1994), Ghosh et al. (1997), Mukhopadhyay et al. (2004), Mukhopadhyay and de Silva (2009), Zacks (2009), and other sources.

Willson and Folks (1983) and Willson et al. (1984) contained a wide variety of sequential problems arising from estimation of a NB mean. One may refer to Mukhopadhyay and Diaz (1985) for an overview of a two-stage point estimation problem. Mukhopadhyay and Banerjee (2014,2015) included an extensive set of literature review addressing sequential problems on estimating the mean of a NB population. A majority of sources had dealt with problems of sequential fixed-width or fixed-accuracy confidence intervals, point estimation under a *squared error loss* (SEL), or tests for μ. Under a Bayesian perspective to sequential estimation of a NB mean, Marcus and Talpaz (1985) introduced a one-step look ahead procedure.

One of the loss functions considered in this chapter is the customary Linex loss introduced by Varian (1975), which was discussed widely in Chapters 2 and 3. In order to reiterate, a Linex loss in estimating a generic parameter θ with $\widehat{\theta}_n$ is defined as follows:

$$L_n \equiv L_n\left(\widehat{\theta}_n, \theta\right) = \exp\left\{a(\widehat{\theta}_n - \theta)\right\} - a(\widehat{\theta}_n - \theta) - 1, a \in R. \tag{4.1.3}$$

One may refer to Chapters 2 and 3 for further practical applications under appropriate modifications of this loss function (4.1.3) under negative exponential models. One may also refer to

Zellner (1986) and Chattopadhyay (1998,2000) to gain a broader perspective. Xiao et al. (2005) constructed a minimax confidence bound for a normal mean and Baran and Magiera (2010) dealt with a generalized form of the Linex loss used in sequential estimation of a function of probability of success under a Bernoulli distribution.

We emphasize that we have embarked upon developing only purely sequential methodologies for bounded-risk point estimation.

4.1.2. Layout of This Chapter

Section 4.2 develops an estimation strategy under a slightly modified Linex loss (4.2.2) which is a close variant of (4.1.3). This modification takes into account the CV approach as introduced in Willson and Folks (1983) and further developed by Mukhopadhyay and Diaz (1985). Here, thatch parameter τ was assumed known.

Section 4.3 develops an estimation strategy under SEL (4.3.1) assuming that the thatch parameter τ is known.

Section 4.4 is on a different note in the sense that the thatch parameter τ is now assumed unknown. It deals with estimation of μ under SEL (4.5.1).

Section 4.5 develops a simultaneous estimation strategy for a k sample problem under a modified Linex loss (4.6.2), again utilizing the CV approach. Here, thatch parameters τ_i, $i = 1, 2, ..., k$ are assumed known.

Section 4.6 develops a two-sample estimation strategy under SEL (4.7.1) assuming that the thatch parameters τ_1, τ_2 are known.

Section 4.7 is on a different note in the sense that the thatch parameters τ_1, τ_2 are now assumed unknown. It deals with estimation of $\mu_1 - \mu_2$ under SEL (4.9.1).

Section 4.8 provides brief conclusions.

Section 4.2 through Section 4.7 present summaries obtained from extensive sets of simulation studies based on a variety of parameter configurations highlighting both small and moderate sample-size performances of the proposed estimation strategies with associated risk analyses. They are followed by appropriate sets of illustrations obtained from implementations of each proposed estimation strategy using real data from statistical ecology. We have emphasized (i) weed count data of different species from a field in Netherlands (ii) count data of migrating woodlarks at the Hanko bird sanctuary in Finland and (iii) raptor count data at the Hawk mountain sanctuary, Pennsylvania.

4.2. LINEX LOSS UNDER CV APPROACH AND KNOWN THATCH PARAMETER

In this section, we introduce an appropriate modification to the conventional form of the Linex loss function shown in (4.1.3). We assume that we have available a sequence $\{X_1, X_2, ..., X_n, ...\}$ of *independent and identically distributed* (i.i.d.) random variables from a NB(μ, τ) population. Then, we develop a purely sequential bounded-risk methodology to estimate the NB mean μ under this modified Linex loss (4.2.2) when τ is assumed known.

4.2.1. A Modified Linex Loss

We revisit the *coefficient of variation* (CV) approach and originated by Willson and Folks (1983) which was further developed by Mukhopadhyay and Diaz (1985) and Mukhopadhyay and de Silva (2005). The customary Linex loss from (4.1.3) may be mildly modified with a *generic* notation by taking the CV approach into account. We may let:

$$W_n^* \equiv W_n^*\left(\widehat{\theta}_n, \theta\right) = \exp\left\{\frac{a(\widehat{\theta}_n - \theta)}{\theta}\right\} - \frac{a(\widehat{\theta}_n - \theta)}{\theta} - 1, a \in R, \tag{4.2.1}$$

pretending that the unknown θ parameter is non-zero.

Having recorded $X_1, ..., X_n$, we denote the sample mean $\overline{X}_n = n^{-1}\Sigma_{i=1}^n X_i$ which estimates μ and in the light of (4.2.1), we propose a modified Linex loss function for the point estimation problem on hand as follows and define:

$$L_n \equiv L_n\left(\overline{X}_n, \mu\right) = \exp\left\{\frac{a(\overline{X}_n - \mu)}{\mu}\right\} - \frac{a(\overline{X}_n - \mu)}{\mu} - 1,\ a \in R. \tag{4.2.2}$$

Next, we express the risk function as follows:

$$
\begin{aligned}
E_\mu[L_n] &= E_\mu\left[\exp\left\{\tfrac{a(\overline{X}_n-\mu)}{\mu}\right\}\right] - E_\mu\left[a\left(\tfrac{\overline{X}_n-\mu}{\mu}\right)\right] - 1 \\
&= e^{-a}\left[1+\tfrac{\mu}{\tau}(1-e^t)\right]^{-n\tau} - 1, \text{ using m.g.f. of } \overline{X}_n \qquad (4.2.3) \\
&= \exp\left(\tfrac{a^2}{2n\mu}\right) + \tfrac{a^2}{2n\tau} + o\left(\tfrac{1}{n}\right) - 1.
\end{aligned}
$$

Thus, from (4.2.3), the risk associated with the modified Linex loss function (4.2.2) reduces to:

$$R_n \equiv E_{\boldsymbol{\mu}}[L_n] = \tfrac{a^2}{2n}\left(\tau^{-1}+\mu^{-1}\right) + o(n^{-1}). \qquad (4.2.4)$$

4.2.2. A Sequential Bounded Risk Estimation

The idea here is to bound the risk R_n given in (4.2.4) from above by a suitable constant, namely the risk bound, $\omega(>0)$ and require that $R_n \leq \omega$ for all μ. This leads us to obtain the optimal fixed sample size n^* approximately as follows:

$$n \geq \tfrac{a^2}{2\omega}\left\{\tau^{-1}+\mu^{-1}\right\} = \tfrac{a^2}{2\omega}\tfrac{\sigma^2}{\mu^2} = n^*, \text{ say.} \qquad (4.2.5)$$

The magnitude of n^* remains unknown even though its expression is given by (4.2.5). Hence, we resort to developing a purely sequential bounded risk estimation strategy next. Recall that the thatch parameter τ is assumed known.

Now, Looking back at the expression of n^* from (4.2.5), we see clearly that

$$n^* > \tfrac{a^2}{2\omega\tau}, \qquad (4.2.6)$$

and hence, we determine the pilot size m as follows: Let

$$m \equiv m(\omega) = \left\lfloor \tfrac{a^2}{2\omega\tau} \right\rfloor + 1, \qquad (4.2.7)$$

where $\lfloor u \rfloor$ denotes the largest integer $<$ than $u(>0)$ and we first gather pilot data $X_i, i = 1, ..., m$. After pilot data, we gather one additional observation at-a-time, as needed, according to the stopping rule that is defined next.

Since $\overline{X}_n$ may be zero with a positive probability, whatever be n, we fix a number $\gamma(>\frac{1}{2})$ and define:

$$N = \inf \left\{ n \geq m : n \geq \tfrac{a^2}{2\omega} \left[\tau^{-1} + \left(\overline{X}_n + n^{-\gamma} \right)^{-1} \right] \right\}. \tag{4.2.8}$$

Here, $\left(\overline{X}_n + n^{-\gamma} \right)$ is treated as an estimator of μ which ensures that this estimator is positive with (P_μ) probability one and that the classical central limit theorem (CLT) will remain in effect for $n^{1/2} \left(\overline{X}_n + n^{-\gamma} - \mu \right)$ since $\gamma > \frac{1}{2}$.

At termination, based on the fully gathered data $\{N, X_1, ..., X_{m,} ..., X_N\}$, we propose to estimate μ by the sample mean $\overline{X}_N$. Now, we prove a lemma which will be useful in the sequel in proving the risk efficiency property for the estimation strategy $(N, \overline{X}_N)$.

Lemma 4.2.1. *For the estimation strategy $(N, \overline{X}_N)$ defined via (4.2.8), for each fixed $\mu \in R^+$, $\tau \in R^+$, $\gamma > \frac{1}{2}$, and $s \in R^+$, we have as $\omega \rightarrow 0$:*

$$E_\mu \left[\left| \overline{X}_N - \mu \right|^s \right] \rightarrow 0.$$

Proof of Lemma 4.2.1: Consider s fixed. We begin with the following inequality:

$$0 \leq \overline{X}_N \leq \sup_{n \geq 1} \overline{X}_n = W, \text{ say,} \tag{4.2.9}$$

where certainly all positive powers of W are integrable in view of Wiener's (1939) ergodic theorem because all positive moments of the X's are finite. Obviously, $\overline{X}_N \rightarrow \mu$ in probability (P_μ) as $\omega \rightarrow 0$. Hence, we can claim:

$$E_\mu \left[\overline{X}_N^k \right] \rightarrow \mu^k, \tag{4.2.10}$$

since $\overline{X}_N^k$ is uniformly integrable in view of (4.2.9), for all $k > 0$.

Now, for a fixed positive integer $r (> s)$, we apply Jensen's inequality to write:

$$E_\mu \left[\left| \overline{X}_N - \mu \right|^r \right] = \mu^r E_\mu \left[\left| \tfrac{\overline{X}_N}{\mu} - 1 \right|^r \right] \leq \mu^r E_\mu^{1/2} \left[\left| \tfrac{\overline{X}_N}{\mu} - 1 \right|^{2r} \right]. \tag{4.2.11}$$

But, since $2r$ is an even positive integer, we can express:

$$E_\mu \left[\left| \tfrac{\overline{X}_N}{\mu} - 1 \right|^{2r} \right] = \Sigma_{i=0}^{2r} (-1)^i \binom{2r}{i} E_\mu \left[\left(\tfrac{\overline{X}_N}{\mu} \right)^{2r-i} \right] < \infty, \tag{4.2.12}$$

in view of binomial theorem and (4.2.10).

Combining (4.2.11)-(4.2.12), we can immediately claim uniform integrability of $\left|\overline{X}_N - \mu\right|^s$ which completes the proof. ■

We now move forward to establish a number of attractive first-order asymptotic properties in Theorem 4.2.1 for the proposed purely sequential estimation strategy $(N, \overline{X}_N)$. Their proofs are outlined in Section 4.2.3.

Theorem 4.2.1. *With loss function L_N, pilot size m, and terminal sample size N defined in (4.2.2), (4.2.7) and (4.2.8) respectively, under the purely sequential estimation rule $(N, \overline{X}_N)$ from (4.2.8), for each fixed $\mu \in R^+$* and *$\tau \in R^+$ we have as $\omega \to 0$:*

(i) $N/n^* \xrightarrow{P_\mu} 1$ *if* $\gamma > \frac{1}{2}$;

(ii) $E_\mu\left[(N/n^*)^s\right] \to 1$ *for all* s, *if* $\gamma > \frac{1}{2}(\text{or } > 1)$ *when* $s < (\text{or } >)0$

[*asymptotic first-order efficiency*];

(iii) $E_\mu\left[L_N\right]/\omega \to 1$ *if* $\gamma > 1$ [*asymptotic risk efficiency*];

where n^* comes from (2.5) and τ is assumed known.

Part (ii) shows that the sequential methodology (4.2.8) is efficient along the lines of Chow and Robbins (1965) and first-order efficient in the sense of Ghosh and Mukhopadhyay (1981). That is, we may expect the terminal sample size N to hover around the optimal fixed sample size n^* when n^* is large. Part (iii) shows that the achieved risk $E_\mu\left[L_N\right]$ may be expected to hover around the preassigned risk-bound ω when n^* is large.

4.2.3. Proof of Theorem 4.2.1

Part (i):

From (4.2.8), we get the following inequality:

$$\begin{aligned}\tfrac{a^2}{2\omega}\left\{\left(\overline{X}_N + N^{-\gamma}\right)^{-1} + \tau^{-1}\right\} \le N \le \tfrac{a^2}{2\omega}\left\{\left(\overline{X}_{N-1} + (N-1)^{-\gamma}\right)^{-1} + \tau^{-1}\right\} I(N > m) \\ + (m-1)I(N = m) + 1 \quad \text{w.p.1}(P_\mu).\end{aligned} \tag{4.2.13}$$

Dividing (4.2.13) throughout by n^* we get:

$$\left\{\left(\overline{X}_N + N^{-\gamma}\right)^{-1} + \tau^{-1}\right\}(\mu^{-1}+\tau^{-1})^{-1} \le N/n^* \le \left\{\left(\overline{X}_{N-1} + (N-1)^{-\gamma}\right)^{-1} + \tau^{-1}\right\}$$
$$\times(\mu^{-1}+\tau^{-1})^{-1} + (m-1)n^{*-1}I(N=m) + \tfrac{2\omega}{a^2}(\mu^{-1}+\tau^{-1})^{-1} \ \text{ w.p.1}(P_\mu). \tag{4.2.14}$$

Next, taking limits on all sides of (4.2.14) and noting the following facts: $N \to \infty$ w.p.1(P_μ), $\overline{X}_N \to \mu$ w.p.1(P_μ), $m/n^* = O(1)$, and $P_\mu(N = m) \to 0$ as $\omega \to 0$, completes the proof of Part (i).

Part (ii):

Case 1: $s < 0$.

From (4.2.8) we get the following inequality w.p.1(P_μ):

$$\frac{N}{n^*} \ge \frac{1}{n^*}\frac{a^2}{2\omega}\left(\frac{1}{\overline{X}_N+N^{-\gamma}}\right) = \frac{\mu^2}{\sigma^2}\left(\frac{1}{\overline{X}_N+N^{-\gamma}}\right)$$
$$\Rightarrow \frac{n^*}{N} \le \frac{\sigma^2}{\mu^2}\left(\overline{X}_N + N^{-\gamma}\right) \le \frac{\sigma^2}{\mu^2}\sup_{n\ge1}\left(\overline{X}_n + 1\right) = W, \text{ say}$$
$$\Rightarrow \left(\frac{N}{n^*}\right)^s \le W^s, \text{ for } s<0.$$

But, again W^s is clearly integrable in view of Wiener's (1939) ergodic theorem. Thus, using part (i), we conclude:

$$E_\mu\left[\left(\frac{N}{n^*}\right)^s\right] \to 1 \text{ as } \omega \to 0 \text{ when } s<0. \tag{4.2.15}$$

Note that $\gamma > \frac{1}{2}$ suffices when $s < 0$.

Case 2: $s > 0$.

This part follows along the lines of Willson and Folks (1983) who improvised upon some of the original techniques from Mukhopadhyay (1974). Accordingly, we first fix some arbitrarily small $\epsilon > 0$ and define $\beta = (1+\epsilon)^{1/s}n^*$. We tacitly disregard that β may not be an integer. Then, we may write:

$$\begin{aligned} E_\mu[N^s] &= \Sigma_{n=m}^{\infty} n^s P_\mu(N=n) \\ &\le \Sigma_{n=m}^{\beta+1}(\beta+1)^s n^{*s} P_\mu(N=n) + \Sigma_{n>\beta+1}^{\infty} n^s P_\mu(N=n) \\ &\le (\beta+1)^s n^{*s} P_\mu(N \le \beta+1) + T(\beta), \text{ say, .} \end{aligned}$$

where $T(\beta) = \Sigma_{n>\beta+1}^{\infty} n^s P_\mu(N = n)$. Further,

$$E_\mu\left[\left(\frac{N}{n^*}\right)^s\right] \leq (\beta+1)^s P_\mu(N \leq \beta+1) + n^{*-s}T(\beta).$$

Let $[N = n]$ denote the even $N = n$. Now, from (4.2.8), we note the fact that

$$[N = n] \subset \left\{n - 1 < \frac{a^2}{2\omega}\left(\frac{1}{\tau} + \frac{1}{\overline{X}_{n-1} + (n-1)^{-\gamma}}\right)\right\},$$

so that we may express:

$$[N = n] \Rightarrow \Sigma_{i=1}^{n-1} X_i < \frac{a^2\tau(n-1)}{2\omega\tau(n-1) - a^2} - (n-1)^{-\gamma+1}.$$

Obviously, $(n-1)^{-\gamma+1}$ ought to be negligible for large n which will require us to assume that $\gamma > 1$.

That is, we may rewrite:

$$n \geq \beta \Rightarrow \frac{a^2\tau n}{2\omega\tau n - a^2} - n^{-\gamma+1} \leq \frac{\tau n}{\left(\frac{\sigma}{\mu}\right)^2(1+\epsilon)\tau - 1} - n^{-\gamma+1} = a(n), \text{ say.}$$

Hence, we obtain:

$$\begin{aligned} T(\beta) &= \Sigma_{n>\beta+1}^{\infty} n^s P_\mu(N = n) \\ &\leq \Sigma_{n>\beta+1}^{\infty}(n+1)^s P_\mu\left\{\Sigma_{i=1}^{n} X_i < a(n)\right\} \\ &\leq \Sigma_{n>\beta+1}^{\infty}(n+1)^s \inf_{t<0} P_\mu\left\{e^{-ta(n)}e^{t\Sigma_{i=1}^{n}X_i} > 1\right\}. \end{aligned}$$

Next, using the moment generating function of $\Sigma_{i=1}^{n} X_i$, which is distributed as $NB(n\mu,\, n\tau)$, we can show that $T(\beta) \leq \Sigma_{n=1}^{\infty} d_n$ where $d_n > 0$ and $d_n^{1/n} \to d$, $0 < d < 1$, as $n \to \infty$. Hence, we can claim:

$$\limsup_{\omega\to 0} E_\mu\left[(N/n^*)^s\right] \leq 1 + \epsilon, \tag{4.2.16}$$

but $\epsilon(>0)$ is arbitrary. Also, from part (i) and Fatou's lemma, we have:

$$\liminf_{\omega\to 0} E_{\mu,\tau}\left[(N/n^*)^s\right] \geq E_{\mu,\tau}\left[\liminf_{\omega\to 0}(N/n^*)^s\right] = 1. \tag{4.2.17}$$

We omit further details for brevity. Now, combining (4.2.15)-(4.2.17) completes the proof of

Part (ii).

Part (iii):

We will improvise upon some of the techniques developed recently by Mukhopadhyay and Zacks (2017). For a clear understanding, we will split the proof of part (iii) into a number of (main) steps as follows:

Step 1:

From (4.2.2), we may express $E_\mu[L_N]/\omega$ as:

$$\begin{aligned}
&\omega^{-1}E_\mu\left[\exp\left\{\tfrac{a(\overline{X}_N-\mu)}{\mu}\right\} - a\left(\tfrac{\overline{X}_N-\mu}{\mu}\right) - 1\right] \\
&\quad = \omega^{-1}E_\mu\left[\tfrac{a^2}{2}\left(\tfrac{\overline{X}_N-\mu}{\mu}\right)^2 + \tfrac{a^3}{6}\left(\tfrac{\overline{X}_N-\mu}{\mu}\right)^3 e^{\frac{a}{\mu}\xi_N}\right] \\
&\qquad\qquad \text{where } \xi_N \text{ is a random variable between 0 and } (\overline{X}_N-\mu),
\end{aligned}$$

so that we have:

$$\begin{aligned}
&\omega^{-1}E_\mu[L_N] \\
&\quad = \tfrac{n^*}{\sigma^2}E_\mu\left[(\overline{X}_N-\mu)^2\right] + \tfrac{a}{3\mu}E_\mu\left[\tfrac{n^*}{\sigma^2}(\overline{X}_N-\mu)^3 e^{\frac{a}{\mu}\xi_N}\right] \\
&\quad = \tfrac{n^*}{\sigma^2}E_\mu[I_1] + \tfrac{a}{3\mu\sigma^2}n^*E_\mu[I_2], \text{ say }.
\end{aligned} \tag{4.2.18}$$

Here, n^* comes from (4.2.5).

Step 2:

We first address the term $n^*E_\mu[I_2]$ from (4.2.18) and show that it is $o(1)$. From Anscombe's (1952) *random central limit theorem*, with $W_N = \Sigma_{i=1}^N X_i$, we can claim:

$$U \equiv U_N = \tfrac{W_N - N\mu}{\sigma\sqrt{n^*}} \xrightarrow{\pounds} N(0,1), \text{ as } \omega \to 0. \tag{4.2.19}$$

The term n^*I_2 is rewritten as follows (w.p.1):

$$\begin{aligned}
n^*I_2 &= n^*(\overline{X}_N-\mu)^3 e^{\frac{a}{\mu}\xi_N} = \tfrac{n^*}{N^3}(W_N - N\mu)^3 e^{\frac{a}{\mu}\xi_N} = \tfrac{n^{*5/2}}{N^3}\sigma^3 U^3 e^{\frac{a}{\mu}\xi_N} \\
&= \sigma^3 e^{\frac{a}{\mu}\xi_N}\left(\tfrac{n^*}{N}\right)^3 n^{*-1/2}U^3, \text{ with } U \text{ from (2.19)}.
\end{aligned} \tag{4.2.20}$$

Case 1: On the set where $\frac{a}{\mu}\xi_N < 0$.

Note that (4.2.20) gives w.p.1(P_μ):

$$n^* |I_2| I\left(\tfrac{a}{\mu}\xi_N < 0\right) \le \sigma^3 \left(\tfrac{n^*}{N}\right)^3 n^{*-1/2} |U|^3. \tag{4.2.21}$$

From Theorem 2 of Chow et al. (1979), we can claim: $|U|^s$ is uniformly integrable for all $s > 0$ when $\gamma > 1$. Also, from Theorem 4.2.1, part (ii), it follows that $(n^*/N)^s$ is uniformly integrable for fixed $s > 0$ when $\gamma > \frac{1}{2}$. But, using Cauchy-Schwartz inequality, we can claim:

$$E_\mu \left[\left|\left(\tfrac{n^*}{N}\right)^3 U^3\right|\right] \le E_\mu^{1/2}\left[\left(\tfrac{n^*}{N}\right)^6\right] E_\mu^{1/2}\left[\left|U^6\right|\right] = O(1),$$

which combined with (4.2.21) shows that

$$E_\mu \left[n^* |I_2| I\left(\tfrac{a}{\mu}\xi_N < 0\right)\right] = o(1), \tag{4.2.22}$$

when $\gamma > 1$.

Case 2: On the set where $\frac{a}{\mu}\xi_N \ge 0$

Again, (4.2.20) gives w.p.1(P_μ):

$$n^* |I_2| I\left(\tfrac{a}{\mu}\xi_N \ge 0\right) = \sigma^3 e^{\frac{a}{\mu}\xi_N} \left(\tfrac{n^*}{N}\right)^3 {n^*}^{-1/2} |U|^3 I\left(\tfrac{a}{\mu}\xi_N \ge 0\right), \tag{4.2.23}$$

and we will now show that

$$E_\mu \left[\left(\tfrac{n^*}{N}\right)^3 e^{\frac{a}{\mu}\xi_N} |U|^3 I\left(\tfrac{a}{\mu}\xi_N \ge 0\right)\right] = O(1).$$

Upon repeated uses of Holder's inequality to split the expectations of various terms within (4.2.23), with appropriate choices of $\alpha > 1, \beta > 1,\ \alpha^{-1} + \beta^{-1} = 1$ and $\alpha' > 1, \beta' > 1,\ \alpha'^{-1} + \beta'^{-1} = 1$, we claim:

$$E_\mu \left[\left(\tfrac{n^*}{N}\right)^3 e^{\frac{a}{\mu}\xi_N} |U|^3 I\left(\tfrac{a}{\mu}\xi_N \ge 0\right)\right] \le O(1) E_\mu^{1/\beta\alpha'}\left[e^{\frac{a}{\mu}\beta\alpha'\xi_N} I\left(\tfrac{a}{\mu}\xi_N \ge 0\right)\right], \tag{4.2.24}$$

when $\gamma > 1$.

We now verify that

$$E_\mu\left[e^{\lambda\xi_N} I\left(\frac{a}{\mu}\xi_N \geq 0\right)\right] = O(1),$$

where we see $\lambda = \frac{a}{\mu}\beta\alpha'$ from (4.2.24) and thus write w.p.1(P_μ):

$$e^{\lambda\xi_N} I\left(\frac{a}{\mu}\xi_N \geq 0\right) \leq e^{\left|\lambda\left(\overline{X}_N-\mu\right)\right|} = \Sigma_{s=0}^{\infty}\frac{1}{s!}\left|\lambda\left(\overline{X}_N-\mu\right)\right|^s, \tag{4.2.25}$$

with each term being positive and integrable. So, by applying the monotone convergence theorem, we have from (4.2.25):

$$E_\mu\left[e^{\lambda\xi_N} I\left(\frac{a}{\mu}\xi_N \geq 0\right)\right] \leq \Sigma_{s=0}^{\infty}\frac{1}{s!}E_\mu\left[\left|\lambda\left(\overline{X}_N-\mu\right)\right|^s\right]. \tag{4.2.26}$$

Then, after using Lemma 4.2.1, from (4.2.26), we can claim:

$$\lim_{\omega\to 0} E_\mu\left[e^{\lambda\xi_N} I\left(\frac{a}{\mu}\xi_N \geq 0\right)\right] = 0, \tag{4.2.27}$$

when $\gamma > 1$. A combination of (4.2.20), (4.2.22), and (4.2.27) leads us to conclude:

$$n^* E_\mu[I_2] = o(1), \tag{4.2.28}$$

when $\gamma > 1$.

Step 3:

Next, we go back to (4.2.18) and address the first term, namely $\frac{n^*}{\sigma^2}E_\mu[I_1]$, and show that it converges to 1 as $\omega \to 0$. We improvise the proof found in Mukhopadhyay (1978), a precursor of Ghosh and Mukhopadhyay (1979). We recall $U \equiv U_N = \frac{W_N - N\mu}{\sigma\sqrt{n^*}}$ from (4.2.19) with $W_N = \Sigma_{i=1}^N X_i$, and we rewrite $\frac{n^*}{\sigma^2}E_\mu[I_1]$ as:

$$\frac{n^*}{\sigma^2}E_\mu\left[(\overline{X}_N-\mu)^2\right] = E_\mu\left[U_N^2\right] + E_\mu\left[U_N^2\left\{\frac{n^{*2}}{N^2}-1\right\}\right] = E_\mu[I_{11}] + E_\mu[I_{12}], \text{ say.} \tag{4.2.29}$$

From (4.2.19), we know that $U_N^2 \xrightarrow{\pounds} \chi_1^2$ as $\omega \to 0$. Then, in view of Wald's second equation (Theorem 2.4.5, Ghosh et al. 1997; Theorem 3.5.5, Mukhopadhyay and de Silva 2009), we can claim:

$$E_\mu[I_{11}] = E_\mu[U_N^2] = E_\mu\left[\tfrac{N}{n^*}\right] \Rightarrow \lim_{\omega\to 0} E_\mu[I_{11}] = 1, \tag{4.2.30}$$

using Part (ii) when $\gamma > 1$.

Now,then, U_N^2 is also uniformly integrable. Also, in view of (4.2.7), we can claim that $\left|\frac{n^{*2}}{N^2} - 1\right|$ is bounded from above by $\max\{1, \tau^2\sigma^4\mu^{-4} - 1\}$. In other words,

$$|I_{12}| \le \max\{1, \tau^2\sigma^4\mu^{-4} - 1\} U_N^2,$$

so that $|I_{12}|$ is also uniformly integrable. However, $I_{12} \xrightarrow{P_\mu} 0$ as $\omega \to 0$, so that $E_\mu[I_{12}] = o(1)$.

Combining this with (4.2.29)-(4.2.30), we conclude that $\lim_{\omega\to 0} \frac{n^*}{\sigma^2} E_\mu[I_1] = 1$. This completes the proof of Part (iii) and Theorem 4.2.1 in view of (4.2.18). ■

4.2.4. Simulations and Real Data Illustrations

We begin with a summary (Section 4.2.4.1) from a set of simulation studies to examine the performances of our proposed purely sequential estimation strategy (4.2.8) for small and moderate values of n^*. Sections 4.2.4.2-4.2.4.3 highlight performances of our estimation strategy using real data from statistical ecology. We have emphasized (i) weed count data of different species from a field in Netherlands and (ii) count data of migrating woodlarks at the Hanko bird sanctuary in Finland.

4.2.4.1. A Summary from Simulations

We first generated pseudorandom observations from the distribution (4.1.1) with combinations of choices for μ and τ. We fixed the values $a = 1$ and $\gamma = 1.5$. We determined m from (4.2.7). Each row in Table 4.1 corresponds to averages from 10000 replications which were run under a given configuration. In order to represent varying sample sizes, we show results for fixed values of $n^* = 50$ (small), 200 (medium).

Since our stopping rule (4.2.8) is very similar to that of Willson and Folks (1983), we have also included a set of simulations based on the rule given by Willson and Folks (1983), however, we evaluated our risk function at termination. This is done for comparing the results obtained from the two sampling methods. The first block in Table 4.1 shows performances of our purely sequential method (4.2.8) whereas the second block shows performances of Willson and Folks's (1983) rule. The results are shown under the following setting:

$$\boxed{n^* = 50, 200 \text{ and } (\mu, \tau) = (2, 3), (3, 4)} \tag{4.2.31}$$

Each block in Table (4.1) shows n^* (column 3), ω (column 4), the estimated values $\overline{x}$ and $s_{\overline{x}}$ (column 5), estimated values $\overline{n}$, $s_{\overline{n}}$ (column 6), the ratio $\overline{n}/n^*$ (column 7) plus values of $\overline{z}$ and $s_{\overline{z}}$ (column 8) with

$$\begin{gathered} N = n_i, r_i = \tfrac{a^2}{2n_i}\left(\tfrac{1}{\mu} + \tfrac{1}{\tau}\right) \text{ as in (4.2.4) under the } i^{\text{th}} \text{ replication,} \\ \text{and } \overline{r} = H^{-1}\Sigma_{i=1}^{H} r_i,\ s_{\overline{r}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(r_i - \overline{r})^2}, \\ \text{so that } \overline{z} = \overline{r}/\omega,\ s_{\overline{z}} = s_{\overline{r}}/\omega,\ H = 10000. \end{gathered} \tag{4.2.32}$$

Under both methods, the $\overline{x}$ values are very close to the corresponding means in each case with very small standard error value $s_{\overline{x}}$. These become even closer for $n^* = 200$. The values of $\overline{n}$ seem to estimate n^* very accurately across the rows. We see that the $\overline{n}$ values overestimate n^* by a small margin. The last column shows that both sequential methodologies tend to provide a risk bound close (or smaller) to preset goal ω for small (moderate) n^*. For both small or moderate n^*, the methodologies of ours and Willson and Folks's are in sync by producing comparable results by taking into account the estimated values $\overline{x}, s_{\overline{x}}$ (column 5), estimated values $\overline{n}, s_{\overline{n}}$ (column 6), the ratio $\overline{n}/n^*$ (column 7), and $\overline{z}, s_{\overline{z}}$ (column 8) from (4.2.32).

4.2.4.2. Illustration 4.2.1: Weed Count Data

We now present our first real data Illustration 4.2.1. We resort to an ecological count dataset. We make use of the weed count data presented by Heijting (2013). They recorded weed count data from quadrats on part of an arable maize field in Wageningen, Netherlands, prior to herbicide application. We looked at data from the year 2001 between 18-21 June, and applied our methodology (4.2.8) on one of the species of weed, *Capsella bursa-pastoris L.* (Shephard's purse). The field was cultivated and sown in north-south direction and the observation area was divided into $16 \times 67 = 1072$ quadrats of 0.75×0.75 meters. One can find further information from:

http://dx.doi.org/10.17026/dans-zu9-r7y8

and Heijting et al. (2007).

Figure 4.1 shows the types of weed included in this dataset which consist of 1072 rows and a NB fit was seen as appropriate with a p-value of 0.84. We found $\widehat{\mu} = 0.3$ and $\widehat{\tau} = 3.98$ from full data. In other words, we considered this dataset as our population with unknown mean μ but with a known value 3.98 for τ. We emphasize that our implementation of the sampling strategy (4.2.8) did not exploit the number $\widehat{\mu}(= 0.3)$ obtained from full dataset.

We fixed $a = 1$ and $\gamma = 1.5$, implemented the methodology (4.2.8) by drawing observations from the full set of data without replacement. We found that with or without replacement made practically no difference.

Table 4.2 shows results from implementing the estimation strategy (4.2.8) in single runs with 4 different preset values of the risk-bound ω. The terminal estimated value of μ are overall close to the value of $\widehat{\mu} = 0.3$, obtained from full data. The n^* values (obtained by pretending that $\mu = 0.3$ and $\tau = 3.98$) are merely provided as a vehicle for ad-hoc comparison with the observed n values. We have not used these n^* values in our implementation. We note that the ratio n/n^* is reasonably close to 1 which is desired. In the last column of Table 4.2, we show a value $\widetilde{z}$ obtained from a single run:

$$\begin{gathered} N = n, \widetilde{r} = \tfrac{a^2}{2n}\left(\tfrac{1}{\overline{x}_n} + \tfrac{1}{\tau}\right) \text{ in the spirit of (4.2.4) under} \\ \text{one replication, so that } \widetilde{z} = \widetilde{r}/\omega. \end{gathered} \tag{4.2.33}$$

Again, column 6 shows $\widetilde{z}$ values mainly to grasp a sense of how close the estimated risk may or may not be when compared with the preset goal ω. The erratic behavior is due to the fact that $\widetilde{z}$ values were obtained from single runs.

4.2.4.3. Illustration 4.2.2: Count Data of Migrating Woodlarks at the Hanko Bird Sanctuary

We now present our next real data Illustration 4.2.2. We resort to an ecological count dataset on migrating woodlarks at Hanko bird sanctuary situated in southwestern Finland. This data has been used in Linden and Mantyniemi (2011) and is available in the *Ecological Archives* E092-120-S2. One should also refer to Supplement 2 in Linden and Mantyniemi (2011) for more details regarding this dataset.

We will use the migration data for Autumn season, 1 September - 10 November, 2009. We worked with the daily counts of migrating birds during first four hours of daylight after sunrise. The dataset included 71 rows and a NB fit was seen as appropriate with a p-value of 0.35. We found $\widehat{\mu} = 3.05$ and $\widehat{\tau} = 0.23$ from full data. In other words, we considered this dataset as our population with unknown μ but with known value 0.23 for τ. We emphasize that our implementation of the sampling strategy (4.2.8) did not exploit the number $\widehat{\mu}(= 3.05)$ obtained from full dataset. We picked $a = 1$ and $\gamma = 1.5$, implemented the methodology (4.2.8) by drawing observations from the full set of data without replacement. We found that with or without replacement made practically no difference.

Table 4.3 shows results from implementing the estimation strategy (4.2.8) in single runs with 3 different preset values of the risk-bound ω. The terminal estimated values of μ look bit erratic and these are not too close to the value $\widehat{\mu} = 3.05$, obtained from full data. The n^* values (obtained by pretending that $\mu = 3.05$ and $\tau = 0.23$) are merely provided as a vehicle for ad-hoc comparison with the observed n values. We have not used these n^* values in our implementation. We note that the sample sizes (Table 4.2.3) are small in the range of $30 - 50$, however the ratio n/n^* remains close to 1 which is desired. In the last column of Table 4.3, we show associated value $\widetilde{z}$ obtained along the line of (4.2.33) from a single run and observe that $\widetilde{z}$ values are smaller than one which indicates that the estimated risk is smaller than the preset goal ω. Again, we emphasize that each row in Table 4.3 summarizes a single run.

4.3. SQUARED ERROR LOSS APPROACH UNDER KNOWN THATCH PARAMETER

In this section, we develop a purely sequential estimation strategy for estimating the mean of a NB(μ, τ) population under a squared error loss function. We assume that the thatch parameter τ is known. Having recorded $X_1, ..., X_n$, recall that the sample mean $\overline{X}_n = n^{-1}\Sigma_{i=1}^{n} X_i$ estimates μ

and we propose a squared error loss function:

$$L_n \equiv L_n(\overline{X}_n, \mu) = b(\overline{X}_n - \mu)^2,\ b > 0. \tag{4.3.1}$$

We express the risk function as:

$$R_n \equiv E_\mu[L_n] = \tfrac{b}{n}\left(\mu + \tau^{-1}\mu^2\right). \tag{4.3.2}$$

4.3.1. A Sequential Bounded Risk Estimation

We will again bound the risk R_n given in (4.3.2) from above by $\omega(>0)$. This leads us to the optimal fixed sample size n^* approximately as follows:

$$n \geq \tfrac{b}{\omega}\left(\mu + \tau^{-1}\mu^2\right) = n^*,\ \text{say}. \tag{4.3.3}$$

The magnitude of n^* remains unknown even though its expression is given by (4.3.3). Hence, we resort to developing a purely sequential bounded risk estimation strategy next. We will continue to assume that the thatch parameter τ is known.

We first fix $m(\geq 1)$ and gather pilot data $X_i, i = 1, ..., m$ of size m from the NB population. Again, since $\overline{X}_n$ may be zero with a positive probability, whatever be n, we fix a number $\gamma(> \frac{1}{2})$ and define:

$$N = \inf\left\{n \geq m : n \geq \tfrac{b}{\omega}\left[\overline{X}_n + \tau^{-1}\overline{X}_n^2 + n^{-\gamma}\right]\right\}. \tag{4.3.4}$$

Here, $n^{-\gamma}$ ensures that the estimator is positive with probability one. Based on the fully gathered data $\{N, X_1, ..., X_{m,} ..., X_N\}$, we propose to estimate μ by the sample mean $\overline{X}_N$.

Theorem 4.3.1 gives a set of attractive first-order asymptotic properties for the proposed purely sequential estimation methodology $(N, \overline{X}_N)$ obtained from (4.3.4).

Theorem 4.3.1. *With loss function L_N and terminal sample size N defined in (4.3.1) and (4.3.4) respectively, under the purely sequential estimation rule $(N, \overline{X}_N)$ from (4.3.4), for each fixed $\mu \in R^+$* and *$\tau \in R^+$ we have as $\omega \to 0$:*

(*i*) $N/n^* \xrightarrow{P_\mu} 1$ *if* $\gamma > \frac{1}{2}$;

(*ii*) $E_\mu\left[(N/n^*)^s\right] \to 1$ *for all* $s > 0$, *if* $\gamma > \frac{1}{2}$ [*asymptotic first-order efficiency*];

(*iii*) $E_\mu\left[L_N\right]/\omega \to 1$ *if* $\gamma > 1$ [*asymptotic risk efficiency*];

where n^* *comes from* (*4.3.3*) *and* τ *is assumed known.*

Part (ii) shows that the sequential methodology (4.3.4) is efficient along the lines of Chow and Robbins (1965) and first-order efficient in the sense of Ghosh and Mukhopadhyay (1981). That is, we may expect N to hover around the optimal fixed sample size n^* when n^* is large. Part (iii) shows that the achieved risk $E_\mu\left[L_N\right]$ may be expected to hover around the preassigned risk-bound ω when n^* is large.

4.3.2. Proof of Theorem 4.3.1

Part (*i*):

From (4.3.4), we write down the following inequality:

$$\begin{aligned} \tfrac{b}{\omega}\left(\overline{X}_N + \tau^{-1}\overline{X}_N^2 + N^{-\gamma}\right) \le N \le \tfrac{b}{\omega}\left(\overline{X}_{N-1} + \tau^{-1}\overline{X}_{N-1}^2 + (N-1)^{-\gamma}\right) \\ + (m-1) + 1 \quad \text{w.p.1}(P_\mu). \end{aligned} \tag{4.3.5}$$

Dividing (4.3.5) throughout (4.3.5) by n^*, we get:

$$\begin{aligned} \left(\overline{X}_N + \tau^{-1}\overline{X}_N^2 + N^{-\gamma}\right)(\mu + \tau^{-1}\mu^2)^{-1} \le N/n^* \le \left(\overline{X}_{N-1} + \tau^{-1}\overline{X}_{N-1}^2 + (N-1)^{-\gamma}\right) \\ \times(\mu + \tau^{-1}\mu^2)^{-1} + (m-1)n^{*-1} + n^{*-1} \quad \text{w.p.1}(P_\mu). \end{aligned} \tag{4.3.6}$$

Taking limits on all sides of (4.3.6) and noting the following facts: $N \to \infty$ w.p.1(P_μ), $\overline{X}_N \to \mu$ w.p.1(P_μ), and $m/n^* = o(1)$ as $\omega \to 0$ completes the proof. Here, $\gamma > \frac{1}{2}$ suffices.

Part (*ii*):

From the right-hand side of (4.3.6), we have the following inequality (for sufficiently large n^*):

$$N/n^* \le \left(\overline{X}_{N-1} + \tau^{-1}\overline{X}_{N-1}^2 + (m-1)^{-\gamma}\right)(\mu + \tau^{-1}\mu^2)^{-1} + m. \tag{4.3.7}$$

Now, denoting $\sup_{n \geq 2} \left(\overline{X}_n + \tau^{-1} \overline{X}_n^2 \right)$ as W we can claim:

$$N/n^* \leq (\mu + \tau^{-1}\mu^2)^{-1}[W + 1] + m. \tag{4.3.8}$$

Thus, the right-hand side of (4.3.8) is free from ω and using Wiener's (1939) ergodic theorem we can claim the uniform integrability of all positive powers of N/n^*. Next, appealing to Part (i), we complete the proof. Here, $\gamma > \frac{1}{2}$ suffices.

Part (iii): In this proof which is split into a number of steps for clarity, we improvise on the techniques that were originally developed by Ghosh and Mukhopadhyay (1979) and then moved further along by Sen and Ghosh (1981). Throughout, we fix an arbitrary ϵ in $(0, 1)$.

Step 1:

We note:

$$E_\mu[L_N]/\omega = \tfrac{b}{\omega} E_\mu \left[(\overline{X}_N - \mu)^2 \right] = \tfrac{n^*}{\sigma^2} E_\mu \left[(\overline{X}_N - \mu)^2 \right].$$

Now, we need to verify that $\frac{n^*}{\sigma^2} E_\mu \left[\left(\overline{X}_N - \mu \right)^2 \right] \to 1$ as $\omega \to 0$ when $\gamma > 1$. This can be shown as follows:

Along the lines of (4.2.29), we get:

$$\tfrac{n^*}{\sigma^2} E_\mu \left[\left(\overline{X}_N - \mu \right)^2 \right] = E_\mu \left[U_N^2 \right] + E_\mu \left[U_N^2 \left\{ \tfrac{{n^*}^2}{N^2} - 1 \right\} \right] = E_\mu \left[I_{11} \right] + E_\mu \left[I_{12} \right], \text{ say.} \tag{4.3.9}$$

where

$$U_N = \tfrac{W_N - N\mu}{\sigma \sqrt{n^*}} \text{ with } W_N = \Sigma_{i=1}^N X_i \text{ and } N \text{ comes from (4.3.4).}$$

We can claim that

$$E_\mu \left[I_{11} \right] \to 1 \text{ as } \omega \to 0 \text{ when } \gamma > \tfrac{1}{2}, \tag{4.3.10}$$

in the spirit of (4.2.30). Thus, in view of (4.3.9), we have to verify:

$$E_\mu \left[I_{12} \right] \to 0 \text{ as } \omega \to 0. \tag{4.3.11}$$

Step 2:

On the set $|N-n^*| \le \epsilon n^*$, we can express:

$$
\begin{gathered}
-\epsilon \le \tfrac{N}{n^*} - 1 \le \epsilon \Rightarrow \tfrac{1}{1+\epsilon} \le \tfrac{n^*}{N} \le \tfrac{1}{1-\epsilon} \\
\Rightarrow \tfrac{-\epsilon(2+\epsilon)}{(1+\epsilon)^2} \le \left(\tfrac{{n^*}^2}{N^2} - 1\right) \le \tfrac{\epsilon(2-\epsilon)}{(1-\epsilon)^2} \Rightarrow \tfrac{-\epsilon(2+\epsilon)}{(1-\epsilon)^2} \le \tfrac{-\epsilon(2+\epsilon)}{(1+\epsilon)^2} \\
\le \left(\tfrac{{n^*}^2}{N^2} - 1\right) \le \tfrac{\epsilon(2-\epsilon)}{(1-\epsilon)^2} \le \tfrac{\epsilon(2+\epsilon)}{(1-\epsilon)^2},
\end{gathered}
$$

which shows:

$$\left|\left(\tfrac{n^*}{N}\right)^2 - 1\right| I_{[|N-n^*| \le \epsilon n^*]} \le \epsilon(2+\epsilon)(1-\epsilon)^{-2}. \tag{4.3.12}$$

Step 3:

We recall that $U_N \xrightarrow{\pounds} N(0,1)$ by Anscombe's (1952) random CLT so that $U_N^2 \xrightarrow{\pounds} \chi_1^2$ as $\omega \to 0$, and we showed that $E_\mu\left[U_N^2\right] \to 1$ as $\omega \to 0$. when $\gamma > \frac{1}{2}$. One may go back and see (4.2.30). Hence, U_N^2 is uniformly integrable. Then, in view of (4.3.9), we can claim:

$$U_N^2\left|\left(\tfrac{n^*}{N}\right)^2 - 1\right| I_{[|N-n^*| \le \epsilon n^*]} \le \epsilon(2+\epsilon)(1-\epsilon)^{-2} U_N^2,$$

that is, $U_N^2\left|\left(\frac{n^*}{N}\right)^2 - 1\right| I_{[|N-n^*| \le \epsilon n^*]}$ is also uniformly integrable when $\gamma > \frac{1}{2}$.

But, in view of Part (i), we have:

$$U_N^2\left|\left(\tfrac{n^*}{N}\right)^2 - 1\right| I_{[|N-n^*| \le \epsilon n^*]} \to 0 \text{ in probability}(P_\mu) \text{ as } \omega \to 0,$$

so that we have:

$$E_\mu\left[U_N^2\left|\left(\tfrac{n^*}{N}\right)^2 - 1\right| I_{[|N-n^*| \le \epsilon n^*]}\right] \to 0 \text{ as } \omega \to 0. \tag{4.3.13}$$

Step 4:

Next, on the set $N > n^*(1+\epsilon)$, we observe w.p. $1(P_\mu)$:

$$
\begin{gathered}
\tfrac{N}{n^*} - 1 > \epsilon \Rightarrow 0 < \left(\tfrac{n^*}{N}\right)^2 < \left(\tfrac{1}{1+\epsilon}\right)^2 < 1 < 2 \\
\Rightarrow \left|\left(\tfrac{n^*}{N}\right)^2 - 1\right| < 1,
\end{gathered}
$$

so that we can express:

$$U_N^2\left|\tfrac{{n^*}^2}{N^2} - 1\right| I_{[N-n^* > \epsilon n^*]} \le U_N^2,$$

w.p. $1(P_\mu)$

That is, $U_N^2 \left|\left(\frac{n^*}{N}\right)^2 - 1\right| I_{[N-n^*>\epsilon n^*]}$ is uniformly integrable when $\gamma > \frac{1}{2}$. But, in view of Part (i), we have:

$$U_N^2 \left|\left(\frac{n^*}{N}\right)^2 - 1\right| I_{[N-n^*>\epsilon n^*]} \to 0 \text{ in probability}(P_\mu) \text{ as } \omega \to 0,$$

so that we conclude:

$$E_\mu\left[U_N^2 \left|\left(\frac{n^*}{N}\right)^2 - 1\right| I_{[N-n^*>\epsilon n^*]}\right] \to 0 \text{ as } \omega \to 0, \tag{4.3.14}$$

since $P_\mu\{N > n^*(1+\epsilon)\} \to 0$ as $\omega \to 0$.

Step 5:

Now, we will proceed to show:

$$E_\mu\left[U_N^2 \left|\frac{{n^*}^2}{N^2} - 1\right| I_{[N-n^*<-\epsilon n^*]}\right] \to 0 \text{ as } \omega \to 0. \tag{4.3.15}$$

Let q be a generic positive constant that does not involve ω. From (4.3.4), we can claim that $N \geq \frac{b}{\omega} N^{-\gamma}$ w.p.1(P_μ) so that we immediately have $N \geq \left(\frac{b}{\omega}\right)^{1/(1+\gamma)} = O(n^{*1/(1+\gamma)})$ w.p.1(P_μ). This implies that $\left(\frac{n^*}{N}\right)^2 \leq q{n^*}^2 \left(\frac{1}{n^*}\right)^{2/(1+\gamma)} = qn^{*2\gamma/(1+\gamma)}$ w.p.1(P_μ) and then we can write:

$$\begin{aligned}
&\left|E_\mu\left[U_N^2 \left|\frac{{n^*}^2}{N^2} - 1\right| I_{[N-n^*<-\epsilon n^*]}\right]\right| \\
&\quad \leq qn^{*-1+(2\gamma/(1+\gamma))} E_\mu\left[(W_N - N\mu)^2 I_{[N-n^*<-\epsilon n^*]}\right] \\
&\quad = qn^{*(2\gamma-1)/(1+\gamma)} E_\mu\left[(W_N - N\mu)^2 I_{[N<(1-\epsilon)n^*]}\right] \\
&\quad \leq qn^{*(2\gamma-1)/(1+\gamma)} E_\mu\left[(W_{N_k} - N_k\mu)^2 I_{[N\leq k]}\right],
\end{aligned} \tag{4.3.16}$$

where, and in what follows, we denote: $k = \lfloor n^*(1-\epsilon) \rfloor = O(n^*)$, $N_k = \min(N, k)$, $\gamma' = E_\mu[(X_1 - \mu)^3]$ and $\gamma'' = E_\mu[(X_1 - \mu)^4]$.

Now, using Cauchy-Schwartz inequality and Wald's fourth lemma (Theorem 7, Chow et al.

1965; Theorem 2.4.7, Ghosh et al. 1997), we have:

$$
\begin{aligned}
& E_\mu \left[(W_{N_k} - N_k\mu)^2 \, I_{[N\le k]} \right] \\
& \quad \le E_\mu^{1/2} \left[(W_{N_k} - N_k\mu)^4 \right] P_\mu^{1/2}(N \le k) \\
& \quad \le \left\{ 6\sigma^2 E_\mu \left[N_k \left(W_{N_k} - N_k\mu\right)^2 \right] + 4\gamma' E_\mu \left[N_k \left(W_{N_k} - N_k\mu\right) \right] \right. \\
& \qquad \left. + \, \gamma'' E_\mu[N_k] \right\}^{1/2} P_\mu^{1/2}(N \le k).
\end{aligned} \tag{4.3.17}
$$

But, we have $N_k \le k$ w.p.1(P_μ). Next, we further use Wald's second equation (Theorem 8, Chow et al. 1965; Theorem 2.4.5, Ghosh et al. 1997) and Cauchy-Schwartz inequality in (4.3.17) to obtain:

$$
\begin{aligned}
& E_\mu \left[(W_{N_k} - N_k\mu)^2 \, I_{[N\le k]} \right] \\
& \quad \le \left\{ 6\sigma^2 k E_\mu \left[(W_{N_k} - N_k\mu)^2 \right] + 4\gamma' E_\mu^{1/2} \left[N_k^2 \right] E_\mu^{1/2} \left[(W_{N_k} - N_k\mu)^2 \right] \right. \\
& \qquad \left. + \, \gamma'' E_\mu[N_k] \right\}^{1/2} P_\mu^{1/2}(N \le k) \\
& \quad \le \left\{ 6\sigma^4 k^2 + 4\gamma' k^{3/2}\sigma + \gamma'' k \right\}^{1/2} P_\mu^{1/2}(N \le k).
\end{aligned} \tag{4.3.18}
$$

Step 6:

Next, we let $h = \left\lfloor \left(\frac{b}{\omega}\right)^{1/(1+\gamma)} \right\rfloor + 1$. We may pick ω small enough so that $h \le k$ where k was defined underneath (4.3.16). Then, we can write:

$$
\begin{aligned}
P_\mu \left\{ N \le n^*(1-\epsilon) \right\} &= P_\mu(N \le k) = \Sigma_{n=h}^{k} P_\mu(N = n) \\
&\le \Sigma_{n=h}^{k} P_\mu \left\{ n \ge \tfrac{n^*}{\sigma^2} \overline{X}_n \right\}.
\end{aligned} \tag{4.3.19}
$$

Now, with fixed but otherwise arbitrary $\nu(\ge 1)$, we may rewrite (4.3.19) and obtain:

$$
\begin{aligned}
& P_\mu \left\{ N \le n^*(1-\epsilon) \right\} \\
& \quad \le \Sigma_{n=h}^{k} P_\mu \{ \left| \overline{X}_n - \mu \right| \ge \epsilon\mu \} \\
& \quad \le (\epsilon\mu)^{-\nu} \Sigma_{n=h}^{k} E_\mu \left[\left| \overline{X}_n - \mu \right|^{2\nu} \right], \text{ by Tchebyshev's inequality} \\
& \quad \le q \Sigma_{n=h}^{k} n^{-\nu}.
\end{aligned} \tag{4.3.20}
$$

The last step in (4.3.20) follows from the lemma of Sen and Ghosh (1981). One may also refer to Lemma 9.2.3 in Ghosh et al. (1997, pp. 275-276).

Next, (4.3.20) leads to:

$$P_\mu\left\{N \le n^*(1-\epsilon)\right\} \le qkh^{-\nu} \le O(n^*)O(n^{*-\nu/(1+\gamma)}) = O\left(n^{*(1+\gamma-\nu)/(1+\gamma)}\right), \tag{4.3.21}$$

whatever be the fixed but otherwise arbitrary number $\nu(\ge 1)$.

Thus, by combining (4.3.16)-(4.3.18) with (4.3.20)-(4.3.21), we can claim:

$$E_\mu\left[U_N^2\left|\frac{{n^*}^2}{N^2}-1\right|I_{[N-n^*<-\epsilon n^*]}\right] = O(n^{*(3\gamma-\nu)/(1+\gamma)}). \tag{4.3.22}$$

Now, (4.3.22) validates (4.3.15) for any $\gamma > 1/2$, provided that we pick $\nu > \max(2, 3\gamma)$ which is certainly possible to do.

Our proof of Theorem 4.3.1 is thus complete. ■

Remark 4.3.1 (Negative Moments of N/n*). Now, we are in a position to summarize asymptotic behavior of $E_\mu\left[(N/n^*)^s\right]$ when $s < 0$. Clearly, $E_\mu\left[(N/n^*)^s I\left(N > \frac{1}{2}n^*\right)\right] \to 1$. But, then, in view of (4.3.21), we also have:

$$E_\mu\left[(N/n^*)^s I\left(N \le \frac{1}{2}n^*\right)\right] = O\left(n^{*-s+((1+\gamma-\nu)/(1+\gamma))}\right) = o(1),$$

when we pick $\nu > \max(2, (1+\gamma)(1-s))$ which is certainly possible to do. Thus, it follows that $E_\mu\left[(N/n^*)^s\right] \to 1$ when $\gamma > \frac{1}{2}, s < 0$.

4.3.3. Simulations and Real Data Illustrations

We begin with a summary (Section 4.3.3.1) from a set of simulation studies to examine the performances of our proposed purely sequential estimation strategy (4.3.4) for small and moderate values of n^*. Sections 4.3.3.2-4.3.3.3 highlight performances of our estimation strategy using real data from statistical ecology. Again, we have emphasized (i) weed count data of different species from a field in Netherlands and (ii) count data of migrating woodlarks at the Hanko bird sanctuary in Finland.

4.3.3.1. A Summary from Simulations

We generated pseudorandom observations from the distribution (4.1.1) with some combinations of choices for μ and τ. We fixed the values $b = 1$, $m = 5$, and $\gamma = 1.5$. Each row in Table 4.3.1 corresponds to averages from 10000 replications which were run under a given configuration. The choices of μ, τ, and n^* are consistent with those fixed in Section 4.2.4.1. In the spirit of (4.2.32), column 8 in Table 4.4 shows values of $\overline{z}$ and $s_{\overline{z}}$ where

$$\begin{gathered} N = n_i, r_i = \tfrac{b}{n_i}\left(\mu + \tau^{-1}\mu^2\right) \text{ as in (3.2) under the } i^{\text{th}} \text{ replication,} \\ \text{and } \overline{r} = H^{-1}\Sigma_{i=1}^{H} r_i,\ s_{\overline{r}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(r_i - \overline{r})^2}, \\ \text{so that } \overline{z} = \overline{r}/\omega,\ s_{\overline{z}} = s_{\overline{r}}/\omega,\ H = 10000. \end{gathered} \tag{4.3.23}$$

It is encouraging to observe that the summary highlighted in Table 4.4 looks generally consistent with those presented in Section 4.2.4.1 even though Section 4.2 took into account a different loss function. Table 4.4 shows clearly that our proposed sequential strategy (4.3.4) performs very well. In order to represent varying sample sizes, we show results for fixed values of $n^* = 50$ (small), 200 (medium).

4.3.3.2. Illustration 4.3.1: Weed Count Data

We return to use the weed count data discussed in Section 4.2.4.2 in order to highlight our first real data illustration. One may refer to Section 4.2.4.2 for a detailed description of the dataset. We had applied the methodology (4.3.4) on one of the species of weed, *Capsella bursa-pastoris L.* (Shephard's purse) from the full set of data without replacement. We found that with or without replacement made practically no difference.

Table 4.5 shows real data Illustration 4.3.1. The choices of γ and n^* are consistent with those in Section 4.2.4.2 along with $b = 1$ and $m = 10$. In column 6 of Table 4.5, we show a value $\widetilde{z}$ obtained from a single run:

$$\begin{gathered} N = n, \widetilde{r} = \tfrac{b}{n}\left(\overline{x}_n + \tau^{-1}\overline{x}_n^2\right) \text{ in the spirit of (4.2.33) under} \\ \text{one replication, so that } \widetilde{z} = \widetilde{r}/\omega. \end{gathered} \tag{4.3.24}$$

Again, column 6 shows $\widetilde{z}$ values mainly to grasp a sense of how close the estimated risk may or may not be when compared with the preset goal ω. A slight erratic behavior is due to the fact that

$\widetilde{z}$ values were obtained from single runs.

4.3.3.3 Illustration 4.3.2: Count Data of Migrating Woodlarks at the Hanko Bird Sanctuary

We return to use the ecological count dataset on migrating woodlarks at Hanko bird sanctuary situated in southwestern Finland that we discussed in Section 4.2.4.3 in order to highlight our second real data illustration. We applied the methodology (4.3.4) on the daily counts of migrating birds during first four hours of daylight after sunrise by drawing observations from the full set of data without replacement. We found that with or without replacement made practically no difference. Table 4.6 provides the results.

4.4. SQUARED ERROR LOSS APPROACH UNDER UNKNOWN THATCH PARAMETER

In this section, we develop a purely sequential estimation strategy for estimating the mean μ of a NB(μ, τ) population, under a squared error loss function assuming that **both** mean parameter μ and the thatch parameter τ are unknown. Having recorded $X_1, ..., X_n$, recall that the sample mean $\overline{X}_n = n^{-1}\Sigma_{i=1}^{n} X_i$ estimates μ and we propose a squared error loss function for our estimation problem as follows:

$$L_n \equiv L_n(\overline{X}_n, \mu) = b(\overline{X}_n - \mu)^2, \ b > 0. \tag{4.4.1}$$

The loss function (4.5.1) looks same as in (4.3.1), but in the present formulation we again emphasize that τ remains unknown.

Next, we may express the risk function as follows:

$$R_n \equiv E_{\mu,\tau}[L_n] = \tfrac{b}{n}\sigma^2, \tag{4.4.2}$$

where once again we recall that $\sigma^2 \equiv \mu + \tau^{-1}\mu^2$ from (4.1.2).

4.4.1. A Sequential Bounded Risk Estimation

Our goal is to bound the risk R_n given in (4.4.2) from above by $\omega(> 0)$. This leads us to the optimal fixed sample size n^* as follows:

$$n \geq \tfrac{b}{\omega}\sigma^2 = n^*, \text{ say.} \tag{4.4.3}$$

The magnitude of n^* remains unknown. Hence, we resort to developing a purely sequential bounded risk estimation strategy in the fundamental spirits of Chow and Robbins (1965) under the assumption that the mean parameter μ and the thatch parameter τ are both unknown.

We first fix $m(\geq 2)$ and gather pilot data $X_i, i = 1, ..., m$ of size m from the NB population. Now since σ^2 is unknown, we estimate it using the customary sample variance $S_n^2 = (n-1)^{-1}\Sigma_{i=1}^n(X_i - \overline{X}_n)^2, n \geq m$. However since S_n^2 may be zero with a positive probability, whatever be n, we also fix a number $\gamma(> \frac{1}{2})$ and define:

$$N = \inf\left\{n \geq m : n \geq \tfrac{b}{\omega}\left[S_n^2 + n^{-\gamma}\right]\right\}. \tag{4.4.4}$$

Again, $S_n^2 + n^{-\gamma}$ is treated as an estimator of σ^2 having gathered data $X_i, i = 1, ..., n$. This ensures that our estimated variance is positive with probability one for all $n \geq m$.

Next, based on the fully gathered data $\{N, X_1, ..., X_{m,}..., X_N\}$, we propose to estimate μ by the sample mean $\overline{X}_N$. The following Theorem gives a set of attractive first-order asymptotic properties for the proposed purely sequential methodology (4.4.4).

Theorem 4.4.1. *With loss function L_N and terminal sample size N defined in (4.4.1) and (4.4.4) respectively, under the purely sequential estimation rule $(N, \overline{X}_N)$ from (4.4.4), for each fixed $\mu \in R^+$* and $\tau \in R^+$ *we have as* $\omega \to 0$:

(*i*) $N/n^* \overset{P_{\mu,\tau}}{\to} 1$;

(*ii*) $E_{\mu,\tau}\left[(N/n^*)^s\right] \to 1$ for all s [*asymptotic first-order efficiency*];

(*iii*) $E_{\mu,\tau}\left[L_N\right]/\omega \to 1$ *if* $\gamma > 1$ [*asymptotic risk efficiency*];

where n^* comes from (4.4.3) and μ, τ are both assumed unknown.

4.4.2. An Outline of a Proof of Theorem 4.4.1

Part (i) follows from the inequality:

$$\left(S_N^2 + N^{-\gamma}\right)\sigma^{-2} \leq N/n^* \leq \left(S_{N-1}^2 + (N-1)^{-\gamma}\right)\sigma^{-2} + n^{*-1} \quad \text{w.p.1}(P_{\mu,\tau}), \tag{4.4.5}$$

and the facts that $N \to \infty$, $S_N^2 \to \sigma^{-2}$, $S_{N-1}^2 \to \sigma^{-2}$ w.p.1$(P_{\mu,\tau})$ as $\omega \to 0$.

Next, for small enough ω, observe that the right-hand side of (4.4.5) implies w.p.1$(P_{\mu,\tau})$:

$$N/n^* \leq \sigma^{-2}\left\{2N^{-1}\Sigma_{i=1}^{N}(X_i-\mu)^2+1\right\}+1 \leq \sigma^{-2}\sup_{n\geq 2}\left\{2n^{-1}\Sigma_{i=1}^{n}(X_i-\mu)^2+1\right\}+1.$$

Now, when $s > 0$, Part (ii) follows in the spirit of our proof of Theorem 4.3.1, part (ii).

When $s < 0$, a proof similar to that in (4.3.21) can be easily put together since the sample variance is a U-statistic. One needs to improvise upon the proof of an analogous result from Sen and Ghosh (1981). Finally, when $s < 0$, Part (ii) follows in the spirit of the proof outlined previously in our Remark 4.3.1.

Part (iii) can be proved along the line of the proof of Theorem 4.3.1, part (iii). More details are left out for brevity. ■

4.4.3. Simulations and Real Data Illustrations

We begin with a summary (Section 4.4.3.1) from a set of simulation studies to examine the performances of our proposed purely sequential estimation strategy (4.4.4) for small and moderate values of n^*. Sections 4.4.3.2-4.4.3.3 highlight performances of our estimation strategy using real data from statistical ecology. Again, we have emphasized (i) weed count data of different species from a field in Netherlands and (ii) count data of migrating woodlarks at the Hanko bird sanctuary in Finland.

4.4.3.1. A Summary from Simulations

We generated pseudorandom observations from the distribution (4.1.1) with some combinations of choices for μ and τ. We fixed the values $a = 1$, $m = 5$, and $\gamma = 1.5$. Each row in Table 4.7 corresponds to averages from 10000 replications which were run under a given configuration. The choices of μ, τ, and n^* are consistent with those fixed in previous simulations.

We now present a set of simulation studies and real data analysis to check the sample performance of our proposed purely sequential strategy (4.5.4). The results are as expected and they are consistent with those from earlier sections. Our proposed sequential strategy (4.5.4) is seen to perform well. In the spirits of (4.3.23), column 8 in Table 4.7 shows values of $\overline{z}$ and $s_{\overline{z}}$ where

$$\begin{gathered}N = n_i, r_i = \tfrac{b}{n_i}\sigma^2 \text{ with } \sigma^2 = \mu + \tau^{-1}\mu^2 \text{ as in (4.5.2) under the} \\ i^{\text{th}} \text{ replication, and } \overline{r} = H^{-1}\Sigma_{i=1}^{H} r_i,\ s_{\overline{r}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(r_i - \overline{r})^2}, \\ \text{so that } \overline{z} = \overline{r}/\omega,\ s_{\overline{z}} = s_{\overline{r}}/\omega,\ H = 10000.\end{gathered} \tag{4.4.6}$$

4.4.3.2. Illustration 4.4.1: Weed Count Data

We again return to use the weed count data in order to highlight our first real data illustration. This time, however, we had applied our methodology (4.4.4) on another species of weed, *Polygonum aviculare L.* (knotweed) from the full set of data without replacement. We found that with or without replacement made practically no difference.

A NB fit was seen as appropriate with a p-value of 0.94. We found $\widehat{\mu} = 0.91$ and $\widehat{\tau} = 4.11$ from full data. We emphasize that our implementation of the sampling strategy (4.4.4) did not exploit the numbers $\widehat{\mu}(= 0.91)$ and $\widehat{\tau}(= 4.11)$ obtained from full dataset.

Table 4.8 shows real data Illustration 4.4.1. The choices of γ and n^* are consistent with those in Section 4.2.4.2 along with $b = 1$ and $m = 10$. In column 6 of Table 4.8, we show a value $\widetilde{z}$ obtained from a single run:

$$\begin{gathered}N = n, \widetilde{r} = \tfrac{b}{n}s_n^2 \text{ in the spirit of (4.2.33) under} \\ \text{one replication, so that } \widetilde{z} = \widetilde{r}/\omega.\end{gathered} \tag{4.4.7}$$

Again, column 6 shows $\widetilde{z}$ values mainly to grasp a sense of how close the estimated risk may or may not be when compared with the preset goal ω. A slight erratic behavior is due to the fact that $\widetilde{z}$ values were obtained from single runs.

4.4.3.3 Illustration 4.4.2: Count Data of Migrating Woodlarks at the Hanko Bird Sanctuary

We again return to use the ecological count dataset on migrating woodlarks at Hanko bird sanctuary situated in southwestern Finland in order to highlight our second real data illustration.

We applied the methodology (4.4.4) on the daily counts of migrating birds during first four hours of daylight after sunrise by drawing observations from the full set of data without replacement. We found that with or without replacement made practically no difference.

Table 4.4.3 shows real data Illustration 4.4.2. The choices of γ and n^* are consistent with those in Section 4.2.4.2 along with $b = 1$ and $m = 7$. In column 6 of Table 4.9, we show a value $\widetilde{z}$ obtained from a single run. Again, column 6 shows $\widetilde{z}$ values mainly to grasp a sense of how close the estimated risk may or may not be when compared with the preset goal ω. A slight erratic behavior is due to the fact that $\widetilde{z}$ values were obtained from single runs.

4.5. SIMULTANEOUS MEANS ESTIMATION UNDER MODIFIED LINEX LOSS: KNOWN THATCH PARAMETERS

In this section, we first introduce an appropriate modification to the conventional form of the Linex loss function under a k-sample scenario. We then move on to construct a suitable stopping rule.

4.5.1. A Modified Linex Loss

We have available a data sequence $X_{ij}, j = 1, 2, ..., n, ...$ recorded independently from a $\mathrm{NB}(\mu_i, \tau_i)$ population, assumed to be independent of each other, $i = 1, ..., k$. We denote the i^{th} sample mean $\overline{X}_{in} = n^{-1}\Sigma_{j=1}^{n} X_{ij}$ which estimates $\mu_i, i = 1, ..., k$, $n = 1, 2, ...$ and thereby giving rise to the corresponding sample mean vector, $\overline{\mathbf{X}}_n = (\overline{X}_{1n}, ..., \overline{X}_{kn})$, which simultaneously estimates the unknown mean parameter vector, $\boldsymbol{\mu} \equiv (\mu_1, ..., \mu_k)$. We may also denote $\boldsymbol{\tau} \equiv (\tau_1, ..., \tau_k)$ which is assumed known.

In the light of (4.2.1), we propose a modified Linex loss function for our simultaneous estimation problem as follows and define:

$$\begin{aligned} L_n \equiv L_n\left(\overline{\mathbf{X}}_n, \boldsymbol{\mu}\right) &= \textstyle\sum_{i=1}^{k} W_{in}\left(\overline{X}_{in}, \mu_i\right) \\ &= \textstyle\sum_{i=1}^{k} \left[\exp\left\{\frac{a_i(\overline{X}_{in}-\mu_i)}{\mu_i}\right\} - \frac{a_i(\overline{X}_{in}-\mu_i)}{\mu_i} - 1\right], \text{ with fixed } a_i \in R. \end{aligned} \tag{4.5.1}$$

Next, for a fixed i, we may express:

$$E_{\mu_i}[W_{in}\left(\overline{X}_{in}, \mu_i\right)] = \frac{1}{2}\frac{a_i^2(\mu_i+\tau_i^{-1}\mu_i^2)}{n\mu_i^2} + o\left(\frac{1}{n}\right), i = 1, ..., k. \tag{4.5.2}$$

One may refer to Section 4.2 for additional details. From (4.5.2), the risk associated with the modified Linex loss function (4.5.1) will reduce to:

$$R_n \equiv E_{\boldsymbol{\mu}}[L_n\left(\overline{\mathbf{X}}_n, \boldsymbol{\mu}\right)] = \tfrac{1}{2}n^{-1}\sum_{i=1}^{k} a_i^2\left(\tau_i^{-1} + \mu_i^{-1}\right) + o(n^{-1}). \quad (4.5.3)$$

4.5.2. A Purely Sequential Approach

The idea here is to bound the risk R_n given in (4.5.3) from above. We introduce a suitable constant, namely the risk bound, $\omega(>0)$ and require that $R_n \leq \omega$ for all $\boldsymbol{\mu}, \boldsymbol{\tau}$. This leads us to an approximate optimal fixed sample size n^* as follows:

$$n \geq \tfrac{1}{2\omega}\left\{\Sigma_{i=1}^{k} a_i^2\tau_i^{-1} + \Sigma_{i=1}^{k} a_i^2\mu_i^{-1}\right\} = n^*, \text{ say.} \quad (4.5.4)$$

The magnitude of n^* remains unknown. Hence, we resort to developing a k-sample purely sequential bounded risk simultaneous estimation strategy.

Looking back at the expression of n^* from (4.5.4), we observe:

$$n^* > \frac{1}{2\omega}\Sigma_{i=1}^{k} a_i^2\tau_i^{-1}, \quad (4.5.5)$$

and hence, we determine the pilot size m as follows: Let

$$m \equiv m(\omega) = \left\lfloor \frac{1}{2\omega}\Sigma_{i=1}^{k} a_i^2\tau_i^{-1} \right\rfloor + 1, \quad (4.5.6)$$

where $\lfloor u \rfloor$ denotes the largest integer $< u(>0)$. We first gather the pilot data $X_{ij}, i = 1, ..., k$, $j = 1, ..., m$ of size m from each NB population. Then, we continue to record one additional observation vector $(X_{1j}, ..., X_{kj})$ at-a-time as needed according to the stopping rule that is defined next.

Since $\overline{X}_{in}$ may be zero with a positive probability, whatever be n, in the spirit of Mukhopadhyay and Diaz (1985), we now fix a number $\gamma(>1/2)$ and define:

$$N = \inf\left\{n \geq m : n \geq \tfrac{1}{2\omega}\Sigma_{i=1}^{k} a_i^2\left[\tau_i^{-1} + \left(\overline{X}_{in} + n^{-\gamma}\right)^{-1}\right]\right\}. \quad (4.5.7)$$

Here, $\left(\overline{X}_{in} + n^{-\gamma}\right)$ is treated as an estimator of μ_i which ensures that this estimator is positive with (P_{μ_i}) probability one and that the classical central limit theorem (CLT) will remain in effect for $n^{1/2}\left(\overline{X}_{in}+n^{-\gamma}-\mu_i\right)$.

Next, based on the fully gathered data upon termination, namely $\{X_{i1},...,X_{im,}X_{im+1},...,X_{iN}\}$ from the NB(μ_i,τ_i) population, we propose to estimate μ_i by the sample mean $\overline{X}_{iN}, i=1,...,k$. Finally,

$$\begin{gathered}\text{we simultaneously estimate the unknown parameter vector } \boldsymbol{\mu} \\ \equiv(\mu_1,...,\mu_k) \text{ with the sample mean vector } \overline{\mathbf{X}}_N=(\overline{X}_{1N},...,\overline{X}_{kN}).\end{gathered} \tag{4.5.8}$$

Theorem 4.5.1. *Under the Linex loss (4.5.1), with m and N defined in (4.5.6) and (4.5.7) respectively, and under the purely sequential simultaneous estimation rule (4.5.8), for each fixed* $\boldsymbol{\mu}\in R^{+k}$ and $\boldsymbol{\tau}\in R^{+k}$ *we have as* $\omega\to 0$:

(*i*) $N/n^*\to 1$ *w.p.1*$(P_{\boldsymbol{\mu}})$;

(*ii*) $E_{\boldsymbol{\mu}}\left[N/n^*\right]\to 1$, $E_{\boldsymbol{\mu}}\left[n^*/N\right]\to 1$ [*asymptotic first-order efficiency*];

(*iii*) $E_{\boldsymbol{\mu}}\left[L_N\right]/\omega\to 1$ [*asymptotic risk efficiency*];

where n^* comes from (4.5.5) and $\boldsymbol{\tau}$ vector is assumed known.

Part (ii) shows that the sequential methodology (4.5.7) is efficient along the lines of Chow and Robbins (1965) and first-order efficient in the sense of Ghosh and Mukhopadhyay (1981). That is, we may expect the terminal sample size N from each NB population to hover around the optimal fixed sample size n^* when n^* is large. Also, part (iii) shows that the achieved risk $E_{\boldsymbol{\mu}}\left[L_N\right]$ may be expected to hover around the preassigned risk-bound ω when n^* is large.

4.5.3. Proof of Theorem 4.5.1

Part (i):

From (4.5.7), we get the following inequality w.p.1$(P_{\boldsymbol{\mu}})$:

$$\begin{gathered}\tfrac{1}{2\omega}\sum_{i=1}^k a_i^2\left\{\tau_i^{-1}+\left(\overline{X}_{iN}+N^{-\gamma}\right)^{-1}\right\}\le N\le\tfrac{1}{2\omega}\sum_{i=1}^k a_i^2\left\{\tau_i^{-1}\right. \\ \left.+\ \left(\overline{X}_{i(N-1)}+(N-1)^{-\gamma}\right)^{-1}\right\}+m\Rightarrow\sum_{i=1}^k U_{iN}\le N\le\sum_{i=1}^k V_{iN}+m,\end{gathered} \tag{4.5.9}$$

where:

$$U_{iN}=\tfrac{a_i^2}{2\omega}\left\{\tau_i^{-1}+\left(\overline{X}_{iN}+N^{-\gamma}\right)^{-1}\right\} \text{ and } V_{iN}=\tfrac{a_i^2}{2\omega}\left\{\tau_i^{-1}+\left(\overline{X}_{i(N-1)}+(N-1)^{-\gamma}\right)^{-1}\right\},$$

$i = 1, 2, ..., k$.

Dividing (4.5.9) throughout by n^* we obtain w.p.1($P_{\boldsymbol{\mu}}$):

$$
\begin{aligned}
&\tfrac{1}{n^*}\Sigma_{i=1}^k U_{iN} \le N/n^* \le \tfrac{1}{n^*}\Sigma_{i=1}^k V_{iN} + mn^{*-1} \\
&\Rightarrow \Sigma_{i=1}^k \tfrac{U_{iN}}{n_i^*}\tfrac{n_i^*}{n^*} \le N/n^* \le \Sigma_{i=1}^k \tfrac{V_{iN}}{n_i^*}\tfrac{n_i^*}{n^*} + mn^{*-1},
\end{aligned}
\tag{4.5.10}
$$

with $n_i^* = \frac{1}{2\omega}a_i^2\left(\tau_i^{-1} + \mu_i^{-1}\right)$, $i = 1, 2, ..., k$.

Now, along the line of Section 4.2.3, part (i), we can conclude that both U_{iN}/n_i^*, $V_{iN}/n_i^* \to 1$ w.p.1($P_{\boldsymbol{\mu}}$) for each $i = 1, 2, ..., k$. Hence, proof of part (i) is complete since $\Sigma_{i=1}^k n_i^*/n^* = 1$ and $mn^{*-1} \to 0$.

Part (ii):

We first fix some arbitrarily small $\epsilon > 0$ and define $\beta = (1+\epsilon)n^*$. We tacitly disregard the fact that β may not be an integer. Then, we may write:

$$
E_{\boldsymbol{\mu}}\left[\tfrac{N}{n^*}\right] \le (\beta+1)P_{\boldsymbol{\mu}}(N \le \beta+1) + n^{*-1}T(\beta), \text{ say,}
$$

where $T(\beta) = \Sigma_{n>\beta+1}^{\infty} nP_{\boldsymbol{\mu}}(N = n)$.

Let $[N = n]$ denote the event $N = n$. Now, from (4.5.7), we note the fact that (for $n > m$):

$$
[N = n] \subset \left\{n - 1 < \tfrac{1}{2\omega}\sum_{i=1}^k a_i^2\left[\tau_i^{-1} + \left(\overline{X}_{in-1} + (n-1)^{-\gamma}\right)^{-1}\right]\right\}, \tag{4.5.11}
$$

so that we may claim:

$$
\begin{aligned}
&T(\beta) \\
&\le \sum_{n=\beta}^{\infty}(n+1)P_{\boldsymbol{\mu}}\left\{\sum_{i=1}^k a_i^2\left(\overline{X}_{in} + n^{-\gamma}\right)^{-1} > 2\omega n - \sum_{i=1}^k \frac{a_i^2}{\tau_i^{-1}}\right\} \\
&\le \sum_{n=\beta}^{\infty}(n+1)P_{\boldsymbol{\mu}}\left\{\sum_{i=1}^k a_i^2\left(\overline{X}_{in} + n^{-\gamma}\right)^{-1} > 2\omega - n^{-1}\sum_{i=1}^k \frac{a_i^2}{\tau_i^{-1}}\right\} \\
&\le \sum_{i=1}^k\sum_{n=\beta}^{\infty}(n+1)P_{\boldsymbol{\mu}}\left\{a_i^2\left(\textstyle\sum_{j=1}^n X_{ij} + n^{-\gamma+1}\right)^{-1} > \frac{1}{k}2\omega - n^{-1}\tau_i^{-1}\right\}
\end{aligned}
\tag{4.5.12}
$$

Next, using the moment generating function of $\sum_{j=1}^n X_{ij} \sim NB(n\mu_i, n\tau_i)$, $i = 1, ..., k$, we can show that $T(\beta) \leq \Sigma_{n=1}^{\infty} d_n$ where $d_n > 0$ and $d_n^{1/n} \to d$, $0 < d < 1$, as $n \to \infty$. Hence, in the spirit of Theorem 4.3.1, we can claim:

$$\limsup_{\omega \to 0} E_{\boldsymbol{\mu}}\left[(N/n^*)\right] \leq 1 + \epsilon, \tag{4.5.13}$$

but $\epsilon(> 0)$ is arbitrary. Also, from part (i) and Fatou's lemma, we have:

$$\liminf_{\omega \to 0} E_{\boldsymbol{\mu}}\left[N/n^*\right] \geq E_{\boldsymbol{\mu}}\left[\liminf_{\omega \to 0} N/n^*\right] = 1. \tag{4.5.14}$$

Now, combining (4.5.13)-(4.5.14) completes the proof of the first result in part (ii).

Also, in view of the definition of m from (4.5.6), we observe:

$$0 \leq n^*/N \leq n^*/m = O(1) \text{ w.p.1}(P_{\boldsymbol{\mu}}),$$

that is, n^*/N is bounded. The second result in part (ii) follows by appealing to part (i). Now, the proof of part (ii) is complete.

Part (iii):

From (4.5.1), we may express $E_{\boldsymbol{\mu}}[L_N]/\omega$ as:

$$\omega^{-1} \sum_{i=1}^{k} E_{\boldsymbol{\mu}}\left[\exp\left\{\tfrac{a_i(\overline{X}_{iN} - \mu_i)}{\mu_i}\right\} - a_i\left(\tfrac{\overline{X}_{iN} - \mu_i}{\mu_i}\right) - 1\right] = \sum_{i=1}^{k} \tfrac{R_i}{\omega} = \sum_{i=1}^{k} \tfrac{R_i}{\omega_i}\tfrac{\omega_i}{\omega},$$

where we note:

$$\omega = \tfrac{1}{2n^*}\sum_{i=1}^{k} \tfrac{a_i^2 \sigma_i^2}{\mu_i^2}, \omega_i = \tfrac{1}{2n^*}\tfrac{a_i^2\sigma_i^2}{\mu_i^2} \text{ and } R_i = \\ E_{\boldsymbol{\mu}}\left[\exp\left\{\tfrac{a_i(\overline{X}_{iN} - \mu_i)}{\mu_i}\right\} - a_i\left(\tfrac{\overline{X}_{iN} - \mu_i}{\mu_i}\right) - 1\right].$$

Next, one may follow along the lines of Section 4.2.3, part (iii) to show that $R_i/\omega_i \to 1$ as $\omega \to 0$ for all $i = 1, ..., k$.

This completes the proof of Theorem 4.5.1. ■

4.5.4. Simulations and Real Data Illustrations

Section 4.5.4.1 presents a summary from a set of simulation studies to examine the performances of our proposed purely sequential estimation strategy (4.5.7) for a varying set of values of n^*. Section 4.5.4.2 highlights performances of our estimation strategy using real data from statistical ecology. We have emphasized raptor count data from the Hawk Mountain sanctuary.

4.5.4.1. A Summary from Simulations

We first generated a pair of pseudorandom observations at-a-time from the distribution (4.1.1) with combinations of choices for μ_1, μ_2 and τ_1, τ_2. We fixed the values $a_1 = 1$, $a_2 = 1.2$ and $\gamma = 1.5$. We determined m from (4.5.6). Each row in Table 4.10 corresponds to averages from 10000 replications which were run under each given configuration. We considered a wide variety of values for n^*. In order to represent varying sample sizes, we include results obtained only for fixed values of $n^* = 50$ (small) and $n^* = 200$ (medium).

The results are shown under the following setting:

$$\boxed{n^* = 50, 200 \text{ and } (\mu_1, \mu_2, \tau_1, \tau_2) = (3,1,3,2), (7,3,2,1)}$$

Each block in Table (4.10) shows n^* (column 3), ω (column 4), the estimated values $\overline{x}_1, \overline{x}_2$ along with their estimated standard errors $s_{\overline{x}_1}, s_{\overline{x}_2}$ (columns 5 and 6), estimated values $\overline{n}$ along with its estimated standard error $s_{\overline{n}}$ (column 7), the ratio $\overline{n}/n^*$ (column 8), and the values of $\overline{z}$ and $s_{\overline{z}}$ (column 9) where we denote:

$$\begin{gathered} N = n_i, r_i = \textstyle\sum_{j=1}^{2} \frac{a_j^2}{2n_i}\left(\frac{1}{\mu_j} + \frac{1}{\tau_j}\right) \text{ as in (4.5.3) under the } i^{\text{th}} \text{ replication,} \\ \text{and } \overline{r} = H^{-1}\Sigma_{i=1}^{H} r_i,\ s_{\overline{r}} = \sqrt{(H^2 - H)^{-1}\Sigma_{i=1}^{H}(r_i - \overline{r})^2}, \\ \text{so that } \overline{z} = \overline{r}/\omega,\ s_{\overline{z}} = s_{\overline{r}}/\omega,\ H = 10000. \end{gathered} \tag{4.5.15}$$

Under our proposed methodology (4.5.7), we observe that the $\overline{x}_1, \overline{x}_2$ values are very close to the corresponding population means μ_1, μ_2 in each case with very small standard error values $s_{\overline{x}_1}, s_{\overline{x}_2}$. These become more accurate when $n^* = 200$. The values of $\overline{n}$ seem to estimate n^* very accurately across the rows. We note that the $\overline{n}$ values overestimate n^* by a small margin. The last column appears to validate our sentiment that the sequential methodology tends to provide a risk bound close (or smaller) to (than) the preset goal ω for all values of n^*.

4.5.4.2. Illustration 4.5.1: Raptor Count Data at the Hawk Mountain Sanctuary

We use ecological count dataset on two species of raptors, *American Kestrel* (X_1) and *Golden Eagle* (X_2), from the Hawk Mountain Sanctuary in Pennsylvania. This data had been collected at the Hawk Mountain Sanctuary and is available on its website

$$\boxed{\text{www.hawkmountain.org.}}$$

We worked with the daily count data of raptors for 2 years, 2015 and 2016 (both Spring and Fall). The dataset included 357 rows and a NB fit was seen as appropriate with p-values of 0.29 and 0.43 for X_1 and X_2 datasets respectively. Treating these two datasets as the universe, we first found $\widehat{\mu}_1 = 1.43$, $\widehat{\tau}_1 = 0.28$ and $\widehat{\mu}_2 = 0.7$, $\widehat{\tau}_2 = 0.18$ from full data. In other words, we considered this dataset as our population with unknown μ_1, μ_2 but with known values for τ_1, τ_2. We emphasize that our implementation of the sampling strategy (4.5.7) did not exploit the numbers $\widehat{\mu}_1, \widehat{\mu}_2$ obtained from full datasets.

We picked $a_1 = 1$, $a_2 = 1.2$ and $\gamma = 1.5$, implemented the methodology (4.5.7) by drawing observations from the full set of data without replacement. We found that with or without replacement made practically no difference.

Table 4.11 shows results from implementing the estimation strategy (4.6.7).in single runs with 3 different preset values of the risk bound ω. The terminal estimated values of μ_1, μ_2 are reasonably close to the values $\widehat{\mu}_1 = 1.43$, $\widehat{\mu}_2 = 0.7$ obtained from full data. The n^* values are merely provided as a vehicle for ad-hoc comparison with the observed n values. We have not used these n^* values in our implementation. The sample sizes (Table 4.11) considered are small/moderate and the ratio n/n^* remains close to 1 which is desired. In the last column of Table 4.11, we show associated value $\widetilde{z}$ obtained from a single run:

$$\begin{gathered} N = n, \widetilde{r} = \textstyle\sum_{j=1}^{2} \frac{a_j^2}{2n} \left(\frac{1}{\overline{x}_j} + \frac{1}{\tau_j} \right) \text{ in the spirit of (4.5.3) under} \\ \text{one replication, so that } \widetilde{z} = \widetilde{r}/\omega. \end{gathered} \tag{4.5.16}$$

and observe that $\widetilde{z}$ values are smaller than one which indicates that the estimated risk is smaller than the preset goal ω. We should emphasize that each row in Table 4.11 summarizes a single run.

4.6. COMPARING TWO MEANS UNDER SEL: KNOWN THATCH PARAMETERS

In this section, we fix $k = 2$ and develop a purely sequential strategy for estimating $\Delta \equiv \mu_1 - \mu_2$, the difference of means in two independent NB populations under a squared error loss function. We assume that the thatch parameters τ_1 and τ_2 are known. As in Section 2, we denote $\boldsymbol{\mu} \equiv (\mu_1, \mu_2)$ and $\boldsymbol{\tau} \equiv (\tau_1, \tau_2)$.

We assume that the sample sizes are equal. Thus, having recorded $X_{11}, ..., X_{1n}$ and $X_{21}, ..., X_{2n}$ from NB(μ_1, τ_1) and NB(μ_2, τ_2) respectively, we propose the difference in the sample means

$$T_n \equiv \overline{X}_{1n} - \overline{X}_{2n} = n^{-1}(\Sigma_{i=1}^n X_{1i} - \Sigma_{i=1}^n X_{2i})$$

to estimate Δ as we proceed to work under a *squared error loss function* (SEL):

$$\begin{aligned} L_n \equiv L_n(T_n, \Delta) &= b\left[(\overline{X}_{1n} - \overline{X}_{2n}) - (\mu_1 - \mu_2)\right]^2 \\ &= b\left[(\overline{X}_{1n} - \mu_1) - (\overline{X}_{2n} - \mu_2)\right]^2, b > 0. \end{aligned} \tag{4.6.1}$$

We express the risk function associated with (4.6.1) as:

$$R_n \equiv E_{\boldsymbol{\mu}}[L_n] = \tfrac{b}{n}\left(\sigma_1^2 + \sigma_2^2\right) = \tfrac{b}{n}\left(\mu_1 + \tfrac{\mu_1^2}{\tau_1} + \mu_2 + \tfrac{\mu_2^2}{\tau_2}\right). \tag{4.6.2}$$

4.6.1. A Purely Sequential Approach

We will again bound the risk R_n given in (4.6.2) from above by $\omega(> 0)$. This leads us to the optimal fixed sample size n^* approximately as follows:

$$n \geq \tfrac{b}{\omega}\left(\mu_1 + \tfrac{\mu_1^2}{\tau_1} + \mu_2 + \tfrac{\mu_2^2}{\tau_2}\right) = n^*, \text{ say.} \tag{4.6.3}$$

The magnitude of n^* remains unknown even though its expression is given by (4.6.3). Hence, we resort to developing a purely sequential bounded risk estimation strategy next. We will continue to assume that the thatch parameters τ_1 and τ_2 are known.

We first fix $m(\geq 1)$ and gather a set of pilot data X_{1i}, X_{2i}, $i = 1, ..., m$ of size m from the NB populations. Again, since both $\overline{X}_{1n}, \overline{X}_{2n}$ may be zero with a positive probability, whatever be n, we fix a number $\gamma(> \frac{1}{2})$ and define:

$$N = \inf\left\{n \geq m : n \geq \tfrac{b}{\omega}\left(\overline{X}_{1n} + \overline{X}_{2n} + \tau_1^{-1}\overline{X}_{1n}^2 + \tau_2^{-1}\overline{X}_{2n}^2 + n^{-\gamma}\right)\right\}. \tag{4.6.4}$$

Here, the term $n^{-\gamma}$ ensures that the estimator used on the right-hand side of (4.6.4) is positive w.p.1($P_{\boldsymbol{\mu}}$). Based on the fully gathered data $\{N, X_{i1}, ..., X_{im}, ..., X_{iN}\}$, $i = 1, 2$, we propose to estimate

$$\Delta \equiv \mu_1 - \mu_2 \text{ by } T_N \equiv \overline{X}_{1N} - \overline{X}_{2N}. \tag{4.6.5}$$

Theorem 4.6.1 gives a set of attractive first-order asymptotic properties for the proposed purely sequential estimation methodology obtained from (4.6.4)-(4.6.5).

Theorem 4.6.1. *Under the SEL function L_N from (4.6.1) and the purely sequential estimation strategy (4.6.4)-(4.6.5), for each fixed $\mu_1, \mu_2 \in R^+$* and $\tau_1, \tau_2 \in R^+$ *we have as $\omega \to 0$:*

(*i*) $N/n^* \to 1$ w.p.1($P_{\boldsymbol{\mu}}$) *if* $\gamma > \frac{1}{2}$;

(*ii*) $E_{\boldsymbol{\mu}}\left[(N/n^*)^s\right] \to 1$ *for all* $s > 0$, *if* $\gamma > \frac{1}{2}$ [*asymptotic first-order efficiency*];

(*iii*) $E_{\boldsymbol{\mu}}\left[L_N\right]/\omega \to 1$ *if* $\gamma > 1$ [*asymptotic risk efficiency*];

where n^* comes from (4.6.3) and τ_1, τ_2 are assumed known.

Part (ii) shows that the sequential methodology (4.6.4) is efficient along the lines of Chow and Robbins (1965) and first-order efficient in the sense of Ghosh and Mukhopadhyay (1981). That is, we may expect N to hover around the optimal fixed sample size n^* when n^* is large. Part (iii) shows that the achieved risk $E_{\boldsymbol{\mu}}\left[L_N\right]$ may be expected to hover around the preassigned risk-bound ω when n^* is large.

4.6.2. Proof of Theorem 4.6.1

Part (i):

From (4.6.4), we may write down the following inequality:

$$\tfrac{b}{\omega}\left(\overline{X}_{1N}+\overline{X}_{2N}+\tau_1^{-1}\overline{X}_{1N}^2+\tau_2^{-1}\overline{X}_{2N}^2+N^{-\gamma}\right)\le N\le \tfrac{b}{\omega}\left(\overline{X}_{1N-1}+\overline{X}_{2N-1}+\tau_1^{-1}\overline{X}_{1N-1}^2\right.$$

$$\left.+\tau_2^{-1}\overline{X}_{2N-1}^2+(N-1)^{-\gamma}\right)+m \text{ w.p.1}(P_{\boldsymbol{\mu}}). \tag{4.6.6}$$

Now, dividing (4.6.6) throughout by n^*, and then taking limits on all sides of (4.6.6), we conclude the desired result. Here, $\gamma > \frac{1}{2}$ suffices.

Part (*ii*):

From the right-hand side of (4.6.6), we have the following inequality (for sufficiently large n^*):

$$N/n^* \le \left(\overline{X}_{1N-1}+\overline{X}_{2N-1}+\tau_1^{-1}\overline{X}_{1N-1}^2+\tau_2^{-1}\overline{X}_{2N-1}^2+1\right)(\sigma_1^2+\sigma_2^2)^{-1}+m, \text{ w.p.1}(P_{\boldsymbol{\mu}}). \tag{4.6.7}$$

Now, denoting $\sup_{n\ge 2}\left\{\overline{X}_{1n}+\overline{X}_{1n}+\tau_1^{-1}\overline{X}_{1n}^2+\tau_2^{-1}\overline{X}_{2n}^2\right\}$ by W, we can claim w.p.1$(P_{\boldsymbol{\mu}})$:

$$N/n^* \le (\sigma_1^2+\sigma_2^2)^{-1}\{W+1\}+m. \tag{4.6.8}$$

Thus, the right-hand side of (4.6.8) is free from ω and using Wiener's (1939) ergodic theorem we can claim the uniform integrability of all positive powers of N/n^*. Next, appealing to part (i), we complete the proof. Here, again, $\gamma > \frac{1}{2}$ suffices.

Part (*iii*): In this proof, we may improvise on the techniques that were originally developed by Ghosh and Mukhopadhyay (1979) and further extended by Sen and Ghosh (1981), in the spirit of Section 4.3.2. Details are omitted.

Our proof of Theorem 4.6.1 is now complete. ■

Remark 4.6.1 (Negative Moments of N/n^*). Following along Remark 4.3.1, we can conclude that $E_{\boldsymbol{\mu}}\left[(N/n^*)^s\right] \to 1$ when $\gamma > \frac{1}{2}, s < 0$.

Remark 4.6.2 (Unequal Sample Sizes). In the case of unequal sample sizes, under the corresponding SEL function in the spirit of (4.6.1), we could develop an associated purely sequential bounded risk estimation strategy. This would then involve an allocation scheme as well as a terminal estimator for Δ with a set of desirable asymptotic first-order properties along the lines of Theorem

4.6.1. Specific details are omitted for brevity. However, one will find related analyses in Section 4.7 when thatch parameters are also assumed unknown.

Remark 4.6.3 (A k-Sample consideration). In the case of a k-sample scenario, one could set out to estimate the parameter Δ that is defined as $\Sigma_{i=1}^{k} d_i \mu_i$ by the estimator $T_n \equiv \Sigma_{i=1}^{k} d_i \overline{X}_{in}$ where d_i's are assumed non-zero constants. Under the corresponding SEL function in the spirit of (4.6.1), we could develop an associated purely sequential bounded risk estimation strategy and a set of desirable asymptotic first-order properties along the lines of Theorem 4.6.1. Further details are left out for brevity.

4.6.3. Simulations and Real Data Illustrations

Section 4.6.3.1 presents a summary from a set of simulation studies to examine the performances of our proposed purely sequential estimation strategy (4.6.4)-(4.6.5) for a varying set of values of n^*. Section 4.6.3.2 highlights performances of our estimation strategy using real data from statistical ecology. Again, we utilize raptor count data from the Hawk Mountain sanctuary.

4.6.3.1. A Summary from Simulations

We generated a pair of pseudorandom observations at-a-time from the distribution (4.1.1) with combinations of choices for μ_1, μ_1 and τ_1, τ_2. We fixed the values $b = 1$, $m = 5$, and $\gamma = 1.5$. Each row in Table 4.12 corresponds to averages from 10000 replications which were run under a given configuration. The choices of $\mu_1, \mu_1, \tau_1, \tau_2$, and n^* are consistent with those fixed in Section 4.5.4.1. Column 7 in Table 4.12 shows $\overline{\Delta}$ values which were computed from the observed sample averages, namely, $\overline{x}_1 - \overline{x}_2$ values under each case. In the spirit of (4.5.15), column 10 in Table 4.12 shows values of $\overline{z}$ and $s_{\overline{z}}$ where

$$\begin{gathered} N = n_i, r_i = \frac{b}{n_i} \sum_{j=1}^{2} \left(\mu_j + \tau_j^{-1} \mu_j^2 \right) \text{ as in (3.2) under the } i^{\text{th}} \text{ replication,} \\ \text{and } \overline{r} = H^{-1} \Sigma_{i=1}^{H} r_i, \; s_{\overline{r}} = \sqrt{(H^2 - H)^{-1} \Sigma_{i=1}^{H} (r_i - \overline{r})^2}, \\ \text{so that } \overline{z} = \overline{r}/\omega, \; s_{\overline{z}} = s_{\overline{r}}/\omega, \; H = 10000. \end{gathered} \tag{4.6.9}$$

It is encouraging that the summary highlighted in Table 4.12 looks generally consistent with those presented in Section 4.5.4.1 even though Section 4.5 took into account a different loss function.

The $\overline{\Delta}$ values also accurately estimate the true difference Δ under each case. Table 4.12 shows clearly that our proposed sequential strategy (4.6.4)-(4.6.5) performs very well. In order to represent varying sample sizes, we show selected results for fixed values of $n^* = 50$ (small) and $n^* = 200$ (medium).

4.6.3.2. Illustration 4.6.1: Raptor Count Data at the Hawk Mountain Sanctuary

We return to use the raptor count data discussed as discussed in Section 4.5.4.2 in order to highlight our real data Illustration 4.6.1 in Table 4.13. We applied the methodology (4.6.4)-(4.7.5) on two species of raptors, *American Kestrel* (X_1) and *Golden Eagle* (X_2) from the full set of data without replacement. Again, with or without replacement sampling made practically no difference.

Table 4.13 shows real data Illustration 4.6.1. The choices of γ and n^* are consistent with those in Section 4.5.4.2 along with $b = 1$ and $m = 5$. Column 5 shows $\overline{\Delta}(= \overline{x}_{1n} - \overline{x}_{2n})$ values which estimates $\Delta = 0.73$. In column 8 of Table 4.13, we show a value $\widetilde{z}$ obtained from single runs:

$$N = n, \widetilde{r} = \tfrac{b}{n}\left(\overline{x}_{1n} + \tau_1^{-1}\overline{x}_{1n}^2 + \overline{x}_{2n} + \tau_2^{-1}\overline{x}_{2n}^2\right) \text{ in the spirit of (4.5.16) under one replication, so that } \widetilde{z} = \widetilde{r}/\omega. \tag{4.6.10}$$

Again, column 8 shows $\widetilde{z}$ values mainly to grasp a sense of how close the estimated risk may or may not be when compared with the preset goal ω. A slightly erratic observed behavior is due to the fact that $\widetilde{z}$ values were obtained from single runs.

4.7. COMPARING TWO MEANS UNDER SEL WITH UNEQUAL SAMPLE SIZES: UNKNOWN THATCH PARAMETERS

The purely sequential sampling designs considered in this section will be fundamentally different from the rest of the paper on two important counts:

(a) In contrast with Section 4.6, we now handle estimation of $\Delta \equiv \mu_1 - \mu_2$ from two NB populations when the sample sizes are allowed to be unequal; and

(b) This is the only section within this paper where thatch parameters τ_1, τ_2 are allowed to remain unknown.

But, before we move forward with the NB problems, let us first review very briefly the literature on purely sequential analogues of the Behrens-Fisher problem just so that a reader may get some proper perspective. In the case of two independent normal populations, purely sequential fixed-width confidence interval estimation problems for the difference in means with unequal sample sizes were originally formulated and developed by Robbins et al. (1967) and Srivastava (1970) when both sets of mean and variance parameters remained unknown. Mukhopadhyay (1976,1977) developed associated point estimation problems and expanded fixed-width confidence interval estimation problems while comparing two or three independent normal populations having unknown sets of mean and variance parameters.

The literature is extensive. However, we may refer readers to Ghosh and Mukhopadhyay (1980), Ramkaran et al. (1986), and Mukhopadhyay and Liberman (1989), including other sources, for a broader range of review. The following monographs may also help in this regard: Ghosh et al. (1997, Sections 7.3, 8.3, pp. 222-223, pp. 256-259) and Mukhopadhyay and de Silva (2009, Chapter 13).

In this section, we develop a purely sequential estimation strategy to estimate the difference in means $\Delta \equiv \mu_1 - \mu_2$ of two NB populations, under SEL assuming that both mean parameters in $\mu = (\mu_1, \mu_2)$ and both thatch parameters in $\tau = (\tau_1, \tau_2)$ are unknown. We will also assume that the two sample sizes are unequal.

Having recorded $X_{11}, ..., X_{1n_1}$ NB(μ_1, τ_1) and $X_{21}, ..., X_{2n_2}$ NB(μ_2, τ_2), we continue to use the difference in sample means Δ, namely

$$T_{\mathbf{n}} \equiv \overline{X}_{1n_1} - \overline{X}_{2n_2} = n_1^{-1}\Sigma_{j=1}^{n_1} X_{1j} - n_2^{-1}\Sigma_{j=1}^{n_2} X_{2j} \text{ with } \mathbf{n} = (n_1, n_2),$$

to estimate Δ. We propose SEL for our estimation problem as follows:

$$\begin{aligned} L_{\mathbf{n}} \equiv L_{\mathbf{n}}(\overline{X}_{in_i}, \mu_i) &= b\left[(\overline{X}_{1n_1} - \overline{X}_{2n_2}) - \Delta\right]^2 \\ &= b\left[(\overline{X}_{1n_1} - \mu_1) - (\overline{X}_{2n_2} - \mu_2)\right]^2, \ b > 0, \ i = 1, 2. \end{aligned} \tag{4.7.1}$$

Next, we express the risk function associated with 4.7.1) as follows:

$$R_{\boldsymbol{n}} \equiv E_{\boldsymbol{\mu},\boldsymbol{\tau}}[L_{\boldsymbol{n}}] = b\left(\tfrac{\sigma_1^2}{n_1} + \tfrac{\sigma_2^2}{n_2}\right), \tag{4.7.2}$$

where one may recall that $\sigma_i^2 \equiv \mu_i + \tau_i^{-1}\mu_i^2, i = 1, 2.$

4.7.1. Purely Sequential Approaches

Again, our goal is to have $R_{\boldsymbol{n}}$ from (4.7.2) not to exceed $\omega(>0)$. We will minimize the total sample size $n \equiv n_1 + n_2$ subject to a restriction that we have $b\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \leq \omega$. Hence, the restricted optimization problem is as follows:

$$\text{Minimize } n \equiv n_1 + n_2 \text{ under the restriction } \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \leq \frac{\omega}{b}. \tag{4.7.3}$$

The solution to this problem reduces to the following optimal fixed-sample-size choices:

$$\begin{gathered} n_1 \equiv n_1^* = \tfrac{b}{\omega}\sigma_1(\sigma_1+\sigma_2),\ n_2 \equiv n_2^* = \tfrac{b}{\omega}\sigma_2(\sigma_1+\sigma_2), n_1^* n_2^{*-1} = \sigma_1\sigma_2^{-1}, \\ \text{and } n^* \equiv n_1^* + n_2^* = \tfrac{b}{\omega}(\sigma_1+\sigma_2)^2. \end{gathered} \tag{4.7.4}$$

One may refer to Mukhopadhyay and deSilva (2009, Section 13.5.1) for a proof. Now, since the sample sizes are unequal, we need to come up with a purely sequential sampling involving (i) an allocation scheme and (ii) a termination rule at every step of sampling. We will adopt an *allocation scheme* along the lines of Robbins et al. (1967) and Srivastava (1970) as follows:

An Allocation Scheme:

We first fix $m(\geq 2)$, $\gamma > \frac{1}{2}$ and then gather a set of pilot data $X_{1i}, X_{2i}, i = 1, 2, ..., m$ from the NB populations. Suppose that at some point, we have already recorded n_1, n_2 observations from the two NB populations respectively, but we require one more additional observation. The question is: Which population the next additional observation come from?

Since σ_1^2, σ_2^2 are both unknown, we respectively estimate them using the customary sample variances, namely,

$$S_{1n_1}^2 = (n_1-1)^{-1}\Sigma_{j=1}^{n_1}(X_{1j} - \overline{X}_{1n_1})^2,\ S_{2n_2}^2 = (n_2-1)^{-1}\Sigma_{j=1}^{n_2}(X_{2j} - \overline{X}_{2n_2})^2, n_1, n_2 \geq m.$$

We continue sampling by taking one additional observation, where the next observation would come from:

$$\begin{aligned} &\text{population NB}(\mu_1,\tau_1) \text{ if } \tfrac{n_1}{n_2} \leq \tfrac{S_{1n_1} + n_1^{-\gamma}}{S_{2n_2} + n_2^{-\gamma}}; \\ &\text{population NB}(\mu_2,\tau_2) \text{ if } \tfrac{n_1}{n_2} > \tfrac{S_{1n_1} + n_1^{-\gamma}}{S_{2n_2} + n_2^{-\gamma}}; \end{aligned} \tag{4.7.5}$$

which clearly mimics the ratio of the two optimal fixed sample sizes as shown in (4.7.4).

Purely Sequential Stopping Rules:

In what follows, we define a set of nearly equivalent stopping rules along the lines of Robbins et al. (1967) and Srivastava (1970) as follows:

Rule 1: In conjunction with the allocation scheme (4.7.5), we stop with the first total sample size $n = n_1 + n_2 (\geq 2m)$ when n_i observations from the i^{th} NB population have been taken, $i = 1, 2$, and,

$$n \geq \tfrac{b}{\omega} \left\{ \left(S_{1n_1} + S_{2n_2} \right)^2 + n^{-\gamma} \right\}, \tag{4.7.6}$$

in the light of the expression of n^* from (4.7.4).

Rule 2: In conjunction with the allocation scheme (4.7.5), we stop with the first total sample size $n = n_1 + n_2 (\geq 2m)$ when n_i observations from the i^{th} NB population have been taken, $i = 1, 2$, and,

$$\tfrac{S_{1n_1}^2 + n_1^{-\gamma}}{n_1} + \tfrac{S_{2n_2}^2 + n_2^{-\gamma}}{n_2} \leq \tfrac{\omega}{b}, \tag{4.7.7}$$

in the light of the expression from (4.7.3).

Rule 3: In conjunction with the allocation scheme (4.7.5), we stop with the first total sample size $n = n_1 + n_2 (\geq 2m)$ when n_i observations from the i^{th} NB population have been taken, $i = 1, 2$, and,

$$n_1 \geq \tfrac{b}{\omega} \left\{ S_{1n_1} \left(S_{1n_1} + S_{2n_2} \right) + n^{-\gamma} \right\} \text{ and } n_2 \geq \tfrac{b}{\omega} \left\{ S_{2n_2} \left(S_{1n_1} + S_{2n_2} \right) + n^{-\gamma} \right\}, \tag{4.7.8}$$

in the light of the expression of n_1^*, n_2^* from (4.7.4).

Purely Sequential Estimators of Δ:

Let $N^{(1)}, N^{(2)}, N^{(3)}$ denote the terminal total sample sizes corresponding to the stopping rules $1, 2, 3$ respectively from (4.9.6)-(4.9.8). We can easily check that $P_{\boldsymbol{\mu},\boldsymbol{\tau}} \left\{ N^{(j)} < \infty \right\} = 1, j = 1, 2, 3$. We can also verify:

$$N^{(1)} \leq N^{(2)} \leq N^{(3)} \text{ w.p.1} (P_{\boldsymbol{\mu},\boldsymbol{\tau}}). \tag{4.7.9}$$

Based on the fully gathered data

$$\left\{N^{(j)} \equiv N_1^{(j)} + N_2^{(j)}, X_{i1}, ..., X_{im,...}, X_{iN_i^{(j)}}\right\},$$

using the stopping rule j, $i = 1, 2$, $j = 1, 2, 3$, we propose to estimate:

$$\Delta \equiv \mu_1 - \mu_2 \text{ by } T_{\mathbf{N}^{(j)}} \equiv \overline{X}_{1N_1^{(j)}} - \overline{X}_{2N_2^{(j)}}, \text{ with } \mathbf{N}^{(j)} = (N_1^{(j)}, N_2^{(j)}),\ j = 1, 2, 3. \tag{4.7.10}$$

The following Theorem gives a set of attractive first-order asymptotic properties for the proposed purely sequential estimation methodologies (4.7.4) when $\mu_1, \mu_2, \tau_1, \tau_2$ *are all assumed unknown.*

Theorem 4.7.1. *With* $\mathbf{N}^{(j)} = (N_1^{(j)}, N_2^{(j)})$, *the loss function* $L_{\mathbf{N}^{(j)}}$ *in the spirit of (4.7.1), the allocation scheme (4.7.5), under any purely sequential estimation strategy* $\left(\mathbf{N}^{(j)}, T_{\mathbf{N}^{(j)}}\right)$ *from (4.7.6)-(4.7.8) and (4.7.10), for each fixed* $\boldsymbol{\mu} \in R^{+2}$ and $\boldsymbol{\tau} \in R^{+2}$ *we have as* $\omega \to 0$:

(*i*) $N^{(j)}/n^* \to 1$ *w.p.1*$(P_{\boldsymbol{\mu},\boldsymbol{\tau}})$;

(*ii*) $E_{\boldsymbol{\mu},\boldsymbol{\tau}}\left[(N^{(j)}/n^*)^s\right] \to 1$ *for all* $s > 0$, *if* $\gamma > \frac{1}{2}$[*asymptotic first-order efficiency*];

(*iii*) $E_{\boldsymbol{\mu},\boldsymbol{\tau}}\left[L_{\mathbf{N}^{(j)}}\right]/\omega \to 1$ *if* $\gamma > 1$[*asymptotic risk efficiency*];

where $n^* = n_1^* + n_2^*$ *comes from (4.7.4).*

4.7.2. Proof of Theorem 4.7.1

For brevity, we will provide a brief outline of the proof. Details are available in Ghosh et al. (1997, Sections 7.3, 8.3, pp. 222-223, pp. 256-259), Mukhopadhyay and de Silva (2009, Chapter 13), Robbins et al. (1967), and Srivastava (1970), and in other sources.

Part (*i*):

From (4.7.6), we write down the following inequality w.p.1$(P_{\boldsymbol{\mu},\boldsymbol{\tau}})$:

$$N^{(1)} \geq \tfrac{b}{\omega}\left\{\left(S_{1N_1^{(1)}} + S_{2N_2^{(1)}}\right)^2 + N^{(1)^{-\gamma}}\right\}. \tag{4.7.11}$$

Dividing (4.7.11) throughout by n^*, we write w.p.1$(P_{\boldsymbol{\mu},\boldsymbol{\tau}})$:

$$N^{(1)}/n^* \geq (\sigma_1 + \sigma_2)^{-2}\left\{\left(S_{1N_1^{(1)}} + S_{2N_2^{(1)}}\right)^2 + N^{(1)^{-\gamma}}\right\}. \tag{4.7.12}$$

Now, noting the facts that $N^{(1)} \to \infty$, $N_1^{(1)} \to \infty$, $N_2^{(1)} \to \infty$, $S_{1N_1^{(1)}} \to \sigma_1$, and $S_{2N_2^{(1)}} \to \sigma_2$ w.p.1($P_{\boldsymbol{\mu},\boldsymbol{\tau}}$) as $\omega \to 0$, we have:

$$\liminf_{\omega\to 0} N^{(1)}/n^* \geq 1 \text{ w.p.1}(P_{\boldsymbol{\mu},\boldsymbol{\tau}})$$

Since $N^{(1)} \leq N^{(2)} \leq N^{(3)}$, it follows that:

$$\liminf_{\omega\to 0} N^{(3)}/n^* \geq \liminf_{\omega\to 0} N^{(2)}/n^* \geq \liminf_{\omega\to 0} N^{(1)}/n^* \geq 1 \text{ w.p.1}(P_{\boldsymbol{\mu},\boldsymbol{\tau}}). \tag{4.7.13}$$

Next, from (4.7.8) we have the following inequalities w.p.1($P_{\boldsymbol{\mu},\boldsymbol{\tau}}$):

$$\begin{aligned} N_1^{(3)}/n^* &\leq (\sigma_1+\sigma_2)^{-2}\left\{S_{1N_1^{(3)}-1}\left(S_{1N_1^{(3)}-1}+S_{2N_2^{(3)}-1}\right)+\left(N_1^{(3)}-1\right)^{-\gamma}\right\}+m/n^* \text{ and} \\ N_2^{(3)}/n^* &\leq (\sigma_1+\sigma_2)^{-2}\left\{S_{2N_2^{(3)}-1}\left(S_{1N_1^{(3)}-1}+S_{2N_2^{(3)}-1}\right)+\left(N_2^{(3)}-1\right)^{-\gamma}\right\}+m/n^*. \end{aligned} \tag{4.7.14}$$

Combining the inequalities from (4.7.14) we can write w.p.1($P_{\boldsymbol{\mu},\boldsymbol{\tau}}$):

$$N^{(3)}/n^* \leq (\sigma_1+\sigma_2)^{-2}\left\{\left(S_{1N_1^{(3)}-1}+S_{2N_2^{(3)}-1}\right)^2+\left(N_1^{(3)}-1\right)^{-\gamma}+\left(N_2^{(3)}-1\right)^{-\gamma}\right\}+m/n^*, \tag{4.7.15}$$

which clearly shows that

$$\limsup_{\omega\to 0} N^{(3)}/n^* \leq 1 \text{ w.p.1}(P_{\boldsymbol{\mu},\boldsymbol{\tau}}).$$

Now, since $N^{(1)} \leq N^{(2)} \leq N^{(3)}$ w.p.1($P_{\boldsymbol{\mu},\boldsymbol{\tau}}$), it follows that

$$\limsup_{\omega\to 0} N^{(1)}/n^* \leq \limsup_{\omega\to 0} N^{(2)}/n^* \leq \limsup_{\omega\to 0} N^{(3)}/n^* \leq 1 \text{ w.p.1}(P_{\boldsymbol{\mu},\boldsymbol{\tau}}). \tag{4.7.16}$$

Proof of part (i) is complete by combining (4.7.13) and (4.7.16).

Part (ii):

From (4.7.15), we have the following inequality w.p.1($P_{\boldsymbol{\mu},\boldsymbol{\tau}}$) with large enough n^*:

$$
\begin{aligned}
N^{(3)}/n^* &\le (\sigma_1+\sigma_2)^{-2}\left\{\left(S_{1N_1^{(3)}-1}+S_{2N_2^{(3)}-1}\right)^2+2\right\}+m \\
&\le (\sigma_1+\sigma_2)^{-2}\sup\nolimits_{n_1,n_2\ge 2}\left\{\left(S_{1n_1-1}+S_{2n_2-1}\right)^2+2\right\}+m.
\end{aligned}
\tag{4.7.17}
$$

Now, denoting $\sup_{n_1\ge 2,n_2\ge 2}\left\{\left(S_{1n_1-1}+S_{2n_2-1}\right)^2\right\}$ by W, we can claim w.p.1($P_{\boldsymbol{\mu},\boldsymbol{\tau}}$):

$$
N^{(3)}/n^* \le (\sigma_1+\sigma_2)^{-2}\{W+2\}+m, \tag{4.7.18}
$$

from (4.7.17).

Since the right-hand side of (4.7.18) is free from ω, using Wiener's (1939) ergodic theorem we can claim uniform integrability of all positive powers of N/n^*. Next, appealing to part (i), we complete the proof. Here, $\gamma > \frac{1}{2}$ suffices.

Part (iii): A proof can be improvised along the lines of how Ghosh and Mukhopadhyay (1979) handled their distribution-free two-sample problem. One may also refer to Ghosh et al. (1997, pp. 280-284). Further details are omitted. ■

Remark 4.7.1. (A k-Sample consideration). In the case of a k-sample scenario, one could set out to estimate the parameter Δ that is defined as $\Sigma_{i=1}^k d_i\mu_i$ by the estimator $T_n \equiv \Sigma_{i=1}^k d_i\overline{X}_{in}$ where d_i's are assumed non-zero constants. Under the corresponding SEL function in the spirit of (3.1), we could develop an associated purely sequential bounded risk estimation strategy and a set of desirable asymptotic first-order properties along the lines of Theorem 5.1. Obviously, an appropriate allocation scheme will become much more complicated than (4.7.5). One may refer to Mukhopadhyay and Liberman (1989) for some insight. Further details are left out for brevity.

Remark 4.7.2. One could think of implementing both allocation and stopping rules (4.7.5)-(4.7.8) by replacing $S_{in_i}^2$ with $\widehat{\mu}_{in_i,\text{MLE}}+\widehat{\tau}_{in_i,\text{MLE}}^{-1}\widehat{\mu}_{in_i,\text{MLE}}^2$ where $\widehat{\mu}_{in_i,\text{MLE}}$ and $\widehat{\tau}_{in_i,\text{MLE}}$ are respectively the maximum likelihood estimators (MLE) of μ_i,τ_i obtained from $X_{i1},...,X_{in_i}, n_i \ge m, i = 1,2$. In other words, at every stage of sampling one would update $\widehat{\mu}_{in_i,\text{MLE}}$ and $\widehat{\tau}_{in_i,\text{MLE}}$ sequentially. We *conjecture* that purely sequential estimation strategies revised along those lines will continue to remain asymptotically first-order efficient and risk-efficient. At the same time, however, the small and moderate sample performances will remain nearly similar without any appreciable changes. But, such revised purely sequential estimation strategies will become extremely time-consuming.

4.7.3. Simulations and Real Data Illustrations

We begin with a summary (Section 4.7.3.1) from a set of simulation studies to examine the performances of our proposed purely sequential estimation strategies (4.7.5)-(4.7.8) and (4.7.10). We fixed the values of the risk bound ω which gave us the values of n_1^*, n_2^* and hence the final optimal fixed-sample-size n^*. Section 4.7.3.2 highlights performances of our estimation strategies using real raptor count data from the Hawk Mountain sanctuary.

4.7.3.1. A Summary from Simulations

We generated pseudorandom observations from the distribution (4.1.1) with some combinations of choices for $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$. We fixed the values $b = 1$, $m = 5$, and $\gamma = 1.5$. Each row in Table 4.14 corresponds to averages from 10000 replications which were run under a given configuration. The results are shown under the following settings:

$$\boxed{\omega = 0.05, 0.1 \text{ and } (\mu_1, \mu_2, \tau_1, \tau_2) = (3, 1, 3, 2).} \tag{4.7.19}$$

We simply use notations in Table 4.14 which are very similar to what we had used earlier, but naturally with obvious changes that must reflect having unequal sample sizes.

We have observed clearly that $\overline{n}^{(1)} < \overline{n}^{(2)} < \overline{n}^{(3)}$ in all our simulations. This feature is consistent with our theoretical claim stated in (4.7.9). For brevity, we omit other comments since they line up with our previous comments made in the contexts of Tables 4.10 and 4.12.

4.7.3.2. Illustration 4.7.1: Raptor Count Data

We again return to use the raptor count data in order to highlight our real data Illustration 4.7.1 in Table 4.15. As we did earlier, we applied the methodologies (4.7.5)-(4.7.8) and (4.7.10) on two of the species of raptors, *American Kestrel* (X_1) and *Golden Eagle* (X_2) from the full set of data without replacement. We found that with or without replacement made practically no difference.

However, to illustrate varying sample sizes, we use X_1 data only from a single year (2016). The datasets thus consisted of 179 and 357 rows respectively for X_1, X_2 and NB fits were seen rather appropriate with p-values of 0.34 and 0.43 respectively. Treating these two datasets as the universe, we found $\widehat{\mu}_1 = 1.28$, $\widehat{\tau}_1 = 0.21$, $\widehat{\mu}_2 = 0.7$, $\widehat{\tau}_2 = 0.18$.

Table 4.15 shows real data Illustration 4.7.1. Each block in Table 4.15 shows results from stopping rules $N^{(j)}$, $j = 1, 2, 3$. We fixed $\gamma = 1.5$ along with $b = 1$ and $m = 5$. We also fixed certain values for our risk bound ω which are shown in the table. Column 3 shows $\overline{\Delta}^{(j)} = \overline{x}_1^{(j)} - \overline{x}_2^{(j)}$ values which estimate $\Delta = 0.58$, $j = 1, 2, 3$. In column 8 of Table 4.15, we show a value $\widetilde{z}$ obtained from a single run:

$$N^{(j)} = n^{(j)},\ \widetilde{r}^{(j)} = b\left(\frac{\overline{x}_1^{(j)} + \tau_1^{(j)-1}\overline{x}_1^{(j)2}}{n_1^{(j)}} + \frac{\overline{x}_2^{(j)} + \tau_2^{(j)-1}\overline{x}_2^{(j)2}}{n_2^{(j)}}\right) \text{ in the spirit of (4.5.17)} \\ \text{under one replication, so that } \widetilde{z}^{(j)} = \widetilde{r}^{(j)}/\omega,\ j = 1, 2, 3. \tag{4.7.20}$$

We show $\widetilde{z}$ mainly to grasp a sense of how close the estimated risk may or may not be when compared with the preset goal ω. A slight erratic behavior is due to the fact that $\widetilde{z}$ values were obtained from single runs. Columns 4,5,6 in Table 4.15 give the estimated individual sample sizes $n_1^{(j)}, n_2^{(j)}$ and the estimated total sample size $n^{(j)}$, for each rule $N^{(j)}$, $j = 1, 2, 3$.

These can be seen close to the corresponding n_1^*, n_2^* and n^* values respectively which are shown exclusively for purposes of comparisons. We again notice that $n^{(1)} < n^{(2)} < n^{(3)}$ under each configuration, that is a feature which is consistent with our theoretical claim stated in (4.7.9).

4.8. A BRIEF SUMMARY OF CHAPTER 4

We have proposed and developed a set of purely sequential methodologies to estimate a negative binomial mean μ, under several useful loss functions both when the thatch parameter $\tau(> 0)$ may be assumed known or unknown. We should emphasize that in the case of sequential point estimation problems for μ when τ remains unknown, the literature has been rather scarce. One may look back at Mukhopadhyay and Diaz (1985) for a lone treatment in this case and we have not been aware of any other substantial approaches. Sections 4.4 and 4.7 have proposed and investigated new and substantial sequential estimation strategies for μ when τ remains unknown. Illustrations using both simulations and real datasets put this contribution in a proper perspective.

4.8.1. Specific Thoughts on Section 4.2

In Section 4.2, we noted a straightforward and naturally arriving positive and known lower bound for n^* in (4.2.6) which led to the specific choice of a pilot sample size m from (4.2.7) used in Section 4.2. This lower bound for n^* was so noted first in Willson and Folks (1983) in developing their purely sequential methodology. However, the same lower bound for n^* led Mukhopadhyay and Diaz (1985) to propose an associated two-stage estimation methodology.

In a different vein, Mukhopadhyay and Duggan (1997) developed a remarkable two-stage fixed-width confidence interval methodology for the normal mean when the unknown population variance had a known positive lower bound. It was truly remarkable because Mukhopadhyay and Duggan (1997) could develop asymptotic second-order properties for an appropriately modified two-stage estimation methodology. It was a clear vindication of Stein's (1945,1949) original two-stage fixed-width confidence interval methodology.

The proliferation of the core ideas from Mukhopadhyay and Duggan (1997) in many directions has been rather widespread and that continues to spread in areas including big data problems as well as small n large p problems. For brevity, we only mention some of the important references in order to connect the dots: Mukhopadhyay and Aoshima (1998), Aoshima and Mukhopadhyay (1998,1999), Mukhopadhyay (1999a,b), Mukhopadhyay and Duggan (2000,2001), Aoshima and Takada (2000), and Aoshima and Yata (2010).

We could easily develop appropriate two-stage methodologies in Section 4.2 in the light of Mukhopadhyay and Diaz (1985). However, such methodologies and techniques would be well-understood by now and their second-order asymptotic efficiency properties would also clearly follow along the lines of the detailed asymptotic second-order analysis from Mukhopadhyay and de Silva (2005) developed in the context of the NB problem of Mukhopadhyay and Diaz (1985). Thus, in Sections 4.2 and 4.3, we keep out the corresponding two-stage estimation strategies primarily for brevity.

4.8.2. Specific Thoughts on Sections 4.3

In Section 4.3, a straightforward and natural positive and known lower bound m for n^* such that $m \to \infty$ and $m/n^* = O(1)$ was **not** available. This is in direct contrast with our Section 4.2.

But, from the proof of Theorem 4.3.1, one notes that the terminal sample size N was at least $(b/\omega)^{1/(1+\gamma)}$ w.p.1. Thus, one could immediately propose appropriately modified two-stage estimation strategies in the spirit of Mukhopadhyay (1980). See also Mukhopadhyay and Solanky (1994, pp. 25-27), Ghosh et al. (1997, pp.156-157), and Mukhopadhyay and de Silva (2009, pp. 114-118) for details. However, since such two-stage estimation methodologies will fail to achieve asymptotic second-order properties, associated discussions are again kept out for brevity.

4.8.3. Specific Thoughts on Section 4.4

Going back to Section 4.4, one could think of stopping rules different from (4.4.4) by replacing the sample variance S_n^2 with the following estimator:

$$\widehat{\mu}_{n,\mathrm{MLE}} + \widehat{\tau}_{n,\mathrm{MLE}}^{-1}\widehat{\mu}_{n,\mathrm{MLE}}^2,$$

where $\widehat{\mu}_{n,\mathrm{MLE}}$ and $\widehat{\tau}_{n,\mathrm{MLE}}$ are respectively the maximum likelihood estimators (MLE) of μ, τ obtained from $X_1, ..., X_n, n \geq m$. In other words, at every stage of sampling one must update $\widehat{\mu}_{n,\mathrm{MLE}}$ and $\widehat{\tau}_{n,\mathrm{MLE}}$ sequentially. But, then such revised versions of sequential estimation strategies will become extremely time-consuming.

We *conjecture* that those revised sequential estimation strategies will be asymptotically first-order efficient and risk-efficient. At the same time, however, the small and moderate sample performances will remain nearly similar without any appreciable changes. Hence, in Section 4.4, we continued using S_n^2 in defining the stopping boundaries for its own versatility, for example, the sample variance is a U-statistic.

4.8.4. Specific Thoughts on Sections 4.5-4.7

Similar to earlier thoughts, one can notice that in Section 4.5, we have a natural, positive and known lower bound for n^* which led to a specific choice of pilot sample size m. One can hence construct an appropriate two-stage methodology along the lines of Mukhopadhyay and Diaz (1985) which may be expected to have asymptotic second-order properties along the lines of Mukhopadhyay and de Silva (2005).

Also, in Sections 4.6-4.7, one may notice that the terminal sample size N is at least $O\left(\omega^{1/(1+\gamma)}\right)$ and one can easily construct appropriate modified two-stage estimation strategies along the lines of Mukhopadhyay (1980). One should refer to Mukhopadhyay and Solanky (1994, pp. 25-27), Ghosh et al. (1997, pp. 156-157), and Mukhopadhyay and de Silva (2009, pp. 114-118) for further details. These two-stage strategies however will fail to achieve asymptotic second-order properties and hence we leave them out for brevity.

Table 4.1. Simulation results from $10,000$ replications with m from (4.2.7), $a = 1$, $\gamma = 1.5$

μ	τ	n^*	ω	$\overline{x}$ $s_{\overline{x}}$	$\overline{n}$ $s_{\overline{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$
Our purely sequential procedure (4.2.8)							
2	3	50	0.0083	2.0178 0.0026	51.65 0.0683	1.0330	1.0155 9.72×10^{-6}
		200	0.0020	2.0032 0.0013	201.70 0.1275	1.0085	1.0084 1.24×10^{-6}
3	4	50	0.0058	3.0178 0.0032	51.21 0.0504	1.0242	1.0072 5.32×10^{-6}
		200	0.0014	3.0080 0.0016	201.33 0.0973	1.0066	1.0040 6.79×10^{-7}
Sequential procedure by Willson and Folks (1983)							
2	3	50	0.0083	1.9989 0.0025	50.59 0.0392	1.0119	0.9940 6.30×10^{-6}
		200	0.0020	2.0010 0.0012	200.44 0.0770	1.0022	0.9992 7.97×10^{-7}
3	4	50	0.0058	2.9976 0.0032	50.46 0.0310	1.0092	0.9945 3.55×10^{-6}
		200	0.0014	2.9993 0.0016	200.59 0.0619	1.0029	0.9979 4.48×10^{-7}

Capsella bursa-pastoris L.

Polygonum aviculare L.

Figure 4.1. Pictures of the types of weed included in the weed count data under consideration.

Photo Credits: Max Licher (photographer) and SCINet

Table 4.2. Analysis of weed count data using purely sequential strategy (4.2.8) assuming $\tau = 3.98$ with $a = 1$, $\gamma = 1.5$

n^*	ω	$\hat{\mu}$: $\overline{x}_n$	n	n/n^*	$\widetilde{z}$
75	0.0182	0.91	82	1.09	1.2069
200	0.0068	0.44	151	0.75	1.1339
500	0.0027	0.41	607	1.21	0.9052
700	0.0019	0.42	782	1.11	1.0078

Table 4.3. Analysis of woodlarks data using purely sequential strategy (4.2.8) assuming $\tau = 0.23$ with $a = 1$, $\gamma = 1.5$

n^*	ω	$\hat{\mu}$: $\overline{x}_n$	n	n/n^*	$\widetilde{z}$
30	0.0779	4.03	32	1.06	0.9218
40	0.0584	2.78	41	1.02	0.9830
50	0.0467	3.15	51	1.02	0.9794

Table 4.4. Simulation results from $10,000$ replications of the purely sequential procedure (4.3.4) with $a = 1$, $\gamma = 1.5$, $m = 5$

μ	τ	n^*	ω	$\overline{x}$ / $s_{\overline{x}}$	$\overline{n}$ / $s_{\overline{n}}$	$\overline{n}/n^*$	$\overline{z}$ / $s_{\overline{z}}$
2	3	50	0.0667	2.0035	49.53	0.9907	1.0626
				0.0026	0.0943		2.58×10^{-4}
		200	0.0167	2.0001	199.37	0.9968	1.0148
				0.0013	0.1817		5.61×10^{-5}
3	4	50	0.1050	3.0056	49.90	0.9981	1.0307
				0.0032	0.0783		2.11×10^{-4}
		200	0.0262	3.0006	200.14	1.0007	1.0055
				0.0016	0.1568		2.11×10^{-5}

Table 4.5. Analysis of weed count data using purely sequential strategy (4.3.4) assuming $\tau = 3.98$ with $b = 1$, $\gamma = 1.5$, $m = 10$

n^*	ω	$\hat{\mu}$: $\overline{x}_n$	n	n/n^*	$\widetilde{z}$
75	0.0043	0.28	74	0.98	0.8350
200	0.0016	0.32	210	1.05	1.1332
500	0.0006	0.32	540	1.08	1.2116
700	0.0004	0.28	667	0.95	1.2524

Table 4.6. Analysis of woodlarks data using purely sequential strategy (4.3.4) assuming $\tau = 0.23$ with $b = 1$, $\gamma = 1.5$, $m = 7$

n^*	ω	$\hat{\mu}$: $\overline{x}_n$	n	n/n^*	$\widetilde{z}$
30	1.4498	2.61	31	1.03	0.7490
40	1.0873	2.84	39	0.97	0.8939
50	0.8699	3.47	46	0.92	1.3950

Table 4.7. Simulation results from $10,000$ replications of the purely sequential procedure (4.4.4) with $b = 1$, $\gamma = 1.5$, $m = 5$

μ	τ	n^*	ω	$\overline{x}$ $s_{\overline{x}}$	$\overline{n}$ $s_{\overline{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$
2	3	50	0.0667	1.9989 0.0029	46.15 0.1491	0.9231	1.2634 4.72×10^{-4}
		200	0.0167	1.9972 0.0013	197.69 0.2939	0.9884	1.0358 2.79×10^{-5}
3	4	50	0.1050	3.0036 0.0036	46.10 0.1438	0.9220	1.2731 8.54×10^{-4}
		200	0.0262	2.9995 0.0016	196.92 0.2758	0.9846	1.0369 4.11×10^{-5}

Table 4.8. Analysis of weed count data using purely sequential procedure (4.4.4) with $b = 1$, $\gamma = 1.5$, $m = 10$

n^*	ω	$\hat{\mu}$: $\overline{x}_n$	n	n/n^*	$\widetilde{z}$
75	0.0148	1.06	65	0.87	1.3860
200	0.0055	0.87	224	1.12	0.8406
500	0.0022	0.92	538	1.07	0.9584
700	0.0015	0.87	765	1.09	0.9186

Table 4.9. Analysis of woodlarks data using purely sequential procedure (4.4.4) with $b = 1$, $\gamma = 1.5$, $m = 7$

n^*	ω	$\hat{\mu}$: $\overline{x}_n$	n	n/n^*	$\widetilde{z}$
30	1.4498	2.87	27	0.90	0.9881
40	1.0873	2.54	28	0.70	1.0047
50	0.8699	3.10	59	1.18	0.8744

Table 4.10. Simulation results for the purely sequential estimation methodology (4.5.7) from $10,000$ replications when $k = 2$ with m from (4.5.6), $a_1 = 1$, $a_2 = 1.2$, $\gamma = 1.5$

(μ_1, μ_2)	(τ_1, τ_2)	n^*	ω	$\overline{x}_1$ / $s_{\overline{x}_1}$	$\overline{x}_2$ / $s_{\overline{x}_2}$	$\overline{n}$ / $s_{\overline{n}}$	$\overline{n}/n^*$	$\overline{z}$ / $s_{\overline{z}}$
$(3, 1)$	$(3, 2)$	50	0.0246	3.0014	0.9994	51.06	1.0213	0.9844
				0.0034	0.0017	0.0380		1.77×10^{-5}
		200	0.0061	3.0018	1.0004	201.12	1.0056	0.9957
				0.0017	0.0008	0.0745		2.25×10^{-6}
$(7, 3)$	$(2, 1)$	50	0.0225	7.0056	3.0046	50.73	1.0147	0.9862
				0.0078	0.0048	0.0138		6.01×10^{-6}
		200	0.0056	6.9949	2.9984	200.75	1.0037	0.9964
				0.0039	0.0024	0.0265		7.42×10^{-7}

Table 4.11. Analysis of raptor count data when $k = 2$ using purely sequential estimation strategy (4.5.7) with m from (4.6.6) assuming $\tau_1 = 0.28$, $\tau_2 = 0.18$ and $a_1 = 1$, $a_2 = 1.2$, $\gamma = 1.5$

n^*	ω	$\hat{\mu}_1$: $\overline{x}_{1n}$	$\hat{\mu}_2$: $\overline{x}_{2n}$	n	n/n^*	$\widetilde{z}$: (4.5.16)
50	0.1313	1.57	0.61	57	1.14	0.8772
100	0.0656	1.41	0.76	101	1.01	0.9900
200	0.0328	1.48	0.66	201	1.00	0.9950

Table 4.12. Simulation results for the purely sequential estimation methodology (4.6.4)-(4.6.5) from $10,000$ replications with $m = 5$, $b = 1$, $\gamma = 1.5$

(μ_1, μ_2)	(τ_1, τ_2)	n^*	ω	$\overline{x}_1$ $s_{\overline{x}_1}$	$\overline{x}_2$ $s_{\overline{x}_2}$	$\overline{\Delta}$ $s_{\overline{\Delta}}$	$\overline{n}$ $s_{\overline{n}}$	$\overline{n}/n^*$	$\overline{z}$ $s_{\overline{z}}$
$(3, 1)$	$(3, 2)$	50	0.15	3.0016 0.0034	1.0003 0.0017	2.0013 0.0025	50.09 0.0732	1.0018	1.0219 2.55×10^{-4}
		200	0.0375	3.0036 0.0017	1.0001 0.0008	2.0035 0.0012	200.05 0.1465	1.0002	1.0051 2.81×10^{-5}
$(7, 3)$	$(2, 1)$	50	0.87	6.9945 0.0080	2.9959 0.0049	3.9986 0.0068	50.09 0.0828	1.0019	1.0291 1.84×10^{-3}
		200	0.2175	6.9988 0.0039	3.0040 0.0024	3.9948 0.0030	199.75 0.1653	0.9987	1.0082 1.85×10^{-4}

Table 4.13. Analysis of raptor count data using purely sequential estimation strategy (4.6.4)-(4.6.5) assuming $\tau_1 = 0.28$, $\tau_2 = 0.18$ with $b = 1$, $\gamma = 1.5$

n^*	ω	$\hat{\mu}_1$: $\overline{x}_{1n}$	$\hat{\mu}_2$: $\overline{x}_{2n}$	$\overline{\Delta}$	n	n/n^*	$\widetilde{z}$: (4.6.10)
50	0.2431	1.60	0.58	1.02	41	0.82	1.2195
100	0.1215	1.68	0.71	0.97	108	1.08	0.9259
200	0.0607	1.47	0.82	0.65	221	1.10	0.9049

Table 4.14. Simulation results for the purely sequential estimation methodologies (4.7.5)-(4.7.8) and (4.7.10) from 10,000 replications with $(\mu_1, \mu_2) = (3, 1)$, $(\tau_1, \tau_2) = (3, 2)$, $m = 5$, $b = 1$, $\gamma = 1.5$

Rule j Eq #	$\overline{x}_1^{(j)}$ $s_{\overline{x}_1^{(j)}}$	$\overline{x}_2^{(j)}$ $s_{\overline{x}_2^{(j)}}$	$\overline{\Delta}^{(j)}$ $s_{\overline{\Delta}^{(j)}}$	$\overline{n}_1^{(j)}$ $s_{\overline{n}_1^{(i)}}$	$\overline{n}_2^{(j)}$ $s_{\overline{n}_2^{(i)}}$	$\overline{n}^{(j)}$ $s_{\overline{n}^{(i)}}$	$\overline{n}^{(j)}/n^*$ $s_{\overline{n}^{(i)}}$	$\overline{z}^{(j)}$ $s_{\overline{z}^{(i)}}$
			$\omega = 0.05$, $n_1^* = 180$, $n_2^* = 90$, $n^* = 270$					
$j = 1$	3.0041	1.0058	1.9983	189.35	85.23	274.58	1.0169	1.0358
Eq (4.7.6)	0.0065	0.0035	0.0051	1.9589	1.0091	1.1151		0.0026
$j = 2$	3.0031	1.0050	1.9981	196.86	87.47	284.33	1.0530	1.0605
Eq (4.7.7)	0.0069	0.0047	0.0059	2.0358	1.1025	1.2384		0.0030
$j = 3$	3.0033	1.0042	1.9991	196.96	89.91	286.87	1.0625	1.1121
Eq (4.7.8)	0.0078	0.0036	0.0070	2.1254	1.3002	1.5488		0.0035
			$\omega = 0.1$, $n_1^* = 90$, $n_2^* = 45$, $n^* = 135$					
$j = 1$	3.0068	1.0090	1.9978	91.76	46.38	138.14	1.0233	0.9924
Eq (4.7.6)	0.0028	0.0020	0.0014	1.2121	0.5247	0.7122		0.0034
$j = 2$	3.0103	1.0121	1.9982	92.51	49.87	142.38	1.0547	0.9789
Eq (4.7.7)	0.0054	0.0037	0.0058	1.3029	0.6069	0.8098		0.0036
$j = 3$	3.0135	1.0144	1.9991	97.23	51.05	148.28	1.0984	0.9712
Eq (4.7.8)	0.0059	0.0048	0.0064	1.4781	0.7814	0.9947		0.0035

Table 4.15. Analysis of raptor count data using purely sequential estimation methodologies (4.7.5)-(4.7.8) and (4.7.10) with $m = 5$, $b = 1$, $\gamma = 1.5$

Rule j Eq #	$\overline{x}_1^{(j)}$	$\overline{x}_2^{(j)}$	$\overline{\Delta}^{(j)}$	$n_1^{(j)}$	$n_2^{(j)}$	$n^{(j)}$	$n^{(j)}/n^*$	$\widetilde{z}^{(j)}$ (4.9.20)
		$\omega = 0.1,\ n_1^* = 89.97,\ n_2^* = 146.56,\ n^* = 236.53$						
$j = 1$ Eq (4.7.6)	1.3021	0.6568	0.6453	79.09	138.77	217.86	0.9211	1.3100
$j = 2$ Eq (4.7.7)	1.3212	0.6920	0.6292	95.82	144.92	240.74	1.0178	1.2018
$j = 3$ Eq (4.7.8)	1.3844	0.7215	0.6629	97.37	150.89	248.26	1.0496	1.0987
		$\omega = 0.2,\ n_1^* = 44.98,\ n_2^* = 73.28,\ n^* = 118.26$						
$j = 1$ Eq (4.7.6)	1.2608	0.6921	05687	37.36	69.02	106.38	0.8996	1.4252
$j = 2$ Eq (4.7.7)	1.3258	0.7412	0.5846	48.20	74.88	123.08	1.0408	1.2214
$j = 3$ Eq (4.7.8)	1.3378	0.7503	0.5875	53.17	76.30	129.47	1.0948	1.1287

Chapter 5
Concluding Thoughts

This thesis involves problems surrounding two important but interesting distributions, namely negative exponential and negative binomial. We have designed appropriate multistage methodologies to estimate the unknown parameters under each of these distributions. One of the loss functions we have made use of is the Linex loss which was introduced by Varian in 1975. This loss is asymmetric in nature and addresses estimation error by penalizing over-estimation and under-estimation unequally.

Chapter 2 covered two-stage and purely sequential estimation of a negative exponential location under a Linex loss. The existing literature was rather scarce and mainly concentrated on a normal distribution and/or a squared error loss function. Chattopadhyay (1998) first developed a sequential estimation problem under a Linex loss. As seen in this chapter, a negative exponential model proves to be useful in many reliability problems such as to depict the failure times of electrical components among many others. It is also used widely to model survival data under different situations. We developed a modified two-stage procedure along the lines of Mukhopadhyay and Duggan (1997) which proved to have exciting properties as opposed to the usual Stein type two-stage methodology. Our proposed methods were also seen to be both operationally convenient and theoretically sound. They were supported by analyses from simulations and real data, namely infant mortality and bone marrow data and were seen to perform remarkably well.

In Chapter 3 we decided to extend our ideas to a two-sample problem which involved multistage estimation of the difference in locations of two independent negative exponential populations. It was again carried out under an appropriate Linex loss. An interesting aspect to this problem was that we assumed that the scale parameters are σ and $b\sigma$, where b is known a'priori. A motivational scenario was also introduced to support this problem structure. Our proposed methodologies were again seen to enjoy interesting efficiency and consistency properties. These were applied to real datasets from cancer studies and reliability analysis and were seen to perform exceptionally well.

Chapter 4 involved multistage estimation under a different distribution, namely negative binomial. Being a discrete distribution, it becomes operationally different than the negative exponential.

It was first parametrized by Anscombe in 1949, 1950 and is heavily applied to model count data. It proves to handle over-dispersion more adequately than a Poisson model. Willson and Folks (1983) and Willson et al. (1984) contains a wide variety of sequential problems arising from estimation of a NB mean. We developed appropriate purely sequential methodologies to estimate 1) a NB mean associated with a single sample, 2) a k-mean vector of NB means and 3) difference in means of two independent NB populations. The estimation was again carried out under appropriate Linex loss and was also compared under a squared error loss. Different situations pertaining to known or unknown thatch parameter/s were discussed. All of our proposed methodologies were seen to enjoy interesting efficiency and consistency properties. These were heavily supported by real datasets from ecology, namely weed count data, bird count data and raptor count data. For completeness, we also discussed possible modifications and ramifications to construct a two-stage methodology under different situations.

Chapter 6
Future Research Directions

6.1. THREE-STAGE AND ACCELERATED SEQUENTIAL STOPPING RULES

We have witnessed a number of multistage and or purely sequential sampling strategies throughout this thesis. These mainly comprised of two-stage, modified two-stage and purely sequential stopping rules. Following along these lines, one can extend further to construct three-stage and accelerated sequential methodologies under appropriate situations. Three-stage methodologies were largely developed by Hall (1981). and discussed in Mukhopadhyay (1990). Hall (1983) built ideas on an accelerated sequential rule and outlined some basic properties. Mukhopadhyay and Solanky (1991) further built a unified theory. Both of these methods prove to have their advantages over two-stage and purely sequential rules in some aspects. They are much more operationally convenient to apply and enjoy interesting efficiency properties at the same time. We hope to construct these methodologies to estimate the location parameter of a negative exponential distribution, again under an appropriate Linex loss function. One can also extend these further to a two-sample scenario to estimate the difference in location parameters of two independent negative exponential distributions. We can support our sampling designs with interesting applications from health studies.

6.2. NON-PARAMETRIC METHODS

A vast literature has been developed surrounding distribution free or non-parametric methodologies concerned with estimation. One may refer to Mukhopadhyay (1978), Ghosh and Mukhopadhyay (1979), Sen (1981,1985) for more detailed literature. These methods do not assume any crucial characteristics of any distribution under concern. Let us assume that we have a sequence of independent observations $X_1, X_2, ..., X_n, ...$ from a population having a distribution function F. We hope to construct suitable multistage stopping rules solely for estimation purposes, where we assume very little about the distribution function F. Point estimation can be carried out under specified loss functions including the squared error or Linex loss. These methodologies hope to revolve around appropriate U-statistics which would estimate the unknown parameters.

6.3. CHANGE POINT DETECTION PROBLEMS

The field of change point problems is well developed over the years. We still hope to cover untrodden grounds in relation with these. Let us assume that $X_1, X_2, ..., X_n$ be an independent sequence and τ be a change point such that $X_1, ..., X_{\tau-1} \sim NExpo(\mu_0, \sigma_0)$ and $X_\tau, ..., X_n \sim NExpo(\mu_1, \sigma_1)$. We also assume that $\mu_1 > \mu_0$. The underlying problem could be to test and detect the changepoint τ, and draw inferences such as estimation of the unknown parameters both before and after the change.

List of Tables

List of Figures

Bibliography

Anscombe, F. J. (1949). The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution, *Biometrics* 5: 165-173.

Anscombe, F. J. (1950). Sampling Theory of the Negative Binomial and Logarithmic Series Distributions, *Biometrika* 37: 358-382.

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

Aoshima, M. and Mukhopadhyay, N. (1998). Fixed-Width Simultaneous Confidence Intervals for Multinormal Means in Several Intraclass Correlation Models, *Journal of Multivariate Analysis* 66: 46-63.

Aoshima, M. and Mukhopadhyay, N. (1999). Second-Order Properties of a Two-Stage Fixed-Size Confidence Region When the Covariance Matrix Has a Structure, *Statistical Inference and Data Analysis* (Tokyo, 1997), *Communications in Statistics - Theory & Methods* 28: 839–855.

Aoshima, M., Mukhopadhyay, N., and Kobayashi, Y. (2011). Two-Stage Procedures for Estimating the Difference of Means when the Sampling Cost is Different, *Sequential Analysis* 30: 160-171.

Aoshima, M. and Takada, Y. (2000). Second-Order Properties of a Two-Stage Procedure for Comparing Several Treatments with a Control, *Journal of Japan Statistical Society* 30: 27-41.

Aoshima, M. and Yata, K. (2010). Asymptotic Second-Order Consistency for Two-Stage Estimation Methodologies and Its Applications, *Annals of Institute of Statistical Mathematics* 62: 571-600.

Bain, L. J. (1978). *Statistical Analysis of Reliability and Life Testing Models*, New York: Dekker.

Balakrishnan, N. and Basu, A. P. (1995). *The Exponential Distribution: Theory, Methods and Application*, edited volume, Amsterdam: Gordon and Breach.

Banerjee, S. and Mukhopadhyay, N. (2016). A General Sequential Fixed-Accuracy Confidence Interval Estimation Methodology for a Positive Parameter: Illustrations Using Health and Safety Data, *Annals of Institute of Statistical Mathematics* 68: 541-570. DOI: 10.1007/s10463-015-0504-2

Baran, J. and Magiera, R. (2010). Optimal Sequential Estimation Procedures of a Function of a Probability of Success Under a Linex Loss, *Statistical Papers* 51: 511. doi: 10.1007/s00362-008-0137-0.

Barrigossi, J. A. F. (1997). Development of an IPM System for the Mexican Bean Beetle (*Epilachna varivestis Mulsant) A*s a Pest of Dry Bean (*Phaseolus vulgaris L.*), Ph.D. diss., University of Nebraska-Lincoln.

Basu, A. P. (1971). On a Sequential Rule for Estimating the Location Parameter of an Exponential Distribution, *Naval Research Logistics Quarterly* 18: 329-337.

Bechhofer, R. E. (1954). A Singke Stage Multiple Decision Procedure for Ranking Means of Normal Populations with Variances, *Annals of Mathematical Statistics* 25: 16-39.

Bechhofer, R. E., Kiefer, J., and Sobel, M. (1968). *Sequential Identification and Ranking Procedures,* Chicago: University of Chicago Press.

Bechhofer, R. E., Santner, T. J., and Goldsman, D. M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons,* New York: Wiley.

Bliss, C. I. and Owen, A. R. G. (1958). Negative Binomial Distributions with a Common κ, *Biometrika* 45: 37-58.

Chattopadhyay, S. (1998). Sequential Estimation of Normal Mean Under Asymmetric Loss Function with a Shrinkage Stopping Rule, *Metrika* 48: 53-59.

Chattopadhyay, S. (2000). Sequential Estimation of Exponential Location Parameter Using an Asymmetric Loss Function, *Communications in Statistics-Theory & Methods* 29: 783-795.

Chattopadhyay, S., Chaturvedi, A., and Sengupta, R. N. (2000). Sequential Estimation of a Linear Function of Normal Means Under Asymmetric Loss Function, *Metrika* 52: 225-235.

Chattopadhyay, S., Datta, S., and Sengupta, R. N. (2005). Asymmetric Penalized Prediction Using Adaptive Sampling Procedures, *Sequential Analysis* 24: 23-43.

Chattopadhyay, S. and Sengupta, R. N. (2006). Three-Stage and Accelerated Sequential Point Estimation of the Normal Mean Using LINEX Loss Function, *Statistics* 40: 39 – 49.

Chernoff, H. (1972). *Sequential Analysis and Optimal Design,* CBMS 8, Philadelphia: SIAM.

Chow, Y. S., Hsiung, C., and Lai, T. L. (1979). Extended Renewal Theory and Moment Convergence in Anscombe's Theorem, *Annals of Probability* 7: 304-318.

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., Robbins, H., and Teicher, H. (1965). Moments of Randomly Stopped Sums, *Annals of Mathematical Statistics* 36: 789-799.

Crowder, M. J. (2000). Tests for a Family of Survival Models Based on Extremes, in *Recent Advances in Reliability Theory,* N. Limnios and N. Nikulin, eds., pp. 307-321, Boston: Birkhauser.

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

Efron, B. (1988). Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve, *Journal of American Statistical Association* 83: 414-425.

Ghosh, M. and Mukhopadhyay, N. (1975). Asymptotic Normality of Stopping Times in Sequential Analysis, Unpublished Report.

Ghosh, M. and Mukhopadhyay, N. (1976). On Two Fundamental Problems of Sequential Estimation, *Sankhya, 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 & Methods* 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, *Sankhya, Series A* 43: 220-227.

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

Ghosh, B. K. and Sen, P. K. (1981). *Handbook of Sequential Analysis,* edited volume, New York: Dekker.

Ghurye, S. G. (1958). Note on Sufficient Statistics and Two-Stage Procedures, *Annals of Mathematical Statistics* 29: 155-166.

Gibbons, J. D., Olkin, I., and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology,* New York: Wiley.

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selection and Ranking Populations,* New York: Wiley.

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 Statistics*

Society, Series B 45: 219-223.

Heijting, S. (2013). *Weed Count Data Heijting et al.* DANS. http://dx.doi.org/10.17026/dans-zu9-r7y8.

Heijting, S., Van Der Werf, W., Stein, A., and Kropff, M. J. (2007). Are Weed Patches Stable in Location? Application of an Explicitly Two-Dimensional Methodology, *Weed Research* 47: 381-395. http://dx.doi.org/10.17026/dans-zu9-r7y8v.

Hwang, L.-C. and Lee, C.-H. (2012). Bayes Sequential Estimation for One-Parameter Exponential Family Under Asymmetric LINEX Loss Function, *Sequential Analysis* 31: 3-21.

Isogai, E., Kobayashi, K., and Uno, C. (2011). Higher Order Approximations by a Two-Stage Procedure for a Negative Exponential Distribution, *Journal of Statistical Planning and Inference* 141: 3304-3312.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions-2*, New York: Wiley.

Jokiel-Rokita, A. (2008). A Sequential Estimation Procedure for the Parameter of an Exponential Distribution Under Asymmetric Loss Function, *Statistics and Probability Letters* 78: 3091–3095.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis*, second edition, New York: Springer.

Kuno, E. (1969). A New Method of Sequential Sampling to Obtain the Population Estimates with a Fixed Level of Precision, *Researches on Population Ecology* 11: 127-136.

Kuno, E. (1972). Some Notes on Population Estimation by Sequential Sampling, *Researches on Population Ecology,* 14: 58-73.

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.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data,* second edition, Hoboken: Wiley.

Lawless, J. F. and Singhal, K. (1980). Analysis of Data from Life Test Experiments Under an Exponential Model, *Naval Research Logistics Quarterly* 27: 323-334.

Lehmann, E. L. (1951). *Notes on the Theory of Estimation,* Berkeley: University of California Press.

Leinhardt, S. and Wasserman, S. S. (1979). Exploratory Data Analysis: An Introduction to Selected Methods, in *Sociological Methodology*, K. Schuessler, ed., pp. 311-365, San Francisco: Jossey-Bass.

Linden, A. and Mantyniemi, S. (2011). Using Negative Binomial Distribution to Model Overdispersion in Ecological Count Data, *Ecology* 92: 1414-1421.

Lombard, F. and Swanepoel, J. W. H. (1978). On Finite and Infinite Confidence Sequences, *South African Statistical Journal* 12: 1-24.

Marcus, R. and Talpaz, H. (1985). A Bayesian Procedure for the Sequential Estimation of the Mean of a Negative Binomial Distribution, *Statistische Hefte* 26: 303. doi: 10.1007/BF02932540.

Mukhopadhyay, N. (1974). Sequential Estimation of Location Parameter in Exponential Distributions, *Calcutta Statistical Association Bulletin* 23: 85-95.

Mukhopadhyay, N. (1976). Sequential Estimation of a Linear Function of Means of Three Normal Populations, *Journal of American Statistical Association* 71: 149-153.

Mukhopadhyay, N. (1977). Remarks on Sequential Estimation of a Linear Function of Two Means: The Normal Case, *Metrika* 24: 197-201.

Mukhopadhyay, N. (1978). Sequential Point Estimation of the Mean when the Distribution is Unspecified, *Technical Report* #312, School of Statistics, Department of Theoretical Statistics, University of Minnesota, Minneapolis, Minnesota, USA .

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. (1984). Sequential and Two-Stage Procedures for Selecting the Better Exponential Population Covering the Case of Unknown and Unequal Scale Parameters, *Journal of Statistical Planning and Inference* 9: 33-43.

Mukhopadhyay, N. (1988). Sequential Estimation Problems for Negative Exponential Populations, *Communications in Statistics-Theory & Methods, Series A* 17: 2471-2506.

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

Mukhopadhyay, N. (1993). Non Renewal Theory and Beyond: Reviewing the Roles in Selection and Ranking, *Multiple Comparisons, Selection and Applications in Biometry,* F. M. Hoppe ed., pp. 293-313, New York: Marcel Dekker.

Mukhopadhyay, N. (1995). Two-Stage and Multi-Stage Estimation, in *The Exponential Distribution: Theory, Methods and Application*, Chapter 26, N. Balakrishnan and A. P. Basu, eds., pp. 429-452, Amsterdam: Gordon and Breach.

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

Mukhopadhyay, N. (1999a). Second-Order Properties of a Two-Stage Fixed-Size Confidence Region for the Mean Vector of a Multivariate Normal Distribution, *Journal of Multivariate Analysis* 68: 250-263.

Mukhopadhyay, N. (1999b). Higher Than Second-Order Approximations Via Two-Stage Sampling, *Sankhya, Series A* 61: 254-269.

Mukhopadhyay, N. (2002). Sequential Sampling, in *The Encyclopedia of Environmetrics,* Sharaawi, A. H. and Piegorsch, W. W., eds., Vol. 4, pp. 1983-1988, Chichester: Wiley.

Mukhopadhyay, N. and Aoshima, M. (1998). Multivariate Multistage Methodologies for Simultaneous All Pairwise Comparisons, *Metrika* 47: 185-201.

Mukhopadhyay, N. and Banerjee, S. (2014). Purely Sequential and Two-Stage Fixed-Accuracy Confidence Interval Estimation Methods for Count Data for Negative Binomial Distributions in Statistical Ecology: One-Sample and Two-Sample Problems, *Sequential Analysis* 33: 251-285.

Mukhopadhyay, N. and Banerjee, S. (2015). Sequential Negative Binomial Problems and Statistical Ecology: A Selected Review with New Directions, *Statistical Methodology* 26: 34-60.

Mukhopadhyay, N. and Bapat, S. R. (2016a). Multistage Point Estimation Methodologies for a Negative Exponential Location Under a Modified Linex Loss Function: Illustrations with Infant Mortality and Bone Marrow Data, *Sequential Analysis* 35: 175-206.

Mukhopadhyay, N. and Bapat, S. R. (2016b). Multistage Estimation of the Difference of Locations of Two Negative Exponential Populations under a Modified Linex Loss Function: Real Data Illustrations from Cancer Studies and Reliability Analysis, *Sequential Analysis* 35: 387-412.

Mukhopadhyay, N. and Bapat, S. R. (2017a). Purely Sequential Bounded-Risk Point Estimation of the Negative Binomial Mean Under Various Loss Functions: Part I - One-Sample Problem, *Submitted for publication.*

Mukhopadhyay, N. and Bapat, S. R. (2017b). Purely Sequential Bounded-Risk Point Estimation of the Negative Binomial Means Under Various Loss Functions: Part II - Multi-Sample Problems, *Submitted for publication.*

Mukhopadhyay, N. and Darmanto, S. (1988). Sequential Estimation of the Difference of Means of Two Negative Exponential Populations, *Sequential Analysis* 7: 165-190

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

Mukhopadhyay, N. and de Silva, B. M. (2005). Two-Stage Estimation of the Mean of the Mean of a Negative Binomial Distribution with Applications to Mexican Bean Beetle Data, *Sequential Analysis* 24: 99-137.

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

Mukhopadhyay, N., de Silva, B. M. and Waikar, V. B. (2010). On Two-Stage Confidence Interval Procedures and Their Comparisons for Estimating the Difference of Normal Means, *Sequential Analysis* 29: 44-68.

Mukhopadhyay, N. and Diaz, J. (1985). Two-Stage Sampling for Estimating the Mean of a Negative Binomial Distribution, *Sequential Analysis* 4: 1-18.

Mukhopadhyay, N. and Duggan, W. T. (1997). Can a Two-Stage Procedure Enjoy Second-Order Properties? *Sankhya, 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. and Duggan, W. T. (2000). On a Two-Stage Procedure Having Second-Order Properties with Applications, *Annals of Institute of Statistical Mathematics* 51: 621-636.

Mukhopadhyay, N. and Duggan, W. T. (2001). A Two-Stage Point Estimation Procedure for the Mean of an Exponential Distribution and Second-Order Results, *Statistics & Decisions* 19: 155-171.

Mukhopadhyay, N. and Hamdy, H. I. (1984). On Estimating the Difference of Location Parameters of Two Negative Exponential Distributions, *Canadian Journal of Statistics* 12: 67-76

Mukhopadhyay, N. and Hilton, G. F. (1986). Two-Stage and Sequential Procedures for Estimating the Location Parameter of a Negative Exponential Distribution, *South African Statistical Journal* 20: 117-136.

Mukhopadhyay, N. and Liberman, S. (1989). Sequential Estimation of a Linear Function of

Mean Vectors, *Sequential Analysis* 8: 381-395.

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. and Zacks, S. (2017). Modified Linex Two-Stage and Purely Sequential Estimation of the Variance in a Normal Distribution with Illustrations Using Horticultural Data, *Submitted for publication.*

Mulekar, M. S. and Young, L. J. (1993). A Fixed Sample Size Selection Procedure for Negative Binomial Populations, *Metrika* 40: 25-35.

Mulekar, M. S. and Young, L. J. (2004). Sequential Estimation in the Agricultural Sciences, in *Applied Sequential Methodologies,* Mukhopadhyay, N., Datta, S., and Chattopadhyay, S., eds., pp. 293-318, New York: Dekker.

Nyrop, J. P. and Binns, M. R. (1991). Quantitative Method for Designing and Analyzing Sampling Program for Use in Pest Management, in *Handbook of Pest Management,* Pimental, D., ed., Vol. 2, pp. 67-132, Boca Raton: CRC.

R Core Team (2014). *R: A Language and Environment for Statistical Computing,* R Foundation for Statistical Computing, Vienna, Austria.

Ramkaran, Chaturvedi, A. C., and Akbar, S. A. (1986). Sequential Estimation of a Linear Function of k-Normal Means, *Journal of Indian Society of Agricultural Statistics* 38: 395-402.

Robbins, H. (1959). Sequential Estimation of the Mean of a Normal Population, *Probability and Statistics,* H. Cramer Volume, U. Grenander ed., pp. 235-245, Uppsala: Almquist and Wiksell.

Robbins, H., Simons, G., and Starr, N. (1967). A Sequential Analogue of the Behrens-Fisher Problem, *Annals of Mathematical Statistics* 38: 1384-1391.

Robbins, H. and Starr, N. (1965). Remarks on Sequential Hypothesis Testing, *Statistics Technical Report. No 68,* University of Minnesota, Minneapolis.

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

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

Shanker, R., Fesshaye, H., and Selvaraj, S. (2016). On Modeling of Lifetime Data Using One Parameter Akash, Lindley and Exponential Distributions, *Biometrics and Biostatistics International Journal* 2: 1-10. http://dx.doi.org/10.15406/bbij.2016.03.00061

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

Srivastava, M. S. (1970) On a Sequential Analogue of the Behrens-Fisher Problem, *Journal of Royal Statistical Society, Series B* 32: 144-148.

Starr, N. (1966b). On the Asymptotic Efficiency of a Sequential Procedure for Estimating the Mean, *Annals of Mathematical Statistics* 36: 1173-1185.

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.

Swanepoel, J. W. H. and van Wyk, J. W. J. (1982). Fixed Width Confidence Intervals for the Location Parameter of an Exponential Distribution, *Communications in Statistics-Theory & Methods, Series A* 11: 1279-1289.

Takada, Y. (2000). Sequential Point Estimation of Normal Mean Under Linex Loss Function, *Metrika* 52: 163-171.

Takada, Y. (2001). Bayes Sequential Estimation of Poisson Mean Under a LINEX Loss Function, *Sequential Analysis* 20: 55–64.

Takada, Y. (2006). Multistage Estimation Procedures with Bounded Risk for the Normal Mean Under LINEX Loss Function, *Sequential Analysis* 25: 227–239.

Takada, Y. and Nagao, H. (2004). Asymptotic Improvement of the Sample Mean Vector for Sequential Point Estimation of a Multivariate Normal Mean with a Linex Loss Function, *Sientiae Mathematicae Japonicae Online* 10: 71-79.

The New York Times, 28 September, 1975, p. E-3, Table 3.

Varian, H. R. (1975). A Bayesian Approach to Real Estate Assessment, in *Studies in Bayesian Econometrics and Statistics*, Leonard J. Savage, Stephen E. Fienberg, and Arnold Zellner, eds., pp. 195-208, North Holland: Elsevier.

Wald, A. (1947). *Sequential Analysis,* New York: Wiley.

Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test, *Annals of Mathematical Statistics* 19: 326-339.

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

Willson, L. J. and Folks, J. L. (1983). Sequential Estimation of the Mean of the Negative Binomial Distribution, *Communications in Statistics-Sequential Analysis* 2: 55-70.

Willson, L. J., Folks, J. L., and Young, J. H. (1984). Multistage Estimation Compared with Fixed-Sample Size Estimation of the Negative Binomial Parameter κ, *Biometrics* 40: 109-117.

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

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

Xiao, Y., Takada, Y., and Shi, N. (2005). Minimax Confidence Bound of the Normal Mean under an Asymmetric Loss Function, *Annals of Institute of Statistical Mathematics* 57: 167-182.

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

Zacks, S. and Mukhopadhyay, N. (2009). On Exact and Asymptotic Properties of Two-Stage and Sequential Estimation of the Normal Mean Under Linex Loss, *Communications in Statistics-Theory & Methods* 38: 2992-3014.

Zelen, M. (1966). Application of Exponential Models to Problems in Cancer Research, *Journal of Royal Statistical Society* 129: 368-398.

Zellner, A. (1986). Bayesian Estimation and Prediction Using Asymmetric Loss Functions, *Journal of American Statistical Association* 81: 446-451