

Sensitivity of Normal-based Triple Sampling Sequential Point Estimation to the Normality Assumption

A. S. Yousef^{1,2}, A. C. Kimber¹, H. I. Hamdy²
University of Southampton¹ and Kuwait University²

Abstract

This article discusses the sensitivity of the sequential normal-based triple sampling procedure for estimating the population mean to departures from normality. We assume that the underlying population has finite absolute sixth moment and find that asymptotically the behaviour of the estimator and of the sample size depend on the skewness and kurtosis of the underlying distribution when using a squared error loss function with linear sampling cost. These results enable the effects of non-normality easily to be assessed both qualitatively and quantitatively. We supplement our asymptotic results with a simulation experiment to study the performance of the estimator and the sample size in a range of conditions.

Keywords: Asymptotic relative efficiency, kurtosis, regret, sampling cost, simulation, skewness, squared error loss function, Taylor expansion.

Corresponding author:

A.C.Kimber, School of Mathematics and Southampton Statistical Sciences Research Institute,
University of Southampton, SO17 1BJ, U.K.

Email A.C.Kimber@soton.ac.uk

Tel +44 (0)23 8059 5124

Fax +44 (0)23 8059 5147

1. Introduction

Hall (1981) introduced the triple sampling sequential technique to achieve the operational savings made possible by sampling in batches and at the same time to guarantee the asymptotic efficiency of the one by one sequential sampling of Anscombe (1953), Robbins (1959), and Chow and Robbins (1965), though see also Liu (1997a) for a discussion of higher order efficiency. Although Hall's techniques were specifically devised to construct fixed width confidence for the normal mean with a prescribed coverage, his results may be modified to treat point estimation problems. Mukhopadhyay (1985) used Hall's (1981) results to derive a triple sampling sequential point estimation technique for the normal mean with bounded cost function. Further progress was made in Mukhopadhyay *et. al.* (1987) and Mukhopadhyay (1990) discussed triple sampling mores generally; see also the texts by Mukhopadhyay and Solanky (1994), Ghosh *et. al.* (1997) and Mukhopadhyay and de Silva (2009). Vik and Mukhopadhyay (1988) used the asymptotic properties of U -statistics to obtain properties of a triple sampling scheme for an unknown distribution.

In developing specific triple sampling schemes, a common approach is to assume that the form of the distribution from which observations will be sampled is known and to construct an appropriate triple sampling scheme accordingly. The properties of such a scheme are then derived using the known distributional form. For example, Hamdy and Pallotta (1987) assumed a Pareto underlying distribution, while Hamdy (1988) considered the normal case with various loss functions. Our approach, however, may be motivated by the following scenario.

Suppose a matched pair clinical trial is to be run where one member of each pair receives treatment A and the other receives treatment B. The mean difference in treatment effect is the parameter of interest. In the protocol, drawn up in advance of running the trial, a triple sampling scheme is proposed under the assumption that the differences are normally distributed with unknown variance. In this paper we attempt to answer questions such as: what are the properties of the point estimator of the mean difference and the distribution of the sample size when the true underlying distribution is not normal? The results we derive hold for a wide class of distributions (those with finite absolute sixth moment) and are presented in terms of skewness and kurtosis so that the sensitivity of the normal-based triple sampling scheme to the presence of non-normality may be seen both qualitatively and quantitatively.

Early work on sensitivity to assumptions in sequential inference was conducted by Bhattacharjee and Nagendra (1964) in the context of the Wald sequential test for the mean, and by Bhattacharjee (1965), Blumenthal and Govindarajulu (1977) and Ramkaran (1983) for Stein's two stage sampling procedure. Jureckova and Sen (1996) devoted several chapters to the sensitivity of point and interval estimation methods in sequential statistical inference. See also Liu *et. al.* (2002) for results on the robustness of one-by-one sequential sampling in confidence interval estimation.

We first set up some basic notation. Let X_1, X_2, X_3, \dots be a sequence of independent and identically distributed random variables from a continuous distribution function $F(\cdot)$ with mean μ , variance θ , skewness γ and kurtosis β , all unknown but finite. The main focus in this study is the estimation of μ in the presence of the nuisance parameter θ or some continuously differentiable and bounded function $g(\theta)$. Throughout the paper we require that $E(|X_1|^6) < \infty$ in order to obtain our asymptotic results.

In the literature on sequential sampling for inference about the mean and for most distributions, it is assumed that the sample size required to satisfy the above conditions takes the general form

$$n \geq n^* = \lambda g(\theta), \quad (1.1)$$

where λ is a function of some predetermined constants A and C , which may appear in a cost function incurred in point estimation of μ . Moreover, the function λ is also permitted to approach infinity such that the optimal sample size $n^* \rightarrow \infty$. Having observed a random sample X_1, X_2, \dots, X_n from the distribution $F(\cdot)$ with $n \geq 2$, we use the sample mean $\bar{X}_n = \sum_{i=1}^n X_i/n$ and the sample variance

$S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n-1)$ as estimators of μ and θ respectively.

The triple sampling scheme is introduced in section 2. The main asymptotic results are derived in section 3. The asymptotic distribution of the stopping sample size is obtained in section 4. The squared error loss function is discussed in the present context in section 5 and some asymptotic regret and efficiency results for the triple sampling scheme with squared error loss are presented in section 6. Finally, results of a simulation study are presented in section 7.

2. Triple sampling procedure for inference

Since n^* in (1.1) is unknown, no fixed sample size procedure provides the above point estimation for μ uniformly for all θ . Therefore, we use a sequential sampling procedure to estimate μ via estimation of the optimal sample size n^* . We now give an account of the triple sampling procedure as described by Hall (1981), which may be described by the following three phases:

Pilot Phase: An initial random sample X_1, X_2, \dots, X_m of size $m \geq 2$ is taken from the distribution $F(\cdot)$. We use \bar{X}_m and S_m^2 as our initial estimators for μ and θ respectively.

Main Study Phase: Let δ ($0 < \delta < 1$) be a fixed design factor. The sample size required to complete the main study phase is defined by

$$N_1 = \max\{m, [\delta \lambda g(S_m^2)] + 1\}, \quad (2.1)$$

where $[x]$ is the integer part of x . If $m \geq [\delta \lambda g(S_m^2)] + 1$, we stop sampling at this stage. Otherwise, we continue to observe an additional random sample $X_{m+1}, X_{m+2}, \dots, X_{N_1}$ of size $N_1 - m$ from $F(\cdot)$. We augment the $N_1 - m$ observations with the previous m observations and estimate the parameters using \bar{X}_{N_1} and $S_{N_1}^2$.

Fine Tuning Phase: Let

$$N = \max\{N_1, [\lambda g(S_{N_1}^2)] + 1\}. \quad (2.2)$$

If $N_1 \geq [\lambda g(S_{N_1}^2)] + 1$, we stop at this stage. Otherwise we continue and take $N - N_1$ more observations from $F(\cdot)$, $X_{N_1+1}, X_{N_1+2}, \dots, X_N$, after which sampling is terminated and we propose \bar{X}_N as a point estimator of μ .

Throughout the following sections, the asymptotic characteristics of the triple sampling scheme are developed under the assumption made by Hall (1981) that

$$\lim \text{Sup}(m/n^*) < \delta \text{ as } m \rightarrow \infty, \text{ and } n^* = O(m^r), \text{ for } r > 1 \quad (2.3)$$

Before developing the theory of the triple sampling procedure specified above, we state the following preliminary results in the fixed sample size case which will facilitate proofs of the theorems to follow.

Lemma 1

Let X be a random variable with distribution function $F(\cdot)$ as above and let $Z = X - \mu$. For a random sample of Z_1, Z_2, \dots, Z_m we have:

- i) $E(Z) = 0, E(Z^2) = \theta, E(Z^3) = \gamma \theta^{3/2}, E(Z^4) = \beta \theta^2$.
- ii) $E(\bar{Z}_m) = 0, E(\bar{Z}_m^2) = \frac{\theta}{m}, E(\bar{Z}_m^3) = \frac{\gamma \theta^{3/2}}{m^2}, E(\bar{Z}_m^4) = \frac{\theta^2(\beta + 3m - 3)}{m^3}$
- iii) $E(\bar{Z}_m S_m^2) = \frac{\gamma \theta^{3/2}}{m}, E(\bar{Z}_m^2 S_m^2) = \frac{\theta^2(\beta + m - 3)}{m^2}, E(\sum_{i \neq j}^m Z_i Z_j S_m^2) = -2\theta^2$.

Proof of (i) is immediate and for (ii) the two terms $E(\bar{Z}_m^3)$ and $E(\bar{Z}_m^4)$ follow immediately from Rohatgi (1976, p303). The proof of (iii) follows by taking the expectation over the identities:

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

$$\bar{Z}_m^2 S_m^2 = (m-1)^{-1} \left\{ m^{-2} \left[\sum Z_i^4 + \sum_{i \neq j} \sum Z_i^2 Z_j^2 + \sum_{i \neq j} Z_i^2 \sum \sum Z_i Z_j \right] - m \bar{Z}_m^4 \right\}$$

and

$$\begin{aligned} \sum_{i \neq j} \sum Z_i Z_j S_m^2 &= m^{-1} \left[2 \sum_{i \neq j} \sum Z_i^3 Z_j + \sum_{i \neq j \neq k} \sum \sum Z_i^2 Z_j Z_k \right] \\ &- (m(m-1))^{-1} \left[2 \sum_{i \neq j} \sum Z_i^2 Z_j^2 + 4 \sum_{i \neq j \neq k} \sum \sum Z_i^2 Z_j Z_k + \sum_{i \neq j \neq k \neq l} \sum \sum \sum Z_i Z_j Z_k Z_l \right]. \end{aligned}$$

3. Asymptotic characteristics of the triple sampling procedure

The asymptotic characteristics of the triple sampling procedure are thoroughly discussed through the following theorems. Theorem 1 provides results regarding the asymptotic characteristics of the main study phase. Specifically, second order approximations of the expectation and the variance of the second stage sample mean are given as the initial sample size gets large.

Theorem 1

For the triple sampling rule (2.1)-(2.2), if condition (2.3) holds, then as $m \rightarrow \infty$, we have

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

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

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

$$E(\bar{X}_{N_1}) = E \left(N_1^{-1} \sum_{i=1}^{N_1} Z_i \right) + \mu = E \left\{ E \left(N_1^{-1} \sum_{i=1}^{N_1} Z_i \mid N_1 \right) \right\} + \mu. \quad (3.1)$$

Then, conditioning on the σ -field generated by the random variables Z_1, Z_2, \dots, Z_m , we have

$$E(\bar{X}_{N_1}) = E \left\{ N_1^{-1} E \left(\sum_{i=1}^m Z_i + \sum_{i=m+1}^{N_1} Z_i \mid Z_1, Z_2, \dots, Z_m \right) \right\} + \mu. \text{ Given } Z_1, Z_2, \dots, Z_m, \text{ the first sum, } \sum_{i=1}^m Z_i, \text{ is non-random,}$$

and the second sum has an expectation zero. Hence:

$$E(\bar{X}_{N_1}) = E \left\{ N_1^{-1} \sum_{i=1}^m Z_i \right\} + \mu \quad (3.2)$$

A Taylor series expansion of N_1^{-1} around δn^* gives:

$$N_1^{-1} = (\delta n^*)^{-1} - (\delta n^*)^{-2} (N_1 - \delta n^*) + \eta^{-3} (N_1 - \delta n^*)^2, \quad (3.3)$$

where η is a random variable lying between N_1 and δn^* . For large λ we approximate $N_1 \approx \delta \lambda g(S_m^2)$ as $\lambda \rightarrow \infty$, where g is a differentiable bounded function around θ . By expanding $g(S_m^2)$ and $g^2(S_m^2)$ around θ and substituting in (3.3), we obtain:

$$E\left(N_1^{-1} \sum_{i=1}^m Z_i\right) = -(\delta n^*)^{-1} \left[\frac{d}{d\theta} \ln g(\theta) \right] \gamma \theta^{3/2} + o(\lambda^{-1}), \quad (3.4)$$

where the remainder term is of order $o(\lambda^{-1})$. By substituting this in (3.2), (i) follows.

To prove (ii), we also write

$$\begin{aligned} \text{Var}(\bar{X}_{N_1}) &= E(\bar{X}_{N_1} - \mu)^2 \\ &= E\left(E\left[N_1^{-2} \left(\sum_{i=1}^{N_1} Z_i^2 + \sum_{i \neq j}^{N_1} \sum Z_i Z_j \right) \middle| Z_1, Z_2, \dots, Z_m \right] \right). \end{aligned} \quad (3.5)$$

The first term in (3.5), conditioned on the σ -field generated by the random variables Z_1, Z_2, \dots, Z_m , can be written as

$$E\left(N_1^{-2} \sum_{i=1}^{N_1} Z_i^2 \right) = E\left\{ N_1^{-2} E\left(\sum_{i=1}^m Z_i^2 + \sum_{i=m+1}^{N_1} Z_i^2 \middle| Z_1, Z_2, \dots, Z_m \right) \right\} \quad (3.6)$$

and therefore,

$$E\left(N_1^{-2} \sum_{i=1}^{N_1} Z_i^2 \right) = E\left(N_1^{-2} \sum_{i=1}^m Z_i^2 \right) + \theta E(N_1^{-1} - m N_1^{-2}). \quad (3.7)$$

The first term on the right hand side of (3.7) can be treated along the same lines that led to (3.4). The second term of (3.7) can be written as

$$\begin{aligned} E\{N_1^{-2}(N_1 - m)\} &\leq E(N_1^{-1}) \\ &= o(\lambda^{-1}) \quad \text{as } m \rightarrow \infty \end{aligned}$$

Finally,

$$E\left(N_1^{-2} \sum_{i=1}^{N_1} Z_i^2 \right) = \theta (\delta n^*)^{-1} - 2\theta^2 (\beta - 1) (\delta n^*)^{-2} \left[\frac{d}{d\theta} \ln g(\theta) \right] + o(\lambda^{-1}). \quad (3.8)$$

Similarly, the second term of (3.5) yields

$$E\left(N_1^{-2} \sum_{i \neq j}^{N_1} \sum Z_i Z_j \right) = 4\theta^2 (\delta n^*)^{-2} \left[\frac{d}{d\theta} \ln g(\theta) \right] + o(\lambda^{-2}), \quad (3.9)$$

where we have used Lemma 1 and the assumption that g and its derivative are bounded. Hence, the proof of part (ii) is complete.

From part (i) of Theorem 1 the bias of \bar{X}_{N_1} depends on the variance and skewness of the underlying distribution and the form of the function g , as well as on the optimal sample size n^* and the design factor δ . In the normal case, where $\gamma = 0$, we observe that the bias is zero.

The variance of the second stage sample mean $Var(\bar{X}_{N_1})$ depends on $g'(\theta)$, the kurtosis, the variance θ , n^* and δ . Suppose $g'(\theta) > 0$. Then platykurtic, leptokurtic and mesokurtic distributions yield estimators of the mean with variances larger than, less than and equal to $\theta(\delta n^*)^{-1}$ respectively. Note that, for large n^* , the variance of \bar{X}_{N_1} tends to zero, as expected.

It is of interest to see whether the fine tuning stage (third stage) reduces the amount of bias evident from Theorem 1. The following results in Theorem 2 provide asymptotic characteristics of the second stage sample size N_1 which are essential for proving Theorems 3 and 4 later.

Lemma 2

For the triple sampling rule (2.1)-(2.2), if condition (2.3) holds, then conditioning on the σ -field generated by Z_1, Z_2, \dots, Z_m , we have

$$(i) E \left[\left(S_{N_1}^2 - \theta \right) \sum_{i=1}^{N_1} \sum_{i \neq j}^{N_1} Z_i Z_j \mid Z_1, Z_2, \dots, Z_m \right] = -2N_1^{-2} \left(\sum_{i=1}^m \sum_{i \neq j}^m Z_i^2 Z_j^2 + \theta^2 (N_1 - m)(N_1 - m - 1) \right)$$

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

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

$$(iv) E \left[N_1 \left(S_{N_1}^2 - \theta \right) \mid Z_1, Z_2, \dots, Z_m \right] = -N_1^{-1} \sum_{i=1}^m \sum_{i \neq j}^m Z_i Z_j$$

$$(v) E \left[N_1 \left(S_{N_1}^2 - \theta \right)^2 \mid Z_1, Z_2, \dots, Z_m \right] = \theta^2 N_1 + N_1^{-1} \left(\sum_{i=1}^m Z_i^4 + \beta \theta^2 (N_1 - m) \right) + N_1^{-1} \left(\sum_{i=1}^m \sum_{i \neq j}^m Z_i^2 Z_j^2 + \theta^2 (N_1^2 - 2mN_1 - N_1 + m^2 + m) \right) + N_1^{-3} \left(\sum_{i=1}^m \sum_{i \neq j}^m Z_i^2 Z_j^2 + \theta^2 (N_1 - m)(N_1 - m - 1) \right)$$

Theorem 2

For the triple sampling rule (2.1)-(2.2), if condition (2.3) holds, then as $m \rightarrow \infty$, we have

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

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

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

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

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

Proof:

To prove (i), we write

$$E \left[\left(S_{N_1}^2 - \theta \right) \sum_{i=1}^{N_1} \sum_{j \neq i}^{N_1} Z_i Z_j \right] = E \left\{ E \left[\left(S_{N_1}^2 - \theta \right) \sum_{i=1}^{N_1} \sum_{j \neq i}^{N_1} Z_i Z_j \mid Z_1, Z_2, \dots, Z_m \right] \right\}$$

Hence, from (i) of lemma 2, we have

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

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

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

and the second term leads to

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

where we have used the assumptions in (2.3) and the fact that g and its derivatives are bounded.

Therefore, (i) follows.

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

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

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

To prove (iv), recall (iv) of Lemma 2 and part (iii) of Lemma 1.

The proof of Theorem 2 is completed. We delete details for brevity. The following Theorem 3 presents the asymptotic characteristics of the third stage sample.

Theorem 3

For the triple sampling rule (2.1)-(2.2), if condition (2.3) holds, then as $m \rightarrow \infty$, we have

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

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

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

$$E(\bar{X}_N) = E \left\{ N^{-1} E \left(\sum_{i=1}^{N_1} Z_i + \sum_{i=N_1+1}^N Z_i \mid Z_1, Z_2, \dots, Z_{N_1} \right) \right\} + \mu. \quad (3.10)$$

Again the first sum $\sum_{i=1}^{N_1} Z_i$ in (3.10) is non-random. Thus, (3.10) reduces to

$$E(\bar{X}_N) = E \left(N^{-1} \sum_{i=1}^{N_1} Z_i \right) + \mu \quad (3.11)$$

Also, $N \approx \lambda g(S_{N_1}^2)$, as $\lambda \rightarrow \infty$, where $g(\cdot)$ is a differentiable and bounded function around θ . By expanding N^{-1} around n^* , and $g(S_{N_1}^2)$ around θ , using Taylor series we obtain

$$E \left(N^{-1} \sum_{i=1}^{N_1} Z_i \right) = - \left[\frac{d}{d\theta} \ln g(\theta) \right] (n^*)^{-1} E \left\{ (S_{N_1}^2 - \theta) \sum_{i=1}^{N_1} Z_i \right\} + o(\lambda^{-1}). \quad (3.12)$$

Application of part (iii) of Theorem 2 completes the proof of (i) of Theorem 3.

Proof of (ii) can be obtained directly from the following

$$\text{Var}(\bar{X}_N) = E \left(N^{-2} E \left(\sum_{i=1}^{N_1} Z_i^2 + \sum_{i \neq j} \sum_{i,j} Z_i Z_j \mid Z_1, Z_2, \dots, Z_{N_1} \right) \right) \\ + \theta E \left(E \left(N^{-1} - N_1 N^{-2} \mid Z_1, Z_2, \dots, Z_{N_1} \right) \right) \quad (3.13)$$

By conditioning on the σ -field generated by Z_1, Z_2, \dots, Z_{N_1} and expanding N^{-2} and $g(S_{N_1}^2)$ around n^* and θ respectively, it can be shown that

$$E \left(N^{-2} \left(\sum_{i=1}^{N_1} Z_i^2 \mid Z_1, Z_2, \dots, Z_{N_1} \right) \right) = \theta (n^*)^{-2} E(N_1) - 2\theta^2 (\beta - 1) (n^*)^{-2} \left[\frac{d}{d\theta} \ln g(\theta) \right] + o(\lambda^{-2}) \quad (3.14)$$

$$E \left(N^{-2} \left(\sum_{i \neq j} \sum_{i,j} Z_i Z_j \mid Z_1, Z_2, \dots, Z_{N_1} \right) \right) = 4\theta^2 \left[\frac{d}{d\theta} \ln g(\theta) \right] (n^*)^{-2} \left(1 - (\delta n^*)^{-1} \right) + o(\lambda^{-3}). \quad (3.15)$$

Also

$$\begin{aligned}
E(N^{-1} - N_1 N^{-2}) &= -(n^*)^{-2} E(N_1) + (n^*)^{-1} \\
&\quad + \theta^2 (\beta - 1) (n^*)^{-2} \left[\frac{g''(\theta)}{g(\theta)} \left(4 - \frac{1}{2} \delta^{-1} \right) + 2\delta^{-1} \left(\frac{g'(\theta)}{g(\theta)} \right)^2 - \frac{3g^{2''}(\theta)}{2g^2(\theta)} \right] + o(\lambda^{-2}) \quad (3.16)
\end{aligned}$$

By adding (3.14), (3.15) and (3.16) the proof of Theorem 3 is completed.

In view of (i) and (ii) of Theorem 3, it is worth mentioning that the third stage has indeed reduced the bias noticed in (i) and (ii) of Theorem 1.

Theorem 4 gives asymptotic results regarding the second stage estimator of the unknown variance θ .

Theorem 4

For the triple sampling rule (2.1)-(2.2), if condition (2.3) holds, then as $m \rightarrow \infty$, we have:

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

$$(ii) E(S_{N_1}^4) = \theta^2 + \theta^2 (\beta - 1) (\delta n^*)^{-1} + o(\lambda^{-1})$$

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

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

Proof: To prove (i), we write

$$\begin{aligned}
E(S_{N_1}^2) &= E \left(N_1^{-1} E \left(\sum_{i=1}^{N_1} Z_i^2 \mid Z_1, Z_2, \dots, Z_m \right) \right) \\
&\quad - E \left((N_1 (N_1 - 1))^{-1} E \left(\sum_{i \neq j}^{N_1} \sum Z_i Z_j \mid Z_1, Z_2, \dots, Z_m \right) \right)
\end{aligned}$$

Consequently

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

while

$$\begin{aligned}
E \left((N_1 (N_1 - 1))^{-1} E \left(\sum_{i \neq j}^{N_1} \sum Z_i Z_j \mid Z_1, Z_2, \dots, Z_m \right) \right) &= E \left((N_1 (N_1 - 1))^{-1} \sum_{i=1}^m \sum_{i \neq j}^m Z_i Z_j \right) \\
&\leq E \left(\sum_{i=1}^m \sum_{i \neq j}^m Z_i Z_j \right) = 0 \quad (3.17)
\end{aligned}$$

To prove (ii), we write

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

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

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

while

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

and

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

The proof of (iii) follows immediately from parts (i) and (ii).

A more general second order approximation of the expectation of a continuously differentiable function g of $S_{N_1}^2$ as $m \rightarrow \infty$ is presented in (iv). The proof of (iv) follows by expanding the function $g(S_{N_1}^2)$ around $g(\theta)$ and using (i), (ii) and (iii) of Theorem 4.

Clearly, (i) of Theorem 4 illustrates that $S_{N_1}^2$ is biased for θ . The bias depends on the kurtosis of the underlying distribution, the form of g , δ and n^* . Similar arguments can be applied to discuss (ii) and (iii) above.

Next, the expectation of the final stage sample size N and other asymptotic characteristics may be easily obtained from (iv) of Theorem 4 above as given in the following Theorem 5.

Theorem 5

Let N be defined as in (2.2) and assume that condition (2.3) holds, then as $m \rightarrow \infty$, we have:

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

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

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

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

Proof: To prove part (i), note that $N = \lceil \lambda g(S_{N_1}^2) \rceil + 1$, *a.s.* except possibly on a set

$\xi = \left\{ \left(N_1 > \lceil \lambda g(S_{N_1}^2) \rceil + 1 \right) \cup \left(m > \lceil \delta \lambda g(S_m^2) \rceil + 1 \right) \right\}$ of measure zero, such that $\int_{\xi} N dP = o(1)$; see, for

example, Hall (1981) for details. Further discussion of this will be given later in section 7. Hence,

$$\begin{aligned} N &= \lceil \lambda g(S_{N_1}^2) \rceil + 1 \\ &= \lambda g(S_{N_1}^2) - \left\{ \lambda g(S_{N_1}^2) - \lceil \lambda g(S_{N_1}^2) \rceil \right\} + 1 \\ &= \lambda g(S_{N_1}^2) + \varepsilon_{N_1} \end{aligned} \tag{3.18}$$

Thus, $E(N) = \lambda E(g(S_{N_1}^2)) + E(\varepsilon_{N_1})$, *as* $m \rightarrow \infty$.

Using Theorem 4 part (iv), we obtain the result.

Proof of (ii): From (3.16), $Var(N) \approx \lambda^2 Var(g(S_{N_1}^2))$ *as* $m \rightarrow \infty$,

$$\begin{aligned} Var(g(S_{N_1}^2)) &= Var\left(g(\theta) + g'(\theta)(S_{N_1}^2 - \theta) + \dots + \frac{1}{2}g''(\tau)(S_{N_1}^2 - \theta)\right) \\ &= (g'(\theta))^2 Var(S_{N_1}^2) + o(\lambda^{-1}), \end{aligned}$$

where τ is a random variable between $S_{N_1}^2$ and θ . By using Theorem 4 part (iii), and the assumption that g and its derivatives are bounded, part (iii) of Theorem 5 follows similarly.

Remarks

If the underlying distribution is normal, Hall (1981, pg. 1237) proved that $\varepsilon_{N_1} \xrightarrow{L} U(0,1)$ as $\lambda \rightarrow \infty$. More generally, ε_{N_1} is continuous over the interval (0,1) and is independent of λ . This can be shown easily from (2.2) and the inequality $\lambda g(S_{N_1}^2) < \lceil \lambda g(S_{N_1}^2) \rceil + 1 \leq \lambda g(S_{N_1}^2) + 1$. We were unable to prove that $\varepsilon_{N_1} \xrightarrow{L} U(0,1)$ for general underlying distribution F . A simulation study to investigate the behaviour of ε_{N_1} for non-normal underlying distributions is given in section 7 and the results tend to support the conjecture that Hall's result applies for any continuous distribution. In particular ε_{N_1} only enters our results via its expectation and this is very close to 0.5 in all our simulations.

It is also evident from Theorem 5 above that Hall's (1981) Theorem 1 is obtained for the normal distribution case, when the optimal sample size $n^* = \lambda\theta$.

We also stress that both the expectation of N and its variance depend on the kurtosis of the underlying distribution and accordingly will reflect the amount of departure from normality while estimating the optimal sample size n^* . To reiterate, for distributions with kurtosis close to 3, $E(N)$

and $Var(N)$ will be little affected by such departures from normality. However, a more substantial effect will accrue for underlying distributions that are either very flat or very peaked.

It is also of interest to consider the general form of the expectation of a real valued continuously differential function on the final stage sample size N to be able to derive asymptotic results for all moments of N , provided they exist. We also stress that we have not assumed independence between \bar{X}_{N_1} and S_m^2 or \bar{X}_N and $S_{N_1}^2$, and therefore the above results are more general in that sense.

Theorem 6

Let $h(> 0)$ be a continuously differentiable real valued function in a neighborhood around n^* such that $Sup_{n \geq m} h(n) = O(h'''(n^*))$. Then, as $\lambda \rightarrow \infty$

$$E(h(N)) = h(n^*) - \theta^2(\beta-1)(\delta n^*)^{-1} \left\{ \left(\frac{dn^*}{d\theta} \right)^2 \left[\left(n^* \right)^{-1} h'(n^*) - (1/2)h''(n^*) \right] - (1/2)h'(n^*) \frac{d^2 n^*}{d\theta^2} \right\} + h'(n^*)E(\varepsilon_{N_1}) + o(\lambda^{-1}(|h'''(n^*)|)).$$

Proof: The proof follows by expanding the differentiable function $h(N)$ around $h(n^*)$ using Taylor series and applying the results of Theorem 5. The general form of the second order asymptotic expansion of the expectation of a real valued continuously differentiable function $h(> 0)$ enables one to obtain the expectations of positive and negative moments of N in subsequent analysis.

4. Asymptotic normality of the stopping sample size N

Theorem 7

Let N be defined as in (2.2) and assume that condition (2.3) holds. Then as $m \rightarrow \infty$, N is asymptotically normal with mean n^* and variance

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

Proof: The proof of Theorem 7 is a straightforward application of Theorem 6 above with $h(v) = e^{tv}$ and $v = (N - n^*) / \sqrt{Var(N)}$.

5. Squared error loss function to estimate the mean

In this section our main objective is to develop a triple sampling point estimation procedure to estimate the mean μ of the population. In particular, if a point estimate of the unknown μ is required, we assume that the incurred cost of estimating mean μ by the corresponding sample mean \bar{X}_n can be approximated by the following squared error loss function in (5.1) with a linear sampling cost. The literature in sequential sampling has considered several forms of higher order cost functions to

model estimation cost. However, squared error loss functions are recommended and commonly used in sequential point estimation problems (see, for example, Degroot, 1962). Therefore, we write the cost (loss) function as

$$L_n(A) = A(\bar{X} - \mu)^2 + Cn, \quad (5.1)$$

where C is the known cost per unit sample and A is a constant permitted to approach ∞ . We elaborate further on determination of A in subsequent developments. The risk associated with (1.1) can be written as

$$R_n(A) = E(L_n(A)) = AE(\bar{X}_n - \mu)^2 + Cn = A\theta/n + Cn \quad (5.2)$$

Treating n as a continuous variable in (5.2), we differentiate (5.2) with respect to n and equate the results to zero to obtain the optimal sample size as

$$n \geq n^* = \sqrt{\frac{A\theta}{C}}. \quad (5.3)$$

The value of n^* in (5.3) is unknown because the population variance θ is unknown. It has been shown by Dantzig (1940), Stein (1945, 1949) and Seelbinder (1953) that no fixed sample size procedure exists to achieve the above optimal requirement uniformly over θ . In light of (5.3), $n^* \rightarrow \infty$ in two ways, first, either A is permitted to approach infinity (extremely high cost of estimation error) or the cost of sampling is cheap.

Since the optimal sample size in (5.3) depends on the unknown variance θ , no fixed sample size procedure can be used to estimate μ uniformly over all θ . Therefore, the triple sampling procedure in (2.1)-(2.2) can be used to provide a point estimator of μ with $\lambda = \sqrt{A/C}$ and $g(\theta) = \sqrt{\theta}$. The question arises: how efficient is this estimator?

6. The asymptotic regret and efficiency of triple sampling point estimation under squared error loss function

In the literature on sequential point estimation several measures have been developed of the efficiency of the sequential procedures (triple sampling, or accelerated sequential schemes) relative to the fixed sample size counterpart had the form of $g(\theta)$ in (1.1) been completely specified. The regret reflects the expected cost of missed opportunity which measures the risk in using the triple sampling procedures to perform point estimation of the population mean instead of the fixed sample size procedure, had the nuisance parameter(s) been known. Other weaker measures like the asymptotic relative efficiency, $\eta(A) = E(L_N(A)) / E(L_{n^*}(A))$, which is the ratio of the triple sampling risk to the optimal risk are also used. For an efficient sampling procedure we expect $\eta(A) \rightarrow 1$ and that $\omega(A)$ is bounded as $n^* \rightarrow \infty$.

Theorem 8

The risk associated with the squared error loss equation (5.1) as $m \rightarrow \infty$ is given by

$$\begin{aligned}
R_N(A) &= E[L_N(A)] \\
&= 2Cn^* - C(\beta - 3) + (1/4)(\beta - 1)C\delta^{-1} + CE(\varepsilon_{N_1}) + o(1)
\end{aligned}$$

Moreover, the asymptotic relative efficiency of the triple sampling scheme and the asymptotic regret are given by

- i) $\eta(A) = 1 + o(\lambda^{-1})$,
- ii) $\omega(A) = -C(\beta - 3) + (1/4)(\beta - 1)C\delta^{-1} + CE(\varepsilon_{N_1}) + o(1)$, as $m \rightarrow \infty$.

Proof of Theorem 8:

This is immediate if we recall (ii) of Theorem 3 and (i) of Theorem 5. Moreover, the regret of triple sampling associated with equation (5.1) is given by

$$\begin{aligned}
\omega(A) &= R_N(A) - R_n^*(A) \\
&= -C(\beta - 3) + (1/4)(\beta - 1)C\delta^{-1} + CE(\varepsilon_{N_1}) + o(1)
\end{aligned}$$

If the cost of unit sampling tends to zero, then we expect zero regret. However, for C non-zero, the regret is bounded, as illustrated in Theorem 8, and depends on the kurtosis β .

Theorem 8 has several consequences. First, the case of normal distributions treated by Mukhopadhyay et al. (1987) and Hamdy (1988) are special cases. Secondly, for distributions with $\beta < 3$ a non-vanishing positive regret is expected. In addition, for distributions with $\beta > 3$ we expect either positive or negative non-vanishing regret, depending on the values of the kurtosis and the design factor δ . Specifically, for distributions with $\beta > 6$, negative regret is expected with $\delta = 1/2$. Martinsek (1988) argued that for one-by-one sequential procedures negative regret is expected when $\beta > 3$. It is also worth mentioning that the regret of purely sequential procedures depends on both the kurtosis and skewness of the underlying distribution, as indicated by Martinsek (1988), while our findings in Theorem 8 emphasize that the triple sampling procedure depends only on the kurtosis. This could be due to the nature of one-by-one purely sequential procedure which filters data. This filtration may cause either acceleration or delay. On the other hand, triple sampling is based on bulks (batches). Therefore, if an extreme observation presents, its effect on such the decision to stop or continue sampling will be diluted by the rest of the bulk at that stage, which may cause the triple sampling procedure to be less sensitive to extreme observations than one-by-one sequential procedures.

However, a general formula for the regret incurred in estimating μ with squared error loss function (6.1) can be written as

$$\omega(A) = -2C\theta(\beta - 3) \left(\frac{g'(\theta)}{g(\theta)} \right) + C\theta^2(\beta - 1)\delta^{-1} \left(\frac{g'(\theta)}{g(\theta)} \right)^2 + CE(\varepsilon_{N_1}) + o(1) \quad (6.1)$$

Obviously, the non-vanishing regret in (6.1) depends on β , δ , C and the form of $g(\theta)$. To elaborate further, consider the loss function in Martinsek (1988) of the form

$$L_n(A) = A\theta^{b-1}(\bar{X}_n - \mu)^2 + n.$$

This, under the triple sampling scheme, provides the following regret:

$$\omega(A) = -(\beta - 3)b + b^2(\beta - 1) / 4\delta + E(\varepsilon_{N_1}) + o(1) \quad (6.2)$$

The regret of the triple sampling procedure in (6.2) is the same as Martinsek's (1988) equation (7) for symmetric underlying distributions.

The above results are asymptotic and therefore to study the small to moderate sample size performance of triple sampling a series of Monte Carlo simulation experiments were performed and are presented in section 7.

7. Simulation results

Since the results obtained above are asymptotic in nature, a series of simulation studies, each based on 50,000 replications were undertaken to investigate the small, moderate, and large sample size performance of the normal-based triple sampling procedure (2.1) and (2.2) under squared error loss with (without loss of generality) $C=1$ and consequently from (5.3) $A = (n^*)^2 / \theta$. The optimal sample size n^* was allowed to vary from small to large (24, 43, 61, 76, 96, 125, 171, 246, 500); $\delta = 0.3, 0.5, 0.8$; $m = 10, 15$ and the underlying distributions were $U(0, 1)$, $N(0, 1)$ and exponential with mean 1, thereby giving $\theta = 1/12, 1$ and 1 respectively. For each replicate the stopping sample size, the stage at which sampling was stopped, the estimate of the mean, the loss and the value of ε_{N_1} were recorded. In addition, in order to investigate the practically important case of symmetric underlying distributions with thicker than normal tails, we used contaminated normal and t distributions. Relevant summary statistics for each combination of conditions were recorded across replicates and these results form the basis of the discussion given below in sections 7.1 and 7.2. To save space the results for $\delta = 0.5$ and $m = 15$ are reported; other combinations gave qualitatively similar results.

7.1 Termination stage, the distribution of ε_{N_1} and regret

Tables 1, 2 and 3 show the estimated probabilities of stopping after the first and second stages, summary statistics for the mean and standard deviation of ε_{N_1} , the p-value of a Kolmogorov-Smirnov test of uniformity for the distribution of ε_{N_1} and the estimated regret.

For the normal and uniform cases (Tables 1 and 2) there is little early stopping unless n^* is small, whereas for the exponential case (Table 3) the proportion of early stopping declines more slowly.

For all three distributions the hypothesis that the ε_{N_1} follow a $U(0, 1)$ distribution mostly cannot be rejected even on the basis of 50,000 replicates when n^* is large, which suggests that the asymptotic result of Hall(1981) for the distribution of ε_{N_1} may also hold for other underlying distributions. In particular note that the estimate of $E(\varepsilon_{N_1})$ is very close to 0.5 in all our simulations.

The estimated regret is rather volatile in the simulation results but it is interesting that in Table 3 the estimated regret approaches the limiting value, -1.5 , from above. It appears that the asymptotic

results in Theorem 8(ii), which let $m \rightarrow \infty$, are poor approximations to the actual regret when m is not large, such as in our case with $m=15$.

(Tables 1, 2, 3 near here)

7.2 The estimator for the mean and the stopping sample size

Table 4 shows estimates obtained from the simulations for the mean and standard deviation of \bar{X}_N together with the corresponding asymptotic values obtained from Theorem 3 for the $N(0, 1)$, $U(0, 1)$ and exponential(1). There is good agreement with the asymptotic results for the estimator of the mean: it is clearly unbiased for symmetric underlying distributions and the estimated standard deviation is close to the asymptotic value except for $n^*=24$, where m/n^* is not less than the design factor δ (as required by (2.3)). For the exponential distribution the slight negative bias for the estimator of the mean predicted by Theorem 3 is clearly present but the asymptotic standard deviation is lower than the actual estimated variance for intermediate values of n^* , probably a result of the early stopping seen in Table 3.

Table 5, for the same set of underlying distributions, shows estimates obtained from the simulations for the mean and standard deviation of the stopping sample size N together with the corresponding asymptotic results obtained from Theorem 5. Table 6 shows the corresponding results for two contaminated normal mixture distributions ($N(0, 1)$ with probability 0.9 or 0.95 and $N(0, 9)$ with probability 0.1 or 0.05), which model moderately thick tailed underlying distributions, and for the $t(5)$ distribution, which models a heavy tailed underlying distribution. Indeed since $t(5)$ has only four moments, it can be thought of as the most extreme symmetric underlying distribution for our asymptotic results. Apart from the case $n^*=24$ the mean stopping sample size is well approximated by the asymptotic result which has been calculated assuming that $E(\varepsilon_{N_i})=0.5$. For the variability in N the situation is more complicated but when $n^*>24$ for the lighter tailed underlying distributions (normal, uniform) the asymptotic standard deviation is an underestimate, whereas for the heaviest tailed distribution $t(5)$ the asymptotic standard deviation is an overestimate.

(Tables 4 to 6 near here)

References

- Anscombe, F.J., 1953. Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1-21.
- Bhattacharjee, G.P., 1965. Effect of non-normality on Stein's two sample test. *Ann. Math. Statist.* **36**, No. 2, 651-663.
- Bhattacharjee, G.P., and Nagendra, Y., 1964. Effect of non-normality on a sequential test for mean. *Biometrika*, **51** 281-287.
- Blumenthal, S., and Govindarajulu, Z., 1977. Robustness of Stein's two-stage procedure for mixtures of normal populations. *J. Amer. Statist. Assoc.* **72** 192-196.
- Chow, Y.S., and Robbins, H., 1965. On the asymptotic theory of fixed width confidence intervals for the mean. *Ann. Math. Statist.* **36** 457-462.
- Dantzig, G.B., 1940. On the non existence of tests of "students" hypothesis having power functions independent of σ . *Ann. Math. Statist.* **11** 186- 192.

- Ghosh, M., Mukhopadhyay, N. and Sen, P.K., 1997. Sequential Estimation. Wiley, New York.
- Hall, P., 1981. Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.*, **9** 1229-1238.
- Hamdy, H.I., 1988. Remarks on the asymptotic theory of triple stage estimation of the normal mean. *Scand. J. Statist.* **15** 303-310.
- Hamdy, H.I., and Palotta, W. J., 1987. Triple sampling procedure for estimating the scale parameter of Pareto distribution. *Commun. Statist., Theory and Methods*, **16** 2155-2164.
- Johnson, N.L., Kotz, S., and Balakrishnan, N., 1994. Continuous Univariate Distributions, Vol. I. Wiley, New York.
- Jureckova, J., and Sen, P.K., 1996. Robust Statistical Procedures, Asymptotics and Interrelations. Wiley, New York.
- Liu, W., 1997a. A k-stage sequential sampling procedure for estimation of normal mean. *J. Statist. Planning and Inference*, **65** 109-127.
- Liu, W., 1997b. Improving the fully sequential estimation method of Anscombe-Chow-Robbins. *Ann. Statist.*, **25**, 2164-2171.
- Liu, W., Wang, N. and Wang, S., 2002. Approximations of distributions of some standardized partial sums in sequential analysis. *Austral. New Zealand J. Statist.*, **44** 109-119.
- Martinsek, A.T., 1988. Negative regret, optimal stopping, and the elimination of outliers. *J. Amer. Statist. Assoc.* **83** 160-163.
- Mukhopadhyay, N., 1985. A note on three stage and sequential point estimation procedures for a normal mean. *Sequential Anal.*, **4** 311-319.
- Mukhopadhyay, N. and de Silva, B.M., 2009. Sequential Methods and Their Applications. CRC, Boca Raton.
- Mukhopadhyay, N., Hamdy H.I., AL-Mahmeed M., and Costanza M.C., 1987. Three stage point estimation procedures for a normal mean. *Sequential Anal.*, **6** 21-36.
- Mukhopadhyay, N., and Mavromoustakos, A., 1987. Three stage estimation procedures of the negative exponential distribution. *Metrika*, **34** 83-93.
- Mukhopadhyay, N. and Solanky, T.K.S., 1994. Multistage Selection and Ranking Procedures. Dekker, New York.
- Ramkaran, 1983. The robustness of Stein's two-stage procedure. *Ann. Statistic.* **11** 1251-1256.
- Robbins, H., 1959. Sequential estimation of the mean of a normal population. Probability and Statistics (Harald Cramer Volume) 235-245. Almqvist and Wiksell, Uppsala, Sweden.
- Rohatgi, V.K., 1976. An Introduction to Probability Theory and Mathematical Statistics. Wiley, New York, 303.
- Seelbinder, D.M., 1953. On stein's two-stage procedure. *Ann. Math. Statist.* **24** 640-649. [458, 459]
- Stein, C., 1945. A two-stage sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243-258.
- Stein, C. 1949. Some problems in sequential estimation. *Econometrika.* **17** 77-78.
- Vik, G. and Mukhopadhyay, N. 1988. Triple sampling to construct fixed-width confidence intervals for estimable parameters based on U-statistics. *Metron* **46** 165-174

n^*	$P(N = m)$	$P(N = N_1)$	$E(\varepsilon_{N_1})$	$Sd(\varepsilon_{N_1})$	p-value	ω
24	0.768	0.000	0.497	0.288	0.007	5.232
43	0.221	0.001	0.495	0.289	0.001	3.938
61	0.061	0.003	0.498	0.288	0.032	1.652
76	0.024	0.005	0.497	0.288	0.003	1.842
96	0.007	0.007	0.500	0.289	0.821	0.633
125	0.002	0.009	0.501	0.289	0.507	1.320
171	0.000	0.010	0.499	0.288	0.507	1.413
246	0.000	0.012	0.497	0.289	0.015	2.564
500	0.000	0.013	0.498	0.288	0.154	-0.101

Table 1 Estimated probabilities of early stopping, estimated mean and standard deviation of ε_{N_1} and the estimated regret (asymptotic value 1.5) for the $N(0, 1)$ case with $m = 15$ and $\delta = 0.5$.

n^*	$P(N = m)$	$P(N = N_1)$	$E(\varepsilon_{N_1})$	$Sd(\varepsilon_{N_1})$	p-value	ω
24	0.839	0.000	0.496	0.288	0.006	5.471
43	0.112	0.000	0.497	0.289	0.006	5.219
61	0.016	0.000	0.501	0.288	0.250	2.970
76	0.005	0.000	0.501	0.288	0.759	2.154
96	0.001	0.000	0.498	0.290	0.113	1.710
125	0.000	0.000	0.500	0.288	0.400	1.287
171	0.000	0.000	0.498	0.289	0.151	1.466
246	0.000	0.000	0.501	0.289	0.211	5.923
500	0.000	0.000	0.500	0.288	0.303	7.641

Table 2 Estimated probabilities of early stopping, estimated mean and standard deviation of ε_{N_1} and the estimated regret (asymptotic value 2.1) for the $U(0, 1)$ case with $m = 15$ and $\delta = 0.5$.

n^*	$P(N = m)$	$P(N = N_1)$	$E(\varepsilon_{N_1})$	$Sd(\varepsilon_{N_1})$	p-value	ω
24	0.748	0.010	0.495	0.289	<0.001	-1.494
43	0.413	0.035	0.499	0.289	0.131	17.779
61	0.232	0.056	0.498	0.288	0.076	22.727
76	0.150	0.065	0.496	0.289	0.013	18.896
96	0.084	0.072	0.499	0.289	0.164	15.153
125	0.040	0.078	0.499	0.288	0.145	8.634
171	0.014	0.085	0.499	0.289	0.324	7.505
246	0.005	0.083	0.498	0.289	0.019	9.994
500	0.000	0.084	0.501	0.288	0.439	9.809

Table 3 Estimated probabilities of early stopping, estimated mean and standard deviation of ε_{N_1} and the estimated regret (asymptotic value -1.5) for the exponential case with mean 1 with $m = 15$ and $\delta = 0.5$.

Underlying distribution						
n^*	Normal (0, 1)		Uniform (0, 1)		Exponential (1)	
	$E(\bar{X}_N)$	$Sd(\bar{X}_N)$	$E(\bar{X}_N)$	$Sd(\bar{X}_N)$	$E(\bar{X}_N)$	$Sd(\bar{X}_N)$
24	0.0008 (0)	0.253 (0.214)	0.5003 (0.5)	0.074 (0.062)	0.970 (0.958)	0.219 (0.220)
43	0.0007 (0)	0.162 (0.157)	0.5000 (0.5)	0.047 (0.045)	0.951 (0.977)	0.183 (0.159)
61	-0.0005 (0)	0.131 (0.131)	0.4999 (0.5)	0.038 (0.038)	0.971 (0.984)	0.152 (0.132)
76	-0.0008 (0)	0.116 (0.117)	0.5003 (0.5)	0.034 (0.034)	0.981 (0.987)	0.130 (0.118)
96	-0.0004 (0)	0.103 (0.103)	0.5001 (0.5)	0.030 (0.030)	0.986 (0.990)	0.111 (0.104)
125	0.0001 (0)	0.090 (0.090)	0.5002 (0.5)	0.026 (0.026)	0.991 (0.992)	0.095 (0.091)
171	0.0005 (0)	0.077 (0.077)	0.5000 (0.5)	0.022 (0.022)	0.993 (0.994)	0.079 (0.077)
246	0.0004 (0)	0.064 (0.064)	0.4999 (0.5)	0.019 (0.018)	0.996 (0.996)	0.066 (0.064)
500	-0.0002 (0)	0.045 (0.045)	0.5000 (0.5)	0.013 (0.013)	0.998 (0.998)	0.045 (0.045)

Table 4 Estimated mean and standard deviation (asymptotic values in parentheses) of \bar{X}_N for $N(0, 1)$, $U(0, 1)$ and the exponential case with mean 1 with $m = 15$ and $\delta = 0.5$.

Underlying distribution						
n^*	Normal (0, 1)		Uniform (0, 1)		Exponential (1)	
	$E(N)$	$Sd(N)$	$E(N)$	$Sd(N)$	$E(N)$	$Sd(N)$
24	16.36(23)	4.64(4.90)	15.24 (23.9)	1.96 (3.10)	18.00(18.5)	7.11 (9.80)
43	41.11 (42)	8.83 (6.56)	42.52 (42.9)	5.55 (4.15)	36.70(37.5)	14.28(13.11)
61	59.78 (60)	8.54 (7.81)	60.77 (60.9)	5.49 (4.94)	55.42(55.5)	16.22(15.62)
76	74.90 (75)	9.33 (8.72)	75.82 (75.9)	5.97 (5.51)	70.67(70.5)	18.11(17.44)
96	94.88 (95)	10.42(9.80)	95.76 (95.9)	6.64 (6.20)	90.71(90.5)	20.44(19.60)
125	123.90(124)	11.74(11.18)	124.86(124.9)	7.47 (7.07)	119.59(119.5)	23.36(22.36)
171	169.83(170)	13.72(13.08)	170.86(170.9)	8.60 (8.27)	165.67(165.5)	27.64(26.15)
246	244.95(245)	16.34(15.68)	245.78(245.9)	10.26 (9.92)	241.02(240.5)	33.55(31.37)
500	498.82(499)	23.19(22.36)	499.86(499.9)	14.54(14.14)	495.23(494.5)	48.71(44.72)

Table 5 Estimated mean and standard deviation (asymptotic values in parentheses) of N for $N(0, 1)$, $U(0, 1)$ and the exponential case with mean 1 with $m = 15$ and $\delta = 0.5$.

Underlying distribution						
0.95N(0, 1)+0.05N(0, 9)		0.9N(0, 1)+0.1N(0, 9)			t(5)	
n^*	$E(N)$	$Sd(N)$	$E(N)$	$Sd(N)$	$E(N)$	$Sd(N)$
24	17.48(18.76)	6.48(8.94)	17.95(19)	6.93(9.38)	17.33(18.5)	6.37(9.80)
43	38.26(37.76)	12.07(11.96)	36.97(38)	13.42(12.55)	38.65(37.5)	12.22(13.11)
61	57.40(55.76)	12.75(14.24)	56.38 (56)	14.09(14.96)	57.65(55.5)	12.75(15.62)
76	72.37(70.76)	14.53(15.90)	71.32(71)	16.10(16.69)	72.70(70.5)	14.38(17.44)
96	92.42(90.76)	16.77(17.87)	91.53(91)	18.56(18.76)	92.81(90.5)	16.66(19.60)
125	121.47(119.76)	19.68(20.39)	120.52(120)	21.24(21.41)	121.68(119.5)	19.65(22.36)
171	167.40(165.76)	23.70(23.85)	166.59(166)	25.71(25.04)	167.73(165.5)	23.30(26.15)
246	242.25(240.76)	28.85(28.61)	241.57(241)	31.30(30.03)	242.64(240.5)	28.68(31.37)
500	496.56(494.76)	42.26(40.78)	495.91(495)	45.39(42.82)	496.85(494.5)	43.31(44.72)

Table 6 Estimated mean and standard deviation (asymptotic values in parentheses) of N for two contaminated normal underlying distributions and $t(5)$ with $m = 15$ and $\delta = 0.5$.