The Korean Communications
in Statistics Vol. 5, No. 2, 1998
pp. 539-557

Three Stage Estimation for the Mean of a One-Parameter Exponential Family

M. AlMahmeed[1], A. Al-Hessainan[2], M. S. Son[3] and H. I. Hamdy[4]

Abstract

This article is concerned with the problem of estimating the mean of a one-parameter exponential family through sequential sampling in three stages under quadratic error loss. This more general framework differs from those considered by Hall (1981) and others. The differences are : (i) the estimator and the final stage sample size are dependent; and (ii) second order approximation of a continuously differentiable function of the final stage sample size permits evaluation of the asymptotic regret through higher order moments. In particular, the asymptotic regret can be expressed as a function of both the skewness ρ and the kurtosis β of the underlying distribution. The conditions on ρ and β for which negative regret is expected are discussed. Further results concerning the stopping variable N are also presented. We also supplement our theoretical findings with simulation results to provide a feel for the triple sampling procedure presented in this study.

1. Introduction

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables from the one parameter natural exponential family

$$dF_a(x) = \int_R e^{ax - \psi(a)} dP(x), \quad x \in R, \quad a \in \Omega \tag{1.1}$$

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

1) Associate Professor of Statistics, Head of Department of Quantitative Methods and Information Systems, Kuwait University, Kuwait, P.O. Box 5486, Saffat, Kuwait.
2) Kuwait University, Kuwait, P.O. Box 5486, Saffat, Kuwait.
3) Professor of Statistics, Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05401, USA
4) Professor of Statistics, Head of Department of Quantitative Methods and Information Systems, Kuwait University, Kuwait, P.O. Box 5486, Saffat, Kuwait.

$$\int_R e^{\alpha x} dP(x) < \infty , \quad \alpha \in \Omega.$$

The function $\Psi(\cdot)$ is convex on Ω (e.g. see Lehmann (1986, pg. 57)), satisfying the moment relation $E(e^{tx}) = e^{\Psi(t+\alpha) - \Psi(\alpha)}$.
Hence

$$E(X) = \Psi'(\alpha) = \theta, \quad Var(X) = \Psi''(\alpha),$$

$$E(X-\theta)^3 = \Psi'''(\alpha), \quad E(X-\theta)^4 = \Psi''''(\alpha) + 3\Psi''^2(\alpha),$$

with

$$\rho = \Psi'''(\alpha)/\Psi''^{3/2}(\alpha) \text{ and } \beta = \Psi''''(\alpha)/\Psi''^2(\alpha) + 3,$$

where primes mean derivatives and ρ and β denote the coefficients of skewness and kurtosis of the distribution F_α respectively.

Let $X_1, X_2, \ldots, X_n$, $(n \geq 2)$ be a random sample from the distribution F_α. It is not hard to prove that

$$E(\overline{X}_n) = \theta, \quad Var(\overline{X}_n) = n^{-1}\Psi''(\alpha),$$

$$E(\overline{X}_n - \theta)^3 = n^{-2}\Psi'''(\alpha), \quad E(\overline{X}_n - \theta)^4 = n^{-3}\{\Psi''''(\alpha) + 3(n-1)\Psi''^2(\alpha)\}.$$

Naturally, we propose to use the maximum likelihood estimator (MLE) $\overline{X}_n$ to estimate the unknown parameter θ subject to the following quadratic loss function with linear sampling cost,

$$L_n(A) = A^2(\overline{X}_n - \theta)^2 + n, \tag{1.2}$$

where $A > 0$ is a known constant. It follows from (1.2) that the risk of estimating θ by $\overline{X}_n$ is given by

$$E(L_n(A)) = A^2\Psi''(\alpha)n^{-1} + n \tag{1.3}$$

Minimizing the risk in (1.3) results in the optimal sample size, $n^* = A\sqrt{\Psi''(\alpha)}$. The variance function $\Psi''(\alpha)$ depends on $\theta = \Psi'(\alpha)$ and so does every function of α, see Morris (1982) for details.
Therefore, we write n^* as

$$n^* = Ag(\theta) \tag{1.4}$$

Further, it is possible to relate $g'(\theta)$ and $g''(\theta)$ to ρ and β of the underlying distribution

F_α.

Recall $\theta = \Psi(\alpha)$, then

$$\Psi'(\alpha)\left(\frac{d\alpha}{d\theta}\right) = 1,$$

and

$$\Psi''\left(\frac{d\alpha}{d\theta}\right)^2 + \Psi'(\alpha)\left(\frac{d^2\theta}{d\theta^2}\right) = 0.$$

Thus,

$$g'(\theta) = (1/2)\,\Psi'(\alpha)\ \rho\ \frac{d\alpha}{d\theta} \qquad (1.5)$$
$$= (1/2)\,\rho,$$

and

$$g''(\theta) = (1/2)(\beta - (3/2)\rho^2 - 3)/\sqrt{\Psi'(\alpha)}. \qquad (1.6)$$

It is obvious from (1.6) that if $g''(\theta) = 0$ then $\beta = (3/2)\rho^2 + 3$.
Had θ been known, then the optimal risk would be

$$E(L_{n^*}(A)) = 2n^*.$$

In applications, however n^* is unknown because θ is unknown. Therefore, in (1.4) we mimic $g(\theta)$ by $g(\overline{X}_n)$ to develop the following three stage sampling procedure to estimate θ via estimate of n^*.

2. Sequential Sampling in Three Stage

The idea of sequential sampling in three stages was first introduced by Hall (1981) to construct a fixed width confidence interval for a normal mean. Unlike one-by-one, purely sequential sampling, three-stage sampling involves sampling in consecutive groups. This sampling technique was designed to combine the operational savings made possible by sampling in batches with the efficiency of purely sequential procedures.

The three-stage procedure begins with an initial random sample of size $m(\geq 2)$ from the distribution F_α to compute the estimate $g(\overline{X}_m)$ of $g(\theta)$. Then, a fraction $\gamma \in (0,1)$ is selected to determine the percentage of n^* to be estimate in the second stage. Accordingly, the second stage sample size is determined by the stopping rule

$$N_1 = \max\{m,\ [\gamma A g(\overline{X}_m)] + 1\}, \qquad (2.1)$$

where $[\cdot]$ denotes the largest integer function. If the decision is to continue sampling, the initial sample is augmented with a second randomly selected batch of size $N_1 - m$ to determine the final sample size from the stopping rule

$$N = \max\{N_1,\ [Ag(\overline{X}_{N_1})] + 1\}, \tag{2.2}$$

If necessary, a third batch of size $N - N_1$ is randomly selected and combined with the previous N_1 observations to compute the sample mean $\overline{X}_N$ as an estimate for the unknown parameter θ.

Decision rules similar to (2.1)-(2.2) were considered by Woodroofe (1987) in studying the asymptotic local minimax regret of the three-stage point estimation procedure. The three sampling stages were aptly termed *the pilot study, the main study, and the fine-tuning.*

In this study we provide rigorous derivations of the asymptotic second order characteristics of the three-stage scheme (2.1)-(2.2). In particular, we thoroughly develop large sample properties of a continuously differentiable function of N instead of focusing solely on rational powers of N. Moreover, we obtain a sharp estimate for the asymptotic regret associated with the above sampling technique within the quadratic error loss function (1.2) framework. The phenomenon of negative regret outlined by Martinsek (1998) is highlighted in this context as well. The asymptotic normality of the stopping variable N is also explored through the moment generating function of N. Applications to some standard distributions are illustrated for comparison with some previous known results. The present problem also differs from those considered by many authors including Hall (1981), Mukhopadhyay et al. (1987), Hamdy (1988), Hamdy et al. (1989), Hamdy and Son (1991), Hamdy (1995) and Hamdy et al. (1995) in that the sample mean and the final stage sample size are highly correlated and therefore our approach took a different turn.

Throughout the following section the asymptotic characteristics of the three stage sampling scheme are developed under Hall's (1981) assumptions that $A = O(m^r)$, $r > 1$ and lim sup $(m/n^*) < \gamma$ as $m \to \infty, A \to \infty$. This setup implies that as $m \to \infty, A \to \infty$ at a faster rate. Lemma 1 gives preliminary results concerning the above triple sampling procedure.

Lemma 1. For the three stage rule (2.1)-(2.2), if g and its derivatives are bounded, then as $m \to \infty$

(i) $AE(\overline{X}_{N_1} - \theta) = -(1/2)\rho\gamma^{-1} + o(1)$;

(ii) $AE(\overline{X}_{N_1}-\theta)^2=\sqrt{\Psi''(\alpha)}\ \gamma^{-1}+o(1)$;

(iii) $AE(g(\overline{X}_{N_1})-g(\theta))=(1/4)(\beta-(5/2)\rho^2-3)\gamma^{-1}+o(1)$.

Proof of Lemma 1. Conditioning on the σ-field generated by $X_1,\dots,X_m$, we have

$$E(\overline{X}_{N_1})=EE(\overline{X}_{N_1}|X_1,X_2,\dots X_m)=E(N_1^{-1}(\sum_{i=1}^{m}X_i+\theta(N_1-m))),$$

and

$$E(\overline{X}_{N_1}-\theta)=m\ E\{N_1^{-1}(\overline{X}_m-\theta)\}.$$

We then expand N_1^{-1} around $g(\theta)$ and obtain

$$N_1^{-1}=(\gamma n^*)^{-1}-(g(\overline{X}_m)-g(\theta))\ g^{-1}(\theta)\ (\gamma n^*)^{-1}+(\gamma A)^{-1}(g(\overline{X}_m)-g(\theta))^2\ \tau^{-3},$$

where τ is a random variable lying between $g(\overline{X}_m)$ and $g(\theta)$. It follows that

$$\begin{aligned}E(\overline{X}_{N_1}-\theta)&=-mg^{-1}(\theta)(\gamma n^*)^{-1}E\{(\overline{X}_m-\theta)(g(\overline{X}_m)-g(\theta))\}\\&\quad+m(\gamma A)^{-1}E\{(\overline{X}_m-\theta)(g(\overline{X}_m)-g(\theta))^2\tau^{-3}\}\\&=I+II.\end{aligned}$$

Say.

Now in I, we expand $g(\overline{X}_m)$ around θ to obtain

$$I=-mg^{-1}(\theta)(\gamma n^*)^{-1}\ E\{(\overline{X}_m-\theta)\{g(\theta)+(\overline{X}_m-\theta)g'(\theta)+(1/2)(\overline{X}_m-\theta)^2g''(\upsilon)-g(\theta)\}\},$$

where υ is a random variable lying between $\overline{X}_m$ and θ. Hence,

$$I=mg^{-1}(\theta)(\gamma n^*)^{-1}\ g'(\theta)E(\overline{X}_m-\theta)^2-(1/2)mg^{-1}(\theta)(\gamma n^*)^{-1}E\{(\overline{X}_m-\theta)^3g''(\upsilon)\}$$

Next we evaluate $E\{(\overline{X}_m-\theta)^3g''(\upsilon)\}$. Since $g''\le K_1$ is bounded, where K_1 is a generic constant independent of $\overline{X}_m$, υ, m, and A, it follows that

$$\begin{aligned}(1/2)mg^{-1}(\theta)(\gamma n^*)^{-1}E\{(\overline{X}_m-\theta)^3g''(\upsilon)\}&\le(1/2)K_1mg^{-1}(\theta)(\gamma n^*)^{-1}E(\overline{X}_m-\theta)^3\\&=(1/2)K_1mg^{-1}(\theta)(\gamma n^*)^{-1}\Psi'''(\alpha)/m^2\end{aligned}$$

$$\leq (1/2) K_1 m g^{-2}(\theta)(\gamma A)^{-1} \Psi'(\alpha)$$
$$= o(A^{-1}) \text{ as } m \to \infty$$

Hence, we obtain

$$I = -(1/2)(\gamma A)^{-1}\rho + o(A^{-1}), \qquad \text{by (1.5).}$$

Next, we proceed to evaluate II. The function $g' \leq \sqrt{K_2}$ is bounded by some generic constant $\sqrt{K_2}$ independent of $\overline{X}_m$, τ, m and A. Therefore,

$$|(g(\overline{X}_m) - g(\theta))^2| \leq K_2(\overline{X}_m - \theta)^2$$

Thus,

$$II \leq K_2 m(\gamma A)^{-1} E\{(\overline{X}_m - \theta)^3 \tau^{-3}\}.$$

Recall that τ lies between $g(\theta)$ and $g(\overline{X}_m)$, so we discuss the following two cases. First, if $g(\theta) \leq \tau \leq g(\overline{X}_m)$, then $\tau^{-1} \leq g^{-1}(\theta)$. Subsequently, we have

$$\begin{aligned} K_2 m(\gamma A)^{-1} E\{(\overline{X}_m - \theta)^3 \tau^{-3}\} &\leq K_2 m(\gamma A)^{-1} g^{-3}(\theta) E(\overline{X}_m - \theta)^3 \\ &= K_2 m(\gamma A)^{-1} g^{-3}(\theta) \Psi''(\alpha)/m^2 \\ &\leq K_2 m(\gamma A)^{-1} g^{-3}(\theta) \Psi''(\alpha) \\ &= o(A^{-1}) \text{ as } m \to \infty \end{aligned}$$

Second, if $g(\overline{X}_m) \leq \tau \leq g(\theta)$, we have $(A\tau)^{-1} \leq (N_1 - 1)^{-1} \leq (m-1)^{-1}$, which gives

$$\begin{aligned} K_2 m(\gamma A)^{-1} E\{(\overline{X}_m - \theta)^3 \tau^{-3}\} &\leq K_2(\gamma)^{-1} A^2 \Psi''(\alpha)/(m-1)^4 \\ &= o(A^{-1}) \text{ as } m \to \infty, \end{aligned}$$

where we have used the assumption that $A = o(m^{-r})$ for $r > 1$. Thus, $II = o(A^{-1})$ as $m \to \infty$.

Collecting terms, we have

$$A\, E(\overline{X}_{N_1} - \theta) = I + II = -(1/2)\gamma^{-1}\rho + \theta(1) \text{ as } m \to \infty.$$

This proves (i) of Lemma 1. Part (ii) can be justified in a similar manner and therefore we omit further details.

To prove (iii), we expand $g(\overline{X}_{N_1})$ in a Taylor series around θ and take the expectation throughout to obtain

$$Eg(\overline{X}_{N_1}) = g(\theta) + E(\overline{X}_{N_1} - \theta)g'(\theta) + (1/2)E(\overline{X}_{N_1} - \theta)^2 g''(\theta) + E(R_{1N_1}). \tag{2.3}$$

Similar arguments to those used above and Lemma (4.1) of Woodroofe (1987) yield $E(R_{1N_1}) = o(A^{-1})$. Now, substitute equations (1.5), (1.6), and (i) and (ii) of Lemma 1 into (2.3) to obtain (iii) of Lemma 1, which completes the proof.

It remains to elaborate on our findings in Lemma 1. It is clear from (i) of Lemma 1 that $\overline{X}_{N_1}$ is biased for θ. The amount for bias depends on the skewness ρ of the underlying distribution F_α as well as the design factor γ, which represents the fraction n^* of we wish to estimate in the second stage, and the known constant A which reflects the decision maker's attitude towards estimation risk. Martinsek (1988) obtained a result similar to (i) of Lemma 1 for the non-parametric case for purely sequential sampling. It is of interest to study whether the third stage information about θ will indeed reduce the bias noticed in (i) of Lemma 1. It also follows from (iii) of Lemma 1 that

$$A\,Var(g(\overline{X}_{N_1})) = (1/4)\rho^2\gamma^{-1}\sqrt{\Psi''(\alpha)} + o(1). \tag{2.4}$$

As we proceed to develop theory for triple sampling, we focus upon studying the asymptotic features of this group sampling technique following Chow and Robbins (1965). Some preliminary results concerning the stopping variable N are given in Lemma 2.

Lemma 2. Under the conditions in Lemma 1, we have as $m \to \infty$

(i) $E(N) = n^* + (1/2)(1 + (1/2)(\beta - (5/2)\rho^2 - 3)\gamma^{-1}) + o(1)$;

(ii) $A^{-1}Var(N) = (1/4)\sqrt{\Psi''(\alpha)}\rho^2\gamma^{-1} + o(1)$;

(ii) $A^{-2}E(|N - n^*|^3) = o(1)$.

Proof of Lemma 2. To prove (i), note that $N = [Ag(\overline{X}_{N_1})] + 1$ except perhaps on a set

$$\xi \equiv \{(N_1 > ([Ag(\overline{X}_{N_1})] + 1)) \cup (m > [\gamma A\, g(\overline{X}_m)] + 1)\}$$

such that $\int_{\xi} N\, dP = o(1)$; see for example Hall (1981) for details. Hence,

$$E(N) = A\ E(g(\overline{X}_{N_1}) + E(\phi) + o(1) \tag{2.5}$$

where $\phi \equiv 1 - A(g(\overline{X}_{N_1}) - [A\ g(\overline{X}_{N_1})])$ If the underlying distribution is smooth as in the case of the normal distribution, Hall (1981, pg.1237) proved that ϕ is asymptotically uniform over (0,1) and therefore $E(\phi) = 1/2$. Generally however, the random variable ϕ is continuous over (0,1) independent of A. This can be shown easily from the rule (2.2) and the inequality

$$Ag(\overline{X}_{N_1}) < [A\ g(\overline{X}_{N_1})] + 1 \le A\ g(\overline{X}_{N_1}) + 1.$$

However, if m is large enough, a simple application of Anscombe's (1952) central limit theorem shows that $g(\overline{X}_{N_1}) \sim N\,(g(\theta), \delta)$, where $\delta = (1/4)\rho^2 \Psi'(\alpha)(\gamma n^*)^{-1}$. Nevertheless, from a practical standpoint, it seems reasonable to set $E(\phi) = 1/2$. We then substitute (iii) of Lemma 1 in (2.5); thus (i) of Lemma 2 is straightforward.

To prove (ii) we write the variance of N as

$$Var(N) \approx A^2\ Var(g(\overline{X}_{N_1})) + 2Cov(\phi, g(\overline{X}_{N_1})) + o(A) \text{ as } m \to \infty \tag{2.6}$$

Next, the Cauchy-Schwarz inequality yields

$$\begin{aligned} Cov^2(\phi, g(\overline{X}_{N_1})) &\le Var(\phi)\, Var(g(\overline{X}_{N_1})) \\ &= o(A^{-1}) \quad \text{as } m \to \infty, \end{aligned} \tag{2.7}$$

by (2.4) Thus, substitute (2.4) and (2.7) in (2.6) and then (ii) of Lemma 1 is immediate.
To prove (iii) of Lemma 2 we write

$$\begin{aligned} A^{-2}\ |N - n^*|^3 &= A\left|g(\overline{X}_{N_1}) - g(\theta)\right|^3 \\ &\le AK_3\left|\overline{X}_{N_1} - \theta\right|^3, \end{aligned}$$

where K_3 is another generic constant. Part (iii) of Lemma 1 can be used to prove that $E(|\overline{X}_{N_1} - \theta|^3) = o(A^{-1})$ and therefore, $A^{-2}\ E(|N - n^*|^3) = o(1)$. This completes the proof of Lemma 2.

The following Theorem 1 provides a second order asymptotic expansion for the expectation of a continuously differentiable function h of N.

Theorem 1. Let h be a continuously differentiable function in a neighborhood of n^*, such that $\sup_{n\geq m}|h'''(n^*)|=O\ (|h'''(n^*)|)$. then, as $m\rightarrow\infty$,

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

Proof of Theorem 1. A Taylor expansion of $h(N)$ gives

$$E(h(N))=h(n^*)+E(N-n^*)h'(n^*)+(1/2)E(N-n^*)^2h''(n^*)+E(R_{1N}), \tag{2.8}$$

where $|R_{1N}|=(1/6)|(N-n^*)^3h'''(\eta)|$. The random variable η lies between N and n^*. Therefore,

$$E|R_{1N}|\ \leq\ (1/6)\ \sup_{n\geq m}\ |h'''(n)|\ E(|N-n^*|^3)$$
$$=\ o(A^2|h'''(A)|)$$

as $m\rightarrow\infty$ by (iii) of Lemma 2 and the assumption that h''' is bounded. Equation (2.8) and (ii) of Lemma 2 complete the proof of Theorem 1.

Lemma 3. Under the conditions in Lemma 1, we have as $m\rightarrow\infty$,

(i) $E(\overline{X}_N)=\theta-(1/2)\rho A^{-1}+o(A^{-1})$;

(ii) $A^2E(\overline{X}_N-\theta)^2=n^*+(1/4)\{(1/2)(7+16\gamma)\rho^2-(1+4\gamma)\beta+(3+10\gamma)\}\gamma^{-1}+o(1)$.

Proof of Lemma 3. Conditioning on the σ-field generated by $X_1,\ldots,X_{N_1}$, we write

$$E(\overline{X}_N-\theta)=E\{N_1N^{-1}(\overline{X}_{N_1}-\theta)\}.$$

We then expand N^{-1} around θ with a remainder R_{2N},

$$N^{-1}=n^{*-1}-(\overline{X}_{N_1}-\theta)g'(\theta)g^{-1}(\theta)n^{*-2}+R_{2N}.$$

Consequently, we have

$$E(\overline{X}_N-\theta)=-E\{N_1(\overline{X}_{N_1}-\theta)^2\}g'(\theta)g^{-1}(\theta)n^{*-1}+E(R_{3N}).$$

Lemma (3.4) of Woodroofe (1987) gives $E(N_1(\overline{X}_{N_1}-\theta)^2)=\Psi'(a)$. Thus,

$$E(\overline{X}_N-\theta) = -(1/2)\rho A^{-1}+E(R_{3N}).$$

It remains to estimate $E(R_{3N})$. Now,

$$(R_{3N}) \le K_4 A^2\{N_1(\overline{X}_{N_1}-\theta)(g(\overline{X}_{N_1})-g(\theta))^2\zeta^{-3}\},$$

for some generic constant K_4 independent of N and m, where ζ is a random variable lying between N and n^*. If $n^* \le \zeta \le N$ we have

$$E(R_{3N}) \le K_4 A^{-1} E\{ N_1^2(\overline{X}_{N_1}-\theta)^3\}=o(A^{(-1)}) \text{ as } A\to\infty$$

where $E\{ N_1^2(\overline{X}_{N_1}-\theta)^3\}=\Psi''(a)$ as $A\to\infty$ by Lemma (3.4) of Woodroofe (1987). On the other hand, if $N < \zeta < n^*$, we obtain

$$\begin{aligned} E(R_{3N}) &\le K_4 A^2 E\{ N_1^2(\overline{X}_{N_1}-\theta)^3\}/m^3 \\ &= o(A^{-1}) \text{ as } A\to\infty, \end{aligned}$$

where we have used Lemma (3.4) of Woodroofe (1987) and the condition $n^*=o(m^r)$ for $r>1$.

We now turn to prove (ii) of Lemma 3. Consider the expression

$$\begin{aligned} A^2E(\overline{X}_N-\theta)^2 &= A^2E\{E(\overline{X}_{N_1}-\theta)^2|X_1,\ldots,X_{N_1}\} \\ &= A^2\ E(N_1^2N^{-2}(\overline{X}_{N_1}-\theta)^2)+n^{*2}E(N^{-1})-n^{*2}E(N_1N^{-2}) \\ &= I + II - III. \end{aligned}$$

Also, consider the following expansion of N^{-2} around with θ a remainder R_{4N} :

$$\begin{aligned} N^{-2}=n^{*-2}&-2(\overline{X}_{N_1}-\theta)A^{-2}g'(\theta)g^{-3}(\theta) \\ &-(\overline{X}_{N_1}-\theta)^2A^{-2}\{g^{-3}(\theta)g''(\theta)-3g^{-4}(\theta)g'^2(\theta)\}+R_{4N}. \end{aligned}$$

For the first term we get,

$$I=\gamma n^{*}-2g^{-3}(\theta)g'(\theta)E\{ N_1^{2}(\overline{X}_{N_1}-\theta)^{3}\}$$
$$-(g^{-3}(\theta)g''(\theta)-3g^{-4}(\theta)g'^{2}(\theta))\ E\{ N_1^{2}(\overline{X}_{N_1}-\theta)^{4}\}+E(R_{5N})$$

Moreover, Lemma (3.4) of Woodroofe (1987) also provides

$$E\{ N_1^{2}(\overline{X}_{N_1}-\theta)^{4}\} = 3\Psi''^{2}(\alpha),\ \text{as}\ m\rightarrow\infty.$$

Hence,

$$I = \gamma n^{*}+(1/2)\{(7/2)\rho^{2}-3\beta+9\}+o(1).$$

Lemma (4.1) of Woodroofe (1987) is used to show that $E(R_{5N})=o(1)$ as $A\rightarrow\infty$.
For the second term we use Theorem 1 with $h(t)=t^{-1}$, $t>0$, to get

$$II=n^{*}-(1/2)\{\gamma+(1/2)\beta-(7/8)\rho^{2}-(3/2)\}\gamma^{-1}+o(1).$$

The third term is obtained through arguments similar to those used to derive I above, as well as another application of Woodroofe's (1987) Lemma (3.4) to prove that $E\{N_1(\overline{X}_{N_1}-\theta)^{2}\}=\Psi''(\alpha)$.

Finally we arrive at

$$III=\gamma n^{*}(1/2)\{3\rho^{2}-\beta+3\}+o(1),$$

which proves (ii) of Lemma 3, and the proof is complete.

Interestingly, (i) of Lemma 3 is the same expression obtained by Martinsek (1988) for estimating the mean of a nonparametric non-lattice distribution by purely sequential sampling under a quadratic loss function similar to (1.2). It is obvious from (i) of Lemma 3 that the information gained by the third stage reduces the bias noticed in (i) of Lemma 1 by proportion γ. We expect $\overline{X}_N$ to be biased downward if F_α has positive skewness and upward if F_α has negative skewness.

The main result of this study is given in the following Theorem 2 to obtain the asymptotic regret of the triple sampling procedure (2.1)-(2.2). As in Chow and Robbins (1965), we define the asymptotic regret of the triple sampling procedure (2.1)-(2.2) as

$$\omega = E(L_N(A)) - E(L_{n^*}(A)) \text{ as } m \to \infty.$$

Theorem 2. Under the quadratic loss function in (1.2), the regret of the procedure (2.1)-(2.2) is given by

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

(The special case of $\gamma \to 1$ in the above regret gives the corresponding pure sequential regret $\omega = 2.25\rho^2 - \beta + 3 + o(1)$, as $m \to \infty$.)

It is obvious that the above asymptotic regret is a non-vanishing quantity independent of m and A. The amount of regret encountered in using a triple sampling procedure instead of the fixed sample counterpart depends on two main sources of information. The first source is due to the nature of the underlying distribution F_a through the coefficients ρ and β. The second source of information is related to the role played by the second stage (the main study) through factor γ.

Proof of Theorem 2. Consider the quadratic error loss function in (1.2) to obtain

$$\omega = A^2 E(\overline{X}_N - \theta)^2 + E(N) - 2n^*,$$

and then apply (ii) of Lemma 3 and (i) of Lemma 2 to arrive at the statement of Theorem 2.

Martinsek (1988) discussed the possibility of sequential procedures resulting in negative regret. Here, we also argue the possibility that three stage schemes may surpass the fixed sample size procedures. Following the result in Theorem 2, one would expect negative regret when $\beta > (1/4)(8\gamma + 1)\rho^2\gamma^{-1} + 3$. Hall (1981) and others recommended the use of $\gamma = 1/2$ for practical applications. In that sense three stage sampling procedures result in negative regret if $\beta > 2.5\rho^2 + 3$. The previous conclusion is similar to that given in Martinsek (1988) for the nonparametric distributions using a completely different approach. On the other hand, had the purely sequential procedure been used we would expect negative regret if $\beta > 2.5\rho^2 + 3$.

2.1 The Asymptotic Normality of the Stopping Rule

Finally, we explore the asymptotic distribution of the stopping rule (2.2). Specifically, we investigate the large sample performance of the expression $(N - n^*)/\sqrt{N^*}$. This can be as a direct application of Theorem 2. Let $h(N) = e^{t(N-n^*)/\sqrt{n^*}}$ to obtain the limiting moment generating function, $E(e^{t(N-n^*)/\sqrt{n^*}}) = 1 + (1/4)\rho^2\gamma^{-1}(t^2/2) + o(A^{-1/2})$ as $m \to \infty$, which is the moment generating function of the normal distribution with mean 0 and variance $(1/4)\rho^2\gamma^{-1}$.

3. Applications

In this section we examine some samples in which the estimation process is subjected to the loss function in (1.2). For each case we apply the main results to study the regret associated with the triple sampling procedure. Other important results are also presented for completeness.

3.1 The normal case

Let $X_1, X_2, \ldots$ be a sequence of independent and identically normally distributed random variables with known mean μ and unknown variance θ. Without loss of generality we can assume the normal mean $\mu = 0$. The main interest is estimation of the normal variance θ, under the loss function in (1.2). The above distribution is a member of the exponential family. Hence, we write

$$dF_a(x) = (2\pi)^{-1/2} e^{-x^2/2\theta - (1/2)\ln(\theta)} dx \text{ , for } -\infty < x < \infty \text{ and } \theta > 0.$$

In view of equation (1.1) with "x" replaced by "x^2" we have,

$$\Psi(a) = (1/2)\,a = -(2\theta)^{-1}$$

which results in $\rho = \sqrt{8}$ and $\beta = 15$. Also, minimization of the loss function in (1.2) provides an optimal sample size $n^* \approx \sqrt{2}A\theta$. If the triple sampling procedure (2.1)-(2.2) is used we propose the sample variance S^2_N to estimate θ . Now, (i) of Lemma 3 gives

$$E(S^2_N) = \theta - \sqrt{2}A^{-1} + o(A^{-1}),$$

which shows that S^2_N is underestimating the unknown variance θ. The amount of bias decreases as A increase. Also, (i) of Lemma 2 yields

$$E(N)=n^{*}+(1/2)-2\gamma^{-1}+o(1) \text{ and } Var(N)=(1/4)\gamma^{-1}n^{*}+o(A).$$

Theorem 2 measures the asymptotic regret of the three-stage procedure (2.1)-(2.2) to estimate θ as $\omega=4+2\gamma^{-1}+o(1)$.
Therefore, the "opportunity cost" of using the procedure (2.1)-(2.2) is equivalent to $4+2\gamma^{-1}$ observation in comparison to the fixed sample size procedure had θ been known. Letting $\gamma\to1$ in the above cost we obtain the regret $\omega=6+o(1)$ of the corresponding purely one-by-one sequential procedure. It is obvious that

$$(N-n^{*})/\sqrt{n^{*}} \approx N(0,2\gamma^{-1}) \text{ as } A\to\infty.$$

3.2 The gamma case

Let $X_1, X_2, \ldots$ be a i.i.d. random variables with the common density

$$dF_{\alpha}(x)=\Gamma^{-1}(\alpha)x^{\alpha-1}e^{-\alpha x/\theta+\alpha\ln(\alpha/\theta)}dx \qquad \text{for } x>0,\ \theta>0,\ \alpha>0,$$

where α is known. The goal is to estimate θ under the loss function in (1.2). In view of (1.1) we conclude that $\Psi(\alpha)=-\alpha\ln(\alpha/\theta)$ and $\alpha=-\alpha/\theta$. Direct computation shows that $\rho=2/\sqrt{\alpha},\ \beta=(6/\alpha)+3$ and $n^{*}=A\sqrt{\alpha}\theta$. If the triple sampling procedure (2.1)-(2.2) is employed to estimate θ by $\overline{X}_N$, we have from (i) of Lemma 3 that $E(\overline{X}_N)=\theta-(A\sqrt{\alpha})^{-1}+o(A^{-1})$, and from (i) of Lemma 2 that $E(N)=n^{*}+(1/2)-2\gamma^{-1}+o(1)$. By Theorem 2 the asymptotic regret turns out to be

$$\omega=2\alpha^{-1}+(\alpha\gamma)^{-1}+o(1).$$

It is evident from the above regret that the two parameters α and γ play a role in determining the cost of using the three stage estimation procedure instead of the fixed sample size procedure. When $0<\alpha<1$ we expect higher opportunity costs than those for $\alpha>1$. For $\alpha=1$ the distribution is exponential for which the opportunity cost. In particular, if we let $\gamma\to1$ in the above expression, we get the corresponding pure sequential asymptotic regret obtained by Woodroofe (1977) for the same problem. For practical use many authors have suggested to consider $\gamma=1/2$.

Finally, we point out that results concerning other exponential family distributions can be derived along the lines of the above examples using our findings in Lemma 2, Lemma 3, and Theorem2.

3.3 Simulations Results

In this section we assume estimating the scale parameter θ of the exponential distribution of the form

$$dF_{\alpha}(x) = e^{-x/\theta - \ln(\theta)},$$

where the result in section 3.2 still valid for $\alpha = 1$ as well.

A series of 5000 simulation runs were generated to provide a feel for small, moderate to large sample size performance of the proposed three stage point estimation sampling procedure (2.1)-(2.2) under squared error loss function. The optimal sample size n^* were allowed to vary from small to large (5, 10, 15, 20, 25, 30, 50, 75, 100, 150, 200, 500, 1000) and we took $\theta = 1$ in all cases. The design factor γ, which represents the proportion of n^* to be estimated in the second stage phase ranged between 0.3, 0.5, 0.7, and 0.8 to study the effect of increasing the factor on the final stage performance. The starting sample size m were permitted to vary from 10 to 15 to study the impact of increasing m on the final estimation of θ, n^* and regret.

Table (I) presents the simulation results regarding the procedure (2.1)-(2.2) for different values of γ, m and n^*. Each row in the table presents an estimate $\overline{N}$ of n^* and its standard error $S.E.(\overline{N})$. The point estimate $\hat{\theta}_N$ of θ and its standard error $S.E.(\hat{\theta}_N)$ and finally the regret $\hat{\omega}$ of ω.

Close inspection of the numerical findings shows perfect agreement with our theoretical developments, presented in theorems 1 and 2 and Lemmas 1, 2 and 3. And the expectation of the final stage sample approaches the optimal sample size n^* as expected, where for small values of n^* the standard error of $\overline{N}$ increases as n^* increases. The triple stage point estimation $\hat{\theta}_N$ approaches 1 in all cases and as noticed, the standard error $S.E.(\hat{\theta}_N)$ decreases as n^* increases. The estimated regret was negative in most cases which indicates that the triple sampling procedure outperform the fixed sample size procedure had θ been known. We recommend the use of $m = 10$ and $\gamma = 0.5$ for practical implementation of the procedure.

Table(I) Three-stage procedure to estimate the scale parameter of the exponential distribution. (number of simulations=5000)

$\gamma=0.3$ $m=10$					
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	10.009	0.002	1.0010	0.3135	0.0088
10	11.465	0.033	0.9922	0.2629	-8.5352
15	15.809	0.064	1.0076	0.2580	-14.1914
20	20.477	0.086	0.9990	0.2362	-19.5230
25	24.958	0.102	0.9997	0.2178	-25.0416
30	29.503	0.117	0.9981	0.2020	-30.4972
50	47.696	0.178	0.9989	0.1535	-52.3044
75	72.368	0.235	1.0012	0.1234	-77.6316
100	97.301	0.273	1.0001	0.1042	-102.6994
150	147.171	0.340	1.0014	0.0847	-152.8292
200	197.049	0.391	1.0005	0.0727	-202.9512
500	497.075	0.605	0.9990	0.0444	-502.9250
1000	996.812	0.848	1.0000	0.0313	-1003.1876
$\gamma=0.3$ $m=15$					
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	15.000	0.000	0.9947	0.2547	5.0000
10	15.082	0.007	1.0026	0.2513	-4.9182
15	16.780	0.038	1.0007	0.2226	-13.2202
20	20.811	0.068	1.0028	0.2214	-19.1886
25	25.395	0.089	0.9948	0.2048	-24.6054
30	30.535	0.109	0.9983	0.1935	-29.4646
50	49.162	0.157	1.0004	0.1501	-50.8382
75	72.723	0.220	1.0013	0.1213	-77.2774
100	97.148	0.269	0.9990	0.1050	-102.8518
150	147.143	0.327	1.0004	0.0832	-152.8568
200	197.243	0.386	1.0007	0.0744	-202.7570
500	497.486	0.608	0.9996	0.0456	-502.5136
1000	995.186	0.856	0.9991	0.0319	-1004.8144
$\gamma=0.5$ $m=10$					
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	10.007	0.002	1.0015	0.3102	0.0074
10	11.516	0.032	1.0089	0.2721	-8.4840
15	15.410	0.056	1.0022	0.2585	-14.5902
20	19.652	0.073	0.9982	0.2397	-20.3476
25	24.282	0.092	1.0031	0.2133	-25.7182
30	28.597	0.107	0.9976	0.1974	-31.4034
50	48.433	0.149	0.9997	0.1546	-51.5672
75	73.139	0.184	0.9988	0.1208	-76.8610
100	97.906	0.214	0.9982	0.1040	-102.0940
150	148.202	0.259	1.0013	0.0844	-151.7978
200	198.003	0.302	0.9989	0.0714	-201.9970
500	497.591	0.475	0.9993	0.0459	-502.4090
1000	997.771	0.683	0.9993	0.0316	-1002.2290

Table(I) - continue

		$\gamma=0.5$	$m=15$		
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	15.000	0.000	0.9926	0.2562	5.0000
10	15.063	0.005	0.9986	0.2463	-4.9366
15	16.774	0.037	0.9981	0.2243	-13.2256
20	20.758	0.066	1.0002	0.2195	-19.2420
25	25.153	0.082	0.9983	0.2083	-24.8470
30	29.681	0.092	1.0011	0.1889	-30.3190
50	48.634	0.145	1.0016	0.1521	-51.3664
75	73.459	0.178	1.0005	0.1204	-76.5408
100	98.394	0.207	1.0002	0.1040	-101.6060
150	148.829	0.256	1.0019	0.0840	-151.1712
200	198.238	0.293	0.9998	0.0723	-201.7620
500	497.944	0.470	0.9994	0.0449	-502.0562
1000	998.951	0.661	1.0000	0.0317	-1001.0490
		$\gamma=0.7$	$m=10$		
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	10.008	0.002	1.0055	0.3161	0.0080
10	11.374	0.028	1.0010	0.2672	-8.6258
15	15.043	0.053	1.0027	0.2576	-14.9566
20	19.465	0.072	0.9987	0.2366	-20.5350
25	24.228	0.087	0.9968	0.2164	-25.7716
30	29.328	0.098	1.0033	0.1982	-30.6724
50	49.460	0.129	0.9987	0.1472	-50.5404
75	74.675	0.168	0.9976	0.1203	-75.3250
100	100.106	0.191	0.9992	0.1020	-99.8938
150	150.899	0.242	1.0015	0.0822	-149.1014
200	201.305	0.298	0.9999	0.0709	-198.6954
500	505.172	0.552	1.0005	0.0441	-494.8284
1000	1010.733	0.981	0.9993	0.0322	-989.2668
		$\gamma=0.7$	$m=15$		
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	15.000	0.000	1.0011	0.2564	5.0000
10	15.069	0.006	0.9961	0.2485	-4.9306
15	16.687	0.034	0.9987	0.2252	-13.3132
20	20.342	0.060	0.9989	0.2249	-19.6578
25	24.683	0.076	1.0011	0.2082	-25.3172
30	29.303	0.091	1.0020	0.1948	-30.6972
50	49.253	0.124	1.0019	0.1490	-50.7468
75	74.184	0.156	0.9979	0.1200	-75.8160
100	99.520	0.180	1.0013	0.1015	-100.4796
150	149.828	0.222	1.0009	0.0828	-150.1718
200	200.294	0.264	1.0012	0.0700	-199.7056
500	502.640	0.469	1.0015	0.0454	-497.3602
1000	1005.080	0.747	1.0000	0.0322	-994.9196

Table(I) - continue

$\gamma=0.8$ $m=10$					
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	10.007	0.002	1.0008	0.3145	0.0074
10	11.313	0.027	1.0009	0.2723	-8.6868
15	15.106	0.054	1.0057	0.2537	-14.8942
20	19.704	0.072	0.9953	0.2312	-20.2958
25	24.871	0.087	1.0000	0.2135	-25.1292
30	29.860	0.099	0.9986	0.1991	-30.1400
50	50.664	0.140	1.0008	0.1490	-49.3364
75	76.305	0.179	0.9977	0.1186	-73.6952
100	102.401	0.213	1.0000	0.1002	-97.5992
150	154.481	0.303	0.9998	0.0809	-145.5188
200	206.369	0.373	1.0016	0.0712	-193.6312
500	518.088	0.854	1.0003	0.0437	-481.9116
1000	1034.826	1.551	1.0002	0.0312	-965.1738
$\gamma=0.8$ $m=15$					
N	$\overline{N}$	$S.E.(\overline{N})$	$\hat{\theta}_N$	$S.E.(\hat{\theta}_N)$	$\hat{\omega}$
5	15.000	0.000	1.0019	0.2583	5.0000
10	15.065	0.006	0.9969	0.2459	-4.9346
15	16.564	0.032	0.9936	0.2251	-13.4364
20	20.169	0.058	1.0009	0.2189	-19.8310
25	24.760	0.077	1.0049	0.2103	-25.2396
30	29.601	0.090	1.0014	0.1916	-30.3986
50	49.915	0.125	1.0004	0.1481	-50.0852
75	75.596	0.159	1.0006	0.1188	-74.4044
100	100.650	0.188	0.9990	0.1013	-99.3500
150	152.209	0.245	0.9989	0.0810	-147.7906
200	203.225	0.308	0.9983	0.0701	-196.7746
500	509.750	0.601	1.0004	0.0440	-490.2498
1000	1022.431	1.134	0.9998	0.0311	-977.5692

References

[1] Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc* Vol. 45, 600-607.

[2] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* Vol. 36, 1203-1212.

[3] Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of mean. *Ann. Statist.* Vol. 9, 1229-1238.

[4] Hamdy, H. I. (1988). Remarks on the asymptotic theory of triple stage estimation of the normal mean. *Scand. J. of Stat.* Vol. 15, 303-310.

[5] Hamdy, H. I. (1995). Performance of fixed width intervals under type II error: the exponential case. *South African Statist. J.* Vol. 31, 1-11.

[6] Hamdy, H. I, AlMahmeed, M. and Al-Zalzalah, Y. (1995). A certain accelerated procedure to construct simultaneous confidence region: the exponential case. *Statistics* Vol. 28, 57-71.

[7] Hamdy, H. I. Mukhopadhyay, N., Costanza, M. C. and Son, (1988). Triple stage point estimation for the exponential location parameter. *Ann. Inst. Statis. Math.* Vol. 40, 785-797.

[8] Hamdy, H. I, Son, M. S. (1990) On accelerating sequential procedures for point estimation: the normal case. *Statistical LI,* 437-446.

[9] Hamdy, H. I, M. S. AlMahmeed, M. and Nigm, A. (1989) Three stage estimation procedure for the exponential location parameters. *Metron XLVII,* 279-294.

[10] Lehmann, E. L. (1986). *Testing Statistical Hypotheses*; 2ed. John Wily & Sons, New York.

[11] Martinsek, A. T. (1988). Negative regret, optimal stopping, and the elimination of outliers. *J. Am. Stat. Assoc.* Vol. 83, 160-163.

[12] Morris, C. N. (1982). Natural exponential families with quadratic variance function. *Ann Statist.* Vol. 10, 65-80.

[13] Mukhopadhyay, N., Hamdy, H. I. AlMahmeed, M., Costanza, M. C. (1987). Three-stage point estimation procedure for a normal mean. *Sequential Analysis* Vol. 6, 21-36.

[14] Woodroofe, M. (1977). Second order approximations for sequential point and interval estimation. *Ann Statist.* Vol. 5, 984-995.

[15] Woodroofe, M. (1987). Asymptotically optimal sequential point estimation in three stages. In *New Perspectives in Theoretical and Applied Statistics.* John Wiley and Sons, Inc. New York. 397-411.