ACCELERATED SEQUENTIAL PROCEDURE TO ESTIMATE THE MEAN OF UNKNOWN DISTRIBUTION

H. I. Hamdy *M. S. Son*
University of Vermont

Keywords: *Sample size, Purely Sequential Estimation, Minimum Risk Point Estimation, Fixed Width Confidence Intervals, Fixed size Confidence Regions, Exponential Distribution, Intra-Class Model, Regression Model.*

Summary. Consider the accelerated sequential procedure of Hall (1983). Second order asymptotic expressions are obtained for the expectation of well behaved functions of the stopping variable. The wide applicability of the results is demonstrated by working out several point and interval estimation problems.

1. Introduction. Consider the problem of estimating some unknown parameter $\mu \varepsilon R$ in the presence of a nuisance parameter $\theta > 0$ using sequential procedures. Compared with other procedures, the purely sequential procedure due to Anscombe (1953), Robbins (1959) and Chow and Robbins (1965) has greater efficiency. Yet, it is rather slow and can be costly to perform, e.g., see Hall (1983) who proposed an accelerated sequential procedure to reduce, by an arbitrary predetermined factor, the number of sampling operations needed to construct a fixed width confidence interval for the mean of a normal population with unknown variance. Hall (1983) obtained asymptotic expressions for the mean and variance of the stopping variable.

An accelerated sequential procedure, which can be used for a wider class of populations, is developed in the sequel along the lines of Hall (1983), and Hamdy and Son (1991), and Son and Hamdy (1990). A second order asymptotic expression is then obtained for the expectation of an arbitrary, but well behaved (see Assumption 1 below), function h of the stopping variable. Such an expression can be useful in both point and interval estimation situations as illustrated in section 4. Further, as a by-product, a similar expression is obtained for the purely sequential procedure.

Our approach takes advantage of the fact that the optimal sample sizes arising in most sequential estimation problems have the following structure

$n^* = \lambda g(\theta)$, (1)

where n^* is the optimal sample size, $\lambda > 0$ is a known constant and $g > 0$ is a given function of the unknown nuisance parameter θ. Obviously $n^* \to \infty$ *as* $\lambda \to \infty$. However, we make the following assumptions.

Assumption 1. Let h be continously differentiable function such that

$\underset{n \geq m}{Sup}\, h'''(n) = O(h'''(\lambda))$ *as* $m \to \infty$.

Assumption 2. Let g a continously differentiable bijective function such that as $m \to \infty$

$\underset{n \geq m\lambda}{Sup} \left| \dfrac{d^3(nf(n))}{dn^3} \right| < \infty$, where $f = g^{-1}$

Let $\hat{\mu}_n$ and $\hat{\theta}_n$ be the usual estimators of μ and θ, respectively, based on a random sample of fixed size $n \geq 2$.

Assumption 3. There exists a sequence W_1, W_2, ... of positive i.i.d. random variables each having mean θ and variance τ^2 such that $E(W_1{}^4) < \infty$, where

$$\hat{\theta}_n = \overline{W}_n = S_n / n \text{ , and } S_n = \sum_{i=1}^{n} W_i \text{ , } n \geq 2.$$

Assumption 1 is made to facilitate the calculation of the order of the remainder terms in the Taylor series expansions in sequel, while Assumption 2 enables us to explicitly define the inverse function of g and evaluate the order of the remainder term in eqn (7) in the appendix. Asumption 3 is made following Woodroofe (1977). Nevertheless, it is shown in section 4 that these assumptions hold true in many applications.

Regarding notation, we use f', f'', f''' to denote the first, second, and third derivatives, respectively, of a given function f.

2. A Purely Sequential Procedur. Since n^*

depends on the unknown nuisance parameter θ, we resort to the following purely sequential procedure to estimate n^*.

Assume that an initial sample of size $m \geq 2$ has been taken from the underlying distribution. Then, we define the following stopping rule

$$M = inf \ \{n > m : n \geq \lambda g\ (\overline{W}_n)\}. \quad (2)$$

Once M is determined, we proceed to estimate the unknown parameter μ.

The main result concerning the one-by-one sequential procedure (2) is given in Theorem 1 whose proof is based on the following lemma.

Lemma 1. Let M be defined by (2). Then, under Assumptions 2-3, we have

$$(i)\ E(M - n^*)^2 = \frac{\tau^2}{n^*}\left(\frac{dn^*}{d\theta}\right)^2 + o(\lambda).$$

$$(ii) E\ |M - n^*|^3 = o(\lambda^2)$$

$$(iii) E(M) = n^* - \frac{\tau^2}{2n^*}\left\{\frac{2}{n^*}\left(\frac{dn^*}{d\theta}\right)^2 - \frac{d^2n^*}{d\theta^2} - \frac{2\nu}{\tau^2}\left(\frac{dn^*}{d\theta}\right)\right\} + o(1).$$

where ν is given by Woodroofe (1977), see also Finster (1983).

Theorem 1. Let M be defined by (2). Then, under Assumptions 1-3, we have

$$E\{h(M)\} = h(n^*) - \varsigma h'(n^*) + o(h'(\lambda)),$$

where

$$\varsigma = \frac{\tau^2}{2n^*}\left\{\left(\frac{2}{n^*} - \frac{h''(n^*)}{h'(n^*)}\right)\left(\frac{dn^*}{d\theta}\right)^2 - \frac{d^2n^*}{d\theta^2} - \frac{2\nu}{\tau^2}\left(\frac{dn^*}{d\theta}\right)\right\}.$$

Both Lemma 1 and Theorem 1 are proved in the appendix.

Remark 1. In the above expression, ς represents the cost of not knowing the nuisance parameter θ under the purely sequential set-up, see Simons (1968) for more details.

3. An Accelerated Sequential Procedure. Hall (1983) suggested accelerating purely sequential procedures by a predetermined factor $\gamma \in (0, 1)$. We adopt his idea and outline the procedure in three phases. In the pilot phase an initial sample of size $m \geq 2$ is taken to start the sequential phase, where observations are taken one-by-one, to estimate only a fraction γ of n^* according to the following rule

$$N_1 = inf\ \{n \geq m : n \geq \gamma \lambda g\ (\overline{W}_n)\}. \quad (3)$$

The final sample size is then define by

$$N = \max\ \{N\ ,\ [\lambda g\ (\overline{W}_{N_1})] + 1\}, \quad (4)$$

where [x] denotes the largest integer less than or equal to x. In the accelerated phase, the remaining observations $(N - N_1)$ are taken in one last bulk and we then proceed to estimate μ using the N observations.

The following lemma is an immediate consequence of Theorem 1.

Lemma 2. Let N_1 be defined by (3). Then, under Assumptions 1- 3, we have as $\lambda \to \infty$

$$E\ \{h(\gamma^{-1} N_1)\} = h(n^*) - \gamma^{-1} \varsigma h'(n^*) + o(h'(\lambda)),$$

The main result concerning the accelerated sequential procedure defined by (3) and (4) is given in Theorem 2 which is proved in the appendix.

Theorem 2. Let N be defined by (4). Then, under Assumptions 1-3, we have as $\lambda \to \infty$

$$E\ \{h(N)\ \} = h(n^*) - \eta h'(n^*) + o(h'(\lambda)),$$

where

$$\eta = \frac{\tau^2}{2\gamma n^*}\left\{\left(\frac{2}{n^*} - \frac{h''(n^*)}{h'(n^*)}\right)^2 - \frac{d^2 n^*}{d\theta^2}\right\} - \frac{1}{2}.$$

Remark 2. As in Remark 1, the above expression for η represents the cost of not knowing the nuisance parameter θ if we accelerate the purely sequential procedure by a predetermined coefficient $\gamma \in (0, 1)$.

4 Appliciations. It is shown here, by way of examples,

how to use theorem 2 to solve several point and interval estimation problems. For point estimation, the optimal sample size is that value of n which minimizes the risk $R_n\ (A) = E\ \{L_n\ (A)\}$, where $L_n(A)$ is the following squared error loss plus a linear cost function

$L_n(A) = A(\hat{\mu}_n - \mu)^2 + n,$

and $A > 0$ is a known constant.

As for interval estimation, given $\alpha \epsilon (0,1)$, the optimal sample size is that value of n for which $Pr\{\mu \epsilon I_n\} \geq 1 - \alpha$, where I_n is a confidence interval of fixed width $d > 0$.

In each case, we begin by finding the optimal sample size n^*. Since n^* depends on the unknown nuisance paramete θ, we resort to the accelerated sequential procedure defined by (3) and (4) to estimate n^*. Once N is determined, we propose $\hat{\mu}_N$ as a point estimator for μ and compute the accelerated sequential risk, $E\{L_N(A)\}$, the optimal risk, $R_n^*(A)$, and the regret, $\omega(A) = E\{L_N(A)\} - R_n^*(A)$.

On the other hand, we propose I_N as a confidence interval for μ and compute $Pr\{\mu \in I_N\}$.

Example 1. An Exponential Model.

Let $Y_1, Y_2, ...$ be a sequence of i.i.d. random variables having the following exponential distribution

$f(y: \mu, \theta) = (1/\theta)e^{-(y-\mu)/\theta},\ y > \mu,$

where the location parameter $\mu \in R$ and the scale parameter $\theta > 0$ are assumed unknown.

For a random sample of fixed size $n (\geq 2)$, the usual estimators of μ and θ are given by

$$\hat{\mu}_n = \min_{1 \leq i \leq n} Y_i, \qquad \hat{\theta}_n = (n-1)^{-1} \sum_{i=1}^{n} (Y_i - \hat{\mu}_n),$$

respectively. Now, let $W_1, \ldots, W_{n-1}$ be i.i.d. random variables from $f(\cdot : 0, \theta)$. Then, it follows from Lemma 6 in Lombard and Swanepoel (1978) that the distribution of $\{\overline{W}_n, n \geq 2\}$ is identical to that of $\{\hat{\theta}_n, n \geq 2\}$, which is equivalent to Assumption 3. Further, it is easily verified that $\tau^2 = \theta^2$.

Point estimation of μ. Since the risk is given by $R_n(A) = 2A\theta^2 / n^2 + n.$

The optimal sample size takes the form in (1) with $\lambda = (4A)^{1/3}$ and $g(\theta) = \theta^{2/3}$. Now, the accelerated sequential risk is given by

$$E\{L_N(A)\} = (1/2)(n^*)^3 E(N^{-2}) + E(N).$$

Hence, using Theorem 2 with h (N) = N and $h(N) = N^{-2}$, we obtain

$$E(N) = n^* - \left(\frac{5}{9}\lambda^{-1} - 1/2\right) + o(1),$$

and

$$E(N^{-2}) = (n^*)^{-2} + 2(n^*)^{-3}\left(\frac{11}{9\gamma} - 1/2\right) + o(A^{-1});$$

$$R_n^*(A) = \frac{3}{2}n^*;$$

respectively, so that

$$E\{L_N(A)\} = \frac{3}{2}n^* + \frac{2}{3}\gamma^{-1} + o(1).$$

Therefore, the regret is given by $\omega(A) = \frac{2}{3}\gamma^{-1} + o(1)$, which is bounded and independent of A.

Interval estimation of μ. Consider a fixed width confidence interval for μ of the form $I_n = (\hat{\mu}_n - d, \hat{\mu}_n)$. The associated confidence coefficient is given by

$Pr\{\mu \in I_N\} = 1 - e^{(-nd/\theta)}.$

Let $a = -\ln(\alpha)$, since we require the confidence coefficient to be at least $(1-\alpha)$ then the optimal sample size would be as in (1) with $\lambda = a/d$ and $g(\theta) = \theta$. Further, since

$Pr\{\mu \in I_N\} = 1 - E\{e^{(-Nd/\theta)}\},$

then, using Theorem 2 with $h(N) = e^{(-Nd/\theta)}$, we obtain

$$Pr\{\mu \in I_N\} = (l - \alpha) - \frac{a\alpha}{n^*}\left(\frac{a - \lambda + 2}{2\lambda}\right) + o(d),$$

where $\eta = (a - \gamma + 2)/(2\gamma)$ represents the cost of not knowing the nuisance parameter θ.

Example 2. An Intra-Class Model.

Let $Y_1, Y_2 ....$ be a sequence of i.i.d. random variables such that $Y_i = \mu + \epsilon_i$, where the ϵ_i are normally distributed with $E(\epsilon_i) = 0$ and

$$cov(\epsilon_i, \epsilon_j) = \sigma^2, \quad i = j,$$
$$= \rho\sigma^2, \quad i \neq j.$$

The parameters $\mu \in R$, $\sigma^2 > 0$ and $\rho \in (-1, 0)$ are assumed unknown.

For a random sample of fixed size $n (\geq 2)$, the usual estimator of μ is

$\hat{\mu}_n = n^{-1} \sum_{i=1}^{n} Y_i$.

Further, it can be shown that $\hat{\mu}_n \sim N(\mu, \theta/n + \rho\sigma^2)$, where $\theta=\sigma^2(1-\rho)$. On the other hand, the usual estimator of θ is given by

$$\hat{\theta}_n = (n-1)^{-1} \sum_{i=1}^{n} (Y_i - \hat{\mu}_n)^2.$$

Consider the following Helmert's transformation

$$Z_i = \frac{1}{\sqrt{i(i+1)}} \left(\sum_{j=1}^{i} Y_j - iY_{i+1} \right), \quad i = 1, \cdots n-1.$$

Define $W_i = Z_i^2$. Then, $W_i \sim \theta\chi^2_{(1)}$. It is easily verified that Assumption 3 hold true with $\tau^2 = 2\theta^2$.

Point estimation of μ. The risk here is given by
$R_n(A) = A\theta/n + A\rho\sigma^2 + n$.

Thus, the optimal sample size is of the form in (1) with $\lambda = A^{1/2}$ and $g(\theta) = \theta^{1/2}$. Further, since the accelerated sequential risk is given by
$E\{L_N(A)\} = (n^*)^2 E(N^{-1}) + A\rho\sigma^2 + E(N)$,
then, using Theorem 2 with $h(N)=N$ and $h(N)=N^{-1}$, we obtain

$$E(N) = n^* - \left(3/4\gamma^{-1} - 1/2 \right) + o(1),$$

and

$$E(N^{-1}) = (n^*)^{-1} + (n^*)^{-2} \left(\gamma^{-1} - 1/2 \right) + o(A^{-1});$$

$R_n^*(A) = 2n^* + A\rho\sigma^2,$
respectively, so that
$E\{L_N(A)\} = 2n^* + A\rho\sigma^2 + 1/2\gamma^{-1} + o(1)$.

Hence, the regret is given by $\omega(A) = (1/2)\gamma^{-1} + o(1)$.

Interval estimation of μ. Consider a fixed width confidence interval for μ of the form $I_n = (\hat{\mu}_n - d, \hat{\mu}_n + d)$. Let $\Phi(\cdot)$ denote the standard normal distribution function and set $a = \Phi^{-1}(1-\alpha/2)$. Thus, the confidence coefficient is given by

$$Pr\{\mu \in I_n\} = 2\Phi(d/\sqrt{\theta/n + \rho\sigma^2}) - 1 \geq 2\Phi(d/\sqrt{\theta/n}) - 1.$$

Since the confidence coefficient is required to be at least $(1-\alpha)$, then the optimal sample size would be as in (1) with $\lambda = a^2/d^2$ and $g(\theta) = \theta$. Now, it can be shown that

$$Pr\{\mu \in I_N\} = 2E\{\Phi(d/\sqrt{\theta/N + \rho\sigma^2})\} - 1 \geq 2E\{\Phi(d/\sqrt{\theta/N})\} - 1.$$

Hence, using Theorem 2 with $h(N) = \Phi(d/\sqrt{\theta/N})$. We obtain

$$Pr\{\mu \in I_N\} \geq (1-\alpha) - \frac{a\phi(a)}{n^*} \left(\frac{a^2 - \gamma + 5}{2\gamma} \right) + o(d^2).$$

Here the cost of not knowing the nuisance parameter θ is given by
$\eta = (a^2 - \gamma + 5)/(2\gamma)$.

Remark 3. The results of Theorems (1) and (2) still hold true for $\mu \in R^k, k \geq 2$. However, the initial sample size should now be $m \geq k+1$.

Example 3. A Fixed Size Confidence Region For The Regression Parameters.

Consider the model $Y_n = X_n\beta + \varepsilon_n$, where Y_n is an observed $n\times1$ vector, X_n is a known $n\times p$ matrix of rank p, β is a $p\times1$ vector of unknown regression parameters and ε_n is an $n\times1$ random vector distributed as $N_n(0, \theta I)$, where I is the $n\times n$ identity matrix and $\theta>0$ is unknown. We assume that $n > p \geq 2$.

Let G^T denote the transpose of a matrix G. Then, the usual estimators of β and θ are given by

$$\hat{B}_n = (X_n^T X_n)^{-1} X_n^T Y_n,$$

and

$$\hat{\theta}_n = (n-p)^{-1} (Y_n - X_n\hat{B}_n)^T (Y_n - X_n\hat{B}_n),$$

respectively. Further, let $Z_1, \ldots, Z_{n-p}$ be an orthogonal basis for the error space, i.e., the null space of X_n^T. It is easily verified that Assumption 3 holds true with $W_i = Z_i^2 \sim \theta\chi^2_{(1)}$ and $\tau^2 = 2\theta^2$.

In order to study the large sample properties of $\hat{B}_n$, it is usually assumed that the matrix $(n^{-1}X_n^T X_n)$ converges to a positive definite matrix as $n \to \infty$. Thus, we use $(n^{-1}X_n^T X_n)$ as the weight matrix in the following fixed size ellipsoidal confidence region for β

$$B_n = \{b \in R^p : (\hat{B}_n - b)^T (n^{-1}X_n^T X_n)(\hat{B}_n - b) \leq d^2\}.$$

The confidence coefficient associated with B_n is given by

$$Pr\{\beta \in B_n\} = F\,(n\,d^2/\theta),$$

where F(.) is the distribution function of a chi-square random variable with p degrees of freedom. Let a_p be such that $F(a_p)=1\text{-}\alpha$. Then, since we require the confidence coefficient to be at least $(1\text{-}\alpha)$, the optimal sample size would be as in (1) with $\lambda = a_p/d^2$ and $g(\theta)=\theta$.

We start with the sample size $m \geq p+1$ and use the accelerated sequential procedure defined by (3) and (4) to determine N. Then, we compute $\bar{B}_N$ and propose the confidence region B_N for β. It can be shown that

$$Pr\{\beta \in B_N\} = E\,\{F(N\,d^2/\theta)\}.$$

Hence, using Theorem 2 with $h\,(N) = F\,(Nd^2/\theta)$, we obtain

$$Pr\{\beta \in B_N\} = (1\text{-}\alpha) - \frac{a_p F'(a_p)}{n^*}\left(\frac{a_p - \gamma + 6 - p}{2\gamma}\right) + o\,(d^2).$$

The cost of not knowing the nuisance parameter θ in this case is given by $\eta = (a_p - \gamma + 6 - p)\,/\,(2\,\gamma)$.

Appendix

Proof of Lemma 1. Consider (2) and expand $\overline{W}_M$ around θ in a Taylor series and carry out some algebraic manipulations to obtain

$$E(M\text{-}n^*)^2 = \lambda^2\, g'(\theta)^2\, E(\overline{W}_M - \theta)^2 + E(R_1),$$

where R_1 is the remainder term. It is easily verified that $E(R_1)$ is of order $o(\lambda)$. Further, using the asymptotic normality of $\overline{W}_n$, $n \geq 2$, the results of Anscombe (1952) and the uniform integrability of $(\overline{W}_M - \theta)^2$, we obtain

$$E(\overline{W}_M - \theta)^2 = \tau^2/n^* + o(\lambda^{-1}),$$

and (i) is established. To prove (ii) let f denote the inverse function of g. Then, recalling (2), we can write

$$S_M = M\,f\,(M/\lambda) - D_M, \tag{5}$$

where D_M represents the excess under the boundary at the stopping time. Using the results of Woodroofe (1977) we have that $E(D_M) = \nu$ while Wald's lemma yields $E\,(S_M) = \theta\, E\,(M)$. Thus, taking the expectation of both sides of (5), we get

$$\theta\, E(M) = E\{M\,f\,(M/\lambda)\} - \nu. \tag{6}$$

Let R_2 denote the remainder term in a Taylor series second order expansion of $\{Mf(M/\lambda)\}$ around n^* and take the expectation to get

$$E\{Mf(M/\lambda)\}_2 = \theta E(M) + g(\theta) f'(g(\theta)) E(M\text{-}n^*) + \frac{\tau^2}{2\lambda n^*}\left(\frac{dn^*}{d\theta}\right)\{g(\theta) f''(g(\theta)) + 2f'(g(\theta))\} + E(R_2) \tag{7}$$

where use has been made of (i), and (ii) is established upon substituting (7) into (6), rearranging terms and observing that

$$R_2 = \frac{1}{3!}\lambda^{-2}(M\text{-}n^*)^3((\nu/\lambda f'''(\nu/\lambda) + 3f''(\nu/\lambda))$$

where v is a random variable lies between M and n^*. It follows that,

$$E\,(R_2) = \frac{1}{3!}\lambda^{-2}\, E\,|M\text{-}n^*|\, Sup_{n \geq m\lambda}\, 3\left(nf'''(n) + 3f''(n/\lambda)\right)$$

$$= \frac{1}{3!}\,\lambda^{-2}\, E\,|M - n^*|^3 \,\underset{n \geq m\lambda}{Sup}\, \frac{d^3\,(n\,f\,(n))}{dn^3}$$

= o(1) by (iii) of Lemma 1 and assumption 2.

Proof of Theorem 1. Let R_3 denote the remainder term in a Taylor series second order expansion of h(M) around n^*. It can be shown that $E(R_3)$ is of order $o(h'(\lambda))$. Thus, taking the expectation of the above mentioned expansion and using Lemma 1, and assumption 1 establishes the theorem.

Proof of Theorem 2. Recalling (3) and (4) we can write

$$N_1 = Y\lambda g(\overline{W}_{N_I}) + V_{N_I}, \tag{8}$$

where V_{N_I} is a random variable representing the excess over the boundary after the termination of the sequential phase, and

$$N = \lambda\, g(\overline{W}_{N_I}) + U_{N_I}, \tag{9}$$

where U_{N_I} is asymptotically uniform over (0,1) as in Hall (1983). Substituting (8) into (9) we obtain

$$N = Y^{-1} N_1 \text{-} Y^{-1} V_{N_I} + U_{N_I}$$

Hence, using Taylor's Theorem, we get

$$h(N) = h(Y^{-1} N_1) + (U_{N_I} \text{-} Y^{-1} V_{N_I}) h'(Y^{-1} N_1) + R_4,$$

where R_4 is the remainder term. Now, similar arguments to those used to evaluate $E(R_2)$ can be used to show that $E\,(R_5)$ is of order $o\,(h(\lambda))$. Further, it follows from Theorem 2.1 in Woodroofe (1977) that N_1 and V_{N_I} are asymptotically independent. Furthermore, going along the lines of Hamdy (1988), it can be shown that N_1 and U_{N_I} are asymptotically uncorrelated. Moreover,

$$E\{h'(Y^{-1} N_1)\} = h'(n^*) + o\,(h''(\lambda)).$$

Hence, using Lemma 2, we get
$E(h(N))=h(n^*)-\{Y^{-1}\{\varsigma+E(V_N)\}-0.5\}h'(n^*)$
$+o(h'(\lambda))$.
The theorem is established upon proving that

$$E(V_{N_I}) = \frac{\nu}{n^*}\left(\frac{dn^*}{d\theta}\right) + o(1). \qquad (10)$$

To this end recall (8) and write
$S_{N_I} = N_1 f''(N_1 - V_{N_I})/\lambda\gamma)$. (11)
But, similar to (5), we have
$S_{N_I} = N_1 f(N_1/\lambda\gamma) - D_{N_I}$. (12)
Therefore, from (11) and (12), we obtain
$D_{N_I}=N_1\{f(N_1/\lambda\gamma)-f(N_1/\lambda\gamma-V_{N_I}/\lambda\gamma)\}$,
by Taylor Theorem.
$= (N_1/\lambda\gamma)f'(N_1/\lambda\gamma)V_{N_I} + R_5$,
where R_5 is the remainder term. But,
$E\{N_1 f'(N_1/\lambda\gamma)\} = \gamma n^* f'(n^*/\lambda) + o(\lambda)$.
Hence, upon observing that $E(R_5)$ is of order o(1), we get
$\nu = (n^*/\lambda)f'(n^*/\lambda)E(V_{N_I}) + o(1)$,
observe also that $f\,g(\theta) = \theta$ taking the derivative of both sides with respect to θ we get
$f'\,g(\theta)g'(\theta)=1$.
Therefore,
$f'\,g(\theta) = 1/g'(\theta)$,
from which (10) follows. This completes the proof.

References

Anscombe, F.J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Phil. Soc., 48, 600-607.*

(1953). Sequential estimation. *JRSS, 15, 1-21.*

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.*

Finster, M. (1983). A frequentistic approach to sequential estimation in the general linear model. *JASA, 78, 403-407.*

Hall, P. (1983). Sequential estimation saving sampling operations. *JRSS, B 45, 219-223.*

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 Son, M. (1991). On accelerating sequential procedures for point estimation: the normal case. Statistica, 437-446.

Lombard, F. and Swanepoel, J. W. H. (1978). On finite and infinite confidence sequences. *S. Afr. Statist. J., 12, 1-24.*

Robbins, H. (1959). Sequential estimation of the mean of a normal population. *In probability and Statistics (Harold Cramer Volume), Almquist and Wiksell, Upsala.*

Simons, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist., 39, 1946- 1952.*

Son, M. and Hamdy, H. I. (1990). On accelerating sequential procedures for estimating exponential distributions. *Commun. Statist. Theor. Meth., 9, 3315-3324.*

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