The Annals of Statistics
1981, Vol. 9, No. 6, 1229–1238

ASYMPTOTIC THEORY OF TRIPLE SAMPLING FOR SEQUENTIAL ESTIMATION OF A MEAN

BY PETER HALL

Australian National University

We describe the asymptotic theory of triple sampling as it pertains to the estimation of a mean. We obtain limit theorems for the case of the normal distribution. Our results show that triple sampling combines the simplicity of Stein's double sampling technique with the efficiency of the fully sequential Anscombe-Chow-Robbins procedure.

1. Introduction. Stein's (1945) well known double sampling procedure has an advantage over some other methods of sequential sampling in that it requires only two sampling operations and can be used to obtain a fixed width confidence interval with a precisely known coverage probability for a normal mean. However, the inexact procedure of Anscombe (1953) and Chow and Robbins (1965) is more efficient than Stein's since it uses a significantly smaller sample size to achieve a confidence interval with very nearly the same coverage. If a third stage is appended to Stein's method it loses its exactness but becomes strongly competitive with Chow and Robbins' procedure from the point of view of efficiency. As the name "triple sampling" implies, the technique involves only three sampling operations. In many real situations significant savings in time and money may be achieved by sampling in bulk, and in these circumstances a triple sampling procedure is more attractive than Anscombe, Chow and Robbins' "one-by-one" sampling. Our aim in this paper is to give a rigorous account of the large sample properties of triple sampling.

Let $\mathscr{P}$ be a population with mean μ and finite variance σ^2, both unknown. Suppose we wish to find a two-sided confidence interval for μ of width 2d and coverage probability very nearly $1 - \alpha$, and we have available a "pilot" sample of size m. Calculate its variance σ_m^2 and let

$$M = \max\{[(\eta_{m-1}\sigma_m/d)^2] + 1, m\},$$

where $[x]$ denotes the integer part of x and η_{m-1} is the $\alpha/2$ critical point of Student's t distribution with $m - 1$ degrees of freedom. Draw a second sample of size $M - m$, pool it with the first, and construct the mean $\bar{X}_M$ of the pooled sample. Stein (1945) suggested that we employ a confidence interval of width $2d$ centered on $\bar{X}_M$. As $d \to 0$ the coverage probability approaches $1 - \alpha$ (Anscombe, 1952), and in the case where $\mathscr{P}$ is normal, the coverage probability will not be less than $1 - \alpha$.

This procedure works very well if the pilot sample size happens to be chosen close to the "optimal" sample size, n_0, which would have been used had the population been normal with known variance σ^2. However, if $m \ll n_0$ then Stein's method is likely to lead to significant oversampling, with the result that if $m/n_0 \to 0$ as $d \to 0$ then $E(M) - n_0 \to +\infty$ (Cox, 1952). To overcome this difficulty we propose the following procedure, which is considerably more robust against the possibility of m being chosen too small.

Fix c in the range $0 < c < 1$ and let the second sample be of size $M_1 - m$, where

$$M_1 = \max\{[c(\eta\sigma_m/d)^2] + 1, m\}$$

Received May 27, 1980; revised February 1981.

AMS 1970 *subject classification*. Primary 62L12, 62E20; secondary 62F10.

Key words and phrases. Confidence interval, efficiency, normal distribution, sequential methods, triple sampling.

and η is the $\alpha/2$ critical point of the $N(0, 1)$ distribution. Calculate the variance $\sigma^2_{M_1}$ for the pooled sample of size M_1, and let the third sample be of size $M_2 - M_1$, where

$$M_2 = \max\{[(\eta\sigma_{M_1}/d)^2] + 1, M_1\}.$$

Let $\bar{X}_{M_2}$ be the mean of the pooled sample of size M_2. Then an approximate $(1 - \alpha)$-level confidence interval for μ is $(\bar{X}_{M_2} - d, \bar{X}_{M_2} + d)$.

Of course, the performance of the triple sampling technique depends very much on the accuracy of this approximation. In Section 2 we shall slightly modify the procedure by adding a small, *bounded* number of observations to our sample. It follows easily from the argument given there that if we increase the third sample size to $M_2' - M_1$, where

$$M_2' = \max\{[(\eta\sigma_{M_1}/d)^2 + (5 + \eta^2 - c)/2c + \varepsilon] + 1, M_1\}$$

for any $\varepsilon > 0$, then the coverage probability for a normal population will be strictly greater than $1 - \alpha$ for all sufficiently small d. Since we only need a *fixed* increase in sample size to achieve this inequality, our procedure is considerably more economical than Stein's when m is small relative to n_0.

If we are prepared to accept a coverage probability of $1 - \alpha + o(d^2)$ as $d \to 0$, the expected cost in extra observations of not knowing σ^2 is only $(\eta^2 + 1)/2c$. This quantity is less than 5 if $\eta = 1.96$ and $c = ½$.

In practice it seems a safe bet to choose $c = ½$; see Section 3. In theory, the efficiency of our procedure can be improved if we let c depend on m and increase slowly to 1 as $m \to \infty$. In the case of the normal distribution the results of the following sections remain true if we set $c = 1 - Cd^{1-\varepsilon}$ for positive constants C and ε. The value of c appearing in the limit results should then be replaced by 1.

In Section 2 we shall describe the asymptotic theory of triple sampling, and give the proofs in Section 4. The theory of Anscombe, Chow and Robbins' procedure has been presented by Anscombe (1952), and more recently by Woodroofe (1977). The arguments used by these authors depend on properties of stopping times for sums of independent, nonnegative random variables, and are very different from our own. In this context we must mention also the work of Simons (1968) and Starr and Woodroofe (1968, 1969, 1972). The techniques we have employed are closer to those used by Cox (1952) to analyse Stein's double sampling procedure.

The results of a series of Monte Carlo trials are presented in Section 3. These indicate that the difference between the expected sample size and the optimal sample size remains fairly static as d decreases, even if m is fixed. The Monte Carlo study also shows that if m is fixed then as d decreases the coverage probability tends at first to decrease, and then to increase. This suggests that the value of m is relatively important for large values of d. Our asymptotic study shows that m is unimportant for small d.

Our application of triple sampling to the construction of fixed width confidence intervals serves to demonstrate the idea, and of course it has other applications. For example, the results of Section 2 may be used to treat the problem of sequential point estimation, as in Woodroofe (1971). However, it does not seem possible to give a global theory for all such applications, like that given by Cox (1952) for double sampling.

2. Asymptotic theory. We begin with a slight reparametrization of the problem. Let $0 < c < 1$, $r \geq 1$ and $\lambda = \lambda(m)$, $m \geq 1$ be constants with the properties

$$\lambda(m) \to \infty, \ \limsup_{m\to\infty} m/\lambda(m) < c\sigma^2 \text{ and } \lambda(m) = 0(m^r). \tag{2.1}$$

Since σ^2 is unknown this effectively means that m must be chosen to be a smaller order of magnitude than $\lambda(m)$, in which case the lim sup in (2.1) will equal zero. Assume that X_1, $X_2, \cdots$ are independent normal $N(\mu, \sigma^2)$ variables, and define $N_1 = [c\lambda\sigma^2_m] + 1$, $M_1 = \max(N_1, m)$, $N_2 = [\lambda\sigma^2_{M_1}] + 1$ and $M_2 = \max(N_2, M_1)$. The pooled sample size is M_2. It is readily proved that $M_2/\lambda \to 1$ with probability 1, and in L^1. Our next result provides a more detailed account of the asymptotic behaviour of the moments of M_2.

THEOREM 1. *If* (2.1) *holds then*

$$E(M_2) = \lambda\sigma^2 + \tfrac{1}{2} - 2c^{-1} + o(1), \tag{2.2}$$

$$\mathrm{Var}(M_2) = 2c^{-1}\lambda\sigma^2 + o(\lambda) \tag{2.3}$$

(2.4) *and*

$$E\,|M_2 - EM_2|^3 = o(\lambda^2)$$

as $m \to \infty$.

It is very easy to prove a central limit theorem for M_2. Indeed, from condition (2.1) it follows that $P(M_2 = [\lambda\sigma^2_{N_1}] + 1) \to 1$ and $N_1/c\lambda \to 1$ in probability as $m \to \infty$.

A result of Anscombe (1952) now implies that

$$N_1^{1/2}(\sigma^2_{N_1} - 1) \to_{\mathscr{D}} N(0, 2),$$

whence

$$(c/\lambda)^{1/2}(M_2 - \lambda) \to_{\mathscr{D}} N(0, 2);$$

see also Bhattacharya and Mallik (1973). If we let c depend on m and converge to 1 sufficiently slowly, specifically if $\lambda^{1/2}(m)[1 - c(m)] \to \infty$, then

$$\lambda^{-1/2}(M_2 - \lambda) \to_{\mathscr{D}} N(0, 2).$$

Suppose σ^2 is unknown, and we wish to derive a confidence interval for μ of width $2d$ and coverage $1 - \alpha$. Let Φ and ϕ denote the standard normal distribution and density functions, and η be the solution of $1 - \Phi(\eta) = \alpha/2$. Let $\varepsilon = \varepsilon(d)$ be a sequence of constants converging to zero, to be determined explicitly very shortly. Carry out the triple sampling procedure above with an initial sample of size $m = m(d)$, and

$$\lambda = \lambda(d) = (\eta/d)^2(1 + \varepsilon), \tag{2.5}$$

obtaining a combined sample of size M_2. The optimal sample size is approximately $\lambda\sigma^2$, and condition (2.1) asks that the ratio of the pilot sample size to the optimal sample size be less than c. The interval $I = (\bar{X}_{M_2} - d, \bar{X}_{M_2} + d)$ is an approximate $(1 - \alpha)$-level confidence interval for μ (Anscombe, 1952), and we shall apply Theorem 1 to obtain a measure of the order of this approximation.

Let $\Psi(x) = 2\{1 - \Phi(x^{1/2})\}$, $x > 0$, and $\ell = d/\sigma$. The exact coverage probability of I is $1 - \alpha(d)$, where

$$\alpha(d) = E\{\Psi(\ell^2 M_2)\} = \Psi(\ell^2 EM_2) + \tfrac{1}{2}\ell^4 E(M_2 - EM_2)^2 \Psi''(\ell^2 EM_2) + r_1(d),$$

where $|r_1(d)| \le C\ell^6 E\,|M_2 - EM_2|^3 = o(d^2)$ by Theorem 1. Also,

$$\Psi(\ell^2 EM_2) = \alpha + (\ell^2 EM_2 - \eta^2)\Psi'(\eta^2) + r_2(d),$$

where $|r_2(d)| = o(d^2 + |\varepsilon|)$. From Theorem 1 we see that

$$\ell^2 EM_2 - \eta^2 = \eta^2\varepsilon + (1 - 4/c)d^2/2\sigma^2 + o(d^2)$$

and

$$\tfrac{1}{2}\ell^4 E(M_2 - EM_2)^2 = \eta^2 d^2/c\sigma^2 + o(d^2),$$

whence

$$\alpha(d) = \alpha + \phi(\eta)\{(d/\sigma)^2(5 + \eta^2 - c)/2\eta c - \varepsilon\eta\} + o(d^2 + |\varepsilon|).$$

This expansion suggests setting $\varepsilon = (d/\sigma)^2(5 + \eta^2 - c)/2\eta^2 c$ in (2.5). Of course σ^2 is unknown, and the logical alternative is to slightly modify the triple sampling procedure. Starting with the same initial sample of size m, gather a second sample of size $M_1 - m$ where $M_1 = \max(m, [c(\eta/d)^2\sigma^2_m] + 1)$, and a third sample of size $M - M_1$ where

$$M = \max(M_1, [(\eta/d)^2\sigma^2_{M_1} + (5 + \eta^2 - c)/2c] + 1). \tag{2.6}$$

The interval $(\bar{X}_M - d, \bar{X}_M + d)$ is an approximate $(1 - \alpha)$-level confidence interval for μ, and minor modifications to the argument above will show that $P(|\bar{X}_M - \mu| > d) = \alpha + o(d^2)$. The expected size of the pooled sample is

$$E(M) = (\eta/d)^2\sigma^2 - 2c^{-1} + (5 + \eta^2 - c)/2c + ½ + o(1)$$
$$= (\eta/d)^2\sigma^2 + (\eta^2 + 1)/2c + o(1).$$

This compares favourably with the expected sample size needed in the Anscombe-Chow-Robbins sequential procedure to achieve a similar order of accuracy, (Anscombe, 1953).

$$E(M_{ACR}) = (\eta/d)^2\sigma^2 + (\eta^2 + 1)/2 + o(1).$$

3. Monte Carlo trials. We conducted a series of Monte Carlo trials using a wide range of values of the parameters. Varying c from 0.5 to 0.8 led to no detectable change in coverage probabilities. However, large values of c (0.7 or 0.8) led to a slight increase in average sample sizes when the optimal sample size, n_0, was small to moderate ($\le$200). This was due to a tendency for N_1 to exceed n_0 when c was large and m small. The increase with $c = 0.8$ was generally less than 5% of the average sample size with $c = 0.5$.

We shall report in detail on a series of trials with $c = 0.5$. Set $\eta = 1.96$ and $m = 10$, and let $M_1 = \max(10, [0.5(1.96/d)^2\sigma_m^2] + 1)$ and

$$M = \max(M_1, [(1.96/d)^2\sigma_m^2 + 8.3416] + k + 1), \tag{3.1}$$

where $k \ge 0$ is an integer. Note that $(5 + \eta^2 - c)/2c = 8.3416$, so that formulae (2.6) and (3.1) coincide when $k = 0$. The confidence interval $(\bar{X}_M - d, \bar{X}_M + d)$ should cover the mean with probability about 0.95, the exact coverage probability increasing with k.

Table 1 presents the results of Monte Carlo trials using the standard normal distribution. For each of 9 values of d in the range 0.1 to 0.4 we conducted 1,000 trials using the

TABLE 1

Results of 1000 Monte Carlo trials with $c = 0.5$, $\eta = 1.96$, $m = 10$

d	n_0	k = 0				k = 3			
		$\bar{M}$	$\bar{M}-n_0$	s_M	p	$\bar{M}$	$\bar{M}-n_0$	s_M	p
0.4	24	29.3	5.3	11.0	0.950	32.3	8.3	10.9	0.964
0.3	43	46.8	3.8	16.1	0.956	49.1	6.1	16.5	0.949
0.25	61	65.2	4.2	18.8	0.949	68.5	7.5	18.9	0.948
0.225	76	84.3	8.3	20.4	0.953	82.5	6.5	21.6	0.959
0.2	96	100.4	4.4	22.4	0.930	103.4	7.4	23.5	0.942
0.175	125	137.2	12.2	27.5	0.948	133.4	8.4	27.3	0.955
0.15	171	174.9	3.9	32.8	0.936	184.2	13.2	30.7	0.970
0.125	246	252.9	6.2	39.0	0.959	252.3	6.3	38.5	0.952
0.1	384	389.2	5.2	47.4	0.958	393.4	9.4	48.3	0.958

d	n_0	k = 5				k = 8			
		$\bar{M}$	$\bar{M}-n_0$	s_M	p	$\bar{M}$	$\bar{M}-n_0$	s_M	p
0.4	24	34.1	10.1	11.1	0.973	37.8	13.8	11.0	0.978
0.3	43	52.0	9.0	15.8	0.958	55.1	12.1	15.9	0.963
0.25	61	70.8	9.8	18.1	0.955	73.4	12.4	18.8	0.962
0.225	76	88.9	12.9	20.8	0.945	88.1	12.1	21.4	0.952
0.2	96	104.6	8.6	24.1	0.951	108.5	12.5	23.3	0.953
0.175	125	142.3	17.3	27.8	0.949	137.8	12.8	27.6	0.951
0.15	171	180.6	9.6	30.4	0.954	182.5	11.5	31.7	0.954
0.125	246	256.1	10.1	38.4	0.959	258.5	12.5	38.6	0.958
0.1	384	394.9	10.9	47.3	0.954	398.7	14.7	47.7	0.954

technique described above, and repeated these for different values of k. For each set of 1,000 values of M we computed the mean $\bar{M}$ and standard deviation s_M, as well as the proportion of times p that the confidence interval covered the origin. The results indicate that taking $k = 5$ or 8 should lead to a confidence interval with coverage probability very nearly equal to 0.95.

The value of n_0 in Table 1 equals the integer nearest to $(1.96/d)^2$. It is clear that n_0 is close to the average sample size, $\bar{M}$, even when the initial sample size is only one fortieth of n_0. Indeed, the difference $\bar{M} - n_0$ remains fairly constant as n_0 increases.

4. Proofs. It clearly suffices to consider the case $\mu = 0$, and we may also suppose that $\sigma^2 = 1$. Indeed, to obtain the general results of Section 2 from the special case $\sigma^2 = 1$, it is necessary only to replace λ by $\lambda\sigma^2$ in (2.1)–(2.4). Thus, we may assume below that our population is $N(0, 1)$. We shall prove Theorem 1 as a corollary of

THEOREM 2. *Under the conditions of Theorem* 1,

$$\lambda E(\sigma_{M_1}^2) = \lambda - c^{-1}\operatorname{var}(X_1^2) + o(1), \tag{4.1}$$

$$E(M_2) = E(N_2) + o(1), \tag{4.2}$$

$$\operatorname{Var}(M_2) = \operatorname{Var}(N_2) + o(1) \tag{4.3}$$

and

$$\operatorname{Var}(N_2) = c^{-1}\lambda \operatorname{Var}(X_1^2) + o(\lambda). \tag{4.4}$$

Interestingly, Theorem 2 remains true under condition (1) and the assumption that $E|X_1|^{4r} < \infty$; the normality of X_1 is not necessary. However, we shall not prove it in this generality.

We first state

LEMMA 1. *For any $\delta > 0$ there exists $\xi = \xi(\delta) > 0$ such that*

$$\{|\sigma_m^2 - 1| > \delta\} \subseteq \{|\textstyle\sum_1^m (X_i^2 - 1)| > m\xi\} \cup \{|\textstyle\sum_1^m X_i| > m\xi\}$$

for all m.

LEMMA 2. *With $\Delta = c\lambda\sigma_m^2 - [c\lambda\sigma_m^2]$ we have*

$$\begin{aligned} c\lambda E(\sigma_{M_1}^2 \mid X_1, \cdots, X_m) &= c\lambda + mc\lambda(\sigma_m^2 - 1)/M_1 + (1 - \Delta)(M_1 - m)/M_1(M_1 - 1) \\ &\quad + \{c\lambda(m^{-1/2} \textstyle\sum_1^m X_i)^2/(M_1 - 1) - 1\} \\ &\quad + \{(m - 1)/(M_1 - 1) - mc\lambda(m^{-1/2} \textstyle\sum_1^m X_i)^2/M_1(M_1 - 1)\} \end{aligned}$$

almost surely on the set $\{M_1 = N_1\}$.

PROOF. This follows after some manipulation of the relation

$$\begin{aligned} &(M_1 - 1)E(\sigma_{M_1}^2 \mid X_1, \cdots, X_m) \\ &\quad = (M_1 - 1)(1 - m/M_1) + (m - 1)\sigma_m^2 + (M_1 - m)(\textstyle\sum_1^m X_j)^2/mM_1. \end{aligned} \tag{4.5}$$

Note that in view of (2.1),

$$\{N_1 < m\} \subseteq \{c\lambda\sigma_m^2 < m\} \subseteq \{|\sigma_m^2 - 1| > \delta\} \tag{4.6}$$

for some $\delta > 0$ and all large m.

We are now in a position to prove (4.1). We shall let $r_{im}(\delta)$ and $r_i(\delta)$ denote real constants indexed by m and $\delta > 0$, and with the property

$$\limsup_{m\to\infty} |r_{\iota m}(\delta)| = r_\iota(\delta) \to 0 \quad \text{as} \quad \delta \to 0. \tag{4.7}$$

Fix $\delta > 0$. For large m, $\{|M_1 - 1 - c\lambda| > c\lambda\delta\} \subseteq \{N_1 < m\} \cup \{|\sigma_m^2 - 1| > \delta/2\}$, and so in view of (4.6) and Lemma 1,

$$\begin{aligned} A_m &\equiv \{|\sigma_m^2 - 1| > \delta\} \cup \{|M_1 - 1 - c\lambda| > c\lambda\delta\} \cup \{M_1 \neq N_1\} \\ &\subseteq \{|\textstyle\sum_1^m (X_i^2 - 1)| > m\xi\} \cup \{|\sum_1^m X_i| > m\xi\} \equiv B_m \end{aligned} \tag{4.8}$$

for some $\xi > 0$. The set A_m is measurable in the σ-field generated by $X_1, \cdots, X_m$, and so from Lemma 2,

$$\begin{aligned} c\lambda E\{\sigma_{M_1}^2 I(\tilde{A}_m)\} &= c\lambda E\{I(\tilde{A}_m) E(\sigma_{M_1}^2 | X_1, \cdots, X_m)\} \\ &\leq c\lambda + mc\lambda E\{I(\tilde{A}_m)(\sigma_m^2 - 1)/M_1\} + E(M_1^{-1}) \\ &\quad + \{c\lambda/c\lambda(1-\delta) - 1\} + \{m/c\lambda(1-\delta) - mc\lambda/(c\lambda(1+\delta)+1)^2\} \\ &\quad + mc\lambda E\{(m^{-1/2} \textstyle\sum_1^m X_i)^2 I(B_m)\}/(c\lambda(1-\delta))^2. \end{aligned}$$

The sum of the last four terms may be made arbitrarily small by choosing δ small and then m large, and a reverse inequality may be obtained in a similar way. Therefore

$$c\lambda E\{\sigma_{M_1}^2 I(\tilde{A}_m)\} = c\lambda + mc\lambda E\{I(\tilde{A}_m)(\sigma_m^2 - 1)/M_1\} + r_{1m}(\delta),$$

where $r_{1m}(\delta)$ satisfies (4.7). For $|\sigma_m^2 - 1| \leq \delta < 1/2$ and large m we have

$$c\lambda(\sigma_m^2 - 1)/M_1 = (\sigma_m^2 - 1)\{1 - (\sigma_m^2 - 1) + R_m\}$$

where $|R_m| \leq C\{\delta|\sigma_m^2 - 1| + (c\lambda)^{-1}\}$. Consequently

$$mc\lambda E\{I(\tilde{A}_m)(\sigma_m^2 - 1)/M_1\} = -mE(\sigma_m^2 - 1)^2 + r_{2m}(\delta),$$

where

$$|r_{2m}(\delta)| \leq m\{E((|\sigma_m^2 - 1| + |\sigma_m^2 - 1|^2)I(A_m)) + C(\delta E|\sigma_m^2 - 1|^2 + (c\lambda)^{-1}E|\sigma_m^2 - 1|)\}.$$

But $mE(\sigma_m^2 - 1)^2 \sim \operatorname{Var} X_1^2$,

$$mE\{|\sigma_m^2 - 1| I(A_m)\} \leq m(E|\sigma_m^2 - 1|^2)^{1/2}(P(B_m))^{1/2} = o(1),$$

and since $\{m(\sigma_m^2 - 1)^2, m \geq 1\}$ is uniformly integrable then $mE\{(\sigma_m^2 - 1)^2 I(A_m)\} = o(1)$. Therefore $r_{2m}(\delta)$ satisfies (4.7), and combining these estimates we deduce that

$$\lambda E\{\sigma_{M_1}^2 I(\tilde{A}_m)\} = \lambda - c^{-1} \operatorname{Var} X_1^2 + r_{3m}(\delta).$$

We shall complete the proof of (4.1) by showing that with B_m as in (4.8),

$$\lambda E\{\sigma_{M_1}^2 I(B_m)\} = o(1).$$

From (4.5) we see that for $m \geq 2$, $E(\sigma_{M_1}^2 | X_1, \cdots, X_m) \leq C\{1 + m^{-2}(\sum_1^m X_i)^2\}$; note that $c\lambda\sigma_m^2 \leq M_1$. Therefore

$$\begin{aligned} \lambda E\{\sigma_{M_1}^2 I(B_m)\} &= \lambda E\{I(B_m) E(\sigma_{M_1}^2 | X_1, \cdots, X_m)\} \\ &\leq \lambda C\{P(B_m) + m^{-4r}\xi^{-4r+2} E|\textstyle\sum_1^m X_i|^{4r} \\ &\quad + m^{-2} \displaystyle\int_{\{|\sum_1^m (X_i^2 - 1)| > m\xi\}} |\textstyle\sum_1^m X_i|^2 \, dP\} = o(1), \end{aligned}$$

as required.

LEMMA 3. *There exist positive constants ξ and η, and integers $p = p(m) > (1 + \xi)m$, such that for all sufficiently large m,*

$$\{\lambda\sigma_{N_1}^2 < c\lambda\sigma_m^2 + 1\} \subseteq \{|\textstyle\sum_1^m (X_i^2 - 1)| > m\xi\} \cup \{|\sum_1^m X_i| > m\xi\}$$
$$\cup \{|\textstyle\sum_{m+1}^p (X_i^2 - 1)| > p\xi\} \cup \{\sup_{n>\lambda\eta} |n^{-1} \sum_1^n X_i| > \xi\}.$$

Lemma 3 follows in a fairly straightforward way from Lemma 1, noting that

$$\{\lambda\sigma_{N_1}^2 < c\lambda\sigma_m^2 + 1\} \cap \{|N_1 - c\lambda| \le c\lambda\delta\} \cap \{\bar{X}_{N_1}^2 \le \delta\}$$
$$\subseteq \{(m-1)\lambda(\textstyle\sum_1^{c\lambda(1-\delta)} X_i^2 - c\lambda(1+\delta)\delta) < c\lambda(1+\delta)c\lambda \sum_1^m X_i^2 + c\lambda(1+\delta)(m-1)\}$$
$$\subseteq \{\textstyle\sum_1^{c\lambda(1-\delta)} X_i^2 < c\lambda(1-2\delta)\} \cup \{\sum_1^m X_i^2 > mc^{-1}(1-5\delta)\},$$

provided m is large.

From Lemmas 1 and 3 we conclude that

$$P(|\sigma_m^2 - 1| > \delta) + P(\lambda\sigma_{N_1}^2 < c\lambda\sigma_m^2 + 1) = o(m^{-2r+1}). \tag{4.9}$$

Since $M_2 = N_2$ except possibly on the set $E = \{N_1 < m\} \cup \{\lambda\sigma_{N_1}^2 < c\lambda\sigma_m^2 + 1\}$, condition (4.2) will follow if we prove that

$$\int_E (m + \lambda\sigma_m^2 + \lambda\sigma_{M_1}^2)\, dP = o(1), \tag{4.10}$$

and hence if we show that

$$mP(N_1 < m) + mP(\lambda\sigma_{N_1}^2 < c\lambda\sigma_m^2 + 1) = o(1) \tag{4.11}$$

(4.12) and
$$\lambda \int_{\{\lambda\sigma_{N_1}^2 < c\lambda\sigma_m^2 + 1\}} \sigma_m^2\, dP = o(1).$$

Condition (4.11) follows from (4.9) using (4.6). To prove (4.12) observe that for any event E and $m \ge 2$,

$$\int_E \sigma_m^2\, dP \le 2P(E) + 2m^{-1} \int_E |\textstyle\sum_1^m (X_i^2 - 1)|\, dP,$$

and so it suffices to show that

$$R_i \equiv (\lambda/m) \int_{E_i} |\textstyle\sum_1^m (X_i^2 - 1)|\, dP = o(1), \qquad i = 1, 2, 3,$$

where $E_1 = \{|\sum_1^m (X_i^2 - 1)| > m\xi\}$, $E_2 = \{|\sum_{m+1}^p (X_i^2 - 1)| > p\xi\}$ and

$$E_3 = \{|\textstyle\sum_1^m X_i| > m\xi\} \cup \{\sup_{n>\lambda\eta} |n^{-1} \sum_1^n X_i| > \xi\}.$$

This is easily accomplished using the moment bounds and, in the case of R_3, the Cauchy-Schwartz inequality.

The techniques leading to (4.10) may be used to prove that

$$\int_E (m^2 + \lambda^2\sigma_m^4 + \lambda^2\sigma_{M_1}^4)\, dP = o(\lambda). \tag{4.13}$$

We shall prove shortly that $\mathrm{Var}(M_1) = O(\lambda)$, and from this and (4.13) we deduce (4.3).

It remains to establish (4.4). We first prove

LEMMA 4. *Under the conditions of Theorem* 1,

$$E\{(M_1(M_1 - 1))^{-2}(\textstyle\sum_1^{M_1} X_i)^4\} = o(\lambda^{-1}).$$

PROOF. Expanding and taking first the mean conditional on $X_1, \cdots, X_m$ we see that the left side is dominated by

$$CE\{M_1^{-4}(\sum_1^m X_i)^4 + M_1^{-3}(|\sum_1^m X_i| + |\sum_1^m X_i|^2) + M_1^{-2}\}$$
$$\le C\{(c\lambda(1-\delta))^{-4}E(\sum_1^m X_i)^4 + (c\lambda(1-\delta))^{-3}(E|\sum_1^m X_i| + E|\sum_1^m X_i|^2) + (c\lambda(1-\delta))^{-2}\}$$
$$+ C\left\{m^{-4}\int_{D_m}(\sum_1^m X_i)^4\,dP + m^{-3}\int_{D_m}(|\sum_1^m X_i| + |\sum_1^m X_i|^2)\,dP + m^{-2}P(D_m)\right\},$$

where

$$D_m \equiv \{|M_1 - c\lambda| > c\lambda\delta\} \subseteq \{|\sum_1^m (X_i^2 - 1)| > m\xi\} \cup \{|\sum_1^m X_i| > m\xi\}$$

for some $\xi > 0$. The proof may now be completed as before.

LEMMA 5. *Under the conditions of Theorem* 1,

$$\begin{aligned}\operatorname{Var}\{(M_1-1)^{-1}(\textstyle\sum_1^{M_1} X_i^2)\} &= E\{(M_1-1)^{-1}\}\operatorname{Var}X_1^2 + \operatorname{Var}\{(M_1-1)^{-1}\}\\ &\quad + E\{(M_1-1)^{-2}\}\operatorname{Var}X_1^2 + \operatorname{Var}\{(M_1-1)^{-1}\textstyle\sum_1^m (X_i^2-1)\}\\ &\quad - m(\operatorname{Var}X_1^2)E\{(M_1-1)^{-2}\}\\ &\quad + 2\operatorname{Cov}\{(M_1-1)^{-1}\textstyle\sum_1^m (X_i^2-1), (M_1-1)^{-1}\}.\end{aligned}$$

The techniques used to derive (4.1) may be employed to prove that $E\{(M_1-1)^{-1}\} \sim (c\lambda)^{-1}$ and $E\{(M_1-1)^{-2}\} \sim (c\lambda)^{-2}$. Therefore (4.4) will follow from Lemmas 1 and 3 if we prove that

$$|\operatorname{Var}\{(M_1-1)^{-1}\textstyle\sum_1^m (X_i^2-1)\} - m(\operatorname{Var}X_1^2)E\{(M_1-1)^{-2}\}| = o(\lambda^{-1}) \tag{4.14}$$

and

$$|\operatorname{Cov}\{(M_1-1)^{-1}\textstyle\sum_1^m (X_i^2-1), (M_1-1)^{-1}\}| = o(\lambda^{-1}). \tag{4.15}$$

With $E_m = \{|M_1 - 1 - c\lambda| > c\lambda\delta\}$ we deduce that

$$\begin{aligned}a_m \equiv |E\{(M_1-1)^{-1}\textstyle\sum_1^m (X_i^2-1)\}| &= |E\{(c\lambda(M_1-1))^{-1}(c\lambda - (M_1-1))\textstyle\sum_1^m (X_i^2-1)\}|\\ &\le (c\lambda(1-\delta))^{-1}\delta E|\textstyle\sum_1^m (X_i^2-1)|\\ &\quad + Cm^{-1}\int_{E_m}|\textstyle\sum_1^m (X_i^2-1)|\,dP = o(\lambda^{-1/2}).\end{aligned}$$

A similar argument will show that

$$|E\{(M_1-1)^{-2}\textstyle\sum_1^m (X_i^2-1)\}| = o(\lambda^{-1}).$$

This proves (4.15), and (4.14) will follow if

$$|E\{(M_1-1)^{-2}(\textstyle\sum_1^m (X_i^2-1))^2\} - (c\lambda)^{-2}m\operatorname{Var}(X_1^2)| = o(\lambda^{-1}). \tag{4.16}$$

In view of condition (2.1) and the classical subsequence argument, it suffices to consider the case $m/\lambda(m) \to \ell$ where $0 \le \ell < 1$. If $\ell = 0$ then $r > 1$, and the left side of (4.16) is dominated by the sum of the two terms, which is readily seen to be $o(\lambda^{-1})$; note that $m/\lambda^2 = o(\lambda^{-1})$. If $\ell > 0$ then for any $\delta > 0$,

$$E\{(M_1-1)^{-2}(\textstyle\sum_1^m (X_i^2-1))^2\} \le (c\lambda(1-\delta))^{-2}m\operatorname{Var}X_1^2 + 2m^{-2}\int_{E_m}|\textstyle\sum_1^m (X_i^2-1)|\,dP.$$

The last term is $o(m^{-1}) = o(\lambda^{-1})$, and since a reverse inequality may be established in the same way then (4.16) is true.

We now derive Theorem 1 from Theorem 2. The result (2.3) follows from (4.3) and (4.4), and (2.4) may be proved using a similar argument. Condition (2.2) will follow from (4.1) and (4.2) if we show that $U_m \equiv \lambda\sigma_{N_1}^2 - [\lambda\sigma_{N_1}^2]$ is asymptotically uniform on (0, 1).

We may write $\sigma_n^2 = (n-1)^{-1} \sum_1^{n-1} Y_i^2$, $n \geq 2$, where the variables $Y_i = \{i(i+1)\}^{-1/2}(iX_{i+1} - \sum_{j=1}^{i} X_j)$ are independent $N(0, 1)$. Let $J = [c\lambda\sigma_m^2]$ and $V = c\lambda\sigma_m^2 - J$. Conditional on $J = j \geq m$ and $V = v$ we have $\lambda\sigma_{N_1}^2 = (\lambda/j) \sum_m^j Y_i^2 + a$, where $a = (m-1)(j+v)/cj$. Therefore with F_n and f_n denoting the distribution and density functions of a χ_n^2 variable,

$$P(U_m \leq x \mid J = j, V = v) = \textstyle\sum_k \{F_{j-m+1}(j(k+x-a)/\lambda) - F_{j-m+1}(j(k-a)/\lambda)\}$$
$$= (jx/\lambda) \textstyle\sum_k f_{j-m+1}(j(k-a)/\lambda) + r_{1m}(j, v)$$

where

$$|r_{1m}| \leq \tfrac{1}{2}(jx/\lambda)^2 \textstyle\sum_k \sup_{0<y<1} |f'_{j-m+1}(j(k+y-a)\lambda)|.$$

The function $|f'_{j-m+1}|$ has at most two local maxima on $(0, \infty)$, and so

$$|r_{1m}| \leq (j/\lambda)^2 \textstyle\sum_k |f'_{j-m+1}(j(k-a)/\lambda)| + (j/\lambda)^2 \sup_z |f'_{j-m+1}(z)|.$$

Integration by parts shows that for a piecewise differentiable function g,

$$\tfrac{1}{2}\{g(k) + g(k+1)\} = \int_k^{k+1} g(z)\, dz + \int_k^{k+1} g'(z)(z - k - \tfrac{1}{2})\, dz.$$

Consequently

$$\textstyle\sum_k f_{j-m+1}(j(k-a)/\lambda) = \int_{-\infty}^{\infty} f_{j-m+1}(j(z-a)/\lambda)\, dz + r_{2m}$$

and

$$\textstyle\sum_k |f'_{j-m+1}(j(k-a)/\lambda)| = \int_{-\infty}^{\infty} |f'_{j-m+1}(j(z-a)/\lambda)|\, dz + r_{3m},$$

where

$$|r_{2m}| \leq (j/\lambda) \int_{-\infty}^{\infty} |f'_{j-m+1}(j(z-a)/\lambda)|\, dz = \int_0^{\infty} |f'_{j-m+1}(z)|\, dz,$$

and

$$|r_{3m}| \leq \int_0^{\infty} |f''_{j-m+1}(z)|\, dz.$$

Combining these estimates we see that

$$P(U_m \leq x \mid J = j, V = v) = x + r_{4m},$$

where for $j \geq m$,

$$|r_{4m}| \leq 2(j/\lambda) \int_0^{\infty} |f'_{j-m+1}(z)|\, dz + (j/\lambda)^2 \left\{ \int_0^{\infty} |f''_{j-m+1}(z)|\, dz + \sup_z |f'_{j-m+1}(z)| \right\}.$$

It follows from elementary calculus that

$$\int_0^{\infty} |f'_n(z)|\, dz = O(n^{-1/2}) \qquad \text{and} \qquad \int_0^{\infty} |f''_n(z)|\, dz + \sup_z |f'_n(z)| = O(n^{-1})$$

as $n \to \infty$. Therefore uniformly in $0 < v < 1$ and $j > (1+\varepsilon)m$ we have $|r_{4m}| \leq C(j^{1/2}/\lambda + j/\lambda^2)$, where C depends only on $\varepsilon > 0$. Consequently $P(U_m \leq x) = x + r_{5m}$ where

$$|r_{5m}| \leq C\{P(J \leq (1+\varepsilon)m) + E(J^{1/2}/\lambda + J/\lambda^2)\}.$$

The proof is completed by noting that $E(J)/\lambda^2 \to 0$, and $P(J \leq (1+\varepsilon)m) \to 0$ if ε is sufficiently small.

Acknowledgments. I am indebted to the referee and the associate editor for comments which led to a considerable improvement in presentation.

REFERENCES

ANSCOMBE, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48** 600–607.

ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc.* **15** 1–21.

BERNSTEIN, S. (1939). Quelques remarques sur le théorème limite Liapounoff. *Dokl. Akad. Nauk SSSR (Comptes rendus)* **24** 3–8.

BHATTACHARYA, P. K. and MALLIK, A. (1973). Asymptotic normality of the stopping times of some sequential procedures. *Ann. Statist.* **1** 1203–1211.

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.

COX, D. R. (1952). Estimation by double sampling. *Biometrika* **39** 217–227.

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

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.

STARR, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.

STARR, N. and WOODROOFE, M. (1968). Remarks on a stopping time. *Proc. Nat. Acad. Sci. U.S.A.* **61** 1215–1218.

STARR, N. and WOODROOFE, M. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. U.S.A.* **63** 285–288.

STARR, N. and WOODROOFE, M. (1972). Further remarks on sequential point estimation: The exponential case. *Ann. Math. Statist.* **43** 1147–1154.

STEIN, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

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

DEPARTMENT OF STATISTICS
FACULTY OF ECONOMICS
THE AUSTRALIAN NATIONAL UNIVERSITY
POST OFFICE BOX 4
CANBERRA, A.C.T. 2600
AUSTRALIA