IMPROVING THE FULLY SEQUENTIAL SAMPLING SCHEME OF ANSCOMBE–CHOW–ROBBINS

By Wei Liu

University of Southampton

A new sequential sampling scheme is proposed in which, after an initial batch sample, sampling is continued in batches of data-dependent sizes (at most k such batches), and then one-at-a-time with a data-dependent stopping rule. This new scheme requires about the same sample size as the fully sequential Anscombe–Chow–Robbins (ACR) sampling scheme but substantially fewer sampling operations. The problem of constructing fixed-width confidence intervals for the mean of a normal population with unknown variance is used as an illustration.

1. Introduction. In many statistical inference problems, some predetermined accuracy is required of a procedure used, and the "optimal" fixed-sample-size procedure to meet this accuracy requirement often depends on some unknown nuisance parameter. For example, we wish to construct a confidence interval for the unknown mean θ of a normal population $N(\theta, \sigma^2)$ with preassigned accuracy "width $2d$ and confidence level γ" for given $d > 0$ and $\gamma \in (0, 1)$; the optimal fixed-sample-size procedure requires a sample of size $n_0 = (z\sigma/d)^2$, where $z = \Phi^{-1}((1 + \gamma)/2)$ and Φ is the cdf of a $N(0, 1)$ random variable, and constructs confidence interval $\bar{Y} \pm d$ for θ, where $\bar{Y}$ is the sample mean. Note, however, the sample size $n_0 = (z\sigma/d)^2$ depends on σ^2, which is often (and is assumed in this paper) unknown. To solve such problems, it is necessary to use a sequential sampling scheme.

A two-stage sampling scheme of Stein (1945) requires only two samples to achieve this end. The basic idea is to take a first sample of size m to get an estimate of the unknown parameter and, hence, an estimate of the required total sample size by using the fixed-sample-size formula; the second sample is then taken to make good the shortfall of this estimated total sample size. When the size of the first sample m is too small, however, the estimates after the first sample may not be accurate enough and, as a result, the two-stage scheme can perform poorly. On the other hand, if the size of the first sample m is set to be too large, then m itself may already exceed the required total sample size.

To improve the two-stage scheme, Hall (1981) proposed a three-stage scheme. Instead of taking just one sample to make up the shortfall of the projected total sample size after the first sample, this scheme takes a second sample to make up only a proportion (half, say) of the projected total sample size. After this second sample, the unknown parameter and the required

---

total sample size can then be reestimated, and they are most likely to be more accurate than those estimates calculated after the first sample. A third sample is then taken to make up the shortfall of the newly estimated total sample size. As one might expect, the three-stage scheme is robust to the small value of m, since it has the extra opportunity to update the estimates after the second sample, which is often of reasonably large size.

The most frequently used sequential sampling scheme is the fully sequential sampling scheme due to Anscombe (1953), Robbins (1959) and Chow and Robbins (1965), the ACR scheme. After the first sample of size m, the ACR scheme takes observations one by one. It renews the estimates of the unknown parameter and the total sample size after each new observation, and checks whether enough observations have already been drawn. Not surprisingly, the ACR scheme is very efficient in terms of sample size. The stopping time of the ACR scheme may be written in the form

$$(1.1) \qquad t^* = \inf\big\{n \geq m \colon S_n < cn^\alpha L(n)\big\}$$

[see Woodroofe (1977)], where S_n, $n \geq 1$, are the partial sums of certain i.i.d. positive random variables $X_1, X_2, \ldots,$ $L(n)$ is a sequence of numbers given by $1 + L_0/n + o(1/n)$ as $n \to \infty$, $\alpha > 1$, $m \geq 1$ and c is a positive parameter (which is often allowed to approach zero). The properties of t^* have been studied by Woodroofe (1977) and, in particular, it has been shown that under certain assumptions

$$(1.2) \qquad E(t^*) = \lambda + \beta\mu^{-1}\nu - \beta L_0 - \tfrac{1}{2}\alpha\beta^2\tau^2\mu^{-2} + o(1)$$

as $c \to 0$, where $\beta = 1/(\alpha - 1)$, $\lambda = (\mu/c)^\beta$, $\mu = E(X_1)$, $\tau^2 = \mathrm{Var}(X_1)$ and

$$(1.3) \qquad \nu = (2\mu)^{-1}\beta[(\alpha - 1)^2\mu^2 + \tau^2] - \sum_{n=1}^{\infty} n^{-1}E\{\max(S_n - n\alpha\mu, 0)\}.$$

Despite its great efficiency in terms of sample size, the ACR scheme can be expensive to carry out since, after the first sample, it is fully sequential and so requires a lot of sampling operations. In many real situations significant savings can be achieved by gathering many observations together. The purpose of this paper is to propose a new sequential sampling scheme which needs about the same sample size as the ACR scheme but substantially fewer sampling operations. This new scheme starts with up to $k + 1$, $k \geq 1$, random samples followed by fully sequential sampling. The motivation is that, when we are far away from the target, the truly required total sample size, we can leap forward by taking clusters of observations, and when we are getting closer to the target we should approach carefully by taking one observation at a time. The new sampling scheme is given in Section 2 with the proofs outlined in Section 4. Section 3 applies the general theory to construct a fixed-width confidence interval for θ of a normal population $N(\theta, \sigma^2)$. Throughout this paper, $\langle x \rangle$ denotes the largest integer no larger than x.

2. The new sampling scheme. The new sequential sampling scheme is defined in terms of a stopping time, as for the ACR scheme. Assume that integer k (≥ 1) and constants $0 < \rho_1 < \cdots < \rho_k < 1$ are prefixed, and that $m \geq 1$ is an integer which will be allowed to approach infinity. Under the notation of Section 1, define

$$N_0 = m,$$

(2.1) $$N_j = \max\{\langle \rho_j(\bar{X}_{N_{j-1}}/c)^\beta \rangle, N_{j-1}\}, \qquad 1 \leq j \leq k,$$

$$t_m = \inf\{n \geq N_k: S_n < cn^\alpha L(n)\}.$$

The new scheme is to first take a random sample of size N_0, then successive random samples of sizes $N_j - N_{j-1} \geq 0$, $j = 1, \ldots, k$, and then to sample one at a time until the stopping time t_m. It is clear that this t_m is similar to the stopping time t^* in (1.1) of the ACR scheme, except that $t_m \geq N_k$.

The limit operation under which the asymptotic results have been established is similar to those in Hall (1981):

(2.2) $$m \to \infty, \qquad c = c(m) \to 0, \qquad \lambda = O(m^r), \qquad \limsup \frac{m}{\lambda} < \rho_1,$$

where $r \geq 1$ is a fixed constant. In addition to the assumption that $X_1, X_2, \ldots$ are i.i.d. positive random variables with both mean μ and variance τ^2 positive and finite, we suppose also that X_1 has a density f which is continuous a.e. and that some power of the characteristic function of X_1 is integrable. Finally, we suppose that

(2.3) $$F(x) \leq Bx^a \quad \text{for all } x > 0$$

and for some $B > 0$ and $a > 0$, where F is the distribution function of X_1. Now we have the following major result of this paper.

THEOREM 2.1. *Suppose that (2.3) holds and that $E|X_1|^\xi < \infty$ for some constant $\xi > 2r + 1/(\alpha - 1)$. Then, under limit operation (2.2),*

(2.4) $$E(t_m) = \lambda + \beta\mu^{-1}\nu - \beta L_0 - \tfrac{1}{2}\alpha\beta^2\tau^2\mu^{-2} + o(1).$$

It is clear that the asymptotic expansions of $E(t_m)$ and $E(t^*)$ are of the same form, though the limit operations are different. Although the constants ρ_j do not appear in the asymptotic expansion of $E(t_m)$, they play vital roles in determining the performance of this new sequential sampling scheme. In actual fact, if ρ_k is set close to zero, then the new scheme is similar to the ACR scheme, since most observations will be taken in the stage of fully sequential sampling. On the other hand, if ρ_1 is set close to unity, then the new scheme behaves like Stein's two-stage scheme, since most observations will be taken in the first two samples.

3. Fixed-width confidence interval for a normal mean.   Assume that random observations $Y_1, Y_2, \ldots$ are from $N(\theta, \sigma^2)$. Then the ACR procedure continues sampling until

$$T^* = \inf\{n \geq m + 1 \colon n > l_n(z\sigma_n/d)^2\},$$

where $m + 1 \ (\geq 2)$ is the size of the first sample, l_n is a sequence of numbers given by $l_n = 1 + 2\Delta/n + o(1/n)$ as $n \to \infty$, and $\sigma_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$ with $\bar{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$. On stopping sampling, a confidence interval for θ is defined to be $I_{T^*} \equiv \bar{Y}_{T^*} \pm d$. The following asymptotic result can be derived from (1.2) [see, e.g. Woodroofe (1977), Theorem 4.1].

THEOREM 3.1.   *Suppose that $m \geq 3$, then, as $n_0 = (z\sigma/d)^2 \to \infty$,*

$$E(T^*) = n_0 + \nu + 2\Delta - 2 + o(1),$$

where $\nu = 3/2 - \sum_{n=1}^{\infty} n^{-1} E\{(\chi_n^2 - 2n)^+\} \approx 0.817$. Moreover, if $m \geq 6$, then

$$(3.1) \qquad P\{\theta \in I_{T^*}\} = \gamma + n_0^{-1}\{z^2\psi'(z^2)(\nu + 2\Delta - 2) + z^4\psi''(z^2)\} + o(n_0^{-1})$$

as $n_0 \to \infty$, where $\psi(x) = 2\Phi(\sqrt{x}) - 1$, and ψ' and ψ'' denote the first and second order derivatives of ψ, respectively.

The new procedure operates as follows. Fix the values of the integer $k(\geq 1)$ and the constants $0 < \rho_1 < \cdots < \rho_k < 1$. Take a first sample of size $m + 1$, and take the next k samples sequentially with the ith, $i = 2, \ldots, k + 1$, sample having size $M_{i-1} - M_{i-2}$, where

$$M_j = \max\{\langle \rho_j(z/d)^2\sigma_{M_{j-1}}^2 \rangle + 1, M_{j-1}\}, \qquad 1 \leq j \leq k,$$

and $M_0 = m + 1$. Then, continue sampling until

$$T = \inf\{n \geq M_k \colon n > l_n(z\sigma_n/d)^2\}.$$

On stopping sampling, construct confidence interval $I_T \equiv \bar{Y}_T \pm d$ for θ. Denote this procedure by $\mathscr{P}(k, \rho_1, \ldots, \rho_k)$.

The stopping time T may be written in the form $t_m + 1$ by applying the Helmert transformation in the usual way [see, e.g., Woodroofe (1977), (3.2)], where

$$N_0 = m,$$
$$N_j = \max\{\langle \rho_j(z\sigma/d)^2 \bar{X}_{N_{j-1}} \rangle, N_{j-1}\}, \qquad 1 \leq j \leq k,$$
$$t_m = \inf\{n \geq N_k \colon S_n < (z\sigma/d)^{-2} n^2 L(n)\},$$

where $X_1, X_2, \ldots$ are i.i.d. χ_1^2 random variables and $L(n) = 1 + (1 - 2\Delta)/n + o(1/n)$. So, t_m is of form (2.1) with $\alpha = 2$, $\beta = 1$, $\mu = 1$, $\tau^2 = 2$, $c = (z\sigma/d)^{-2}$ and $\lambda = 1/c = (z\sigma/d)^2 = n_0$. Now from Theorem 2.1 we have the following theorem.

THEOREM 3.2. *Suppose that*

$$(3.2) \quad m \to \infty, \qquad n_0 = n_0(m) \to \infty, \qquad n_0 = O(m^r), \qquad \limsup \frac{m}{n_0} < \rho_1,$$

where $r \geq 1$ is a constant. Then

$$
\begin{aligned}
(3.3) \qquad & E(T) = n_0 + \nu + 2\Delta - 2 + o(1), \\
& P\{\theta \in I_T\} = \gamma + n_0^{-1}\{z^2 \psi'(z^2)(\nu + 2\Delta - 2) + z^4 \psi''(z^2)\} + o(n_0^{-1}).
\end{aligned}
$$

To investigate the suitable choice of $(k, \rho_1, \ldots, \rho_k)$, a series of Monte Carlo trials have been conducted, in which we set $l_n = 1 + 2\Delta/n$ and choose $2\Delta = 2 - \nu + (1 + z^2)/2$ so that the confidence levels of both the new and ACR procedures are equal to $\gamma + o(n_0^{-1})$. Varying γ from 0.90 to 0.99 leads to similar results. So we shall report in detail only for $\gamma = 0.95$, which implies that $z = 1.96$ and $2\Delta = 3.604$. It is noteworthy that both the new and ACR procedures depend on d and σ only through $d/\sigma = z/\sqrt{n_0}$, and so the results are presented in terms of n_0 instead of d and σ^2. The initial sample size has been set at $m + 1 = 10, 20$, and a wide range of values of n_0 has been used.

From the simulation results, it becomes clear that the following choices of $(k, \rho_1, \ldots, \rho_k)$ work well: $(1, 0.5), (2, 0.5, 0.8), (3, 0.5, 0.8, 0.975)$. Table 1 presents the simulation results of these three new procedures, $\mathscr{P}_1 \equiv \mathscr{P}(1, 0.5)$, $\mathscr{P}_2 \equiv \mathscr{P}(2, 0.5, 0.8)$ and $\mathscr{P}_3 \equiv \mathscr{P}(3, 0.5, 0.8, 0.975)$, and the ACR procedure $\mathscr{P}_0$, with each entry based on 10,000 trials. For each procedure we computed the

TABLE 1
Results of 10,000 *Monte Carlo trials with* $\gamma = 0.95$ *and* $2\Delta = 3.604$

		$m + 1 = 10$				$m + 1 = 20$			
n_0	$\mathscr{P}$	$\bar{M}$	$\bar{M} - n_0$	s_M	p	$\bar{M}$	$\bar{M} - n_0$	s_M	p
24	$\mathscr{P}_0$	25.8	1.8	7.2	0.943	27.0	3.0	5.7	0.959
	$\mathscr{P}_1$	25.9	1.9	7.2	0.945	27.0	3.0	5.7	0.957
	$\mathscr{P}_2$	25.8	1.8	7.3	0.944	26.9	3.0	5.7	0.960
	$\mathscr{P}_3$	25.8	1.8	7.1	0.955	26.8	2.8	5.6	0.963
61	$\mathscr{P}_0$	63.0	2.0	11.7	0.950	63.0	2.0	11.5	0.948
	$\mathscr{P}_1$	63.2	2.2	11.8	0.947	63.1	2.1	11.4	0.947
	$\mathscr{P}_2$	63.0	2.0	11.8	0.949	62.9	1.9	11.6	0.950
	$\mathscr{P}_3$	63.5	2.5	10.9	0.963	63.4	2.4	10.6	0.965
125	$\mathscr{P}_0$	127.4	2.4	15.9	0.950	127.5	2.5	16.0	0.952
	$\mathscr{P}_1$	127.8	2.8	16.5	0.950	127.4	2.4	16.0	0.948
	$\mathscr{P}_2$	127.5	2.5	17.0	0.949	127.0	2.0	16.2	0.950
	$\mathscr{P}_3$	128.1	3.1	15.5	0.962	127.6	2.6	14.8	0.965
384	$\mathscr{P}_0$	386.5	2.5	27.8	0.948	386.4	2.4	27.8	0.950
	$\mathscr{P}_1$	388.1	4.1	32.0	0.951	386.4	2.4	27.8	0.951
	$\mathscr{P}_2$	387.7	3.7	31.1	0.950	386.1	2.1	27.9	0.952
	$\mathscr{P}_3$	388.6	4.6	30.5	0.964	386.7	2.7	25.9	0.963

average total sample size $\bar{M}$, the standard deviation of the total sample size s_M and the proportion of times p that θ is covered by the confidence intervals.

From Table 1, it can be seen that all the confidence levels are close to the target value $\gamma = 0.95$, and the expected total sample sizes of the four procedures are hardly different. The four procedures require significantly different numbers of sampling operations, however. For example, when $m+1 = 10$ and $n_0 = 125$, the average number of sampling operations is about 119 for $\mathscr{P}_0$, 67 for $\mathscr{P}_1$, 33 for $\mathscr{P}_2$ and 11 for $\mathscr{P}_3$. So $\mathscr{P}_3$ needs only about 9% of the number of sampling operations of the ACR procedure; $\mathscr{P}_3$, as well as $\mathscr{P}_1$ and $\mathscr{P}_2$, is more advantageous than the ACR procedure when n_0 is large. Asymptotically, the ACR procedure requires on average $n_0 + \nu + 2\Delta - 2 - m + o(1)$ sampling operations while $\mathscr{P}(k, \rho_1, \ldots, \rho_k)$ requires on average $(1-\rho_k)n_0 + \nu + 2\Delta + 2\rho_k/\rho_{k-1} - 3/2 + k + o(1)$ sampling operations.

Finally, the problem of constructing fixed-width confidence intervals serves only to demonstrate the idea, which can be used to deal with many other problems, for example, the sequential point estimation [see, e.g., Woodroofe (1977)], hypotheses testing [see, e.g., Liu (1997)], ranking and selection [see, e.g., Mukhopadhyay and Solanky (1994)] and simultaneous confidence intervals [see, e.g., Liu (1995)].

4. Proofs. Limit operation (2.2) is assumed unless otherwise stated, and C and C_1 denote some generic constants. We first have the following lemma.

LEMMA 4.1. *If $E|X_1|^\xi < \infty$ for some $\xi > 2$ and (2.3) holds, then for $0 < \gamma < 1$ we have $P\{t^* < N_k\} = O(\lambda^{-\min(\gamma, \, 1/r)\xi/2})$.*

PROOF. Letting $\delta > 0$, it is straightforward to show that $P\{|\bar{X}_{N_0} - \mu| > \delta\} = O(m^{-\xi/2})$ by using the Markov inequality and Von Bahr's theorem [Von Bahr, (1965)]. This then implies $P\{|N_1 - \rho_1\lambda| > \delta\rho_1\lambda\} = O(m^{-\xi/2})$. These two results together imply $P\{|\bar{X}_{N_1} - \mu| > \delta\} = O(m^{-\xi/2})$ and, then, $P\{|N_2 - \rho_2\lambda| > \delta\rho_2\lambda\} = O(m^{-\xi/2})$. Continuing in this way, we can show that $P\{|N_k - \rho_k\lambda| > \delta\rho_k\lambda\} = O(m^{-\xi/2})$ and, hence, $P\{N_k > (1+\delta)\rho_k\lambda\} = O(m^{-\xi/2}) = O(\lambda^{-\xi/(2r)})$.

From Lemma 2.3 of Woodroofe (1977), we have, for $0 < \delta, \gamma < 1$, $P\{t^* \leq \delta\lambda\} = O(\lambda^{-\gamma\xi/2})$. Combining the two results above gives

$$P\{t^* < N_k\} \leq P\{t^* \leq \delta\lambda\} + P\{\delta\lambda < N_k\}$$
$$= O(\lambda^{-\gamma\xi/2}) + O(\lambda^{-\xi/(2r)})$$
$$= O(\lambda^{-\min(\gamma, \, 1/r)\xi/2}), \qquad \rho_k < \delta < 1,$$

as required. □

LEMMA 4.2. *If $E|X_1|^\xi < \infty$ for some $\xi > 2$, then $\{(t_m/\lambda)^{\xi(\alpha-1)}\}$ are dominated.*

PROOF.   First note that $\sup_{n \geq 1}(S_n/n)^\xi$ is integrable [see, e.g., Woodroofe (1977), Lemma 2.1]. From the definition of N_i,

$$N_k/\lambda \leq \rho_k(\bar{X}_{N_{k-1}}/\mu)^\beta + \cdots + \rho_1(\bar{X}_{N_0}/\mu)^\beta + N_0/\lambda$$
$$\leq C_0 \sup_{n \geq 1}(S_n/n)^\beta + C_1$$

for some constants C_0 and C_1, and so $\{(N_k/\lambda)^{\xi(\alpha-1)}\}$ are dominated. The required result now follows directly from

$$(t_m/\lambda)^{\xi(\alpha-1)} = (t_m/\lambda)^{\xi(\alpha-1)}(I_{\{t_m = N_k\}} + I_{\{t_m > N_k\}})$$
$$\leq (N_k/\lambda)^{\xi(\alpha-1)} I_{\{t_m = N_k\}} + C_0 \sup_{n \geq 1}(S_n/n)^\xi I_{\{t_m > N_k\}},$$

since, on $\{t_m > N_k\}$,

$$(t_m/\lambda)^{\alpha-1} = c\mu^{-1}(t_m)^{\alpha-1} \leq C_0 c(t_m - 1)^{\alpha-1} L(t_m - 1)$$
$$\leq C_0 S_{t_m-1}/(t_m - 1) \leq C_0 \sup_{n \geq 1}(S_n/n). \qquad \square$$

PROOF OF THEOREM 2.1.   Observe that (1.2) still holds under the assumptions of Theorem 2.1 and limit operation (2.2). So the theorem follows directly from

$$0 \leq Et_m - Et^* = E(t_m - t^*)I_{\{t^* < N_k\}} \leq Et_m I_{\{t^* < N_k\}}$$
$$\leq \left(E(t_m/\lambda)^{\xi(\alpha-1)}\right)^{1/[\xi(\alpha-1)]} \lambda \left(P\{t^* < N_k\}\right)^{1-1/[\xi(\alpha-1)]}$$
$$= O(\lambda^{1-\min(\gamma,1/r)\{1-1/[\xi(\alpha-1)]\}\xi/2}) = o(1). \qquad \square$$

PROOF OF (3.3).   Observe that (3.1) still holds under limit operation (3.2), and so (3.3) follows by noting that

$$|P\{\theta \in I_T\} - P\{\theta \in I_{T^*}\}| \leq E|\psi(z^2 n_0^{-1} T) - \psi(z^2 n_0^{-1} T^*)|$$
$$= E|\psi(z^2 n_0^{-1} T) - \psi(z^2 n_0^{-1} T^*)|I_{\{T^* < N_k + 1\}}$$
$$\leq C_0 P\{T^* < N_k + 1\} = o(n_0^{-1})$$

where the last equality is due to Lemma 4.1. $\square$

REFERENCES

ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **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.

HALL, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* **9** 1229–1238.

LIU, W. (1995). Fixed-width simultaneous confidence intervals for all pairwise comparisons. *Comput. Statist. Data Anal.* **20** 35–44.

LIU, W. (1997). On some sample size formulae for controlling both size and power in clinical trials. *J. Roy. Statist. Soc. Ser. D Statistician* **46** 239–251.

MUKHOPADHYAY, N. and SOLANKY, T. K. S. (1994). *Multistage Selection and Ranking Procedures: Second-Order Asymptotics*. Dekker, New York.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics* (Harold Cramér Volume) 235–245. Almquist and Wiksell, Stockholm.

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

VON BAHR, B. (1965). On convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36** 808–818.

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

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHAMPTON
SOUTHAMPTON, SO17 1BJ
ENGLAND
E-MAIL: wl@maths.soton.ac.uk