Z. Wahrscheinlichkeitstheorie verw. Gebiete
70, 411–416 (1985)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Comparison of Some Sequential Procedures With Related Optimal Stopping Rules*

Adam T. Martinsek
University of Illinois at Urbana-Champaign, Department of Mathematics, Urbana, IL 61801, USA

Summary. For the problem of estimating the mean of independent, identically distributed random variables, with loss equal to a linear combination of squared error and sample size, certain sequential procedures have been shown to be asymptotically optimal when compared with the best fixed sample size rule. In this paper it is shown that these procedures are asymptotically suboptimal when compared with a closely related optimal stopping rule.

1. Introduction

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) with finite mean μ and variance $\sigma^2 \in (0, \infty)$. Suppose one wishes to estimate μ by the sample mean $\bar{X}_n = n^{-1} \sum_1^n X_i$, based on a sample of size n, with the following loss structure. If one stops with n observations and estimates μ by $\bar{X}_n$, the loss incurred is

$$L_n = A(\bar{X}_n - \mu)^2 + n, \qquad A > 0. \tag{1.1}$$

The object is to minimize the risk in estimation by choosing an appropriate sample size.

If σ is known, and a fixed sample size n is used, the risk

$$R_n = E(L_n) = A\sigma^2 n^{-1} + n \tag{1.2}$$

is minimized by the optimal fixed sample size

$$n_0 \simeq A^{1/2} \sigma \tag{1.3}$$

(that is, $[A^{1/2}\sigma] \leqq n_0 \leqq [A^{1/2}\sigma] + 1$), with corresponding minimum risk

$$R_{n_0} \simeq 2A^{1/2}\sigma. \tag{1.4}$$

---

* A preliminary version of this paper was presented at the International Meeting on Sample Survey Analysis and on Sequential Analysis, Jerusalem, Israel, June 1982

However, if σ is unknown the optimal fixed sample size n_0 cannot be used, and there is no fixed sample size procedure that will achieve the risk R_{n_0}.

For the case of unknown σ, Robbins (1959) proposed the following type of sequential procedure:
let

$$T_A=\inf\left\{n\geqq n_A: n\geqq A^{1/2}\left[n^{-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right]^{1/2}\right\}$$
$$=\inf\left\{n\geqq n_A: n^{-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\leqq A^{-1}n^2\right\}, \tag{1.5}$$

where n_A is a positive integer that may depend on A, and estimate μ by $\bar{X}_{T_A}$. A considerable amount of work has been done on proving the asymptotic risk efficiency of such procedures (i.e., $R_{T_A}/R_{n_0}\to 1$ as $A\to\infty$), establishing bounded regret ($R_{T_A}-2A^{1/2}\sigma=O(1)$ as $A\to\infty$), and obtaining second order approximations to the risk R_{T_A}. For references, see Martinsek (1983). In particular, Woodroofe (1977) showed that for normal X_i, if n_A is a fixed integer greater than or equal to 4,

$$R_{T_A}-2A^{1/2}\sigma=\tfrac{1}{2}+o(1) \quad \text{as} \quad A\to\infty. \tag{1.6}$$

Woodroofe's result has been generalized recently in Martinsek (1983), where it is shown that if $E|X_1|^{8p}<\infty$ for some $p>1$, the X_i are non-lattice, and $\delta A^{1/4}\leqq n_A=o(A^{1/2})$ as $A\to\infty$ for some $\delta>0$, then

$$R_{T_A}-2A^{1/2}\sigma=2-(3/4)\operatorname{Var}(Z_1^2)+2E^2(Z_1^3)+o(1) \tag{1.7}$$

as $A\to\infty$, where $Z_1=(X_1-\mu)/\sigma$ (related results are also given for the lattice case).

One interesting feature of the expansion (1.7) is that for some distributions the non-vanishing term on the right hand side is negative; in fact, it is shown in Martinsek (1983) that this non-vanishing term can take arbitrarily large negative values as the distribution of the X_i varies. It follows that in some cases the "regret" due to using the sequential procedure with stopping rule T_A in ignorance of σ (rather than the best fixed sample size n_0 when σ is known) is actually negative. In particular, n_0 is not the optimal stopping rule which minimizes

$$AE(\bar{X}_\tau-\mu)^2+E\tau \tag{1.8}$$

over all stopping rules for the sequence $X_1, X_2, \ldots$ (that such an optimal rule τ_0 exists for each $A>0$ follows from Theorem 4.5 of Chow et al., 1971).

This note considers the question of the performance of stopping rules such as T_A, when compared with the optimal stopping rule τ_0 that minimizes (1.8). It will be shown in Sect. 2 that

$$\limsup_{A\to\infty}\{\inf_\tau[AE(\bar{X}_\tau-\mu)^2+E\tau]\}/R_{N_A}$$
$$=\limsup_{A\to\infty}\{AE(\bar{X}_{\tau_0}-\mu)^2+E\tau_0\}/R_{N_A}$$
$$<1, \tag{1.9}$$

i.e., the sequential procedure with stopping rule N_A is asymptotically suboptimal as $A\to\infty$, for every N_A which is asymptotically risk efficient with respect to the best fixed sample size n_0. In particular, this holds for T_A defined by (1.5).

A few remarks are in order about the statistical meaning of this result. First, the optimal stopping rule τ_0, and the infimum in (1.9), are for the class of all stopping rules for the sequence $X_1, X_2, \ldots$, including those for which the distribution of the X_i is known. Indeed, (1.9) will be proved by comparing N_A with a stopping rule that depends on μ. Such a stopping rule is of course not available in the original estimation problem (after all, if μ is known there is no problem), but the question of minimizing (1.8) is a different matter. In fact, one can think of n_0 as minimizing (1.8) over all non-random stopping times, even if everything about the distribution of the X_i is known. The value R_{n_0} is then a standard against which the performance of a sequential estimation procedure can be measured. From this point of view, (1.9) says that any sequential procedure that is asymptotically optimal when compared with the standard R_{n_0}, is asymptotically suboptimal when compared with the standard that results from allowing random sample sizes in the minimization of (1.8). Moreover, for many distributions (e.g., exponential, Poisson, chi-square) knowledge of σ, which is required to determine n_0, is equivalent to knowledge of μ. Thus, even R_{n_0} is not necessarily an achievable risk in the estimation problem, but merely a "yardstick" for measuring the performance of various estimation procedures.

2. Asymptotic Suboptimality of Certain Sequential Estimation Procedures

Theorem. *Let $X_1, X_2, \ldots$ be i.i.d. with finite mean μ and variance $\sigma^2\in(0,\infty)$. If a sequential estimation procedure with stopping rule N_A satisfies*

$$\lim_{A\to\infty} (R_{N_A}/R_{n_0})=1,$$

where

$$R_{N_A}=AE(\bar{X}_{N_A}-\mu)^2+EN_A,$$

then

$$\limsup_{A\to\infty} \{\inf_{\tau} [AE(\bar{X}_\tau-\mu)^2+E\tau]\}/R_{N_A}<1,$$

where the infimum is over all stopping rules for the sequence of X_i's. That is, the sequential procedure with stopping rule N_A is asymptotically suboptimal.

Proof. Because $\lim_{A\to\infty} (R_{N_A}/R_{n_0})=1$, it suffices to show

$$\limsup_{A\to\infty} \{\inf_{\tau} [AE(\bar{X}_\tau-\mu)^2+E\tau]\}/(2A^{1/2}\sigma)<1.$$

For $A>0$ and $S_n=\sum_{i=1}^{n} X_i$, define

$$\tau_A^*=\inf\{n\geqq 1: |S_n-n\mu|\leqq A^{-1/2}n^{3/2}\}. \tag{2.1}$$

From results of Chow and Lai (1975), since

$$\tau_A^* \leqq \sup\{n \geqq 1: |S_n - n\mu| > A^{-1/2} n^{3/2}\} + 1,$$
$$E\tau_A^* < \infty \quad \text{for all} \quad A > 0.$$

By the definition of τ_A^*,

$$\begin{aligned}
&\limsup_{A\to\infty} \{\inf_\tau [AE(\bar{X}_\tau - \mu)^2 + E\tau]\}/(2A^{1/2}\sigma) \\
&\quad \leqq \limsup_{A\to\infty} [AE(\bar{X}_{\tau_A^*} - \mu)^2 + E\tau_A^*]/(2A^{1/2}\sigma) \\
&\quad \leqq \limsup_{A\to\infty} [2E\tau_A^*]/(2A^{1/2}\sigma).
\end{aligned} \tag{2.2}$$

Because $E\tau_A^*$ is nondecreasing in A, if $E\tau_A^* \not\to \infty$ as $A \to \infty$ the theorem follows immediately. We may therefore assume that $E\tau_A^* \to \infty$ as $A \to \infty$. Then, abbreviating τ_A^* as τ^*,

$$\begin{aligned}
A^{-1} E[(\tau^* - 1)^3] &\leqq E[(S_{\tau^*-1} - (\tau^* - 1)\mu)^2] \\
&= E[(S_{\tau^*} - \tau^*\mu - X_{\tau^*} + \mu)^2] \\
&= E[(S_{\tau^*} - \tau^*\mu)^2] - 2E[(X_{\tau^*} - \mu)(S_{\tau^*} - \tau^*\mu)] + E[(X_{\tau^*} - \mu)^2].
\end{aligned} \tag{2.3}$$

By a result of Gundy and Siegmund (1967) (see Chow and Teicher, 1978, p. 148, Lemma 2),

$$E[(X_{\tau^*} - \mu)^2] = o(E\tau^*). \tag{2.4}$$

By the Wald equation for the second moment and (2.4),

$$\begin{aligned}
E[|(X_{\tau^*} - \mu)(S_{\tau^*} - \tau^*\mu)|] &\leqq E^{1/2}(X_{\tau^*} - \mu)^2 \cdot E^{1/2}(S_{\tau^*} - \tau^*\mu)^2 \\
&= o(E^{1/2}\tau^*) \cdot (\sigma^2 E\tau^*)^{1/2} \\
&= o(E\tau^*),
\end{aligned}$$

and hence from (2.3),

$$A^{-1} E[(\tau^* - 1)^3] \leqq \sigma^2 E\tau^* + o(E\tau^*) = \sigma^2 E(\tau^* - 1) + o(E\tau^*).$$

Application of Jensen's inequality yields

$$E^3[\tau^* - 1] \leqq E[(\tau^* - 1)^3] \leqq A\sigma^2 E(\tau^* - 1) + o(AE(\tau^* - 1))$$

and therefore

$$E^2[\tau^* - 1] \leqq A\sigma^2 + o(A)$$

for A sufficiently large.

It follows that

$$\limsup_{A\to\infty} E[\tau^* - 1]/(A^{1/2}\sigma) \leqq 1. \tag{2.5}$$

Also, by Jensen's inequality and (2.5),

$$\begin{aligned}
&\limsup_{A\to\infty} E[(\tau^* - 1)^2]/(A\sigma^2) \leqq \limsup_{A\to\infty} E^{2/3}[(\tau^* - 1)^3]/(A\sigma^2) \\
&\quad \leqq \limsup_{A\to\infty} \{E(\tau^* - 1)/(A^{1/2}\sigma) + o[E(\tau^* - 1)/A^{1/2}]\}^{2/3} \\
&\quad \leqq 1.
\end{aligned} \tag{2.6}$$

Suppose

$$\limsup_{A\to\infty} E(\tau^*-1)/(A^{1/2}\sigma)=1;$$

then there is a subsequence $\{A_k\}$ along which (writing τ^* for $\tau^*_{A_k}$)

$$E(\tau^*-1)/(A_k^{1/2}\sigma)\to 1 \quad \text{as} \quad k\to\infty. \tag{2.7}$$

But then by (2.6), as $k\to\infty$

$$\operatorname{Var}[(\tau^*-1)/(A_k^{1/2}\sigma)]\to 0. \tag{2.8}$$

Combining (2.7) and (2.8),

$$(\tau^*-1)/(A_k^{1/2}\sigma)\xrightarrow{p} 1 \tag{2.9}$$

and Anscombe's theorem (Anscombe, 1952) yields

$$[S_{\tau^*-1}-(\tau^*-1)\mu]/[\sigma(\tau^*-1)^{1/2}]\xrightarrow{d} N(0,1) \quad \text{as} \quad k\to\infty.$$

Therefore

$$\begin{aligned}&P[(\tau^*-1)/(A_k^{1/2}\sigma)<\tfrac{1}{2}]\\ &\quad\geqq P[|[S_{\tau^*-1}-(\tau^*-1)\mu]/[\sigma(\tau^*-1)^{1/2}]|<\tfrac{1}{2}]\to P[|N(0,1)|<\tfrac{1}{2}]>0,\end{aligned}$$

contradicting (2.9). It follows that

$$\limsup_{A\to\infty} E(\tau^*-1)/(A^{1/2}\sigma)<1,$$

which together with (2.2) finishes the proof.

Remark on the Proof. The stopping time τ_A^* used in the proof above is motivated by the idea of "balancing" the two components of the loss function. Just as the optimal fixed sample size n_0 does this in expectation ($AE(\bar{X}_{n_0}-\mu)^2\simeq n_0$), the rule τ_A^* does it pointwise (by the definition of τ_A^*, if the undershoot is small, $A(\bar{X}_{\tau_A^*}-\mu)^2$ is close to τ_A^*). The same idea is used in Bickel and Yahav (1967, 1968) to derive asymptotically pointwise optimal (A.P.O.) and asymptotically optimal rules, and it would be interesting to know whether τ_A^* is asymptotically optimal in the present case. The work of Bickel and Yahav applies to the problem of minimizing

$$AE(Y_\tau)+E\tau=A[E(Y_\tau)+A^{-1}E\tau],$$

over all stopping times for a sequence of random variables $Y_1, Y_2, \ldots,$ when $n^\beta Y_n$ converges almost surely to a positive random variable for some $\beta>0$. The loss function (1.1) is of the right form, however in this (non-Bayesian) case $Y_n=(\bar{X}_n-\mu)^2$ and nY_n converges in distribution (to $\sigma^2\chi_1^2$). Hence $n^\beta Y_n$ does not converge a.s. to a positive random variable for any $\beta>0$, and the results of Bickel and Yahav do not apply.

In addition to showing that the stopping rule T_A defined by (1.5) is asymptotically suboptimal with respect to the optimal stopping rule τ_0, the Theorem shows that the same is true for any stopping rule which is asymptotically risk efficient with respect to n_0. The result therefore applies to the stopping rules of Starr and Woodroofe (1972) and Vardi (1979) for the gamma and Poisson cases (respectively), among others.

Finally, it should be mentioned again that the stopping rule τ_A^* depends on μ and so is unavailable in the original estimation problem. Asymptotic suboptimality of the sequential estimation procedures is therefore not such a serious failing.

References

Anscombe, F.: Large sample theory of sequential estimation. Proc. Cambridge Philos. Soc. **48**, 600–607 (1952)

Bickel, P., Yahav, J.: Asymptotically pointwise optimal procedures in sequential analysis. Proc. Fifth Berkeley Sympos. Math. Stat. Probab. **1**, 401–413 (1967)

Bickel, P., Yahav, J.: Asymptotically optimal Bayes and minimax procedures in sequential estimation. Ann. Math. Stat. **39**, 442–456 (1968)

Chow, Y.S., Lai, T.L.: Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. Trans. Am. Math. Soc. **208**, 51–72 (1975)

Chow, Y.S., Robbins, H., Siegmund, D.: Great Expectations: The Theory of Optimal Stopping. Boston: Houghton Mifflin Company 1971

Chow, Y.S., Teicher, H.: Probability Theory: Independence, Interchangeability, Martingales. New York: Springer 1978

Gundy, R., Siegmund, D.: On a stopping rule and the central limit theorem. Ann. Math. Stat. **38**, 1915–1917 (1967)

Martinsek, A.: Second order approximation to the risk of a sequential procedure. Ann. Stat. **11**, 827–836 (1983)

Robbins, H.: Sequential estimation of the mean of a normal population. Probability and Statistics (Harald Cramer Volume). Uppsala, Sweden: Almquist and Wiksell 1959

Robbins, H., Siegmund, D.: Boundary crossing probabilities for the Wiener process and sample sums. Ann. Math. Stat. **41**, 1410–1429 (1970)

Robbins, H., Siegmund, D., Wendel, J.: The limiting distribution of the last time $S_n \geqq n\varepsilon$. Proc. Natl. Acad. Sci. USA **61**, 1228–1230 (1968)

Starr, N., Woodroofe, M.: Further remarks on sequential point estimation: the exponential case. Ann. Math. Stat. **43**, 1147–1154 (1972)

Vardi, Y.: Asymptotic optimal sequential estimation: the Poisson case. Ann. Stat. **7**, 1041–1051 (1979)

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

Received March 1, 1983