SEQUENTIAL ANALYSIS
Vol. 23, No. 1, pp. 11–32, 2004

Principles of Optimal Sequential Planning

Claudia Schmegner[1] and Michael I. Baron[2,3,*]

[1]Department of Mathematical Sciences, DePaul University, Chicago, Illinois, USA
[2]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas, USA
[3]IBM Research Division, Yorktown Heights, New York, USA

ABSTRACT

A concept of *sequential planning* is presented as extension and generalization of "pure" sequential procedures. According to it, observations are collected in groups of variable sizes. The article discusses optimality of sequential plans in terms of a suitable risk function that balances an observation cost and a group cost. It is shown that only non-randomized sequential plans based on a sufficient statistic need to be considered in order to achieve optimality. Performance of several classes of plans is evaluated.

Key Words: Cost; Risk; Sequential plan; Sequentially planned probability ratio test.

*Correspondence: Michael I. Baron, IBM T. J. Watson Research Center, Department of Mathematical Sciences, P.O. Box 218, 32-243 Yorktown Heights, NY 10598, USA; E-mail: mbaron@utdallas.edu.

DOI: 10.1081/SQA-120030192 0747-4946 (Print); 1532-4176 (Online)

1. INTRODUCTION

The main principle of classical sequential sampling is to collect one observation at a time. After each observation is collected, the experimenter decides whether to make a decision immediately (and stop data collection) or to postpone it until the next observation is obtained. One may prefer sequential (on-line) schemes to retrospective (off-line) procedures in order to reduce the cost of collected observations by not sampling unnecessary data. Also, a number of statistical problems in quality control, clinical trials, change-point detection, target tracking, etc., allow only a sequential solution (see e.g., Ref.[13]).

At the same time, classical sequential sampling is found impractical in applications where it is cheaper to take a batch of n observations than to collect these observations one by one (see Sec. 7.8 of Ref.;[3] Sec. 2 of Refs.;[5,9,13] Sec. 1.6 of Refs.;[15,21] Sec. 8 of Ref.[23]). This is often caused by the delay between treatments and responses, as in many medical studies,[12] that makes standard sequential sampling schemes highly time consuming.

As another example, measuring surface reflectance by means of a bidirectional reflectance distribution function (BRDF) sequentially is time consuming because it requires physical movement of equipment for every BRDF sample. Therefore, it is proposed to speed up the process by measuring multiple points simultaneously.[16]

For the same reason, monitoring seed banks is only reasonable by testing groups of at least 40 seeds.[24]

In the mentioned applications, one cannot neglect a considerable fixed cost of each measuring stage comparing with the cost of an individual observation. In other words, the linear cost function, traditionally considered in classical sequential analysis, does not reflect the actual costs even approximately in a number of applications. In addition to the observation cost, one should also take into account the cost of each batch or each sampling stage (See Remark 3.12.1 of Ref.[7]). This additional term may correspond to the cost of equipment and personnel required to conduct each sampling stage, the time between stages, the number of times the decision of stopping or continuing sampling is made, and other *fixed* costs that are roughly independent of the number of individual observations in each stage.

Under these costs, one would agree to collect a few additional, possibly unnecessary observations in exchange for the convenience of sampling in batches and significant reduction of the overall time and cost of experiment. The main problem is then to design a sequential experiment in the most economic way by choosing optimal batch sizes.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

The batches are sampled sequentially, therefore, at any moment, all the collected data can be used by practitioners to determine an optimal size of the next batch. The number of batches can also be determined sequentially. Thus, after a batch is sampled, an experimenter has a choice of terminating sampling and making a decision, or sampling one more observation, or sampling two more observations, etc.

In this most general setting, such sequential sampling schemes were introduced by Schmitz[21] and called *sequentially planned procedures*. This monograph extended the classical Wald's sequential probability ratio test (SPRT) to a sequentially planned probability ratio test (SPPRT) and proved existence of Bayes sequentially planned schemes in various general situations. A continuous-time version was treated in Ref.[19]. Application of the developed optimal procedures is still difficult because of the absence of explicit formulae for determining the batch sizes sequentially.

Our goal is to develop general principles and guidelines that can be used to optimize sequential planning and to improve the performance of existing sequentially planned schemes (here and later–*sequential plans*).

Obviously, the classical "pure" sequential procedures constitute a very special case of sequential plans where each batch has size 1. Thus, one will never lose in terms of minimizing the risk function by considering sequential plans in place of pure sequential sampling. Rather, one can warrant some definite profit by implementing the following simple scheme that we will call a *conservative plan*.

At every sampling stage, let q_n be the probability that the next batch of size n forces the experimenter to cease sampling (say, by exceeding the stopping boundaries). Then, instead of sampling one observation, let us take a batch of size $N = \min\{n \geq 0 | q_n > 0\}$. Here $N = 0$ if sampling is already terminated. This plan is conservative in a sense that it always results in exactly the same total number of observations as the pure sequential plan. Thus, it never leads to a higher risk, and it improves the pure sequential plan in all the cases where $q_1 = 0$ is possible. The latter occurs in SPRT with bounded likelihood ratios, binomial and inverse binomial sampling schemes (see Sec. 1.2 of Ref.[6]), and other situations with bounded variables and fixed stopping boundaries. For example, if we decide to reject a shipment of items once 15 defective items are found, or to accept it when 86 non-defective items are found, then it is certainly safe to start sampling by taking a batch of size 15. If, say, this batch contains 3 defective items, then it is equally safe to take the second batch of size 12, etc.

It is clear from this example that a rather simple sequential plan can often dominate the pure sequential scheme uniformly. It is also clear

that sequential plans generalize and unify a number of other popular and commonly used sampling designs,

- *Group sequential sampling* (e.g., in Refs.[11,14,18]), where all group sizes are equal, and the number of sampled groups is determined sequentially by a stopping rule.
- *Multistage procedures* (e.g., Ref.[6]), and Chap. 6 of Ref.,[17] where the group sizes may vary and be chosen based on the observed data, and the number of groups (stages) is fixed and pre-determined.
- *Accelerated sequential procedure* (see Ref.[6] and Sec. 6.4 of Ref.[9]), where a pure sequential sample is followed by one final batch.

Two other sequential plans are proposed in Ref.[10]. One *(expectation-based)* plan chooses each group size to be proportional to the expected number of observations it takes a pure sequential scheme to terminate. The other (*quantile-based*) plan chooses them based on the pre-determined probability to terminate sampling after the next group. Similarly to the introduced conservative plan, both the group sizes and the number of groups are determined sequentially, based on the observed data.

A related scheme is the *dynamic sampling approach*,[1,2] designed for quick change-point detection in a continuous-time process, where one controls the sampling effort at each stage while keeping the average sampling rate at the desired constant.

Obtaining the best optimal sequential plan is not a trivial task because one has to minimize the risk over a vast set of possible plans, including classical sequential and all the discussed procedures as special cases. In this paper, we establish basic principles of optimal sequential planning and apply them to optimize existing sequential plans and to propose the new schemes with enhanced performance. After introducing formal notations and concepts, we show that only non-randomized sequential plans based on a sufficient statistic have to be considered in order to achieve optimality. Application of the established principles significantly reduces the set of plans that should be considered. We then compare some classes of sequential plans according to the asymptotic behavior of their risks. All the proofs are given in Appendix.

2. OPTIMALITY CRITERIA

Let $\boldsymbol{X} = (X_1, X_2, \dots)$ be a stochastic sequence that can be observed according to a chosen sequential plan, and let $\{\mathcal{F}_n, n \geq 0\}$ be the corresponding filtration with $\mathcal{F}_0$ being the trivial sigma-field. One would like to find the optimal way of sampling from $\boldsymbol{X}$, i.e., the optimal sequential plan. Suppose that the first k observations are used to determine the size of the jth sample, N_j. This makes N_j and $\mathcal{F}_k$-measurable integer-valued random variable, i.e.,

$$N_j = N_j(\boldsymbol{X}^k), \quad j = 1, 2, \dots,$$

where $\boldsymbol{X}^k = (X_1, \dots, X_k)$.

Thus, a *non-randomized sequential plan* is a family of integer-valued functions

$$N = \{N^{(k)} : \boldsymbol{X}^k \to \{0, 1, 2, \dots\}\}, \tag{1}$$

that return the size of the next batch given a sample of size k.

A *randomized sequential plan* is then a family of distributions with a non-negative integer-valued support

$$N = \left\{\mathcal{P}^k : \boldsymbol{X}^k \to \{p_0, p_1, p_2, \dots\}, \ \sum p_j = 1\right\} \tag{2}$$

that return a distribution $\mathcal{P}^{(k)}(\boldsymbol{X}^k)$ (a sequence of probabilities $\{p_j\}$) with P_j being the probability that the next batch size is j.

During sequential sampling, a sample of size $M_{j-1} = \sum_{i=1}^{j-1} N_i$ is available at the time when N_j, the size of the next sample, is chosen (with a convention that $M_0 = 0$ and $\boldsymbol{X}^0 = \emptyset$).

Sampling stops as soon as $N_j = 0$ for some j. The total number of sampled groups

$$T = \inf\{j \geq 1 | N_j = 0\} - 1$$

then plays a role of a "stopping time" (it is time indeed if it takes exactly one unit of time to collect one batch). In general, $0 \leq T \leq \infty$, that is, it may be proper or improper.

In these notations, classical sampling schemes are described as follows,

(a) $T = 1$ for retrospective (non-sequential) sampling;
(b) $T = const > 1$ for multistage sampling;

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

(c) $N_1 = \cdots = N_T$ for group sequential sampling;
(d) $N_1 = \cdots = N_T = 1$ for pure sequential sampling;
(e) $N_1 = \cdots = N_{T-1} = 1$ or, more generally, $N_k = \cdots = N_{T-1} = 1$ for accelerated sequential sampling; Also,
(f) $N^{(k)}(\boldsymbol{X}^k) \approx \gamma \boldsymbol{E} \min_n\{n|N^{(k+n)}(\boldsymbol{X}^{k+n}) = 0\}$ for the expectation-based plan;
(g) $N^{(k)}(\boldsymbol{X}^k) \in \{n|P\{N^{(k+n)}(\boldsymbol{X}^{k+n}) = 0|\boldsymbol{X}^k\} \approx \gamma\}$ for the quantile-based plan, and
(h) $N^{(k)}(\boldsymbol{X}^k) = \min_n\{n|P\{N^{(k+n)}(\boldsymbol{X}^{k+n}) = 0|\boldsymbol{X}^k\} > 0\}$ for the introduced conservative plan.

Selection of batch sizes in (f) and (g) depends on the distribution of $\boldsymbol{X}$, which is usually unknown. It is natural to implement these plans assuming the null distribution in testing problems, or some estimator of the unknown parameter. Implementation of plan (h) is parameter-free, as long as all the considered distributions have a common support.

We consider decision theoretic aspects of sequential planning. Thus, we assume a loss function $L(\theta, \delta)$ of a parameter (state of nature) $\theta \in \Theta$ and a terminal decision $\delta \in \mathcal{A}$ and a cost function

$$C(N) = cM_T + aT = \sum_1^T (cN_j + a),$$

where c is the cost of one observation and a is the additional cost of one batch. As before, T is the number of sampled batches and $M_T = M(N) = \sum_1^T N_j$ is the total sample size. Thus, $C(N)$ consists of two parts representing the variable and the fixed cost of experiments. The variable cost is proportional to the number of sampled observations, as in classical sequential analysis. The fixed cost aT forces an experimenter to search for an optimal sequential plan. On the extremes, sampling a large number of small groups would lead to high fixed costs while selecting a small number of large groups is likely to cause a large overshoot (in a sense of many unnecessary observations) and to increase the variable cost. Clearly, the former scheme may be optimal only when the *cost ratio* a/c is low whereas the latter scheme may be optimal when a/c is high.

Then, different sequential plans can be evaluated and compared according to their risk function

$$R(\theta, N, \delta) = \boldsymbol{E}^{\boldsymbol{X}}\{L(\theta, \delta) + cM_T + aT\} \tag{3}$$

and their Bayes risk

$$r(\pi, N, \delta) = \boldsymbol{E}^{\pi} R(\theta, N, \delta),$$

if θ has a prior distribution $\pi(\theta)$. The risk of a randomized plan N with a family of distributions $\{\mathcal{P}\}$ is naturally defined as

$$R(\theta, N, \delta) = \boldsymbol{E}^{\boldsymbol{X}}\boldsymbol{E}^{\{\mathcal{P}\}}\{L(\theta, \delta) + C(N)\}.$$

As in classical sequential analysis, the optimal terminal decision rule is the classical Bayes procedure (see Theorem 4.4. of Ref.[21]). Thus, we will only consider the problem of optimal sampling, i.e., the optimal choice of $N^{(k)}(\boldsymbol{X}^k)$, and will often omit δ as an argument of L, R, or r, assuming that the Bayes rule δ^B is used.

Classical decision theoretic concepts are easily translated into the language of sequential plans. A plan N is called

- *R-better* than a plan $\overline{N}$ if

$$\begin{cases} R(\theta, N) \le R(\theta, \overline{N}) & \text{for any } \theta \in \Theta, \\ R(\theta, N) < R(\theta, \overline{N}) & \text{for some } \theta \in \Theta. \end{cases}$$

- *Admissible* if no plan is R-better than N;
- *Minimax* if $\sup_\theta R(\theta, N) \le \sup_\theta R(\theta, \overline{N})$ for any plan $\overline{N}$;
- *Preferred* to a plan $\overline{N}$ if $r(\pi, N) < r(\pi, \overline{N})$;
- *Bayes* if no plan is preferred to N.

In general, computing a Bayes plan is not trivial. However, in a very special case when the minimum posterior risks at each sampling stage are independent of the collected sample, a simple fixed-sample plan (where $T = 1$ a.s.) is Bayes. The following theorem extends the analogous result from classical sequential analysis (see Theorem 5.3.1 of Ref.[6]) requiring a weaker assumption about the posterior risk ρ_n. A problem will be called *truncated at* $\overline{M}$ if the number of potentially available observations is limited by $\overline{M} < \infty$, and thus only the plans with $M(T) < \overline{M}$ are permitted. The following result is analogous to Theorem 5.3.1 of Ref.[6].

Theorem 1 (Fixed sample plans)**.** *Let*

$$\rho_n(\boldsymbol{X}^n) = \inf_\delta \boldsymbol{E}^{\pi(\theta)}\{L(\theta, \delta) \mid \boldsymbol{X}^n\}$$

be the minimum posterior expected loss based on a sample of size n.

(i) *If $\rho_n(\boldsymbol{X}_n)$ is independent of $\boldsymbol{X}_n$ for any n, then the optimal fixed-sample plan is a Bayes sequential plan for the truncated problem.*

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

(ii) *If in addition, $\rho_n < c$ for some n, then the optimal fixed-sample plan is Bayes for the general problem.*

Thus, if conditions of Theorem 1 are met, the search of optimal sequential plans is restricted to fixed sample plans. Then, naturally, the size of the Bayes fixed sample plan is determined by minimizing the risk

$$r(\pi, n) = \boldsymbol{E}^{\pi}\boldsymbol{E}\{L(\theta, \delta(\boldsymbol{X}^n)) + cn + aT\} = \begin{cases} \rho_n + cn + a & \text{if } n > 0 \\ \rho_0 & \text{if } n = 0 \end{cases}$$

in n for $n = 0, 1, \ldots, \overline{M}$ in (i) or $n = 0, 1, \ldots$ in (ii).

In fact, conditions of Theorem 1 represent a very special situation because the minimum posterior risk ρ_n is usually a function of data $\boldsymbol{X}^n$. Classical examples when ρ_n is independent of $\boldsymbol{X}^n$ are estimation of the Normal mean under the squared-error loss and estimation of the binomial proportion under the weighted squared-error loss (Examples 5.3.1 and 5.3.2 of Ref.[6]). In general, conditions of Theorem 1 are too strong. In the next Section, we establish two general results that substantially reduce the set of sequential plans where the optimal plan can be found, therefore, simplifying the problem of optimal sequential planning.

3. RANDOMIZATION AND SUFFICIENCY

The next two results show that the search of optimal sequential plans, according to criteria introduced in Sec. 2, should be confined to non-randomized plans that are based on a sufficient statistic.

Theorem 2 (Randomization principle)**.** *Let N be a randomized plan. Then there exists a non-randomized plan whose Bayes risk does not exceed the Bayes risk of N.*

Corollary 1. *There exist non-randomized Bayes plans.*

According to this theorem, the search for the Bayes sequential plan should be restricted to non-randomized plans. Alternatively, searching for a minimax plan, one may need to randomize. The next theorem reduces the space of search for the optimal plan even further.

Let $\{S_n = S_n(\boldsymbol{X}^n), n \geq 1\}$ be a family of sufficient statistics. That is, for each sample size n, we consider some (arbitrary) sufficient statistic $S_n(\boldsymbol{X}^n)$.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Theorem 3 (Sufficiency principle)**.** *For any sequential plan N, there exists a sequential plan N^* such that*

(i) *N^* depends on the collected data only through the sufficient statistic, i.e.,*

$$N^*(\boldsymbol{X}^n) = N^*(n, S_n(\boldsymbol{X}^n));$$

(ii) $r(\pi, N^*) \leq r(\pi, N)$.

Corollary 2. *There exist Bayes sequential plans that are based on the sufficient statistic.*

In particular, since the set of order statistics is sufficient in the i.i.d. case,

Corollary 3. *In optimal sequential planning with i.i.d. observations, the choice of each group size is independent of the order in which the collected data were obtained.*

Corollary 4 (Irrelevance of past costs)**.** *In optimal sequential planning, the choice of each group size is independent of the amount already paid in sampling costs for the collected groups of observations.*

Certainly, a permutation of data may lead to a different selection of group sizes and different sampling costs. The last corollary states that the amount already spent on sampling should not be taken into account when one searches for the optimal continuation of a sequential plan.

The stated randomization and sufficiency principles are rather similar to the general principles in classical Bayesian analysis with convex loss functions (see Sec. 1.8 of Ref.[3]). However, the usual Rao-Blackwell arguments cannot be applied to sequential planning because the expectation of the next group size (with respect to the distribution $\mathscr{P}$ of a randomized plan or the conditional distribution of $\boldsymbol{X}^n$ given S_n) may be non-integer and thus cannot be the size of a group. Also, Theorems 2 and 3 do not require convexity of the loss function.

In the next section, we consider specific plans that satisfy both of the stated principles.

4. CONSERVATIVE SPPRT AND OTHER SEQUENTIAL PLANS

A *sequentially planned probability ratio test* or SPPRT[21] is a sequentially planned extension of the popular Wald's *sequential probability ratio test* (SPRT).

Consider testing a simple hypothesis $H_0 : F = F_0$ against a simple alternative $H_A : F = F_1$ based on a sequence of i.i.d. $X_1, X_2, \ldots$ for some mutually absolutely continuous distributions F_0 and F_1. The test and all the sequentially chosen group sizes are based on log-likelihood ratios

$$\Lambda_n = \sum_{j=1}^{n} z_j = \sum_{j=1}^{n} \log \frac{dF_1(X_j)}{dF_0(X_j)} = \log \frac{dF_1(\boldsymbol{X}^n)}{dF_0(\boldsymbol{X}^n)}.$$

That is, each group size $N_j(n, \boldsymbol{X}^n) = N(\Lambda_n)$ is a function of the current log-likelihood ratio Λ_n. Sampling terminates as soon as Λ_n crosses one of the chosen *stopping boundaries* A or B, $A > 0 > B$. Thus,

$$T = \min\{k | \Lambda_{M_k} \notin (B, A)\}.$$

Remark. Notice that sampling continues if Λ_n crosses a stopping boundary but returns to the interval (B, A) before the batch of observations is collected. This is contrary to the classical SPRT. We can offer two explanations to this. First, such a situation can be interpreted as having a seemingly strong evidence supporting H_0 or H_A in the middle of a batch that does not look that strong once the whole batch is observed. We should be thankful for having collected additional observations because now the test is not conclusive, and we are happy to collect more data. Second, our decision should be based on a sufficient statistic, according to Theorem 2. Since the set of order statistics is sufficient, the scheme should not depend on the order in which observations are collected. With a suitable permutation of the same batch, Λ_n does not cross the boundary, thus, sampling continues.

Some vital properties of SPPRT are established below.

Lemma 1. *SPPRT satisfies the sufficiency principle.*

The error probabilities of SPPRT

$$\alpha = \boldsymbol{P}\{\text{Type I error}\} = \boldsymbol{P}_0\{\Lambda_{M_T} \geq A\}$$

and

$$\beta = \boldsymbol{P}\{\text{Type II error}\} = \boldsymbol{P}_1\{\Lambda_{M_T} \leq B\}$$

are controlled by the stopping boundaries.

Lemma 2 (see Lemma 3.7 of Ref.[21]). *For an SPPRT,*

$$A \leq \log \frac{1-\beta}{\alpha}, \quad B \geq \log \frac{\beta}{1-\alpha}. \tag{4}$$

This result and its proof are similar to the case of SPRT. We note, however, that inequalities (4) are not as sharp as in the case of SPRT due to a possibility of collecting observations in addition to the purely sequential procedure. Since SPPRT collects more observations, in general, than the SPRT with the same stopping boundaries, the actual error probabilities of SPPRT are generally lower.

Corollary 5. *For an SPPRT,*

$$\alpha \le e^{-A}, \quad \beta \le e^{B}.$$

Only *conservative SPPRT* is guaranteed to collect the same number of observations and result in the same decision as the corresponding SPRT.

Conservative SPPRT can be used when log-likelihood ratios z_j are bounded a.s. Then, if both boundaries are "far" from the current value of Λ_n, it is possible to collect a batch of observations that does not use more data than the corresponding SPRT.

Suppose that

$$P\{l < z_j < u\} = 1 \tag{5}$$

under F_0 and F_1. Then, after $\boldsymbol{X}^n$ is observed, let

$$N(\boldsymbol{X}^n) = N(\Lambda_n) = \min\left\{ \left\lceil \frac{A - \Lambda_n}{u} \right\rceil^+, \left\lceil \frac{\Lambda_n - B}{l} \right\rceil^+ \right\}$$

be the next group size, where $x^+ = \max\{x, 0\}$ and $\lceil x \rceil = \min\{y \in Z | y \ge x\}$. Clearly, $\Lambda_j \in (B, A)$ for all $j = M_k + 1, \ldots, M_{k+1} - 1$ and only $\Lambda_{M_{k+1}}$ can cross the boundary.

Thus, the total sample size of the conservative SPPRT coincides with that of the corresponding SPRT. It terminates simultaneously with SPRT and results in the same decision (acceptance or rejection of H_0). Hence, the error probabilities of this SPPRT and the SPRT are equal too. The difference in risks (3) is caused therefore by the difference in the number of groups. For the SPRT, the number of groups equals the number of observations, $T = M_T$. Its expectation under $F_1(F_0)$ is *asymptotically* linear in $A(B)$ as $A \to \infty$ ($B \to -\infty$).

Lemma 3. *Let $\mu = \boldsymbol{E}_F(z_j)$. Then there exist limits*

$$\lim_{A \to \infty, A = O(B)} \frac{\boldsymbol{E}_{SPRT}(T|H_1)}{A} = \frac{1}{\mu}$$

and

$$\lim_{B\to-\infty, B=O(A)} \frac{\boldsymbol{E}_{SPRT}(T|H_0)}{|B|} = \frac{1}{|\mu|}.$$

This lemma follows directly from classical results of the renewal theory of random walks (see e.g., Refs.,[8,22] Chap. II).

Contrary to an SPRT, the introduced conservative SPPRT requires sampling an *asymptotically logarithmic* expected number of groups.

Theorem 4. *For a conservative SPPRT with stopping boundaries A and B,*

$$\boldsymbol{E}_{SPPRT}(T|H_1) \le \frac{1+1/\mu}{|\log(1-\mu/u)|}(\log A - \log\log A) + O(1), \tag{6}$$

as $A \to \infty$, $A + B = O(1)$, *and*

$$\boldsymbol{E}_{SPPRT}(T|H_0) \le \frac{1-1/\mu}{|\log(1+\mu/l)|}(\log|B| - \log\log|B|) + O(1),$$

as $B \to -\infty$, $A + B = O(1)$.

According to this theorem, conservative SPPRT provides a major reduction of cost of experiment. This extends to *all reasonable* sequentially planned procedures! Indeed, sampling fewer observations than a conservative plan will necessarily cause to sample at least one more batch. Such a procedure will only involve a larger number of batches and will never be optimal under the risk (3).

Whenever there is a fixed cost a of collecting each batch of observations, the purely sequential SPRT requires an expense proportional to A for sampling cost whereas this amount is at most proportional to $\log A$.

Notice again that the conservative SPPRT uses the same total sample size and results in the same decision as the corresponding SPRT. Thus, the difference in $E(T)$ is the only difference in risk functions (3) for these two schemes. This makes the conservative SPPRT an R-better procedure than SPRT.

Corollary 6. *Under the Condition* (5) *of bounded log-likelihood ratios, any SPRT with* $B < 2l$ *or* $A > 2u$ *is inadmissible.*

Certainly, this does not imply that the conservative SPPRT is an optimal procedure. It is not difficult to design a sequential plan that

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

achieves a *bounded* expected number of sampled batches even if the stopping boundaries diverge to infinity. This can be done at the expense of some increase in the total number of sampled observations.

Theorem 5. *Let N be a sequential plan satisfying*

$$\boldsymbol{P}\{T = k+1 | \boldsymbol{X}^{M_k}, T > k\} \geq p \tag{7}$$

a.s., for some $p > 0$ and all k. Then

(i) *T is proper, i.e., $P(T = \infty) = 0$;*
(ii) *$\boldsymbol{E}(T) \leq 1/p$.*

Condition (7) means that the sequential plan is designed in such a way that any batch has a probability of at least p to be the final batch. A wide range of proposed sequential plans satisfies this condition. For example, it is always satisfied by any quantile-based plan. Under mild additional conditions on z_j, an expectation-based plan also satisfies (7). For example, Theorem 5 is applicable to an expectation-based SPPRT for testing a simple hypothesis about the normal mean.[20]

Implementation of these plans in place of pure sequential procedures significantly reduces sampling costs without sacrificing the accuracy of statistical results.

5. APPENDIX

Proof of Theorem 1. (i) The set of all possible truncated plans is finite, thus the minimum Bayes risk is attained. If $\rho_n(\boldsymbol{X}^n)$ is independent of $\boldsymbol{X}^n$, then

$$\begin{aligned}
\min_{\{N,\delta\}} r(\pi, N, \delta) &= \min_{\{N,\delta\}} \boldsymbol{E}\boldsymbol{E}\{L(\theta, \delta) + cM_T + aT | \theta\} \\
&= \min_{\{N\}} \boldsymbol{E}(\rho_{M_T} + cM_T + aT) \\
&\geq \boldsymbol{E} \min_{\{N\}} (\rho_{M_T} + cM_T + aT) \\
&= \min\left\{\rho_0, \min_{n=1,2\ldots} (\rho_n + cn + a)\right\},
\end{aligned}$$

which is the Bayes risk of the optimal fixed sample plan.

(ii) Suppose that $\rho_{n_0} < c$. For any given sequential plan N, define a plan N^*,

$$N^*(\boldsymbol{X}^n) = \min\{N(\boldsymbol{X}^n), n_0 - n\}.$$

Clearly, N^* is truncated at n_0. Also, it coincides with N on all sequences X where $M(N) \le n_0$. On all the other sequences, $C(N) - C(N^*) \ge c$, $T(N^*) \le T(N)$ and $M(N^*) = n_0$. Hence,

$$r(\pi, N) - r(\pi, N^*) \ge \boldsymbol{P}(M(N) > n_0)\boldsymbol{E}\{\rho_{M(N)} - \rho_{n_0} + c\} \ge 0.$$

Thus, the problem reduces to truncated plans. By (i), the optimal fixed-sample plan is Bayes.

Proof of Theorem 2. Let $\mathscr{D}_k$ be the set of all sequential plans that are not randomized before k observations are collected. That is, it is a set of possibly randomized plans whose associated distributions $\mathscr{P}^{(j)}$ are degenerate for $j < k$ (see (2)). Then $\mathscr{D}_0$ is the set of all sequential plans and $\mathscr{D}_\infty = \bigcap_0^\infty \mathscr{D}_k$ is the set of non-randomized plans. Without loss of generality, assume that all the mentioned sequential plans use the Bayes terminal decision rule δ^B.

(1) For any plan $N^{(k)} \in \mathscr{D}_k$, we construct a plan $N^{(k+1)} \in \mathscr{D}_{k+1}$ as follows.

Let $N^{(k+1)}(\boldsymbol{X}^n) = N^{(k)}(\boldsymbol{X}^n)$ for any $n \ne k$. For a sample $\boldsymbol{X}^k$ of size k, consider plans $\overline{N}^j$, $j = 0, 1, \ldots$ such that $\overline{N}^j(\boldsymbol{X}^k) = j$ a.s. and $\overline{N}^j(\boldsymbol{Y}^k) = N^{(k)}(\boldsymbol{Y}^k)$ for any vector $\boldsymbol{Y}^k \ne \boldsymbol{X}^k$.

Define posterior risks

$$r_k^{(j)}(\boldsymbol{X}_k) = \boldsymbol{E}\Big(\inf_\delta \boldsymbol{E}\{L(\theta, \delta(\boldsymbol{X}^k)) + C(\overline{N}^j)|\boldsymbol{X}^k\}\Big)$$

and

$$r_k(\boldsymbol{X}_k) = \boldsymbol{E}\Big(\inf_\delta \boldsymbol{E}\{L(\theta, \delta(\boldsymbol{X}^k)) + C(N^{(k)})|\boldsymbol{X}^k\}\Big) = \boldsymbol{E}^{\mathscr{P}(\boldsymbol{X}^k)}\big(r_k^{(j)}(\boldsymbol{X}^k)\big) \tag{8}$$

and let

$$N^{(k+1)}(\boldsymbol{X}_k) = \arg\min_j r_k^{(j)}. \tag{9}$$

Notice that $N^{(k+1)}$ is well defined because

$$r_k^{(j)}(\boldsymbol{X}_k) \ge c\boldsymbol{E}\big\{M_T(\overline{N}^j)|\boldsymbol{X}_k\big\} \ge cj \uparrow +\infty$$

as $j \uparrow \infty$. Thus, the minimum of $r_k^{(j)}$ is always attained.

The new plan $N^{(k+1)}$ is not randomized before $(k+1)$ observations are collected. Comparing Bayes risks of $N^{(k)}$ and $N^{(k+1)}$, from (8), we obtain

$$r(\pi, N^{(k+1)}) = \boldsymbol{E} \min_j \left\{ r_k^{(j)}(\boldsymbol{X}^k) \right\} \le \boldsymbol{E}\boldsymbol{E}^{\mathcal{P}(\boldsymbol{X}^k)} \left\{ r_k^{(j)}(\boldsymbol{X}^k) \right\} = r(\pi, N^{(k)}).$$

Thus, the new plan belongs to $\mathcal{D}_{k+1}$ and does not lose to $N^{(k)}$ in terms of the Bayes risk.

(2) We now construct a non-randomized plan $N^{(\infty)} \in \mathcal{D}_\infty$ whose Bayes risk does not exceed $r(\pi, N)$. First, we use the algorithm above to define a sequence of plans $N^{(0)} = N \in \mathcal{D}_0$, $N^{(1)} \in \mathcal{D}_1$, etc. Then let

$$N^{(\infty)}(\boldsymbol{X}^n) = N^{(n+1)}(\boldsymbol{X}^n)$$

for any n. Since $N^{(n+1)}(\boldsymbol{X}^n)$ is not randomized, so is $N^{(\infty)}(\boldsymbol{X}^n)$ for all n, thus $N^{(\infty)}(\boldsymbol{X}^n) \in \mathcal{D}_\infty$. It remains to show that $r(\pi, N^{(\infty)}) \le r(\pi, N)$.

The case $r(\pi, N) = \infty$ is trivial. Let us assume that $r(\pi, N) < \infty$. Then $r(\pi, N^{(\infty)}) < \infty$. Indeed, for any θ and any sequence $\{X_1, X_2, \dots\}$, one has $\lim_{k\to\infty} M(N^{(k)}) = M(N^{(\infty)})$ where $M(N)$ is the total number of observations under the plan N.

Since

$$r(\pi, N^{(k)}) = \boldsymbol{E}\{L(\theta, \delta^B) + cM(N^{(k)}) + a\boldsymbol{E}T(N^{(k)})\} \le r(\pi, N) < \infty,$$

we have

$$\boldsymbol{E}M(N^{(\infty)}) = \lim_{k\to\infty} \boldsymbol{E}M(N^{(k)}) \le r(\pi, N) < \infty,$$

from where

$$r(\pi, N^{(\infty)}) \le \boldsymbol{E}\{L(\theta, \delta^B) + (a+c)\boldsymbol{E}M(N^{(\infty)})\} < \infty.$$

Thus, the introduced plan $N^{(\infty)}$ has a finite risk.

Next, if the total number of observations $M = M(N)$ does not exceed m, the plans $N^{(\infty)}$ and $N^{(m+1)}$ coincide, and so do their loss and cost functions. Hence,

$$\begin{aligned} r(\pi, N^{(\infty)}) - r(\pi, N) &\le r(\pi, N^{(\infty)}) - r(\pi, N^{(m+1)}) \\ &\le \boldsymbol{E}\{I_{\{M>m\}}(L(\theta, \delta^B) + C(N^{(\infty)})\} \to 0, \end{aligned}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

as $m \to \infty$, because $r(\pi, N^{(\infty)}) = \boldsymbol{E}\{L(\theta, \delta^B) + C(N^{(\infty)})\}$ is finite. Hence $r(\pi, N^{(\infty)}) \le r(\pi, N)$.

Proof of Theorem 3. By Theorem 2, it suffices to consider non-randomized plans only. Let ε_k be the set of all sequential plans that choose group sizes based on the observed sufficient statistic before k observations are collected. That is, $N \in \mathscr{E}_k \Leftrightarrow N(\boldsymbol{X}^j) = N(S_j(\boldsymbol{X}^j))$ for any $j < k$. Then $\mathscr{E}_0 = \mathscr{D}_0$ is the set of all sequential plans and $\mathscr{E}_\infty = \bigcap_0^\infty \mathscr{E}_k$ is the set of plans that are based on $\{S_n\}$ for all n.

(1) Choose an arbitrary plan $N^{(k)} \in \mathscr{E}_k$ and let $N^{(k+1)}(\boldsymbol{X}^n) = N^{(k)}(\boldsymbol{X}^n)$ for all $n \ne k$. Notice that $N^{(k)}(\boldsymbol{X}^n) = N^{(k)}(S_n(\boldsymbol{X}^n))$ if $n < k$. For a sample $\boldsymbol{X}^k$ of size k, consider sequential plans $\overline{N}^j$,

$$\overline{N}^j(\boldsymbol{Y}^n) = \begin{cases} j & \text{if } n = k,\ S_k(\boldsymbol{Y}^k) = S_k(\boldsymbol{X}^k) \\ N^{(k)}(\boldsymbol{Y}^k) & \text{if } n = k,\ S_k(\boldsymbol{Y}^k) \ne S_k(\boldsymbol{X}^k) \\ N^{(k)}(\boldsymbol{Y}^n) & \text{if } n \ne k \end{cases} \tag{10}$$

and their posterior risks

$$\begin{aligned} r_k^{(j)}(\boldsymbol{X}_k) &= \boldsymbol{E}\Big(\inf_\delta \boldsymbol{E}\{L(\theta, \delta(\boldsymbol{X}^k)) + C(\overline{N}^j) | \boldsymbol{X}^k\}\Big) \\ &= \boldsymbol{E}\Big(\inf_\delta \boldsymbol{E}\{L(\theta, \delta(\boldsymbol{X}^k)) + C(\overline{N}^j) | S_k(\boldsymbol{X}^k)\}\Big) \\ &= r_k^{(j)}(S_k(\boldsymbol{X}^k)). \end{aligned} \tag{11}$$

Similarly, let

$$r_k(\boldsymbol{X}_k) = E\Big(\inf_\delta \boldsymbol{E}\{L(\theta, \delta(\boldsymbol{X}^k)) + C(N^{(k)}) | \boldsymbol{X}^k\}\Big)$$

be the posterior risk of $N^{(k)}$. Define

$$N^{(k+1)}(\boldsymbol{X}_k) = \arg\min_j r_k^{(j)}$$

and notice from (10) and (11) that $N^{(k+1)}(\boldsymbol{Y}^n)$ is based on a sufficient statistic for $n \le k$, i.e., $N^{(k+1)} \in \varepsilon_{k+1}$. Also,

$$r(\pi, N^{(k+1)}) = E^{S_k} \min_j \{r_k^{(j)}(X_k)\} \le E^{S_k} E\{r_k(X_k) | S_k\} = r(\pi, N^{(k)}).$$

(2) Define a plan $N^{(\infty)}$ by constructing the sequence of plans $N^{(0)} = N \in \mathscr{E}_0$, $N^{(1)} \in \mathscr{E}_1$, etc., by the algorithm introduced in the first

part of the proof and letting

$$N^{(\infty)}(\boldsymbol{X}^n) = N^{(n+1)}(\boldsymbol{X}^n).$$

The rest of the proof coincides with part 2 of the proof of Theorem (2). By the same arguments, one shows that $N^{(\infty)} \in \mathscr{E}_\infty$ and $r(\pi, N^{(\infty)}) \leq r(\pi, N)$. □

Proof of Lemma 1. By the Factorization theorem,

$$\exp\{\Lambda_n\} = \frac{f(\boldsymbol{X}^n|\theta_1)}{f(\boldsymbol{X}^n|\theta_0)} = \frac{g(S_n(\boldsymbol{X}^n)|\theta_1)h(\boldsymbol{X}^n)}{g(S_n(\boldsymbol{X}^n)|\theta_0)h(\boldsymbol{X}^n)} = \frac{g(S_n(\boldsymbol{X}^n)|\theta_1)}{g(S_n(\boldsymbol{X}^n)|\theta_0)}.$$

Therefore, $N(\Lambda_n) = N(S_n)$ is based on the sufficient statistic.

Proof of Theorem 4. Consider the case of $F = F_1$, so that $\mu = \boldsymbol{E}(z_j) = K(F_1, F_0)$ is the Kullback information between F_1 and F_0, hence it is positive.

Without loss of generality, we can assume $B = -\infty$, that is, there is only one stopping boundary A. Indeed, for any B satisfying $A + B = O(1)$, as $A \to \infty$, consider

$$\tau = \max\left\{n|\Lambda_n \leq \frac{Bu - Al}{u - l}\right\}.$$

Since

$$\frac{Bu - Al}{u - l} = A - \frac{A - B}{u - l}u = B + \frac{A - B}{u - l}(-l),$$

after collecting τ observations, the random walk will always be "closer" to A than to B, that is the batch sizes are selected based on the distance to A. We show that τ is proper, and $\boldsymbol{E}\tau < \infty$. Thus, the error in $\boldsymbol{E}(T)$ caused by setting $B = -\infty$ does not exceed $\boldsymbol{E}\tau = O(1)$.

Indeed, let $U(x) = \sum_{n=0}^{\infty} P\{\Lambda_n \leq x\}$ be the renewal measure of $(-\infty, x]$ which equals the expected time the random walk Λ_n spends below level x.[22]

Boundedness of z_1 guarantees existence of a moment generating function $\phi(t) = \boldsymbol{E}e^{tz_1}$ for all t. Since $\phi(0) = 1$ and $\phi'(0) = \mu > 0$, we have $\phi(t^*) < 1$ for some $t^* < 0$. Then, by Markov inequality (and essentially, by Chernoff inequality, see Theorem 9.3 of Ref.[4]),

$$\boldsymbol{P}\{\Lambda_n \leq x\} = \boldsymbol{P}\{e^{t^*\Lambda_n} \geq e^{t^*x}\} \leq e^{-t^*x}\boldsymbol{E}e^{t^*\Lambda_n} = e^{-t^*x}\phi^n(t^*)$$

so that

$$U(x) \le e^{-t^* x} \sum_{n=0}^{\infty} \phi^n(t^*) < \infty$$

(similar result in Ref.[8] concerns renewal measures of bounded intervals only).

Letting $x = (Bu - Al)/(u - l)$, we get

$$\boldsymbol{E}\tau = U\left(\frac{Bu - Al}{u - l}\right) < \infty.$$

For any $n \ge 0$, let $D_n = A - \sum_1^n z_j$ be the distance from the current log-likelihood ratio statistic to the upper stopping boundary ($D_0 = A$). A conservative SPPRT will then choose batch sizes as follows,

$$N_{k+1}(\boldsymbol{X}^{M_k}) = N(D_{M_k}) = \left\lceil \frac{D_{M_k}}{u} \right\rceil^+.$$

Consider the sequence of D_{M_k}. For any k, we have $D_{M_{k+1}} = A - \sum_1^{M_{k+1}} z_j$ and $D_{M_k} = A - \sum_1^{M_k} z_j$, and thus,

$$D_{M_{k+1}} = D_{M_k} - \sum_{M_k+1}^{M_{k+1}} z_j.$$

Therefore, since $M_{k+1} - M_k = N_{k+1} = \lceil D_{M_k}/u \rceil^+$,

$$\boldsymbol{E}\{D_{M_{k+1}} | D_{M_k}\} = D_{M_k} - \mu \left\lceil \frac{D_{M_k}}{u} \right\rceil^+ \le \left(1 - \frac{\mu}{u}\right) D_{M_k},$$

and

$$\boldsymbol{E}D_{M_k+1} = \boldsymbol{E}\boldsymbol{E}\{D_{M_k+1} | D_{M_k}\} \le \left(1 - \frac{\mu}{u}\right) \boldsymbol{E}D_{M_k}.$$

Thus, for any k, $\boldsymbol{E}D_{M_k+1} \le (1 - \mu/u)\boldsymbol{E}D_{M_k}$. By induction, we obtain $ED_{M_k} \le (1 - \mu/u)^k \boldsymbol{E}D_{M_1}$. Since $D_0 = A$, it follows that,

$$\boldsymbol{E}D_{M_k} \le A\left(1 - \frac{\mu}{u}\right)^k. \tag{12}$$

Next, let

$$T_A(y) = \min\left\{k \middle| \sum_1^{M_k} z_j \geq y\right\} = \min\{k | D_{M_k} \leq A - y\}$$

be the number of sampled batches required for $\log \Lambda_n$ to cross level $y\,(y \leq A)$. In particular, the total number of batches is $T = T_A(A) = \min\left\{k | \sum_1^{M_k} z_j \geq A\right\}$.

Consider $y = A - c\log A + c\log\log A$, where $c = 1/|\log(1-\mu/u)|$. Then,

$$\boldsymbol{E}(T) \leq \boldsymbol{E}T_A(y) + \boldsymbol{E}T_{A-y}(A-y) + O(1). \tag{13}$$

Noticing that $y < A$ and using Markov inequality and (12), we get

$$\begin{aligned}
\boldsymbol{E}T_A(y) &\leq \sum_{k<A-y} \boldsymbol{P}\{D_{M_k} \geq A-y\} + \sum_{k\geq A-y} \boldsymbol{P}\{D_{M_k} \geq A-y\} \\
&\leq A - y + \sum_{k\geq A-y} \frac{\boldsymbol{E}D_{M_k}}{A-y} \leq A - y + \sum_{k\geq A-y} \frac{A(1-\mu/u)^k}{A-y} \\
&\leq A - y + \frac{A(1-\mu/u)^{c\log A - c\log\log A}}{(\mu/u)(A-y)} \\
&= c(\log A - \log\log A) + \frac{\log A}{c(\mu/u)(\log A - \log\log A)} \\
&= \frac{\log A - \log\log A}{|\log(1-\mu/u)|} + O(1).
\end{aligned} \tag{14}$$

Let ϵ be a real number such that $0 < \epsilon < A - y$. Then the second term in the right-hand side of (13) is bounded above by the average run length of the pure SPRT with lower stopping boundary $-\epsilon$ and upper stopping boundary $A - y - \epsilon$, where $\epsilon \downarrow 0$. Thus,

$$\begin{aligned}
\boldsymbol{E}T_{A-y}(A-y) &= \frac{(1-\beta)(A-y-\epsilon) + \beta\cdot(-\epsilon)}{\mu} \\
&= \frac{(1-\beta)(A-y) - \epsilon}{\mu} \leq \frac{A-y}{\mu} \\
&= \frac{\log A - \log\log A}{|\log(1-\mu/u)|\mu}
\end{aligned} \tag{15}$$

Using inequalities (14) and (15) in (13), we obtain (6).

Proof of Theorem 5. Theorem 5 allows the following generalized formulation.

Lemma 4. *Consider an arbitrary sequence* $\{\xi_n, n \geq 1\}$ *of random variables and a sequence* $\{E_n, n \geq 1\}$ *of events,* $E_j \in \sigma(\xi_1, \ldots, \xi_j)$. *Let* $T = \inf\{j | E_j\}$ *be a stopping time. If* $\boldsymbol{P}\{E_j | \xi_1, \ldots, \xi_{j-1}\} \geq p$ *a.s., then*

(i) *T is proper*;
(ii) $ET \leq 1/p$.

Proof. Let $A_n = \overline{E}_1 \cap \cdots \cap \overline{E}_n$. Since $\boldsymbol{P}\{E_j | \xi_1, \ldots, \xi_{j-1}\} \geq p$ implies $\boldsymbol{P}\{\overline{E}_j | A_{j-1}\} \leq 1 - p$ for all j,

$$\boldsymbol{P}(A_n) = \boldsymbol{P}(\overline{E}_n | A_{n-1}) \boldsymbol{P}(A_{n-1}) \leq (1 - p) \boldsymbol{P}(A_{n-1}). \tag{16}$$

By induction, $\boldsymbol{P}(A_n) \leq (1 - p)^n$ for all $n \geq 1$, hence $\sum_1^\infty \boldsymbol{P}(A_n)$ converges. Therefore, $\boldsymbol{P}(T = \infty) = P(\limsup_n A_n) = 0$ by the first Borel–Cantelli lemma.

Also,

$$E(T) = \sum_{n=0}^{\infty} P(T > n) = P(T > 0) + \sum_{n=1}^{\infty} P(A_n) \leq \frac{1}{p}.$$

Theorem 5 follows with $\xi_n = X_n$ and $E_n = \{N(\boldsymbol{X}^n) = 0\}$.

REFERENCES

1. Assaf, D. A dynamic sampling approach for detecting a change in distribution. Ann. Statist. **1988**, *16*, 236–253.
2. Assaf, D.; Pollak, M.; Ritov, Y.; Yakir, B. Detecting a change of a normal mean by dynamic sampling with a probability bound on a false alarm. Ann. Statist. **1993**, *21*, 1155–1165.
3. Berger, J.O. *Statistical Decision Theory*; Springer-Verlag: New York, 1985.
4. Billingsley, P. *Probability and Measure*; Wiley: New York, 1986.
5. Emerson, S.S.; Fleming, T.R. Symmetric group sequential test designs. Biometrics **1989**, *45* (3), 905–923.
6. Ghosh, M.; Mukhopadhyay N.; Sen, P.K. *Sequential Estimation*; Wiley: New York, 1997.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

7. Govindarajulu, Z. *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*; American Sciences Press: Columbus, Ohio, 1981.
8. Gut, A. *Stopped Random Walks. Limit Theorems and Applications*; Springer-Verlag: New York, 1988.
9. Hall, P. Sequential estimation saving sampling operations. J. Royal Statist. Soc. B **1983**, *45* (2), 219–223.
10. Hayre, L.S. Group sequential sampling with variable group sizes. J. Royal Statist. Soc. B **1985**, *47* (1), 90–97.
11. Jennison, C.; Turnbill, B.W. *Group Sequential Methods with Applications to Clinical Trials*; Chapman & Hall: Boca Raton, FL, 2000.
12. Jennison, C.; Turnbull, B.W. Interim analyses: The repeated confidence interval approach. J. Royal Statist. Soc. B **1989**, *51* (3), 305–361.
13. Lai, T.L. Sequential analysis: Some classical problems and new challenges. Statistics Sinica **2001**, *11*, 303–408.
14. Lewis, R.J.; Berry, D.A. Group sequential clinical trials: A classical evaluation of Bayesian decision-theoretic designs. J. Amer. Statist. Assoc. **1994**, *89* (428), 1528–1534.
15. Lin, D.Y.; Wei, L.J.; DeMets, D.L. Exact statistical inference for group sequential trials. Biometrics **1991**, *47* (4), 1399–1408.
16. Marschner, S.R.; Westin, S.H.; Lafortune, E.P.F.; Torrance, K. Image-based brdf measurement. Applied Optics **2000**, *39* (16), 2592–2600.
17. Mukhopadhyay N.; Solanky, T.K.S. *Multistage Selection and Ranking Procedures. Second-Order Asymptotics*; Marcel Dekker: New York, 1994.
18. Pocock, S.J. Group sequential methods in the design and analysis of clinical trials. Biometrika **1977**, *64*, 191–199.
19. Roters, M. A sequentially planned Bayesian multiple decision problem in continuous time. Sequential Analysis **2002**, *21* (1–2), 59–86.
20. Schmegner, C. Decision Theoretic Results for Sequentially Planned Statistical Procedures. Ph.D. dissertation, The University of Texas at Dallas, 2003.
21. Schmitz, N. *Optimal Sequentially Planned Decision Procedures*; Springer-Verlag: New York, 1993.
22. Siegmund, D. *Sequential Analysis: Tests and Confidence Intervals*; Springer-Verlag: New York, 1985.
23. Wetherill, G.B. The most economical sequential sampling scheme for inspection by variables. J. Royal Statist. Soc. B **1959**, *21* (2), 400–408.

24. Whitehead, J. Sequential methods for monitoring declining quality, with application to the long-term storage of seeds. Biometrics **1989**, *45* (1), 13–22.

Received February 2003
Revised October 2003
Accepted October 2003
Recommended by N. Mukhopadhyay

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016