C. Schmegner and M. Baron. Sequential Plans and Risk Evaluation.
Sequential Analysis, 26(4), 335–354, 2007.

Sequential Plans and Risk Evaluation

Claudia Schmegner

Department of Mathematical Sciences, DePaul University, Chicago, Illinois, USA

Michael Baron

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas, USA

Abstract: Introduced in 1990s, sequential plans generalize and optimize the classical sequential procedures allowing sampling in groups of variables sizes. Sequential plans are preferred to pure sequential procedures when the cost function is non-linear so that each observed group implies an additional fixed cost. Optimal plans minimize the risk, a weighted sum of the expected loss, cost of observations, and cost of groups. Although risk evaluation is a challenging problem in sequential planning, risk functions of sequentially planned probability ratio tests (SPPRT) can be obtained as roots of certain functional equations. Explicit solutions are derived for SPPRT on a lattice, allowing practitioners to compare exact risks and choose an optimal procedure.

Keywords: Lattice, Markov chain, Sequential plan, Sequentially planned probability ratio test, Stopping boundaries

Subject classification: 62H30, 62-07, 65U05, 68T05

1 Introduction

Abraham Wald [11] introduced the *sequential probability ratio test (SPRT)* that marked the beginning of modern sequential analysis ([7, 12]). In SPRT, observations are sampled one by one until the gained information guarantees desired error probabilities. At that moment data collection is terminated and a decision is made. The total sample size is not predetermined, it depends on data. After each observation, one decides whether to stop sampling and make an inference based on the data already collected, or to postpone the

decision until one more observation is collected.

Flexibility of sequential schemes allows to search for an optimal balance between reducing the total sampling cost and increasing accuracy of terminal decisions. Other attractive features of sequential algorithms include, in general terms, a relative reduction of the required sample size compared to classical statistics, and, unlike non-sequential procedures, control over both probabilities of Type I and Type II errors.

At the same time, sequential procedures are often criticized as impractical because of their requirement to take only one observation at a time. In a number of applications, such a sampling plan is expensive or not feasible because of additional costs (time, cost of equipment, cost of personnel, etc.) associated with each experiment besides the actual cost of each observation ([1], Sect. 7.8, [5, 6, 13] and others, see [8] for more references and discussion).

This problem can be resolved by expanding the set of sampling schemes from sequential to *sequentially planned* procedures ([10, 9]). Sequential planning is a more general sampling design, which combines flexibility of classical sequential analysis in its choice of a sample size with the convenience of sampling in batches.

1.1 A concept of sequential planning

The main idea is to sample observations in groups instead of collecting them one by one as in *pure* (classical) sequential analysis. After sampling a batch, all the collected data are used to decide, (1) whether or not to continue sampling, and (2) what the next group size is. In this general scheme, the number of batches and their sizes are determined sequentially.

The universal way in which *sequential planning* is defined makes it a generalization and unification of a number of established and commonly used sequential procedures: classical sequential sampling, group sequential sampling, multistage procedures, accelerated sequential procedures (see [9] for details).

Let $X_1, X_2, \ldots$ represent the incoming sequence of data, and let N_j be the size of the j-th sampled group. The size of the initial group N_1 is chosen before any data are collected, and each N_{j+1} for $j \geq 1$ is a function of

$$M_j = \sum_{i=1}^{j} N_i$$

already collected observations, $N_{j+1} = N_{j+1}(\boldsymbol{X}^{M_j})$, where $\boldsymbol{X}^k = (X_1, \ldots, X_k)$.

As soon as $N_j = 0$, sampling terminates. Thus,

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

is the number of sampled groups, M_T is the total sample size, and both quantities are random, dependent on the observed data.

This adaptive sampling approach is the main advantage of sequential planning and at the same time the main difficulty in related optimization problems. Indeed, now one has to optimize over a huge set of sequential plans $\{\mathcal{N}\}$ where each sequential plan $\mathcal{N}$ is a family of integer-valued functions $N_j = N_j^{(d)}(\boldsymbol{X}^d)$ on $\boldsymbol{R}^d$ for all j and d.

Examples of sequentially planned procedures are:

– *expectation-based plans* ([4]), where the size of each batch is proportional to the expected number of observations needed to complete the classical sequential sampling (say, the expected sample size needed to cross the stopping boundary), i.e.,

$$N_j^{(k)}(\boldsymbol{X}^k) \approx \gamma \boldsymbol{E} \min_n \left\{ n | N_j^{(k+n)}(\boldsymbol{X}^{k+n}) = 0 \right\};$$

– *quantile-based plans* ([4]), where the batch sizes are chosen so that the data collection stops (the boundary is crossed) immediately after the next batch with a given probability, i.e.,

$$N_j^{(k)}(\boldsymbol{X}^k) \in \left\{ n | P\{N^{(k+n)}(\boldsymbol{X}^{k+n}) = 0 | \boldsymbol{X}^k\} \approx \gamma \right\};$$

– *conservative plans* ([9]), where the maximum possible batch is taken that guarantees that no excessive observations (beyond a boundary) are collected, i.e.,

$$N_j^{(k)}(\boldsymbol{X}^k) = \min_n \left\{ n | P\{N^{(k+n)}(\boldsymbol{X}^{k+n}) = 0 | \boldsymbol{X}^k\} > 0 \right\};$$

– *m-conservative plans* ([9]), that guarantee no more than m excessive observations, i.e.,

$$N_j^{(k)}(\boldsymbol{X}^k) \leq \min_n \left\{ n | P\{N^{(k+n)}(\boldsymbol{X}^{k+n}) = 0 | \boldsymbol{X}^k\} > 0 \right\} + m.$$

The main objective is to choose batch sizes $N_1, N_2, \ldots, N_T$ in an optimal way, to achieve the desired optimal balance between sampling costs and accuracy of the inferences.

1.2 Decision Theoretic Approach

In order to reflect the cost of experiments (groups), sequential planning is based on a non-linear cost function

$$C(\mathcal{N}) = cM_T + c_0 T = \sum_1^T (cN_j + c_0), \tag{1}$$

where c is the cost of one observation and c_0 is the additional cost of one batch. Certainly, the two terms in the sum in (1) correspond to the *variable cost* and the *fixed cost* of the j-th experiment. This cost function not only considers the expense due to the observations collected, as in classical sequential analysis, but also accounts for fixed expenditure associated with each group. The fixed batch cost c_0 is assumed to be independent of the number of observations in each group.

Then, the *risk function* is defined as

$$R(\theta, \mathcal{N}, \delta) = \boldsymbol{E}^{\boldsymbol{X}} \left\{ L(\theta, \delta) + C(\mathcal{N}) \right\} = \boldsymbol{E}^{\boldsymbol{X}} \left\{ L(\theta, \delta) + cM_T + c_0 T \right\}, \tag{2}$$

where $\theta \in \Theta$ is a parameter, $\delta \in \mathcal{A}$ a terminal decision rule, $L(\theta, \delta)$ is a loss function, and $\boldsymbol{X} = (X_1, X_2, \ldots)$.

Plans are examined and compared with respect to their risk function $R(\theta, \mathcal{N}, \delta)$ and their *Bayes risk*

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

with $\pi(\theta)$ being the prior distribution of θ.

As in classical decision theory, a plan $\mathcal{N}$ is *R-better* than a plan $\tilde{\mathcal{N}}$ if

$$\begin{cases} R(\theta, \mathcal{N}, \delta) \leq R(\theta, \tilde{\mathcal{N}}, \delta) \text{ for any } \theta \in \Theta, \\ R(\theta, \mathcal{N}, \delta) < R(\theta, \tilde{\mathcal{N}}, \delta) \text{ for some } \theta \in \Theta; \end{cases}$$

it is *admissible* if no plan is R-better than $\mathcal{N}$, and it is *Bayes* if it minimizes the Bayes risk among all sequential plans.

In this general setting of sequential planning, we search for (a) optimal sequential schemes, and (b) strategies to improve existing sequential plans. This is not effortless because the class of all sequential plans is large and irregular due to the variety of possible choices of group size functions N_j.

Two general principles of optimal sequential planning ([9]) significantly reduce the set of competing sequential plans, thus simplifying our search.

First, one does not need to randomize, i.e., we can only consider deterministic functions $N_j^{(k)}(\boldsymbol{X}^d)$. Otherwise, there exists a non-randomized sequential plan whose Bayes risk does not exceed the risk of the original plan.

Second, group sizes $N_j^{(k)}$ should be functions of a sufficient statistic of $\boldsymbol{X}^k$. Again, there will always exist a plan satisfying this sufficiency principle and having an equal or lower Bayes risk.

In general, computation and comparison of risks of sequential plans is not trivial. To do it, expected number of groups $\boldsymbol{E}(T)$ and expected total sample size $\boldsymbol{E}(M_T)$ have to be computed. Clearly, it includes computation of the expected stopping time as a small special case ($N_j = 1$ for $j \leq T$).

In this paper, we study the risk of the sequentially planned probability ratio test (SPPRT), which is a sequentially planned generalization of Wald's SPRT that satisfies both above mentioned principles. It is shown in Section 3 that the risk of SPPRT and each risk component can be obtained as roots of certain functional equations. Section 4 defines SPPRT on a lattice. Explicit solutions for the risk components and the total risk are derived in this case. Results are based on the elementary theory of random walks.

2 Sequentially Planned Probability Ratio Test

The sequentially *planned* probability ratio test (SPPRT) was introduced by Schmitz ([10]) as a generalization of Wald's sequential probability ratio test (SPRT). Similarly to the SPRT, it is used to test a simple null hypothesis H_0: $F = F_0$ against a simple alternative H_1: $F = F_1$ based on a sequence of iid random variables $X_1, X_2, \ldots$ from distribution F. In many applications (e.g., clinical trials), it is also used to test between composite hypotheses in presence of an "indifference region", for example, for testing H_0: $\theta \leq \theta_0$ vs H_1: $\theta \geq \theta_1$ for $\theta_0 < \theta_1$ ([3]). It is the case when improvement by less than $(\theta_1 - \theta_0)$ is not sufficient for marketing a new treatment or implementing a new program. Technically, the test procedure is equivalent to the test of simple H_0 and H_1.

A sequential version of the likelihood ratio test, SPRT is based on log-likelihood ratios

$$\Lambda_n = \log \frac{dF_1(\boldsymbol{X}^n)}{dF_0(\boldsymbol{X}^n)} = \sum_{j=1}^{n} \log \frac{dF_1(X_j)}{dF_0(X_j)} = \sum_{j=1}^{n} z_j. \tag{3}$$

Then, Λ_n is a random walk with a negative drift ($\boldsymbol{E}_{F_0} z_j < 0$) under H_0 , and it is a random walk with a positive drift ($\boldsymbol{E}_{F_1} z_j > 0$) under the alternative, H_1. Stopping boundaries A and B ($A > 0 > B$) are

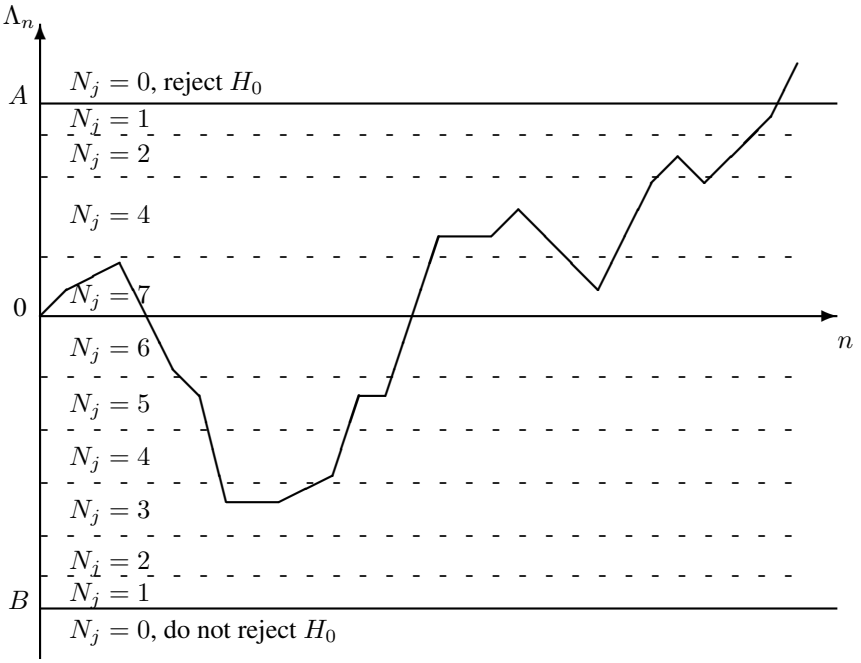

Figure 1: Example of an SPPRT. Horizontal regions define the next sampling step.

chosen in accordance with the desired probabilities of Type I and Type II errors. Then, after observing X_n,

– sampling is terminated and H_0 is rejected if $\Lambda_n \geq A$;

– sampling is terminated and H_0 is not rejected if $\Lambda_n \leq B$;

– sampling continues and X_{n+1} is collected if $B < \Lambda_n < A$.

This scheme naturally translates into the setting of sequential plans. In SPPRT, after j batches totaling $M_j = \sum_1^j N_k$ observations are collected,

– sampling is terminated and H_0 is rejected if $\Lambda_{M_j} \geq A$;

– sampling is terminated and H_0 is not rejected if $\Lambda_{M_j} \leq B$;

– sampling continues and a new *batch* of $N_{j+1} = N(\Lambda_{M_j})$ observations is collected if $B < \Lambda_n < A$.

Thus, an SPPRT divides the interval (B, A) into subsets that define the next sampling step N_j based on Λ_{M_j} (Fig. 2.1). Sampling terminates when Λ_{M_j} leaves this interval.

Remark 1. The key idea is that the size of the next batch is determined based on the current log-likelihood ratio only. For example, if Λ_k exceeds A for some $k \in (M_{j-1}, M_j)$, before the end of the batch,

but falls between B and A at the end of it, the null hypothesis is *not* rejected at this point, and sampling continues. This may be argued because a pure sequential scheme, SPRT, would terminate and reject H_0 at k. However, the following two arguments justify the given definition of SPPRT. First, the order of observations should not affect the result, by the sufficiency principle. If $\Lambda_{M_{j-1}} < A$ and $\Lambda_{M_j} < A$, then there always exists an ordering of $X_{M_{j-1}+1}, \ldots, X_{M_j-1}$ such that $\Lambda_n < A$ for all $n \in (M_{j-1}, M_j)$. Second, on a practical language, looking back at the j-th batch, we see that $\Lambda_k > A$ suggests significant evidence against the null hypothesis. The rest of the batch, however, denies this evidence. Trusting a larger sample, we should be glad that sampling did not stop at k, preventing a possibly wrong decision.

Remark 2. Being a function of the log-likelihood ratio, the size of each batch is independent of the number j of previously collected batches and the number M_j of previously collected observations. This also agrees with the sufficiency principle. Our next step is independent of the actual sampling routine that led to collection of $\boldsymbol{X}^{M_j}$ and independent of the price already paid for the collected data.

Remark 3. All SPPRT satisfy the sufficiency and non-randomization principles of optimal sequential planning ([9]), according to the Factorization Theorem.

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 2.1 *([10], Lemma 3.7) For an SPPRT,*

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

Corollary 2.2 *For an SPPRT,*

$$\alpha \leq e^{-A}, \;\; \beta \leq e^{B}.$$

3 Evaluation of Risk Components of SPPRT

The main problem we address in this paper is calculation of the risks associated with different sequential plans. Once these values are obtained, the plan with the smallest risk can be selected.

As a first step, a more general problem is solved, namely functional equations for integer-valued functions on random walks are derived. The log-likelihood ratio test statistic, Λ_n, being a random walk in the case of i.i.d. data, the derived equations can be applied directly to the computation of the risk components of the SPPRT, i.e., its error probabilities, the expected number of batches, and the expected total sample size.

Therefore, a new notation will not be introduced.

Consider a random walk, $\Lambda_n = \sum_{j=1}^{n} z_j$, with $\Lambda_0 = 0$ and z_j i.i.d. random variables. Let $\boldsymbol{Z} = (z_1, z_2, \ldots)$, $\boldsymbol{Z}^k = (z_1, z_2, \ldots, z_k)$, $\boldsymbol{Z}_k = (z_k, z_{k+1}, \ldots)$, and let A and B be constants such that $B < 0 < A$.

Given N, a nonnegative, integer-valued non-random function such that $N(u) = 0$ for any $u \notin (B, A)$ and $N(u) \geq 1$ for $B < u < A$, define a sequence of functions $N_1(u) = N(u)$, $N_2(u) = N(u + \Lambda_{N_1(u)})$, $N_3(u) = N(u + \Lambda_{N_1(u)+N_2(u)}), \ldots, N_k(u) = N(u + \Lambda_{N_1(u)+\ldots+N_{k-1}(u)}), \ldots$. This sequence corresponds to group sizes.

Likewise, the function $M_k(u) = \sum_1^k N_j(u)$, where $k = 1, 2, \ldots$, represents the total number of observations collected in k groups, and the function $T(u)$ defined below signifies the total number of groups,

$$T(u) = \inf\{k : \ N_k(u) = 0\} - 1. \tag{5}$$

It follows that $T(u)$ is random since it is dependent on $\{z_j\}$. Also, $T(u) = 0$ for $u \notin (B, A)$ and $T(u) > 0$ if $u \in (B, A)$.

Theorem 3.1 *(Expected number of sampled groups) The function* $g(u) = \boldsymbol{E}^{\boldsymbol{Z}} T(u)$ *solves a functional equation,*

$$g(u) = 1 + \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} g(u + \Lambda_{N(u)}), \tag{6}$$

with a boundary condition, $g(u) = 0$ *for* $u \notin (B, A)$.

Theorem 3.2 *(Expected total sample size) The function* $\nu(u) = \boldsymbol{E}^{\boldsymbol{Z}} M_{T(u)}$ *solves a functional equation,*

$$\nu(u) = N(u) + \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \nu(u + \Lambda_{N(u)}), \tag{7}$$

with a boundary condition, $\nu(u) = 0$ *for* $u \notin (B, A)$.

Theorem 3.3 *(Power function) The function* $\rho(u) = \boldsymbol{P}\{\Lambda_{M_T(u)} \geq A\}$ *solves a functional equation,*

$$\rho(u) = \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \rho(u + \Lambda_{N(u)}). \tag{8}$$

with boundary conditions, $\rho(u) = 1$ *if* $u \geq A$ *and* $\rho(u) = 0$ *if* $u \leq B$.

Theorem 3.4 *(Probability of accepting the null hypothesis) The function* $\tau(u) = \boldsymbol{P}\{\Lambda_{M_T(u)} \leq B\}$ *solves a functional equation*

$$\tau(u) = \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \tau(u + \Lambda_{N(u)}). \tag{9}$$

with boundary conditions, $\tau(u) = 0$ *if* $u \geq A$ *and* $\tau(u) = 1$ *if* $u \leq B$.

From Theorems 3.1-3.4, derivation of the functional equation for the risk of SPPRT is straightforward.

3.1 Functional Equation for the Risk

Let θ be a parameter characterizing the distribution of z, and consider a hypothesis testing problem regarding θ. Also, let $L(\theta, \delta)$ be a loss function corresponding to the decision rule δ,

$$L_u(\theta, \delta) = \begin{cases} 0 & \text{if a correct decision is made} \\ w_\alpha & \text{if a type I error is made} \\ w_\beta & \text{if a type II error is made} \end{cases}$$

Then,

Theorem 3.5 *(The risk of SPPRT) The risk function,* $R_u(\theta, \mathcal{N}, \delta) = \boldsymbol{E}^{\boldsymbol{Z}}\{L_u(\theta, \delta) + c_0 T(u) + cM_{T(u)}\}$, *solves a functional equation,*

$$R_u(\theta, \mathcal{N}, \delta) = c_0 + cN(u) + \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} R_{u+\Lambda_{N(u)}}(\theta, \mathcal{N}, \delta), \tag{10}$$

with the following boundary conditions:

(i) If H_0 *is true,* $R_u(\theta, \mathcal{N}, \delta) = w_\alpha$ *for* $u \geq A$, *and* $R_u(\theta, \mathcal{N}, \delta) = 0$ *for* $u \leq B$.

(ii) If H_0 *is false,* $R_u(\theta, \mathcal{N}, \delta) = 0$ *for* $u \geq A$, *and* $R_u(\theta, \mathcal{N}, \delta) = w_\beta$ *for* $u \leq B$.

For a prior distribution $\pi(\theta)$, a functional equation for the Bayes risk follows directly from equation (10), by taking expectations of both sides with respect to π (also, see Corollary 4.9).

Proofs of all the results are given in Appendix.

4 Sequentially Planned Probability Ratio Test on a Lattice

In general, the functional equations obtained in the previous section do not have closed-form solutions. However, exact solutions can be derived for *SPPRT on a lattice*.

4.1 SPPRT on a Lattice

An SPPRT is said to be *on a lattice* if there exists a positive real number Δ such that $\Lambda_n \in \{k\Delta, k \in \mathbb{Z}\} = \mathcal{L}_\Delta$. Without loss of generality, the stopping boundaries can be assumed to belong to $\mathcal{L}_\Delta$. Then, let $a = A/\Delta$ and $b = B/\Delta \in \mathbb{Z}$. Using the random walk nature of Λ_n, the SPPRT can be connected with a certain Markov chain. Define

$$Y_n = \begin{cases} \Lambda_{M_n}/\Delta & \text{if} \quad \Lambda_{M_k} \in (B, A)\ \forall k \le n, \\ b & \text{if} \quad \Lambda_{M_k} \le B \text{ and } \Lambda_{M_j} < A\ \forall j < k \text{ for some } k \le n, \\ a & \text{if} \quad \Lambda_{M_k} \ge A \text{ and } \Lambda_{M_j} > B\ \forall j < k \text{ for some } k \le n. \end{cases}$$

Thus, ΔY_n is the value of the log-likelihood ratio process after n groups of observations are collected, until the process is stopped, i.e., for all $n < T$. As soon as the sampling stops and the decision is taken, ΔY_n freezes at the upper stopping boundary in case of rejection of H_0, and it freezes at the lower boundary if H_0 is not rejected.

Lemma 4.1 *$\{Y_n\}$ is a stationary Markov chain with absorbing states a and b.*

This representation of the SPPRT in terms of a Markov chain results in a simplification of the functional equations for the risk and risk components derived in section 3. Specifically, these equations become matrix equations for which explicit solutions are available.

Before any data are collected, the Markov chain initializes at 0: $Y_0 = \Lambda_0/\Delta = 0$. As in the previous section, all possible starting values of the random walk are considered, $Y_0 = b, b+1, \ldots, a$, and a risk vector $\boldsymbol{R}(\theta, \mathcal{N}, \delta)$ is defined,

$$\boldsymbol{R}(\theta, \mathcal{N}, \delta) = \begin{pmatrix} R_{b+1}(\theta, \mathcal{N}, \delta) \\ \vdots \\ R_{a-1}(\theta, \mathcal{N}, \delta) \end{pmatrix}.$$

$R_u(\theta, \mathcal{N}, \delta)$ represents the risk corresponding to a Markov chain starting at u. In the end, only $R_0(\theta, \mathcal{N}, \delta)$ is needed. The solution of a matrix equation for the whole risk vector will enable calculation of $R_0(\theta, \mathcal{N}, \delta)$ as well.

Matrix equations for the risk components (expected number of groups, expected number of observations, and error probabilities), and finally for the risk itself will be derived.

Let

$$\boldsymbol{P} = \boldsymbol{P}(\theta) = \begin{pmatrix} p_{b+1,b+1} & p_{b+1,b+2} & \cdots & p_{b+1,a-1} \\ p_{b+2,b+1} & p_{b+2,b+2} & \cdots & p_{b+2,a-1} \\ \vdots & \vdots & \vdots & \vdots \\ p_{a-1,b+1} & p_{a-1,b+2} & \cdots & p_{a-1,a-1} \end{pmatrix}$$

be the matrix of transition probabilities for Y_n for states *between* the absorbing barriers, $b+1, \ldots, a-1$. Here, $p_{ij} = P\{Y_{n+1} = j | Y_n = i\}$ for $i, j \in \{b+1, \ldots, a-1\}$. Matrix $\boldsymbol{P}$ is only the central portion of the transition probability matrix of Y_n since states a and b have not been included.

Lemma 4.2 *For any SPPRT, where z_j has finite variance, the number of groups T is finite with probability one.*

Finite T implies that the Markov chain Y_n will eventually enter either state a or state b and never come back to visit the states between b and a. In other words, the states between the boundaries, $b+1, b+2, \ldots, a-2, a-1$, are transient, while the stopping boundaries, a and b are persistent.

Lemma 4.3 *For any SPPRT on a lattice, the matrix $(\boldsymbol{I} - \boldsymbol{P})$ is invertible.*

Invertibility of $(\boldsymbol{I} - \boldsymbol{P})$ is necessary in order to prove existence and uniqueness of solutions for the matrix equations for the risk components and the risk itself.

Let $\mathbf{g}$ be a vector of expected number of groups corresponding to different starting points of the SPPRT,

$$\mathbf{g} = (g(u)), \text{ where } u = b+1, \ldots, a-1, \text{ and } g(u) = \boldsymbol{E}^{\boldsymbol{Z}} T(u).$$

Beyond the stopping boundaries, the expected number of groups is 0, $g(a) = g(b) = 0$.

An explicit expression for the vector of expected number of groups is obtained.

Theorem 4.4 *The expected number of groups $\boldsymbol{E}(T)$ of SPPRT equals*

$$\mathbf{g} = (\boldsymbol{I} - \boldsymbol{P})^{-1}\mathbf{1}, \tag{11}$$

where $\mathbf{1}$ *is a vector of ones.*

A similar matrix equation is derived for the expected number of observations. Denote by $\underline{\nu}$ the vector of expected numbers of observations corresponding to different starting points of the SPPRT,

$$\underline{\nu} = (\nu(u)), \text{ where } u = b+1, \ldots, a-1, \text{ and } \nu(u) = \boldsymbol{E}^{\boldsymbol{Z}} M_{T(u)}.$$

Of course, $\nu(a) = \nu(b) = 0$, that is, no more observations are taken once the boundaries are crossed. Also, let $\boldsymbol{N}$ be the vector of batch sizes based on the current value of Λ_{M_j}, i.e., on the current state of Y_j,

$$\boldsymbol{N} = (N(u)), \text{ where } u = b+1, \ldots, a-1.$$

Theorem 4.5 *The expected number of observations $E(M_T)$ equals*

$$\underline{\nu} = (\boldsymbol{I} - \boldsymbol{P})^{-1}\boldsymbol{N}. \tag{12}$$

In the same way, matrix equations for the expected loss, that is, for the error probabilities, are obtained. General expressions for the probabilities of accepting and rejecting the null hypotheses are derived.

Denote by $\underline{\rho}$ the vector of probabilities of rejecting the null hypothesis,

$$\underline{\rho} = (\rho(u)), \text{ where } u = b+1, \ldots, a-1, \text{ and } \rho(u) = P(\text{reject } H_0).$$

and by $\gamma(u)$ the transition probability from a state u to the absorbing barrier a in one batch,

$$\underline{\gamma} = (\gamma(u)) \text{ where } u = b+1, \ldots, a-1, \text{ and } \gamma(u) = P(u \to a).$$

Theorem 4.6 *The probability of rejecting the null hypothesis equals*

$$\underline{\rho} = (\boldsymbol{I} - \boldsymbol{P})^{-1}\underline{\gamma}. \tag{13}$$

Let $\underline{\tau}$ be the vector of probabilities of accepting the null hypothesis,

$$\underline{\tau} = (\tau(u)) \text{ where } u = b+1, \ldots, a-1, \text{ and } \tau(u) = P(\text{accept } H_0).$$

and let $\xi(u) = P(u \to b)$ be the one-batch transition probability from state u to the absorbing state b and by $\underline{\xi}$,

$$\underline{\xi} = (\xi(u)) \text{ where } u = b+1, \ldots, a-1, \text{ and } \xi(u) = P(u \to b).$$

Theorem 4.7 *The probability of accepting the null hypothesis equals*

$$\underline{\tau} = (\boldsymbol{I} - \boldsymbol{P})^{-1}\underline{\xi}. \tag{14}$$

Combining the results of Theorems 4.4-4.7, an expression is derived for the (frequentist) risk for the general class of SPPRT on a lattice.

Corollary 4.8 *Under the null hypothesis,*

$$\boldsymbol{R}(\theta_0, \mathcal{N}, \delta) = (\boldsymbol{I} - \boldsymbol{P}(\theta_0))^{-1}\{w_\alpha \underline{\gamma} + c_0\mathbf{1} + c\boldsymbol{N}\}.$$

Under the alternative hypothesis,

$$\boldsymbol{R}(\theta_1, \mathcal{N}, \delta) = (\boldsymbol{I} - \boldsymbol{P}(\theta_1))^{-1}\{w_\beta \underline{\xi} + c_0\mathbf{1} + c\boldsymbol{N}\}.$$

Calculation of the Bayes risk of SPPRT is immediate.

Corollary 4.9 *The Bayes risk of SPPRT $\mathcal{N}$ with prior probabilities $\pi_j = \boldsymbol{P}(H_j)$ for $j = 0, 1$ equals*

$$\boldsymbol{r}(\pi, \mathcal{N}, \delta) = \pi_0(\boldsymbol{I} - \boldsymbol{P}(\theta_0))^{-1}\{w_\alpha \underline{\gamma} + c_0\mathbf{1} + c\boldsymbol{N}\} + \pi_1(\boldsymbol{I} - \boldsymbol{P}(\theta_1))^{-1}\{w_\beta \underline{\xi} + c_0\mathbf{1} + c\boldsymbol{N}\}. \tag{15}$$

Again, vector $\boldsymbol{r}$ consists of Bayes risks for all possible starting points of the random walk Λ_n. The value of r_0 is the Bayes risk of interest in the problem at hand.

4.2 SPPRT for Common Distributions

The theory of SPPRT on a lattice, that we just developed, can be applied to a number of common discrete and continuous underlying distribution families. Typically, a certain relationship between θ_0 and θ_1 needs to be met. However, accurate approximations are always available even if this condition is not satisfied exactly.

Example 4.10 The SPPRT of the null hypothesis $\theta = \theta_0$ versus the alternative $\theta = \theta_1$ for observations coming from a Binomial(m, θ) distribution is on a lattice if and only if $r = \log\frac{1-\theta_1}{1-\theta_0} / \log\frac{\theta_1}{\theta_0}$ is rational.

Example 4.11 The SPPRT of the null hypothesis $\theta = \theta_0$ versus the alternative $\theta = \theta_1$ for observations coming from a Negative Binomial(k, θ) distribution is on a lattice if and only if $r = \log \frac{1-\theta_1}{1-\theta_0} / \log \frac{\theta_1}{\theta_0}$ is rational.

Example 4.12 The SPPRT of the null hypothesis $\theta = \theta_0$ versus the alternative $\theta = \theta_1$ for observations coming from a Poisson(θ) distribution is on a lattice if and only if $r = (\theta_1 - \theta_0)/ \log \frac{\theta_1}{\theta_0}$ is rational.

Remark 4.13 If the conditions of the above theorems are not satisfied, r can always be approximated by a rational number with an error as small as desired.

Although being on a lattice is a discrete property, some SPPRT satisfy this condition even if the underlying distribution of data is continuous.

Lemma 4.14 *The SPPRT of the null hypothesis $\theta = \theta_0$ versus the alternative $\theta = \theta_1$ for observations coming from a distribution with density function of the form*

$$f(x|\theta) = g(x)h(\theta)\mathbf{1}_{x \in A(\theta)},$$

where $A(\theta_0) \neq A(\theta_1)$, is on a lattice.

Clearly, the following two examples satisfy the assumed form of the distribution in Lemma 4.14 above. Thus, the corresponding SPPRT is on a lattice, and its risk can be computed by Corollary 4.8.

Example 4.15 The SPPRT of the null hypothesis, $\theta = \theta_0$ versus the alternative $\theta = \theta_1$ for observations coming from a Uniform$(0, \theta)$ distribution is on a lattice.

Example 4.16 The SPPRT of the null hypothesis, $\theta = \theta_0$ versus the alternative $\theta = \theta_1$ for observations coming from a Pareto(θ) distribution with the density function

$$f_\alpha(x|\theta) = \alpha\theta^\alpha x^{-(\alpha+1)}\mathbf{1}_{x>\theta},$$

for any fixed α, is on a lattice.

The derived theory can be applied, for instance, to evaluate the sampling cost of an SPPRT for testing proportions with given and fixed probabilities of Type I and Type II errors, as the next subsection exemplifies.

4.3 Application: SPPRT for Testing Proportions

In this subsection, risks of competing sequential plans for testing proportions are calculated using matrix equations. Comparison of the of these risks will enable the practitioner to choose the best plan.

Consider the Markov chain Y_n introduced above, with states $b, b+1, \ldots, a-1, a$, where a and b are absorbing barriers. For states $i, j \in \{b, b+1, \ldots, a-1, a\}$, the transition probabilities from i to j are calculated as follows,

$$\begin{aligned} p_{ij} &= P\{Y_{n+1} = j \mid Y_n = i\} = P\{\Lambda_{M_{n+1}}/\Delta = j \mid \Lambda_{M_n}/\Delta = i\} \\ &= P\left\{\frac{1}{\Delta}\sum_{k=M_n+1}^{M_n+N(i)} \Lambda_k = j - i\right\} = P\left\{\frac{1}{\Delta}\sum_{k=1}^{N(i)} \Lambda_k = j - i\right\} \\ &= P\left\{\frac{1}{\Delta}\sum_{k=1}^{N(i)} \log\frac{f(X_k|\theta_1)}{f(X_k|\theta_0)} = j - i\right\}. \end{aligned}$$

Now, consider the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, where the underlying distribution is Bernoulli(θ) and $\theta_1 = 1 - \theta_0$ (for simplicity; in fact, only the condition of Lemma 4.10 needs to be satisfied). Then, if X is a Bernoulli(θ) random variable,

$$\frac{f(X|\theta_1)}{f(X|\theta_0)} = \frac{\theta_1^X(1-\theta_1)^{1-X}}{\theta_0^X(1-\theta_0)^{1-X}} = \left(\frac{\theta_1}{\theta_0}\right)^{2X-1}.$$

Supposing $\theta_0 > 1/2$, let $\Delta = \log\frac{\theta_0}{\theta_1} = -\log\frac{\theta_1}{\theta_0}$. Then,

$$\frac{1}{\Delta}\log\frac{f(X|\theta_1)}{f(X|\theta_0)} = \frac{2X-1}{\Delta}\log\frac{\theta_1}{\theta_0} = 1 - 2X = \begin{cases} -1 & \text{if } X = 1 \\ 1 & \text{if } X = 0 \end{cases}.$$

Thus,

$$\begin{aligned} p_{ij} &= P\left\{\sum_{k=1}^{N(i)}(1 - 2X_k) = j - i\right\} = P\left\{2\sum_{k=1}^{N(i)} X_k - N(i) = i - j\right\} \\ &= P\left\{\sum_{k=1}^{N(i)} X_k = \frac{1}{2}[i - j + N(i)]\right\}. \end{aligned}$$

If $i - j + N(i)$ is odd, the transition from i to j is impossible, hence $p_{ij} = 0$. Alternatively, if $i - j + N(i)$ is even, since $\sum_{k=1}^{N(i)} X_k$ has a Binomial($N(i), \theta$) distribution, p_{ij} is the Binomial probability of having $[i - j + N(i)]/2$ successes in $N(i)$ trials, which can be easily calculated once $N(i)$ is chosen. Now we need

to define $N(i), i \in \{b, b+1, \ldots, a-1, a\}$. Of course, we have the boundary conditions $N(a) = N(b) = 0$, and only need to determine $N(i)$ for $i \in \{b+1, \ldots, a-1\}$.

For different SPPRTs, sample sizes $N(i)$ will be determined exactly. For certain choices of θ_0 and error probabilities, the sampling costs under H_0 will be calculated. Sequential plans introduced in Section 1 take the following forms in the case of SPPRT for Bernoulli family:

(a) Pure sequential plan: $N(i) = 1$, for any $i \in \{b+1, \ldots, a-1\}$;

(b) Conservative plan: $N(i) = \min\{a-i, i-b\}$;

(c) Expectation based plan: under $\theta = \theta_0$ ($\theta_0 > 1/2$), using Wald's lemma, the sample size $N(i)$ is given by

$$\begin{aligned} N(i) &= \min\left\{ \left\lceil \gamma \frac{(a-i)\Delta}{(2\theta_0 - 1)\Delta} \right\rceil, \left\lceil \gamma \frac{(i-b)\Delta}{(2\theta_0 - 1)\Delta} \right\rceil \right\} \\ &= \min\left\{ \left\lceil \gamma \frac{a-1}{2\theta_0 - 1} \right\rceil, \left\lceil \gamma \frac{i-b}{2\theta_0 - 1} \right\rceil \right\}, \end{aligned}$$

because $\boldsymbol{E} \log \frac{f(X|\theta_1)}{f(X|\theta_0)} = (2\theta - 1) \log \frac{\theta_1}{\theta_0}$. In general,

$$N(i) = \begin{cases} \min\left\{ \lceil \gamma \frac{a-i}{1-2\theta} \rceil, \lceil \gamma \frac{i-b}{1-2\theta} \rceil \right\} & \text{for } \theta < 1/2 \\ \min\left\{ \lceil \gamma \frac{a-i}{2\theta-1} \rceil, \lceil \gamma \frac{i-b}{2\theta-1} \rceil \right\} & \text{for } \theta > 1/2; \end{cases}$$

(d) m-conservative plans:

$$N(i) = \begin{cases} \min\{a-i, i-b\} & \text{if } \min\{a-i, i-b\} \geq m \\ \text{either } m, \text{ or } m-1, \text{ or} \ldots \text{or} \min\{a-i, i-b\} & \text{otherwise} \end{cases}$$

There are $(m!)^2$ such m-conservative plans including the k-conservative plans for $k < m$ and $(m!)^2 - (k!)^2$ m-conservative plans not including the k-conservative ones.

Using the theory developed in Sections 3-4, we calculate sampling costs of these sequential plans for the following practical example. Then, the risks of these plans can be compared, and the optimal plan will be chosen.

Example 4.17 The goal is to find an inexpensive sequential plan for testing if a certain proportion is at least 50%, with error probabilities at most $\alpha = \beta = 0.05$. Suppose that the fixed cost of each sample c_0 is \$1000 and the cost of each observation c is \$10, and consider the indifference region $(0.48, 0.52)$.

Table 1: Sampling Costs for Different Sequential Plans

Sequential plan	$\boldsymbol{E}(M)$	$\boldsymbol{E}(T)$	$c_0\boldsymbol{E}(T) + c\boldsymbol{E}(M)$	Savings
Pure Sequential	804	804	812,511	0
Conservative	804	44.8	52,857	759,654
4-Conservative	811	44.6	52,719	759,792
5-Conservative	812	44.37	52,499	760,012
Expectation Based, $\gamma = 0.05$	829	32	40,434	772,077
Expectation Based, $\gamma = 0.14$	854	16	25,015	787,494
Expectation Based, $\gamma = 0.17$	888	9.8	18,716	793,795
Expectation Based, $\gamma = 0.18$	892	9.3	18,254	794,257

Thus, we test

$$H_0 : \theta = 0.52 \text{ versus } H_1 : \theta = 0.48.$$

Using inequalities (4), we find the stopping boundaries $A \doteq (1 - 0.05)/0.05 = 19$ and $B = 0.05/(1 - 0.05) = 1/19$. As above, $\Delta = \log(\theta_0/\theta_1) = 0.08$, and thus $a = 36$ and $b = -36$.

Now we are ready to calculate sampling costs of the listed sequential plans. Since it involves rather large matrices, Matlab programs were used to compute the expected total costs according to (11) and (12).

Results are summarized in Table 1. The Savings column presents the difference in total expected costs between the SPRT and the other sequential plans. For a low observation cost relative to the fixed sample cost, we expect to find the optimal plan among m-conservative plans. Still, the average savings brought by the m-conservative plans are not big. With a small m, these plans are still conservative. In contrast, expectation based plans are more aggressive taking a smaller number of larger groups. It is interesting to notice that the considered expectation based plans progressively reduce the sampling cost, as γ increases, up to a point where increasing γ does not bring further improvement. The expectation based plan with $\gamma = 0.19$ does not yield improvement over previous plans. Thus, among expectation based plans, the optimal plan for this problem is the one with γ close to 0.18.

The above example illustrates how one can use the proposed methods in practice. Practitioners can

easily use the developed tools to evaluate performance of different sequential plans and choose the optimal one.

Of course, more general problems of testing θ_0 against θ_1, with θ_0 not necessarily equal to $1-\theta_1$, can be approached similarly. Also, other common discrete distributions discussed in this chapter may be assumed, and the same method for comparing the risks of different plans can be used.

5 Proofs

Proof of Theorem 3.1: It can easily be seen that $N_k(u+\Lambda_{N(u)}) = N_{k+1}(u)$, for any $k \geq 1$.

Then,

$$T(u+\Lambda_{N(u)}) = \inf\{k:\ N_k(u+\Lambda_{N(u)}) = 0\} - 1 = \inf\{k:\ N_{k+1}(u) = 0\} - 1 = T(u) - 1.$$

Solving for $T(u)$, and taking expected values of both sides,

$$\begin{aligned} g(u) &= 1 + \boldsymbol{E}^{\boldsymbol{Z}} T(u+\Lambda_{N(u)}) = 1 + \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \boldsymbol{E}^{\boldsymbol{Z}_{N(u)+1}} \{T(u+\Lambda_{N(u)}) | \boldsymbol{Z}^{N(u)}\} \\ &= 1 + \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} g(u+\Lambda_{N(u)}). \end{aligned}$$

The boundary condition results from the fact that $T(u) = 0$ for $u \notin (B, A)$. □

Proof of Theorem 3.2: Clearly, $M_{T(u)} = N(u) + M_{T(u+\Lambda_{N(u)})}$. The rest of the proof follows as above. □

Proof of Theorem 3.3: Let

$$I(u) = \begin{cases} 1 & \text{if } \Lambda_{M_T(u)} \geq A \\ 0 & \text{if } \Lambda_{M_T(u)} \leq B \end{cases}$$

Then, $\rho(u) = \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \boldsymbol{E}^{\boldsymbol{Z}_{N(u)+1}} (I(u+\Lambda_{N(u)}) | \boldsymbol{Z}^{N(u)}) = \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \rho(u+\Lambda_{N(u)})$. □

Proof of Theorem 3.4: Let $J(u) = 1 - I(u)$, and proceed similarly to the above, by noticing that

$$\tau(u) = \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \boldsymbol{E}^{\boldsymbol{Z}_{N(u)+1}}(J(u+\Lambda_{N(u)})|\boldsymbol{Z}^{N(u)}) = \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \tau(u+\Lambda_{N(u)}). \qquad \square$$

Proof of Theorem 3.5: Under H_0, $\boldsymbol{E}^{\boldsymbol{Z}} L_u(\theta,\delta) = w_\alpha \rho(u)$. Under H_1, $\boldsymbol{E}^{\boldsymbol{Z}} L_u(\theta,\delta) = w_\beta \tau(u)$. Then,

$$R_u(\theta,N,\delta) = w_\alpha \rho(u)\mathbf{1}_{\{H_0\ true\}} + w_\beta \tau(u)\mathbf{1}_{\{H_0\ false\}} + c_0 g(u) + c\nu(u) \tag{16}$$

Using the above results, we get

$$\begin{aligned} R_u(\theta,N,\delta) &= w_\alpha \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \rho(u+\Lambda_{N(u)}) I\{H_0\ true\} + w_\beta \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \tau(u+\Lambda_{N(u)}) I\{H_0\ false\} \\ &\quad + c_0 \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} g(u+\Lambda_{N(u)}) + c_0 + c\boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} \nu(u+\Lambda_{N(u)}) + cN(u) \\ &= c_0 + cN(u) + \boldsymbol{E}^{\boldsymbol{Z}^{N(u)}} R_{u+\Lambda_{N(u)}}(\theta,N,\delta) \end{aligned}$$

Evidently, the boundary conditions hold. Intuitively, it is clear that if the null hypothesis is true and the random walk starts on the boundary A, the null hypothesis is falsely rejected and no observations are taken. On, the other hand, if the random walk starts at B, the null hypothesis is rightly accepted and thus the loss is 0. The reasoning is similar for the case in which the null hypothesis is false. $\square$

Proof of Lemma 4.1: For all n, the size of group $n+1$, N_{n+1}, is completely determined by Y_n. Also,

$$Y_{n+1} = \begin{cases} Y_n + \left(\Lambda_{M_{n+1}} - \Lambda_{M_n}\right)/\Delta & \text{if} \quad Y_n + \left(\Lambda_{M_{n+1}} - \Lambda_{M_n}\right)/\Delta \in (b,a) \\ b & \text{if} \quad Y_n \in (b,a),\ Y_n + \left(\Lambda_{M_{n+1}} - \Lambda_{M_n}\right)/\Delta \leq b \\ a & \text{if} \quad Y_n \in (b,a),\ Y_n + \left(\Lambda_{M_{n+1}} - \Lambda_{M_n}\right)/\Delta \geq a \\ Y_n & \text{if} \quad Y_n \notin (b,a) \end{cases}$$

is a function of Y_n and $\left(\Lambda_{M_{n+1}} - \Lambda_{M_n}\right)$ only. Since

$$\Lambda_{M_{n+1}} - \Lambda_{M_n} = \sum_{k=M_n+1}^{M_{n+1}} \frac{dF_1}{dF_0}(X_k)$$

is independent of $(Y_1,\ldots,Y_n)$,

$$P\{Y_{n+1} = u|Y_1,\ldots,Y_n\} = P\{Y_{n+1} = u|Y_n\}$$

for all integer n and u. Hence, $\{Y_n\}$ is a stationary Markov chain with absorbing states a and b. □

Proof of Lemma 4.2: Since the log-likelihood ratio is a random walk, under H_0 it is sufficient to show that a random walk with a negative drift a.s. spends finite time above any level B (and therefore, it eventually stays below B forever), and, under the alternative, that a random walk with a positive drift a.s. spends finite time below any level A (and therefore, it eventually stays above A forever).

It is enough to solve one case, the other one will follow similarly. Let us a assume the alternative, H_1, that is the case where the log-likelihood ratio, $\Lambda_n = \sum z_j$, is a random walk with positive drift.

Under these conditions, according to Woodroofe (1982) (Corollary 2.5 of [14]), if z_j have finite variance,

$$\sum_{n=1}^{\infty} P\{\Lambda_n \le 0\} < \infty.$$

Also, as stated by the *renewal theorem* (Theorem 2.1 of [14]), under the above conditions (even without the requirement of finite variance)

$$\sum_{n=1}^{\infty} P\{0 \le \Lambda_n \le A\} < \infty.$$

for any finite A.

It follows that

$$\sum_{n=1}^{\infty} P\{\Lambda_n \le A\} < \infty.$$

The series converges, hence by Borel-Cantelli lemma, the event $\{\Lambda_n \le A\}$ occurs only a finite number of times with probability 1. Moreover, this series, which is called *renewal measure* of $(-\infty, A]$, is nothing but the expected number of visits of the random walk to the interval $(-\infty, A]$:

$$\nu(-\infty, A] = \boldsymbol{E}(\text{card}\{n : \Lambda_n \le A\}) = \boldsymbol{E}\sum_{0}^{\infty} 1_{\Lambda_n \le A} = \sum_{0}^{\infty} P\{\Lambda_n \le A\} < \infty.$$

□

Proof of Lemma 4.3: Consider the entire transition probability matrix for the Markov chain Y_n, $\tilde{P} = (p_{ij})$, where $i, j = b, b+1, \ldots, a-1, a$, which is actually, matrix $\boldsymbol{P}$ augmented with two rows and two

columns corresponding to the transition probabilities to and from states a and b. States a and b being absorbing states, the transition probability matrix, can be written in a simplified form:

$$\tilde{P} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ p_{b+1,b} & & & & p_{b+1,a} \\ \vdots & & P & & \vdots \\ p_{a-1,b} & & & & p_{a-1,a} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

It follows by induction that the k-th order transition probability matrix is given by

$$\tilde{P}^k = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ p^{(k)}_{b+1,b} & & & & p^{(k)}_{b+1,a} \\ \vdots & & P^k & & \vdots \\ p^{(k)}_{a-1,b} & & & & p^{(k)}_{a-1,a} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

and thus,

$$\boldsymbol{P}^k = (p^{(k)}_{ij}), \text{ where } i, j = b+1, \dots, a-1.$$

It is shown that $\boldsymbol{P}^k \to \boldsymbol{O}$ as $k \to \infty$, where $\boldsymbol{O}$ is the zero matrix.

Let i and j be any states between the stopping boundaries and thus transient states (due to the fact that the stopping time is proper). Also, let $f^{(\nu)}_{ij}$ be the probability of a first visit to j at time ν for a system that starts at i. Using the same argument as in the proof of theorem 8.3 in Billingsley (1986),

$$\sum_{k=1}^{\infty} p^{(k)}_{ij} = \sum_{k=1}^{\infty} \sum_{\nu=1}^{k} f^{(\nu)}_{ij} p^{(k-\nu)}_{jj} = \sum_{\nu=1}^{\infty} f^{(\nu)}_{ij} \sum_{m=0}^{\infty} p^{m}_{jj} \leq \sum_{m=0}^{\infty} p^{m}_{jj}.$$

By theorem 8.2 in Billingsley (1986), transience of state j is equivalent to $\sum_m p^{(m)}_{ii} < \infty$. It follows immediately that $\lim_{k\to\infty} p^{(k)}_{ij} = 0.$ for any stationary states i and j. Moreover, this convergence is uniform.

It will be shown that $\sum_{k=0}^{\infty} \boldsymbol{P}^k$ is convergent.

Evidently $\lim_{k\to\infty} \|\boldsymbol{P}^k\| = 0$. Then, there exists a positive integer K such that $\|\boldsymbol{P}^k\| \leq 1/2$ for any $k \geq K$.

For $k = 0, \dots, K-1,\ \|\boldsymbol{P}^k\| \leq C = \max_{0,\dots,K-1} \|\boldsymbol{P}^k\|$.

For $k = K, \ldots, 2K-1,\ \|\boldsymbol{P}^k\| \leq 1/2$.

For $k = 2K, \ldots, 3K-1,\ \|\boldsymbol{P}^k\| \leq \|\boldsymbol{P}^K\| \cdot \|\boldsymbol{P}^{k-K}\| \leq (1/2) \cdot (1/2) = 1/4$.

For $k = 3K, \ldots, 4K-1,\ \|\boldsymbol{P}^k\| \leq \|\boldsymbol{P}^K\| \cdot \|\boldsymbol{P}^{k-K}\| \leq (1/2) \cdot (1/4) = 1/8$.

$\vdots$

For $k = nK, \ldots, (n+1)K-1,\ \|\boldsymbol{P}^k\| \leq \|\boldsymbol{P}^K\| \cdot \|\boldsymbol{P}^{k-K}\| \leq (1/2) \cdot (1/2)^{n-1} = (1/2)^n$.

$\vdots$

Then,

$$\sum_{k=0}^{\infty} \|P^k\| \leq K\left(C + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots\right) = KC + K\sum_{k=1}^{\infty}\left(\frac{1}{2}\right)^k < \infty$$

Since the series $\sum_{k=0}^{\infty} \|P^k\|$ is convergent, it is Cauchy and thus $\sum_{k=0}^{n+m} \|\boldsymbol{P}^k\| - \sum_{k=0}^{n} \|\boldsymbol{P}^k\| = \sum_{k=n+1}^{n+m} \|\boldsymbol{P}^k\|$ goes to 0 as $n \to \infty$. Also, $\|\sum_{k=0}^{n+m} \boldsymbol{P}^k - \sum_{k=0}^{n} \boldsymbol{P}^k\| = \|\sum_{k=n+1}^{n+m} \boldsymbol{P}^k\| \leq \sum_{k=n+1}^{n+m} \|\boldsymbol{P}^k\|$. Then, $\|\sum_{k=0}^{n+m} \boldsymbol{P}^k - \sum_{k=0}^{n} \boldsymbol{P}^k\|$ goes to 0 as well, as $n \to \infty$, which implies that $\|\sum_{k=0}^{\infty} \boldsymbol{P}^k\|$ is Cauchy and thus convergent. Hence, $\sum_{k=0}^{\infty} \boldsymbol{P}^k$ is convergent. Then,

$$\begin{aligned}
(\boldsymbol{I} - \boldsymbol{P})\sum_{k=0}^{\infty} \boldsymbol{P}^k &= \sum_{k=0}^{\infty}(\boldsymbol{P}^k - \boldsymbol{P}^{k+1}) = \lim_{n\to\infty}\sum_{k=0}^{n}(\boldsymbol{P}^k - \boldsymbol{P}^{k+1}) \\
&= \lim_{n\to\infty}[(\boldsymbol{I} - \boldsymbol{P}) + (\boldsymbol{P} - \boldsymbol{P}^2) + \cdots + (\boldsymbol{P}^n - \boldsymbol{P}^{n+1})] \\
&= \boldsymbol{I} - \lim_{n\to\infty} \boldsymbol{P}^{n+1} = \boldsymbol{I} - \boldsymbol{O} = \boldsymbol{I}
\end{aligned}$$

Therefore $\boldsymbol{I} - \boldsymbol{P}$ is invertible and $(\boldsymbol{I} - \boldsymbol{P})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{P}^k$. □

Proof of Theorem 4.4: Using equation (6) and the fact that the SPPRT is on a lattice,

$$g(u) = 1 + \boldsymbol{E}g(u + \Lambda_{N(u)}) = 1 + \sum_{k=-\infty}^{\infty} p_{uk}g(k) = 1 + \sum_{k=b+1}^{a-1} p_{uk}g(k).$$

In matrix notation,

$$\mathbf{g} = \mathbf{1} + \mathbf{P}\mathbf{g} \Rightarrow \mathbf{g} = (\boldsymbol{I} - \boldsymbol{P})^{-1}\mathbf{1}.$$

□

The proofs for Theorems 4.5, 4.6, and 4.7 are analogous to the proof above.

Proof of Example 4.10: Suppose the SPPRT is on a lattice.

As the following lemma shows, in order to prove that an SPPRT is on a lattice, it is enough to show that Λ_1 is on a lattice.

Lemma 5.1 *If $X_1, \ldots, X_n$ are iid random variables, then Λ_n is on a lattice for any n if and only if Λ_1 is on a lattice.*

Proof: It is obvious that Λ_1 is on a lattice if Λ_n is on a lattice for any n (let $n = 1$). Now suppose that Λ_1 is on a lattice. Then, $\Lambda_1 = \frac{f(X_1|\theta_1)}{f(X_1|\theta_0)} \in \mathcal{L}_\Delta = \{k\Delta, k \in \mathbb{Z}\}$. Since $X_1, \ldots, X_n$ are iid, $\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \in \mathcal{L}_\Delta$ for any $i = 1, \ldots, n$ and thus, $\Lambda_n = \sum_{i=1}^n \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \in \mathcal{L}_\Delta$. □

Then, $\Lambda_1 \in \mathcal{L}_\Delta = \{k\Delta | k \in \mathbb{Z}\}$, where Δ is a real number. For binomial X_1, $\Lambda_1 = \log \frac{\theta_1^{X_1}(1-\theta_1)^{m-X_1}}{\theta_0^{X_1}(1-\theta_0)^{m-X_1}} = X_1 \log \frac{\theta_1}{\theta_0} + (m - X_1) \log \frac{1-\theta_1}{1-\theta_0}$, where $X_1 = 0, 1, \ldots, m$. If $X_1 = 0$, then $\Lambda_1 = m \log \frac{1-\theta_1}{1-\theta_0} \in \mathcal{L}_\Delta$. Thus $m \log \frac{1-\theta_1}{1-\theta_0} = u\Delta$, for some nonzero integer u. If $X_1 = m$, then $\Lambda_1 = m \log \frac{\theta_1}{\theta_0} \in \mathcal{L}_\Delta$. Thus $m \log \frac{\theta_1}{\theta_0} = v\Delta$, for some nonzero integer v. Therefore, $r = \frac{u}{v} \in \boldsymbol{Q}$.

Conversely, let us suppose that $r = \log \frac{1-\theta_1}{1-\theta_0} / \log \frac{\theta_1}{\theta_0}$ is rational number and show that the SPPRT is on a lattice. Notice that $r < 0$ for any $\theta_0 \neq \theta_1$, and let $r = -\frac{p}{q}$, where p and q are positive integers. Then, $\Lambda_1 = X_1 \log \frac{\theta_1}{\theta_0} + r(m - X_1) \log \frac{\theta_1}{\theta_0} = [qX_1 - p(m - X_1)] \frac{1}{q} \log \frac{\theta_1}{\theta_0} \in \mathcal{L}_\Delta$, where $\Delta = \frac{1}{q} \log \frac{1-\theta_1}{1-\theta_0}$. Thus, the SPPRT is on a lattice. □

Examples 4.10, 4.11, and 4.12 follow similarly.

Proof of Lemma 4.14: Using the definition of the log-likelihood ratio for the density function $f(x|\theta) = g(x)h(\theta)\mathbf{1}_{x \in A(\theta)}$, the following is obtained.

$$\Lambda_n = \begin{cases} n \log \frac{h(\theta_1)}{h(\theta_0)} & \text{if } x_i \in A(\theta_1) \cap A(\theta_0) \text{ for all } i = 1, \ldots, n \\ \infty & \text{if } x_n \in A(\theta_1) - A(\theta_0), \text{ and } \Lambda_{n-1} = (n-1) \log \frac{h(\theta_1)}{h(\theta_0)} \text{ or } \Lambda_{n-1} = \infty \\ -\infty & \text{if } x_n \in A(\theta_0) - A(\theta_1), \text{ and } \Lambda_{n-1} = (n-1) \log \frac{h(\theta_1)}{h(\theta_0)} \text{ or } \Lambda_{n-1} = -\infty \end{cases}$$

Then, $\Lambda_n = n\Delta$, where $\Delta = \frac{h(\theta_1)}{h(\theta_0)}$, and thus any SPPRT for this problem is on a lattice. □

References

[1] J. O. Berger. *Statistical Decision Theory*. Springer-Verlag, New York, 1985.

[2] P. Billingsley. *Probability and Measure*. Wiley, New York, 1986.

[3] Z. Govindarajulu. *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*. American Sciences Press, Columbus, Ohio, 1981.

[4] L. S. Hayre. Group sequential sampling with variable group sizes. *J. Royal Statist. Soc. B*, 47(1):90–97, 1985.

[5] C. Jennison and B. W. Turnbull. Interim analyses: the repeated confidence interval approach. *J. Royal Statist. Soc. B*, 51(3):305–361, 1989.

[6] S. R. Marschner, S. H. Westin, E. P.F. Lafortune, and K. Torrance. Image-based brdf measurement. *Applied Optics*, 39(16):2592–2600, 2000.

[7] I. Olkin. A conversation with W. Allen Wallis. *Statistical Science*, 6:121–140, 1991.

[8] C. Schmegner. *Decision Theoretic Results for Sequentially Planned Statistical Procedures*. Ph. D. Dissertation, The University of Texas at Dallas, 2003.

[9] C. Schmegner and M. Baron. Principles of optimal sequential planning. *Sequential Analysis*, 23(1):11–32, 2004.

[10] N. Schmitz. *Optimal Sequentially Planned Decision Procedures*. Springer-Verlag, New York, 1993.

[11] A. Wald. *Sequential Analysis*. Wiley, New York, 1947.

[12] W. A. Wallis. The statistical research group, 1942-1945. *J. Amer. Statist. Assoc.*, 75:320–330, 1980.

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

[14] M. Woodroofe. *Nonlinear Renewal Theory in Sequential Analysis*. CBMS-NSF Regional Conference Series in Applied Mathematics, Philadehia, PA, 1982.