Optimal group sequential tests with groups of random size

A. Novikov and X. I. Popoca-Jiménez

Department of Mathematics, Metropolitan Autonomous University–Iztapalapa, Mexico City, Mexico

ABSTRACT
We consider sequential hypothesis testing based on observations that are received in groups of random size. The observations are assumed to be independent both within and between the groups. We assume that the group sizes are independent and their distributions are known and that the groups are formed independent of the observations.

We are concerned with a problem of testing a simple hypothesis against a simple alternative. For any (group) sequential test, we take into account the following three characteristics: its type I and type II error probabilities and the average cost of observations. Under mild conditions, we characterize the structure of sequential tests minimizing the average cost of observations among all sequential tests whose type I and type II error probabilities do not exceed some prescribed levels.

ARTICLE HISTORY
Received 27 October 2021
Revised 28 February 2022
Accepted 1 April 2022

KEYWORDS
Hypothesis testing; optimal sequential tests; optimal stopping; sequential analysis

SUBJECT CLASSIFICATIONS
62L10; 62L15; 62F03; 60G40

1. INTRODUCTION
In this article, we consider sequential hypothesis testing when the observations are received in groups of a random size, rather than on a one-at-a-time basis (we adhere to the statistical model proposed by Mukhopadhyay and de Silva [2008] for this context). There are many practical situations where the random group size model comes into question and a lot of theoretical problems arise (see Mukhopadhyay and de Silva 2008).

In this article, we address the problems of optimality of the random group sequential tests for the case of two simple hypotheses, covering theoretical aspects of their optimality.

In the case of random groups of observations, there are different ways to quantify the volume of observations taken for the analysis; for example, a special interest can be put on the total number of observations or on the number of groups taken (see Mukhopadhyay and de Silva 2008). To tackle the possible differences, in this article we introduce a natural concept of cost of observations accounting for the number of groups and/or for the number of observations within the groups and use the average cost as one of the characteristics to be taken into account.

Our main objective is to characterize all of the tests minimizing the average cost of the experiment among all group sequential tests whose type I and type II error probabilities do not exceed some given levels.

CONTACT X. I. Popoca-Jiménez dextera80@yahoo.es Departamento de Matemáticas, UAM-Iztapalapa, San Rafael Atlixco 186, col. Vicentina, Mexico City, C.P. 09340, Mexico.
© 2022 Taylor & Francis Group, LLC
With respect to these, we should start with the classical framework of one-per-group observations where the average sample number is minimized under restrictions on the probabilities of the first and the second kinds. Wald and Wolfowitz (1948) showed that Wald’s sequential probability ratio test (SPRT) has a minimum average sample number among all tests whose error probabilities do not exceed those of the SPRT. The minimum is reached both under the null hypothesis and under the alternative. This strong optimality property is known as the Wald-Wolfowitz optimality.

For the group sequential model we adhere to in this article, Mukhopadhyay and de Silva (2008) proposed an extension of the classical SPRT called RSPRT. In this article, we want to characterize the structure of optimal sequential tests and, in particular, show the optimality of the RSPRT, in the Wald-Wolfowitz sense, when the group sizes are identically distributed.

For a more general case, when the group sizes do not necessarily have the same distribution, we use a weaker approach related to the minimization of the average cost under one hypothesis (see Lorden 1980). If an optimal, in the Wald-Wolfowitz sense, test is ever found, for any specific group size distribution model, it should minimize the average cost under each one of the hypotheses; thus, it should be of the particular form we find here.

Thus, our main concern in this article is the characterization of optimal group sequential tests that minimize the average cost of the observations given restrictions on the error probabilities.

In Section 2, the main definitions and assumptions are presented. In Section 3, the problem of finding an optimal test is reduced to an optimal stopping problem. In Section 4, characterizations of optimal sequential rules are given. In Section 5, the optimality of the random sequential probability ratio tests (RSPRT) is demonstrated.

The proofs of the main results are provided in the appendix.

2. NOTATION AND ASSUMPTIONS

We assume that independent and identically distributed (i.i.d.) observations $X_{kj}$, $j = 1, 2, \ldots, n_k$, are available to the statistician sequentially, in groups numbered by $k = 1, 2, \ldots$. The group sizes $n_k$ are assumed to be values of some independent integer-valued random variables $\nu_k$. The distributions of $\nu_k$ are assumed to be fixed and known to the statistician. But the distribution of $X_{kj}$ (we denote it by $P_{\theta}$) depends on a parameter $\theta$, and the goal of the statistician is to test two simple hypotheses, $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. The number of groups to be taken for the analysis is up to the statistician and is to be determined on the basis of observations available up to the moment of stopping.

We formalize this procedure in detail below.

For any natural $k$, we denote by $X_k^{(n_k)}$ the vector of observations in the $k$th group, $X_k^{(n_k)} = (X_{k1}, \ldots, X_{kn_k})$. If $n_k = 0$ (no observations in the group), we will formally write it as (). Group sizes $n_1, \ldots, n_k$ are assumed to be values of independent variables $\nu_1, \ldots, \nu_k$, with respective probability mass functions (p.m.f.s)

$$P(\nu_i = n) = p_i(n), \quad n \in G \subset \{0, 1, 2, \ldots\},$$

(2.1)

$i = 1, 2, \ldots, k; \ k = 1, 2, \ldots$. Then the joint distribution of the first $k$ consecutive group sizes (let us denote $\nu^{(k)} = (\nu_1, \ldots, \nu_k)$ and $n = (n_1, \ldots, n_k) \in G^k$) is given by
where for all \( l \) with respect to the product measure \( \mu \), it is assumed that the observations \( X_{n_l} \) are independent and identically distributed, given group sizes \( \nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_k = n_k \), for all \( n = (n_1, \ldots, n_k) \in G^k \), \( k = 1, 2, \ldots \).

Let \( X_{n_l} \) take its “values” in a measurable space \((X, \mathcal{X})\), and assume that its distribution \( P_0 \) has a “density” function \( f_0 \) on \( X \) (the Radon-Nikodym derivative of \( P_0 \) with respect to a \( \sigma \)-finite measure \( \mu \) on \( X \)).

In this way, for any number of groups \( k = 1, 2, \ldots \), given any consecutive group sizes \( n = (n_1, \ldots, n_k) \in G^k \), the random vector of observations \( X^{(n)} = (X_1^{(n_1)}, \ldots, X_k^{(n_k)}) \) has a joint density

\[
    f_0^{(n)}(x^{(n)}) = f_0^{(n)}(x_1^{(n_1)}, \ldots, x_k^{(n_k)}) = \prod_{i=1}^{k} \prod_{j=1}^{n_k} f_0(x_{ij})
\]

with respect to the product measure \( \mu^n = \mu^{n_1} \otimes \cdots \otimes \mu^{n_k} \), on \( X^n = X_1^{n_1} \otimes \cdots \otimes X_{n_k}^{n_k} \), where \( \mu^{n_k} = \mu \otimes \cdots \otimes \mu \) (\( n_k \) times). By definition, we assume here that \( \prod_{j=1}^{0} (\cdot) = 1 \) and that \( \mu^0 \) is a probability measure on the (trivial) \( \sigma \)-algebra \( X^0 \) on \( \mathbb{X}^0 = \{ () \} \).

Throughout the article, it will be assumed that the distributions \( P_{\theta_0} \) and \( P_{\theta_i} \) are distinct:

\[
    \mu\{ x : f_{\theta_0}(x) \neq f_{\theta_i}(x) \} > 0. \tag{2.2}
\]

We define a (randomized) stopping rule \( \psi \) as a family of measurable functions \( \psi_n : \mathbb{X}^n \to [0, 1], n \in G^k, k \geq 1 \), where \( \psi_n(x^{(n)}) \) represents the conditional probability to stop, given number of groups \( k \), group sizes \( n = (n_1, \ldots, n_k) \), and the data \( x^{(n)} \) observed until the stopping time.

In a similar manner, a (randomized) decision rule \( \phi \) is a family of measurable functions \( \phi_n : \mathbb{X}^n \to [0, 1], n \in G^k, k \geq 1 \), where \( \phi_n(x^{(n)}) \) represents the conditional probability to reject \( H_0 \), given a number \( k \) of the groups observed, group sizes \( n = (n_1, \ldots, n_k) \), and data \( x^{(n)} \) observed until the time of final decision (to accept or reject \( H_0 \)).

A group sequential test is a pair \((\psi, \phi)\) of a stopping rule \( \psi \) and a decision rule \( \phi \).

For \( n = (n_1, \ldots, n_k) \in G^k \), with any \( k \geq 1 \), let

\[
    t_n^{\psi}(x^{(n)}) = \left( 1 - \psi_{(n_1)}(x_1^{(n_1)}) \right) \cdots \left( 1 - \psi_{(n_1, \ldots, n_{k-1})}(x_1^{(n_1)}, \ldots, x_{k-1}^{(n_{k-1})}) \right), \tag{2.3}
\]

\[
    s_n^{\psi}(x^{(n)}) = t_n^{\psi}(x^{(n)}) \psi_{(n)}(x^{(n)}). \tag{2.4}
\]

(by definition, \( t_n^{\psi}(x^{(n)}) \equiv 1 \) for \( n \in G \)).

Let us also denote for \( n \in G^k, k \geq 1 \),

\[
    S_n^{\psi} = \left\{ x^{(n)} \in \mathbb{X}^n : s_n^{\psi}(x^{(n)}) > 0 \right\} \quad \text{and} \quad T_n^{\psi} = \left\{ x^{(n)} \in \mathbb{X}^n : t_n^{\psi}(x^{(n)}) > 0 \right\}.
\]

In the latter expression, we assume \( s_n^{\psi} = s_n^{\psi}(X^{(n)}) \), despite its initial definition as \( s_n^{\psi}(x^{(n)}) \). Throughout the article, we will use this kind of double interpretation for any function of observations, according to the following rule. If \( F_n, n \in G^k \), is some function of observations and its arguments are omitted, then \( F_n \) is interpreted as \( F_n(X^{(n)}) \) when it is under the probability or expectation sign and as \( F_n(x^{(n)}) \) otherwise.
Any stopping rule $\psi$ generates a random variable $\tau_\psi$ (stopping time), with a p.m.f.
\[
P_\theta(\tau_\psi = k) = \sum_{n \in G} p(n) E_0 s_n^\psi, \quad k = 1, 2, \ldots,
\]
where $E_0(\cdot)$ is expectation with respect to $P_\theta$. Under $H_n$ a stopping rule $\psi$ terminates the testing procedure with probability 1 if
\[
P_\theta_i(\tau_\psi < \infty) = \sum_{k=1}^\infty \sum_{n \in G} p(n) E_0 s_n^\psi = 1. \tag{2.5}
\]
Let us denote $F_i$ as the set of stopping rules satisfying (2.5), and let $S_i$ be the set of all group sequential tests $(\psi, \phi)$ such that $\psi \in F_i$, $i = 0, 1$.

For brevity, we will write $E_0, f_0, P_i, F_0, F_1, S_0, S_1$ instead of $E_0, f_0, P_i, F_0, F_1, S_0, S_1$, respectively, $i = 0, 1$. For a group sequential test $(\psi, \phi)$, the type I and type II error probabilities are defined as
\[
\alpha(\psi, \phi) = P_0(\text{reject } H_0 \text{ using } (\psi, \phi)) = \sum_{k=1}^\infty \sum_{n \in G} p(n) E_0 s_n^\psi \phi_n, \quad \text{and}
\]
\[
\beta(\psi, \phi) = P_1(\text{accept } H_0 \text{ using } (\psi, \phi)) = \sum_{k=1}^\infty \sum_{n \in G} p(n) E_1 s_n^\psi (1 - \phi_n).
\]
Let us assume that the cost of $m$ observations is $c(m) \geq 0, m \in G$. The cost of the first $k$ stages of the experiment for a given $n \in G_k$ is then $c(n) = \sum_{i=1}^k c(n_i)$. The average cost of obtaining the $k$th group of observations is
\[
\tau_k = \sum_{n \in G} p_k(n)c(n).
\]
We assume that $0 \leq \tau_k < \infty$ for all $k \geq 1$. The average total cost of the sequential sampling according to a stopping rule $\psi \in F_i$ is, under $H_i$,
\[
K_i(\psi) = \sum_{k=1}^\infty \sum_{n \in G} p(n)c(n) E_i s_n^\psi, \quad i = 0, 1.
\]
Here are some natural particular cases of the cost structure: if $c(m) = 1$ for all $m \in G$, then $K_i(\psi)$ corresponds to the average number of the groups taken; another particular case is $c(m) = m$ for all $m \in G$, which accounts for the average total number of observations. A combination of these two seems to be quite useful for some applications: $c(m) = a + bm$, with $a, b > 0$ (see, for example, Schmitz 1993).

It is seen from the above definitions that our group sequential experiment always starts with one group of observations (i.e., stage $k = 1$ is always present). The usual (for the sequential analysis) case when the experiment admits stopping without taking observations can be easily incorporated in this scheme by supposing that the first group always has size $n_1 = 0$ (and stopping at this stage means no observations will be taken).

The usual context for sequential hypothesis testing is to minimize average costs under restrictions on the type I and type II error probabilities. We are concerned in this article with minimizing $K_0(\psi)$ and/or $K_1(\psi)$ under restrictions
\[
\alpha(\psi, \phi) \leq \alpha \quad \text{and} \quad \beta(\psi, \phi) \leq \beta, \tag{2.6}
\]
over $(\psi, \phi) \in S_0$ (and/or $(\psi, \phi) \in S_1$), where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ are some given numbers.
3. REDUCTION TO OPTIMAL STOPPING PROBLEM

The problem of minimizing average cost $K_0(\psi)$ under restrictions on the error probabilities $\alpha(\psi, \phi) \leq \alpha$ and $\beta(\psi, \phi) \leq \beta$ is routinely reduced to a nonconstrained optimization problem using the Lagrange multipliers method.

Let us define the Lagrangian function

$$L(\psi, \phi; \lambda_0, \lambda_1) := K_0(\psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi),$$

where $\lambda_0 > 0$ and $\lambda_1 > 0$ are constant multipliers.

The following lemma is the essence of the reduction and is almost trivial. It is placed here for convenience of references.

**Lemma 3.1.** Let $\lambda_0, \lambda_1 > 0$ and a test $(\psi, \phi) \in S_0$ be such that

$$L(\psi, \phi; \lambda_0, \lambda_1) \leq L(\psi', \phi'; \lambda_0, \lambda_1)$$

for all $(\psi', \phi') \in S_0$. Then for every test $(\psi', \phi') \in S_0$ such that

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi),$$

it holds that

$$K_0(\psi') \geq K_0(\psi).$$

The inequality in (3.4) is strict if at least one of the inequalities in (3.3) is strict.

Let us define $g(z) = g(z; \lambda_0, \lambda_1) = \min\{\lambda_0, \lambda_1 z\}$ for all $0 \leq z < \infty$, and let $\mathcal{P}_k = \{n \in G^k : P(n) > 0\}$, for $k \geq 1$.

For all $k \geq 1$ and all $n \in G^k$, let

$$z_n = z_n(x^{(n)}) = \begin{cases} f_1^{(n)}(x^{(n)}) / f_0^{(n)}(x^{(n)}) & \text{if } f_0^{(n)}(x^{(n)}) > 0, \\ \infty & \text{if } f_0^{(n)}(x^{(n)}) = 0, \text{ but } f_1^{(n)}(x^{(n)}) > 0, \\ 0 \text{ (or whatever)} & \text{otherwise.} \end{cases}$$

From this time on, we will use the following notation.

Let us write $\psi \sim I_{\{F_1 \leq F_2\}}$ (say) when $I_{\{F_1 < F_2\}} \leq \psi \leq I_{\{F_1 \leq F_2\}}$. It is easy to see that this is an equivalent way to say that $\psi = 1$ if $F_1 < F_2$, $\psi = 0$ if $F_1 > F_2$, and $0 \leq \psi \leq 1$ when $F_1 = F_2$.

If $\psi, F_1,$ and $F_2$ are functions of some arguments, this agreement should be applied to their values calculated at any given arguments.

**Theorem 3.1.** For all tests $(\psi, \phi) \in S_0$ it holds that

$$L(\psi, \phi; \lambda_0, \lambda_1) \geq \sum_{k=1}^{\infty} \sum_{n \in G^k} P(n) E_0 s_n^\psi(c(n) + g(z_n; \lambda_0, \lambda_1)).$$

There is an equality in (3.5) if and only if the following condition is satisfied:

**Condition $\mathfrak{T}(\psi, \phi).** For all $k \geq 1$ and for all $n \in \mathcal{P}_k$, $\phi_n \sim I_{\{\lambda_0/\lambda_1 = z_n\}}$ $P_0$-a.s. on $S_n^\psi$.  

$$\phi_n \sim I_{\{\lambda_0/\lambda_1 = z_n\}} \quad P_0\text{-a.s. on } S_n^\psi. $$
Inequality (3.5) provides a lower bound for the Lagrangian function of any sequential test in $S_0$. Thus, any test pretending to minimize the Lagrangian function should attain the right-hand side of (3.5). And Condition $D(\psi, \phi)$ is necessary and sufficient for this: the decision should be in favor of $H_1$ whenever the likelihood $z_n$ is greater than (or equal to) $\lambda_0/\lambda_1$.

The proof of Theorem 3.1 can be carried out along the lines of the proof of theorem 2.2 in Novikov (2009).

**Theorem 3.1** is the first step in the Lagrange minimization we discussed earlier in this section. It reduces the problem of minimization of $L(\psi, \phi; \lambda_0, \lambda_1)$ to that of minimization of

$$
L(\psi; \lambda_0, \lambda_1) = \inf_\phi L(\psi, \phi; \lambda_0, \lambda_1) = \sum_{k=1}^{\infty} \sum_{n \in G_k} p(n) E_0 s_w^k (c(n) + g(z_n; \lambda_0, \lambda_1))
$$

(3.7)

over all stopping rules $\psi \in F_0$.

Indeed, if we have a stopping rule $\psi$ minimizing $L(\psi; \lambda_0, \lambda_1)$ over all stopping rules in $F_0$, then, combining the optimal $\psi$ with any decision rule $\phi$ satisfying

$$
\phi_n \simeq I_{(\lambda_0/\lambda_1 < z_n)} \quad \text{for all } n \in G^k \text{ and } k = 1, 2, ...,$$

(cf. (3.6)), we obtain

$$
L(\psi, \phi; \lambda_0, \lambda_1) = L(\psi; \lambda_0, \lambda_1) \leq L(\psi'; \lambda_0, \lambda_1) = L(\psi', \phi; \lambda_0, \lambda_1) \leq L(\psi', \phi'; \lambda_0, \lambda_1).
$$

Consequently, (3.2) is satisfied for all $(\psi', \phi') \in S_0$.

In such a way, starting from this moment, our focus will be on the problem of minimizing $L(\psi; \lambda_0, \lambda_1)$ over all stopping rules $\psi$.

In the meanwhile, a useful consequence of Theorem 3.1 can already be drawn for a particular (in fact, nonsequential) case when the number of groups $N$ is fixed in advance (and $s_w^k \equiv 1$ for all $n \in G^N$). Combining the result of Theorem 3.1 with Lemma 3.1, one immediately obtains an alternative proof of theorem 2.1 in Mukhopadhyay and de Silva (2008).

### 4. OPTIMAL RANDOM GROUP SEQUENTIAL TESTS

#### 4.1. Optimal stopping on a finite horizon

For any $N \geq 1$, we define the class of truncated stopping rules $F^N$ as

$$
F^N = \{ \psi : (1 - \psi_{n_1})(1 - \psi_{n_2})... (1 - \psi_n) \equiv 0 \quad \text{for all } n \in G^N \}.
$$

(4.1)

Let $S^N$ be the set of all group sequential tests $(\psi, \phi)$ with $\psi \in F^N$.

For $\psi \in F^N$, let us denote $L_N(\psi) = L_N(\psi; \lambda_0, \lambda_1) = L(\psi; \lambda_0, \lambda_1)$ (see 3.7).

In this section, we characterize the structure of all stopping rules $\psi \in F^N$ that minimize $L_N(\psi)$ over all $\psi \in F^N$.

It follows from (4.1) that

$$
L_N(\psi) = \sum_{k=1}^{N-1} \sum_{n \in G_k} p(n) E_0 s_w^k (c(n) + g(z_n)) + \sum_{n \in G_N} p(n) E_0 s_w^k (c(n) + g(z_n)).
$$

(4.2)
This expression is just a result of plugging the optimal decision rule $\phi$ from Theorem 3.1 into the Lagrangian function $L_N(\psi, \phi; \lambda_0, \lambda_1)$.

Because the expression in (4.2) depends only on the stopping rule $\psi$, its minimization is an optimal stopping problem, so the solution is through the following variant of the backward induction.

Let the functions $V^N_k = V^N_k(z), z \geq 0, k = 1, 2, \ldots, N$, be defined in the following way: starting from

$$V^N_N(z) = g(z),$$

(4.3)

define recursively, for all $z \geq 0$,

$$V^N_{k-1}(z) = \min \{g(z), \overline{\tau}_k + \sum_{n \in G} p_k(n) E_0 V^N_k(z z_n)\}, \quad k = N, N - 1, \ldots, 2.$$  

(4.4)

Functions $V^N_i(z)$ are obtained by a variant of “backward induction” and provide a minimum value to the partial sum starting from $k = i$ in the expression for $L_N(\psi)$ in (4.2), for a given value $z$ of $z_i, i = 1, 2, \ldots, N - 1$ (see the details in Novikov 2009).

Let us denote

$$V^N_k(z) = \sum_{n \in G} p_k(n) E_0 V^N_k(z z_n), \quad z \geq 0, \quad k = 1, 2, \ldots, N$$

(4.5)

(see the expression for optimal stopping rule in 4.4).

It is important to bear in mind that all of the functions above are constructed on the basis of the constants $\lambda_0 > 0$ and $\lambda_1 > 0$ and some definite cost structure ($c(m)$). Unfortunately, there is no satisfactory way to make them all explicit in the notation, so we leave them implicit in all of the elements of the construction.

The following theorem characterizes the structure of all truncated stopping rules minimizing $L_N(\psi)$.

**Theorem 4.1.** For every $\psi \in F^N$,

$$L_N(\psi) \geq \overline{\tau}_1 + V^N_1(1).$$

(4.6)

There is an equality in (4.6) if $\psi \in F^N$ satisfies the following:

**Condition $\Xi_N(\psi)$.** For all $1 \leq k < N$ and all $n \in \mathcal{P}_k$,

$$\psi_n \simeq I_{\{g(z_n) < \overline{\tau}_{k+1} + V^N_{k+1}(z_n)\}} P_0\text{-a.s. on } T^\psi_n.$$  

(4.7)

Conversely, if there is an equality in (4.6) for some $\psi \in F^N$, then $\psi$ satisfies Condition $\Xi_N(\psi)$.

From Lemma 5.1, it will be seen that, typically, the continuation region of the optimal stopping rule in (4.7) will have a form of interval, not necessarily the same for each $k = 1, \ldots, N - 1$.

The proof of Theorem 4.1 is provided in the appendix.

**Corollary 4.1.** Let a truncated sequential test $(\psi, \phi) \in S^N$ be such that Condition $\Xi_N(\psi)$ of Theorem 4.1 and Condition $\Xi(\psi, \phi)$ of Theorem 3.1 are satisfied.

Then for all truncated tests $(\psi', \phi') \in S^N$ such that

$$x(\psi', \phi') \leq x(\psi, \phi) \quad \text{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi),$$

(4.8)
it holds that

\[ K_0(\psi') \geq K_0(\psi). \]  

(4.9)

The inequality in (4.9) is strict if at least one of the inequalities in (4.8) is strict.

If there are equalities in all of the inequalities in (4.8) and (4.9), then Condition \( \Xi_N(\psi') \) and Condition \( \Xi(\psi', \phi') \) are satisfied for \( (\psi', \phi') \).

**Corollary 4.1** is a direct consequence of Theorem 3.1 and Theorem 4.1 in combination with Lemma 3.1.

There exists a very broad field for quite practical applications of Corollary 4.1. It is very important to highlight that our theoretical framework includes, in particular, group sequential tests with fixed group sizes. Such tests (with \( \nu_i = n_i \), where \( n_i \) are some fixed numbers, \( i = 1, 2, \ldots, K \), with a fixed number \( K \) of groups) are widely used in the practice of clinical trials (see, for example, Jennison and Turnbull 2000). Corollary 4.1, in particular, provides all of the optimal designs for group sequential tests in such a case. If the cost function is defined as \( n_1 + \ldots + n_k \) (the number of observations in the \( k \) observed groups), then the average cost \( K_0(\psi) \) minimized in (4.9) is an average sample number, under \( H_0 \), taken by the whole test, which is always important for applications (see, e.g., Jennison and Turnbull 2000). The important (for ethical reasons) case of minimization of the average sample number under \( H_1 \) is covered by Corollary 4.1 as well, because of the interchangeability of \( \theta_0 \) and \( \theta_1 \). A linear combination of the two average sample numbers may also be used as a criterion of minimization, slightly changing (3.1) by adding a term accounting for the average cost under \( H_1 \), and introducing the respective changes in (4.4) (cf. Novikov [2009] for a more general treatment).

From this point of view, the random size of groups may be useful for modeling the case of missing at random observations practitioners have always been concerned about.

We are currently working on applications of the results of this article to the optimal planning of group sequential tests for binary data (Fleming 1982), based on the algorithms developed for “continuous” sampling from a Bernoulli population (Novikov, Novikov, and Farkhshatov 2022).

### 4.2. Optimal stopping on an infinite horizon

Similar to Novikov (2009), the idea of this part is to pass to the limit, as \( N \to \infty \), on both sides of (4.6) in order to obtain a lower bound for the Lagrangian function and conditions to attain it.

Let us first analyze the right-hand side of (4.6) supposing that \( N \to \infty \).

We have

\[ V_N^N(z) \geq V_{N+1}^N(z) \]  

(4.10)

for all \( z \geq 0 \), because, by (4.4),

\[ V_N^N(z) = g(z) \geq V_{N+1}^N(z) = \min \left\{ g(z), \bar{c}_{N+1} + \sum_{n \in G} p_{N+1}(n) E_0 V_{N+1}^N(zn) \right\}. \]

Applying (4.4) to (4.10) again, we obtain \( V_{N-1}^N(z) \geq V_{N-1}^{N+1}(z) \), and so on, getting finally to
\[ V_k^N(z) \geq V_k^{N+1}(z), \quad z \geq 0, \quad (4.11) \]

for any \( k \) fixed. It follows from (4.11) that there exists \( \lim_{N \to \infty} V_k^N(z) = V_k(z), \quad z \geq 0, \) and, by the Lebesgue theorem of dominated convergence, \( \lim_{N \to \infty} \nabla_k^N(z) = \nabla_k(z), \quad z \geq 0. \)

To pass to the limit on the left-hand side of (4.6), let us define truncation of any \( \psi \in F_0 \) at any level \( N \) as \( \psi^N = (\psi_1, \psi_2, \ldots, \psi_{n_N}, 1, \ldots) \) for all \( n = (n_1, n_2, \ldots, n_{N-1}) \in \mathbb{N}^{N-1}. \)

Because \( \psi^N \in F_N \), we can apply (4.6) with \( L_N(\psi) = L(\psi^N). \) It is easy to see that \( L_N(\psi) \) can be calculated directly over (4.2), whatever the stopping rule \( \psi \in F_0. \)

**Lemma 4.1.** For any \( \psi \in F_0, \)
\[
\lim_{N \to \infty} L_N(\psi) = L(\psi). \quad (4.12)
\]

**Lemma 4.2.**
\[
\inf_{\psi \in F_0} L(\psi) = c_1 + \nabla_1(1).
\]

The proofs of Lemma 4.1 and Lemma 4.2 are provided in the appendix.

Now we are able to characterize optimal stopping rules on infinite horizon.

**Theorem 4.2.** For every \( \psi \in F_0, \)
\[
L(\psi) \geq c_1 + \nabla_1(1). \quad (4.13)
\]

There is an equality in (4.13) if \( \psi \in F_0 \) satisfies

**Condition \( \Xi_{\infty}(\psi).** For all \( 1 \leq k < \infty \) and all \( n \in \mathcal{P}_k, \)
\[
\psi_n \cong I_{\{g(z_n) \leq c_{k+1} + \nabla_{k+1}(z_n)\}} P_0 \text{ a.s. on } T_n^\psi. \quad (4.14)
\]

Conversely, if there is an equality in (4.13) for some \( \psi \in F_0, \) then \( \psi \) satisfies Condition \( \Xi_{\infty}(\psi). \)

The expression (4.14) characterizes all of the optimal stopping rules, because Condition \( \Xi_{\infty}(\psi) \) is necessary and sufficient for the optimality. Below, we give some properties of the set of \( z_n \) for which the inequality in (4.14) is satisfied. In the case of i.i.d. observations and group sizes, it will be equivalent to \( z_n \not\in (A, B) \) (with possible randomization when \( z_n = A \) or \( z_n = B \)).

The proof of Theorem 4.2 is laid out in the appendix.

**Corollary 4.2.** Let a sequential test \( (\psi, \phi) \in S_0 \) be such that Condition \( \Xi_{\infty}(\psi) \) of Theorem 4.2 and Condition \( \Xi(\psi, \phi) \) of Theorem 3.1 are satisfied.

Then for all sequential tests \( (\psi', \phi') \in S_0 \) such that
\[
\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi), \quad (4.15)
\]

it holds that
\[
K_0(\psi') \geq K_0(\psi). \quad (4.16)
\]

The inequality in (4.16) is strict if at least one of the inequalities in (4.15) is strict.
If there are equalities in all of the inequalities in (4.15) and (4.16), then Condition $S_1(w_0)$ and Condition $D(w_0, \lambda_0)$ are satisfied for $(\psi', \phi')$.

5. OPTIMALITY OF THE RANDOM SEQUENTIAL PROBABILITY RATIO TEST

In this section we apply the general results of preceding sections to a particularly important model, assuming that the groups for the group sequential test are formed in a stationary way. More precisely, we assume in this section that the group sizes $\nu_1, \nu_2, \ldots$, are identically distributed (and their common distribution is given by $p(n) = P(\nu_k = n)$, for $n \in G$ and $k = 1, 2, \ldots$). Respectively, the average group costs $\tau_i = \tau$ keep the same value over the experiment time. We will characterize the structure of optimal group sequential tests in the case of infinite horizon. In particular, we will prove the optimal property of the random sequential probability ratio test (RSPRT) proposed by Mukhopadhyay and de Silva (2008) for the random group sequential model.

There are three constants involved in the construction of optimal tests in this case: $\tau, \lambda_0$, and $\lambda_1$. It is easy to see that only two of them suffice to obtain all of the optimal sequential tests of Corollary 4.2. Let these be $c = \tau > 0$ and $\lambda = \lambda_0 > 0$ (and just assume that $\lambda_1 = 1$). The structure of optimal tests of Theorem 4.2 now acquires a simpler form.

Let

$$\rho_0(z) = \rho_0(z; c, \lambda) = g(z; \lambda) = \min\{\lambda, z\}, \quad z \geq 0,$$

and, recursively over $k = 1, 2, \ldots$,

$$\rho_k(z) = \rho_k(z; c, \lambda) = \min\{g(z; \lambda), c + \sum_{n \in G} p(n) E_0 \rho_{k-1}(zn; c, \lambda)\}, \quad z \geq 0. \quad (5.2)$$

Also let

$$\bar{\rho}_k(z) = \bar{\rho}_k(z; c, \lambda) = \sum_{n \in G} p(n) E_0 \rho_k(zn; c, \lambda). \quad (5.3)$$

Functions $\rho_k(z)$, in the case of i.i.d. observations and group sizes, give a simpler form to the functions $V_k^N$ in (4.3)–(4.4), which determine the structure of optimal stopping.

Indeed, it follows from (4.3)–(4.4) that

$$V_k^N(z) = \rho_{N-k}(z; c, \lambda), \quad z \geq 0, \quad (5.4)$$

and from (4.5) that

$$\nabla_k^N(z) = \bar{\rho}_{N-k}(z; c, \lambda). \quad (5.5)$$

Let us define

$$\rho(z) = \rho(z; c, \lambda) = \lim_{k \to \infty} \rho_k(z; c, \lambda), \quad z \geq 0. \quad (5.6)$$

If we take the limit, as $N \to \infty$, in (5.4), then

$$V_k(z) = \rho(z; c, \lambda), \quad z \geq 0, \quad (5.7)$$

for all $k = 1, 2, \ldots$, and by (5.5)
\[ \nabla_k(z) = \bar{p}(z; c, \lambda), \quad z \geq 0. \quad (5.8) \]

Stopping rule (4.14) in Condition \( \mathfrak{S}_\infty(\psi) \) now transforms to
\[ \psi_n \simeq I_{\{g(z_n; \lambda) \leq c + \bar{p}(z_n; c, \lambda)\}}, \quad (5.9) \]
so the form of optimal stopping rules entirely depends on whether the inequality
\[ g(z; \lambda) \leq c + \bar{p}(z; c, \lambda) \quad (5.10) \]
(and/or its strict variant) is fulfilled or not at \( z = z_n \).

First of all, it is easy to see that if
\[ \lambda < c + \bar{p}(\lambda; c, \lambda), \quad (5.11) \]
then (5.9) implies that \( \psi_n \equiv 1 \) for all \( n \in G \) (the optimal test stops after the first group is taken). Therefore, nontrivial optimal sequential tests are only obtained if
\[ \lambda > c + \bar{p}(\lambda; c, \lambda), \quad (5.12) \]
which will be assumed in what follows.

**Lemma 5.1.** If
\[ P_0(f_1(X) > 0) = 1, \quad (5.13) \]
then for any positive \( c \) and \( \lambda \) satisfying (5.12) there exist \( 0 < A < \lambda \) and \( B > \lambda \) such that
\[ g(A; \lambda) = c + \bar{p}(A; c, \lambda), \quad g(B; \lambda) = c + \bar{p}(B; c, \lambda), \quad (5.14) \]
and
\[ g(z; \lambda) < c + \bar{p}(z; c, \lambda) \quad \text{for all} \quad 0 \leq z < A \quad \text{and all} \quad z > B, \quad (5.15) \]
and
\[ g(z; \lambda) > c + \bar{p}(z; c, \lambda) \quad \text{for all} \quad A < z < B. \quad (5.16) \]

It follows from Lemma 5.1 that if (5.13) holds, then (5.9) is equivalent to
\[ I_{\{z_n \in (A, B)\}} \leq 1 - \psi_n \leq I_{\{z_n \in [A, B]\}}, \quad (5.17) \]
that is, any optimal test is a randomized version of the RSPRT by Mukhopadhyay and de Silva (2008), which, in our terms, can be described as \((\psi, \phi)\) with
\[ \psi_n = I_{\{z_n \in (A, B)\}} \quad \text{and} \quad \phi_n = I_{\{z_n \geq B\}}, \quad (5.18) \]
for all \( n \in G^k \) and \( k \geq 1 \).

Obviously, \( \psi_n \) in (5.18) is a particular case of (5.17), and \( \phi_n \) satisfies Condition \( \mathfrak{D}((\psi, \phi)) \). In addition, by virtue of theorem 3.1 in Mukhopadhyay and de Silva (2008), \((\psi, \phi) \in S_0 \) (the details can be found in the proof of Theorem 5.1). Consequently, it follows that this RSPRT is optimal in the sense of Corollary 4.2.

In the same way, all of the sequential tests \((\psi, \phi)\) with \( \psi \) satisfying (5.17) and \( \phi \) satisfying (5.18) (for all \( n \in G^k \) and \( k \geq 1 \)) share the optimum property with the RSPRT when (5.13) is satisfied. In particular, obviously, this is the case when the hypothesized distributions belong to a Koopman-Darmois family.

If (5.13) is not satisfied, optimal tests with stopping rules (5.9) are not necessarily of the RSPRT type. This can be seen from the following simple example.
Let \( H_0 \) state that the (one per group) observations follow a uniform distribution on [0,1], whereas under \( H_1 \) they are assumed to be uniform on [0,0.5]. Using the definition of \( \rho \) in (5.6), on the basis of \( \rho_k \) defined in (5.1) and (5.2), with \( \lambda = 2 \) and \( c = 1 \), one easily sees that \( \rho(z; c, \lambda) = \rho(z; 1, 2) = g(z; 2) = \min\{z, 2\} \) and \( c + \rho(z; c, \lambda) = 1 + \rho(z; 1, 2) = 1 + \min\{z, 1\} \). Let us consider an (optimal) test corresponding to (5.10), with a strict inequality. It is immediate that, with the above definitions of \( g \) and \( \rho \), (5.10) holds if and only if \( z < 2 \). But the consecutive values of the probability ratios \( z_1, z_2, \ldots, z_k \) are, respectively, 2, 4, 8, \ldots, \( 2^k \), whenever \( X_1, X_2, \ldots, X_k \leq 0.5 \), so the test (minimizing \( K_0 \)) only stops when \( X_i > 0.5 \), for the first time, in which case \( z_i = 0 \).

It is seen from this example, first, that the test minimizing \( K_0 \) is not an RSPRT (because an RSPRT should also stop when \( z_i \geq B > 0 \), which happens under \( H_0 \) with a positive probability; thus, there would be a positive \( \alpha \)-error) and, second, that, in no way does it minimize \( K_1 \), because under \( H_1 \) it never stops.

Nevertheless, the following theorem shows that, even if (5.13) is not satisfied, not only the RSPRT tests with \( \psi \) satisfying (5.18) but also their “randomized” versions with \( \psi \) satisfying (5.17) are optimal in the sense of Wald and Wolfowitz (1948); that is, they minimize the average cost under both \( H_0 \) and \( H_1 \), given restrictions on the error probabilities.

**Theorem 5.1.** Let \( A < B \) be two positive constants. Let \( \psi \) be any stopping rule satisfying

\[
I_{\{z_n \in (A, B)\}} \leq 1 - \psi_n \leq I_{\{z_n \in [A, B]\}},
\]

for all \( n \in G^k \) and \( k = 1, 2, \ldots \), and let \( \phi \) be a decision rule defined as

\[
\phi_n = I_{\{z_n \geq B\}}.
\]

Then \( (\psi, \phi) \in S_0 \cap S_1 \), and it is optimal in the following sense: for any sequential test \( (\psi', \phi') \in S_0 \cap S_1 \) such that

\[
\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi),
\]

it holds that

\[
K_0(\psi) \leq K_0(\psi') \quad \text{and} \quad K_1(\psi) \leq K_1(\psi').
\]

Both inequalities in (5.22) are strict if at least one of the inequalities in (5.21) is strict.

**Remark 5.1.** The optimum property stated in Theorem 5.1 in the case of one-per-group observations is known as the Wald-Wolfowitz optimality (see Wald and Wolfowitz 1948). Burkholder and Wijsman (1963) proved that all “extended” SPRTs—that is, those admitting a randomized decision between stopping and continuing in case \( z_n = A \) or \( z_n = B \)—share the same optimum property with the SPRT. Our Theorem 5.1 states the same in the case of random group sequential tests: the “extended” group sequential tests—that is, those with stopping rules satisfying (5.19)—minimize the average cost under both \( H_0 \) and \( H_1 \).

**Remark 5.2.** Very much like in the classical one-per-group case, when taking no observations is permitted, only the case \( 1 \in [A, B] \) is meaningful for the RSPRT, because otherwise a trivial test (namely, the one that, without taking any observations, accepts
or rejects $H_0$ depending on whether $1 < A$ or $1 > B$) performs better than the optimal tests of Theorem 5.1 ($\min\{1, \lambda\} < c + \bar{p}(1; c, \lambda)$ in terms of optimal stopping of Theorem 4.2).

**APPENDIX**

Lengthy and very technical proofs are gathered in this appendix. For simplicity, it is assumed throughout this section that all group sizes are identically distributed, even when this is not explicitly required in the respective statement.

**Lemma A.1.** If a stopping rule $\psi$ is such that

$$\sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} \to 0 \text{ as } r \to \infty, \quad (A.1)$$

for some $n \in G^k, k \geq 1$, then

$$E_0 \delta_n^{\psi} + \sum_{r=1}^{\infty} \sum_{m \in G} p(m)E_0 s_{\psi, n, m}^{\psi} = E_0 t_n^{\psi}. \quad (A.2)$$

**Proof of Lemma A.1.** Let $r$ be any natural number. Then

$$\sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} - \sum_{m \in G^{r+1}} p(m)E_0 t_{\psi, n, m}^{\psi}$$

$$= \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} - \sum_{m \in G^{r+1}} \sum_{i \in G} p(m)p(i)E_0 t_{\psi, n, m, i}^{\psi}$$

$$= \sum_{m \in G} p(m)(E_0 t_{\psi, n, m}^{\psi} - \sum_{i \in G} p(i)E_0 t_{\psi, n, m, i}^{\psi})$$

$$= \sum_{m \in G} p(m)(E_0 t_{\psi, n, m}^{\psi} - E_0(1 - \psi_{n_1})... (1 - \psi_{n_r})(1 - \psi_{n_{r+1}})... (1 - \psi_{n_m}))$$

$$= \sum_{m \in G} p(m)E_0 s_{\psi, n, m}^{\psi},$$

so

$$\sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} = \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} - \sum_{m \in G^{r+1}} p(m)E_0 t_{\psi, n, m}^{\psi}. \quad (A.3)$$

Applying the sum over $r$ from $r = 1$ to $r = k$ on both sides of (A.3), we obtain

$$\sum_{r=1}^{k} \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} = \sum_{r=1}^{k} \left( \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} - \sum_{m \in G^{r+1}} p(m)E_0 t_{\psi, n, m}^{\psi} \right), \quad (A.4)$$

Passing to the limit, as $k \to \infty$, in (A.4) and making use of (A.1) on the right-hand side, we get

$$\sum_{r=1}^{\infty} \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi} = \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi}. \quad (A.5)$$

On the other hand, it is easy to see (by definitions 2.3–2.4) that

$$E_0 \delta_n^{\psi} = E_0 t_n^{\psi} - \sum_{m \in G} p(m)E_0 t_{\psi, n, m}^{\psi}. \quad (A.6)$$

Adding the expressions on both sides of (A.6) to the respective sides of (A.5), we obtain (A.2). \qed
**Lemma A.2.** For a stopping rule \( \psi \), it holds that
\[
\sum_{m \in G} p(m)E_{\psi} t_{m}^{\psi} \to 0, \quad \text{as } r \to \infty,
\]
if and only if
\[
\sum_{r=1}^{\infty} \sum_{m \in G} p(m)E_{\psi} s_{m}^{\psi} = 1.
\]

**Proof of Lemma A.2.** Let us suppose that (A.7) is satisfied. Repeating the steps of the proof of Lemma A.1, but with \( t_{n, m}^{\psi} \) instead of \( t_{m}^{\psi} \), we obtain, in place of (A.5),
\[
\sum_{r=1}^{\infty} \sum_{m \in G} p(m)E_{\psi} s_{m}^{\psi} = \sum_{m \in G} p(m)E_{\psi} t_{m}^{\psi} = 1,
\]
the latter because \( t_{m}^{\psi} = 1 \) for all \( m \in G \), by definition. (A.8) is proved.

Let now (A.8) be fulfilled. Acting as in the proof of Lemma A.1 again, we obtain
\[
\sum_{r=1}^{k} \sum_{m \in G} p(m)E_{\psi} s_{m}^{\psi} \leq \sum_{m \in G} p(m)E_{\psi} t_{m}^{\psi} - \sum_{r=1}^{k} \sum_{m \in G} p(m)E_{\psi} t_{m}^{\psi},
\]
the left-hand side of (A.10) tends to 1, as \( k \to \infty \), by virtue of (A.8). The first term on the right-hand size is equal to 1, for the same reason as in (A.9). Therefore, the second term on the right-hand side of (A.10) tends to 0, so (A.7) follows.

**Lemma A.3.** Let \( \psi \in F_{0} \) be such that \( K_{0}(\psi) < \infty \). Then
\[
\sum_{n \in \mathbb{G}^{k}} p(n)c(n)E_{\psi} t_{n}^{\psi} \to 0, \quad \text{as } N \to \infty.
\]

**Proof of Lemma A.3.** Let \( \psi \in F_{0} \) be such that \( K_{0}(\psi) < \infty \). By definition,
\[
K_{0}(\psi) = \sum_{k=1}^{\infty} \sum_{n \in \mathbb{G}^{k}} p(n)c(n)E_{\psi} s_{n}^{\psi},
\]
thus,
\[
\sum_{k=N}^{\infty} \sum_{n \in \mathbb{G}^{k}} p(n)c(n)E_{\psi} s_{n}^{\psi} \to 0,
\]
as \( N \to \infty \). Let \( c(n|N) = c(n_{1}) + \ldots + c(n_{N}) \) for any \( n \in \mathbb{G}^{k} \) and \( k \geq N \).

It follows from Lemma A.1 that
\[
\sum_{n \in \mathbb{G}^{k}} c(n|N)p(n)E_{\psi} t_{n}^{\psi} = \sum_{n \in \mathbb{G}^{k}} c(n)\sum_{m \in G} p(m)(E_{\psi} s_{m}^{\psi} + \sum_{r=1}^{\infty} \sum_{m \in G} p(m)E_{\psi} t_{m}^{\psi},
\]
\[
= \sum_{k=N}^{\infty} \sum_{n \in \mathbb{G}^{k}} c(n|N)p(n)E_{\psi} s_{n}^{\psi} \leq \sum_{k=N}^{\infty} \sum_{n \in \mathbb{G}^{k}} p(n)c(n)E_{\psi} s_{n}^{\psi} \to 0,
\]
as \( N \to \infty \), by virtue of (A.12).

**Proof of Theorem 4.1.** Let \( \psi \in F_{N}^{k} \) be any truncated stopping rule. For any \( k = 1, 2, \ldots, N \), let us define
\[
Q_{k}^{N}(z) = \sum_{i=1}^{k-1} \sum_{n \in \mathbb{G}^{k}} p(n)E_{\psi} s_{n}^{\psi}(c(n) + g(z_{n})) + \sum_{n \in \mathbb{G}^{k}} p(n)E_{\psi} t_{n}^{\psi}(c(n) + V_{k}^{N}(z_{n})),
\]
for all \( z \geq 0 \). In particular, we have \( Q_{N}^{N}(1) = L_{N}(\psi) \) and \( Q_{1}^{N}(1) = \tau_{1} + \nabla_{1}^{N}(1) \).

We want to prove
\[
L_{N}(\psi) = Q_{N}^{N}(1) \geq Q_{N-1}^{N}(1) \geq \ldots \geq Q_{1}^{N}(1) = \tau_{1} + \nabla_{1}^{N}(1)
\]
and determine the form of \( \psi \in F_{N}^{k} \) turning all of the inequalities in (A.16) into equalities.
Let us first prove that for 1 \( \leq k \leq N - 1 \) it holds that
\[
Q_{k+1}^N(1) \geq Q_k^N(1).
\] (A.17)

By definition (A.15), inequality (A.17) is equivalent to
\[
\sum_{n \in G^k} p(n)E_0 t_n^\psi(c(n) + g(z_n)) + \sum_{n \in G^{k+1}} p(n)E_0 t_n^\psi(c(n) + V_{k+1}^N(z_n)) \\
\geq \sum_{n \in G^k} p(n)E_0 t_n^\psi(c(n) + V_k^N(z_n)).
\] (A.18)

By the Fubini theorem, the left-hand side of (A.18) is equal to
\[
\sum_{n \in G^k} p(n) \int t_n^\psi(c(n) + \psi_n g(z_n) + (1 - \psi_n) \left( \sum_{m \in G} p_{k+1}(m)c(m) \\
+ \sum_{m \in G} p_{k+1}(m) \left[ V_{k+1}^N(z_n z_m) f_0^{(m)} d\mu^m \right] f_0^{(n)} d\mu^n \right).
\] (A.19)

By virtue of lemma 5.1 in Novikov (2009), the minimum value of (A.19), over all \( \psi_n, n \in G^k \), is equal to
\[
\sum_{n \in G^k} p(n) \int t_n^\psi(c(n) + \min\{g(z_n), \sum_{m \in G} p_{k+1}(m)c(m) \\
+ \sum_{m \in G} p_{k+1}(m) \left[ V_{k+1}^N(z_n z_m) f_0^{(m)} d\mu^m \right] f_0^{(n)} d\mu^n \right).
\] (A.20)

and is attained if and only if for all \( n \in \mathcal{P}_k \),
\[
\psi_n \simeq I_{\{g(z_n) < z_{k+1} + V_{k+1}^N(z_n)\}}
\] (A.21)

\( \mu^\psi \)-a.e. on \( T_n^\psi \cap \{c^{(n)} > 0\} \).

This completes the proof of (A.17).

Applying (A.17) consecutively for \( k = 1, 2, \ldots, N - 1 \), we also obtain the proof for (A.16). In addition, there are equalities in all of the inequalities in (A.16) if and only if (A.22) is satisfied for all \( n \in \mathcal{P}_k \), \( \mu^\psi \)-a.e. on \( T_n^\psi \cap \{c^{(n)} > 0\} \), for all \( k = 1, 2, \ldots, N - 1 \), which coincides with Condition \( \mathcal{E}_N(\psi) \) of Theorem 4.1.

**Proof of Lemma 4.1.** Let us suppose first that \( L(\psi) < \infty \). By (3.7) and (4.2) we have
\[
L(\psi) - L_N(\psi) = \sum_{k=N}^{\infty} \sum_{n \in G^k} p(n)E_0 t_n^\psi(c(n) + g(z_n)) \\
- \sum_{n \in G^k} p(n)c(n)E_0 t_n^\psi - \sum_{n \in G^k} p(n)E_0 t_n^\psi g(z_n).
\] (A.22)

Because the series \( L(\psi) \) is converging, by supposition, the first term on the right-hand side of (A.22) tends to zero, as \( N \to \infty \).

In addition, this implies that \( K_0(\psi) < \infty \), so by Lemma A.3 (see A.11), the second term on the right-hand side of (A.22) also tends to zero as \( N \to \infty \).

Finally, \( g(z) \leq \lambda_0 \) for all nonnegative \( z \). Thus,
\[
\sum_{n \in G^k} p(n)E_0 t_n^\psi g(z_n) \leq \lambda_0 \sum_{n \in G^k} p(n)E_0 t_n^\psi \to 0, \quad \text{as} \quad N \to \infty.
\]

This latter holds by Lemma A.2, because the supposition \( \psi \in F \) implies (A.8). Thus, we proved (4.12) in case \( L(\psi) < \infty \).
Now let $L(\psi) = \infty$; then

$$L_N(\psi) \geq \sum_{k=1}^{N-1} \sum_{n \in G} p(n) E_0 s_n^\psi (c(n) + g(z_n)) \to L(\psi) = \infty,$$

as $N \to \infty$, so that (4.12) holds in this case as well. \hfill \Box

Proof of Lemma 4.2. Denote $U = \inf_{\psi \in F} L(\psi)$, $U_N = \inf_{\psi \in F^N} L(\psi)$.

By Theorem 4.1, $U_N = \overline{\tau_1} + \nabla_1^N (1)$ for any $N = 1, 2, \ldots$. It is obvious that $U_N \geq U$ for any $N = 1, 2, \ldots$, so, $\lim_{N \to \infty} U_N \geq U$. Let us show that in fact $\lim_{N \to \infty} U_N = U$.

Suppose the contrary: $\lim_{N \to \infty} U_N = U + 4\epsilon$ with some $\epsilon > 0$; this would imply

$$U_N \geq U + 3\epsilon.$$ (A.23)

for all large enough $N$. By definition of $U$, there exists a stopping rule $\psi \in F_0$, such that $U \leq L(\psi) \leq U + \epsilon$. Because, by Lemma 4.1, $L_N(\psi) \to L(\psi)$, as $N \to \infty$, we have $L_N(\psi) \leq U + 2\epsilon$ for all large enough $N$. Because, by definition, $U_N \leq L_N(\psi)$, we have that $U_N \leq U + 2\epsilon$ for all large enough $N$, which contradicts (A.23).

Hence, $U = \lim_{N \to \infty} U_N = \overline{\tau_1} + \lim_{N \to \infty} \nabla_1^N (1) = \overline{\tau_1} + \nabla_1 (1)$. \hfill \Box

A.1. Proof of Theorem 4.2

Let $\psi \in F_0$ be any stopping rule. For $k = 1, 2, \ldots$, let us define for all $z \geq 0$,

$$Q_k(z) = \sum_{i=1}^{k-1} \sum_{n \in G} p(n) E_0 s_n^\psi (c(n) + g(z_n))$$

(A.24)

$$+ \sum_{n \in G} p(n) E_0 t_n^\psi (c(n) + V_k(z_n))$$

(cf. A.15). By the Lebesgue theorem of dominated convergence, it follows that

$$\lim_{N \to \infty} Q_N^\psi (z) = Q_k(z).$$ (A.25)

For the same reason, we have from (A.17)

$$Q_{k+1}(1) \geq Q_k(1)$$ (A.26)

for any $k = 1, 2, \ldots$.

Completely analogous to obtaining (A.21) from (A.17), through the steps of (A.18), (A.19), and (A.20), we can obtain that there is an equality in (A.26) if for all $n \in P_k$

$$\psi_n \simeq I_{\{g(z_n) \leq \overline{\tau_1} + \nabla_1 (z_n)\}}$$ (A.27)

$\mu^\ast$-a.e. on $T^\psi \cap \{f_0^{(n)} > 0\}$. Applying (A.26) consecutively for $k = 1, 2, \ldots$, we obtain

$$Q_k(1) \geq Q_{k-1}(1) \geq \ldots \geq Q_1(1) = \overline{\tau_1} + \nabla_1 (1),$$ (A.28)

and there are all equalities in (A.28), for any natural $k$, in case Condition $\mathcal{E}_\infty (\psi)$ is satisfied.

In particular, we have that for all natural $k$,

$$Q_k(1) = \sum_{i=1}^{k-1} \sum_{n \in G} p(n) E_0 s_n^\psi (c(n) + g(z_n))$$

(A.29)

$$+ \sum_{n \in G} p(n) E_0 t_n^\psi (c(n) + V_k(z_n)) = \overline{\tau_1} + \nabla_1 (1).$$

It follows from this, first, that $K_0(\psi) < \infty$, because otherwise

$$\lim_{k \to \infty} \sum_{i=1}^{k-1} \sum_{n \in G} p(n) E_0 s_n^\psi (c(n) + g(z_n)) \geq K_0(\psi)$$
would be infinite and, hence, would contradict \((A.29)\). Thus, \(K_0(\psi) < \infty\), so, by virtue of Lemma A.3,
\[
\sum_{n \in G^k} p(n)c(n)E_0 t_n^\psi \to 0, \text{ as } k \to \infty.
\]
Furthermore,
\[
\sum_{n \in G^k} p(n)E_0 t_n^\psi V_k(z_n) \leq \lambda_0 \sum_{n \in G^k} p(n)E_0 t_n^\psi \to 0, \text{ as } k \to \infty,
\]
because \(\psi \in F_0\) by supposition.

It follows from \((A.29), (A.30), and (A.31)\) that
\[
\lim_{k \to \infty} \frac{X_n}{C_0} = 1
\]
thus, the sufficiency part of Theorem 4.2 is proved.

Let us prove the necessity part.

Passing to the limit in \((A.16)\) as \(N \to \infty\), we have for any \(\psi \in F_0\)
\[
L(\psi) \geq Q_k(1) \geq Q_{k-1}(1) \geq ... \geq Q_1(1) = \bar{z}_1 + \nabla_1(1),
\]
for all natural \(k\). Let us suppose now that there is an equality in \((4.13)\) for a stopping rule \(\psi \in F_0\).

It follows from \((A.32)\) now that all of the inequalities in \((A.32)\) are in fact equalities, so
\[
Q_{k+1}(1) = Q_k(1) \text{ for all natural } k.
\]
This implies, once again, that \(\Phi_\infty(\psi)\) is satisfied.

### A.2. Proof of Theorem 5.1

Before starting with the proof of the theorem, we need some previous work to be done.

Let us denote
\[
R(z; c, \lambda) = \inf_{(\psi, \phi) \in S_k} (cE_0 T_\psi + \lambda x(\psi, \phi) + z\beta(\psi, \phi))
\]
for any \(z \geq 0, \ c \geq 0, \text{ and } \lambda \geq 0\). It follows from Theorem 4.2 and Theorem 3.1 that
\[
R(1; c, \lambda) = c + \bar{p}(1; c, \lambda)
\]
for all \(c > 0\) and \(\lambda > 0\).

The following lemma shows that \((A.34)\) is in fact a particular case of a much more general relationship.

**Lemma A.4.** For all \(z > 0, \ c > 0, \text{ and } \lambda > 0\),
\[
R(z; c, \lambda) = c + \bar{p}(z; c, \lambda).
\]

**Proof of Lemma A.4.** It follows from \((A.34)\) that
\[
R(z; c, \lambda) = z \inf_{(\psi, \phi) \in S_k} \left( \frac{c}{z} E_0 T_\psi + \frac{\lambda}{z} x(\psi, \phi) + \beta(\psi, \phi) \right)
\]
\[
= z(c/z + \bar{p}(1; c/z, \lambda/z)) = c + z\bar{p}(1; c/z, \lambda/z).
\]
Following the definitions of \((5.1)–(5.3)\) and \((5.8)\), it is not difficult to see that
\[
z\bar{p}(1; c/z, \lambda/z) = \bar{p}(z; c, \lambda) \text{ for all } z, c, \lambda > 0.
\]
Applying \((A.37)\) on the right-hand side of \((A.36)\), we have \((A.35)\).

**Lemma A.5.** Function \(R(z; c, \lambda)\) defined by \((A.33)\) is concave and jointly continuous on \(\{z \geq 0, c \geq 0, \lambda \geq 0\}\).
Proof of Lemma A.5. Because \( R \) is defined as an infimum of a family of linear functions (see A.33), it is concave on \( \{ z \geq 0, c \geq 0, \lambda \geq 0 \} \), and it follows from theorem 10.1 in Rockafellar (1970) that it is (jointly) continuous on \( \{ z > 0, c > 0, \lambda > 0 \} \).

It remains to show that \( R(z; c, \lambda) \) is continuous at any point with \( c = 0, \lambda = 0 \), or \( z = 0 \).

The most nontrivial part is when \( c = 0 \) with \( \lambda > 0 \) and \( z > 0 \).

Let \( k \) be any natural number. Let us define the stopping rule \( \psi \) in such a way that \( s^k_n = 1 \) for all \( n \in G^k \), and let \( \phi \) be a decision rule with \( \phi_n = I_{\{z_n \geq \lambda / z\}} \), \( n \in G^k \). Then, by the Markov inequality,

\[
\lambda x(\psi, \phi) = \lambda \sum_{n \in G} p(n) P_0(Z_n \geq \lambda / z) \leq \sqrt{\lambda z} \sum_{n \in G} p(n) E \sqrt{Z_n^{1/2}} = \sqrt{\lambda z^k},
\]

(A.38)

where

\[
r = \sum_{n \in G} p(n) E_Z^{1/2} < 1
\]

(this latter inequality holds by virtue of the well-known fact that the Hellinger divergence \( E_0 Z_n^{1/2} = \int f_0^{1/2} f_1^{1/2} d\mu < 1 \) whenever 2.2 is satisfied).

Analogously,

\[
z \beta(\psi, \phi) = z \sum_{n \in G} p(n) P_1(Z_n^{-1} \geq (\lambda / z)^{-1}) \leq \sqrt{\lambda z} \sum_{n \in G} p(n) E_1 Z_n^{-1/2} = \sqrt{\lambda z^k}.
\]

(A.39)

Now let \( \varepsilon \) be any positive number. Taking \( c = 0 \) in (A.33) and

\[
k > \left( \ln \varepsilon - \ln \left( \sqrt{\lambda z} \right) \right) / \ln r,
\]

(A.40)

we see from (A.38) and (A.39) that

\[
R(z; 0, \lambda) \leq \lambda x(\psi, \phi) + z \beta(\psi, \phi) < 2 \varepsilon.
\]

Because \( \varepsilon \) is arbitrarily small, it follows now that \( R(z; 0, \lambda) = 0 \), whatever \( z > 0, \lambda > 0 \).

To prove the continuity at any point \((z, 0, \lambda)\), with \( \lambda, z > 0 \), let us start with some \( \lambda_n \to \lambda \), \( z_n \to z \), \( c_n \to 0 \), as \( n \to \infty \).

Let \( \varepsilon > 0 \) be an arbitrary number again. If \( k \) is large enough to satisfy (A.40), we have

\[
R(z_n; c_n, \lambda_n) \leq c_n k + \lambda_n x(\psi, \phi) + z_n \beta(\psi, \phi) < 5 \varepsilon
\]

for all \( n \) such that \( |\lambda_n - \lambda| < \varepsilon \), \( |z_n - z| < \varepsilon \), and \( c_n < \varepsilon / k \); that is,

\[
\lim_{n \to \infty} R(z_n; c_n, \lambda_n) = R(z; 0, \lambda) = 0.
\]

Now, if \( \lambda = 0 \) and \( z \geq 0 \), \( c \geq 0 \), with the same definition of \((\psi, \phi)\) as above, and \( k = 1 \), we have that \( \phi_n = I_{\{z_n \geq 0\}} = 1 \), so \( x(\psi, \phi) = 1 \) and \( \beta(\psi, \phi) = 0 \); thus, \( R(z; c, 0) = \inf_{(\psi, \phi) \in \mathcal{S}_c} (c E_0 T_\psi + z \beta(\psi, \phi)) \leq c \). On the other hand, \( E_0 T_\psi \geq 1 \), so \( R(z; c, 0) = c \). If now \((z_n, c_n, \lambda_n) \to (z, c, 0)\) as \( n \to \infty \), then \( R(z_n; c_n, \lambda_n) \leq c_n + \lambda_n \to c \) as \( n \to \infty \). On the other hand, \( R(z_n; c_n, \lambda_n) = \inf_{(\psi, \phi) \in \mathcal{S}_c} (c_n E_0 T_\psi + \lambda_n x(\phi, \psi) + z_n \beta(\psi, \phi)) \geq c_n \to c \), as \( n \to \infty \).

The case \( z = 0 \) can be treated analogously.

In view of (A.35), it is convenient to extend the definition of \( \overline{p}(z) = \overline{p}(z; c, \lambda) \) (defined initially for positive \( c, \lambda \)) in such a way that (A.35) holds for all \( z \geq 0, c \geq 0, \lambda \geq 0 \); that is, defining \( \overline{p}(z; c, \lambda) = 0 \) whenever any of its arguments is 0. Let us do so.

The following corollary is a direct consequence of Lemmas A.4 and A.5.

Corollary A.1. The function \( \overline{p}(z; c, \lambda) \) is concave and continuous on \( \{ z \geq 0, c \geq 0, \lambda \geq 0 \} \).

The following lemma is almost obvious but will be very useful in what follows.

Lemma A.6. Let \( F(z) \) be any convex nonnegative function on \( \{ z \geq 0 \} \) such that \( F(0) = 0 \). Then

\[
F(z_1) \leq F(z_2)
\]

(A.41)

for all \( 0 \leq z_1 < z_2 \), and the inequality in (A.42) is strict whenever \( F(z_1) > 0 \).
Proof of Lemma 5.1. For any \( c > 0 \) and \( \lambda > 0 \), let us define the functions \( D_1(z) = D_1(z; c, \lambda) = z - \bar{p}(z; c, \lambda) \) and \( D_2(z) = D_2(z; c, \lambda) = \lambda - \bar{p}(z; c, \lambda), \) \( z \geq 0 \). Obviously, \( D_1(\lambda) = D_2(\lambda) \). By virtue of the properties of \( \bar{p}(z; c, \lambda) \), we have that \( D_1(z) \) and \( D_2(z) \) are convex and continuous functions on \([0, \infty)[/itex].

By Jensen’s inequality,
\[
\bar{p}(z) = \sum_{n \in \mathbb{C}} p(n)E_0 \rho(zz_n) \leq \sum_{n \in \mathbb{C}} p(n) \rho(z) = \rho(z) \leq g(z), \tag{A.42}
\]
so \( g(z; \lambda) = \bar{p}(z; c, \lambda) = \min\{D_1(z), D_2(z)\} \geq 0 \).

By virtue of Lemma A.6, if a positive \( c \) is such that \( c < D_1(\lambda) \) (which is equivalent to \( \lambda > c + \bar{p}(\lambda; c, \lambda) \); i.e., 5.12), then there exists a unique \( A < \lambda \) such that \( D_1(A) = c \), and \( D_1(z) < c \) for \( z < A \) and \( D_1(z) > c \) for \( z > A \).

It is not difficult to see that (5.13) implies that \( \lim_{z \to \infty} D_2(z) = 0 \).

Quite analogous to the above property of \( D_1 \), if \( 0 < c < D_2(\lambda) = D_1(\lambda) \), then there exists a unique \( B > \lambda \) such that \( D_2(B) = c \), \( D_2(z) > c \) for \( z < B \), and \( D_2(z) < c \) for \( z > B \).

Thus, (5.14), (5.15), and (5.16) follow.

Lemma A.7. Let \( A \) and \( B \) be such that \( 0 < A < B < \infty \). Then there exist \( \lambda \) and \( c \) such that \( A \leq \lambda < B \) and \( c > 0 \) and (5.14) is satisfied.

Proof of Lemma A.7. Let \( A \) and \( B \) be such that \( 0 < A < B < \infty \). For any \( \lambda \in [A, B] \), let us define \( c = c(\lambda) \) as a solution to the equation
\[
c + \bar{p}(A; c, \lambda) = A. \tag{A.43}
\]

The existence of a unique solution \( c = c(\lambda) \) of (A.43) follows from the fact that the left-hand side of (A.43) is a continuous strictly increasing function of \( c \) taking values from 0 (at \( c = 0 \)) to \( \infty \). Furthermore, \( c(\lambda) \) is a continuous function of \( \lambda \), as an implicit function (A.43) defined by a function that is continuous in all of its variables (by Corollary A.1). In addition, \( c(\lambda) > 0 \) for all \( \lambda \in [A, B] \) because the contrary would imply, by virtue of (A.43), that \( \bar{p}(A; 0, \lambda) = A \); that is, \( A = 0 \), a contradiction.

Let us now define
\[
G(\lambda) = \lambda - \bar{p}(B; c(\lambda), \lambda) - c(\lambda),
\]
which is a continuous function of \( \lambda \) as a composition of two continuous functions.

Let us show that
\[
G(A) \leq 0, \quad \text{and} \quad G(B) > 0. \tag{A.44}
\]

Indeed,
\[
G(A) = A - \bar{p}(B; c(A), A) - c(A) \leq A - \bar{p}(A; c(A), A) - c(A) = 0 \tag{A.45}
\]
(by A.43).

Let us now show that
\[
G(B) = B - \bar{p}(B; c(B), B) - c(B) > 0. \tag{A.46}
\]
Taking into account that, by (A.43),
\[
c(B) + \bar{p}(A; c(B), B) = A,
\]
we see that (A.46) is equivalent to
\[
B - \bar{p}(B; c(B), B) > A - \bar{p}(A; c(B), B). \tag{A.47}
\]
Because \( F(z) = z - \bar{p}(z; c(B), B) \) satisfies the conditions of Lemma A.6, the converse to (A.47) would imply that \( F(A) = 0 \); that is, \( A - \bar{p}(A; c(B), B) = 0 \) or, in view of (A.43), \( c(B) = 0 \), a contradiction.
Thus, (A.44) is proved, so there exists $k \in [A, B]$ such that $G(k) = 0$; that is, we found $k \in [A, B]$ and $c = c(k) > 0$ such that

$$c + \Phi(A; c, k) = A \quad \text{and} \quad c + \Phi(B; c, k) = B,$$

which is equivalent to (5.14). $\square$

**Proof of Theorem 5.1.** It follows from Lemma A.7 by virtue of Lemma 5.1 that Condition $S_1(\psi)$ is satisfied. Condition $D(\psi, \phi)$ is also satisfied because $B \geq \lambda$. To apply Corollary 4.2, it remains to show that $(\psi, \phi) \in S_0$; that is, that

$$\sum_{n \in G} P(n) E_n^{\psi} = P_0(\tau_{\psi} \geq k) \to 0, \quad k \to \infty. \tag{A.48}$$

It follows from theorem 3.1 of Mukhopadhyay and de Silva (2008) that there exist $a > 0$ and $0 < r < 1$ such that $P_0(\tau_{\psi} \geq k) \leq ar^k$ for all natural $k$, so that (A.48) follows. The conditions of theorem 3.1 of Mukhopadhyay and de Silva (2008) are satisfied because of our model assumptions and (2.2).

Hence, by Corollary 4.2, $(\psi, \phi)$ has a minimum value of $K_0(\psi)$ among all tests $(\psi', \phi')$ satisfying the restrictions (4.15) on the error probabilities.

To prove that the same test minimizes $K_1(\psi)$, we can apply Lemma A.7 again, just interchanging the hypothesized distributions. Let us consider two simple hypotheses $H_0^0$ : “the true distribution is given by $f_0^0$” vs. $H_1^0$ : “the distribution corresponds to $f_1^0$”.

In a very natural way, any test $(\psi, \phi)$ from (5.19) and (5.20) is immediately adapted to the problem of testing $H_0^0$ vs. $H_1^0$ :

$$\psi_n^s(x^n) = \psi_n(x^n), \quad \phi_n^s(x^n) = 1 - \phi_n(x^n), \tag{A.49}$$

so that

$$I_{\{z_n^{-1} \in \{B^{-1}, A^{-1}\}\}} \leq 1 - \psi_n^s \leq I_{\{z_n^{-1} \in [B^{-1}, A^{-1}]\}}, \tag{A.50}$$

and

$$\phi_n^s = I_{\{z_n^{-1} > B^{-1}\}} \tag{A.51}$$

for all $n \in G_k$ and $k = 1, 2, \ldots$.

Let $\psi^*$ and $\beta^*$ denote the error probabilities in the problem of testing $H_0'$ vs. $H_1'$ and $K_0'$ the average cost of observations under $H_0'$. Then, obviously,

$$\alpha^*(\psi^*, \phi^*) = \beta(\psi, \phi), \quad \text{and} \quad \beta^*(\psi^*, \phi^*) = \alpha(\psi, \phi), \tag{A.52}$$

and $K_0'(\psi^*) = K_1(\psi)$.

Applying now Lemma A.7 to the problem of testing $H_0'$ vs. $H_1'$ in the same way we applied it above for testing $H_0$ vs. $H_1$, we get, by Corollary 4.2, that $(\psi, \phi)$ minimizes $K_1(\psi)$ among the tests $(\psi', \phi') \in S_1$ satisfying (4.15).

**ACKNOWLEDGMENTS**

The authors are grateful to the anonymous reviewers and editors for their valuable comments that contributed to the improvement of this work.

**DISCLOSURE**

The authors have no conflicts of interest to report.
FUNDING

A. Novikov thanks SNI by CONACyT, Mexico, for partial support for this work. X. I. Popoca-Jiménez thanks CONACyT, Mexico, for scholarships for her studies.

ORCID

A. Novikov [15] http://orcid.org/0000-0003-1034-8951

REFERENCES

Burkholder, D. L., and R. A. Wijsman. 1963. “Optimum Properties and Admissibility of Sequential Test.” The Annals of Mathematical Statistics 34 (1):1–17. doi:10.1214/aoms/1177704238

Fleming, T. R. 1982. “One-Sample Multiple Testing Procedure for Phase ii Clinical Trials.” Biometrics 38 (1):143–51. doi:10.2307/2530297

Jennison, C., and B. W. Turnbull. 2000. Group Sequential Methods with Applications to Clinical Trials. Boca Raton, FL: Chapman and Hall/CRC.

Lorden, G. 1980. “Structure of Sequential Tests Minimizing an Expected Sample Size.” Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete 51 (3):291–302. doi:10.1007/BF00587355

Mukhopadhyay, N., and B. M. de Silva. 2008. “Theory and Applications of a New Methodology for the Random Sequential Probability Ratio Test.” Statistical Methodology 5 (5):424–53. doi:10.1016/j.stamet.2007.10.002

Novikov, A. 2009. "Optimal Sequential Tests for Two Simple Hypotheses." Sequential Analysis 28 (2):188–217. doi:10.1080/07474940902816809

Novikov, A., A. Novikov, and F. Farkhshatov. “A computational approach to the Kiefer-Weiss problem for sampling from a Bernoulli population.” Sequential Analysis, 2022. arXiv:2110.04802 [stat.ME].

Rockafellar, R. T. 1970. Convex Analysis. New Jersey: Princeton Univ. Press.

Schmitz, N. 1993. “Optimal Sequentially Planned Decision Procedures.” In Lecture Notes in Statistics, edited by N. Schmitz, vol. 79. New York: Springer-Verlag.

Wald, A., and J. Wolowitz. 1948. “Optimum Character of the Sequential Probability Ratio Test.” The Annals of Mathematical Statistics 19 (3):326–39. doi:10.1214/aoms/1177730197