Design and performance evaluation in Kiefer-Weiss problems when sampling from discrete exponential families

Andrey Novikov and Fahil Farkhshatov

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

ABSTRACT
In this article, we deal with problems of testing hypotheses in the framework of sequential statistical analysis. The main concern is the optimal design and performance evaluation of sampling plans in the Kiefer-Weiss problems. For the case of observations following a discrete exponential family, we provide algorithms for optimal design in the modified Kiefer-Weiss problem and obtain formulas for evaluating their performance, calculating operating characteristic function, average sample number, and some related characteristics. These formulas cover, as a particular case, sequential probability ratio tests (SPRTs) and their truncated versions, as well as optimal finite-horizon sequential tests. On the basis of the developed algorithms we propose a method of construction of optimal tests and their performance evaluation for the original Kiefer-Weiss problem. All algorithms are implemented as functions in the R programming language and can be downloaded from https://github.com/tosina-base/Kiefer-Weiss, where the code for the binomial, Poisson, and negative binomial distributions is readily available. Finally, we make numerical comparisons of the Kiefer-Weiss solution with the SPRT and the fixed-sample-size test having the same levels of the error probabilities.

1. INTRODUCTION
In a sequential statistical experiment, a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ is potentially available to the statistician on a one-by-one basis. The observed data bear the information about the underlying distribution $P_\theta$, where $\theta$ is an unknown parameter whose true value is of interest to the statistician. In this article, we are concerned with testing a simple hypothesis $H_0 : \theta = \theta_0$ against a simple alternative $H_1 : \theta = \theta_1$, in the framework of sequential analysis and, more specifically, in the Kiefer-Weiss setting (see Kiefer and Weiss 1957).

In its simplest form, a sequential hypothesis test is a pair $(\tau, \delta)$ consisting of a stopping time $\tau$ and a (terminal) decision rule $\delta$. Formally, it is required that $\{\tau = n\} \in \sigma(X_1, \ldots, X_n)$ and $\{\tau = n, \delta = i\} \in \sigma(X_1, \ldots, X_n)$, for any natural $n$ and for $i = 0, 1$. The
performance characteristics of a sequential test are the type I and type II error probabilities, \( \alpha(\tau, \delta) = P_{\theta_0}(\delta = 1) \) and \( \beta(\tau, \delta) = P_{\theta_1}(\delta = 0) \), and the average sample number, \( E_\theta \tau \).

The Kiefer-Weiss problem consists in finding a test \( \langle \tau, \delta \rangle \) with a minimum value of \( \sup_\theta E_\theta \tau \) among all tests satisfying the constraints on the type I and type II error probabilities:

\[
\alpha(\tau, \delta) \leq \alpha \quad \text{and} \quad \beta(\tau, \delta) \leq \beta.
\]

This article is meant to be a continuation of our recent publication (Novikov, Novikov, and Farkhshatov 2022) on the Kiefer-Weiss problem, where we investigated the case of sampling from a Bernoulli population. A brief review of relevant general results concerning the Kiefer-Weiss problem can be found in the Introduction to Novikov, Novikov, and Farkhshatov (2022).

In this article we consider models of independent and identically distributed (i.i.d.) observations that follow a discrete exponential (Koopman-Darmois) family. We are concerned with the Kiefer-Weiss and related problems, where we propose a unified approach to the optimal design and performance evaluation of sampling plans. We develop a whole set of computational algorithms for this. The algorithms are fully implemented in the R programming language (R Core Team 2013) for the binomial, Poisson, and negative binomial (Pascal) distributions. Performance evaluation for the sequential probability ratio test (SPRT) and truncated SPRT is implemented as well. The program code can be downloaded from the GitHub repository at https://github.com/tosinabase/Kiefer-Weiss.

Using the developed computer code, in the implemented models, for a range of error probabilities \( \alpha, \beta \), we calculate the parameters of optimal sampling plans and their characteristics, as well as those of Wald’s SPRT and fixed-sample-size tests with the same levels of type I and type II error probabilities.

In Section 2, we lay out general results our solutions to the Kiefer-Weiss problem are based on for the models we are concerned with. In Section 3, we develop computing algorithms for optimal design and performance evaluation of the respective sequential sampling plans and present the numerical results. Section 4 provides a discussion of the results and marks some direction for further work.

2. KIEFER-WEISS PROBLEM

In this section, we formulate general results on the structure of the optimal tests in the Kiefer-Weiss problem and its modified version.

2.1. Optimal Sampling Plans

In this section we follow the notation and general assumptions of Novikov (2009).

In particular, we consider sequential tests \( \langle \psi, \phi \rangle \), with \( \psi = (\psi_1, \psi_2, ...) \) being a stopping rule and \( \phi = (\phi_1, \phi_2, ...) \) being a (terminal) decision rule. After some data \((x_1, ..., x_n)\) have been observed, \( \psi_n = \psi_n(x_1, ..., x_n) \) means the (conditional) probability to stop at stage \( n \), and \( \phi_n = \phi_n(x_1, ..., x_n) \) means the probability to reject \( H_0 \) given the data \((x_1, ..., x_n)\), after the decision to stop has been reached, \( n = 1, 2, ... \).
Denoting $c_n = c_n(x_1, ..., x_n) = (1 - \psi_1)(1 - \psi_2)...(1 - \psi_{n-1})$ and $s_n = c_n\psi_n$, we have for any test $\langle \psi, \phi \rangle$ the definition of the error probability of type I

$$
\alpha(\psi, \phi) = \sum_{n=1}^{\infty} E_{\theta_0} s_n^\psi \phi_n,
$$
of the error probability of type II

$$
\beta(\psi, \phi) = \sum_{n=1}^{\infty} E_{\theta_0} s_n^\psi (1 - \phi_n),
$$
and the average sample number

$$
N(\theta; \psi) = \sum_{n=1}^{\infty} n E_{\theta_0} s_n^\psi
$$
when the true parameter value is $\theta$ (provided that $\sum_{n=1}^{\infty} E_{\theta_0} s_n^\psi = 1$; otherwise, it is infinite).

It is common to express the error probabilities in terms of the operating characteristic curve, defined as

$$
OC_{\theta}(\psi, \phi) = \sum_{n=1}^{\infty} E_{\theta} s_n^\psi (1 - \phi_n),
$$
where $\alpha(\psi, \phi) = 1 - OC_{\theta_0}(\psi, \phi)$, $\beta(\psi, \phi) = OC_{\theta_1}(\psi, \phi)$.

Let $\mathcal{S}(\alpha, \beta)$ be the set of all tests such that

$$
\alpha(\psi, \phi) \leq \alpha, \quad \beta(\psi, \phi) \leq \beta,
$$
where $\alpha, \beta \in [0, 1]$ are some fixed real numbers.

We are interested in finding tests minimizing $\sup_\theta N(\theta; \psi)$ over all of the tests in $\mathcal{S}(\alpha, \beta)$. This problem is known as the Kiefer-Weiss problem.

The respective modified Kiefer-Weiss problem is to minimize $N(\theta; \psi)$ over all $\langle \psi, \phi \rangle \in \mathcal{S}(\alpha, \beta)$, for a given fixed value of $\theta$.

All known solutions of the Kiefer-Weiss problem, for particular models, have been obtained through the modified version of the problem (see Kiefer and Weiss 1957; Weiss 1962; Freeman and Weiss 1964; Lai 1973; Lorden 1980; Huffman 1983; Zhitlukhin, Muravlev, and Shiryaev 2013; Tartakovsky, Nikiforov, and Basseville 2014, among many others).

The solutions to the modified Kiefer-Weiss problem can be obtained, at least in theory, in a very general situation using the following variant of the Lagrange multipliers method.

Let

$$
L_\theta(\psi, \phi) = N(\theta; \psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi),
$$
where $\theta$ is some fixed value of the parameter and $\lambda_0, \lambda_1$ are some nonnegative constants (called Lagrange multipliers).

Then the tests minimizing $N(\theta; \psi)$ subject to (2.1) can be obtained through an unconstrained minimization of $L_\theta(\psi, \phi)$ over all $\langle \psi, \phi \rangle$, using an appropriate choice of the Lagrange multipliers (see Novikov 2009, section 2).
Lorden (1980) showed that in the case of i.i.d. observations the problem of minimizing the Lagrangian function is reduced to an optimal stopping problem for a Markov process.

It is easy to see that finding Bayesian tests used in Kiefer and Weiss (1957) is mathematically equivalent to the minimization of (2.2).

To construct the optimal tests we need an additional assumption on the distribution of the observations. Let \( f_n^n = f_n^n(x_1, ..., x_n) \) be the Radon-Nikodym derivative of the distribution of \( X_1, ..., X_n \) with respect to a product measure \( \mu^n = \mu \otimes ... \otimes \mu \) (\( n \) times \( \mu \) by itself), \( n = 1, 2, ... \).

In Lorden (1980), it is shown that, in the case of i.i.d. observations following a distribution from a Koopman-Darmois family, the tests giving solution to the modified problem have bounded with probability one stopping times when \( h = (h_0, h_1) \).

Let us describe the construction of tests minimizing the Lagrangian function calculated at some \( h \), over all truncated tests; that is, those not taking more than a fixed number \( H \) of observations (\( H \) is also called horizon in this case).

Formally, let \( \mathcal{S}^H = \{ (\psi, \phi) : c_H^\psi = 0 \} \) be the class of all such tests.

The structure of tests minimizing the Lagrangian function in \( \mathcal{S}^H \) can be characterized in the following way.

Let us define
\[
L_{\theta}^H(\psi, \phi) = \min \{ L_{\theta}^H(\lambda_0 f^n_{\theta_0}, \lambda_1 f^n_{\theta_1}) \},
\]
and, recursively over \( n = H - 1, ..., 1 \),
\[
L_{\theta,n}^H = \min \{ L_{\theta,n}^H(\lambda_0 f^n_{\theta_0}, \lambda_1 f^n_{\theta_1} + \mathcal{I} V_{\theta,n+1}^H) \},
\]
where
\[
\mathcal{I} V_{\theta,n+1}^H = (\mathcal{I} V_{\theta,n+1}^H)(x_1, ..., x_n) = \int V_{\theta,n+1}^H(x_1, ..., x_{n+1}) d\mu(x_{n+1}).
\]

Note that functions \( V_{\theta,n}^H \) defined above, as well as \( L_{\theta}(\psi, \phi) \) in (2.2), implicitly depend on \( \lambda_0, \lambda_1 \).

From proposition 1 of Novikov, Novikov, and Farkhshatov (2022), we obtain that for any \( (\psi, \phi) \in \mathcal{S}^H \),
\[
L_{\theta}(\psi, \phi) \geq 1 + 3 V_{\theta,1}^H,
\]
and the right-hand side in (2.6) is attained if
\[
\psi_n = I_{(\min \{ \lambda_0 f^n_{\theta_0}, \lambda_1 f^n_{\theta_1} \}) \leq f^n_{\theta_0} + \mathcal{I} V_{\theta,n+1}^H} \]
for \( n = 1, 2, ..., H - 1 \), and \( \psi_H \equiv 1 \), and
\[
\phi_n = I_{\lambda_0 f^n_{\theta_0} \leq \lambda_1 f^n_{\theta_1}} \]
for \( n = 1, 2, ..., H \).

The tests with strict inequalities in some (or all) of the inequalities in (2.7) and/or (2.8) also attain the lower bound (2.6), as do the respectively randomized tests (i.e., those taking some values between 0 and 1 in case the corresponding equality in (2.7) and/or (2.8) takes place).
Let us denote $\mathcal{M}_0^H$ the set of all such tests, and let $\mathcal{M}^H = \cup_0 \mathcal{M}_0^H$.

Under very mild conditions, the optimal nontruncated tests are obtained on the basis of limits $V_{0,n} = \lim_{H \to \infty} V_{0,n}^H$. Optimal stopping rules for the nontruncated tests are obtained substituting $V_{0,n}$ for $V_{0,n}^H$ in (2.7):

$$\psi_n = I\{\min\{\zeta_0 f_{H_0}^n, \zeta f_{0}^n\} \leq f_{H_0}^n + IV_{0,n+1}\} \quad (2.9)$$

for all natural $n$, applying (2.8) for all $n$ as the terminal decision rule (see section 3 in Novikov 2009). When the optimal test is truncated, it holds that $V_{0,n} = V_{0,n}^H$ for large enough $H$.

Again, all of the randomized tests (taking values between 0 and 1 in case of an equality in (2.8) and/or (2.9)) share the same optimality property minimizing $L_\theta(\psi, \phi)$.

Let $\mathcal{M}_0$ be the class of tests $\langle \psi, \phi \rangle$ satisfying (2.8) and (2.9) for all $n$ (including those with strict inequalities and those respectively randomized), and $\mathcal{M} = \cup_0 \mathcal{M}_0$.

Proposition 2 in Novikov, Novikov, and Farkhshatov (2022) offers the following way of solving the Kiefer-Weiss problem.

Take some $\langle \psi, \phi \rangle \in \mathcal{M}_0$, and let

$$\sup_{\theta} N(\theta; \psi^*) - N(\theta; \psi) = \Delta. \quad (2.10)$$

Then

$$\sup_{\theta} N(\theta; \psi^*) \leq \inf_{\theta} \sup_{\theta} N(\theta; \psi) + \Delta, \quad (2.11)$$

where the infimum is taken over all tests $\langle \psi, \phi \rangle$ such that $\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*)$ and $\beta(\psi, \phi) \leq \beta(\psi^*, \phi^*)$. It follows from (2.11) that $\langle \psi^*, \phi^* \rangle$ is as close as to within $\Delta$ to the solution to the Kiefer-Weiss problem. In symmetrical cases, $\Delta = 0$ if one takes $\theta = (\theta_0 + \theta_1)/2$ (see Kiefer and Weiss 1957; Weiss 1962; Lai 1973; Lorden 1980, among others).

In Novikov, Novikov, and Farkhshatov (2022) we proposed a numerical solution seeking a minimum value of $\Delta$ over $\mathcal{M}$ by means of a computer program. For the case of Bernoulli observations, we wrote a program code in R language for the numerical minimization and ran it for a broad range of $\theta_0$, $\theta_1$ and $\alpha$ and $\beta$. In all evaluated examples the minimum value of $\Delta$ was very close to 0 (see Novikov, Novikov, and Farkhshatov 2022). Later in this article, we show that the same happens when sampling from binomial, Poisson, and geometric distributions.

Below we describe the computer algorithms for implementation of the proposed method for these, and more general, discrete one-parameter exponential families of distributions.

### 3. ALGORITHMS FOR COMPUTING OPTIMAL SAMPLING PLANS

In this section, we describe computational algorithms for the modified Kiefer-Weiss problem in the case when the hypothesized distributions come from a discrete exponential family. They are based on the algorithms we developed for sampling from a Bernoulli population (see Novikov, Novikov, and Farkhshatov 2022). The application to
the Kiefer-Weiss problem straightforwardly follows from proposition 2 therein; in fact, the algorithmic part stays unchanged.

The proposed algorithms are implemented in the form of a program code in the R programming language (R Core Team 2013) and are available as a part of GitHub repository at https://github.com/tosinabase/Kiefer-Weiss.

3.1. Optimal Tests in Modified Kiefer-Weiss Problem

Let the observations $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. random variables following a distribution from a discrete one-parameter exponential family. More specifically, we assume that

$$f_h(x) = \exp \left( h(x) \right), x \in \mathbb{Z}^+, \quad (3.1)$$

where $h(x)$ is a nonnegative function on $\mathbb{Z}^+$. Then the joint probability

$$f_n^h(x_1, \ldots, x_n) = \exp \left( \theta s_n - nb(\theta) \right) h_n(x_1, \ldots, x_n),$$

where $s_n = \sum_{i=1}^n x_i$, and $h_n(x_1, \ldots, x_n) = \prod_{i=1}^n h(x_i)$, $n = 1, 2, 3, \ldots$ It is well known that under very general assumptions $E h X_i = b'(\theta)$ and $\text{Var} h X_i = b''(\theta) > 0$ for all $\theta$.

Let $\theta_0 < \theta_1$ be two fixed parameter values defining the hypotheses $H_0$ and $H_1$.

Let us construct a solution to the modified Kiefer-Weiss problem for a given $\theta$ when sampling from a distribution of type (3.1).

Let us express (2.3)–(2.5) in a more “computer-friendly” form, namely, using the distribution of the sufficient statistic $\sum_{i=1}^n X_i$ (see Novikov, Novikov, and Farkhshatov 2022). This helps to avoid dealing with very small numbers representing the joint probabilities in (2.3), etc., in case the truncation level $H$ is high.

Let

$$g^0_n(s) = C_n(s) \exp (\theta s - nb(\theta)), \quad s \in \mathbb{Z}^+, \quad (3.1)$$

be the probability mass function of the statistic $\sum_{i=1}^n X_i$, for any natural $n$.

Now, let us define

$$U^H_{\theta, H}(s) = \min \{ \lambda_0 g^H_{\theta_0}(s), \lambda_1 g^H_{\theta_1}(s) \}, \quad s \in \mathbb{Z}^+,$$

and, recursively over $n = H, \ldots, 2$,

$$U^H_{\theta, n-1}(s) = \min \{ \lambda_0 g_{\theta_0}^{n-1}(s), \lambda_1 g_{\theta_1}^{n-1}(s), g_0^{n-1} + J_n U^H_{\theta, n}(s) \}, \quad s \in \mathbb{Z}^+,$$

where $J_n U(s) = \sum_x U(s+x) d_n(x, s), \quad s \in \mathbb{Z}^+$, and

$$d_n(x, s) = C_1(x) \frac{C_{n-1}(s)}{C_n(s+x)}, \quad x, s \in \mathbb{Z}^+.$$

Proposition 3.1.

$$V^H_{\theta, n}(x_1, \ldots, x_n) = U^H_{\theta, n}(s_n) h_n(x_1, \ldots, x_n) / C_n(s_n), \quad (3.2)$$

where $s_n = \sum_{i=1}^n x_i$, for $n = 1, \ldots, H$.

Proof. By induction over $n = H, H - 1, \ldots, 1$. 

It follows from (2.3) that
\[
V_{0,H}^H(x_1, \ldots, x_H) = \min \{ \lambda_0 f_{00}^H(x_1, \ldots, x_H), \lambda_1 f_{01}^H(x_1, \ldots, x_H) \}
\]
\[
= \min \{ \lambda_0 g_{00}^H(s_H), \lambda_1 g_{01}^H(s_H) \} \frac{h_H(x_1, \ldots, x_H)}{C_H(s_H)}
\]
\[
= U_{0,H}^H(s_H) \frac{h_H(x_1, \ldots, x_H)}{C_H(s_H)}.
\]

Let us suppose that (3.2) holds for some \(2 \leq n \leq H\). Then, by virtue of (2.4),
\[
V_{0,n-1}^H = \min \left\{ \lambda_0 f_{00}^{n-1}, \lambda_1 f_{01}^{n-1}, f_{0n}^{n-1} + \sum_{x_n} U_{0,n}^H(x_1 + \ldots + x_n) \frac{h_n(x_1, \ldots, x_n)}{C_n(s_n)} \right\}
\]
\[
= \min \left\{ \lambda_0 g_{00}^{n-1}, \lambda_1 g_{01}^{n-1}, g_{0n}^{n-1} + \sum_{x_n} U_{0,n}^H(s_n + x_n)C_1(x_n) \frac{C_{n-1}(s_{n-1})}{C_n(s_{n-1} + x_n)} \right\}
\]
\[
\times \frac{h_{n-1}(x_1, \ldots, x_{n-1})}{C_{n-1}(s_{n-1})} = U_{0,n-1}^H(s_{n-1}) \frac{h_{n-1}(x_1, \ldots, x_{n-1})}{C_{n-1}(s_{n-1})},
\]

Now, for any truncation level \(H\), we have optimal tests for the modified Kiefer-Weiss problem in the form of (3.3)–(3.4) (along with their strict inequality versions), for any fixed \(\theta\). Respectively, (2.9) acquires the form
\[
\psi_n = I\{ \lambda_0 g_{00}^n, x_1, \ldots, x_n \leq \lambda_1 g_{01}^n + \sum_{x_n} U_{0,n}^H \}
\]
\[
= I\{ \lambda_0 g_{00}^n, x_1, \ldots, x_n \leq \lambda_1 g_{01}^n \}.
\]

It is easily seen from the proof that the optimal stopping rule (2.7) is equivalent to
\[
\psi_n = I\{ \lambda_0 g_{00}^n, x_1, \ldots, x_n \leq \lambda_1 g_{01}^n + \sum_{x_n} U_{0,n}^H \},
\]
\[
\phi_n = I\{ \lambda_0 g_{00}^n, x_1, \ldots, x_n \leq \lambda_1 g_{01}^n \}.
\]

Now, for any truncation level \(H\), we have optimal tests for the modified Kiefer-Weiss problem in the form of (3.3)–(3.4) (along with their strict inequality versions), for any fixed \(\theta\). Respectively, (2.9) acquires the form
\[
\psi_n = I\{ \lambda_0 g_{00}^n, x_1, \ldots, x_n \leq \lambda_1 g_{01}^n + \sum_{x_n} U_{0,n}^H \},
\]
\[
\phi_n = I\{ \lambda_0 g_{00}^n, x_1, \ldots, x_n \leq \lambda_1 g_{01}^n \}.
\]

3.2. Bounds for Continuation Regions

In this part, we find bounds for continuation regions of optimal tests in the modified Kiefer-Weiss problem.

Let us start with the case \(\theta \in (\theta_0, \theta_1)\).

It follows from (3.5) that \(\psi_n(s) = 1\) if \(\lambda_0 g_{00}^n (s) < g_{01}^n (s)\), and this is equivalent to
\[
s > \frac{\log \lambda_0}{\theta - \theta_0} + n \frac{b(\theta) - b(\theta_0)}{\theta - \theta_0}.
\]

So \(\psi_n(s)\) can only be 0 when \(s \leq B_n(\lambda_0, \theta_0, \theta)\), where \(B_n(\lambda_0, \theta_0, \theta)\) is a maximum integer less than or equal to the right-hand side of (3.6).
In the same way, \( \psi_n(s) = 1 \) if \( \lambda_1 g^n_{\hat{\theta}_1}(s) < g^n_{\theta}(s) \), and this is equivalent to

\[
\frac{-\log \lambda_1}{\theta_1 - \theta} + n \left( \frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} - \frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} \right) < 0,
\]

so \( \psi_n(s) \) can only be 0 when \( s \geq A_n(\lambda_1, \theta_1, \theta) \), where \( A_n(\lambda_1, \theta_1, \theta) \) is a minimum integer greater than or equal to the right-hand side of (3.7).

Therefore,

\[
\{s : \psi_n(s) < 1\} \subset \{s : A_n(\lambda_1, \theta_1, \theta) \leq s \leq B_n(\lambda_0, \theta_0, \theta)\},
\]

whenever

\[
A_n(\lambda_1, \theta_1, \theta) \leq B_n(\lambda_0, \theta_0, \theta).
\]

Then, the set of \( s \) for which the test continues is finite. In addition, there is no such \( s \) that \( \psi_n(s) < 1 \) (implying that \( \psi_n(s) \equiv 1 \)) if

\[
A_n(\lambda_1, \theta_1, \theta) > B_n(\lambda_0, \theta_0, \theta),
\]

which is equivalent, due to (3.6) and (3.7), to

\[
\log \lambda_0 + \frac{\log \lambda_1}{\theta_1 - \theta} + n \left( \frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} - \frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} \right) < 0.
\]

Let us show that in case \( \theta \in (\theta_0, \theta_1) \) the expression in parentheses in (3.9) is negative. Denoting

\[
G(\theta) = (b(\theta) - b(\theta_0))(\theta_1 - \theta) - (b(\theta_1) - b(\theta))(\theta - \theta_0),
\]

we have \( G(\theta_0) = G(\theta_1) = 0 \), and calculating the second derivative, \( G''(\theta) = b''(\theta)(\theta_1 - \theta_0) \), which is positive because \( b''(\theta) = \text{Var}_X > 0 \). Thus, \( G(\theta) \) is convex, resulting in \( G(\theta) < 0 \) for \( \theta \in (\theta_0, \theta_1) \). Therefore, (3.9) is equivalent to

\[
n > \left( \frac{\log \lambda_0}{\theta - \theta_0} + \frac{\log (\lambda_1)}{\theta_1 - \theta} \right) / \left( \frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} - \frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} \right).
\]

Let us denote \( H = H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta) \) the maximum integer less than or equal to the right-hand side of (3.11). We know now that \( H \) is the last step when an optimal \( \psi_n \) would be allowed to continue, when \( \theta \in (\theta_0, \theta_1) \).

There is no bound for \( H \) if \( \theta \not\in (\theta_0, \theta_1) \).

If \( \theta \geq \theta_1 \), then

\[
\{s : \phi_n(s) < 1\} \subset \{s : 0 \leq s \leq B_n(\lambda_0, \theta, \theta)\}
\]

for all natural \( n \).

Finally, if \( \theta \leq \theta_0 \), then

\[
\{s : \phi_n(s) < 1\} \subset \{s : s \geq A_n(\lambda_1, \theta_1, \theta)\}
\]

for all natural \( n \).

To conclude this part, let us note that the optimal decision rule in (3.4) can be defined as

\[
\phi_n = 1 - I_{[0, B_n(\lambda_0/\lambda_1, \theta_0, \theta_1)]}
\]

for all natural \( n \) (cf. equation 3.6 and the definition of \( B_n \)).
3.3. Computational Algorithms for Designing Optimal Tests

The theoretical basis is formulas (3.3) and (3.4).

Let us consider first the case when \( \theta \in (\theta_0, \theta_1) \).

It follows from (3.8) that the continuation region is finite (or empty) for any natural \( n \). Therefore, the optimal test will be completely characterized by a set of continuation intervals \([a_n, b_n] \), \( n = 1, \ldots, H - 1 \), where \( H \) is the maximum number of observations the test may take. We know from the previous subsection that \( H \leq H(\hat{\lambda}_0, \hat{\lambda}_1, \theta_0, \theta_1, \theta) \).

Thus, an optimal test can be obtained “working backward” from \( H = H(\hat{\lambda}_0, \hat{\lambda}_1, \theta_0, \theta_1, \theta) \) in the following way:

1. Set \( n = H \).
2. Define \( a_{n-1} \) and \( b_{n-1} \), respectively, as a minimum and a maximum \( s \in \mathbb{Z}^+ \), for which
   \[
   g_0^{n-1}(s) + \mathcal{J}_n U_{\hat{\lambda}, n}(s) < \min\{g_{\hat{\lambda}}^{n-1}(s), g_{\hat{\lambda}_1}^{n-1}(s)\}. \tag{3.12}
   
   \]
   If no such \( s \) exist, set \( H = n - 1 \). If \( H = 1 \), declare “stop after the first step” state and stop; else, go to Step 1.
3. For \( s \in [a_{n-1}, b_{n-1}] \), store \( v_{n-1}(s) = \min\{g_0^{n-1}(s) + \mathcal{J}_n U_{\hat{\lambda}, n}(s), g_{\hat{\lambda}}^{n-1}(s), g_{\hat{\lambda}_1}^{n-1}(s)\} \).
   Take into account, for future use, that
   \[
   U_{\hat{\lambda}, n-1}(s) = \begin{cases} 
   v_{n-1}(s), & \text{if } s \in [a_{n-1}, b_{n-1}] \\
   \min\{g_0^{n-1}(s), g_{\hat{\lambda}_1}^{n-1}(s)\}, & \text{otherwise}.
   \end{cases}
   \]
4. If \( n = 2 \), then stop; else, set \( n = n - 1 \), and go to Step 2.

“Stop after the first step” as a result of this algorithm means that the optimum test has to stop after the first observation is taken (thus it is not sequential). Usually, this means that \( \hat{\lambda}_0 \) and/or \( \hat{\lambda}_1 \) are too small to produce a meaningful sequential test. An optimal test can be defined in this case as \( \psi_1 \equiv 1 \) and \( \phi_1 = 1 - I_{[0, b_1(\hat{\lambda}_0, \hat{\lambda}_1, \theta_0, \theta_1)]}(s) \).

If the algorithm does not terminate in “stop after the first step,” we have an optimal test with continuation intervals \([a_n, b_n] \), \( n = 1, 2, \ldots, H - 1 \), with a maximum number of observations equal to \( H \).

If \( \theta \notin (\theta_0, \theta_1) \), the above algorithm can be used, after some modifications, to obtain optimal tests in the class of truncated tests \( \mathcal{M}^H_{\hat{\lambda}} \). The modifications concern the way the “backward induction” in Step 2 is performed, because there is only one bound for the continuation region in any one of the cases seen in Subsection 3.2, when \( \theta \notin (\theta_0, \theta_1) \). For example, if \( \theta > \theta_1 \), we seek \( s \) satisfying (3.12) downwards, starting from \( B_{n-1}(\lambda_0, \theta_0, \theta) \), and then, starting from \( s = 0 \), upwards, obtaining \( b_{n-1} \) and \( a_{n-1} \) (if any), respectively.

If \( \theta = \theta_1 \), we can narrow the search region in Step 2, because it is known (see Novikov and Popoca Jiménez 2022) that the continuation region (if not empty) is always an interval when sampling from an exponential family, so, after \( b_{n-1} \) has been found, we may want to keep searching for \( a_{n-1} \) downwards.
There is a very similar situation in the case $\theta = \theta_0$, where we search for $a_{n-1}$ upwards, starting from $A_{n-1}(\lambda_1, \theta_1, \theta_0)$, and then keep searching upwards until $b_{n-1}$ is found.

The only remaining case $\theta < \theta_0$ is the hardest one, because there is no upper bound for the continuation region, unless the distribution of $S_{n-1}$ is bounded (as in the binomial case). So, in our computer implementation, we search for $a_{n-1}$ upwards, starting from $A_{n-1}(\lambda_1, \theta_1, \theta_0)$, and, after $a_{n-1}$ has been found, keep searching upwards for $b_{n-1}$ indefinitely. When $b_{n-1}$ is found, it will correspond to the optimal test unless there is an $s > b_{n-1}$ satisfying (3.12). Unfortunately, there is no theoretical result that could guarantee that the continuation region is an interval in this case. Nevertheless, we implement the algorithm as described above, so one has to take into account that the test obtained in the modified Kiefer-Weiss problem in case $\theta < \theta_0$ can be suboptimal.

The exact algorithm we implemented for the Bernoulli case in Novikov, Novikov, and Farkhshatov (2022) in particular cases we experimented with gives the same results.

As shown in Hawix and Schmitz (1998), the optimal tests for modified Kiefer-Weiss problem with $\theta > \theta_1$ or $\theta < \theta_0$ may not stop with probability one under $H_0$ and/or $H_1$, so they are not of particular importance for applications. Our finite-horizon implementation in this section may be useful in these situations, because it terminates with probability one under each one of the hypotheses and gets close to the optimal modified Kiefer-Weiss test as $H \to \infty$. Nevertheless, high values of the average sample number should be expected where they are infinite in the infinite-horizon case. Instead, one may want to include the average sample number, whose large value causes the problem, as an additional criterion for the Lagrangian minimization, with some small weight. For example, if one wants to minimize $N(\theta; \psi)$ with $\theta < \theta_0$, the Lagrangian function may also include a term $cN(\theta_1; \psi)$ with some small $c$. The only change to the algorithm above would be using

$$g^{n-1}_\theta(s) + c g^{n-1}_{\theta_1}(s) + J_n U_{\theta, n}(s) < \min\{g^{n-1}_{\theta_0}(s), g^{n-1}_{\theta_1}(s)\}.$$  

in place of (3.12). This could be helpful in the applied problem mentioned by Hawix and Schmitz (1998), where the modified Kiefer-Weiss problem is shown to be largely useless. This will no longer be a Kiefer-Weiss problem, so we do not go into detail. Interested readers can easily modify our code on their own to obtain a meaningful test for this problem. The rest of the algorithms for performance evaluation we develop below will be applicable without any change.

### 3.4. Operating Characteristic, Average Sample Number, and Related Formulas

In this paragraph, we obtain formulas for calculating error probabilities, average sample number, and some related probabilities for truncated sequential tests.

Let $(\psi, \phi) \in \mathcal{F}^H$. This test will be held fixed within this subsection, so we will suppress it in the notation.

Let

$$a_H^{H}(s) = 1 - \phi_H(s), \ s \in \mathbb{Z}^+,$$  

and, recursively over $n = H - 1, H - 2, \ldots, 1$,  

\[(3.13)\]
\[ a_0^n(s) = \psi_n(s)(1 - \phi_n(s)) + (1 - \psi_n(s))E_0 a_0^{n+1}(s + X_{n+1}), \quad s \in \mathbb{Z}^+. \]

Let us denote \( D_n^H = D_n^H(\psi, \phi) \) the event meaning “\( H_0 \) is accepted at or after step \( n \) as a result of applying the test \( \langle \psi, \phi \rangle \)” in particular, the operating characteristic curve is \( OC_\theta(\psi, \phi) = P_\theta(D_n^H(\psi, \phi)) \).

Let \( S_n = \sum_{i=1}^n X_i \) for any natural \( n \), and \( S_0 = 0 \).

**Proposition 3.2.** For any \( 1 \leq n \leq H \),
\[
a_0^n(S_n) = P_\theta(D_n^H|X_1, \ldots, X_n), \tag{3.14}
\]
and
\[
OC_\theta(\psi, \phi) = E_0 a_0^1(X_1).
\]

**Proof.** By induction over \( n = H, H-1, \ldots, 1 \).

For \( n = H \), it follows from (3.13) that \( a_0^H(S_H) = 1 - \phi_H(S_H) = P_\theta\{D_n^H|X_1, \ldots, X_H\} \) (this latter equality follows from the definition of the decision function \( \phi \)).

Let us suppose now that (3.14) holds for some \( 2 \leq n \leq H \).

Then
\[
a_0^{n-1}(S_{n-1}) = \psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1}))
+ (1 - \psi_{n-1}(S_{n-1}))E_0\{a_0^n(S_{n-1} + X_n)|X_1, \ldots, X_{n-1}\}
= \psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1}))
+ (1 - \psi_{n-1}(S_{n-1}))E_0\{P_\theta\{D_n^H|X_1, \ldots, X_n\}|X_1, \ldots, X_{n-1}\}
= \psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1}))
+ (1 - \psi_{n-1}(S_{n-1}))E_0\{P_\theta\{D_n^H|X_1, \ldots, X_n\}|X_1, \ldots, X_{n-1}\}
= E_0\{\psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1})) + (1 - \psi_{n-1}(S_{n-1}))ID_{H_n^2}|X_1, \ldots, X_{n-1}\}
= P_\theta\{ID_{H_{n-1}}|X_1, \ldots, X_{n-1}\}
\]

In a similar way, for any \( k < H \), let
\[
b_{0,k}^k(s) = (1 - \psi_k(s)), \quad s \in \mathbb{Z}^+, \tag{3.15}
\]
and, recursively over \( n = k-1, k-2, \ldots, 1 \),
\[
b_{0,n}^k(s) = E_0 b_{0,n+1}^k(s + X_{n+1})(1 - \psi_n(s)), \quad s \in \mathbb{Z}^+. \tag{3.16}
\]

**Proposition 3.3.** For all \( n \leq k < H \),
\[
b_{0,n}^k(S_n) = E_0\{(1 - \psi_n)\cdots(1 - \psi_k)|X_1, \ldots, X_n\}. \tag{3.17}
\]

In particular, \( P_\theta(\tau_\psi > k) = E_0 b_{0,1}^k(X_1) \).

**Proof.** By induction over \( n = k, k-1, \ldots, 1 \).

For \( n = k \), it follows from (3.15) that \( b_{0,k}^k(S_k) = 1 - \psi_k(S_k) = E_0\{1 - \psi_k(S_k)|X_1, \ldots, X_k\} \).

Let us suppose now that (3.17) holds for some \( 2 \leq n \leq k \).
Then
\[
\beta_{j,n-1}(S_{n-1}) = (1 - \psi_{n-1}(S_{n-1}))E_{0}\{b_{0,n}^j(S_{n-1} + X_n)|X_1, \ldots, X_{n-1}\} \\
= (1 - \psi_{n-1}(S_{n-1}))E_{0}\{(1 - \psi_n)(1 - \psi_k)|X_1, \ldots, X_n|X_1, \ldots, X_{n-1}\} \\
= (1 - \psi_{n-1}(S_{n-1}))E_{0}\{(1 - \psi_n)(1 - \psi_k)|X_1, \ldots, X_{n-1}\} \\
= E_0\{(1 - \psi_{n-1})(1 - \psi_k)|X_1, \ldots, X_{n-1}\}
\]

The average sample number can now be calculated as \(N(\theta; \psi) = \sum_{k=1}^{H} P_{\theta}(\tau_\psi \geq k)\).

A more direct way is to incorporate the calculation of this sum into (3.16). In fact, we can apply the algorithm to any stopping rule \(\psi = (\psi_1, \psi_2, \ldots)\) by truncating it at level \(H\); that is, defining \(\psi^H = (\psi_1, \ldots, \psi_H \equiv 1, \ldots)\).

**Corollary 3.1.** Let
\[
d_H^{H-1}(s) = (1 - \psi_{H-1}(s)), \ s \in \mathbb{Z}^+,
\]
and, recursively over \(n = H - 2, H - 3, \ldots, 1,\)
\[
d_H^{H,n}(s) = (1 + E_0d_H^{H,n+1}(s + X_{n+1}))(1 - \psi_n(s)), \ s \in \mathbb{Z}^+.
\]

Then \(N(\theta; \psi^H) = 1 + E_0d_H^{H,1}(X_1)\).

We consider meaningful for applications only the tests for which \(N(\theta; \psi^H) \rightarrow N(\theta; \psi)\) as \(H \rightarrow \infty\). Thus, **Corollary 3.1** provides a method for numerical evaluation of the average sample number \(N(\theta; \psi)\) as \(\lim_{H \rightarrow \infty} (1 + E_0d_H^{H,1}(X_1))\). In particular, we use this fact to calculate the average sample size of the SPRT and of the optimal tests in the modified Kiefer-Weiss problems when \(\theta \not\in (\theta_0, \theta_1)\) (see the code in [https://github.com/tosinabase/Kiefer-Weiss](https://github.com/tosinabase/Kiefer-Weiss)).

### 3.5. Numerical Results

The main goal of this part is to illustrate the use of the developed algorithms on concrete examples of discrete exponential families of distributions, to show that the obtained sequential tests in any one of the examples provide numerical solutions to the Kiefer-Weiss problem, and to analyze the efficiency of the obtained tests with respect to the classical sequential probability ratio tests and the fixed-sample-size tests, provided that these have the same level of error probabilities.

We use a series of concrete examples of hypothesis tests for binomial, Poisson, and geometric distributions. We test the following hypotheses: \(\theta_0 = 0.05\) versus \(\theta_1 = 0.08\) for the binomial distribution corresponding to \(n = 3\) Bernoulli trials with success probability \(\theta\), \(\theta_0 = 0.5\) versus \(\theta_1 = 0.7\) for the Poisson distribution with mean \(\theta\), and \(\theta_0 = 1\) versus \(\theta_1 = 2\) for the geometric distribution with mean \(\theta\). For each pair, we employ a range of error probabilities widely used in practice: \(\alpha(= \beta) = 0.1, 0.05, 0.025, 0.01, 0.005, 0.001, \) and \(0.0005\).

In addition, to see the effect of asymmetric error probabilities, we made the same evaluations for \(\alpha = 0.1\) and \(\beta = 0.0005\) and for \(\alpha = 0.0005\) and \(\beta = 0.1\) for each pair of the hypotheses.
Table 1. Poisson distribution with mean $\theta_0 = 0.5$ versus $\theta_1 = 0.7$.

| $\alpha = \beta$ | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|------------------|-----|------|-------|------|-------|-------|--------|
| $\lambda_0$      | 305.94 | 691.65 | 1,495.23 | 3,993.13 | 8,275.87 | 43,707.77 | 88,888.07 |
| $\lambda_1$      | 326.39 | 737.05 | 1,596.43 | 4,262.75 | 8,815.25 | 46,564.03 | 94,690.75 |
| $\theta$         | 0.58464 | 0.58794 | 0.58953 | 0.59072 | 0.59130 | 0.59206 | 0.59225 |
| $H$              | 353 | 442 | 533 | 649 | 745 | 935 | 1,024 |
| $N(\theta; \psi^*)$ | 67.93 | 114.80 | 166.46 | 239.77 | 297.96 | 439.60 | 502.73 |
| $N(\theta_0; \psi^*)$ | 57.06 | 87.92 | 117.97 | 156.92 | 185.92 | 252.90 | 281.79 |
| $N(\theta_1; \psi^*)$ | 51.66 | 79.55 | 106.80 | 141.98 | 168.15 | 228.59 | 254.63 |
| $Q_{\alpha}(\psi^*)$ | 165 | 247 | 331 | 440 | 522 | 712 | 794 |
| $\Delta$         | 1E-06 | 2E-06 | 2E-05 | 9E-06 | 3E-06 | 8E-06 | 4E-05 |
| log(A)           | −0.916 | −1.240 | −1.553 | −1.957 | −2.260 | −2.961 | −3.262 |
| log(B)           | 0.868 | 1.191 | 1.504 | 1.908 | 2.212 | 2.912 | 3.214 |
| $N(\theta; W)$   | 72.28 | 129.16 | 199.74 | 313.98 | 416.52 | 708.62 | 857.74 |
| $N(\theta_0; W)$ | 55.55 | 83.53 | 109.64 | 141.86 | 165.01 | 217.05 | 239.08 |
| $N(\theta_1; W)$ | 50.06 | 75.11 | 98.48 | 127.26 | 148.04 | 194.56 | 214.22 |
| $Q_{\alpha}(W)$  | 281 | 504 | 780 | 1,229 | 1,632 | 2,779 | 3,364 |
| FSS              | 98.07 | 161.05 | 229.01 | 322.43 | 395.11 | 568.75 | 644.84 |
| $R(\psi^*)$      | 1.44 | 1.40 | 1.38 | 1.34 | 1.33 | 1.29 | 1.28 |
| $R_0(\psi^*)$    | 1.72 | 1.83 | 1.94 | 2.05 | 2.13 | 2.25 | 2.29 |
| $R_1(\psi^*)$    | 1.90 | 2.02 | 2.14 | 2.27 | 2.35 | 2.49 | 2.53 |
| $QR(\psi^*)$     | 0.59 | 0.65 | 0.69 | 0.73 | 0.76 | 0.80 | 0.81 |
| $R(W)$           | 1.36 | 1.25 | 1.15 | 1.03 | 0.95 | 0.80 | 0.75 |
| $R_0(W)$         | 1.77 | 1.93 | 2.09 | 2.27 | 2.39 | 2.62 | 2.70 |
| $R_1(W)$         | 1.96 | 2.14 | 2.33 | 2.53 | 2.67 | 2.92 | 3.01 |
| $QR(W)$          | 0.35 | 0.32 | 0.29 | 0.26 | 0.24 | 0.20 | 0.19 |

For each combination of $\theta_0, \theta_1$ and $\alpha$ and $\beta$ we ran the computer code corresponding to the implementation of the method of section 2 in Novikov, Novikov, and Farkhshatov (2022; using option 1 with the bound $H$ defined by the right-hand side of equation 3.11). For the solution of the corresponding modified Kiefer-Weiss problem, we use the algorithms of Subsections 3.3 and 3.4.

To comply with the restrictions on the error probabilities, we seek the Lagrange multipliers $\lambda_0$ and $\lambda_1$ that provide the best approximation to the nominal values of $\alpha$ and $\beta$. We use the numerical minimization over $\lambda_0$ and $\lambda_1$ searching for a minimum value of

$$\max\{|\alpha(\psi^*, \phi^*) - \alpha|/\alpha, |\beta(\psi^*, \phi^*) - \beta|/\beta\}.$$  (3.18)

We employed the fminsearch function of the neldermead R package for the gradient-free Nelder-Mead method (see Bihorel 2022). A reasonably good formula to start the numerical minimization of (3.18) is $\lambda_0 = \kappa_0/\alpha$, $\lambda_1 = \kappa_1/\beta$, where $\kappa_0$ and $\kappa_1$ are empirical coefficients in the range of 10 to 60 depending on the model.

Due to the discrete nature of the probabilities involved in the evaluation, there is no guarantee, generally speaking, that the minimum of (3.18) can be 0. Nevertheless, in all of the evaluated cases the real and the nominal error probability are approximately within 0.002 of relative distance (3.18) to each other for the range of $\alpha$ and $\beta$ evaluated.

The respective numerical results are presented in Tables 1 to 3. For each test, the table contains the corresponding values of $\theta, \lambda_0, \lambda_1$; its average sample number $N(\theta; \psi^*)$; and its corresponding $\Delta$ (see equation 2.10), as well as the average sample number under the two hypotheses $N(\theta_0; \psi^*)$ and $N(\theta_1; \psi^*)$ and the 0.99th quantile $Q_{0.99}(\psi^*)$ of the distribution of the sample number under $\theta$. We also present the maximum sample number (denoted $H$ in the table) the test actually takes.

In the second part of each table are the calculated characteristics of the corresponding SPRT with the closest values of $\alpha$ and $\beta$ to the nominal ones. The same numerical
minimization procedure as above has been employed, this time over the continuation bounds $A$ and $B$ of the SPRT. The values of the average sample number of Wald’s SPRT $N(\theta; W)$, along with the average sample number under both hypotheses, $N(\theta_0; W)$ and $N(\theta_1; W)$, and the 0.99th quantile $Q_{0.99}(W)$ of the distribution of the

### Table 2. Geometric distribution with mean $\theta_0 = 1$ versus $\theta_1 = 2$.

| $x = \beta$ | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|----------|-----|------|-------|------|-------|-------|--------|
| $\lambda_0$ | 69.00 | 154.09 | 333.06 | 893.28 | 1,848.87 | 9,762.77 | 19,882.77 |
| $\lambda_1$ | 84.38 | 189.88 | 408.32 | 1,092.24 | 2,258.64 | 11,940.38 | 24,318.97 |
| $\theta$ | 1.27794 | 1.31841 | 1.33953 | 1.35556 | 1.36336 | 1.37428 | 1.37741 |
| $H$ | 74 | 96 | 118 | 145 | 167 | 213 | 233 |
| $N(\theta; \psi^*)$ | 15.43 | 23.60 | 31.62 | 41.92 | 49.66 | 67.51 | 75.212 |
| $Q_{0.99}(\psi^*)$ | 11.63 | 17.53 | 23.35 | 30.75 | 36.32 | 49.11 | 54.57 |
| $\Delta$ | 40 | 61 | 80 | 106 | 126 | 172 | 192 |
| $\log A$ | -0.8920 | -1.2098 | -1.5170 | -1.9213 | -2.2260 | -2.9269 | -3.2278 |
| $\log B$ | 0.7318 | 1.0452 | 1.3615 | 1.7649 | 2.0668 | 2.7692 | 3.0698 |
| $N(\theta_0; W)$ | 18.24 | 31.59 | 48.32 | 75.44 | 99.78 | 168.99 | 204.24 |
| $N(\theta_1; W)$ | 15.30 | 22.62 | 29.64 | 38.30 | 45.43 | 58.60 | 64.53 |
| $N(\theta_1; W)$ | 11.42 | 16.61 | 21.48 | 27.51 | 31.87 | 41.60 | 45.71 |
| $Q_{0.99}(W)$ | 67 | 119 | 184 | 290 | 384 | 654 | 791 |
| FSS | 23.83 | 39.00 | 55.29 | 77.74 | 95.21 | 136.89 | 155.21 |

### Table 3. Binomial distribution with mean $\theta_0 = 0.05$ versus $\theta_1 = 0.08$ and $n = 3$.

| $x = \beta$ | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|----------|-----|------|-------|------|-------|-------|--------|
| $\lambda_0$ | 450.00 | 1,020.19 | 2,201.68 | 5,880.48 | 12,177.81 | 64,446.61 | 131,326.7 |
| $\lambda_1$ | 489.75 | 1,110.18 | 2,393.14 | 6,398.75 | 13,250.02 | 70,018.20 | 142,274.9 |
| $\theta$ | 0.06193 | 0.06263 | 0.06296 | 0.06316 | 0.06328 | 0.06345 | 0.06350 |
| $H$ | 484 | 630 | 771 | 943 | 1,068 | 1,355 | 1,480 |
| $N(\theta_0; \psi^*)$ | 101.13 | 171.00 | 247.90 | 357.06 | 443.72 | 654.74 | 748.81 |
| $N(\theta_0; \psi^*)$ | 86.28 | 133.00 | 178.58 | 237.31 | 281.04 | 382.50 | 426.12 |
| $N(\theta_1; \psi^*)$ | 75.46 | 116.41 | 156.41 | 208.08 | 246.49 | 335.21 | 373.45 |
| $Q_{0.99}(\psi^*)$ | 247 | 369 | 493 | 654 | 778 | 1,060 | 1,184 |
| $\Delta$ | -3E-07 | -1E-06 | 2E-04 | 2E-04 | 4E-04 | 7E-08 | 5E-05 |
| $\log(A)$ | -2.1517 | -2.8990 | -3.6164 | -4.5949 | -5.2469 | -6.8607 | -7.5540 |
| $\log(B)$ | 2.0034 | 2.7527 | 3.4713 | 4.4033 | 5.1012 | 6.7147 | 7.4086 |
| $N(\theta_0; W)$ | 107.24 | 192.16 | 296.92 | 466.76 | 618.96 | 1,053.14 | 1,275.06 |
| $N(\theta_0; W)$ | 83.91 | 126.52 | 166.12 | 215.01 | 250.13 | 329.02 | 362.47 |
| $N(\theta_0; W)$ | 73.07 | 109.94 | 144.16 | 186.36 | 216.82 | 285.00 | 313.86 |
| $Q_{0.99}(W)$ | 414 | 748 | 1,159 | 1,826 | 2,423 | 4,124 | 4,997 |
| FSS | 146.62 | 240.94 | 341.25 | 480.98 | 589.32 | 847.40 | 961.23 |
| $R(\psi^*)$ | 1.45 | 1.41 | 1.38 | 1.35 | 1.33 | 1.29 | 1.28 |
| $R_0(W)$ | 1.70 | 1.81 | 1.91 | 2.03 | 2.10 | 2.22 | 2.26 |
| $R_1(W)$ | 1.94 | 2.07 | 2.18 | 2.31 | 2.39 | 2.53 | 2.57 |
| $QR(W)$ | 0.59 | 0.65 | 0.69 | 0.74 | 0.76 | 0.80 | 0.81 |
| $R(W)$ | 1.37 | 1.25 | 1.15 | 1.03 | 0.95 | 0.80 | 0.75 |
| $R_0(W)$ | 1.75 | 1.90 | 2.05 | 2.24 | 2.36 | 2.58 | 2.65 |
| $R_1(W)$ | 2.01 | 2.19 | 2.37 | 2.58 | 2.72 | 2.97 | 3.06 |
| $QR(W)$ | 0.35 | 0.32 | 0.29 | 0.26 | 0.24 | 0.21 | 0.19 |
sample number, calculated at \( \theta \), are provided in the tables. All of the characteristics are calculated using the exact formulas in Subsection 3.4, with appropriately large horizon \( H \). \( \log (A) \) and \( \log (B) \) are the endpoints of the continuation interval of the corresponding SPRT.

Finally, FSS is the minimum value of the sample number required by the optimal fixed-sample-size test with error probabilities \( \alpha \) and \( \beta \). For given \( \alpha \) and \( \beta \), FSS is calculated as

\[
\frac{n^* + \frac{\beta(n^*)}{\beta(n^*) - 1}}{\beta(n^*) - 1},
\]

where \( n^* = n^*(\alpha, \beta) \) is a maximum integer \( n \) for which the type II error probability \( \beta(n) \) of the most powerful level \( \alpha \) (Neyman-Pearson) test is greater than or equal to \( \beta \).

In the last part of each table are calculated values of efficiency of each test with respect to the FSS tests. The efficiency is calculated as the ratio of FSS to other characteristics of the respective test: \( R(\psi^*) = \frac{\text{FSS}/N(\theta; \psi^*)}{\text{FSS}/N(0_1; \psi^*)} \), \( R_0(\psi^*) = \frac{\text{FSS}/N(\theta_0; \psi^*)}{\text{FSS}/N(0_1; \psi^*)} \), \( R_1(\psi^*) = \frac{\text{FSS}/N(0_1; \psi^*)}{\text{FSS}/N(\theta_1; \psi^*)} \), and \( QR(\psi^*) = \frac{\text{FSS}/Q_{99}(\psi^*)}{\text{FSS}/Q_{99}(0_1)} \) for the optimal Kiefer-Weiss test and \( R(W) = \frac{\text{FSS}/N(\theta; W)}{\text{FSS}/N(\theta_0; W)} \), \( R_0(W) = \frac{\text{FSS}/N(\theta_0; W)}{\text{FSS}/N(\theta_1; W)} \), \( R_1(W) = \frac{\text{FSS}/N(\theta_1; W)}{\text{FSS}/N(\theta_1; W)} \), and \( QR(W) = \frac{\text{FSS}/Q_{99}(W)}{\text{FSS}/Q_{99}(W)} \) for Wald’s SPRT. For example, \( R_0(\psi^*) = 2 \) means that the optimal Kiefer-Weiss test takes two times fewer observations, on average, under \( H_0 \) compared to the corresponding fixed-sample-size test.

### Table 4. Comparison of tests for asymmetric error probabilities.

| Model | Binomial | Poisson | Geometric | Binomial | Poisson | Geometric |
|-------|----------|---------|-----------|----------|---------|-----------|
| \((\theta_0, \theta_1)\) | (0.05, 0.08) | (0.5, 0.7) | (1.0, 2.0) | (0.05, 0.08) | (0.5, 0.7) | (1.0, 2.0) |
| \(\gamma_0\) | 948.57 | 640.92 | 150.43 | 80,909.41 | 54,753.29 | 11,788.63 |
| \(\gamma_1\) | 91,786.79 | 60,360.17 | 15,823.13 | 978.27 | 656.25 | 161.15 |
| \(\theta\) | 0.05551 | 0.53918 | 1.12992 | 0.07145 | 0.64532 | 1.62774 |
| \(H\) | 1,172 | 806 | 200 | 1,058 | 741 | 151 |
| \(N(\theta; \psi^*)\) | 350.27 | 232.95 | 60.47 | 324.14 | 219.70 | 50.07 |
| \(N(0_1; \psi^*)\) | 320.30 | 211.46 | 56.82 | 135.48 | 89.49 | 24.16 |
| \(N(\theta_1; \psi^*)\) | 116.76 | 79.88 | 17.36 | 279.23 | 190.56 | 40.83 |
| \(Q_{99}(\psi^*)\) | 695 | 464 | 118 | 655 | 443 | 102 |
| \(\Delta\) | –8E-05 | 6E-06 | –5E-08 | 6E-05 | 3E-05 | –4E-09 |
| \(\log(A)\) | –7.4481 | –3.2167 | –3.1829 | –2.2541a | –0.9620 | –0.9276 |
| \(\log(B)\) | 2.1115 | 0.9132 | 0.7673 | 7.1464a | 3.1680 | 3.0242 |
| \(N(\theta; W)\) | 371.78 | 263.20 | 65.89 | 373.89a | 258.21 | 61.57 |
| \(N(\theta_0; W)\) | 311.04 | 205.19 | 55.50 | 190.62a | 72.37 | 19.56 |
| \(N(\theta_1; W)\) | 95.35 | 65.17 | 14.54 | 263.52a | 183.89 | 39.27 |
| \(Q_{99}(W)\) | 1,293 | 879 | 200 | 1,372a | 937 | 241 |
| \(FSS\) | 478.61 | 319.13 | 81.10 | 449.66 | 303.78 | 69.14 |
| \(R(\psi^*)\) | 1.37 | 1.37 | 1.34 | 1.39 | 1.38 | 1.38 |
| \(R_0(\psi^*)\) | 1.49 | 1.51 | 1.43 | 3.32 | 3.39 | 2.86 |
| \(R_1(\psi^*)\) | 4.10 | 4.00 | 4.67 | 1.61 | 1.59 | 1.69 |
| \(QR(\psi^*)\) | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 |
| \(R(W)\) | 1.22 | 1.21 | 1.23 | 1.20a | 1.18 | 1.12 |
| \(R_0(W)\) | 1.54 | 1.56 | 1.46 | 4.10a | 4.20 | 3.53 |
| \(R_1(W)\) | 3.02 | 4.90 | 5.38 | 1.71a | 1.65 | 1.76 |
| \(QR(W)\) | 0.37 | 0.36 | 0.41 | 0.38b | 0.32 | 0.29 |

This column corresponds to the actual value of \( \alpha = 0.000585 \); no better approximation for \( \alpha = 0.0005 \) could be achieved.
4. DISCUSSION AND CONCLUSIONS

4.1. Analysis of Numerical Results

Tables 1 to 4 are convenient for analyzing the relative efficiency of the optimal tests with respect to the fixed-sample-size test.

It is clearly seen that all three efficiencies \( R(\psi^*) \), \( R_0(\psi^*) \), \( R_1(\psi^*) \) do not vary much in the whole range of \( \alpha \) and \( \beta \) computed. The lowest value of \( R(\psi^*) \) is slightly below 1.3 and is attained at the minimum values of \( \alpha = \beta \) considered. There is a clear tendency of the relative efficiency decreasing with \( \alpha = \beta \to 0 \). As a reference, one can bear in mind that the relative efficiency \( R(\psi^*) \), by definition, cannot drop below 1.

Both \( R_0(\psi^*) \) and \( R_1(\psi^*) \) slightly vary between approximately 1.8 and 2.8 for \( \alpha = \beta \), with their maximum values achieved for small \( \alpha = \beta \).

It is clearly seen from Table 4 that \( R(\psi^*) \) maintains its level at approximately 1.4, even for very asymmetric \( \alpha \) and \( \beta \).

The relative efficiency under \( H_1 \), \( R_1 \), tends to have higher values (up to approximately 4.0 to 4.5) for the very asymmetric case of \( \alpha = 0.1 \) and \( \beta = 0.0005 \), and \( R_0(\psi^*) \) stays at approximately 1.5, which is slightly lower than in the case of equal \( \alpha \) and \( \beta \). The behavior of \( R_0(\psi^*) \) and \( R_1(\psi^*) \) for the other pair of extreme values (when \( \alpha = 0.0005 \) and \( \beta = 0.1 \)) is nearly reversed.

The relative efficiency \( QR(\psi^*) \) based on the 0.99th quantile of the optimal Kiefer-Weiss test behaves quite well maintaining the approximate level of 0.6 to 0.8 for all levels of \( \alpha \), \( \beta \) computed.

The relative efficiency of the SPRT based on the average sample number evaluated at \( \theta \) drops to approximately 0.7–0.8 for lower levels of \( \alpha \) and \( \beta \), which is still comparable to the efficiency of approximately 1.3 of the optimal Kiefer-Weiss test. But SPRT shows a very low efficiency \( QR(W) \) based on the 0.99th quantile of the sample number distribution under \( \theta \), which is as low as approximately 0.3–0.2, meaning that the 0.99th quantile can reach a three to five times higher level than the fixed sample size.

In general, it seems remarkable that the pattern of the relative efficiency is largely the same between the three models we consider in this article and between these models and the Bernoulli model in Novikov, Novikov, and Farkhshatov (2022). Comparing the relative efficiency numbers of \( R(\psi^*) \) for the respective \( \alpha = \beta \) level between the four models, we observe almost identical results. \( R_0(\psi^*) \) and \( R_1(\psi^*) \) show very similar behavior.

The same happens with the efficiencies calculated for the SPRT (comparing \( R(W) \) or \( R_0(W) \) or \( R_1(W) \) between the models for the same level of \( \alpha = \beta \)). Even quantile-based efficiencies \( QR(\psi^*) \) are practically identical between the models, and the same is valid for \( QR(W) \).

Approximately the same happens for both cases of asymmetric error probabilities (see Table 4): all of the relative efficiency numbers maintain largely the same level between the models.

4.2. Further Work

The most immediate work to be done is to develop the computer algorithms for solving the Kiefer-Weiss problem for sampling from continuous exponential families of
distributions. For a normal distribution and \( \alpha = \beta \) there are various known numerical results in the literature (see, for example, Lorden 1976).

Another important aspect is the efficiency of Lorden’s 2-SPRT with respect to the optimal Kiefer-Weiss test. The numerical results of Lorden (1976) show excellent performance of the 2-SPRT in the symmetric normal case. Our method should permit not only investigating the nonsymmetric normal case but also the performance of the 2-SPRT for other exponential families of distributions. In particular, assessing the nonasymptotic rate of approximation for the asymptotically optimal 2-SPRT constructed by Huffman (1983) could be of interest for applications (see Mulekar, Young, and Young 1992, for example).

**DISCLOSURE**

The authors have no conflicts of interest to report.

**Funding**

A. Novikov gratefully acknowledges partial support of SNI by CONACyT (Mexico) for this work. F. Farkhatov thanks CONACyT (Mexico) for scholarship for his doctoral studies.

**ORCID**

Andrey Novikov [http://orcid.org/0000-0003-1034-8951](http://orcid.org/0000-0003-1034-8951)

**REFERENCES**

Bihorel, S. 2022. “neldermead: R Port of the 'Scilab' Neldermead Module.” R package version 1.0-12. [https://CRAN.R-project.org/package=neldermead](https://CRAN.R-project.org/package=neldermead).

Freeman, D., and L. Weiss. 1964. “Sampling Plans Which Approximately Minimize the Maximum Expected Sample Size.” *Journal of the American Statistical Association* 59 (305): 67–88. doi:10.1080/01621459.1964.10480701

Hawix, A., and N. Schmitz. 1998. “Remark on the Modified Kiefer-Weiss Problem for Exponential Families.” *Sequential Analysis* 17 (3-4):297–303. doi:10.1080/07474949808836415

Huffman, M. D. 1983. “An Efficient Approximate Solution to the Kiefer-Weiss Problem.” *The Annals of Statistics* 11 (1):306–16. doi:10.1214/aos/1176346081

Kiefer, J., and L. Weiss. 1957. “Some Properties of Generalized Sequential Probability Ratio Tests.” *The Annals of Mathematical Statistics* 28 (1):57–75. doi:10.1214/aoms/1177707037

Lai, T. L. 1973. “Optimal Stopping and Sequential Tests Which Minimize the Maximum Expected Sample Size.” *The Annals of Statistics* 1 (4):659–73. doi:10.1214/aos/1176342461

Lorden, G. 1976. “2-SPRT’S and The Modified Kiefer-Weiss Problem of Minimizing an Expected Sample Size.” *The Annals of Statistics* 4 (2):281–91. doi:10.1214/aos/1176343407

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

Mulekar, M. S., L. J. Young, and J. H. Young. 1992. “Using an Approximate Kiefer-Weiss Solution for Testing Insect Population Densities.” *Metrika* 39 (1):219–26. doi:10.1007/BF02614005

Novikov, A. 2009. “Optimal Sequential Tests for Two Simple Hypothesis.” *Sequential Analysis* 28 (2):188–217. doi:10.1080/07474940902816809
Novikov, A., A. Novikov, and F. Farkhshatov. 2022. “A Computational Approach to the Kiefer-Weiss Problem for Sampling from a Bernoulli Population.” Sequential Analysis 41 (2):198–219. doi:10.1080/07474946.2022.2070212

Novikov, A., and X. I. Popoca Jiménez. 2022. “Optimal Group-Sequential Tests with Groups of Random Size.” Sequential Analysis 41 (2):220–40. doi:10.1080/07474946.2022.2070213

R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Tartakovsky, A. G., I. V. Nikiforov, and M. Basseville. 2014. Sequential Analysis: Hypothesis Testing and Changepoint Detection. Boca Raton: Chapman & Hall/CRC Press.

Weiss, L. 1962. “On Sequential Tests Which Minimize the Maximum Expected Sample Size.” Journal of the American Statistical Association 57 (299):551–66. doi:10.1080/01621459.1962.10500543

Zhitlukhin, M. V., A. A. Muravlev, and A. N. Shiryaev. 2013. “The Optimal Decision Rule in the Kiefer-Weiss Problem for a Brownian Motion.” Russian Mathematical Surveys 68 (2):389–94. doi:10.1070/RM2013v068n02ABEH004834