Generalization of Sequential Wald’s Test for More Than Two Hypotheses

BORIS DARKHOVSKY
Institute for Systems Analysis RAS
Department of of Macrosystems Dynamics
Prospekt 60-letya Oktyabria 9, 117312 Moscow
RUSSIA
darbor2004@mail.ru

Abstract: The sequential testing problem for more than two simple hypotheses is considered. Generalization of classical Wald’s procedure is proposed. It is proved that the procedure minimizes the greatest average run length in the assumption of validity of any hypothesis among all sequential decision rules with given constraints on the greatest error probabilities. The received results pass to the classical ones for the case of two simple hypotheses sequential testing problem.

Key–Words: Sequential hypotheses testing, Wald’s procedure

1 Introduction

Since Wald’s seminal ideas in sequential testing of a simple hypothesis $H_0 : P = P_0$ against a simple alternative $H_1 : P = P_1$ more than 60 years have passed (see [1]). The classic Wald’s sequential probability ratio test (SPRT) was shown by Wald and Wolfowitz ([2]) to be optimal in the following sense: among tests which sample size $T$ has a finite expectation under both $H_0$ and $H_1$ and error probabilities satisfy

$$P_0\{\text{Reject } H_0\} \leq \alpha \quad (1)$$

and

$$P_1\{\text{Reject } H_1\} \leq \beta, \quad (2)$$

the SPRT minimizes both $E_0T$ and $E_1T$.

In 1950-1990s Wald’s ideas were generalized in different directions. In [3] the lower bounds for average sample size in multihypothesis testing was given. In [4, 5, 6, 7] the class of tests motivated by the Bayesian framework was considered.

The second class of tests represents a modification of the matrix SPRT (the combination of one-sided SPRT’s). It was considered in [8, 9, 10].

In [11, 12, 13, 14] the most advanced problems of sequential testing of many simple hypotheses for the case of independent observations was considered. Different ramifications for dependent observations and many composite hypotheses are also considered.

In [11] the problem of testing many composite hypotheses with indifference zone for an unknown parameter was considered. Two sequential tests were proposed with given constraints on error probabilities. These tests generalized earlier results of [15] and [6].

Although there is a lot of publications in the field, the following question is still open (as far as the author knows): is it possible to generalize the non-asymptotic Wald’s result for more than two hypotheses?

In [16] a new way (the minimax-Bayesian statement) to formalize the problem of sequential testing for two composite hypotheses was proposed. As a result, the non-asymptotic generalization of Wald’s test was given. In [17] more checkable assumptions for this case were discussed. Based on [16, 17], in this paper we get the non-asymptotic generalization of the classical results for the case of more than two simple hypotheses.

This paper is organized as follows. In Section 2 we give the foundation of the main result. Namely, we describe the minimax-Bayesian statement of sequential testing for two composite hypotheses. The optimal strategy and the optimal sequential test are given. The test generalizes the classical result without utilization of the asymptotic. In Section 3 we consider the problem of sequential testing for three simple hypotheses. Using the results of Section 2, we propose a non-asymptotic optimal sequential test generalizing the classical result. In Section 4 we give a sketch description for more than three simple hypotheses case.
2 Foundation: Minimax-Bayesian Statement of Sequential Testing for Two Composite Hypotheses

2.1 Problem Statement

Let \( \Theta = \{ \theta = (\theta_0, \theta_1) \} \) be a parametric set in a finite-dimensional space, \( \Theta = \Theta_0 \cup \Theta_1, \Theta_0 \cap \Theta_1 = \emptyset. \) Consider two family of probabilistic distributions \( \mathcal{P}_1 = \{ f_i(x, \theta_i) \} \) \( \theta_i \in \Theta_i, i = 0, 1 \), where \( f_i(x, \theta_i) \) is a density function (d.f.) w.r.t. some \( \sigma \)-finite measure \( \mu. \) The problem of sequential testing of two composite hypotheses on the base of independent and identically distributed (with d.f. \( f_i(x, \theta_i) \)) risk:

\[
\mathcal{P}_1 = \{ f_i(x, \theta_i) \} \}
\]

where

\[
\mathcal{P}_1 \]

is a family of probabilistic distributions on \( \Theta \) such that \( f_{\theta_0}(x, \theta_0) \) and a terminal decision \( d(\theta) = d(\theta) \) (d = 1 means that \( H_1 \) is true, \( i = 0, 1 \)).

Consider “differential” error probabilities for the procedure \( \delta: P(\theta_1, 0) = \alpha(\theta_0, \theta_1), P(\theta_1, 0) = \beta(\theta_0, \theta_1) \) and “differential” average rung before the decision making

\[
E_{\theta_0}(T(\theta_0, \theta_1)) = T_0(\theta_0, \theta_1), E_{\theta_1}(T(\theta_0, \theta_1)) = T_1(\theta_0, \theta_1). \]

Then Bayesian risk of procedure \( \delta \) under prior distribution \( F(\cdot) \) of the parameter is equal to

\[
R(\delta, F) = \int_\Theta [w_0 \alpha(\delta, \theta) + cT_0(\delta, \theta)] dF(\theta) + \int_{\Theta_1} [w_1 \alpha(\delta, \theta) + cT_1(\delta, \theta)] dF(\theta),
\]

where \( w_0 > 0, w_1 > 0 \) are losses from the false decision and \( c > 0 \) is a cost of one observation.

We say that the procedure is a minimax-Bayesian if it minimizes the greatest (over a given class of prior distributions) risk:

\[
\mathcal{R}(\delta, q) = \sup_{F \in F_q} R(\delta, F).
\]

In particular, we get the definition of Bayesian criterion for classical Wald’s problem in case each parametric set consists of one point.

The following assumptions will be used:

A1. The d.f. are defined on some open set containing \( \Theta \) and \( \mu(x : f_0(x, \theta_0) \neq f_1(x, \theta_1)) > 0 \) for all \( \theta \) \( \in \Theta \).

A2. For \( \mu-a.s. \) \( x \) the d.f. \( f_i(x, \cdot) \) is continuous and positive and \( \Theta \) is a compact.

Denote by \( \rho(a, \theta_0) = \int f_i(x, \theta_0)d\mu(x), a, b = 0, 1 \) the Kullback-Leibler distance between the parametric points \( \theta_a, \theta_b. \)

A3. \( \min_{a, b} \rho(\theta_i^*, \theta_j) > \max_{a, b} \rho(\theta_i^*, \theta_j) \) for any \( \theta_i^* \) \( \in \Theta_i, (i, j) = 0, 1, i \neq j \).

A4. Uniform Cramer condition: for any \( \theta_0 \) \( \in \Theta_0, \theta_1 \in \Theta_1 \) and \( \eta(\omega, \theta) = \ln \frac{f_0(\omega, \theta)}{f_1(\omega, \theta)} \) \( \zeta(\omega, \theta) = \ln \frac{f_0(\omega, \theta)}{f_1(\omega, \theta)} \)

\[
\sup_{\theta \in \Theta} E_{\theta_0} \exp(\eta(\theta)) < \infty, \sup_{\theta \in \Theta} E_{\theta_0} \exp(\zeta(\theta)) < \infty \text{ for } |t| < H, H > 0.
\]

A5. For any \( \theta_j \) \( \in \Theta_j \) the functions \( \kappa_j(t, \theta_j^*) = \int f_i(\omega, \theta_j^*) \mu(d\omega) \) has only two zeros: 0 and \( t_j^*(\theta_j^*) > 0 \), and the function \( t_j^*(\cdot, \theta_j^*) \) is continuous for any \( \theta_j^* \) \( \in \Theta_j \) and \( \min_j t_j^*(\theta_j^*) > 0, (i, j) = 0, 1, i \neq j \).

2.2 Minimax-Bayesian procedure

Denote \( L_i(X(n), \theta_i) = \prod_{k=1}^n f_i(x_k, \theta_i), i = 0, 1 \). Let \( q^*(n), q_*(n) \) be, correspondingly, an upper and lower posterior probability that hypothesis \( H_0 \) is true after \( n \) observations. Then

\[
q^*(n) = \frac{qH^*(n)}{1-q+qH^*(n)}, q_*(n) = \frac{qH^*(n)}{1-q+qH^*(n)}, q_*(0) = q, \]

where \( H^*(n) = \sup_{\theta_0 \in \Theta_0} L_0(X(n), \theta_0) \), \( H_*(n) = \inf_{\theta_1 \in \Theta_1} L_1(X(n), \theta_1) \), \( n = 1, 2, \ldots \).

Theorem 1 Let assumptions A1-A2 hold. Then there exist numbers \( 0 < B < A \) such that the optimal (minimax-Bayesian) strategy has the form:

i) if \( H_*(n) \geq A \), then hypothesis \( H_0 \) must be accepted;

ii) if \( H^*(n) \leq B \), then hypothesis \( H_1 \) must be accepted;

iii) otherwise the observations must be continued.

Note that in the case of simple hypotheses \( \Theta_0 = \{ \theta_0 \}, \Theta_1 = \{ \theta_1 \} \) the classical Wald’s rule follows from here because \( H_*(n) = H^*(n) = H(n), n = 1, 2, \ldots \) where \( H(n) \) is the usual likelihood ratio \( \prod_{k=1}^n f(x_k, \theta_0)/\prod_{k=1}^n f(x_k, \theta_1). \)
2.3 Sequential criterion

Let \(0 < C < 1 < D\) are given numbers. Consider the criterion described above with boundaries \(C\) (instead of \(B\)) and \(D\) (instead of \(A\)). Denote this criterion by \(\delta_{C,D}\). Consider the following class of sequential procedures \(K_{a,b} = \{\delta : \sup_{\theta_0 \in \Theta_0} \alpha_0(\delta, \theta_0) \leq a, \sup_{\theta_1 \in \Theta_1} \alpha_1(\delta, \theta_1) \leq b\} \).

**Theorem 2** Let assumptions A1-A5 hold. If numbers \(C^{-1}, D\) are sufficiently large, then procedure \(\delta_{C,D}\) is optimal one in the following sense:

for any procedure \(\delta \in K_{a,b}\) such that
\[
\sup_{\theta_0 \in \Theta_0} T_0(\delta, \theta_0) < \infty, \sup_{\theta_1 \in \Theta_1} T_1(\delta, \theta_1) < \infty,
\]
the inequalities hold
\[
\sup_{\theta_0 \in \Theta_0} T_0(\delta, \theta_0) \leq \sup_{\theta_0 \in \Theta_0} T_0(\delta, \theta_0),
\sup_{\theta_1 \in \Theta_1} T_1(\delta, \theta_1) \leq \sup_{\theta_1 \in \Theta_1} T_1(\delta, \theta_1)
\]
where \(a = \sup_{\theta_0 \in \Theta_0} \alpha_0(\delta, \theta_0), b = \sup_{\theta_1 \in \Theta_1} \alpha_1(\delta, \theta_1)\).

Besides, for any \(\theta_0^* \in \Theta_0, \theta_1^* \in \Theta_1\)
\[
\lim_{D \to \infty} \left[ \frac{\ln D}{\ln a(\theta_0^*, \theta_1^*)} \right] = \min_{\theta \in \Theta} \left( T_0(\delta, \theta_0^*), T_1(\delta, \theta_1^*) \right),
\]
\[
\lim_{C \to 0} \left[ \frac{\ln C}{\ln a(\theta_0^*, \theta_1^*)} \right] = \min_{\theta \in \Theta} \left( T_0(\delta, \theta_0^*), T_1(\delta, \theta_1^*) \right),
\]
\[
\lim_{D \to \infty} \left[ \frac{\ln C}{\ln a(\theta_0^*, \theta_1^*)} \right] = \min_{\theta \in \Theta} \left( T_0(\delta, \theta_0^*), T_1(\delta, \theta_1^*) \right),
\]
Further, if \(D = (1 - \alpha)/\beta, C = \alpha/(1 - \beta), \alpha > 0, \beta > 0, \alpha + \beta < 1\) then \(P_1(\text{accept } H_0) \leq D^{-1}, P_0(\text{accept } H_1) \leq C\).

3 Generalization of Wald’s procedure for three simple hypotheses

3.1 Problem statement and assumptions

Let \(\{f_i(x)\}, i \in J = \{1, 2, 3\}\) are d.f. w.r.t. some \(\sigma\)-finite measure \(\mu\).

The problem of sequential testing for three simple hypotheses on the base of independent and identically distributed (with d.f. \(f(x)\)) observations \(X^n = (x_1, x_2, \ldots, x_n)\) is considered: \(\{H_i : f(x) = f_i(x)\}, i \in J\).

Let
\[
\rho(f_i, f_j) = \int \ln \frac{f_i(x)}{f_j(x)} f_i(x) \mu(dx), \quad i \neq j, \quad (i, j) \in J
\]
are corresponding Kullback-Leibler distances (we assume that \(0 < \rho(f_i, f_j) < \infty, i \neq j\)).

We assume that at least one of the following conditions hold:

i) \(\rho(f_2, f_1) > \rho(f_2, f_3), \rho(f_3, f_1) > \rho(f_3, f_2)\);

ii) \(\rho(f_1, f_2) > \rho(f_1, f_3), \rho(f_3, f_2) > \rho(f_3, f_1)\);

iii) \(\rho(f_2, f_3) > \rho(f_2, f_1), \rho(f_1, f_3) > \rho(f_1, f_2)\).

Without loss of generality we suppose below that condition i) holds. Note that condition i) corresponds to condition A3 if we consider two (composite) hypotheses: \(H_0 : \{f = f_1\}\) and \(H_1 : \{f = (f_2 \text{ or } f_3)\}\). This is the crucial point for using the result from section 3.

We use the following assumptions:

B1. Cramer condition: for any \((i, j, k) \in J\) and any \(|t| < H, \quad H > 0\)
\[
\int \left( \frac{f_i(x)}{f_j(x)} \right)^t f_k(x) \mu(dx) < \infty.
\]

B2. The functions
\[
\int f_i(x) f_j(x) f_k(x) \mu(dx), \int \left( \frac{f_i(x)}{f_j(x)} \right)^t f_k(x) \mu(dx)
\]
\[
\int \left( \frac{f_i(x)}{f_j(x)} \right)^t f_k(x) \mu(dx), \int \left( \frac{f_i(x)}{f_j(x)} \right)^t f_k(x) \mu(dx)
\]
\[
\int \left( \frac{f_i(x)}{f_j(x)} \right)^t f_k(x) \mu(dx)
\]
has only two zeros: 0 and the positive one.

3.2 Optimal stopping

Put
\[
U(n) = \min \left( \sum_{k=1}^{n} \ln \frac{f_k(x_k)}{f_2(x_k)}, \sum_{k=1}^{n} \ln \frac{f_2(x_k)}{f_3(x_k)} \right)
\]
\[
V(n) = \max \left( \sum_{k=1}^{n} \ln \frac{f_k(x_k)}{f_2(x_k)}, \sum_{k=1}^{n} \ln \frac{f_2(x_k)}{f_3(x_k)} \right)
\]
\[
G(n) = \sum_{k=1}^{n} \ln \frac{f_2(x_k)}{f_3(x_k)}
\]

Let \(0 < C < 1 < D\) are given numbers. Define the following stopping times:
\[
\tau_1 = \min \{n : U(n) \geq \ln D\},
\tau_2 = \min \{n : V(n) \leq \ln C\}
\]
\[
\tau = \min(\tau_1, \tau_2),
\tau_3 = \min \{n \geq \tau : G(n) \geq \ln D\},
\tau_4 = \min \{n \geq \tau : G(n) \leq \ln C\}
\]
\[
T^* = \tau_1 I(\tau = \tau_1) + \min(\tau_3, \tau_4) I(\tau = \tau_2),
\]
where \( I(A) \) is the indicator of a set \( A \).

Define the sequential procedure \( \delta^* = (T^*, d^*) \), where \( d^* = i \) means "accept d.f. \( f_i \), \( i = 1, 2, 3 \) and

\[
d^* = \begin{cases} 
1 & \text{if } T^* = \tau_1 \\
2 & \text{if } \tau = \tau_2, T^* = \tau_3 \\
3 & \text{if } \tau = \tau_2, T^* = \tau_4.
\end{cases}
\]

**Theorem 3** Let numbers \( C^{-1}, D \) are sufficiently large. Then the stopping time \( T^* \) is optimal one in a following sense:

for any sequential procedure \( \delta = (R, d) \) with the stopping time \( R \) such that \( \max_{i \in J} E_i R < \infty \) and

\[
P\{d = 2 \text{ or } 3 | f_1\} \leq \alpha, \quad P\{d = 3 | f_2\} \leq \alpha \quad \text{max} \left\{ P\{d = 1 | f_3\}, \ P\{d = 1 | f_2\} \right\} \leq \beta, \\
P\{d = 2 | f_3\} \leq \beta
\]

the inequality holds

\[
\max_{i \in J} E_i T^* \leq \max_{i \in J} E_i R.
\]

Here

\[
\alpha = P\{d^* = 2 \text{ or } 3 | f_1\} \\
\beta = \max\{P\{d^* = 1 | f_2\}, P\{d^* = 1 | f_3\}\}.
\]

Besides, if

\[
\hat{\alpha} > 0, \hat{\beta} > 0, \hat{\alpha} + \hat{\beta} < 1, \\
D = (1 - \hat{\alpha})/\hat{\beta}, \ C = \hat{\alpha}/(1 - \hat{\beta})
\]

then

\[
P\{d^* = 1 | f_2 \text{ or } f_3\} \leq D^{-1}, \ P\{d^* = 2 | f_3\} \leq D^{-1} \\
P\{d^* = 2 \text{ or } 3 | f_1\} \leq C, \ P\{d^* = 3 | f_2\} \leq C.
\]

Obviously, the classic result follows from here in the case of two hypotheses.

### 4 The case of more than three hypotheses

Let us consider the general case with more than three simple hypotheses. The optimal sequential procedure can be constructed step by step as follows.

Step 1. We divide the set of hypotheses on two subsets so that for these subsets one of the condition of type i) - iii) (see Section 3) is satisfied. Therefore, we get two composite hypotheses.

Step 2. Using the optimal procedure for two composite hypotheses (see Section 2), we get the first pair of stopping times (compare with \( \tau_1, \tau_2 \) from Section 3).

Step 3. Each of two subsets again divide on two parts so that for these parts one of the condition of type i) - iii) (see Section 3) is satisfied. Therefore, we again get two composite hypotheses.

Step 4. Using the optimal procedure for two composite hypotheses (see Section 2), we get the second pairs of stopping times for each subset (compare with \( \tau_3, \tau_4 \) from Section 3), and so on.

The optimal stopping time can be constructed from the collection \( \{\tau_i\} \) by logical operations.

### References:


