Asymptotic Optimality of Certain Multihypothesis Sequential Tests: Non-i.i.d. Case *

Alexander G. Tartakovsky
Center for Applied Mathematical Sciences
University of Southern California

Abstract. It is known that certain combinations of one-sided sequential probability ratio tests are asymptotically optimal (relative to the expected sample size) for problems involving a finite number of possible distributions when probabilities of errors tend to zero and observations are independent and identically distributed according to one of the underlying distributions. The objective of this paper is to show that two specific constructions of sequential tests asymptotically minimize not only the expected time of observation but also any positive moment of the stopping time distribution under fairly general conditions for a finite number of simple hypotheses. This result appears to be true for general statistical models which include correlated and non-homogeneous processes observed either in discrete or continuous time. For statistical problems with nuisance parameters, we consider invariant sequential tests and show that the same result is valid for this case. Finally, we apply general results to the solution of several particular problems such as a multi-sample slippage problem for correlated Gaussian processes and for statistical models with nuisance parameters.

Keywords: Multihypothesis sequential tests, one-sided SPRT, invariant sequential tests, asymptotic optimality, r-quick convergence, correlated and non-homogeneous processes, moments of a stopping time, multi-sample slippage problem.

Primary 62L10; secondary 62F05, 62M02:

1. Introduction

The goal of testing statistical hypotheses is to relate an observed stochastic process to one of N possible classes based on some knowledge of distributions of the observations under each class (hypothesis). In sequential setting, the number of observations is allowed to be random (a function of the observations). A sequential procedure (test) includes a stopping time and a terminal decision to achieve a tradeoff between the average observation time and quality of the decision.

The problem of sequential testing of many hypotheses is substantially more difficult than that of testing two hypotheses. For multiple-decision testing problems it is usually very difficult if even possible to

* This research was supported in part by the U.S. Army Research Office Grant DAAH04-95-1-0164

obtain optimal solutions. For this reason a substantial part of the development of sequential multihypothesis testing in the last three decades has been directed towards the study of suboptimal procedures, basically multihypothesis modifications of sequential probability ratio test (SPRT), when observations are independent and identically distributed (i.i.d.). See, e.g., Armitage (1950), Chernoff (1959), Kiefer and Sacks (1963), Lorden (1967, 1977), Pavlov (1984, 1990), Sosulin and Fishman (1985), Fishman (1987), Dragalin (1987), Dragalin and Novikov (1994). The generalization for the case of non-stationary processes with independent increments was made by Tartakovskii (1981), Golubev and Khas'minskii (1983), Verdenskaya and Tartakovskii (1991), Tartakovsky (1991, 1998). The condition on independence of the log-likelihood ratio (LLR) increments was crucial in these works. Particularly, in Tartakovsky (1998) the moment inequalities for martingales are used as one of the main parts of the proof of asymptotic optimality. Moreover, the determinacy of the quadratic variation of the corresponding martingales is a very important property which is not generally fulfilled for the processes with dependent increments. On the other hand, Lai (1981) considered the case of discrete time, $t = 1, 2, \dots$, without the assumption on independence of the LLR increments but asymptotically homogeneous in a certain sense and presented sufficient conditions for asymptotic optimality of Wald's SPRT as probabilities of errors tend to zero. These sufficient conditions have been formulated in terms of finiteness of moments of the last exit time of the normalized LLR $t^{-1}Z(t)$ from the neighborhood of some constant q. In the terminology of Strassen (1967) and Lai (1976) this means r-quick convergence of $t^{-1}Z(t)$ as $t \to \infty$ to a constant q (namely in this sense the observation process is asymptotically homogeneous). Lai (1981) has also applied this result to prove asymptotic optimality of certain invariant sequential tests for two hypotheses.

In the present paper we continue to investigate the asymptotic properties of multihypothesis sequential tests that are specific combinations of one-sided SPRT's. While the idea of proofs is relied on the method proposed by Lai (1981), we generalize previous results, including that obtained in Lai (1981), to the multihypothesis case and substantially more general statistical models. In fact, Theorem 2.3 and Corollary 2.2 show that certain multihypothesis sequential tests asymptotically minimize not only the expected sample size but also all the moments of the stopping time distribution under very general conditions. These conditions allow us to consider models that include correlated and non-homogeneous (even asymptotically) processes observed either in discrete or continuous time.

The paper is organized as follows. In Section 2 we obtain a first-order asymptotic optimality property for two kinds of sequential tests of multiple simple hypotheses for a large class of statistical models. The sequential tests considered are based on Markov "accepting" and "rejecting" times for hypotheses and, in essence, are specific combinations of one-sided SPRT's. The conditions of asymptotic optimality are formulated in terms of r-quick convergence of LLR's normalized to $f(t)$ where $f(t)$ is some positive increasing function. The power function $f(t) = t^\lambda$, $\lambda > 0$, is of special interest for many applications. In Section 3 we present two examples related to the multi-sample slippage problem for correlated Gaussian processes where the corresponding conditions are verified. In Section 4 we show how the results may be extended to the case of multiple composite hypotheses for models with nuisance parameters or even for non-parametric models in the framework of invariant tests and give two additional examples that illustrate asymptotic optimality of invariant sequential tests.

2. Asymptotic Optimality of Multihypothesis Sequential Tests

2.1. Notation, test procedures and preliminaries

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \boldsymbol{P})$, $t \in \mathcal{R}$, where $\mathcal{R} = \{0, 1, \dots\}$ or $[0, \infty)$, be a stochastic basis with standard assumptions about monotonicity and right-continuity (in the continuous time case) of the σ-algebras $\mathcal{F}_t$. The sub-σ-algebra $\mathcal{F}_t = \sigma(X^t)$ of $\mathcal{F}$ is assumed to be generated by the process $X^t = \{X(u),\ 0 \le u \le t\}$ observed up to time t, which is defined on the space $(\Omega, \mathcal{F})$. Consider the problem of sequential testing of $N+1$ simple hypotheses "$H_i:\ \boldsymbol{P} = \boldsymbol{P}_i$", $i = 0, 1, \dots, N$, where $N \ge 1$ and $\boldsymbol{P}_i$ are completely known probability measures, which are mutually locally absolutely continuous.

A pair $D = (\tau, d)$ is said to be a sequential test of hypotheses if τ is a Markov stopping time with respect to the family $\{\mathcal{F}_t\}_{t \ge 0}$, i.e. $\{\tau \le t\} \in \mathcal{F}_t$, and $d = d(X^\tau)$ is an $\mathcal{F}_\tau$-measurable function (terminal decision function) with values in the set $\{0, 1, \dots, N\}$. Therefore $d = i$ is identified with accepting the hypothesis H_i, that is $\{d = i\} = \{\tau < \infty, D \,\text{accepts}\, H_i\}$.

Let us introduce two classes of tests:

$$\boldsymbol{\Delta}(\boldsymbol{\alpha}) = \{D:\ \boldsymbol{P}_i(d \ne i) \le \alpha_i, \quad i = 0, 1, \dots, N\},$$
$$\boldsymbol{\Delta}(\boldsymbol{\beta}) = \{D:\ \boldsymbol{P}_i(d = j) \le \beta_{ij}, \quad i, j = 0, 1, \dots, N,\ i \ne j\},$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_N)$, $\boldsymbol{\beta} = \|\beta_{ij}\|$ $(i \ne j)$ and where α_i and β_{ij} are positive values less than one. In other words, the classes $\boldsymbol{\Delta}(\boldsymbol{\alpha})$ and

$\mathbf{\Delta}(\boldsymbol{\beta})$ include the tests for which the probabilities of errors, $\Pr(\text{reject } H_i \mid H_i \text{ true})$ and $\Pr(\text{accept } H_j \mid H_i \text{ true})$, are not greater than the predefined numbers α_i and β_{ij}, respectively.

Let $\boldsymbol{P}_i^t(\cdot)$ be the restriction of the measure $\boldsymbol{P}_i$ to the σ-algebra $\mathcal{F}_t = \sigma(X^t)$ and let

$$Z_{ij}(t) = \log \frac{d\boldsymbol{P}_i^t}{d\boldsymbol{P}_j^t}(X^t), \qquad i, j = 0, 1, \dots, N, \ i \neq j$$

denote the LLR processes. For the sake of brevity we omit the index 0 when $j = 0$ and write simply $Z_i(t)$ instead of $Z_{i0}(t)$ $(Z_0(t) = 0)$.

In the sequel we consider the two following multihypothesis sequential tests for both discrete and continuous time cases simultaneously.

Test D_1. The test $D_1 = (\nu, d_1)$ represents the matrix SPRT (the combination of one-sided SPRT's) that is constructed from $(N+1)N/2$ extended SPRT's between hypotheses H_i and H_j. It is defined as follows

$$\nu = \inf\{t \in \mathcal{R} : Z_{ij}(t) \geq a_{ji} \quad \text{for all } j \neq i \text{ and some } i\}, \tag{2.1}$$

$$d_1 = i \quad \text{for which (2.1) holds,} \tag{2.2}$$

where a_{ji} are some positive finite numbers[1].

Alternatively, the test of (2.1), (2.2) may be written in the following, more convenient for asymptotic study form. Let

$$\tau_i = \inf\{t \in \mathcal{R} : Z_i(t) \geq \max_{j \neq i}[Z_j(t) + a_{ji}]\} \tag{2.3}$$

be the Markov "accepting" time for the hypothesis H_i. Then

$$\nu = \min(\tau_0, \tau_1, \dots, \tau_N), \qquad d_1 = i \quad \text{if} \quad \nu = \tau_i. \tag{2.4}$$

Thus in this test each component SPRT is extended until for some $i = 0, 1, \dots, N$ all N SPRT's involving H_i accept H_i. The test of this kind was considered earlier by Armitage (1950) and Lorden (1977) for i.i.d. observations, by Verdenskaya and Tartakovskii (1991) and Tartakovsky (1991) in the case of non-homogeneous Gaussian processes and by Tartakovsky (1998) for general statistical models that include arbitrary asymptotically power non-homogeneous processes with independent increments.

Test D_2. For any $i, j = 0, 1, \dots, N$, $i \neq j$, define the Markov times

$$\eta_{ij} = \inf\{t \in \mathcal{R} : Z_{ji}(t) \geq b_{ij}\}, \tag{2.5}$$

[1] In what follows in definitions of Markov times we always assume that $\inf\{\varnothing\} = \infty$.

where b_{ij} are positive thresholds. The second test is based on constructing, for any hypothesis H_i, the Markov "rejecting" time

$$\eta_i = \max_{\substack{0 \le j \le N \\ j \ne i}} \eta_{ij}.$$

Let $\eta_{(0)} \le \eta_{(1)} \le \cdots \le \eta_{(N-1)} \le \eta_{(N)}$ be a time-ordered set of rejecting times $\eta_0, \ldots, \eta_N$. The test $D_2 = (\eta, d_2)$ is defined as

$$\eta = \eta_{(N-1)}, \qquad d_2 = \arg \max_{i=0,1,\ldots,N} \eta_i. \tag{2.6}$$

Thus the observation process is continued up to rejection of all except one hypotheses, and in this instant the remaining hypothesis is accepted.

The adaptive rejecting type sequential tests for composite hypotheses were considered by Pavlov (1985, 1990) for the i.i.d. case when $b_{ji} = b$ and by Dragalin and Novikov (1994) for the processes with independent stationary increments [see also Tartakovsky (1998) in the case of the processes with independent non-stationary increments].

The following theorem allows for choosing the thresholds a_{ji} and b_{ij} to embed the tests D_1 and D_2 in the classes $\boldsymbol{\Delta}(\boldsymbol{\alpha})$ and $\boldsymbol{\Delta}(\boldsymbol{\beta})$. We use $\boldsymbol{E}_i$ to denote the expectation under the measure $\boldsymbol{P}_i$.

Theorem 2.1. *The following assertions hold*:

(i) $b_{ij} \ge |\log \beta_{ij}|$ *for* $i, j = 0, 1, \ldots, N$, $i \ne j$ *implies* $D_2 \in \boldsymbol{\Delta}(\boldsymbol{\beta})$;

(ii) $\sum_{k \ne i} \exp(-b_{ik}) \le \alpha_i$ *for* $i = 0, 1, \ldots, N$ *implies* $D_2 \in \boldsymbol{\Delta}(\boldsymbol{\alpha})$;

(iii) $a_{ji} \ge |\log \beta_{ji}|$ *for* $i, j = 0, 1, \ldots, N$, $i \ne j$ *implies* $D_1 \in \boldsymbol{\Delta}(\boldsymbol{\beta})$;

(iv) $\sum_{k \ne j} \exp(-a_{jk}) \le \alpha_j$ *for* $j = 0, 1, \ldots, N$ *implies* $D_1 \in \boldsymbol{\Delta}(\boldsymbol{\alpha})$.

Proof. We prove only (i) and (ii). The proof of (iii) and (iv) may be found in Tartakovsky (1998). By $1\!\!1_{\{\cdot\}}$ everywhere below we denote an indicator function.

(i) Since (by definition of the Markov time η_{ij}) $Z_{ji}(\eta_{ij}) - b_{ij} \ge 0$ on the set $\{\eta_{ij} < \infty\}$ and since

$$\{d_2 = j\} \Longrightarrow \{\eta_j < \infty\} \Longrightarrow \{\eta_{ij} < \infty\},$$

we have

$$\begin{aligned} \boldsymbol{P}_i(d_2 = j) &\le \boldsymbol{P}_i(\eta_{ij} < \infty) = \boldsymbol{E}_j \left\{ 1\!\!1_{\{\eta_{ij} < \infty\}} e^{-Z_{ji}(\eta_{ij})} \right\} \\ &= e^{-b_{ij}} \boldsymbol{E}_j \left\{ 1\!\!1_{\{\eta_{ij} < \infty\}} e^{-[Z_{ji}(\eta_{ij}) - b_{ij}]} \right\} \le e^{-b_{ij}}, \end{aligned}$$

from which (i) follows.

Assertion (ii) follows immediately from (i) and the obvious equality

$$\boldsymbol{P}_i(d_2 \neq i) = \sum_{\substack{0 \le j \le N \\ j \neq i}} \boldsymbol{P}_i(d_2 = j).$$

□

Corollary 2.1. *Let*

$$b_{ij} = b_i = \log(N/\alpha_i), \qquad a_{ji} = a_j = \log(N/\alpha_j). \tag{2.7}$$

Then both tests belong to the class $\boldsymbol{\Delta}(\boldsymbol{\alpha})$. *If*

$$b_{ij} = |\log \beta_{ij}|\,, \qquad a_{ji} = |\log \beta_{ji}|\,, \tag{2.8}$$

then $D_1, D_2 \in \boldsymbol{\Delta}(\boldsymbol{\beta})$.

In the sequel we are interested in asymptotic behavior of finite moments of a stopping time, $\boldsymbol{E}_i\tau^r$, $r > 0$, for small values of error probabilities. We first present a heuristic outline of the approach to finding asymptotics for the moments of the stopping time distribution. We concentrate on the test D_1, since the corresponding asymptotics for the D_2 follow from those for D_1 due to the following fact:

$$\eta(\|b_{ij}\|) \le \nu(\|a_{ji}\|) \quad \text{whenever } b_{ji} \le a_{ji}. \tag{2.9}$$

Indeed, the stopping time η of test D_2 can written in the form

$$\eta = \min_k \widetilde{\eta}_k, \qquad \widetilde{\eta}_k = \widetilde{\eta}_k(\|b_{ij}\|) = \max_{j \neq k} \eta_{jk},$$

where the Markov times η_{jk} are defined in (2.5). It is easily seen that

$$\widetilde{\eta}_i(\|b_{ij}\|) \le \inf\bigl\{t \in \mathcal{R} : \min_{j \neq i}[Z_{ij}(t) - b_{ji}] \ge 0\bigr\},$$

where the right hand side is nothing but the stopping time $\tau_i(\|b_{ji}\|)$ when $a_{ji} = b_{ji}$ (see (2.3)). Since $\eta(\|b_{ij}\|) = \min_k \widetilde{\eta}_k(\|b_{ij}\|)$, $\nu(\|a_{ji}\|) = \min_k \tau_k(\|a_{ji}\|)$ and since the corresponding stopping times are monotone increasing functions of thresholds the inequality (2.9) follows.

Assume that $\boldsymbol{E}_i Z_{ij}(t) = q_{ij} f(t)$ (at least approximately for large t), where q_{ij} are some positive finite numbers and $f(t)$ is an increasing function. Introduce the following notation: $\tilde{a}_{ji} = q_{ij}^{-1} a_{ji}$, $\widetilde{Z}_{ij}(t) = q_{ij}^{-1} Z_{ij}(t)$, and

$$\widetilde{\tau}_i = \inf\bigl\{t \in \mathcal{R} : \min_{j \neq i} \widetilde{Z}_{ij}(t) \ge \max_{j \neq i} \tilde{a}_{ji}\bigr\}.$$

If the thresholds a_{ji} are sufficiently large, one can neglect the overshoot and write

$$\min_{j \neq i} \widetilde{Z}_{ij}(\widetilde{\tau}_i) = \min_{j \neq i} [f(\widetilde{\tau}_i) + S_{ij}(\widetilde{\tau}_i)] \approx \max_{j \neq i} \tilde{a}_{ji},$$

where $S_{ij}(t) = \widetilde{Z}_{ij}(t) - f(t)$ is the random part of $\widetilde{Z}_{ij}(t)$. Next, if "fluctuations" of $Z_{ij}(t)$ are not too large in some sense, then it may be expected that in average $S_{ij}(\widetilde{\tau}_i) \ll f(\widetilde{\tau}_i)$. In this case $f(\widetilde{\tau}_i) \approx \max_{j \neq i} \widetilde{a}_{ji}$ and hence for large a_{ji} and any $r > 0$ we can expect that

$$\boldsymbol{E}_i \widetilde{\tau}_i^r \approx \left[F\big(\max_{j \neq i} \tilde{a}_{ji}\big) \right]^r,$$

where $F(\cdot)$ is the inverse function for $f(t)$. Since $\nu \leq \tau_i \leq \widetilde{\tau}_i$, one may also expect

$$\boldsymbol{E}_i \nu^r \lessapprox \left[F\big(\max_{j \neq i} \tilde{a}_{ji}\big) \right]^r.$$

On the other hand, if we choose the numbers a_{ji} according to the formulas (2.7) and (2.8), the right hand side of the latter inequality turns out to be the lower bound for $\boldsymbol{E}_i \tau^r$ in the corresponding classes (see Theorem 2.2). Thus we expect that Test D_1 is asymptotically optimal when the probabilities of errors vanish.

In mathematical terms the aforementioned property "fluctuations of $Z_{ij}(t)$ are not too large" will be expressed by an r-quick convergence of $S_{ij}(t)/f(t)$ to 0 (see Definition 2.1 below).

2.2. Lower bounds for moments of the stopping time

Theorem 2.1 does not require any specific conditions relative to the structure of the observed process. However, to obtain lower estimates for the moments of a stopping time and to prove asymptotic optimality of the above introduced tests some restrictions should be imposed. For derivation of the lower bounds the following almost sure (a.s.) convergence is sufficient:

$$\tfrac{1}{f(t)} Z_{ij}(t) \xrightarrow[t \to \infty]{\boldsymbol{P}_i - \text{a.s.}} q_{ij} \quad \text{for all } i, j,\ i \neq j, \tag{2.10}$$

where q_{ij} are positive finite constants and $f(t)$ is an increasing function that characterizes a degree of nonhomogeneity of the LLR's[2].

The asymptotic optimality result requires stronger constraints, which we express in terms of the so called r-quick convergence of the normalized LLR processes

$$Y_{ij}^{(f)}(t) = \frac{1}{f(t)} q_{ij}^{-1} Z_{ij}(t) \xrightarrow[t \to \infty]{\boldsymbol{P}_i - r\text{-quickly}} 1.$$

[2] If $f(t) = t$, we will say that the LLR processes are asymptotically homogeneous.

Definition 2.1. For $r > 0$, a stochastic process $\{\xi(t), t \in \mathcal{R}\}$ is said to converge r-quickly to a constant q (under $\boldsymbol{P}$) if and only if $\boldsymbol{E}T_h^r < \infty$ for all $h > 0$, where $T_h = \sup\{t \in \mathcal{R} : |\xi(t) - q| \geq h\}$ ($\sup\{\varnothing\} = 0$) and where $\boldsymbol{E}$ is the expectation with respect to $\boldsymbol{P}$ (cf. Lai (1976, 1981)). If the corresponding r-quick convergence condition holds for any positive r, we will say that the process $\xi(t)$ converges strongly completely[3] to q.

We will use the following notation for the r-quick convergence

$$\boldsymbol{P} - \mathrm{rQ} - \lim_{t\to\infty} \xi(t) = q \quad \text{or} \quad \xi(t) \xrightarrow[t\to\infty]{\boldsymbol{P}-r\text{-quickly}} q,$$

and the abbreviation s.c. will be used for the strong complete convergence. So $\xi(t) \to q$ $\boldsymbol{P}$−s.c. means that the process $\xi(t)$ converges to q strongly completely under $\boldsymbol{P}$ (i.e. r-quickly for all positive r).

Let h be a real positive number and let

$$T_{i,j}^{(f)}(h) = \sup\left\{t \in \mathcal{R} : \left|Y_{ij}^{(f)}(t) - 1\right| > h\right\}, \quad \sup\{\varnothing\} = 0$$

be the last time when the process $Y_{ij}^{(f)}(t)$ leaves the region $(1-h, 1+h)$. Suppose that $\boldsymbol{E}_i[T_{i,j}^{(f)}(h)]^r < \infty$ for all $h > 0$ and some $r > 0$. Then

$$\tfrac{1}{f(t)} Z_{ij}(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_i-r\text{-quickly}} q_{ij}. \tag{2.11}$$

Note that (2.10) is nothing but the strong law of large numbers for the LLR's with the rate $f(t)$ (generally $f(t) \neq t$), while (2.11) strengthens this law into the r-quick version. Both conditions do not require neither homogeneity nor independence of the LLR increments (cf. Tartakovsky (1998) where the case of the LLR's with non-homogeneous but independent increments was considered). Also, in terms of $T_{i,j}^{f}(h)$ the a.s. convergence (2.10) is expressed as $\boldsymbol{P}_i(T_{i,j}^{(f)}(h) < \infty) = 1$ for all $h > 0$.

By $D_{\boldsymbol{\alpha}} = (\tau_{\boldsymbol{\alpha}}, d_{\boldsymbol{\alpha}})$ and $D_{\boldsymbol{\beta}} = (\tau_{\boldsymbol{\beta}}, d_{\boldsymbol{\beta}})$ we denote any tests from the classes $\boldsymbol{\Delta}(\boldsymbol{\alpha})$ and $\boldsymbol{\Delta}(\boldsymbol{\beta})$, respectively. Recall that $F(t)$ is used to denote the inverse function for $f(t)$. Also write

$$\alpha_{\max} = \max_{0\leqslant i\leqslant N} \alpha_i, \qquad \beta_{\max} = \max_{\substack{0\leqslant i,j\leqslant N \\ i\neq j}} \beta_{ij}.$$

[3] The introduced definition of strong complete convergence should not be mixed with the notion of complete convergence [see Hsu and Robbins (1947), Lukacs (1975)], which is equivalent to 1-quick convergence in our terminology.

Lemma 2.1. *Let the observed process* $\{X(t),\, t \in \mathcal{R}\}$ *be such that the condition* (2.10) *is satisfied and, in addition, for all finite* $L > 0$

$$\boldsymbol{P}_i\Big(\sup_{t \le L} Z_{ij}^{+}(t) < \infty\Big) = 1. \tag{2.12}$$

Then the stopping times $\tau_{\boldsymbol{\alpha}}$ *and* $\tau_{\boldsymbol{\beta}}$ *go to infinity in probability as* $\alpha_{\max} \to 0$ *and* $\beta_{\max} \to 0$*, respectively. Moreover, for every* $0 < \delta < 1$ *and all* $i = 0, 1, \dots, N$

$$\lim_{\alpha_{\max}\to 0} \inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{P}_i\Big\{\tau > \delta \cdot F\Big(\max_{j \ne i} \tfrac{1}{q_{ij}} |\log \alpha_j|\Big)\Big\} = 1, \tag{2.13}$$

$$\lim_{\beta_{\max}\to 0} \inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\beta})} \boldsymbol{P}_i\Big\{\tau > \delta \cdot F\Big(\max_{j \ne i} \tfrac{1}{q_{ij}} |\log \beta_{ji}|\Big)\Big\} = 1. \tag{2.14}$$

Proof. Write $\Omega_{i,L} = \{d = i\} \cap \{\tau \le L\}$. By Wald's likelihood ratio identity

$$\boldsymbol{P}_j(d = i) = \boldsymbol{E}_i\big\{\mathbb{1}_{\{d=i\}} \exp[Z_{ji}(\tau)]\big\} = \boldsymbol{E}_i\big\{\mathbb{1}_{\{d=i\}} \exp[-Z_{ij}(\tau)]\big\}$$

and hence for any $L > 0$, $B > 0$

$$\begin{aligned} \boldsymbol{P}_j(d = i) &\ge \boldsymbol{E}_i\big\{\mathbb{1}_{\{\Omega_{i,L}, Z_{ij}(\tau) < B\}} \exp[-Z_{ij}(\tau)]\big\} \\ &\ge \exp(-B)\boldsymbol{P}_i\Big(\Omega_{i,L},\, \sup_{t \le L} Z_{ij}(t) < B\Big) \\ &\ge \exp(-B)\Big\{\boldsymbol{P}_i(\Omega_{i,L}) - \boldsymbol{P}_i\Big(\sup_{t \le L} Z_{ij}(t) \ge B\Big)\Big\}. \end{aligned}$$

Since $\boldsymbol{P}_i(\Omega_{i,L}) \ge \boldsymbol{P}_i(d = i) - \boldsymbol{P}_i(\tau > L)$, it follows that

$$\boldsymbol{P}_i(\tau > L) \ge \boldsymbol{P}_i(d = i) - \boldsymbol{P}_j(d = i)\exp(B) - \boldsymbol{P}_i\Big(\sup_{t \le L} Z_{ij}(t) \ge B\Big),$$

which for the class $\boldsymbol{\Delta}(\boldsymbol{\alpha})$ yields

$$\boldsymbol{P}_i(\tau_{\boldsymbol{\alpha}} > L) \ge 1 - \alpha_i - \alpha_j e^{B} - \boldsymbol{P}_i\Big(\sup_{t \le L} Z_{ij}(t) \ge B\Big). \tag{2.15}$$

Now, assume $B = c q_{ij} f(L)$ with $c > 1$. Then

$$\begin{aligned} &\boldsymbol{P}_i\Big\{\sup_{t \le L} Z_{ij}(t) \ge B\Big\} = \boldsymbol{P}_i\Big\{\sup_{t \le L} Z_{ij}(t) \ge c q_{ij} f(L)\Big\} \\ &\le \boldsymbol{P}_i\Big\{\sup_{t \le K} Z_{ij}^{+}(t) + \sup_{K < t \le L} Z_{ij}^{+}(t) \ge c q_{ij} f(L)\Big\} \\ &\le \boldsymbol{P}_i\Big\{\sup_{t \le K} Z_{ij}^{+}(t) + f(L) \sup_{K < t \le L}\Big(\tfrac{1}{f(t)} Z_{ij}^{+}(t) - q_{ij}\Big) \ge (c-1) q_{ij} f(L)\Big\} \\ &\le \boldsymbol{P}_i\Big\{\tfrac{1}{f(L)} \sup_{t \le K} Z_{ij}^{+}(t) + \sup_{t > K}\Big|\tfrac{1}{f(t)} Z_{ij}^{+}(t) - q_{ij}\Big| \ge (c-1) q_{ij}\Big\}. \end{aligned}$$

By the condition (2.10), for any $\varepsilon > 0$ there exists a finite with probability 1 random variable $K_{ij}(\varepsilon)$ such that

$$\left| \frac{Z_{ij}^{+}\big(K_{ij}(\varepsilon)\big)}{f\big(K_{ij}(\varepsilon)\big)} - q_{ij} \right| \le \varepsilon$$

(in fact, one may take $K_{ij}(\varepsilon) = T_{ij}^{(f)}(\varepsilon) + 1$) and hence

$$\boldsymbol{P}_i\Big\{\sup_{t\le L} Z_{ij}(t) \ge c q_{ij} f(L)\Big\} \le \boldsymbol{P}_i\Big\{\tfrac{1}{f(L)} \sup_{t\le K_{ij}(\varepsilon)} Z_{ij}^{+}(t) \ge (c-1)q_{ij} - \varepsilon\Big\}.$$

By the conditions (2.10), (2.12), the right side of the latter inequality approaches zero when $L \to \infty$, $c > 1 + \varepsilon/q_{ij}$ and therefore, as $L \to \infty$,

$$\boldsymbol{P}_i\Big\{\sup_{t\le L} Z_{ij}(t) \ge c q_{ij} f(L)\Big\} \to 0 \quad \text{for every } c > 1 \text{ and all } j \ne i. \tag{2.16}$$

Next, assuming $L = L_\alpha = F(\delta q_{ij}^{-1}|\log \alpha_j|)$ with $0 < \delta < 1/c$ and using (2.15), one obtains

$$\begin{aligned}\boldsymbol{P}_i\Big\{\tau_{\boldsymbol{\alpha}} > F\big(\delta q_{ij}^{-1}|\log \alpha_j|\big)\Big\} \ge 1 - \alpha_i - \alpha_j^{1-\delta c}& \\ - \boldsymbol{P}_i\Big\{\sup_{t\le L_\alpha} Z_{ij}(t) \ge c q_{ij} L_\alpha\Big\}&.\end{aligned} \tag{2.17}$$

Since (2.17) holds regardless of the choice of a specific sequential test $D_{\boldsymbol{\alpha}}$, we may write (for all $j \ne i$)

$$\begin{aligned}\inf_{D\in\boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{P}_i\Big\{\tau_{\boldsymbol{\alpha}} > F\Big(\delta q_{ij}^{-1}|\log \alpha_j|\Big)\Big\} \ge 1 - \alpha_i - \alpha_j^{1-\delta c}& \\ - \boldsymbol{P}_i\Big\{\sup_{t\le L_\alpha} Z_{ij}(t) \ge c q_{ij} L_\alpha\Big\}&,\end{aligned}$$

which along with (2.16) shows that for every $0 < \delta < 1$ and all $j \ne i$

$$\lim_{\alpha_{\max}\to 0} \inf_{D\in\boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{P}_i\Big\{\tau_{\boldsymbol{\alpha}} > F\Big(\delta q_{ij}^{-1}|\log \alpha_j|\Big)\Big\} = 1,$$

and the assertion (2.13) follows.

Analogously one can obtain (2.14) for $\boldsymbol{\Delta}(\boldsymbol{\beta})$. □

The following theorem determines the lower estimates for moments of any positive order of the stopping time.

Theorem 2.2. *Assume* (2.10) *holds. Then for any* $k > 0$ *and all* $i = 0, 1, \dots, N$ *as* $\alpha_{\max} \to 0$ *and* $\beta_{\max} \to 0$

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_i \tau^k \geq \left[\max_{j \neq i} F\Big(\frac{|\log \alpha_j|}{q_{ij}} \Big) \right]^k (1 + o(1)); \tag{2.18}$$

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\beta})} \boldsymbol{E}_i \tau^k \geq \left[\max_{j \neq i} F\Big(\frac{|\log \beta_{ji}|}{q_{ij}} \Big) \right]^k (1 + o(1)), \tag{2.19}$$

where $o(1) \to 0$ *as* $\alpha_{\max} \to 0$ $(\beta_{\max} \to 0)$.

Proof. Consider the class $\boldsymbol{\Delta}(\boldsymbol{\alpha})$. Denoting

$$M_i(\boldsymbol{\alpha}) = \frac{\tau_{\boldsymbol{\alpha}}}{\max_{j \neq i} F(|\log \alpha_j| / q_{ij})}$$

and applying the generalized Chebyshev inequality, we obtain

$$\boldsymbol{E}_i \big[M_i(\boldsymbol{\alpha}) \big]^k \geq \delta^k \boldsymbol{P}_i \big\{ M_i(\boldsymbol{\alpha}) > \delta \big\} \quad \text{for any } k > 0 \text{ and } \delta > 0,$$

where by Lemma 2.1

$$\lim_{\alpha_{\max} \to 0} \inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{P}_i \big\{ M_i(\boldsymbol{\alpha}) > \delta \big\} = 1 \quad \text{for every } 0 < \delta < 1.$$

This shows that

$$\underline{\lim}_{\alpha_{\max} \to 0} \inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_i \big[M_i(\boldsymbol{\alpha}) \big]^k \geq 1 \quad \text{for any } k > 0,$$

which is the same as (2.18).

The inequality (2.19) is proved in an analogous manner. □

2.3. Asymptotic optimality

Now everything is prepared to prove a first-order asymptotic optimality of the tests D_1 and D_2 relative to any positive moment of the observation time in respective classes when error probabilities go to zero. Everywhere below the notation $x_\alpha \sim y_\alpha$ means that $\lim_{\alpha \to 0} (x_\alpha / y_\alpha) = 1$.

Definition 2.2. A sequential test $D^* = (\tau^*, d^*)$ will be called asymptotically optimal in a first-order sense with respect to the kth moment of the stopping time in the class $\boldsymbol{\Delta}(\boldsymbol{\alpha})$ if $D^* \in \boldsymbol{\Delta}(\boldsymbol{\alpha})$ and

$$\frac{\boldsymbol{E}_i \big(\tau^* \big)^k}{\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_i \big(\tau \big)^k} \sim 1 \qquad \text{as} \quad \alpha_{\max} \to 0.$$

Write $\alpha_{\min} = \min_i \alpha_i$, $\beta_{\min} = \min_{i,j} \beta_{ij}$ and let γ_i and γ_{ij} be some numbers, $0 < \gamma_i \le 1$, $0 < \gamma_{ij} \le 1$. In our asymptotic consideration it is assumed that α_i and β_{ij} approach zero in such a way that

$$\log \alpha_i / \log \alpha_{\min} = \gamma_i, \qquad i = 0, 1, \dots, N; \tag{2.20}$$

$$\log \beta_{ij} / \log \beta_{\min} = \gamma_{ij}, \qquad i \neq j,\ i, j = 0, 1, \dots, N. \tag{2.21}$$

Note that we do not assume the constraints of the form $\alpha_i = c_i \alpha$ and $\beta_{ij} = c_{ij}\alpha$ with $\alpha \to 0$ and c_i, c_{ij} being positive constants [see, e.g., Lorden (1977); Pavlov (1990)]. In this particular case $\gamma_i = \gamma_{ij} = 1$ and it will be called asymptotically symmetric. We focus on the general asymmetric case assuming only that the numbers γ_i and γ_{ij} are bounded from zero. The importance of this general case is motivated by numerous applications [see, e.g., Tartakovsky (1991, 1998), Verdenskaya and Tartakovsky (1991) and examples below].

Additional notation:

$$R_i = \min_{\substack{0 \le j \le N \\ j \neq i}} (q_{ij}/\gamma_j), \quad \widetilde{R}_i = \min_{\substack{0 \le j \le N \\ j \neq i}} (q_{ij}/\gamma_{ji}).$$

We also recall that $\tilde{a}_{ji} = q_{ij}^{-1} a_{ji}$, $\widetilde{Z}_{ij}(t) = q_{ij}^{-1} Z_{ij}(t)$, and

$$\widetilde{\tau}_i = \inf\big\{t \in \mathcal{R} : \min_{j \neq i} \widetilde{Z}_{ij}(t) \ge \max_{j \neq i} \tilde{a}_{ji}\big\}.$$

The following theorem formalizes the heuristic argument given in the end of Section 2.1.

Theorem 2.3. *Suppose the conditions* (2.11) *hold. Then*

(i) $\boldsymbol{E}_i \nu^r < \infty$ *and* $\boldsymbol{E}_i \eta^r < \infty$ *for any finite* $\{a_{ji}\}$ *and* $\{b_{ij}\}$;

(ii) *if the numbers* a_{ji} *and* b_{ij} *are determined by* (2.7), *then both tests are asymptotically optimal in the class* $\boldsymbol{\Delta}(\boldsymbol{\alpha})$ *as* $\alpha_{\max} \to 0$: *for all* $i = 0, 1, \dots, N$

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_i \tau^r \sim \boldsymbol{E}_i \nu^r \sim \boldsymbol{E}_i \eta^r \sim \left[F\Big(\frac{|\log \alpha_{\min}|}{R_i}\Big)\right]^r; \tag{2.22}$$

(iii) *if the numbers* a_{ji} *and* b_{ji} *are determined by* (2.8), *then the tests* D_1 *and* D_2 *are asymptotically optimal in the class* $\boldsymbol{\Delta}(\boldsymbol{\beta})$ *as* $\beta_{\max} \to 0$: *for all* $i = 0, 1, \dots, N$

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\beta})} \boldsymbol{E}_i \tau^r \sim \boldsymbol{E}_i \nu^r \sim \boldsymbol{E}_i \eta^r \sim \left[F\Big(\frac{|\log \beta_{\min}|}{\widetilde{R}_i}\Big)\right]^r. \tag{2.23}$$

Proof. Let $h \in (0,1)$ and write $T_i^{(f)}(h) = \max_{j\neq i} T_{i,j}^{(f)}(h)$. From the one hand, by the definition of the stopping time $\widetilde{\tau}_i$

$$\min_{j\neq i} \widetilde{Z}_{ij}(\widetilde{\tau}_i - 1) < \max_{j\neq i} \tilde{a}_{ji}$$

and from the other hand by (2.11) on the event $\{\widetilde{\tau}_i > T_i^{(f)}(h) + 1\}$

$$\widetilde{Z}_{ij}(\widetilde{\tau}_i - 1) > f(\widetilde{\tau}_i - 1)(1-h) \quad \text{for all} \ \ j \neq i.$$

These two inequalities yield

$$\widetilde{\tau}_i < 1 + F\Big(\frac{\max_{j\neq i} \tilde{a}_{ji}}{1-h}\Big) \quad \text{on } \big\{\infty > \widetilde{\tau}_i > T_i^{(f)}(h) + 1\big\}$$

and hence for $h \in (0,1)$

$$\begin{aligned} \widetilde{\tau}_i \le 1 + \mathbb{1}_{\{\widetilde{\tau}_i > T_i^{(f)}(h)+1\}} F\Big(\frac{\max_{j\neq i} \tilde{a}_{ji}}{1-h}\Big) \\ + \mathbb{1}_{\{\widetilde{\tau}_i \le T_i^{(f)}(h)+1\}} T_i^{(f)}(h) \le 1 + F\Big(\frac{\max_{j\neq i} \tilde{a}_{ji}}{1-h}\Big) + T_i^{(f)}(h). \end{aligned} \tag{2.24}$$

Since

$$\nu = \min_k \tau_k \le \tau_i = \inf\big\{t \in \mathcal{R} : \min_{j\neq i} [\widetilde{Z}_{ij}(t) - \tilde{a}_{ji}] \ge 0\big\} \le \widetilde{\tau}_i$$

and by virtue of (2.11)

$$\boldsymbol{E}_i \big[T_i^{(f)}(h)\big]^r \le N \max_{j\neq i} \boldsymbol{E}_i \big[T_{i,j}^{(f)}(h)\big]^r < \infty,$$

it follows from (2.24) that $\boldsymbol{E}_i \nu^r < \infty$, i.e. the assertion (i) holds for D_1.

Also it follows from (2.24) that as $\min a_{ji} \to \infty$

$$\boldsymbol{E}_i \nu^r \le \Big[F\Big(\frac{\max_{j\neq i} \tilde{a}_{ji}}{1-h}\Big)\Big]^r (1 + o(1)) \quad \forall\, 0 < h < 1.$$

Passing $h \to 0$, we obtain the upper estimate

$$\boldsymbol{E}_i \nu^r \le \Big[\max_{j\neq i} F\Big(\frac{a_{ji}}{q_{ij}}\Big)\Big]^r (1 + o(1)) \quad \text{as} \ \ \min a_{ji} \to \infty. \tag{2.25}$$

Now, if we choose $a_{ji} = \log(N/\alpha_j)$, then by Corollary 2.1 $D_1 \in \boldsymbol{\Delta}(\boldsymbol{\alpha})$ and by (2.25)

$$\boldsymbol{E}_i \nu^r \le \Big[F\Big(\frac{|\log \alpha_{\min}|}{R_i}\Big)\Big]^r (1 + o(1)) \quad \text{as} \ \ \alpha_{\max} \to 0.$$

Since this upper bound is asymptotically the same as the lower one (2.18), we get (2.22) for Test D_1.

Next, if we choose $a_{ji} = |\log \beta_{ji}|$, then by Corollary 2.1 $D_1 \in \boldsymbol{\Delta}(\boldsymbol{\beta})$ and by (2.25)

$$\boldsymbol{E}_i \nu^r \le \left[F\Big(\frac{|\log \beta_{\min}|}{\widetilde{R}_i}\Big)\right]^r (1+o(1)) \quad \text{as} \quad \beta_{\max} \to 0.$$

Again the upper bound coincides with the lower one (2.19) and (2.23) follows for the class $\boldsymbol{\Delta}(\boldsymbol{\beta})$.

To prove assertions (i)-(iii) for Test D_2 it suffices to use the inequality (2.9) and Corollary 2.1. The proof is complete. □

Remark 2.1. While Theorem 2.2 requires only the strong law of large numbers (2.10), the r-quick convergence condition (2.11) cannot be weakened into the a.s. convergence in Theorem 2.3. (Note once more that $Z_{ij}(t)/f(t) \to q_{ij}$ $\boldsymbol{P}_i$−a.s. iff $\boldsymbol{P}_i\{T_{i,j}^{(f)}(h) < \infty\} = 1$ for all $h > 0$.) In fact, the a.s. convergence does not even guarantee the finiteness of moments of the stopping times ν and η.

The case where $f(t) = t^\lambda$, $\lambda > 0$ (asymptotically power nonhomogeneity) is especially important for many practical applications (see examples in the following sections).

Corollary 2.2. *Let $f(t) = t^\lambda$ and assume that $t^{-\lambda} Z_{ij}(t) \to q_{ij}$ $\boldsymbol{P}_i$−s.c. Then the tests D_1 and D_2 minimize all the positive moments of the observation time: for any $r > 0$ and all $i = 0, 1, \dots, N$*

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_i \tau^r \sim \boldsymbol{E}_i \nu^r \sim \boldsymbol{E}_i \eta^r \sim \left| \frac{\log \alpha_{\min}}{R_i} \right|^{r/\lambda} \quad \textit{as} \quad \alpha_{\max} \to 0, \quad (2.26)$$

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\beta})} \boldsymbol{E}_i \tau^r \sim \boldsymbol{E}_i \nu^r \sim \boldsymbol{E}_i \eta^r \sim \left| \frac{\log \beta_{\min}}{\widetilde{R}_i} \right|^{r/\lambda} \quad \textit{as} \quad \beta_{\max} \to 0. \quad (2.27)$$

Theorem 2.3 allows us to deduce the following optimality result in the i.i.d. case.

Corollary 2.3. *Suppose that in the discrete time case $X(t)$, $t = 1, 2, \dots$ are i.i.d. random variables with respect to one of the measures $\boldsymbol{P}_0, \dots, \boldsymbol{P}_N$, while in the continuous time case $\{X(t), t \ge 0\}$ is the process with independent stationary increments. Assume also that for some $r > 0$, $\boldsymbol{E}_i |Z_{ij}(1)|^{r+1} < \infty$. Then the asymptotic formulas* (2.26) *and* (2.27) *hold true with $\lambda = 1$, $R_i = \min_{j \ne i}(I_{ij}/\gamma_j)$, $\widetilde{R}_i = \min_{j \ne i}(I_{ij}/\gamma_{ij})$, where $I_{ij} = \boldsymbol{E}_i Z_{ij}(1)$ are the Kullback-Leibler information numbers. If $\boldsymbol{E}_i |Z_{ij}(1)|^r < \infty$ for all positive r, then the corresponding tests minimize any positive moment of the stopping time distribution.*

Proof. In the i.i.d. case the condition $\boldsymbol{E}_i|Z_{ij}(1)|^{r+1} < \infty$ is both necessary and sufficient for the r-quick convergence of $t^{-1}Z_{ij}(t)$ to I_{ij} (see Chow and Lai (1975), Lai (1975, 1976)). Thus the desired result is a direct consequence of Theorem 2.3. □

Remark 2.2. In the i.i.d. case the conditions $\boldsymbol{E}_i\,|Z_{ij}(1)|^{r+1} < \infty$ can be perhaps relaxed to the finiteness of the first absolute moments. To be specific, we conjecture that the assertion of Corollary 2.3 (the tests minimize all finite moments of the stopping time) holds whenever $0 < I_{ij} < \infty$. This result will be proved elsewhere.

Remark 2.3. Define the one-sided SPRT's

$$\tau_{ij}(a_{ji}) = \inf\{t \in \mathcal{R} : Z_{ij}(t) \ge a_{ji}\}, \quad i \neq j,\ i,j = 0,1,\ldots,N,$$

and let $D^* = (\nu^*, d^*)$ be the "accepting" multihypothesis (matrix) SPRT of the form

$$\nu^*(\|a_{ji}\|) = \min_i \max_{j\neq i} \tau_{ij}(a_{ji}), \quad d^* = \arg\min_i \max_{j\neq i} \tau_{ij}(a_{ji}).$$

Assertions of Theorem 2.3, Corollary 2.2, and Corollary 2.3 hold for sequential test D^*. To see this it is sufficient to note that the assertion of Corollary 2.1 for the test D_1 is equally true for the test D^* and that $\tau_i(\|a_{ji}\|) \le \max_{j\neq i} \tau_{ij}(a_{ji})$, hence $\nu(\|a_{ji}\|) \le \nu^*(\|a_{ji}\|)$.

The following lemma contains some inequalities and sufficient conditions for r-quick convergence that are useful for applications.

Lemma 2.2. *Let $\xi(t)$, $t \in \mathcal{R}$, be a random process, $\xi(0) = 0$. Define $M_\xi(u) = \sup_{0\le t\le u} |\xi(t)|$ in the continuous time case and $M_\xi(u) = \max_{1\le t\le \lceil u\rceil} |\xi(t)|$ in the discrete time case ($\lceil x\rceil$ is an integer part of x). Also define*

$$\begin{aligned} T(\varepsilon, r, f) &= \sup\Big\{t : \tfrac{1}{f(t)}\,|\xi(t)| > \varepsilon\Big\}, \quad \sup\{\varnothing\} = \infty; \\ \boldsymbol{J}_1(\varepsilon, r, f) &= \int_0^\infty u^{r-1}\boldsymbol{P}\Big\{|\xi(u)| \ge \varepsilon f(u)\Big\}\,du; \\ \boldsymbol{J}_2(\varepsilon, r, f) &= \int_0^\infty u^{r-1}\boldsymbol{P}\Big\{M_\xi(u) \ge \varepsilon f(u)\Big\}\,du. \end{aligned}$$

(i) *For any positive number r and any increasing function $f(t)$, $f(0) = 0$,*

$$\begin{aligned} r\boldsymbol{J}_1(\varepsilon, r, f) &\le \boldsymbol{E}\big[T(\varepsilon, r, f)\big]^r \\ &\le r\int_0^\infty t^{r-1}\boldsymbol{P}\Big\{\sup_{t\ge u} \tfrac{1}{f(u)}\,|\xi(u)| \ge \varepsilon\Big\}\,dt. \end{aligned} \tag{2.28}$$

(ii) *If* $f(t) = t^\lambda$, $\lambda > 0$, *then for any* $r > 0$

$$\Big\{\boldsymbol{J}_2(\varepsilon, r, \lambda) < \infty \;\forall\, \varepsilon > 0\Big\} \Longrightarrow \Big\{\boldsymbol{E}\big[T(\varepsilon, r, \lambda)\big]^r < \infty \;\forall\, \varepsilon > 0\Big\}. \quad (2.29)$$

Proof. Obviously,

$$\boldsymbol{P}\left\{|\xi(t)| \ge \varepsilon f(t)\right\} \le \boldsymbol{P}\big\{T(\varepsilon, r, f) \ge t\big\} \le \boldsymbol{P}\left\{\sup_{u \ge t} \tfrac{1}{f(u)} |\xi(u)| \ge \varepsilon\right\}$$

from which the inequalities (2.28) follow immediately.

To prove (ii) we note that

$$\begin{aligned}
&\boldsymbol{E}\big[T(2\varepsilon, r, f)\big]^r \le r \int_0^\infty t^{r-1} \boldsymbol{P}\left\{\sup_{u \ge t} u^{-\lambda} |\xi(u)| \ge 2\varepsilon\right\} dt \\
&\le r \int_0^\infty t^{r-1} \boldsymbol{P}\left\{\sup_{u \ge t} \big[|\xi(u)| - \varepsilon u^\lambda\big] \ge \varepsilon t^\lambda\right\} dt \\
&\le r \int_0^\infty t^{r-1} \boldsymbol{P}\left\{\sup_{u > 0} \big[|\xi(u)| - \varepsilon u^\lambda\big] \ge \varepsilon t^\lambda\right\} dt \\
&\le r \sum_{n=1}^\infty \int_0^\infty t^{r-1} \boldsymbol{P}\left\{\sup_{(2^{n-1}-1)t^\lambda < u^\lambda \le (2^n-1)t^\lambda} \Big[|\xi(u)| - \varepsilon u^\lambda\Big] \ge \varepsilon t^\lambda\right\} dt \\
&\le r \sum_{n=1}^\infty \int_0^\infty t^{r-1} \boldsymbol{P}\left\{\sup_{u^\lambda \le 2^n t^\lambda} |\xi(u)| \ge 2^{n-1} \varepsilon t^\lambda\right\} dt \\
&= r \sum_{n=1}^\infty \int_0^\infty t^{r-1} \boldsymbol{P}\left\{M_\xi(2^{n/\lambda} u) \ge 2^{n-1} \varepsilon t^\lambda\right\} dt \\
&= r \left[\sum_{n=1}^\infty 2^{-n/\lambda}\right] \int_0^\infty u^{r-1} \boldsymbol{P}\left\{M_\xi(u) \ge (\varepsilon/2) u^\lambda\right\} du \\
&= r \big(2^{1/\lambda} - 1\big)^{-1} \boldsymbol{J}_2(\varepsilon/2, r, \lambda).
\end{aligned}$$

Now the implication (2.29) follows in an obvious manner. □

3. Examples

In this section we give two examples of detection of signals in "multichannel" systems with correlated and non-stationary observations that illustrate the general results of the previous section.

Let $X(t) = (X_1(t), \ldots, X_N(t))$, $N \ge 2$, be an N-component process. Consider the following multiple decision problem known as the

N-sample slippage problem (see, e.g., Ferguson (1967), Tartakovsky (1997)). On the basis of observation of the N populations we wish to decide whether or not the populations are equal or one of them has slipped to the right of the rest and, if so, which one. In other words, under hypothesis H_0, $X_1(t), \dots, X_N(t)$ are mutually independent and distributed according to the measure $\boldsymbol{P}_0$, and, under H_i, all $X_k(t)$ are mutually independent, $X_1(t), \dots, X_{i-1}(t)$, $X_{i+1}(t), \dots, X_N(t)$ are distributed according to $\boldsymbol{P}_0$ and $X_i(t)$ according to the measure $\boldsymbol{P}_i$. The processes $X_m(t)$, $t \in \mathcal{R}$ may have fairly general structure. Since $Z_{ij}(t) = Z_i(t) - Z_j(t)$ and

$$Z_m(t) = \log \frac{d\boldsymbol{P}_m^t}{d\boldsymbol{P}_0^t} \left(X^t\right) = \log \frac{d\boldsymbol{P}_m^t}{d\boldsymbol{P}_0^t} \left(X_m^t\right),$$

the LLR $Z_{ij}(t)$ depends on observation process X^t through the components X_i^t and X_j^t.

One of the interesting applications of this model is a signal detection in a multi-channel (multi-resolution) system[4]. There may be no useful signal at all (hypothesis H_0) or a signal may be present in one of the N channels, in the ith, say (hypothesis H_i). It is necessary to detect a signal as soon as possible and to indicate the number of the channel where the signal is located. Consider a commonly used additive model (Bakut *et al.* (1963), Sosulin (1978), Tartakovsky (1991)) where an observed process in the jth channel, $X_j(t)$, represents either additive mixture of a useful signal $S_j(t)$ with a noise $\xi_j(t)$ or only noise:

$$X_j^{(i)}(t) = \begin{cases} S_j(t) + \xi_j(t) & \text{if} \quad i = j, \ j = 1, \dots, N, \\ \xi_j(t) & \text{if} \quad i \neq j, \ i = 0, 1, \dots, N, \end{cases} \tag{3.1}$$

where the superscript i means that the process $X_j(t)$ is regarded under H_i, i.e. when a signal is present in the ith channel. Generally the signal $S_j(t)$ is random and its structure is different for different channels.

In the following two examples we verify the conditions of Theorem 2.3 for the two particular models which are popular in the detection theory (see Sosulin (1978)).

3.1. Detection of deterministic signals in correlated Gaussian noise

The functions $S_1(t), \dots, S_N(t)$ are deterministic and $\xi_j(t) = v_j(t) + \dot{w}_j(t)$, where $\{v_j(t)\}$ are mutually independent L_2-continuous Gaussian

[4] General models considered below are valid for different sensors: radar, infrared sensor, ladar, or sonar.

processes, $\dot{w}_j(t)$ are mutually independent white Gaussian noises (i.e., $\{w_j(t)\}$ are mutually independent standard Wiener processes).

Write $\mathcal{F}_{mt} = \sigma(y_m(u),\ u \in [0,t])$,

$$y_m(t) = \int_0^t [1\!\!1_{\{i=m\}} S_m(u) + v_m(u)]\, du + w_m(t).$$

Let $\hat{v}_m(t) = \boldsymbol{E}_m [v_m(t) \mid \mathcal{F}_{mt}]$ be an optimal (in the mean-square-error sense) filtering estimate of the process $v_m(t)$ observed in a white Gaussian noise. Since $\xi_m(t)$ is a Gaussian process, $\hat{v}_m(t)$ is a linear functional

$$\hat{v}_m(t) = \int_0^t C_m(t,u) \left[dy_m(u) - 1\!\!1_{\{i=m\}} S_m(u)\, du\right],$$

where $C_m(t,u)$ is a characteristic of an optimal filter that satisfies the well-known Wiener-Hopf equation (see, e.g., Liptser and Shiryaev (1977)). Additional notation:

$$\widetilde{S}_m(t) = S_m(t) - \int_0^t S_m(u) C_m(t,u)\, du, \quad m = 1, \dots, N; \quad \widetilde{S}_0(t) = 0;$$
$$d\widetilde{y}_m(t) = dy_m(t) - \left[\int_0^t C_m(t,u)\, dy_m(u)\right] dt.$$

With the use of Theorems 7.12 and 7.15 of Liptser and Shiryaev (1977) it can be shown that the process $\widetilde{y}_m(t)$ may be represented via an innovation (standard Wiener) process $\widetilde{w}_m(t)$ as

$$\widetilde{y}_m(t) = \int_0^t 1\!\!1_{\{i=m\}} \widetilde{S}_m(u)\, du + \widetilde{w}_m(t) \tag{3.2}$$

and that the LLR's are of the form

$$\begin{aligned} Z_{ij}(t) = \int_0^t \widetilde{S}_i(u)\, d\widetilde{y}_i(u) - \int_0^t \widetilde{S}_j(u)\, d\widetilde{y}_j(u) \\ - \frac{1}{2} \int_0^t \left[\widetilde{S}_i^2(u) - \widetilde{S}_j^2(u)\right] du \end{aligned} \tag{3.3}$$

(cf. Tartakovskii (1981), Verdenskaya and Tartakovskii (1991), Tartakovsky (1991)).

Using (3.2) and (3.3), we obtain that under H_i

$$Z_{ij}^{(i)}(t) = \frac{1}{2}[\mu_i(t) + \mu_j(t)] + \int_0^t \widetilde{S}_i(u)\, d\widetilde{w}_i(u) - \int_0^t \widetilde{S}_j(u)\, d\widetilde{w}_j(u), \tag{3.4}$$

where $\{\widetilde{w}_i(t)\}$ and $\{\widetilde{w}_i(t)\}$ are mutually independent standard Wiener processes and where $\mu_m(t) = \int_0^t \widetilde{S}_m^2(u)\,du$.

Write

$$W_m(t) = \int_0^t \widetilde{S}_m(u)\,d\widetilde{w}_m(u)$$

and assume that $\mu_m(t) = q_m f(t)$ with $q_m > 0$ and $f(t)$ being an increasing function. Then

$$\boldsymbol{P}_i - \mathrm{rQ} - \lim_{t\to\infty} \tfrac{1}{f(t)} Z_{ij}(t) = (q_i + q_j)/2$$

if and only if

$$\boldsymbol{P}_i - \mathrm{rQ} - \lim_{t\to\infty} \tfrac{1}{f(t)} W_m(t) = 0, \quad m = 0, 1, \dots, N. \tag{3.5}$$

By Lemma 2.2 the following implication holds:

$$\left\{\int_0^\infty t^{r-1}\boldsymbol{P}_i\Big\{\sup_{u\le t}|W_m(u)| > \varepsilon f(t)\Big\}\,dt < \infty \ \forall\varepsilon > 0\right\} \Longrightarrow \\ \tfrac{1}{f(t)} W_m(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_i - r\text{-quickly}} 0. \tag{3.6}$$

Assume that $f(t) = t^\lambda$, $\lambda > 0$, i.e. $S_m(t) = \lambda q_m t^{\lambda-1}$ or, more generally,

$$\lim_{t\to\infty} t^{-\lambda}\mu_m(t) = q_m. \tag{3.7}$$

Then denoting by $\Phi(y)$ a standard normal distribution function, we obtain

$$\int_0^\infty t^{r-1}\boldsymbol{P}_i\big\{|W_m(t)| > \varepsilon t^\lambda\big\}\,dt = 2\int_0^\infty t^{r-1}\Phi\big(-\varepsilon t^{\lambda/2}/\sqrt{q_m}\big)\,dt \\ \le \frac{4q_m^{(r-\lambda/2+1)/2}}{\lambda\varepsilon^{r-\lambda/2+1}} \int_{-\infty}^\infty t^{r-\lambda/2}\,d\Phi\big(t\big) < \infty \quad \text{for all } \varepsilon, \lambda, r > 0. \tag{3.8}$$

Therefore under the condition (3.7)

$$t^{-\lambda} Z_{ij}(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_i - \text{s.c.}} (q_i + q_j)/2. \tag{3.9}$$

By Corollary 2.2, the asymptotic equalities (2.26) and (2.27) hold with $q_{ij} = (q_i + q_j)/2, i, j \ne 0$; $q_{i0} = q_i/2$, $q_{0j} = q_j/2$, $i, j = 1, \dots, N$, and the tests D_1 and D_2 are asymptotically optimal within the corresponding classes for small probabilities of errors. The condition (3.7), which turns out to be sufficient for optimality of the multihypothesis sequential tests, is satisfied for most applications.

However, if the function $f(t)$ grows too slow, then the expected observation times may be infinite. For illustration consider the following example. Let $\widetilde{S}_m(t) = \sqrt{q_m/(1+t)}$ and hence

$$\mu_m(t) = q_m \ln(1+t), \quad f(t) = \ln(1+t).$$

Then

$$\frac{1}{\ln(1+t)} Z_{ij}(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_i-\text{a.s.}} (q_i+q_j)/2,$$

but not r-quickly, since the condition (3.6) is not satisfied and, as a result, $\boldsymbol{E}_i T_{ij}^{(f)}(h) = \infty$. Indeed,

$$\int_0^\infty \boldsymbol{P}_i\big\{|W_m(t)| > \varepsilon \ln(1+t)\big\}\, dt = 2\int_0^\infty \Phi\Big(-\varepsilon\sqrt{\tfrac{\ln(1+t)}{q_m}}\Big)\, dt$$
$$= 4\int_0^\infty v\big(e^{v^2}-1\big)\Phi\big(-\varepsilon v/\sqrt{q_m}\big)\, dv = \infty \quad \text{for sufficiently small } \varepsilon.$$

Thus Theorem 2.3 may not be applied. In fact, in this case $\boldsymbol{E}_i\nu = \infty$ for sufficiently small α (see, e.g., Golubev and Khas'minskii (1983)).

Similar results may be obtained in the discrete time case when the noises $\xi_i(t)$, $t = 1, 2, \dots$ in (3.1) are pth order autoregressive (stable) Gaussian processes and the signals $\widetilde{S}_m(n)$ at the output of "autoregressive filters" are such that

$$\lim_{t\to\infty} t^{-\lambda} \sum_{n=1}^{t} \widetilde{S}_m^2(n) = q_m > 0.$$

3.2. Detection of Gaussian Markov signals in white noise

The noises $\xi_j(t) = \sqrt{B}\dot{w}_j(t)$ are mutually independent and statistically equivalent white Gaussian processes ($\{w_j(t)\}$ are standard Wiener processes) and the signals $S_i(t)$, $i = 1, \dots, N$, are stochastic processes ($t \in [0,\infty)$). Let us verify the r-quick convergence condition (2.11) in the case where $S_i(t)$ are statistically independent stationary Markov Gaussian processes,

$$\boldsymbol{E}S_i(t) = 0;\ \boldsymbol{E}S_i(t)S_i(t+u) = \delta^2 \exp(-\rho\,|u|);\ \boldsymbol{E}S_i(t)S_j(t) = 0,\ i \neq j.$$

We introduce the following additional notation

$$y_m(t) = \int_0^t 1\!\!1_{\{i=m\}} S_m(u)\, du + w_m(t), \quad \mathcal{F}_{mt} = \sigma(y_m(u), 0 \le u \le t),$$
$$\widehat{S}_m(t) = \boldsymbol{E}_m\big[S_m(t) \mid \mathcal{F}_{mt}\big],$$

where the index i corresponds to the hypothesis H_i (the signal is located in the ith channel).

Since $y_m(t)$ is the L_2-continuous Gaussian process and the processes $y_m(t)$ and $y_k(t)$ are mutually independent, applying Theorem 7.15 of Liptser and Shiryaev (1977) we obtain

$$\begin{aligned} Z_{ij}(t) = \frac{1}{B}\left[\int_0^t \widehat{S}_i(u)\,dy_i(u) - \int_0^t \widehat{S}_j(u)\,dy_j(u)\right] \\ - \frac{1}{2B}\int_0^t \left[\widehat{S}_i^2(u) - \widehat{S}_j^2(u)\right] du, \end{aligned} \tag{3.10}$$

where $\widehat{S}_0(t) = 0$.

Using (3.10) and Theorem 7.12 of Liptser and Shiryaev (1977), it is easy to show that

$$\begin{aligned} Z_{ij}^{(i)}(t) = \frac{1}{\sqrt{B}}\left[\int_0^t \widehat{S}_i(u)\,d\widetilde{w}_i(u) - \int_0^t \widehat{S}_j(u)\,dw_j(u)\right] \\ + \frac{1}{2B}\int_0^t \left[\widehat{S}_i^2(u) + \widehat{S}_j^2(u)\right] du \end{aligned} \tag{3.11}$$

where the superscript means that the LLR process is considered under H_i and where $\{\widetilde{w}_i(t), t \ge 0\}$ is a standard Wiener process.

The functional $\widehat{S}_m(t)$ coincides with the optimal mean-square-error estimate of the signal $S_m(t)$ and satisfies the system of Kalman-Bucy equations (see, e.g., Sosulin (1978), Liptser and Shiryaev (1977))

$$d\widehat{S}_m(t) = -(\rho + K_i(t)/B)\widehat{S}_m(t)\,dt + K_m(t)\,dy_m(t), \tag{3.12}$$

$$\dot{K}_m(t) = -2\rho K_m(t) - K_m^2(t)/B + \sigma, \qquad t \ge 0, \tag{3.13}$$

with the initial conditions $\widehat{S}_m(0) = 0$, $K_m(0) = \delta^2$. Here $K_m(t) = \boldsymbol{E}_m[S_m(t) - \widehat{S}_m(t)]^2$ is the mean-square filtering error and $\sigma = 2\rho\delta^2$ is the diffusion coefficient of $S_m(t)$. It is easy to see that

$$\boldsymbol{E}_i\widehat{S}_i^2(t) = \boldsymbol{E}_i S_i^2(t) - K_i(t) = \delta^2 - K_i(t). \tag{3.14}$$

Solving the Ricatti equation (3.13), after simple algebra we obtain

$$\frac{1}{t}\int_0^t K_i(u)\,du = \frac{\sigma}{\rho + C_1} + o(1) \qquad \text{as} \quad t \to \infty, \tag{3.15}$$

where $C_1 = \rho\sqrt{1 + \sigma/(\rho^2 B)}$. It follows from (3.14) and (3.15) that for all $i = 1, \ldots, N$

$$\lim_{t\to\infty} \frac{1}{2Bt}\int_0^t \boldsymbol{E}_i\widehat{S}_i^2(u)\,du = \frac{\delta^2}{B}\frac{Q}{\left(1 + \sqrt{1 + 2Q}\right)^2} = Q_1^2, \tag{3.16}$$

where $Q = \sigma/(2\rho^2 B) = \delta^2/(\rho B)$ is a parameter which characterizes a signal-to-noise ratio. Analogously one can show that for all $i = 0, 1, \dots, N$, $i \neq j$, $j \neq 0$

$$\lim_{t\to\infty} \frac{1}{2Bt} \int_0^t \boldsymbol{E}_i \widehat{S}_j^2(u)\, du = \frac{Q_1^2}{\sqrt{1+2Q}}. \tag{3.17}$$

By Taraskin (1970), if a random process $x(t) \in L_2[0,\infty]$ is such that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \boldsymbol{E}[x^2(u)]\, du = c^2 \neq 0,$$

then the normalized stochastic integral (over a standard Wiener process $W(t)$)

$$t^{-1/2} \int_0^t x(u)\, dW(u)$$

converges to a Gaussian variable $\mathcal{N}(0, c^2)$. From this and (3.16), (3.17) immediately follows that for all $i, j = 0, 1, \dots, N$ (including $i = j$)

$$\frac{1}{t} \int_0^t \widehat{S}_j(u)\, dw_j(u) \xrightarrow[t\to\infty]{\boldsymbol{P}_i-\text{s.c.}} 0. \tag{3.18}$$

Indeed, using the above facts one may show that

$$\int_0^\infty t^{r-1} \boldsymbol{P}_i \Big\{ \sup_{u\le t} |\widehat{W}_m(u)| > \varepsilon t \Big\}\, dt < \infty \quad \text{for any} \quad r > 0,\, \varepsilon > 0,$$

which implies (3.18) (see Lemma 2.2). Here $\widehat{W}_m(t) = \int_0^t \widehat{S}_m(u)\, dw_m(u)$.

Next, since $\widehat{S}_i(t)$ is a Gaussian process (see (3.12)), for any $p > 0$

$$\boldsymbol{E}_i \widehat{S}_j^{2p}(t) = 2^{p-1/2} \Gamma(p+1/2) \sqrt{2/\pi} \left\{ \boldsymbol{E}_i \left[\widehat{S}_j^2(t) \right] \right\}^p$$

and similar to (3.16), (3.17) we obtain that the expectations

$$\boldsymbol{E}_i \int_0^t \widehat{S}_j^{2p}(u)\, du \sim t\, 2^{p-1/2} \Gamma(p+1/2) \sqrt{2/\pi} \begin{cases} Q_1^{2p}, & i = j, \\ Q_1^{2p} (1+2Q)^{-p/2}, & i \neq j \end{cases}$$

are asymptotically linear functions of t as $t \to \infty$ for any $p > 0$. Therefore

$$\int_0^\infty t^{r-1} \boldsymbol{P}_i \big\{ \widehat{Y}_m(t) > \varepsilon t \big\}\, dt < \infty \quad \text{for any} \quad r > 0, \;\; \varepsilon > 0,$$

where $\widehat{Y}_m(t) = \int_0^t \widehat{S}_m^2(u)\, du$. Hence

$$\frac{1}{t} \int_0^t \widehat{S}_m^2(u)\, du \xrightarrow[t\to\infty]{\boldsymbol{P}_i-\text{s.c.}} \lim_{t\to\infty} \frac{1}{t} \int_0^t \left[\boldsymbol{E}_i \widehat{S}_m^2(u) \right] du.$$

This fact and (3.16), (3.17) imply that

$$\frac{1}{2Bt}\int_0^t \left[\widehat{S}_i^2(u)+\widehat{S}_j^2(u)\right]\,du$$
$$\xrightarrow[t\to\infty]{\boldsymbol{P}_i-\text{s.c.}} \begin{cases} Q_1^2(1+2Q)^{-1/2}\left(1+\sqrt{1+2Q}\right) & \text{for } i\neq 0,\\ Q_1^2(1+2Q)^{-1/2} & \text{for } i=0.\end{cases}$$

Now, using this last relation along with (3.11) and (3.18), we obtain that for every $r>0$

$$\frac{1}{t}Z_{ij}(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_i-r\text{-quickly}} \begin{cases} Q_1^2(1+2Q)^{-1/2}\left(1+\sqrt{1+2Q}\right), & i\neq 0,\\ Q_1^2(1+2Q)^{-1/2}, & i=0.\end{cases}$$

Thus the condition (2.11) holds with $f(t)=t$ and

$$q_{ij}=\begin{cases} Q_1^2(1+2Q)^{-1/2}\left(1+\sqrt{1+2Q}\right), & i\neq 0,\\ Q_1^2(1+2Q)^{-1/2}, & i=0.\end{cases}$$

Applying Theorem 2.3 we may conclude that the sequential tests of (2.4) and (2.6) asymptotically minimize all the moments of the observation time.

4. Other Examples: Problems with Nuisance Parameters

So far we considered the case where the measures $\boldsymbol{P}_0,\boldsymbol{P}_1,\dots,\boldsymbol{P}_N$ were completely known. However, more practical situation is that when these measures are unknown partially (parametric uncertainty) or completely (non-parametric uncertainty). In many cases one may overcome prior uncertainty by using the principle of invariance with respect to nuisance parameters. In this section we first explain how the general results of Section 2 can be used in sequential testing of composite hypotheses in the presence of nuisance parameters and then present two examples to illustrate the procedure.

Let $\{X_t, t\in\mathcal{R}\}$ be a stochastic process defined on the probability space $(\Omega,\mathcal{F},\boldsymbol{P})$, where we assume that the probability measure $\boldsymbol{P}$ belongs to a family $\mathfrak{P}$ (parametric or non-parametric), and the hypotheses H_i, $i=0,1,\dots,N$, to be tested are described by "$H_i:\boldsymbol{P}\in\mathfrak{P}_i$", where $\mathfrak{P}_i\in\mathfrak{P}$. Next, suppose that the family $\mathfrak{P}$ is invariant under a group of measurable transformations $\mathcal{G}$ on the sample space $\mathcal{X}$. The invariance property implies that the distribution of any invariant statistic depends on $\boldsymbol{P}$ only through its orbits. Therefore if there exists a group $\mathcal{G}$ that leaves the problem invariant and we are interested only in invariant sequential tests $D \in \mathcal{I}$, for which $D(G(X)) = D(X)\ \forall\ X \in \mathcal{X}$, $G \in \mathcal{G}$, then the hypotheses become simple. For more details, see Lehmann (1986), Ghosh (1970), Ferguson (1967).

Let $\mathbb{M}(X)$ be a maximal invariant statistic with respect to the group $\mathcal{G}$. By Theorem 5.6.1 of Ferguson (1967), instead of restricting attention to the class of invariant tests $\mathcal{I}$, we can restrict attention to the class of all tests, which are the functions of maximal invariant $\mathbb{M}$ (these two classes are equivalent). Thus it turns out that we can use all the results obtained in Section 2 by simple replacing the observed data X with a maximal invariant statistic $\mathbb{M}(X)$. In other words, we can again consider the tests D_1 and D_2 defined in (2.4) and (2.6), but with $Z_{ij}(t)$ now defined by

$$Z_{ij}(t) = \log \frac{d\boldsymbol{Q}_i^t}{d\boldsymbol{Q}_j^t}(\mathbb{M}_t), \qquad i \neq j,\ i,j = 0,1,\dots,N,$$

where $\boldsymbol{Q}_m$ is the measure corresponding to the maximal invariant under H_m.

Theorems 2.1, 2.2, and 2.3 remain true in the classes of invariant tests $\boldsymbol{\Delta}_{\mathcal{I}}(\boldsymbol{\alpha})$ and $\boldsymbol{\Delta}_{\mathcal{I}}(\boldsymbol{\beta})$ for which $D \in \mathcal{I}$ and

$$\Pr(\text{rejecting } H_i \mid H_i \text{ true}) \leq \alpha_i, \quad \Pr(\text{accepting } H_j \mid H_i \text{ true}) \leq \beta_{ij},$$

respectively.

4.1. The slippage problem for the exponential distribution with a scale nuisance parameter

Let $\mathcal{E}(\theta)$ be the exponential distribution with the density

$$p_\theta(x) = \theta \exp(-\theta x) \mathbb{1}_{\{[0,\infty)\}}(x), \tag{4.1}$$

where θ is an unknown positive scale parameter and $\mathbb{1}_{\{A\}}(x)$ is an indicator of a set A. The hypotheses H_i, $i = 0,1,\dots,N$, are of the form

$$\text{“}H_0:\ X_k(t) \in \mathcal{E}(\theta),\ k = 1,\dots,N\text{”};$$

$$\text{“}H_i:\ X_i(t) \in \mathcal{E}(\theta_i) \quad \text{and} \quad X_k(t) \in \mathcal{E}(\theta),\ k \neq i\text{”}, \qquad i = 1,\dots,N.$$

Write $\theta_i/\theta = c_i$ and assume that c_i, $i = 1,\dots,N$, are the given numbers from the interval $(0,1)$. Thus, under H_0, the observations $X_1(t),\dots,X_N(t)$, $t = 1,2,\dots$ are i.i.d. according to $\mathcal{E}(\theta)$. Under H_i, $X_j(t)$, $t = 1,2,\dots$, $j \neq i$ are i.i.d. again according to $\mathcal{E}(\theta)$ and $X_i(t)$, $t = 1,2,\dots$ are i.i.d. according to $\mathcal{E}(c_i\theta)$ ($X_i(t)$ and $X_j(t)$, $i \neq j$ are mutually independent), where θ is unknown and $c_i \in (0,1)$, $i = 1,\dots,N$, are given.

Write

$$Y_k(n) = X_k(n)/X_1(1),\ Y_k^t = (Y_k(1), \dots, Y_k(t)),\ \mathbb{M}_t = (Y_1^t, \dots, Y_N^t).$$

The problem is invariant under the group of scale changes $\mathcal{G} : G_b(x) = b\,x,\ b > 0$ and $\mathbb{M}_t$ is the maximal invariant. The statistic $\mathbb{M}_t$ has distributions with the densities (under H_i)

$$\begin{aligned} p_{it}(\boldsymbol{m}_t) &= c_i^t \int_0^\infty v^{Nt-1} \exp\left(-v c_i \sum_{n=1}^t y_i(n) - v \sum_{n=1}^t \sum_{\substack{k=1 \\ k\neq i}}^N y_k(n) \right) dv \\ &= \frac{c_i^t\,\Gamma(Nt)}{\left[\sum_{k=1}^N \sum_{n=1}^t y_k(n) - (1-c_i) \sum_{n=1}^t y_i(n)\right]^{Nt}}, \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function and where $c_0 = 1$ $(y_1(1) = 1)$. Using these relations we obtain that the LLR's $Z_{ij}(t) \stackrel{\text{def}}{=} \log[p_{it}(\mathbb{M}_t)/p_{jt}(\mathbb{M}_t)]$ are equal to

$$Z_{ij}(t) = t\left\{\log(c_i/c_j) + N\,\log\left[\frac{1-(1-c_j)M_j(t)}{1-(1-c_i)M_i(t)}\right]\right\}, \tag{4.2}$$

where

$$M_k(t) = \frac{t^{-1}\sum_{n=1}^t X_k(n)}{\sum_{m=1}^N t^{-1}\sum_{n=1}^t X_m(n)}.$$

Since $\boldsymbol{E}_i\,|X_k(1)|^s < \infty\ \forall\ s > 0$, we have

$$M_k(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\theta}-r\text{-quickly}} \frac{\boldsymbol{E}_{i,\theta}X_k(1)}{\boldsymbol{E}_{i,\theta}X_i(1) + \sum_{m\neq i}\boldsymbol{E}_{i,\theta}X_m(1)} \tag{4.3}$$

for every $r > 0$ and $i = 0, 1, \dots, N$. Next, since $\boldsymbol{E}_{i,\theta}X_i(1) = (c_i\theta)^{-1}$ and $\boldsymbol{E}_{i,\theta}X_m(1) = \theta^{-1}$ for $m \neq i$, we obtain from (4.3) that

$$M_k(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\theta}-\text{s.c.}} \begin{cases} [1+(N-1)c_i]^{-1}, & k = i, \\ c_i[1+(N-1)c_i]^{-1}, & k \neq i,\ i \neq 0, \\ N^{-1}, & i = 0. \end{cases} \tag{4.4}$$

Now using (4.2) and (4.4) it is easy to show that the conditions (2.11) are fulfilled for every $r > 0$ with $\lambda = 1$ and

$$q_{ij} = \begin{cases} \log\left(\frac{c_i}{c_j}\right) + N\log\left\{1 + \frac{(1-c_i)^2 + c_i(c_j-c_i)}{c_iN}\right\}, & i \neq 0,\ j \neq 0, \\ N\log\left(1 + \frac{1-c_i}{c_iN}\right) - |\log c_i|\,, & i \neq 0,\ j = 0, \\ N\log\left(1 - \frac{1-c_j}{N}\right) + |\log c_j|\,, & i = 0,\ j \neq 0. \end{cases} \tag{4.5}$$

It is also easy to check that q_{ij} are positive and finite for all $i \neq j$ and for arbitrary $0 < c_i < 1$, $i = 1, \dots, N$. Thus, applying Theorem 2.3, we conclude that the invariant tests D_1 and D_2 based on the LLR's (4.2) asymptotically minimize all the moments of the observation time distribution within the classes $\boldsymbol{\Delta}_{\mathcal{I}}(\boldsymbol{\alpha})$ and $\boldsymbol{\Delta}_{\mathcal{I}}(\boldsymbol{\beta})$ of invariant tests. The asymptotic relations (2.22), (2.23) are true with $F(y) = y$ and q_{ij}, which are defined in (4.5). If, furthermore, N tends to infinity so that $\log(N)/|\log\alpha| = o(1)$, then

$$q_{ij} = \begin{cases} \log(\frac{c_i}{c_j}) + \frac{(1-c_i)^2}{c_i} + c_j - c_i + O(1/N), & i, j \neq 0, \\ (1-c_i)c_i^{-1} - |\log c_i| + O(1/N), & i \neq 0,\ j = 0, \\ -(1-c_j) + |\log c_j| + O(1/N), & i = 0,\ j \neq 0. \end{cases} \qquad (4.6)$$

It is interesting to compare these asymptotic characteristics with those in the case of known in advance value of σ^2. Suppose for simplicity that $\alpha_1 = \cdots = \alpha_N = \alpha$ and $\alpha_0 \le \alpha$ (this case is typical for many applications). As shown by Tartakovsky (1991), in the latter case (with known σ^2)

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_i \tau^r \sim \boldsymbol{E}_i \nu^r \sim \boldsymbol{E}_i \eta^r \sim \left| \frac{\log \alpha_0}{Q - \log(1+Q)} \right|^r, \qquad i = 1, \dots, N;$$

$$\inf_{D \in \boldsymbol{\Delta}(\boldsymbol{\alpha})} \boldsymbol{E}_0 \tau^r \sim \boldsymbol{E}_0 \nu^r \sim \boldsymbol{E}_0 \eta^r \sim \left| \frac{\log \alpha}{\log(1+Q) - Q/(1+Q)} \right|^r.$$

One can see that asymptotic performance of the tests are always better in the case of complete prior knowledge. However, if $N \to \infty$ such that $\log(N)/|\log\alpha| = o(1)$, then by (4.6) the invariant tests have asymptotically the same characteristics as the asymptotically optimal sequential tests in the case of known σ^2. At the same time if N is fixed and $Q \to 0$, then the invariant tests provide $[(N-1)/N]^r$ times worth characteristics compared with the case of completely known σ^2 (for all H_i, $i = 0, 1, \dots, N$). By contrast, as $Q \to \infty$ and N is fixed, then

$$\boldsymbol{E}_0 \nu^r(\sigma^2 \text{ unknown}) \sim \boldsymbol{E}_0 \nu^r(\sigma^2 \text{ known}),$$

$$\frac{\boldsymbol{E}_i \nu^r(\sigma^2 \text{ unknown})}{\boldsymbol{E}_i \nu^r(\sigma^2 \text{ known})} \sim (Q/[(N-1)\log Q])^r \to \infty, \qquad i = 1, \dots, N.$$

4.2. The N-sample slippage problem for the Gaussian distribution with unknown mean and variance

Suppose that under H_0 the components $X_k(t)$, $t = 1, 2, \dots$ of the vector $X(t) = (X_1(t), \dots, X_N(t))$ are of the form

$$X_k(t) = \theta + \xi_k(t), \qquad k = 1, \dots, N$$

while under $H_i,\ i = 1, \dots, N$,

$$X_i(t) = \theta + \mu_i + \xi_i(t) \quad \text{and} \quad X_k(t) = \theta + \xi_k(t) \quad \text{for } k \neq i,$$

where $\xi_i(t)$, $t = 1, 2, \dots$ are i.i.d. Gaussian random variables with zero mean and variance σ^2. Both parameters, the mean θ and the variance σ^2, are supposed to be unknown and we wish to test $N+1$ hypotheses

$$\text{“}H_0 : \mu_k/\sigma = 0 \quad \text{for all } k = 1, \dots, N\text{”};$$

$$\text{“}H_i : \mu_k/\sigma = 0 \quad \text{for } k \neq i \quad \text{and} \quad \mu_i/\sigma = \delta_i\text{”}, \quad i = 1, \dots, N,$$

where $\delta_1, \dots, \delta_N$ are given positive numbers (the modification for arbitrary, possibly negative numbers is straightforward). This problem is invariant under change in location and scale and the maximal invariant is $\mathbb{M}_t = (Y_1^t, \dots, Y_N^t)$, where

$$Y_k^t = (Y_k(1), \dots, Y_k(t)), \quad Y_k(n) = [X_k(n) - X_1(1)]/[X_2(1) - X_1(1)],$$
$$Y_1(1) = 0, \quad Y_2(1) = 1.$$

It may be shown that under H_i the distribution of $\mathbb{M}_t$ has the density

$$\begin{aligned} p_{it}(\boldsymbol{m}_t) = (2\pi)^{-(Nt-1)/2}(Nt)^{-1/2} s_t^{-(Nt-1)} \exp\left\{-\frac{\delta_i^2(N-1)t}{2N}\right\} \\ \cdot \int_0^\infty v^{Nt-2} \exp\left\{-\frac{Nt}{2}v^2 + \delta_i \frac{\sum_{n=1}^t [y_i(n) - \overline{y}_t]}{s_t} v\right\}, \end{aligned} \tag{4.7}$$

where $y_1(1) = 0$, $y_2(1) = 1$,

$$\overline{y}_t = (Nt)^{-1} \sum_{n=1}^t \sum_{k=1}^N y_k(n), \quad s_t^2 = (Nt)^{-1} \sum_{n=1}^t \sum_{k=1}^N [y_k(n) - \overline{y}_t]^2.$$

Define

$$f(v, z) = -\tfrac{1}{2}v^2 + zv + \log v, \tag{4.8}$$

$$U_N(z, t) = \int_0^\infty v^{-2} \exp\bigl[tNf(v, z)\bigr]\, dv,$$

$$T_i(t) = \frac{\delta_i (Nt)^{-1} \sum_{n=1}^t [X_i(n) - \overline{X}_t]}{\left\{(Nt)^{-1} \sum_{n=1}^t \sum_{k=1}^N [X_k(n) - \overline{X}_t]^2\right\}^{1/2}}. \tag{4.9}$$

By (4.7) the LLR's for the maximal invariant are of the form

$$Z_{ij}(t) = -\frac{N-1}{2N}(\delta_i^2 - \delta_j^2)t + \log \frac{U_N(T_i(t), t)}{U_N(T_j(t), t)}. \tag{4.10}$$

The LLR's $Z_{ij}(n)$ are too complicated for direct use. In fact, it is rather difficult, if possible, to calculate $Z_{ij}(n)$ explicitly (although it can be done locally for small differences $|d_i - d_j|$, i.e. for close hypotheses). However, for questions as general as convergence it is permissible to replace $Z_{ij}(n)$ by a suitable approximation, say $Z'_{ij}(n)$. If $\boldsymbol{P}_i(|Z_{ij}(n) - Z'_{ij}(n)| < C) = 1$ for some $n \geq n_0$ $(n_0 \geq 1)$ with C a constant, then the r-quick convergence of $n^{-1}Z'_{ij}(n)$ to q_{ij} will imply the corresponding convergence $n^{-1}Z_{ij}(n) \to q_{ij}$.

Following the lines of Wijsman (1971) we now show that

$$\left|\tfrac{1}{t}Z_{ij}(t) - \Psi_{ij}\big(T_i(t), T_j(t)\big)\right| \leq C_{ij}/t, \quad i \neq j, \quad t = 1, 2, \dots, \qquad (4.11)$$

where C_{ij} are finite positive constants. The functions $\Psi_{ij}(z_1, z_2)$ will be defined later on (see (4.15) below). The proof is based, in essence, on the application of the Laplace asymptotic integration method (see, e.g., Copson (1965)) and its "uniform version" (due to Wijsman (1971)).

Using conventional Laplace's asymptotic method, we obtain

$$U_N(z_k, t) \sim 2\left[\frac{\pi}{2Ntv_0^2(1+v_0^2)}\right]^{1/2} \exp\{Ntf(v_0, z_k)\}, \quad t \to \infty, \quad (4.12)$$

where $v_0 = v_0(z_k, N) = 0.5[z_k + (4 + z_k^2)^{1/2}]$ is the point of maximum of $f(v, z_k)$, $f(v_0, z_k) = \max_{v \geq 0} f(v, z_k)$. This formula is valid for every fixed z_k. However, we are interested in behavior of $U_N(T_k(t), t)$ for a random $T_k(t)$ rather for a fixed z_k and hence we need the convergence in (4.12) to be uniform for $|z_k| \leq \delta_k N^{-1/2}$. (Note that $|T_k(t)| \leq \delta_k N^{-1/2}$, see (4.9).) Thus we need the extension of the conventional Laplace method to the uniform version that gives Theorem 4.1 of Wijsman (1971). It is easily checked that all the conditions of this theorem are justified in our case and, in fact, (4.12) holds uniformly in $|z_k| \leq \delta_k N^{-1/2}$. This means that uniformly in $|z_k| \leq \delta_k/\sqrt{N}$

$$\log U_N(z_k, t) = Ntf(v_0(z_k), z_k) - \tfrac{1}{2}\log t + C_k(z_k) + o(1) \quad \text{as} \quad t \to \infty,$$

where

$$C_k(z_k) = \frac{1}{2}\log\frac{2\pi}{Nv_0^2(z_k)[1 + v_0^2(z_k)]}.$$

Now using this last expression and noting that

$$f\big(v_0(z), z\big) = \max_{v \geq 0} f\big(v, z\big) = \phi(z) - \tfrac{1}{2} - \log 2,$$

where

$$\phi(z) = z\big(z + \sqrt{4 + z^2}\big)/4 + \log\big(z + \sqrt{4 + z^2}\big), \qquad (4.13)$$

we obtain that

$$\log \frac{U_N(z_i,t)}{U_N(z_j,t)} = Nt\left[\phi(z_i) - \phi(z_j)\right] + R(z_i,z_j) + o_t(z_i,z_j), \qquad (4.14)$$

where $o_t(z_i,z_j) \to 0$ as $t \to \infty$ uniformly in $|z_k| \le \delta_k N^{-1/2}$, $k = i,j$, and where

$$R(z_i,z_j) = \frac{1}{2}\log\left[\frac{\left(z_j + \sqrt{4+z_j^2}\right)^2\left[1 + \left(z_j + \sqrt{4+z_j^2}\right)^2\right]}{\left(z_i + \sqrt{4+z_i^2}\right)^2\left[1 + (z_i + \sqrt{4+z_i^2}\right)^2\right]}\right].$$

Write

$$\Psi_{ij}(x,y) = N\left[\phi(x) - \phi(y)\right] - \frac{N-1}{2N}\left(\delta_i^2 - \delta_j^2\right). \qquad (4.15)$$

Exploiting (4.10) and (4.14), we conclude that

$$|Z_{ij}(t) - t\Psi_{ij}(T_i(t),T_j(t)) - R(T_i(t),T_j(t))| = o(1) \quad \text{as} \quad t \to \infty.$$

Since $|T_k(t)| \le \delta_k/\sqrt{N}$, the term $R(T_i(t),T_i(t))$ is bounded by a constant, which along with the previous equality shows that there is a finite positive constant C_{ij} for which the approximation (4.11) holds with the function Ψ_{ij} defined by (4.13), (4.15).

Thus if under $\boldsymbol{P}_{i,\boldsymbol{a}}$ ($\boldsymbol{a} = (\theta,\sigma^2)$) the statistics $T_k(t)$ converge r-quickly to some constants Λ_{ik}, the value of $t^{-1}Z_{ij}(t)$ will converge to $\Psi(\Lambda_{ii},\Lambda_{ij})$ and to obtain the final result it suffices to investigate the limiting behavior of $T_k(t)$. Since $\boldsymbol{E}_i\,|X_k(1)|^s < \infty$ for all $s > 0$, it follows that for any $r > 0$

$$t^{-1}\sum_{n=1}^{t} X_k(n) \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\boldsymbol{a}}-r\text{-quickly}} \boldsymbol{E}_{i,\boldsymbol{a}}X_k(1) = \begin{cases}\theta, & i=0,\ i\ne k,\\ \theta+\mu_i, & k=i;\end{cases}$$

$$\overline{X}_t \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\boldsymbol{a}}-r\text{-quickly}} N^{-1}\sum_{k=1}^{N}\boldsymbol{E}_{i,\boldsymbol{a}}X_k(1) = \begin{cases}\theta, & i=0,\\ \theta+\mu_i/N, & i\ne 0;\end{cases}$$

$$(Nt)^{-1}\sum_{k=1}^{N}\sum_{n=1}^{t} X_k^2(n) \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\boldsymbol{a}}-r\text{-quickly}} N^{-1}\sum_{k=1}^{N}\boldsymbol{E}_{i,\boldsymbol{a}}X_k^2(1)$$

$$= \begin{cases}\sigma^2+\theta^2, & i=0,\\ \sigma^2+\theta^2+(\mu_i^2+2\mu_i\theta)/N, & i\ne 0;\end{cases}$$

$$\overline{X}_t^2 \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\boldsymbol{a}}-r\text{-quickly}} \begin{cases}(\theta+\mu_i/N)^2, & i\ne 0,\\ \theta^2, & i=0.\end{cases}$$

Since $T_k(t)$ may be rewritten in the form

$$T_k(t) = \frac{(\delta_k/N)\left[t^{-1}\sum_{n=1}^t X_k(n) - \overline{X}_t\right]}{\left[N^{-1}\sum_{k=1}^N t^{-1}\sum_{n=1}^t X_k^2(n) - \overline{X}_t^2\right]^{1/2}},$$

it follows from these last relations that $T_k(t) \to 0$ $\boldsymbol{P}_{0,\boldsymbol{a}}$-s.c. for $k = 1, \dots, N$ and that

$$T_k(t) \xrightarrow[t\to\infty]{\boldsymbol{P}_{i,\boldsymbol{a}}-\text{s.c.}} \begin{cases} -\delta_i\delta_k/[N^2\{1+(N-1)\delta_i^2/N^2\}^{1/2}], & k \neq i, \\ \delta_i^2(N-1)/[N^2\{1+(N-1)\delta_i^2/N^2\}^{1/2}], & k = i \end{cases}$$

for $i = 1, \dots, N$. Using this fact together with (4.11), (4.14), and (4.15), we finally obtain that the conditions (2.11) hold for every $r > 0$ with $f(t) = t$ and

$$q_{ij} = N\left[\phi(\Lambda_{ii}) - \phi(\Lambda_{ij})\right] - \frac{N-1}{2N}(\delta_i^2 - \delta_j^2), \quad i, j \neq 0; \tag{4.16}$$

$$q_{0j} = \frac{N-1}{2N}\delta_j^2, \qquad q_{i0} = N\left[\phi(\Lambda_{ii}) - \log 2\right] - \frac{N-1}{2N}\delta_i^2, \tag{4.17}$$

where

$$\begin{aligned} \Lambda_{ii} &= \frac{(N-1)\delta_i^2}{N^2\sqrt{1+(N-1)\delta_i^2/N^2}}, \\ \Lambda_{ij} &= -\frac{\delta_i\delta_j}{N^2\sqrt{1+(N-1)\delta_i^2/N^2}}. \end{aligned} \tag{4.18}$$

An analysis of (4.16)-(4.18) shows that $q_{ij} > 0$ and hence the conditions of Theorem 2.3 are satisfied. Thus the tests considered are asymptotically optimal. Furthermore, it follows from (4.16)-(4.18) that $\min_{j\neq i} q_{ij} = q_{i0}$ for $i = 1, \dots, N$ and hence in the case where $\alpha_i = \alpha$ for $i = 1, \dots, N$ and $\alpha_0 \leq \alpha$, we obtain from (2.22)

$$\begin{aligned} \inf_{D\in\boldsymbol{\Delta}_{\mathcal{I}}(\boldsymbol{\alpha})} \boldsymbol{E}_i\tau^r \sim \boldsymbol{E}_i\nu^r \sim \boldsymbol{E}_i\eta^r &\sim \left|\frac{2N\log\alpha_0}{(N-1)\delta_i^2(\rho_{i,N}-1)}\right|^r, \quad i \neq 0, \\ \inf_{D\in\boldsymbol{\Delta}_{\mathcal{I}}(\boldsymbol{\alpha})} \boldsymbol{E}_0\tau^r \sim \boldsymbol{E}_0\nu^r \sim \boldsymbol{E}_0\eta^r &\sim \left|\frac{2N\log\alpha}{(N-1)(\min_j\delta_j)^2}\right|^r, \end{aligned}$$

where $\rho_{i,N} = 2N^2[\phi(\Lambda_{ii}) - \log 2]/[(N-1)\delta_i^2]$.

It should be pointed out that the problem of testing the hypotheses $H_i : \mu = \mu_i$ does not have an invariant solution when there is the additional nuisance parameter – variance. In fact, following the lines of the proof of Theorem 3.3.1 of Linnik (1968) one can establish that the hypothesis "$H : (\theta/\sigma^\epsilon) = \delta$" does not admit an invariant testing for $\epsilon \in [0, 1)$. To avoid this complication adaptive sequential tests may be constructed, based on estimates of nuisance parameters. This approach will be considered elsewhere.

Also, for $N = 1$ the problems considered in these two examples do not have an invariant solution. Thus the condition $N \geq 2$ appears to be substantial.

Acknowledgements

The author thanks Thomas Ferguson and Robert Liptser for useful discussions.

References

P. Armitage, Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis, *J. Royal Statist. Soc. B*, **12** (1950) 137–144.

P.A. Bakut, I.A. Bolshakov, *et al.*, *Statistical Radar Theory*, Vol. 1, Sovetskoe Radio, Moscow, 1963 (In Russian).

H. Chernoff, Sequential design of experiments, *Ann. Math. Statist.* **30** (1959) 755–770.

Y.S. Chow and T.L. Lai, Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossing, *Trans. Amer. Math. Soc.*, **208** (1975) 51–72.

E.T. Copson, *Asymptotic Expansions*, University Press, Cambridge, 1965.

V.P. Dragalin, Asymptotic solution of a problem of detecting a signal from k channels, *Russian Math. Surveys*, **42** (1987) 213–214.

V.P. Dragalin and A.A. Novikov, Adaptive sequential tests for composite hypotheses, *Tech. Report* No. 94.4, Institute for Applications of Mathematics and Informatics (Milan, 1994).

T.S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York, London, 1967.

M.M. Fishman, Average duration of asymptotically optimal multi-alternative sequential procedure for recognition of processes, *Soviet Journ. Communic. Technol. Electron.*, **30** (1987) 2541–2548.

G.K. Golubev and R.Z. Khas'minskii, Sequential testing for several signals in Gaussian white noise, *Theory Prob. Appl.* **28** (1983) 573–584.

B.K. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Reading, 1970.

P.L. Hsu and H. Robbins, Complete convergence and the law of large numbers, *Proc. Nat. Acad. Sci. U.S.A.*, **33** (1947) 25–31.

J. Kiefer and J. Sacks, Asymptotically optimal sequential inference and design, *Ann. Math. Statist.*, **34** (1963) 705–750.

T.L. Lai, Termination, moments and exponential boundedness of the stopping rule for certain invariant sequential probability ratio tests, *Ann. Statist.*, **3** (1975) 581–598.

T.L. Lai, On r-quick convergence and a conjecture of Strassen, *Ann. Probability*, **4** (1976) 612–627.

T.L. Lai, Asymptotic optimality of invariant sequential probability ratio tests, *Ann. Statist.*, **9** (1981) 318–333.

E.L. Lehmann, *Testing Statistical Hypotheses*, John Wiley & Sons, New York, 1986.

Yu.V. Linnik, *Statistical Problems with Nuisance Parameters*, American Math. Soc., Providence, RI, 1968.

R.Sh. Liptser and A.N. Shiryaev, *Statistics of Random Processes*, Vol. 1, Springer-Verlag, New York, 1977.

G. Lorden, Integrated risk of asymptotically Bayes sequential tests, *Ann. Math. Statist.*, **38** (1967) 1399–1422.

G. Lorden, Nearly-optimal sequential tests for finitely many parameter values, *Ann. Statist.*, **5** (1977) 1–21.

E. Lukacs, *Stochastic Convergence*, Academic Press, New York, 1975.

I.V. Pavlov, A sequential decision rule for the case of many composite hypotheses, *Engineering Cybernetics*, **22** (1984) 19–23.

I.V. Pavlov, Sequential procedure of testing composite hypotheses with applications to the Kiefer-Weiss problem, *Theory Prob. Appl.* **35** (1990) 280–292.

Yu.G. Sosulin, *Theory of Detection and Estimation of Stochastic Signals*, Sovetskoe Radio, Moscow, 1978 (In Russian).

Yu.G. Sosulin and M.M. Fishman, *Theory of Sequential Decisions and Its Applications*, Radio i Svyaz', Moscow, 1985 (In Russian).

V. Strassen, Almost sure behavior of sums of independent random variables and martingales, *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, **2** (1967) 315–343.

A.F. Taraskin, On asymptotic normality of stochastic integrals and estimates of drift coefficients of diffusion processes. In *Mathematical Physics* **8** (1970) 149–163. Naukova Dumka, Kiev (In Russian).

A.G. Tartakovskii, Sequential composite hypothesis testing with dependent non-stationary observations, *Probl. Information Transmis.*, **17** (1981) 18–28.

A.G. Tartakovskii, Sequential testing of many simple hypotheses with dependent observations, *Probl. Information Transmis.*, **24** (1988) 299–309.

A.G. Tartakovsky, *Sequential Methods in the Theory of Information Systems*, Radio i Svyaz', Moscow, 1991 (In Russian).

A.G. Tartakovsky, Minimax-invariant regret solution to the N-sample slippage problem, *Math. Methods Statist.*, **6** (1997) 491–508.

A.G. Tartakovsky, Asymptotically optimal sequential tests for nonhomogeneous processes, *Sequential Analysis*, **17** (1998) 33–62.

N.V. Verdenskaya and A.G. Tartakovskii, Asymptotically optimal sequential testing of multiple hypotheses for nonhomogeneous Gaussian processes in an asymmetric situation, *Theory Prob. Appl.*, **36** (1991) 536–547.

R.A. Wijsman, Exponentially bounded stopping time of sequential probability ratio tests for composite hypotheses, *Ann. Math. Statist.*, **42** (1971) 1859–1869.