TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Robust Sequential Tests for Memoryless Discrimination from Dependent Observations

by E. Geraniotis

TR 91-25

Robust Sequential Tests for Memoryless Discrimination from Dependent Observations

by E. Geraniotis

TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Research support for this report has been partially provided by the Office of Naval Research under Contract N00014-89-J-1375

TR 91-25

ROBUST SEQUENTIAL TESTS FOR MEMORYLESS DISCRIMINATION FROM DEPENDENT OBSERVATIONS

Evaggelos Geraniotis

Copyright c 1991. Evaggelos Geraniotis. All Rights Reserved.

ROBUST SEQUENTIAL TESTS FOR MEMORYLESS DISCRIMINATION FROM DEPENDENT OBSERVATIONS

Evaggelos Geraniotis

Department of Electrical Engineering
and Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

The problem of robust sequential discrimination from two dependent observation sequences with uncertain statistics is addressed. As in Part I ([1]) of this study, which treated asymptotically optimal sequential discrimination for stationary observations characterized by m-dependent or mixing type of dependence, sequential tests based on memoryless nonlinearities are employed. In particular, the sequential tests robustified in this paper employ linear test statistics of the form $S_n = \overline{A}\sum_{i=1}^{n} g(X_i) + \overline{B}n$, where $\{X_i\}_{i=1}^{n}$ is the observation sequence, the coefficients $\overline{A}$ and $\overline{B}$ are selected so that the normalized drifts of S_n are antipodal under the two hypotheses, and the nonlinearity g solves a linear integral equation. As shown in Part I, the performance of these tests is very close to that of the asymptotically optimal memoryless sequential tests when the statistics of the observations are known. The above tests are robustified in terms of the error probabilities and the expected sample numbers under the two hypotheses, for statistical uncertainty determined by 2-alternating capacity classes for the marginal (univariate) pdfs and upper bounds on the correlation coefficients of time-shifts of the observations sequence for the bivariate pdfs. Finally, the robustification of sequential tests based on a test statistic similar to S_n defined above is carried out for detecting a weak-signal in stationary m-dependent or mixing noise with uncertainty in the univariate and bivariate pdfs.

This research was supported in part by the Office of Naval Research under contract N00014-89-J-1375 and in part by the Systems Research Center at the University of Maryland, College Park, through the National Science Foundation's Engineering Research Centers Program: NSF CDR 8803012.

1. Introduction

In Part I of this study [1], we addressed the problem of sequential discrimination between two arbitrary stationary sequences of observations characterized by m-dependent or mixing type conditions. The necessary for the development of memoryless sequential discriminators statistics, namely the marginal and bivariate pdfs of the observations, were assumed to be known. The discriminators derived in [1] employed memoryless nonlinearities and were optimal among the class of such structures. Different types of sequential tests employing linear or quadratic test statistics were considered and a minimization of the expected sample numbers of these tests under the two hypotheses for fixed desirable error probabilities was carried out with respect to the coefficients of the test statistics and the nonlinearities. The performance of the various sequential tests and nonlinearities derived was evaluated via simulation for several situations of practical interest encountered in radar discrimination and involving envelope observations with ρ-mixing dependence.

Before discussing the problem addressed here and the contributions made by this paper (which constitutes Part II of this study) we summarize the most relevant conclusions from Part I (see [1]). These are the following:

(a) The sequential discriminators employing a linear test statistic of the form $S_n = \bar{A}T_n + \bar{B}n$, where $T_n = \sum_{i=1}^{n} g(X_i)$, $\{X_i\}_{i=1}^{n}$ is the observation sequence, $\bar{A}$ and $\bar{B}$ are coefficients selected so that the normalized drifts of S_n under the two hypotheses are antipodal, and the nonlinearity g solves a linear integral equation, which depends on the marginal and bivariate pdfs of the observation sequences under the two hypotheses (refer to equation (59) of [1]), perform only slightly worse than the discriminators employing quadratic tests and nonlinearities g solving the appropriate nonlinear integral equation (refer to equation (62) of [1]). The memoryless discriminators with linear test statistics are easy to implement; we only need to solve a linear integral equation, which is easily accomplished via discretization and reduction to a linear system of equations, as discussed in Section 4 of [1], to obtain the optimal nonlinearity, and then form the sum T_n and the test statistic S_n in a straightforward manner. In this paper, it will be established that they are also amenable to robustification. The memoryless discriminators with quadratic test statistics are optimal within the class of memoryless sequential discriminators structures, because the quadratic processing that follows the formation of the sum T_n is asymptotically optimal, as it corresponds to the likelihood ratio. Therefore, performance is compromised very little, if one uses the sequential discriminators based on linear test statistics and the nonlinearity solving a linear integral equation.

(b) The sequential discriminators described in (a) provide significant gains in performance (meaning that they achieve the same discrimination reliability faster, with fewer samples) when compared to the i.i.d. sequential discriminators or to the block memoryless discriminators of [2] (the companion paper to Parts I and II) for identical desirable error probabilities. This justifies our recommendation for their use in situations characterized by m-dependent or mixing types of dependence and our interest in robustifying them for situations characterized by statistical uncertainty.

In this part of this study, we robustify the sequential discriminators above against uncertainty in the marginal and the bivariate pdfs. The robustification may be necessary for many situations of practical interest, in which the statistics of the observations are unknown or at best partially known. The literature on the subject contains a considerable amount of research in robust signal processing as attested by the references cited in the tutorial of [3]. However, most of the work on robust detection is concerned with fixed-sample-size (or block) schemes. The work of [4] constitutes of an exception; it considers robust detection of weak signals in additive i.i.d. noise for uncertainty in the noise pdf within p-point classes.

This paper makes a twofold contribution. On the one hand, it robustifes the sequential discriminators of [1], which employ linear test statistics and nonlinearities solving linear integral equations for two arbitrary stationary sequences of observations with m-dependent or mixing type of dependence and uncertainty in the marginal and bivariate pdfs. Emphasis is placed on situations that can not be characterized as weak signals in additive dependent noise. As will be shown in the following section, the uncertainty class for the marginal pdfs is determined by 2-alternating capacities; this is a very general model that includes several popular uncertainty models as subcases. The uncertainty for the bivariates is determined by bounds on the correlation coefficients between time-shifts of the observation sequences. On the other hand, this paper derives robust sequential detectors for weak signals in additive m-dependent or mixing noise. Here the uncertainty on the marginal pdfs is of the ε-contaminated or total variation type, whereas the uncertainty in the bivariate pdfs is determined by bounds on the correlation coefficients between time-shifts of the noise sequence.

Consequently, the first part of the paper, which is concerned with the robustification of sequential memoryless discrimination schemes, extends naturally the work of [1] and [2] that dealt with optimal memoryless sequential and block discrimination schemes, respectively, for known observation statistics. The second part of the paper, which is concerned with robust sequential memoryless schemes for the detection of weak signals, extends the results of [5] and [6] for the robust fixed-sample-size detection of weak signals in additive dependent noise to sequential detection schemes, while at the same time extending the results of [4] for the sequential detection of weak-signals in i.i.d. noise to the case of dependent noise. Although [4] deals with a detection problem, the weak signal is first estimated using Huber's M-estimators and then the estimate is used to form a likelihood ratio, on which a probability ratio test (SPRT) with Wald's thresholds is performed. In this context, since combined estimation and detection are used, the process is somewhat complicated: not only the nonlinearity used in the estimator needs to be derived, but also the estimate needs to be computed from the n observations for each step of the SPRT, as the latter progresses. The sequential detector proposed in this paper does not rely on any estimation process; the task is accomplished with a simpler structure. This comparison will be elaborated upon in the following.

Our approach is that of **minimax robustness**, according to which we derive sequential discrimination schemes that guarantee a desirable level of performance in terms of the error probabilities (false alarm and miss) and the expected sample numbers under the two hypotheses for the least-favorable elements in the uncertainty classes (i.e., for marginal distributions in capacity classes and bivariate distributions satisfying the bounds on the correlation coefficients) and show that, for any other elements in these classes, the performance of the robust sequential schemes is superior.

This paper is organized as follows. In Sections 2 and 3, we develop and analyze robust sequential memoryless schemes for the cases of general discrimination from two arbitrary stationary dependent observations sequences and of the detection of a weak signal in stationary dependent noise, respectively. In each of these two Sections, we first introduce the necessary notation and the uncertainty classes for the marginal and bivariate pdfs of the observation sequences or the noise sequence; then we derive expressions for the error probabilities and the expected sample numbers under mismatch for the sequential test employed; finally, we derive the robust sequential memoryless discriminators or detectors for the uncertainty calsses of interest. In Section 4, a summary of the paper and conclusions are presented.

2. Robust Sequential Memoryless Discriminators

2.1 Preliminaries

As in [1] and [2] the general hypothesis testing problem of this paper is formulated as the need to discriminate between the two hypotheses

$$H_k:\ \underline{X} \sim f_k^{(n)} \quad \text{for} \quad k = 0, 1, \tag{1}$$

where $\underline{X} = (X_1, X_2, \ldots, X_n)$ denotes the vector of n dependent observations and $f_k^{(n)}(\underline{X})$ the n-dimensional joint pdf of $\underline{X}$. For many situations of practical interest characterized by dependence and non-Gaussian statistics $f_k^{(n)}$ is hard or even impossible to obtain in closed form, as density estimation in n dimensions can be a truly formidable task. Therefore, we resort to models of dependence that are as non-restictive as possible and at the same time make the analysis of discrimination schemes possible.

In [1] and [2], various models of dependence were reviewed. Here we cite only the most relevant definitions so that we can introduce the necessary notation and make the presentation in this paper self-sufficient. The simplest model of dependence assumes that, under both hypotheses, the observations are **stationary** and **m-dependent**, meaning that (see [13]) the stationary data X_l and $X_{l'}$ are correlated with known correlation for $|l-l'| \le m_k$ and independent for $|l-l'| > m_k$, under hypothesis H_k. The least restrictive dependence model of interest is that of **ρ-mixing** which is characterized by

$$cov_k\{X,Y\} \le \rho_{k,n} \tag{2a}$$

for all real $X \in L_2(A)$ and $Y \in L_2(B)$. X and Y are random variables measurable with respect to A and B, respectively, where A is an event from X_1^l, the latter being the σ-algebra generated by the random variables $\{X_1, X_2, \ldots, X_l\}$, and B an event from X_{l+n}^{∞}, which is the σ-algebra generated by the random variables $\{X_{l+n}, X_{l+n+1}, \ldots\}$. $\rho_{k,n}$ are sequences of real numbers, such that $\rho_{k,n} \to 0$, as $n \to \infty$, for $k = 0, 1$. Equation (2a) implies the weaker but more intuitive result

$$cov_k\{X_l, X_{l+n}\} \le \rho_{k,n} \tag{2b}$$

and represents a good model for a time series of data that are asymptotically uncorrelated.

The main component of the test statistic for the discriminators of [1] and [2] is of the form $T_n(\underline{X}) = \sum_{i=1}^{n} g(X_i)$, where the number of samples n is large (e.g., $n \gg m_k$) and g is a nonlinearity chosen to maximize a suitable performance measure. The means of $T_n(\underline{X})$ under the two hypothesis H_1 and H_0 are given by $n\mu_1$ and $n\mu_0$, respectively, where

$$\mu_k = E_k\{g(X_1)\} = \int g(x) f_k(x) dx, \quad k = 0, 1 . \tag{3}$$

E_k denotes expectation under hypothesis H_k and $f_k(x)$ is the corresponding marginal density. The asymptotic variance of $T_n(\underline{X})$ under hypothesis H_k is $n\sigma_k^2$, where $\sigma_k^2 = \lim_{n\to\infty} n^{-1} var_k\{T_n\}$ (k=0,1) is given by

$$\begin{aligned}\sigma_k^2(g) &= var_k\{g(X_1)\} + 2\sum_{j=1}^{m_k} cov_k\{g(X_1)g(X_{j+1})\} \\ &= E_k\{g(X_1)^2\} + 2\sum_{j=1}^{m_k} E_k\{g(X_1)g(X_{j+1})\} - (2m+1)[E_k\{g(X_1)\}]^2 \end{aligned} \tag{4a}$$

for m-dependent observations, and by

$$\sigma_k^2(g) = E_k\{g(X_1)^2\} - [E_k\{g(X_1)\}]^2 + 2\sum_{j=1}^{\infty}\left[E_k\{g(X_1)g(X_{j+1})\} - [E_k\{g(X_1)\}]^2\right] \tag{4b}$$

for ρ-mixing observations.

In addition, $T_n(\underline{X})$ is **asymptotically Gaussian distributed** under hypothesis H_k with mean $n\mu_k$ and variance $n\sigma_k^2$ cited above, provided that $\sigma_k^2 > 0$ and some other conditions hold. This follows from the validity of the Central Limit Theorem (CLT) for dependent observations (see [13] and particularly the tutorial in [14], which provides CLTs for various mixing types of stationary observations). For example, in the ρ-mixing case, the condition for a CLT to hold is that the variance in (4b) is $\sigma_k^2 > 0$ and $\sum_{n=0}^{\infty} \rho_{k,2^n} < \infty$ (see [12]). Actually, the existence and validity of Central Limit Theorems for quantities like T_n formed from the dependent observations of (1) constitutes the basis for the remainder of this paper.

The sequential test (SPRT) to be robustified is based on the linear test statistic of (33a) of [1], which employs the nonlinearity solving the linear integral equation of (59)-(60) of [1]. The reasons for this choice of test statistic and nonlinearity have already been discussed in Section 1 of this paper and in Section 5 of Part I [1]. The test statistic of interest is expressed as

$$\hat{S}_n = \frac{2(\hat{\mu}_1-\hat{\mu}_0)}{\hat{\sigma}_1^2+\hat{\sigma}_0^2}\left[\sum_{i=1}^{n}\hat{g}(X_i) - \frac{\hat{\mu}_1\hat{\sigma}_0^2+\hat{\mu}_0\hat{\sigma}_1^2}{\hat{\sigma}_1^2+\hat{\sigma}_0^2}n\right] \tag{5}$$

and it is compared to Wald's thresholds

$$\hat{a} = ln\frac{\hat{\beta}}{1-\hat{\alpha}} < 0 \tag{6a}$$

and

$$\hat{b} = ln\frac{1-\hat{\beta}}{\hat{\alpha}} > 0 \tag{6b}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the desirable error probabilities for the SPRT. In (5), the means $\hat{\mu}_k$ and variances $\hat{\sigma}_k^2$, for $k = 0, 1$, are obtained from (3) and (4a)-(4b), respectively, upon substitution for the nonlinearity $\hat{g}$ and the marginal and bivariate pdfs $\hat{f}_k$ and $\{\hat{f}_k^{(1,j+1)}\}_{j=1}^{\infty}$. These pdfs (i) may represent estimates of the statistics that govern the observation sequences under the two hypotheses ($k = 0, 1$) and thus could be different from the actual statistics of the observations, or (ii) they may be chosen to characterize the least-favorable conditions for the operation of the test of (5) within certain uncertainty classes (as will be done in Section 2.2.) below. For notational convenience, we use $\hat{f}_k^{(j)}(x,y)$, instead of $\hat{f}_k^{(1,j+1)}(x,y)$, heretoforth. We assume that these marginal and bivariate pdfs, from which $\hat{g}$ is determined, always exist and the corresponding distributions (cdfs) are denoted by $\hat{F}_k$ and $\hat{F}_k^{(j)}$. We denote by $\hat{F}_k^*$ the pair ($\hat{F}_k$, $\{\hat{F}_k^{(j)}\}_{j=1}^{\infty}$).

The linear integral equation that $\hat{g}$ solves was derived in [1] but is cited here for the sake of completeness:

$$\frac{\hat{f}_1(x) - \hat{f}_0(x)}{w(\hat{\alpha},\hat{\beta})\hat{f}_1(x) + w(\hat{\beta},\hat{\alpha})\hat{f}_0(x)} - \int \overline{K}(x,y)\hat{g}(y)dy = \hat{g}(x) \tag{7}$$

where

$$\overline{K}(x,y) = \frac{w(\hat{\alpha},\hat{\beta})\hat{K}_1(x,y) + w(\hat{\beta},\hat{\alpha})\hat{K}_0(x,y)}{w(\hat{\alpha},\hat{\beta})\hat{f}_1(x) + w(\hat{\beta},\hat{\alpha})\hat{f}_0(x)} \; . \tag{8}$$

The kernels $\hat{K}_k(x,y)$ for $k = 0$, 1 are defined as

$$\hat{K}_k(x,y) = \sum_{j=1}^{m_k} \left[\hat{f}_k^{(j)}(x,y) + \hat{f}_k^{(j)}(y,x) - 2\hat{f}_k(x)\hat{f}_k(y)\right] - \hat{f}_k(x)\hat{f}_k(y) \tag{9}$$

for m-dependent observations; m_k should be replaced by ∞ for ρ-mixing observations. The function $w(x,y)$ is defined as

$$w(x,y) = (1-x)\; ln\frac{1-x}{y} + x\; ln\frac{x}{1-y} \tag{10}$$

and, as shown in [1] for desired error probabilities $\hat{\alpha}$ and $\hat{\beta}$, $w(\hat{\alpha},\hat{\beta}) > 0$ and $w(\hat{\beta},\hat{\alpha}) > 0$.

The operating conditions of the above test statistic are determined for our analysis, which involves only first and second order statistics, by the actual marginal $F_k(x)$ and bivariate $\{F_k^{(j)}\}_{j=1}^{\infty}$ distributions ($k = 0$, 1) of the observations, which are generally different from the ones involved in (5) and (7)-(9). This situation is called **mismatch** and plays an important role in the robustification of the test in (5). The above distributions may or may not have densities; when they have pdfs, these are denoted by f_k and $f_k^{(j)}$. We use F_k^* to denote collectively the pair $(F_k, \{F_k^{(j)}\}_{j=1}^{\infty})$ of the actual cdfs of the observations under hypothesis H_k $(k = 0, 1)$. Clearly, the asymptotic means of the test statistic $\hat{S}_n$ under H_k depend on F_k and its asymptotic variances on F_k^*.

We now describe the uncertainty classes, to which belong the marginal and bivariates pdfs necessary for solving (7) and obtaining $\hat{g}$, as well as the pdfs characterizing the operating conditions of the test statistic in (5). The uncertainty classes are identical to those considered in [2] for the robustification of block memoryless discriminators. These classes constitute an extension to those considered in [5] and [6] for memoryless block discrimination, in that the classes for the marginals pdfs treated in this paper and in [1]-[2] are broader. Specifically, the **marginal pdfs** are assumed to belong to uncertainty classes determined by 2-alternating capacities, also termed Huber-Strassen classes. These classes include many popular models of uncertainty, such as the ε-contaminated classes (see [7]), the total variation classes (see [7]), the band classes (see [8]), and the p-point classes (see [9]). These classes are characterized by either a degree of deviation from a nominal (known) pdf or by known upper and lower bounds (confidence limits) on the members of the class. They can be considered as special cases of a general uncertainty model involving a capacity as the upper measure of each specific class. Generalized capacities [10] can also be considered in this context. The basic theory of minimax robustness for these was developed in [11]. The least favorable elements with respect to the Bayes risk of these classes have been identified in closed form for each one of the four uncertainty classes enumerated above. In Appendix A, we review the most relevant to our problem results of this theory and provide a complete example based on the ε-contaminated model. We assume that the nominal distributions determining the uncertainty classes of (A-1) have densities (pdfs) and so do the least-favorable distributions singled out by Lemma 1 of Appendix A.

As in [5] and [6], the **bivariate pdfs** are assumed to belong to classes determined by bounds on the correlation coefficients between time-shifts of the observation sequence (assumed to be ρ-mixing in the less restrictive case), then

$$\sup_g \frac{|cov_k\{g(X_1)g(X_{1+j})\}|}{var_k\{g(X_1)\}var_k\{g(X_{1+j})\}} \leq r_{k,j} \tag{11}$$

where g ranges over all measurable functions satisfying $E_k\{g^2(X_1)\} < \infty$ under hypothesis H_k, for $k = 0, 1$. Since we assume stationarity, the denominator in (11) is $var_k\{g(X_1)\}$. The parameters $r_{k,j}$ can be obtained from the parameters $\rho_{k,j}$ of the ρ-mixing process $\{X_i\}_{i=1}^{\infty}$ under H_k. As proved in Proposition 6 of [2], for processes $\{X_i\}_{i=1}^{\infty}$ with bivariate distributions having diagonal expansions involving an orthonormal set of polynomials, the supremum of the correlation coefficient of the process $\{g(X_i)\}_{i=1}^{\infty}$ in (11) can be directly related to the correlation coefficient and the moments up to order four of the original process $\{X_i\}_{i=1}^{\infty}$. Examples of processes with such expansions are those with Gausssian or Gamma distributions and processes obtained from them via memoryless transformations.

For a given marginal distribution F_k, equality holds in (11) for all g, if the bivariate distribution function is

$$F_k^{(j)}(x,y) = (1 - r_{k,j})F_k(x)F_k(y) + r_{k,j}F_k(x \Lambda y) \tag{12a}$$

where $x \Lambda y$ is the minimum of x and y. If F_k has a density f_k, then we may write for the bivariate pdf

$$f_k^{(j)}(x,y) = (1 - r_{k,j})f_k(x)f_k(y) + r_{k,j}f_k(x)\delta(x-y) \tag{12b}$$

where $\delta(x)$ is Dirac's δ function. In [6] it was shown by two constructions that there exist processes with bivariate distributions given by (12). Notice that, if the condition (11) is satisfied, then (4a) and (4b) imply

$$\sigma_k^2(g) \le (1 + 2R_k)\ var_k\{g(X_1)\} \tag{13}$$

with $R_k = \sum_{j=1}^{m_k} r_{k,j}$, where for the m-dependent model, m_k is the number of samples it takes the signal to decorrelate (meaning that $cov_k\{X_1, X_{j+1}\} = 0$ for $j > m_k$) under hypothesis H_k; for the ρ-mixing model, $m_k = \infty$. The equality is achieved in (13) for the cdfs defined by (12) and thus (12) has maximum variance among all cdfs in the class defined by (11). In this formulation, it turns out that the value of the sum R_k, rather than the individual terms of the sum, are relevant to the robustification that follows.

2.2 Robustification of Sequential Memoryless Nonlinear Discriminators

Before robustifying the performance measures of interest, we establish the following result, which provides the error probabilities and the expected sample numbers of the sequential test under mismatch and is used extensively in the sequel.

Proposition 1: Let $\overline{P}_k(\hat{g}, F_k^*)$ denote the probability of error under mismatch and $E_k\{N \mid \hat{g}, F_k)\}$ the required average sample number, when hypothesis H_k ($k = 0, 1$) is true. Let us assume that the sequential test of (5), with thresholds $\hat{a}$ and $\hat{b}$ defined by (6a)-(6b) for desired error probabilities $\hat{\alpha}$ and $\hat{\beta}$, is employed. Then the following identities hold

$$\overline{P}_0(\hat{g}, F_0^*) = \alpha = P_0\{\hat{S}_N \text{ up-crosses } \hat{b} \text{ before it down-crosses } \hat{a}\} = \frac{1-(e^{\hat{a}})^{\frac{2\mu_0(\hat{g}, F_0)}{\sigma_0^2(\hat{g}, F_0^*)}}}{1-(e^{\hat{a}-\hat{b}})^{\frac{2\mu_0(\hat{g}, F_0)}{\sigma_0^2(\hat{g}, F_0^*)}}} \tag{14a}$$

$$\overline{P}_1(\hat{g}, F_1^*) = \beta = P_1\{\hat{S}_N \text{ down-crosses } \hat{a} \text{ before it up-crosses } \hat{b}\} = \frac{1-(e^{\hat{b}})^{\frac{2\mu_1(\hat{g}, F_1)}{\sigma_1^2(\hat{g}, F_1^*)}}}{1-(e^{\hat{b}-\hat{a}})^{\frac{2\mu_1(\hat{g}, F_1)}{\sigma_1^2(\hat{g}, F_1^*)}}} \tag{14b}$$

and

$$E_0\{N \mid \hat{g}, F_0)\} = \frac{\omega(\hat{\alpha},\hat{\beta};\alpha)}{-\bar{\mu}_0(\hat{g},F_0)} \tag{15a}$$

$$E_1\{N \mid \hat{g}, F_1)\} = \frac{\omega(\hat{\beta},\hat{\alpha};\beta)}{\bar{\mu}_1(\hat{g},F_1)} \tag{15b}$$

where

$$\bar{\mu}_k(\hat{g},F_k) = \frac{2(\hat{\mu}_1-\hat{\mu}_0)}{\hat{\sigma}_1^2+\hat{\sigma}_0^2}\left[\mu(\hat{g},F_k) - \frac{\hat{\mu}_1\hat{\sigma}_0^2+\hat{\mu}_0\hat{\sigma}_1^2}{\hat{\sigma}_1^2+\hat{\sigma}_0^2}\right] \tag{16}$$

and

$$\bar{\sigma}_k^2(\hat{g},F_k^*) = \frac{4(\hat{\mu}_1-\hat{\mu}_0)^2}{(\hat{\sigma}_1^2+\hat{\sigma}_0^2)^2}\sigma^2(\hat{g},F_k^*) \tag{17}$$

for $k = 0, 1$. In (16)-(17), $\mu(\hat{g},F_k) = \int \hat{g}(x)dF_k(x) = \lim_{n\to\infty} n^{-1}E_k\{\hat{T}_n\}$ denotes the asymptotic mean and $\sigma^2(\hat{g},F_k^*) = \lim_{n\to\infty} n^{-1}var_k\{\hat{T}_n\}$ the asymptotic variance [obtained from (4a) or (4b) for $\hat{g}$ and marginal/bivariate pair F_k^*] under mismatch of $\hat{T}_n = \sum_{i=1}^{n}\hat{g}(X_i)$, when hypothesis H_k is true. The corresponding means and variances under matched conditions are denoted by $\hat{\mu}_k = \mu(\hat{g},\hat{F}_k) = \mu(\hat{g},\hat{f}_k)$ and $\hat{\sigma}_k^2 = \sigma^2(\hat{g},\hat{F}_k^*) = \sigma^2(\hat{g},\hat{f}_k^*)$. In this case, the pdfs are assumed to exist and have already been used in the definition of the test statistic in (5). The rest of the quantities involved in (14)-(15) are the error probabilities under matched conditions

$$\hat{\alpha} = \hat{P}_0\{\hat{S}_N \geq \hat{b}\} \tag{18a}$$

and

$$\hat{\beta} = \hat{P}_1\{\hat{S}_N \leq \hat{a}\} \tag{18b}$$

and the quantity

$$\omega(\hat{x},\hat{y};x) = (1-x)\ln\frac{1-\hat{x}}{\hat{y}} + x\ln\frac{\hat{x}}{1-\hat{y}} \tag{19}$$

which, under matched conditions, reduces to $\omega(\hat{x},\hat{y};\hat{x}) = w(\hat{x},\hat{y})$ defined by (10).

Remark 1: Equations (14a)-(14b) and (15a)-(15b) are approximations which become tight, when the desired probabilities of false alarm and miss are sufficiently small, so that a large number of samples is required to achieve the desired reliability. If N (the required sample number) is large under both hypotheses, then (i) the overshoot phenomenon present in Wald's approximations can be neglected and (ii) the diffusion (Brownian motion) approximation used in the computation of the means and variances of the test statistic becomes accurate.

Remark 2: The numerators of (15a)-(15b) are usually positive for all situations of interest and so are the denominators. These facts are established in Proposition 3 below.

Proof: Expressions (14a)-(14b) were obtained following similar steps, as for deriving (25a)-(25b) of [1] (Part I), essentially by applying the Brownian motion approximation to the linear test statistic of (5) operating under mismatch conditions. The expressions in (16) and (17) actually represent the drift and the variance of the diffusion, i.e., for large n

$$E_k\{\hat{S}_n\} \cong n\overline{\mu}_k(\hat{g},F_k)$$

and

$$var_k\{\hat{S}_n\} \cong n\overline{\sigma}_k^2(\hat{g},F_k^*) .$$

Furthermore, the expressions (15a)-(15b) can be obtained in a manner similar to that used for deriving equations (15a)-(15b) and (18a)-(18b) of [1]. Specifically, by neglecting the overshoot phenomenon and using Wald's approximations we obtain that under mismatch

$$E_0\{\hat{S}_N\} \cong -\omega(\hat{\alpha},\hat{\beta};\alpha)$$

and

$$E_1\{\hat{S}_N\} \cong \omega(\hat{\beta},\hat{\alpha};\beta) .$$

Moreover, we can easily show that, for N taking large values,

$$E_k\{\hat{S}_N\} \cong \overline{\mu}_k(\hat{g},f_k)E_k\{N\}$$

Then (15a)-(15b) follows from a combination of the above equations.

Next we establish the first of the two main results of this section pertaining to the error probabilities of the sequential test of (5).

Proposition 2: Consider the sequential test which is based on the thresholds $\hat{a}$ and $\hat{b}$, obtained from (6a)-(6b) for desired error probabilities $\hat{\alpha}$ and $\hat{\beta}$, and on the linear test statistic of (5), which employs the nonlinearity $\hat{g}$ solving (7)-(8) with the kernels of (9) modified according to (12b) as

$$\hat{K}_k(x,y) = 2R_k\hat{f}_k(x)\delta(x-y) - (1+2R_k)\hat{f}_k(x)\hat{f}_k(y) \tag{20}$$

where $\hat{F}_k$ ($\hat{f}_k$) are the cdfs (pdfs) singled out by Lemma 1 of Appendix A for the capacity class of (A-1), and $\hat{F}_k^{(j)}$ satisfy (12a), for all j and $k = 0, 1$. Then this test is **least-favorable for the error probabilities under the two hypotheses**, that is,

$$\overline{P}_k(\hat{g},F_k^*) \leq \overline{P}_k(\hat{g},\hat{F}_k^*) \tag{21}$$

for any $F_k^* = (F_k,\{F_k^{(j)}\}_{j=1}^{\infty})$ with F_k in the capacity class $\boldsymbol{F}_k$ of (A-1) and $F_k^{(j)}$ satisfying (11). In (21), $\overline{P}_k(\hat{g},F_k^*)$, the error probabilities under mismatch, are as defined in Proposition 1 by (14a)-(14b).

Remark 3: This Proposition holds under the same conditions that Proposition 1. The issues raised by Remark 1 about the accuracy of Wald's approximations and the Brownian-motion (diffusion) approximation are also valid here. (21) can be expressed as $\alpha \leq \hat{\alpha}$ and $\beta \leq \hat{\beta}$ in the notation of Proposition 1. The result in (21) is valid under the assumptions (A4) stated in Remark 4 below.

Proof: Proving (21) requires several steps. We start by using the fact that, since $\hat{F}_k^{(j)}$ satisfies

(12a), the equality in (13) is achieved, and we can write

$$\hat{\sigma}_k^2 = \sigma^2(\hat{g},\hat{F}_k^*) = (1 + 2R_k)\sigma^2(\hat{g},\hat{F}_k) \ . \tag{22}$$

The quantity $\sigma^2(\hat{g},\hat{F}_k^*)$ defined after (17) (for matched conditions) now depends only on the marginal cdf $\hat{F}_k$; this is equivalent to removing the dependence from the observations sequence and modifying the kernels of (9) according to (12b), so that they are given by (20). We return to this important point later in this proof. We use (22) to define

$$P_k(\hat{g},\hat{F}_k) = \overline{P}_k(\hat{g},\hat{F}_k^*) \tag{23}$$

where $P_k(\hat{g},\hat{F}_k)$, for $k = 0, 1$, is obtained from (14a)-(14b) by using (22) in (17) with $\hat{F}_k$ replacing F_k. The left-hand-side in (23) depends only on the univariate (marginal) distribution $\hat{F}_k$. Furthermore, because of (13)

$$\sigma^2(\hat{g},F_k^*) \leq (1 + 2R_k)\sigma^2(\hat{g},F_k) \tag{24}$$

for any F_k^* with bivariates $F_k^{(j)}$ satisfying (11), for $j = 1, 2, \cdots$, and arbitrary marginals F_k. Therefore, since $\overline{P}_k(\hat{g},F_k^*)$ is an increasing function of $\overline{\sigma}_k^2(\hat{g},F_k^*)$ given by (17) and the latter is an increasing function of $\sigma^2(\hat{g},F_k^*)$, which is the left-hand-side of (24), we obtain

$$\overline{P}_k(\hat{g},F_k^*) \leq P_k(\hat{g},F_k) \ . \tag{25}$$

In (25), $P_k(\hat{g},F_k)$, for $k = 0, 1$, can be obtained from (14a)-(14b) by using the right-hand-side member of (24), for $\sigma^2(\hat{g},F_k^*)$ in (17). The right-hand-side of (25) now depends only on the marginal cdf F_k.

Upon substitution from (23) and (25) in (21), we find that (21) is valid, if the following inequality holds

$$P_k(\hat{g},F_k) \leq P_k(\hat{g},\hat{F}_k) \tag{26}$$

for all the marginal cdfs F_k in the class $\boldsymbol{F_k}$ given by (A-1) and $\hat{F}_k$ singled out by Lemma 1 of

Appendix A. This inequality corresponds to the least-favorability condition for the new error probabilities $P_k(\hat{g},F_k)$ obtained from the original ones $\overline{P}_k(\hat{g},F_k^*)$ by removing the dependence of the observations sequence through the use of the bounds of (11).

To prove the inequality in (26), we use some of the results on minimax robustness reviewed in Appendix A. Several steps are involved in the proof. First we exploit the fact that the mismatch error probability of (14a) is an increasing function of the normalized drift c_0, whereas (14b) is a decreasing function of the drift c_1, defined as $c_k = 2\overline{\mu}_k(\hat{g},F_k) \,/\, \overline{\sigma}_k^2(\hat{g},F_k^*)$ for $k = 0, 1$, and, for the worst-case, as $\hat{c}_k = 2\overline{\mu}_k(\hat{g},\hat{F}_k) \,/\, \overline{\sigma}_k^2(\hat{g},\hat{F}_k^*)$. Removing the dependence in the observations through the bounds of (11) implies that $\hat{c}_0 = -1$ and $\hat{c}_1 = 1$. Using this and the definitions (16)-(17) we establish that the inequalities of (26), for $k = 0, 1$, are equivalent to the following inequalities characterizing the mismatch and matched worst-case situations:

$$c_0 = \frac{[\mu(\hat{g},F_0)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},\hat{F}_1)+[\mu(\hat{g},F_0)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0)}{[\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_0)\sigma^2(\hat{g},F_0)} \leq -1 = \hat{c}_0 \quad (27a)$$

and

$$c_1 = \frac{[\mu(\hat{g},F_1)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0)+[\mu(\hat{g},F_1)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},\hat{F}_1)}{[\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},F_1)} \geq 1 = \hat{c}_1 \quad (27b)$$

In (27a)-(27b), the means and variances involved depend only on the marginal cdfs $\hat{F}_k$ and F_k, as has already been shown above. In particular, both the matched worst-case variances $\hat{\sigma}_k^2$ and the mismatch variances σ_k^2 involve the same factor $(1 + 2R_k)$ after the dependence in the observations is removed in the expressions for $\overline{P}_k(\hat{g},F_k^*)$ and subsequently in the expressions for c_k. Under the assumption that

$$\mu(\hat{g},\hat{F}_1) \geq \mu(\hat{g},\hat{F}_0) \quad (A1)$$

the conditions

$$\mu(\hat{g},F_0) \le \mu(\hat{g},\hat{F}_0) \quad \text{and} \quad \sigma^2(\hat{g},\hat{F}_0) \ge \sigma^2(\hat{g},F_0) \tag{28a}$$

are sufficient for (27a). To prove this notice that, if (28a) holds, then

$$[\mu(\hat{g},F_0)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},\hat{F}_1)+[\mu(\hat{g},F_0)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},F_0)$$
$$\le [\mu(\hat{g},F_0)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0) \le [\mu(\hat{g},\hat{F}_0)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0)$$
$$= -[\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0) \le -[\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_0)\sigma^2(\hat{g},F_0)$$

and thus $c_0 \le -1 = \hat{c}_0$. Similarly, the conditions

$$\mu(\hat{g},F_1) \ge \mu(\hat{g},\hat{F}_1) \quad \text{and} \quad \sigma^2(\hat{g},\hat{F}_1) \ge \sigma^2(\hat{g},F_1) \tag{28b}$$

are sufficient for (27b). To prove this notice that, if (28b) holds, then

$$[\mu(\hat{g},F_1)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0)+[\mu(\hat{g},F_1)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},\hat{F}_1)$$
$$\ge [\mu(\hat{g},F_1)-\mu(\hat{g},\hat{F}_1)](1+2R_0)\sigma^2(\hat{g},\hat{F}_0)+[\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},\hat{F}_1)$$
$$\ge [\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},\hat{F}_1) \ge [\mu(\hat{g},\hat{F}_1)-\mu(\hat{g},\hat{F}_0)](1+2R_1)\sigma^2(\hat{g},F_1)$$

and thus $c_1 \ge 1 = \hat{c}_1$.

Now we show that conditions (28a)-(28b) are satisfied for the $\hat{g}$ that solves the linear integral equation (7), after the removal of dependence in the observations through the bounds of (11). As already discussed, we substitute $\hat{f}_k^{(j)}$ from (12b) for the $\hat{f}_k^{(j)}$ in (9) to obtain the kernels $\hat{K}_k$ in (20); after some further manipulations, the linear integral equation (7) becomes

$$\frac{\hat{f}_1(x)-\hat{f}_0(x)}{w(\hat{\alpha},\hat{\beta})(1+2R_1)\hat{f}_1(x)+w(\hat{\beta},\hat{\alpha})(1+2R_0)\hat{f}_0(x)}$$
$$= \hat{g}(x) - \int\left[\frac{w(\hat{\alpha},\hat{\beta})(1+2R_1)\hat{f}_1(x)\hat{f}_1(y)+w(\hat{\beta},\hat{\alpha})(1+2R_0)\hat{f}_0(x)\hat{f}_0(y)}{w(\hat{\alpha},\hat{\beta})(1+2R_1)\hat{f}_1(x)+w(\hat{\beta},\hat{\alpha})(1+2R_0)\hat{f}_0(x)}\right]\hat{g}(y)dy$$

or, equivalently, since $\hat{g}$ scales both members of the integral equation,

$$\hat{g}(x) = \frac{\hat{f}_1(x)-\hat{f}_0(x)}{A\hat{f}_1(x)+\hat{f}_0(x)} + \int\left[\frac{A\hat{f}_1(x)\hat{f}_1(y)+\hat{f}_0(x)\hat{f}_0(y)}{A\hat{f}_1(x)+\hat{f}_0(x)}\right]\hat{g}(y)dy \tag{29}$$

where

$$A = \frac{w(\hat{\alpha},\hat{\beta})(1 + 2R_1)}{w(\hat{\beta},\hat{\alpha})(1 + 2R_0)} \,. \tag{30}$$

An integral equation of the form (29) was solved in [6, Appendix C] for a different A. The solution to (30) is shown to be

$$\hat{g}(x) = \frac{\hat{f}_1(x)}{A\hat{f}_1(x) + \hat{f}_0(x)} = \frac{\pi_v(x)}{A\pi_v(x) + 1} \tag{31}$$

for all $x \in \Omega$ (the sample space), where $\pi_v(x) = \hat{f}_1(x)/\hat{f}_0(x) > 0$ is the Huber-Strassen derivative defined in Lemma 1 of Appendix A of this paper.

In Appendix B, we use the dominance properties (A-3)-(A-4) of Lemma 1 of Appendix A to prove that the sufficient conditions for the minimax robustness of the error probabilities in (26) [namely (28a) and (28b)] are satisfied for the nonlinearity $\hat{g}$ given by (31).

Remark 4: Sufficient conditions for robustification are the assumptions (A2) and (A3) of Appendix B, which can be summarized as

$$\mu(\hat{g},\hat{F}_0) \le 0 \le \mu(\hat{g},\hat{F}_1) \tag{A4-1}$$

where the two equalities are not allowed to hold simultaneously (resulting in $\mu(\hat{g},\hat{F}_1) - \mu(\hat{g},\hat{F}_0) > 0$) and

$$A\,\mu(\hat{g},\hat{F}_1) \ge 1 \,. \tag{A4-2}$$

These conditions are not particularly restrictive for most practical situations. Specifically, A [given by (30)] is typically a relatively large positive number (as is the case for the realistic discrimination scenaria considered in Section 4 of [1]), because, under H_1, the observations are usually more strongly positively correlated than under H_0; this implies that (A4-2) is easily satisfied. Furthermore, (A4-1) is satisfied in most situations in which a good choice of $\hat{g}$ has been made; it represents a good condition for adequate separation of the means $\mu(\hat{g},\hat{F}_k)$ $(k = 0, 1)$ under the two hypotheses and, consequently, for a better performance of the sequential test of (5).

Finally, we prove the second main result of this section, which pertains to the robustification of the expected sample numbers of the sequential memoryless test of (5).

Proposition 3: Suppose that the same sequential test as in Proposition 2 is employed. A notation identical to that of Propositions 1 and 2 is used. Assume that, besides the assumptions (A4) of Remark 4, the following additional assumptions hold:

$$\ln \frac{1-\hat{\beta}}{\hat{\alpha}} \gg \hat{\beta} \ln \frac{(1-\hat{\alpha})(1-\hat{\beta})}{\hat{\alpha}\hat{\beta}} \tag{A5-1}$$

$$\ln \frac{1-\hat{\alpha}}{\hat{\beta}} \gg \hat{\alpha} \ln \frac{(1-\hat{\alpha})(1-\hat{\beta})}{\hat{\alpha}\hat{\beta}} . \tag{A5-2}$$

Then the sequential test of (5) is **minimax robust for the expected sample numbers under the two hypotheses**, that is,

$$E_k\{N \mid \hat{g},F_k\} \le E_k\{N \mid \hat{g},\hat{F}_k\} , \quad \text{for } k = 0, 1 , \tag{32a}$$

and

$$E_1\{N \mid \hat{g},F_1\} + E_0\{N \mid \hat{g},F_0\} \le E_1\{N \mid \hat{g},\hat{F}_1\} + E_0\{N \mid \hat{g},\hat{F}_0\} \le E_1\{N \mid g,\hat{F}_1\} + E_0\{N \mid g,\hat{F}_0\} \tag{32b}$$

for all marginal cdfs F_k in the capacity class $\boldsymbol{F}_k$ of (A-1) with bivariates $F_k^{(j)}$ satisfying (11) and any measurable function g satisfying $E_k\{g^2(X_1)\} < \infty$. $\hat{F}_k$ $(\hat{f}_k)$ is the cdf (pdf) singled out by Lemma 1 of Appendix A with bivariates $\hat{F}_k^{(j)}$ satisfying (12a), for all j. The expected sample numbers under mismatch $E_k\{N \mid \hat{g},F_k\}$ are as defined by (15a)-(15b) of Proposition 1.

Remark 5: The assumptions (A5) are not so restrictive, since they can be easily satisfied, if both $\hat{\alpha}$ and $\hat{\beta}$ (the desirable error probabilities under worst-case conditions) are smaller than

10^{-3}. Therefore, our results are valid for sequential tests with high reliability; recall that, according to Remarks 1 and 3, the sufficiently small probabilities of false alarm and miss are necessary for Wald's approximations and the diffusion approximation used in Proposition 1 to be accurate.

Remark 6: The choice of the linear test statistic in (5) and of $\hat{g}$ solving the linear integral equation of (29) restricts the validity of the right-hand-side inequality in (32b) to the classes of sequential tests employing linear test statistics and solving integral equations. However, as already discussed at the beginning of Section 4, these choices are well justified by practical considerations.

Remark 7: The inequalities in (32) are not inequalities in the strict sense; this becomes clear in the proof that follows immediately below and is related to the assumptions (A5-1) and (A5-2). However, these inequalities are satisfied, for all practical purposes, when $\hat{\alpha}$ and $\hat{\beta}$ are sufficiently small (refer to Remark 5). In particular, as $\hat{\alpha} \to 0$ and $\hat{\beta} \to 0$, the required number of samples $N \to \infty$, under both hypotheses, the thresholds $\hat{a}$ and $\hat{b}$ obtained from (6a)-(6b) for the desirable error probabilities $\hat{\alpha}$ and $\hat{\beta}$ become $\hat{a} \to -\infty$ and $\hat{b} \to \infty$, and (32) reduces to the asymptotic result

$$E_0\{N_\infty \mid \hat{g},F_0\} = \frac{-\hat{a}}{-\bar{\mu}_0(\hat{g},F_0)} \leq E_0\{N_\infty \mid \hat{g},\hat{F}_0\} = \frac{-\hat{a}}{-\bar{\mu}_0(\hat{g},\hat{F}_0)} \tag{33a}$$

$$E_1\{N_\infty \mid \hat{g},F_1\} = \frac{\hat{b}}{\bar{\mu}_1(\hat{g},F_1)} \leq E_1\{N_\infty \mid \hat{g},\hat{F}_1\} = \frac{\hat{b}}{\bar{\mu}_1(\hat{g},\hat{F}_1)} \tag{33b}$$

$$E_1\{N_\infty \mid \hat{g},F_1\} + E_0\{N_\infty \mid \hat{g},F_0\} \leq E_1\{N_\infty \mid \hat{g},\hat{F}_1\} + E_0\{N_\infty \mid \hat{g},\hat{F}_0\} \leq E_1\{N_\infty \mid g,\hat{F}_1\} + E_0\{N_\infty \mid g,\hat{F}_0\} \tag{33c}$$

The quantities involved in these inequalities are termed **asymptotic speeds** of the SPRT (matched and mismatched ones), the asymptotic nature being denoted by the subscript ∞ of the required sample number N.

Remark 8: As promised in Remark 2, the numerators of (15a)-(15b) are positive and so are the denominators for the situations of interest in this paper. This definitely holds for the robust sequential test of (5) and observations with marginal cdfs within the capacity classes of (A-1) and bivariates satisfying (11), and is established in the proof below.

Proof: The right-hand-side in the inequality (32b) follows from the fact that $\hat{g}$ is selected to optimize the sum of the average sample numbers under the two hypotheses (for desirable error probabilities smaller than $\hat{\alpha}$ and $\hat{\beta}$) of the sequential test (5), when the cdfs are $\hat{F}_k$ ($k = 0, 1$) and the dependence of the observations has been removed through the bounds of (11) and (20). Under these conditions, $\hat{g}$ solves (29), which is a version of the linear integral equation of (7). In this context, $\hat{g}$ is the optimal such nonlinearity for a sequential test employing a linear test statistic and for solving a linear integral equation. This optimization was discussed in detail in Section 2 of Part I of this study (see [1]).

The left-hand-side inequality in (32a) is established as follows. We prove the result for $k = 1$; a similar proof holds for $k = 0$, as well. From (15b) and the definition of $\omega(\hat{x},\hat{y};x)$ in (19) we can rewrite the left-hand-side of (32a), for $k = 1$, in the equivalent form

$$E_1\{N \mid \hat{g},F_1\} = \frac{\ln[(1-\hat{\beta})/\hat{\alpha}] - \beta \ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})]}{\hat{\mu}_1(\hat{g},F_1)} \cong \frac{\ln[(1-\hat{\beta})/\hat{\alpha}]}{\hat{\mu}_1(\hat{g},F_1)}$$

$$\leq \frac{\ln[(1-\hat{\beta})/\hat{\alpha}]}{\hat{\mu}_1(\hat{g},\hat{F}_1)} \cong \frac{\ln[(1-\hat{\beta})/\hat{\alpha}] - \hat{\beta} \ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})]}{\hat{\mu}_1(\hat{g},\hat{F}_1)} = E_1\{N \mid \hat{g},\hat{F}_1\} \tag{34}$$

In proving (34), we first establish that the the numerators of (15a)-(15b) are positive for all elements in the uncertainty classes considered in this paper, as predicted in Remarks 8 and 2. We use the facts that $\alpha \leq \hat{\alpha}$ and $\beta \leq \hat{\beta}$, as established in Proposition 2 [refer to inequality (21) and Remark 3], and that $\alpha + \hat{\beta} < 1$ to show that

$$\omega(\hat{\beta},\hat{\alpha};\beta) \geq \omega(\hat{\beta},\hat{\alpha};\hat{\beta}) = w(\hat{\beta},\hat{\alpha}) > 0 \tag{35a}$$

and

$$\omega(\hat{\alpha},\hat{\beta};\alpha) \geq \omega(\hat{\alpha},\hat{\beta};\hat{\alpha}) = w(\hat{\alpha},\hat{\beta}) > 0 \tag{35b}$$

where the final inequalities come from the discussion following (10). Moreover, under assumption (A4-1) and, as (28a)-(28b) hold for all F_k in the capacity class of (A-1), we have that

$$\bar{\mu}_1(\hat{g},F_1) \geq \bar{\mu}_1(\hat{g},\hat{F}_1) = \frac{2(\hat{\mu}_1 - \hat{\mu}_0)^2\hat{\sigma}_1^2}{(\hat{\sigma}_1^2 + \hat{\sigma}_0^2)^2} > 0 \tag{36a}$$

and

$$\bar{\mu}_0(\hat{g},F_0) \leq \bar{\mu}_0(\hat{g},\hat{F}_0) = -\frac{2(\hat{\mu}_1 - \hat{\mu}_0)^2\hat{\sigma}_0^2}{(\hat{\sigma}_1^2 + \hat{\sigma}_0^2)^2} < 0 \tag{36b}$$

where $\hat{\sigma}_k^2 = (1 + 2R_k)\sigma^2(\hat{g},\hat{f}_k)$ and $\hat{\mu}_k = \mu(\hat{g},\hat{f}_k)$, for $k = 0, 1$, as defined in the proof of Proposition 2. This establishes that the denominators in (15a)-(15b) are strictly positive, for all elements in the uncertainty classes of interest.

Returning to the proof of (34), we use (21), for $k = 0, 1$, which is equivalent to $\beta \leq \hat{\beta}$, the fact that condition (A5-1) holds for $\hat{\beta}$ sufficiently small, and (36a) to obtain the two approximations in (34). Then the inequality in (34) follows from (36a), since all terms involved are positive. Thus, although the initial numerators of (34) satisfy the opposite inequality from the desirable one [see (35a)], assumption (A5-1) and the correct inequality satisfied by the denominators [see (36a)] prevail to render the desirable inequality in the middle of (34) and thus the left-hand-side of (32a). The left-hand-side of (32b) follows trivially. Notice that, for the asymptotic results of Remark 7, the numerators of (33a)-(33b) are stricIty positive and the denominators satisfy the correct inequalities, so that the inequalities in (33a)-(33b) are strict and do not require the assumptions (A5), which are trivially satisfied when $\hat{\alpha} \to 0$ and $\hat{\beta} \to 0$.

3. Robust Sequential Memoryless Detectors for Weak Signals

3.1 Preliminaries

In this section, we consider the following special case of (1) pertaining to the detection of a weak signal in dependent non-Gaussian noise: we must decide between the two hypotheses

$$H_0: \; X_i = N_i \;, \qquad \text{for } \; i = 1, 2, \ldots, n \tag{37a}$$

and

$$H_1: \; X_i = N_i + \theta \;, \; \text{ for } \; i = 1, 2, \ldots, n \,, \tag{37b}$$

where $\{N_i\}_{i=1}^{n}$ is the noise sequence assumed to be m-dependent or ρ-mixing and θ is a known weak signal, i.e., $\theta \to 0$.

As in [5], the stationary noise sequence $\{N_i\}_{i=1}^{n}$ has a **symmetric marginal pdf** $f(x) = f(-x)$ belonging to an **ε-contaminated uncertainty class** (see [7]):

$$f(x) = (1 - \varepsilon) f^0(x) + \varepsilon \tilde{f}(x) \tag{38}$$

where $f^0(x)$ is a known symmetric pdf (termed nominal), ε $(0 < \varepsilon < 1)$ the known degree of uncertainty, and $\tilde{f}(x)$ an arbitrary symmetric pdf.

The following conditions are assumed to hold about the nonlinearity g and the marginal pdf of the noise f:

$$g(-x) = -\, g(x) \;, \tag{39}$$

$$\frac{\partial}{\partial\theta}\Big[\int g(x) f(x - \theta)\, dx\Big]\Big|_{\theta=0} = \int \frac{\partial}{\partial\theta}\big[g(x) f(x - \theta)\big]\Big|_{\theta=0}\, dx \;, \tag{40a}$$

$$\lim_{l \to \infty} \int g(x) f'(x - \theta_l)\, dx = \int g(x) f'(x)\, dx \tag{40b}$$

for $\theta_l \to 0$ as $l \to \infty$,

$$\lim_{t \to 0} E\{[g(N_1 + t) - g(N_1)]^2\} = 0 \;, \tag{40c}$$

$$\int g(x)f'(x)dx < 0 \tag{41}$$

and

$$\sigma_0^2(g) = E\{[g(N_1)]^2\} + 2\sum_{j=1}^{\infty} E\{g(N_1)g(N_{j+1})\} > 0 \tag{42}$$

for a ρ-mixing stationary noise sequence; the ∞ in the sum of (42) should be replaced by the parameter m for an m-dependent stationary noise sequence.

The **bivariate pdf** of the noise sequence, denoted by $f^{(j)}(x,y)$ for the pair (N_1,N_{j+1}), satisfies an inequality similar to (11) (see [6]), for $k = 0$, that is,

$$\sup_g \frac{|cov\{g(N_1)g(N_{1+j})\}|}{var\{g(N_1)\}var\{g(N_{1+j})\}} = \sup_g \frac{|E\{g(N_1)g(N_{1+j})\}|}{E\{[g(N_1)]^2\}} \le r_j \tag{43}$$

where g ranges over all measurable functions satisfying $E\{[g(N_1)]^2\} < \infty$. The parameters r_j can be obtained from the parameters ρ_j of the ρ-mixing process $\{N_i\}_{i=1}^{\infty}$ (refer to the discussion following (11) in Section 2.1 for more details). We denote by f^* the collection $(f,\{f^{(j)}\}_{j=1}^{\infty})$. For a given marginal pdf f, equality holds in (43) for all g, if the bivariate pdf takes the form

$$f^{(j)}(x,y) = (1-r_j)f(x)f(y) + r_j f(x)\delta(x-y) \tag{44}$$

where $\delta(x)$ is Dirac's δ function. In [6] random processes $\{N_i\}_{i=1}^{\infty}$ with pdfs of the form (44) have been constructed. Similarly to (13), if the condition (43) is satisfied,

$$\sigma_0^2(g) \le (1+2R)E\{[g(N_1)]^2\} \tag{45}$$

where $R = \sum_{j=1}^{\infty} r_j$, for a ρ-mixing noise sequence; ∞ should be replaced by m for an m-dependent noise sequence. The equality in (45) is satisfied for all g, if $f^{(j)}$ is given by (44); thus, the pdf of (44) has maximum variance among all pdfs in the class defined by (11).

As discussed in [4], when defining the sequence $\theta_l = c/\sqrt{l} \to 0$ as $l \to \infty$, we use an index other than n (the sample number) for the sequential detection problem; thus, $l \to \infty$ implies $\theta_l \to 0$ and the required sample number for the SPRT $N \to \infty$, as well.

The sequential test (SPRT) to be robustified is based on the linear test statistic of (5) and the thresholds (6a)-(6b) of Section 2.1 with the necessary adjustments for the weak-signal in additive noise case. Actually, as proved in Section 2 of Part I (see equation (12) of [1] and the subsequent discussion) the modified test statistic:

$$\hat{S}_n = \frac{\hat{\mu}_\theta}{\hat{\sigma}_0^2}\left[\sum_{i=1}^{n}\hat{g}(X_i) - \frac{\hat{\mu}_\theta}{2}n\right] \tag{46}$$

is optimal within the class of SPRTs based on memoryless nonlinearities, in the sense that it is a likelihood ratio sequential test performed on $\sum_{i=1}^{n}\hat{g}(X_i)$. To minimize the expected sample numbers under the two hypotheses for desirable error probabilities $\hat{\alpha}$ and $\hat{\beta}$, the nonlinearity $\hat{g}$ in (46) must maximize the asymptotic relative efficiency (ARE) $[\int\hat{g}(x)\hat{f}'(x)dx]^2/[\hat{\sigma}_0^2(\hat{g})]$ (refer to equation (22) of [1]). Therefore, $\hat{g}$ is the solution to the linear integral equation

$$\hat{g}(x) = -\frac{\hat{f}'(x)}{\hat{f}(x)} - \int\hat{K}(x,y)\hat{g}(y)dy \tag{47}$$

where the kernel $\hat{K}(x,y)$ is defined by

$$\hat{K}(x,y) = \sum_{j=1}^{\infty}[\hat{f}^{(j)}(x,y) + \hat{f}^{(j)}(y,x)] \tag{48}$$

for a ρ-mixing sequence; ∞ should be replaced by m for an m-dependent noise sequence. In (47) and (48) $\hat{f}(x)$ and $\{\hat{f}^{(j)}(x,y)\}_{j=1}^{\infty}$ denote marginal and bivariate pdfs of the noise sequence that either represent estimates of the statistics of the noise equence and thus are different from the actual statistics of the noise sequence, or may be chosen to characterize the least-favorable conditions for the operation of the test of (46) within specific uncertainty classes [like those of

(38) and (42)] as will be done in Section 3.2. below. For notational convenience, we use $\hat{f}^* = (\hat{f}, \{\hat{f}^{(j)}\}_{j=1}^{\infty})$. The above situation is clearly one of **mismatch**, since the operating conditions of the above test statistic are determined by the actual statistics of the noise sequence, namely the marginal pdfs f in the class (38) and the bivariate pdfs $\{f^{(j)}\}_{j=1}^{\infty}$ in the class (42), which are generally different form the ones involved in (46) and (47)-(48). Finally, in (46) the mean $\hat{\mu}_\theta$ is given by

$$\hat{\mu}_\theta = \int \hat{g}(x)\hat{f}(x-\theta)dx \neq 0 \tag{49}$$

whereas the mean $\hat{\mu}_0 = \int \hat{g}(x)\hat{f}(x)dx = 0$ since $\hat{g}$ and $\hat{f}$ are odd and even functions, respectively. This last fact justifies why $\hat{\mu}_0$ is not present in (46), which was directly derived from (5). The variance $\hat{\sigma}_0^2$ is obtained from (42) upon substitution for $\hat{g}$ and $\hat{f}^*$.

3.2 Robustification of Sequential Memoryless Detectors for Weak Signals

First, we evaluate the error probabilities and the expected sample numbers of the sequential test of (46) under mismatch.

Proposition 4: Let $\overline{P}_k(\hat{g}, f^*)$ denote the probability of error under mismatch and $E_k\{N \mid \hat{g}, f)\}$ the required average sample number, when hypothesis H_k $(k = 0, 1)$ of (37a) or (37b) is true. Let us assume that the sequential test of (46) with thresholds $\hat{a}$ and $\hat{b}$ defined by (6a)-(6b) for desired error probabilities $\hat{\alpha}$ and $\hat{\beta}$, is employed. Then the following identities hold, under the assumptions (39)-(42):

$$\overline{P}_0(\hat{g}, F_0^*) = \alpha = P_0\{\hat{S}_N \text{ up-crosses } \hat{b} \text{ before it down-crosses } \hat{a}\} = \frac{1-(e^{\hat{a}})^{\frac{2\mu_0(\hat{g},f)}{\sigma_0^2(\hat{g},f^*)}}}{1-(e^{\hat{a}-\hat{b}})^{\frac{2\mu_0(\hat{g},f)}{\sigma_0^2(\hat{g},f^*)}}} \tag{50a}$$

$$\bar{P}_1(\hat{g},F_1^*) = \beta = P_1\{\hat{S}_N \text{ down–crosses } \hat{a} \text{ before it up–crosses } \hat{b}\} = \frac{1-(e^{\hat{b}})^{\frac{2\bar{\mu}_1(\hat{g},f)}{\bar{\sigma}_0^2(\hat{g},f^*)}}}{1-(e^{\hat{b}-\hat{a}})^{\frac{2\bar{\mu}_1(\hat{g},f)}{\bar{\sigma}_0^2(\hat{g},f^*)}}} \tag{50b}$$

and

$$E_0\{N\mid\hat{g},f)\} = \frac{\omega(\hat{\alpha},\hat{\beta};\alpha)}{-\bar{\mu}_0(\hat{g},f)} \tag{51a}$$

$$E_1\{N\mid\hat{g},f)\} = \frac{\omega(\hat{\beta},\hat{\alpha};\beta)}{\bar{\mu}_1(\hat{g},f)} \tag{51b}$$

where

$$\bar{\mu}_1(\hat{g},f) = \frac{\hat{\mu}_\theta}{\hat{\sigma}_0^2}\left[\mu_\theta(\hat{g},f) - \frac{\hat{\mu}_\theta}{2}\right] \tag{52a}$$

$$\bar{\mu}_0(\hat{g},f) = -\frac{\hat{\mu}_\theta^2}{2\hat{\sigma}_0^2} \tag{52b}$$

and

$$\bar{\sigma}_0^2(\hat{g},f^*) = \frac{\hat{\mu}_\theta^2}{\hat{\sigma}_0^4}\sigma^2(\hat{g},f^*)\ . \tag{53}$$

In (52)-(53), $\mu_\theta(\hat{g},f) = \int\hat{g}(x)f(x-\theta)dx = \lim_{n\to\infty} n^{-1}E\{\hat{T}_n\}$ denotes the asymptotic mean and $\sigma^2(\hat{g},f^*) = \lim_{n\to\infty} n^{-1}E\{[\hat{T}_n]^2\}$ denotes the asymptotic variance [obtained from (42) for $\hat{g}$ and marginal/bivariate pair f^*] under mismatch of $\hat{T}_n = \sum_{i=1}^{n}\hat{g}(X_i)$. The corresponding means and variances under matched conditions are denoted by $\hat{\mu}_\theta = \mu_\theta(\hat{g},\hat{f})$ and $\hat{\sigma}_0^2 = \sigma^2(\hat{g},\hat{f}^*)$. The rest of the quantities involved in (50)-(51) are $\hat{\alpha}$ and $\hat{\beta}$, the error probabilities under matched conditions still given by (18a)-(18b), and the quantity $\omega(\hat{x},\hat{y};x)$ defined by (19).

Remark 9: Remarks similar to Remarks 1 and 2 made for Proposition 1 are valid here. Also, the proof of Proposition 4 is very similar to that of Proposition 1 and is omitted.

Next we establish the first of the two main results of this section pertaining to the robustness of the sequential test of (46) with respect to the error probabilities.

Proposition 5: Consider a sequential test based on the thresholds $\hat{a}$ and $\hat{b}$, obtained from (6a)-(6b) for desired error probabilities $\hat{\alpha}$ and $\hat{\beta}$, and on the linear test statistic of (46), where the nonlinearity $\hat{g}$ solves (47) with the kernels of (48) modified according to (44) as

$$\hat{K}(x,y) = 2R\hat{f}(x)\delta(x-y) \tag{54}$$

and is given by

$$\hat{g}(x) = -\hat{f}'(x)/\hat{f}(x) \ . \tag{55}$$

In (54) $\hat{f}$ is the least-favorable pdf for the ARE and the mean-square estimation error for uncertainty in the marginal pdf of the noise within the class of (38); $\hat{f}$ has been evaluated by Huber in [7]; moreover, $\hat{f}^{(j)}$ satisfies (44), for all j. Then this test is **least-favorable for the error probabilities under the two hypotheses**, that is,

$$\overline{P}_k(\hat{g},f^*) \leq \overline{P}_k(\hat{g},\hat{f}^*) \ , \ \text{for} \ k = 0, 1 \ , \tag{56}$$

for any $f^* = (f, \{f^{(j)}\}_{j=1}^{\infty})$ with f in the class of (38) and $f^{(j)}$ satisfying (43). In (56), $\overline{P}_k(\hat{g},f^*)$, the error probabilities under mismatch, are as defined in Proposition 4 by (50a)-(50b).

Remark 10: This Proposition holds under the same conditions as Proposition 4.

Proof: The sequence of steps necessary for Proving (56) is similar to that used for the proof of Proposition 2, but the individual steps differ. Here we sketch the proof and cite the points that are different.

On the basis of (42)-(45) we can derive the equality

$$\hat{\sigma}_0^2 = \sigma^2(\hat{g}, \hat{f}^*) = (1 + 2R)\sigma^2(\hat{g}, \hat{f}) \tag{57}$$

where $\sigma^2(\hat{g}, \hat{f})$ depends only on $\hat{f}$ and the inequality

$$\sigma^2(\hat{g}, f^*) \leq (1 + 2R)\sigma^2(g, f) \tag{58}$$

for all f^* with bivariates $f^{(j)}$ in (43), where $\sigma^2(\hat{g}, f)$ depends only on f belonging to (38). Using these two results and the fact that $\overline{P}_k(\hat{g}, f^*)$ of (50a)-(50b) is an increasing function of $\overline{\sigma}_0^2(\hat{g}, f^*)$, and thus of $\sigma^2(\hat{g}, f^*)$, we deduce that (56) is equivalent to

$$P_k(\hat{g}, f) \leq P_k(\hat{g}, \hat{f}) \ , \ \text{ for } k = 0, 1 \ , \tag{59}$$

where $P_k(\hat{g}, f)$ is obtained from (50a)-(50b) by using the right-hand member of (58) in (53). Equation (59) involves only the marginal pdfs f and $\hat{f}$ and this simplifies considerably the final part of the proof.

To prove (59) we observe that (50a) is an increasing function of the normalized drift c_0, whereas (50b) is a decreasing function of the drift c_1, defined as $c_k = 2\overline{\mu}_k(\hat{g}, f)/\overline{\sigma}_0^2(\hat{g}, f^*)$, for $k = 0, 1$, and for the worst case $\hat{c}_k = 2\overline{\mu}_k(\hat{g}, \hat{f}) \ / \ \overline{\sigma}_0^2(\hat{g}, \hat{f}^*)$. Removing the dependence in the observations through the bounds of (43) and (44) implies that $\hat{c}_0 = -1$ and $\hat{c}_1 = 1$. Thus (59) becomes equivalent to

$$c_0 = -\frac{(1 + 2R)\sigma^2(\hat{g}, \hat{f})}{(1 + 2R)\sigma^2(\hat{g}, f)} \leq -1 = \hat{c}_0 \tag{60a}$$

and

$$c_1 = \frac{[2\mu_\theta(\hat{g}, f) - \hat{\mu}_\theta](1 + 2R)\sigma^2(\hat{g}, \hat{f})}{\hat{\mu}_\theta(1 + 2R)\sigma^2(\hat{g}, f)} \geq 1 = \hat{c}_1 \tag{60b}$$

These inequalities are satisfied, if

$$\sigma^2(\hat{g}, \hat{f}) \geq \sigma^2(\hat{g}, f) \tag{61a}$$

and

$$\mu_\theta(\hat{g},f) \geq \mu_\theta(\hat{g},\hat{f}) = \hat{\mu}_\theta > 0 \ . \tag{61b}$$

The last two conditions can be put in the following equivalent forms

$$\int \hat{g}^2(x)\hat{f}(x)dx \geq \int \hat{g}^2(x)f(x)dx \tag{62a}$$

and

$$-\int \hat{g}(x)f'(x)dx \geq -\int \hat{g}(x)\hat{f}'(x)dx > 0 \tag{62b}$$

Obtaining (62a) from (61a) is trivial. To obtain (62b) from (61b) we subtract $\mu_0(\hat{g},f)$ from the left-hand side and $\mu_0(\hat{g},\hat{f})$ from the right-hand side of (61b). Since both these terms are 0 (due to $\hat{g}$ being an odd and any f in (38) being an even function) we obtain

$$\int \hat{g}(x)[f(x-\theta) - f(x)]dx \geq \int \hat{g}(x)[\hat{f}(x-\theta) - \hat{f}(x)]dx$$

which after dividing by $\theta > 0$ and taking the limit as $\theta \to 0$ yields (62b).

Notice that the conditions (62a)-(62b) are satisfied for all f in the class (38) if $\hat{f}$ is the least-favorable pdf for the ARE derived in [7] and $\hat{g}$ is given by (55). Indeed, in the proof of [7] the ARE was robustified by minimzing its numerator and maximizing its denominator; the former corresponds to (62b) and the latter to (62a) in our situation. The inequality $\hat{\mu}_\theta > 0$ or equivalently $-\int \hat{g}(x)\hat{f}'(x)dx > 0$ follows from assumption (41). This completes the proof or Proposition 5.

Remark 11: Proposition 5 holds not only for classes of the form (38) for the marginal pdfs of the noise sequence, but also for any other class of pdfs for which the numerator and denominator of the ARE are respectively minimized and maximized simultaneously by the same least-favorable pdf $\hat{f}$ in the class. For example, for the total variation uncertainty class, also introduced in [7], (62a)-(62b) hold and so do Propositions 5 and 6.

Finally, we prove the second main result of this section, which pertains to the robustness of the sequential memoryless test of (46) with respect to the expected sample numbers.

Proposition 6: Suppose that the same sequential test as in Proposition 5 is employed. A notation identical to that of Propositions 4 and 5 is used. Assume that conditions (62a)-(62b) and assumptions (A5-1)-(A5-2) hold. Then the sequential test of (46) is **minimax robust for the expected sample numbers under the two hypotheses**, that is,

$$E_k\{N \mid \hat{g}, f\} \leq E_k\{N \mid \hat{g}, \hat{f}\}\ , \quad \text{for } k = 0, 1\ , \tag{63a}$$

and

$$E_1\{N \mid \hat{g}, f\} + E_0\{N \mid \hat{g}, f\} \leq E_1\{N \mid \hat{g}, \hat{f}\} + E_0\{N \mid \hat{g}, \hat{f}\} \leq E_1\{N \mid g, \hat{f}\} + E_0\{N \mid g, \hat{f}\} \tag{63b}$$

for all marginal pdfs f in the class (38) with bivariates $f^{(j)}$ satisfying (43) and any measurable function g satisfying $E\{g^2(X_1)\} < \infty$; $\hat{f}$ is the pdf derived by Huber in [7] and has, in this case, bivariates $\hat{f}^{(j)}$ satisfying (44), for all j. The expected sample numbers under mismatch $E_k\{N \mid \hat{g}, f\}$ are as defined by (50a)-(50b) of Proposition 4.

Remark 12: In contrast to Remark 6 following Proposition 3, the choice of the linear test statistic in (46) and of $\hat{g}$ solving the linear integral equation of (47) [the solution being given by (55)] does not restrict the validity of the right-hand-side inequality in (63b). This is due to the fact that the test test statistic of (46) is optimal within the class of memoryless structures.

Remark 13: Remarks 7 and 8 following Proposition 3 are also valid here. In particular, results similar to the asymptotic results of (33a)-(33b) regarding the asymptotic speeds of the SPRT hold for the situation described by Proposition 6.

Proof: The right-hand-side in the inequality (63b) follows from the fact that $\hat{g}$ is selected to optimize the sum of the average sample numbers under the two hypotheses (for desirable error probabilities smaller than α and β) of the sequential test (46), when the noise pdf is $\hat{f}$ and the dependence of the observations has been removed through the bounds of (43) and (54). Under these conditions, $\hat{g}$ is given by (55) and maximizes the ARE for the matched worst case. Also refer to the beginning of Section 2.1 of Part I [1].

The left-hand-side inequality in (63a) is established as follows. We prove the result for $k = 1$; a similar proof holds for $k = 0$, as well. From (51b) and the definition of $\omega(\hat{x},\hat{y};x)$ in (19) we can rewrite the left-hand-side of (63a), for $k = 1$, in the equivalent form

$$E_1\{N \mid \hat{g},f\} = \frac{\ln[(1-\hat{\beta})/\hat{\alpha}] - \hat{\beta}\ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})]}{\bar{\mu}_1(\hat{g},f)} \cong \frac{\ln[(1-\hat{\beta})/\hat{\alpha}]}{\bar{\mu}_1(\hat{g},f)}$$

$$\leq \frac{\ln[(1-\hat{\beta})/\hat{\alpha}]}{\bar{\mu}_1(\hat{g},\hat{f})} \cong \frac{\ln[(1-\hat{\beta})/\hat{\alpha}] - \hat{\beta}\ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})]}{\bar{\mu}_1(\hat{g},\hat{f})} = E_1\{N \mid \hat{g},\hat{f}\} \tag{64}$$

Regarding the denominators of (64) we can apply (61b) to obtain

$$\bar{\mu}_1(\hat{g},f) \geq \bar{\mu}_1(\hat{g},\hat{f}) = \frac{\hat{\mu}_\theta^2}{2(1+2R)\sigma^2(\hat{g},\hat{f})} > 0 \tag{65a}$$

and

$$\bar{\mu}_0(\hat{g},f) = \bar{\mu}_0(\hat{g},\hat{f}) = -\frac{\hat{\mu}_\theta^2}{2(1+2R)\sigma^2(\hat{g},\hat{f})} < 0\,. \tag{65b}$$

After establishing (65a), we follow for the proof of (64) similar arguments, as we did for the proof of (34) for Proposition 3. We do not repeat them here.

Remark 14: The robust sequential test of (46) which uses a test statistic $S_n = \frac{\hat{\mu}_\theta}{\hat{\sigma}_0^2}\left[\sum_{i=1}^{n}\hat{g}(X_i) - \frac{\hat{\mu}_\theta}{2}n\right]$ with $\hat{g}$ given by (55), $\hat{\mu}_\theta$ given by (49), and $\hat{\sigma}_0^2$ given by (42) upon substitution for $\hat{g}$, $\hat{f}$, and $\hat{f}^{(j)}$ from (44), is easier to implement than the sequential test of [4]. The latter first estimates θ by an M-estimator for each step n of the SPRT; this involves solving the nonlinear equation $\sum_{i=1}^{n} l(X_i - \theta_n) = 0$ for the estimate θ_n, where $l(x)$ is an appropriate nonlinearity and $\{X_i\}_{i=1}^{n}$ are the n observations collected untill the n-th step of the sequential test. Then it performs an SPRT based on the likelihood ratio of θ_n.

4. Conclusions

In this paper, we robustified sequential tests based on memoryless nonlinearities. We developed robust sequential tests for (i) memoryless dicrimination from two arbitrary stationary m-dependent or mixing observations, and (ii) memoryless detection of a weak signal in additive stationary m-dependent or mixing noise. In both cases, the marginal pdfs of the two observation sequences or of the noise sequence belong to uncertainty classes, such as ε-contaminated classes and total variation classes, whereas the bivariate pdfs satisfy bounds on the correlation coefficients of time-shifts of the observation sequences or the noise sequence.

The robust sequential tests derived have the form of (5) for the discrimination problem and of (46) for the problem of detecting a weak signal. They consis of SPRTs based on simple linear test statistics involving nonlinearities $\hat{g}$ associated with the least-favorable pdf in the uncertainty class of marginal pdfs and with bivariates which achieve the aforementioned bounds on the correlation coeffcients of time-shifts of the observation or noise sequences. In the case of detection of weak signals, the test of (46) is considerably easier to implement that the test proposed in [4] for the i.i.d. case (refer to Remark 14).

Coupled with the results of the first part of this study (see [1]), which derived optimal sequential discrimination schemes based on memoryless nonlinearities and established their superiority to the conventional i.i.d. discriminators and to fixed-sample-size memoryless schemes for environments characterized by strongly correlated observations, this paper strengthened further the usefulness of these sequential tests by establishing that they can be rendered relatively immune to statistical uncertainty within certain popular classes of distributions.

References

[1] E. Geraniotis. ''Sequential Tests for Memoryless Discrimination from Dependent Observations--Part I: Optimal Tests." Submitted to the *IEEE Transactions on Information Theory*, 1989, under review.

[2] D. Sauder and E. Geraniotis. ''Optimal and Robust Memoryless Block Discrimination from Dependent Observations." Submitted for publication to the *IEEE Transactions on Information Theory*, 1989, under review.

[3] S. A. Kassam and H. V. Poor. "Robust Techniques for Signal Processing." *Proceedings of the IEEE*, Vol. 73, pp. 433-481, March 1985.

[4] A. H. El-Sawy and V. D. Vandelinde. ''Robust Sequential Detection of Signals in Noise." *IEEE Transactions on Information Theory*, Vol. IT-25, pp. 346-353, May 1979.

[5] G. Moustakides and J. B. Thomas. ''Min-Max Detection of Weak Signals in ϕ-Mixing Noise." *IEEE Transactions on Information Theory*, Vol. IT-30, pp. 529-537, May 1984.

[6] J. S. Sadowsky. ''A Maximum Variance Model for Robust Detection and Estimation with Dependent Data." *IEEE Transactions on Information Theory*, Vol. IT-32, pp. 220-226, March 1986.

[7] P. J. Huber. "A Robust Version of the Probability Ratio Test." *Annals of Mathematical Statistics*, Vol. 36, pp. 1753-1758, 1965.

[8] S. A. Kassam. "Robust Hypothesis Testing for Bounded Classes of Probability Densities." *IEEE Transactions on Information Theory*, Vol. IT-27, pp. 242-247, 1981.

[9] K. S. Vastola and H. V. Poor. "On the p-Point Uncertainty Class." *IEEE Transactions on Information Theory*, Vol. IT-30, pp.374-376, 1984.

[10] V. Strassen. " The Existence of Probability Measures with Given Marginals." *Annals of Mathematical Statistics*, Vol. 36, No. 2, April 1965.

[11] P. J. Huber and V. Strassen. "Minimax Tests and the Neyman-Pearson Lemma for Capacities." *Annals of Statistics*, Vol. 1, pp. 251-265, 1973.

[12] E. Geraniotis and Y. A. Chau. "On Robust Continuous-Time Discrimination." Submitted for publication to the *IEEE Transactions on Information Theory*, 1988, under review.

[13] P. Billingsley. *Convergence of Probability Measures*. New York: John Wiley & Sons, 1968.

[14] M. Peligrad. "Recent Advances in the Central Limit Theorem and its Weak Invariance Principle for Mixing Sequences of Random Variables (A Survey)." *Dependence in Probability and Statistics,* Birkhauser, 1985, p. 193.

[15] A. Wald. *Sequential Analysis.* New York: John Wiley & Sons, 1947.

Appendix A

Review of Uncertainty Models and of Basic Results from the Theory of Minimax Robustness

In this Appendix we present a review of the uncertainty models based on 2-alternating capacities and of Huber's basic theory of minimax robustness associated with these models.

Definition: A positive finite set function υ on a sample space Ω with a complete, separable, and metrizable topology and associated Borel field F is called a **2-alternating capacity** if it is increasing, continuous from below, continuous form above on closed sets, and satisfies the conditions $\upsilon(\emptyset) = 0$, and

$$\upsilon(A \cup B) + \upsilon(A \cap B) \leq \upsilon(A) + \upsilon(B).$$

Suppose now that M is the class of measures on (Ω, F) and $m \in M$ is a measure. Consider the uncertainty class which is determined by the 2-alternating capacity υ as follows

$$M_{\upsilon} = \{m \in M \mid m(A) \leq \upsilon(A),\ \forall A \in F,\ m(\Omega) = \upsilon(\Omega)\}. \tag{A-1}$$

When Ω is compact several popular uncertainty models like ε-contaminated neighborhoods [7], total variation neighborhoods [7], band classes [8], and p-point classes [9], are special cases of this model. The most general uncertainty classes of the form (A-1) are determined by generalized capacities of the form $\upsilon(A) = \int \upsilon_{\underline{a}}(A)\mu(d\underline{a})$ for $\forall A \in F$ (see [10]), where $\upsilon_{\underline{a}}$ is a capacity when conditioned on the parameter vector $\underline{a}$ and μ is the measure induced by the joint distribution of these parameters. Fundamental properties of the uncertainty model (A-1) have been studied by Huber and Strassen (see [11]). We state the relevant properties as Lemma 1.

Lemma: Suppose υ_0 and υ_1 are 2-alternating capacities on (Ω, F) and (M_0, M_1) are the uncertainty classes determined by (υ_0, υ_1) as in (1). In that case, there exists a Borel-measurable function $\pi_{\upsilon}:\Omega \rightarrow [0, \infty]$ such that the average (Bayes) risk $\theta\upsilon_0(A) + \upsilon_1(A^c)$ is minimized for $A^* = \{\pi_{\upsilon} > \theta\}$, i.e.,

$$\theta\upsilon_0(\{\pi_{\upsilon} > \theta\}) + \upsilon_1(\{\pi_{\upsilon} \leq \theta\}) \leq \theta\upsilon_0(A) + \upsilon_1(A^c) \tag{A-2}$$

for all $A \in F$ and $\theta \geq 0$; A denotes an arbitrary decision test, A^c its complement, θ can be interpreted as the ratio of the prior probabilities of the two hypotheses H_0 and H_1, and $\{\pi_{\upsilon} > \theta\}$ can be interpreted as the **likelihood ratio test** for υ_0 versus υ_1. Clearly, (A-2) together with (A-1) imply that

$$\theta m_0(\{\pi_{\upsilon} > \theta\}) + m_1(\{\pi_{\upsilon} \leq \theta\}) \leq \theta\upsilon_0(\{\pi_{\nu} > \theta\}) + \upsilon_1(\{\pi_{\upsilon} \leq \theta\}) \leq \theta\upsilon_0(A) + \upsilon_1(A^c) \tag{*}$$

for all $m_0 \in M_0$, $m_1 \in M_1$, $\theta \geq 0$, and $A \in F$; this inequality establishes the **minimax robustness** of the test based on π_{υ}. Furthermore, there exist measures $(\hat{m}_0, \hat{m}_1)$ in $M_0 \times M_1$ such that

$$\hat{m}_0(\{\pi_{\upsilon} > \theta\}) = \upsilon_0(\{\pi_{\upsilon} > \theta\}) \geq m_0(\{\pi_{\upsilon} > \theta\}) \tag{A-3}$$

$$\hat{m}_1(\{\pi_{\upsilon} \leq \theta\}) = \upsilon_1(\{\pi_{\upsilon} \leq \theta\}) \geq m_1(\{\pi_{\upsilon} \leq \theta\}) \tag{A-4}$$

for all $\theta \geq 0$; these inequalities imply that π_{υ} is stochastically largest over M_0 under $\hat{m}_0$ and stochastically smallest over M_1 under $\hat{m}_1$. The quantity π_{υ} is sometimes termed the **Huber-Strassen derivative** of the classes M_0 and M_1, is denoted by $d\upsilon_1/d\upsilon_0$, is given by $\pi_{\upsilon} = d\hat{m}_1/d\hat{m}_0$, and is unique a.e. $[\hat{m}_1 + \hat{m}_0]$; it plays the role of the **worst-case likelihood ratio** for the two uncertainty classes. The dominance properties (A-3)-(A-4) establish the existence of measures in the classes M_0 and M_1 that achieve the upper values provided by υ_0

and υ_1 for sets of the form $\{\pi_\upsilon \le \theta\}$ and their complements. The measures $(\hat{m}_0, \hat{m}_1)$ are termed the **least-favorable** measures over $M_0 \times M_1$. For the aforementioned four uncertainty classes (ε-contaminated mixtures, total variation classes, band classes, and p-point classes), which are special cases of the general model (A-1) when $\mathbf{\Omega}$ is compact, the least-favorable pairs of probability measures (actually, the corresponding probability density functions) have been derived in closed form ([7]-[9]). Depending on the form of the joint distributions of the parameter vector $\underline{a}$ under the two hypotheses, even the most general uncertainty model that can be obtained from (A-1), that is, when $\upsilon(A)_i = \int \upsilon_{\underline{a},i}(A)\mu_i(d\underline{a})$ for $\forall A \in \boldsymbol{F}$ are the generalized capacities of [10]), can result in closed form expressions for the least-favorable probability measures.

Example: Consider the ε-contaminated mixture uncertainty classes of probability measures described in [7]

$$M_j = \{m_j \in M \mid m_j(A) = (1-\varepsilon_j)m_j^0(A) + \varepsilon_j \tilde{m}_j(A) \text{ for all } A \in \boldsymbol{F},\ \tilde{m}_j(\Omega) = \tilde{m}_j^0(\Omega) = 1\}, \quad \text{(A-5)}$$

for $j = 0, 1$, which are determined by the known **nominal probability measures** m_0^0 and m_1^0 and the **degrees of uncertainty** ε_0 and ε_1 ($0 \le \varepsilon_j \le 1$ for $j = 0, 1$) the unknown probability measures $\tilde{m}_j$ are allowed to take any arbitrary values. This uncertainty class is appropriate for modeling situations in which the probability measures governing the observations are convex combinations of known probability measures and arbitrary probability measures. Then the associated 2-alternating capacities are

$$\upsilon_j(A) = \begin{cases} (1-\varepsilon)m_j^0(A) + \varepsilon_j, & A \neq \varnothing \\ 0, & A = \varnothing \end{cases} \quad \text{(A-6)}$$

and the least-favorable distributions are

$$d\hat{m}_0/d\lambda = \begin{cases} (1-\varepsilon_0)\, dm_0^0/d\lambda\,, & dm_1^0/dm_0^0 \le c_0 \\ [(1-\varepsilon_0)/c_0)]dm_1^0/d\lambda\,, & c_0 < dm_1^0/dm_0^0 \end{cases} \quad \text{(A-7a)}$$

$$d\hat{m}_1/d\lambda = \begin{cases} (1-\varepsilon_1)\, dm_1^0/d\lambda\,, & c_1 < dm_1^0/dm_0^0 \\ c_1(1-\varepsilon_1)\, dm_0^0/d\lambda\,, & dm_1^0/dm_0^0 \le c_1 \end{cases} \quad \text{(A-7b)}$$

where λ is the Lebesgue measure and $0 \le c_1 \le c_0 < \infty$ are constants such that $\hat{m}_1(\Omega) = \hat{m}_0(\Omega) = 1$, and the Huber-Strassen derivative π_υ has the form

$$\pi_\upsilon = d\hat{m}_1/d\hat{m}_0 = \frac{1-\varepsilon_1}{1-\varepsilon_0}\min\{c_0, \max\{c_1, dm_1^0/dm_0^0\}\} \quad \text{(A-8)}$$

which consists of a censored version of the nominal likelihood-ratio dm_1^0/dm_0^0.

Recently, in [12] the dominance properties (A-3) and (A-4) were exploited to extend the Huber-Strassen theory to more general objective functions than the Bayes risk of (A-2). We cite the following proposition from [12] without the proof provided there as Lemma 2.

Lemma 2: Suppose that the measures (m_0, m_1) on $(\mathbf{\Omega}, \boldsymbol{F})$ belong to $\boldsymbol{M}_0 \times \boldsymbol{M}_1$ characterized by (A-1) and that x is a real variable:

(i) If one of the following situations holds:

(a) both $g(\pi_\upsilon)$ and $h(x)$ are nonnegative, increasing functions of π_υ (the Huber-Strassen

derivative) and x, respectively;
(b) $g(\pi_\upsilon)$ is a nonnegative, decreasing function of π_υ and $h(x)$ is a nonpositive and increasing function of x;
(c) $g(\pi_\upsilon)$ is a nonpositive, increasing function of π_υ and $h(x)$ is a nonnegative and decreasing function of x;
(d) both $g(\pi_\upsilon)$ and $h(x)$ are nonpositive and decreasing functions of π_υ and x, respectively; then

$$\int_\Omega g(\pi_\upsilon(x))h(x)m_0(dx) \leq \int_\Omega g(\pi_\upsilon(x))h(x)\hat{m}_0(dx) \tag{A-9}$$

$$\int_\Omega g(\pi_\upsilon(x))h(x)m_1(dx) \geq \int_\Omega g(\pi_\upsilon(x))h(x)\hat{m}_1(dx) \tag{A-10}$$

(ii) If one of the following situations holds:
(a) both $g(\pi_\upsilon)$ and $h(x)$ are nonnegative, decreasing functions of π_υ and x, respectively;
(b) $g(\pi_\upsilon)$ is a nonnegative, increasing function of π_υ and $h(x)$ is a nonpositive and decreasing function of x;
(c) $g(\pi_\upsilon)$ is a nonpositive, decreasing function of π_υ and $h(x)$ is a nonnegative and increasing function of x;
(d) both $g(\pi_\upsilon)$ and $h(x)$ are nonpositive and increasing functions of π_υ and x, respectively; then

$$\int_\Omega g(\pi_\upsilon(x))h(x)m_0(dx) \geq \int_\Omega g(\pi_\upsilon(x))h(x)\hat{m}_0(dx) \tag{A-11}$$

$$\int_\Omega g(\pi_\upsilon(x))h(x)m_1(dx) \leq \int_\Omega g(\pi_\upsilon(x))h(x)\hat{m}_1(dx) \tag{A-12}$$

where $\hat{m}_0$ and $\hat{m}_1$ are singled out by Lemma 1.

Remark 1: If either $g(x) = 1$ or $h(x) = 1$ for all x, i.e., one of the two functions g or h is absent from the integrands of (A-9)-(A-12), the inequalities in (A-9)-(A-12) still hold; in this case the nonnegativity of the function involved is not a necessary condition.

Remark 2: Lemmas 1 and 2 hold even if the 2-alternating capacity υ_0 is itself a measure. In this case, the uncertainty class M_0 has a single element υ_0.

Remark 3: If the nominal measures m_k^0 characterizing the uncertainty class (e.g., the ε-contaminated or total variation classes or the upper and lower bounds in the case of the band class) are absolutely continuous with respect to the Lebesgue measure λ on $(\Omega, \boldsymbol{F})$, that is $m_0^0 \ll \lambda$, then for the least-favorable measures singled out by Lemma 1 $\hat{m}_0 \ll \lambda$ and $\hat{m}_1 \ll \lambda$ as well. In other words, if the nominal distributions have densities (pdfs), so do the least-favorable ones, although many elements of the uncertainty class in (A-1) may not have pdfs.

Appendix B

Establishing the Sufficient Conditions (28a)-(28b) for the Minimax Robustness of the Sequential Test of (5)

In this Appendix we establish (28a)-(28b), the sufficient conditions for (26) which expresses the minimax robustness (actually least-favorability) of the sequential test of (5) for the error probabilities.

Since $\frac{\partial \hat{g}}{\partial \pi_v} = 1/[A\pi_v+1]^2 > 0$, $\hat{g}$ is an increasing function of π_v. Consequently, from (A-3) of Appendix A we obtain

$$\mu(\hat{g},F_0) = \int \hat{g}(x)dF_0(x) \le \int \hat{g}(x)d\hat{F}_0(x) = \mu(\hat{g},\hat{F}_0) \tag{B-1a}$$

and from (A-4) we obtain

$$\mu(\hat{g},F_1) = \int \hat{g}(x)dF_1(x) \ge \int \hat{g}(x)d\hat{F}_1(x) = \mu(\hat{g},\hat{F}_1) \tag{B-1b}$$

which establish the desirable inequalities involving the means in conditions (28a) and (28b).

The corresponding proof for the variances in (28a) and (28b) is more complicated. We actually show that

$$\int [\hat{g}(x)-\mu(\hat{g},F_0)]^2 dF_0(x) \le \int [\hat{g}(x)-\mu(\hat{g},\hat{F}_0)]^2 dF_0(x) \le \int [\hat{g}(x)-\mu(\hat{g},\hat{F}_0)]^2 d\hat{F}_0(x) \tag{B-2a}$$

and

$$\int [\hat{g}(x)-\mu(\hat{g},F_1)]^2 dF_1(x) \le \int [\hat{g}(x)-\mu(\hat{g},\hat{F}_1)]^2 dF_1(x) \le \int [\hat{g}(x)-\mu(\hat{g},\hat{F}_1)]^2 d\hat{F}_1(x) \,. \tag{B-2b}$$

The left-hand-side inequalities in (B-2a)-(B-2b) follow from an application of the minimum variance principle of estimating a random variable $\hat{g}(X)$ by its mean $\mu(\hat{g},F_k)$ under cdf F_k. The right-hand-side inequalities in (B-2a)-(B-2b) follow from the dominance properties (A-3) and (A-4) of Appendix A, provided that the function $[\hat{g}(x) - \mu(\hat{g},\hat{F}_0)]^2$ is increasing in the Huber-Strassen derivative π_v and the function $[\hat{g}(x) - \mu(\hat{g},\hat{F}_1)]^2$ is decreasing in π_v. These last facts are established as follows. We notice that

$$\mu(\hat{g},\hat{F}_0) = \int \frac{\hat{f}_1(x)\hat{f}_0(x)dx}{A\hat{f}_1(x) + \hat{f}_0(x)} \quad \text{and} \quad \mu(\hat{g},\hat{F}_1) = \int \frac{\hat{f}_1^2(x)dx}{A\hat{f}_1(x) + \hat{f}_0(x)} \tag{B-3}$$

which implies

$$\hat{g}(x)-\mu(\hat{g},\hat{F}_0) = \frac{[1 - A\mu(\hat{g},\hat{F}_0)]\pi_v(x) - \mu(\hat{g},\hat{F}_0)}{A\pi_v(x) + 1} \tag{B-4a}$$

and

$$\frac{\partial}{\partial \pi_v}\left\{[\hat{g}(x) - \mu(\hat{g},\hat{F}_0)]^2\right\} = \frac{2\{[1 - A\mu(\hat{g},\hat{F}_0)]\pi_v(x) - \mu(\hat{g},\hat{F}_0)\}}{[A\pi_v(x) + 1]^3} \ge 0 \tag{B-5a}$$

if the conditions

$$\mu(\hat{g},\hat{F}_0) \le 0 \quad \text{and} \quad A > 0 \tag{A2}$$

hold. Similarly,

$$\hat{g}(x) - \mu(\hat{g},\hat{F}_1) = \frac{[1 - A\,\mu(\hat{g},\hat{F}_1)]\pi_v(x) - \mu(\hat{g},\hat{F}_1)}{A\,\pi_v(x) + 1} \qquad \text{(B-4b)}$$

and

$$\frac{\partial}{\partial \pi_v}\left\{[\hat{g}(x) - \mu(\hat{g},\hat{F}_1)]^2\right\} = \frac{2\{[1 - A\,\mu(\hat{g},\hat{F}_1)]\pi_v(x) - \mu(\hat{g},\hat{F}_1)\}}{[A\,\pi_v(x) + 1]^3} \le 0 \qquad \text{(B-5b)}$$

if the following conditions hold

$$\mu(\hat{g},\hat{F}_1) \ge 0 \quad \text{and} \quad A\,\mu(\hat{g},\hat{F}_1) \ge 1 \ . \qquad \text{(A3)}$$