SRC TR 87-157

Robust Distributed Discrete-Time Block and Sequential Detection in Uncertain Environments

by

E. Geraniotis

ROBUST DISTRIBUTED DISCRETE-TIME
BLOCK AND SEQUENTIAL DETECTION IN UNCERTAIN ENVIRONMENTS

Evaggelos Geraniotis

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

ABSTRACT

Two detectors making independent observations must decide which one of two hypotheses is true. Both fixed-sample-size (block) detection and sequential detection are considered. The decisions are coupled through a common cost function which for tests with fixed sample size consists of the sum of the error probabilities while for sequential tests it comprises the sum of the error probabilities and the expected sample sizes. The probability measures which govern the statistics of the i.i.d. observations belong to uncertainty classes determined by 2-alternating capacities.

A minimax robust (worst-case) design is pursued according to which the two detectors employ fixed-sample-size tests or sequential probability ratio tests whose likelihood ratios and thresholds depend on the least-favorable probability measures over the uncertainty class. For the aforementioned cost function the optimal thresholds of the two detectors turn out to be coupled. It is shown that, despite the uncertainty, the two detectors are thus guaranteed a minimum level of acceptable performance.

This research was supported in part by the Systems Research Center at the University of Maryland, College Park, through National Science Foundation CDR-85-00108 and in part by the Office of Naval Research under contract N00014-86-K-0013.

I. INTRODUCTION

In [1] and [2] distributed discrete-time fixed-sample-size (block) detection and sequential detection problems, respectively, were formulated and solved. The two detectors collect independent observations and make decisions which are coupled through a common cost function. Then, the optimal decisions are characterized by thresholds which are coupled. The hypothesis testing models considered in [1] and [2] assume perfect knowledge of the statistics of the observations.

In this paper we formulate similar problems for the case in which the observations are characterized by statistical uncertainty. Both fixed-sample-size (block) and sequential discrete-time robust detection problems are considered. Continuous-time distributed detection problems with known statistics are considered in [3] while similar problems with statistical uncertainty are treated in [4], the companion to this paper.

In particular the observations are assumed to have probability distributions (measures) which belong to 2-alternating capacity classes. The 2-alternating Choquet capacities classes include several useful uncertainty models like the ϵ-contaminated class [5], the total variation class [5], the band class [6] and the p-point class [7], which have been popular among the statisticians.

The design philosophy that we pursue for the problem above is that of minimax robustness. According to it a worst-case situation (operational conditions) are specified in terms of a performance criterion is identified and the optimal decision design for this situation is derived. Then, this decision design is employed independent of the actual conditions (which are not known, except for the fact that they belong to some strucured uncertainty class, e.g., the 2-alternating capacity class) and its performance under any other situation is better than that under the worst-case operational conditions.

Minimax robust signal processing techniques have received considerable attention in the last 15 years (see the tutorial in [8]). The selection of uncertainty classes determined by 2-alternating capacities is motivated by the fact that for the uncertainty models defined in [5]-[7] the least-favorable operational conditions (here probability measures) can be obtained in closed form as the general results of [9] indicate. In [9] the performance criterion is the Bayes risk or the error probabilities of the Neyman Pearson formulation of the hypothesis testing problem. The results of [10] complemented these of [9] by considering the Chernoff bounds on the error probabilities and by studying their asymptotic properties in the presence of uncertainty within 2-alternating capacities.

This paper is organized as follows. In Section II we formulate and solve the problem of robust distributed discrete-time detection with fixed sample size. Then in Section III we treat the case of robust distributed discrete-time sequential detection. In each section the distributed system and the uncertainty model are introduced first, then the case of detection under mismatch is considered, then the case of robust detection for finite sample sizes (which are fixed in Section II and random variables in Section III) is treated, and, finally, asymptotic results for large sample sizes are derived.

In all cases the robust tests are based on the likelihhod ratios between the least-favorable measures in the uncertainty class and the optimal decision making of the two detectors is coupled through their thresholds. For both the block and the sequential detection case we show that as the number of observations increases the joint cost function decreases exponentially to zero despite the uncertainty.

II. MINIMAX ROBUST DISTRIBUTED FIXED-SAMPLE-SIZE DETECTION

II.A Problem Formulation and Models of Uncertainty

Consider the following hypothesis testing problem of two simple hypotheses H_0 and H_1 with two decision-makers. Decision-maker i $(i = 1, 2)$ is equipped with a sensor and is faced with testing the hypotheses H_1 versus H_0:

$$H_0: \quad X_{i,l} \sim m_{0,i}, \quad l = 1,2,...,n$$
$$H_1: \quad X_{i,l} \sim m_{1,i}, \quad l = 1,2,...,n \tag{1}$$

In (1) $X_{i,l}$ denotes the l-th observation (sample), n is the number of samples, and $m_{j,i}$ (for $j = 0, 1$) defined on the sample space (Ω_i, B_i), and σ-filed is the probability measure which governs the statistics of the i.i.d. observations of the decision maker i under hypothesis H_j. It is assumed that the two decision-makers make independent observations so that the probability measures ($m_{0,1}$ and $m_{0,2}$) are mutually independent and so are ($m_{1,1}$ and $m_{1,2}$).

The probability measures $m_{0,i}$, $m_{1,i}$, for the two detectors ($i = 1, 2$), are only known to belong to uncertainty classes $M_{0,i}$ and $M_{1,i}$, respectively, which are determined by the 2-alternating capacities $v_{0,i}$ and $v_{1,i}$ (defined below) as

$$M_{j,i} = \left\{ m_{j,i} \in M_i \mid m_{j,i}(A) \leq v_{j,i}(A) \ , \ VA \in B_i \ , m_{j,i}(\Omega) = v_{j,i}(\Omega) \right\}, \tag{2}$$

where M_i is the class of measures on $(f\,at\,\Omega, B_i)$ and $j = 0, 1$ for the two hypotheses.

The decision making of detectors 1 and 2 is coupled through the following cost structure:

$$C(d_1, d_2; h) = \begin{cases} 0 \text{ for } & d_1 = d_2 = h \\ e \text{ for } & d_1 \neq d_2 \\ f \text{ for } & d_1 = d_2 \neq h \end{cases} , \tag{3}$$

where $d_1, d_2, h \in \{0,1\}$, e and f are non-negative constants, and we assume that $f > 2e$. Since the cost $[C(1,1;0) = C(0,0;1)]$ of wrong decisions by both detectors is expected to be considerably larger than the cost $[C(0,1;0) = C(1,0;0) = C(0,1;1) = C(1,0;1)]$ of a wrong decision by one of the detectors, this assumption does not impose a serious restriction on the generality of our problem formulation.

Next we define the 2-alternating capacities:

Definition: A positive set function v on a sample space Ω and assosciated σ-field B is called a **2-alternating capacity** if it is increasing, continuous from below, continuous from above on closed sets, and satisfies the conditions $v(\phi) = 0$, $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$. Suppose now that M is the class of measures on (Ω, B) and $m \in M$ is any such measure. Consider the uncertainty class which is determined by the 2-alternating capacity v as follows [compare with (2)]:

$$M_v = \left\{ m \in M \mid m(A) \leq v(A), \ \forall A \in B, m(\Omega) = v(\Omega) \right\}. \tag{4}$$

When Ω is compact several popular uncertainty models like ϵ-contaminated neighborhoods [5], total variation neighborhoods [5], band classes [6] and p-point classes [7] are special cases of this model.

Example: The ϵ-contaminated model [5]

$$M_\epsilon = \left\{ m \in M \mid m(A) = (1-\epsilon)\, m^0(A) + \epsilon \tilde{m}(A), \ \forall A \in B, m^0(\Omega) = \tilde{m}(\Omega) \right\}, \tag{5}$$

for $\epsilon \in [0, 1]$. Then $v(A) = (1-\epsilon)\, m^0(A) + \epsilon\, m^0(\Omega)$

Fundamental properties of these uncertainty models have been studied by Huber and Strassen [9]. We will state the relevant properties as a Lemma.

Lemma 1: Suppose v_0 and v_1 are 2-alternating capacities on (Ω, B) and M_0 and M_1 are the uncertainty classes determined by them as in (1). Then there exists a Lebesgue-measurable function $\pi_v : \Omega \to [\,0, \infty\,]$ such that

$$\theta\, v_0(\{\pi_v > \theta\}) + v_1(\{\pi_v \leq \theta\}) \leq \theta\, v_0(A) + v_1(A^c) \tag{6}$$

for all $A \in B$ and all $\theta \geq 0$. Furthermore there exist measures $(\hat{m}_0, \hat{m}_1)$ in $M_0 \times M_1$ such that

$$\hat{m}_0(\{\pi_v > \theta\}) = v_0(\{\pi_v > \theta\}) \tag{7}$$

$$\hat{m}_1(\{\pi_v \leq \theta\}) = v_1(\{\pi_v \leq \theta\}) \tag{8}$$

(that is, π_v is stochastically largest over M_0 under $\hat{m}_0$ and stochastically smallest over M_1 under $\hat{m}_1$) and π_v is a version of $d\hat{m}_1/d\hat{m}_0$ and is unique a.e. $[\,\hat{m}_0\,]$. The measures $(\hat{m}_0, \hat{m}_1)$ are termed the **least-favorable** measures over $M_0 \times M_1$.

Example: The ϵ-contaminated mixture uncertainty classes described by

$$M_j = \left\{\, m_j \in M \;\middle|\; m_j = (1-\epsilon_j)\, m_j^0 + \epsilon_j \tilde{m}_j \,,\ \tilde{m}_j(\Omega) = m^0(\Omega) \right\},\ j = 0, 1 \tag{9}$$

associated with the 2-alternating capacities

$$v_j(A) = \begin{cases} (1-\epsilon_j)\, m_j^0(A) + \epsilon_j & ,\ A \neq \phi \\ 0 & ,\ A = \phi \end{cases}, \tag{10}$$

have the least-favorable distributions

$$d\hat{m}_0/d\lambda = \begin{cases} (1-\epsilon_0)\, dm_0^0/d\lambda\ , & dm_1^0/dm_0^0 \leq c_2 \\ \dfrac{1-\epsilon_0}{c_2}\, dm_1^0/d\lambda\ , & c_2 < dm_1^0/dm_0^0 \end{cases} \tag{11}$$

$$d\hat{m}_1/d\lambda = \begin{cases} (1-\epsilon_1)\, dm_1^0/d\lambda\ , & c_1 < dm_1^0/dm_0^0 \\ c_1(1-\epsilon_1)\, dm_0^0/d\lambda\ , & dm_1^0/dm_0^0 \leq c_1 \end{cases} \tag{12}$$

and the Huber-Strassen derivative π_v

$$\pi_v = d\hat{m}_1/d\hat{m}_0 = \frac{1-\epsilon_1}{1-\epsilon_0} \min \left\{ c_2 , \max (c_1 , dm_1^0/dm_0^0) \right\} \tag{13}$$

where $0 \leq c_1 \leq c_2 < \infty$ are such that $\hat{m}_1(\Omega) = \hat{m}_0(\Omega) = 1$.

Let us now return to the hypothesis testing problem (1). Assuming that the a priori probabilities for the hypotheses H_0 and H_1 are λ and $1-\lambda$, respectively, and that likelihood ratio tests are employed, the average cost is

$$\begin{aligned} J(L_1^{(n)},L_2^{(n)},\tilde{\eta}_1,\tilde{\eta}_2) = \lambda\{ & e\,[m_{0,1}^{(n)}(\{L_1^{(n)}(\underline{X}_1)>\tilde{\eta}_1\}) + m_{0,2}^{(n)}(\{L_2^{(n)}(\underline{X}_2)>\tilde{\eta}_2\})] \\ & + (f-2e)m_{0,1}^{(n)}(\{L_1^{(n)}(\underline{X}_1)>\tilde{\eta}_1\})\cdot m_{0,2}^{(n)}(\{L_2^{(n)}(\underline{X}_2)>\tilde{\eta}_2\})\} \\ + (1-\lambda)\{ & e\,[m_{1,1}^{(n)}(\{L_1^{(n)}(\underline{X}_1)\leq\tilde{\eta}_1\}) + m_{1,2}^{(n)}(\{L_2^{(n)}(\underline{X}_2)\leq\tilde{\eta}_2\})] \\ & + (f-2e)m_{1,1}^{(n)}(\{L_1^{(n)}(\underline{X}_1)\leq\tilde{\eta}_1\})\cdot m_{1,2}^{(n)}(\{L_2^{(n)}(\underline{X}_2)\leq\tilde{\eta}_2\})\} \end{aligned} \tag{14}$$

In (14) $m_{j,i}^{(n)}$ are the n-th order extensions of the probability measures $m_{j,i}$ and characterize the observations $\underline{X}_i = (X_{1,i},X_{2,i},\ldots,X_{n,i})$ of the i-th decision-maker $(i = 1, 2)$ under hypothesis H_j $(j = 0, 1)$. By $L_i^{(n)}(\underline{X}_i) = (dm_{1,i}^{(n)}/dm_{0,i}^{(n)})(\underline{X}_i) = \prod_{l=1}^{n} (dm_{1,i}/dm_{0,i})(X_{l,i})$ we denote the likelihood ratio based on $\underline{X}_i$ of the i-th decision-maker and by $\tilde{\eta}_i$ its threshold.

The optimal thresholds for (14) are the pair (η_1,η_2) which minimizes the average cost function $J(L_1^{(n)},L_2^{(n)},\tilde{\eta}_1,\tilde{\eta}_2)$, that is

$$(\eta_1,\eta_2) = arg \min_{\tilde{\eta}_1,\tilde{\eta}_2} J(L_1^{(n)},L_2^{(n)},\tilde{\eta}_1,\tilde{\eta}_2) \tag{15}$$

Actually the likelihood ratio tests (LRTs) are the optimal policies for the two-decision-maker problem formulated above as stated in the following proposition

Proposition 1: Likelihood ratio tests (LRTs) with thresholds which minimize

$J(L_1^{(n)}, L_2^{(n)}, \tilde{\eta}_1, \tilde{\eta}_2)$ of (14) are optimal over all tests for the aforementioned common cost structure

Proof: The proof follows closely the corresponding proof of [1] about the optimality of the one-detector strategy (i.e., the likelihood ratio test) in this case of decision makers with independent observations, and will be ommitted.

II.B Robust Distributed Block Detection

The expression for the average cost function in (14) is valid for the case that there is no uncertainty in the statistics of the observations of the two decision makers. In the presence of uncertainty within the 2-alternating classes $M_{j,i}$ of (1), the likelihood ratios $\hat{L}_i^{(n)}$ and the thresholds $\hat{\eta}_i$, $i = 1, 2$, which are matched to the least-favorable measures $\hat{m}_{j,i}$ (singled out by Lemma 1) of the classes $M_{j,i}$ are employed. In this case the **average cost function under mismatch**--that is, when the statistics of the observations are actually governed by $m_{j,i} \in M_{j,i}$--is given by $J(\hat{L}_1^{(n)}, \hat{L}_2^{(n)}, \hat{\eta}_1, \hat{\eta}_2)$ which is obtained from (14), if we replace $L_i^{(n)}$ by $\hat{L}_i^{(n)}$ and η_i by $\hat{\eta}_i$, for $i = 1, 2$, and these thresholds are the solution to the minimization problem:

$$(\hat{\eta}_1, \hat{\eta}_2) = arg \min_{\tilde{\eta}_1, \tilde{\eta}_2} \hat{J}(\hat{L}_1^{(n)}, \hat{L}_2^{(n)}, \tilde{\eta}_1, \tilde{\eta}_2) , \tag{16}$$

where $\hat{J}(\hat{L}_1^{(n)}, \hat{L}_2^{(n)}, \tilde{\eta}_1, \tilde{\eta}_2)$ is the average cost when the likelihood ratios $\hat{L}_i^{(n)}$ ($i = 1, 2$ for the two detectors) are employed and the observations are distributed according to $\hat{m}_{j,i}$ ($j = 0, 1$ for the two hypotheses).

Lemma 1 provides the robust test and the least-favorable distributions for the one-dimensional (single observation) case and a single detector. For the case of n independent identically distributed (i.i.d) observations, we denote by $m_j^{(n)}$ $j = 0, 1$ the measures on (Ω^n, B^n) which are the n-th order extensions of the measures $m_j \in M_j$ of the classes defined in (4) and by $\hat{m}_j^{(n)}$ $j = 0, 1$ the n-th order extensions of the measures $\hat{m}_j$ singled out by Lemma 1. Then, the following result holds

Lemma 2: For any threshold $\eta > 0$ and any decision statistic $g^{(n)}$:

$$m_0^{(n)}(\{\hat{L}^{(n)}(\underline{X}) > \eta\}) \leq \hat{m}_0^{(n)}(\{\hat{L}^{(n)}(\underline{X}) > \eta\}) \leq \hat{m}_0^{(n)}(\{g^{(n)}(\underline{X}) > \eta\}) \tag{17}$$

$$m_1^{(n)}(\{\hat{L}^{(n)}(\underline{X}) \leq \eta\}) \leq \hat{m}_1^{(n)}(\{\hat{L}^{(n)}(\underline{X}) \leq \eta\}) \leq \hat{m}_1^{(n)}(\{g^{(n)}(\underline{X}) \leq \eta\}) \tag{18}$$

Where $\hat{L}^{(n)}(\underline{X}) = \dfrac{d\hat{m}_1^{(n)}}{d\hat{m}_0^{(n)}}(\underline{X}) = \prod_{l=1}^{n} \dfrac{d\hat{m}_1}{d\hat{m}_0}(X_l)$ is the likelihood ratio, $X_l \in \Omega$ is the l-th observation and $\underline{X} = (X_1, X_2, \cdots, X_n) \in \Omega^n$. Equations (17) and (18) imply that the test based on $\hat{L}^{(n)}$ is minimax robust.

Proof: For i.i.d. observations this Lemma was first proved in [5, section 4] and [9]; very recently a more straihtforward proof was given in [11].

For the case of two detectors and uncertainty within 2-alternating capacity classes the following result holds:

Proposition 2: The LRTs based on the least-favorable pairs of distributions $(\hat{m}_{0,i}, \hat{m}_{1,i})$ in the classes $(M_{0,i}, M_{1,i})$, $i = 1,2$ (for the two detectors) are minimax robust with respect to the average cost function defined in (14), that is

$$J(\hat{L}_1^{(n)}, \hat{L}_2^{(n)}, \hat{\eta}_1, \hat{\eta}_2) \leq \hat{J}(\hat{L}_1^{(n)}, \hat{L}_2^{(n)}, \hat{\eta}_1, \hat{\eta}_2) \leq \hat{J}(g_1^{(n)}, g_2^{(n)}, \eta_1, \eta_2) \tag{19}$$

where $g_i^{(n)}$ $(i = 0, 1)$ is any decision statistic operating on the observations $\underline{X}_i$.

Proof: The right-hand-side inequality in (19) is a straightforward application of Proposition 1 to the case characterized by $\hat{m}_{j,i}$ ($j = 0,1$ and $i = 1, 2$) for which the likelihood ratios are $\hat{L}_i^{(n)}$ and the optimal thresholds are $\hat{\eta}_i$.

The left-hand-side inequality in (19) is a consequence of Lemma 2. Specifically we apply the left-hand-side inequalities in (17) and (18) to the probability measures $(m_{0,i}, \hat{m}_{0,i})$ and $(m_{1,i}, \hat{m}_{1,i})$, respectively, of the two detectors $(i = 0, 1)$, and then use the definitons of the mismatch average cost function J and the average cost function $\hat{J}$ matched to the least-favorable pair of probability measures $(\hat{m}_{0,i}, \hat{m}_{1,i})$.

Note: The optimal thresholds $(\hat{\eta}_1, \hat{\eta}_2)$ can be determined from the error probabilities $\hat{\alpha}_i, \hat{\beta}_i$ $(i=1,2)$ for the least-favorable case of problem (1) by minimizing

$$\min\left\{\lambda\left[e(\hat{\alpha}_1+\hat{\alpha}_2)+(f-2e)\hat{\alpha}_1\hat{\alpha}_2\right]+(1-\lambda)\left[e(\hat{\beta}_1+\hat{\beta}_2)+(f-2e)\hat{\beta}_1\hat{\beta}_2\right]\right\}$$

under the constraints $\hat{\beta}_i = f_i(\hat{\alpha}_i)$ [operating receiver characteristic (ROC) for detector i], $0 \le \hat{\alpha}_i \le 1$, $0 \le \hat{\beta}_i \le 1$, and $\hat{\alpha}_i + \hat{\beta}_i \le 1$ for $i=1,2$.

II.C Asymptotic Performance

We will need the following two Lemmas which are concerned with the Chernoff upper bounds on the error probabilitties of hypothesis testing problems in the presence of uncertainty within 2-alternating capacity classes:

Lemma 3: Suppose that in the presence of uncertainty about the statistics of the aforementioned i.i.d. observations $\underline{X} = (X_1, X_2, \ldots, X_n)$ we employ a likelihood ratio test based on $\hat{L}^{(n)}$ defined above, then the error probabilities of the hypothesis testing problem of H_1 versus H_0 can be upperbounded by the **Chernoff bounds**:

$$m_0^{(n)}\{\hat{L}^{(n)}(\underline{X}) > n\gamma\} \le \exp\{-n[s\gamma + C_0(s,\hat{L})]\} \tag{20}$$

$$m_1^{(n)}\{\hat{L}^{(n)}(\underline{X}) \le n\gamma\} \le \exp\{-n[-s\gamma + C_1(s,\hat{L})]\} \tag{21}$$

where $\eta = n\gamma$ is the threshold, $\hat{L} = d\hat{m}_1/d\hat{m}_0$, and for all $s \in (0,1)$ the **Chernoff distances** $C_j(s,\hat{L})$ are given by

$$C_0(s,\hat{L}) = -\ln E_0\{\hat{L}^{s}\} \quad (22)$$

$$C_1(s,\hat{L}) = -\ln E_1\{\hat{L}^{-s}\}\ . \quad (23)$$

In (22)-(23) the expectations are with respect to the probability measures m_0 and m_1, respectively.

Proof: See [10].

Lemma 4: As the number of observations increases the Chernoff bounds of (20)-(21) converge exponentially to zero for all probability measures m_j $j = 0, 1$ belonging to uncertainty classes of the form (4).

Proof: See [10].

Note: Lemmas 3 and 4 are also valid for discrete-time stationary Gaussian observations with spectral uncertainty determined by 2-alternating capacity classes; see [10] for details.

The following proposition provides the desired asymptotic result for the mismatch average cost function $J(\hat{L}_1^{(n)},\hat{L}_2^{(n)},\hat{\eta}_1,\hat{\eta}_2)$ as the number of observations n increases:

Proposition 3: Under the assumptions of Proposition 2, the average cost function under mismatch converges to zero exponentially as the number of observations increases, despite the uncertainty; that is, $J(\hat{L}_1^{(n)},\hat{L}_2^{(n)},\hat{\eta}_1,\hat{\eta}_2)\to 0$, as $n\to\infty$ for all probability measures $m_{j,i}$ in the uncertainty class $M_{j,i}$ given by (1).

Proof: By applying Lemma 3 to the error probabilities of the hypothesis testing problem of each of the two detectors and using the definition of $J(\hat{L}_1^{(n)},\hat{L}_2^{(n)},\hat{\eta}_1,\hat{\eta}_2)$ we derive an upper bound on the average cost under mismatch in terms of the Chernoff bounds. This takes the form

$$J(\hat{L}_1^{(n)},\hat{L}_2^{(n)},\hat{\eta}_1,\hat{\eta}_2) \le \lambda\{e\,[\exp\{-n\,[s\,\hat{\gamma}_1+C_{0,1}(s,\hat{L}_1)]\} + \exp\{-n\,[s\,\hat{\gamma}_2+C_{0,2}(s,\hat{L}_2)]\}]$$

$$+ (f - 2e)\exp\{-n[s\hat{\gamma}_1 + C_{0,1}(s,\hat{L}_1)]\}\exp\{-n[s\hat{\gamma}_2 + C_{0,2}(s,\hat{L}_2)]\}\}$$

$$+ (1-\lambda)\{e[\exp\{-n[-s\hat{\gamma}_1 + C_{1,1}(s,\hat{L}_1)]\} + \exp\{-n[-s\hat{\gamma}_2 + C_{1,2}(s,\hat{L}_2)]\}]$$

$$+ (f - 2e)\exp\{-n[-s\hat{\gamma}_1 + C_{1,1}(s,\hat{L}_1)]\}\exp\{-n[-s\hat{\gamma}_2 + C_{1,2}(s,\hat{L}_2)]\}\} \quad (24)$$

where $\hat{\eta}_i = n\hat{\gamma}_i$ for $i=1,2$, is the threshold for the i-th detector, $\hat{L}_i = d\hat{m}_{1,i}/d\hat{m}_{0,i}$, and for all $s \in (0,1)$ the Chernoff distances $C_{j,i}(s,\hat{L}_i)$ for $j = 0, 1$ are given by $C_{0,i}(s,\hat{L}_i) = -\ln E_{0,i}\{\hat{L}_i^s\}$ and $C_{1,i}(s,\hat{L}_i) = -\ln E_{1,i}\{\hat{L}_i^{-s}\}$, where the expectations are with respect to the probability measures $m_{0,i}$ and $m_{1,i}$, respectively. Finally we apply Lemma 4 to (24) to complete the proof of Proposition 3.

III. MINIMAX ROBUST DISTRIBUTED SEQUENTIAL DETECTION

III.A Problem Formulation

The distributed sequential detection problem that we consider in this section has a lot of sismilarities with the problem considered in the previous section. The two decision makers are faced with the same hypothesis testing problem described in (1), where the uncertainty classes of (2) and the cost function of (3) remain the same. However, now there is also a cost for collecting data, which for the i-th decision maker ($i = 1, 2$) is defined by:

$$k_i \left[\lambda E_{0,i}\{N_i\} + (1-\lambda) E_{1,i}\{N_i\}\right], \tag{25}$$

where k_i ($i = 1, 2$) are nonnegative constants, $E_{j,i}$ denotes expectation with respect to the probability measure $m_{j,i}$ (under the hypothesis H_j, $j = 0, 1$, and for the i-th detector, $i = 1, 2$), the a priori probabilities for the hypotheses H_0 and H_1 are λ and $1-\lambda$, respectively, and the random variable N_i is the (discrete) stopping time (sample size) of the i-th detector; i.e., the number of samples necessary in order to reach a decision in favor of one of the two hypotheses.

Recall [12] that in the sequential detection of a single detector, the optimal test, termed the sequential probability ratio test (SPRT), consists of keep sampling till the likelihood ratio $L^{(n)}$ based on n samples of the observations exceeds B or falls below A--the two thresholds--in which case a decision is made in favor of H_1 or H_0, respectively.

Assuming that SPRTs are employed by both detectors, we can write the average cost as

$$J(L_1^{(N_1)}, L_2^{(N_2)}, \tilde{A}_1, \tilde{B}_1, \tilde{A}_2, \tilde{B}_2)$$

$$= \lambda \Big\{ k_1 E_{0,1}\{N_1 \mid L_1\} + k_2 E_{0,2}\{N_2 \mid L_2\}$$

$$+e\,[m_{0,1}^{(*)}(\{L_1^{(N_1)}(\underline{X}_1)\geq\tilde{B}_1\}) + m_{0,2}^{(*)}(\{L_2^{(N_2)}(\underline{X}_2)\geq\tilde{B}_2\})]$$

$$\left. +(f-2e)\,m_{0,1}^{(*)}(\{L_1^{(N_1)}(\underline{X}_1)\geq\tilde{B}_1\})\cdot m_{0,2}^{(*)}(\{L_2^{(N_2)}(\underline{X}_2)\geq\tilde{B}_2\})\right\}$$

$$+(1-\lambda)\left\{ k_1 E_{1,1}\{N_1 \mid L_1\} + k_2\, E_{1,2}\{N_2 \mid L_2\}\right.$$

$$+e\,[m_{1,1}^{(*)}(\{L_1^{(N_1)}(\underline{X}_1)\leq\tilde{A}_1\}) + m_{1,2}^{(*)}(\{L_2^{(N_2)}(\underline{X}_2)\leq\tilde{A}_2\})]$$

$$\left. +(f-2e)\,m_{1,1}^{(*)}(\{L_1^{(N_1)}(\underline{X}_1)\leq\tilde{A}_1\})\cdot m_{1,2}^{(*)}(\{L_2^{(N_2)}(\underline{X}_2)\leq\tilde{A}_2\})\right\} \tag{26}$$

where N_i **($i = 1, 2$) are (discrete) stopping times for the two detectors, that is, if** $L_i^{(n)}(\underline{X}_i)$, which is based on the n observations $\underline{X}_i$, is larger than or equal to $\tilde{B}_i$, it is decided that H_1 is true, the test terminates and $N_i = n$; if it is smaller than or equal to $\tilde{A}_i$, it is decided that H_0 is true, the test again terminates and $N_i = n$; otherwise, one more sample (observation) is collected and the procedure continues. $m_{j,i}^{(*)}$ is the probability measure which governs the observations of the i-th detector under hypothesis H_j ($j = 0, 1$) when the SPRT terminates after N_i samples, $L_i^{(N_i)}(\underline{X}_i) = \prod_{l=1}^{N_i} L_i(X_{l,i})$ is the likelihood ratio of the i-th detector based on the N_i samples of the i.i.d. observations $\underline{X}_i = (X_{1,i}, X_{2,i}, \ldots, X_{N_i,i})$, $L_i(X_{l,i}) = \frac{dm_{1,i}}{dm_{0,i}}(X_{l,i})$ is the likelihood ratio for one-sample, and $0 < \tilde{A}_i < 1 < \tilde{B}_i$ are the two thresholds for the SPRT of detector i. The notation $E_{j,i}\{N_i \mid L_i\}$ has been preferred over the notation $E_{j,i}\{N_i\}$ for the expected value of N_i under probability measure $m_{j,i}$ and an SPRT employing the likelihood ratio $L_i^{(n)} = dm_{1,i}^{(n)}/dm_{0,i}^{(n)}$, because it allows us to consider situations of mismatch, that is, when the likelihood ratio employed is not the one corresponding to the operating probability measures.

The optimal thresholds for (26) are the quadruple (A_1, B_1, A_2, B_2) which minimizes the average cost function $J(L_1^{(N_1)}, L_2^{(N_2)}, \tilde{A}_1, \tilde{B}_1, \tilde{A}_2, \tilde{B}_2)$, that is

$$(A_1, B_1, A_2, B_2) = arg \min_{\tilde{A}_1, \tilde{B}_1, \tilde{A}_2, \tilde{B}_2} J(L_1^{(N_1)}, L_2^{(N_2)}, \tilde{A}_1, \tilde{B}_1, \tilde{A}_2, \tilde{B}_2) \tag{27}$$

Actually the sequential probability ratio tests (SPRTs) are the optimal policies for the two-decision-maker problem formulated above as stated in the following proposition

Proposition 4: SPRTs with thresholds which minimize $J(L_1^{(N_1)}, L_2^{(N_2)}, \tilde{A}_1, \tilde{B}_1, \tilde{A}_2, \tilde{B}_2)$ of (26) are optimal over all tests for the aforementioned common cost structure.

Proof: The proof is provided in [2] for discrete-time sequential detection and in [3] for continuous-time sequential detection and it establishes the optimality of the one-detector strategy (i.e., the SPRT) in this case of decision makers with independent observations. It will be ommitted.

III.B. Robust Distributed Sequential Detection

The expression for the average cost function in (26) is valid for the case that there is no uncertainty in the statistics of the observations of the two decision makers. In the presence of uncertainty within the 2-alternating classes $M_{j,i}$ of (1), the likelihood ratios $\hat{L}_i^{(n)}$ and the thresholds $(\hat{A}_i, \hat{B}_i)$, $i = 1, 2$, which are matched to the least-favorable measures $\hat{m}_{j,i}$ (singled out by Lemma 1) of the classes $M_{j,i}$ are employed. In this case the **average cost function under mismatch**--that is, when the statistics of the observations are actually governed by $m_{j,i} \in M_{j,i}$--is given by $J(\hat{L}_1^{(N_1)}, \hat{L}_2^{(N_2)}, \hat{A}_1, \hat{B}_1, \hat{A}_2, \hat{B}_2)$ which is obtained from (26), if we replace $L_i^{(N_i)}$ by $\hat{L}_i^{(N_i)}$ and (A_i, B_i) by $(\hat{A}_i, \hat{B}_i)$, for $i = 1, 2$. These thresholds are the solution to the minimization problem:

$$(\hat{A}_1,\hat{B}_1,\hat{A}_2,\hat{B}_2) = arg \min_{\tilde{A}_1,\tilde{B}_1,\tilde{A}_2,\tilde{B}_2} \hat{J}(\hat{L}_1^{(N_1)},\hat{L}_2^{(N_2)},\tilde{A}_1,\tilde{B}_1,\tilde{A}_2,\tilde{B}_2) , \tag{28}$$

where $\hat{J}(\hat{L}_1^{(N_1)},\hat{L}_2^{(N_2)},\tilde{A}_1,\tilde{B}_1,\tilde{A}_2,\tilde{B}_2)$ is the average cost when SPRTs based on the likelihood ratios $\hat{L}_i^{(N_i)}$ and the thresholds $(\tilde{A}_i,\tilde{B}_i)$ ($i = 1, 2$ for the two detectors) are employed and the observations are distributed according to $\hat{m}_{j,i}$ ($j = 0, 1$ for the two hypotheses).

For sequential detection the following result also holds

Lemma 5: For i.i.d. observations with probability measures belonging to uncertainty classes of the form (4), **the sequential probability ratio test (SPRT) based on the likelihood ratio $\hat{L}^{(n)}(\underline{X})$ defined in Lemma 2 above and on the thresholds $\hat{A}$ and $\hat{B}$ ($\hat{A} < 1 < \hat{B}$) is** minimax robust for the error probabilities; that is, it satisfies the equations

$$m_0^{(*)}(\{\hat{L}^{(N)}(\underline{X})>\hat{B}\}) \le \hat{m}_0^{(*)}(\{\hat{L}^{(N)}(\underline{X})>\hat{B}\}) \le \hat{m}_0^{(*)}(\{g^{(N)}(\underline{X})>\hat{B}\}) \tag{29}$$

$$m_1^{(*)}(\{\hat{L}^{(N)}(\underline{X})\le\hat{A}\}) \le \hat{m}_1^{(*)}(\{\hat{L}^{(N)}(\underline{X})\le\hat{A}\}) \le \hat{m}_1^{(*)}(\{g^{(N)}(\underline{X})\le\hat{A}\}) \tag{30}$$

In (29)-(30) N is a stopping time for the SPRT. The measures $m_j^{(*)}$ and $\hat{m}_j^{(*)}$ for $j = 0, 1$ are the multi-dimensional extensions of the original measures m_j and $\hat{m}_j$, respectively, which are induced by the stopping time N defined above. Finally, $g^{(n)}$ is any other decision statistic, based on the n observations $\underline{X}$, which could be used in the aforementioned sequential test instead of the likelihood ratio.

Proof: See section 5 of [4].

Before stating and proving the basic result of this section we need to prove the following Lemma about the expected values of the sample sizes (stopping times) of the SPRTs under mismatch:

Lemma 6: Let $E_{j,i}[N_i \mid \hat{L}_i]$ represent the expected sample size of the SPRT of the i-th detector under hypothesis H_j based on the likelihood ratio $\hat{L}_i^{(n)}(\underline{X}_i) = \prod_{l=1}^{n} \hat{L}_i(X_{l,i})$, where

$\hat{L}_i(X_{l,i}) = \frac{d\hat{m}_{1,i}}{d\hat{m}_{0,i}}(X_{l,i})$, and on the thresholds $(\hat{A}_i, \hat{B}_i)$, when the i.i.d. observations are distributed according to the probability measure $m_{j,i}$. Then

$$E_{0,i}\{N_i \mid \hat{L}_i\} = \frac{\omega(\hat{\alpha}_i, \hat{\beta}_i, \alpha_i)}{E_{0,i}\{-\ln\hat{L}_i\}} \tag{31}$$

$$E_{1,i}\{N_i \mid \hat{L}_i\} = \frac{\omega(\hat{\beta}_i, \hat{\alpha}_i, \beta_i)}{E_{1,i}\{\ln\hat{L}_i\}}, \tag{32}$$

where

$$\omega(\hat{x}, \hat{y}, x) = (1-x)\ \ln\ \frac{1-\hat{x}}{\hat{y}} + x\ \ln\ \frac{\hat{x}}{1-\hat{y}}\ , \tag{33}$$

α_i, β_i are the mismatch error probabilities for the i-th detector under hypotheses H_0 and H_1, respectively, given by

$$\alpha_i = m_{0,i}^{(*)}(\{\hat{L}_i^{(N_i)} \geq \hat{B}\}) \tag{34}$$

and

$$\beta_i = m_{1,i}^{(*)}(\{\hat{L}_i^{(N_i)} \leq \hat{A}\})\ , \tag{35}$$

while $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the corresponding matched error probabilities, which are given by

$$\hat{\alpha}_i = \hat{m}_{0,i}^{(*)}(\{\hat{L}_i^{(N_i)} \geq \hat{B}\}) \tag{36}$$

and

$$\hat{\beta}_i = \hat{m}_{1,i}^{(*)}(\{\hat{L}_i^{(N_i)} \leq \hat{A}\})\ . \tag{37}$$

Proof: We prove only (31); (32) can be proved in a similar way. We write two different expressions for $E_{0,i}\{\ln\hat{L}_i^{(N_i)}(\underline{X}_i)\}$:

$$E_{0,i}\left\{\ln\hat{L}_i^{(N_i)}(\underline{X}_i)\right\} = E_{0,i}\left\{\ln\prod_{l=1}^{N_i}\hat{L}_i(X_{l,i})\right\} = E_{0,i}\left\{\sum_{l=1}^{N_i}\ln\hat{L}_i(X_{l,i})\right\}$$

$$= \bar{E}_{0,i}\left\{\tilde{E}_{0,i}\left\{\sum_{l=1}^{N_i} \ln\hat{L}_i(X_{l,i}) \mid N_i\right\}\right\}$$

$$= \bar{E}_{0,i}\left\{N_i \tilde{E}_{0,i}\left\{\ln\hat{L}_i(X_{l,i})\right\}\right\}$$

$$= \bar{E}_{0,i}\{N_i \mid \hat{L}_i\}\, E_{0,i}\{\ln\hat{L}_i\}, \tag{38}$$

where $\tilde{E}_{0,i}$ denotes expectation with respect to the measure $m_{0,i}$ governing the observations, whereas $\bar{E}_{0,i}$ denotes expectation with respect to the measure induced by the stopping-time variable N_i; since $\tilde{E}_{0,i}\{ln\hat{L}_i\}$ is a constant, it can be pulled out of the expectation $\bar{E}_{0,i}\{N_i \cdot\}$ in the equation proceeding the last one in (38)]; and

$$E_{0,i}\left\{\ln\hat{L}_i^{(N_i)}(\underline{X}_i)\right\} \approx m_{0,i}^{(*)}\left\{\hat{L}_i^{(N_i)} \geq \hat{B}\right\}\ln\hat{B} + m_{0,i}^{(*)}\left\{\hat{L}_i^{(N_i)} \leq \hat{A}\right\}\ln\hat{A}$$

$$= \alpha_i \ln\frac{1-\hat{\beta}_i}{\hat{\alpha}_i} + (1-\alpha_i)\ln\frac{\hat{\beta}_i}{1-\hat{\alpha}_i}$$

$$= -\omega(\hat{\alpha}_i, \hat{\beta}_i, \alpha_i). \tag{39}$$

In deriving (39) we used the definition of the SPRT, the definitions (34)-(35), the Wald's approximations

$$\hat{B}_i \approx \frac{1-\hat{\beta}_i}{\hat{\alpha}_i} \tag{40}$$

and

$$\hat{A}_i \approx \frac{\hat{\beta}_i}{1-\hat{\alpha}_i}\,, \tag{41}$$

and neglected the oveshoot phenomena [12]. Then (31) follows from (38) and (39).

For the case of uncertainty within 2-alternating capacity classes the following result holds:

Proposition 5: The SPRTs which employ thresholds (A_i, B_i) and a likelihood ratio $\hat{L}_i^{(n)}$ defined as in Lemma 6 which is based on the least-favorable pairs of distributions $(\hat{m}_{0,i}, \hat{m}_{1,i})$ in the classes $(\boldsymbol{M}_{0,i}, \boldsymbol{M}_{1,i})$, $i = 1,2$ (for the two detectors) are minimax robust with respect to the average cost function defined in (26), that is

$$J(\hat{L}_1^{(N_1)}, \hat{L}_2^{(N_2)}, \hat{A}_1, \hat{B}_1, \hat{A}_2, \hat{B}_2) \leq \hat{J}(\hat{L}_1^{(N_1)}, \hat{L}_2^{(N_2)}, \hat{A}_1, \hat{B}_1, \hat{A}_2, \hat{B}_2) \leq \hat{J}(g_1^{(N_1)}, g_2^{(N_2)}, A_1, B_1, A_2, B_2) \tag{42}$$

where $g_i^{(N_i)}$ ($i = 1, 2$) is any decision statistic operating on the observations $\underline{X}_i$, if for $i = 1, 2$ $\hat{\alpha}_i$ and $\hat{\beta}_i$ of (36)-(37) satisfy the following condition:

$$\ln \frac{1-\hat{\beta}_i}{\hat{\alpha}_i} >> -\hat{\beta}_i \ln \frac{\hat{\alpha}_i \hat{\beta}_i}{(1-\hat{\alpha}_i)(1-\hat{\beta}_i)} . \tag{43}$$

Proof: To prove the right-hand-side inequality in (42) we only need to use Proposition 4 for the optimality of the one-person stategies (the SPRTs). To prove the left-hand-side inequality in (42) it suffices, because of the definition of $J(\hat{L}_i^{(N_1)}, \hat{L}_i^{(N_2)}, \hat{A}_1, \hat{B}_1, \hat{A}_2, \hat{B}_2)$ to show that for $j = 0, 1$ and $i = 1, 2$

$$E_{j,i}\{N_i \mid \hat{L}_i\} \leq \hat{E}_{j,i}\{N_i \mid \hat{L}_i\} \tag{44}$$

and

$$m_{0,i}^{(*)}(\{\hat{L}^{(N_i)}(\underline{X}_i) > \hat{B}_i\}) \leq \hat{m}_{0,i}^{(*)}(\{\hat{L}^{(N_i)}(\underline{X}_i) > \hat{B}_i\}) \tag{45}$$

$$m_{1,i}^{(*)}(\{\hat{L}^{(N_i)}(\underline{X}_i) \leq \hat{A}_i\}) \leq \hat{m}_{1,i}^{(*)}(\{\hat{L}^{(N)}(\underline{X}) \leq \hat{A}_i\}) . \tag{46}$$

Since (45) and (46) follow from an application of Lemma 5 to the robust SPRT of the i-th detector, we only need to prove (44) in order to complete the proof of (42). Next, we prove

(44) for $j = 1$; a similar proof holds for $j = 0$. We write

$$E_{1,i}\{N_i \mid \hat{L}_i\} = \frac{1}{E_{1,i}\{\ln\hat{L}_i\}}\left[\beta_i \ln\frac{\hat{\alpha}_i\hat{\beta}_i}{(1-\hat{\alpha}_i)(1-\hat{\beta}_i)} + \ln\frac{1-\hat{\beta}_i}{\hat{\alpha}_i}\right]$$

$$\approx \frac{1}{E_{1,i}\{\ln\hat{L}_i\}}\ln\frac{1-\hat{\beta}_i}{\hat{\alpha}_i}$$

$$\leq \frac{1}{\hat{E}_{1,i}\{\ln\hat{L}_i\}}\ln\frac{1-\hat{\beta}_i}{\hat{\alpha}_i}$$

$$\approx \frac{1}{\hat{E}_{1,i}\{\ln\hat{L}_i\}}\left[\hat{\beta}_i \ln\frac{\hat{\alpha}_i\hat{\beta}_i}{(1-\hat{\alpha}_i)(1-\hat{\beta}_i)} + \ln\frac{1-\hat{\beta}_i}{\hat{\alpha}_i}\right]$$

$$= \hat{E}_{1,i}\{N_i \mid \hat{L}_i\} \tag{47}$$

In proving (47) we successively used the definition (32) of Lemma 6, condition (43), the inequality $E_{1,i}\{\ln\hat{L}_i\} \geq \hat{E}_{1,i}\{\ln\hat{L}_i\}$ $(i = 1, 2)$ which follows from the stochastic dominance property (8) of Lemma 1 when applied to the increasing function $\ln(\cdot)$ and the probability measures $m_{1,i}$ and $\hat{m}_{1,i}$, condition (43) again, and the definition (32) for the matched case $m_{1,i} = \hat{m}_{1,i}$.

Note: The optimal thresholds $(\hat{A}_i, \hat{B}_i)$ $(i = 1, 2)$ can be determined form the Wald's approximations (40)-(41) where the error probabilities $\hat{\alpha}_i$ and $\hat{\beta}_i$ are solutions to the minimization problem:

$$\min\left\{\lambda\left[k_1\frac{\omega(\hat{\alpha}_1,\hat{\beta}_1,\hat{\alpha}_1)}{\hat{E}_{0,1}\{-\ln\hat{L}_1\}} + k_2\frac{\omega(\hat{\alpha}_2,\hat{\beta}_2,\hat{\alpha}_2)}{\hat{E}_{0,2}\{-\ln\hat{L}_2\}} + e(\hat{\alpha}_1+\hat{\alpha}_2) + (f-2e)\hat{\alpha}_1\hat{\alpha}_2\right]\right.$$
$$\left. + (1-\lambda)\left[k_1\frac{\omega(\hat{\beta}_1,\hat{\alpha}_1,\hat{\beta}_1)}{\hat{E}_{1,1}\{\ln\hat{L}_1\}} + k_2\frac{\omega(\hat{\beta}_2,\hat{\alpha}_2,\hat{\beta}_2)}{\hat{E}_{1,2}\{\ln\hat{L}_2\}} + e(\hat{\beta}_1+\hat{\beta}_2) + (f-2e)\hat{\beta}_1\hat{\beta}_2\right]\right\}$$

under the constraints $0 \leq \hat{\alpha}_i \leq 1$, $0 \leq \hat{\beta}_i \leq 1$, and $\hat{\alpha}_i + \hat{\beta}_i \leq 1$ for $i = 1, 2$.

III.C Asymptotic Performance

III.C Asymptotic Performance

The following proposition provides a result on the asymptotic speed--which is defined as the sum of the asymptotic (for small error probabilities) stopping times of the two detectors--of the robust sequential test.

Proposition 6: Suppose that for the problem (1) with the uncertainty classes (2) and under the mismatch conditions of Proposition 5 above, the error probabilities $\hat{\alpha}_i$ and $\hat{\beta}_i$ (for $i=1,2$) approach zero. Then the sum of the asymptotic expected stoppong times--under mismatch and for the least-favorable case--satisfy the inequalities

$$\lambda\left[k_1\frac{-ln\hat{A}_1}{E_{0,1}\{-ln\hat{L}_1\}}+k_2\frac{-ln\hat{A}_2}{E_{0,2}\{-ln\hat{L}_2\}}\right]+(1-\lambda)\left[k_1\frac{ln\hat{B}_1}{E_{1,1}\{ln\hat{L}_1\}}+k_2\frac{ln\hat{B}_2}{E_{1,2}\{ln\hat{L}_2\}}\right]$$

$$\leq\lambda\left[k_1\frac{-ln\hat{A}_1}{\hat{E}_{0,1}\{-ln\hat{L}_1\}}+k_2\frac{-ln\hat{A}_2}{\hat{E}_{0,2}\{-ln\hat{L}_2\}}\right]+(1-\lambda)\left[k_1\frac{ln\hat{B}_1}{\hat{E}_{1,1}\{ln\hat{L}_1\}}+k_2\frac{ln\hat{B}_2}{\hat{E}_{1,2}\{ln\hat{L}_2\}}\right]$$

$$\leq\lambda\left[k_1\underline{\hat{E}}_{0,1}\{N_1\mid G_1\}+k_2\underline{\hat{E}}_{0,2}\{N_2\mid G_2\}\right]+(1-\lambda)\left[k_1\underline{\hat{E}}_{1,1}\{N_1\mid G_1\}+k_2\underline{\hat{E}}_{1,2}\{N_2\mid G_2\}\right] \tag{48}$$

where G_1 and G_2 are any other sequential tests different from the SPRT; $\underline{\hat{E}}_{j,i}$ denotes the limit of the expectation $\hat{E}_{j,i}$ as $\hat{\alpha}_i\rightarrow 0$ and $\hat{\beta}_i\rightarrow 0$.

Proof: As $\hat{\alpha}_i\rightarrow 0$ and $\hat{\beta}_i\rightarrow 0$, then α_i and β_i approach zero as well, since $\alpha_i\leq\hat{\alpha}_i$ and $\beta_i\leq\hat{\beta}_i$. Thus $J(\hat{L}_1,\hat{L}_2,\hat{A}_1,\hat{B}_1,\hat{A}_2,\hat{B}_2)$ (under mismatch) reduces to the first sum in (48), whereas $\hat{J}(\hat{L}_1,\hat{L}_2,\hat{A}_1,\hat{B}_1,\hat{A}_2,\hat{B}_2)$ reduces to the second sum in (48). The first sum is smaller than the second sum since $E_{0,i}\{-ln\hat{L}_i\}\geq\hat{E}_{0,i}\{-ln\hat{L}_i\}$ and $E_{1,i}\{ln\hat{L}_i\}\geq\hat{E}_{1,i}\{ln\hat{L}_i\}$ for $i=1,2$ because of the stochastic dominance inequalities (7) and (8) of Lemma 1. The second inequality (48) holds because of a theorem by Wald [12] (for the matched single detector case) which states that the SPRT has the minimum asymptotic speed (expected stopping time) among all sequential tests.

V. CONCLUSIONS

In this paper we considered two detectors making independent observations and trying decide which one of two hypotheses is true. Both fixed-sample-size (block) detection and sequential detection were considered. The decisions were coupled through a common cost function which for fixed-sample-size tests consisted of the sum of the error probabilities while for sequential tests it comprised the sum of the error probabilities and the expected sample sizes. The probability measures which govern the statistics of the i.i.d. observations belonged to uncertainty classes determined by 2-alternating capacities.

We were able to derive minimax robust (worst-case) designs according to which the two detectors employ fixed-sample-size tests or sequential probability ratio tests whose likelihood ratios and thresholds depend on the least-favorable probability measures over the uncertainty class (actually, the Huber-Strassen derivative and the least-favorable elements of the 2-alternating capacity class). For the aforementioned cost function the optimal thresholds of the two detectors turn out to be coupled. It was shown that, despite the uncertainty, the two detectors are guaranteed a minimum level of acceptable performance. In the case of block detection it was also shown, via Chernoff bounds, that for the aforementioned robust likelihood ratio test the two-detector cost function decreases exponentially to zero as the number of observations increases for all elements in the uncertainty class.

The results of this paper can be extended to several directions. First, they can be extended to situations of distributed detection where the two detectors are still making independent observations but the obseravtions for each detector are not i.i.d (they could be stationary Gaussian with spectral uncertainty, first-order Markov with uncertainty in the transition probabilities, or more generally dependendent observations). Second, we can formulate and solve similar problems in continuous-time (see [4]). Third, we can formulate and investigate problems of data fusion from distributed sensors in uncertain environments. Finally, we should relax the assumption of independent observations for the two detectors and formulate and attempt to solve similar problems for the case in which the observations of the two detectors are correlated.

REFERENCES

[1] R. R. Tenney, and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. on Aerospace and Electronic Systems,*, Vol. AES-17, pp. 501-510, July 1981.

[2] D. Teneketzis and Y.-C. Ho, "The decentralized Wald problem," 1985, preprint.

[3] A. LaVigna, A. M. Makowski and J. S. Baras, "A continuous-time distributed version of Wald' s sequential hypothesis testing problem," 1986, preprint.

[4] E. Geraniotis, "Robust distributed block and sequential continuous-time detection", submitted for presentation at the *26th Conference on Decision and Control,*, Los Angeles, December 1987.

[5] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758, 1965.

[6] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 242-247, 1981.

[7] K. S. Vastola and H. V. Poor, "On the p-point uncertainty class," *IEEE Trans. Information Theory,* vol. IT-30, pp.374-376, 1984.

[8] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing," *Proc. IEEE,* vol. 73, pp. 433-481, March 1985.

[9] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, vol. 1, pp. 251-265, 1973.

[10] E. Geraniotis, "Performance bounds for discrimination problems with uncertain statistics," *IEEE Trans. on Information Theory,* Vol. IT-31, pp. 703-707, September 1985.

[11] J. S. Sadofsky, "On the robust discriminatrion of Poisson measures," submitted to the *IEEE Trans. on Information Theory,* 1986, preprint.

[12] A. Wald, *Sequential Analysis,* Wiley, New York, 1947.