



**TECHNICAL
RESEARCH
REPORT**

SRC TR 88-99

On Minimax Robust Data Fusion

by

E. Geraniotis and Y. A. Chau

SYSTEMS RESEARCH CENTER

UNIVERSITY OF MARYLAND

COLLEGE PARK, MARYLAND 20742

ON MINIMAX ROBUST DATA FUSION

E. Geraniotis and Y. A. Chau

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

ABSTRACT

In this paper, minimax robust data fusion schemes based on discrete-time observations with statistical uncertainty are considered. The observations are assumed to be i.i.d and the decisions of all sensors independent when conditioned on the either of two hypotheses.. The statistics of the observations are only known to belong to uncertainty classes determined by 2-alternating Choquet capacities. Both cases of fixed-sample-size (block) data fusion and sequential data fusion are examined. For specific performance measures, three robust fusion rules: suboptimal, optimal and asymptotically optimal --as the number of sensors increases--are derived for the block data fusion case, and an asymptotically robust fusion rule is derived for the sequential data fusion case; these fusion rules are optimal in the class of rules employing likelihood ratio tests. In all situations the robust fusion rule makes use of likelihood ratios and thresholds which depend on the least-favorable probability distributions in the uncertainty class. In the limit of a large number of sensors, it is shown that the same threshold can be used by all sensors, which in turn simplifies the overall computation.

I. INTRODUCTION AND PROBLEM FORMULATION

Recently, distributed detection with or without a fusion center has attracted considerable attention([1]-[3]). In [1] it is shown that the optimal decision statistic for fixed-sample-size (block) detection is a likelihood ratio test with dependent thresholds which satisfy coupled equations. The computation of these thresholds is a very difficult task and becomes even more difficult when sequential test is used ([2]). Further in [1]-[3] the statistics of the observations are assumed to be known a priori, which can be unrealistic in practical situations. Frequently only imperfect knowledge is available in real situations. In [4] minimax robust distributed detection schemes without fusion were developed for a general uncertainty class, the class of 2-alternating Choquet capacities. In this paper minimax robust data fusion schemes are derived for the binary discrimination problem. The uncertainty class considered here is the same as that of [4], it contains many uncertainty models ([5]-[7]) and is described by

$$\mathcal{P} = \{P|P(S) \leq v(S), \forall S \in \mathcal{B}, P(\Omega) = v(\Omega)\} \quad (1)$$

where Ω is a sample space, \mathcal{B} the associated σ -field and v a 2-alternating capacity [5]. The properties relevant to our paper of this uncertainty model are stated in the following.

Lemma 1: Suppose v_0 and v_1 are two alternating capacities on (Ω, \mathcal{B}) and \mathcal{P}_0 and \mathcal{P}_1 are the uncertainty classes determined by them as in (1). Then there exists a Lebesgue-measurable function $\pi : \Omega \rightarrow [0, \infty]$ such that

$$\theta v_0(\{\pi_v > \theta\}) + v_1(\{\pi_v \leq \theta\}) \leq \theta v_0(S) + v_1(S^c) \quad (2)$$

for all $S \in \mathcal{B}$ and all $\theta \geq 0$. Furthermore there exist measures (\hat{P}_0, \hat{P}_1) in $\mathcal{P}_0 \times \mathcal{P}_1$ such

that

$$\hat{P}_0(\{\pi_v > \theta\}) = v_0(\{\pi_v > \theta\}) \quad (3)$$

$$\hat{P}_1(\{\pi_v \leq \theta\}) = v_1(\{\pi_v \leq \theta\}) \quad (4)$$

i.e., π_v is stochastically largest over \mathcal{P}_0 under \hat{P}_0 and stochastically smallest over \mathcal{P}_1 under \hat{P}_1 . The function π_v is a version of $d\hat{P}_1/d\hat{P}_0$ and unique a.e. $[\hat{P}_0 + \hat{P}_1]$. The measures (\hat{P}_0, \hat{P}_1) are termed the least-favorable measures over $\mathcal{P}_0 \times \mathcal{P}_1$.

The system models used here are similar to those of [2] and [9]. Each sensor makes a decision based on the observations $y_i(t)$ and sends a binary message u_i ($i = 1, 2, \dots, K$) which represents its decision ; then the fusion center makes the final decision according to a specific fusion rule. Hereafter the following assumption is made.

Assumption 1(A1): (i) The statistics of the observations conditioned on either hypothesis, $p_i(y_j(t))$ ($i = 0, 1; j = 1, 2, \dots, K$), are the same for each sensor j and all time instants t . (stationarity) (ii) $p_i(y_j(t))$ are independent across sensors and time instants.

Both cases of **fixed-sample-size (block)** data fusion and **sequential** data fusion are considered. Denote the likelihood ratio and the optimal threshold of sensor i by L_i and η_i , respectively, ($i = 1, \dots, K$), and by η_0 the optimal threshold of the fusion center. In the case of block data fusion, when K is finite, the performance index adopted here has the form

$$\begin{aligned} & C_K(L_i, \eta_i, i = 0, \dots, K) \\ &= c_1 P_{F_0}(L_i, \eta_i; i = 0, \dots, K) + c_2 P_{M_0}(L_i, \eta_i; i = 0, \dots, K) \end{aligned} \quad (5)$$

where P_{F_0} and P_{M_0} are the probabilities of false alarm and missing of the fusion center and c_1, c_2 are positive constants representing cost. Further let $p_m(m = 0, 1)$ denote the prior

probability of hypothesis H_m then $P_e = p_0 P_{F_0} + p_1 P_{M_0}$ is the error probability. When $K \rightarrow \infty$, since the error probability will approach zero, the performance is measured by

$$J_\infty(L_i, \eta_i = 0, 1, \dots,) = \lim_{K \rightarrow \infty} \frac{\ln P_e(L_i, \eta_i, i = 0, \dots, K)}{K} \quad (6)$$

which is the exponential rate of decrease of the error probability and is useful for asymptotic analysis ([2]&[8]). For sequential data fusion, besides the error probability, the expected value of the sample size of the fusion center must also be included as part of the performance measure.

This paper is organized as follows. In section II, robust block data fusion algorithms are derived for suboptimal fusion and for optimal fusion with equal local thresholds. In particular as the number of sensors $K \rightarrow \infty$, the asymptotically robust fusion rule is derived. In section III, an asymptotically robust sequential data fusion rule is developed. Finally,

II. MINIMAX ROBUST BLOCK DATA FUSION

II.A “AND”, “OR” and Optimal Fusion Rules

Two simple data fusion rules are : (1) The fusion center decides H_1 if all sensors transmit $u_i = 1$ for all i —termed “AND” rule. (2) The fusion center decides H_1 if at least one sensor transmits $u_i = 1$ for $1 \leq i \leq K$ —termed “OR” rule. Notice that for these two suboptimal fusion rules, there is no η_0 in the usual sense of thresholds. These two policies are suboptimal but easy to implement, especially when the number of sensors become very large.

Under (A1), the error probabilities of false alarm and missing can be expressed as

$$P_{F_0} = P(u_i = 1 ; i = 1, \dots, K | H_0) = \prod_{i=1}^K P_{F_i} \quad (7)$$

and

$$P_{M_0} = P(u_i = 0 ; 1 \leq i \leq K | H_1) = \sum_{i=1}^K P_{M_i} \quad (8)$$

for the “AND” fusion rule; and as

$$P_{F_0} = P(u_i = 1 ; 1 \leq i \leq K | H_0) = \sum_{i=1}^K P_{F_i} \quad (9)$$

and

$$P_{M_0} = P(u_i = 0 ; i = 1, \dots, K | H_1) = \prod_{i=1}^K P_{M_i} \quad (10)$$

for the “OR” fusion rule.

Usually for the purpose of performance evaluation the distribution of the observations is assumed to be known a priori. However, what we may have in practice is some imperfect knowledge about the statistics of observations. Let the statistics of data belong to the uncertainty set defined by equation (1). Then by lemma 1 the error probabilities P_{F_i} and

P_{M_i} of each sensor ($1 \leq i \leq K$) satisfy ² [5]

$$P_{F_i}(\hat{L}_i, \hat{\eta}_i) \leq \hat{P}_{F_i}(\hat{L}_i, \hat{\eta}_i) \leq \hat{P}_{F_i}(D_i, \eta_i) \quad (11)$$

and

$$P_{M_i}(\hat{L}_i, \hat{\eta}_i) \leq \hat{P}_{M_i}(\hat{L}_i, \hat{\eta}_i) \leq \hat{P}_{M_i}(D_i, \eta_i) \quad (12)$$

where $\hat{L}_i = d\hat{P}_{1i}/\hat{P}_{0i}$ is the worst-case likelihood ratio of the i th sensor (the measures \hat{P}_{1i} and \hat{P}_{0i} are singled out by lemma 1) and D_i is any other potential test of that sensor with the same observed data. For finite summations ([8]&[9]) or multiplications ([7]) over i , the direction of inequalities still holds the same as in (11) and (12). Thus the following proposition holds.

Proposition 1: Suppose the performance index of data fusion with “AND” or “OR” fusion rules is defined by equation (5). Then under (A1)

$$\begin{aligned} & C_K(\hat{L}_i, \hat{\eta}_i; i = 1, \dots, K) \\ & \leq \hat{C}_K(\hat{L}_i, \hat{\eta}_i; i = 1, \dots, K) \\ & \leq \hat{C}_K(D_i, \eta_i; i = 1, \dots, K) \end{aligned} \quad (13)$$

Remark 1: The threshold η_0 or $\hat{\eta}_0$ of the fusion center are missing from (13) since the “AND” or “OR” fusion rules do not involve common thresholds.

Next we consider the **optimal** robust fusion rule. Here the optimality is in the sense that the center uses the Neyman-Pearson test on the transmitted binary signals by the sensors. When all the sensors are alike and with the same level of error probabilities, i.e.,

²a variable with a hat means that it is a function of worst-case values; a probability with a hat means that the corresponding sensor operates under worst-case conditions.

$P_{F_i} = P_F$ and $P_{M_i} = P_M$, thus $\eta_i = \eta (i = 1, \dots, K)$, the Neyman-Pearson test becomes k - out - of - K rule [3] under (A1). Therefore the optimal fusion rule is

$$k \underset{H_0}{\overset{H_1}{>}} k^* \quad (14)$$

where k^* is a suitable threshold related to η_0 . In this case the overall error probabilities of false alarm and missing under mismatch [that is, \hat{k}^* is the threshold matched to the worst-case (\hat{P}_F, \hat{P}_M) , but (P_F, P_M) characterize the operating conditions] are

$$P_{F_0} = \sum_{i=\lceil k^* \rceil}^K \binom{K}{i} P_F^i (1 - P_F)^{K-i} \quad (15)$$

$$P_{M_0} = \sum_{i=1}^{\lceil k^* \rceil - 1} \binom{K}{i} (1 - P_M)^i P_M^{K-i} \quad (16)$$

Notice that k^* is a function of P_F , P_M and K ; P_F and P_M are functions of the number of samples \bar{n} that each sensor collects for making its decision. For proposition 2, another assumption is made (see Appendix)

Assumption 2(A2): $\hat{P}_F \leq \lceil k^* \rceil / K$ and $\hat{P}_M \leq 1 - \lceil k^* \rceil / K$ for all K .

Then from the above definition of P_{F_0} and P_{M_0} , proposition 2 follows.

Proposition 2: Suppose the thresholds of all sensors are equal, then under (A1) and (A2) the following inequalities are true

$$\begin{aligned} & C_K(\hat{L}_i, \hat{\eta}; i = 0, \dots, K) \\ & \leq \hat{C}_K(\hat{L}_i, \hat{\eta}; i = 0, \dots, K) \\ & \leq \hat{C}_K(D_i, \eta; i = 0, \dots, K) \end{aligned} \quad (17)$$

Proof: Taking derivative of (15) with respect to P_F , we have

$$\frac{dP_{F_0}}{dP_F} = \sum_{i=\lceil k^* \rceil}^K \binom{K}{i} P_F^{(i-1)} (1 - P_F)^{(K-i-1)} (i - K P_F)$$

If (A2) is satisfied, then $P_F \leq \hat{P}_F \leq \lceil \hat{k}^* \rceil / K$ as well and subsequently $dP_{F_0}/dP_F \geq 0$; thus P_{F_0} is an increasing function of P_F . By a similar argument, it can be shown that $dP_{M_0}/dP_M \geq 0$, thus P_{M_0} is also an increasing function of P_M . Hence C_K will be an increasing function of P_F and P_M under (A1) and (A2). Thus by (11) and (12), (17) holds.

Remark 2: The assumption (A2) is nonrestrictive for practical cases (refer to Appendix). If it is not satisfied, data fusion is of no use because P_{F_0} and P_{M_0} do not offer any improvement over P_F and P_M .

The above proposition only covers a special case [9]; in general the computation complexity of the optimal thresholds of the sensors is N-P complete. Thus our interest in the limiting case, $K \rightarrow \infty$, is justified.

II.B Asymptotic Behavior

When the number of sensors increases to infinity, the probability of error decreases to zero. Therefore the performance index adopted in asymptotic analysis should be the exponential rate of decrease of the error probability (6), where $P_\epsilon = p_0 P_{F_0} + p_1 P_{M_0}$. As stated in theorem 1 of [2], when $K \rightarrow \infty$ under the assumption 1 of [2], the same threshold η could be used in each sensor for optimal data fusion (to minimize $\ln P_\epsilon / K$). Hence $J_\infty(D_i, \eta_i; i = 0, 1, \dots) = J_\infty(D_i, \eta; i = 0, 1, \dots)$ and the error probabilities P_{F_0} and P_{M_0} satisfy (15) and (16). Then proposition 3 follows.

Proposition 3: Under (A1) and (A2), as $K \rightarrow \infty$, the performance index defined by (6) satisfies

$$J_\infty(\hat{L}_i, \hat{\eta}_0, \hat{\eta}; i = 0, 1, \dots)$$

$$\begin{aligned}
&\leq \hat{J}_\infty(\hat{L}_i, \hat{\eta}_0, \hat{\eta}; i = 0, 1, \dots) \\
&\leq \hat{J}_\infty(D_i, \eta_0, \eta; i = 0, 1, \dots)
\end{aligned} \tag{18}$$

Proof: Based on the above discussion, a similar proof to that of proposition 2 can be developed by replacing C_K with J_∞ .

III. ASYMPTOTICALLY ROBUST SEQUENTIAL DATA FUSION

In this section the SPRT (sequential probability ratio test) is adopted as the decision policy of the fusion center whereas the policies of all the sensors remain the same, i.e., likelihood ratio tests. The data used for fusion are the binary signals transmitted by the sensors.

The scheme of sequential data fusion is used to minimize the following cost function : the exponent of the sum of the error probability and the average sample sizes under both hypotheses. This kind of fusion rule has not yet been shown to be the optimal one, but for robustness in the class of likelihood ratio tests, it is a reasonable choice for the sequential fusion rule. Thus, in this case, the expected value of the sample size N_0 of the fusion center (that is, the number of times the fusion center collects the decisions of the sensors before reaching its decision) enters into the performance measure.

Again let the statistics of observations conditioned on either hypothesis belong to the uncertainty model defined by (1). When $K \rightarrow \infty$, under suitable conditions, it can be shown that the worst-case pair of statistics of observations is still the one singled out by lemma 1, if the performance index for robustness has the following form

$$\begin{aligned}
 & J_{\infty}(L_i, \eta_i; i = 0, 1, \dots) \\
 &= \lim_{K \rightarrow \infty} \frac{1}{K} \ln \left\{ \sum_{m=0}^1 p_m E_m \{N_0 | L_i, \eta_i; i = 0, \dots, K\} - 1 \right. \\
 & \quad \left. + P_e(L_i, \eta_i; i = 0, \dots, K) \right\}
 \end{aligned} \tag{19}$$

where $\eta_0 = (A, B)$, $(A < 1 < B)$ are the optimal thresholds of the SPRT used by the fusion center and $\eta_i, i = 1, \dots, K$ are the optimal thresholds used by the sensors. As $K \rightarrow \infty$,

the expected sample size approaches 1 and the error probability approaches 0. Thus (19) is the exponent of the sum of these two terms (i.e., the sum is of the form $\exp(J_\infty K)$) and is an extension of the performance measure in [2]. In the following, we abbreviate $(L_i, \eta_i; i = 0, \dots, K)$ as $\underline{\eta}$ where $\underline{\eta} = (\eta_1, \eta_2, \dots, \eta_K)$ and $(L_i, \eta_0, \eta_i; i = 0 \dots, K)$ as η for those expressions inside the conditioning in the expected sample size and inside the error probability. Define for $0 < s < 1$

$$\mu_i(\eta_i, s_i) = \ln \sum_{l=0}^1 f_{i0}(l)^{1-s_i} f_{i1}(l)^{s_i}$$

where $f_{im}(l)$ is the probability mass function under hypothesis $H_m (m = 0, 1)$ of the binary message $l \in \{0, 1\}$ that the i -th sensor sends to the fusion center; $f_{im}(\cdot)$ is stationary, independent of time. Let (η, s) minimize $\mu_i(\eta_i, s_i)$ for $i = 1, 2, \dots$ (i.e., $\mu(\eta, s) \leq \mu_i(\eta_i, s_i)$ for all i) and $r = e^\mu$. Before we proceed, we state the assumption 1 of [2].

Assumption 3(A3): (i) $|\mu(\eta, s)| < \infty, \forall \eta, s$. (ii) There exists a constant R such that $|\ddot{\mu}(\eta, s)| \leq R, \forall (\eta, s) \in [0, \infty] \times [0, 1]$, where a dot stands for differentiation with respect to s .

Notice that the function μ has the same form as that in [2] and [10]. In fact, we can obtain a similar result to theorem 1 of [2] for (19). First we derive the upper and the lower bounds on the expected sample size.

Lemma 2: Suppose (A1) and (A3) are satisfied , then for $m = 0, 1$

$$E_m[N_0|\eta] \leq 1 + \frac{e^{-sd_m} r(\eta, s)^K}{1 - r(\eta, s)^K} \quad (20)$$

$$E_m[N_0|\underline{\eta}] \geq 1 + c_1 \exp[K\mu(\eta, s) - (2KR)^{1/2}] \quad (21)$$

where $d_0 = a, d_1 = -b$ and c_1 is a constant defined in the course of the proof.

Proof: Denote by u_{ij} the message send by sensor i at time instant j and by f_{im} its (stationary) distribution under hypothesis $H_m (m = 0, 1)$. Further let

$$Z_i(n) = \sum_{j=1}^n \ln \frac{f_{i1}(u_{ij})}{f_{i0}(u_{ij})} ; i = 1, \dots, K$$

be the log-likelihood ratio based on n decisions of the i -th sensor and define by $g(K, n) = \sum_{i=1}^K Z_i(n)$ the overall sufficient statistic used by the fusion center .Then the left-hand side of (20) becomes ($a = \ln A, b = \ln B$)

$$\begin{aligned} E_m[N_0|\eta] &= \sum_{n=0}^{\infty} P_m(N_0 > n) \\ &= 1 + \sum_{n=1}^{\infty} P_m(N_0 > n) \\ &= 1 + \sum_{n=1}^{\infty} P_m(a < \sum_{i=1}^K Z_i(n) < b) \\ &= 1 + \sum_{n=1}^{\infty} P_m(a < g(K, n) < b) \end{aligned}$$

Only the proof of the case $m = 0$ will be given, since a similar proof can be obtained for $m = 1$. Suppose the same threshold η is used in all sensors, then

$$\begin{aligned} &\sum_{n=1}^{\infty} P_0(a < g(K, n) < b) \\ &\leq \sum_{n=1}^{\infty} P_0(a < g(K, n)) \\ &\leq \sum_{n=1}^{\infty} E_0[e^{s(g(K, n) - a)}] \\ &= \sum_{n=1}^{\infty} e^{-sa} r(\eta, s)^{Kn} \\ &= \frac{e^{-sa} r(\eta, s)^K}{1 - r(\eta, s)^K} \end{aligned}$$

where the second inequality is the Chernoff bound, and the equality is from the sum of a geometric series ($r(\eta, s) < 1$). Hence we have (20).

On the other hand, for the proof of (21), suppose different thresholds $\underline{\eta}$ are used by the sensors, then

$$\begin{aligned} & \sum_{n=1}^{\infty} P_0(a < g(K, n) < b) \\ & \geq P_0(a < g(K, 1) < b) \\ & = P_0(a < \sum_{i=1}^K \ln \frac{f_{i1}(u_{i1})}{f_{i0}(u_{i1})} < b) \end{aligned}$$

where u_{ij} are *i.i.d* for all j (across time). Let $\mathcal{U}_K = (a < g(K, 1) < b)$. Then by (3.20)&(3.40) of [10] and under assumption 1 of [2]

$$\begin{aligned} & P_0(\mathcal{U}_K) \\ & \geq c_1 \exp\left[\sum_{i=1}^K \mu_i(\eta_i, s_i) - (2 \sum_{i=1}^K \ddot{\mu}_i(\eta_i, s_i))^{1/2}\right] \\ & \geq c_1 \exp[K\mu(\eta, s) - (2KR)^{1/2}] \end{aligned}$$

where Q_s and \mathcal{Y}_s have been defined in [10] and

$0 < c_1 = \sum_{\mathcal{U}_K \cap \mathcal{Y}_s} Q_s(Y) < \infty$. Thus we have (21).

Next, let us consider the bounds on the error probability.

Lemma 3: Suppose (A1) and (A3) are satisfied,

$$P_e(\eta) \leq e^{K\mu(\eta, s)} \quad (22)$$

$$P_e(\underline{\eta}) \geq c_2 \exp\{K\mu(\eta, s)E[N_0] + (K\mu(\eta, s))^{1/2}E[N_0^{1/2}]\} \quad (23)$$

where $E[N_0] = \sum_{m=0}^1 p_m E_m[N_0]$ and c_2 is a constant defined in the course of the proof.

Proof: Because $A < 1 < B$,

$$P_0(L_0 \geq B|N_0 = n) < P_0(L_0 > 1|N_0 = n);$$

$$P_1(L_0 \leq A|N_0 = n) < P_1(L_0 < 1|N_0 = n)$$

Suppose the same threshold η is used for all sensors, then by section III of [10], we have

$$\begin{aligned}
& P_e(\eta) \\
&= \sum_{n=1}^{\infty} P_e(\eta|N_0 = n)P(N_0 = n) \\
&\leq \sum_{n=1}^{\infty} P(N_0 = n)[P_0(L_0 > 1|N_0 = n) + P_1(L_0 \leq 1|N_0 = n)] \\
&\leq \sum_{n=1}^{\infty} P(N_0 = n)e^{Kn\mu(\eta,s)} \\
&\leq E[e^{KN_0\mu(\eta,s)}] \\
&\leq e^{K\mu(\eta,s)}
\end{aligned}$$

since $\mu(\eta, s) < 0$ and $N_0 \geq 1$. Then we show (23). Denote by $\tilde{\eta}$ the optimal threshold for block fusion under known $N_0 = n$, then

$$\begin{aligned}
& p_0P_0(L_0 \geq B|N_0 = n) + p_1P_1(L_0 \leq A|N_0 = n) \\
&= p_0P_0(L_0 > A|N_0 = n) + p_1P_1(L_1 \leq A|N_0 = n) \\
&\geq p_0P_0(L_0 > \tilde{\eta}|N_0 = n) + p_1P_1(L_0 \leq \tilde{\eta}|N_0 = n)
\end{aligned}$$

According to III of [10], we have

$$\begin{aligned}
& P_e(\underline{\eta}) \\
&= \sum_{n=1}^{\infty} [p_0P_0(\underline{\eta}|N_0 = n)P_0(N_0 = n) \\
&\quad + p_1P_1(\underline{\eta}|N_0 = n)P_1(N_0 = n)] \\
&\geq \sum_{n=1}^{\infty} [p_0P_0(L_0 > \tilde{\eta}|N_0 = n)P_0(N_0 = n) \\
&\quad + p_1P_1(L_0 \leq \tilde{\eta}|N_0 = n)P_1(N_0 = n)] \\
&\geq \sum_{n=1}^{\infty} \sum_{m=0}^1 \{p_mP_m(N_0 = n) \exp[\sum_{i=1}^K \mu_i(\eta_i, s_i)]
\end{aligned}$$

$$\begin{aligned}
& -(2 \sum_{i=1}^K \ddot{\mu}_i(\eta_i, s_i))^{1/2} \sum_{\mathcal{Y}_m \cap \mathcal{Y}_s} Q_s(Y) \} \\
& \geq \sum_{m=0}^1 c_2 p_m E_m \{ \exp [K N_0 \mu(\eta, s) - (2 \sum_{i=1}^K \ddot{\mu}(\eta_i, s_i))^{1/2}] \}
\end{aligned}$$

where $\mathcal{Y}_i (i = 0, 1)$ were defined in [10]. Let

$$c_2 = \min \{ \exp \{ E[\sum_{\mathcal{Y}_0 \cap \mathcal{Y}_s} Q_s(Y)] \}, \exp \{ E[\sum_{\mathcal{Y}_1 \cap \mathcal{Y}_s} Q_s(Y)] \} \}$$

Using (A3) and the convexity of exponential function, we have

$$\begin{aligned}
P_\varepsilon(\underline{\eta}) & \geq c_2 \exp \{ E[K N_0 \mu(\eta, s) - (2 K N_0 R)^{1/2}] \} \\
& = c_2 \exp \{ K \mu(\eta, s) E[N_0] - (2 K R)^{1/2} E[N_0^{1/2}] \}
\end{aligned}$$

which is (24). According to lemmas 2 and 3, it can be shown the performance (measured by J_∞) will not be hurt if the same threshold is used by all sensors, and this value could be computed via a simple method (optimize μ_i with respect to η, s). Thus the following proposition holds.

Proposition 4: Under (A1) and (A3),

$$\begin{aligned}
& \lim_{K \rightarrow \infty} \frac{1}{K} \ln \{ E[N_0 | \underline{\eta}] - 1 + P_\varepsilon(\underline{\eta}) \} \\
& = \lim_{K \rightarrow \infty} \frac{1}{K} \ln \{ E[N_0 | \eta] - 1 + P_\varepsilon(\eta) \}
\end{aligned} \tag{24}$$

Proof: This is a direct result from (20)-(23) of the above two lemmas. By using the lower bounds in (21) and (23) and the facts that

$$\ln(a_1 + a_2 + a_3) \geq \frac{1}{3}(\ln a_1 + \ln a_2 + \ln a_3) + \ln 3$$

and $\lim_{K \rightarrow \infty} E_m[N_0 | \eta] = 1$, we have

$$\lim_{K \rightarrow \infty} \frac{1}{K} \ln \{ E[N_0 | \underline{\eta}] - 1 + P_\varepsilon(\underline{\eta}) \}$$

$$\begin{aligned}
&\geq \lim_{K \rightarrow \infty} \frac{1}{K} \left\{ \sum_{m=0}^1 \{ \ln p_m + \ln [E_m[N_0|\underline{\eta}] - 1] \} + \ln [P_\varepsilon(\underline{\eta})] \right\} \\
&\geq \mu(\eta, s)
\end{aligned}$$

Then by substituting from the upper bounds in (20) and (22) and taking the limit, we will have

$$\begin{aligned}
&\lim_{K \rightarrow \infty} \frac{1}{K} \ln \{ E[N_0|\underline{\eta}] - 1 + P_\varepsilon(\underline{\eta}) \} \\
&\leq \mu(\eta, s) \\
&\leq \lim_{K \rightarrow \infty} \frac{1}{K} \ln \{ E[N_0|\underline{\eta}] - 1 + P_\varepsilon(\underline{\eta}) \}
\end{aligned}$$

The reverse inequality also holds by definition.

Thus the same threshold can be used for all sensors; consequently $P_{F_i} = P_F$ and $P_{M_i} = P_M$ for $i = 1, \dots, K$. Let n_0 denote the number of times that the fusion center collects the decision of the K sensors before making its own decision. Then in analogy to the k - out - of - K rule of block data fusion, the decision test of the fusion center becomes

$$\begin{aligned}
&\text{decide } H_1 \quad \text{if } k(n_0) \geq b^* \\
&\text{defer} \quad " \quad b^* > k(n_0) > a^* \\
&\text{decide } H_0 \quad " \quad a^* \geq k(n_0)
\end{aligned}$$

where $k(n_0)$ is the number of sensors which have transmitted '1' by the n_0 -th time the fusion center has collected the decisions of the sensors and the thresholds (a^*, b^*) are

induced by (a, b) and are functions of P_F , P_M and K . Then P_{F_0} and P_{M_0} have forms

$$P_{F_0} = \sum_{i=\lceil b^* \rceil}^K \binom{K}{i} P_F^i (1 - P_F)^{(K-i)} \quad (25)$$

$$P_{M_0} = \sum_{i=1}^{\lceil a^* \rceil - 1} \binom{K}{i} P_M^{(K-i)} (1 - P_M)^i \quad (26)$$

Similar to (A2) the following assumption is made for robustness (see Appendix).

Assumption 4(A4): For sequential data fusion $\hat{P}_F \leq \lceil \hat{b}^* \rceil / K$ and $\hat{P}_M \leq 1 - \lceil \hat{a}^* \rceil / K$ for all K .

Finally, the following assumption is necessary for robustness (see [4]).

Assumption 5(A5): $\ln \frac{1 - \hat{P}_M}{\hat{P}_F} \gg \hat{P}_M \ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F}$.

Proposition 5: Under (A1), (A3) and assumption 1 of [2], the performance index defined by (20) satisfies

$$\begin{aligned} & J_\infty(\hat{L}_i, \hat{\eta}_0, \hat{\eta}; i = 0, 1, \dots) \\ & \leq \hat{J}_\infty(\hat{L}_i, \hat{\eta}_0, \hat{\eta}; i = 0, 1, \dots) \\ & \leq \hat{J}_\infty(D_i, \eta_0, \eta; i = 0, 1, \dots) \end{aligned} \quad (27)$$

Proof: By proposition 4, the same threshold can be used by all sensors to obtain asymptotically optimal sequential data fusion. Then P_{F_0} and P_{M_0} will have the forms described by (22) and (23). Therefore a similar proof to that of proposition 2 can be developed corresponding to the exponent of the error probability. According to proposition 5 of [4], the least-favorable pair of distributions for expected sample sizes are the same as those for the error probabilities under (A4). Thus the saddle point of performance measure (19) can be reached by the same least-favorable pair of proposition 2 in section II.

IV. CONCLUSION

In this paper we formulated and solved several problems of robust data fusion for situations characterized by incomplete knowledge of the probability distributions of the discrete-time observations. The uncertainty in the probability distributions was modeled by 2-alternating capacity classes. For robust block data fusion, suboptimal, optimal, and asymptotically optimal—as the number of sensors increases—schemes were derived. For sequential data fusion an asymptotically robust scheme was developed. This scheme is robust in the sense that the sensors use likelihood ratio tests and the fusion uses the sequential probability ratio test (SPRT). In all situations the robust fusion rule use likelihood ratios and thresholds that depend on the pair of least-favorable distributions in the uncertainty class. The advantage of asymptotic analysis manifests itself in that the optimal thresholds of the various sensors turn out to be the same, which in turn reduces considerably the required computation effort.

REFERENCES

- [1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-17, pp. 501-510, July 1981.
- [2] J. N. Tsitsiklis, "On threshold rules in decentralized detection," *Proceedings of 25th CDC*, pp. 232-236, December, 1986.
- [3] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-22, pp. 98-101, January 1986.
- [4] E. Geraniotis, "Robust distributed discrete-time block and sequential detection," *Proceeding of the 1987 Conference on Information Sciences and Systems*, pp. 354-360, John Hopkins Univ., March 1987.
- [5] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities", *Ann. Statist.*, Vol. 1, pp. 251-265, 1973.
- [6] S. A. Kassam, "Robust hypothesis testing for bounded classes of probabilities densities," *IEEE Transactions on Information Theory*, Vol. IT-27, pp. 242-247, 1981.
- [7] 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.
- [8] J. S. Sadowsky, "An asymptotically least-favorable Chernoff bound for a large class of dependent data processes," *IEEE Transactions on Information Theory*, Vol. IT-33, pp. 52-61, January 1987.
- [9] V. V. S. Sarma and K. A. Rao, "Decentralized detection and estimation in distributed sensor systems," *Proceedings of the IEEE Cybernetics and Society Conference*, pp. 438-441, 1983.
- [10] C. E. Shannon, R. G. Gallager and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Information and Control*, pp. 65-103, 1967.

APPENDIX

In (A2), it follows from [2] that

$$k^* = \left[\ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F} \right]^{-1} (\eta_0 + K \ln \frac{1 - \hat{P}_F}{\hat{P}_M})$$

and for large K

$$[k^*]/K \approx \left[\ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F} \right]^{-1} \ln \frac{1 - \hat{P}_F}{\hat{P}_M}$$

If $\eta_0 = 0$, $\hat{P}_M = \hat{P}_F$, then $k^* = K/2$ and $[k^*]/K \approx 1/2$. Thus (A2) is not restrictive.

Similarly, in (A3)

$$a^* = \left[\ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F} \right]^{-1} (a + K \ln \frac{1 - \hat{P}_F}{\hat{P}_M})$$

$$b^* = \left[\ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F} \right]^{-1} (b + K \ln \frac{1 - \hat{P}_F}{\hat{P}_M})$$

and for large K

$$[a^*] \approx \left[\ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F} \right]^{-1} \ln \frac{1 - \hat{P}_F}{\hat{P}_M}$$

$$[b^*] \approx \left[\ln \frac{(1 - \hat{P}_M)(1 - \hat{P}_F)}{\hat{P}_M \hat{P}_F} \right]^{-1} \ln \frac{1 - \hat{P}_F}{\hat{P}_M}$$

If $\hat{P}_M = \hat{P}_F$, as $K \rightarrow \infty$, $[a^*]/K \approx 1/2$, $[b^*]/K \approx 1/2$.

