Distributed Detection and Data Fusion

Peter Willett
University of Connecticut
peter.willett@uconn.edu

I. Introduction

The initial paper on the subject of distributed detection, by Tenney and Sandell [27], showed that under a fixed fusion rule, for two sensors with one bit outputs, the optimal Bayes sensor decision rule is a likelihood ratio test. In [5], it is shown that the optimal fusion rule for N sensors is a likelihood ratio test on the data received from the sensors. Reibman and Nolte [23] and Hoballah and Varshney [10] have generalized the results in [27] to N sensors with optimal fusion, again with the restriction of one bit sensor outputs. Hoballah and Varshney [11] have also investigated system optimization under the aforementioned conditions for a mutual information criterion. The restriction of the sensor outputs to single bits seems unduly harsh since it implies either very rapid decision rates or extremely narrowband channels. In this paper, we remove this restriction and assume instead that the i^{th} sensor produces a $\log_2 M_i$ bit output. The authors have derived some results concerning this more general case in [35]; using different techniques, Tsitsiklis has come to conclusions similar to those presented in this paper [29], [30], as have, for example Thomopoulos, Viswanathan, et al. [28], [32].

As always, when speaking of optimality, it is necessary to impose a criterion for judgment. For detectors, several such measures have been used. In this paper, we study the structure of optimal decentralized detectors for the Neyman-Pearson, Bayes, Ali-Silvey, and mutual (Shannon) information criteria. We assume that, conditioned on the actual hypothesis (state of nature), the random processes observed at any two different sensors are independent. Given this assumption we show that for each criterion, the optimal strategy is to quantize the local likelihood ratio at the sensors (to the maximum number of bits allowable), and transmit this result to the fusion center. The fusion center then performs a likelihood ratio test on this received data.

That to quantize the likelihood ratio is the optimal thing to do is scarcely a surprising result. Even prior to a proof of its optimality it was used by a number of researchers; and most certainly an excellent and rigorous proof is available in [29]. In that work, however, reference is made to the different and equally-rigorous proof in [35], and since [35] has never been published archivally, it seems appropriate to offer it in the present paper. Certainly the field has not stood still since [35], [29], and some papers we particularly admire are [3], [6], [4]; but to present a complete bibliography is not the aim of this offering.

Following presentation of background material in section II, section III, demonstrating optimality of likelihood ratio quantization under Neyman-Pearson, Bayes, and Ali-Silvey distance criteria, are more detailed versions of the material in [35]. Section IV provides discussion and a number of examples and comparisons. One of these will deal with a case in which sensors observe (conditionally) iid data, yet optimally should not use an identical quantization; in section V we discuss this pathology in some depth.

II. Statement of the Problem

We consider the binary detection problem with N sensors. The purpose of the detector is to discriminate optimally between two states of nature H_0 and H_1. For example, H_0 may represent a noise only hypothesis whereas H_1 represents signal plus additive noise. For the Bayes and mutual information criteria, we

consider the state of nature to be a random variable; for the Neyman-Pearson and Ali-Silvey criteria, it is assumed to be deterministic. The i^{th} sensor input is a realization of the random vector $X_i \in (\Omega_i, B_i)$ (of arbitrary dimension). The statistical behavior of this random vector is defined by

$$\begin{aligned} H_0 &: X_i \sim P_{0i} \\ H_1 &: X_i \sim P_{1i} \end{aligned}$$

where H_0 and H_1 are the null and alternative hypotheses, respectively, and the P_{ki}, $i = 1, 2, \ldots, N$, $k = 0, 1$ are probability distributions. We assume throughout that P_{0i} and P_{1i} are absolutely continuous with respect to each other; hence, the same σ-algebra is used under both hypotheses.

If the channels between the sensors and the fusion center are not bandlimited, the decentralized detection problem becomes, in effect, centralized. The sensors will then transmit the received data directly. Let $x = \{x_1, x_2, \ldots, x_N\}$ be a realization of the complete (at all sensors) random data $X = \{X_1, X_2, \ldots, X_N\}$ with distribution P_k under hypothesis H_k. The optimal centralized test for the Neyman-Pearson and Bayes criteria is well known [31]. As shown in [18], the same test is also optimal for the mutual information criterion [2]. This test is given by

$$\phi(x) = \begin{cases} 1 & \text{if } L(x) > \tau \\ \gamma & \text{if } L(x) = \tau \\ 0 & \text{if } L(x) < \tau \end{cases} \tag{1}$$

where $\phi(x)$ is the probability of deciding for H_1 when x is observed, L is the likelihood ratio dP_1/dP_0, and γ and τ are chosen to conform to the specific performance goals of the system designer. For Bayes detection, γ is irrelevant and for mutual information it is optimally 0 or 1. If we assume no bandwidth constraint and that the data at different sensors is conditionally independent, then the sensors can transmit the local likelihood ratio (since it is then a sufficient statistic for the detection problem [24, pages 145-147]), rather than the data itself. We assume henceforth that the sensor data is, in fact, conditionally independent.

In general, for decentralized detection, the channels between the sensors and the fusion center will be bandlimited. Suppose that, for a given decision, the maximum number of bits that can be transmitted by sensor i is $b_i = \log_2 M_i$. Suppose also that the number of possible values of L_i (the local likelihood ratio at the i^{th} sensor) is greater than M_i. Under these circumstances, we must transmit a non-sufficient statistic U_i from the i^{th} sensor to the fusion center. The major subject of this paper is the optimal form of such a statistic. Since the channel bit rate is limited to b_i, it is apparent that U_i must belong to a set containing at most M_i elements. The elements themselves are irrelevant, for convenience we assume that they form the set $\{0, 1, \ldots, M_i - 1\}$ and we define the probabilities $\alpha_{ij} = \Pr\{U_i = j | H_0\}$ and $\beta_{ij} = \Pr\{U_i = j | H_1\}$. Hence, the situation we are studying is the following: the i^{th} sensor receives the random data X_i; according to some yet to be determined decision rule, it produces the statistic U_i, which takes on one of the values in the set $\{0, 1, \ldots, M_i - 1\}$ in accordance with the probabilities $\{\alpha_{ij}\}$ and $\{\beta_{ij}\}$; the set of statistics $\{U_i\}$, $i = 1, 2, \ldots, N$, is transmitted to the fusion center where the final binary decision U_0 is made. The objective is to have a procedure (at both the sensors and the fusion center) that selects the value for U_0 in an optimal way.

We assume throughout this paper that X_i has a density. This can be done without loss of generality since we are concerned with the relative probabilities under the two hypotheses of sets of values for X_i but never with the actual values of X_i itself. Hence, for any X_i that does not have a density and any partition U_i of the set Ω_i, we can substitute the random variable X_i' and the partition U_i' so that X_i' has a density and U_i' has the same distribution as U_i under both hypotheses. We do not, however, assume that the likelihood ratio of X_i has a density since this excludes many situations of interest (for example,

additive signals in Laplace noise). One of the main points to be demonstrated in this paper is that it is the distributions of the likelihood ratio with which we are most interested.

Lastly, as a notational convenience we let I_P^+ be the set of positive integers less than or equal to P and we let I_Q be the set of non-negative integers less than or equal to Q.

III. Neyman-Pearson and Bayes Criteria

A binary detection system makes a decision that can take on one of two values. The selection of this value amounts to deciding, based on the available data, that one of the two possible states of nature H_0 and H_1 is true. Since the data used in the decision process is random, there is generally a non-zero probability of error associated with the system decision. Throughout this paper, we will refer to the probability of deciding H_1 when H_0 is actually true as the false alarm probability, level, or size of the test. We will refer to the probability of correctly deciding H_1 as the probability of detection or the power of the test. These two probabilities exhaustively define the statistical properties of the test.

A. Decentralized Neyman-Pearson Detection

The objective for systems designed under the Neyman-Pearson criterion is to achieve maximum power for fixed test level. For an arbitrary fusion rule (with given sensor decision rules for the N sensors) and maximum overall level α_0, the ultimate decision U_0 is determined by a partition of the observation space D of the vector $\{U_1, U_2, \ldots, U_N\}$. The space D is the Cartesian product of the sets I_{M_i-1} for $i \in I_N^+$. In the following, we denote a decision for H_k by $U_0 = k$. Let $\mathbf{u} = \{u_1, u_2, \ldots, u_N\}$ be a realization of $\mathbf{U}$. If the fusion rule is chosen according to the Neyman-Pearson criterion (that is, the rule which maximizes the resulting power of the test), then we use a likelihood ratio test given by

$$\phi(\mathbf{u}) \;=\; \begin{cases} 1 & \text{if } L(\mathbf{u}) > \tau \\ \gamma & \text{if } L(\mathbf{u}) = \tau \\ 0 & \text{if } L(\mathbf{u}) < \tau \end{cases} \tag{2}$$

where $\phi(\mathbf{u}) = \Pr\{U_0 = 1 | \mathbf{U} = \mathbf{u}\}$,

$$L(\mathbf{u}) \;=\; \prod_{n=1}^{N} \beta_{nu_n} \left[\prod_{n=1}^{N} \alpha_{nu_n} \right]^{-1} \tag{3}$$

and $\gamma \in [0, 1]$. If we define $D_1 = \{\mathbf{u} : L(\mathbf{u}) > \tau\}$, $D_\gamma = \{\mathbf{u} : L(\mathbf{u}) = \tau\}$, and $D_0 = D \cap \overline{D_1 \cup D_\gamma}$, then for the fixed test size α_0, γ and τ are chosen so that

$$\alpha_0 \;=\; \sum_{D_1} \prod_{n=1}^{N} \alpha_{nu_n} \;+\; \gamma \sum_{D_\gamma} \prod_{n=1}^{N} \alpha_{nu_n}. \tag{4}$$

The power of this test is

$$\beta \;=\; \sum_{D_1} \prod_{n=1}^{N} \beta_{nu_n} \;+\; \gamma \sum_{D_\gamma} \prod_{n=1}^{N} \beta_{nu_n}. \tag{5}$$

It is well known [20] that the optimal Bayes test, with $\gamma = 0$ and an appropriately chosen value for τ, can be written in the form of (2). Hence, for a given set of priors, the optimal Bayes test can be found by choosing an appropriate Neyman-Pearson test. Thus, it is clear that the optimal Bayes decentralized test will be determined by the set of ordered pairs (α_0, β^*) where $\alpha_0 \in (0, 1)$ is the test size and β^* is the maximum power possible at that size. Which particular pair is used will depend on the prior probabilities of H_0 and H_1.

Let $\phi_{i,j}(x_i) = \Pr(U_i = j|X_i = x_i)$, for each $j \in I_{M_i-1}$, and each $i \in I_N^+$. If, for each x_i, $\sum_{j=0}^{M_i-1} \phi_{i,j}(x_i) = 1$, then the set $\{\phi_{i,j}\}$, defines a mapping of Ω_i into I_{M_i-1} and can be used in a decentralized detector as a sensor decision rule. We call such a mapping an M_i level mapping. With β_{ij} and α_{ij} as previously defined we have the following lemma.

Lemma 1: If $\beta_{ij}/\alpha_{ij} = \beta_{ik}/\alpha_{ik}$ for $j \neq k$, then the performance of a decentralized detection system using an M_i level mapping as the i^{th} sensor decision rule is equivalent to that of a system using an $M_i - 1$ level mapping.

Proof. See [35], [36]. ■

Generally, there are an uncountable number of possible decision rules for the sensors. A decision rule of particular interest is given by the following definition.

Definition 1: A monotone partition of the set Ω by the likelihood ratio L is a collection of disjoint, non-trivial sets $\{R_0, \ldots, R_N\}$ with $\cup_{i=0}^{N} R_i = \Omega$ such that if $x \in R_j$, $x' \in R_j$, $y \in R_k$, $y' \in R_k$, and $L(x) > L(y)$, then $L(x') \geq L(y')$ a.e.(P_0).

We henceforth refer to such a partition as a *likelihood ratio partition*. Let $L_i = dP_{1i}/dP_{0i}$ be the likelihood ratio of X_i. We show later that the optimal decision rule for the i^{th} sensor is a likelihood ratio partition of Ω_i by L_i.

Lemma 2: For fixed test size, there exists an M_i level mapping such that the power of a decentralized detection system using this M_i level mapping as the i^{th} sensor decision rule is no smaller than that of a system using any $M_i - 1$ level mapping.

Proof. See [35], [36]. ■

Lemma 3: Assume that for any finite set $F \subset \Re^+$ containing M_i or fewer elements, $P_{0i}\{L_i(X_i) \in F\} < 1$. Then a decentralized detector using a likelihood ratio partition of Ω_i is at best equivalent to one that uses some sensor decision rule of the form

$$Pr(U_i = j|X_i = x_i) = \phi_{i,j}(x_i) = \begin{cases} 1 & \text{if } t_{i,j} < L_i(x_i) < t_{i,j+1} \\ \gamma_{i,j} & \text{if } L_i(x_i) = t_{i,j} \\ 1 - \gamma_{i,j+1} & \text{if } L_i(x_i) = t_{i,j+1} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $j \in I_{M_i-1}$, $t_{i,0} = 0$, $t_{i,M_i} = \infty$, $t_{i,j} \leq t_{i,j+1}$, and $\gamma_{i,k} \in [0, 1]$, $k \in I_{M_i}$. Furthermore, there always exists a likelihood partition for which detector performance will be equivalent to that achieved using the decision rule of (6).

Proof. See [35], [36]. ■

We note in passing that if X_i does *not* have a density, the definition of a likelihood ratio partition may be modified by omitting the word disjoint. In that event, Lemma 3 will hold. We also note that if F is any finite set of real numbers and if $L_i(X_i) \in F$ with probability one, then the performance of a decentralized detector using likelihood ratio partitions cannot be improved by increasing the number of levels at the i^{th} sensor beyond the number of elements in F.

Theorem 1: The Neyman-Pearson optimal sensor decision rule is a likelihood ratio partition.

Proof. We assume that $\beta_{ij}/\alpha_{ij} \geq \beta_{ik}/\alpha_{ik}$. Then $L(\mathbf{u}_i, j) \geq L(\mathbf{u}_i, k)$. Hence $\{\mathbf{u}_i, u_i = k\} \in D_1$ implies that $(\mathbf{u}_i, u_i = j) \in D_1$, and $(\mathbf{u}_i, u_i = k) \in D_\gamma$ implies that $(\mathbf{u}_i, u_i = j) \in D_\gamma \cup D_1$; that is, if there is a non-zero probability that we would decide for H_1 when $u_i = k$, then the probability of so deciding is at least as large if we change u_i from k to j. Assume that the levels are ordered as in (**??**). Then for each $N - 1$ dimensional vector $\mathbf{u}_i$ there is some integer $b(\mathbf{u}_i)$ such that for $j \geq b(\mathbf{u}_i)$, $(\mathbf{u}_i, u_i = j) \in D_1$. Furthermore, there is some integer $a(\mathbf{u}_i) < b(\mathbf{u}_i)$ such that for $a(\mathbf{u}_i) < j < b(\mathbf{u}_i)$, $(\mathbf{u}_i, u_i = j) \in D_\gamma$.

Suppose we are given an arbitrary decentralized detection system which satisfies (2) and (**??**). Let $\phi_{i,j}(x_i) = \Pr(U_i = j|X_i = x_i)$ define the sensor partition rules for this arbitrary system. Here $i \in I_N^+$ and $x_i \in \Omega_i$. For this partition, let $Q_{ki}(j) = \int_{\Omega_i} \phi_{i,j}(x_i) dP_{ki}(x_i)$ be the probability that $U_i = j$ under the

hypothesis H_k, $k = 0, 1$. Also construct a likelihood ratio partition for the i^{th} sensor with the probability functions Q^*_{ki}, $k = 0, 1$. We choose the latter partition so that

$$Q_{0i}(j) \;=\; {Q_{0i}}^*(j), \quad j = 0, 1, \ldots, M_i - 1\,. \tag{7}$$

and assume that it can be described by Eq. (6) with $t_{i,0} = 0$, $t_{i,M_i} = \infty$, $t_{i,j} < t_{i,j+1}$, and the $\gamma_{i,j} \in [0, 1]$, $k \in I_{M_i}$. The $\{t_{i,j}\}$, $j \in I^+_{M_i-1}$ and $\{\gamma_{i,j}\}$ are chosen to satisfy the equality constraints of Eq. (7). Using the Neyman-Pearson fusion rule used for the arbitrary partition, ${Q_{0i}}^*$ and ${Q_{1i}}^*$ will also satisfy (**??**) and the two tests will have the same level. For the likelihood ratio partition we have, for any $n \in I_{M_i-1}$,

$${Q_{1i}}^*\{U_i \;\geq\; n\} \;=\; P_{1i}\{L(X_i) \;>\; t_{i,n}\} \;+\; \gamma_{i,n} P_{1i}\{L(X_i) \;=\; t_{i,n}\}.$$

For Neyman-Pearson fusion and by (**??**), the power of the test using the arbitrary partition is

$$\beta = \sum_{\mathbf{u}_i} \Pr\{\mathbf{u}_i|H_1\}\,[Q_{1i}\{u_i \geq b(\mathbf{u}_i)\} + \gamma Q_{1i}\{a(\mathbf{u}_i) < u_i < b(\mathbf{u}_i)\}]$$

where $\gamma \in [0, 1]$ is a randomization chosen so that the test level may be exactly specified. Suppose that $\mathbf{v}_i$ is some arbitrary value of $\mathbf{u}_i$ and that $b(\mathbf{v}_i) = b$ and $a(\mathbf{v}_i) = a$. Then for $\mathbf{u}_i = \mathbf{v}_i$ the bracketed term in (8) can be written as

$$Q_{1i}\{u_i \geq b\} + \gamma Q_{1i}\{a < u_i < b\} = \sigma + \gamma\rho$$

where

$$\sigma = \int_{\Omega_i} \sum_{j \geq b} \phi_{i,j}(x_i)\, dP_{1i}(x_i)$$

and

$$\rho = \int_{\Omega_i} \sum_{a < j < b} \phi_{i,j}(x_i)\, dP_{1i}(x_i)$$

For the likelihood ratio partition with Neyman-Pearson fusion, the power is

$$\beta^* = \sum_{\mathbf{u}_i} \Pr\{\mathbf{u}_i|H_1\}\,[{Q_{1i}}^*\{u_i \geq b(\mathbf{u}_i)\} \;+\; \gamma {Q_{1i}}^*\{a(\mathbf{u}_i) < u_i < b(\mathbf{u}_i)\}] \tag{8}$$

For $\mathbf{u}_i = \mathbf{v}_i$ the bracketed term in (8) can be written as.

$${Q_{1i}}^*\{u_i \geq b\} \;+\; \gamma {Q_{1i}}^*\{a < u_i < b\} = \sigma^* + \gamma\rho^* \tag{9}$$

where

$$\sigma^* = P_{1i}\{L_i(x_i) > t_{i,b}\} + \gamma_{i,b} P_{1i}\{L_i(x_i) = t_{i,b}\}$$

and

$$\rho^* = P_{1i}\{t_{i,a+1} \leq L_i(x_i) < t_{i,b}\} + (1 - \gamma_{i,b}) P_{1i}\{L_i(x_i) = t_{i,b}\} - (1 - \gamma_{i,a+1}) P_{1i}\{L_i(x_i) = t_{i,a+1}\} \tag{10}$$

Since $\mathbf{v}_i$ is arbitrary, to complete the proof it suffices to show that the quantity in (9) is at least as large as $\sigma + \gamma\rho$ or

$$\sigma^* \;-\; \sigma \;-\; \gamma(\rho \;-\; \rho^*) \;\geq\; 0\,. \tag{11}$$

By Eq. (7), the H_0 probability corresponding to σ is:

$$P_{0i}\{L_i(x_i) > t_{i,b}\} + \gamma_{i,b} P_{0i}\{L_i(x_i) = t_{i,b}\}$$

and the H_0 probability corresponding to ρ is:

$$P_{0i}\{t_{i,a+1} \leq L_i(x_i) < t_{i,b}\} + (1 - \gamma_{i,b}) P_{0i}\{L_i(x_i) = t_{i,b}\} - (1 - \gamma_{i,a+1}) P_{0i}\{L_i(x_i) = t_{i,a+1}\}\,. \tag{12}$$

Hence, by the Neyman-Pearson lemma for randomized tests [15], we know that $\sigma^* \geq \sigma$. Then a necessary condition for (11) not to hold is $\rho^* < \rho$. In that case, the left side of (11) is decreasing in γ and

$$\sigma^* - \sigma - \gamma(\rho - \rho^*) \geq \sigma^* + \rho^* - (\sigma + \rho). \tag{13}$$

But

$$\sigma^* + \rho^* = P_{1i}\{L_i(x_i) > t_{i,a+1}\} + \gamma_{i,a+1}P_{1i}\{L_i(x_i) = t_{i,a+1}\}$$

and, by Eq. (7),

$$\int_{\Omega_i} \sum_{j>a} \phi_{i,j} dP_{0i}(x_i) P_{0i}\{L_i(x_i) > t_{i,a+1}\} + \gamma_{i,a+1}P_{0i}\{L_i(x_i) = t_{i,a+1}\} \tag{14}$$

so by the Neyman-Pearson lemma, the right side of (13) is non-negative and (11) holds. ■

B. Decentralized Bayes Detection

We now turn to the optimal decentralized Bayes test. Let π_0 and π_1 be the prior probabilities of H_0 and H_1, respectively. Also let C_{mn} be the cost assigned to choosing hypothesis H_m when H_n is the true hypothesis. The average or Bayes cost of the detection system is then

$$\mathcal{C} = \sum_{n=0}^{1} \pi_n C_{0n} + \sum_{n=0}^{1} \pi_n (C_{1n} - C_{0n}) P_n(D_1)$$

where D_1 is the fusion center decision region for H_1. The objective of Bayesian detection is to minimize this cost. If, given a true hypothesis, the cost of making an error is greater than the cost of making the correct decision (as it logically should be) then the optimal fusion rule for the Bayes test is given by the following:

$$\phi(\mathbf{u}) = \begin{cases} 1 & \text{if } L(\mathbf{u}) > \tau \\ 0 & \text{if } L(\mathbf{u}) < \tau \end{cases} \tag{15}$$

where $\phi(\mathbf{u})$ is the probability of choosing H_1 given that $\mathbf{U} = \mathbf{u}$. The optimal threshold is

$$\tau = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \tag{16}$$

Note the similarity between Eqs. (2) and (15). Having specified the fusion rule, we now turn to the sensor quantizers.

Theorem 2: The sensor decision rules for an optimal Bayes detector are likelihood ratio partitions of the sensor observation spaces.

Proof. The Bayes test uses, for each set of priors (π_0, π_1), a particular level and the maximum power that can be achieved for a test at that level. Eq. (2) gives the maximum power test for the level achieved by the threshold-randomization pair γ and τ. For Bayes tests, we specify that $\gamma = 0$ and τ is as given by Eq. (16). Hence, by application of Theorem 1, the theorem is proved. ■

We have therefore shown that the optimal decentralized detector under the Neyman-Pearson and Bayes criteria, uses a quantized version of the local likelihood ratio as the sensor decision rule. Hence, in the design of such a system, it is necessary only to partition the likelihood ratio rather than the raw data. This is intuitively satisfying since the likelihood ratio is a sufficient statistic for the detection problem.

IV. Discussion and Examples

It is intuitively appealing that an optimal sensor mapping strategy should be a quantization of the minimal sufficient statistic for detection, the likelihood ratio. Although the result appears to have been widely assumed, up to now it has been proven only for a number of special cases. We mention several here, the first two of which have been cited previously:

- For the Bayes criterion with single-bit channels ($M_i = 2$), Reibman and Nolte have shown that sensors perform local likelihood ratio tests [23].
- Poor and Thomas [21] have derived Ali-Silvey distance optimal thresholds for the case in which the local likelihood ratios are monotonically-increasing functions of the data.
- For the "weak-signal" case, Kassam [12] has demonstrated that the sensor mapping function that maximizes the efficacy (or asymptotic signal-to-noise ratio) is a quantization of the differential log-likelihood ratio.
- For discrimination $E_{H_0}\left[\log\left(\frac{P_1(\mathbf{u})}{P_0(\mathbf{u})}\right)\right]$ (an Ali-Silvey distance measure closely related to Shannon Information) and discrete-valued sensor data, it can be inferred from the *refinement* lemma [2, Theorem 4.2.3] that the optimal sensor mapping is a quantization of the local likelihood ratio.
- For the Chernoff function $-\log\left\{E_{H_0}\left[\left(\frac{P_1(\mathbf{u})}{P_0(\mathbf{u})}\right)^s\right]\right\}, 0 < s < 1$, it has been shown in [13] that the optimal mapping is a quantization of the local likelihood ratio. Note that Bhattacharyya distance is a special case ($s = \frac{1}{2}$) of the Chernoff function.

Which measure is most appropriate is largely a function of the application. If prior probabilities for the hypotheses are known, then Shannon Information may be most appealing; if, in addition, error costs can be assumed, then a Bayes criterion should be used. Without prior probabilities the Neyman-Pearson criterion is most natural; however, the resulting scheme is in general optimal only at a particular false-alarm rate and its derivation can require extensive computation. The Ali-Silvey distance criteria, which utilize the sensor outputs $\mathbf{U}$ rather than the final decision U_0, generally work well for a broad range of false-alarm rate; unfortunately, there is no guarantee of any meaningful optimality. We present the following examples for clarification.

Example 1: Suppose we are to design a constant false-alarm rate (CFAR) system for Swerling II targets. The system is constrained to have three sensors ($N = 3$); each sensor can send one of three possible messages ($M_i = 3$ $i = 1, 2, 3$) to the fusion center. A cell-averaging (CA) scheme on 20 homogeneous reference cells $\{\{Y_{ij}\}_{j=1}^{20}\}_{i=1}^{3}$ is used to estimate the ambient noise level. For each i and j, Y_{ij} is exponentially distributed with parameter λ; that is, the density function of the reference value is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

The output of the test cell Z_i is exponentially distributed with parameter $\frac{\lambda}{2}$ under H_1, and with parameter λ under H_0. All random variables are assumed to be independent. The sensors quantize the normalized data $\{X_i\}_{i=1}^{3}$, where

$$X_i = \frac{Z_i}{\frac{1}{20}\sum_{i=1}^{20} Y_{ij}}$$

It can be shown that the X_i are independent and identically distributed (i.i.d.) with density

$$f_X(x) = \frac{\theta}{(1+\theta x/20)^{21}}$$

where $\theta = 1$ under H_0 and $\theta = 0.5$ under H_1. Since the local likelihood ratios

$$L_i(X_i) = \frac{1}{2}\left(\frac{1+X_i/20}{1+X_i/40}\right)^{21}$$

are monotone-increasing functions of X_i, quantizing the data directly is equivalent to a likelihood-ratio partition.

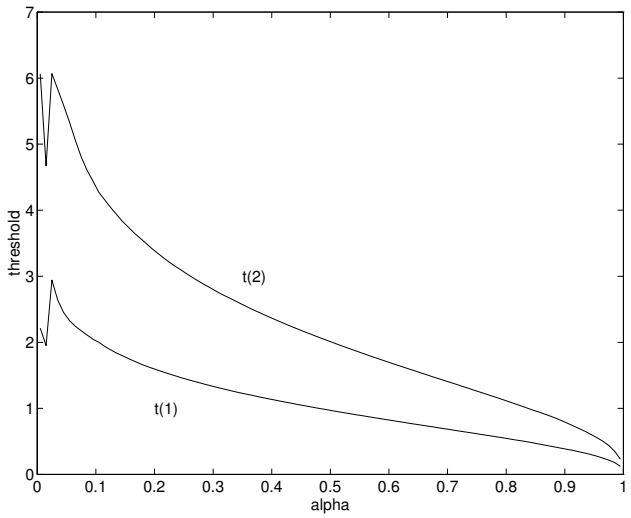

Fig. 1. Plot of Neyman-Pearson optimal thresholds versus α for example 1 (CA-CFAR).

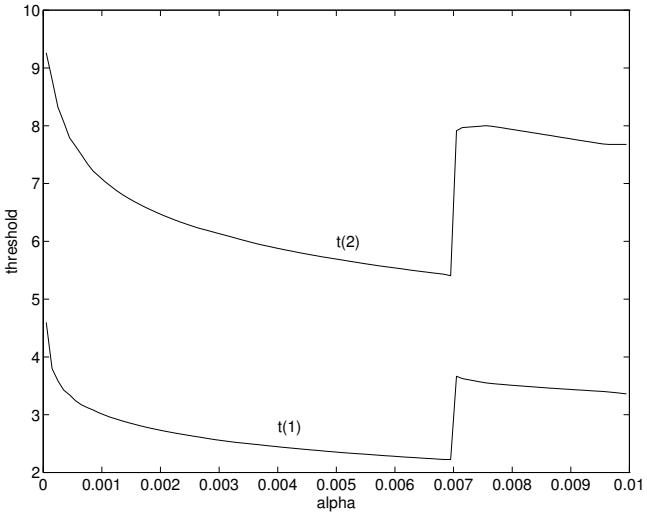

Fig. 2. Plot of Neyman-Pearson optimal thresholds versus α for example 1. This is an expanded version of the low-α behavior in figure 1.

As we discuss later, it is not always true that the optimal sensor mappings are identical likelihood ratio partitions. Obviously, this will not in general be the case if the channel capacity constraints differ between sensors. However, it still may not be true even if all of the sensors observe i.i.d. random processes and face identical communications constraints (i.e. $M_i = M$). However, in the case of Example 1, the partitions are in fact identical. The Neyman-Pearson optimal thresholds are plotted against α in figure

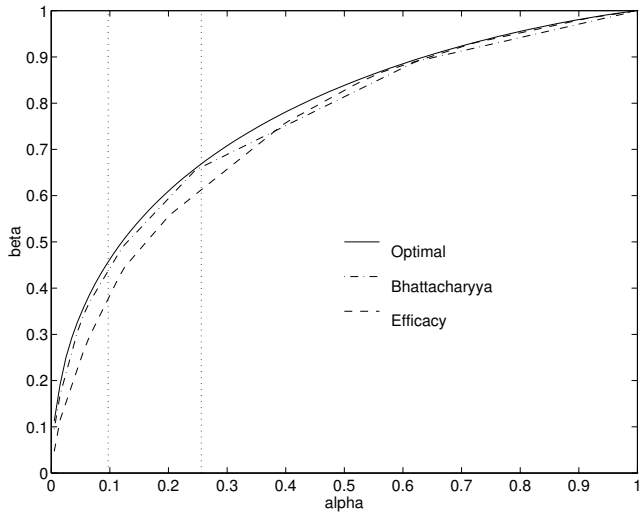

Fig. 3. Performance comparison of thresholds optimized under Neyman-Pearson, efficacy, and Bhattacharyya measures, in situation of example 1. Note that the NP-optimal thresholds vary with α; but those under the other two criteria cannot, hence the piecewise-linear ROCs. The vertical lines denote the region of this ROC obtainable under by maximizing mutual information; see figure 4.

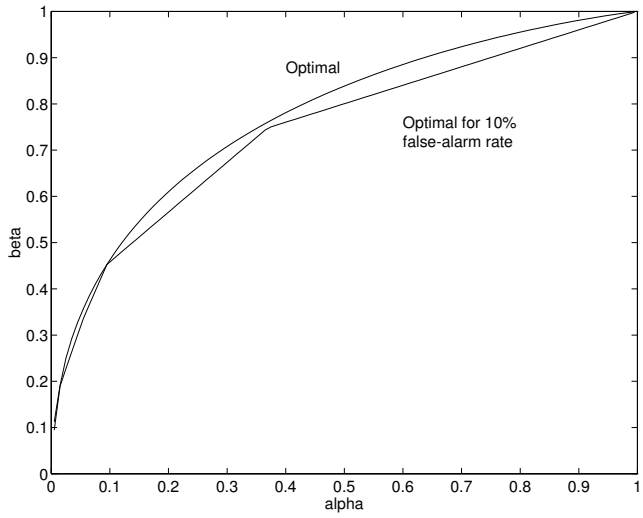

Fig. 4. Performance comparison of thresholds optimized under Neyman-Pearson criterion in situation of example 1. The "Optimal" thresholds vary with α; but that optimized for a fixed $\alpha = 10\%$ naturally do not.

1. Note the discontinuity at $\alpha \approx 0.7\%$. This can be seen more clearly in figure 2, where we have expanded the α axis for $\alpha < 0.01$. The discontinuity is due to a transition in the optimal fusion rule: for $\alpha > 0.007$, $D_1 = \{(2,3,3)\ (3,2,3)\ (3,3,2)\ (3,3,3)\}$; for $\alpha < 0.007$ D_1 is any permutation of $\{(1,3,3)\ (2,2,3)\ (2,3,3)\ (3,3,3)\}$. Here (u_1, u_2, u_3) represent the outputs of the three sensors, and each output is assumed to be ordered as in (**??**). As is true for many quantized detection systems, there

Criterion	t_1	t_2
Bhattacharyya	1.33	3.65
J-Divergence	1.46	4.05
Efficacy	0.58	1.60

TABLE I
OPTIMAL THRESHOLDS (EXAMPLE 1)

is no fusion rule that is optimal for all α.

Optimal thresholds under a number of other criteria are shown in Table 1. In all cases the thresholds are the same at each of the sensors. Receiver operating characteristics (ROCs) for detectors optimized for the Neyman-Pearson, efficacy and Bhattacharyya distance criteria are shown in figure 3. Because optimization for the latter two criteria results in only one set of thresholds for each sensor (i.e. the thresholds do not change as a function of false alarm rate), the resulting fusion rule will use randomization and the detector ROC curves are piecewise linear. In figure 4, we show the ROC curve for a quantizer optimized for a fixed false alarm rate (10%). Note that here again fixed thresholds result in randomization and a piecewise linear ROC.

Figure **??** shows the values of α and β, as a function of π_0 $(= \Pr(H_0))$, achieved with the quantizer that maximizes the information measure $I(U_0; H)$. By comparing this to figures 3 and **??**, we see that any information-optimal system (for $0 < \pi_0 < 1$) is identical to a Neyman-Pearson optimal scheme with $0.1 < \alpha < 0.26$ (approximately). This is consistent with Theorem 3; however, we also note that maximum mutual information is achieved for some test level in the interval $(0.1, 0.26)$ for *all* values of π_0. Thus, the mapping between π_0 and the information-maximizing false alarm rate is invertible but not necessarily one-to-one.

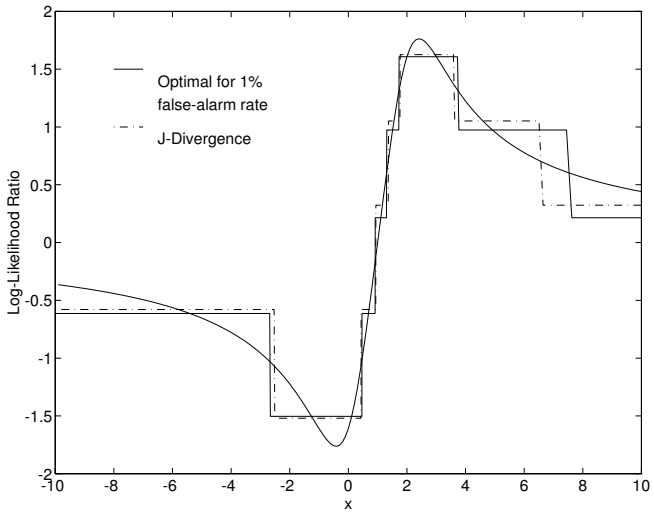

Fig. 5. Quantizer optimized under Neyman-Pearson ($\alpha = .01$) and J-divergence, for additive signal in Cauchy noise situation of example 2.

Example 2: Suppose we have ten sensors ($N = 10$) that take one observation each. The observations

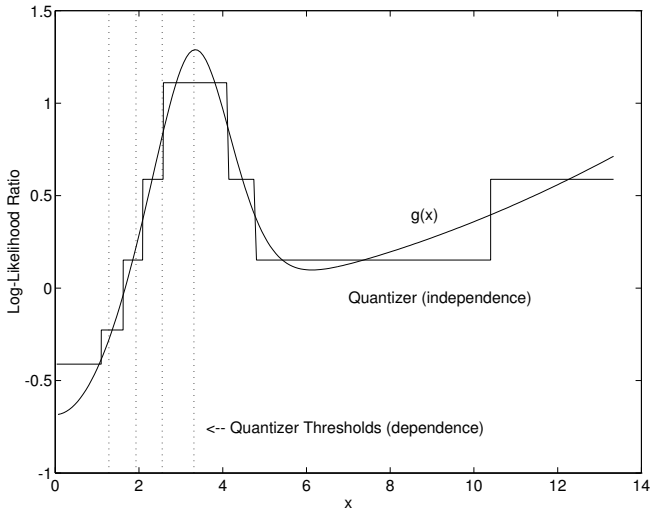

Fig. 6. Quantizers optimized under Bhattacharyya criterion, for dependent situation of example 3. Under an (incorrect) assumption of independence the result is a likelihood ratio quantization; correctly assuming dependence results in a direct data quantization.

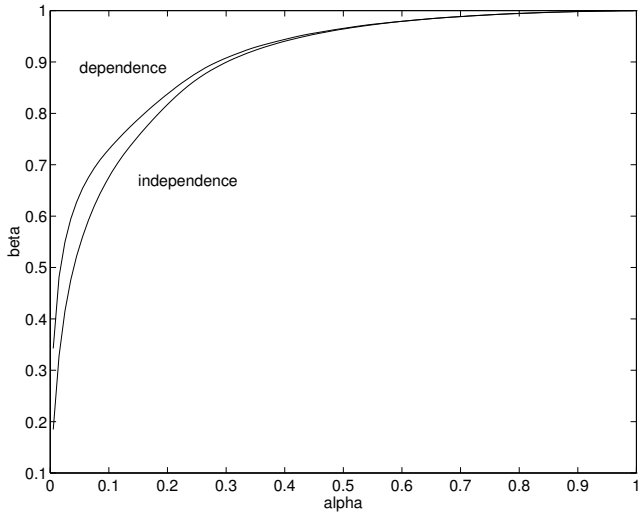

Fig. 7. Performance of quantizers for dependent situation of example 3. The thresholding scheme and an explanation of terms is in figure 6.

are i.i.d. realizations of a unit Cauchy random variable with density

$$f(x) \;=\; \frac{1}{\pi\left(1+(x-\theta)^2\right)}$$

Under the noise-only hypothesis H_0, $\theta = 0$; under H_1, $\theta = 2$. Each of the sensors uses a five level quantizer.

In figure 5, we show the log-likelihood ratio, the quantizer that is optimal for J-Divergence, and the Neyman-Pearson optimal quantizer for a false-alarm rate of 1%. The resulting probabilities of detection for ($\alpha = 0.01$) are $\beta = 0.907$ and $\beta = 0.925$ respectively. Note that in both cases the sensor mappings are likelihood ratio quantizations.

Example 3: Suppose we have ten sensors ($N = 10$) that are capable of transmitting one of five possible messages ($M_i = 5$) each. With Y_i representing the observation at the i^{th} sensor, the detection problem can be written as

$$\begin{aligned} H_0: \quad Y_i &= (1-Z)N_{1i} + ZN_{2i} \\ H_1: \quad Y_i &= S_i + (1-Z)N_{1i} + ZN_{2i} \end{aligned}$$

Here $\{S_i\}_{i=1}^{10}$, $\{N_{1i}\}_{i=1}^{10}$, and $\{N_{2i}\}_{i=1}^{10}$ are all independent, zero-mean and Gaussian, with $E(S_i^2) = E(N_{1i}^2) = 1$ and $E(N_{2i}^2) = 10$. The binary random variable Z takes on values 0 and 1 with respective probabilities 90% and 10% and has the same value at all of the sensors.

This example is used to demonstrate the effect of dependence among the sensors. Without the high power noise process, the model conforms to the unknown (Gaussian) signal in Gaussian noise problem. When the N_{2i} process is included we have added the possibility of jamming: there is a 10% probability that *all* of the sensors will be jammed, and a 90% probability of *no* jamming.

For each sensor the statistic $X_i = Y_i^2$ is sufficient. Since Y_i is always Gaussian, X_i, conditioned on the hypothesis and the presence or absence of jamming, will have a Rayleigh density given by:

$$f_R(x;\sigma^2) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

Here, σ^2 will depend on both the hypothesis and the value of Z. Hence, the univariate density of X_i is

$$f(x_i) = (1-\epsilon) f_R(x_i; 1+\theta) + \epsilon f_R(x_i; 11+\theta) \tag{17}$$

and the joint density of the observations is

$$f(x_1, x_2, \ldots, x_{10}) = (1-\epsilon) \prod_{i=1}^{10} f_R(x_i; 1+\theta) + \epsilon \prod_{i=1}^{10} f_R(x_i; 11+\theta) \tag{18}$$

Here $\epsilon = 0.1$, while $\theta = 0$ under H_0 and $\theta = 1$ under H_1.

We consider two cases of quantizers optimal under the Bhattacharyya criterion. First, we maximize the Bhattacharyya distance of U_i, the (quantized) output of each sensor; this is equivalent to the assumption that the sensor observations are independent with the density given by Eq. (17). Second, we maximize the Bhattacharyya distance for $\mathbf{U}$; that is, for the ensemble input to the fusion center. In figure 6, we plot the logarithm of the local likelihood ratio, the quantizer that results from the first maximization, and we indicate the thresholds that are used to quantize the *data* for the latter maximization. As expected, the independence assumption results in a quantization of the local likelihood ratio as specified by (17). The fact that inclusion of the dependence results in a quantization of the data X_i is not surprising; a large value of X_i is indicative of jamming at the i^{th} sensor which in turn implies jamming at *all* sensors. Note that due to dependence, the quantizer *levels* in the second case are meaningless and are not shown. The ROC curves under both assumptions are shown in figure 7. Since the dependence assumption more accurately models the problem, the resulting ROC dominates that found when independence is assumed.

Example 4: Suppose that three sensors ($N = 3$) observe mutually independent data and that the likelihood ratio of the data has density

$$f_{H_0}(x) = \begin{cases} \frac{c}{1+\left[k(x-\frac{1}{2})\right]^2} + \frac{c}{1+\left[k(x-\frac{3}{2})\right]^2} & 0 \leq x \leq 2 \\ 0 & \text{else} \end{cases}$$

under H_0 and $xf_{H_0}(x)$ under H_1). The constant c is chosen to normalize the density and $k = 100$. Each sensor uses binary quantization ($M_i = 2$).

This example is discussed in [38], [37]. It can be shown that if identical thresholds are used at all of the sensors, the optimal test (given the identical threshold constraint) will require a randomized fusion rule. Such tests were posited to be sub-optimal in [37], although an excellent recent article [9] indicates that this is true over a more restricted class than was previously believed. In any case, when the constraint is removed, randomization is no longer necessary. The optimal thresholds for the unconstrained case are shown in figure 8 as a function of α. In the figure, the partitions labeled AND, OR, and MAJORITY refer to the optimal fusion rule logic for the test levels within the partitions. Note that the three thresholds are not in general equal. Intuition suggests that if the sensors are quantizing i.i.d. random variables, then the quantizers used should be identical; this, however, is not always true. It should also be noted that for any prior probabilities, the operating points for a system optimized with respect to either global mutual information or Bayes cost must lie on the ROC curve for Neyman-Pearson optimal detectors. Hence, for i.i.d. sensor observations and identical channel constraints, the optimal detectors for the Bayes and mutual information criteria may use different sensor quantizers. From this somewhat pathological example we draw some conclusions in the following section V.

V. Identical Quantization for Identical Sensors

In the previous example 4, independent and identically-distributed sensor data with local likelihood ratio densities differentiable to arbitrary order gives rise, optimally, to *different* sensor mapping functions [37]. When does this happen, and when can we reasonably ignore the possibility? Certainly it is known to occur when sensor likelihood ratios contain point masses of probability; yet example 4 does not contain point masses.

Here we provide some answer. It will turn out that there is a threshold in terms of point-mass behavior of the likelihood ratio densities; once these densities are sufficiently "peaky", non-identical sensor quantization functions ought to be used.

A. The Necessary Conditions

With the local decision thresholds equal it is clear that the fusion rule must be of the "k-out-of-n" variety; that is,

$$\mathbf{D} \;=\; \left\{ \begin{array}{ll} H_1 & \sum_{i=1}^{n} U_i \geq k \\ H_0 & \sum_{i=1}^{n} U_i < k \end{array} \right\} \tag{19}$$

This is a familiar form, particularly for $k = 1$ (the *or* rule), for $k = n$ (the *and* rule), and for $k = \frac{n+1}{2}$ (*majority-logic*). If an optimized system is to use identical local decision thresholds, then certainly optimizing under the constraint of a k-out-of-n fusion rule must again result in identical thresholds. Let us check this.

Following standard optimization theory we shall minimize the function

$$\mathbf{C}(t_1, \ldots, t_n) \;\equiv\; 1 - \beta + \lambda(\alpha - \alpha_d) \tag{20}$$

subject to the constraint $\alpha = \alpha_d$. We shall assume for now that $1 < k < n$. Writing β in terms of t_i we get

$$\beta \;=\; F_1(t_i) Pr\left(\sum_{j=1, j\neq i}^{n} U_j \geq k \mid H_1 \right) + \; (1 - F_1(t_i)) Pr\left(\sum_{j=1, j\neq i}^{n} U_j \geq k-1 \mid H_1 \right) \tag{21}$$

or

$$\beta \;=\; Pr\left(\sum_{j=1,j\neq i}^{n} U_j \geq k \mid H_1\right) + (1-F_1(t_i))Pr\left(\sum_{j=1,j\neq i}^{n} U_j = k-1 \mid H_1\right) \tag{22}$$

with a similar expression for α, and in which we have defined

$$F_l(t) \;\equiv\; Pr(L_i \leq t \mid H_l) \tag{23}$$

for $l = 0, 1$.

The first-order necessary conditions for a minimum are that the gradient of $\mathbf{C}$ with respect to the thresholds is zero, or that

$$f_1(t_i)Pr\left(\sum_{j=1,j\neq i}^{n} U_j = k-1 \mid H_1\right) \;=\; \lambda f_0(t_i)Pr\left(\sum_{j=1,j\neq i}^{n} U_j = k-1 \mid H_0\right) \tag{24}$$

for all $i \in \{1, n\}$, and where $f_l(t) \equiv \frac{d}{dt}F_l(t)$ are the likelihood ratio densities. It is clear that equation (24) is satisfied for any set of identical thresholds, hence that value t^* which satisfies the false-alarm constraint is at least a critical point of the cost function. Evaluating the probabilities explicitly we have

$$\lambda^* \;=\; \frac{f_1(t^*)}{f_0(t^*)}\left(\frac{1-F_1(t^*)}{1-F_0(t^*)}\right)^{k-1}\left(\frac{F_1(t^*)}{F_0(t^*)}\right)^{n-k} \tag{25}$$

corresponding to $t_1 = \ldots = t_n = t^*$.

It can be shown that a necessary condition for $t_1 = \ldots = t_n = t^*$ is that

$$(n-k)\frac{d}{dt^*}\ln\left(\frac{f_1(t^*)/f_0(t^*)}{F_1(t^*)/F_0(t^*)}\right) + (k-1)\frac{d}{dt^*}\ln\left(\frac{f_1(t^*)/f_0(t^*)}{(1-F_1(t^*))/(1-F_0(t^*))}\right) \;>\; 0 \tag{26}$$

While this has been proven only for $1 < k < n$, it can be shown via a parallel development that (26) is also a necessary condition for the *and* ($k = n$) and *or* ($k = 1$) fusion rules. It is interesting that similar ROC slope conditions have been used as predictors of pathology in [**?**], in which it is shown that under certain such conditions the probability of error does not go to zero as the number of sensors grows without bound.

B. Interpretation

Let us recapitulate the results of the previous section. We have noted that if an identical threshold is to be used at each sensor, then a k-out-of-n fusion rule must be employed. With such thresholds and fusion rule the first-order necessary conditions for optimality are indeed satisfied; in addition we have presented (second-order) conditions *necessary* for optimality, and *sufficient* for the cost function $\mathbf{C}$ to be at least a local minimum.

Consider the functions

$$g_k(x) \;=\; \sum_{l=k}^{n}\binom{n}{l}x^l(1-x)^{n-l} \tag{27}$$

for which $g_k(0) = 0$, $g_k(1) = 1$, and $\dot{g}_k(x) > 0$ for $0 \leq x \leq 1$. With $\beta_0(\alpha_0)$ representing the receiver operating characteristic (ROC) of an individual sensor, a system optimal under the assumption of identical thresholds uses

$$k^* \;=\; \arg\,\max_k\left\{g_k\left(\beta_0\left(g_k^{-1}(\alpha_d)\right)\right)\right\} \tag{28}$$

for its fusion rule. If the operating point on the individual-sensor ROC corresponding to the design false-alarm rate α_d does *not* satisfy (26), then an optimal design does *not* use identical thresholds. If on the

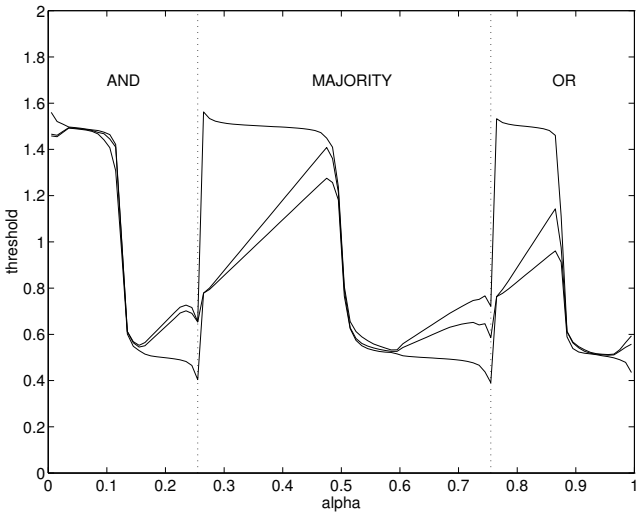

Fig. 8. Thresholds for "pathological" case of example 4. Here the sensors are independent (conditioned on hypothesis) and identical, yet the three sensors employ different thresholds. The fusion rules which happen to be optimal are also given.

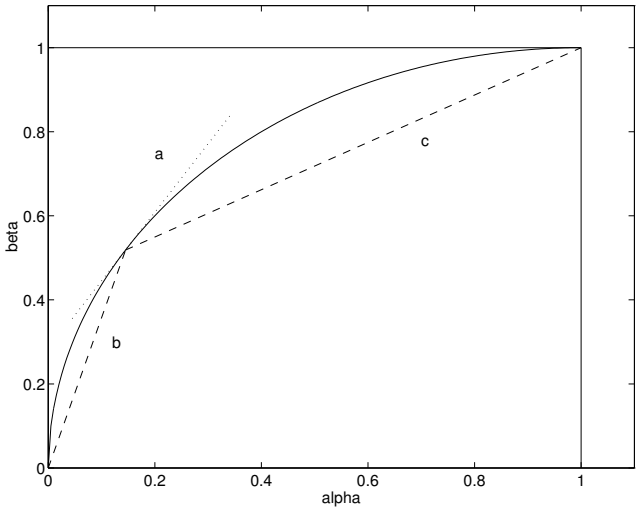

Fig. 9. Interpretation of second-order necessary conditions.

other hand (26) is satisfied, then the design produces at least a *local* minimum of the cost function $\mathbf{C}$; unfortunately there do not appear to be any general statements one can make about convexity, and there may still be a better design using non-identical thresholds, or even with a different (not k-out-of-n) fusion rule.

The *and* and *or* fusion rules are exceptional. In these cases (24) may be written as

$$\frac{f_1(t_i)}{1-F_1(t_i)}\prod_{j=1}^{n}(1-F_1(t_j)) \quad = \quad \lambda\frac{f_0(t_i)}{1-F_0(t_i)}\prod_{j=1}^{n}(1-F_0(t_j)) \tag{29}$$

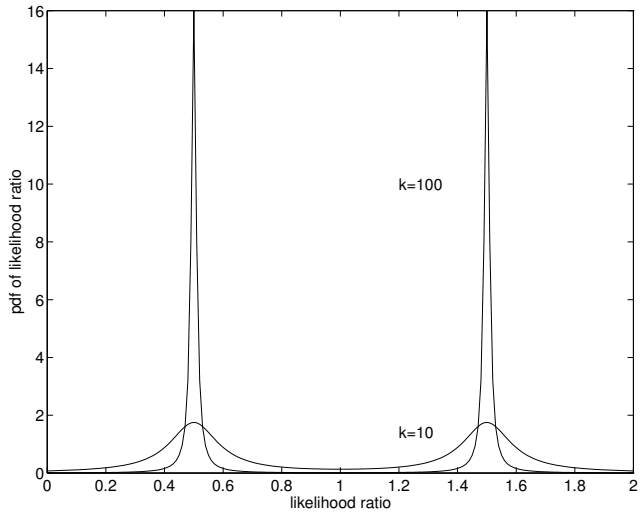

Fig. 10. Likelihood ratio density functions for the "pathological" example 4, different k parameters.

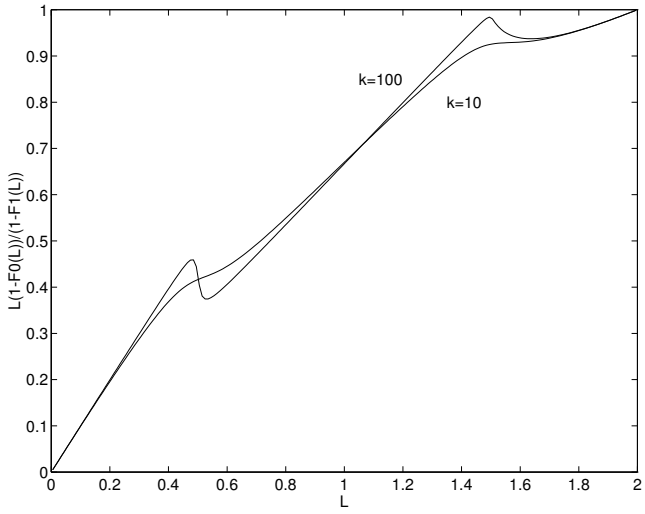

Fig. 11. The version of equation (26) appropriate for example 9 (same as example 4) and an *and* rule. Note that with $k = 10$ this curve is monotone (the derivative is always positive, and hence (26) is always satisfied); while for $k = 100$ this is not so.

and

$$\frac{f_1(t_i)}{F_1(t_i)} \prod_{j=1}^{n} (F_1(t_j)) \quad = \quad \lambda \frac{f_0(t_i)}{F_0(t_i)} \prod_{j=1}^{n} (F_0(t_j)) \tag{30}$$

respectively, for which it is clear that (26) guarantees unique solutions. In fact, the *and* and *or* rules, since they represent inequality (26) in its purest forms, offer the most straightforward interpretation. The

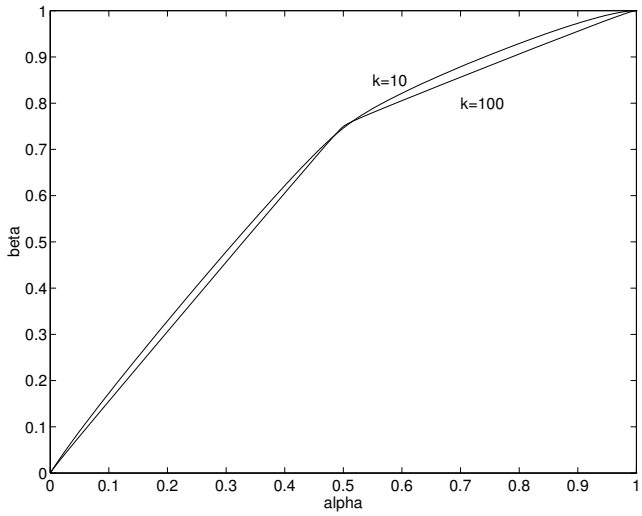

Fig. 12. The single-sensor ROC for example 9 (same as example 4), for use with equation (31) to determine whether the sensors use identical quantizations. The same conclusions can be drawn as in the alternative approach of figure 11: when $k = 100$, this ROC is sufficiently flat to violate (31).

necessary conditions for the *and* rule may be written in terms of the individual-sensor ROC as

$$\frac{d}{d\alpha} \ln\left(\frac{d\beta}{d\alpha}\right) \quad < \quad \frac{d}{d\alpha} \ln\left(\frac{\beta}{\alpha}\right) \tag{31}$$

while for the *or* rule we have

$$\frac{d}{d\alpha} \ln\left(\frac{d\beta}{d\alpha}\right) \quad < \quad \frac{d}{d\alpha} \ln\left(\frac{1-\beta}{1-\alpha}\right) \tag{32}$$

Geometrically the left sides of (31) and (32) represent the relative rate of change of the slope of the individual-sensor ROC at a given operating point. The right sides represent the relative rates of change of the slopes of secants drawn between the operating point and (0,0) and (1,1) respectively; this is shown in figure 9. Here the slope of a is $\frac{d\beta}{d\alpha}$, that of b is $\frac{\beta}{\alpha}$, and that of c is $\frac{1-\beta}{1-\alpha}$. Since all quantities are negative, we see that the second-order necessary conditions will be violated for a sufficiently *flat* (in the sense of constant slope) portion of the ROC — this is indicative of a large (but not necessarily a point) mass of probability.

Example 5: For the known-signal in Gaussian noise hypothesis problem, inequalities (31) and (32) are always satisfied.

Example 6: For the single-sensor ROC described by $\beta_0 = \alpha_0^s$ $(0 < s < 1)$, which may arise from an exponential-density hypothesis-testing problem, the left and right sides of (31) are equal. In fact, the performance of such a test using an n-sensor *and* rule is completely insensitive to the thresholds used, and actually is equal to the single-sensor performance. Not surprisingly the *and* rule is never optimal, and (26) is satisfied for any $k > 1$.

Example 7: (Same as example 4.) The null-hypothesis density f_0 is shown in Figure 10 for $k = 10$ and $k = 100$. It was observed that for $k = 10$ the thresholds optimally should be identical, while they may differ for $k = 100$, the latter case more closely resembling point masses. The actual thresholds for $k = 100$ and three sensors are shown in Figure 8. It turns out that (26) begins to fail for some t when $k > 11$. For example, if an *and* rule is to be used, only the second term in equation (26) must be checked

since $k = n$. This logarithm (that is, without the derivative) is plotted in figure 11. It is clear that with $k = 10$ the slope is always positive, and there is no doubt that sensor quantizations should be identical; yet with $k = 100$ this is not true, and the behavior of figure 8 results. The same conclusion can be drawn from directly from the ROC of figure 12 using equation (31) instead of (26).

In summary, there is, for the binary quantization case, an easy-to-check necessary condition, given in terms of the local ROC, for the sensor mapping functions to be identical. It turns out that the condition is violated when the local likelihood ratio densities become "point-mass like" in that their probability is sharply concentrated; they need lose their continuity and differentiability, however. The condition was derived in the context of Neyman-Pearson detection systems, but the proof is valid for Bayesian schemes as well. The necessary condition is known to be sufficient only for *and* and *or* fusion rules. Excellent further results on the relative performance of identical versus non-identical quantization are available in [6]; generally the difference is not great.

VI. The Good, Bad and Ugly: Decentralized Detection with Correlation

We are particularly concerned in [39] with hypotheses of correlated Gaussian data with different mean vectors

$$\begin{aligned} H &: x_1, x_2 \sim N(0, 0, 1, 1, \rho) \\ K &: x_1, x_2 \sim N(s_1, s_2, 1, 1, \rho) \end{aligned}$$

in which

$$\begin{aligned} N(s_1, s_2, \sigma_1^2, \sigma_2^2, \rho) &\equiv frac12\pi(1-\rho^2)\sigma_1\sigma_2 \\ &\times e^{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x_1-s_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-s_1}{\sigma_1}\right)\left(\frac{x_2-s_2}{\sigma_2}\right) + \left(\frac{x_2-s_2}{\sigma_2}\right)^2\right]} \end{aligned}$$

is the usual bivariate Gaussian density function. Note that there is no loss of generality in assuming zero means under H and unit standard deviations since we could translate x_1 and x_2 by any means under H and scale them both by any non-unity standard deviations and this would not change any of the results in this paper. Further, we may assume that ρ is non-negative (if $\rho < 0$, just multiply one of the x_i by -1). The advantage of these assumptions is that a convenient representation of our results occurs on the (s_1, s_2) plane.

Recall that there are mappings $x_1 \to U_1$ and $x_2 \to U_2$ according to the sets $\{A_i\}$ in (33):

$$U_i = U_i(x_i) = \begin{cases} 1 & ; \quad x_i \in A_i \\ 0 & ; \quad x_i \notin A_i \end{cases} \tag{33}$$

where each A_i is a (possibly infinite) union of intervals

$$A_i = \bigcup_{j=1}^{M_i} (l_{i,j}, u_{i,j}) \tag{34}$$

There are three fusion rules to consider:

$$AND: \quad U = \begin{cases} 1 & U_1 = U_2 = 1 \\ 0 & else \end{cases}$$

$$OR: \quad U = \begin{cases} 0 & U_1 = U_2 = 0 \\ 1 & else \end{cases}$$

$$XOR: \quad U = \begin{cases} 1 & U_1 = 0, U_2 = 1 \\ 1 & U_1 = 1, U_2 = 0 \\ 0 & else \end{cases}$$

From DeMorgan's laws it is straightforward to identify the equivalence between a test using an OR rule and prior π_H, and one using an AND rule having prior $1 - \pi_H$, after appropriate complementation of U_1, U_2, and U. Thus one problem's OR is another's AND, and although one rule may dominate for different parameter values, it is sufficient to restrict analysis to the AND case. It is sometimes mistakenly assumed that the XOR rule is "non-monotone" and hence always suboptimal. While this is true for cases of conditionally-independent sensor data, with dependence it is not. As a somewhat trivial example, consider the sensor data (already binary) given by Table II.

y_1	y_2	$Pr(y_1, y_2 \mid H)$	$Pr(y_1, y_2 \mid K)$
0	0	.7	.2
0	1	.1	.4
1	0	.1	.4
1	1	.1	0

TABLE II

AN EXAMPLE (ALREADY BINARY) OF A SITUATION IN WHICH AN XOR RULE MAY BE OPTIMAL. NOTE THAT BOTH y_1 AND y_2 ARE MONOTONE IN THEIR MARGINAL LIKELIHOOD RATIOS.

To find the individual sensor mapping rules we begin with necessary conditions on the A_i. Assuming an AND rule and that one of the A_i is known, we can calculate the best A_j ((i, j) is either $(1, 2)$ or $(2, 1)$) as [25]

$$A_j = \left\{ x_j : \frac{\int_{A_i} P_K(x_i, x_j)\, dx_i}{\int_{A_i} P_H(x_i, x_j)\, dx_i} \geq \frac{\pi_H}{\pi_K} \equiv \tau \right\} \tag{35}$$

where τ is our threshold. For the OR rule we similarly have

$$A_j = \left\{ x_j : \frac{1 - \int_{A_i} P_K(x_i, x_j)\, dx_i}{1 - \int_{A_i} P_H(x_i, x_j)\, dx_i} \geq \frac{\pi_H}{\pi_K} \right\} \tag{36}$$

for which care must sometimes be taken to avoid round-off error. For the XOR rule we have

$$A_j = \left\{ x_j : \left[1 - 2 \int_{A_i} P_K(x_i, x_j)\, dx_i \right] \geq \frac{\pi_H}{\pi_K} \left[1 - 2 \int_{A_i} P_H(x_i, x_j)\, dx_i \right] \right\} \tag{37}$$

Note that it is perilous to write (37) in terms of a likelihood ratio test, since the denominator can change sign and reverse the direction of the inequality.

We concern ourselves with the case that the densities P_K and P_H be multivariate Gaussian. For the most part we restrict attention to the two-sensor (bivariate Gaussian) situation. We ask, by way of our quest for canonical "rules":

1) Assuming an AND rule, are the A_i always simply-connected? That is, is it in all cases possible to compute U_i using a single threshold on x_i?
2) If not, are the A_i *never* simply-connected?
3) With respect to the above two questions, what happens for the OR and XOR rules?
4) Can we avoid consideration of one or more of the fusion rules (XOR, for example), since it is never optimal?
5) Is there any situation in which the data from one sensor is ignored?
6) Can we at least say that no more than *two* thresholds are needed for quantization?

It will be proven that the answer to the first two questions is negative. We shall decompose the "space" of mean-shifts into three regions, coined "good", "bad", and "ugly". In the first of these it *is* always possible to compute U_i using a single threshold on the data; and in the second it is *never* possible. Thus, we have a rule similar to that governing the conditionally-independent case; unfortunately, no rule is

forthcoming for the "ugly" region, although we are able to show that a single-threshold quantization rule offers a person-by-person optimal (PBPO) solution, meaning that many optimization routines, including the Gauss-Seidel scheme we use, will get "stuck" if initialized with single thresholds or if a single-threshold solution is found. As regards the third question, the behavior of the OR rule is similar to that of the AND rule. Again, unfortunately, there is little to be done for XOR – we show numerically that whatever "regions" may pertain, they do not easily correspond to those for the AND/OR rules. In fact, we are able to prove that single-threshold quantizers *never* form an optimal quantization for an XOR fusion rule: the XOR rule is in a sense always ugly. It is thus tempting to hope that the fourth question may be answered in the affirmative, particularly since it is easily seen to be true for the conditionally-independent case (the XOR rule would not be a likelihood ratio test from the fusion center's point of view). It is generally difficult to *prove* superiority of one fusion rule over another unless the latter is clearly a bad idea; however, we show numerically a number of situations in which the XOR rule *is* optimal. As regards the fifth question, it should be understood that in the conditionally-independent (Gaussian) situation this is *never* true, since regardless of SNR one sensor's observation can always be of sufficient certainty to contradict the other. It is thus perhaps surprising that we are able to prove the answer to the fifth question is "yes". It may be clear that the last question may be asked in a tone of frustration; unfortunately the answer, "no", is scarcely placatory.

VII. Feedback

A group of human decision-makers, each in possession of an informative observation but constrained to offer only *yes/no* opinions to one another, would most probably engage in successive rounds of communication characterized by the implicit inquiries: "Are you sure?"; or "Are you very, very sure, or just very sure?" We have termed this discussion/operation *parley*, and have examined ways to automate it for use in sensor networks.

To be specific, we have assumed in [26] that sensors observe data which is informative for hypothesis-testing purposes, and communicate to each other binary data in a broadcast fashion. We have examined two ways to formulate the procedure. Under the first, which we term "greedy," each sensor tries to be as correct as possible, conditioned on its current knowledge, at every round of decision-making. Under the second, the decision rule is such that if all sensors agree, then the decision that they agree on would match that of a (hypothetical) centralized decision-maker with access to all data. Both schemes are straightforward to implement; the first is probably closer in flavor to the human example.

Among our findings is that in both cases *a consensus always occurs*. In fact, in the first scheme the consensus occurs almost too quickly: correctness is sacrificed, and the probability-of-error performance is not particularly impressive. We do not recommend this scheme. Under the second, consensus is delayed somewhat (although never, in our examples, for too long), in such a way that the decision is the best possible given the data. If there is a reason to consider the use of feedback, then it is because correctness is more important than the use of resources for communication – as such, it seems that this would be the scheme of choice.

VIII. Censoring

The censoring scheme [22] appears to represent a valid practical alternative to likelihood-ratio quantization, and is particularly well-suited to situations in which the prior probability of a target being present in a given resolution cell is small. The result that the no-send region is a single interval facilitates the processing of the received signal at the local sensors, and moreover, the design of the overall system is greatly simplified. The performance of the censoring scheme is very near to optimal for small Π_K even with quite severe communication rate constraints. The same conclusions, with an appropriate redefinition

of the average communication rate, are valid for the Neyman-Pearson case. Optimization in these cases is still difficult, and hence in this paper we have presented results of a more practical nature. We have shown that the same result is true under the easily-optimized Ali-Silvey distance measures. We have also shown that under certain conditions the "no-send" regions correspond to the local likelihood ratio values being lower than a threshold.

We have further shown by example that censoring can involve impressive saving in communication. A feedback scheme to retrieve lost data (when necessary) was also presented, and it was demonstrated by example that optimal (unquantized) performance is possible with very little communication.

IX. Networks that Learn

In [17] was considered a parametric statistical model for social data. Elaborating on the latent-variables model for social sensing put forth in [33], [34], we performed a detailed characterization of the estimation performance bounds in terms of Fisher information, and proposed different likelihood-based inference algorithms (Expectation-Maximization and Fisher scoring) that are able to meet such bounds asymptotically. The theoretical analysis revealed the basic scaling laws of the addressed social sensing model: i) the MSE performance scales inversely with the number of tasks accomplished by the network agents, while ii) as a function of the number of agents, the MSE saturates to the performance of a *clairvoyant* system that has exact knowledge of the underlying hypotheses, i.e., of the latent variables. Application of the method to a couple of classical estimation problems revealed that the Expectation-Maximization algorithm is well suited to the latent-variables structure of the model, as noticed in [33], [34], and that for realistic number of agents' tasks the Fisher scoring method is effective in delivering the best achievable performance. In any case, the Fisher scoring approach would provide an excellent initialization ("warm start") for EM.

References

[1] S.M. Ali and S.D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Royal Stat. Soc.*, Series B, vol. 28, pp. 131-142, 1966.

[2] R. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley: Reading, MA, 1987.

[3] R. Blum, "Quantization in Multi-sensor Random Signal Detection," *IEEE Transactions on Information Theory*, January 1995.

[4] R. Blum, "Necessary Conditions for Optimum Distributed Detectors Under the Neyman-Pearson Criterion," *IEEE Transactions on Information Theory*, May 1996.

[5] Z. Chair and P.K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace Electron. Systems*, AES-22, pp. 98-101, January 1986.

[6] P.-N. Chen and A. Papamarcou, "New Asymptotic Results in Parallel Distributed Detection", *IEEE Transactions on Information Theory*, November 1993.

[7] T.L. Grettenberg, "Signal selection in communication and radar systems," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 265-275, October 1963.

[8] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Tech.*, vol. COM-15, pp. 52-60, February 1967.

[9] Y. Han and T. Kim, "Randomized Fusion Rules can be Optimal in Distributed Neyman-Pearson Detectors", *IEEE Transactions on Information Theory*, pp.1281-1288, July 1997.

[10] I.Y. Hoballah and P.K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inform. Theory*, vol. IT-35, pp.995-1000, September 1989.

[11] I.Y. Hoballah and P.K. Varshney, "An information theoretic approach to the distributed detection problem," *IEEE Trans. Inform. Theory*, vol. IT-35, pp.988-994, September 1989.

[12] S. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, 1987.

[13] D. Kazakos, V. Vannicola, and M. Wicks, "On Optimal Quantization for Detection," *Proceedings of 1989 Conference on Information Sciences and Systems*, March 1989.

[14] H. Kobayashi and J.B. Thomas, "Distance measures and related criteria," *Proc. Fifth Annual Allerton Conf. Circuit and System Theory*, pp. 491-500, Monticello, IL, October 1967.

[15] E.L. Lehmann, *Testing Statistical Hypotheses*, Wiley: New York, 1986.

[16] D. Luenberger, *Linear and Nonlinear Programming,*, Addison-Wesley, 1984. bibitemtapk Z. Tang, K. Pattipati, and D. Kleinman, "A distributed M-ary hypothesis testing problem with correlated observations," *IEEE transactions on automatic control*, AC-37, pp.1042-1046, July 1992.
[17] S. Marano, V. Matta and P. Willett, "The Importance of Being Earnest: Social Sensing With Unknown Agent Quality," *IEEE Transactions on Signal and Information Processing over Networks*, Vol. 2, No. 3, pp. 306-320, September 2016.
[18] A. Martinez, "Detector optimization using Shannon's information," *Proc. 20th Annual Conf. Information Sciences and Systems*, Princeton University, Princeton, NJ, March 1986.
[19] H.V. Poor, "Robust decision design using a distance criteria," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 575-587, September 1980.
[20] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag: New York, 1988.
[21] H.V. Poor and J.B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, pp. 893-900, September 1977.
[22] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring Sensors: A Low Communication Rate Scheme for Distributed Detection," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 32-2, pp. 554-568, April 1996.
[23] A.R. Reibman and L.W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerospace Electron. Systems*, AES-23, pp. 24-30, January 1987.
[24] S.D. Silvey, *Statistical Inference*, Chapman and Hall: London, 1974.
[25] Z. Tang, K. Pattipati, and D. Kleinman, "A distributed M-ary hypothesis testing problem with correlated observations," *IEEE transactions on automatic control*, AC-37, pp.1042-1046, July 1992.
[26] P. Swaszek and P. Willett, "Parley as an Approach to Distributed Detection," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 447-457, January 1995.
[27] R.R. Tenney and N.R. Sandell, Jr., "Detection with distributed sensors," *IEEE Trans. Aerospace Electron. Systems*, AES-17, pp.501-510, July 1981. pp.772-775, May 1989.
[28] S. Thomopoulos, R. Viswanathan, D. Bougoulias, "Optimal Distributed Decision Fusion," *IEEE Transactions on Aerospace and Electronic Systems*, pp.761-765, September 1989.
[29] J.N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing, vol. 2 – Signal Detection*, H. V. Poor and J. B. Thomas, editors, JAI Press, 1990.
[30] J.N. Tsitsiklis, "Extremal properties of likelihood ratio quantizers," Tech. Report #LIDS-P-1923, Laboratory for Information and Decision Sciences, Massachusetts Institute of Technology, Cambridge, MA, November, 1989.
[31] H.L. Van Trees, *Detection, Estimation, and Modulation Theory,* vol. 1, Wiley: New York, 1968.
[32] R. Viswanathan, A. Ansari, S. Thomopoulos, "Optimal Partitioning of Observations in Distributed Detection,", *Proceedings of International Symposium on Information Theory*, p197, 1988. bibitemviaa R. Viswanathan and V. Aalo, "On Counting Rules in Distributed Detection," 1989. *IEEE Transactions on Acoustics, Speech, and Signal Processing*,
[33] D. Wang, L. Kaplan, and T. Abdelzaher, "Maximum Likelihood Analysis of Conflicting Observations in Social Sensing," *ACM Trans. Sensor Networks*, vol. 10, no. 2, article 30, pp. 1–27, Jan. 2014.
[34] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On Credibility Estimation Tradeoffs in Assured Social Sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 6, pp. 1026–1037, Jun. 2013.
[35] D. Warren and P. Willett, "Optimal decentralized detection for conditionally independent sensors," *Proc. 1989 Amer. Control Conf.*, vol. 2, pp.1326-1329, Pittsburgh, June 1989.
[36] D. Warren and P. Willett, "Optimum Quantization for Detector Fusion: Some Proofs, Examples, and Pathology," *Journal of the Franklin Institute*, No. 336, pp. 323-359, 1999.
[37] P. Willett and D. Warren, "The Suboptimality of Randomized Tests in Quantized Detection Systems", *IEEE Transactions on Information Theory*, March 1992.
[38] P. Willett and D. Warren, "Randomization and Scheduling in Decentralized Detection," *Proceedings of* 27^{th} *Annual Allerton Conference on Communications, Control, and Computing*, September 1989.
[39] P. Willett, P. Swaszek, and R. Blum, "The Good, Bad, and Ugly: Distributed Detection of a Known Signal in Dependent Gaussian Noise," *IEEE Transactions on Signal Processing*, pp. 3266-3280, December 2000.