Fundamental Structures and Asymptotic Performance Criteria in Decentralized Binary Hypothesis Testing

Hakan Delic, P. Papantoni-Kazakos, *Fellow, IEEE,* and Demetrios Kazakos, *Fellow, IEEE*

***Abstract*—In this paper, two fundamental distributed decision network structures are considered: the first system consists of finite number of sensors, each collecting asymptotically many data, while the second one employs asymptotically many sensors, each collecting a single datum. For binary hypothesis testing, the Neyman–Pearson criterion is utilized and justified via information theoretic arguments. Asymptotic relative efficiency performance measure is used to establish tradeoffs between the two structures, by comparing the performance characteristics of the decentralized detection systems to their centralized counterparts.**

I. Introduction

THE ADVANCES in computer networks and computer technology, in general, have given rise to interest in distributed statistical techniques. In hypothesis testing, such techniques induce decentralized decision making at the sensor-level. Specifically, upon collecting the data, sensors transmit their decisions to a central processor, called the fusion center, instead of the whole string of observations. The central processor then declares the final decision based on the local decisions, and its own observations if there are any. The resulting distributed detection systems provide reliability and savings in communication costs, as well as the capability of integrating different sensing techniques, such as sonics, microwave, infrared, and x-ray sensors [2]. The advantages of distributed detection are achieved at the expense of suboptimal performance, due to information loss in local decision making.

Results and optimization procedures for centralized detection are well-established [8], [21]. Decentralized detection is not a straightforward extension of the classical hypothesis testing. In fact, the simultaneous optimization of the local detectors, along with the fusion rule, generally results in an NP-complete problem [20]. Early efforts concentrated on finding the optimal local decision rules within a Bayesian framework, assuming a fixed fusion rule [14], [16], and it was shown that the computations that yield the optimal thresholds are coupled. In [2], on the other hand, the optimal fusion rule is derived, assuming fixed local processors. More general approaches have been undertaken in [6] and [13], where the local decision rules and the fusion rule are optimized simultaneously; in both papers, a global cost function is minimized. The Neyman–Pearson criterion is used in [15], [18], and [19], for finding the optimal local detector structures. In particular, in [18], it is proven that with three or more sensors, the decentralized system can outperform its best sensor, in terms of detection and false alarm probabilities.

In all the works above, numerical examples have been utilized to compare centralized and decentralized detection systems. In this paper, we attempt to make a thorough and theoretical study of different distributed detection structures by comparing their performances to the optimal centralized scheme. In doing so, our main tool will be relative efficiencies. Relative efficiency of a test, and its large-sample counterpart asymptotic relative efficiency, have been widely used as a performance measure in nonparametric detection [17], and detection in dependent noise [12]. In particular, we will consider suitable versions of a definition due to Pitman [17], which describes the asymptotic relative efficiency as the ratio of the number of data needed by one test to achieve the same performance as a second test, as the hypotheses tend to get closer and sample size grows.

Fundamental concepts on hypothesis testing, as well as asymptotic performance measures, can be found in [5], [7], [10], and [21]. Concepts connecting hypothesis testing and information theory, which are used in this paper to establish the optimal decision rule are treated in [1]. The organization of the paper is as follows. In Section II, we provide an overview of distributed detection using the Neyman–Pearson criterion. Some modeling assumptions and asymptotic results on the performance of decentralized detection are established in Section III, and these results are extended to the location parameter case in Section IV. Robust test functions are considered in Section V. Conclusions can be found in Section VI. Proofs of the propositions and lemmata are presented in the Appendix.

II. Distributed Neyman–Pearson Detection

Let us consider the distributed decision network depicted in Fig. 1 whose function is binary hypothesis testing. The two hypotheses, H_0 and H_1, are assumed to be generated by discrete-time processes possessing density functions for all dimensionalities N of data sequences $x^N \triangleq \{x_1, \cdots, x_N\}$, where $f_{ij}(x_j^N)$, $i = 0, 1$, denotes the density function of hypothesis H_i at sensor j and the vector point x_j^N, $j =$

Paper approved by N. A. Zervos, the Editor for Transmission Systems of the IEEE Communications Society. Manuscript received August 8, 1991; revised March 3, 1992 and May 5, 1993. This work was supported by the National Science Foundation under Grants NCR-9003513 and MSS-9216372, and EPRI Contract RP8030-08. This paper was presented in part at the IEEE International Symposium on Information Theory, Budapest, Hungary, June 1991.

H. Delic and D. Kazakos are with the Department of Electrical and Computer Engineering, University of Southwestern Louisiana, Lafayette, LA 70504-3890 USA.

P. Papantoni-Kazakos is with the Department of Electrical Engineering, University of Alabama, Tuscaloosa, AL 35487 USA.

IEEE Log Number 9406248.

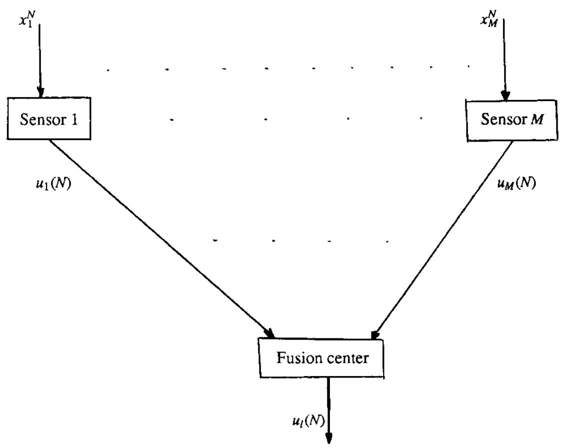


Fig. 1. A fundamental decision network.

$1, \cdots, M$. The output produced by each sensor at time N is a decision regarding the acting hypothesis and is a binary number. For sensor j, this binary number is denoted $u_j(N)$, and equals i, $i = 0, 1$, if the latter sensor decides in favor of hypothesis H_i, at time N. It is assumed that the same hypothesis is acting throughout the collection of all the data. The fusion center receives at time N the binary numbers $u_1(N), \cdots, u_m(N)$ from M other sensors; it produces as output (at time N) the binary number $u_l(N)$. For soft decision fusion, see [9]. Capital letters will denote random variables, while lower case letters will denote their realizations.

At time N, the fusion center performs a function on its inputs, $u_1(N), \cdots, u_M(N)$, to produce its output $u_l(N)$. This function is actually a decision rule, $\delta_l(N) = \delta_l(N; \{u_j(N)\}_{j=1}^M)$, and stands for the probability that H_1 is decided by the fusion center, at time N, given $\{u_j(N)\}_{j=1}^M$. Equivalently, the output, $u_i(N)$, of the fusion center at time N is as follows:

$$u_l(N) = \begin{cases} 1, & \text{with probability } \delta_l(N) \\ 0, & \text{with probability } 1 - \delta_l(N). \end{cases} \tag{1}$$

Regardless of the optimization criterion adopted, Bayesian or Neyman–Pearson, the optimal decision rule $\delta_l(N)$ consists of a log-likelihood test function, $T_l(N; \{u_j(N)\}_{j=1}^M)$ and a threshold, $\lambda_l(N)$, and generally has the following form:

$$\begin{aligned}\delta_l(N) &= \delta_l(N; \{u_j(N)\}_{j=1}^M) \\ &= \begin{cases} 1 & \text{if } T_l(N; \{u_j(N)\}_{j=1}^M) > \lambda_l(N) \\ r & \text{if } T_l(N; \{u_j(N)\}_{j=1}^M) = \lambda_l(N) \\ 0 & \text{if } T_l(N; \{u_j(N)\}_{j=1}^M) < \lambda_l(N) \end{cases}\end{aligned} \tag{2}$$

where $r \in [0, 1]$, and

$$T_l(N; \{u_j(N)\}_{j=1}^M) = \log \frac{P(\{u_j(N)\}_{j=1}^M | H_1)}{P(\{u_j(N)\}_{j=1}^M | H_0)}. \tag{3}$$

If $r \neq 0$, the decision rule $\delta_l(N)$ is called randomized; a randomized decision rule may be optimal only if the Neyman–Pearson optimization criterion is used [5], [8].

Let $\beta_j(N)$ be the power probability induced by sensor j at time N; that is, the probability that H_1 is decided given that H_1 is true. Also, let $\alpha_j(N)$ denote the false alarm probability induced by sensor j at time N; that is, the probability that H_1 is decided given that H_0 is true. Let us now assume that the external data received by different sensors are mutually independent, and that the M sensors whose outputs the fusion center receives have no common observations and they do not communicate with each other. Then, $U_1(N), \cdots, U_M(N)$ are all mutually independent, and

$$T_l(N; \{u_j(N)\}_{j=1}^M) = \sum_{j=1}^M w_j(N) u_j(N) \tag{4}$$

where

$$w_j(N) \triangleq \log \frac{\beta_j(N)[1 - \alpha_j(N)]}{\alpha_j(N)[1 - \beta_j(N)]}. \tag{5}$$

While the optimal test function employed by the fusion center is generally as in (3), and is as in (4) in the mutually independent case, the threshold $\lambda_l(N)$ in (2) is determined by the optimization criterion used. In this paper, we will use the Neyman–Pearson criterion, for the following reasons: a) In most applications, it is unreasonable to assume that the *a priori* probabilities of the hypotheses are fixed and known; thus, the Bayesian criterion is not usable then. b) The Neyman–Pearson criterion maximizes the information provided by the sensors about the acting hypothesis. This is formally stated in Proposition 1, whose proof can be found in [3].

Proposition 1: Given the set $\{\alpha_j(N)\}_{j=1}^M$ of false alarm probabilities, at time N, the information provided by the M sensors about the acting hypothesis is minimized, when the power probability induced by each sensor equals the corresponding false alarm probability (in which case, no information about the acting hypothesis is provided). The information about the acting hypothesis is maximized, when the power probability induced by each sensor is maximized. Thus, the optimization criterion, deployed by each sensor, which maximizes the information about the acting hypothesis is the Neyman–Pearson criterion. □

When the sensors use the Neyman–Pearson optimization criterion, each sensor is assigned a fixed false alarm probability or rate, that does not vary with time. Let α_j denote the false alarm rate assigned to sensor j. Furthermore, the decision threshold $\lambda_l(N)$ and the number r in (2), for the fusion center, are determined by the prespecified false alarm rate α_l. In particular, for the test function in (3), the threshold $\lambda_l(N)$ and the number r are such that

$$\begin{aligned}P[T_l(N; \{U_j(N)\}_{j=1}^M) > \lambda_l(N) | H_0] \\ + rP[T_l(N; \{U_j(N)\}_{j=1}^m) = \lambda_l(N) | H_0] = \alpha_l.\end{aligned} \tag{6}$$

As we already discussed, when the sensors are all mutually independent, the test function in (6) takes the form in (4). If the sensors are identical, with the assigned false alarm rate $\alpha_j = \alpha$, $j = 1, \cdots, M$, attaining power probability $\beta(N)$ at time N, then, given false alarm rate α_l for the fusion center, the decision rule $\delta_l(N)$ is generally randomized and has the form [13]

$$\delta_l(N) = \delta_l(N; \{u_j(N)\}_{j=1}^M) = \begin{cases} 1, & \text{if } \sum_{j=1}^M u_j(N) > K \\ r, & \text{if } \sum_{j=1}^M u_j(N) = K \\ 0, & \text{if } \sum_{j=1}^M u_j(N) < K, \end{cases} \tag{7}$$

where K and r are such that

$$\alpha_l = \sum_{k=K+1}^{M} \binom{M}{k} \alpha^k (1-\alpha)^{M-k} + r\binom{M}{K} \alpha^K (1-\alpha)^{M-K}. \tag{8}$$

The power attained by the fusion center at time N in then as follows:

$$\beta_l(N) = \sum_{k=K+1}^{M} \binom{M}{k} [\beta(N)]^k [1-\beta(N)]^{M-k} + r\binom{M}{K} [\beta(N)]^K [1-\beta(N)]^{M-K}. \tag{9}$$

From (7), we also conclude that, with K and r as in (8), the outputs of the fusion center are as follows:

$$u_l(N) = \begin{cases} 1, & \text{with probability } 1, \text{ if } \sum_{j=1}^M u_j(N) > K \\ 1, & \text{with probability } r, \text{ if } \sum_{j=1}^M u_j(N) = K \\ 0, & \text{otherwise.} \end{cases}$$

The randomization in (7) can be avoided by properly assigning the sensors to nonidentical thresholds [22].

III. Performance Studies

In this section, we will study a distributed decision network where the fusion center receives decisions from M sensors. The sensors base their decisions solely on their observations, and do not communicate with each other. We will adopt the assumptions that the external data received by different sensors in the structure are mutually independent, and that the Neyman–Pearson optimization criterion is used by all sensors, with a prespecified false alarm rate α for each sensor. We will consider identical statistics of the external data for different sensors; then, x_j^N and f_i, $i = 0, 1$, will respectively denote a length-N data sequence and the first order density function of the data sequences under hypothesis H_i that correspond to sensor j. We will also assume that the densities f_i are continuous everywhere. Then, (7), (8), and (9) hold, where, for each sensor, we have

$$u_j(N) = \begin{cases} 1, & \text{if } \log \frac{f_1(x_j^N)}{f_0(x_j^N)} \geq \lambda(N) \\ 0, & \text{if } \log \frac{f_1(x_j^N)}{f_0(x_j^N)} < \lambda(N) \end{cases} \tag{10}$$

where $u_j(N)$ is the decision declared by the jth sensor, $j = 1, \cdots, M$, at time N, and where $\lambda(N)$ is such that

$$\alpha = \int_{C_j^N} f_0(x_j^N)\, dx_j^N \tag{11}$$

and

$$C_j^N \triangleq \left\{ x_j^N: \log \frac{f_1(x_j^N)}{f_0(x_j^N)} \geq \lambda(N) \right\}.$$

The power probability $\beta_j(N)$ induced by the sensor j at time N is

$$\beta(N) = \int_{C_j^N} f_1(x_j^N)\, dx_j^N. \tag{12}$$

Referring to (11) and (12), it is well known that the threshold $\lambda(N)$ is uniquely determined, and that the power probability $\beta(N)$ is monotonically nondecreasing (and in nonpathological cases strictly increasing), with increasing N. In addition, it is easily concluded that the pair (K, r) in (8) is uniquely determined for every N, and that the power probability $\beta_l(N)$ attained by the fusion center is monotonically nondecreasing (and in nonpathological cases monotonically increasing), with increasing N [5], [10].

To gain understanding of the issues involved in system performance, we will focus on the case where the hypotheses per element are generated by stationary and memoryless stochastic processes. The distribution of the data at any sensor, generated by hypothesis H_i, $i = 0, 1$, is then completely described by the one-dimensional density function $f_i(x)$ (at scalar point x). We will assume that $f_i(x)$, $i = 0, 1$, are continuous for all x. Since it is assumed that the sensors are identical, in the sequel, the subscript j will be dropped. In the considered case, for each sensor, we have

$$\log \frac{f_1(x^N)}{f_0(x^N)} = \sum_{n=1}^{N} \log \frac{f_1(x_n)}{f_0(x_n)},$$

where the process $\{\log f_1(X_n)/f_0(X_n)\}_{n=1}^N$ is then stationary and memoryless. Let us define

$$Y(N) \triangleq \frac{1}{N} \sum_{n=1}^{N} \log \frac{f_1(X_n)}{f_0(X_n)}. \tag{13}$$

The random variable $Y(N)$ has, respectively, the following mean and variance, under the two hypotheses H_0 and H_1:

$$\mu_i \triangleq E\{Y(N)|H_i\} = \int dx\, f_i(x) \log \frac{f_1(x)}{f_0(x)}; \qquad i = 0, 1. \tag{14}$$

$$[\sigma_i(N)]^2 \triangleq E\{[Y(N) - \mu_i]^2 | H_i\} \triangleq \frac{1}{N}\rho_i^2 = \frac{1}{N}\left\{ \int dx f_i(x) \left[\log \frac{f_1(x)}{f_0(x)}\right]^2 - \left[\int dx\, f_i(x) \log \frac{f_1(x)}{f_0(x)}\right]^2 \right\}. \tag{15}$$

All integrals are taken with respect to the entire real line unless otherwise specified. Let $KL(f_1, f_0)$ denote the Kullback–Leibler distance of the density f_1 with respect to the density f_0; that is,

$$KL(f_1, f_0) \triangleq \int dx\, f_1(x) \log \frac{f_1(x)}{f_0(x)}. \tag{16}$$

We then note that the means μ_0 and μ_1 are actually Kullback–Leibler distances between the corresponding densities. In particular,

$$\mu_1 = KL(f_1, f_0), \quad \mu_0 = -KL(f_0, f_1). \tag{17}$$

Since $KL(f_1, f_0) \geq 0$, with equality if and only if $f_1(x) = f_0(x)$ a.e. (almost everywhere), we conclude that if $f_1(x) \neq f_0(x)$ a.e., then the mean μ_1 is strictly positive, and the mean μ_0 is strictly negative. In addition, from (15), we conclude that, if $\rho_i < \infty$, $i = 0, 1$, then the variances $[\sigma_i(N)]^2$, $i = 0, 1$, converge to zero as N increases to infinity, with rate inversely proportional to N. Finally, applying the central limit theorem, we conclude that, asymptotically ($N \to \infty$), the variable $Y(N)$ in (13) is then Gaussian, with mean equal to μ_i and variance equal to $[\sigma_i(N)]^2$, when the data are generated by hypothesis H_i, $i = 0, 1$. That is,

$$Y(N) \xrightarrow[N \to \infty]{} G(\mu_i, \sigma_i(N)), \quad \text{under hypothesis } H_i, i = 0, 1. \tag{18}$$

Asymptotic Performance—Stationary and Memoryless Processes

Considering the distributed detection structure in Fig. 1 and assuming hypotheses generated by stationary and memoryless processes per sensor, we seek to gain insight into the performance of the system and the involved trade-offs by studying the induced asymptotics. Specifically, we wish to compare the structure to another centralized optimal structure, when asymptotically many data are collected and fixed false alarm and power rates are attained. As well known, requiring that fixed false alarm and power rates be attained by asymptotically many data implies the assumption that the two hypotheses are "close" to each other (or hard to distinguish from each other) in an appropriate sense related to the Kullback–Leibler distance.

To compare the distributed topology in Fig. 1 to a completely centralized structure consisting of a single sensor, it is natural to consider the case where the M sensors are identical to f_1 and f_0 denoting, respectively, the densities that generate the hypotheses H_1 and H_0 per sensor. In the centralized structure, we will assume that the two hypotheses, H_1 and H_0, are generated by the same densities, f_1 and f_0.

Let $f_1(x)$ and $f_0(x)$ both be analytic a.e., bounded and such that

$$|\mu_1 - \mu_0| < \epsilon \tag{19}$$

and

$$\rho_1^2 \approx \rho_0^2 \triangleq \rho^2 < \infty, \tag{20}$$

where $0 < \epsilon \ll 1$.

Let $\Phi(x)$ denote the cumulative distribution function of the zero-mean and unit-variance Gaussian random variable at point x, and let $\Phi^{-1}(y) = x$ imply $\Phi(x) = y$. The proofs of all the following lemmata can be found in the Appendix.

Lemma 1: Consider a centralized system, where a single sensor collects data and makes decisions about the acting hypotheses, H_1 versus H_0, using the Neyman–Pearson optimization criterion. Let each hypothesis be generated by a stationary and memoryless process, where f_1 and f_0 denote then the one-dimensional density functions that generate hypotheses H_1 and H_0, respectively. Let $f_1(x)$ and $f_0(x)$ both be analytic a.e., satisfying the conditions in (19) and (20). Let a power rate β and a false alarm rate α be given. Then, the number $N_c(\alpha, \beta)$, of data needed to attain the (α, β) performance is asymptotically large and is given by the following expression:

$$N_c(\alpha, \beta) = \frac{\rho^2}{(\mu_1 - \mu_0)^2}[\Phi^{-1}(\beta) - \Phi^{-1}(\alpha)]^2. \tag{21}$$

□

Next, assuming the conditions given in Lemma 1 are still valid, we will find the number of data needed by the decentralized structure to attain the (α, β) performance, and then compare the latter number to that in (21) to obtain the asymptotic relative efficiency (ARE) of the decentralized structure. Two cases will be studied: Structure 1, where the number of sensors in the decentralized structure is finite and each element collects asymptotically many data, and Structure 2, where M is asymptotically large and each of the M sensors collects a single datum.

Lemma 2: Consider M mutually independent and identical sensors, where the decision of the jth sensor at time N is denoted $u_j(N)$. For each of the M sensors, the two hypotheses are generated by stationary and memoryless processes, with first order density functions f_1 and f_0, where $f_1(x)$ and $f_0(x)$ are both analytic a.e., satisfying the conditions in Lemma 1. The decisions $\{u_j(N)\}_{j=1}^M$ are transmitted to the fusion center, which assimilates them and produces a final decision $u_l(N)$, at time N. Let M be finite, and let false alarm and power rates, α and β, be given. Then, the performance (α, β) is attained asymptotically, as $N \to \infty$ per sensor. The overall number of data (across all the sensors) needed to attain α and β is given by the following expression:

$$N(M, \alpha, \beta) = M\frac{\rho^2}{(\mu_1 - \mu_0)^2} \cdot \left\{\inf_{(y, z) \in D(\alpha, \beta)}[\Phi^{-1}(z) - \Phi^{-1}(y)]\right\}^2 \tag{22}$$

where

$$\begin{aligned} D(\alpha, \beta) \triangleq \Bigg\{(y, z)\colon\ & 0 < y < z \leq 1,\ 0 \leq r \leq 1,\ 0 \leq K \leq M, \\ & \alpha = \sum_{k=K+1}^{M}\binom{M}{k}y^k(1-y)^{M-k} + r\binom{M}{K}y^K(1-y)^{M-K}, \\ & \beta = \sum_{k=K+1}^{M}\binom{M}{k}z^k(1-z)^{M-k} + r\binom{M}{K}z^K(1-z)^{M-K}\Bigg\}. \end{aligned} \tag{23}$$

For $N(M, \alpha, \beta)$ as in (22) and $N_c(\alpha, \beta)$ as in (21), the ratio $N(M, \alpha, \beta)/N_c(\alpha, \beta)$ determines the ARE of the decentralized structure at α and β. This ARE, denoted by $\text{ARE}(M, \alpha, \beta)$, is given by the following expression:

$$\text{ARE}(M, \alpha, \beta) \triangleq \frac{N(M, \alpha, \beta)}{N_c(\alpha, \beta)} = M\left\{\frac{\inf_{(y, z)\in D(\alpha, \beta)}[\Phi^{-1}(z) - \Phi^{-1}(y)]}{\Phi^{-1}(\beta) - \Phi^{-1}(\alpha)}\right\}^2. \tag{24}$$

□

Lemma 3: Let us consider a decentralized system as in Lemma 2, with the conditions in Lemma 1 holding. Let each of the M sensors collect a single datum, and let the same performance pair (α, β) as in Lemmata 1 and 2 be required. Then, the (α, β) performance is attained by asymptotically many sensors, as $M \to \infty$. The number, $M(\alpha, \beta)$, of sensors needed to attain this performance is given by the following expression:

$$M = M(\alpha, \beta) = \inf_{\lambda}\left\{\frac{[p_{\lambda,1}(1-p_{\lambda,1})]^{1/2}}{p_{\lambda,1}-p_{\lambda,0}}\Phi^{-1}(\beta) - \frac{[p_{\lambda,0}(1-p_{\lambda,0})]^{1/2}}{p_{\lambda,1}-p_{\lambda,0}}\Phi^{-1}(\alpha)\right\}^2, \quad (25)$$

where

$$p_{\lambda,0} \triangleq \int_C dx\, f_0(x),\ p_{\lambda,1} \triangleq \int_C dx\, f_1(x),$$

and

$$C \triangleq \left\{x: \frac{f_1(x)}{f_0(x)} \geq \lambda\right\}.$$

Let in addition, $p_{\lambda,0}(1-p_{\lambda,0}) \approx p_{\lambda,1}(1-p_{\lambda,1})$, $\forall\lambda$. For $M(\alpha, \beta)$ as in (25) and $N_c(\alpha, \beta)$ as in (21), the ratio $M(\alpha, \beta)/N_c(\alpha, \beta)$ is the ARE of the decentralized structure considered in this lemma, at α and β. This ARE, denoted by ARE(∞), is given by the following expression:

$$\text{ARE}(\infty) \triangleq \frac{M(\alpha, \beta)}{N_c(\alpha, \beta)} = \left\{\inf_{\lambda}\frac{p_{\lambda,0}(1-p_{\lambda,0})}{[p_{\lambda,1}-p_{\lambda,0}]^2}\right\}\frac{(\mu_1-\mu_0)^2}{\rho^2}. \quad (26)$$

□

Remarks: Structures 1 and 2, considered in Lemmata 2 and 3, respectively, can be viewed as antipodal cases for asymptotic studies of distributed decision networks. The one in Lemma 2 involves a finite, or even small, number of sensors, while the one in Lemma 3 assumes infinitely many sensors. From the expressions of the asymptotic relative efficiencies, (24) in Lemma 2 and (26) in Lemma 3, we initially observe that the "best" among the two antipodal network structures generally depends on a number of parameters and characteristics; namely, the false alarm and power system performance required, as well as the distribution of the data.

IV. The Location Parameter Case

In this section, we focus on results regarding the location parameter case. In particular, we are seeking the derivations of specific results in Lemma 3, and comparisons between the structures in Lemmata 2 and 3, when the densities f_1 and f_0 are such that $f_1(x) = f_0(x-\theta)$, with $\theta \ll 1$. To pursue our goal, we will need the following proposition, whose proof is in the Appendix:

Proposition 2: Let $f(x)$ be a density function defined on the real line, which is also analytic in x a.e. Let $f'(x)$ denote the first derivative at x and let $f^{(k)}(x)$ denote the kth derivative at x. Define

$$F(x) \triangleq \int_{-\infty}^{x} f(y)\,dy, \quad h(x) \triangleq \frac{F(x)}{f(x)}.$$

Let $f(x)$ be symmetric around zero, and such that

$$f(\infty) = f(-\infty) = 0, \quad f^{(k)}(\infty) = f^{(k)}(-\infty) = 0, \qquad \forall k,$$

$$-\frac{f'(x)}{f(x)} \text{ is monotonically increasing with } x,$$

$$-\frac{f'(x)}{f(x)} > \frac{h(x)-h(-x)}{2h(x)h(-x)}; \qquad \forall x > 0.$$

Let us define

$$Y(x) \triangleq \frac{F(x)F(-x)}{f^2(x)} = \frac{F(x)}{f(x)}\cdot\frac{F(-x)}{f(-x)}.$$

Then, $Y(x)$ attains its minimum at $x = 0$, where

$$Y(0) = \inf_x Y(x) = \min_x Y(x) = \frac{1}{4f^2(0)}.$$ □

We now present a result, which is a special case of Lemma 3.

Lemma 4: Consider the network structure and the assumptions in Lemma 3. Suppose

$$f_0(x) = f(x), \quad f_1(x) = f(x-\theta), \quad (27)$$

where $\theta \ll 1$, and $f(x)$ satisfies all the conditions in Proposition 2. Let also $f^{(k)}(x)f^{(n)}(x)f^{-1}(x)$ be integrable for all $k < \infty$ and $n < \infty$. Then, the ARE in (26) becomes

$$\text{ARE}(\infty) = \frac{1}{4f^2(0)}\int dx\frac{[f'(x)]^2}{f(x)}, \quad (28)$$

where $\int dx\,[f'(x)]^2/f(x)$ is the Fisher information of the density $f(x)$. □

Remarks: Two important and interesting cases arise, when the density $f(x)$ is Gaussian or Laplacian, where we obtain specific numbers for the ARE(∞) in (28), given below.

1) Let $f(x)$ be zero-mean Gaussian with variance $\sigma^2 < \infty$. Then, if $\phi(x)$ denotes the density function of the zero-mean and unit-variance Gaussian random variable at the point x, we have

$$f(x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right), \quad f'(x) = -\frac{x}{\sigma^3}\phi\left(\frac{x}{\sigma}\right),$$
$$f(0) = \frac{1}{\sigma\sqrt{2\pi}}, \quad \int dx\,\frac{[f'(x)]^2}{f(x)} = \frac{1}{\sigma^2}.$$
$$\text{ARE}(\infty) = \frac{\pi}{2}. \quad (29)$$

2) Let $f(x)$ be Laplacian with variance $\sigma^2 < \infty$. That is,

$$f(x) = \frac{1}{\sqrt{2}\sigma}e^{-\sqrt{2}|x|/\sigma}.$$

Then,

$$f'(x) = -\frac{e^{-\sqrt{2}|x|/\sigma}}{\sigma^2}\,\text{sgn}(x), \quad f(0) = \frac{1}{\sqrt{2}\sigma},$$

$$\int dx\,\frac{[f'(x)]^2}{f(x)} = 2\int_0^\infty dx\,\frac{[f'(x)]^2}{f(x)} = \frac{2}{\sigma^2}.$$

$$\text{ARE}(\infty) = 1. \quad (30)$$

From the results in 1) and 2) above, we observe that in the Gaussian and location parameter case, asymptotically, the decentralized structure requires $\pi/2$ times the data required by the centralized structure to attain the same false alarm and power performance. When the data are Laplacian distributed, on the other hand, the decentralized and centralized structures are asymptotically equivalent; they both require the same number of data for identical performance.

The results in 2) above demonstrate that there is something unique about the Laplacian density. Along these lines and in view of the result in Lemma 4, which shows that, in the location parameter case, ARE(∞) is basically determined by the Fisher information of the density function that generates the data, we will state a proposition whose proof can be found in the Appendix.

Proposition 3: Consider the class of first order density functions which are defined on the real line, analytic a.e., symmetric around zero, monotonically decreasing in $(0, \infty)$, converge to zero at infinity, and have a fixed value at zero. Within this class, the Laplacian density attains the minimum Fisher information. □

In view of Proposition 3, and in conjunction with Lemma 4, it is naturally concluded that, in the location parameter case, the Structure 2 attains its best performance when the data are Laplacian distributed. Then, the decentralized structure is asymptotically identical to the "absolutely optimal" centralized structure in terms of performance, while the former induces tremendous savings in communication cost as compared to the latter (binary decisions communicated rather than analog data, or equivalently binary rather than analog communication channels used).

In comparison between the decentralized structures in Lemma 2 and Lemma 3, for the location parameter case, we can conclude that Structure 2 is superior in terms of the ARE measure. However, when the distribution of the data is unknown, or not well defined, or a multipurpose distributed decision network with a predictable performance is sought regardless of the data distributions, the structure with few sensors and many data per sensor has an advantage and may then be desirable. Hence, there is a performance versus robustness tradeoff between the two structures. This kind of tradeoff is typical in many systems; the final selection of the deployed network structure depends on the environment within which it operates and the objectives decided upon.

Some Additional Asymptotic Results and Comparisons

Let us focus on the network structure in Lemma 2, where a limited number of sensors are used and each collects a large number of data. We already discussed some of the advantages and disadvantages of this structure, and we computed its asymptotic relative efficiency precisely, when the number of sensors is two. Revisiting the expression for ARE(M, α, β), in (24) and Lemma 2, two questions arise naturally. Namely, a) how does ARE(M, α, β) vary with changing M, and b) what is the best ARE performance that the structure can attain when the number of sensors varies? Answers to both questions can be provided by the study of the case where the number of sensors is large. Indeed, it is naturally expected that the network will then attain its best performance. In addition, the results of this study will provide inputs as to how ARE(M, α, β) varies when M does. Assuming studies as above, we start with a proposition.

Proposition 4: Consider the network structure in Lemma 2, and assume that $M \gg 1$. Then, ARE(M, α, β) in (24), takes the following form:

$$\text{ARE}(M, \alpha, \beta) = M\left\{\frac{\inf_{(y,z)\in F(\alpha,\beta)}[\Phi^{-1}(z) - \Phi^{-1}(y)]}{\Phi^{-1}(\beta) - \Phi^{-1}(\alpha)}\right\}^2 \tag{31}$$

where

$$\begin{aligned} F(\alpha, \beta) = \Bigg\{y, z\colon \frac{1}{(z-y)}\Bigg[&[z(1-z)]^{1/2} \\ &+ [y(1-y)]^{1/2}\frac{\Phi^{-1}(1-\alpha)}{\Phi^{-1}(\beta)}\Bigg] \\ &= \frac{M^{1/2}}{\Phi^{-1}(\beta)},\ 0 < y < z \le 1\Bigg\}. \end{aligned} \tag{32}$$

□

To proceed with the search for the infimum in Proposition 4, we will adopt the special simpler case where $\beta = 1 - \alpha$ to avoid complications. This case will still provide us with the inputs we are searching for. When $\beta = 1 - \alpha$, the optimization problem takes the following form:

$$\inf_{(y,z)}[\Phi^{-1}(z) - \Phi^{-1}(y)] \tag{33}$$

subject to the constraints

$$0 < y < z \le 1 \tag{34}$$

and

$$\frac{1}{(z-y)}\{[z(1-z)]^{1/2} + [y(1-y)]^{1/2}\} = \frac{M^{1/2}}{\Phi^{-1}(1-\alpha)}. \tag{35}$$

The solution of the optimization problem in (33), (34), and (35) is included in the lemma below.

Lemma 5: The solution of the optimization problem in (33), (34), and (35) is as follows:

$$\inf_{(y,z)}[\Phi^{-1}(z) - \Phi^{-1}(y)] = 2\Phi^{-1}\left(\frac{1}{2} + \frac{1}{2\sqrt{1+A^2}}\right),$$

attained for

$$z = \frac{1}{2} + \frac{1}{2\sqrt{1+A^2}}$$

and

$$y = \frac{1}{2} - \frac{1}{2\sqrt{1+A^2}}$$

where

$$A \triangleq \frac{M^{1/2}}{\Phi^{-1}(1-\alpha)}.$$ □

The result in Lemma 5 provides the means for the computation of ARE$(M, \alpha, 1-\alpha)$, $M \gg 1$, in a relatively straightforward fashion. This is expressed in Lemma 6 below.

Lemma 6: Let α and $\beta = 1 - \alpha$ be given, and let M be such that $M \gg [\Phi^{-1}(1-\alpha)]^2$. Then,

$$\text{ARE}\,(M,\,\alpha,\,1-\alpha) = \frac{\pi}{2}. \qquad \square$$

Remarks: The result in Lemma 6 leads to a number of observations: a) Comparing the result in Lemma 6 to that in (29), we conclude that when the data are Gaussian and the hypotheses are distinguished by a shift in location parameter, the network structure in Lemma 2 is uniformly (for all M in the structure of Lemma 2) inferior to that in Lemma 3, while the two structures become asymptotically equivalent, as $M \to \infty$ in the structure of Lemma 2. Thus, in the Gaussian and location parameter case, the network structure which involves a large number of sensors with a single datum per sensor is preferable. b) Comparing the result in Lemma 6 to that in (30), we observe again the superiority of the network structure in Lemma 3, when the data are Laplacian distributed and the two hypotheses are determined by a location parameter shift.

V. Robust Test Functions for Asymptotically Many Sensors

Let us consider Structure 2 in the presence of hypotheses generated by stationary and memoryless processes. The structure consists of large number of identical sensors, each collecting a single datum. The main assumption here is that each sensor deploys a robust test function. We will specifically consider the adoption of a particular such test function, whose powerful properties are known and have been thoroughly studied [7], [8]. In particular, let the output of the jth sensor be as follows:

$$u_j = \begin{cases} 1 & \text{if } z_d(x_j) \geq \lambda \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1, \cdots, M$, where, for some $d > 0$, $z_d(x)$ is given below.

$$z_d(x) = \begin{cases} d & \text{if } x \geq d \\ x & \text{if } -d < x < d \\ -d & \text{if } x \leq -d. \end{cases} \tag{37}$$

The test function in (37) has strong performance merits and protects the system from performance breakdowns when highly erroneous data occur.

Given the decentralized structure, we present the following lemma to assess the performance of the test function $z_d(x)$ in terms of asymptotic relative efficiency. The proof of the lemma can be found in [4].

Lemma 7: Consider the decentralized structure with large number of sensors, where each sensor deploys the test function $z_d(x)$ in (34). Consider also a centralized, single-element structure whose test function, upon the collection of N data, is $N^{-1}\sum_{j=1}^{N} z_d(x_j)$. For both structures, let the acting hypotheses be generated by stationary and memoryless Gaussian processes. In particular, let the density function per datum be $\sigma^{-1}\phi(x/\sigma)$, under the H_0 hypothesis, and $\sigma^{-1}\phi[(x-\theta)/\sigma]$, under the H_1 hypothesis. Let σ be arbitrary and let $\theta < \epsilon$, where $0 < \epsilon \ll 1$. The asymptotic relative efficiency of the decentralized structure at the Gaussian process, when the test function $z_d(x)$ is deployed, is defined as the ratio of the number of sensors needed by this structure over the number of data needed by the centralized structure to attain the same false alarm and power rates, when the acting hypotheses are generated by the Gaussian processes described above. This asymptotic relative efficiency is denoted by ARE $(\infty,\, z_d,\, G)$, and it is given below.

$$\text{ARE}\,(\infty,\, z_d,\, G) = \frac{\pi}{2}\eta(d/\sigma) \tag{38}$$

where

$$\eta(x) \triangleq \frac{[2\Phi(x)-1]^2}{2[1-x^2]\Phi(x) + 2x^2 - 1 - 2x\phi(x)}. \qquad \square$$

Comparing the result in (38) to (29), we observe the following for hypotheses generated by stationary and memoryless Gaussian processes with location parameter shift: when the structure deploys the robust test function in (37), its ARE is $\eta(d/\sigma)$ times the ARE attained when the structure deploys the optimal-at-Gaussian test function. It can be shown that $\eta(x)$ is monotonically increasing in x, for $x \geq 0$, and it is smaller than one for all x in $(0, \infty)$. Furthermore, $\lim_{x\to\infty}\eta(x) = 1$ and $\lim_{x\to 0}\eta(x) = 2/\pi$. We thus conclude that, for the Gaussian hypotheses considered, Structure 2 has better asymptotic performance when the robust test function in (37) is deployed (by both the centralized and decentralized structures) as compared to the performance when the optimal-at-Gaussian model test function is adopted instead. In addition, when the robust test function is employed by the decentralized structure, powerful resistance to highly erroneous data is accomplished as well. In fact, it can proven that the deployment of the test function $z_d(x)$ in (37) by the decentralized structure presents a powerful advantage for a whole class of analytic, symmetric and unimodal density functions, when the location parameter case is considered [4].

VI. Comments and Conclusions

In Section III, we studied two fundamental distributed decision network structures. Both structures include a number of sensors and a fusion center, where the sensors collect external data and feed the fusion center with their outputs. In one structure, named Structure 1, relatively few sensors are included, each collecting many data. In the other structure, named Structure 2, a large number of sensors are included, each collecting a single datum.

To gain insight into the performance of the two structures, we assumed that the stochastic processes that generate the hypotheses are stationary and memoryless. We then derived asymptotic results regarding the performance of the two network structures, as compared to that of the absolutely optimal, completely centralized single-sensor system, using the asymptotic relative efficiency (ARE) performance measure and adopting the Neyman–Pearson optimization criterion for the decisions performed by each sensor.

We found that the ARE performance of Structure 1 is independent of the processes that generate the hypotheses, while the ARE performance of Structure 2 is directly determined by the characteristics of these processes. This quality makes Structure 1 more "robust" regarding sensitivity to data distributions, generally at the possible expense of reduced performance at particular data distributions, as compared to Structure 2. We also found that Structure 2 is equivalent to the absolutely optimal centralized structure when the two hypotheses are generated by Laplacian processes, while then Structure 1 is by far inferior.

When the processes that generate the two hypotheses are memoryless, the results indicate that Structure 2 may be superior to Structure 1 for at least a class of interesting data distributions, such as the Gaussian and location parameter case. This, however, does not necessarily imply superiority of Structure 2 when the processes that generate the hypotheses have memory (correlated data). Indeed, Structure 2 does not then take advantage of the possibly rich-in-correlation data structure if the sensors are mutually independent, while Structure 1 does. We postulate that, when the sensors in both structures are mutually independent and the processes that generate the two hypotheses are rich in memory, then Structure 1 will be superior to Structure 2 in terms of performance, in addition to being robust. The level of memory richness needed to make Structure 1 preferable to Structure 2 (in terms of performance), for certain data distributions, such as the Gaussian, is an open and highly interesting problem.

The decision regarding the selection of network Structures, 1 versus 2, generally depends on a number of tradeoffs. An important such tradeoff is performance at particular data distributions versus robustness (insensitivity to the specifics of such distributions).

In this paper, we also studied decentralized detection with large number of sensors, where each sensor deployed a robust test function. Our results indicate that if the robust test function is used, even when the acting hypotheses are Gaussian, the decentralized structure exhibits better ARE characteristics, while resistance to outlier data is simultaneously attained.

Appendix

Proof of Lemma 1: The decision performed by the centralized detector is as follows:

$$u(N) = \begin{cases} 1, & \text{if } \frac{1}{N}\sum_{n=1}^{N} \log \frac{f_1(x_n)}{f_0(x_n)} \geq \lambda(N) \\ 0 & \text{if } \frac{1}{N}\sum_{n=1}^{N} \log \frac{f_1(x_n)}{f_0(x_n)} < \lambda(N) \end{cases}$$

where, given the false alarm rate α, the threshold $\lambda(N)$ is determined by the equality

$$\alpha = P\left(\frac{1}{N}\sum_{n=1}^{N} \log \frac{f_1(X_n)}{f_0(X_n)} \geq \lambda(N) | H_0\right).$$

But, given the condition in (19) and given α and β, the pair (α, β) can only be attained asymptotically, as $N \to \infty$. Asymptotically, as discussed in Section III, the test function $N^{-1}\sum_{n=1}^{N} \log\{f_1(X_n)/f_0(X_n)\}$ is Gaussian under both hypotheses. In particular,

$$Y(N) \triangleq \frac{1}{N}\sum_{n=1}^{N} \log \frac{f_1(X_n)}{f_0(X_n)} \xrightarrow[N\to\infty]{} G\left(\mu_i, \frac{\rho}{\sqrt{N}}\right)$$

under hypothesis H_i, $i = 0, 1$. Then, for $N \to \infty$, we have

$$\alpha = 1 - \Phi\left(\frac{\lambda(N) - \mu_0}{\rho}\sqrt{N}\right) \to$$
$$\lambda(N) = \mu_0 + \frac{\rho}{\sqrt{N}}\Phi^{-1}(1-\alpha).$$

$$\beta = 1 - \Phi\left(\frac{\lambda(N) - \mu_1}{\rho}\sqrt{N}\right)$$
$$= \Phi\left(\frac{\mu_1 - \mu_0}{\rho}\sqrt{N} - \Phi^{-1}(1-\alpha)\right). \quad \text{(A.1)}$$

The result follows directly from (A.1). □

Proof of Lemma 2: Due to the closeness of the two hypotheses, any given pair of false alarm and power rates (y, z), $y < z$, is attained by each of the M sensors for asymptotically many data, $N(y, z)$. Directly from (21) we then have

$$N(y, z) = [\Phi^{-1}(z) - \Phi^{-1}(y)]^2 \frac{\rho^2}{(\mu_1 - \mu_0)^2}. \quad \text{(A.2)}$$

From (7), (8), and (9), it also follows that, given the false alarm rate α, the fusion center makes decision $u_l(N)$, and attains power probabilities $\beta(N)$ that are as follows:

$$u_l(N) = \begin{cases} 1, & \text{with probability 1, if } \sum_{j=1}^{M} u_j(N) > K \\ 1, & \text{with probability } r, \text{ if } \sum_{j=1}^{M} u_j(N) = K \\ 0, & \text{otherwise,} \end{cases}$$

where r and K are such that

$$\alpha = \sum_{k=K+1}^{M} \binom{M}{k} y^k (1-y)^{M-k} + r\binom{M}{K} y^K (1-y)^{M-K},$$

and where

$$\beta(N) = \sum_{k=K+1}^{M} \binom{M}{k} z^k (1-z)^{M-k} + r\binom{M}{K} z^K (1-z)^{M-K}.$$

If, in addition to α, a power rate β is required, then we conclude that the smallest number of data needed per sensor to attain (α, β) is $\inf_{(y,z)\in D(\alpha,\beta)} N(y, z)$, for $N(y, z)$ as in (A.2). Considering then the total number of data needed by all the sensors to attain the (α, β) performance, we find $N(M, \alpha, \beta)$ in (22). $\text{ARE}(M, \alpha, \beta)$ is obtained by direct division of (22) by (21). □

Proof of Lemma 3: Considering the decentralized structure, and given α and β, due to the closeness of the hypotheses as described in Lemma 1, the rates α and β can be attained only by asymptotically many sensors ($M \to \infty$). The test function, $M^{-1}\sum_{j=1}^{M} U_j$, used by the fusion center, is then Gaussian, under both hypotheses, by the central limit theorem. In particular, we have

$$\frac{1}{M}\sum_{j=1}^{M} U_j \xrightarrow[M\to\infty]{} G(\mu_i, M^{-1/2}\sigma_i), \quad \text{under } H_i, \quad i = 0, 1$$

where now

$$\mu_i \triangleq E\{U_j | H_i\} = \int_C dx\, f_i(x) = p_{\lambda, i}, \qquad i = 0, 1,$$

$$\sigma_i^2 \triangleq E\{[U_j - \mu_i]^2 | H_i\} = p_{\lambda,i}(1 - p_{\lambda,i}), \qquad i = 0, 1,$$

and

$$u_j = \begin{cases} 1, & \text{if } \frac{f_1(x_j)}{f_0(x_j)} \geq \lambda \\ 0, & \text{if } \frac{f_1(x_j)}{f_0(x_j)} < \lambda. \end{cases}$$

The decision u_l by the fusion center is as follows:

$$u_l = \begin{cases} 1, & \text{if } \frac{1}{M}\sum_{j=1}^{M} u_j \geq \nu \\ 0, & \text{if } \frac{1}{M}\sum_{j=1}^{M} u_j < \nu \end{cases}$$

where, for $M \to \infty$ and for given false alarm rate α, we have

$$\alpha = 1 - \Phi\left(\frac{\nu - p_{\lambda,0}}{[p_{\lambda,0}(1 - p_{\lambda,0})]^{1/2}}\sqrt{M}\right)$$

and thus

$$\nu = p_{\lambda,0} + \left[\frac{p_{\lambda,0}(1 - p_{\lambda,0})}{M}\right]^{1/2} \Phi^{-1}(1 - \alpha). \qquad \text{(A.3)}$$

Then, the power probability β_M attained when the number of sensors is M and asymptotically large, is given by the following expression for ν as in (A.3)

$$\begin{aligned} \beta_M &= \Phi\left(\frac{p_{\lambda,1} - \nu}{[p_{\lambda,1}(1 - p_{\lambda,1})]^{1/2}}\sqrt{M}\right) \\ &= \Phi\left(\frac{p_{\lambda,1} - p_{\lambda,0}}{[p_{\lambda,1}(1 - p_{\lambda,1})]^{1/2}}\sqrt{M} \right. \\ &\quad \left. - \left[\frac{p_{\lambda,0}(1 - p_{\lambda,0})}{p_{\lambda,1}(1 - p_{\lambda,1})}\right]^{1/2} \Phi^{-1}(1 - \alpha)\right) \qquad \text{(A.4)} \end{aligned}$$

Requiring that $\beta_M = \beta$, and solving (A.4) for M, we obtain that the minimum number of sensors needed to attain α and β is as in (25). The ARE(∞) in (26) is obtained by direct division of (25) by (21). □

Proof of Proposition 2: Under the conditions in Proposition 2, $Y(x) = Y(-x)$, and $Y'(x) > 0$, $\forall x > 0$. Due to the symmetry of $Y(x)$, $Y'(x) < 0$, $\forall x < 0$. Thus, the function $Y(x)$ has a minimum attained at $x = 0$. □

Proof of Lemma 4: In the special case of the Lemma and subject to the conditions there, first order approximation via Taylor expansion with respect to θ is valid. Then

$$f_1(x) = f(x - \theta) \approx f(x) - \theta f'(x),$$

$$f_1(x) - f_0(x) = f(x - \theta) - f(x) \approx -\theta f'(x),$$

$$\frac{f_1(x)}{f_0(x)} = \frac{f(x - \theta)}{f(x)} \approx 1 - \theta\frac{f'(x)}{f(x)},$$

$$\frac{(\mu_1 - \mu_0)^2}{\rho^2} \approx \theta^2 \int \frac{[f'(x)]^2}{f(x)}\, dx. \qquad \text{(A.5)}$$

Due to the assumed monotonicity of $-f'(x)/f(x)$, we conclude that $1 - \theta f'(x)/f(x) > \lambda$ corresponds to $x > \zeta$, for some ζ. Thus

$$\begin{aligned} p_{\lambda,0} &= \int_{\left\{x: \frac{f_1(x)}{f_0(x)} \geq \lambda\right\}} dx\, f_0(x) \approx \int_{\left\{x: 1 - \theta\frac{f'(x)}{f(x)} \geq \lambda\right\}} dx\, f(x) \\ &= \int_{\{x: x \geq \zeta\}} dx\, f(x) = 1 - F(\zeta) = F(-\zeta), \end{aligned}$$

$$\begin{aligned} p_{\lambda,1} &= \int_{\left\{x: \frac{f_1(x)}{f_0(x)} \geq \lambda\right\}} dx\, f_1(x) \\ &\approx \int_{\left\{x: 1 - \theta\frac{f'(x)}{f(x)} \geq \lambda\right\}} dx [f(x) - \theta f'(x)] \\ &= \int_{\{x: x \geq \zeta\}} dx\, f(x) - \theta \int_{\{x: x \geq \zeta\}} dx\, f'(x) \\ &= F(-\zeta) + \theta f(\zeta). \end{aligned}$$

$$\begin{aligned} p_{\lambda,1}(1 - p_{\lambda,1}) &\approx p_{\lambda,0}(1 - p_{\lambda,0}) \\ &\approx F(-\zeta)[1 - F(-\zeta)] = F(\zeta)F(-\zeta). \end{aligned}$$

$$p_{\lambda,1} - p_{\lambda,0} \approx \theta f(\zeta).$$

$$\frac{p_{\lambda,0}(1 - p_{\lambda,0})}{[p_{\lambda,1} - p_{\lambda,0}]^2} \approx \frac{F(\zeta)F(-\zeta)}{\theta^2 f^2(\zeta)}. \qquad \text{(A.6)}$$

Substituting (A.5) and (A.6) in (26), we obtain

$$\text{ARE}(\infty) = \left\{\inf_{\zeta} \frac{F(\zeta)F(-\zeta)}{f^2(\zeta)}\right\} \int dx\, \frac{[f'(x)]^2}{f(x)}.$$

But, due to Proposition 2, $\inf_\zeta F(\zeta)F(-\zeta)/f^2(\zeta) = 1/4f^2(0)$, and we thus obtain the result in (28). □

Proof of Proposition 3: a) Given some arbitrary density $f(x)$, $x \in (-\infty, \infty)$, and applying the Schwartz inequality, we obtain

$$\begin{aligned} \left[\int dx\, f(x)\left|\frac{f'(x)}{f(x)}\right|\right]^2 &\leq \int dx\, f(x)\left[\frac{f'(x)}{f(x)}\right]^2 \\ &= \int dx\, \frac{[f'(x)]^2}{f(x)}, \qquad \text{(A.7)} \end{aligned}$$

with equality if and only if

$$\left|\frac{f'(x)}{f(x)}\right| = c, \quad \text{a.e.} \qquad \text{(A.8)}$$

for some constant c. Within the considered class of densities, the only one that satisfies (A.8), and thus satisfies (A.7) with equality, is the Laplacian density.

b) For the class of densities considered, and for some density $f(x)$ within the class, we have

$$\begin{aligned} \int dx\, f(x)\left|\frac{f'(x)}{f(x)}\right| &= -2\int_0^\infty dx\, f(x)\frac{f'(x)}{f(x)} \\ &= -2\int_0^\infty dx\, f'(x) = 2f(0). \end{aligned}$$

c) Combining the results in a) and b), we finally conclude that

$$\int dx \frac{[f'(x)]^2}{f(x)} \geq 4f^2(0) = c,$$

$$\text{ARE}\,(M,\,\alpha,\,1-\alpha) = M\frac{\{\Phi^{-1}\{\frac{1}{2}+\frac{1}{2}\Phi^{-1}(1-\alpha)\{M+[\Phi^{-1}(1-\alpha)]^2\}^{-1/2}\}\}^2}{[\Phi^{-1}(1-\alpha)]^2},\quad M \gg 1. \tag{A.17}$$

for all f belonging to the class described in the Proposition, and with equality if and only if $f(x)$ is the Laplacian density. □

Proof of Proposition 4: When M is large, the binomial distributions in (23) are approximated well by Gaussian distributions with means and variances respectively given by My versus Mz, and $My(1-y)$ versus $Mz(1-z)$, using DeMoivre–Laplace Theorem [11]. The binomial sums in (23) then take the following form, for the appropriate ν value.

$$\alpha = 1 - \Phi\left(\frac{\nu - My}{[My(1-y)]^{1/2}}\right)$$
$$\beta = 1 - \Phi\left(\frac{\nu - Mz}{[Mz(1-z)]^{1/2}}\right). \tag{A.9}$$

From the equations in (A.9) and the equality $\Phi^{-1}(1-z) = -\Phi^{-1}(z)$, we find that z and y belong to the set $F(\alpha,\,\beta)$ in (32). □

Proof of Lemma 5: a) Let us define, for A as in (36),

$$c \triangleq z - y,\; d \triangleq cA. \tag{A.10}$$

Then, from the conditions in (34) and (35), we find in a straightforward manner that

$$y = \frac{1-c}{2} \pm \frac{d}{2}\left[\frac{1}{c^2+d^2} - 1\right]^{1/2}$$
$$z = \frac{1+c}{2} \pm \frac{d}{2}\left[\frac{1}{c^2+d^2} - 1\right]^{1/2}, \tag{A.11}$$

subject to

$$c^2 + d^2 \le 1. \tag{A.12}$$

Considering the expressions in (A.10), we can write (A.11) and (A.12) as follows:

$$y = \frac{1-c}{2} \pm \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2}$$
$$z = \frac{1+c}{2} \pm \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2}, \tag{A.13}$$

for $0 < c \le (1+A^2)^{-1/2}$, where then,

$$A\left[\frac{1}{1+A^2} - c^2\right]^{1/2} < 1 - c. \tag{A.14}$$

b) Using the equality $\Phi^{-1}(1-z) = -\Phi^{-1}(z)$, let us consider the following function:

$$\begin{aligned} Y(x) &\triangleq \Phi^{-1}\left(\frac{1+c}{2}+x\right) - \Phi^{-1}\left(\frac{1-c}{2}+x\right) \\ &= \Phi^{-1}\left(\frac{1+c}{2}+x\right) + \Phi^{-1}\left(\frac{1+c}{2}-x\right), \\ &\qquad -\frac{1+c}{2} \le x \le \frac{1-c}{2}. \end{aligned}$$

We observe that $Y(x) = Y(-x)$, $\forall x > 0$. Thus, to minimize $Y(x)$, we should initially consider negative x values only. Given that, in conjunction with the expressions in (A.13), we conclude that in the search for $\inf[\Phi^{-1}(z) - \Phi^{-1}(y)]$, where y and z are as in (A.13), it suffices to consider the latter expressions with the minus sign only. Then, we are searching for

$$\inf_{c\in B}\left\{\Phi^{-1}\left(\frac{1+c}{2} + \frac{A}{2}\left[\frac{1}{1-A^2} - c^2\right]^{1/2}\right) + \Phi^{-1}\left(\frac{1+c}{2} - \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2}\right)\right\},$$

where

$$B \triangleq \left\{c\colon 0 < c \le \frac{1}{\sqrt{1+A^2}}\right\}.$$

Due to the conditions in (A.13) and (A.14), we conclude that

$$\frac{1+c}{2} - \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2} < \frac{1}{2}, \qquad \forall c \in B.$$

Thus,

$$\Phi^{-1}\left(\frac{1+c}{2} - \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2}\right) = \Phi^{-1}\left(\frac{1}{2} - \frac{c}{2} + \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2}\right), \quad \forall c \in B, \tag{A.15}$$

where

$$\frac{1}{2} - \frac{c}{2} + \frac{A}{2}\left[\frac{1}{1+A^2} - c^2\right]^{1/2} > \frac{1}{2}, \qquad \forall c \in B.$$

c) Due to (A.15), the last step in the search for the infimum of the function $\Phi^{-1}(z) - \Phi^{-1}(y)$ corresponds to finding

$$\inf_{c \in B} \left\{ \frac{1}{2} - \frac{c}{2} + \frac{A}{2} \left[\frac{1}{1+A^2} - c^2 \right]^{1/2} \right\}. \tag{A.16}$$

But the argument in (A.16) is a monotonically decreasing function of c, and it is easily concluded that its minimum is attained at $c = (1 + A^2)^{-1/2}$. This conclusion gives the result in the lemma. □

Proof of Lemma 6: Substituting the result of Lemma 5 into (31), and setting $\beta = 1 - \alpha$, we find (A.17) shown at the top of the previous page. For $M \gg [\Phi^{-1}(1-\alpha)]^2$, we have that

$$\tfrac{1}{2}\Phi^{-1}(1-\alpha)\{M + [\Phi^{-1}(1-\alpha)]^2\}^{-1/2} \ll 1,$$

and thus, using Taylor expansion and first order approximation,

$$\begin{aligned} \Phi^{-1}&\left(\frac{1}{2} + \frac{1}{2}\Phi^{-1}(1-\alpha)\{M + [\Phi^{-1}(1-\alpha)]^2\}^{-1/2}\right) \\ &\approx \frac{1}{2\phi(0)}\Phi^{-1}(1-\alpha)\{M + [\Phi^{-1}(1-\alpha)]^2\}^{-1/2} \\ &\approx \frac{1}{2\phi(0)} M^{-1/2}\Phi^{-1}(1-\alpha) \\ &= \sqrt{\frac{\pi}{2M}}\Phi^{-1}(1-\alpha). \end{aligned} \tag{A.18}$$

Substituting (A.18) in (A.17), we obtain the result in the lemma. □

References

[1] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. 20, pp. 405–417, 1974.

[2] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace Electron. Syst.*, vol. 22, pp. 98–101, 1986.

[3] H. Delic, "Robustness and asymptotic performance studies in distributed detection," Ph.D. dissertation, Univ. Virginia, Charlottesville, 1992.

[4] H. Delic and P. Papantoni-Kazakos, "Robust decentralized detection by asymptotically many sensors," *Signal Processing*, vol. 33, pp. 223–233, 1993.

[5] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach.* New York: Academic, 1967.

[6] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Theory*, vol. 35, pp. 995–1000, 1989.

[7] P. J. Huber, *Robust Statistics.* New York: Wiley, 1981.

[8] D. Kazakos and P. Papantoni-Kazakos, *Detection and Estimation.* New York: Computer Science Press, 1990.

[9] C. C. Lee and J. J. Chao, "Optimum local decision space partitioning for distributed detection," *IEEE Trans. Aerospace Electron. Syst.*, vol. 25, pp. 536–544, 1989.

[10] E. L. Lehmann, *Testing Statistical Hypotheses.* New York: Wiley, 1959.

[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw Hill, 1991.

[12] H. V. Poor and J. B. Thomas, "Memoryless discrete-time detection of a constant signal in m-dependent noise," *IEEE Trans. Inform. Theory*, vol. 25, pp. 54–61, 1979.

[13] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerospace Electron. Syst.*, vol. 23, pp. 24–30, 1987.

[14] F. A. Sadjadi, "Hypothesis testing in a distributed environment," *IEEE Trans. Aerospace Electron. Syst.*, vol. 22, pp. 134–137, 1986.

[15] R. Srinivasan, "Distributed radar detection theory," *IEE Proc.-F, Commun., Radar, Signal Processing*, vol. 133, pp. 55–60, 1986.

[16] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerospace Electron. Syst.*, vol. 17, pp. 501–510, 1981.

[17] J. B. Thomas, "Nonparametric detection," *Proc. IEEE*, vol. 58, pp. 623–631, 1970.

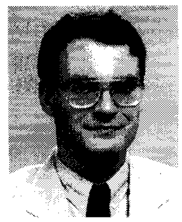
[18] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias, "Optimal decision fusion in multiple sensor systems," *IEEE Trans. Aerospace Electron. Syst.*, vol. 23, pp. 644–653, 1987.

[19] ——, "Optimal distributed decision fusion," *IEEE Trans. Aerospace Electron. Syst.*, vol. 25, pp. 761–765, 1989.

[20] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Automat. Contr.*, vol. 30, pp. 440–446, 1985.

[21] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

[22] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inform. Theory*, vol. 38, pp. 355–361, 1992.



Hakan Delic was born in Bonn, Germany, on October 28, 1966. He received the B.S. degree (with honors) in electrical engineering from Bogazici University, Istanbul, Turkey, in July 1988, and the M.S. and Ph.D. degrees in electrical engineering from the University of Virginia, Charlottesville, in January 1990 and May 1992, respectively.

Between September 1992 and August 1994, he was a Research Associate at the University of Virginia Health Sciences Center, working on neural networks and information theory. In Setember 1994, he joined the Department of Electrical and Computer Engineering of the University of Southwestern Louisiana, Lafayette, where he is presently an Assistant Professor. His research interests include computer communication networks, information theory, detection and estimation theory, distributed information processing, statistical signal processing, neural networks, and robust statistics and its applications.

P. Papantoni-Kazakos (S'73–M'73–SM'79–F'91) was born in Greece on March 28, 1945. She received the B.S. degree in mechanical and electrical engineering from the National Technical University of Athens, Greece, in 1968; and the M.S. degree from Princeton University, Princeton, NJ, in 1970, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1973, both in electrical engineering.

From 1973 to 1978 she was a faculty member at Rice University, Houston, TX. From September 1978 to September 1986 she was with the faculty of the Department of Electrical Engineering at the University of Connecticut, Storrs. In September 1986 she joined the University of Virginia, Charlottesville, where she is currently a Professor of Electrical Engineering and Mathematics. In 1977 she spent one year with Bell Laboratories, on a leave of absence from Rice University. In 1981, she spent one year on special assignment, as a Scientific Officer at the Office of Naval Research, Arlington, VA. In September 1993, she joined the University of Ottawa, Ottawa, Ont., Canada, as an industrial chair holder for high-speed networks. Her research interests include decision theory, distributed processing and neural network structures, statistical communications, information theory, robust statistical and encoding methods, stochastic processes, and computer communications networks. She is coeditor of and contributor to the book: *Nonparametric Methods in Communications* (New York: Marcel Dekker, 1977). She is also coauthor of the book: *Detection and Estimation* (Computer Science Press, 1989).

Dr. Kazakos is a member of the American Mathematical Society, Eta Kappa Nu, Sigma Xi, the Society of Women in Science, and the Society of Women in Engineering. She is also a Fellow of the Institute of Electrical and Electronics Engineers. She has served as the Secretary to the Board of Governors in the IEEE Information Theory Group, she has been editor for Random Access Systems of the IEEE Communications Transactions, and member of the U.S. Army Basic Research Committee of the National Research Council. She is currently a member of the IEEE COMSOC Technical Committee on Computer Communications. She organized and Chaired the 1978 AFOSR Workshop on Communication Theory and Applications.

Demetrios Kazakos (S'69–M'73–SM'88–F'92) was born in Athens, Greece, in 1944. He received the Diploma in electrical, mechanical, and industrial engineering from the Hellenic National Polytechnic University, Greece, in 1967, the M.S. degree from Princeton University, Princeton, NJ, in 1970, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1973, both in electrical engineering, specializing in statistical communication.

From 1973 to 1975 he held postdoctoral positions at Rice University, Houston, TX, and NASA Johnson Space Center, Houston. From September 1975 to August 1980 he was on the faculty of the Department of Electrical Engineering, State University of New York at Buffalo. Since September 1980 he has been at the University of Virginia, where he now holds the position of Professor in the Departments of Electrical Engineering and Mathematics. Since August 1993, he has been the Head of the Department of Electrical and Computer Engineering of the University of Southwestern Louisiana in Lafayette.

Dr. Kazakos is a member of the Institute of Mathematical Statistics, Eta Kappa Nu, and Sigma Xi. He is currently Editor for Random Access and Distributed Communication Systems for IEEE TRANSACTIONS ON COMMUNICATIONS. His research interests include decision theory, statistical communications, information theory, robust statistical methods, stochastic processes, communication networks, and distributed signal processing. He is coeditor and contributor to the book: *Nonparametric Methods in Communications* (New York: Marcel Dekker, 1977). He is also coauthor of the book *Detection and Estimation* (Rockville, MD, Computer Science Press, 1989).