Math. Control Signals Systems (1988) 1: 167–182

Mathematics of Control, Signals, and Systems

Decentralized Detection by a Large Number of Sensors*

John N. Tsitsiklis†

Abstract. We consider the decentralized detection problem, in which N independent, identical sensors transmit a finite-valued function of their observations to a fusion center which then decides which one of M hypotheses is true. For the case where the number of sensors tends to infinity, we show that it is asymptotically optimal to divide the sensors into $M(M-1)/2$ groups, with all sensors in each group using the same decision rule in deciding what to transmit. We also show how the optimal number of sensors in each group may be determined by solving a mathematical programming problem. For the special case of two hypotheses and binary messages the solution simplifies considerably: it is optimal (asymptotically, as $N \to \infty$) to have all sensors perform an identical likelihood ratio test, and the optimal threshold is very easy to determine numerically.

Key words. Decentralized detection, Multiple hypothesis testing, Asymptotic error bounds.

1. Introduction and Problem Definition

The (static) decentralized detection problem is defined as follows. There are M hypotheses $H_1, \ldots, H_M$, with known prior probabilities $P(H_i) > 0$, and there are N sensors. Let Y be a set endowed with a σ-field $\mathcal{F}_Y$ of measurable sets. Let y_i, $i = 1, \ldots, N$, the observation of the ith sensor, be a random variable taking values in Y. We assume that the y_i's are conditionally independent and identically distributed, given either hypothesis, with a known conditional distribution $P(y|H_j)$, $j = 1, \ldots, M$. Let D be a positive integer. Each sensor i evaluates a D-valued message $u_i \in \{1, \ldots, D\}$ as a function of its own observation; that is, $u_i = \gamma_i(y_i)$, where the function γ_i: $Y \mapsto \{1, \ldots, D\}$ is the decision rule of sensor i and is assumed to be a measurable function. The messages $u_1, \ldots, u_N$ are all transmitted to a fusion center which declares one of the hypotheses to be true, based on a decision rule γ_0: $\{1, \ldots, D\}^N \mapsto \{1, \ldots, M\}$. That is, the final decision u_0 of the fusion center is given by $u_0 = \gamma_0(u_1, \ldots, u_N)$. The objective is to choose the decision rules $\gamma_0, \gamma_1, \ldots, \gamma_N$ of the sensors and the fusion center so as to minimize the probability of error in the decision of the fusion center. (An alternative formulation of the problem, of the Neyman–Pearson type, will be considered in the last section).

* Date received: April 21, 1987. Research supported by the Army Research Office under Contract DAAG-29-84-K-0005 and by the Office of Naval Research under Contract N00014-84-K-0519.

† Room 35-214, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U.S.A.

This problem and its variants have been the subject of a fair amount of recent research [TS], [E], [TA], [LS], especially for the case of binary hypotheses ($M = 2$) and binary messages ($D = 2$). For this special case, it is known that any optimal set of decision rules has the following structure. Each one of the sensors evaluates its message u_i using a likelihood ratio test with an appropriate threshold. Then, the fusion center makes its decision by performing a final likelihood ratio test. (Here, the messages received by the fusion center play the role of its observations.) Without something like the conditional independence assumption we have introduced, this result fails to hold and the problem is intractable (NP-hard), even for the case of two sensors [TA]. Assuming conditional independence, the optimal value of the threshold of each sensor may be obtained by finding all solutions of a set of coupled algebraic equations (which are the person-to-person optimality conditions for this problem) and by selecting the solution which results to least cost. Unfortunately (and contrary to intuition), even if the observations of each sensor are identically distributed (given either hypothesis) it is not true that all sensors should use the same threshold (see the Appendix for an example). This renders the computation of the optimal thresholds intractable when the number of sensors is large. To justify this last claim, consider what is involved in just evaluating the probability of error corresponding to a fixed set of decision rules when each sensor uses a different threshold. In order to evaluate the error probability, we have to perform a summation over all possible values of $(u_1, \ldots, u_N)$, which means that there are 2^N terms to be summed. (This is in contrast to the case of equal thresholds in which the u_i's are identically distributed and therefore the binomial formula may be used to obtain a sum with only $N + 1$ summands). Of course, to determine an optimal set of decision rules this effort may have to be repeated a number of times. This suggests that the computational effort grows exponentially with the number N of sensors.

The above discussion motivates the main results of this paper which show that, for the case $M = 2$, $D = 2$, it is asymptotically optimal to have each sensor use the same threshold, also providing a simple method for computing the optimal threshold. For the general case of $M > 2$ hypotheses, it is no longer true, even in the limit as $N \to \infty$, that each sensor should use the same decision rule. Nevertheless, we show that, as $N \to \infty$, at most $M(M-1)/2$ different decision rules need to be used by the sensors. The determination of an asymptotically optimal set of decision rules is still a hard computational problem, except for the case where the observation set Y is finite with small cardinality.

Notation. Throughout, P_i will stand for the (conditional) measure $P(\cdot|H_i)$ on $(Y, \mathcal{F}_Y)$, under hypothesis H_i. Furthermore, $E_i[\cdot]$ will stand for expectation with respect to the measure P_i.

2. The Bayesian Problem

We start by noticing that, having fixed the decision rules $\gamma_1, \ldots, \gamma_N$ of the sensors, the optimal decision for the fusion center is determined by using the maximum *a posteriori* probability (MAP) rule. (The messages to the fusion center are thought of as the measurements available to it.) Thus, γ_0 is straightforward to determine in

terms of $\gamma_1, \ldots, \gamma_N$. For this reason, we shall be concerned only with optimization with respect to $(\gamma_1, \ldots, \gamma_N)$. Any such set of decision rules will be denoted, for convenience, by γ^N.

We introduce some more notation. Let Γ be a set of decision rules from which the decision rules of all sensors are to be selected. In general, we take Γ to be the set of all (measurable) functions from Y into the set $\{1, \ldots, D\}$. However, we may, for some reason, wish to restrict to a smaller class of decision rules, possibly having some simplifying structure. We return to this issue in Section 3. Let Γ^N be the Cartesian product of Γ with itself, N times. For any $\gamma^N \in \Gamma^N$, let $J_N(\gamma^N)$ be the probability of an erroneous final decision by the fusion center (always assuming that the fusion center uses the MAP rule). We are concerned with the minimization of $J_N(\gamma^N)$, over all $\gamma^N \in \Gamma^N$, when N is very large.

It is easy to show that, as the number of sensors grows to infinity, the probability of error goes to zero, for any reasonable set of decision rules, in fact exponentially fast. Consequently, we need a more refined way of comparing different sets of decision rules, as $N \to \infty$. To this effect, for any given value of N and any choice γ^N of decision rules for the N-sensor problem, we consider the exponent of the error probability defined by

$$r_N(\gamma^N) = \frac{\log J_N(\gamma^N)}{N}.$$

Let $R_N = \inf_{\gamma^N \in \Gamma^N} r_N(\gamma^N)$ be the optimal exponent. Let Γ_0^N be the set of all $\gamma^N \in \Gamma^N$ with the property that the set $\{\gamma_1, \ldots, \gamma_N\}$ has at most $M(M-1)/2$ different elements. Let $Q_N = \inf_{\gamma^N \in \Gamma_0^N} r_N(\gamma^N)$ be the optimal exponent, when we restrict to sets of decision rules in Γ_0^N. The following result shows that, asymptotically, optimality is not lost, if we restrict to Γ_0^N.

Theorem 1. *Subject to Assumption 1 below,* $\lim_{N\to\infty}(Q_N - R_N) = 0$.

The rest of this section is devoted to the proof of Theorem 1. We first need to introduce some auxiliary tools.

Let us fix some $\gamma \in \Gamma$. The mapping from the true hypothesis H_i to the decision of a sensor employing the decision rule γ may be thought of as a noisy communication channel which is completely described by the probabilities

$$p_i^\gamma(d) = P_i(\gamma(y) = d).$$

The ability of such a channel to discriminate between hypotheses H_i and H_j $(i \neq j)$ may be quantified by a function $\mu_{ij}(\gamma, s)$, $s \in [0, 1]$, defined by the following formula [SGB]:

$$\mu_{ij}(\gamma, s) = \log\left[\sum_{d=1}^{D} (p_i^\gamma(d))^{1-s}(p_j^\gamma(d))^s\right]. \tag{1}$$

We use the convention $0^0 = 0$; thus, the summation in (1) is to be performed only over those d's for which $p_i^\gamma(d)p_j^\gamma(d) \neq 0$. Assuming that $\mu_{ij}(\gamma, s)$ is not infinite, it is easy to see that $\mu_{ij}(\gamma, s)$ is infinitely differentiable, as a function of s, and its derivatives are continuous on $[0, 1]$, provided that we define the derivative at an endpoint as the limit when we approach the endpoint from the interior.

Notice that, for any fixed γ, the function $\mu_{ij}(\gamma, s)$ is equal to $\log E[e^{sX}]$, where X is the log-likelihood ratio of the distributions associated with $p_j^{\gamma}(\cdot)$ and $p_i^{\gamma}(\cdot)$, and where the expectation is with respect to $p_i^{\gamma}(\cdot)$. As is well known, minimizing the characteristic function of a random variable X yields tight bounds on the probability of large deviations of X from its mean. Since in this case X is the log-likelihood ratio, this method leads to tight bounds on the probability of error. One particular such result that we will use is taken from [SGB]:

Lemma 1. *Let there be two hypotheses H' and H''. Let $x_1, \ldots, x_N$ be measurements taking values in a finite set $\{1, \ldots, D\}$ which are conditionally independent given the true hypothesis, and assume that the conditional distribution of x_i, when H is true, is given by $p_H^i(d) = P(x_i = d|H)$. Let*

$$\mu(i, s) = \log\left[\sum_{d=1}^{D} (p_{H'}^i(d))^{1-s}(p_{H''}^i(d))^s\right]$$

and $\mu(s) = \sum_{i=1}^{N} \mu(i, s)$. Assume that $\mu(i, s)$, $\mu'(i, s)$, $\mu''(i, s)$ exist and are finite, where primes on μ stand for differentiation with respect to s. Let s^* minimize $\mu(s)$, over $s \in [0, 1]$. Then,

(a) *There exists a decision rule for deciding between H' and H'', on the basis of the measurements $x_1, \ldots, x_N$, for which*

$$P(\text{decide } H'|H'' \text{ is true}) + P(\text{decide } H''|H' \text{ is true}) \leq 2\exp\{\mu(s^*)\}.$$

(b) *For any rule for deciding between H' and H'', on the basis of the measurements $x_1, \ldots, x_N$, we have*

$$P(\text{decide } H'|H'' \text{ is true}) + P(\text{decide } H''|H' \text{ is true})$$
$$\geq \tfrac{1}{2}\exp\{\mu(s^*) - [2\mu''(s^*)]^{1/2}\}.$$

Proof. Part (a) of the lemma is the corollary on p. 84 of [SGB]. For part (b), it is shown in [SGB] (equation (3.42), p. 87) that

$$P(\text{decide } H'|H'' \text{ is true}) + P(\text{decide } H''|H' \text{ is true})$$
$$> \tfrac{1}{2}\exp\{\mu(s) - s\mu'(s) - s[2\mu''(s)]^{1/2}\}$$
$$+ \tfrac{1}{2}\exp\{\mu(s) + (1-s)\mu'(s) - (1-s)[2\mu''(s)]^{1/2}\} \qquad \text{for all} \quad s \in (0, 1).$$

If $s^* \in (0, 1)$, we have $\mu'(s^*) = 0$ and the desired result follows immediately. If $s^* = 0$, we may take the limit in the above inequality, as $s \downarrow 0$. Since μ'' is continuous, and therefore bounded, we have $\lim_{s\downarrow 0} s\mu''(s) = 0$, which yields

$$P(\text{decide } H'|H'' \text{ is true}) + P(\text{decide } H''|H' \text{ is true}) \geq \tfrac{1}{2}\exp\{\mu(0)\}$$
$$\geq \exp\{\mu(0) - [2\mu''(0)]^{1/2}\}.$$

The last inequality follows because μ is convex, and therefore $\mu''(s) \geq 0$. The argument for the case $s^* = 1$ is identical. ■

The bounds of parts (a) and (b) of the lemma could be far apart if μ'' is left uncontrolled. For this reason we introduce the following assumption:

Assumption 1. For $i \neq j$, and for all $\gamma \in \Gamma$ and $s \in [0, 1]$:

(a) $|\mu_{ij}(\gamma, s)| < \infty$.
(b) There exists a finite constant A such that $|\mu''_{ij}(\gamma, s)| \leq A$.

The substance of Assumption 1 is explored in Section 5; it is shown there that it corresponds to some minor restrictions on the distribution of the observations which are satisfied in typical situations of practical interest.

As a preview of the remainder of the proof, we use Lemma 1, for each pair of distinct hypotheses, to argue that the decision rules $\gamma_1, \ldots, \gamma_N$ of the sensors should be chosen so as to minimize

$$\max_{\{(i,j):i\neq j\}} \min_{s\in[0,1]} \sum_{k=1}^{N} \mu_{ij}(\gamma_k, s).$$

We reformulate this as a linear programming problem and use linear programming theory to show that a small number of different γ_k's suffices.

Let $\mathcal{F}$ be the set of all finite subsets of Γ. For any $F \in \mathcal{F}$, let

$$\Lambda(F) = \min_{x_\gamma} \max_{\{(i,j):i\neq j\}} \min_{s\in[0,1]} \sum_{\gamma\in F} x_\gamma \mu_{ij}(\gamma, s),$$

where the minimization with respect to x_γ is subject to the constraints

$$x_\gamma \geq 0 \qquad \text{for all} \quad \gamma \in F, \tag{2a}$$

$$\sum_{\gamma\in F} x_\gamma = 1. \tag{2b}$$

Let

$$\Lambda^* = \inf_{F\in\mathcal{F}} \Lambda(F).$$

Let us fix N and some collection $\gamma^N \in \Gamma^N$ of decision rules. Let $\alpha = \min_i P(H_i)$. We then have, using part (b) of Lemma 1,

$$\begin{aligned} J_N(\gamma^N) &= \sum_{\{(i,j):i\neq j\}} P(\text{decide } H_i | H_j) P(H_j) \\ &\geq \frac{\alpha}{2} \max_{\{(i,j):i\neq j\}} \exp\left[\sum_{k=1}^{N} \mu_{ij}(\gamma_k, s_{ij}^*) - \left(2 \sum_{k=1}^{N} \mu''_{ij}(\gamma_k, s_{ij}^*) \right)^{1/2} \right], \end{aligned}$$

where s_{ij}^* minimizes $\sum_{k=1}^{N} \mu_{ij}(\gamma_k, s)$ over $s \in [0, 1]$. Let F be the set of different decision rules (elements of Γ) which are present in the collection γ^N of decision rules. For each $\gamma \in F$, let x_γ be the proportion of the sensors using decision rule γ; that is, x_γ is equal to the number of k's such that $\gamma_k = \gamma$, divided by N. By construction, the coefficients x_γ satisfy the constraints (2a), (2b). Using Assumption 1(b) to bound $\mu''_{ij}(\gamma_k, s_{ij}^*)$, the definition of s_{ij}^*, and the definition of $\Lambda(F)$, we have

$$\begin{aligned} J_N(\gamma^N) &\geq \frac{\alpha}{2} \exp\left(\max_{\{(i,j):i\neq j\}} \min_{s\in[0,1]} \left[N \sum_{\gamma\in F} x_\gamma \mu_{ij}(\gamma, s) \right] - (2NA)^{1/2} \right) \\ &\geq \frac{\alpha}{2} e^{N\Lambda(F) - (2NA)^{1/2}} \geq \frac{\alpha}{2} e^{N\Lambda^* - (2NA)^{1/2}}. \end{aligned}$$

This shows that $R_N \geq \Lambda^* - (2A/N)^{1/2} + (1/N)\log(\alpha/2)$. Taking the limit as $N \to \infty$, we obtain

$$\liminf_{N\to\infty} R_N \geq \Lambda^*. \tag{3}$$

Lemma 2. *$\Lambda^* = \inf_{F\in\mathcal{F}_0} \Lambda(F)$, where $\mathcal{F}_0$ is the collection of all subsets of Γ of cardinality no larger than $M(M-1)/2$.*

Proof. Given some $F \in \mathcal{F}$, let s_{ij}^*, x_γ^* be such that the constraints (2a), (2b) are satisfied and

$$\Lambda(F) = \max_{\{(i,j):i\neq j\}} \sum_{\gamma\in F} x_\gamma^* \mu_{ij}(\gamma, s_{ij}^*).$$

(Such s_{ij}^*, x_γ^* exist because the quantity $\max_{\{(i,j):i\neq j\}} \sum_{\gamma\in F} x_\gamma \mu_{ij}(\gamma, s_{ij})$ is continuous in s_{ij}, x_γ and is defined over a compact set; therefore, the minimum arising in the definition of $\Lambda(F)$ is attained.) In particular, if the s_{ij}^*'s are fixed, then the x_γ^*'s are determined by minimizing $\max_{\{(i,j):i\neq j\}} \sum_{\gamma\in F} x_\gamma \mu_{ij}(\gamma, s_{ij}^*)$, subject to the constraints (2a), (2b). This minimization is equivalent to the following linear programming problem:

$$\begin{aligned}
&\min \lambda \\
&\text{subject to} \\
&\qquad \lambda \geq \sum_{\gamma\in F} x_\gamma \mu_{ij}(\gamma, s_{ij}^*) \qquad \text{for all } i, j, \quad i \neq j, \\
&\qquad x_\gamma \geq 0 \qquad \text{for all} \quad \gamma \in F, \\
&\sum_{\gamma\in F} x_\gamma = 1.
\end{aligned}$$

Let T be the cardinality of the set F. The above-defined linear program has $T+1$ variables and $T + 1 + M(M-1)/2$ constraints. From linear programming theory [PS], we know that there exists an optimal solution at which the number of constraints for which equality holds is no smaller than the number of variables. Therefore, with this optimal solution, at most $M(M-1)/2$ of the constraints hold with a strict inequality, which implies that at most $M(M-1)/2$ of the x_γ's are nonzero. Therefore, for any $F \in \mathcal{F}$, there exists some $F' \in \mathcal{F}_0$ such that $\Lambda(F') \leq \Lambda(F)$. This completes the proof of Lemma 2. ■

Let us fix some N and some $\varepsilon > 0$. Let F be a subset of Γ of cardinality no larger than $M(M-1)/2$ (that is, $F \in \mathcal{F}_0$), such that $\Lambda(F) \leq \Lambda^* + \varepsilon$, which exists, because of Lemma 2. Let x_γ^*, and s_{ij}^* be such that

$$\max_{\{(i,j):i\neq j\}} \sum_{\gamma\in F} x_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) = \Lambda(F) \leq \Lambda^* + \varepsilon.$$

We now define a collection γ^N of decision rules to be used by the N sensors: for each $\gamma \in F$, we let exactly $\lfloor N x_\gamma^* \rfloor$ of them use the decision rule γ; if there are any remaining sensors, which is the case if $N x_\gamma^*$ is not an integer for some γ, we let these sensors use an arbitrary decision rule from the set F. Let N_0 be the number of these remaining sensors.

We now estimate the probability of error under this particular γ^N. The probability of error is bounded above by the probability of error for the case where the fusion center chooses to ignore the messages transmitted by the last N_0 sensors and this is what we will assume. We now have

$$J_N(\gamma^N) \leq \sum_{\{(i,j):i\neq j\}} P(\text{decide } H_i | H_j \text{ is true}) P(H_j)$$

$$\leq M^2 \max_{\{(i,j):i\neq j\}} [P(\text{decide } H_i | H_j \text{ is true}) + P(\text{decide } H_j | H_i \text{ is true})]. \quad (4)$$

The expression inside the brackets on the right-hand side of (4) refers to the probabilities of error for a context in which H_i and H_j are the only hypotheses. Since the fusion center uses the MAP rule, it is using a decision rule which would be optimal even if it had to discriminate only between the two hypotheses H_i and H_j (always assuming that the last N_0 messages are ignored). Thus, for each pair of hypotheses, the upper bound on the probability of error furnished by Lemma 1(a) is applicable. This yields

$$J_N(\gamma^N) \leq 2M^2 \max_{\{(i,j):i\neq j\}} \exp\left[\sum_{\gamma\in\Gamma} \lfloor Nx_\gamma^* \rfloor \mu_{ij}(\gamma, s_{ij}^*)\right]. \quad (5)$$

We now use the inequality $Nx_\gamma^* - \lfloor Nx_\gamma^* \rfloor \leq 1$ to obtain

$$\sum_{\gamma\in\Gamma} \lfloor Nx_\gamma^* \rfloor \mu_{ij}(\gamma, s_{ij}^*) \leq \sum_{\gamma\in\Gamma} Nx_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) + \sum_{\gamma\in\Gamma} |\mu_{ij}(\gamma, s_{ij}^*)| \leq \sum_{\gamma\in\Gamma} Nx_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) + K,$$

where K is a constant independent of N. We substitute the above inequality in the right-hand side of (5), then take logarithms and divide by N to obtain

$$Q_N \leq \frac{\log J_N(\gamma^N)}{N} \leq \frac{2\log M}{N} + \frac{\log K}{N} + \max_{\{(i,j):i\neq j\}} \sum_{\gamma\in\Gamma} x_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) \leq \Lambda^* + \varepsilon + \frac{K'}{N},$$

where K' is another constant independent of N. We take the limit as $N \to \infty$ and use the fact that ε was arbitrary to conclude that $\limsup_{N\to\infty} Q_N \leq \Lambda^*$. We combine this inequality with (3) and the obvious inequality $R_N \leq Q_N$ to complete the proof of Theorem 1. ■

3. Special Cases and Computational Considerations

Let us start by stressing that the proof of Theorem 1 is constructive and suggests a procedure for determining an asymptotically optimal set of decision rules. Namely, we have to solve the optimization problem defining Λ^*. The value of Λ^* is the optimal exponent and the associated optimal values of the x_γ^*'s are the proportions of the sensors who should use each decision rule γ.

Theorem 1 is most useful in the case of binary hypotheses ($M = 2$) and binary messages ($D = 2$). For that case it is known [TS] that, without loss of optimality, we may assume that each sensor decides what to transmit by performing a likelihood ratio test, with an appropriate threshold. We thus let Γ be the set of all such decision rules. Furthermore, in this case we have $M(M-1)/2 = 1$ and Theorem 1 implies that it is asymptotically optimal to let every sensor use the same threshold. In order to compute Λ^* we only need to optimize over all subsets of Γ of cardinality 1.

Therefore, the optimal threshold may be computed by solving the optimization problem

$$\min_{\gamma \in \Gamma} \min_{s \in [0,1]} \mu_{12}(\gamma, s). \tag{6}$$

Notice that each $\gamma \in \Gamma$ can be described by a single real number, the value of the threshold being employed. We are therefore dealing with a nonlinear optimization problem in two dimensions. In typical problems, the probabilities $p_i^{\gamma}(d)$ are given by simple analytical expressions, as a function of the threshold corresponding to γ. Therefore, simple analytical expressions are also available for $\mu_{12}(\gamma, s)$ as well. It is known that $\mu_{12}(\gamma, s)$ is a convex function of s, for every γ [SGB], which makes the optimization with respect to s easier. Unfortunately, we are not aware of any simple but nontrivial examples in which the solution of the above optimization problem and the corresponding value of the optimal threshold may be obtained analytically.

In the case of binary hypotheses ($M = 2$) and messages of arbitrary cardinality $D > 2$, it is known that likelihood ratio tests are again optimal except that each decision rule consists of $D - 1$ thresholds which determine which one of the D messages is to be sent. The same discussion as for the case of $D = 2$ applies here and (asymptotically) each sensor should use the same set of thresholds. The only difference is that γ is parametrized by a $(D - 1)$-dimensional real vector (as opposed to a scalar). Thus, the problem (6), which needs to be solved in order to determine the optimal thresholds, is a D-dimensional optimization problem. This may become quite hard unless D is small, the reason being that, in general, $\mu(\gamma, s)$ is not a convex function of the parameters specifying γ.

For the case where $M > 2$, Theorem 1 is less useful for computing an asymptotically optimal set of decision rules. The reason is that we have to perform an optimization problem over all subsets of Γ of cardinality $M(M - 1)/2$. In principle, it seems possible to reformulate the optimization problem defining Λ^* in a way that avoids having to consider each such subset of Γ (which would be impossible anyway if Γ is infinite). Namely, we might perform the minimization

$$\min_{x \in P} \max_{\{(i,j): i \neq j\}} \min_{s \in [0,1]} \int_{\Gamma} \mu_{ij}(\gamma, s)\, dx(\gamma),$$

where $x(\cdot)$ is a positive measure on Γ with $x(\Gamma) = 1$ and where P is the set of all such measures. Leaving aside the technical difficulties in showing that this is an equivalent problem, it still does not seem particularly promising from a computational point of view. It appears that the only cases in which a numerical solution is possible are those cases in which the set Y is finite and has small cardinality, because in that case Γ is also finite and has small cardinality. Notice that if $F_1 \subset F_2$, then $\Lambda(F_2) \leq \Lambda(F_1)$. Therefore, if Γ is finite, we have $\Lambda^* = \Lambda(\Gamma)$. This suggests that in order to compute Λ^* it is preferable to ignore Theorem 1: instead of computing $\Lambda(F)$ for each F of cardinality $M(M - 1)/2$, and then taking the minimum, we may just compute $\Lambda(\Gamma)$.

Example 1. Let $M = 3$, $D = 2$, and $Y = \{1, 2, 3\}$. Let each hypothesis be equally likely and let the statistics of the observation y be as follows: conditioned on H_i being true, y takes the value i with probability $1 - 2\varepsilon$ and takes each one of the

remaining two values with probability ε $(0 < \varepsilon < 1/4)$. There are three possible decision rules. The ith possible decision rule is: $\gamma_i(y) = 1$ if and only if $y = i$. Notice that γ_1 does not provide any information useful in dsicriminating between H_2 and H_3. Thus, $\mu_{23}(\gamma_1, s) = 0$ for all s; similarly, $\mu_{12}(\gamma_3, s) = \mu_{13}(\gamma_2, s) = 0$ for all s. Furthermore, by symmetry, $\mu_{12}(\gamma_1, s) = \mu_{13}(\gamma_1, s) = \mu_{23}(\gamma_2, s)$, etc. Let α be the value of the minimum of $\mu_{12}(\gamma_1, s)$, over $s \in [0, 1]$. Let x_i be the proportion of sensors using γ_i. The optimal values of x_1, x_2, x_3 are determined by solving the problem

$$\alpha \max_{x_1, x_2, x_3} \min\{x_1 + x_2, x_1 + x_3, x_2 + x_3\},$$

over the unit simplex. It is easy to see that the optimal solution is $x_1 = x_2 = x_3 = \frac{1}{3}$, exactly as expected from the symmetry of the problem, and the corresponding value of the optimal exponent Λ^* is $2\alpha/3$.

4. Alternative Interpretations

Theorem 1 may be restated in a different language referring to a different context. For simplicity, we only consider the case $M = 2$. Suppose that we want to transmit a binary message and that we have a collection of noisy, memoryless, and independent channels at our disposal. We are allowed to transmit a total of N times using any of the available channels each time. A receiver observes the N outputs of the channels, uses its knowledge of which channels were being used, and makes a decision on what was transmitted. The problem consists of finding which channels should be used and how many times each, in order to maximize the probability of correct decoding. For small N, it may be better to use a different channel each time, even if the original message is binary. However, our result states that, for binary messages, as $N \to \infty$, there is a single best channel which should be used for all transmissions. To see the analogy, think of the hypothesis H_1 or H_2 as the value of the binary message which we want to transmit and think of u_i as the output of the ith transmission. A different channel corresponds to a different decision rule and the characteristics of the channel correspond to the quantities $p_i^j(d)$.

A different analogy may be made in the context of optimal design of measurements for failure detection. Suppose that we have a system which may be in one of two states: up or down. We have a collection of devices which may be used for failure detection. They are, however, unreliable and may make errors of both types. Furthermore, the probabilities of either type of error can be different for different devices. Suppose that, in order to increase reliability, we want to use N such devices. Then, our result states that, as $N \to \infty$, there exists a single best device and that we should use N replicas of it, rather than using many devices with different characteristics.

5. The Content of Assumption 1

In this section we explore Assumption 1. Our objective here is to obtain conditions on the distributions P_i under which Assumption 1 can be shown to hold. Proposition 1 below deals with Assumption 1(a).

Proposition 1. *Assumption* 1(a) *fails to hold if and only if there are two hypotheses* H_i, H_j, *such that the corresponding measures* P_i *and* P_j *are mutually singular.*[1]

Proof. Suppose that Assumption 1(a) fails. Then there exist some i, j and some $\gamma \in \Gamma$ for which $p_i^\gamma(d)p_j^\gamma(d) = 0$ for all $d \in \{1, \ldots, D\}$. Thus, for any $d \in \{1, \ldots, D\}$, the set $\{y \in Y: \gamma(y) = d\}$ has nonzero measure under P_i only if it has zero measure under P_j. Since the sets $\{y \in Y: \gamma(y) = d\}$ cover the entire set Y, it follows that P_i and P_j are mutually singular. ■

As a consequence of Proposition 1, we can see that if there are only two hypotheses and Assumption 1(a) fails to hold we are dealing with the uninteresting situation where each sensor is able to determine the true hypothesis on its own, with zero probability of error. For the case of more than two hypotheses, however, there are nontrivial detection problems in which Assumption 1(a) fails to hold. We conjecture that a somewhat modified version of Theorem 1, covering such a case, is possible. We now explore Assumption 1(b) and show that it holds for two interesting situations.

Proposition 2. *Suppose that the observation set* Y *is finite and that Assumption* 1(a) *holds. Then Assumption* 1(b) *also holds.*

Proof. The derivatives of $\mu_{ij}(\gamma, s)$, with respect to s are easily calculated to be [SGB, equations (3.24)–(3.25)]:

$$\mu'_{ij}(\gamma, s) = \sum_{d=1}^{D} \frac{(p_i^\gamma(d))^{1-s}(p_j^\gamma(d))^s}{\sum_{c=1}^{D}(p_i^\gamma(c))^{1-s}(p_j^\gamma(c))^s} \log \frac{p_j^\gamma(d)}{p_i^\gamma(d)}, \tag{7}$$

$$\mu''_{ij}(\gamma, s) = \left[\sum_{d=1}^{D} \frac{(p_i^\gamma(d))^{1-s}(p_j^\gamma(d))^s}{\sum_{c=1}^{D}(p_i^\gamma(c))^{1-s}(p_j^\gamma(c))^s}\left(\log \frac{p_j^\gamma(d)}{p_i^\gamma(d)}\right)^2\right] - [\mu'_{ij}(\gamma, s)]^2, \tag{8}$$

where all summations are over those c's and d's for which $p_i^\gamma(c)p_j^\gamma(c)$ (respectively, $p_i^\gamma(d)p_j^\gamma(d)$) is nonzero.

Let α be the minimum of $p_i^\gamma(c)$, where the minimum is taken over all choices of γ, c, i, such that $p_i^\gamma(c) > 0$. Since Y is finite, the set of all possible decision rules γ is also finite and therefore α is the minimum of finitely many positive quantities and is itself positive. By Assumption 1(a) the denominator in equation (7) must have a nonzero summand and this summand will be bounded below by $\alpha^{1-s}\alpha^s = \alpha$. The numerator is bounded by D. Concerning the logarithmic term, it is bounded, in absolute value, by $|\log \alpha|$, for any d in the range of the summation. We conclude that $\mu'_{ij}(\gamma, s)$ is bounded in absolute value by a constant independent of i, j, γ, s. A similar argument applies to $\mu''(\gamma, s)$ and concludes the proof. ■

Proposition 3. *Suppose that, for any* i, j, *the measure* P_i *is absolutely continuous with respect to* P_j *and let* L_{ij} *denote the Radon–Nikodym derivative* dP_i/dP_j. *Assume*

[1] Two positive measures P_1, P_2, defined on a common (measurable) space Y are called mutually singular if there exists a measurable subset U of Y such that $P_1(U) = P_2(Y - U) = 0$.

that

$$E_i[\log^2 L_{ij}] < \infty \quad \textit{for all } i, j. \tag{9}$$

Then Assumption 1 *holds.*

Proof. The fact that Assumption 1(a) holds is immediate from our assumption of absolute continuity and Proposition 1.

For any decision rule γ: $Y \mapsto \{1, \ldots, D\}$, let $\mathcal{F}^\gamma$ be the smallest σ-field contained in $\mathcal{F}_Y$ with respect to which the function γ is measurable. Let P_i^γ denote the restriction of the measure P_i on the σ-field $\mathcal{F}^\gamma$. It follows from the absolute continuity assumption that P_i^γ is absolutely continuous with respect to P_j^γ. We define L_{ij}^γ to be equal to the Radon–Nikodym derivative dP_i^γ/dP_j^γ. As is well known

$$L_{ij}^\gamma = E_j[L_{ij}|\mathcal{F}^\gamma] \quad \text{a.s. } (P_j). \tag{10}$$

Consider the function ϕ: $(0, \infty) \mapsto (0, \infty)$ defined by $\phi(t) = t \log^2 t$. An easy calculation shows that it is convex. Therefore, using (10) and Jensen's inequality,

$$\begin{aligned} E_i[\log^2 L_{ij}^\gamma] &= E_j[L_{ij}^\gamma \log^2 L_{ij}^\gamma] = E_j[\phi(L_{ij}^\gamma)] = E_j[\phi(E_j[L_{ij}|\mathcal{F}^\gamma])] \\ &\le E_j[E_j[\phi(L_{ij})|\mathcal{F}^\gamma]] = E_j[L_{ij} \log^2 L_{ij}] = E_i[\log^2 L_{ij}]. \end{aligned}$$

Using (9), we conclude that there exists a constant $B < \infty$ such that $E_i[\log^2 L_{ij}^\gamma] \le B$ for all γ, i, j; using the inequality $E[|x|] \le 1 + E[x^2]$, we obtain the same conclusion for $E_i[\log L_{ij}^\gamma]$.

Notice now that $L_{ij}^\gamma(y) = p_i^\gamma(d)/p_j^\gamma(d)$, for every y such that $\gamma(y) = d$, almost surely. Using this observation, equation (7) may be rewritten as

$$\mu_{ij}'(\gamma, s) = \frac{E_i[(L_{ji}^\gamma)^s \log L_{ji}^\gamma]}{E_i[(L_{ji}^\gamma)^s]}, \tag{11}$$

similarly, equation (8) becomes

$$\mu_{ij}''(\gamma, s) = \frac{E_i[(L_{ij}^\gamma)^s \log^2 L_{ji}^\gamma]}{E_i[(L_{ji}^\gamma)^s]} - [\mu_{ij}'(\gamma, s)]^2. \tag{12}$$

Using the obvious inequality $(L_{ji}^\gamma)^s \le (1 + L_{ji}^\gamma)$, for all $s \in [0, 1]$, we obtain the bound

$$|\mu_{ij}'(\gamma, s)| \le \frac{|E_i[\log L_{ji}^\gamma]| + |E_i[L_{ji}^\gamma \log L_{ji}^\gamma]|}{E_i[(L_{ji}^\gamma)^s]} = \frac{|E_i[\log L_{ji}^\gamma]| + |E_j[\log L_{ji}^\gamma]|}{E_i[(L_{ji}^\gamma)^s]}.$$

We have already proved that the numerator is bounded. We now establish a lower bound on $E_i[(L_{ji}^\gamma)^s]$. Since $E_i[L_{ji}] = 1$, it follows that there exists an $\mathcal{F}_Y$-measurable set $Y_0 \subset Y$ and some $\varepsilon > 0$, $\delta > 0$, such that $P_i(Y_0) \ge \varepsilon$ and $L_{ji}(y) \ge \delta$ for all $y \in Y_0$. Since $x^s \ge \min\{1, x\}$, we obtain $E_i[L_{ji}^s] \ge \varepsilon \min\{1, \delta\}$ for all $s \in [0, 1]$. We now use the fact that the function $\phi(x) = x^s$ is concave, for any fixed $s \in [0, 1]$, and Jensen's inequality to obtain

$$E_i[(L_{ji}^\gamma)^s] = E_i[(E_i[L_{ji}|\mathcal{F}^\gamma])^s] \ge E_i[E_i[L_{ji}^s|\mathcal{F}^\gamma]] = E_i[L_{ji}^s] \ge \varepsilon \min\{1, \delta\}.$$

This concludes the proof that $\mu'(\gamma, s)$ is bounded. The proof of the boundedness of $\mu''(\gamma, s)$ is identical and is omitted. ■

6. The Neyman–Pearson Problem

In this section we consider the Neyman–Pearson version of the problem studied in the preceding sections. We are given an observation set Y, endowed with a σ-field $\mathcal{F}_Y$. There are two hypotheses ($M = 2$) and for each hypothesis we are given a measure P_i on $(Y, \mathcal{F})$, $i = 1, 2$. Let D be a fixed positive integer and let Γ be the set of all measurable functions γ: $Y \mapsto \{1, \ldots, D\}$. As before, the ith sensor makes an independent observation y_i whose statistics are described by P_j, assuming that hypothesis H_j is true. Again, the ith sensor transmits a message $\gamma_i(y_i)$ to a fusion center, where $\gamma_i \in \Gamma$, and finally the fusion center makes a final decision using a decision rule γ_0. We allow γ_0 to be randomized. That is, the final decision of the fusion center may depend on the messages it has received as well as an internally generated random variable. Let Γ_0 be the set of all candidate decision rules γ_0 for the fusion center.

For any given $(\gamma_0, \gamma_1, \ldots, \gamma_N) \in \Gamma_0 \times \Gamma^N$, consider the probabilities of error defined by

$$J_N^1(\gamma_0, \gamma_1, \ldots, \gamma_N) = P_1(\gamma_0(\gamma(y_1), \ldots, \gamma(y_N)) = 2), \tag{13}$$

$$J_N^2(\gamma_0, \gamma_1, \ldots, \gamma_N) = P_2(\gamma_0(\gamma(y_1), \ldots, \gamma(y_N)) = 1), \tag{14}$$

Let us fix a constant β belonging to $(0, 1)$. We would like to minimize $J_N^1(\gamma_0, \ldots, \gamma_N)$, over all $\gamma_0, \ldots, \gamma_N$ satisfying

$$J_N^2(\gamma_0, \gamma_1, \ldots, \gamma_N) \leq 1 - \beta. \tag{15}$$

The optimal value of J_N^1 decreases exponentially with N and we define

$$r_N(\gamma_0, \ldots, \gamma_N) = \frac{1}{N} \log J_N^1(\gamma_0, \ldots, \gamma_N).$$

Let

$$R_N = \inf r_N(\gamma_0, \ldots, \gamma_N), \tag{16}$$

where the infimum is taken over all $(\gamma_0, \ldots, \gamma_N) \in \Gamma_0 \times \Gamma^N$ satisfying (15). We will use the following assumption:

Assumption 2.

(a) P_2 is absolutely continuous with respect to P_1.

(b)
$$E_2\left[\log^2\left(\frac{dP_2}{dP_1}\right)\right] = A < \infty, \tag{17}$$

where dP_2/dP_1 is the Radon–Nikodym derivative of the two measures.

We define $\mathcal{F}^\gamma$ and P_i^γ as in Section 5: $\mathcal{F}^\gamma$ is the σ-field on Y generated by γ and P_i^γ is the measure P_i restricted to $\mathcal{F}^\gamma$. The argument in the proof of Proposition 3, in Section 5, applies here and shows that $E_2[\log^2(dP_2^\gamma/dP_1^\gamma)] \leq A$ for all $\gamma \in \Gamma$.

The latter inequality also implies that there exists some $B < \infty$ such that

$$K(\gamma) = E_2\left[\log \frac{dP_2^\gamma}{dP_1^\gamma}\right] \leq B \qquad \text{for all} \quad \gamma \in \Gamma. \tag{18}$$

The quantity $K(\gamma)$ defined by equation (18) may be recognized as the Kullback–Liebler [KL] information distance between the distributions of the random variable $\gamma(y)$ under the two alternative hypotheses. It is guaranteed to be nonnegative. Furthermore, Stein's lemma [B] states that $K(\gamma)$ is the asymptotic error exponent if all sensors are using the same decision rule γ and if the fusion center chooses γ_0, according to the Neyman–Pearson lemma. In light of this fact, the following result should be expected.

Theorem 2. *If Assumption 2 holds, then:*

(i) $\lim_{N\to\infty} R_N = -\sup_{\gamma\in\Gamma} K(\gamma)$.

(ii) *The value of R_N stays the same if in the definition* (16) *we impose the additional constraint $\gamma_1 = \cdots = \gamma_N$.*

Proof. (Outline) Fix some $\varepsilon > 0$ and let $\gamma^* \in \Gamma$ be such that $K(\gamma^*) \geq \sup_{\gamma\in\Gamma} K(\gamma) - \varepsilon$. Let the fusion center choose γ_0 optimally, subject to (15). From Stein's lemma we obtain $\lim_{N\to\infty} r_N(\gamma_0, \gamma^*, \ldots, \gamma^*) = -K(\gamma^*)$. In particular, $\limsup_{N\to\infty} R_N \leq -K(\gamma^*) \leq -\sup_{\gamma\in\Gamma} K(\gamma) + \varepsilon$. Since ε was arbitrary, we conclude that $\limsup_{N\to\infty} R_N \leq -\sup_{\gamma\in\Gamma} K(\gamma)$, and we have shown this bound to be valid under the additional constraint $\gamma_1 = \cdots = \gamma_N$.

In order to complete the proof, it is sufficient to show that for any $\gamma_0, \ldots, \gamma_N$ satisfying (15) we have

$$r_N(\gamma_0, \ldots, \gamma_N) \geq -\frac{1}{N}\sum_{i=1}^{N} K(\gamma_i) + f(N) \geq -\sup_{\gamma\in\Gamma} K(\gamma) + f(N), \tag{19}$$

where f is a function with the property $\lim_{N\to\infty} f(N) = 0$ and which does not depend on $\gamma_0, \ldots, \gamma_N$. While this result does not follow from the usual formulation of Stein's lemma (which uses the assumption $\gamma_1 = \cdots = \gamma_N$), it may be proved by a small variation of the proof of that lemma, and for this reason the proof will be omitted. Suffice it to say that we may follow the proof of Stein's lemma given in [B]. Wherever in that proof convergence in probability of a log-likelihood ratio to its mean is asserted, we replace such a statement with an inequality which bounds the probability of a deviation of a log-likelihood ratio from its mean. Such an inequality is obtained from Chebychev's inequality. Because of (17) the variance of the log-likelihoods of interest admits the same bound, regardless of the choice of the γ_i's. For this reason, the function f in (19) may be taken independent of the γ's. The proof is then completed by taking the infimum of both sides of (19) over all $\gamma_0, \ldots, \gamma_N$ and then letting N tend to infinity. ■

We continue with a few observations. For simplicity we restrict our discussion to the case of binary messages ($D = 2$).

It is easy to prove that there is no loss of optimality if we constrain the γ_i's to correspond to likelihood ratio tests [HV]. If we are only interested in asymptotics, the same conclusion may be obtained from Theorem 2: it is not hard to show that if a decision rule does not have the form of a likelihood ratio test, then another decision rule can be found for which $K(\gamma)$ is even larger. This leads to the conclusion that asymptotically optimality is not lost by assuming that each γ_i consists of a comparison of the likelihood ratio computed by that sensor with a threshold.

As is well known, randomization is generally required in optimal hypothesis testing, under the Neyman–Pearson formulation. For this reason we allowed the decision rule of the fusion center to employ an internally generated random variable. We may ask whether anything can be gained by allowing the sensors as well to use randomized decision rules. The answer is generally positive. For example, if $N = 1$, then the best strategy is to let the single sensor perform an optimal Neyman–Pearson test (for which randomization is needed) and have the fusion center adopt the decision of the sensor. Interestingly enough, however, randomization does not help asymptotically as $N \to \infty$, which we now prove. For any two measures P, Q on $(Y, \mathcal{F}_Y)$, let $K(Q, P) = E[\log(dQ/dP)]$, where the expectation is with respect to Q. With this notation, $K(\gamma) = K(P_2^\gamma, P_1^\gamma)$ for all $\gamma \in \Gamma$. It is known, and easy to show, that $K(Q, P)$ is a convex function of (Q, P). Suppose now that a sensor uses a decision rule which involves randomization. The pair (P_2^γ, P_1^γ) of the probability distribution of the message transmitted by a sensor using a randomized decision rule γ lies in the convex hull of such pairs of probability distributions corresponding to non-randomized decision rules. Using the convexity of K, it follows that randomization cannot help in increasing the supremum of $K(\gamma)$ and, therefore, does not help asymptotically.

From a computational point of view, the problem of this section is a little easier than the problem of Section 2, the reason being that we do not have the additional free parameter s of Section 2. In particular, with decision rules parametrized by a scalar threshold, maximization of $K(\gamma)$ is equivalent to a one-dimensional optimization problem. As there may be multiple local optima, some form of exhaustive search may be required.

As an illustration we study the performance of a naive selection of the decision rule γ of each sensor. We let each sensor perform a maximum likelihood test and transmit its decision to the fusion center. This is certainly a bad idea if $N = 1$ because in that case the sensor should perform a Neyman–Pearson test which is, generally, different from a maximum likelihood test. Still, one may wonder whether such a naive prescription has any performance guarantees, as $N \to \infty$. The answer is negative, as the following example shows. Let P_1 and P_2 be as in Fig. 1. A decision rule γ corresponding to a maximum likelihood test is to let $\gamma(y) = 1$ if and only if $y > \frac{1}{2}$. For this choice of γ, if we assume that ε is small enough and use a Taylor series expansion, we obtain

$$K(\gamma) = \tfrac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{2}-\varepsilon}\right) + \tfrac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{2}+\varepsilon}\right) \le A\varepsilon^2,$$

where A is some positive constant. Let us now consider the decision rule γ given by $\gamma(y) = 1$ if and only if $y > 1$. We then have $K(\gamma) = \log(1/(1 - \varepsilon/2)) \ge \varepsilon/2 + B\varepsilon^2$ for

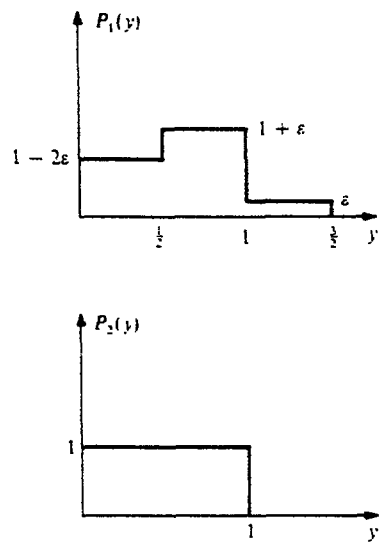

Fig. 1

some constant B. We conclude from this example that the naive decision rule suggested above can be far from optimal (in terms of error exponent) by an arbitrary multiplicative factor.

Acknowledgments. The author is grateful to Professor Robert Gallager who suggested that the results of [SGB] could be used in proving Theorem 1.

Appendix

We consider here the problem introduced in Section 2, with two hypotheses ($M = 2$), binary messages ($D = 2$), two sensors ($N = 2$), and with y_1, y_2 identically distributed and conditionally independent given either hypothesis. We present an example which shows that it is possible that different sensors may have to use different decision rules even if their observations are identically distributed. An example of this type was presented in [TS]. However, that example used a special cost function which introduced a large penalty if both sensors send the same message and the wrong decision is made by the fusion center. Naturally, this creates an incentive for the sensors to try to transmit different messages, and therefore use different decision rules. Thus, the asymmetry of the optimal decision rules of the two sensors can be ascribed to this particular aspect of the cost function and does not prove that asymmetrical decision rules may be optimal for our cost function (probability of error).

Our example is the following. We let H_1 and H_2 be equally likely. The observations y_1, y_2 are conditionally independent, given either hypothesis, take values in

$\{1, 2, 3\}$, and have the following common distribution:

$$P(y = 1|H_1) = \tfrac{4}{5}, \qquad P(y = 2|H_1) = \tfrac{1}{5}, \qquad P(y = 3|H_1) = 0,$$
$$P(y = 1|H_2) = \tfrac{1}{3}, \qquad P(y = 2|H_2) = \tfrac{1}{3}, \qquad P(y = 3|H_2) = \tfrac{1}{3}.$$

An optimal set of decision rules may be found by exhaustive enumeration. Since each sensor has to perform a likelihood ratio test, there are only two candidate decision rules for each sensor:

(A) $u_i = 1$ if and only if $y_i = 1$.
(B) $u_i = 1$ if and only if $y_i \in \{1, 2\}$.

Thus, we need to consider three possibilities:

(i) Both sensors use (A).
(ii) Both sensors use (B).
(iii) Sensor 1 uses (A) and sensor 2 uses (B).

Naturally, we assume that the fusion center is using the maximum *a posteriori* probability rule.

Explicit evaluation of the expected cost for each possibility shows that the optimal set of decision rules consists of one sensor using decision rule A, one sensor using decision rule B, and the fusion center deciding H_1 if and only if $u_1 = u_2 = 1$, for an expected cost of 19/90.

References

[B] R. R. Bahadur, *Some Limit Theorems in Statistics*, SIAM, Philadelphia, 1971.

[E] L. T. Ekchian, Distributed Detection and Communication Problems, Ph.D. Dissertation, Department of EECS, Massachusetts Institute of Technology, Cambridge, MA, 1982.

[HV] I. Y. Hoballah and P. K. Varshney, Decentralized Neyman–Pearson detection, *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, (1986), pp. 237–241.

[KL] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. of Math. Statist.* **22** (1951), 279–280.

[LS] G. S. Lauer and N. R. Sandell, Jr., Decentralized Detection Given Waveform Observations, Technical Paper TP-131, Alphatech Inc., Burlington, MA, 1982.

[PS] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[R] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1974.

[SGB] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channels, I, *Inform. and Control*, **10** (1967), 65–103.

[TS] R. R. Tenney and N. R. Sandell, Jr., Detection with distributed sensors, *IEEE Trans. Aerospace Electron. Systems*, **17** (1981), 501–510.

[TA] J. N. Tsitsiklis and M. Athans, On the complexity of decentralized decision making and detection problems, *IEEE Trans. Automat. Control*, **30** (1985), 440–446.

CORRECTION

Decentralized Detection with a Large Number of Sensors

MCSS, Vol. 1, 1988, pp. 167–182

John N. Tsitsiklis

Example 1 in page 174 of the paper is incorrect. This is because calculations were carried out as if $\mu_{ij}(\gamma, s)$ was the same for all i, j, γ for which $\mu_{ij}(\gamma, s)$ is not identically equal to zero. However, it is seen from the definition of $\mu_{ij}(\gamma, s)$ [cf. Eq. (1)] that $\mu_{ij}(\gamma, s) = \mu_{ji}(\gamma, 1-s)$, which is different than $\mu_{ji}(\gamma, s)$, in general.

The correct solution to the problem considered in Example 1 proceeds as follows. We have $\mu_{23}(\gamma_1, s) = \mu_{13}(\gamma_2, s) = \mu_{12}(\gamma_3, s) = 0$. Also, there exists a function $\nu(s)$ such that $\mu_{ij}(\gamma_i, s) = \nu(s)$ and $\mu_{ij}(\gamma_j, s) = \nu(1-s)$, for every i, j, and s. Thus, the optimization problem that has to be solved in order to compute the optimal exponent Λ^* is

$$\begin{aligned}
\Lambda^* = \min_{x_1,x_2,x_3} \max\Big\{ & \min_{s\in[0,1]} \big[x_1\mu_{12}(\gamma_1,s) + x_2\mu_{12}(\gamma_2,s) + x_3\mu_{12}(\gamma_3,s)\big], \\
& \min_{s\in[0,1]} \big[x_1\mu_{13}(\gamma_1,s) + x_2\mu_{13}(\gamma_2,s) + x_3\mu_{13}(\gamma_3,s)\big], \\
& \min_{s\in[0,1]} \big[x_1\mu_{23}(\gamma_1,s) + x_2\mu_{23}(\gamma_2,s) + x_3\mu_{23}(\gamma_3,s)\big] \Big\} \\
= \min_{x_1,x_2,x_3} \max\Big\{ & \min_{s\in[0,1]} \big[x_1\nu(s) + x_2\nu(1-s)\big], \\
& \min_{s\in[0,1]} \big[x_1\nu(s) + x_3\nu(1-s)\big], \\
& \min_{s\in[0,1]} \big[x_2\nu(s) + x_3\nu(1-s)\big] \Big\}.
\end{aligned}$$

(The outer minimization is over all nonnegative x_1, x_2, x_3 which sum to 1.) Unfortunately, symmetry considerations alone are not sufficient to ascertain that the symmetrical solution ($x_i = 1/3$ for each i) is the optimal one, as we now indicate. The exponent corresponding to the symmetrical solution is seen to be $\frac{1}{3}\min_{s\in[0,1]}[\nu(s)+\nu(1-s)] = \frac{2}{3}\nu(\frac{1}{2})$. On the other hand, the nonsymmetric solution $x_1 = x_2 = \frac{1}{2}$, $x_3 = 0$, results to an exponent equal to

$$\max\Big\{\nu(\frac{1}{2}), \frac{1}{2}\min_{s\in[0,1]}\nu(s)\Big\}.$$

In particular, if $\frac{1}{2}\min_{s\in[0,1]}\nu(s) < \frac{2}{3}\nu(\frac{1}{2})$, then the symmetric solution is not optimal. An analytical method for determining whether this is the case is not apparent.

17