June 1986

LIDS-P-1570

ON THRESHOLD RULES IN DECENTRALIZED DETECTION[1]

John N. Tsitsiklis[2]

ABSTRACT

We consider a decentralized detection problem in which a number of identical sensors transmit a binary function of their observations to a fusion center which then decides which one of two alternative hypotheses is true. We show that, when the number of sensors grows to infinity, optimality is not lost (in terms of the probability of error) if we constrain the sensors to use the same decision rule in deciding what to transmit. This results in considerable simplification of the problem. We also discuss the case where the messages may take more than two values and the case of M–ary ($M > 2$) hypotheses. Next we consider two variants of a decentralized sequential detection problem. For one variant we show that each sensor should decide what to transmit based on a likelihood ratio test; for the other, we demonstrate that such a result fails to hold and that more complicated decision rules are required.

1. 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 (NR 649–003).

2. Room 35–214, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

I. STATIC DECENTRALIZED DETECTION WITH MANY SENSORS.

The static decentralized detection problem is defined as follows. There are two hypotheses H_0 and H_1, with given prior probabilities and N sensors. Let y_i, $i = 1, ..., N$, the observation of the i-th sensor, be a random variable taking values in a set Y. We assume that the y_i's are conditionally independent, given either hypothesis, with a known conditional distribution $P(y_i|H_j)$, $j = 1, 2$. Each sensor i evaluates a binary message $u_i \in \{0, 1\}$ as a function of its own observation; that is $u_i = \gamma_i(y_i)$, where the function $\gamma_i : Y \mapsto \{0, 1\}$ is the decision rule of sensor i. The messages $u_1, ..., u_N$ are all transmitted to a fusion center which declares hypothesis H_0 or H_1 to be true, based on a decision rule $\gamma_0 : \{0, 1\}^N \mapsto \{0, 1\}$. That is, the final decision u_0 of the fusion center is given by $u_0 = \gamma_0(u_1, ..., u_N)$. To any set of decision rules $\Gamma_N = \{\gamma_0, \gamma_1, ..., \gamma_N\}$ we associate a cost $J_N(\Gamma_N)$ which is equal to the probability that the fusion center declares true the wrong hypothesis. The problem consists of finding a set of decision rules Γ_N which minimzes the cost $J(\Gamma_N)$.

The above defined problem and its variants have been the subject of a fair amount of recent research [TeSa, Ek, TsAt, LaSa]. 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 t_i. Then, the fusion center makes its decision by performing a final likelihood ratio test. (Here, the messages received by the center play the role of its observations [TeSa]). Without the conditional independence assumption we introduced, this result fails to hold and the problem is intractable, even for the case of two sensors [TsAt].

Concerning the appropriate value of the threshold t_i of each sensor, it 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 Appendix A 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 cost associated to a fixed set Γ_N of decision rules if each sensor uses a different threshold. In order to evaluate the expected cost, we have to perform a summation over all possible values of $(u_1, ..., u_N)$, 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 strategy, 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 result of this section. From now on we assume that the y_i's are identically distributed (given either hypothesis) and we show that, as the number N of sensors grows to infinity, it is asymptotically optimal to have each sensor use the same threshold and that there is a very simple method for computing the optimal value of that threshold.

Notice that as the number of sensors grows to infinity, the probability of error goes to zero under any reasonable strategy, in fact exponentially fast. Consequently, we need a more refined way of comparing different strategies, as $N \to \infty$. To this effect, for any given value of N and any set 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 $R_N = \inf_{\Gamma_N} r_N(\Gamma_N)$ be the optimal exponent.

To any threshold $t \in [0, \infty]$, we associate a decision rule $\gamma^t : Y \mapsto \{0, 1\}$, defined by $u = \gamma^t(y) = 0$, if $P(y|H_1)/P(y|H_0) < t$, and $u = \gamma^t(y) = 1$, otherwise. We define $f_i^t(j) = P(\gamma^t(y) = j \,|\, H_i)$, $i, j = 0, 1$, which describes the statistics of the message to be sent by any sensor using the decision rule γ^t. We define, for $s \in [0, 1]$,

$$\mu(t, s) = \log \left[\sum_{j=0}^{1} f_0^t(j)^{1-s} f_1^t(j)^s \right]. \tag{1}$$

We use here the convention $0^0 = 0$; equivalently, the summation in (1) has to be performed only over those j's for which $f_0^t(j) f_1^t(j) \neq 0$.

Assumption 1: a) $|\mu(t, s)| < \infty$, $\forall t, s$.

b) There exists a constant A such that $|\mu''(t, s)| \leq A$, $\forall (t, s) \in [0, \infty] \times [0, 1]$, where a prime stands for differentiation with respect to s.

Part (a) of the assumption requires that for any t, either $f_0^t(0) f_1^t(0) \neq 0$, or $f_0^t(1) f_1^t(1) \neq 0$. If this fails to hold, then observation y_i of a single sensor determines which hypothesis is true, with zero error probability, and the problem is vacuous. Part (b) of the Assumption is explored in Appendix B where it shown that it corresponds to some minor restrictions on the distribution of the y_i's which are satisfied in many situations of practical interest.

The function μ has been used in [ShGaBe] in order to obtain the error exponent in certain coding problems. As it turns out, it may be used to obtain the optimal error exponent for our problem as well.

Let us define s^* and t^* by $\mu(t^*, s^*) \leq \mu(t, s)$, $\forall (t, s) \in [0, \infty] \times [0, 1]$, assuming that such s^*, t^* exist. Given a number N of sensors, let Γ_N^* be the set of decision rules obtained by having each sensor use the threshold t^* and the fusion center use the optimal likelihood ratio test, given its data $u_1, ..., u_N$. Let $r_N^* = r_N(\Gamma_N^*)$.

Theorem 1: Under the above assumptions, $\lim_{N\to\infty}(r_N^* - R_N) = 0$.

Proof: The upper bound in [ShGaBe, eqn. (3.7)] shows that $\limsup_{N\to\infty} r_N^* \leq \mu(t^*, s^*)$. (In fact this lower bound was proved in [ShGaBe] under the assumption that the decoder – fusion center – used the maximum likelihood rule. In our case where the fusion center is assumed to perform an optimal likelihood ratio test, the associated cost cannot be larger, so the result still holds.)

Consider now an arbitrary set Γ_N of decision rules in which sensors $1, ..., N$ use thresholds $t_1, ..., t_N$, respectively. Using the corollary in p.84 of [ShGaBe], we obtain

$$\log J_N(\Gamma_N) \geq c + \sum_{i=1}^{N} \mu(t_i, \hat{s}) - [\sum_{i=1}^{N} \mu''(t_i, \hat{s})]^{1/2},$$

where c is an absolute constant and where $\hat{s}$ is chosen so that it minimizes $\sum_{i=1}^{N} \mu(t_i, s)$ over all $s \in [0,1]$. Using the definition of (t^*, s^*), we have $\mu(t^*, s^*) \leq \mu(t_i, \hat{s})$, $\forall i$. Using also the bound A on μ'', we obtain $\log J_N(\Gamma_N) \geq c + N\mu(t^*, s^*) - (AN)^{1/2}$. Taking the infimum over all Γ_N, we obtain $R_N \geq (c/N) + \mu(t^*, s^*) - (A/N)^{1/2}$. Taking the limit as $N \to \infty$, we obtain $\liminf_{N\to\infty} R_N \geq \mu(t^*, s^*) \geq \limsup_{N\to\infty} r_N^*$. The reverse inequality also holds by definition. This concludes the proof of the theorem. •

The above theorem demonstrates that having each sensor use the same threshold t^* is asymptotically optimal, as $N \to \infty$, in the sense that it achieves the best error exponent. Furthermore, the definition of t^* demonstrates that it can be computed fairly easily. In typical applications, $f_i^t(j)$, and therefore $\mu(t, s)$, is given by a simple analytical expression. Thus, we only need to solve a nonlinear optimization problem in two dimensions which is not hard to do numerically. This task is facilitated further by the fact that $\mu(t, s)$ is convex, as a function of s [ShGaBe]. Unfortunately, there do not seem to be any simple examples for which t^* can be evaluated analytically.

Our result may be restated in a different language refering to a different context. Suppose that we want to transmit a binary message and that we have a collection of noisy binary memoryless and independent channels in our disposal. We are allowed to transmit a total of N times using any of the available channels. A receiver observes the N outputs of the channels, uses its knowledge of which channels were being used, and decides whether a zero or a one has been transmitted. The problem consists of finding which channels should be used and how many times, in order to maximize the probability of correct decoding. For small N, it may be better to use a different channel each time. However, our result states that, as $N \to \infty$, there is a single best channel which may be used for each transmission. To see the analogy, think of the hypotheses H_0 and H_1 as the value of the binary message which we want to transmit and think of u_i as the output of the i–th transmission. A different channel corresponds to a different choice of the threshold and the characteristics of the channel correspond to the quantities $f_i^t(j)$ (see Figure 1).

A different analogy may be made in the context of optimal design of measurements for failure detection. 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.

So far we have restricted ourselves to the case of two hypotheses and binary messages from the sensors to the fusion center. The case of K–valued messages (taking values in $\{1, ..., K\}$ may be handled in the same way. Given a candidate decision rule γ for the sensors (in this case γ may be parameterized by $K-1$ thresholds), we define $f_i^{\gamma}(j) = P(\gamma(y) = j|H_i)$ and $\mu(\gamma, s) = \sum_{k=1}^{K} f_1^{\gamma}(j)^{1-s} f_2^{\gamma}(j)^s$. An asymptotically optimal decision rule γ to be used by all sensors is determined by minimizing μ over γ and s. Notice that the minimization is now over a K–dimensional space. Given that $\mu(\gamma, s)$ does not seem to be a convex function of γ (no matter how we choose the $K-1$–dimensional parametrization of γ), finding a global minimum of $\mu(\gamma, s)$ is likely to become intractable as K becomes large. (Local minima are of course always easy to find, but there are no guarantees about their quality.)

The case of M–ary $(M > 2)$ hypotheses, is substantially different. Suppose that the messages are still constrained to be binary. Then, the analog of Theorem 1 fails to hold, as demonstrated by the following trivial example. Let there be three equally likely hypotheses H_1, H_2, H_3 and assume that for each sensor i, its observation y_i is equal to j with probability one if H_j holds. Clearly binary messages from two sensors are sufficient for the center to attain zero error probability. On the other hand, if all sensors use the same decision rule the probability of error is at least $1/3$, no matter how many sensors are involved. For example, if the decision rule employed is $u_i = 0$, if $y_i = 1$, and $u_i = 1$, otherwise, then the center has no way of distinguishing H_2 from H_3.

For the case of M–ary hypotheses and with the messages allowed to be M–valued (rather than binary), we do not know whether an analog of Theorem 1 holds and we conjecture it doesn't.

II. DECENTRALIZED SEQUENTIAL DETECTION.

Let there be again two hypotheses H_0 and H_1 with known prior probabilities. At each time $t \in \{0, 1, 2, ...\}$ we have N_t sensors which obtain observations $y_{i,t}$, $i = 1, ..., N_t$. We assume that the random variables $y_{i,t}$ are independent, conditioned on either hypothesis. Each sensor evaluates a message $u_{i,t}$ according to a rule $u_{i,t} = \gamma_{i,t}(y_{i,t})$ which it transmits to a fusion center. The fusion center at each time t receives the messages $u_{i,t}$ and has three options: declare H_0 true, declare H_1 true or defer the decision at a cost of C units. The decision of the fusion center at time t is constrained to be a function of the information available to it which is $\mathcal{F}_t = \{u_{i,\tau} : 0 \le \tau \le t, 1 \le i \le N_\tau\}$. The objective is to choose the decision rules $\gamma_{i,t}$ of the sensors, as well as the decision rule of the fusion center so as to minimize the probability of a wrong decision plus the total cost of deferring the decision (the latter being equal to C times the time at which is the decision is made). It is shown below that each of the sensors should use a likelihood ratio rule for deciding what to transmit and that the decision rule of the fusion center has the standard form encountered in centralized sequential detection. This result is similar to the results of [TeHo, TeVa, PaAt] and the proof is straightforward. Our interest in this result is in comparing it with the counterexample in the end of this section.

Theorem 2: Under the above assumptions, the decision rules for the sensors and the fusion center may be constrained to have the following structure, without increasing the value of the optimal expected cost:

a) For any t, i, there exists a threshold $T_{i,t}$ such that the decision of the i–th sensor at time t has the following form: $u_{i,t} = 0$ if and only if $P(y_{i,t}|H_1)/P(y_{i,t}|H_0) < T_{i,t}$;

b) For any t, there exist thresholds $A_t \leq B_t$ such that the decision of the fusion center at time t is equal to H_0, "defer", or H_1 according to whether $P(H_1|\mathcal{F}_t)/P(H_0|\mathcal{F}_t)$ lies below A_t, between A_t and B_t, or above B_t, respectively.

Proof: For any fixed decision rules of the sensors, the fusion center is faced with a classical (centralized) sequential detection problem. Thus, part (b) of the theorem follows immediately. For part (a), the argument goes as follows. Consider the i–th sensor at time t and suppose that the decision rules of all the other sensors, as well as of the fusion center are fixed. Given its observation $y_{i,t}$, if that sensor decides to transmit the value 0, the expected cost will be $E[Cost|H_0, u_{i,t} = 0]P(H_0|y_{i,t}) + E[Cost|H_1, u_{i,t} = 0]P(H_1|y_{i,t})$. There is an analogous expression for the expected cost if it decides to transmit the value 1. Obtaining the minimum of two such expressions, which are linear in $P(H_0|y_{i,t})$, $P(H_1|y_{i,t})$, is equivalent to comparing the ratio $P(H_0|y_{i,t})/P(H_1|y_{i,t})$ to a threshold, which is a likelihood ratio test. •

Remark: Typically, the results in the literature, together with the above characterization of the optimal decision rules, also provide a set of coupled equations satisfied by the optimal thresholds. These are straightforward to write down in our case but not particularly enlightening and we therefore omit them. Furthermore, such equations are not particularly useful in practice since they are almost impossible to solve numerically. The only exception arises if one considers the infinite horizon limit of a time–invariant problem. In that case, we have a smaller set of equations which in some cases may be solved approximately, using asymptotic formulas from sequential analysis, as done in [TeHo].

The above studied problem may be unrealistic in certain occasions. The main deficiency is the assumption that the random variables $y_{i,t}$ are conditionally independent. While it may be realistic that observations are independent across different sensors at a given time, in many situations it is not true that observations at different times are independent. Suppose, for example that the i-th sensor at time t is the same physical entity as the i-th sensor at time $t+1$. In that case, the information available to that sensor at time $t+1$ is the information available to it at time t together with any new measurements. Clearly, these two pieces of information are not conditionally independent. The question then arises whether likelihood ratio threshold rules are still optimal when the conditional independence assumption is dropped. The answer is negative, in general, and also in the special case referred to above, in which a fixed set of sensors accumulate information over time.

In more detail, the special case mentioned above is the following: There are N sensors. Each sensor i at time t observes a random variable $y_{i,t}$ and is allowed to transmit a binary message to the fusion center which is constrained to be a function of the information $\{y_{i,0}, ..., y_{i,t}\}$ available to that sensor. We assume that the $y_{i,t}$'s are conditionally independent given either hypothesis. Concerning the fusion center, the assumptions are as before.

Consider now the following particular example of the last described problem. There are two sensors ($N = 2$) and the two hypotheses are assumed to be equally likely. The random variable $y_{1,0}$ observed by sensor 1 at time zero has the distribution depicted in Figure 2a. The random variables $y_{1,t}$, for $t > 0$ are assumed to be deterministic, so that they carry no information. Concerning the second sensor, the random variable $y_{2,0}$ observed at time 0 has the conditional distribution depicted in Figure 2b and the random variables $y_{2,t}$ for $t > 0$ are assumed to be deterministic. We assume that $y_{1,0}$ and $y_{2,0}$ are conditionally independent. Notice that all the information ever obtained by the sensors may be transmitted to the fusion center by time $t = 1$. Also, there is an essentially unique decision rule for the second sensor. Let us assume that the cost C of deferring the decision is very small. In that case the problem becomes one of choosing the decision rule of the first sensor so as to minimize the expected time at which the final decision is made subject to the constraint that the final decision is the optimal one, given the random variables $y_{1,0}$ and $y_{2,0}$. Thus, the decison of the fusion center has to be: declare H_0 true if $y_{1,0} = 1$ or if $y_{1,0} = 2$ and $y_{2,0} = 1$; declare H_1 true otherwise. We first consider the threshold rule in which $u_{1,0} = 0$ if $y_{1,0} = 1$ and $u_{1,0} = 1$, otherwise. The fusion center cannot make the final decision at time $t = 0$ if $u_{1,0} = 1$ and $y_{2,0} = 1$ because it must wait one more time unit to learn whether $y_{1,0}$ equals 2 or 3. In particular, it has to wait with probability $(1/6) + (p/3)$. By symmetry, the same conclusion is obtained if sensor 1 uses the decision rule $u_{1,0} = 0$, if $y_{1,0} \in \{1, 2\}$ and $u_{1,0} = 1$, otherwise.

Consider now a non–threshold decision rule for sensor 1, where $u_{1,0} = 0$ if $y_{1,0} \in \{1, 3\}$ and $u_{1,0} = 1$, otherwise. If $u_{1,0} = 1$, then the fusion center may make an optimal decision at time zero, based on the value of $y_{2,0}$. Therefore, it has to wait with probability only $1 - p$. By letting p be large enough, this non–threshold rule is better than any threshold rule; therefore, threshold rules are not, in general, optimal for problems of this type.

We may conclude from the above discussion that threshold rules are optimal in sequential problems if each message to be sent is a function of a random variable which is conditionally independent from the other random variables. If on the other hand the same sensor has more than one chance of transmitting information, then threshold rules are no more optimal. Rather, such a sensor should try to code information in the messages so that the most useful pieces of information are transmitted first. How to choose such a coding rule in an optimal way is related to the problems studied in [PaTs, Wi] and we conjecture that it is an intractable combinatorial problem.

ACKNOWLEDGEMENTS: I am grateful to Prof. Robert Gallager who suggested that the results of [ShGaBe] could be used in proving Theorem 1.

REFERENCES

[Ek] Ekchian, L.T., "Distributed detection and communication problems", Ph.D. dissertation, Dept. of EECS, MIT, Cambridge, MA, 1982.

[LaSa] Lauer, G.S., and Sandell, N.R., Jr., "Decentralized detection given waveform observations", Alphatech Inc. Technical Paper TP–131, Burlington, MA, 1982.

[PaTs] Papadimitriou, C.H., Tsitsiklis, J.N., "On the complexity of designing distributed protocols", *Information and Control,* 53, 1982, pp. 211–218.

[PaAt] Papastavrou, J., Athans, M., "A Distributed Hypothesis–Testing Team Decision Problem with Communications Cost", Technical Report LIDS–P–1538, Laboratory for Information and Decision Systems, M.I.T., 1986. [ShGaBe] Shannon, C.E., Gallager, R.G., Berlekamp, E.R., "Lower bounds to error probability for coding on discrete memoryless channels, I", *Information and Control,* 10, 1967, pp. 65–103.

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

[TeVa] Teneketzis, D., Varaiya, P., "The decentralized quickest detection problem", *IEEE Transactions on Automatic Control,* AC–29, 7, 1984.

[TeSa] Tenney, R.R., Sandell, N.R., Jr., "Detection with distributed sensors", it IEEE Transactions on Aerospace and Electronic Systems, AES–17, 4, 1981.

[TsAt] Tsitsiklis, J.N., Athans, M., "On the Complexity of Decentralized Decision Making and Detection Problems", IEEE Transactions on Automatic Control, AC–30, 5, 1985.

[Wi] Witsenhausen, H.S., "The zero error side information problem and chromatic numbers", *IEEE Transactions on Information Theory,* IT–22, 5, 1976, pp. 592–593.

APPENDIX A

We consider here the problem introduced in Section 1, with two agents ($N = 2$) and with y_1, y_2 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 [TeSa]. However, that example had a different cost function which (roughly speaking) introduced a large penalty if both sensors send the same message and the wrong decision is made by the fusion center. Th asymmetry of the optimal decision rules of the two sensors could be ascribed to this particular aspect of the cost function and therefore did not prove that asymmetrical decision rules may be optimal for our cost function (probability of error).

Our example is the following. We let H_0 and H_1 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:

$P(y = 1|H_0) = 4/5, \qquad P(y = 2|H_0) = 1/5 \qquad P(y = 3|H_0) = 0,$

$P(y=1|H_1)=1/3, \qquad P(y=2|H_1)=1/3 \qquad P(y=3|H_1)=1/3.$

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) $u_i = 0$ iff $y_i = 1$,

(B) $u_i = 1$ iff $y_i \in \{1,2\}$.

Thus, we need to consider three possibilities: (i) both sensors use (A); (ii) both sensors use (B); sensor 1 uses (A) and sensor 2 uses (B). For each possibility we also need to consider the two alternative decision rules available to the center: decide H_0 if and only if $u_1 = u_2 = 0$, versus, decide H_0 if and only if $u_1u_2 = 0$.

Explicit evaluation of the expected cost for each of the six combinations 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_0 if and only if $u_1 = u_2 = 0$, for an expected cost of 19/90.

APPENDIX B:

In this appendix we explore Assumption 1(b) and present some simple sufficient conditions on the distribution of the y_i's which force this assumption to hold.

Proposition: Suppose that Assumption 1(a) holds and that there exist $B > 0$, $\alpha > 0$, $\beta \in (0,1)$ such that for any t and for $j = 0,1$ we have either:

(i) $f_0^t(j) = f_1^t(j) = 0$, or

(ii) $\frac{1}{B} f_0^t(j)^{1+\alpha} \leq B f_1^t(j)$.

Then, there exists some A such that $\mu''(t,s) \leq A$, $\forall (t,s) \in [0,\infty] \times [0,1]$.

Proof: (Outline) The derivatives of $\mu(t,s)$, with respect to s are easily calculated to be [ShGaBe, equations (3.24)–(3.25)]:

$$\mu'(t,s) = \sum_j \frac{f_0^t(j)^{1-s} f_1^t(j)^s}{\sum_k f_0^t(k)^{1-s} f_1^t(k)^s} \log \frac{f_1^t(j)}{f_0^t(j)}, \tag{2}$$

$$\mu''(t,s) = \left[\sum_j \frac{f_0^t(j)^{1-s} f_1^t(j)^s}{\sum_k f_0^t(k)^{1-s} f_1^t(k)^s} \left(\log \frac{f_1^t(j)}{f_0^t(j)} \right)^2 \right] - [\mu'(t,s)]^2, \tag{3}$$

where all summations are made over only those j's and k's for which $f_0^t(j) f_1^t(j)$, (respectively, $f_0^t(k) f_1^t(k)$) is nonzero.

We first consider

$$\sum_k f_0^t(k)^{1-s} f_1^t(k)^s. \tag{4}$$

It is positive, because of Assumption 1(a), and we will show that is bounded away from zero, uniformly in t, s. We first consider the case of $f_0^t(0) = 0$. Let $Y_0^t = \{y : \gamma^t(y) = 0\}$. We then have $P(Y_0^t|H_0) = 0$. We partition the set of all y's into partitioned in three disjoint sets Y_0, Y_1, Y_2, such that $P(Y_0|H_0) = 0$, $P(Y_1|H_1) = 0$ and each of the measures $P(\cdot|H_0)$, $P(\cdot|H_1)$ is absolutely continuous with respect to the other on Y_2. It follows that Y_0^t consists of a subset of Y_0 together with a subset of Y_2 which has zero probability under $P(\cdot|H_1)$. Therefore, $f_1^t(0) \leq P(Y_0|H_1)$. Furthermore, because of Assumption 1(a), Y_2 has nonzero probability under either hypothesis, which shows that $f_1^t(0) = P(Y_0^t|H_1) \leq P(Y_0|H_1) < 1$. We now use the relation $f_1^t(0) + f_1^t(1) = 1$; the sum in (4) equals $f_1^t(1)^s$ which is bounded below by $f_1^t(1)$. The latter quantity is bounded below by $1 - P(Y_0|H_1) > 0$. This provides the desired lower bound on the sum in (4), under the assumption $f_0^t(0) = 0$. The same argument works for the cases where $f_0^t(1)$ or $f_1^t(0)$ or $f_1^t(1)$ is zero.

We now turn to the case where $f_i^t(j) \neq 0$, $\forall i, j$. Then, the sum in (4) may be written as $f_0^t(0)^{1-s} f_1^t(0)^s + f_0^t(1)^{1-s} f_1^t(1)^s$. Using the bounds assumed for this proposition, the above expression is bounded below by

$$f_0^t(0)^{1-s}\frac{1}{B}f_0^t(0)^{s(1+\alpha)} + f_0^t(1)^{1-s}\frac{1}{B}f_0^t(1)^{s(1+\alpha)} \geq \frac{1}{B}[f_0^t(0)^{1+s\alpha} + (1 - f_0^t(0))^{1+s\alpha}] \geq \frac{1}{B} > 0.$$

(The last step is the Minkowski inequality.) This completes the proof of the desired bound for the expression in (4).

To complete the proof of the proposition, it suffices to show that if $f_0^t(j) f_1^t(j) \neq 0$, then

$$f_0^t(j)^{1-s} f_1^t(j)^s \log \frac{f_1^t(j)}{f_0^t(j)} \tag{5}$$

and

$$f_0^t(j)^{1-s} f_1^t(j)^s \log^2 \frac{f_1^t(j)}{f_0^t(j)} \tag{6}$$

are bounded, uniformly over all t, s. (Boundedness of the first expression provides a uniform bound on $\mu'(t,s)$; boundedness of the second expression provides the desired bound on $\mu''(t,s)$.)

Let $C = \max_{x\in(0,1]} |x \log x|$, which is known to be finite. Let us assume, without loss of generality that $f_1^t(j) \leq f_0^t(j)$. We have $f_1^t(j)/f_0^t(j) \geq \frac{1}{B} f_0^t(j)^\alpha$. Therefore, the expression in (5) is bounded (in absolute value) by $f_0^t(j)[\log B + \alpha |\log f_0^t(j)|] \leq \log B + \alpha C$.

An almost identical argument yields the same conclusion for the expression in (6), provided that we define C as $\max_{x\in(0,1]} |x \log^2 x|$. This completes the proof of the proposition. •

We mention a few examples where the conditions assumed in the above proposition are easily verified:

(a) If the supports of $P(\cdot|H_0)$, $P(\cdot|H_1)$ are finite sets with nonempty intersection.

(b) If $P(\cdot|H_i)$ are exponential distributions with different means.

(c) If $P(\cdot|H_i)$ are normal distributions with different means and variances.

However, the conditions assumed would not hold if $P(\cdot|H_0)$ was exponential and $P(\cdot|H_1)$ was normal. Roughly speaking, these conditions require that the tail of the distribution of y behaves in roughly the same way under either hypothesis. Let us close by conjecturing that Theorem 1 remains valid even if this assumption is removed. The reason behind this conjecture is that, even as the number N of sensors tends to infinity, the optimal thresholds may be expected to lie in a compact subset of $(0, \infty)$ (for nondegenerate problems), in which case the nature of the tails of the distributions does not matter.

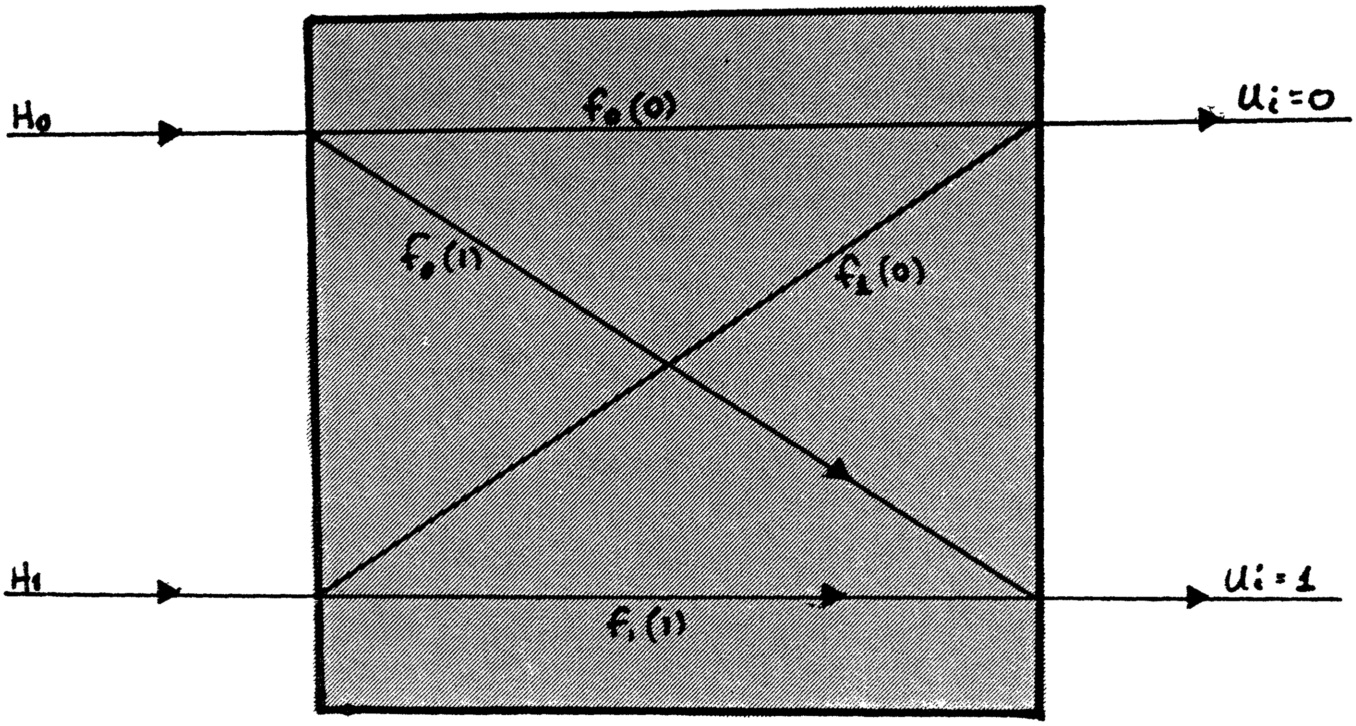

Fig. 1

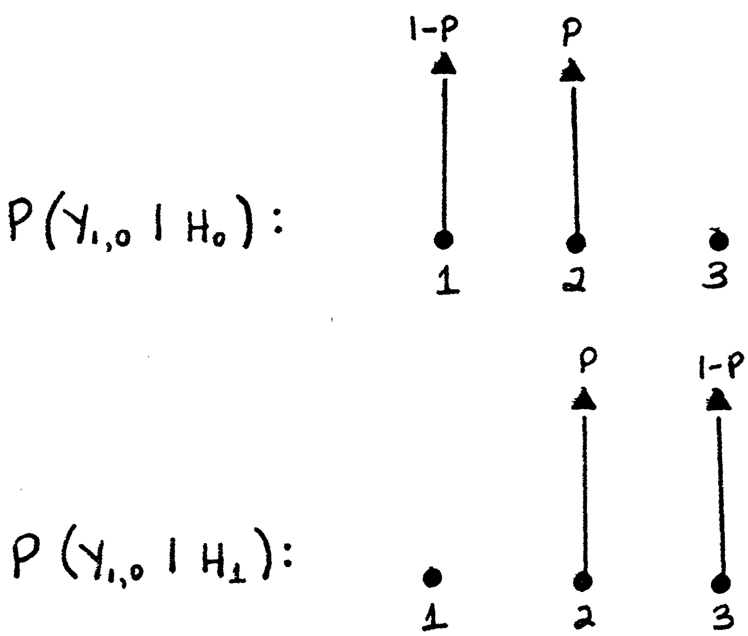

Fig. 2a

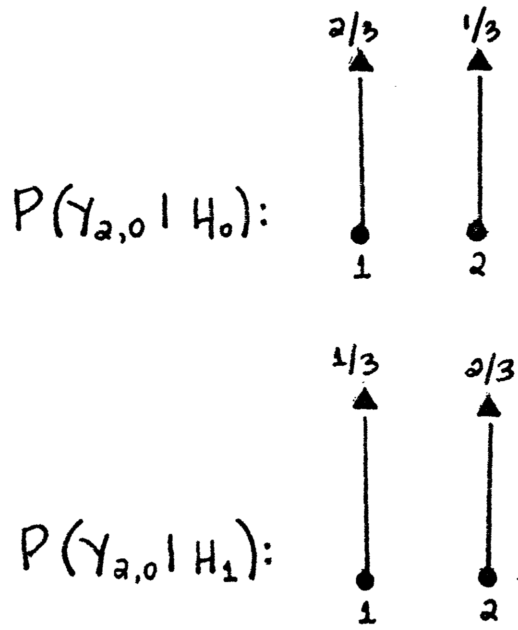

Fig.2b