

**TECHNICAL  
RESEARCH  
REPORT**

*Institute for  
Systems  
Research*

**Signal Detection Games with  
Power Constraints**

*by D. Sauder and E. Geraniotis*

*The Institute for Systems  
Research is supported by the  
National Science Foundation  
Engineering Research Center  
Program (NSFD CD 8803012),  
Industry and the University*

TR 93-16

# SIGNAL DETECTION GAMES WITH POWER CONSTRAINTS

Doug Sauder and Evaggelos Geraniotis

## ABSTRACT

In this paper we formulate mathematically and solve maximin and minimax detection problems for signals with power constraints. These problems arise whenever it is necessary to distinguish between a genuine signal and a spurious one designed by an adversary with the principal goal of deceiving the detector. The spurious (or deceptive) signal is usually subject to certain constraints, such as limited power, which preclude it from replicating the genuine signal exactly.

The detection problem is formulated as a zero-sum game involving two players: the detector designer and the deceptive signal designer. The payoff is the probability of error of the detector which the detector designer tries to minimize and the deceptive signal designer to maximize. For this detection game, saddle point solutions—whenever possible—or otherwise maximin and minimax solutions are derived under three distinct constraints on the deceptive signal power; these distinct constraints involves lower bounds on (i) the peak power, (ii) the probabilistic average power, and (iii) the time average power. The cases of i.i.d. and correlated signals are both considered.

---

D. Sauder was with the Department of Electrical Engineering, University of Maryland, College Park, MD, he is now with Locus, Inc, Alexandria Virginia, VA.

E. Geraniotis is the Department of Electrical Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742.



# **SIGNAL DETECTION GAMES WITH POWER CONSTRAINTS**

Douglas W. Sauder and Evaggelos Geraniotis

Copyright c 1993. Douglas W. Sauder and Evaggelos Geraniotis. All Rights Reserved.



## I. INTRODUCTION

In certain areas of communications, it is necessary to employ a detector to distinguish between a genuine signal and a spurious one. An adversary may be the designer of the spurious signal, whose principal goal is to defeat the detector. The spurious (or deceptive) signal is subject to certain constraints, such as limited power, which preclude it from exactly replicating the genuine signal.

It is sometimes reasonable to assume that the detector designer has limited knowledge (or no knowledge at all) of the statistics of the deceptive signal. It is also reasonable to assume that information about the detector is kept confidential from the designer of the deceptive signal. Accordingly, the appropriate mathematical formulation is that of a mathematical game. The game we consider here involves two players, whom we call the detector designer and the deceptive signal designer, in a zero-sum game. The strategy of the detector designer is a detector (or decision rule) which decides whether the received signal is genuine or spurious. The strategy of the deceptive signal designer is a probability distribution, subject to certain constraints, that governs the statistics of the spurious signal. The payoff is the probability of error of the detector, which the detector designer tries to minimize and the deceptive signal designer tries to maximize.

In this paper we derive the optimal strategies for the detector designer and the deceptive signal designer when constraints are imposed on the peak power or average power of the deceptive signal. The paper is organized as follows: In Section II, the problem is formally introduced as a mathematical game involving a detector designer and a deceptive signal designer. In Section III, we consider the selection of an optimal deceptive signal and an optimal detector when average power constraints imposed involve only time averaging. It is shown that a saddle point to the game exists if the constraints are fully known by the detector designer as well as the signal designer. In Section IV, we consider the problem when power constraints involving ensemble averages are imposed. The optimal density for the deceptive signal is obtained for the case of univariate densities. The formal proof of the main result in Section VI (*viz* Theorem 2) is provided in Section V. In Section VI, we provided briefly some numerical examples and compare the different deceptive signal densities, while in Section VII we discuss some extensions of the results. Finally, in Section VIII conclusions are drawn.

## II. PROBLEM FORMULATION

The task of the detector is to decide whether the signal it receives is genuine or spurious. For our purposes, it is sufficient to model this task as a binary hypothesis testing problem with  $H_0$

being the hypothesis that the signal is spurious and  $H_1$  the hypothesis that the signal is genuine. The signal received by the detector is assumed to be discrete-time, possibly obtained by sampling a continuous-time signal; multidimensional, with dimensionality denoted by  $m$ ; and stationary. The detector collects a sample of size  $n$  and makes a decision concerning the authenticity of the signal. Thus the data available to the detector is denoted  $\mathbf{x} = (x_1, \dots, x_n)$ , where each  $x_t = (x_{t1}, \dots, x_{tm})$  is an  $m$ -dimensional vector. The hypotheses are formally stated thus:

$$\begin{aligned} H_0 : \quad \mathbf{X} \text{ has a distribution } F_0^{(n)} \quad & \text{(spurious)} \\ H_1 : \quad \mathbf{X} \text{ has a distribution } F_1^{(n)} \quad & \text{(genuine)} \end{aligned} \tag{1}$$

where  $\mathbf{X} = (X_1, \dots, X_n)$  is a random vector with each component  $X_t$  being itself an  $m$ -dimensional random vector. While the distribution  $F_1^{(n)}$  of the genuine signal is known to both the detector designer and the deceptive signal designer, the distribution  $F_0^{(n)}$  is chosen by the deceptive signal designer and is unknown to the detector designer. We use  $F_i$  to denote the  $m$ -variate marginal density of  $F_i^{(n)}$  corresponding to any one of the components of  $\mathbf{X}$ . We assume that the distribution under  $H_1$  (the genuine signal) is continuous, having a density function  $f_1^{(n)}$ . In cases where there is also a density under  $H_0$ , it is denoted by  $f_0^{(n)}$ . The marginal density of  $X_t$  under  $H_i$  is denoted  $f_i$ .

The detector is represented by a randomized decision rule  $d$  such that, when  $\mathbf{x}$  is observed,  $H_1$  is decided with probability  $d(\mathbf{x})$ . No restrictions are placed on  $d$ ; any measurable function from  $\mathbb{R}^{nm}$  to  $[0, 1]$  is permitted. One approach for the detector designer is to design a detector based on the assumption of a particular distribution  $F_0^{(n)}$ . Since the detector designer cannot know the true density which governs the spurious signal, a detector designed in this manner typically operates in a mismatch situation.

While the detector may be chosen without restriction, the deceptive signal designer's strategy, which is the  $nm$ -dimensional distribution  $F_0^{(n)}$  of the spurious signal, must conform to certain constraints. First, there is an upper bound on the magnitude of the signal, given by

$$P\{q_U(X_t) > A_{\max}\} = 0 \tag{2}$$

where  $q_U(\cdot)$  is a quadratic form. Because of our assumption that the process is stationary, the bound given by (2) is understood to hold for any  $t$ . Note that if the signal dimension is  $m = 2$ , with the components of  $x_t$  being in-phase and quadrature components and  $q_U(x_t) = x_{t1}^2 + x_{t2}^2$ , then (2) has the interpretation of a constraint on the peak power or signal amplitude.

A lower bound is also imposed on the signal. We consider three such lower bounds. Analogous to (2), we have first

$$P\{q_L(X_t) < A_{\min}\} = 0 \quad (3)$$

where  $q_L$  is also a quadratic form. This constraint also has an interpretation as a constraint on signal amplitude. A more general form of (3) is

$$P\{q_L(X_1, \dots, X_N) < D\} = 0 \quad (4)$$

where  $q_L$  is again a quadratic form. As a particular case, we may have  $m = 2$  and

$$q_L(x_1, \dots, x_N) = \frac{1}{N} \sum_{t=1}^N (x_{t1}^2 + x_{t2}^2) \quad (5)$$

which has the interpretation of a lower bound on time-averaged power over a window of length  $N$ . Of course, (4) reduces to (3) when  $N = 1$ . Finally, a lower bound may be imposed instead on the expected power of the signal, given by

$$E_0 q_L(X_t) \geq E_{\min} \quad (6)$$

where  $E_0$  denotes expectation taken under the distribution  $F_0$ . Whereas upper-bound constraints, such as (2), typically arise from physical constraints on a transmitter, the lower-bound constraints, such as (3), (4), and (6), are subjectively defined, their purpose being to guarantee a minimum amplitude or power level. A weak signal risks the possibility of being dismissed as pure noise by the detector's receiver; hence, a minimum power level is important.

Loosely speaking, the payoff of the game is the error probability of the detector. More precisely, the payoff is a function of the error probabilities  $P_0$  and  $P_1$ , which denote errors occurring when  $H_0$  and  $H_1$  are true, respectively. Such a function may be defined according to the Bayes, Neyman-Pearson, minimax, or other performance criterion; in general, the performance measure need only satisfy certain necessary properties. The performance measure is denoted  $S(F_0^{(n)}, d)$ , as a function of the distribution  $F_0^{(n)}$  and the decision rule  $d$ .

In the theory of mathematical games, the goal is usually to obtain strategies that are either maximin or minimax, depending on whether a player's objective is to maximize or minimize the payoff value. We shall not consider the so-called mixed strategies, where the strategy is chosen randomly according to a specified distribution at each play. A *maximin* solution is defined as a pair  $(d^M, F_0^M)$  which satisfies

$$S(d^M, F_0^M) = \max_{F_0} \min_d S(d, F_0) \quad (7)$$



while a *minimax* solution is defined as a pair  $(d^m, F_0^m)$  which satisfies

$$S(d^m, F_0^m) = \min_d \max_{F_0} S(d, F_0) \quad (8)$$

A *saddle point* exists when the maximin pair and the minimax pair are the same or, alternatively, when the following inequality holds:

$$S(\hat{d}, F_0) \leq S(\hat{d}, \hat{F}_0) \leq S(d, \hat{F}_0). \quad (9)$$

In (9), the pair  $(\hat{d}, \hat{F}_0)$  is called a saddle point and  $\hat{d} = d^M = d^m$  and  $\hat{F}_0 = F_0^M = F_0^m$ . It is clearly the goal of the deceptive signal designer to employ a maximin density and of the detector designer to employ a minimax detector.

The performance measure  $S$  may be defined according to a Bayes, Neyman-Pearson, or other criterion. Rather than considering several different performance measures, we merely state the properties which are required of  $S$ . First,  $S$  must be bounded and a continuous function of the error probabilities  $P_0$  and  $P_1$ . Second, we require that the optimal detector for  $S$  be given by the likelihood ratio test (LRT). Some of our proofs depend on showing that a given distribution is not a maximin distribution. Thus we require the following property: Let  $d$  be a decision rule and  $F$  a distribution that are candidates for being a maximin pair, that is,  $F$  is a potential maximin distribution and  $d$  is its minimizing decision rule. We may verify that  $d$  and  $F$  are not a maximin pair with respect to  $S$ , if we can find a decision rule  $d'$  and a distribution  $F'$  such that one of the following two conditions is true:

- (1)  $P_1(d) = P_1(d')$  and  $P_0(d, F) < P_0(d', F') = \min_{r \in \mathcal{D}_0} P_0(r, F')$   
where  $\mathcal{D}_0 = \{r : P_1(r) \leq P_1(d)\}$
- (2)  $P_0(d, F) = P_0(d', F')$  and  $P_1(d) < P_1(d') = \min_{r \in \mathcal{D}_1} P_1(r)$   
where  $\mathcal{D}_1 = \{r : P_0(r, F') \leq P_0(d, F)\}$ .

It can be shown that any performance measure defined under the Bayes or Neyman-Pearson criterion satisfies the required properties.

### III. SIGNAL DETECTION GAME UNDER A TIME AVERAGE CONSTRAINT

#### A. The Optimal Strategies

We are concerned here with the detection game when the constraints (2) and (4) are imposed on the deceptive signal. The constraint given by (3) is included as a special case of (4) with  $N = 1$ .

The interpretation we are most interested in is that expressed by (5), namely, that of a time-average power constraint.

The constraints define a region  $B \in \mathbb{R}^{mn}$  which is permitted to have nonzero probability under  $F_0^{(n)}$ . Specifically, this region is

$$B = \{\mathbf{x} : \max_{1 \leq t \leq n} q_U(x_t) \leq A_{\max}, \min_{1 \leq t \leq n-N+1} q_L(x_t, \dots, x_{t+N-1}) \geq D\}. \quad (10)$$

The optimal strategies are stated in Theorem 1.

**THEOREM 1.** The maximin distribution  $\hat{F}_0^{(n)}$  and minimax detector  $\hat{d}$  form a saddle point of the game. The distribution  $\hat{F}_0^{(n)}$  has a density

$$\hat{f}_0^{(n)}(\mathbf{x}) = \begin{cases} c f_1^{(n)}(\mathbf{x}) & \text{if } \mathbf{x} \in B \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where  $c$  is given by

$$c = \left[ \int_B f_1^{(n)}(\mathbf{x}) d\mathbf{x} \right]^{-1}. \quad (12)$$

The detector  $\hat{d}$  is given by

$$\hat{d}(\mathbf{x}) = \begin{cases} \alpha & \text{if } \mathbf{x} \in B \\ 1 & \text{otherwise.} \end{cases} \quad (13)$$

*Proof.* The first inequality in (9) holds because the detector does not depend on the shape of the density  $f_0^{(n)}$ , but only on the properties that  $f_0^{(n)}(\mathbf{x}) = 0$ , whenever  $\mathbf{x} \notin B$ . Thus, for any density  $f_0^{(n)}$  which conforms to the lower and upper bounds, there will actually be equality for the first inequality in (9). The second inequality holds because  $\hat{d}$  is the likelihood ratio test for the densities  $\hat{f}_0^{(n)}$  and  $f_1^{(n)}$ .  $\square$

### B. Case of I.I.D. Observations

Although the result stated in Theorem 1 is for correlated observations, there is no significant simplification in the case of i.i.d. observations. If the genuine signal is i.i.d., the spurious signal given by Theorem 1 is not i.i.d. in general. However, it may be desired to generate the spurious signal as an i.i.d. signal, if the genuine signal is i.i.d. It turns out that the deceptive signal is i.i.d. if  $N = 1$ , so that the constraint (4) reduces to (3). The density (11) becomes a product density

$$\hat{f}_0^{(n)}(\mathbf{x}) = \prod_{t=1}^n \hat{f}_0(x_t) \quad (14)$$

where

$$\hat{f}_0(x) = \begin{cases} p^{-1} f_1(x) & \text{if } q_U(x) \leq A_{\max} \text{ and } q_L(x) \geq A_{\min} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

and  $p$  is defined by

$$p = P\{q_U(X_t) \leq A_{\max}, q_L(X_t) \geq A_{\min}\}.$$

No simplification of the detector rule (13) is possible.

### C. Method for Generating the Optimal Deceptive Signal

In theory, it is quite simple to generate the optimal deceptive signal whenever it is possible to generate data which have the same distribution as the genuine signal. To generate a vector  $\mathbf{x}$  from the multivariate density  $\hat{f}_0^{(n)}$ , first generate a vector  $\mathbf{x}$  from the density  $f_1^{(n)}$ . Then test the vector for acceptance by checking whether  $q_L(\mathbf{x}) \geq A_{\min}$  and  $q_U(x_t) \leq A_{\max}$ , for  $t = 1, \dots, n$ . If the conditions hold, then  $\mathbf{x}$  is accepted and is the desired vector from  $\hat{f}_0^{(n)}$ . Otherwise, the vector is rejected. The process is repeated until a vector is finally accepted.

The acceptance-rejection method works well if the probability of acceptance is large enough. In fact, the probability of acceptance is precisely  $c^{-1}$ , with  $c$  given by (12). The number of vectors  $M$  that must be generated before one is accepted has a geometric distribution, with the probability of success being  $c^{-1}$ . Therefore, the average value of  $M$  is  $c$ . If  $c^{-1}$  is very small, then the average value of  $M$  may be too large to make the acceptance-rejection method feasible. If  $A_{\min}$  and  $A_{\max}$  remain fixed and the dimensionality  $n$  is increased,  $c^{-1}$  decreases. Thus, if  $n$  is large, it is likely that the acceptance-rejection method may be too expensive to use.

If  $n$  is too large to make the method feasible, suboptimal methods are possible. Consider, for example, an ARMA model for the genuine signal,

$$X_t + \sum_{i=1}^L a_i X_{t-i} = \sum_{j=0}^{M-1} b_j U_{t-j} \quad (17)$$

where  $\{U_t\}$  is an i.i.d. process and the signal dimensionality is  $m = 1$ . After each  $X_t$  is generated, we check for the upper bound (2) and the lower bound (4). If the the recently generated  $X_t$  is rejected, then we may generate  $X_t$  again by generating another  $U_t$  and using the same  $X_{t-L}, \dots, X_{t-1}$  and  $U_{t-M+1}, \dots, U_{t-1}$  that were previously used. The method, of course, generates an approximately optimal signal. If i.i.d. data are to be generated, then each component  $x_i$  may be generated and tested independently of all other components.

#### IV. SIGNAL DETECTION GAME UNDER AN EXPECTED POWER CONSTRAINT

When we consider the lower bound (6) on the expected power, we are confronted with a problem which is significantly more difficult than the one considered in the preceding section. Consequently, we must take the dimensionality to be  $m = 1$  and the length of the observed data segment to be  $n = 1$ . We provide here the optimal (maximin) deceptive signal distribution, but not the optimal (minimax) detector. The density of the genuine signal is assumed to be non-zero only for positive numbers; in Section VII we discuss the application to densities which allow positive and negative values.

The solution to the problem is obtained in two phases. First we approximate the problem by a finite dimensional one and obtain its solution. Then we consider the limit as the dimensionality of the finite dimensional problem goes to infinity, thus obtaining the solution to the original problem.

##### A. The Finite Dimensional Problem

Throughout this section, we use the notation  $I_A$  for the indicator function

$$I_A(x) \triangleq \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

and  $F_i(u, v]$  and  $G_i(u, v]$  to denote the integrals

$$F_i(u, v] \triangleq \int_{(u, v]} dF_i(x)$$

$$G_i(u, v] \triangleq \int_{(u, v]} x^2 dF_i(x).$$

The integration is actually performed over the interval  $(u, v]$ , which includes the point  $v$ . More generally,  $F_i I$  and  $G_i I$  involve integration over the interval  $I$ , which may or may not include the endpoints. Since  $F_1$  does not contain any point masses, this distinction is only necessary when  $F_0$  is involved.

To make the analysis manageable, we assume that the distribution  $F_0$  has a density of the form

$$f_0(x) = \sum_{i=1}^M c_i I_{(u_{i-1}, u_i]}(x) f_1(x) \tag{18}$$

with  $c_i \geq 0$  and  $0 = u_0 < u_1 < \dots < u_M = A_{\max}$ . A justification for this particular form is given in the next section. (See Proposition 1.) As observed in (18), in each region the density  $f_0$

matches the density  $f_1$  as closely as possible, their ratio being constant. By restricting attention to densities of the form (18), we make the problem more manageable in that we must find only the optimal values of the parameters  $\{c_i\}$  and  $\{u_i\}$ . We also allow limiting forms of (18), so that, if  $u_{M-1} \rightarrow A_{\max}$ , for example, we may obtain a point mass at  $A_{\max}$ .

For the specific form of the density (18), the constraint (6) becomes

$$\sum_{i=1}^M c_i G_1(u_{i-1}, u_i] \geq E_{\min}. \quad (19)$$

This constraint imposes further constraints on the values of the parameters. Note that when

$$E_{\min} > \frac{G_1(0, A_{\max}]}{F_1(0, A_{\max}]} \quad (20)$$

does *not* hold, the solution is trivial: we take  $f_0$  to be a scaled version of  $f_1$  over the interval  $[0, A_{\max}]$ . In this case, the lower bound on expected power is superfluous. If (20) *does* hold, then probability mass must be moved to the right to increase the average power under  $H_0$ , that is, we cannot have  $c_1 = c_2 = \dots = c_M$ . Consequently, we assume that (20) is true.

The detector which is optimal against the density (18) is the likelihood ratio test (LRT), which has the form

$$d(x) = \sum_{i=1}^M \alpha_i I_{(u_{i-1}, u_i]}(x) + I_{(A_{\max}, \infty)}(x) \quad (21)$$

where  $0 \leq \alpha_i \leq 1$  for  $i = 1, \dots, M$ . Actually, the  $\alpha_i$  parameters in (21) may take only the three values 0,  $\alpha$ , and 1. This is because in a LRT, randomization is necessary only when  $f_1(x)/f_0(x)$  is precisely equal to the threshold. The randomization parameter  $\alpha$  determines the particular operating point on the receiver operating characteristic (ROC).

#### B. Case of $M = 2$

For the case of  $M = 2$ , the density  $f_0$  given by (18) has the form

$$f_0(x) = c_1 I_{(0, u_1]} f_1(x) + c_2 I_{(u_1, A_{\max}]} f_1(x). \quad (22)$$

The optimal (maximin) density takes one of the following two forms:

$$f_0(x) = c_1 I_{(0, A_{\max})} f_1(x) + h \delta(A_{\max} - x) \quad (23)$$

or

$$f_0(x) = c_2 I_{(u_{\min}, A_{\max}]}(x) f_1(x). \quad (24)$$

In the first density (23),  $u_1$  is taken as far to the right as possible, resulting in a point mass (Dirac  $\delta$  function) at the point  $A_{\max}$ . The value of  $c_1$  is taken as large as possible and the point mass  $h$  as small as possible, while still maintaining the lower bound (6) on expected power. The graph of such a density is shown in Figure 2, which may be compared to the graph of  $f_1$  in Figure 1. (N.B. We did not try to show the proper scaling on the  $y$ -axis.)

For this particular form of the density, the constraint (6) becomes

$$c_1 G_1(0, A_{\max}) + h A_{\max}^2 \geq E_{\min}. \quad (25)$$

Since  $f_0$  is a probability density, we have also

$$c_1 F_1(0, A_{\max}] + h = 1. \quad (26)$$

From (25) and (26), we obtain the inequalities

$$c_1 \leq \frac{A_{\max} - E_{\min}}{A_{\max} F_1(0, A_{\max}] - G_1(0, A_{\max})} \quad (27)$$

$$h \geq \frac{E_{\min} F_1(0, A_{\max}] - G_1(0, A_{\max})}{A_{\max} F_1(0, A_{\max}] - G_1(0, A_{\max})}. \quad (28)$$

Note that the best performance results are obtained when we take equality in (27) and (28).

In the second density (24),  $u_1$  is taken as far to the left as possible,  $u_1 = u_{\min}$ , causing  $c_1$  to take the value 0 and  $c_2$  to take its minimum value. The minimum value  $u_{\min}$ , obtained from (6) and the fact that  $f_0$  is a probability density, is given by  $u_{\min} = Z^{-1}(E_{\min})$ , where  $Z$  is the increasing function

$$Z(t) \triangleq \frac{G_1(t, A_{\max})}{F_1(t, A_{\max})} \quad (29)$$

and  $Z^{-1}$  its inverse. The minimum value of  $c_2$  is given by  $c_2 = F_1(u_{\min}, A_{\max})^{-1}$ . This density is of the same form as the one given in Section 3, where a lower bound on the amplitude was imposed. Figure 3 shows the graph of such a density.

Which of these forms the optimal density takes depends on the specific performance measure, the density  $f_1$ , and the value  $E_{\min}$ . It may be necessary to determine the value of the performance measure for both densities and compare the results in order to determine which density is optimal. As a particular example, we take the performance measure  $S$  to be the maximum of the two error probabilities. Because of the randomization in the decision rule, this means that  $S = P_0 = P_1$ , that is, the performance measure is the minimum value of  $P_0$  under the constraint that  $P_1 = P_0$ .

First consider the form given by (23) with equality taken in (27) and (28). The optimal decision rule (LRT) is

$$d(x) = \begin{cases} \alpha & \text{if } 0 < x < A_{\max} \\ 0 & \text{if } x = A_{\max} \\ 1 & \text{if } x > A_{\max}. \end{cases} \quad (30)$$

Therefore the error probabilities are given by

$$\begin{aligned} P_0 &= \alpha c_1 F_1(0, A_{\max}) \\ P_1 &= (1 - \alpha) F_1(0, A_{\max}). \end{aligned} \quad (31)$$

Taking  $\alpha$  so that  $P_0 = P_1$  yields

$$\begin{aligned} Q_1(E_{\min}) &\triangleq P_0 = P_1 \\ &= \frac{(A_{\max}^2 - E_{\min}) F_1(0, A_{\max})}{A_{\max}^2 - E_{\min} + A_{\max}^2 F_1(0, A_{\max}) - G_1(0, A_{\max})}. \end{aligned} \quad (32)$$

Next consider the form given by (24). The optimal decision rule against (24) is given by

$$d(x) = \begin{cases} \alpha & \text{if } u_{\min} < x \leq A_{\max} \\ 1 & \text{otherwise.} \end{cases} \quad (33)$$

Evaluating  $P_0$  and  $P_1$ , and choosing  $\alpha$  so that  $P_0 = P_1$ , yields

$$Q_2(E_{\min}) \triangleq P_0 = P_1 = \frac{F_1(u_{\min}, A_{\max})}{F_1(u_{\min}, A_{\max}) + 1}. \quad (34)$$

For a given value of  $E_{\min}$ , we evaluate the two expressions  $Q_1$  and  $Q_2$ . If  $Q_1 > Q_2$  then the optimal density is (23); otherwise, (24) is optimal.

### C. Case of $M = 3$

For the case of  $M = 3$ , the density  $f_0$  given by (18) takes the form

$$f_0(x) = c_1 I_{(0, u_1]} f_1(x) + c_2 I_{(u_1, u_2]} f_1(x) + c_3 I_{(u_2, u_3]} f_1(x). \quad (35)$$

The optimal density takes the form

$$f_0(x) = c_2 I_{(u_1, A_{\max})}(x) f_1(x) + h \delta(A_{\max} - x) \quad (36)$$

and is shown in Figure 4. Comparing (36) and (35), we see that in the optimization  $u_2 \rightarrow A_{\max}$ , causing a point mass  $h$  at  $A_{\max}$ ; and  $u_1$  decreases, forcing  $c_1$  to take the value 0. It is possible

and, in many cases actually true, that  $u_1 = 0$  or that  $h = 0$  in (36). In such cases, the density reduces to one of the ones which is optimal for two regions. It is not possible, however, to have  $u_1 = 0$  and  $h = 0$  simultaneously, if the condition (20) holds. Since we must have  $\int f_0(x) dx = 1$  and  $\int x^2 f_0(x) dx = E_{\min}$ , only one of the three parameters,  $c_2$ ,  $u_1$ , and  $h$ , may be chosen independently. This one independent parameter must be determined by a one-dimensional optimization; the other parameters may then be determined from the independent parameter.

As a particular example, we consider again the performance measure  $S = P_0 = P_1$ . The optimal decision rule (LRT) against the density (36) is

$$d(x) = \begin{cases} 1 & \text{if } 0 < x \leq u_1 \text{ or } x > A_{\max} \\ \alpha & \text{if } u_1 < x < A_{\max} \\ 0 & \text{if } x = A_{\max}. \end{cases} \quad (37)$$

The error probabilities are given by

$$\begin{aligned} P_0 &= \alpha c_2 F_1(u_1, A_{\max}) \\ P_1 &= (1 - \alpha) F_1(u_1, A_{\max}). \end{aligned} \quad (38)$$

Taking  $u_1$  as the independent parameter and choosing  $\alpha$  so that  $P_0 = P_1$  we have

$$Q(u_1, E_{\min}) \triangleq P_0 = P_1 = \frac{(A_{\max}^2 - E_{\min}) F_1(u_1, A_{\max})}{A_{\max}^2 F_1(u_1, A_{\max}) - G_1(u_1, A_{\max}) + A_{\max}^2 - E_{\min}}. \quad (39)$$

The function  $Q$  must be optimized over  $u_1$ . Since this function is differentiable, we may set the derivative equal to zero to find any potential maximum points in the interval  $(0, u_{\min})$ . The derivative is given by

$$\frac{\partial}{\partial u} Q(u, E_{\min}) = \frac{-f_1'(u)(A_{\max}^2 - E_{\min})\{A_{\max}^2 - E_{\min} + u^2 F_1(u, A_{\max}) - G_1(u, A_{\max})\}}{[A_{\max}^2 F_1(u_1, A_{\max}) - G_1(u_1, A_{\max}) + A_{\max}^2 - E_{\min}]^2}. \quad (40)$$

The maximizing value of  $u_1$  will occur either at a point in  $(0, u_{\min})$  where the derivative is zero, or at one of the endpoints 0 or  $u_{\min}$ . Whenever  $u_1$  is taken as one of the endpoints, density is the same as the density that is optimal for  $M = 2$ .

#### D. Case of $M \geq 4$

Nothing is to be gained by taking  $M \geq 4$ , since the optimal density will have the form (36). This is the gist of Theorem 2, which is proved in the following section.



THEOREM 2. Let  $F_0$  be the maximin distribution in  $\mathcal{C}_M$ . Then  $F_0$  takes one of these three forms: (a)  $F_0 \in \mathcal{C}_3$ , with a point mass at  $A_{\max}$  and  $c_1 = 0$ ; (b)  $F_0 \in \mathcal{C}_2$ , with a point mass at  $A_{\max}$ ; (c)  $F_0 \in \mathcal{C}_2$ , with  $c_1 = 0$ .

Evidently, the density  $f_0$  follows  $f_1$  as closely as possible over the “center” region  $(u_1, A_{\max}]$ , while a zero region is allowed to the left (at  $(0, u_1]$ ) and a point mass is allowed to the right (at  $A_{\max}$ ). The zero region and the point mass are required to increase the average power; however, introducing another “level” in the center region would make  $f_0$  look even less like  $f_1$ . We think that this intuitive reasoning makes good sense. Our proof, however, relies on mathematics and not intuition.

The form of the densities given for  $M = 2$  and  $M = 3$  actually define the form of the maximin distribution in the class of all distributions which satisfy the required constraints. This is the result of Theorem 3.

THEOREM 3. Let  $F_0$  be the maximin distribution in the class of all distributions satisfying the constraints (2) and (6). Then  $F_0$  has one of the forms given in Theorem 2.

*Proof.* Let  $\{F_{0M}\}$  be any sequence of d.f.s, where  $F_{0M}$  has the form given by (41), that converges to  $F_0$ . Let  $F_{0M}^*$  be the maximin d.f. in the class  $\mathcal{C}_M$ . From Theorem 2, we know that  $F_{0M}^*$  has one of the forms (a), (b), or (c). Furthermore, since  $F_{0M}^* = F_{03}^*$  for  $M > 3$ , the sequence  $F_{0M}^*$  converges to a distribution  $F_0^*$ . Let  $S(F)$  denote the minimum value of the performance measure:  $S(F) = \inf_d S(d, F)$ . Then  $S(F_{0M}) \geq S(F_{0M}^*)$  for all  $M$ . Since  $S$  is continuous, the inequality hold in the limit:  $S(F_0) \geq S(F_0^*)$ .

## V. PROOF OF THEOREM 2

In this section, we include a series of lemmas, propositions, and so on, which, had they been included in the preceding section, would have made it considerably less readable. Essentially, these results prove the statements made in the preceding section.

The first matter which deserves attention is the justification of the form of the density given by (18). From a purely engineering viewpoint, the approach to the problem based on  $M$ -regions can be justified by the argument that it is a practical approach to a difficult problem. However, a mathematical justification is also possible. It is possible to approximate any  $n$ -dimensional distribution function (d.f.) arbitrarily closely by a distribution function which has a density of the form

$$f(\mathbf{x}) = \sum_{i=1}^M c_i I_{R_i}(\mathbf{x}) \quad (41)$$

where the  $R_i$  are rectangles in  $n$ -dimensions. The approximation improves as  $M$  increases. Note that the density in (41) has flat tops, that is,  $f$  is constant over the region  $R_i$ . A mathematically precise statement of this approximation states that there is a sequence of d.f.s having densities of the form (41) that converge to the required d.f. in the Levy Metric (or weak topology), as  $M \rightarrow \infty$ . Thus we may conclude that we lose little in terms of optimality if we take  $M$  quite large.

When it comes to approximating the maximin distribution, we may, however, improve on the approximation (41). Let the regions  $R_1, R_2, \dots, R_M$  and the amount of mass  $\mu_i$  that is assigned to  $R_i$  be fixed. Call the set of d.f.s that satisfy this property  $\mathcal{B}_M$ , that is,

$$\mathcal{B}_M \triangleq \{F : F(R_i) = \mu_i, i = 1, \dots, M\}.$$

Then we have the following proposition.

**PROPOSITION 1.** If  $F$  is the maximin d.f. in  $\mathcal{B}_M$ , then  $F$  has a density of the form

$$f(\mathbf{x}) = \sum_{i=1}^M I_{R_i}(\mathbf{x}) c_i f_1(\mathbf{x}). \quad (42)$$

*Proof.* Let  $S$  be the performance measure involved and  $d$  the minimizing decision rule corresponding to  $F$ ,  $S(d, F) = \inf_r S(r, F)$ . Note that because of the form of  $d$ , it has the same performance for every distribution in  $\mathcal{B}_M$ . If  $F'$  is any other d.f. in  $\mathcal{B}_M$ , then  $\inf_r S(r, F') < S(d, F') = S(d, F)$ .  $\square$

The next issue we address is the proof that the maximin density in the class of  $M$ -region densities, where  $M > 3$ , is in the class of 3-region densities. We consider only univariate d.f.s which have densities of the form

$$f(x) = \sum_{i=1}^M I_{(u_{i-1}, u_i]}(x) c_i f_1(x), \quad 0 \leq c_i \leq K \text{ for } i = 1, \dots, M. \quad (43)$$

The upper bound  $K$  on the  $c_i$ s is necessary for now to avoid delta functions in the densities. Define the set  $\mathcal{C}_M(K)$  of d.f.s by

$$\mathcal{C}_M(K) = \{F : F \text{ has a density (43) and } \int x^2 f(x) dx \geq E_{\min}\}.$$

Note that the  $c_i$ 's need not all be distinct; in this sense,  $\mathcal{C}_{M-1}(K) \subset \mathcal{C}_M(K)$ . We make the following distinction, however: If  $F$  and  $F'$  are two different distributions, such that  $F \in \mathcal{C}_L(K)$ ,

$F' \in \mathcal{C}_M(K)$ ,  $F' \notin \mathcal{C}_L(K)$ ,  $M > L$ , and  $S(F, d) = S(F', d')$ , where  $d$  minimizes  $S$  with respect to  $F$  and  $d'$  minimizes  $S$  with respect to  $F'$ , then  $F'$  is not a maximin distribution. In other words, when the maximin distribution is not unique, only the distribution (or distributions) which is “minimal” is considered a maximin distribution.

The LRT for any d.f. in  $\mathcal{B}_M(K)$  has the form

$$d(x) = \sum_{i=1}^M \alpha_i I_{(u_{i-1}, u_i]}(x) + I_{(A_{\max}, \infty)}(x). \quad (44)$$

This is used in the results which follow.

LEMMA 1. If  $\{c_i\}$  are the parameters of the maximin density in  $\mathcal{C}_M(K)$ , then  $c_1 \leq c_2 \leq \dots \leq c_M$ .

*Proof.* In this proof, all symbols with a prime (') correspond to  $F'$ , while all the unprimed symbols correspond to  $F$ . Let  $d$  be the minimizing decision rule corresponding to a chosen  $F \in \mathcal{C}_M(K)$ . Suppose  $c_{j+1} < c_j$ . Let  $F'$  be a d.f. in  $\mathcal{C}_M(K)$  with  $u'_i = u_i$  for all  $i$  and  $c'_i = c_i$  for  $i \neq j, j+1$ . Assume  $c'_j = c'_{j+1}$ . Let  $d'$  be a decision rule such that  $P_1(d', F') = P_1(d, F)$  and  $\inf_r P_0(r, F') = P_0(d', F')$ . Note that  $\alpha'_j = \alpha'_{j+1}$ , where  $\alpha'_i$ ,  $i = 1, \dots, M$ , correspond to  $d'$  [See (44)]. Since  $F(u_{j-1}, u_{j+1}] = F'(u_{j-1}, u_{j+1}]$ , we have  $P_0(d, F) < P_0(d', F) = P_0(d', F')$ . Thus  $(d, F)$  cannot be a maximin pair.  $\square$

Now consider the interval  $(u_{j-1}, u_{j+1}]$ , which is the union of the two intervals  $(u_{j-1}, u_j]$  and  $(u_j, u_{j+1}]$ . For now, we want to optimize over only the parameters  $u_j$ ,  $c_j$ , and  $c_{j+1}$ , while keeping the other parameters fixed. Since  $f$  is a probability density, we must have

$$\int f(x) dx = \sum_{i=1}^M c_i F_1(u_{i-1}, u_i] = 1. \quad (45)$$

Furthermore, the lower bound on average power requires that

$$\int x^2 f(x) dx = \sum_{i=1}^M c_i G_1(u_{i-1}, u_i] \geq E_{\min}. \quad (46)$$

If we solve (45) for  $c_{j+1}$  and substitute into (46), we obtain the bound

$$c_j \leq \ell_1(u_j) \triangleq \frac{UG_1(u_j, u_{j+1}] - VF_1(u_j, u_{j+1})}{F_1(u_{j-1}, u_j]G_1(u_j, u_{j+1}] - F_1(u_j, u_{j+1})G_1(u_{j-1}, u_j]} \quad (47)$$

with

$$U \triangleq 1 - \sum_{i \neq j-1, j} c_i F_1(u_{i-1}, u_i]$$

$$V \triangleq E_{\min} - \sum_{i \neq j-1, j} c_i G_1(u_{i-1}, u_i].$$

Note that with all the other parameters fixed,  $U$  represents the portion of the probability mass assigned to the interval  $(u_{j-1}, u_{j+1}]$ , while  $V$  represents the contribution from  $(u_{j-1}, u_{j+1}]$  to the expected power. If we solve (45) for  $c_j$  and substitute into (46), we obtain the bound

$$c_{j+1} \geq \ell_2(u_j) \triangleq \frac{V F_1(u_{j-1}, u_j] - U G_1(u_{j-1}, u_j]}{F_1(u_{j-1}, u_j] G_1(u_j, u_{j+1}] - F_1(u_j, u_{j+1}] G_1(u_{j-1}, u_j]}. \quad (48)$$

Note that the denominator in  $\ell_1$  and  $\ell_2$  is always positive. The inequality  $\ell_1 < \ell_2$  is equivalent to the inequality

$$\frac{V}{U} > \frac{G_1(u_{j-1}, u_{j+1}]}{F_1(u_{j-1}, u_{j+1}]} \quad (49)$$

which is analogous to (37). The inequality (37) is global, while (49) is local.

Equality holds in (47) if and only if inequality holds in (48), and as  $E_{\min}$  increases with  $u_j$  fixed,  $\ell_1$  decreases and  $\ell_2$  increases. Note that  $\ell_1(u_j) \rightarrow 0$  as  $u_j$  moves to the left, and  $\ell_2(u_j) \rightarrow +\infty$  as  $u_j$  moves to the right. These constraints quantify the idea that probability mass must be moved to the right in order to meet the constraint on expected power.

LEMMA 2. If (49) holds, then  $\ell_1(u_j)$  and  $\ell_2(u_j)$  are increasing functions of  $u_j$ .

*Proof.* The derivatives of  $\ell_1$  and  $\ell_2$  are

$$\frac{d\ell_1}{du_j} = \frac{f_1(u) \{V F_1(u_{j-1}, u_{j+1}] - U G_1(u_{j-1}, u_{j+1}]\} \{G_1(u_j, u_{j+1}] - u_j^2 F_1(u_j, u_{j+1}]\}}{|F_1(u_{j-1}, u_j] G_1(u_j, u_{j+1}] - F_1(u_j, u_{j+1}] G_1(u_{j-1}, u_j)]|^2} \quad (50)$$

$$\frac{d\ell_2}{du_j} = \frac{f_1(u) \{V F_1(u_{j-1}, u_{j+1}] - U G_1(u_{j-1}, u_{j+1}]\} \{u_j^2 F_1(u_{j-1}, u_j] - G_1(u_{j-1}, u_j]\}}{|F_1(u_{j-1}, u_j] G_1(u_j, u_{j+1}] - F_1(u_j, u_{j+1}] G_1(u_{j-1}, u_j)]|^2}. \quad (51)$$

It is easy to see that all of the factors in the derivatives are positive; therefore the derivatives are positive.  $\square$

If we choose our interval  $(u_{j-1}, u_{j+1}]$  appropriately and optimize over  $u_j$ ,  $c_j$ , and  $c_{j+1}$ , then we get either  $u_j$  increasing, with  $c_{j+1} \rightarrow c_{j+2}$ , or  $u_j$  decreasing, with  $c_j \rightarrow c_{j-1}$ , thus reducing the number of intervals from  $M$  to  $M - 1$ . This is the gist of Lemma 3.

LEMMA 3. If  $F_0 \in \mathcal{C}_M(K)$ ,  $M > 3$ , and  $F_0 \notin \mathcal{C}_{M-1}(K)$ , then  $F_0$  is not maximin in  $\mathcal{C}_M(K)$ .

*Proof.* We prove the contrapositive of the statement, that is, if  $F_0$  is not in  $\mathcal{C}_{M-1}(K)$ , then  $F_0$  cannot be maximin. Our proof will consist of constructing a distribution  $F' \in \mathcal{C}_{M-1}(K)$  such that  $P_1(d, F_0) = P_1(d', F')$  and  $P_0(d, F_0) \leq P_0(d', F')$ , where  $d$  and  $d'$  are the minimizing decision rules for  $F_0$  and  $F'$ , respectively. The construction involves moving  $u_j$  to the right or left while keeping  $u_i$ ,  $i \neq j$ , and  $c_i$ ,  $i \neq j, j+1$ , fixed. The parameters  $c_j$  and  $c_{j+1}$  are allowed to change.

Suppose that  $F_0 \notin \mathcal{C}_{M-1}(K)$ , which means that  $c_1 < c_2 < \dots < c_M$ . The LRT for testing  $F_0$  against  $F_1$  is given by (44) with each  $\alpha_i$  equal to 0,  $\alpha$ , or 1. Since the  $c_i$ s are all distinct, only one of the  $\alpha_i$ s can be equal to  $\alpha$ , say  $\alpha_k$ . We may assume without loss of generality that  $c_k > 0$ . (Otherwise, if  $c_k = 0$ , we must have  $\alpha = \alpha_k = 1$ . Take  $\alpha = \alpha_{k+1} = 0$ .) The case of  $k \geq M-1$  must be treated as a special case; for now assume that  $k < M-1$ . The error probabilities are given by

$$P_0 = \alpha c_k F_1(u_{k-1}, u_k] + \sum_{i=1}^{k-1} c_i F_1(u_{i-1}, u_i] \quad (52)$$

$$P_1 = F_1(u_{k-1}, A_{\max}] - \alpha F_1(u_{k-1}, u_k]. \quad (53)$$

If we solve (53) for  $\alpha$ ,

$$\alpha = \frac{F_1(u_{k-1}, A_{\max}] - P_1}{F_1(u_{k-1}, u_k]}, \quad (54)$$

and substitute into (52), we have

$$P_0 = c_k \{F_1(u_{k-1}, A_{\max}] - P_1\} + \sum_{i=1}^{k-1} c_i F_1(u_{i-1}, u_i] \quad (55)$$

which allows us to see the effects of the parameters on  $P_0$  while  $P_1$  is fixed. Suppose  $\alpha > 0$ . Then from (54) we know that the coefficient of  $c_k$  in (55) is positive. Therefore,  $P_0$  is as large as possible only if  $c_k$  is as large as possible, which means that  $c_k = \ell_1(u_k)$  and  $c_{k+1} = \ell_2(u_k)$ . Furthermore, if we increase  $u_k$  while keeping  $P_1$  fixed, then  $c_k = \ell_1(u_k)$ ,  $c_{k+1} = \ell_2(u_k)$ , and  $P_0$  all increase. Construct  $F'$  from  $F_0$  by moving  $u'_k$  to the right (from  $u_k$ ), keeping  $c'_k = \ell_1(u'_k)$  and  $c'_{k+1} = \ell_2(u'_k)$ , until  $c'_{k+1} = c'_{k+2}$ . If  $\alpha = 0$ , then the coefficient of  $c_k$  in (55) is zero. Therefore, we may construct  $F'$  as before, the difference being that  $P_1(d, F) = P_1(d', F')$  and  $P_0(d, F_0) = P_0(d', F')$ . By our assumption concerning uniqueness,  $F_0$  cannot be maximin in  $\mathcal{C}_M(K)$ .

Now consider the case where  $k \geq M-1$ . Using

$$\sum_{i=1}^{k-1} c_i F_1(u_{i-1}, u_i] = 1 - \sum_{i=k}^M c_i F_1(u_{i-1}, u_i] \quad (56)$$

in (55), we have

$$P_0 = c_k \{F_1(u_k, A_{\max}) - P_1\} + 1 - \sum_{i=k+1}^M c_i F_1(u_{i-1}, u_i]. \quad (57)$$

Suppose  $\alpha < 1$ . Then from (54) the coefficient of  $c_k$  in (57) is negative. Thus if we decrease  $u_{k-1}$  while keeping  $P_1$  fixed, then  $c_{k-1} = \ell_1(u_{k-1})$  and  $c_k = \ell_2(u_{k-1})$  both decrease while  $P_0$  increases. Construct  $F'$  from  $F_0$  by moving  $u'_{k-1}$  to the left (from  $u_{k-1}$ ), keeping  $c'_{k-1} = \ell_1(u'_{k-1})$  and  $c'_k = \ell_2(u'_{k-1})$ , until  $c'_{k-1} = c'_{k-2}$ . If  $\alpha = 1$ , then the coefficient of  $c_k$  is zero. Therefore, we may construct  $F'$  as before. In either case, we conclude that  $F_0$  is not maximin in  $\mathcal{C}_M(K)$ .  $\square$

LEMMA 4. Let  $F_0$  be the maximin distribution in  $\mathcal{C}_M(K)$ . Then  $F_0$  takes one of these three forms: (a)  $F_0 \in \mathcal{C}_3(K)$ , with  $c_1 = 0$  and  $c_3 = K$ ; (b)  $F_0 \in \mathcal{C}_2(K)$ , with  $c_2 = K$ ; or (c)  $F_0 \in \mathcal{C}_2(K)$ , with  $c_1 = 0$ .

*Proof.* From Lemma 3 we know that  $F_0 \in \mathcal{C}_3(K)$ . The minimizing decision rule has the form (44). If  $\alpha_1 = 0$ , then  $\alpha_2 = \alpha_3 = 0$ . If  $\alpha_3 = 1$ , then  $\alpha_1 = \alpha_2 = 1$ . In both cases, the performance does not depend on the density involved. Thus we have to consider the cases (i)  $0 < \alpha_1 < 1$ ; (ii)  $\alpha_1 = 1$ ,  $\alpha_2 = \alpha_3 = 0$ ; (iii)  $0 < \alpha_2 < 1$ ; (iv)  $\alpha_1 = \alpha_2 = 1$ ,  $\alpha_3 = 0$ ; and (v)  $0 < \alpha_3 < 1$ .

Case (i). The error probabilities are given by (55) with  $j = 1$ . Since the coefficient of  $c_1$  is positive, we may increase  $P_0$ , while keeping  $P_1$  fixed, by increasing  $c_1$  and  $u_1$ . The result is a density of the form (b).

Case (ii). As long as  $u_1$  and  $c_1$  remain fixed,  $P_0$  and  $P_1$  will not change. Therefore, we may take  $c_3 = K$ , moving  $u_2$  to the right if necessary. This results in a density of the form (b).

Case (iii). The error probabilities are given by either (55) or (57) with  $j = 2$ . Since the coefficient of  $c_2$  in (55) is positive, we may increase  $P_0$ , while keeping  $P_1$  fixed, by increasing  $c_2$  and  $u_2$ . Similarly, since the coefficient of  $c_2$  in (57) is negative, we may increase  $P_0$ , while keeping  $P_1$  fixed, by decreasing  $c_2$  and  $u_1$ . We may not increase  $P_0$  if  $c_1 = 0$  ( $u_1$  cannot be moved any farther to the left) or if  $c_3 = K$  ( $u_2$  cannot move any farther to the right). Thus we have a density of the form (a).

Case (iv). As long as  $u_2$  and  $c_3$  remain fixed,  $P_0$  and  $P_1$  will not change. Therefore, we may take  $c_1 = 0$ , moving  $u_1$  to the left if necessary. This results in a density of the form (c).

Case (v). The error probabilities are given by (57) with  $j = 3$ . Since the coefficient of  $c_3$  is negative, we may increase  $P_0$ , while keeping  $P_1$  fixed, by decreasing  $c_3$  and  $u_2$ . This results in a density of the form (c).  $\square$

We are now ready to remove the restriction that  $c_i \leq K$ . Define

$$\mathcal{C}_M \triangleq \overline{\bigcup_{K \geq 1} \mathcal{C}_M(K)}$$

where the bar denotes the closure. By taking the closure of the union, we allow point masses in the distributions (delta functions in the densities).

*Proof of Theorem 2.*

Let  $F_{0,K}$  denote the maximin distribution in  $\mathcal{C}_M(K)$ , and  $S(F)$  the minimum value of the performance measure when testing  $F$  against  $F_1$ . Certainly  $S(F_{0,K}) \leq S(F_{0,L})$  for  $K \leq L$ . Thus  $\{S(F_{0,K})\}_{K=1}^{\infty}$  is a bounded increasing sequence which has a limit  $S^*$ . If  $F_0$  is any limit point of the sequence  $\{F_{0,K}\}_{K=1}^{\infty}$ , then by Lemma 3 it must have one of the three forms (a), (b), or (c), as stated in the theorem. Let  $F$  be any other distribution in  $\mathcal{C}_M$  which is a candidate for being the maximin distribution. There must be a sequence  $\{F_K\}$  of distributions, with  $F_K \in \mathcal{C}_M(K)$ , which converges to  $F$ . Since  $S(F_K) \leq S(F_{0,K})$  for each  $K$ , and since  $S$  is continuous, it follows that the inequality holds in the limit, too. Thus  $S(F) \leq S(F_0) = S^*$ . By our assumption concerning uniqueness,  $F$  is maximin in  $\mathcal{C}_M$  only if it is “minimal,” that is,  $F$  has one of the forms (a), (b), or (c).  $\square$

## VI. EXAMPLES

We mentioned briefly in Section II that the purpose of the lower bound constraints on the spurious signal is to increase the average or minimum power received by the detector. One of the prime concerns of the designer of the system that generates the spurious signal is the trade-off between average the power received by the detector and the probability of error of the detector. It is desired to generate the strongest signal possible (up to the physical limit of the transmitter) while keeping the detector error probability as high as possible. When the deceptive signal source generates continuously at peak transmitter power (a constant signal), an intelligently designed detector can discriminate with zero error probability. On the other hand, if no lower bound constraint is placed on the power of the transmitted deceptive signal, the detector error probability is relatively high, though not equal to 1 if the upper bound on the deceptive signal is lower than the maximum possible amplitude of the genuine signal. We now present briefly some results, which compare the performance of the deceptive signal under some of the distributions given in earlier sections with the average power/error probability trade-off shown explicitly. The error probabilities obtained in the simulation results are those of the optimal matched detectors.

We assume that the genuine signal has a lognormal distribution. Thus we have  $X_i = \exp(\sigma Y_i + \mu)$ , where  $\{X_i\}$  is the observed genuine signal, and we assume  $\{Y_i\}$  to be a first-order Gaussian autoregressive process with

$$Y_{i+1} = \rho Y_i + \sqrt{1 - \rho^2} \epsilon_i \quad (58)$$

where  $\{\epsilon_i\}$  is an i.i.d.  $N(0, 1)$  sequence. The parameters are  $\mu = 0.8$ ,  $\sigma^2 = 0.25$ , and  $\rho = 0.9$  for the correlated case and  $\rho = 0$  for the i.i.d. case. We take  $A_{\max} = 4$ , which is the 88th percentile of the genuine distribution. Thus, the maximum power which may be transmitted by the deceptive signal source is  $A_{\max}^2 = 16$ . In the included figures, we normalize the average deceptive signal power by dividing by the maximum power.

Performance results for the i.i.d. case are shown in Figure 5. The detector employed in each case is given by Theorem 1. The deceptive signal actually generated was the one described in the last part of Section 3 for a first order AR process. The ordinates of the plots show the common value of the error probabilities  $P_0 = P_1$ , obtained by properly choosing the randomization parameter  $\alpha$  in the test (13). The abscissas show the ratio of the average power actually transmitted to the maximum possible power  $A_{\max}^2$ . The two plots represent two different lengths for the window over which the power is averaged [see Eqn. (10)]. The case of  $N = 1$  involves a lower-bound constraint on the signal amplitude. The curves are parameterized by the value of the lower bound  $D$ . The plots show that the longer window length improves the deceptive signals performance significantly. Recall that the signal with  $N = 1$  is i.i.d., while the other is not.

Figure 6 shows a similar situation, but with the genuine signal being correlated. Comparing Figures 5 and 6 reveals that it is harder for the detector to correctly distinguish the signals when the genuine signal is correlated. As before, increasing the window length improves the performance of the deceptive signal.

Finally, we look at the performance of the density (36) obtained under an expected power constraint. For a single sample, the optimal decision rule (LRT) is given by (37), and value of the performance measure  $\max(P_0, P_1)$  is given by (39). For multiple i.i.d. samples, the performance measure under the optimal detector is similar to (39) and is given by

$$Q^{(n)}(u_1, E_{\min}) \triangleq \frac{(A_{\max}^2 - E_{\min})^n F_1(u_1, A_{\max})^n}{(A_{\max}^2 F_1(u_1, A_{\max}) - G_1(u_1, A_{\max}))^n + (A_{\max}^2 - E_{\min})^n} \quad (59)$$

where  $n$  is the sample size. Figure 7 shows the value of  $Q^{(n)}$ , with  $u_1$  chosen optimally, plotted as a function of  $E_{\min}/A_{\max}^2$ ; this is shown by the curve labeled EP. The other curves are the same as those shown in Figure 5. The univariate density (36) obtained under an expected power constraint



and used to generate an i.i.d. signal shows a significantly better performance than the densities which were derived under a lower bound on the time-averaged power.

## VII. EXTENSIONS OF THE RESULTS

One of the key factors necessary to obtain the results in Sections VI and V is the fact the the number of parameters to be dealt with is minimal. When we try to extend the results to a dimension larger than 1 or to densities which admit positive as well as negative values, we find that the number of parameters to optimize over becomes significantly larger. The two constraints available to us—that the density integrate to 1 and that the expected power be equal to  $E_{\min}$ —allow us to reduce the number of independent parameters by at most two.

Consider first the extension to a dimensionality greater than 1, say 2. If we divide the region  $(0, A_{\max}]^2$  into four regions, by taking two break points in each dimension, for example, then we have a total of six parameters to optimize over. Using the constraints we have available, we may reduce this to four independent parameters. However, if we take three break points per dimension, then we have eleven independent parameters to optimize over.

There are other problems associated with a higher dimensionality. Unless the density is fairly simple, generating a signal which has the distribution defined by the multivariate density may be too difficult to be practical. Furthermore, we would not necessarily be achieving our primary goal of obtaining a *stationary* signal which is effective at deceiving the detector while maintaining an increased power level. What we would actually have, if we obtain an optimal  $n$ -variate density, is a method to generate optimal blocks of  $n$ .

One extension is worth mentioning, however. Suppose the signal dimension  $m$  is greater than 1. If the genuine signal is i.i.d., then we may generate blocks of  $m$  independently from an optimal (or nearly optimal)  $m$ -variate density, if such a density can be found. That the result would be good may be expected, considering the comparison of the i.i.d. signals in the preceding section. Now, suppose that we begin with a finite dimensional problem, as in Section VI, where the regions in  $(0, A_{\max}]^m$  are defined by the parameters  $0 = u_0 < u_1 < \dots < u_M = A_{\max}$  as

$$R_i = \{\mathbf{x} : x_j > u_{i-1} \text{ for } 1 \leq j \leq n, \text{ and } x_j \leq u_i \text{ for at least one } x_j\}$$

for  $i = 1, \dots, M$ . In this case, the lemmas and theorems in Sections VI and V go through as before, the final result being a point mass at  $(A_{\max}, \dots, A_{\max})$ , a scaled region over  $R_2$ , and a zero region over  $R_1$ . A signal with this density may be easily generated.

Finally, we consider the extension to densities with positive and negative support. It is clear that, for nonsymmetric densities, we may approximate the problem by a finite dimensional one, as in Section IV. If we fix the amount of mass provided for positive values, then we can apply Theorems 2 and 3 to the “positive” and “negative” parts of the density to obtain a density of the form [compare (36)]

$$f_0(x) = c_2^- I_{(A_{\max}^-, u_1^-)}(x) f_1(x) + h^- \delta(A_{\max}^- - x) + c_2^+ I_{(u_1^+, A_{\max}^+)}(x) f_1(x) + h^+ \delta(A_{\max}^+ - x) \quad (60)$$

where the first two terms on the right hand side correspond to the “negative” part and the last two to the “positive” part. Recall that (36) involves three parameters,  $c_2$ ,  $u_1$ , and  $h$ . Using the two available constraints, we reduce this to one independent parameter to optimize over. In (60), however, we have six parameters, and we must optimize over four independent ones. In the case where  $f_1$  is symmetric,  $f_0$  is symmetric, and we have only one independent parameter to optimize over.

## VIII. CONCLUSION

In this paper, we have formulated a detection game which pits a detector designer against a deceptive signal designer. Different constraints on the probability distribution governing the deceptive signal were imposed. In the case of a lower bound on the time-averaged power of the signal, we showed that if the detector designer knows the parameters of the constraints on the deceptive signal, then the optimal detector and optimal distribution form a saddle point of the detection game. We provided both the optimal detector and the optimal distribution. Given a lower bound on the expected power, rather than the time-averaged power, we were unable to obtain the optimal detector, but provided the optimal distribution for the case where the signal dimension is one and the length of the data segment is one.

The results for the time-averaged power constraint are fairly complete. They are also the very practical. We presented a method for generating the optimal deceptive signal and an approximately optimal deceptive signal. It was shown that even if the genuine signal is i.i.d., the optimal deceptive signal is not i.i.d. in general. The exception to this occurs when the constraints are imposed on the signal amplitude only.

For the case of a lower-bound on the expected power of the deceptive signal, the game does not have a saddle point. We obtained the optimal density for the case of a univariate density and showed that the result is applicable in a situation where the genuine signal is i.i.d. Our numerical

results which compare the signals obtained under a time-averaged power constraint and an expected power constraint show that the latter performs significantly better.

We discussed the possibility of extending our results. We showed that it is possible to obtain an  $m$ -variate density which resembles the optimal univariate density, and suggested that such a density be used to generate an i.i.d. signal having dimension  $m$ . We also showed that the results for an expected power constraint are valid for nonsymmetric densities which allow both positive and negative values.

Our results, particularly those in Section III, assume that the detector designer knows the parameters of the constraints placed on the deceptive signal. This assumption is unrealistic and pessimistic from the viewpoint of the deceptive signal designer. One area for further research could be to relax or remove this assumption. For example, the detector designer may presuppose a probability distribution on the parameters, and thereby pursue a Bayesian approach to the detector design, or, if no assumption is made concerning the parameters, a minimax approach may be pursued. In either case, we expect that the research will lead into the area of *mixed strategies*, where a player chooses his strategy according to a probability distribution. In the case of an expected power constraint on the signal, where there is no saddle point, it seems that mixed strategies may also be appropriate.

## REFERENCES

- [1] J. M. Danskin, *The Theory of Max-Min*, New York: Springer-Verlag, 1967.
- [2] M. V. Hegde, W. E. Stark, and D. M. Teneketzis, "On the Capacity of Channels with Unknown Interference," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 770–783, July 1989.
- [3] L. C. Thomas, *Games, Theory and Applications*, New York: Wiley, 1984.
- [4] H. V. Poor, *An Introduction to Signal Detection and Estimation*, New York: Springer-Verlag, 1988.

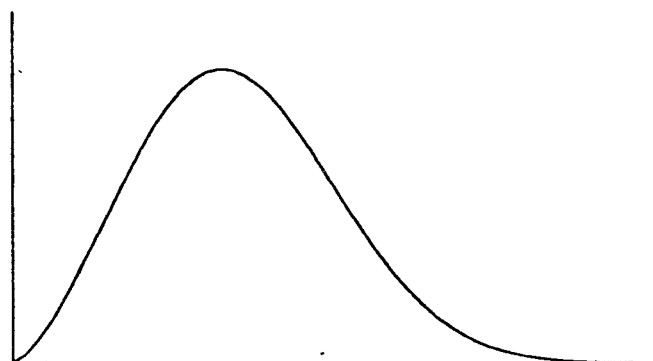


Figure 1. Graph of  $f_1$ .

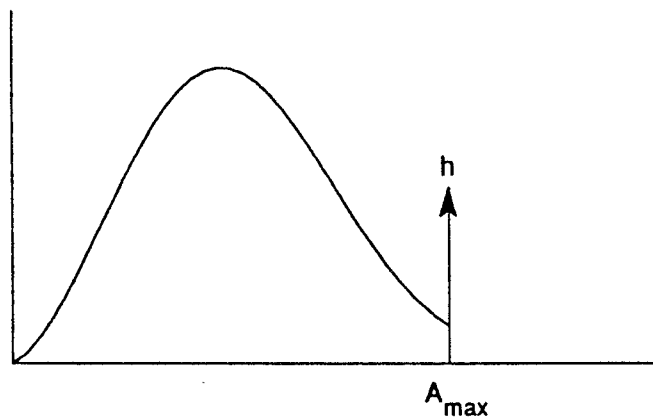


Figure 2. Graph of  $f_0$  given by Eqn. (23).

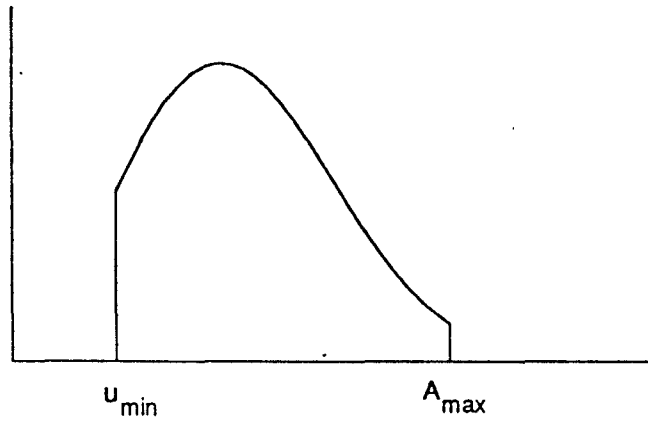


Figure 3. Graph of  $f_0$  given by Eqn. (24).

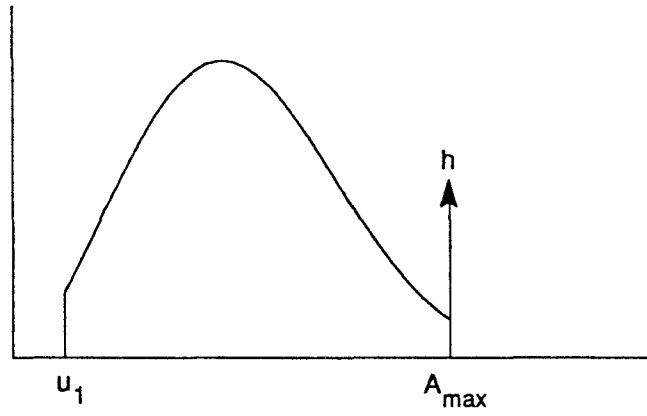


Figure 4. Graph of  $f_0$  given by Eqn. (36).

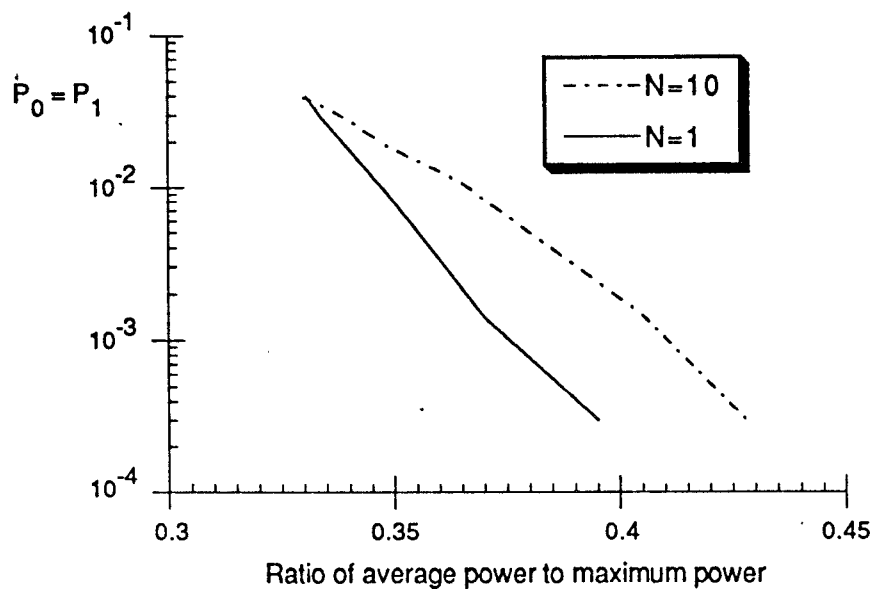


Figure 5. Plot of the common value of the error probabilities vs. average power for i.i.d. signals.

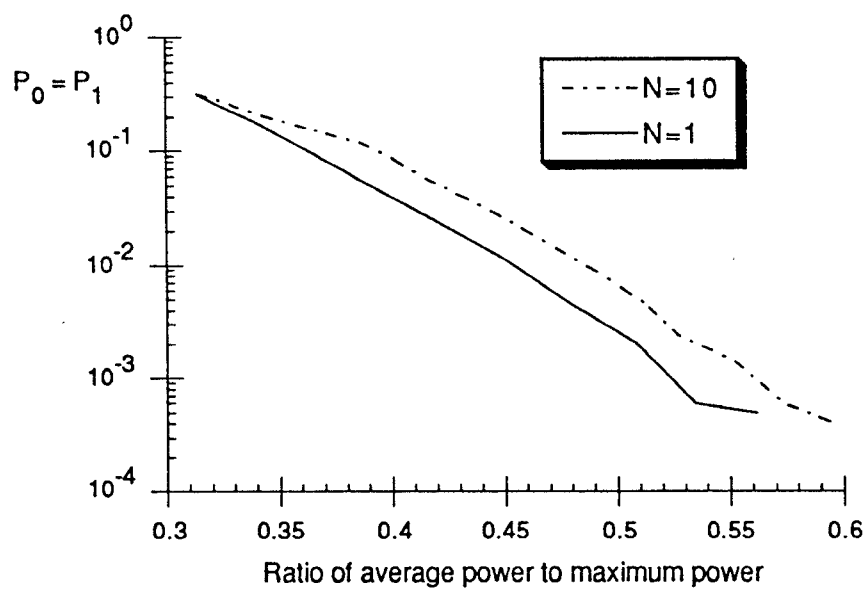


Figure 6. Plot of the common value of the error probabilities vs. average power for correlated signals.

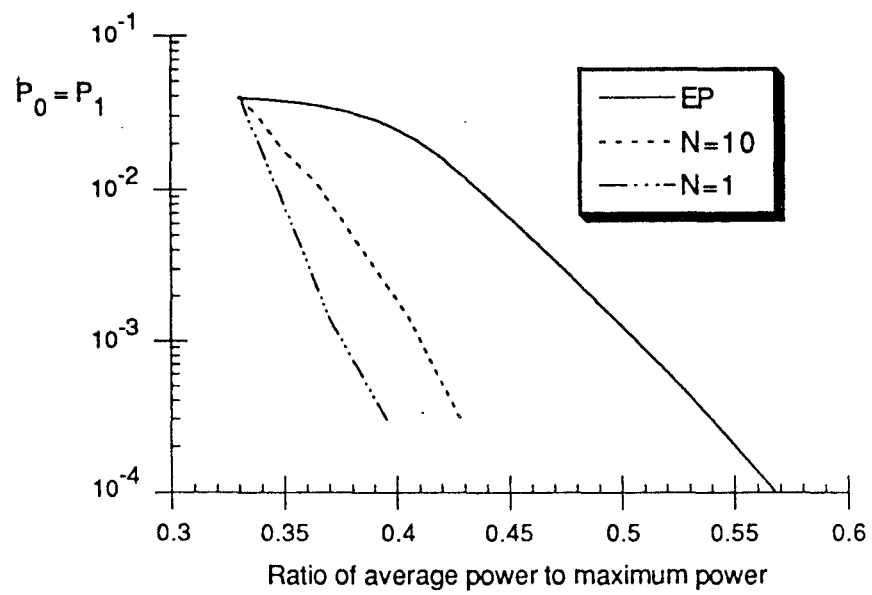


Figure 7. Plot of the common value of the error probabilities vs. average power for i.i.d. signals.





