




Optimal Decision Rules for Simple Hypothesis Testing Under General Criterion Involving Error Probabilities

Berkan Dulek , Cuneyd Ozturk , *Student Member, IEEE*, and Sinan Gezici , *Senior Member, IEEE*

Abstract—The problem of simple M -ary hypothesis testing under a generic performance criterion that depends on arbitrary functions of error probabilities is considered. Using results from convex analysis, it is proved that an optimal decision rule can be characterized as a randomization among *at most* two deterministic decision rules, each of the form reminiscent to Bayes rule, if the boundary points corresponding to each rule have zero probability under each hypothesis. Otherwise, a randomization among *at most* $M(M-1)+1$ deterministic decision rules is sufficient. The form of the deterministic decision rules are explicitly specified. Likelihood ratios are shown to be sufficient statistics. Classical performance measures including Bayesian, minimax, Neyman-Pearson, generalized Neyman-Pearson, restricted Bayesian, and prospect theory based approaches are all covered under the proposed formulation. A numerical example is presented for prospect theory based binary hypothesis testing.

Index Terms—Hypothesis testing, optimal tests, convexity, likelihood ratio, randomization.

I. PROBLEM STATEMENT

CONSIDER a detection problem with M simple hypotheses:

$$\mathcal{H}_j : \mathbf{Y} \sim f_j(\cdot), \text{ with } j = 0, 1, \dots, M-1, \quad (1)$$

where the random observation \mathbf{Y} takes values from an observation set Γ with $\Gamma \subset \mathbb{R}^N$. Depending on whether the observed random vector $\mathbf{Y} \in \Gamma$ is continuous-valued or discrete-valued, $f_j(\mathbf{y})$ denotes either the probability density function (pdf) or the probability mass function (pmf) under hypothesis \mathcal{H}_j . For compactness of notation, the term *density* is used for both pdf and pmf. In order to decide among the hypotheses, we consider the set of pointwise randomized decision functions, denoted by \mathbf{D} , i.e., $\boldsymbol{\delta} := (\delta_0, \delta_1, \dots, \delta_{M-1}) \in \mathbf{D}$ such that $\sum_{i=0}^{M-1} \delta_i(\mathbf{y}) = 1$ and $\delta_i(\mathbf{y}) \in [0, 1]$ for $0 \leq i \leq M-1$ and $\mathbf{y} \in \Gamma$. More explicitly, given the observation \mathbf{y} , the detector decides in favor of hypothesis \mathcal{H}_i with probability $\delta_i(\mathbf{y})$. Then, the probability of

choosing hypothesis \mathcal{H}_i when hypothesis \mathcal{H}_j is true, denoted by p_{ij} with $0 \leq i, j \leq M-1$, is given by

$$p_{ij} := \mathbb{E}_j[\delta_i(\mathbf{y})] = \int_{\Gamma} \delta_i(\mathbf{y}) f_j(\mathbf{y}) \mu(d\mathbf{y}), \quad (2)$$

where $\mathbb{E}_j[\cdot]$ denotes expected value under hypothesis \mathcal{H}_j and $\mu(d\mathbf{y})$ is used in (2) to denote the N -fold integral and sum for continuous and discrete cases, respectively. Let $\mathbf{p}(\boldsymbol{\delta})$ denote the (column) vector containing all pairwise error probabilities p_{ij} for $0 \leq i, j \leq M-1$ and $i \neq j$ corresponding to the decision rule $\boldsymbol{\delta}$. It is sufficient to include only the pairwise error probabilities in $\mathbf{p}(\boldsymbol{\delta})$, i.e., p_{ij} with $i \neq j$. To see this, note that (2) in conjunction with $\sum_{i=0}^{M-1} \delta_i(\mathbf{y}) = 1$ imply $\sum_{i=0}^{M-1} p_{ij} = 1$, from which we get the probability of correctly identifying hypothesis \mathcal{H}_i as $p_{ii} = 1 - \sum_{i=0, i \neq j}^{M-1} p_{ij}$.

For M -ary hypothesis testing, we consider a generic decision criterion that can be expressed in terms of the error probabilities as follows:

$$\begin{aligned} & \underset{\boldsymbol{\delta} \in \mathbf{D}}{\text{minimize}} && g_0(\mathbf{p}(\boldsymbol{\delta})) \\ & \text{subject to} && g_i(\mathbf{p}(\boldsymbol{\delta})) \leq 0, \quad i = 1, 2, \dots, m \\ & && h_j(\mathbf{p}(\boldsymbol{\delta})) = 0, \quad j = 1, 2, \dots, p \end{aligned} \quad (3)$$

where g_i and h_j denote arbitrary functions of the pairwise error probability vector. Classical hypothesis testing criteria such as Bayesian, minimax, Neyman-Pearson (NP) [1], generalized Neyman-Pearson [2], and restricted Bayesian [3] are all special cases of the formulation in (3). For example, in the restricted Bayesian framework, the Bayes risk with respect to (w.r.t.) a certain prior is minimized subject to a constraint on the maximum conditional risk [3]:

$$\begin{aligned} & \underset{\boldsymbol{\delta} \in \mathbf{D}}{\text{minimize}} && r_B(\boldsymbol{\delta}) \\ & \text{subject to} && \max_{0 \leq j \leq M-1} R_j(\boldsymbol{\delta}) \leq \alpha \end{aligned} \quad (4)$$

for some $\alpha \geq \alpha_m$, where α_m is the maximum conditional risk of the minimax procedure [1]. The conditional risk when hypothesis \mathcal{H}_j is true, denoted by $R_j(\boldsymbol{\delta})$, is given by $R_j(\boldsymbol{\delta}) = \sum_{i=0}^{M-1} c_{ij} p_{ij}$ and the Bayes risk is expressed as $r_B(\boldsymbol{\delta}) = \sum_{j=0}^{M-1} \pi_j R_j(\boldsymbol{\delta})$, where π_j denotes the *a priori* probability of hypothesis \mathcal{H}_j and c_{ij} is the cost incurred by choosing hypothesis \mathcal{H}_i when in fact hypothesis \mathcal{H}_j is true. Hence, (4) is a special case of (3).

In this letter, we consider a generic M -ary simple hypothesis testing framework and do not make any specific assumptions on the employed optimization criterion expect that it can be specified using functions of error probabilities. Not only does

Manuscript received July 24, 2019; revised January 2, 2020; accepted January 8, 2020. Date of publication January 13, 2020; date of current version February 12, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ashish Pandharipande. (*Corresponding author: Sinan Gezici.*)

B. Dulek is with the Department of Electrical and Electronics Engineering, Hacettepe University, Beytepe Campus, Ankara 06800, Turkey (e-mail: berkan@ee.hacettepe.edu.tr).

C. Ozturk and S. Gezici are with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey (e-mail: cuneyd@ee.bilkent.edu.tr; gezici@ee.bilkent.edu.tr).

Digital Object Identifier 10.1109/LSP.2020.2966330

this allow us to account for several classical performance criteria that are mentioned above but also to generalize prospect theory based approaches developed in [4] for behavioral (e.g., human) decision makers who may have a distorted view of probabilities and costs. Our approach to this problem is to characterize the set of all achievable pairwise error probabilities and the corresponding optimal decision rule that delivers any given feasible pairwise error probability vector. To that aim, we first specify the optimal decision rule that yields an extreme point of the set of all achievable pairwise error probabilities. Randomization is required only if the solution of the specific optimization problem under consideration occurs at an interior point or a boundary (but not an extreme) point of the feasible set. In this way, for the first time in the literature, we provide a unified characterization of optimal decision rules for simple hypothesis testing under a general criterion involving error probabilities.

II. PRELIMINARIES

Let \mathbf{v} be a real (column) vector of length $M(M-1)$ whose elements are denoted as v_{ij} for $0 \leq i, j \leq M-1$ and $i \neq j$. Next, we present an optimal deterministic decision rule that minimizes the weighted sum of p_{ij} 's with arbitrary real weights \mathbf{v} .¹

A. Optimal Decision Rule That Minimizes $\mathbf{v}^T \mathbf{p}(\delta)$

The corresponding weighted sum of pairwise error probabilities can be written as

$$\begin{aligned} \mathbf{v}^T \mathbf{p}(\delta) &= \sum_{i=0}^{M-1} \sum_{j=0, j \neq i}^{M-1} v_{ij} p_{ij} \\ &= \int_{\Gamma} \sum_{i=0}^{M-1} \delta_i(\mathbf{y}) \left(\sum_{j=0, j \neq i}^{M-1} v_{ij} f_j(\mathbf{y}) \right) \mu(d\mathbf{y}), \end{aligned} \quad (5)$$

where (2) is substituted for p_{ij} in (5). Defining $V_i(\mathbf{y}) := \sum_{j=0, j \neq i}^{M-1} v_{ij} f_j(\mathbf{y})$, we get

$$\begin{aligned} \mathbf{v}^T \mathbf{p}(\delta) &= \int_{\Gamma} \sum_{i=0}^{M-1} \delta_i(\mathbf{y}) V_i(\mathbf{y}) \mu(d\mathbf{y}) \\ &\geq \int_{\Gamma} \min_{0 \leq i \leq M-1} \{V_i(\mathbf{y})\} \mu(d\mathbf{y}) \end{aligned} \quad (6)$$

The lower bound in (6) is achieved if, for all $\mathbf{y} \in \Gamma$, we set

$$\delta_{\ell}(\mathbf{y}) = 1 \text{ for } \ell = \underset{0 \leq i \leq M-1}{\operatorname{argmin}} V_i(\mathbf{y}) \quad (7)$$

(and hence, $\delta_i(\mathbf{y}) = 0$ for all $i \neq \ell$), i.e., each observed vector \mathbf{y} is assigned to the corresponding hypothesis that minimizes $V_i(\mathbf{y})$ over all $0 \leq i \leq M-1$. In case where there are multiple hypotheses that achieve the same minimum value of $V_{\ell}(\mathbf{y})$ for a given observation \mathbf{y} , the ties can be broken by arbitrarily selecting one of them since the boundary decision does not affect the decision criterion $\mathbf{v}^T \mathbf{p}(\delta)$. However, pairwise probabilities for erroneously selecting hypotheses \mathcal{H}_i and \mathcal{H}_j will change if the set of boundary points

$$\begin{aligned} \mathbf{B}_{i,j}(\mathbf{v}) &:= \{\mathbf{y} \in \Gamma : V_i(\mathbf{y}) = V_j(\mathbf{y}) \leq V_k(\mathbf{y}) \\ &\text{for all } 0 \leq k \leq M-1, k \neq i, k \neq j\} \end{aligned} \quad (8)$$

occurs with nonzero probability. We also define the set of all boundary points

$$\mathbf{B}(\mathbf{v}) := \bigcup_{\substack{0 \leq i \leq M-1 \\ i < j \leq M-1}} \mathbf{B}_{i,j}(\mathbf{v}) \quad (9)$$

and the complimentary set where $V_i(\mathbf{y})$ for some $0 \leq i \leq M-1$ is strictly smaller than the rest:

$$\begin{aligned} \bar{\mathbf{B}}(\mathbf{v}) &:= \Gamma \setminus \mathbf{B}(\mathbf{v}) = \{\mathbf{y} \in \Gamma : V_i(\mathbf{y}) < V_j(\mathbf{y}), \text{ for some} \\ &0 \leq i \leq M-1 \text{ and all } 0 \leq j \leq M-1, j \neq i\} \end{aligned} \quad (10)$$

B. The Set of Achievable Pairwise Error Probability Vectors

Let \mathbf{P} denote the set of all pairwise error probability vectors that can be achieved by randomized decision functions $\delta \in \mathbf{D}$, i.e., $\mathbf{P} := \{\mathbf{p}(\delta) : \delta \in \mathbf{D}\}$. In this part, we present some properties of \mathbf{P} .

Property 1: \mathbf{P} is a convex set.

Proof: Let $\mathbf{p}^1(\delta^1)$ and $\mathbf{p}^2(\delta^2)$ be two pairwise error probability vectors obtained by employing randomized decision functions δ^1 and δ^2 , respectively. Then, for any θ with $0 \leq \theta \leq 1$, $\mathbf{p}_{\theta} = \theta \mathbf{p}^1(\delta^1) + (1-\theta) \mathbf{p}^2(\delta^2) \in \mathbf{P}$ since \mathbf{p}_{θ} is the pairwise error probability vector corresponding to the randomized decision rule $\theta \delta^1 + (1-\theta) \delta^2$ as seen from (2). ■

Property 2: Let \mathbf{p}_0 be a point on the boundary of \mathbf{P} . There exists a hyperplane $\{\mathbf{p} : \mathbf{v}^T \mathbf{p} = \mathbf{v}^T \mathbf{p}_0\}$ that is tangent to \mathbf{P} at \mathbf{p}_0 and $\mathbf{v}^T \mathbf{p} \geq \mathbf{v}^T \mathbf{p}_0$ for all $\mathbf{p} \in \mathbf{P}$.

Proof: Follows immediately from the supporting hyperplane theorem [5, Sec. 2.5.2]. ■

III. CHARACTERIZATION OF OPTIMAL DECISION RULE

In order to characterize the solution of (3), we first present the following lemma.

Lemma: Let \mathbf{p}_0 be a point on the boundary of \mathbf{P} and $\{\mathbf{p} : \mathbf{v}^T \mathbf{p} = \mathbf{v}^T \mathbf{p}_0\}$ be a supporting hyperplane to \mathbf{P} at the point \mathbf{p}_0 .

Case 1: Any deterministic decision rule of the form given in (7) corresponding to the weights specified by \mathbf{v} yields \mathbf{p}_0 if $\mathbf{B}(\mathbf{v})$, defined in (9), has zero probability under all hypotheses.

Case 2: \mathbf{p}_0 is achieved by a randomization among at most $M(M-1)$ deterministic decision rules of the form given in (7), all corresponding to the same weights specified by \mathbf{v} , if $\mathbf{B}(\mathbf{v})$, defined in (9), has nonzero probability under some hypotheses.

Proof: See Appendix A. ■

It should be noted that the condition in case 1 of the lemma, i.e., $\mathbf{B}(\mathbf{v})$ has zero probability under all hypotheses, is not difficult to satisfy. A simple example is when the observation under hypothesis \mathcal{H}_i is Gaussian distributed with mean μ_i and variance σ^2 for all $0 \leq i \leq M-1$. Furthermore, the lemma implies that any extreme point of the convex set \mathbf{P} , i.e., any point on the boundary of the convex set \mathbf{P} that is not a convex combination of any other points in the set, can be achieved by a deterministic decision rule of the form (7) without any randomization. The points that are on the boundary but not extreme points can be obtained via randomization as stated in case 2.

Next, we present a unified characterization of the optimal decision rule for problems that are in the form of (3). We suppose that the problem in (3) is feasible and let δ^* and $\mathbf{p}^*(\delta^*)$ denote an optimal decision rule and the corresponding pairwise error probabilities, respectively.

Theorem: An optimal decision rule that solves (3) can be obtained as

¹In classical Bayesian M -ary hypothesis testing, $v_{ij} = \pi_j(c_{ij} - c_{jj})$.

Case 1: A randomization among at most two deterministic decision rules of the form given in (7), each specified by some real v , if $B(v)$, defined in (9), has zero probability under all hypotheses for all real v ; otherwise

Case 2: A randomization among at most $M(M-1) + 1$ deterministic decision rules of the form given in (7), one specified by some real v and the remaining $M(M-1)$ correspond to the same weights specified by another real v .

Proof: If the optimal point $\mathbf{p}^*(\delta^*)$ is on the boundary of \mathbf{P} , then the lemma takes care of the proof. Here, we consider the case when $\mathbf{p}^*(\delta^*)$ is an interior point of \mathbf{P} . First, we pick an arbitrary $\mathbf{v}^1 \in \mathbb{R}^{M(M-1)}$ and derive the optimal deterministic decision rule according to (7). Let \mathbf{p}^1 denote the pairwise error probability vector corresponding to the employed decision rule. Then, we move along the ray that originates from \mathbf{p}^1 and passes through $\mathbf{p}^*(\delta^*)$. Since \mathbf{P} is bounded, this ray will intersect with the boundary of \mathbf{P} at some point, say \mathbf{p}^2 . If the condition in case 1 is satisfied, then by lemma-case 1, there exists a deterministic decision rule of the form given in (7) that yields \mathbf{p}^2 . Otherwise, by lemma-case 2, \mathbf{p}^2 is achieved by a randomization among at most $M(M-1)$ deterministic decision rules of the form given in (7), all sharing the same weight vector \mathbf{v}^2 . Since $\mathbf{p}^*(\delta^*)$ resides on the line segment that connects \mathbf{p}^1 to \mathbf{p}^2 , it can be attained by appropriately randomizing among the decision rules that yield \mathbf{p}^1 and \mathbf{p}^2 . ■

When the optimization problem in (3) possesses certain structure, the maximum number of deterministic decision rules required to achieve optimal performance may be reduced below those given in the theorem. For example, suppose that the objective is a concave function of \mathbf{p} and there are a total of n constraints in (3) which are all linear in \mathbf{p} (i.e., the feasible set, denoted by \mathbf{P}' , is the intersection of \mathbf{P} with halfspaces and hyperplanes). It is well known that the minimum of a concave function over a closed bounded convex set is achieved at an extreme point [5]. Hence, in this case, the optimal point \mathbf{p}^* is an extreme point of \mathbf{P}' . By Dubin's theorem [6], any extreme point of \mathbf{P}' can be written as a convex combination of $n+1$ or fewer extreme points of \mathbf{P} . Since any extreme point of \mathbf{P} can be achieved by a deterministic decision rule of the form (7), the optimal decision rule is obtained as a randomization among at most $n+1$ deterministic decision rules of the form (7). If there are no constraints in (3), i.e., $n=0$, the deterministic decision rule given in (7) is optimal and no randomization is required with a concave objective function.

An immediate and important corollary of the theorem is given below.

Corollary: Likelihood ratios are sufficient statistics for simple M -ary hypothesis testing under any decision criterion that is expressed in terms of arbitrary functions of error probabilities as specified in (3).

Proof: It is stated in the theorem that a solution of the generic optimization problem in (3) can be expressed in terms of decision rules of the form given in (7). These decision rules only involve comparisons among $V_i(\mathbf{y})$'s, which are linear w.r.t. the density terms $f_i(\mathbf{y})$'s. Normalizing $f_i(\mathbf{y})$'s with $f_0(\mathbf{y})$ and defining $L_i(\mathbf{y}) := f_i(\mathbf{y})/f_0(\mathbf{y})$, we see that an optimal decision rule that solves the problem in (3) depends on the observation \mathbf{y} only through the likelihood ratios. ■

IV. NUMERICAL EXAMPLES

In this section, numerical examples are presented by considering a binary hypothesis testing problem (i.e., $M=2$ in

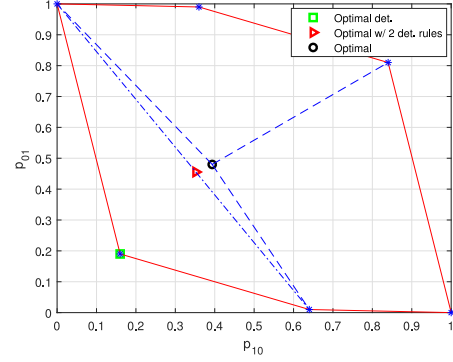


Fig. 1. Convex hull of pairwise error probability vectors corresponding to deterministic decision rules in (7), and pairwise error probability vectors corresponding to decision rules which yield the minimum objectives attained via no randomization (marked with square), randomization of two (marked with triangle) and three deterministic decision rules (marked with circle), where $p_{10}^{(1)} = p_{10}^{(2)} = 0.4$, $p_{01}^{(1)} = p_{01}^{(2)} = 0.1$, and $\kappa = 5$.

(1)) in order to illustrate the theoretical results. Suppose that a bit (0 or 1) is sent over two independent binary channels to a decision maker, which aims to make an optimal decision based on the binary channel outputs. The output of binary channel k is denoted by $y_k \in \{0, 1\}$, $k = 1, 2$, and the decision maker declares its decision based on $\mathbf{y} = [y_1, y_2]$. The probability that the output of binary channel k is i when bit j is sent is denoted by $p_{ij}^{(k)}$ for $0 \leq i, j \leq 1$ with $p_{0j}^{(k)} + p_{1j}^{(k)} = 1$. Then, the pmf of \mathbf{y} under \mathcal{H}_j is given by

$$f_j(\mathbf{y}) = p_{ij}^{(1)} p_{\ell j}^{(2)} \text{ if } \mathbf{y} = [i, \ell] \quad (11)$$

for $i, \ell \in \{0, 1\}$ and $j \in \{0, 1\}$. As in the previous sections, the pairwise error probability vector of the decision maker for a given decision rule δ is represented by $\mathbf{p}(\delta)$, which is expressed as $\mathbf{p}(\delta) = [p_{10}, p_{01}]^T$ in this case. It is assumed that the decision maker knows the conditional pdfs in (11).

In this section, a special case of (3) is considered based on prospect theory by focusing on a behavioral decision maker [4], [7]–[9]. In particular, there exist no constraints (i.e., $m = p = 0$ in (3)) and the objective function in (3) is expressed as

$$g_0(\mathbf{p}(\delta)) = \sum_{i=0}^1 \sum_{j=0}^1 w(P(\mathcal{H}_i \text{ is selected} \& \mathcal{H}_j \text{ is true})) v(c_{ij}) \quad (12)$$

where $w(\cdot)$ is a weight function and $v(\cdot)$ is a value function, which characterize how a behavioral decision maker distorts probabilities and costs, respectively [4], and $P(\cdot)$ denotes the probability of its argument. In the numerical examples, the following weight function is employed: $w(p) = \frac{p^\kappa}{(p^\kappa + (1-p)^\kappa)^{1/\kappa}}$ [4], [7]–[9]. In addition, the other parameters are set as $v(c_{00}) = 3$, $v(c_{01}) = 10$, $v(c_{10}) = 20$, and $v(c_{11}) = 7$. Furthermore, the prior probabilities of bit 0 and bit 1 are assumed to be equal.

The aim of the decision maker is to obtain a decision rule that minimizes (12). In the first example, κ is set to 5, and the parameters of the binary channels are selected as $p_{10}^{(1)} = p_{10}^{(2)} = 0.4$ and $p_{01}^{(1)} = p_{01}^{(2)} = 0.1$. In this case, it can be shown via (11) that there exist 6 different deterministic decision rules in the form of (7), which achieve the pairwise error probability vectors marked with blue stars in Fig. 1. The convex hull of these pairwise error probability vectors is also illustrated in the figure. Over these deterministic decision rules (i.e., in the absence of

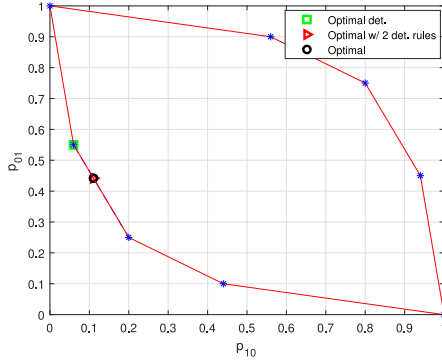


Fig. 2. Convex hull of pairwise error probability vectors corresponding to deterministic decision rules in (7), and pairwise error probability vectors corresponding to decision rules which yield the minimum objectives attained via no randomization (marked with square), randomization of two (marked with triangle) and three deterministic decision rules (marked with circle), where $p_{10}^{(1)} = 0.3$, $p_{10}^{(2)} = 0.2$, $p_{01}^{(1)} = 0.4$, and $p_{01}^{(2)} = 0.25$, and $\kappa = 1.5$.

randomization), the minimum achievable value of (12) becomes 0.1901, which corresponds to the pairwise error probability vector shown with the green square in Fig. 1. If randomization between two deterministic decision rules in the form of (7) is considered, the resulting minimum objective value becomes 0.0422, and the corresponding pairwise error probability vector is indicated with the red triangle in the figure. On the other hand, in compliance with the theorem (case 2), the minimum value of (12) is achieved via randomization of (at most) three deterministic decision rules in the form of (7) (since $M(M-1)+1=3$). In this case, the optimal decision rule randomizes among δ_1 , δ_2 , and δ_3 , with randomization coefficients of 0.41, 0.51, and 0.08, respectively, as given below:

$$\begin{aligned} \delta_1(\mathbf{y}) &= 0 \text{ for all } \mathbf{y} \\ \delta_2(\mathbf{y}) &= \begin{cases} 0, & \text{if } \mathbf{y} \in \{[0, 1], [1, 0], [1, 1]\} \\ 1, & \text{if } \mathbf{y} = [0, 0] \end{cases} \\ \delta_3(\mathbf{y}) &= \begin{cases} 0, & \text{if } \mathbf{y} = [1, 1] \\ 1, & \text{if } \mathbf{y} \in \{[0, 0], [0, 1], [1, 0]\} \end{cases} \end{aligned} \quad (13)$$

This optimal decision rule achieves the lowest objective value of 0.0400, and the corresponding pairwise error probability vector is marked with the black circle in Fig. 1. Hence, this example shows that randomization among three deterministic decision rules may be required to obtain the solution of (3).

In the second example, the parameters are taken as $\kappa = 1.5$, $p_{10}^{(1)} = 0.3$, $p_{10}^{(2)} = 0.2$, $p_{01}^{(1)} = 0.4$, and $p_{01}^{(2)} = 0.25$. In this case, there exist 8 different deterministic decision rules in the form of (7), which achieve the pairwise error probability vectors marked with blue stars in Fig. 2. The minimum value of (12) among these deterministic decision rules is 3.9278, which corresponds to the pairwise error probability vector shown with the green square in the figure. In addition, the pairwise error probability vectors corresponding to the solutions with randomization of two and three deterministic decision rules are marked with the red triangle and the black circle, respectively. In this scenario, the minimum objective value (3.8432) can be achieved via randomization of two deterministic decision rules, as well. This is again in compliance with the theorem (case 2), which states that an optimal decision rule can be obtained as a randomization among at most $M(M-1)+1$ deterministic decision rules of the form given in (7).

V. CONCLUDING REMARKS

This letter presents a unified characterization of optimal decision rules for simple M -ary hypothesis testing under a generic performance criterion that depends on arbitrary functions of error probabilities. It is shown that optimal performance with respect to the design criterion can be achieved by randomizing among at most two deterministic decision rules of the form reminiscent (but not necessarily identical) to Bayes rule when points on the decision boundary do not contribute to the error probabilities. For the general case, the solution for an optimal decision rule is reduced to a search over two weight coefficient vectors, each of length $M(M-1)$. Likelihood ratios are shown to be sufficient statistics.

Finally, we point out that the form of optimal local sensor decision rules for the problem of distributed detection [10]–[13] with conditionally independent observations at the sensors and an arbitrary fusion rule can be characterized using the proposed framework.

APPENDIX A PROOF OF LEMMA

Since $\{\mathbf{p} : \mathbf{v}^T \mathbf{p} = \mathbf{v}^T \mathbf{p}_0\}$ is a supporting hyperplane to \mathbf{P} at the point \mathbf{p}_0 , we get $\mathbf{v}^T \mathbf{p} \geq \mathbf{v}^T \mathbf{p}_0$ for all $\mathbf{p} \in \mathbf{P}$. Furthermore, the deterministic decision rule given in (7), denoted here by δ^* , minimizes $\mathbf{v}^T \mathbf{p}$ among all decision rules $\delta \in \mathbf{D}$ (and consequently over all $\mathbf{p} \in \mathbf{P}$). Since $\mathbf{p}_0 \in \mathbf{P}$ as well, the deterministic decision rule given in (7) achieves a performance score of $\mathbf{v}^T \mathbf{p}_0$. Any other decision rule that does not agree with δ^* on any subset of $\bar{\mathbf{B}}(\mathbf{v})$ with nonzero probability measure will have a strictly greater performance score than $\mathbf{v}^T \mathbf{p}_0$ (due to the optimality of δ^*), and hence, cannot be on the supporting hyperplane.

Case 1: We prove the first part by contrapositive. Suppose that the deterministic decision rule δ^* given in (7) yields $\mathbf{p}^* \neq \mathbf{p}_0$ meaning that \mathbf{p}_0 is achieved by some other decision rule $\delta^0 \in \mathbf{D}$. Since δ^* minimizes $\mathbf{v}^T \mathbf{p}$ over all $\mathbf{p} \in \mathbf{P}$, $\mathbf{v}^T \mathbf{p}^* = \mathbf{v}^T \mathbf{p}_0$ holds and both \mathbf{p}^* and \mathbf{p}_0 are located on the supporting hyperplane $\{\mathbf{p} : \mathbf{v}^T \mathbf{p} = \mathbf{v}^T \mathbf{p}_0\}$. This implies that δ^* and δ^0 must agree on any subset of $\bar{\mathbf{B}}(\mathbf{v})$ with nonzero probability measure. As a result, the difference between the pairwise probability vectors \mathbf{p}^* and \mathbf{p}_0 must stem from the difference of δ^* and δ^0 over $\mathbf{B}(\mathbf{v})$. Consequently, the set $\mathbf{B}(\mathbf{v})$ cannot have zero probability under all hypotheses.

Case 2: Suppose that the set of boundary points specified by $\mathbf{B}(\mathbf{v})$ has nonzero probability under some hypotheses. In this case, each point in $\mathbf{B}_{i,j}(\mathbf{v})$ can be assigned arbitrarily (or in a randomized manner) to hypotheses \mathcal{H}_i and \mathcal{H}_j . Since the way the ties are broken does not change $\mathbf{v}^T \mathbf{p}$, the resulting error probability vectors are all located on the intersection of the set \mathbf{P} with the $M(M-1)-1$ dimensional supporting hyperplane $\{\mathbf{p} : \mathbf{v}^T \mathbf{p} = \mathbf{v}^T \mathbf{p}_0\}$. By Carathéodory's Theorem [14], any point (including \mathbf{p}_0) in the intersection set, whose dimension is at most $M(M-1)-1$, can be represented as a convex combination of at most $M(M-1)$ extreme points of this set. Since these extreme points can only be obtained via deterministic decision rules which all agree with δ^* on the set $\bar{\mathbf{B}}(\mathbf{v})$, \mathbf{p}_0 can be achieved by a randomization among at most $M(M-1)$ deterministic decision rules of the form given in (7), all corresponding to the weights specified by \mathbf{v} . ■

REFERENCES

- [1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
- [2] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed. New York, USA: Chapman & Hall, 1986.
- [3] J. L. Hodges Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, vol. 23, no. 3, pp. 396–407, Sep. 1952.
- [4] S. Gezici and P. K. Varshney, "On the optimality of likelihood ratio test for prospect theory-based binary hypothesis testing," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1845–1849, Dec. 2018.
- [5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.
- [6] H. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 265–271, May 1980.
- [7] R. Gonzales and G. Wu, "On the shape of the probability weighting function," *Cogn. Psychol.*, vol. 38, no. 1, pp. 129–166, 1999.
- [8] D. Prelec, "The probability weighting function," *Econometrica*, vol. 66, no. 3, pp. 497–527, 1998.
- [9] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk Uncertainty*, vol. 5, pp. 297–323, 1992.
- [10] C. Altay and H. Delic, "Optimal quantization intervals in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 38–48, Feb. 2016.
- [11] C. A. M. Sotomayor, R. P. David, and R. Sampaio-Neto, "Adaptive nonassisted distributed detection in sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 6, pp. 3165–3174, Dec. 2017.
- [12] A. Ghobadzadeh and R. S. Adve, "Separating function estimation test for binary distributed radar detection with unknown parameters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1357–1369, Jun. 2019.
- [13] D. Warren and P. Willett, "Optimum quantization for detector fusion: Some proofs, examples, and pathology," *J. Franklin Inst.*, vol. 336, no. 2, pp. 323–359, 1999.
- [14] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1968.