

Review

A Framework for Characterization of Optimal Decision Rules in Hypothesis-Testing Problems

Emre Efendi ¹, Berkan Dulek ¹ , Sinan Gezici ^{2,*} and Yanglei Song ³

¹ Department of Electrical and Electronics Engineering, Hacettepe University, Beytepe Campus, Ankara 06800, Turkey; eefendi@ee.hacettepe.edu.tr (E.E.); berkan@ee.hacettepe.edu.tr (B.D.)

² Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey

³ Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada; yanglei.song@queensu.ca

* Correspondence: gezici@ee.bilkent.edu.tr; Tel.: +90-312-290-1219

Abstract

In this review paper, we present a framework for the characterization of optimal decision rules in M -ary hypothesis-testing problems where the performance metric is defined as a function of pairwise error probabilities. This framework is based on the approaches developed in several recent studies in the literature, which are unified and presented in a tutorial fashion in this paper. A pairwise error probability represents the probability of selecting a specific hypothesis when a different hypothesis is true, and can be stacked into a pairwise probability vector for a given problem. In the considered framework, instead of optimizing the performance metric of interest over the infinite-dimensional set of all possible decision rules, the optimization is performed directly over the compact and convex set of all achievable pairwise probability vectors. We demonstrate that any pairwise probability vector within this feasible set can be realized via a randomization of at most two likelihood ratio quantizers (LRQs) with different sets of parameters. While one of these LRQs can always be selected as a deterministic LRQ, the other one is possibly a randomized LRQ, which can be written as a randomization of at most $M(M - 1)$ deterministic LRQs, with M denoting the number of hypotheses. The main advantage of this framework is that it allows for the attainment of pairwise probability vectors that do not reside on the boundary of the feasible set and that are fundamentally inaccessible via LRQs, which are optimal for classical performance metrics such as the Bayes risk or the Neyman–Pearson criterion. Furthermore, we show that the characterization of decision rules with the presented framework is particularly advantageous for performance metrics based on prospect theory (PT), such as behavioral utility. Specifically, it is demonstrated that the optimal pairwise probability vector for a PT-based metric is not guaranteed to lie on the boundary of the feasible set of pairwise probability vectors. This results in suboptimal performance achieved by LRQs for such performance metrics. On the other hand, the randomized decision rules characterized in this paper can achieve pairwise probability vectors located in the interior of the feasible set, thereby yielding optimal performance. Numerical results corroborate these findings, demonstrating that the decision rules characterized within our framework yield optimal behavioral utility-based performance scores.

Keywords: decision rule; hypothesis testing; Bayes risk; Neyman–Pearson criterion; prospect theory



Academic Editors: Onur Günlü, Holger Boche, Giuseppe Caire, Shlomo Shamai (Shitz) and Osvaldo Simeone

Received: 27 April 2026

Revised: 1 June 2026

Accepted: 5 June 2026

Published: 9 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In his seminal book [1], H. Vincent Poor characterizes detection and estimation theory as the study of drawing inferences about a phenomenon of interest (POI) from noisy observations. This paper focuses primarily on detection problems (also called hypothesis-testing problems), which involve identifying the true underlying state of the POI. In such problems, observations whose probability distributions depend on the true state of the POI are processed by a decision rule, or detector. Traditional performance metrics frequently employed to evaluate such decision rules include the *Bayes risk*, the *Neyman–Pearson (NP) criterion*, and the *minimax criterion*. As demonstrated in [1–3], likelihood ratio tests (LRTs) are proven to be optimal for these classical performance metrics. These classical metrics rely on the assumption that the decision agent (DA) is rational, much like physical sensors, which function as rigid, objective measurement devices.

Classical performance metrics may be inadequate in scenarios where human operators are integrated into detection systems, as human judgment is inherently influenced by cognitive load, fatigue, and varying risk preferences. This contrasts sharply with the rigid, objective behavior assumed by classical performance metrics [4–6]. Given that human operators are increasingly utilized as primary decision-makers in diverse fields, such as medicine, surveillance, and crowdsourcing [3,7], and that the subjective experiences of human users are now a central consideration in the design of many modern systems [8–12], a fundamentally different analytical framework is required. Such a framework must accommodate the nuances of human participation, moving beyond the traditional models developed for detection systems comprised solely of physical sensors.

One of the most popular methods to incorporate the effects of a behavioral DA, i.e., a human, in detection systems is to adopt a performance metric that models human behavioral biases using prospect theory (PT), which is the most frequently used method since it captures decision-making behaviors of human DAs [13–21]. PT posits that behavioral biases arise from the inconsistent perception of objective probability and utility values. To model these perceptions, PT employs a probability weighting function and a utility valuation function [22–26]. By devising a performance metric that utilizes the perceived probability and utility values, it is possible to incorporate the effects of human DAs into the detection system. For example, a PT-based subjective utility function is used as a performance metric in [13–15,17,18]. The behavioral risk and behavioral utility have been employed as behavioral variants of the classical Bayes risk [16,19,20]. Additionally, a PT-based formulation of the NP problem was recently proposed in [21].

It has been established that, due to the non-linearity of the probability weighting function, direct analytical approaches often fail to reveal the structure of optimal detection rules for PT-based performance metrics. Moreover, it is reported that LRTs or likelihood ratio quantizers (LRQs), which are the canonical optimal rules for rational performance metrics, may be suboptimal for PT-based performance metrics [16,19–21]. As a remedy, a novel approach to characterize the optimal decision rules for any performance metric that can be expressed as a function of the *pairwise error probabilities* for *M*-ary hypothesis testing was developed in [16,19,20]. In this context, a pairwise error probability vector denotes the likelihood of selecting a specific hypothesis given that a different hypothesis is true. We represent these probabilities collectively as a single column vector, referred to throughout this study as the *pairwise probability vector*.

In this work, we provide a comprehensive review of the framework that was developed in our previous works with application to PT-based performance metrics. The presented framework is built on an indirect approach to characterize the optimal decision rules. Rather than performing optimization over the infinite-dimensional space of all possible randomized decision rules, the performance metric is optimized directly over

the set of all achievable pairwise probability vectors. This shift is justified by the fact that decision rules influence performance metrics solely through their constituent pairwise error probabilities. By leveraging the compactness and convexity of this feasible set, the optimal decision rules can be characterized based on the geometric location of their corresponding pairwise probability vectors. Specifically, we demonstrate that any element within the set of all achievable pairwise probability vectors can be realized through a randomization of two LRQs constructed with different parameter sets. More explicitly, one of these LRQs can always be chosen as a deterministic LRQ, while the second may be either deterministic or randomized. In this context, a deterministic LRQ partitions the space of likelihood ratios into regions where a hypothesis is selected via convex separating hyperplanes. A randomized LRQ utilizes randomization only if the observation falls on a separating hyperplane where multiple hypotheses yield the same weighted cost. Furthermore, any randomized LRQ can be expressed as a randomization of deterministic LRQs that are constructed with the same parameter set but favor different hypotheses when an observation lies on a separating hyperplane. While this approach is particularly well-suited for non-linear PT-based performance metrics, the proposed method also provides a comprehensive toolkit for characterizing optimal decision rules for any metric that is a function of pairwise error probabilities.

In this paper, we also apply the proposed framework to a PT-based performance metric, namely the behavioral utility. The pairwise probability vectors of LRQs are fundamentally restricted to the boundary of the set of all achievable pairwise probability vectors. However, due to the non-linearity of the probability weighting function in PT, the behavioral utility may be maximized at a point residing in the interior of this feasible set. We demonstrate that by leveraging the randomized decision rules characterized in our framework, such interior points of the feasible set can be realized, whereas they remain inaccessible to standard deterministic LRQs. We corroborate these theoretical findings and demonstrate the efficacy of the proposed framework through several numerical examples, where the behavioral utility serves as the performance metric for the decision rules.

In this paper, we focus on the following points:

- We provide a general framework, which is developed in several papers [16,19,20], and present a unified approach for the characterization of optimal decision rules for any performance metric that can be expressed as a function of the pairwise probability vector. This framework leverages the geometric properties of the set of all achievable pairwise probability vectors. Specifically, we show that for M -ary hypothesis testing, any pairwise probability vector can be realized through a randomization of at most two (possibly randomized) LRQs constructed with different parameter sets.
- We present a method to sweep the set of all achievable pairwise probability vectors for the binary hypothesis-testing case.
- We apply the unified framework to a PT-based performance metric, namely the *behavioral utility*, for binary hypothesis-testing problems.

The remainder of this paper is organized as follows. In Section 2, the problem formulation for M -ary hypothesis testing is provided, and optimal decision rules for classical performance metrics, such as the Bayes risk and the NP criterion, are presented for review purposes. Section 3 combines the results of several works to form a framework for the characterization of optimal decision rules. Section 4 applies this framework to the PT-based performance metric, i.e., behavioral utility. The numerical results are provided in Section 5, and Section 6 concludes the paper.

2. Optimal Decision Rules for Classical Performance Metrics

In detection problems, the underlying state of a POI is determined by processing the observation of a DA by a decision rule. The observation space of the DA and the observation sampled from the observation space of the DA are denoted by $\mathcal{Y}$ and $\mathbf{Y}$, respectively, where $\mathcal{Y} \subseteq \mathbb{R}^N$. The various possible states of the POI are represented by a finite set of M hypotheses, denoted as $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{M-1}$. The probability distribution of the observation $\mathbf{Y}$ under each hypothesis is given by

$$\begin{aligned}\mathcal{H}_0 : \mathbf{Y} &\sim f_0 \\ \mathcal{H}_1 : \mathbf{Y} &\sim f_1 \\ &\vdots \\ \mathcal{H}_{M-1} : \mathbf{Y} &\sim f_{M-1}\end{aligned}$$

where f_i represents the probability density of $\mathbf{Y}$ under hypothesis $\mathcal{H}_i$ with respect to some reference measure μ on $\mathcal{Y}$. Depending on whether the observation space $\mathcal{Y}$ is continuous or discrete, f_j denotes either a *probability density function (pdf)* or a *probability mass function (pmf)* under hypothesis $\mathcal{H}_j$, respectively. For conciseness, the term *distribution* is used to refer to both pdfs and pmfs in the remainder of this paper.

The decision-making process of the DA is symbolized with

$$q = \delta(\mathbf{y}) \quad (1)$$

where $\mathbf{y}$ is the DA's observation (i.e., a realization of $\mathbf{Y}$), δ is a possibly randomized decision rule employed by the DA and q is the DA's decision. Given that $\mathbf{y}$ is observed, the decision rule δ is defined as the following vector: $\delta(\mathbf{y}) = [\delta_0(\mathbf{y}), \dots, \delta_{M-1}(\mathbf{y})]^\top$, where $\delta_i(\mathbf{y})$ denotes the probability of choosing $\mathcal{H}_i$ given that $\mathbf{y}$ is observed for $i \in \{0, 1, \dots, M-1\}$. Consequently, the probability of choosing hypothesis $\mathcal{H}_i$ given that hypothesis $\mathcal{H}_j$ is the true state can be computed as

$$p_{ij} = \int_{\mathcal{Y}} \delta_i(\mathbf{y}) f_j(\mathbf{y}) \mu(d\mathbf{y}), \quad (2)$$

where $\int \mu(d\mathbf{y})$ denotes an integral operation if $\mathbf{Y}$ follows a continuous distribution and a summation over the observation space if $\mathbf{Y}$ follows a discrete distribution. These conditional probabilities in (2), often referred to as pairwise (error) probabilities, are utilized in the metrics that quantify the performance of the randomized decision rule $\delta \in \Delta$ under each possible hypothesis, where Δ denotes the set of all possible (randomized) decision rules.

Note that deterministic decision rules can be viewed as a special case of randomized decision rules. Specifically, a deterministic decision rule partitions the observation space $\mathcal{Y}$ into M disjoint subsets $\{\mathcal{Y}_i\}_{i=0}^{M-1}$. If the observation $\mathbf{y}$ falls into the subset $\mathcal{Y}_i$, the hypothesis $\mathcal{H}_i$ is selected deterministically [1]. This implies that for $\mathbf{y} \in \mathcal{Y}_i$, $\delta_i(\mathbf{y}) = 1$ and $\delta_k(\mathbf{y}) = 0$ for all $k \neq i$. Consequently, for a deterministic decision rule, the pairwise probability of choosing $\mathcal{H}_i$ when the true hypothesis is $\mathcal{H}_j$ given in (2) simplifies to

$$p_{ij} = \int_{\mathcal{Y}_i} f_j(\mathbf{y}) \mu(d\mathbf{y}). \quad (3)$$

The performance of a decision rule (detector) is evaluated using a performance metric function. In the literature, the most frequently utilized performance metrics include the Bayes risk, the NP criterion, the minimax criterion, and the restricted Bayes risk [1,2]. An optimal decision rule is defined as one that optimizes these respective performance

metrics. In the following sections, we focus specifically on the Bayes risk and the NP criterion and present the derivations of the optimal decision rules.

2.1. Bayesian Hypothesis Testing

A detection problem in which the performance metric is the Bayes risk is referred to as the Bayesian hypothesis-testing problem. In this problem, the prior probabilities $\{\pi_j\}_{j=0}^{M-1}$ are assumed to be known, and the Bayes risk of a randomized decision rule $\delta \in \Delta$ is expressed as

$$r(\delta) = \sum_{j=0}^{M-1} \pi_j R_j(\delta), \quad (4)$$

where $R_j(\delta)$ denotes the conditional risk under hypothesis $\mathcal{H}_j$, which is defined as

$$R_j(\delta) = \sum_{i=0}^{M-1} c_{ij} P(\mathcal{H}_i | \mathcal{H}_j) = \sum_{i=0}^{M-1} c_{ij} \mathbb{E}_j[\delta_i(\mathbf{Y})], \quad (5)$$

with c_{ij} representing the finite cost incurred when hypothesis $\mathcal{H}_i$ is selected given that $\mathcal{H}_j$ is the true state. Here, $\mathbb{E}_j[\cdot]$ denotes the expectation operator with respect to the distribution $f_j(\mathbf{y})$, and $\mathbb{E}_j[\delta_i(\mathbf{Y})]$ corresponds to the pairwise probability of choosing $\mathcal{H}_i$ given that $\mathcal{H}_j$ is true, i.e., $p_{ij} = \mathbb{E}_j[\delta_i(\mathbf{Y})]$. Thus, the Bayes risk represents the weighted average of the conditional risks, which themselves quantify the average cost incurred by employing decision rule δ when hypothesis $\mathcal{H}_j$ is true.

The Bayes rule is the optimal decision rule that minimizes the Bayes risk. To obtain it, we first rewrite the Bayes risk by expanding the conditional probabilities:

$$r(\delta) = \sum_{j=0}^{M-1} \pi_j \sum_{i=0}^{M-1} c_{ij} p_{ij}. \quad (6)$$

By substituting (2) into the expression above and interchanging the order of summation and integration, we obtain

$$r(\delta) = \int_{\mathcal{Y}} \sum_{i=0}^{M-1} \delta_i(\mathbf{y}) \left(\sum_{j=0}^{M-1} \pi_j c_{ij} f_j(\mathbf{y}) \right) \mu(d\mathbf{y}). \quad (7)$$

Let $f(\mathbf{y}) \triangleq \sum_{j=0}^{M-1} \pi_j f_j(\mathbf{y})$ denote the marginal $\mu(\cdot)$ -density of the observation, and let $\pi_j(\mathbf{y}) \triangleq P(\mathcal{H}_j | \mathbf{y}) = \pi_j f_j(\mathbf{y}) / f(\mathbf{y})$ represent the posterior probability that hypothesis $\mathcal{H}_j$ is true given the observation $\mathbf{y}$. The Bayes risk in (7) can then be expressed as

$$r(\delta) = \int_{\mathcal{Y}} \left(\sum_{i=0}^{M-1} \delta_i(\mathbf{y}) \sum_{j=0}^{M-1} c_{ij} \pi_j(\mathbf{y}) \right) f(\mathbf{y}) \mu(d\mathbf{y}). \quad (8)$$

By defining the *posterior cost* of choosing hypothesis $\mathcal{H}_i$ as

$$c_i(\mathbf{y}) \triangleq \sum_{j=0}^{M-1} c_{ij} \pi_j(\mathbf{y}), \quad (9)$$

the Bayes risk in (8) becomes

$$\begin{aligned} r(\delta) &= \int_{\mathcal{Y}} \left(\sum_{i=0}^{M-1} \delta_i(\mathbf{y}) c_i(\mathbf{y}) \right) f(\mathbf{y}) \mu(d\mathbf{y}) \\ &\geq \int_{\mathcal{Y}} \min_{0 \leq i \leq M-1} \{c_i(\mathbf{y})\} f(\mathbf{y}) \mu(d\mathbf{y}). \end{aligned} \quad (10)$$

The lower bound in (10) is achieved if, for every $\mathbf{y} \in \mathcal{Y}$, we set $\delta_\ell(\mathbf{y}) \geq 0$ for $\ell \in \mathcal{C}(\mathbf{y}) \triangleq \{j \in \{0, 1, \dots, M-1\} : c_j(\mathbf{y}) = \min_i c_i(\mathbf{y})\}$ such that $\sum_{\ell \in \mathcal{C}(\mathbf{y})} \delta_\ell(\mathbf{y}) = 1$, and $\delta_i(\mathbf{y}) = 0$ for all $i \notin \mathcal{C}(\mathbf{y})$. In other words, when multiple hypotheses attain the same minimum posterior cost for a given observation $\mathbf{y}$, choosing a hypothesis from the set $\mathcal{C}(\mathbf{y})$ deterministically or selecting among them randomly yields an identical Bayes risk value. Consequently, randomized decision rules are unnecessary. Hence, the Bayes rule can be expressed by selecting an index $\ell \in \mathcal{C}(\mathbf{y})$ and setting:

$$\delta_\ell^B(\mathbf{y}) = 1, \quad \text{for some } \ell \in \mathcal{C}(\mathbf{y}), \quad (11)$$

with $\delta_i^B(\mathbf{y}) = 0$ for all $i \in \{0, 1, \dots, M-1\} \setminus \{\ell\}$. Namely, any tie in posterior costs is broken deterministically in favor of a hypothesis $\mathcal{H}_\ell$ such that $\ell \in \mathcal{C}(\mathbf{y})$.

For the case of *uniform cost assignment (UCA)*, where the cost of a correct decision is $c_{ii} = 0$ and the cost of any erroneous decision is $c_{ij} = 1$ for $i \neq j$, the Bayes decision rule reduces to the *maximum a posteriori (MAP)* decision rule as follows. Under UCA, the posterior cost in (9) simplifies to $c_i(\mathbf{y}) = 1 - \pi_i(\mathbf{y})$; therefore, minimizing the posterior cost as in (11) becomes equivalent to maximizing the posterior probability. Hence, the MAP rule is given by

$$\delta_\ell^{\text{MAP}}(\mathbf{y}) = 1, \quad \ell = \arg \max_{0 \leq i \leq M-1} \pi_i(\mathbf{y}), \quad (12)$$

with $\delta_k^{\text{MAP}}(\mathbf{y}) = 0$ for all $k \neq \ell$, where ties in $\arg \max$ are broken in a deterministic fashion. Furthermore, if the prior probabilities are assumed to be equal (i.e., $\pi_j = 1/M$ for all j), the MAP rule further reduces to the *maximum likelihood (ML)* decision rule:

$$\delta_\ell^{\text{ML}}(\mathbf{y}) = 1, \quad \ell = \arg \max_{0 \leq i \leq M-1} f_i(\mathbf{y}), \quad (13)$$

where $\delta_k^{\text{ML}}(\mathbf{y}) = 0$ for all $k \neq \ell$.

Binary Hypothesis Testing

In binary hypothesis testing, the POI can be in one of $M = 2$ states, denoted by $\mathcal{H}_0$ and $\mathcal{H}_1$, with prior probabilities π_0 and π_1 , respectively. In this case, the optimal Bayes decision rule simplifies to a comparison of the posterior costs (cf. (11)):

$$c_0(\mathbf{y}) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} c_1(\mathbf{y}), \quad (14)$$

which, by substituting the definitions of posterior cost in (9), can explicitly be written as

$$c_{00}\pi_0(\mathbf{y}) + c_{01}\pi_1(\mathbf{y}) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} c_{10}\pi_0(\mathbf{y}) + c_{11}\pi_1(\mathbf{y}). \quad (15)$$

Assuming $c_{01} > c_{11}$ (i.e., the cost of an error exceeds the cost of a correct decision), we can substitute the posterior probabilities $\pi_j(\mathbf{y}) = \pi_j f_j(\mathbf{y}) / f(\mathbf{y})$, and by rearranging the terms, simplify the Bayes decision rule to a likelihood ratio test (LRT):

$$L(\mathbf{y}) \triangleq \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{(c_{10} - c_{00})\pi_0}{(c_{01} - c_{11})\pi_1} \triangleq \eta, \quad (16)$$

where η denotes the optimal threshold against which the likelihood ratio is compared. The decision rule δ^B is then fully characterized by the indicator function $\delta_1^B(\mathbf{y})$:

$$\delta_1^B(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq \eta \\ 0, & \text{if } L(\mathbf{y}) < \eta, \end{cases} \quad (17)$$

with $\delta_0^B(\mathbf{y}) = 1 - \delta_1^B(\mathbf{y})$.

Similarly, for binary hypothesis testing, the MAP rule in (12) can be expressed in terms of the weighted likelihood functions:

$$\pi_1 f_1(\mathbf{y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \pi_0 f_0(\mathbf{y}), \quad (18)$$

which is equivalent to the following LRT:

$$\delta_1^{\text{MAP}}(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq \frac{\pi_0}{\pi_1} \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Assuming equal prior probabilities ($\pi_0 = \pi_1 = 1/2$), the MAP decision rule further reduces to the ML decision rule, which compares the likelihood ratio to a unit threshold:

$$\delta_1^{\text{ML}}(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

2.2. Neyman–Pearson (NP) Hypothesis Testing

For binary hypothesis testing, the NP criterion is one of the most widely utilized performance metrics alongside the Bayes risk. In the NP framework, the prior probabilities are assumed to be unknown. The objective is to design a decision rule (test) that maximizes the detection probability while maintaining the false alarm probability below a pre-specified threshold α , where $\alpha \in [0, 1]$ is referred to as the *level* of the test.

The optimal NP decision rule is obtained by solving the following constrained optimization problem:

$$\begin{aligned} & \underset{\delta \in \Delta}{\text{maximize}} && P_D(\delta) \\ & \text{subject to} && P_F(\delta) \leq \alpha \end{aligned} \quad (21)$$

where $P_D(\delta) = P(\mathcal{H}_1|\mathcal{H}_1) = p_{11}$ and $P_F(\delta) = P(\mathcal{H}_1|\mathcal{H}_0) = p_{10}$ denote the *detection probability* and *false alarm probability* of the randomized decision rule δ , respectively. In statistics, P_D and P_F are often referred to as the *power* and *size* of the test. Thus, the NP approach seeks to find the *most powerful α -level test* for comparing $\mathcal{H}_0$ against $\mathcal{H}_1$ [1,2].

In the following, we present the NP lemma, which proves that the solution to the optimization problem in (21) is a randomized LRT of the form:

$$\delta_1^{\text{NP}}(\mathbf{y}) = \begin{cases} 1, & \text{if } f_1(\mathbf{y}) > \eta_0 f_0(\mathbf{y}) \\ \gamma_0, & \text{if } f_1(\mathbf{y}) = \eta_0 f_0(\mathbf{y}) \\ 0, & \text{if } f_1(\mathbf{y}) < \eta_0 f_0(\mathbf{y}) \end{cases} \quad (22)$$

where the threshold $\eta_0 \geq 0$ and the randomization constant $\gamma_0 \in [0, 1]$ are chosen to satisfy the size constraint as $P_F(\delta^{\text{NP}}) = \alpha$ ([1] Section II.D).

Lemma 1 (Neyman–Pearson Lemma [1]). *For a binary hypothesis-testing problem with densities $f_0(\mathbf{y})$ and $f_1(\mathbf{y})$, and a given level $\alpha > 0$, the following statements are true:*

Optimality: Let $\tilde{\delta} = (\tilde{\delta}_0, \tilde{\delta}_1)$ be any decision rule satisfying $P_F(\tilde{\delta}) \leq \alpha$, and let $\tilde{\delta}' = (\tilde{\delta}'_0, \tilde{\delta}'_1)$ be of the form

$$\tilde{\delta}'_1(\mathbf{y}) = \begin{cases} 1, & \text{if } f_1(\mathbf{y}) > \eta f_0(\mathbf{y}) \\ \gamma(\mathbf{y}), & \text{if } f_1(\mathbf{y}) = \eta f_0(\mathbf{y}) \\ 0, & \text{if } f_1(\mathbf{y}) < \eta f_0(\mathbf{y}) \end{cases} \quad (23)$$

where $\eta \geq 0$ and $\gamma(\mathbf{y}) \in [0, 1]$ are such that $P_F(\tilde{\delta}') = \alpha$. Then, $P_D(\tilde{\delta}') \geq P_D(\tilde{\delta})$. Hence, size- α decision rule of the form (23) is an NP rule.

Existence: For every $\alpha \in (0, 1)$, there is a decision rule $\tilde{\delta}^N = (\tilde{\delta}_0^N, \tilde{\delta}_1^N)$ of the form of (23) with $\gamma(\mathbf{y}) = \gamma_0$ (a constant) for which $P_F(\tilde{\delta}^N) = \alpha$.

Uniqueness: Suppose $\tilde{\delta}'' = (\tilde{\delta}''_0, \tilde{\delta}''_1)$ is any α -level NP rule for $\mathcal{H}_0$ versus $\mathcal{H}_1$. Then, $\tilde{\delta}''$ must be of the form of (23) except possibly on a subset of $\mathcal{Y}$ having zero probability under $\mathcal{H}_0$ and $\mathcal{H}_1$.

The NP rule in (22) can be interpreted as a convex combination of two deterministic LRTs, $\delta^* = (\delta_0^*, \delta_1^*)$ and $\delta^\dagger = (\delta_0^\dagger, \delta_1^\dagger)$, defined as follows:

$$\delta_1^*(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq \eta_0 \\ 0, & \text{if } L(\mathbf{y}) < \eta_0 \end{cases}, \quad \delta_1^\dagger(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > \eta_0 \\ 0, & \text{if } L(\mathbf{y}) \leq \eta_0 \end{cases}. \quad (24)$$

Specifically, the randomized NP rule in (22) corresponds to selecting the deterministic rule δ^* with probability γ_0 and the rule $\delta^\dagger$ with probability $1 - \gamma_0$. Note that if the event $\{f_1(\mathbf{y}) = \eta_0 f_0(\mathbf{y})\}$ occurs with probability zero under both hypotheses (as is typical for continuous distributions), the randomization constant γ_0 becomes irrelevant, and the rule simplifies to a standard deterministic LRT.

3. Optimal Decision Rules Under General Criterion in Terms of Error Probabilities

In this section, we present a method to characterize optimal decision rules for performance metrics that are functions of pairwise error probabilities. A generic decision criterion expressed in terms of pairwise error probabilities can be formulated as the following optimization problem [19]:

$$\begin{aligned} & \underset{\delta \in \Delta}{\text{minimize}} && g_0(\mathbf{p}(\delta)) \\ & \text{subject to} && g_i(\mathbf{p}(\delta)) \leq 0, \quad i = 1, 2, \dots, m \\ & && h_j(\mathbf{p}(\delta)) = 0, \quad j = 1, 2, \dots, n \end{aligned} \quad (25)$$

where Δ is the set of all possible (randomized) decision rules, and g_i and h_j represent arbitrary functions of the pairwise error probabilities. The vector $\mathbf{p}(\delta)$ denotes the pairwise probability vector, defined as

$$\mathbf{p}(\delta) = [p_{10}, p_{20}, \dots, p_{M-1,0}, p_{01}, p_{21}, \dots, p_{M-2,M-1}]^\top. \quad (26)$$

where p_{ij} denotes the probability of choosing $\mathcal{H}_i$ when the true hypothesis is $\mathcal{H}_j$, as defined in (2). Note that it is sufficient to include only the pairwise error probabilities (i.e., p_{ij} for $i \neq j, i, j \in \{0, 1, \dots, M-1\}$) in $\mathbf{p}(\delta)$, as the correct decision probabilities are linearly dependent: $p_{jj} = 1 - \sum_{i \neq j} p_{ij}$. Classical hypothesis-testing criteria, including Bayesian, minimax, NP, and restricted Bayesian, can all be viewed as special cases of the formulation in (25).

Decision rules influence performance metrics exclusively through their corresponding pairwise probability vectors. Consequently, optimizing a performance metric over the set of all achievable pairwise probability vectors is equivalent to optimizing it over the set of all randomized decision rules. We define the set of all achievable pairwise probability vectors as

$$\mathcal{P}(\Delta) \triangleq \{\mathbf{p}(\delta) : \delta \in \Delta\}. \quad (27)$$

It is important to note that the set $\mathcal{P}(\Delta)$ is both convex and compact, as established in [19,20,27]. Using this property, the general decision criterion in (25) can be reformulated as an optimization over the achievable set of pairwise probability vectors:

$$\begin{aligned} & \underset{\mathbf{p} \in \mathcal{P}(\Delta)}{\text{minimize}} && g_0(\mathbf{p}) \\ & \text{subject to} && g_i(\mathbf{p}) \leq 0, \quad i = 1, 2, \dots, m \\ & && h_j(\mathbf{p}) = 0, \quad j = 1, 2, \dots, n \end{aligned} \quad (28)$$

Next, we leverage the convexity and compactness of $\mathcal{P}(\Delta)$ to elucidate the relationship between pairwise probability vectors and their corresponding decision rules. To develop the framework, we first consider a linear objective of the form $\mathbf{v}^\top \mathbf{p}$, that is, a weighted sum of the pairwise error probabilities, and show that the corresponding optimal rules are likelihood ratio quantizers (LRQs) associated with vector $\mathbf{v}$. We then return to the general optimization problem in (28) and use this characterization to describe the structure of the optimal decision rules.

Consider the unconstrained performance metric $\mathbf{v}^\top \mathbf{p}(\delta)$, where $\mathbf{v} \in \mathbb{R}^{M(M-1)}$ is a weight vector with real-valued entries v_{ij} for $i, j \in \{0, 1, \dots, M-1\}$ and $i \neq j$. Since $\mathcal{P}(\Delta)$ is a convex and compact set, for any $\mathbf{v}$, there exists a supporting hyperplane $\{\mathbf{p} : \mathbf{v}^\top \mathbf{p} = \mathbf{v}^\top \mathbf{p}^*\}$ such that $\mathbf{p}^* \in \mathcal{P}(\Delta)$ and $\mathbf{v}^\top \mathbf{p} \geq \mathbf{v}^\top \mathbf{p}^*$ for all $\mathbf{p} \in \mathcal{P}(\Delta)$ [28]. To derive the decision rule that yields $\mathbf{v}^\top \mathbf{p}^*$, we express the performance metric $\mathbf{v}^\top \mathbf{p}(\delta)$ as follows:

$$\mathbf{v}^\top \mathbf{p}(\delta) = \sum_{i=0}^{M-1} \sum_{j=0, j \neq i}^{M-1} v_{ij} p_{ij}. \quad (29)$$

By substituting p_{ij} with (2) and interchanging the order of summation and integration, (29) becomes

$$\mathbf{v}^\top \mathbf{p}(\delta) = \int_{\mathcal{Y}} \sum_{i=0}^{M-1} \delta_i(\mathbf{y}) \left(\sum_{j=0, j \neq i}^{M-1} v_{ij} f_j(\mathbf{y}) \right) \mu(d\mathbf{y}). \quad (30)$$

By defining the *weighted likelihood sum* for hypothesis $\mathcal{H}_i$ as

$$V_i(\mathbf{y}) \triangleq \sum_{j=0, j \neq i}^{M-1} v_{ij} f_j(\mathbf{y}), \quad (31)$$

the performance metric in (30) becomes

$$\begin{aligned} \mathbf{v}^\top \mathbf{p}(\delta) &= \int_{\mathcal{Y}} \sum_{i=0}^{M-1} \delta_i(\mathbf{y}) V_i(\mathbf{y}) \mu(d\mathbf{y}) \\ &\geq \int_{\mathcal{Y}} \min_{0 \leq i \leq M-1} V_i(\mathbf{y}) \mu(d\mathbf{y}). \end{aligned} \quad (32)$$

The lower bound in (32); hence, the minimizer of the performance metric, can be achieved by a randomized LRQ or deterministic LRQ as specified below.

Let $\mathcal{V}(\mathbf{y})$ denote the set of indices that minimizes $V_i(\mathbf{y})$ in (32). That is,

$$\mathcal{V}(\mathbf{y}) \triangleq \{j \in \{0, 1, \dots, M-1\} : V_j(\mathbf{y}) = \min_{0 \leq i \leq M-1} V_i(\mathbf{y})\}.$$

We call a decision rule δ a *randomized LRQ associated with $\mathbf{v}$* if

$$\sum_{\ell \in \mathcal{V}(\mathbf{y})} \delta_\ell(\mathbf{y}) = 1 \quad (33)$$

and $\delta_i(\mathbf{y}) = 0$ for all $i \notin \mathcal{V}(\mathbf{y})$. It is noted that any randomized LRQ associated with $\mathbf{v}$ attains the lower bound in (32), and hence any such rule is optimal for minimizing $\mathbf{v}^\top \mathbf{p}(\delta)$. As a special case of randomized LRQ, we can set one of the $\delta_\ell(\mathbf{y})$ components to one for a certain ℓ in $\mathcal{V}(\mathbf{y})$ and set all the other elements of δ to zero, yielding a deterministic LRQ. In other words, we call a decision rule δ a *deterministic LRQ associated with $\mathbf{v}$* if

$$\delta_\ell(\mathbf{y}) = 1 \quad \text{for some } \ell \in \mathcal{V}(\mathbf{y}) \quad (34)$$

and $\delta_i(\mathbf{y}) = 0$ for all $i \in \{0, 1, \dots, M-1\} \setminus \{\ell\}$. Therefore, every deterministic LRQ associated with $\mathbf{v}$ is also optimal for the same linear objective, $\mathbf{v}^\top \mathbf{p}(\delta)$.

Remark 1. In this context, a deterministic LRQ is a decision rule that separates the likelihood ratio space into M convex polytopes in which one hypothesis is chosen deterministically. The polytopes in which different hypotheses are selected are separated via convex hyperplanes. It should be noted that, if the observation $\mathbf{y}$ corresponds to a point on a separating hyperplane on the likelihood ratio space, there exists more than one element in the $\mathcal{V}(\mathbf{y})$ set and a deterministic LRQ selects a hypothesis deterministically as given in (34). Conversely, a randomized LRQ chooses a hypothesis randomly from the hypotheses whose indices compose the $\mathcal{V}(\mathbf{y})$ set, if the likelihood ratio of observation, $\mathbf{y}$, corresponds to a separating hyperplane. Naturally, a randomized LRQ can be viewed as a randomization of the deterministic LRQs constructed with the same parameter set, $\mathbf{v}$.

Next, we consider the general optimization problem in (28). We define the set of *boundary points* as a subset of the observation space $\mathcal{Y}$ for a given weight vector $\mathbf{v}$ as [19]

$$\mathbf{B}(\mathbf{v}) \triangleq \bigcup_{0 \leq i < j \leq M-1} \mathbf{B}_{i,j}(\mathbf{v}), \quad (35)$$

where each subset $\mathbf{B}_{i,j}(\mathbf{v})$ represents the points where at least two hypotheses achieve the same minimum weighted likelihood sum:

$$\mathbf{B}_{i,j}(\mathbf{v}) \triangleq \{\mathbf{y} \in \mathcal{Y} : V_i(\mathbf{y}) = V_j(\mathbf{y}) \leq V_k(\mathbf{y}), \forall k \neq i, j\}. \quad (36)$$

The complement of the boundary set, denoted as $\bar{\mathbf{B}}(\mathbf{v})$, contains the observations where a single hypothesis strictly minimizes the weighted likelihood sum; that is,

$$\bar{\mathbf{B}}(\mathbf{v}) := \mathcal{Y} \setminus \mathbf{B}(\mathbf{v}) = \{\mathbf{y} \in \mathcal{Y} : V_i(\mathbf{y}) < V_j(\mathbf{y}), \text{ for some } 0 \leq i \leq M-1 \text{ and all } 0 \leq j \leq M-1, j \neq i\}. \quad (37)$$

If the boundary set $\mathbf{B}(\mathbf{v})$ has zero probability measure under all hypotheses, as is typical for continuous distributions, ties can be broken arbitrarily without affecting the performance metric. However, if $\mathbf{B}(\mathbf{v})$ occurs with non-zero probability (e.g., in discrete or mixed distributions), the specific hypothesis selected when multiple indices yield identical $V_i(\mathbf{y})$ values directly influences the resulting pairwise error probabilities.

In ([19] Lemma), the decision rules which yield pairwise probability vectors that reside on the boundary of $\mathcal{P}(\Delta)$ are characterized. We summarize this lemma below:

Lemma 2 ([19] Lemma). Let $\mathbf{p}^*$ be a point on the boundary of $\mathcal{P}(\Delta)$, and let $\{\mathbf{p} : \mathbf{v}^\top \mathbf{p} = \mathbf{v}^\top \mathbf{p}^*\}$ be a supporting hyperplane to $\mathcal{P}(\Delta)$ at $\mathbf{p}^*$.

- **Case 1 (Zero Probability Boundary):** If the boundary set $B(\mathbf{v})$ defined in (35) has zero probability measure under all hypotheses, then any deterministic LRQ associated with $\mathbf{v}$ of the form in (34) yields the pairwise probability vector $\mathbf{p}^*$.
- **Case 2 (Non-zero Probability Boundary):** If $B(\mathbf{v})$ has non-zero probability measure under at least one hypothesis, then $\mathbf{p}^*$ is achievable by a randomized LRQ associated with $\mathbf{v}$ as defined in (33), which randomizes at most $M(M-1)$ deterministic LRQs, all of which take the form of (34) and correspond to the same weight vector $\mathbf{v}$.

In Case 1 of Lemma 2, the optimal decision rule for $\inf_{\mathbf{p} \in \mathcal{P}(\Delta)} \mathbf{v}^\top \mathbf{p}$ is unique up to sets of probability measure zero. Consequently, the corresponding optimal pairwise probability vector $\mathbf{p}^*$ is the unique minimizer of the linear objective. It follows that $\mathbf{p}^*$ is an exposed point of $\mathcal{P}(\Delta)$, and hence also an extreme point. Since $\mathbf{p}^*$ is an extreme point of $\mathcal{P}(\Delta)$, it cannot be written as a nontrivial convex combination of two distinct points in $\mathcal{P}(\Delta)$. Therefore, a deterministic LRQ in (34) yields $\mathbf{p}^*$.

In Case 2 of Lemma 2, the boundary of $\mathcal{P}(\Delta)$ may contain points that are not extreme points, meaning they can be expressed as a convex combination of the extreme points in the set. Consequently, the linear functional $\mathbf{v}^\top \mathbf{p}$ may not have a unique minimizer. All pairwise probability vectors that minimize $\mathbf{v}^\top \mathbf{p}(\delta)$ lie on the intersection of $\mathcal{P}(\Delta)$ with the supporting hyperplane $\{\mathbf{p} : \mathbf{v}^\top \mathbf{p} = \mathbf{v}^\top \mathbf{p}_0\}$, which has a dimension of at most $M(M-1) - 1$. By Carathéodory's Theorem, any point $\mathbf{p}^*$ within this intersection can be represented as a convex combination of at most $M(M-1)$ extreme points of the set [29]. These extreme points correspond to deterministic decision rules of the form (34) associated with the same weight vector $\mathbf{v}$, [27]. Crucially, these deterministic rules differ only in the specific hypothesis selected for observations $\mathbf{y} \in B(\mathbf{v})$, where the weighted likelihood sums are tied.

In Figures 1 and 2, we provide illustrative examples for both cases of Lemma 2 within a binary hypothesis-testing framework. The set of all achievable pairwise probability vectors, $\mathcal{P}(\Delta)$, is highlighted in the figures.

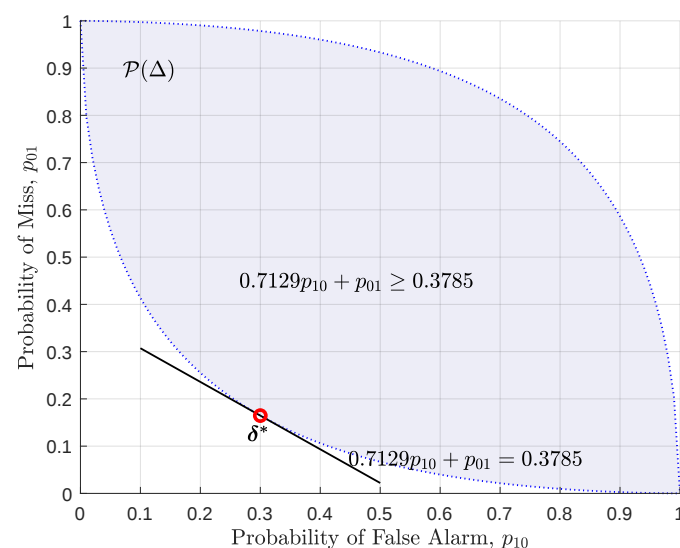


Figure 1. Set of achievable pairwise probability vectors when $\mathbf{Y}$ has a continuous distribution (Lemma 2-Case 1).

First, we consider the case in which the boundary set $B(v)$ has zero probability measure. Let the weights be assigned as $v_{01} = 1$ and $v_{10} = 0.7129$. Under these weights, the weighted likelihood sums in (31) become $V_0(\mathbf{y}) = v_{01}f_1(\mathbf{y}) = f_1(\mathbf{y})$ and $V_1(\mathbf{y}) = v_{10}f_0(\mathbf{y}) = 0.7129f_0(\mathbf{y})$. Since $B(v)$ has zero probability measure, the probability of $V_0(\mathbf{y}) = V_1(\mathbf{y})$ occur with zero probability. Hence, $\mathbf{v}^\top \mathbf{p}$ is minimized if the optimal decision rule selects the hypothesis that yields the minimum $V_i(\mathbf{y})$ value, which can be written as

$$\delta_\ell^*(\mathbf{y}) = 1, \quad \text{for } \ell = \arg \min_{i \in \{0,1\}} V_i(\mathbf{y}). \quad (38)$$

More explicitly, for binary hypothesis testing, the decision rule given above can be written as

$$v_{01}f_1(\mathbf{y}) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\geq}} v_{10}f_0(\mathbf{y}). \quad (39)$$

Since v_{01} and v_{10} are positive, the decision rule δ^* can be expressed as

$$L(\mathbf{y}) = \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} \underset{v_{01}}{\overset{v_{10}}{\geq}}. \quad (40)$$

More explicitly, the optimal decision rule that minimizes $\mathbf{v}^\top \mathbf{p}(\delta)$ is

$$\delta_1^*(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq 0.7129, \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

Since the boundary points of $\mathbf{y}$ occur with zero probability measure, the hypothesis selected when $L(\mathbf{y}) = 0.7129$ does not change the pairwise probabilities. The resulting pairwise probability vector for this rule is $\mathbf{p}^* = [0.3, 0.1646]^\top$, which achieves the minimum performance score of $\mathbf{v}^\top \mathbf{p}^*(\delta^*) = 0.3785$. The optimal pairwise probability vector, $\mathbf{p}^*(\delta^*)$ is at the intersection of $\mathcal{P}(\Delta)$ and the hyperplane $0.7129p_{10} + p_{01} = 0.3785$, as shown in Figure 1.

The boundary of the set of pairwise probability vectors for an example to Case 2 of Lemma 2 is shown by the dotted lines in Figure 2. In that case, the weights are set as $v_{10} = 0.675$ and $v_{01} = 1$, which lead to the weighted likelihood sums $V_0(\mathbf{y}) = f_1(\mathbf{y})$ and $V_1(\mathbf{y}) = 0.675f_0(\mathbf{y})$. The optimal decision rule selects the hypothesis with minimum $V_i(\mathbf{y})$ value for $i \in \{0, 1\}$, which can be written as

$$v_{01}f_1(\mathbf{y}) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\geq}} v_{10}f_0(\mathbf{y}). \quad (42)$$

Since $B(v)$ occur with nonzero probability, the probability of $v_{01}f_1(\mathbf{y})$ and $v_{10}f_0(\mathbf{y})$ being equal has a nonzero probability for these specific weights. Hence, there exist infinitely many pairwise probability vectors that minimize $\mathbf{v}^\top \mathbf{p}(\delta)$, which compose the set that is defined as the intersection of the $\mathcal{P}(\Delta)$ set and the hyperplane (line) given below:

$$0.675p_{10} + p_{01} = 0.37. \quad (43)$$

These decision rules differ only in the probability with which $\mathcal{H}_1$ is selected when $v_{01}f_1(\mathbf{y}) = v_{10}f_0(\mathbf{y})$, a condition that can be equivalently expressed as $L(\mathbf{y}) = \frac{v_{10}}{v_{01}}$. The deterministic decision rules δ^2 and δ^3 , whose pairwise probability vectors are the extreme points of the set of minimizers for $\mathbf{v}^\top \mathbf{p}(\delta)$, are given by

$$\delta_1^2(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > 0.675, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad \delta_1^3(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq 0.675, \\ 0, & \text{otherwise.} \end{cases} \quad (44)$$

The pairwise probability vectors of δ^2 and δ^3 are shown in Figure 2. Any point $\mathbf{p}^*(\delta^*)$ on this segment can be expressed as the convex combination:

$$\mathbf{p}^*(\delta^*) = \lambda \mathbf{p}^2(\delta^2) + (1 - \lambda) \mathbf{p}^3(\delta^3), \quad (45)$$

where $\lambda \in [0, 1]$. The randomized decision rule δ^* that achieves $\mathbf{p}^*(\delta^*)$ is similarly a convex combination $\delta^* = \lambda \delta^2 + (1 - \lambda) \delta^3$, which can be explicitly written as a randomized LRT:

$$\delta_1^*(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > 0.675, \\ 1 - \lambda, & \text{if } L(\mathbf{y}) = 0.675, \\ 0, & \text{if } L(\mathbf{y}) < 0.675. \end{cases} \quad (46)$$

In this case, if the event $\{L(\mathbf{y}) = 0.675\}$ occurs with non-zero probability, hypothesis $\mathcal{H}_1$ is selected with probability $1 - \lambda$ to satisfy the specific performance requirement.

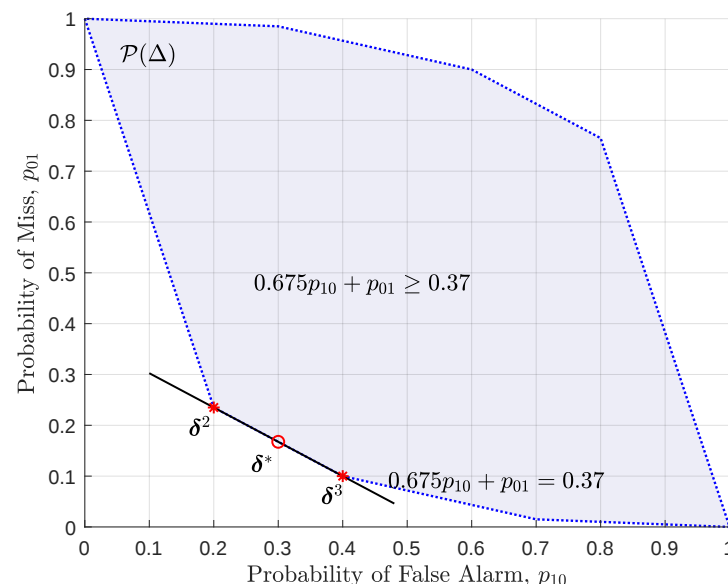


Figure 2. Set of achievable pairwise probability vectors when $\mathbf{Y}$ has a discrete distribution (Lemma 2-Case 2).

Based on the characterization of the decision rules whose pairwise probability vectors lie on the boundary of $\mathcal{P}(\Delta)$, the following theorem, developed in ([19] Theorem), provides a complete characterization of the optimal decision rules according to the location of the optimizer in $\mathcal{P}(\Delta)$, including interior points that are not covered directly by Lemma 2.

Theorem 1 ([19] Theorem). Suppose the optimization problem in (25) is feasible, and let δ^* and $\mathbf{p}^*(\delta^*)$ denote an optimal decision rule and its corresponding pairwise probability vector, respectively. An optimal decision rule that solves (25) can be characterized as follows:

- **Case 1:** If the boundary set $B(\mathbf{v})$ defined in (35) has zero probability measure under all hypotheses for any weight vector $\mathbf{v}$, then δ^* is a randomization among at most two deterministic LRQs of the form given in (34), each specified by a distinct weight vector.
- **Case 2:** Otherwise, δ^* is a randomization of two LRQs with separate $\mathbf{v}$ vectors. Note that one of these LRQs can always be chosen as a deterministic LRQ of the form given in (34),

which corresponds to Case 1 of Lemma 2. The other LRQ is possibly a randomized LRQ that corresponds to Case 2 of Lemma 2.

When the optimization problem in (25) exhibits specific structural properties, the maximum number of deterministic decision rules required to achieve optimal performance may be reduced below the bounds established in this theorem.

If the optimal pairwise probability vector $\mathbf{p}^*$ resides on the boundary of $\mathcal{P}(\Delta)$, Lemma 2 directly characterizes the optimal decision rule. Specifically, if the boundary set $B(v)$ has a zero (Case 1) or non-zero (Case 2) probability measure, $\mathbf{p}^*(\delta^*)$ is achieved by a deterministic LRQ or (possibly) a randomized LRQ, respectively. In the latter case, the LRQ can be written as a randomization among at most $M(M-1)$ rules that are constructed using the same weight vector v [19].

Conversely, if the optimal pairwise probability vector resides in the interior of $\mathcal{P}(\Delta)$, a randomization of two LRQs is required. Firstly, let $B(v)$ have a zero probability measure. Under this assumption, the boundary of $\mathcal{P}(\Delta)$ consists entirely of extreme points, which are achieved by the deterministic rules defined in (34). The optimal decision rule that attains $\mathbf{p}^*(\delta^*)$ is a randomization of two deterministic LRQs characterized by different weight vectors. To see this, let $\mathbf{p}^1(\delta^1)$ be a point on the boundary achieved by rule δ^1 with weight vector v^1 . A ray originating from $\mathbf{p}^1(\delta^1)$ and passing through $\mathbf{p}^*(\delta^*)$ intersects the boundary at another point $\mathbf{p}^2(\delta^2)$. Since $\mathbf{p}^2(\delta^2)$ is also an extreme point, it corresponds to a deterministic rule δ^2 with weight vector v^2 . Thus, $\mathbf{p}^*(\delta^*)$ can be expressed as the convex combination:

$$\mathbf{p}^*(\delta^*) = \lambda \mathbf{p}^1(\delta^1) + (1 - \lambda) \mathbf{p}^2(\delta^2), \quad (47)$$

where $\lambda \in [0, 1]$ is the randomization coefficient. The corresponding optimal decision rule is then

$$\delta^* = \lambda \delta^1 + (1 - \lambda) \delta^2, \quad (48)$$

where δ^1 and δ^2 are selected with probabilities λ and $1 - \lambda$, respectively. An example for binary hypothesis testing is illustrated in Figure 3.

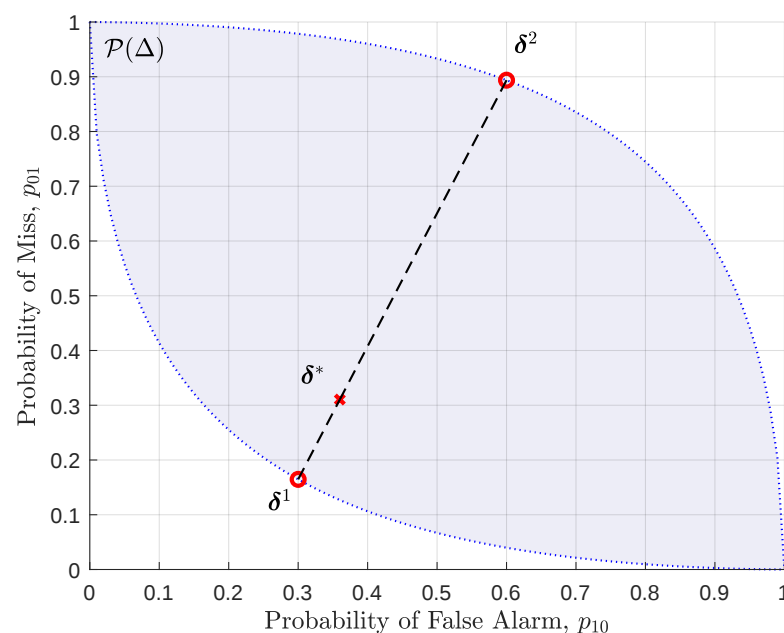


Figure 3. Set of achievable pairwise probability vectors and an example of Theorem 1 for Y with a continuous distribution (Theorem 1-Case 1).

Now, assume that $B(v)$ has a non-zero probability measure (Theorem 1, Case 2) and the p^* is in the interior of $\mathcal{P}(\Delta)$. The optimal pairwise probability vector $p^*(\delta^*)$ is achieved via a randomization between two LRQs with different weight vectors; one of these LRQs can always be chosen as a deterministic LRQ. To show this, let $p^1(\delta^1)$ be an extreme point of $\mathcal{P}(\Delta)$ achieved by decision rule δ^1 characterized by the weight vector v^1 . A ray originating from $p^1(\delta^1)$ and passing through $p^*(\delta^*)$ intersects the boundary of $\mathcal{P}(\Delta)$ at a point $\tilde{p}$. Since points on the boundary are not necessarily extreme points in Case 2, $\tilde{p}$ can be expressed as a convex combination of at most $M(M-1)$ extreme points lying on the same supporting hyperplane. Let these points be $p^2(\delta^2), \dots, p^{M(M-1)+1}(\delta^{M(M-1)+1})$, achieved by deterministic LRQs $\delta^2, \dots, \delta^{M(M-1)+1}$. The optimal pairwise probability vector $p^*(\delta^*)$ can then be written as a convex combination of these $M(M-1) + 1$ extreme points:

$$p^*(\delta^*) = \sum_{k=1}^{M(M-1)+1} \lambda_k p^k(\delta^k), \quad (49)$$

where $\lambda_k \in [0, 1]$ and $\sum_{k=1}^{M(M-1)+1} \lambda_k = 1$. Consequently, the optimal decision rule is a randomized LRQ, which can be written as a randomization of at most $M(M-1) + 1$ deterministic rules as given below:

$$\delta^* = \sum_{k=1}^{M(M-1)+1} \lambda_k \delta^k, \quad (50)$$

where each deterministic LRQ, δ^k , is employed with probability λ_k . Notably, since the pairwise probability vectors $p^2(\delta^2), \dots, p^{M(M-1)+1}(\delta^{M(M-1)+1})$ lie on the same supporting hyperplane, their corresponding decision rules can be constructed using the same weight vector v .

An example illustrating Case 2 of Theorem 1 in a binary hypothesis-testing scenario is depicted in Figure 4 for binary hypothesis testing. Note that LRQs become LRTs for binary hypothesis testing. In this instance, the optimal decision rule is a randomization of two LRQs, δ^1 and δ^{23} . The pairwise probability vector of δ^1 is an extreme point; hence, δ^1 is a deterministic LRQ. On the other hand, the pairwise probability vector of δ^{23} reside at the line segment between the two extreme points that correspond to the deterministic LRQs, δ^2 and δ^3 . Hence, the pairwise probability vector of δ^{23} can be written as a convex combination of these extreme points and thus, δ^{23} is a randomized LRQ that randomizes δ^2 and δ^3 . Note that, since the pairwise probability vectors corresponding to δ^2 and δ^3 lie on the same supporting hyperplane, these two rules are constructed using the same weight vector v . The only distinction between δ^2 and δ^3 is the specific hypothesis selected when the likelihood ratio exactly equals the threshold value. This event occurs with non-zero probability in this case. It should be noted that the decision rule, δ^1 , can always be selected such that it corresponds to an extreme point on a different part of the boundary and is therefore characterized by a different weight vector.

Special Cases

In this section, we demonstrate that optimal decision rules for standard performance metrics, such as Bayes risk and the NP criterion, can be characterized as special cases of the framework presented in the preceding part.

Bayesian M -ary Hypothesis Testing:

We first consider the *Bayesian M -ary hypothesis-testing* problem. By rearranging the summation terms, the Bayes risk in (6) can be expressed as a linear function of the pairwise probability vector:

$$\begin{aligned}
 r(\delta) &= \sum_{j=0}^{M-1} \sum_{\substack{i=0 \\ i \neq j}}^{M-1} \pi_j (c_{ij} - c_{jj}) p_{ij} + \sum_{j=0}^{M-1} \pi_j c_{jj} \\
 &= \mathbf{v}^\top \mathbf{p}(\delta) + \text{constant},
 \end{aligned} \tag{51}$$

where the weights are defined as $v_{ij} \triangleq \pi_j (c_{ij} - c_{jj})$. As established in the previous section, the decision rule that minimizes a performance metric of the form $\mathbf{v}^\top \mathbf{p}(\delta)$ can be expressed as the deterministic rule defined in (34).

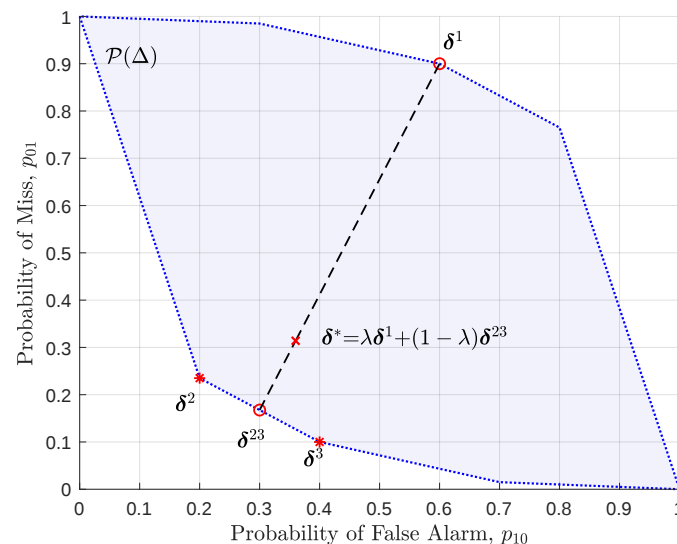


Figure 4. Set of achievable pairwise probability vectors and an example of Theorem 1 for $\mathbf{Y}$ with a discrete distribution (Theorem 1-Case 2).

Binary Bayesian Hypothesis-Testing

For binary hypothesis testing ($M = 2$), the decision rule in (34) simplifies to the following comparison:

$$V_1(\mathbf{y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} V_0(\mathbf{y}). \tag{52}$$

Substituting $V_0(\mathbf{y}) = v_{01}f_1(\mathbf{y})$ and $V_1(\mathbf{y}) = v_{10}f_0(\mathbf{y})$, the optimal Bayesian rule becomes a standard LRT:

$$\delta_1^B(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq \frac{v_{10}}{v_{01}}, \\ 0, & \text{otherwise,} \end{cases} \tag{53}$$

where $v_{10} = \pi_0(c_{10} - c_{00})$ and $v_{01} = \pi_1(c_{01} - c_{11})$. Next, we focus on the probability of error performance metric, which can be viewed as a special case of Bayes risk with $c_{10} = c_{01} = 1$ and $c_{00} = c_{11} = 0$. We derive the lower bound for the probability of error and show that the lower bound can be achieved via the MAP rule.

Remark 2. Consider the scenario in which the performance metric is the minimum probability of error. The probability of error associated with the decision rule δ can be expressed as

$$P_e(\mathbf{p}(\delta)) = \pi_1 p_{01} + \pi_0 p_{10}. \tag{54}$$

The lower bound on the probability of error is given by

$$P_e(\mathbf{p}(\delta)) \geq \pi_1 - d_v(\pi_0 f_0(\mathbf{y}), \pi_1 f_1(\mathbf{y})), \tag{55}$$

where $d_v(\pi_0 f_0(\mathbf{y}), \pi_1 f_1(\mathbf{y}))$ denotes the total variation distance and is computed as [30]:

$$d_v(\pi_0 f_0(\mathbf{y}), \pi_1 f_1(\mathbf{y})) = \sup_{\delta \in \Delta} [\pi_1 p_{11} - \pi_0 p_{10}]. \quad (56)$$

The equality in (55) is achieved if the decision rule that yields the minimum probability of error is employed. This optimal decision rule can be obtained by solving the following optimization problem:

$$\delta^* = \arg \min_{\delta \in \Delta} P_e(\mathbf{p}(\delta)). \quad (57)$$

Since the probability of error is a function of the pairwise probabilities, this optimization problem can equivalently be expressed as

$$\mathbf{p}^*(\delta^*) = \arg \min_{\mathbf{p} \in \mathcal{P}(\Delta)} P_e(\mathbf{p}). \quad (58)$$

By denoting $\mathbf{v} = [\pi_0, \pi_1]^T$, the probability of error can compactly be written as $P_e(\mathbf{p}(\delta)) = \mathbf{v}^T \mathbf{p}(\delta)$. Consequently, the decision rule that minimizes the probability of error and achieves the lower bound in (55) is the LRT given by

$$\delta^*(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq \frac{\pi_0}{\pi_1}, \\ 0, & \text{otherwise.} \end{cases} \quad (59)$$

It should be noted that the decision rule specified above is the MAP rule given in (19). An illustrative example with the location of the pairwise probability vector of the decision rule given above, which results in a minimum probability of error value of 0.2266, is shown in Figure 5 with a red circle.

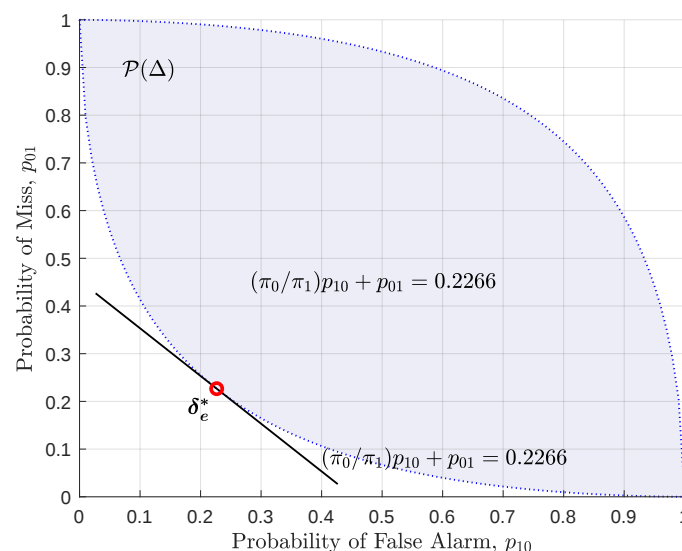


Figure 5. The optimal decision rule and the pairwise probability vector that minimizes the probability of error (Remark 2).

NP Hypothesis Testing

In *NP hypothesis testing* ($M = 2$), the objective is to maximize the detection probability p_{11} while maintaining the false alarm probability p_{10} below a specified level α . Since $p_{01} = 1 - p_{11}$, the NP criterion can equivalently be expressed as the following constrained optimization:

$$\min_{\delta \in \Delta} p_{01} \quad \text{subject to} \quad p_{10} \leq \alpha. \quad (60)$$

Since the miss probability p_{01} is a non-increasing function of the false alarm probability p_{10} , the optimal decision rule achieves the false alarm constraint with equality, $p_{10} = \alpha$, and takes the form of a randomized LRT. By definition, the NP rule achieves the minimum possible miss probability for a given false alarm level; geometrically, this implies that the pairwise probability vector of the NP rule resides on the lower boundary of $\mathcal{P}(\Delta)$. According to Lemma 2, the optimal decision rule is a randomization among at most two deterministic decision rules, both associated with the same weight vector v .

The lower boundary of the set $\mathcal{P}(\Delta)$ consists of pairwise probability vectors corresponding to standard NP decision rules. To characterize the upper boundary, we introduce the *flipped NP* criterion ([20] Lemma 2), which is provided in the lemma given below.

Lemma 3 ([20] Lemma 2 (Flipped Neyman–Pearson Rule)). *Let the observation vector $\mathbf{y}$ be distributed according to $f_0(\mathbf{y})$ and $f_1(\mathbf{y})$ under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. For any $0 \leq \alpha \leq 1$, let δ^{FNP} be a decision rule c.f. (22):*

$$\delta_1^{\text{FNP}}(\mathbf{y}) = \begin{cases} 1, & \text{if } f_1(\mathbf{y}) < \eta_0 f_0(\mathbf{y}) \\ \gamma_0, & \text{if } f_1(\mathbf{y}) = \eta_0 f_0(\mathbf{y}) \\ 0, & \text{if } f_1(\mathbf{y}) > \eta_0 f_0(\mathbf{y}) \end{cases} \quad (61)$$

where $0 \leq \gamma_0 \leq 1$ and $\eta_0 \geq 0$ are chosen such that the false alarm probability $p_{10} = \alpha$. Then, δ^{FNP} maximizes the miss probability p_{01} among all rules δ satisfying $p_{10} \geq \alpha$. That is, δ^{FNP} solves:

$$\max_{\delta \in \Delta} p_{01} \quad \text{subject to} \quad p_{10} \geq \alpha. \quad (62)$$

Under the flipped NP criterion, the objective is to maximize the miss probability p_{01} while maintaining the false alarm probability p_{10} at a specific level α . Since the miss probability is a non-increasing function of the false alarm probability, the FNP rule achieves the constraint $p_{10} = \alpha$ with equality. By maximizing p_{01} for a fixed p_{10} , the pairwise probability vectors of the FNP rules define the upper boundary of $\mathcal{P}(\Delta)$. According to Lemma 2, the FNP rule can be implemented as a randomization between at most two deterministic decision rules sharing the same weight vector v . Hence, the flipped NP rule given in Lemma 3 and the characterization of the decision rule done by using Lemma 2 agree.

For a fixed false alarm probability α , the lower and upper boundaries of the set $\mathcal{P}(\Delta)$ correspond to the minimum and maximum achievable miss probabilities, respectively. These extremes are achieved by the NP and flipped NP rules. Consequently, any arbitrary point $\tilde{\mathbf{p}} = [\tilde{p}_{10}, \tilde{p}_{01}]^\top \in \mathcal{P}(\Delta)$ can be represented as a convex combination of the NP and flipped NP pairwise probability vectors, both evaluated at a false alarm probability $\alpha = \tilde{p}_{10}$:

$$\tilde{\mathbf{p}} = \lambda \mathbf{p}^{\text{NP}}(\delta^{\text{NP}}) + (1 - \lambda) \mathbf{p}^{\text{FNP}}(\delta^{\text{FNP}}), \quad (63)$$

where the randomization coefficient is given by $\lambda = \frac{\tilde{p}_{01} - p_{01}^F}{p_{01}^N - p_{01}^F}$. Thus, the decision rule $\tilde{\delta}$ achieving $\tilde{\mathbf{p}}(\tilde{\delta})$ written as the randomization of NP and flipped NP rules is given below:

$$\tilde{\delta} = \lambda \delta^{\text{NP}} + (1 - \lambda) \delta^{\text{FNP}}, \quad (64)$$

where the NP rule δ^{NP} and the flipped NP rule δ^{FNP} are employed with probabilities λ and $1 - \lambda$, respectively. For binary hypothesis testing, every point in $\mathcal{P}(\Delta)$ is uniquely characterized by the false alarm probability α and the randomization constant λ . As a result, the entire set $\mathcal{P}(\Delta)$ can be spanned by sweeping $\alpha \in [0, 1]$ and $\lambda \in [0, 1]$.

4. Binary Hypothesis Testing with Prospect Theory-Based Behavioral Utility as Performance Metric

The rational performance metrics, such as Bayes risk and NP criterion, utilize the actual probability and utility or cost values; as a result, direct analysis of these performance metrics often reveals the structure of the optimal decision rules. Furthermore, the optimal pairwise probability vectors of rational performance metrics lie on the boundary of the $\mathcal{P}(\Delta)$ set. In some hypothesis-testing problems, however, the DA or the designer of the decision rule may exhibit behavioral biases and therefore may not act as a fully rational agent, such as a physical sensor. In such cases, rational performance metrics may fail to capture the actual decision-making behavior of the DA. To model such behavior, PT-based approaches have been adopted in the literature [13–21], which employ a probability weighting function and a utility valuation function [22–26]. Since the probability weighting function introduces a nonlinear transformation of the objective probabilities, the resulting behavioral utility is generally nonlinear in the pairwise probability vector; therefore, unlike classical criteria, its optimizer need not lie on the boundary of $\mathcal{P}(\Delta)$.

In this section, we focus on the problem of binary hypothesis testing with a DA whose performance metric is the PT-based behavioral utility that is computed using the perceived probability and utility values that are obtained using the probability weighting and utility valuation functions, which are specified in the following.

4.1. Prospect Theory (PT)

Originally developed by Daniel Kahneman and Amos Tversky in 1979, PT provides a robust framework for understanding how behavioral biases influence decision-making under uncertainty. In PT, it is posited that the behavioral biases of human DAs are driven by subjective perceptions of probability and utility. Specifically, unlike rational agents, such as physical sensors that operate on objective probability and utility values, human decision-makers perceive these quantities through subjective probability weighting and utility valuation functions [22,23,26,30].

To model the human perception of probabilities, a probability weighting function, $w(p)$, is employed in PT. A widely adopted formulation for this function is given by

$$w(p) = \frac{p^\vartheta}{(p^\vartheta + (1-p)^\vartheta)^{1/\vartheta}}, \quad (65)$$

where $p \in [0, 1]$ represents the objective probability and $w(p) \in [0, 1]$ denotes its perceived version [31]. Here, ϑ serves as a tuning parameter, which satisfies $\vartheta > 0.279$ [25,32]. This function typically captures the human tendency to *overweigh low probabilities* and *underweigh high probabilities*. Examples are shown in Figure 6 for $\vartheta \in \{0.49, 0.69, 0.89, 1\}$.

The subjective perception of an outcome's utility is modeled by a utility valuation function, $v(u)$, defined relative to a reference point u_r as given below:

$$v(u) = \begin{cases} (u - u_r)^\kappa, & \text{if } u \geq u_r \\ -\beta(u_r - u)^\kappa, & \text{if } u < u_r, \end{cases} \quad (66)$$

where u represents the objective utility viewed as a gain if $u > u_r$ and a loss if $u < u_r$ [23]. Here, $\beta > 1$ is the *loss-aversion* parameter, and $\kappa \in (0, 1)$ reflects *diminishing sensitivity*. The median behavioral parameters are estimated as $\beta = 2.25$ and $\kappa = 0.88$ [23,26]. The resulting utility valuation function, characterized by a steeper slope for losses than for gains, is depicted in Figure 7 for the standard case of $u_r = 0$.

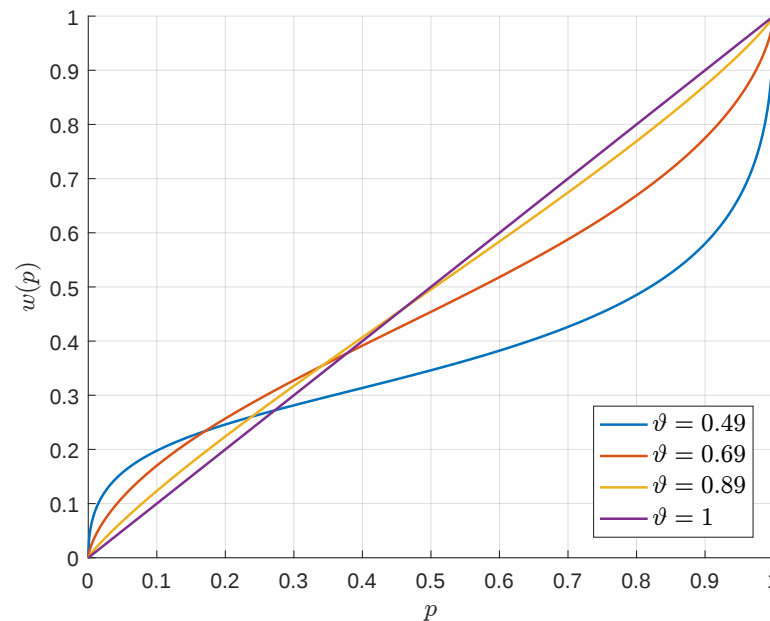


Figure 6. The probability weighting function in (65) for various values of $\vartheta \in \{0.49, 0.69, 0.89, 1\}$.

The median values for the behavioral parameters are estimated as $\vartheta = 0.69$, $\beta = 2.25$, and $\kappa = 0.88$ [23,26]. Furthermore, the behavioral parameters for a rational agent are $\vartheta = 1$, $\beta = 1$, and $\kappa = 1$. In Figure 6, the probability weighting function is depicted for various values of $\vartheta \in \{0.49, 0.69, 0.89, 1\}$. Similarly, the utility valuation function for a behavioral decision-maker with $\beta = 2.25$, $\kappa = 0.88$, and $u_r = 0$ is provided in Figure 7.

4.2. Optimal Decision Rules for Behavioral Utility-Based Binary Hypothesis Testing

We assume that a behaviorally-biased DA, i.e., a human DA, is employed for binary hypothesis testing, where the conditional distributions of the observation vector $\mathbf{Y}$ are given by

$$\mathcal{H}_0 : \mathbf{Y} \sim f_0 \quad (67)$$

$$\mathcal{H}_1 : \mathbf{Y} \sim f_1 \quad (68)$$

The DA evaluates the performance using a behavioral utility metric, expressed as

$$\mathcal{U}(\mathbf{p}(\delta)) = \sum_{j=0}^1 \sum_{i=0}^1 w(p_{ij}\pi_j)v(u_{ij}) \quad (69)$$

where π_j is the prior probability of hypothesis $\mathcal{H}_j$, and u_{ij} denotes the actual utility associated with choosing $\mathcal{H}_i$ when $\mathcal{H}_j$ is true, which can be considered as the negative of the cost value, c_{ij} , in (6). Here, w and v represent the probability weighting and utility valuation functions, respectively (see (65) and (66)).

The behavioral utility performance metric in (69) utilizes perceived probability and utility values rather than their objective counterparts. Consequently, it can be viewed as a behavioral variant of the Bayes risk (with a reverse sign, as it is defined as utility). The fundamental differences are: (i) instead of absolute costs, the metric employs utility values u_{ij} , where negative values represent perceived losses and positive values represent perceived gains, and (ii) unlike a rational agent who utilizes actual probability and cost values, the behavioral agent processes these through non-linear functions w and v , capturing human-like cognitive biases.

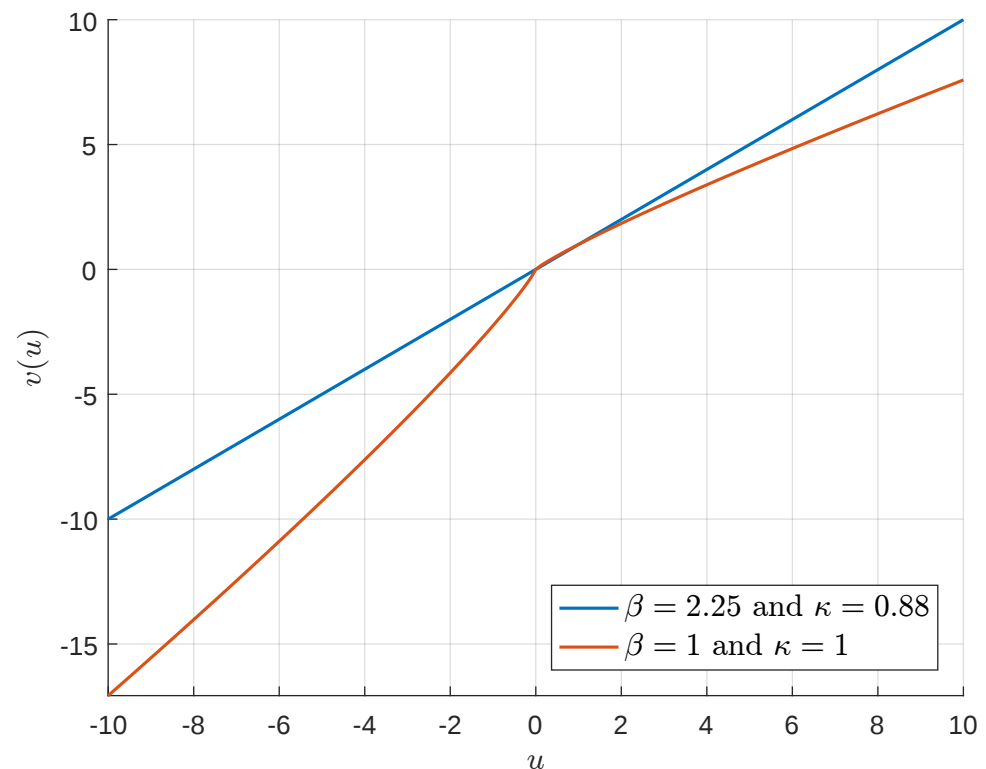


Figure 7. The utility valuation function with parameters $u_r = 0$, $\beta = 2.25$, and $\kappa = 0.88$.

For this problem, the optimal decision rule maximizes the behavioral utility and is obtained by solving the following optimization problem:

$$\delta^* = \arg \max_{\delta \in \Delta} \mathcal{U}(\mathbf{p}(\delta)). \quad (70)$$

Expanding the behavioral utility in (69) in terms of the pairwise error probabilities for the binary case ($p_{00} = 1 - p_{10}$ and $p_{11} = 1 - p_{01}$), we obtain

$$\begin{aligned} \mathcal{U}(\mathbf{p}(\delta)) = & w(\pi_0(1 - p_{10}))v(u_{00}) + w(\pi_0 p_{10})v(u_{10}) \\ & + w(\pi_1 p_{01})v(u_{01}) + w(\pi_1(1 - p_{01}))v(u_{11}). \end{aligned} \quad (71)$$

A decision rule, δ , influences the behavioral utility solely through its pairwise probability vector, $\mathbf{p}(\delta)$. Consequently, maximizing the behavioral utility over the set of all randomized decision rules, Δ , is equivalent to maximizing the function $\mathcal{U}(\mathbf{p})$ over the set of all achievable pairwise probability vectors, $\mathcal{P}(\Delta)$:

$$\mathbf{p}^*(\delta^*) = \arg \max_{\mathbf{p} \in \mathcal{P}(\Delta)} \mathcal{U}(\mathbf{p}). \quad (72)$$

The optimization of behavioral utility is thus a specific instance of the general problem formulated in (25). Therefore, the optimal decision rules can be characterized using Theorem 1. Notably, as detailed in ([16] Proposition 1), certain utility configurations can result in scenarios where a standard LRT is no longer the optimal strategy. ([16] Proposition 1) is summarized in the following proposition.

Proposition 1 ([16] Proposition 1). *Suppose that the probability weighting function w is monotonically increasing.*

- *Case (a): If $v(u_{10})v(u_{00}) < 0$ or $v(u_{11})v(u_{01}) < 0$, then the solution to (72) is a randomized LRT. This rule takes the form of either a standard NP rule, δ^{NP} , or a flipped NP rule, δ^{FNP} :*

$$\delta_1^{\text{NP}}(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > \eta_1 \\ \gamma_1, & \text{if } L(\mathbf{y}) = \eta_1 \\ 0, & \text{if } L(\mathbf{y}) < \eta_1 \end{cases} \quad (73)$$

$$\delta_1^{\text{FNP}}(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) < \eta_2 \\ \gamma_2, & \text{if } L(\mathbf{y}) = \eta_2 \\ 0, & \text{if } L(\mathbf{y}) > \eta_2 \end{cases} \quad (74)$$

- *Case (b): If $v(u_{10})v(u_{00}) \geq 0$ and $v(u_{11})v(u_{01}) \geq 0$, then the optimal decision rule may require a randomization of multiple LRTs.*

Unlike Bayesian risk, which utilizes objective probability and cost values, the optimality of a single LRT cannot be guaranteed in behavioral utility-based binary hypothesis testing. This is primarily due to the non-linearity of the probability weighting function, which overweights low probabilities and underweights high probabilities. Consequently, the position of the optimal pairwise probability vector $\mathbf{p}^*(\delta^*)$ within the achievable set $\mathcal{P}(\Delta)$ depends heavily on the signs of the utility valuation functions. When the conditions for Case (a) are met, i.e., $v(u_{10})v(u_{00}) < 0$ or $v(u_{11})v(u_{01}) < 0$ holds, the optimal pairwise probability vector $\mathbf{p}^*(\delta^*)$ resides on the boundary of $\mathcal{P}(\Delta)$. In this scenario, Lemma 2 applies directly, and the optimal rule is a randomization among at most two deterministic decision rules constructed with the same weight vector $\mathbf{v}$.

Conversely, if the conditions $v(u_{10})v(u_{00}) \geq 0$ and $v(u_{11})v(u_{01}) \geq 0$ holds, i.e., Case (b) is valid, the optimal vector $\mathbf{p}^*(\delta^*)$ may reside in the interior of the $\mathcal{P}(\Delta)$ set. Under these circumstances, the structure of the optimal rule depends on the measure of the boundary set $\mathcal{B}(\mathbf{v})$. If the boundary points of $\mathbf{y}$ occur with a nonzero probability measure, the optimal decision rule is a randomization between two possibly randomized LRTs with distinct weight vectors. More explicitly, one of these LRTs can always be chosen as a deterministic LRT; however, the other LRT may be a randomized LRT, which randomizes two LRTs with the same weight vector. Since the optimization problem given in (72) is a special case of the general framework given in (25), by leveraging Theorem 1, the following corollary can be devised to characterize the optimal decision rules for behavioral utility.

Corollary 1. *For PT-based binary hypothesis testing, the optimal decision rule can be expressed as a randomization among at most three deterministic decision rules based on the likelihood ratio $L(\mathbf{y}) = f_1(\mathbf{y})/f_0(\mathbf{y})$. The first decision rule can always be chosen as*

$$\delta^1(\mathbf{y}) = \begin{cases} 1, & L(\mathbf{y}) > \eta_1 \\ 0 \text{ or } 1, & L(\mathbf{y}) = \eta_1 \\ 0, & L(\mathbf{y}) < \eta_1 \end{cases} \quad (75)$$

and the other two decision rules form either an NP rule or a flipped NP rule. If they form an NP rule, they are characterized as

$$\delta^2(\mathbf{y}) = \begin{cases} 1, & L(\mathbf{y}) \geq \eta_2 \\ 0, & L(\mathbf{y}) < \eta_2 \end{cases} \quad (76a)$$

$$\delta^3(\mathbf{y}) = \begin{cases} 1, & L(\mathbf{y}) > \eta_2 \\ 0, & L(\mathbf{y}) \leq \eta_2 \end{cases} \quad (76b)$$

Conversely, if the other two decision rules form a flipped NP rule, they are characterized as

$$\delta^2(\mathbf{y}) = \begin{cases} 1, & L(\mathbf{y}) \leq \eta_3 \\ 0, & L(\mathbf{y}) > \eta_3 \end{cases} \quad (77a)$$

$$\delta^3(\mathbf{y}) = \begin{cases} 1, & L(\mathbf{y}) < \eta_3 \\ 0, & L(\mathbf{y}) \geq \eta_3 \end{cases} \quad (77b)$$

This corollary is a direct application of Theorem 1 to the behavioral utility. Geometrically, let δ^1 be a deterministic decision rule associated with a weight vector $\mathbf{v}^1$. Due to its structure, its pairwise probability vector resides on the lower boundary of $\mathcal{P}(\Delta)$. A ray originating from its pairwise probability vector $\mathbf{p}^1(\delta^1)$ intersects the boundary of the set $\mathcal{P}(\Delta)$ at a point that is not necessarily an extreme point of the set. If this intersection point is located on the lower boundary of $\mathcal{P}(\Delta)$, it can be achieved via an NP rule; thus, a randomization of the decision rules in (76) yields the intersection point. Conversely, if the intersection point is located on the upper boundary of $\mathcal{P}(\Delta)$, the randomization of the decision rules in (77) is required, i.e., the intersection point corresponds to a pairwise probability vector of a flipped NP rule.

If the observation $\mathbf{Y}$ is governed by a continuous distribution such that the boundary points occur with zero probability measure, every point on the boundary of $\mathcal{P}(\Delta)$ becomes an extreme point. In such cases, the optimal decision rule simplifies to a randomization of at most two deterministic LRTs with different weight vectors. Moreover, if Case 1 of Proposition 1 holds, the optimal decision rule is either a deterministic LRT in the case where the boundary points of the $\mathbf{Y}$ occur with zero probability or a randomized LRT that can be written as a randomization of two deterministic LRTs with the same weight vector, $\mathbf{v}$, in the case the boundary points of $\mathbf{Y}$ occurring with non-zero probabilities.

An alternative approach for characterizing the optimal decision rule that maximizes behavioral utility is presented in ([16] Proposition 2) and also in ([20] Proposition 1), which are provided in the proposition given below. Based on the geometric structure of the achievable set, the boundary of $\mathcal{P}(\Delta)$ consists of points corresponding to either NP rules or flipped NP rules. Consequently, the optimal pairwise probability vector $\mathbf{p}^*$ must be achieved by one of the following: (i) a standard NP rule, (ii) a flipped NP rule, or (iii) a randomization between an NP rule and a flipped NP rule that share the same false alarm probability α . This third case effectively allows for the exploration of the interior of $\mathcal{P}(\Delta)$ by taking convex combinations of the lower and upper boundaries at any given false-alarm level.

Proposition 2 ([16] Proposition 2). Let $\mathbf{p}^*(\delta^*) = [p_{10}^*, p_{01}^*]^\top$ be the optimal pairwise probability vector that solves the behavioral utility maximization problem in (72). The corresponding optimal decision rule δ^* can be implemented as a randomized rule:

$$\delta^* = \lambda \delta^{\text{NP}} + (1 - \lambda) \delta^{\text{FNP}}, \quad (78)$$

where δ^{NP} and δ^{FNP} are the NP and flipped NP rules, respectively, both designed with size $\alpha = p_{10}^*$. Let $p_{01}^{N,*}$ and $p_{01}^{F,*}$ denote the miss probabilities associated with δ^{NP} and δ^{FNP} at this false alarm level. The randomization coefficient is then given by

$$\lambda = \frac{p_{01}^{F,*} - p_{01}^{N,*}}{p_{01}^{F,*} - p_{01}^{N,*}}. \quad (79)$$

Specifically, the optimal rule δ^* selects the decision outcome of δ^{NP} with probability λ and the outcome of δ^{FNP} with probability $1 - \lambda$.

The fundamental intuition behind Proposition 2 is that any point $\tilde{\mathbf{p}} = [\tilde{p}_{10}, \tilde{p}_{01}]^\top$ within the achievable set $\mathcal{P}(\Delta)$ can be realized through a specific randomized decision rule:

$$\tilde{\delta}(\mathbf{y}) = \lambda \delta^{\text{NP}} + (1 - \lambda) \delta^{\text{FNP}}, \quad (80)$$

where δ^{NP} and δ^{FNP} are the NP and flipped NP decision rules of size $\alpha = \tilde{p}_{10}$, respectively. In this framework, the randomization constant λ dictates the probability with which the NP rule is selected over the flipped NP rule. Geometrically, the lower and upper boundaries of the set $\mathcal{P}(\Delta)$ are defined by the NP and flipped NP rules across all possible sizes. Consequently, if the optimal pairwise probability vector $\mathbf{p}^*(\delta^*)$ lies on the lower boundary, we set $\lambda = 1$ to recover the standard NP rule of size $\tilde{p}_{10}$. Conversely, if $\mathbf{p}^*(\delta^*)$ resides on the upper boundary, the randomization constant is set to $\lambda = 0$ to employ the flipped NP rule. For the case where $\mathbf{p}^*(\delta^*)$ is located in the interior of $\mathcal{P}(\Delta)$, the randomization of these two boundary rules with $\lambda = (p_{01}^{F,*} - p_{01}^*) / (p_{01}^{F,*} - p_{01}^{N,*})$ ensures that the desired miss probability is achieved at the target false alarm level.

The approach given in Proposition 2 effectively "scans" the vertical interior of the achievable set, providing a robust implementation of optimal rules. Furthermore, using Proposition 2, it is possible to devise a practical algorithm to obtain the decision rule that maximizes the behavioral utility, which consists of the following steps:

1. **Solve for $\mathbf{p}^*(\delta^*)$:** Solve the optimization problem in (72) to obtain the optimal pairwise probability vector $\mathbf{p}^*(\delta^*) = [p_{10}^*, p_{01}^*]^\top$.
2. **Design Boundary Rules:** Design the standard NP and flipped NP rules as LRTs defined in (73) and (74), both set to a false alarm probability of $\alpha = p_{10}^*$.
3. **Randomize:** Combine these two rules using the randomization constant λ derived in Proposition 2 to achieve the target miss probability p_{01}^* .

5. Numerical Examples

In this section, we provide several numerical examples to corroborate the theoretical results established in the previous sections based on Theorem 1. To this end, we consider a behaviorally biased DA employed to detect the true state of a POI. The POI resides in one of two states, denoted by hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, which occur with prior probabilities $\pi_0 = \pi_1 = 0.5$. The performance of the behavioral DA is evaluated using the behavioral utility function defined in (69).

In the first example, we assume that the DA receives binary reports from two sources; specifically, these sources send their local observations regarding the actual state of the POI to the DA. Consequently, the set of observation vectors available to the DA is $\mathcal{Y} = \{[0, 0]^\top, [0, 1]^\top, [1, 0]^\top, [1, 1]^\top\}$. The conditional probability mass functions (PMFs) under $\mathcal{H}_0$ and $\mathcal{H}_1$ are provided in Table 1. The perceived utilities and the probability weighting parameter of the behavioral DA are set to $v(u_{00}) = 1.1$, $v(u_{01}) = 0.6$, $v(u_{10}) = 0.5$, $v(u_{11}) = 1$, and $\vartheta = 0.45$ (see (65) and (69)).

Table 1. The conditional probability mass functions of $\mathbf{Y}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$.

True Hypothesis	$\mathbf{y} = [0, 0]^\top$	$\mathbf{y} = [0, 1]^\top$	$\mathbf{y} = [1, 0]^\top$	$\mathbf{y} = [1, 1]^\top$
$\mathcal{H}_0$	0.6	0.15	0.2	0.05
$\mathcal{H}_1$	0.015	0.135	0.085	0.765

In Figure 8, we highlight the boundary of $\mathcal{P}(\Delta)$ with solid blue lines. The maximum value of the behavioral utility $\mathcal{U} = 0.8181$ is achieved with a pairwise probability vector of $\mathbf{p}^*(\delta^*) = [0.17, 0.2486]^\top$, which is shown with a red x sign. This pairwise probability vector can be obtained by using a randomized decision rule that is given below:

$$\delta^* = \lambda^1 \delta^1 + \lambda^2 \delta^2 + \lambda^3 \delta^3, \quad (81)$$

where $\lambda^1 = 0.0977$, $\lambda^2 = 0.4605$, and $\lambda^3 = 0.4418$, and the decision rules δ^1 , δ^2 and δ^3 are specified as follows:

$$\delta_1^1 = \begin{cases} 1 & \text{if } L(\mathbf{y}) < 0.425 \\ 0 & \text{if } L(\mathbf{y}) \geq 0.425 \end{cases} \quad (82a)$$

$$\delta_1^2 = \begin{cases} 1 & \text{if } L(\mathbf{y}) > 0.9 \\ 0 & \text{if } L(\mathbf{y}) \leq 0.9 \end{cases} \quad (82b)$$

$$\delta_1^3 = \begin{cases} 1 & \text{if } L(\mathbf{y}) \geq 0.9 \\ 0 & \text{if } L(\mathbf{y}) < 0.9. \end{cases} \quad (82c)$$

Since δ^2 and δ^3 are constructed with the same weight vector, the optimal decision rule given in (81) can be written as

$$\delta^* = \lambda^1 \delta^1 + \lambda^{23} \delta^{23}, \quad (83)$$

where $\lambda^1 = 0.0977$, $\lambda^{23} = 0.9023$, and δ^{23} is

$$\delta_1^{23} = \begin{cases} 1 & \text{if } L(\mathbf{y}) > 0.9 \\ 0.5103 & \text{if } L(\mathbf{y}) = 0.9 \\ 0 & \text{if } L(\mathbf{y}) < 0.9, \end{cases} \quad (84)$$

which randomly selects δ^2 or δ^3 with probabilities 0.5103 and 0.4897, respectively. The pairwise probability vectors achieved by the decision rules specified above are marked in Figure 8.

Alternatively, the decision rules given above can be expressed as

$$\delta_1^1 = \begin{cases} 1 & \text{if } \mathbf{y} = [0, 0]^\top \\ 0 & \text{otherwise} \end{cases} \quad (85a)$$

$$\delta_1^{23} = \begin{cases} 1 & \text{if } \mathbf{y} = [1, 1]^\top \\ 0.5103 & \text{if } \mathbf{y} = [0, 1]^\top \\ 0 & \text{if } \mathbf{y} \in \{[0, 0]^\top, [1, 0]^\top\}. \end{cases} \quad (85b)$$

where δ^{23} can be written as a randomization δ^2 and δ^3 , which are given below:

$$\delta_1^2 = \begin{cases} 1 & \text{if } \mathbf{y} = [1, 1]^\top \\ 0 & \text{otherwise} \end{cases} \quad (86a)$$

$$\delta_1^3 = \begin{cases} 1 & \text{if } \mathbf{y} \in \{[0, 1]^\top, [1, 1]^\top\} \\ 0 & \text{if } \mathbf{y} \in \{[0, 0]^\top, [1, 0]^\top\}. \end{cases} \quad (86b)$$

Note that $\mathbf{y} = [0, 1]^\top$ is a boundary point of $\mathbf{y}$ for $\mathbf{v}^{23} = [1, 0.9]^\top$.

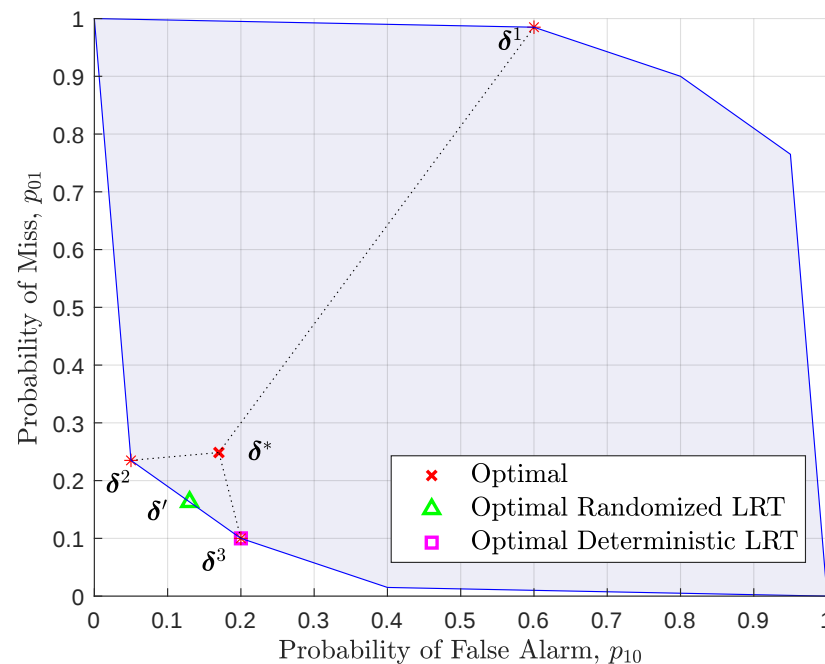


Figure 8. Optimal pairwise probability vectors maximizing the behavioral utility for a case where the optimal solution resides in the *interior* of the feasible set $\mathcal{P}(\Delta)$ shown by the shaded area.

For comparison, we assume a scenario where the DA is restricted to employing either a randomized LRT (i.e., an NP or flipped NP rule) or a deterministic decision rule as defined in (11). In this case, the achievable pairwise probability vectors are limited to the boundary and the extreme points of the set $\mathcal{P}(\Delta)$, respectively. When a randomized decision rule is employed, the maximum behavioral utility is $\mathcal{U} = 0.8155$, achieved at the pairwise probability vector $\mathbf{p}(\delta) = [0.13, 0.163]^\top$. This point is represented by a green triangle in Figure 8 and corresponds to the decision rule δ' specified by

$$\delta'_1(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > 0.9 \\ 0.5333, & \text{if } L(\mathbf{y}) = 0.9, \\ 0, & \text{if } L(\mathbf{y}) < 0.9 \end{cases} \quad (87)$$

which is implemented by randomizing between the deterministic decision rules δ^2 in (82b) and δ^3 in (82c) with probabilities 0.5333 and 0.4667, respectively. Furthermore, if the DA is restricted solely to deterministic LRTs, the maximum behavioral utility drops to $\mathcal{U} = 0.8105$, which is achieved by the decision rule δ^3 defined in (82c). The corresponding pairwise probability vector is indicated by a purple square in Figure 8.

In the second example, we set the DA's probability weighting parameter to $\vartheta = 0.69$ in (65), and the perceived utility values of the DA are $v(u_{00}) = 1.3$, $v(u_{01}) = 0.4$, $v(u_{10}) = 0.4$, and $v(u_{11}) = 1.2$. The optimal pairwise probability vector that yields the maximum behavioral utility value, $\mathcal{U} = 1.1341$, is located on the boundary of the $\mathcal{P}(\Delta)$ set. Furthermore, since the boundary values of $\mathbf{Y}$ occur with non-zero probability, Lemma 2, Case (b) holds, and the optimal decision rule is a randomization of at most two deterministic decision rules. The optimal pairwise probability vector $\mathbf{p}^*(\delta^*) = [0.11, 0.181]^\top$ is indicated by a red 'x' in Figure 9, corresponding to the following decision rule:

$$\delta_1^*(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > 0.9 \\ 0.4, & \text{if } L(\mathbf{y}) = 0.9 \\ 0, & \text{if } L(\mathbf{y}) < 0.9, \end{cases} \quad (88)$$

which employs the deterministic decision rules given below with probabilities 0.4 and 0.6, respectively:

$$\tilde{\delta}_1(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) \geq 0.9 \\ 0, & \text{if } L(\mathbf{y}) < 0.9 \end{cases} \quad (89a)$$

$$\delta_1^\dagger(\mathbf{y}) = \begin{cases} 1, & \text{if } L(\mathbf{y}) > 0.9 \\ 0, & \text{if } L(\mathbf{y}) \leq 0.9. \end{cases} \quad (89b)$$

If the DA is restricted to purely deterministic decision rules, the maximum behavioral utility value drops to $\mathcal{U} = 1.1304$, achieved by the deterministic rule $\tilde{\delta}$ defined in (89a). The pairwise probability vector corresponding to $\tilde{\delta}$ is $\tilde{\mathbf{p}}(\tilde{\delta}) = [0.05, 0.235]^\top$, indicated in Figure 9 by a purple square.

In the third example, we assume that the observation of the DA has a continuous distribution under both hypotheses, and the conditional pdfs of observation are given below:

$$\mathcal{H}_0 : Y \sim \mathcal{N}(0, 1) \quad (90)$$

$$\mathcal{H}_1 : Y \sim \mathcal{N}(m, 1), \quad (91)$$

where m is the mean value of observations under $\mathcal{H}_1$. The probability weighting parameter of the DA is set to $\vartheta = 0.5$ in (65), and the perceived utilities are $v(u_{00}) = 1.2$, $v(u_{01}) = 0.4$, $v(u_{10}) = 0.6$, $v(u_{11}) = 1.1$ in (69). The maximum behavioral utility value that can be achieved with a decision rule as characterized in Theorem 1 and a deterministic decision rule as a function of $m \in [0, 10]$ is given in Figure 10 with blue solid and red dashed lines, respectively. For $m \leq 2$, the maximum behavioral utility that is achieved with decision rules characterized using Theorem 1 and deterministic decision rules is the same. For $m > 2$, the maximum behavioral utility that can be achieved with deterministic decision rules begins to drop while the maximum behavioral utility that can be achieved using the decision rules characterized in Theorem 1 stays the same.

To elucidate the performance gap between deterministic and randomized decision rules, we depict the set $\mathcal{P}(\Delta)$ for $m = 1$, $m = 2$, and $m = 3$ in Figure 11. Deterministic decision rules can only yield pairwise probability vectors that lie on the boundary of the $\mathcal{P}(\Delta)$ set; the deterministic vectors yielding the maximum behavioral utility for $m = 1, 2, 3$ are indicated by a magenta square, a green triangle, and a red star, respectively.

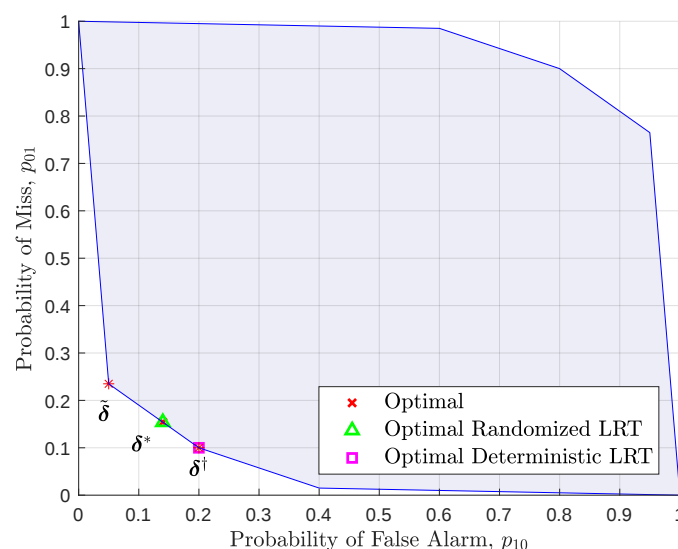


Figure 9. Optimal pairwise probability vectors maximizing behavioral utility for a case where the optimal solution lies on the boundary of $\mathcal{P}(\Delta)$.

For $m \leq 2$, the optimal pairwise probability vector is located on the boundary of the $\mathcal{P}(\Delta)$ set; thus, deterministic rules and the rules characterized via Theorem 1 achieve identical behavioral utility values. However, for $m > 2$, the optimal vector no longer resides on the boundary, rendering it unachievable via deterministic rules. As m increases, the $\mathcal{P}(\Delta)$ set expands, causing the distance between the optimal interior vector and the achievable boundary vectors to grow. This results in the observed decline in the maximum utility achieved by deterministic rules. Conversely, the characterization in Theorem 1 enables the selection of pairwise probability vectors in the interior of $\mathcal{P}(\Delta)$ through randomization. Consequently, the expansion of the achievable set with increasing m does not negatively impact the maximum behavioral utility achieved when using the proposed optimal randomized rules.

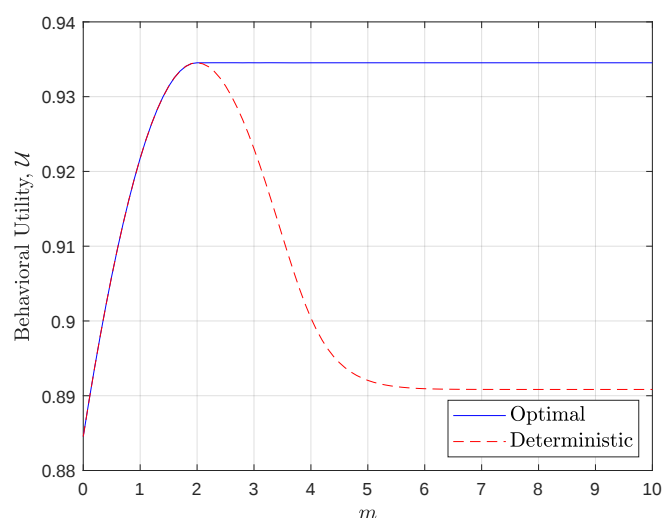


Figure 10. Maximum behavioral utility as a function of the parameter $m \in [0, 10]$, comparing the performance achieved by the optimal (potentially randomized) rules against purely deterministic decision rules.

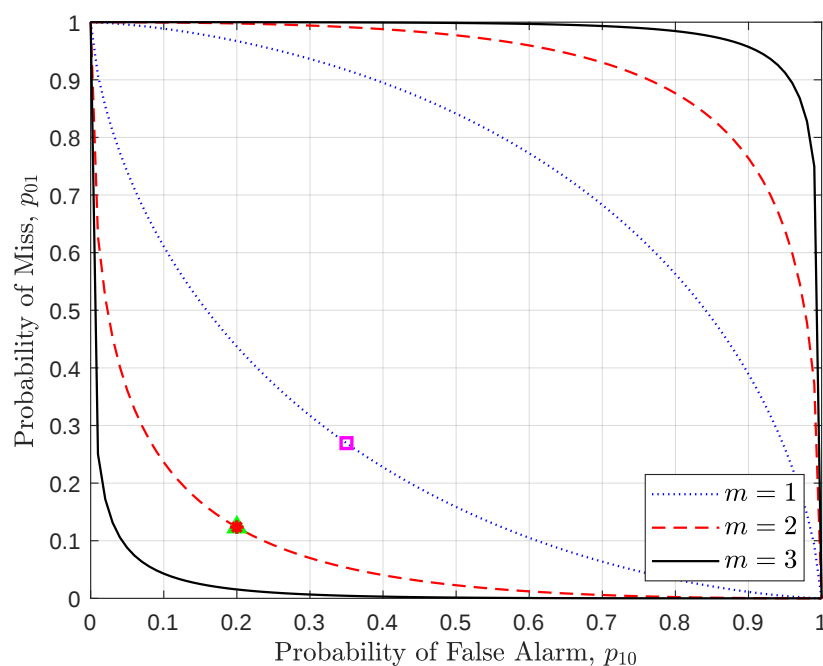


Figure 11. The feasible set of all achievable pairwise probability vectors $\mathcal{P}(\Delta)$, highlighting the optimal vectors attained through randomized rules for specific values of $m \in \{1, 2, 3\}$.

6. Conclusions

In this paper, we present a comprehensive review of the general framework to characterize the optimal decision rules for M -ary hypothesis testing, where the performance metric is defined as a function of the pairwise (error) probabilities. This framework was originally developed in our previous work. The pairwise probabilities represent the probability of selecting a specific hypothesis given that a different hypothesis is true. These pairwise error probabilities are collected into a single vector, referred to as the pairwise probability vector. Then, to identify the optimal decision rules, an indirect approach is adopted, which focuses on the compact and convex set of all achievable pairwise probability vectors, $\mathcal{P}(\Delta)$, instead of optimizing a given performance metric over the infinite-dimensional set of all possible randomized decision rules. By leveraging the topological properties of $\mathcal{P}(\Delta)$, we have presented a theorem that characterizes the structure of the optimal decision rules based on the specific location of the optimal pairwise probability vector. It is established that any achievable pairwise probability vector can be realized through a randomization of at most two possibly randomized LRQs, each constructed with a specific set of parameters.

For M -ary hypothesis testing, an LRQ partitions the likelihood ratio space into M decision regions, which are defined as convex polytopes. The distinction between deterministic and randomized LRQs arises when an observation falls on a hyperplane that separates polytopes corresponding to different hypotheses; in such cases, a deterministic LRQ selects one hypothesis uniquely, whereas a randomized LRQ employs randomization. Furthermore, we have demonstrated that any randomized LRQ can be expressed as a randomization of at most $M(M - 1)$ deterministic LRQs that utilize the same parameter sets and differ only in their tie-breaking rules on the separating hyperplanes.

Several special cases have been highlighted in this work. First, if an optimal pairwise probability vector resides on the boundary of the feasible set, a single LRQ (either deterministic or randomized) is sufficient to achieve that performance. Second, if the likelihood ratio vector is continuously distributed under all hypotheses, any optimal pairwise probability vector in the interior of the feasible set can be achieved via a randomization of two deterministic LRQs corresponding to two distinct parameter sets. Finally, if the optimal pairwise probability vector is an extreme point of the feasible set, a single deterministic LRQ constitutes the optimal decision rule.

Moreover, we have evaluated a special case in which DA's performance is measured using a PT-based behavioral utility function. This metric incorporates perceived probabilities and utilities, which are derived by applying non-linear probability weighting and utility valuation functions to the objective probability and cost values, respectively. Due to the inherent non-linearity of the probability weighting function, direct analytical approaches often fail to reveal the optimal decision rules. Furthermore, it is established that LRTs, which are optimal for rational performance metrics such as Bayes risk and the NP criterion, may be suboptimal in a behavioral context. In these instances, the optimal pairwise probability vector typically resides in the interior of the feasible set. As the pairwise probability vectors of standard LRTs are restricted to the boundary of the feasible set, they yield suboptimal results for behavioral agents. However, this study establishes that the randomized decision rules characterized by the presented theorem can successfully achieve these interior pairwise probability vectors, thereby optimizing performance in non-linear behavioral frameworks. We have corroborated the theoretical results with numerical examples that consider both discrete and continuous observation distributions. In all evaluated scenarios, the performance of the DA has been measured using the behavioral utility function. These results demonstrate that the proposed method for characterizing optimal decision rules yields performance metrics that are either identical or superior to those achieved by standard LRTs.

In summary, by consolidating these diverse scenarios into a single theory that leverages the geometry of the set of achievable pairwise probability vectors, this work bridges the gap between classical rational models, such as Bayes risk and NP criterion, and complex behavioral frameworks. The flexibility of this representation ensures that, whether an agent is operating under Bayesian risk or PT-based distortions, the optimal strategy can be identified and implemented through a consistent, finite-dimensional optimization. A possible avenue for future research is to focus on extending this unified representation to sequential hypothesis testing, where the geometry of the feasible set evolves dynamically over time.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, E.E.; validation, all authors; formal analysis, B.D., S.G. and E.E.; writing—original draft preparation, E.E.; writing—review and editing, B.D., S.G. and Y.S.; visualization, E.E.; supervision, B.D. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: The work of Emre Efendi and Berkan Dulek was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 122E493. Yanglei Song acknowledges the support by NSERC Grant RGPIN-2020-04256.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: Emre Efendi and Berkan Dulek thank TUBITAK for its support. Yanglei Song thanks NSERC for its support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Poor, H.V. *An Introduction to Signal Detection and Estimation*; Springer: New York, NY, USA, 1994.
2. Trees, H.L.V.; Bell, K.L. *Detection, Estimation, and Modulation Theory: Part I*, 2nd ed.; John Wiley & Sons: New Jersey, NJ, USA, 2013.
3. Varshney, P.K. *Distributed Detection and Data Fusion*; Springer: New York, NY, USA, 1997.
4. Bland, A.R.; Schaefer, A. Different Varieties of Uncertainty in Human Decision-Making. *Front. Neurosci.* **2012**, *6*, 85. [[CrossRef](#)]
5. Santos, L.R.; Rosati, A.G. The Evolutionary Roots of Human Decision Making. *Annu. Rev. Psychol.* **2015**, *66*, 321–347. [[CrossRef](#)]
6. Bossaerts, P.; Murawski, C. Computational Complexity and Human Decision-Making. *Trends Cogn. Sci.* **2017**, *21*, 917–929. [[CrossRef](#)] [[PubMed](#)]
7. Raykar, V.C.; Yu, S.; Zhao, L.H.; Valadez, G.H.; Florin, C.; Bogoni, L.; Moy, L. Learning from crowds. *J. Mach. Learn. Res.* **2010**, *11*, 1297–1322.
8. Chen, L.; de Gemmis, M.; Felfernig, A.; Lops, P.; Ricci, F.; Semeraro, G. Human Decision Making and Recommender Systems. *ACM Trans. Interact. Intell. Syst.* **2013**, *3*, 2. [[CrossRef](#)]
9. Belloni, A.; Berger, A.; Besson, V.; Boissier, O.; Bonnet, G.; Bourgne, G.; Chardel, P.A.; Cotton, J.P.; Evreux, N.; Ganascia, J.G.; et al. Towards a framework to deal with ethical conflicts in autonomous agents and multi-agent systems. In Proceedings of the CEPE 2014 Well-Being, Flourishing, and ICTs, Paris, France, 23–25 June 2014.
10. Yu, J.; Cheung, M.H.; Huang, J. Spectrum Investment Under Uncertainty: A Behavioral Economics Perspective. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 2667–2677. [[CrossRef](#)]
11. Schurr, N.; Marecki, J.; Tambe, M. Improving adjustable autonomy strategies for time-critical domains. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*; International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS): Budapest, Hungary, 2009; pp. 353–360.
12. Sanjab, A.; Saad, W.; Başar, T. A Game of Drones: Cyber-Physical Security of Time-Critical UAV Applications with Cumulative Prospect Theory Perceptions and Valuations. *IEEE Trans. Commun.* **2020**, *68*, 6990–7006. [[CrossRef](#)]
13. Xiao, L.; Mandayam, N.B.; Poor, H.V. Prospect Theoretic Analysis of Energy Exchange Among Microgrids. *IEEE Trans. Smart Grid* **2015**, *6*, 63–72. [[CrossRef](#)]
14. Xiao, L.; Xu, D.; Xie, C.; Mandayam, N.B.; Poor, H.V. Cloud Storage Defense Against Advanced Persistent Threats: A Prospect Theoretic Study. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 534–544. [[CrossRef](#)]
15. Saad, W.; Sanjab, A.; Wang, Y.; Kamhoua, C.A.; Kwiat, K.A. Hardware Trojan Detection Game: A Prospect-Theoretic Approach. *IEEE Trans. Veh. Technol.* **2017**, *66*, 7697–7710. [[CrossRef](#)]
16. Gezici, S.; Varshney, P.K. On the Optimality of Likelihood Ratio Test for Prospect Theory-Based Binary Hypothesis Testing. *IEEE Signal Process. Lett.* **2018**, *25*, 1845–1849. [[CrossRef](#)]

17. El Rahi, G.; Etesami, S.R.; Saad, W.; Mandayam, N.B.; Poor, H.V. Managing Price Uncertainty in Prosumer-Centric Energy Trading: A Prospect-Theoretic Stackelberg Game Approach. *IEEE Trans. Smart Grid* **2019**, *10*, 702–713. [[CrossRef](#)]
18. Geng, B.; Brahma, S.; Wimalajeewa, T.; Varshney, P.K.; Rangaswamy, M. Prospect Theoretic Utility Based Human Decision Making in Multi-Agent Systems. *IEEE Trans. Signal Process.* **2020**, *68*, 1091–1104. [[CrossRef](#)]
19. Dulek, B.; Ozturk, C.; Gezici, S. Optimal Decision Rules for Simple Hypothesis Testing Under General Criterion Involving Error Probabilities. *IEEE Signal Process. Lett.* **2020**, *27*, 261–265. [[CrossRef](#)]
20. Dulek, B.; Efendi, E.; Varshney, P.K. Behavioral Utility-Based Distributed Detection with Conditionally Independent Observations. *IEEE Trans. Signal Process.* **2024**, *72*, 3717–3730. [[CrossRef](#)]
21. Efendi, E.; Dulek, B. Optimal Detection for a Prospect Theoretic Variant of the Neyman-Pearson Problem. *IEEE Signal Process. Lett.* **2025**, *32*, 2289–2293. [[CrossRef](#)]
22. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making: Part I*; World Scientific: London, UK, 2013; pp. 99–127.
23. Tversky, A.; Kahneman, D. Advances in prospect theory: Cumulative representation of Uncertainty. *J. Risk Uncertain.* **1992**, *5*, 297–323. [[CrossRef](#)]
24. Wakker, P.; Tversky, A. An axiomatization of cumulative prospect theory. *J. Risk Uncertain.* **1993**, *7*, 147–175. [[CrossRef](#)]
25. Rieger, M.O.; Wang, M. Cumulative prospect theory and the St. Petersburg paradox. *Econ. Theory* **2006**, *28*, 665–679. [[CrossRef](#)]
26. Fox, C.R.; Poldrack, R.A. Prospect Theory and the Brain. In *Neuroeconomics*; Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A., Eds.; Elsevier: London, UK, 2009; pp. 145–173.
27. Tsitsiklis, J. Extremal properties of likelihood-ratio quantizers. *IEEE Trans. Commun.* **1993**, *41*, 550–558. [[CrossRef](#)]
28. Boyd, S.P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: New York, NY, USA, 2004.
29. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1997; Volume 11.
30. Prelec, D. The Probability Weighting Function. *Econometrica* **1998**, *66*, 497–527. [[CrossRef](#)]
31. Gonzalez, R.; Wu, G. On the shape of the probability weighting function. *Cogn. Psychol.* **1999**, *38*, 129–166. [[CrossRef](#)] [[PubMed](#)]
32. Ingersoll, J. Non-Monotonicity of the Tversky-Kahneman Probability-Weighting Function: A Cautionary Note. *Eur. Financ. Manag.* **2008**, *14*, 385–390. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.