Optimal Grouping for Group Minimax Hypothesis Testing

Kush R. Varshney, *Member, IEEE*, and Lav R. Varshney, *Member, IEEE*

Abstract—Bayesian hypothesis testing and minimax hypothesis testing represent extreme instances of detection in which the prior probabilities of the hypotheses are either completely and precisely known, or are completely unknown. Group minimax, also known as Γ-minimax, is a robust intermediary between Bayesian and minimax hypothesis testing that allows for coarse or partial advance knowledge of the hypothesis priors by using information on sets in which the prior lies. Existing work on group minimax, however, does not consider the question of how to define the sets or groups of priors; it is assumed that the groups are given. In this paper, we propose a novel intermediate detection scheme formulated through the quantization of the space of prior probabilities that optimally determines groups and also representative priors within the groups. We show that when viewed from a quantization perspective, group minimax amounts to determining centroids with a minimax Bayes risk error divergence distortion criterion: the appropriate Bregman divergence for this task. In addition, the optimal partitioning of the space of prior probabilities is a Bregman Voronoi diagram. Together, the optimal grouping and representation points are an ϵ-net with respect to Bayes risk error divergence, and permit a rate-distortion type asymptotic analysis of detection performance with the number of groups. Examples of detecting signals corrupted by additive white Gaussian noise and of distinguishing exponentially-distributed signals are presented.

Index Terms—Bayesian hypothesis testing, Bregman divergence, detection theory, minimax hypothesis testing, quantization, Stolarsky mean.

I. INTRODUCTION

BAYESIAN hypothesis testing and minimax hypothesis testing are signal detection formulations for when the prior probabilities of the hypotheses are perfectly and precisely known and for when the prior probabilities of the hypotheses are completely unknown, respectively [1]. Optimal performance in both settings is achieved by likelihood ratio tests with appropriately chosen thresholds. Between these two edge cases, there is an entire set of likelihood ratio tests corresponding to a coarse knowledge of the prior probabilities; these intermediate formulations are explored in this work.

Formulations that lie between Bayesian and minimax hypothesis testing are known as group minimax or Γ-minimax and are of interest because it is difficult to obtain complete information about priors in many decision-making scenarios, but information about priors is also not completely lacking [2]–[9]. Group minimax detection formulations take partial information about priors as input and provide robustness against that partial information, in contrast to minimax hypothesis testing which provides robustness against complete lack of information.

Throughout the long history of group minimax statistical inference, the sets, groups, or Γs in which the true priors lie are treated as inputs and are not optimized for detection or estimation performance. In contrast, our work herein investigates the joint problem of optimizing the groupings within which to find a minimax-optimal representative prior as well as finding those priors for all groupings. With such an output, a system designer can focus efforts on determining the (optimized) group to which a decision problem belongs in place of fully estimating the prior probability as in the Bayesian formulation.

We view the minimax test as one in which knowledge of prior probabilities has been quantized to a single cell encompassing the entire probability simplex, and the Bayesian test as one in which knowledge of prior probabilities has been quantized to an infinite number of cells that finely partition the probability simplex. In the group minimax test, the prior probabilities are quantized to a finite number of cells. The appropriate quantization distortion measure for prior probabilities of hypotheses is Bayes risk error [10], which is a Bregman divergence [11]. However unlike standard quantization, we are interested in minimizing the maximum distortion rather than minimizing the average distortion [12]–[14]. Thus we pursue minimax Bayes risk error quantization of prior probabilities [15].

Group minimax, which provides a means to consider intervals of prior belief rather than exact prior belief, is similar in spirit but differs in details to decision making based on interval-valued probability described in [16]. There are also connections to representative prior distributions [17], the robust Bayesian viewpoint [18], [19], and other areas of decision making in which robustness is desired [20].

To the best of our knowledge, there has been no previous work on the quantization of prior probabilities for hypothesis testing besides our own [10], [21]. Many interesting findings on average distortion clustering with Bregman divergences as the distortion criteria are reported in [22] and [23], but we believe this is the first use of Bregman divergences in

studying group minimax hypothesis testing. (Some new work published after the initial submission of this paper on Bregman geometry relates to hypothesis testing, but for the space of observations, not prior probabilities [24].) Although studies and results in quantization theory typically focus on average distortion, maximum distortion does also appear occasionally, see [14], [25]–[28]. Such a minimax partitioning of a space is known as an ϵ-net or ϵ-covering [29].

In investigating quantization for group minimax hypothesis testing, we derive centroid and nearest neighbor conditions for minimax Bayes risk error distortion and discuss how alternating application of these conditions leads to a locally optimal quantizer. We provide direct derivations in the binary detection case and specialize elegant results from the Bregman divergence literature in the general case. Minimax centroid conditions for Bregman divergences are derived in [30]. The problem of finding the optimal nearest neighbor cell boundaries for a given set of samples, also known as a Voronoi diagram, is addressed for Bregman divergences in [31] and [32]. Advantages of the direct derivations for the binary setting include direct geometric insights, as well as closed-form expressions.

As a further contribution similar in style to rate–distortion theory, we present asymptotic results on detection error as the partialness of information about the prior goes from the minimax hypothesis testing case to the Bayesian hypothesis testing case. Mathematically, this specializes existing results on ϵ-nets for Bregman divergences. Since all previous work on group minimax detection was only concerned with defining a single group as some subset of the simplex rather than considering groups partitioning the entire simplex, this analysis is novel. It shows for example that minimax distortion scales as an inverse square law with the number of groups for binary hypothesis testing, and also provides insight into rate loss when an incorrect criterion is used for performing grouping.

We also present a few examples of group minimax detection with different likelihood models.

The rest of the paper is organized in the following manner. First in Section II, we set forth notation and briefly provide background on Bayesian, minimax and group minimax detection, along with Bayes risk error divergence. We formulate a quantization problem to find optimal groupings for group minimax detection in Section III. Section IV derives the nearest neighbor and centroid optimality conditions for the proposed quantization problem in both the binary and M-ary cases. We analyze the rate–distortion behavior of the groups in Section V. Two examples are presented in Section VI to provide intuition. Section VII provides a summary of the contributions and concludes.

II. PRELIMINARIES

The detection or hypothesis testing problem is the task of accurately determining which of M classes a noisy signal instance belongs to. In the binary ($M = 2$) case, this task is often determining the presence or absence of a target based on a measurement observed through noise. In this section we first discuss binary hypothesis testing and then we consider M-ary

hypothesis testing for $M > 2$. We next present the definition of the Bayes risk error divergence, a quantification of detection performance degradation or regret. Finally, we discuss the objectives of minimax and group minimax detection.

A. Binary Decisions

Consider the binary hypothesis testing problem. There are two hypotheses h_0 and h_1 with prior probabilities $p_0 = \Pr[H = h_0]$ and $p_1 = \Pr[H = h_1] = 1 - p_0$, and a noisy observation Y governed by likelihood functions $f_{Y|H}(y|H = h_0)$ and $f_{Y|H}(y|H = h_1)$. A decision rule $\hat{h}(y)$ that uniquely maps every possible y to either h_0 or h_1 is to be determined. There are two types of error probabilities:

$$p_E^{\mathrm{I}} = \Pr[\hat{h}(Y) = h_1|H = h_0], \text{ and}$$
$$p_E^{\mathrm{II}} = \Pr[\hat{h}(Y) = h_0|H = h_1].$$

Minimizing weighted error, the optimization criterion for the decision rule $\hat{h}(y)$ is the Bayes risk:

$$J = c_{10}p_0 p_E^{\mathrm{I}} + c_{01}(1 - p_0)p_E^{\mathrm{II}}, \qquad (1)$$

where c_{10} is the cost of the first type of error and c_{01} is the cost of the second type of error. The decision rule that optimizes (1) is the following likelihood ratio test [1]:

$$\frac{f_{Y|H}(y|H = h_1)}{f_{Y|H}(y|H = h_0)} \overset{\hat{h}(y)=h_1}{\underset{\hat{h}(y)=h_0}{\gtrless}} \frac{p_0 c_{10}}{(1 - p_0)c_{01}}. \qquad (2)$$

The prior probability p_0 appears on the right side of the rule in the threshold. Since the prior probability is part of the specification of the Bayes-optimal decision rule, the error probabilities p_E^{I} and p_E^{II} are functions of the prior probability. Thus we may write the Bayes risk as a function of p_0:

$$J(p_0) = c_{10}p_0 p_E^{\mathrm{I}}(p_0) + c_{01}(1 - p_0)p_E^{\mathrm{II}}(p_0), \qquad (3)$$

The function $J(p_0)$ is zero at the points $p_0 = 0$ and $p_0 = 1$ and is positive-valued, strictly concave, and continuous in the interval $(0, 1)$ [33]. Under deterministic decision rules, $J(p_0)$ is differentiable everywhere.

The Bayesian hypothesis testing threshold on the right side of (2) relies on the true prior probability p_0, but as discussed in Section I, this value may not be known precisely. When the true prior probability is p_0, but the threshold in $\hat{h}(y)$ uses some other *decision weight* a, there is mismatch. The Bayes risk of the decision rule with threshold

$$\frac{ac_{10}}{(1 - a)c_{01}}$$

is:

$$J(p_0, a) = c_{10}p_0 p_E^{\mathrm{I}}(a) + c_{01}(1 - p_0)p_E^{\mathrm{II}}(a). \qquad (4)$$

The function $J(p_0, a)$ is a linear function of p_0 with slope $(c_{10}p_E^{\mathrm{I}}(a) - c_{01}p_E^{\mathrm{II}}(a))$ and intercept $c_{01}p_E^{\mathrm{II}}(a)$. The function $J(p_0, a)$ is tangent to $J(p_0)$ at a and $J(p_0, p_0) = J(p_0)$. By the point-slope formula of lines, the mismatched Bayes risk is also:

$$J(p_0, a) = J(a) + (p_0 - a)J'(a) \qquad (5)$$

when J is differentiable. An example of how $J(p_0)$ and $J(p_0, a)$ are related is shown in Fig. 1.
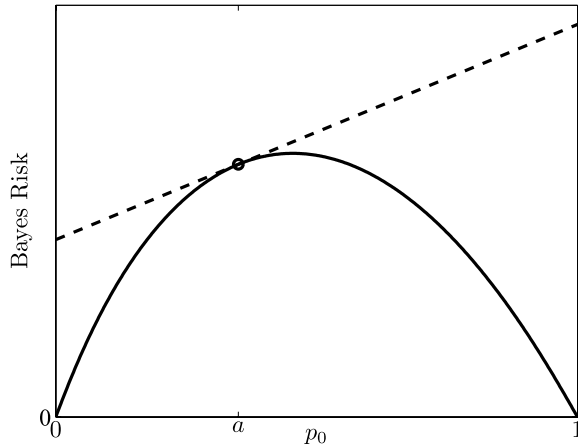
Fig. 1. Example $J(p_0)$ (solid curve) and $J(p_0, a)$ (dashed line).

B. M-Ary Decisions

The basics from the binary case carry over to the M-ary case. With M hypotheses, there are M prior probabilities $p_i > 0, i = 0, \ldots, M-1$ such that $\sum_i p_i = 1$. The collection of priors is denoted by the vector $\mathbf{p}$, which is an element of the M-ary probability simplex. There is also an $M \times M$ matrix of costs c_{ij}. The detection rule $\hat{h}(y)$ in the M-ary case uses priors and costs in an analogous manner to the likelihood ratio test (2). Specifically, the detection rule is $\hat{h}(y) = h_{\hat{i}}$ where

$$\hat{i} = \arg \min_{i \in \{0, \ldots, M-1\}} c_{i0} p_0 + \sum_{j=0}^{M-1} c_{ij} p_j \frac{f_{Y|H}(y|H = h_j)}{f_{Y|H}(y|H = h_0)}. \quad (6)$$

The Bayes risk function is now

$$J(\mathbf{p}) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} c_{ij} p_j \Pr[\hat{h}(Y, \mathbf{p}) = h_i | H = h_j]. \quad (7)$$

With a vector-valued decision weight $\mathbf{a}$, the mismatched Bayes risk function is

$$J(\mathbf{p}, \mathbf{a}) = J(\mathbf{a}) + (\mathbf{p} - \mathbf{a})^T \nabla J(\mathbf{a}) \quad (8)$$

when J is differentiable.

C. Bayes Risk Error Divergence

The Bayes risk $J(\mathbf{p})$ represents the performance of the best possible decision making under uncertainty, whereas the mismatched Bayes risk $J(\mathbf{p}, \mathbf{a})$ represents the degraded decision-making performance due to the decision weight $\mathbf{a}$. Thus, we may quantify the degradation or distortion in detection performance using the difference:

$$d(\mathbf{p} \| \mathbf{a}) = J(\mathbf{p}, \mathbf{a}) - J(\mathbf{p}) \quad (9)$$
$$= -J(\mathbf{p}) + J(\mathbf{a}) + (\mathbf{p} - \mathbf{a})^T \nabla J(\mathbf{a}). \quad (10)$$

This difference is a Bregman divergence termed *Bayes risk error divergence* generated by the convex function $-J(\mathbf{p})$ over a convex domain (the M-ary probability simplex) [10], [11].

The identification of Bayes risk error as a Bregman divergence is important because it allows us to understand its properties ranging from linearity to linear separation to generalized

Pythogorean theorem [23]. Most importantly for this work, the Bregman geometry allows us to understand properties related to smallest enclosing balls [30] and Voronoi diagrams [31], [32].

D. Minimax and Group Minimax Decision Weights

The binary minimax hypothesis testing threshold is determined by finding the decision weight a that minimizes the worst-case $J(p_0, a)$, that is:

$$a_{\text{minimax}}^* = \arg \min_a \max_{p_0} J(p_0, a). \quad (11)$$

Under equivalent notation, the optimal Bayesian decision weight is $a_{\text{Bayesian}}^* = p_0$. In Bayesian hypothesis testing, the decision weight a continually changes with p_0, whereas in minimax hypothesis testing, there is a single decision weight a for all p_0. In the M-ary case, as in the binary case,

$$\mathbf{a}_{\text{minimax}}^* = \arg \min_{\mathbf{a}} \max_{\mathbf{p}} J(\mathbf{p}, \mathbf{a}) \quad (12)$$

and $\mathbf{a}_{\text{Bayesian}}^* = \mathbf{p}$.

The group minimax framework starts with a set Γ within which it is known that the true prior $\mathbf{p}$ is contained. This set Γ may come from various knowledge sources. There are two related objectives for the decision weight in the literature [8] and [9]:

$$\mathbf{a}_{\text{group minimax}}^* = \arg \min_{\mathbf{a}} \max_{\mathbf{p} \in \Gamma} J(\mathbf{p}, \mathbf{a}) \quad (13)$$

and

$$\mathbf{a}_{\text{group minimax}}^* = \arg \min_{\mathbf{a}} \max_{\mathbf{p} \in \Gamma} d(\mathbf{p} \| \mathbf{a}). \quad (14)$$

The extrema for both versions are minimax and Bayesian hypothesis testing when Γ is the entire probability simplex and a single point respectively. We proceed with (14).

III. MINIMAX BAYES RISK ERROR QUANTIZATION

Having described a divergence that quantifies loss in detection performance due to a mismatched decision weight, in this section we describe how that divergence can be utilized within a scalar or vector quantization framework to yield not only the optimal minimax representation point for a given set of priors (the typical group minimax problem), but also the optimal groupings for the group minimax scenario.

A. Quantization for Group Minimax Grouping

The space of all possible decision weights and the space of all true prior probability vectors is the M-ary probability simplex. As discussed in Section II, in the Bayesian case, the decision weight $\mathbf{a}$ changes continuously with the true prior probability vector $\mathbf{p}$ of the detection problem, so that the Bayes risk error divergence $d(\mathbf{p} \| \mathbf{a}) = 0$ for all detection problems. Denoting the mapping from true prior probability $\mathbf{p}$ to decision weight $\mathbf{a}$ as $q(\mathbf{p}) = \mathbf{a}$, this function is the identity function $q(\mathbf{p}) = \mathbf{p}$ in the Bayesian case and has the entire M-ary probability simplex as its range.

On the other hand in the minimax case, there is a single decision weight $\mathbf{a}_{\text{minimax}}^*$ for all detection problems and

$d(\mathbf{p}\|\mathbf{a}^*_{\mathrm{minimax}}) > 0$ for all problems except the one problem in which, by chance, the minimax decision weight is the true prior probability. Here, the mapping from true prior probability to decision weight is a function whose range contains a single point: $q(\mathbf{p}) = \mathbf{a}^*_{\mathrm{minimax}}$.

As discussed in Section I, it may be that the true prior probability and thus the Bayesian decision weight is not exactly known. It may also be, however, that there is some partial information, and thus we need not restrict ourselves to just one decision weight for all detection problems. Therefore, we would like to consider functions $q(\cdot)$ whose range is a finite set of K decision weights $\{\mathbf{a}_1, \ldots, \mathbf{a}_K\}$. With such a range, there are K true prior probabilities for which there is no degradation in detection performance, i.e. $d(\mathbf{p}\|q(\mathbf{p})) = 0$. The function $q(\cdot)$ depends discontinuously on $\mathbf{p}$ such that $q(\mathbf{p}) = \mathbf{a}_k$ for all $\mathbf{p} \in \mathcal{Q}_k$, $k = 1, \ldots, K$. Such a function is a quantizer. The remaining question for the proposed optimal grouping for group minimax hypothesis testing is determining the decision weights $\mathbf{a}_k$ and the quantization cells $\mathcal{Q}_k$.

B. Minimax Bayes Risk Error Quantization Criterion

Robustness is the motivation for both minimax hypothesis testing and group minimax hypothesis testing. We take maximum Bayes risk error divergence as the objective for finding the decision weights and the quantization cells, resulting in the following minimax quantizer design problem:

$$q^*_K = \arg\min_{q_K} \max_{\mathbf{p}} d(\mathbf{p}\|q_K(\mathbf{p})), \qquad (15)$$

where $q_K(\cdot)$ is a quantizer function with $K \geq 1$ cells and decision weights, and K is a fixed parameter. Operationally, knowing in advance that the true prior probability $\mathbf{p}$ falls in cell $\mathcal{Q}_k$ indicates that the decision weight $\mathbf{a}_k$ be used in setting the threshold.

In the $K = 1$ case, it is straightforward to show that the decision weight of $q^*_1(\cdot)$ equals the minimax hypothesis testing value $\mathbf{a}^*_{\mathrm{minimax}}$, and occurs at the peak of $J(\mathbf{p})$. However for $K > 1$, the decision weight $\mathbf{a}_k$ within a cell $\mathcal{Q}_k$ is not the point that minimizes the maximum mismatched Bayes risk $J(\mathbf{p}, \mathbf{a})$; rather it is the point that minimizes the maximum Bayes risk error divergence $d(\mathbf{p}\|\mathbf{a})$. An example of the decision weight as a function of prior probability is shown in Fig. 2.

In this section, we have defined an approach for optimal grouping for group minimax hypothesis testing. This formulation reduces to the two extreme hypothesis testing methodologies: minimax at $K = 1$ and Bayesian as $K \to \infty$.

IV. OPTIMALITY CONDITIONS

This section develops necessary conditions for optimality of a quantizer for the probability simplex under the minimax criterion (15) defined in Section III, first in the scalar quantization (binary hypothesis testing) setting and then in the vector quantization (M-ary hypothesis testing) setting. We find a centroid condition to locally optimize decision weights when the quantization cells are fixed. Then we find a nearest neighbor condition to locally optimize the quantization cells with decision weights fixed. Optimal quantizers can be found by alternately applying the nearest neighbor
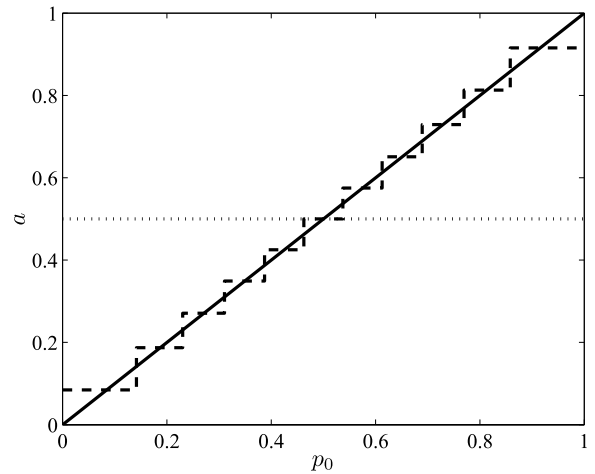

Fig. 2. Example decision weight a as a function of prior probability p_0 for Bayesian (solid), group minimax with eleven groups (dashed), and minimax (dotted) binary hypothesis testing.

and centroid conditions through a version of the iterative Lloyd–Max algorithm [12], [14], [31]. We provide direct derivations for the binary case and specialize more general Bregman divergence results for the M-ary case.

Note that the alternating minimization of the Lloyd–Max algorithm has the same local optimality properties for the minimax setting as for the setting minimizing expected distortion [25]. For Bayes risk error, we previously showed that the convexity properties that yield global optimality from the Lloyd–Max algorithm [34] are typically satisfied [10].

A. Binary Hypothesis Testing Centroid Condition

The K-cell scalar quantizer function $q_K(p_0)$ in the binary hypothesis testing problem has cell notation as follows. The probability simplex $[0, 1]$ is partitioned into K intervals $\mathcal{Q}_1 = [0, b_1]$, $\mathcal{Q}_2 = (b_1, b_2]$, $\mathcal{Q}_3 = (b_2, b_3]$,..., $\mathcal{Q}_K = (b_{K-1}, 1]$. Since $d(p_0\|a)$ increases monotonically with the absolute error, we can observe the convexity of the nearest neighbor cell; consequently each cell must consist of a single interval, see [12, Lemma 6.2.1]. Within a fixed scalar quantization cell $\mathcal{Q}_k$ with boundaries b_{k-1} and b_k, we want an expression for the optimal decision weight:

$$a_k = \arg\min_{a \in \mathcal{Q}_k} \max_{p_0 \in \mathcal{Q}_k} d(p_0\|a). \qquad (16)$$

Theorem 1: In the binary hypothesis testing problem with deterministic likelihood ratio test decision rules, the minimax Bayes risk error divergence optimal decision weight a_k satisfies:

$$J'(a_k) = \frac{J(b_k) - J(b_{k-1})}{b_k - b_{k-1}}. \qquad (17)$$

Proof: Let us first focus on the inner maximization in (16). In the binary hypothesis testing case,

$$d(p_0\|a) = -J(p_0) + J(a) + (p_0 - a)J'(a), \qquad (18)$$

from which we see that the second derivative of $d(p_0\|a)$ with respect to p_0 is $-J''(p_0)$, which is greater than zero due to the

strict concavity of $J(p_0)$. Thus, $d(p_0\|a)$ has no local maxima in the interior of $\mathcal{Q}_k$; the maximum occurs at an endpoint: b_k or b_{k-1}. Consequently,

$$
\max_{p_0 \in \mathcal{Q}_k} d(p_0\|a)
$$
$$
= \max\{d(b_k\|a), d(b_{k-1}\|a)\}
$$
$$
= \frac{d(b_{k-1}\|a) + d(b_k\|a) + \left|d(b_{k-1}\|a) - d(b_k\|a)\right|}{2}. \quad (19)
$$

Substituting (18) into (19) and simplifying, we find that (19) equals

$$
\frac{(b_{k-1} + b_k - 2a)J'(a) - J(b_{k-1}) - J(b_k) + 2J(a)}{2}
$$
$$
+ \frac{\left|(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k)\right|}{2}, \quad (20)
$$

which is to be minimized with respect to $a \in \mathcal{Q}_k$.

Due to the absolute value function, there are two cases to consider:

1) $(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k) \le 0$ and
2) $(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k) > 0$.

Due to the concavity of the Bayes risk function, $J'(a)$ is monotonically decreasing. Therefore, since $(b_{k-1} - b_k)$ is negative, $(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k)$ is a monotonically increasing function of a. Consequently the two cases of the absolute value correspond to the intervals $(b_{k-1}, a^\dagger]$ for case 1 and $(a^\dagger, b_k]$ for case 2, where $a^\dagger$ satisfies:

$$
(b_{k-1} - b_k)J'(a^\dagger) - J(b_{k-1}) + J(b_k) = 0. \quad (21)
$$

In the first case, (20) simplifies to:

$$
(b_k - a)J'(a) + J(a) - J(b_k)
$$

with derivative with respect to a:

$$
(b_k - a)J''(a),
$$

which is less than zero because $(b_k - a) > 0$ and $J''(a) < 0$ due to Bayes risk concavity. Thus the minimization objective is monotonically decreasing in the first case.

In the second case, (20) simplifies to:

$$
(b_{k-1} - a)J'(a) + J(a) - J(b_{k-1}),
$$

which has derivative with respect to a:

$$
(b_{k-1} - a)J''(a),
$$

which is greater than zero because $(b_{k-1} - a) < 0$ and $J''(a) < 0$. In the second case, the minimization objective is monotonically increasing.

Since (20) is decreasing over $(b_{k-1}, a^\dagger]$ and increasing over $(a^\dagger, b_k]$, it is minimized at $a^\dagger$. Therefore $a_k = a^\dagger$. The decision weight satisfies (21). This is equivalently the slope matching condition (17) given in the statement of the theorem. ∎

We can notice that this minimax centroid is a Stolarsky mean [35]; the Stolarsky mean of u and v is in general:

$$
F'^{-1}\left(\frac{F(u) - F(v)}{u - v}\right)
$$

for any reasonable function $F(\cdot)$.

B. Binary Hypothesis Testing Nearest Neighbor Condition

In the binary hypothesis testing nearest neighbor condition, we are to find the cell boundary b_k given the decision weights a_k and a_{k+1}.

Theorem 2: In the binary hypothesis testing problem with deterministic likelihood ratio test decision rules, the minimax Bayes risk error divergence optimal cell boundary b_k is:

$$
b_k = \frac{a_{k+1}J'(a_{k+1}) - a_k J'(a_k) - (J(a_{k+1}) - J(a_k))}{J'(a_{k+1}) - J'(a_k)}. \quad (22)
$$

Proof: As discussed in Section IV-A, the maximum Bayes risk error divergence within a cell occurs at the cell boundary. Therefore, we would like to minimize the Bayes risk error divergence at the cell boundary.

Specifically, b_k should be chosen to minimize the maximum of $d(b_k\|a_k)$ and $d(b_k\|a_{k+1})$. At a given potential boundary point b, the $J(b)$ term is the same in both $d(b\|a_k)$ and $d(b\|a_{k+1})$, so only $J(b, a_k)$ and $J(b, a_{k+1})$ need be considered. Due to the geometry of the problem, b_k should be the abscissa of the point at which the lines $J(p_0, a_k)$ and $J(p_0, a_{k+1})$ intersect. Working with the definitions of $J(p_0, a_k)$ and $J(p_0, a_{k+1})$, we find the point of intersection to be (22). ∎

We can notice that the cell boundary is the tangent line mean of the decision weights [36]. The nearest neighbor condition for minimax Bayes risk error quantization is the same as that for minimum mean Bayes risk error quantization [10].

C. M-Ary Hypothesis Testing Nearest Neighbor Condition

We found the nearest neighbor condition over the binary simplex, i.e., the line segment between zero and one, in Section IV-B. In that case, the cell boundaries were simply two points on the line. The situation is slightly more complicated notationally in the M-ary detection task because of the increased dimensionality. Let us define the M-ary probability simplex as follows:

$$
\mathcal{P}_M = \left\{\boldsymbol{\pi} \in \mathbb{R}_+^{M-1} \;\middle|\; \sum_{i=1}^{M-1} \pi_i \le 1\right\}. \quad (23)
$$

Now in specifying the nearest neighbor condition, we assume that the decision weights $\{\mathbf{a}_1, \ldots, \mathbf{a}_K\}$ are fixed. We denote the set of points in $\mathcal{P}_M$ that are equidistant according to Bayes risk error divergence from $\mathbf{a}_k$ and $\mathbf{a}_{k+1}$ as $\mathcal{B}_{k,k+1}$, such that

$$
\mathcal{B}_{k,k+1} = \{\boldsymbol{\pi} \in \mathcal{P}_M | d(\boldsymbol{\pi}\|a_k) = d(\boldsymbol{\pi}\|a_{k+1})\}. \quad (24)
$$

We show that this bisector $\mathcal{B}_{k,k+1}$ between the two decision weights $\mathbf{a}_k$ and $\mathbf{a}_{k+1}$ is a hyperplane in $\mathcal{P}_M$.

Theorem 3: In the M-ary hypothesis testing problem with deterministic likelihood ratio test decision rules, the Bayes risk error divergence bisector $\mathcal{B}_{k,k+1}$ satisfies the hyperplane equation:

$$
\mathcal{B}_{k,k+1} = \Big\{\boldsymbol{\pi} \in \mathcal{P}_M \mid \boldsymbol{\pi}^T \left(\nabla J(\mathbf{a}_{k+1}) - \nabla J(\mathbf{a}_k)\right) =
$$
$$
\mathbf{a}_{k+1}^T \nabla J(\mathbf{a}_{k+1}) - \mathbf{a}_k^T \nabla J(\mathbf{a}_k) - (J(\mathbf{a}_{k+1}) - J(\mathbf{a}_k))\Big\}.
$$
$$
(25)
$$

Proof: The result follows by specializing [31, Lemma 4], which applies to all Bregman divergences, to Bayes risk error divergence. ∎

It is easy to see that we recover the binary boundary expression for b_k (22) when we set $M = 2$ in (25).

In the binary case, the boundary point bisectors fully specify the quantization cells $\mathcal{Q}_k$, but in the M-ary case, we must go one step further. In particular, the quantization cells are defined as follows:

$$\mathcal{Q}_k = \left\{ \boldsymbol{\pi} \in \mathcal{P}_M \mid d(\boldsymbol{\pi} \| \mathbf{a}_k) \leq d(\boldsymbol{\pi} \| \mathbf{a}_{k'}), \, k' \neq k \right\}. \quad (26)$$

Moreover, as discussed in [31], the cell is a convex polyhedron, which is delineated by the intersection of the bisectors between its decision weight and all other cell decision weights. The set of all minimax quantization cells is the Voronoi diagram of the simplex with the set of fixed decision weights as seeds. If we write the half space induced by $\mathcal{B}_{k,k'}$ such that it contains $\mathbf{a}_k$ and is restricted to $\mathcal{P}_M$ as $\mathcal{H}_{k,k'}$, then

$$\mathcal{Q}_k = \bigcap_{k' \neq k} \mathcal{H}_{k,k'}. \quad (27)$$

Let v_k be the number of vertices of cell $\mathcal{Q}_k$. Each $\mathcal{Q}_k$ has at most $(K-1) + (M-1)$ faces and at most $\binom{K-1}{M-2} + M$ vertices, i.e. $v_k \leq \binom{K-1}{M-2} + M$. (The constant additive terms that correspond to the dimension of the space are due to intersections with the simplex boundary.) Moreover, in the same way as in the binary case that the maximum Bayes risk error divergence within a cell occurs at the cell boundary, the maximum Bayes risk error divergence in the M-ary case occurs at one of the finite v_k vertices [31, Lemma 12].

D. M-Ary Hypothesis Testing Centroid Condition

In Section IV-A, we found the minimax Bayes risk error centroid condition in the binary detection case through a fairly explicit calculation that made use of convexity properties of the Bayes risk function. Here we find the centroid condition in general for M-ary detection, by adapting the minimax centroid results of general Bregman divergences found in [30].

In deriving the centroid condition, the cell $\mathcal{Q}_k$ and its v_k vertices are fixed. Since we know that the maximum divergence occurs at a vertex, we must only examine the vertices of $\mathcal{Q}_k$ in order to find the minimax-optimal decision weight within the cell. Let the vertices of $\mathcal{Q}_k$ be denoted $\{\mathbf{b}_{k,1}, \ldots, \mathbf{b}_{k,v_k}\}$. The optimal decision weight is a functional mean of the vertices.

Theorem 4: In the M-ary hypothesis testing problem with deterministic likelihood ratio test decision rules, the minimax Bayes risk error divergence optimal decision weight $\mathbf{a}_k$ satisfies:

$$\nabla J(\mathbf{a}_k) = \sum_{i=1}^{v_k} w_i \nabla J(\mathbf{b}_{k,i}), \quad (28)$$

where the weights satisfy $w_i \geq 0$ and $\sum_i^{v_k} w_i = 1$.

Putting all of the w_i into a vector $\mathbf{w}$, the optimal weight vector is the solution to the following optimization problem:

$$\max_{\mathbf{w}} \sum_{i=1}^{v_k} w_i d \left(\nabla^{-1} J \left(\sum_{j=1}^{v_k} w_j \nabla J(\mathbf{b}_{k,j}) \right) \Big\| \mathbf{b}_{k,i} \right) \quad (29)$$

subject to the same constraints $w_i \geq 0$ and $\sum_i^{v_k} w_i = 1$.

Proof: The result follows by specializing [30, Sec. 3], which applies to all Bregman divergences, to Bayes risk error divergence. ∎

The optimization problem (29) is similar to that solved in learning support vector machines [30]. The $w_i > 0$ that are found are 'support' vertices that contribute to the location of the decision weight [30].

We note that the centroid condition in the binary case $J'(a_k) = (J(b_k) - J(b_{k-1}))/(b_k - b_{k-1})$ (17) can be expressed as $J'(a_k) = w_1 J'(b_{k-1}) + w_2 J'(b_k)$, with $w_1, w_2 \geq 0$ and $w_1 + w_2 = 1$ due to the concavity of the Bayes risk and the intermediate value theorem of calculus. In this form, we see the correspondence to (28). In contrast to the M-ary case, there is a closed form expression for the decision weight a_k in the binary case without requiring an optimization program to be run. We also note that finding group minimax decision rules is considered to be computationally difficult in general [8], [9]; the optimization program described in this section, however, has favorable computational properties.

V. RATE–DISTORTION ANALYSIS

To understand how quickly or slowly group minimax hypothesis testing approaches the performance of Bayesian hypothesis testing, in this section we examine the maximum achieved distortion as a function of the number of groups K. Asymptotic scaling with number of groups has previously been unexamined in group minimax detection since prior work was only concerned with defining a single group as some subset of the probability simplex. This analysis provides insight into how partialness of information regarding prior probabilities affects detection performance.

Let us denote the minimax distortion overall as:

$$D = \min_{q_K} \max_{\mathbf{p}} d(\mathbf{p} \| q_K(\mathbf{p})). \quad (30)$$

Theorem 5: In the M-ary hypothesis testing problem with deterministic likelihood ratio test decision rules, the maximum Bayes risk error D of the minimax-optimal quantization with K groups satisfies the rate–distortion expression:

$$K = O\left(\frac{1}{(M-1)! D^{\frac{M-1}{2}}} \right) \quad (31)$$

Proof: The result follows from the fact that the volume of the probability simplex in the M-ary detection problem is $\frac{1}{(M-1)!}$ and specializing the results on ϵ-nets for general Bregman divergences given in [31, Lemma 14] to Bayes risk error divergence. ∎

The convergence from the edge case of minimax hypothesis testing to the other edge case of Bayesian hypothesis testing is in proportion to K^{-2} in the binary hypothesis testing case, which is the same scaling seen in the mean Bayes risk error

case presented in [10]. A similar scaling is also noted for detectors based on estimated prior probabilities [37]. The minimax error scaling can be viewed as the asymptotic behavior of the minimum covering radius with respect to Bayes risk error divergence. Note that all Bregman divergences, including squared error, will yield the same scaling behavior for K as a function of D. This implies that grouping by an incorrect Bregman fidelity criterion will incur a constant asymptotic rate loss.

VI. EXAMPLES

We present two signal detection problem examples approached through group minimax hypothesis testing with optimal grouping. The first example is the typical example of detecting a signal through Gaussian noise. The second example is a ternary hypothesis testing problem with three different exponential likelihoods.

A. Detecting Signals in Gaussian Noise

Let us consider the following signal and measurement model:

$$Y = s_m + W, \quad m \in \{0, 1\}, \tag{32}$$

where $s_0 = 0$ and $s_1 = \mu$, and W is a zero-mean, Gaussian random variable with variance σ^2. The parameters μ and σ^2 are known, deterministic quantities. The error probabilities for this signal model are:

$$p_E^{\mathrm{I}}(a) = Q\left(\frac{\mu}{2\sigma} + \frac{\sigma}{\mu} \ln\left(\frac{c_{10}a}{c_{01}(1-a)}\right)\right), \text{ and}$$

$$p_E^{\mathrm{II}}(a) = Q\left(\frac{\mu}{2\sigma} - \frac{\sigma}{\mu} \ln\left(\frac{c_{10}a}{c_{01}(1-a)}\right)\right),$$

where

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-x^2/2} dx.$$

These error probabilities can be put together to obtain the Bayes risk error expression for this detection task:

$$J(p_0, a) = c_{10}p_0 Q\left(\frac{\mu}{2\sigma} + \frac{\sigma}{\mu} \ln\left(\frac{c_{10}a}{c_{01}(1-a)}\right)\right)$$
$$+ c_{01}(1-p_0)Q\left(\frac{\mu}{2\sigma} - \frac{\sigma}{\mu} \ln\left(\frac{c_{10}a}{c_{01}(1-a)}\right)\right). \tag{33}$$

We use the Lloyd–Max algorithm to design quantizers for the proposed group minimax criterion using the centroid and nearest neighbor conditions derived in Section IV: equations (17) and (22).

We show such quantizers for $K = 4$ and different ratios of the Bayes costs c_{10} and c_{01} along with different ratios of μ and σ^2. Fig. 3(a) shows quantizers for equal Bayes costs and equal mean and standard deviation. In the plot, the black curve is $J(p_0)$ and the dashed line is $J(p_0, q_4(p_0))$, with their difference being $d(p_0\|q_4(p_0))$. The circle markers are the representation points that are found. The divergence value $d(p_0\|q_4(p_0))$ is shown in Fig. 3(b). The quantizers and divergences are shown for two other parameter settings—one with a greater level of observation noise ($\mu/\sigma^2 = 1/2$) and
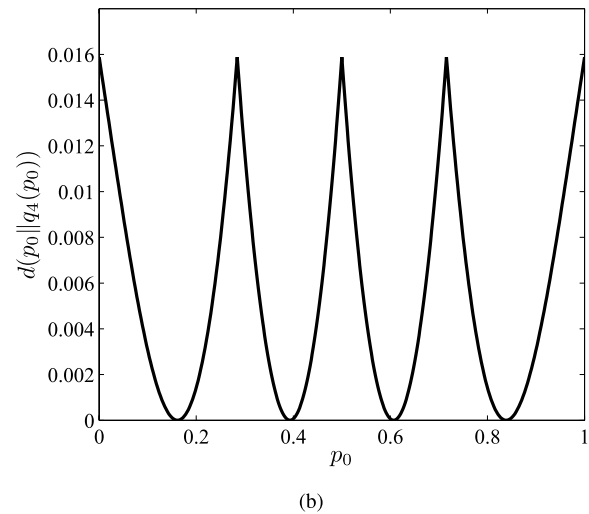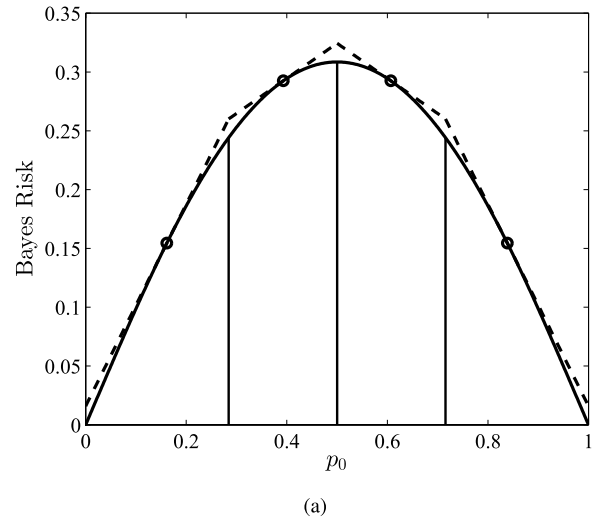

(a)

(b)

Fig. 3. Minimax Bayes risk error (a) quantizers and (b) divergence with $K = 4$ for $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 1$, $c_{01} = 1$. In (a), the black curve is $J(p_0)$, the dashed line is $J(p_0, q_4(p_0))$, the circle markers are the decision weights, and the vertical lines indicate the interval boundaries of the groups.

one with unequal error costs ($c_{10}/c_{01} = 10$)—in Fig. 4 and Fig. 5.

The first thing to notice among the quantizers is that the group intervals are finer closer to the peak of the Bayes risk function $J(p_0)$ than at its ends and that decision weights within the intervals are also closer to the peak. The regime around the peak of the Bayes risk corresponds to the most challenging detection problems; the results show that it is most important to have fine grouping the more difficult the hypothesis test is. Also, it is quite apparent in the divergence plots that the error exhibits an equiripple character, which is to be expected due to the minimax objective.

As we make the detection problem noisier and thus more difficult, $J(p_0)$ gets peakier and the granularity near the center becomes finer. In fact, single decision weights are able to satisfy large ranges of prior probabilities at the edges of the simplex. Unequal error costs skew $J(p_0)$ to be asymmetric; again, the costliest decision regimes require the finest groups. The lesson for system designers is that in a group minimax setting, the most effort should be expended in trying to
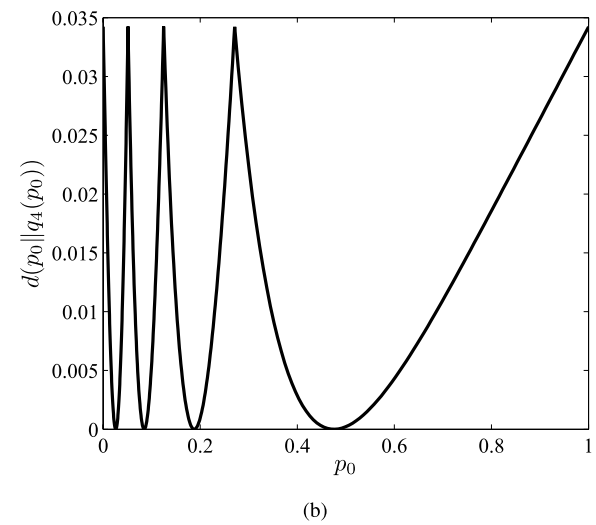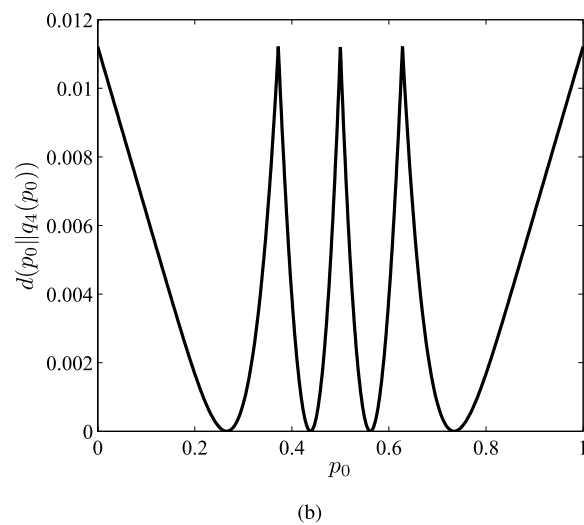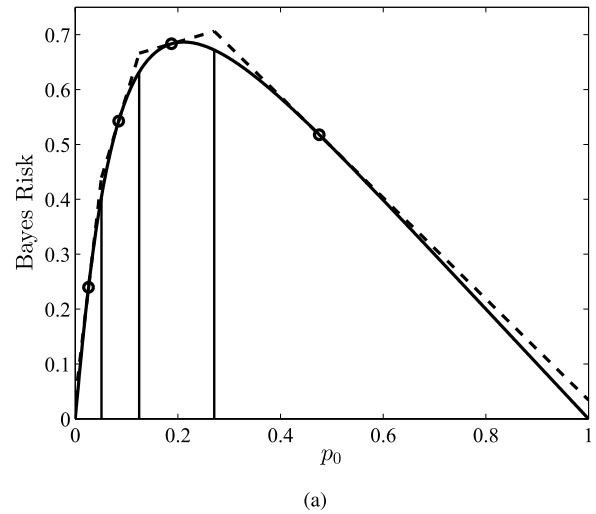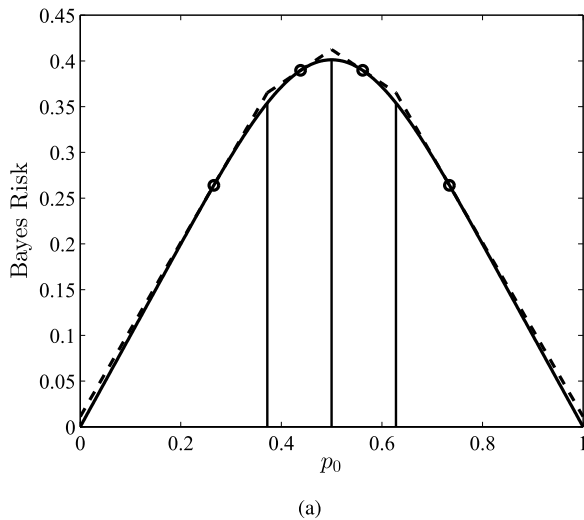
Fig. 4. Minimax Bayes risk error (a) quantizers and (b) divergence with $K = 4$ for $\mu = 1$, $\sigma^2 = 2$, $c_{10} = 1$, $c_{01} = 1$. In (a), the black curve is $J(p_0)$, the dashed line is $J(p_0, q_4(p_0))$, the circle markers are the decision weights, and the vertical lines indicate the interval boundaries of the groups.

Fig. 5. Minimax Bayes risk error (a) quantizers and (b) divergence with $K = 4$ for $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 10$, $c_{01} = 1$. In (a), the black curve is $J(p_0, q_4(p_0))$, the circle markers are the decision weights, and the vertical lines indicate the interval boundaries of the groups.

determine which small range near the largest values of $J(p_0)$ that the true prior lies.

The minimax Bayes risk error as a function of K for this example is shown in Fig. 6 on both linear and logarithmic scales. The curves seen in Fig. 6(b) exactly reflect the behavior expected according to the rate–distortion analysis of Section V. They are almost perfectly linear beyond a couple of small K values. The slopes of the lines are -2 which is the rate predicted for $M = 2$. The degradation between the Bayesian optimal detection performance and the group minimax detection performance drops quite rapidly with K. The improvement going from $K = 1$ (minimax) to $K = 2$ is quite stark.

B. Distinguishing Exponential Likelihoods

In this example, we consider objects that are served at varying rates. Objects are served at rate $\lambda_m > 0$ when in state $H = h_m$ for $m \in \{0, 1, 2\}$, with $\lambda_0 > \lambda_1 > \lambda_2$. The ternary

hypothesis testing task is to determine which state the object is in based on an observation of the time $Y = y$ at which it is served. The likelihood functions take the form:

$$f_{Y|H}(y|H = h_m) = \lambda_m e^{-\lambda_m y}. \tag{34}$$

For simplicity, we only consider the case in which $c_{00} = c_{11} = c_{22} = 0$, and $c_{01} = c_{02} = c_{10} = c_{12} = c_{20} = c_{21} = 1$.

Recall that we denote the prior probabilities p_0 and p_1 through the vector $\mathbf{p}$ and the decision weights a_0 and a_1 through the vector $\mathbf{a}$ (where $p_2 = 1 - p_0 - p_1$ and $a_2 = 1 - a_0 - a_1$). In this example, if we define the following two functions of the decision weights:

$$\gamma_{01}(\mathbf{a}) = \max\left\{0, \frac{1}{\lambda_0 - \lambda_1} \ln\left(\frac{a_0 \lambda_0}{a_1 \lambda_1}\right)\right\}, \tag{35}$$

$$\gamma_{12}(\mathbf{a}) = \max\left\{0, \frac{1}{\lambda_1 - \lambda_2} \ln\left(\frac{a_1 \lambda_1}{(1 - a_0 - a_1)\lambda_2}\right)\right\}, \tag{36}$$
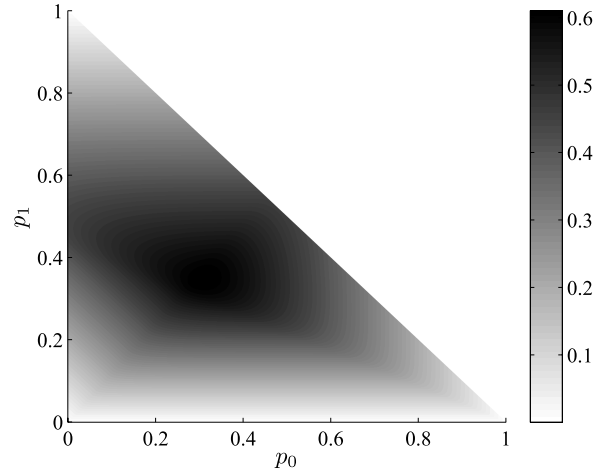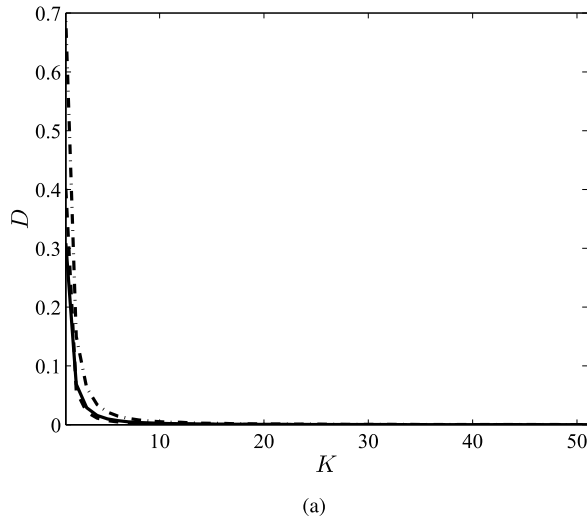
(a)

Fig. 7.   Bayes risk function $J(\mathbf{p})$ for $\lambda_0 = 5$, $\lambda_1 = 4$, $\lambda_2 = 3$.

(b)

Fig. 6.   Minimax Bayes risk error for $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 1$, $c_{01} = 1$ (solid line); $\mu = 1$, $\sigma^2 = 2$, $c_{10} = 1$, $c_{01} = 1$ (dashed line); and $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 10$, $c_{01} = 1$ (dashed and dotted line), on (a) linear and (b) logarithmic scales.

then the mismatched Bayes risk function is:

$$J(\mathbf{p}, \mathbf{a}) = p_0 e^{-\lambda_0 \gamma_{01}(\mathbf{a})} + p_1 \left(1 - e^{-\lambda_1 \gamma_{01}(\mathbf{a})} + e^{-\lambda_1 \gamma_{12}(\mathbf{a})} \right)$$
$$+ (1 - p_0 - p_1) \left(1 - e^{-\lambda_2 \gamma_{12}(\mathbf{a})} \right). \tag{37}$$

We calculate the gradient $\nabla J(\mathbf{a})$ in closed form, but do not write it down here because of its unwieldy nature.

We now examine optimal groupings for group minimax hypothesis testing in the ternary exponential service time example with $\lambda_0 = 5$, $\lambda_1 = 4$, and $\lambda_2 = 3$. The convex Bayes risk function defined over the probability simplex is shown via shading in Fig. 7. The Bayes risk function is zero at all three corners and along the p_0 axis. We apply the alternating nearest neighbor condition and centroid condition of the Lloyd–Max algorithm to find the optimal groupings in the $K = 7$ case. Fig. 8 is a plot of the groups and representation points that are found. The importance of fine group granularity at the peak of the Bayes risk function is even more apparent in this $M = 3$
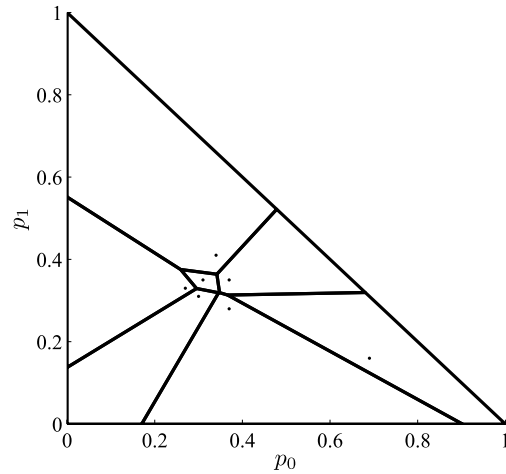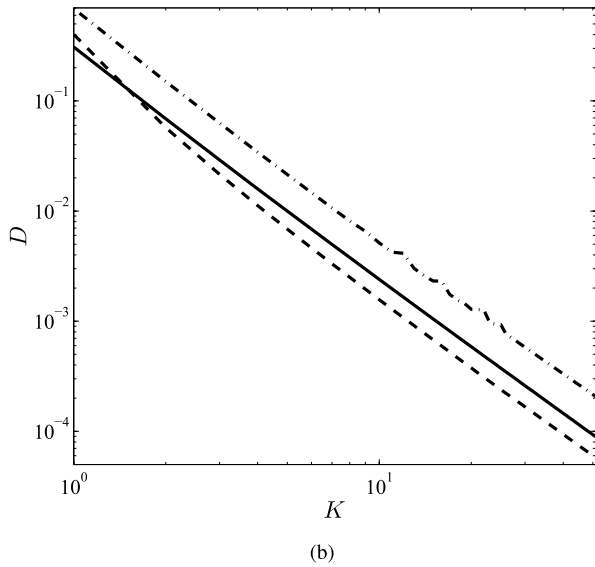

Fig. 8.   Optimal groupings and representation points found for $\lambda_0 = 5$, $\lambda_1 = 4$, $\lambda_2 = 3$ with $K = 7$.
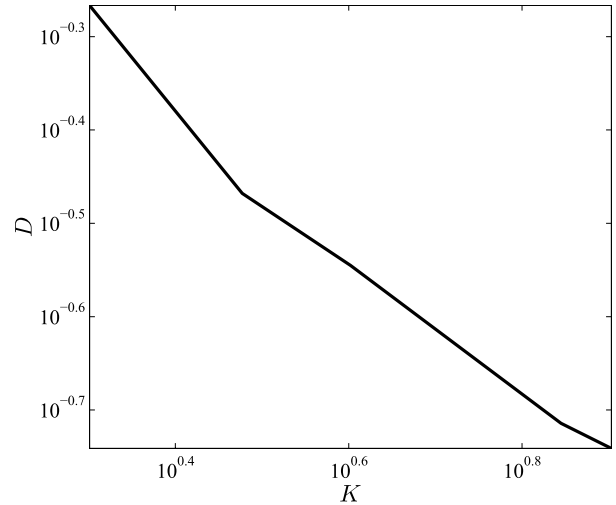

Fig. 9.   Minimax Bayes risk error for $\lambda_0 = 5$, $\lambda_1 = 4$, $\lambda_2 = 3$ on a logarithmic scale.

case than in the Gaussian $M = 2$ case. In Fig. 9, we show the minimax error for this example as a function of K in the logarithm-transformed domain. As expected for $M = 3$, the function is approximately linear with slope -1.

VII. CONCLUSION

The group minimax test as an intermediate formulation between the Bayesian and minimax tests that takes advantage of set-structured, incomplete advanced knowledge of priors was proposed long ago by the early decision theorists. However, results in the literature were obtained under special circumstances and when the sets were predetermined. In this work, we approach group minimax through the emergent theory of quantizing with Bregman divergences and make statements about optimal representative priors that do not rely on any special likelihood functions. By optimizing the minimax Bayes risk error divergence, we obtain a closed-form Stolarsky mean expression for the optimal representative prior within a group in the binary case. In the M-ary case, we present a support vector machine-like program to be solved.

In descriptions of group minimax or Γ-minimax in the literature, no heed is given to determining the best K groups to have in order to maximize detection performance. We solve this problem jointly with finding representative priors within groups through an alternating minimization involving Bregman centroids and Bregman bisectors. The optimal groupings are delineated by a Voronoi diagram or ϵ-net of the space of prior probabilities. We give closed-form expressions for the polyhedral group boundaries. Moreover, in a rate–distortion format, we characterize the rate at which detection performance of group minimax approaches Bayesian detection as the number of optimal groups increases. The outputs of the proposed approach can guide users of signal detection systems as they estimate the prior probabilities of hypotheses, in particular focusing on estimating the cells to which problems of interest belong.

The research described in this paper is for single decision makers. Distributed detection with *multiple* agents working as a team [38]–[40] or with conflicts [41], [42] can also be considered. Additionally, regret theory is closely connected with minimax hypothesis testing [43], [44]; extensions of this paper within the confines of regret theory may be explored. Optimal grouping for *noise-enhanced* group minimax detection is another direction for future work [45].

ACKNOWLEDGMENT

REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 1968.

[2] H. Robbins, "Asymptotically subminimax solutions of compound statistical decision problems," in *Proc. 2nd Berkeley Symp. Math. Statist. Probab.*, Berkeley, CA, USA, Jul./Aug. 1950, pp. 131–148.

[3] I. J. Good, "Rational decisions," *J. Roy. Statist. Soc. B*, vol. 14, no. 1, pp. 107–114, 1952.

[4] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Statist.*, vol. 23, no. 3, pp. 396–407, Sep. 1952.

[5] L. J. Savage, *The Foundations of Statistics*. New York, NY, USA: Wiley, 1954.

[6] H. Robbins, "The empirical Bayes approach to statistical decision problems," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 1–20, Mar. 1964.

[7] J. R. Blum and J. Rosenblatt, "On partial a priori information in statistical inference," *Ann. Math. Statist.*, vol. 38, no. 6, pp. 1671–1678, Dec. 1967.

[8] B. Vidakovic, "Γ-minimax: A paradigm for conservative robust Bayesians," in *Robust Bayesian Analysis*, D. Ríos Insua and F. Ruggeri, Eds. New York, NY, USA: Springer-Verlag, 2000, pp. 241–259.

[9] F. Ruggeri, "Gamma-minimax inference," in *Encyclopedia of Statistical Sciences*, S. Kotz, C. B. Read, N. Balakrishnan, and B. Vidakovic, Eds. Hoboken, NJ, USA: Wiley, 2006.

[10] K. R. Varshney and L. R. Varshney, "Quantization of prior probabilities for hypothesis testing," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4553–4562, Oct. 2008.

[11] K. R. Varshney, "Bayes risk error is a Bregman divergence," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4470–4472, Sep. 2011.

[12] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA, USA: Kluwer, 1992.

[13] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[14] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Berlin, Germany: Springer-Verlag, 2000.

[15] K. R. Varshney and L. R. Varshney, "Multilevel minimax hypothesis testing," in *Proc. IEEE Int. Workshop Statist. Signal Process.*, Nice, France, Jun. 2011, pp. 109–112.

[16] M. Wolfenson and T. L. Fine, "Bayes-like decision making with upper and lower probabilities," *J. Amer. Statist. Assoc.*, vol. 77, no. 377, pp. 80–88, Mar. 1982.

[17] C. Hildreth, "Bayesian statisticians and remote clients," *Econometrica*, vol. 31, no. 3, pp. 422–438, Jul. 1963.

[18] J. O. Berger, "The robust Bayesian viewpoint," Dept. Statist., Purdue Univ., West Lafayette, IN, USA, Tech. Rep. 82-9, Apr. 1982.

[19] L. R. Pericchi and P. Walley, "Robust Bayesian credible intervals and prior ignorance," *Int. Statist. Rev.*, vol. 59, no. 1, pp. 1–23, Apr. 1991.

[20] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Rev.*, vol. 53, no. 3, pp. 464–501, Jul. 2011.

[21] L. R. Varshney, J. B. Rhim, K. R. Varshney, and V. K. Goyal, "Categorical decision making by people, committees, and crowds," in *Proc. Inf. Theory Appl. Workshop*, La Jolla, CA, USA, Feb. 2011.

[22] A. Banerjee, X. Guo, and H. Wang, "On the optimality of conditional expectation as a Bregman predictor," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2664–2669, Jul. 2005.

[23] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, Oct. 2005.

[24] F. Nielsen, "Hypothesis testing, information divergence and computational geometry," in *Proc. Geometric Sci. Inf.*, Paris, France, Aug. 2013, pp. 241–248.

[25] C. Zhu and Y. Hua, "Image vector quantization with minimax L_∞ distortion," *IEEE Signal Process. Lett.*, vol. 6, no. 2, pp. 25–27, Feb. 1999.

[26] N. Sarshar and X. Wu, "Minimax multiresolution scalar quantization," in *Proc. Data Compression Conf.*, Snowbird, UT, USA, Mar. 2004, pp. 52–61.

[27] P. Venkitasubramaniam, L. Tong, and A. Swami, "Minimax quantization for distributed maximum likelihood estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toulouse, France, May 2006, pp. III-652–III-655.

[28] Y. A. Reznik, "An algorithm for quantization of discrete probability distributions," in *Proc. Data Compression Conf.*, Snowbird, UT, USA, Mar. 2011, pp. 333–342.

[29] A. N. Kolmogorov and V. M. Tihomirov, "ϵ-entropy and ϵ-capacity of sets in functional spaces," *Amer. Math. Soc. Translations*, vol. 17, pp. 277–364, 1961.

[30] R. Nock and F. Nielsen, "Fitting the smallest enclosing Bregman balls," in *Proc. Eur. Conf. Mach. Learn.*, Porto, Portugal, Oct. 2005, pp. 649–656.

[31] F. Nielsen, J.-D. Boissonnat, and R. Nock, "Bregman Voronoi diagrams: Properties, algorithms and applications," INRIA Sophia Antipolis Méditerranée, Valbonne, France, Tech. Rep. 6154, Mar. 2007.

[32] J.-D. Boissonnat, F. Nielsen, and R. Nock, "Bregman Voronoi diagrams," *Discrete Comput. Geometry*, vol. 44, no. 2, pp. 281–307, Sep. 2010.

[33] R. A. Wijsman, "Continuity of the Bayes risk," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 1083–1085, Jun. 1970.

[34] A. V. Trushkin, "Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 187–198, Mar. 1982.

[35] K. B. Stolarsky, "Generalizations of the logarithmic mean," *Math. Mag.*, vol. 48, no. 2, pp. 87–92, Mar. 1975.

[36] B. C. Dietel and R. A. Gordon, "Using tangent lines to define means," *Math. Mag.*, vol. 76, no. 1, pp. 52–61, Feb. 2003.

[37] J. Jiao, L. Zhang, and R. D. Nowak, "Minimax-optimal bounds for detectors based on estimated prior probabilities," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 6101–6109, Sep. 2012.

[38] J. B. Rhim, L. R. Varshney, and V. K. Goyal, "Quantization of prior probabilities for collaborative distributed hypothesis testing," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4537–4550, Sep. 2012.

[39] J. B. Rhim, L. R. Varshney, and V. K. Goyal, "Distributed decision making by categorically-thinking agents," in *Decision Making and Imperfection*, T. V. Guy, M. Kárný, and D. H. Wolpert, Eds. Berlin, Germany: Springer-Verlag, 2013.

[40] G. Gül and A. M. Zoubir. (Jun. 2013). *Robust Distributed Detection*. [Online]. Available: http://arxiv.org/pdf/1306.3618

[41] P. D. Grünwald and A. P. Dawid, "Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory," *Ann. Statist.*, vol. 32, no. 4, pp. 1367–1433, Aug. 2004.

[42] J. B. Rhim, L. R. Varshney, and V. K. Goyal, "Conflict in distributed hypothesis testing with quantized prior probabilities," in *Proc. Data Compression Conf.*, Snowbird, UT, USA, Mar. 2011, pp. 313–322.

[43] G. Loomes and R. Sugden, "Regret theory: An alternative theory of rational choice under uncertainty," *Econ. J.*, vol. 92, no. 368, pp. 805–824, Dec. 1982.

[44] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.

[45] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3972–3989, Aug. 2010.

Kush R. Varshney (S'00–M'10) was born in Syracuse, NY, in 1982. He received the B.S. degree (magna cum laude) in electrical and computer engineering with honors from Cornell University, Ithaca, NY, in 2004 and the S.M. and Ph.D. degrees, both in electrical engineering and computer science, from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006 and 2010, respectively.

He is a research staff member in the Mathematical Sciences and Analytics Department at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. While at MIT, he was a research assistant with the Stochastic Systems Group in the Laboratory for Information and Decision Systems and a National Science Foundation Graduate Research Fellow. His research interests include statistical signal processing, statistical learning, data mining, and image processing.

Dr. Varshney is a member of Eta Kappa Nu, Tau Beta Pi, and ACM. He received a Best Student Paper Travel Award at the 2009 International Conference on Information Fusion, the Best Paper Award at the 2013 IEEE International Conference on Service Operations and Logistics, and Informatics, the Best Social Good Paper Award at the 2014 ACM SIGKDD Conference on Knowledge Discovery and Data Mining, and several IBM awards for contributions to business analytics projects. He is on the editorial board of *Digital Signal Processing* and a member of the IEEE Signal Processing Society Machine Learning for Signal Processing Technical Committee.

Lav R. Varshney (S'00–M'10) was born in Syracuse, NY, in 1982. He received the B.S. degree (magna cum laude) in electrical and computer engineering with honors from Cornell University, Ithaca, NY, in 2004 and the S.M. and Ph.D. degrees, both in electrical engineering and computer science, from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006 and 2010, respectively.

He is an assistant professor in the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign. During 2010-2013, he was a research staff member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. His research interests include information and coding theory, statistical signal processing, and collective intelligence in sociotechnical systems.

Dr. Varshney is a member of Eta Kappa Nu, Tau Beta Pi, Sigma Xi, and the Organization for Computational Neurosciences. He received the Ernst A. Guillemin Thesis Award for Outstanding Electrical Engineering S.M. Thesis, the Jin-Au Kong Award Honorable Mention for Electrical Engineering doctoral thesis, and several IBM awards for contributions to computational creativity, crowdsourcing, and data analytics projects.