Quantization of Prior Probabilities for Hypothesis Testing

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

Abstract—In this paper, Bayesian hypothesis testing is investigated when the prior probabilities of the hypotheses, taken as a random vector, are quantized. Nearest neighbor and centroid conditions are derived using mean Bayes risk error (MBRE) as a distortion measure for quantization. A high-resolution approximation to the distortion-rate function is also obtained. Human decision making in segregated populations is studied assuming Bayesian hypothesis testing with quantized priors.

Index Terms—Bayesian hypothesis testing, Bayes risk error, categorization, classification, detection, quantization.

I. INTRODUCTION

CONSIDER a hypothesis testing scenario in which an object is to be observed to determine which one of M states, $\{h_0, \ldots, h_{M-1}\}$, it is in. The object has prior probability p_m of being in state h_m, i.e., $p_m = \Pr[H = h_m]$, and prior probability vector $\boldsymbol{p} = [p_0 \quad \ldots \quad p_{M-1}]^T$, with $\sum_{m=0}^{M-1} p_m = 1$, which is known to the decision maker. M-ary hypothesis testing with known prior probabilities calls for the Bayesian formulation to the problem, for which the optimal decision rule minimizes Bayes risk [2].

Now consider the situation when there is a population of objects, each with its own prior probability vector drawn from the distribution $f_{\boldsymbol{P}}(\boldsymbol{p})$ supported on the $(M-1)$-dimensional probability simplex. If the prior probability vector of each object were known perfectly to the decision maker before observation and hypothesis testing, then the scenario would be no different than that of standard Bayesian hypothesis testing. However, we consider the case in which the decision maker is constrained and can only work with at most K different prior probability vectors. Such a constraint is motivated by scenarios where the decision maker has finite memory or limited information processing resources. Hence, when there are more than K objects in the population, the decision maker must first map the true prior probability vector of the object being observed to one of the K available vectors and then proceed to perform the optimal Bayesian hypothesis test, treating that vector as the prior probabilities of the object.

Although not the only such constrained scenario, one example is that of human decision making. One particular setting is a referee deciding whether a player has committed a foul using his or her noisy observation as well as prior experience. Players commit fouls at different rates; some players are dirtier or more aggressive than others. It is this rate that is the prior probability for the "foul committed" state. Hence, over the population of players, there is a distribution of prior probabilities. If the referee tunes the prior probability to the particular player on whose action the decision is to be made, decision-making performance is improved.

Human decision makers, however, are limited in their information processing capacity and can only carry around seven, plus or minus two, categories without getting confused [3]. Consequently, the referee is limited and categorizes players into a small number of dirtiness levels, with associated representative prior probabilities, exactly the scenario described above.

In this paper, the design of the mapping from prior probability vectors in the population to one of K representative probability vectors is approached as a quantization problem. Mean Bayes risk error (MBRE) is defined as a fidelity criterion for the quantization of $f_{\boldsymbol{P}}(\boldsymbol{p})$ and conditions are derived for a minimum MBRE quantizer. Some examples of MBRE-optimal quantizers are given along with their performance in the low-rate quantization regime. Distortion-rate functions are given for the high-rate quantization regime. Certain human decision-making tasks, as mentioned above, may be modeled by quantized prior hypothesis testing due to certain suboptimalities in human information processing. Human decision making is analyzed in detail for segregated populations, revealing a mathematical model of social discrimination.

Previous work that combines detection and quantization looks at the quantization of observed data, not prior probabilities, and also only approximates the Bayes risk function instead of working with it directly, e.g., [4]–[6] and references cited in [6]. In such work, there is a communication constraint between the sensor and the decision maker, but the decision maker has unconstrained processing capability. Our work deals with the opposite case, where there is no communication constraint between the sensor and the decision maker, however the decision maker is constrained.

A brief look at imperfect priors appears in [7, Sec. 2.E], but optimal quantization is not considered. In [8] and [9], it is shown that small deviations from the true prior yield small deviations in the Bayes risk. We are not aware of any previous work that

has looked at quantization, clustering, or categorization of prior probabilities.

In the remainder of this paper, we focus on binary hypothesis testing, $M = 2$. Section II defines the Bayes risk error distortion and gives some of its properties. Section III discusses low-rate quantization and Section IV discusses high-rate quantization. Some examples with a Gaussian measurement model are given in Section V. Section VI considers the implications on human decision making and Section VII provides a summary and directions for future work.

II. BAYES RISK ERROR

In the binary Bayesian hypothesis testing problem for a given object, there are two hypotheses h_0 and h_1 with prior probabilities $p_0 = \Pr[H = h_0]$ and $p_1 = \Pr[H = h_1] = 1 - p_0$, a noisy observation Y, and likelihoods $f_{Y|H}(y|h_0)$ and $f_{Y|H}(y|h_1)$. Note that we consider a one-shot measurement Y, rather than a set of independent, noisy measurements. A function $\hat{h}(y)$ is designed that uniquely maps every possible y to either h_0 or h_1 in such a way that the function is optimal with respect to Bayes risk $J = E[c(H_i, H_j)]$, an expectation over the nonnegative cost function $c(h_i, h_j)$. This gives the following specification for $\hat{h}(y)$:

$$\hat{h}(\cdot) = \arg\min_{f(\cdot)} E[c(H, f(Y))] \quad (1)$$

where the expectation is over both H and Y. It may be shown that the optimal decision rule $\hat{h}(y)$ is the likelihood ratio test

$$\frac{f_{Y|H}(y|h_1)}{f_{Y|H}(y|h_0)} \underset{\hat{h}(y)=h_0}{\overset{\hat{h}(y)=h_1}{\gtrless}} \frac{p_0(c_{10} - c_{00})}{(1 - p_0)(c_{01} - c_{11})} \quad (2)$$

where $c_{ij} = c(h_i, h_j)$.

There are two types of errors, with the following probabilities:

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

Bayes risk may be expressed in terms of those error probabilities as

$$J = (c_{10} - c_{00})p_0 p_E^{\mathrm{I}} + (c_{01} - c_{11})(1 - p_0)p_E^{\mathrm{II}}$$
$$+ c_{00}p_0 + c_{11}(1 - p_0). \quad (3)$$

It is often of interest to assign no cost to correct decisions, i.e., $c_{00} = c_{11} = 0$, which we assume in the remainder of this paper. In this case, the Bayes risk simplifies to

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

In (4), the dependence of the Bayes risk and error probabilities on p_0 has been explicitly noted. The error probabilities depend on p_0 through $\hat{h}(\cdot)$, given in (2). 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 $(0, 1)$ [2], [10], [11].

In the case when the true prior probability is p_0, but $\hat{h}(y)$ is designed according to (2) using some other value a substituted for p_0, there is a mismatch, and the mismatched Bayes risk is

$$\tilde{J}(p_0, a) = c_{10}p_0 p_E^{\mathrm{I}}(a) + c_{01}(1 - p_0)p_E^{\mathrm{II}}(a). \quad (5)$$

$\tilde{J}(p_0, a)$ is a linear function of p_0 with slope $\left(c_{10}p_E^{\mathrm{I}}(a) - c_{01}p_E^{\mathrm{II}}(a)\right)$ and intercept $c_{01}p_E^{\mathrm{II}}(a)$. Note that $\tilde{J}(p_0, a)$ is tangent to $J(p_0)$ at a and that $\tilde{J}(p_0, p_0) = J(p_0)$.

Definition 1: Let Bayes risk error $d(p_0, a)$ be the difference between the mismatched Bayes risk function $\tilde{J}(p_0, a)$ and the Bayes risk function $J(p_0)$

$$d(p_0, a) = \tilde{J}(p_0, a) - J(p_0)$$
$$= c_{10}p_0 p_E^{\mathrm{I}}(a) + c_{01}(1 - p_0)p_E^{\mathrm{II}}(a)$$
$$- c_{10}p_0 p_E^{\mathrm{I}}(p_0) - c_{01}(1 - p_0)p_E^{\mathrm{II}}(p_0). \quad (6)$$

We now give properties of $d(p_0, a)$ as a function of p_0 and as a function of a.

Theorem 1: The Bayes risk error $d(p_0, a)$ is nonnegative and only equal to zero when $p_0 = a$. As a function of $p_0 \in (0, 1)$, it is continuous and strictly convex for all a.

Proof: Because $J(p_0)$ is a continuous and strictly concave function, and lines $\tilde{J}(p_0, a)$ are tangent to $J(p_0)$, $\tilde{J}(p_0, a) \geq J(p_0)$ for all p_0 and a, with equality when $p_0 = a$. Consequently, $d(p_0, a)$ is nonnegative and only equal to zero when $p_0 = a$. Moreover, $d(p_0, a)$ is continuous and strictly convex in $p_0 \in (0, 1)$ for all a because it is the difference of a continuous linear function and a continuous strictly concave function. ∎

Theorem 2: For any deterministic likelihood ratio test $\hat{h}(\cdot)$, as a function of $a \in (0, 1)$ for all p_0, the Bayes risk error $d(p_0, a)$ has exactly one stationary point, which is a minimum.

Proof: Consider the parameterized curve $(p_E^{\mathrm{I}}, p_E^{\mathrm{II}})$ traced out as a is varied; this is a flipped version of the receiver operating characteristic (ROC). The flipped ROC is a strictly convex function for deterministic likelihood ratio tests. At its endpoints, it takes values $(p_E^{\mathrm{I}} = 0, p_E^{\mathrm{II}} = 1)$ when $a = 1$ and $(p_E^{\mathrm{I}} = 1, p_E^{\mathrm{II}} = 0)$ when $a = 0$ [2], and therefore, has average slope -1. By the mean value theorem and strict convexity, there exists a unique point on the flipped ROC at which $dp_E^{\mathrm{II}}/dp_E^{\mathrm{I}} = -1$. To the left of that point, $-\infty < dp_E^{\mathrm{II}}/dp_E^{\mathrm{I}} < -1$, and to the right of that point, $-1 < dp_E^{\mathrm{II}}/dp_E^{\mathrm{I}} < 0$.

For deterministic likelihood ratio tests, $\beta dp_E^{\mathrm{I}}(a)/da < 0$ and $\gamma dp_E^{\mathrm{II}}(a)/da > 0$ for all $a \in (0, 1)$ and positive constants β and γ [2]. Therefore, if $(\gamma dp_E^{\mathrm{II}})/(\beta dp_E^{\mathrm{I}}) < -1$, i.e., $(\gamma dp_E^{\mathrm{II}}/da)(da/\beta dp_E^{\mathrm{I}}) < -1$, then $\gamma dp_E^{\mathrm{II}}/da > -\beta dp_E^{\mathrm{I}}/da$ and $\beta dp_E^{\mathrm{I}}/da + \gamma dp_E^{\mathrm{II}}/da > 0$. In the same manner, if $(\gamma dp_E^{\mathrm{II}})/(\beta dp_E^{\mathrm{I}}) > -1$, then $\beta dp_E^{\mathrm{I}}/da + \gamma dp_E^{\mathrm{II}}/da < 0$.

Combining the above, we find that the function $\beta p_E^{\mathrm{I}}(a) + \gamma p_E^{\mathrm{II}}(a)$ has exactly one stationary point in $(0, 1)$, which occurs when the slope of the flipped ROC is $-\beta/\gamma$. Denote this stationary point as a_s. For $0 < a < a_s$, $-1 < dp_E^{\mathrm{II}}/dp_E^{\mathrm{I}} < 0$ and the slope of $\beta p_E^{\mathrm{I}}(a) + \gamma p_E^{\mathrm{II}}(a)$ is negative; for $a_s < a < 1$, $-\infty < dp_E^{\mathrm{II}}/dp_E^{\mathrm{I}} < -1$ and the slope of $\beta p_E^{\mathrm{I}}(a) + \gamma p_E^{\mathrm{II}}(a)$ is positive. Therefore, a_s is a minimum. ∎

As a function of a, the Bayes risk error is of the form $\beta p_E^{\mathrm{I}}(a) + \gamma p_E^{\mathrm{II}}(a) + C$. Hence, it also has exactly one stationary point a_s, which is a minimum.

As seen in Section III, the above properties of $d(p_0, a)$ are useful to establish that the Lloyd–Max conditions are not only necessary, but also sufficient for quantizer local optimality.

The third derivative of $d(p_0, a)$ with respect to p_0 is

$$-c_{10}p_0\frac{d^3p_E^{\mathrm{I}}(p_0)}{dp_0^3} - 3c_{10}\frac{d^2p_E^{\mathrm{I}}(p_0)}{dp_0^2}$$
$$-c_{01}(1-p_0)\frac{d^3p_E^{\mathrm{II}}(p_0)}{dp_0^3} + 3c_{01}\frac{d^2p_E^{\mathrm{II}}(p_0)}{dp_0^2} \quad (7)$$

when the constituent derivatives exist. As seen in Section IV, when the third derivative exists and is continuous, $d(p_0, a)$ is locally quadratic, which is useful to develop high-rate quantization theory for Bayes risk error fidelity [12].

III. LOW-RATE QUANTIZATION

The conditions necessary for the optimality of a quantizer for $f_{P_0}(p_0)$ under Bayes risk error distortion are now derived. A K-point quantizer partitions the interval $[0, 1]$ into K regions $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \ldots, \mathcal{R}_K$. For each of these quantization regions $\mathcal{R}_k$, there is a representation point a_k to which the elements are mapped. For regular quantizers, the regions are subintervals $\mathcal{R}_1 = [0, b_1], \mathcal{R}_2 = (b_1, b_2], \mathcal{R}_3 = (b_2, b_3], \ldots, \mathcal{R}_K = (b_{K-1}, 1]$ and the representation points a_k are in $\mathcal{R}_k$.[1] A quantizer can be viewed as a nonlinear function $v_K(\cdot)$ such that $v_K(p_0) = a_k$ for $p_0 \in \mathcal{R}_k$. For a given K, we would like to find the quantizer that minimizes the MBRE

$$D = E[d(P_0, v_K(P_0))] = \int d(p_0, v_K(p_0))f_{P_0}(p_0)dp_0. \quad (8)$$

There is no closed-form solution, but an optimal quantizer must satisfy the nearest neighbor condition, the centroid condition, and the zero probability of boundary condition [13]. The nearest neighbor and centroid conditions are developed for MBRE in Sections III-A and III-B. When $f_{P_0}(p_0)$ is absolutely continuous, the zero probability of boundary condition is always satisfied.

A. Nearest Neighbor Condition

With the representation points $\{a_k\}$ fixed, an expression for the interval boundaries $\{b_k\}$ is derived. Given any $p_0 \in [a_k, a_{k+1}]$, if $\tilde{J}(p_0, a_k) < \tilde{J}(p_0, a_{k+1})$, then Bayes risk error is minimized if p_0 is represented by a_k, and if $\tilde{J}(p_0, a_k) > \tilde{J}(p_0, a_{k+1})$, then Bayes risk error is minimized if p_0 is represented by a_{k+1}. The boundary point $b_k \in [a_k, a_{k+1}]$ is the abscissa of the point at which the lines $\tilde{J}(p_0, a_k)$ and $\tilde{J}(p_0, a_{k+1})$ intersect. The idea is illustrated graphically in Fig. 1.

By manipulating the slopes and intercepts of $\tilde{J}(p_0, a_k)$ and $\tilde{J}(p_0, a_{k+1})$, the point of intersection is found to be

$$b_k = \frac{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1}) - p_E^{\mathrm{II}}(a_k)\right)}{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1}) - p_E^{\mathrm{II}}(a_k)\right) - c_{10}\left(p_E^{\mathrm{I}}(a_{k+1}) - p_E^{\mathrm{I}}(a_k)\right)}. \quad (9)$$

---

[1]Due to the strict convexity of $d(p_0, a)$ in p_0 for all a shown in Theorem 1, quantizers that satisfy the necessary conditions for MBRE optimality are regular; see [13, Lemma 6.2.1]. Therefore, only regular quantizers are considered.
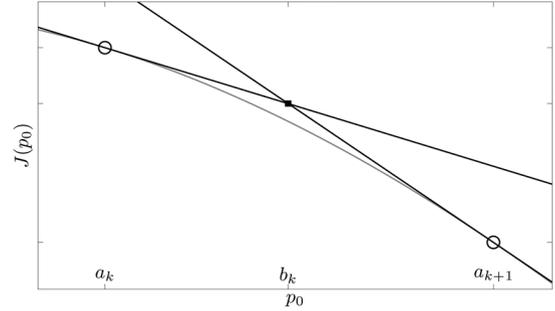

Fig. 1. Intersection of the lines $\bar{J}(p_0, a_k)$, tangent to $J(p_0)$ at a_k, and $\tilde{J}(p_0, a_{k+1})$, tangent to $J(p_0)$ at a_{k+1}, is the optimal interval boundary.

B. Centroid Condition

With the quantization regions fixed, the MBRE is to be minimized over the $\{a_k\}$. Here, the MBRE is expressed as the sum of integrals over quantization regions

$$D = \sum_{k=1}^{K} \int_{\mathcal{R}_k} \left(\tilde{J}(p_0, a_k) - J(p_0)\right)f_{P_0}(p_0)dp_0. \quad (10)$$

Because the regions are fixed, the minimization may be performed for each interval separately.

Let us define $I_k^{\mathrm{I}} = \int_{\mathcal{R}_k} p_0 f_{P_0}(p_0)dp_0$ and $I_k^{\mathrm{II}} = \int_{\mathcal{R}_k}(1 - p_0)f_{P_0}(p_0)dp_0$, which are conditional means. Then

$$a_k = \arg\min_a \left\{c_{10}I_k^{\mathrm{I}}p_E^{\mathrm{I}}(a) + c_{01}I_k^{\mathrm{II}}p_E^{\mathrm{II}}(a)\right\}. \quad (11)$$

Because $\beta p_E^{\mathrm{I}}(a) + \gamma p_E^{\mathrm{II}}(a)$ has exactly one stationary point, which is a minimum (cf. Theorem 2), (11) is uniquely minimized by setting its derivative equal to zero. Thus, a_k is the solution to

$$c_{10}I_k^{\mathrm{I}}\left.\frac{dp_E^{\mathrm{I}}(a)}{da}\right|_{a_k} + c_{01}I_k^{\mathrm{II}}\left.\frac{dp_E^{\mathrm{II}}(a)}{da}\right|_{a_k} = 0. \quad (12)$$

Commonly, differentiation of the two error probabilities is tractable; they are themselves integrals of the likelihood functions and the differentiation is with respect to some function of the limits of integration.

C. Lloyd–Max Algorithm

Alternating between the nearest neighbor and centroid conditions, the iterative Lloyd–Max algorithm can be applied to find minimum MBRE quantizers [13]. The algorithm is widely used because of its simplicity, effectiveness, and convergence properties [14].

In [15], it is shown that the conditions necessary for optimality of the quantizer are also sufficient conditions for local optimality[2] if the following hold. The first condition is that $f_{P_0}(p_0)$ must be positive and continuous in $(0, 1)$. The second condition is that $\int_0^1 d(p_0, a)f_{P_0}(p_0)dp_0$ must be finite for all a. The first and second conditions are met by common distributions such as the beta distribution [16].

---

[2]By local optimality, it is meant that $\{a_k\}$ and $\{b_k\}$ minimize the objective function (8) among feasible representation and boundary points near them.

The third condition is that the distortion function $d(p_0, a)$ must satisfy some properties. It must be zero only for $p_0 = a$, continuous in p_0 for all a, and convex in a; the first two of these hold as discussed in Theorem 1. The third, convexity in a, does not hold for Bayes risk error in general, but the convexity of $d(p_0, a)$ in a is only used by [15] to show that a unique minimum exists. As shown in Theorem 2, $d(p_0, a)$ has a unique stationary point that is a minimum. Therefore, the analysis of [15] applies to Bayes risk error distortion. Thus, if $f_{P_0}(p_0)$ satisfies the first and second conditions, then the algorithm is guaranteed to converge to a local optimum. The algorithm may be run many times with different initializations to find the global optimum.

Further conditions on $d(p_0, a)$ and $f_{P_0}(p_0)$ are given in [15] for there to be a unique locally optimal quantizer, i.e., the global optimum. If these further conditions for unique local optimality hold, then the algorithm is guaranteed to find the globally minimum MBRE quantizer.

In many practical situations, the distribution $f_{P_0}(p_0)$ is not available, but data drawn from it is available. The optimal design of quantizers from data is NP-hard [17], [18]. However, the Lloyd–Max algorithm and its close cousin K-means can be used on data with the Bayes risk error fidelity criterion. In fact, as the size of the data set increases, the sequence of quantizers designed from data converges to the quantizer designed from $f_{P_0}(p_0)$[19], [20]. (Conditions on the distortion function given in [20] except convexity in a are met by the Bayes risk error, but in a similar way to the sufficiency of the Lloyd–Max conditions, the unique minimum property of the Bayes risk error is enough.)

D. Monotonic Convergence in K

Let $D^*(K) = \sum_{k=1}^{K} \int_{\mathcal{R}_k^*} d(p_0, a_k^*) f_{P_0}(p_0) dp_0$ denote the MBRE for an optimal K-point quantizer. We show that $D^*(K)$ monotonically converges as K increases. The MBRE-optimal K-point quantizer is the solution to the following problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{K} \int_{b_{k-1}}^{b_k} d(p_0, a_k) f_{P_0}(p_0) dp_0 \\ \text{such that} \quad & b_0 = 0 \\ & b_K = 1 \\ & b_{k-1} \leq a_k, \qquad k = 1, \ldots, K \\ & a_k \leq b_k, \qquad k = 1, \ldots, K. \end{aligned} \quad (13)$$

Let us add the additional constraint $b_{K-1} = 1$ to (13), forcing $a_K = 1$ and degeneracy of the Kth quantization region. The optimization problem for the K-point quantizer (13) with the additional constraint is equivalent to the optimization problem for the $(K-1)$-point quantizer. Thus, the $(K-1)$-point design problem and the K-point design problem have the same objective function, but the $(K-1)$-point problem has an additional constraint. Therefore, $D^*(K-1) \geq D^*(K)$.

Because $d(p_0, v_K(p_0)) \geq 0$, $D = E[d(P_0, v_K(P_0))] \geq 0$. Because the sequence $D^*(K)$ is nonincreasing and bounded from below, it converges. MBRE cannot get worse when more quantization levels are employed. In typical settings, as in Section V, performance always improves with an increase in the number of quantization levels.

IV. HIGH-RATE QUANTIZATION

Let us apply high-rate quantization theory [14] to the study of minimum MBRE quantization. The distortion function for the MBRE criterion has a positive second derivative in p_0 (due to strict convexity) and for many families of likelihood functions, it has a continuous third derivative; see (7). Thus, it is locally quadratic in the sense of Li *et al.* [12] and in a manner similar to many perceptual, nondifference distortion functions, the high-rate quantization theory is well developed.

At high rate, i.e., K large, if we let

$$\begin{aligned} B(p_0) = & -\frac{1}{2} c_{10} p_0 \frac{d^2 p_E^{\mathrm{I}}(p_0)}{dp_0^2} - c_{10} \frac{dp_E^{\mathrm{I}}(p_0)}{dp_0} \\ & -\frac{1}{2} c_{01}(1-p_0) \frac{d^2 p_E^{\mathrm{II}}(p_0)}{dp_0^2} + c_{01} \frac{dp_E^{\mathrm{II}}(p_0)}{dp_0} \quad (14) \end{aligned}$$

then $d(p_0, a_k)$ is approximated by the following second-order Taylor expansion:

$$d(p_0, a_k) \approx B(p_0)|_{p_0 = a_k} (p_0 - a_k)^2, \qquad p_0 \in \mathcal{R}_k. \quad (15)$$

Assuming that $f_{P_0}(\cdot)$ is sufficiently smooth and substituting (15) into the objective of (13), the MBRE is approximated by

$$D \approx \sum_{k=1}^{K} f_{P_0}(a_k) B(a_k) \int_{\mathcal{R}_k} (p_0 - a_k)^2 dp_0. \quad (16)$$

The MBRE is greater than and approximately equal to the following lower bound, derived in [12] by relationships involving normalized moments of inertia of intervals $\mathcal{R}_k$ and by Hölder's inequality:

$$D_L = \frac{1}{12K^2} \int_0^1 B(p_0) f_{P_0}(p_0) \lambda(p_0)^{-2} dp_0 \quad (17)$$

where the optimal quantizer point density is

$$\lambda(p_0) = \frac{(B(p_0) f_{P_0}(p_0))^{1/3}}{\int_0^1 (B(p_0) f_{P_0}(p_0))^{1/3} dp_0}. \quad (18)$$

Integrating a quantizer point density over an interval yields the fraction of the $\{a_k\}$ that are in that interval. Substituting (18) into (17) yields

$$D_L = \frac{1}{12K^2} \|B(p_0) f_{P_0}(p_0)\|_{1/3}. \quad (19)$$

V. EXAMPLES

As an example, let us consider the following scalar signal and measurement model:

$$Y = s_m + W, \qquad m \in \{0, 1\} \quad (20)$$

where $s_0 = 0$ and $s_1 = \mu$ (a known, deterministic quantity), and W is a zero-mean, Gaussian random variable with variance σ^2. The likelihoods are

$$\begin{aligned} f_{Y|H}(y|h_0) &= \mathcal{N}(y; 0, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} \\ f_{Y|H}(y|h_1) &= \mathcal{N}(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}. \quad (21) \end{aligned}$$

The two error probabilities are

$$p_E^{\mathrm{I}}(p_0) = Q\left(\frac{\mu}{2\sigma} + \frac{\sigma}{\mu}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right)$$
$$p_E^{\mathrm{II}}(p_0) = Q\left(\frac{\mu}{2\sigma} - \frac{\sigma}{\mu}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right) \qquad (22)$$

where

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

Finding the centroid condition, the derivatives of the error probabilities are

$$\frac{dp_E^{\mathrm{I}}(p_0)}{dp_0}\bigg|_{a_k} = -\frac{1}{\sqrt{2\pi}}\frac{\sigma}{\mu}\frac{1}{a_k(1-a_k)}$$
$$\times \exp\left(-\frac{1}{2}\left(\frac{\mu}{2\sigma}+\frac{\sigma}{\mu}\ln\left(\frac{c_{10}a_k}{c_{01}(1-a_k)}\right)\right)^2\right) \tag{23}$$

$$\frac{dp_E^{\mathrm{II}}(p_0)}{dp_0}\bigg|_{a_k} = +\frac{1}{\sqrt{2\pi}}\frac{\sigma}{\mu}\frac{1}{a_k(1-a_k)}$$
$$\times \exp\left(-\frac{1}{2}\left(\frac{\mu}{2\sigma}-\frac{\sigma}{\mu}\ln\left(\frac{c_{10}a_k}{c_{01}(1-a_k)}\right)\right)^2\right). \tag{24}$$

By substituting these derivatives into (12) and simplifying, the following expression is obtained for the representation points:

$$a_k = \frac{I_k^{\mathrm{I}}}{I_k^{\mathrm{I}} + I_k^{\mathrm{II}}}. \tag{25}$$

For high-rate analysis, the second derivatives of the error probabilities are needed. They are

$$\frac{d^2 p_E^{\mathrm{I}}(p_0)}{dp_0^2} = -\frac{1}{\sqrt{8\pi}}\frac{\sigma}{\mu}\frac{1}{p_0^2(1-p_0)^2}$$
$$\times \exp\left(-\frac{1}{8\mu^2\sigma^2}\left(\mu^2+2\sigma^2\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right)^2\right)$$
$$\times \left[-3+4p_0-\frac{2\sigma^2}{\mu^2}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right] \tag{26}$$

and

$$\frac{d^2 p_E^{\mathrm{II}}(p_0)}{dp_0^2} = +\frac{1}{\sqrt{8\pi}}\frac{\sigma}{\mu}\frac{1}{p_0^2(1-p_0)^2}$$
$$\times \exp\left(-\frac{1}{8\mu^2\sigma^2}\left(\mu^2-2\sigma^2\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right)^2\right)$$
$$\times \left[-1+4p_0-\frac{2\sigma^2}{\mu^2}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right]. \tag{27}$$

By inspection, we note that the third derivatives are continuous. Substituting the first derivatives (23) and (24) and second derivatives (26) and (27) into (14), an expression for $B(p_0)$ can be obtained.

Examples with different distributions $f_{P_0}(p_0)$ are presented below. All of the examples use scalar signals with additive Gaussian noise, $\mu = 1$, $\sigma = 1$ (20). As a point of reference, a
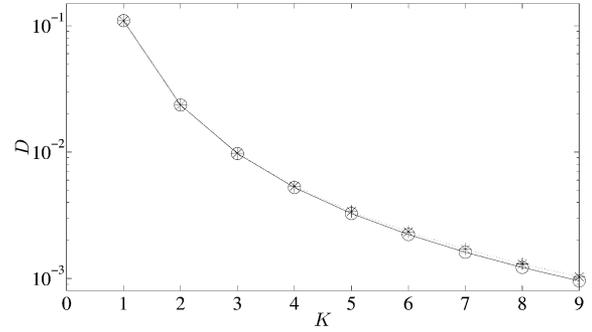

Fig. 2. MBRE for uniformly distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$ plotted on a logarithmic scale as a function of the number of quantization levels K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal uniform quantizer.

comparison is made to quantizers designed under mean absolute error (MAE) [21], i.e., $d(p_0, a) = |p_0 - a|$, an objective that does not account for hypothesis testing.[3]

In the high-rate comparisons, the optimal point density for MAE [23]

$$\lambda(p_0) = \frac{f_{P_0}(p_0)^{1/2}}{\int_0^1 f_{P_0}(p_0)^{1/2}dp_0}$$

is substituted into the high-rate distortion approximation for the MBRE criterion (17). Taking $R = \log_2(K)$, there is a constant gap between the rates using the MBRE point density and the MAE point density for all distortion values. This difference is

$$R_{\mathrm{MBRE}}(D_L) - R_{\mathrm{MAE}}(D_L)$$
$$= \frac{1}{2}\log_2\left(\frac{\|f_{P_0}(p_0)B(p_0)\|_{1/3}}{\|f_{P_0}(p_0)\|_{1/2}\int_0^1 B(p_0)dp_0}\right).$$

The closer the ratio inside the logarithm is to one, the closer the MBRE- and MAE-optimal quantizers are.

A. Uniformly Distributed P_0

We first look at the setting in which all prior probabilities are equally likely. The MBRE of the MBRE-optimal quantizer and a quantizer designed to minimize MAE with respect to $f_{P_0}(p_0)$ are plotted in Fig. 2. (The optimal MAE quantizer for the uniform distribution is the uniform quantizer.) The plot shows MBRE as a function of K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal quantizer. D_L, the high-rate approximation to the distortion-rate function, is plotted in Fig. 3.

The performance of both quantizers is similar, but the MBRE-optimal quantizer always performs better or equally. For $K = 1, 2$, the two quantizers are identical, as seen in Fig. 4(a) and (b).

---

[3]As shown by Kassam [21], minimizing the MAE criterion also minimizes the absolute distance between the cumulative distribution function of the source and the induced cumulative distribution function of the quantized output. Because the induced distribution from quantization is used as the population prior distribution for hypothesis testing, requiring this induced distribution to be close to the true unquantized distribution is reasonable. If distance between probability distributions is to be minimized according to the Kullback–Leibler discrimination between the true and induced distributions (which is defined in terms of likelihood ratios), an application of Pinsker's inequality shows that a small absolute difference is requisite [22]. Although a reasonable criterion, MAE is suboptimal for hypothesis testing performance as seen in the examples.
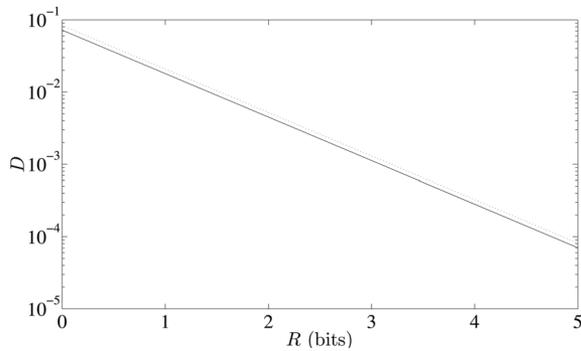
Fig. 3. High-rate approximation of distortion-rate function D_L for uniformly distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$; the solid line is the MBRE-optimal quantizer and the dotted line is the MAE-optimal uniform quantizer.
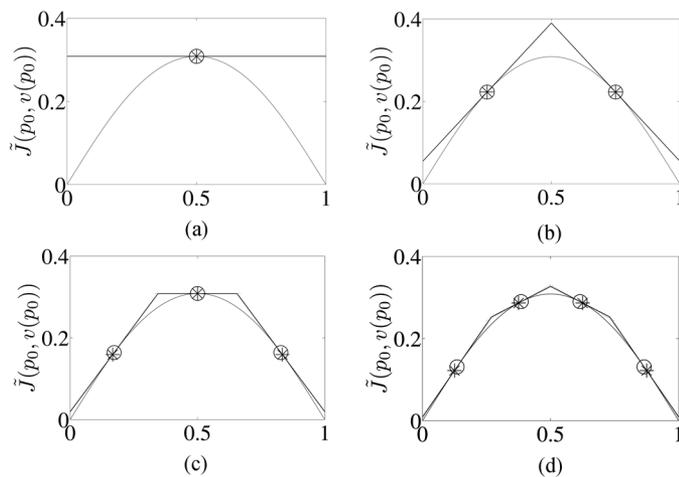

Fig. 5. Optimal MBRE point density for uniformly distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$.

Fig. 4. Quantizers for uniformly distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$. $\tilde{J}(p_0, v_K(p_0))$ is plotted for (a) $K = 1$, (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$; the markers, circle and asterisk for the MBRE-optimal and MAE-optimal quantizers, respectively, are the representation points $\{a_k\}$. The gray line is the unquantized Bayes risk $J(p_0)$.

Fig. 6. MBRE for uniformly distributed P_0 and Bayes costs $c_{10} = 1, c_{01} = 4$ plotted on a logarithmic scale as a function of the number of quantization levels K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal uniform quantizer.

Fig. 7. High-rate approximation of distortion-rate function D_L for uniformly distributed P_0 and Bayes costs $c_{10} = 1, c_{01} = 4$; the solid line is the MBRE-optimal quantizer and the dotted line is the MAE-optimal uniform quantizer.

The plots in Fig. 4 show $\tilde{J}(p_0, v_K(p_0))$ as solid and dotted lines for the MBRE- and MAE-optimal quantizers, respectively; the markers are the representation points. The gray line is $J(p_0)$, the Bayes risk with unquantized prior probabilities. For $K = 3, 4$, the representation points for the MBRE-optimal quantizer are closer to $p_0 = 1/2$ than the uniform quantizer. This is because the area under the point density function $\lambda(p_0)$ shown in Fig. 5 is concentrated in the center. Each increment of K is associated with a large reduction in Bayes risk. There is a very large performance improvement from $K = 1$ to $K = 2$.

In Figs. 6–9, similar plots to those above are given for the case when the Bayes costs c_{10} and c_{01} are unequal. The unequal costs skew the Bayes risk function, and consequently, the representation point locations and point density function. The difference in performance between the MBRE-optimal and MAE-optimal quantizers is greater in this example because the MAE-criterion cannot incorporate the Bayes costs, which factor into MBRE calculation.
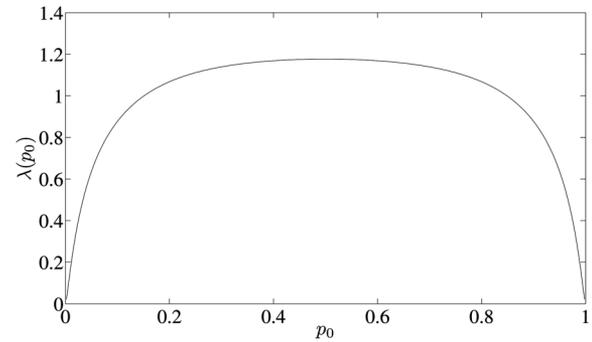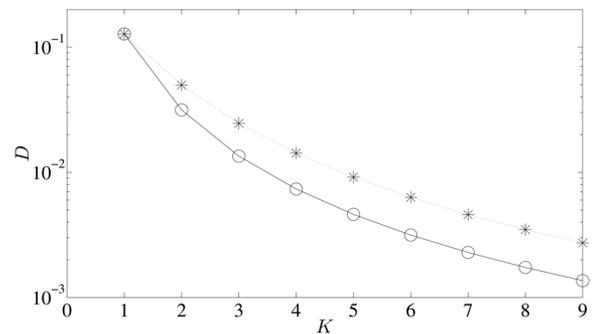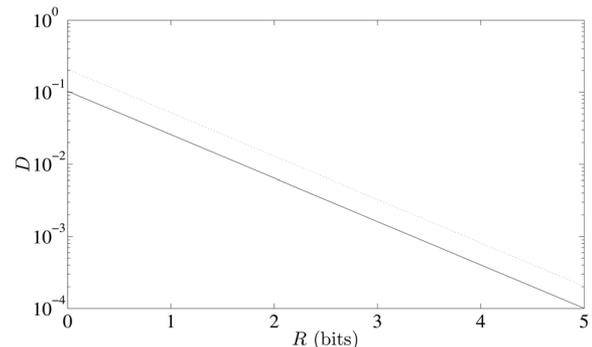
B. Beta Distributed P_0

Now, we look at a nonuniform distribution for P_0, in particular, the Beta(5,2) distribution. The probability density function is shown in Fig. 10. The MBRE of the MBRE-optimal and MAE-optimal quantizers are in Fig. 11. Here, there are also large improvements in performance with an increase in K. The high-rate approximation to the distortion-rate function for this example is given in Fig. 12.

The representation points $\{a_k\}$ are most densely distributed where $\lambda(p_0)$, plotted in Fig. 13, has mass. In particular, more
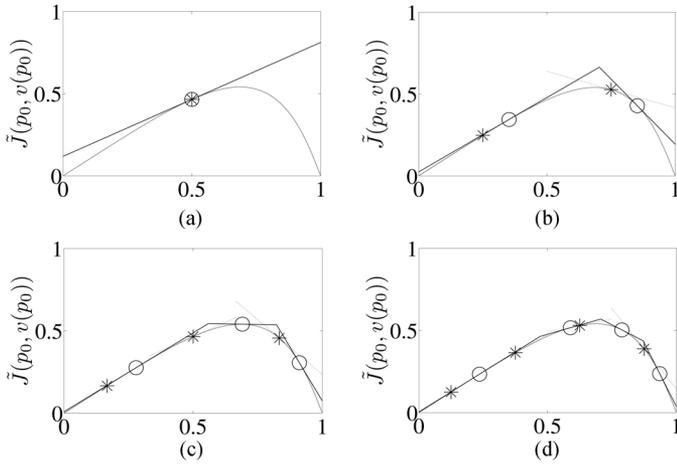
Fig. 8. Quantizers for uniformly distributed P_0 and Bayes costs $c_{10} = 1$, $c_{01} = 4$. $\tilde{J}(p_0, v_K(p_0))$ is plotted for (a) $K = 1$, (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$; the markers, circle and asterisk for the MBRE-optimal and MAE-optimal quantizers, respectively, are the representation points $\{a_k\}$. The gray line is the unquantized Bayes risk $J(p_0)$.
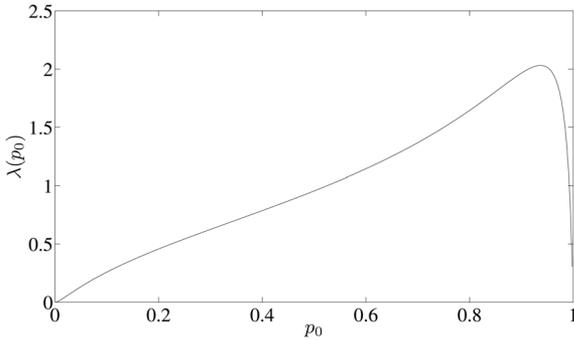

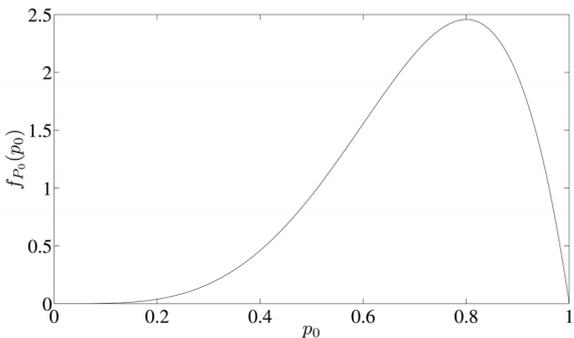Fig. 9. Optimal MBRE point density for uniformly distributed P_0 and Bayes costs $c_{10} = 1$, $c_{01} = 4$.

Fig. 10. Probability density function $f_{P_0}(p_0)$ for the Beta(5,2) distribution.

Fig. 11. MBRE for Beta(5,2) distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$ plotted on a logarithmic scale as a function of the number of quantization levels K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal uniform quantizer.

Fig. 12. High-rate approximation of distortion-rate function D_L for Beta(5,2) distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$; the solid line is the MBRE-optimal quantizer and the dotted line is the MAE-optimal uniform quantizer.

Fig. 13. Optimal MBRE point density for Beta(5,2) distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$.

representation points are in the right half of the domain than in the left, as seen in Fig. 14.

VI. IMPLICATIONS ON HUMAN DECISION MAKING

In the previous sections, we formulated the minimum MBRE quantization problem and discussed how to find the optimal MBRE quantizer. Having established the mathematical foundatio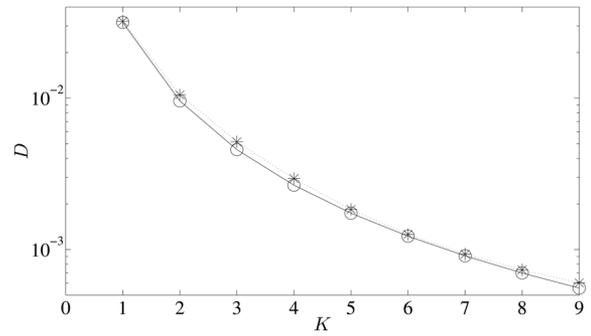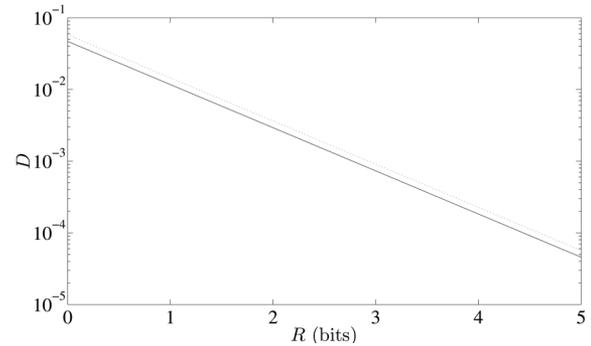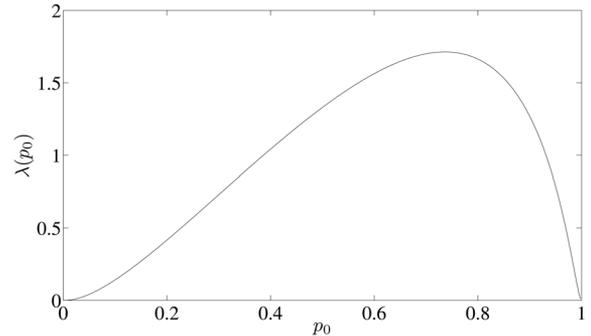ns of hypothesis testing with quantized priors, we may explore the implications of such resource-constrained decision making on human affairs.

Let us consider the particular setting for human decision making mentioned in Section I: a referee determining whether a player has committed a foul using both his or her noisy observation and prior experience. The fraction of plays in which a player commits a foul is that player's prior probability for h_1. Over the population of players, there is a distribution of prior probabilities. Also, as mentioned in Section I, human decision makers categorize into a small number of categories due to limitations in information processing capacity [3]. Decisions by humans may be modeled via quantization of the distribution of
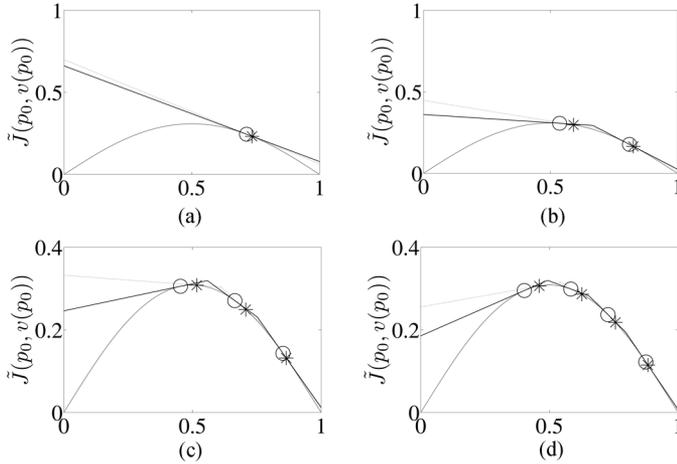
Fig. 14. Quantizers for Beta(5,2) distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$. $\tilde{J}(p_0, v_K(p_0))$ is plotted for (a) $K = 1$, (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$; the markers, circle and asterisk for the MBRE-optimal and MAE-optimal quantizers, respectively, are the representation points $\{a_k\}$. The gray line is the unquantized Bayes risk $J(p_0)$.

prior probabilities and the use of the quantization level centroid of the category in which a player falls as the prior probability when performing hypothesis testing on that player's action.

Therefore, a referee will do a better job with more categories rather than fewer. A police officer confronting an individual with whom he or she has prior experience will make a better decision if he or she has the mental categories "probably violent," "possibly violent or nonviolent," and "probably nonviolent," versus just "violent" and "nonviolent." Similarly, a doctor will have a smaller probability of error when interpreting a blood test if he or she knows the prior probability of the test turning out positive for many categorizations of patients rather than just one for the entire population at large. Additional examples could be given for a variety of decision-making tasks. Implications of this sort are not surprising. However, when one additional component is added to the decision-making scenario, some fairly interesting implications arise. Next, we look at the case when the quantization of two distinct populations is done separately.

We discuss mathematically unavoidable consequences of quantized prior hypothesis testing when quantizing the prior probability for a minority population and the prior probability for a majority population separately, while taking identical prior probability distributions of the two populations $f_{P_0}(p_0)$. Although majority and minority populations can be defined along any socially observable dimension, such as gender or age [24], for ease of exposition, we use race, and more specifically, use "white" and "black" to denote the two populations. Although there is some debate in the social cognition literature [25], it is thought that race and gender categorization is essentially automatic, particularly, when a human actor lacks the motivation, time, or cognitive capacity to think deeply.

We can extend the definition of MBRE to two populations as

$$D^{(2)} = \frac{w}{w+b} E[\tilde{J}(P_0, v_{K_w}(P_0))]$$
$$+ \frac{b}{w+b} E[\tilde{J}(P_0, v_{K_b}(P_0))] - E[J(P_0)] \quad (28)$$

where w is the number of whites encountered, b is the number of blacks encountered, K_w is the number of points in the quantizer for whites, and K_b is the number of points in the quantizer for blacks.[4] To find the optimal allocation of the total quota of representation points $K_t = K_w + K_b$, we minimize $D^{(2)}$ for all $K_t - 1$ possible allocations and choose the best one; more sophisticated algorithms developed for bit allocation to subbands in transform coding may also be used [27].

Fryer and Jackson have previously suggested that it is better to allocate more representation points to the majority population than to the minority population [28]. With two separate scalar quantizers, but a single size constraint, optimizing $D^{(2)}$ over $v_{K_w}(\cdot)$ and $v_{K_b}(\cdot)$ yields the same result. Due to the monotonicity result in Section III-D, the MBRE for members of the minority group is greater than that for the majority group.

Assuming white decision makers have $w > b$ and black decision makers have $b > w$, analysis of quantized prior Bayesian hypothesis testing predicts that there should be own-race bias in decision making. This prediction is in fact born out experimentally. A large body of literature in face recognition shows exactly the predicted own race bias effect, observed colloquially as "they [other-race persons] all look alike." In particular, both parts of the Bayes risk p_E^I and p_E^{II} increase when trying to recognize members of the opposite population [29]. Verification of own race bias in face recognition is due to laboratory experimentation, however similar effects have also been observed in natural experiments through econometric studies.

It has been found that the addition of police officers of a given race is associated with an increase in the number of arrests of suspects of a different race but has little impact on same-race arrests. The effect is more pronounced for minor offenses where the prior probability presumably plays a bigger role than the measurement [30]. There are similar own-race bias effects in the decision by police to search a vehicle during a traffic stop [31], in the decision of human resource professionals to not hire [32], and in the decision of National Basketball Association (NBA) referees to call a foul [33]. The rate of searching, the rate of not hiring, and the rate of foul calling are all greater when the decision-maker is of a different race than the driver, applicant, and player, respectively. A major difficulty in interpreting these econometric studies, however, is that the ground truth is not known. Higher rates may be explained by either greater p_E^I or smaller p_E^{II}.

Because ground truth is lacking in econometric studies, it is not clear how to interpret a finding that white referees call more fouls on black players and that black referees call more fouls on white players. This phenomenon cannot simply be explained by a larger probability of decision error. The Bayes risk must be teased apart into its constituent parts and the Bayes costs must be examined in detail. The measurable quantity in an econometrics study is the probability that a foul is called

$$\Pr[\hat{H}_K = h_1] = 1 - p_0 + p_0 p_E^I(v_K(p_0))$$
$$- (1 - p_0) p_E^{II}(v_K(p_0)). \quad (29)$$

[4]One might assume that w and b are simply the number of whites and blacks in the general population, however these numbers should actually be based on the social interaction pattern of the decision maker. Due to segregation in social interaction, see, e.g., [26] and references therein, there is greater intrapopulation interaction than interpopulation interaction. The decision maker has more training data from intrapopulation interaction.

Looking at the average performance of a white referee over the populations of black and white players, we compare the expected foul rates on whites and blacks ($K_b < K_w$)

$$\Delta = E\left[\Pr[\hat{H}_{K_b} = h_1] - \Pr[\hat{H}_{K_w} = h_1]\right]. \quad (30)$$

If this discrimination quantity Δ is greater than zero, then the white referee is calling more fouls on blacks. If Δ is less than zero, then the referee is calling more fouls on whites. The Δ expression may be written as

$$\Delta(c_{10}, c_{01}) = E\left[p_0 p_E^{\mathrm{I}}(v_{K_b}(p_0)) - (1 - p_0)p_E^{\mathrm{II}}(v_{K_b}(p_0))\right]$$
$$- E\left[p_0 p_E^{\mathrm{I}}(v_{K_w}(p_0)) - (1 - p_0)p_E^{\mathrm{II}}(v_{K_w}(p_0))\right]. \quad (31)$$

The dependence of Δ on c_{10} and c_{01} is explicit on the left-hand side of (31) and is implicit in the error probabilities on the right-hand side. The value of Δ also depends on the unquantized prior distribution $f_{P_0}(p_0)$, the measurement model, and the quantizer.

If the prior distribution and the measurement model are fixed, and the MBRE-optimal quantizer used, we find that the regions in the c_{10}–c_{01} plane where a white referee would call more fouls on blacks and where a white referee would call more fouls on whites are half-planes. For the uniform prior $f_{P_0}(p_0)$, the dividing line between the two regions is exactly $c_{01} = c_{10}$. For the Beta(5,2) prior, the dividing line is $c_{01} = mc_{10}$, where $m > 1$.

Using the division of the c_{10}–c_{01} plane into two parts, we can now interpret the econometric findings in the NBA referee study [33] and related results [30]–[32]. The NBA race bias observations can be generated from the quantized prior hypothesis testing model only if the Bayes risk error has costs $c_{01} > c_{10}$ for a uniform prior or costs $c_{01} \gg c_{10}$ for a Beta(5,2) prior. The choice of Bayes costs with c_{01} greater than c_{10} implies that a referee can tolerate more instances of calling fouls on plays that are not fouls rather than the opposite. This assignment of costs has been called the precautionary principle in some contexts. Very simply, the precautionary principle states "better safe than sorry."

Taken together, the hypothesis testing with quantized priors model, the phenomenon of racial segregation [26], and the results from econometric studies [30]–[33] suggest that referees, police officers, and human resources professionals all follow the precautionary principle.

VII. CONCLUSION AND FUTURE WORK

We have looked at Bayesian hypothesis testing when there is a distribution of prior probabilities, but the decision maker may only use a quantized version of the true prior probability in designing a decision rule. Considering the problem of finding the optimal quantizer for this purpose, we have defined a new fidelity criterion based on the Bayes risk function. For this criterion, MBRE, we have determined the conditions that an optimal quantizer satisfies and worked through a high-rate approximation to the distortion. M-ary hypothesis testing with $M > 2$ requires vector quantization rather than scalar quantization, but determining the Lloyd–Max conditions and high-rate theory is no different conceptually due to the geometry of the Bayes risk function and mismatched Bayes risk function. For the M-ary hypothesis testing case, a multivariate distribution such as the M-dimensional Dirichlet distribution [16] is needed for $f_{\boldsymbol{P}}(\boldsymbol{p})$. Previous, though significantly different, work on quantization for hypothesis testing was unable to directly minimize the Bayes risk, as was accomplished in this work.

The mathematical theory of quantized prior hypothesis testing formulated here leads to a generative model of discriminative behavior when combined with theories of social cognition and empirical facts about social segregation. This biased decision making arises despite having identical distributions for different populations and despite no malicious intent on the part of the decision maker. We also discussed how the choice of Bayes costs affects detection probabilities; in particular, the precautionary principle leads to a higher detection probability for the opposite race, whereas a more optimistic view leads to a higher detection probability for the own race. Such a phenomenon of pessimistic or optimistic attitude fundamentally altering the nature of discrimination seems not to have been described before. Discrimination on the basis of race, gender, and other socially observable characteristics has been a troublesome social problem, but appears to be a permanent artifact of the automaticity of classification and the finite human capacity for information processing.

There are many avenues along which to extend this work, such as dealing with decentralized detection and classification (with possible implications on jury decisions and elections), which may become game theoretic; consideration of additional noise before or after quantization of the prior probabilities; or the development of successively refinable quantizers (for decision makers that possess a memory hierarchy). One can also consider a restricted class of quantizers rather than considering optimal quantization. Such restriction may model further cognitive constraints on human decision makers. In particular, Fryer and Jackson have suggested a heuristic algorithm for quantizer design based on splitting groups [28], which is a rediscovery of the tree-structured vector quantizer (TSVQ) design algorithm given in [34, Fig. 20]. Beyond [34], there has been much recent development in the theory of TSVQ performance and recursive partitioning, which may prove useful.

For the quantizer with $K = 1$, an alternative to the MBRE-optimal representation point

$$\boldsymbol{a}_{\mathrm{MBRE}}^* = \arg\min_{\boldsymbol{a}}\left\{\int \tilde{J}(\boldsymbol{p}, \boldsymbol{a}) f_{\boldsymbol{P}}(\boldsymbol{p}) d\boldsymbol{p}\right\}$$

is the min–max hypothesis testing representation point

$$\boldsymbol{a}_{\mathrm{min-max}}^* = \arg\min_{\boldsymbol{a}}\left\{\max_{\boldsymbol{p}} \tilde{J}(\boldsymbol{p}, \boldsymbol{a})\right\}$$

which is only equivalent in special cases. A distribution on the prior probabilities is needed to specify $\boldsymbol{a}_{\mathrm{MBRE}}^*$, but not to specify $\boldsymbol{a}_{\mathrm{min-max}}^*$. One may consider extending the min–max idea to $K > 1$. This would involve an approach related to ϵ-entropy [35, Sec. 6.1.2], and finding a cover for the unit simplex by K sets of the form $\mathcal{R}_k = \{\boldsymbol{p}|\tilde{J}(\boldsymbol{p}, \boldsymbol{a}_k) \le D\}$, where all $\boldsymbol{p}$ in $\mathcal{R}_k$ map to $\boldsymbol{a}_k$ and D is the same for all $\mathcal{R}_k$.

The general theme of machine learning for the explicit purpose of hypothesis testing, within which this work falls, is receiving increasing attention; framing the hypothesis testing scenario discussed here in terms of probabilistic graphical models of categorization, e.g., the latent Dirichlet allocation model [36] and the hierarchical Dirichlet process mixture model [37], may prove insightful as well.

REFERENCES

[1] K. R. Varshney and L. R. Varshney, "Minimum mean Bayes risk quantization of prior probabilities," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Las Vegas, NV, Apr. 2008, pp. 3445–3448.

[2] A. S. Willsky, G. W. Wornell, and J. H. Shapiro, "Stochastic processes, detection and estimation 6.432 course notes," Dept. Elect. Eng. Comput. Sci., Mass. Inst. Technol., Cambridge, MA, 2003.

[3] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81–97, 1956.

[4] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. COM-25, no. 5, pp. 479–484, May 1977.

[5] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 893–900, Sep. 1977.

[6] R. Gupta and A. O. Hero, III, "High-rate vector quantization for detection," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1951–1969, Aug. 2003.

[7] C. Hildreth, "Bayesian statisticians and remote clients," *Econometrica*, vol. 31, no. 7, pp. 422–438, Jul. 1963.

[8] R. E. Kihlstrom, "The use of approximate prior distributions in a Bayesian decision model," *Econometrica*, vol. 39, no. 6, pp. 899–910, Nov. 1971.

[9] D. C. Gilliland and M. K. Helmers, "On continuity of the Bayes response," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 506–508, Jul. 1978.

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

[11] M. H. DeGroot, *Optimal Statistical Decisions.* Hoboken, NJ: Wiley-Interscience, 2004.

[12] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1082–1091, May 1999.

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

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

[15] 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.

[16] T. L. Fine, *Probability and Probabilistic Reasoning for Electrical Engineering.* Upper Saddle River, NJ: Prentice-Hall, 2006.

[17] M. R. Garey, D. S. Johnson, and H. S. Witsenhausen, "The complexity of the generalized Lloyd-Max problem," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 255–256, Mar. 1982.

[18] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay, "Clustering in large graphs and matrices," in *Proc. 10th Annu. ACM-SIAM Symp. Discrete Algorithms*, Baltimore, MD, 1999, pp. 291–299.

[19] R. M. Gray, J. C. Kieffer, and Y. Linde, "Locally optimal block quantizer design," *Inf. Control*, vol. 45, no. 2, pp. 178–198, 1980.

[20] M. J. Sabin and R. M. Gray, "Global convergence and empirical consistency of the generalized Lloyd algorithm," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 2, pp. 148–155, Mar. 1986.

[21] S. A. Kassam, "Quantization based on the mean-absolute-error criterion," *IEEE Trans. Commun.*, vol. COM-26, no. 2, pp. 267–270, Feb. 1978.

[22] F. Topsøe, "Some inequalities for information divergence and related measures of discrimination," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1602–1609, Jul. 2000.

[23] R. M. Gray and A. H. Gray, Jr, "Asymptotically optimal quantizers," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 143–144, Jan. 1977.

[24] G. A. Akerlof and R. E. Kranton, "Economics and identity," *Quart. J. Econom.*, vol. 115, no. 3, pp. 715–753, Aug. 2000.

[25] C. N. Macrae and G. V. Bodenhausen, "Social cognition: Thinking categorically about others," *Annu. Rev. Psychol.*, vol. 51, pp. 93–120, Feb. 2000.

[26] F. Echenique and R. G. Fryer, Jr, "A measure of segregation based on social interactions," *Quart. J. Econom.*, vol. 122, no. 2, pp. 441–485, May 2007.

[27] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.

[28] R. Fryer and M. O. Jackson, "A categorical model of cognition and biased decision-making," *B. E. J. Theor. Econom.*, vol. 8, no. 1, Jan. 2008.

[29] C. A. Meissner and J. C. Brigham, "Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review," *Psychol. Pub. Pol. L.*, vol. 7, no. 1, pp. 3–35, Jan. 2001.

[30] J. J. Donohue, III and S. D. Levitt, "The impact of race on policing and arrests," *J. Law Econom.*, vol. 44, no. 2, pp. 367–394, Oct. 2001.

[31] K. L. Antonovics and B. G. Knight, "A new look at racial profiling: Evidence from the Boston Police Department," *Rev. Econom. Statist.*, to be published.

[32] M. A. Stoll, S. Raphael, and H. J. Holzer, "Black job applicants and the hiring officer's race," *Ind. Lab. Relat. Rev.*, vol. 57, no. 2, pp. 267–287, Jan. 2004.

[33] J. Price and J. Wolfers, "Racial discrimination among NBA referees," National Bureau of Economic Research, Jun. 2007, Working Paper 13206.

[34] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proc. IEEE*, vol. 73, no. 11, pp. 1551–1588, Nov. 1985.

[35] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[36] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[37] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

Kush R. Varshney (S'00) was born in Syracuse, NY, in 1982. He received the B.S. degree in electrical and computer engineering with honors from Cornell University, Ithaca, NY, in 2004 and the S.M. degree in electrical engineering and computer science from Massachusetts Institute of Technology, Cambridge, in 2006, where he is currently working towards the Ph.D. degree.

He is a National Science Foundation Graduate Research Fellow and a research assistant with the Stochastic Systems Group in the Laboratory for Information and Decision Systems. He was a visiting student at Laboratoire de Mathématiques Appliquées aux Systèmes at École Centrale, Paris, France, in 2006. He was an intern at Sun Microsystems during 2002–2003 and at Sensis Corporation in 2001. His research interests include statistical signal processing, image processing, and machine learning.

Mr. Varshney is a member of Eta Kappa Nu and Tau Beta Pi, and a student member of the Society for Industrial and Applied Mathematics (SIAM).

Lav R. Varshney (S'00) was born in Syracuse, NY, on October 28, 1982. He received the B.S. degree with honors in electrical and computer engineering (*magna cum laude*) from Cornell University, Ithaca, NY in 2004 and the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006, where he is currently working towards the Ph.D. degree in electrical engineering and computer science.

He is a National Science Foundation Graduate Research Fellow and has been a research and teaching assistant at MIT. He was a visitor at the École Polytechnique Fédérale de Lausanne, Switzerland, in 2006, a visiting scientist at Cold Spring Harbor Laboratory in 2005, a research engineering intern at Syracuse Research Corporation during 2002–2003, and an engineering intern at Sensis Corporation in 2001. His research interests include information theory, coding, and theoretical neuroscience.

Mr. Varshney is a member of Tau Beta Pi and Eta Kappa Nu. He received the E. A. Guillemin S.M. Thesis Award, the Capocelli Prize at the 2006 Data Compression Conference, the Best Student Paper Award at the 2003 IEEE Radar Conference, and was a winner of the IEEE 2004 Student History Paper Contest.