2011 Data Compression Conference

Collaboration in Distributed Hypothesis Testing with Quantized Prior Probabilities*

Joong Bum Rhim†, Lav R. Varshney‡, and Vivek K Goyal†

†Research Laboratory of Electronics, Massachusetts Institute of Technology

‡IBM Thomas J. Watson Research Center

Abstract

The effect of quantization of prior probabilities in a collection of distributed Bayesian binary hypothesis testing problems over which the priors themselves vary is studied. In a setting with fusion of local binary decisions by majority rule, optimal local decision rules are discussed. Quantization is first considered under the constraint that agents employ identical quantizers. A method for design is presented that exploits an equivalence to a single-agent problem with a different likelihood function; the optimal quantizers are thus different than in the single-agent case. Removing the constraint of identical quantizers is demonstrated to improve performance. A method for design is presented that exploits an equivalence between agents having diverse K-level quantizers and agents having identical $(3K-2)$-level quantizers.

1 Introduction

In many settings, agents collaborate to make decisions under various uncertainties and information constraints. As an example, consider three physicians consulted by a patient for advice on whether to be treated for a disease. Each physician runs independent and identical diagnostic tests that are not in themselves conclusive. Each physician thus combines the test results with an assessment of the patient's susceptibility to the disease to determine whether to recommend treatment. Note that recommendations naturally incorporate judgments on the relative detriments of leaving the disease untreated and treating despite lack of disease. The patient will undergo treatment if a majority of the physicians recommend it.

In this scenario, the performance of the collaborating physicians is affected by the efficacies of their tests and by their limitation of communicating only binary decisions to the patient; these are well-studied aspects of multi-agent decision making. The focus in this paper is on a different limitation that each physician faces and on the resultant effects on collaborative performance: In interpreting test results, a physician cannot use a decision rule that is perfectly personalized to the patient; instead, the physician must assign the patient to a category and use a decision rule designed for that category. This categorization is justified both by cognitive limitations of the physician [1] and by limitations in rates of learning from data [2, Ch. 5.3]. We study the formation of categories, which may be different for different physicians. We

*This material is based upon work supported by National Science Foundation Grant No. 0729069.

1068-0314/11 $26.00 © 2011 IEEE
DOI 10.1109/DCC.2011.37

assume that the judgments on relative detriments of the two types of misdiagnosis are the same for all physicians; a companion paper removes this assumption [3].

Abstracting from this example, we study Bayesian distributed binary hypothesis testing under an observation model specified by likelihood functions $f_{Y_i|H}(y \mid h_0)$ and $f_{Y_i|H}(y \mid h_1)$, where i indexes the agents. To formalize having an ensemble of problems, we consider the prior probability distribution over hypotheses $\{h_0, h_1\}$ to itself be random with a known distribution. Since $\mathbb{P}(H = h_1) = 1 - \mathbb{P}(H = h_0)$, the pmf of H is described by a single scalar $p_0 = \mathbb{P}(H = h_0)$. We model p_0 as a realization of random variable P_0 with a known pdf f_{P_0}. Categorization of the ensemble of problems is equated with quantization of P_0 to one of K levels. Our focus is on the design of quantizers for P_0 that minimize Bayes risk, which here is a single quantity for the system of agents. It is computed using a single set of Bayes costs and the probabilities of Type I and Type II errors of the collaborative decision made by majority voting.

Related work. Varshney and Varshney [4] recently introduced the precise study of quantization of prior probabilities in Bayesian hypothesis testing. They focus on binary hypothesis testing by a single agent and also briefly discuss M-ary hypothesis testing for $M > 2$. This paper applies and extends results of [4] to collaborative decision making by multiple agents. The introduction of multiple agents highlights the role of the information fusion strategy of agents and introduces a new quantizer design problem with a perhaps-surprising solution.

Most previous work on the effect of quantization in Bayesian distributed detection is focused on the quantization of observations [5–7] or the communication topology and rates among agents [8] and/or to the fusion center [9, 10]. We do not consider quantization of observations here, though it may be noted that quantization outside of the system designer's control could be incorporated into the likelihood functions. We will comment on the impact of simple majority voting based on one-bit communication from agents to the fusion center in Sec. 3.

The use of a single set of Bayes costs is an element of making the agents a *team* in the sense of Marschak and Radner [11], i.e. having a common goal. An alternative is for each agent to have potentially-different Bayes costs. This introduces game-theoretic considerations as described in a companion paper [3].

Paper organization and preview of main results. Sec. 2 formalizes our setting and defines notation; in particular, it defines the Bayes risk objective function of interest and the function that takes its place when the prior probability is quantized. Sec. 3 considers the case in which agents use identical quantizers and identical decision thresholds. It is shown that when the agents observe H corrupted by independent and identically-distributed additive noises, the team performance is equivalent to a single agent with noise given by the median of the noises. This enables optimal design through extension of results from [4]. Sec. 4 considers the case in which agents may use different quantizers. The main result is that optimizing the performance of three agents using K-level quantizers matches the best performance attainable with three agents using $(3K - 2)$-level quantizers that are constrained to be identical. This is dramatic improvement from diversity, and it enables a design methodology relying on the results of Sec. 3.

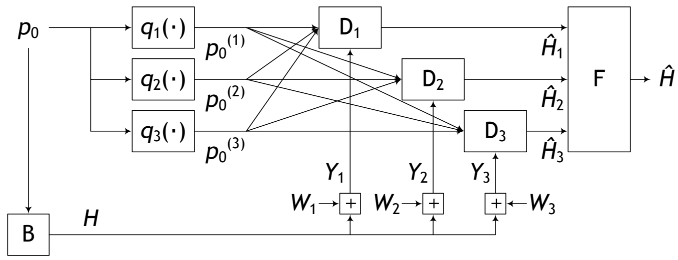


Fig. 1: A schematic diagram depicting the problem information pattern. The environment B generates a Bernoulli signal H. Its prior probability p_0 is quantized by three separate quantizers; the results are used by local agents D_i. Each agent also has access to H corrupted by i.i.d. noise W_i. The fusion center F determines $\hat{H}$ from the local decisions $\hat{H}_i$.

2 Bayesian Binary Hypothesis Testing Problem

Formulation with p_0 given. Consider a group of agents deciding between $H = h_0$ and $H = h_1$ when $\mathbb{P}(H = h_0 \mid P_0 = p_0) = p_0$ is given. The scenario of interest is shown in Fig. 1 (ignoring the $q_i(\cdot)$ blocks) for the case of three agents and additive observation noise. For each i, Agent i (marked D_i) observes Y_i satisfying likelihood function $f_{Y_i|H}$ and sends a local decision $\widehat{H}_i \in \{h_0, h_1\}$ to a fusion center. The observations are assumed to be conditionally independent given H. The fusion center determines $\widehat{H} \in \{h_0, h_1\}$ by majority rule.

Agent i has (local) Type I and Type II error probabilities given by

$$P_{e,i}^{\mathrm{I}} \;=\; \mathbb{P}(\widehat{H}_i = h_1 \mid H = h_0) \qquad \text{and} \qquad P_{e,i}^{\mathrm{II}} \;=\; \mathbb{P}(\widehat{H}_i = h_0 \mid H = h_1).$$

The Type I and Type II error probabilities of the collaborative (global) decision are

$$P_E^{\mathrm{I}} \;=\; \mathbb{P}(\widehat{H} = h_1 \mid H = h_0) \qquad \text{and} \qquad P_E^{\mathrm{II}} \;=\; \mathbb{P}(\widehat{H} = h_0 \mid H = h_1).$$

All four of these quantities depend on p_0 and on the decision rules used by the agents. This dependence will sometimes be shown explicitly. Since global errors occur exactly when the majority of agents make local errors, the global error probabilities can be expressed in terms of the local error probabilities; for example, with three agents

$$P_E^{\mathrm{I}} \;=\; \mathbb{P}(\widehat{H} = h_1 \mid H = h_0) \;=\; P_{e,1}^{\mathrm{I}} P_{e,2}^{\mathrm{I}} + P_{e,2}^{\mathrm{I}} P_{e,3}^{\mathrm{I}} + P_{e,3}^{\mathrm{I}} P_{e,1}^{\mathrm{I}} - 2 P_{e,1}^{\mathrm{I}} P_{e,2}^{\mathrm{I}} P_{e,3}^{\mathrm{I}}, \quad (1)$$

$$P_E^{\mathrm{II}} \;=\; \mathbb{P}(\widehat{H} = h_0 \mid H = h_1) \;=\; P_{e,1}^{\mathrm{II}} P_{e,2}^{\mathrm{II}} + P_{e,2}^{\mathrm{II}} P_{e,3}^{\mathrm{II}} + P_{e,3}^{\mathrm{II}} P_{e,1}^{\mathrm{II}} - 2 P_{e,1}^{\mathrm{II}} P_{e,2}^{\mathrm{II}} P_{e,3}^{\mathrm{II}}. \quad (2)$$

The goal of each agent is to minimize the expected value of the *Bayes risk*

$$R \;=\; c_{10} p_0 P_E^{\mathrm{I}} + c_{01}(1 - p_0) P_E^{\mathrm{II}},$$

where c_{10} and c_{01} are positive constants. Type I and Type II errors incur *Bayes costs* of c_{10} and c_{01}, while correct decisions incur no cost; through the definitions of P_E^{I} and P_E^{II}, it is clear that R is the conditional mean of the Bayes cost given $P_0 = p_0$.

Some of the results and all of the numerical examples are based on the additive white Gaussian noise (AWGN) observation model depicted in Fig. 1, with the noise variables $\{W_i\}_{i=1}^3$ independent Gaussian random variables with mean zero and variance σ^2. In other words, the likelihood functions $f_{Y_i|H}(\cdot \mid h)$ are Gaussian pdfs with mean h and variance σ^2.

Optimal decision rules. For notational simplicity, consider the decision rule for Agent 1 in a setting with three agents. The optimal decision for Agent 1 minimizes Bayes risk (4) with $i = 1$. Suppose that Agents 2 and 3 have some fixed decision rules. By rewriting (1) and (2), we have

$$\begin{aligned} P_E^{\mathrm{I}} &= \left(P_{e,2}^{\mathrm{I}} + P_{e,3}^{\mathrm{I}} - 2P_{e,2}^{\mathrm{I}}P_{e,3}^{\mathrm{I}}\right) P_{e,1}^{\mathrm{I}} + P_{e,2}^{\mathrm{I}}P_{e,3}^{\mathrm{I}} \;\triangleq\; A_{11}P_{e,1}^{\mathrm{I}} + A_{12}, \\ P_E^{\mathrm{II}} &= \left(P_{e,2}^{\mathrm{II}} + P_{e,3}^{\mathrm{II}} - 2P_{e,2}^{\mathrm{II}}P_{e,3}^{\mathrm{II}}\right) P_{e,1}^{\mathrm{II}} + P_{e,2}^{\mathrm{II}}P_{e,3}^{\mathrm{II}} \;\triangleq\; B_{11}P_{e,1}^{\mathrm{II}} + B_{12}, \end{aligned}$$

where A_{11}, A_{12}, B_{11}, and B_{12} are nonnegative quantities that do not depend on the decision rule of Agent 1. The optimal decision rule $\widehat{H}_1(y_1)$ for Agent 1 can be expressed as a likelihood ratio test

$$\frac{f_{Y_1|H}(y_1 \mid h_1)}{f_{Y_1|H}(y_1 \mid h_0)} \underset{\widehat{H}_1(y_1)=h_0}{\overset{\widehat{H}_1(y_1)=h_1}{\gtrless}} \frac{c_{10}p_0A_{11}}{c_{01}(1-p_0)B_{11}} \;\triangleq\; \eta_1,$$

which is similar to the standard optimal decision rule for a single agent.

Under the AWGN observation model, Agent 1 has optimal decision rule

$$y_1 \underset{\widehat{H}_1(y_1)=h_0}{\overset{\widehat{H}_1(y_1)=h_1}{\gtrless}} \frac{h_1 - h_0}{2} + \frac{\sigma^2}{h_1 - h_0} \ln \eta_1 \;\triangleq\; \lambda_1. \tag{3}$$

The optimal decision rules of Agents 2 and 3 are analogous, with thresholds λ_2 and λ_3. Note that $(P_{e,1}^{\mathrm{I}}, P_{e,1}^{\mathrm{II}})$, $(P_{e,2}^{\mathrm{I}}, P_{e,2}^{\mathrm{II}})$, and $(P_{e,3}^{\mathrm{I}}, P_{e,3}^{\mathrm{II}})$, depend directly on λ_1, λ_2, and λ_3, respectively. The optimal value of λ_1 in (3) is a function of λ_2 and λ_3 through η_1 (hence A_{11} and B_{11}), and similarly for λ_2 and λ_3. Thus, the λ_is cannot be optimized independently.

Quantization of p_0. As depicted in Fig. 1, the decision rule of Agent i is based on a quantized version of the prior probability $p_0^{(i)} = q_i(p_0)$. Thus, Agent i makes decisions to minimize *perceived Bayes risk*

$$\bar{R}_i = c_{10}p_0^{(i)} P_E^{\mathrm{I}} + c_{01}\left(1 - p_0^{(i)}\right) P_E^{\mathrm{II}}. \tag{4}$$

The agent's decision rule, and consequently $P_{e,i}^{\mathrm{I}}$ and $P_{e,i}^{\mathrm{II}}$, are determined based on $\bar{R}_i$. The (true) Bayes risk with decision rules impacted by quantization of the prior probability is denoted $\tilde{R}_i$. Based on this discussion, the mean Bayes risk (MBR)

$$\mathbb{E}[\tilde{R}] = \int_0^1 (c_{10}p_0P_E^{\mathrm{I}}(q(p_0)) + c_{01}(1-p_0)P_E^{\mathrm{II}}(q(p_0)))f_{P_0}(p_0)\,dp_0$$

is the appropriate fidelity criterion for the quantizers $q_i(\cdot)$ applied to P_0; see also [4].

3 Agents with Identical Prior-Probability Quantizers

In this section, we assume that all agents use the same quantizer. This causes all perceived Bayes risks to be equal to

$$\bar{R} = c_{10}q(p_0)P_E^{\mathrm{I}} + c_{01}(1 - q(p_0))P_E^{\mathrm{II}},$$

so we further assume that all decision rules are identical. This in turn implies that all local probabilities of error are equal for both Type I and Type II. In the AWGN observation model, identical decision rules correspond to all λ_i thresholds taking one common value λ.

An equivalent single-agent model. Consider any odd number of agents using the decision rule

$$y_i \underset{\widehat{H}_i(y_i)=h_0}{\overset{\widehat{H}_i(y_i)=h_1}{\gtrless}} \lambda \tag{5}$$

for some λ. With this rule, $\widehat{H}_i = h_1$ implies $\widehat{H}_j = h_1$ for any j such that $y_j > y_i$; similarly, $\widehat{H}_i = h_0$ implies $\widehat{H}_j = h_0$ for any j such that $y_j < y_i$. Thus, comparing the median y_i to λ determines the majority-rule decision. In fact, the performance of the collaborating group is the same as that of a single agent with observation likelihood function $f_{Y|H}$ determined by the median of the group's observations. This equivalence is most easily understood in the case of an additive noise observation model.

Theorem 1. *Suppose $2n+1$ agents make decisions using the rule (5) for some λ, where the observation of Agent i is $Y_i = H + W_i$. The collaborative performance by majority rule is equal to the performance of a single agent using rule (5) with observation $Y = H + V$, where $V = \mathrm{median}(\{W_i\}_{i=1}^{2n+1})$.*

When the W_is are independent and identically distributed, their median is well understood from the theory of order statistics [12]. In general, the variance of the median is bounded as $\mathrm{var}(V) \leq \mathrm{var}(W_i)$, with equality attained uniquely by Bernoulli W_is [13]. For continuous noise distributions, as of interest here, the inequality is strict and the variance of V decreases with n. If the W_is are continuous with pdf f_W,

$$f_V(v) = \frac{(2n+1)!}{(n!)^2}[F_W(v)]^n[1 - F_W(v)]^n f_W(v), \tag{6}$$

where $F_W(v) = \int_{-\infty}^{v} f_W(w)\,dw$ is the cdf of W. For the case of Gaussian W_is, $\mathrm{var}(V) \approx 0.449\,\mathrm{var}(W)$ for 3 agents and $\mathrm{var}(V) \approx 0.287\,\mathrm{var}(W)$ for 5 agents. A natural comparison is to $\frac{1}{3}\,\mathrm{var}(W)$ and $\frac{1}{5}\,\mathrm{var}(W)$, which are equivalent noise variances if agents average observations rather than sharing only their (one-bit) decisions.

Optimal quantization. Consider collaboration by $2n+1$ agents under a continuous additive noise observation model. The agents use decision rule (5) and a K-level quantizer $q(\cdot)$, which partitions the support $[0, 1]$ of P_0 into cells $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_K$ and has representation points $a_1, a_2, \ldots, a_K$. Using Theorem 1, this team of agents can be replaced by a single agent with additive noise defined by (6). Thus, optimization of the quantizer in the multi-agent model can be converted to the optimization problem for a single agent discussed in [4].

As discussed in [4], MBR-optimal quantizers are regular. Each cell is an interval: $\mathcal{R}_1 = [b_0, b_1]$, $\mathcal{R}_2 = (b_1, b_2]$, $\ldots$, $\mathcal{R}_K = (b_{K-1}, b_K]$, where $0 = b_0 < b_1 < b_2 <$

$\cdots < b_{K-1} < b_K = 1$, and each representation point a_k belongs to the cell $\mathcal{R}_k$, $b_{k-1} < a_k < b_k$. The true Bayes risk averaged over P_0 is

$$\mathbb{E}[\tilde{R}] = \sum_{k=1}^{K} \int_{b_{k-1}}^{b_k} \left(c_{10} p_0 P_E^{\mathrm{I}}(a_k) + c_{01}(1-p_0) P_E^{\mathrm{II}}(a_k)\right) f_{P_0}(p_0)\, dp_0,$$

where $P_E^{\mathrm{I}}(a)$ and $P_E^{\mathrm{II}}(a)$ are determined based on the perceived Bayes risk when agents use a as the prior probability.

For any fixed partitioning, MBR-optimal quantizers should satisfy

$$a_k = \operatorname*{arg\,min}_{a \in (b_{k-1}, b_k]} \int_{b_{k-1}}^{b_k} (c_{10} p_0 P_E^{\mathrm{I}}(a) + c_{01}(1-p_0) P_E^{\mathrm{II}}(a)) f_{P_0}(p_0)\, dp_0 \tag{7}$$

for every k, which is the centroid condition with respect to MBR. Since the integral on the right side of (7) has only one stationary point that is a minimum extremum [4, Thm. 2], a_k is the unique solution to

$$\left(\int_{b_{k-1}}^{b_k} c_{10} p_0 f_{P_0}(p_0)\, dp_0\right) \left.\frac{P_E^{\mathrm{I}}(a)}{da}\right|_{a_k} + \left(\int_{b_{k-1}}^{b_k} c_{01}(1-p_0) f_{P_0}(p_0)\, dp_0\right) \left.\frac{P_E^{\mathrm{II}}(a)}{da}\right|_{a_k} = 0. \tag{8}$$

For any fixed set of reproduction points, MBR-optimal encoding should map p_0 to the a_k value that minimizes MBR:

$$k = \operatorname*{arg\,min}_{k' \in \{1,2,\ldots,K\}} \left[c_{10} p_0 P_E^{\mathrm{I}}(a_{k'}) + c_{01}(1-p_0) P_E^{\mathrm{II}}(a_{k'})\right].$$

This yields the nearest-neighbor condition with respect to MBR: For $p_0 \in [a_k, a_{k+1}]$,

$$c_{10} p_0 P_E^{\mathrm{I}}(a_k) + c_{01}(1-p_0) P_E^{\mathrm{II}}(a_k) \underset{p_0 \in \mathcal{R}_k}{\overset{p_0 \in \mathcal{R}_{k+1}}{\gtrless}} c_{10} p_0 P_E^{\mathrm{I}}(a_{k+1}) + c_{01}(1-p_0) P_E^{\mathrm{II}}(a_{k+1}).$$

Thus, the boundary between $\mathcal{R}_k$ and $\mathcal{R}_{k+1}$ is b_k yielding equality above at $p_0 = b_k$:

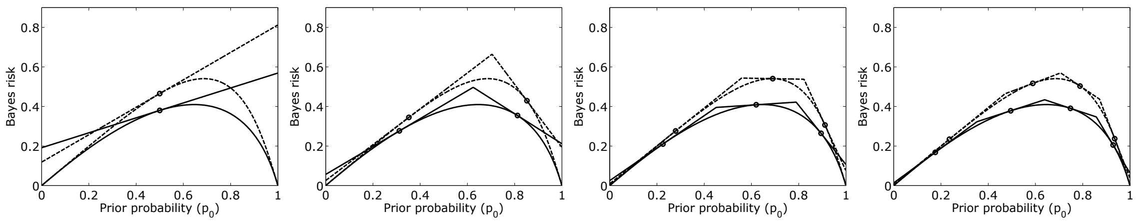
$$b_k = \frac{c_{01}\left(P_E^{\mathrm{II}}(a_{k+1}) - P_E^{\mathrm{II}}(a_k)\right)}{c_{01}\left(P_E^{\mathrm{II}}(a_{k+1}) - P_E^{\mathrm{II}}(a_k)\right) - c_{10}\left(P_E^{\mathrm{I}}(a_{k+1}) - P_E^{\mathrm{I}}(a_k)\right)}. \tag{9}$$

Though MBR-optimal quantizers do not have a closed form, the Lloyd-Max algorithm can find a quantizer that meets the centroid and nearest neighbor conditions by alternating (8) and (9). As given in [4, 14], the algorithm converges to an optimal quantizer if $f_{P_0}(p_0)$ is positive and continuous in $(0, 1)$ and

$$\int_0^1 (c_{10} p_0 P_E^{\mathrm{I}}(a) + c_{01}(1-p_0) P_E^{\mathrm{II}}(a)) f_{P_0}(p_0)\, dp_0$$

is finite for all a.

Fig. 2 provides a set of numerical examples. Optimal quantizers were designed using the Lloyd-Max algorithm for one and three agents and for $K = 1, 2, 3, 4$. The results are shown as plots of Bayes risk as a function of p_0; since P_0 is uniform in this example, MBRs are given by areas under the plotted curves. The examples demonstrate improvement with increasing K and increasing the number of agents. They also demonstrate that the optimal quantizer for a single agent is different than the optimal quantizer to be used identically by a team of agents.



(a) $K = 1$ (b) $K = 2$ (c) $K = 3$ (d) $K = 4$

Fig. 2: Optimal performance with identical K-level quantizers under AWGN observation model, where P_0 is uniformly distributed on $[0, 1]$, $h_0 = 0$, $h_1 = 1$, $\sigma = 1$, and Bayes costs $c_{10} = 1$ and $c_{01} = 4$. Mismatched Bayes risk and unquantized Bayes risk are plotted for one (dashed) and three (solid) agents. Marked points of tangency are the quantizer reproduction points, and points of discontinuity are the quantizer cell boundaries.

4 Agents with Diverse Prior-Probability Quantizers

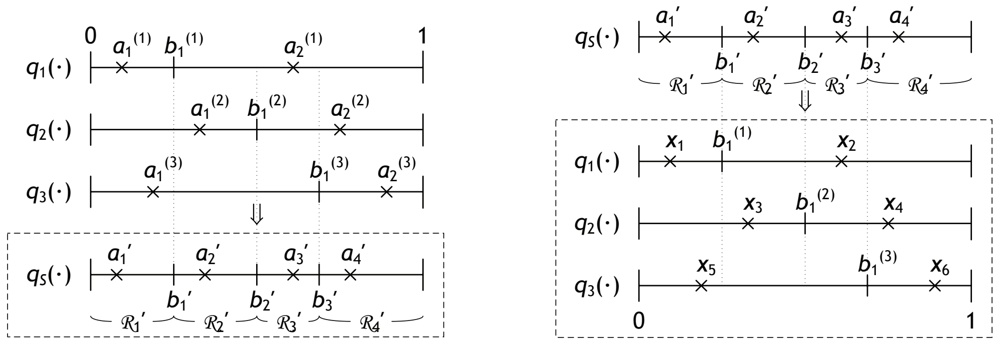
In this section, we remove the restriction that agents use the same quantizer. Differently quantized prior probabilities make the agents' perceived Bayes risks differ even though the agents have the same Bayes costs. The added complication in analysis and optimization is worth it: performance of the team is improved.

We henceforth limit attention to the case of three agents and an equal number of levels K for each agent's quantizer. The quantizers q_1, q_2, and q_3 could be as shown in Fig. 3a. The superscript (i) indicates association with the quantizer of Agent i so that q_i has reproduction points $\{a_k^{(i)}\}_{k=1}^K$ and cell boundaries $\{b_k^{(i)}\}_{k=0}^K$. Since we assume that the agents do not know the true prior p_0, the decision rule is optimized based on the quantized prior probabilities $p_0^{(i)} = q_i(p_0)$. One collaborative way is to minimize the perceived Bayes risk averaged over the three agents, which we call the *perceived common risk*:

$$\bar{R}_C = \frac{1}{3}(\bar{R}_1 + \bar{R}_2 + \bar{R}_3) = \frac{1}{3}c_{10}\left(p_0^{(1)} + p_0^{(2)} + p_0^{(3)}\right)P_E^{\mathrm{I}} + \frac{1}{3}c_{01}\left(3 - p_0^{(1)} - p_0^{(2)} - p_0^{(3)}\right)P_E^{\mathrm{II}}.$$

An equivalence to agents with identical quantizers. Our key result is a performance equivalence between three agents using quantizers (q_1, q_2, q_3) and three agents using a shared quantizer q_{S}. The equivalence enables the optimization of (q_1, q_2, q_3) and shows that when these quantizers are different, the performance achieved is commensurate with having finer quantizers.

Consider again Fig. 3a, which for simplicity has $K = 2$. Each quantizer q_i divides the interval $[0, 1]$ into two cells, but the whole quantization system divides $[0, 1]$ into four cells: $\mathcal{R}'_1 = [0, b_1^{(1)}]$, $\mathcal{R}'_2 = (b_1^{(1)}, b_1^{(2)}]$, $\mathcal{R}'_3 = (b_1^{(2)}, b_1^{(3)}]$, and $\mathcal{R}'_4 = (b_1^{(3)}, 1]$. We thus associate with (q_1, q_2, q_3) a quantizer q_{S} with this partition (and reproduction points yet to be defined). More generally, three regular K-level quantizers can split the entire interval into at most $(3K - 2)$ subintervals. Thus, it is certainly impossible for three agents using different K-level quantizers to achieve performance better than three agents using a shared $(3K-2)$-level quantizer. The equivalence can be achieved.



(a) (b)

Fig. 3: K-level quantizers (q_1, q_2, q_3) and an equivalent $(3K-2)$-level quantizer q_{S}. (a) Conversion from (q_1, q_2, q_3) to q_{S}. (b) Conversion from q_{S} to (q_1, q_2, q_3).

Theorem 2. *Consider three K-level quantizers (q_1, q_2, q_3) and a $(3K-2)$-level quantizer q_{S}. Agents that use (q_1, q_2, q_3) and those that use $(q_{\mathrm{S}}, q_{\mathrm{S}}, q_{\mathrm{S}})$ achieve the same performance for any p_0 if $\bigcup_{i=1}^{3} B_i = B_S$ and $q_{\mathrm{S}}(p_0) = \frac{1}{3}\sum_{i=1}^{3} q_i(p_0)$, where B_1, B_2, B_3, and B_S are the sets of cell boundaries of q_1, q_2, q_3, and q_{S}, respectively.*

Proof. Let T_D denote a team of agents using different quantizers (q_1, q_2, q_3) and T_S denote a team of agents using a shared quantizer $(q_{\mathrm{S}}, q_{\mathrm{S}}, q_{\mathrm{S}})$. Both teams use Bayes costs c_{10} and c_{01}. If $\bigcup_{i=1}^{3} B_i = B_S$, the whole interval $[0,1]$ is divided into the same subintervals by (q_1, q_2, q_3) and by q_{S}. The agents in T_D and T_S will have the same true Bayes risk $c_{10}p_0P_E^{\mathrm{I}} + c_{01}(1-p_0)P_E^{\mathrm{II}}$ for any p_0 if they use the same decision rule in any subinterval $\mathcal{R}_k'$. Since the decision rule should minimize the perceived common risk of agents in T_D and the perceived Bayes risk of agents in T_S, the agents in both groups will use the same decision threshold if

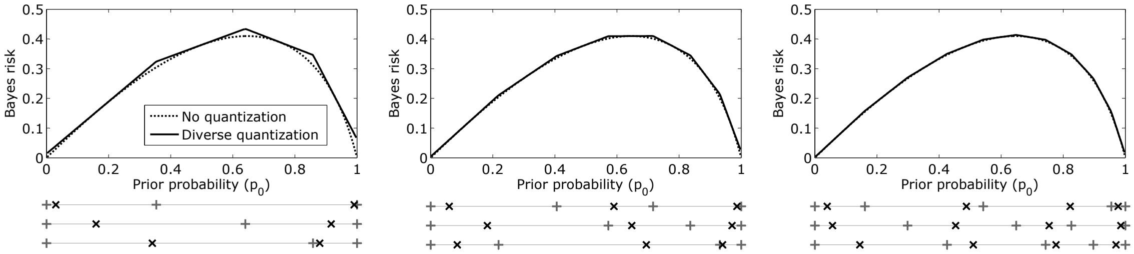
$$\begin{aligned} &\tfrac{1}{3}c_{10}\left[\textstyle\sum_{i=1}^{3} q_i(p_0)\right] P_E^{\mathrm{I}}(\lambda) + \tfrac{1}{3}c_{01}\left[\textstyle\sum_{i=1}^{3}(1-q_i(p_0))\right] P_E^{\mathrm{II}}(\lambda) \\ &\quad = \; z\left(c_{10}q_{\mathrm{S}}(p_0)P_E^{\mathrm{I}}(\lambda) + c_{01}(1-q_{\mathrm{S}}(p_0))P_E^{\mathrm{II}}(\lambda)\right) \end{aligned}$$

for some constant z and any p_0 and λ. Therefore, using different quantizers (q_1, q_2, q_3) is equivalent to using $(q_{\mathrm{S}}, q_{\mathrm{S}}, q_{\mathrm{S}})$ if

$$\frac{\sum_{i=1}^{3} q_i(p_0)}{\sum_{i=1}^{3}(1-q_i(p_0))} = \frac{q_{\mathrm{S}}(p_0)}{1-q_{\mathrm{S}}(p_0)}, \tag{10}$$

or $q_{\mathrm{S}}(p_0) = \frac{1}{3}\sum_{i=1}^{3} q_i(p_0)$ for all $p_0 \in [0,1]$. □

Optimal quantization. Recall that for any K, the optimization of $(3K-2)$-level q_{S} to be shared by three identical agents can be accomplished as outlined in Sec. 3. In light of Theorem 2, agents with diverse K-level quantizers (q_1, q_2, q_3) cannot perform better than the identical agents using q_{S}. If we can create a mapping from the optimized q_{S} to (q_1, q_2, q_3), we will have designed optimal diverse quantizers.



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$

Fig. 4: Performance with optimally designed diverse K-level quantizers with system parameters as in Fig. 2. Mismatched Bayes risk and unquantized Bayes risk are plotted for three values of K. The $K = 2$ performance matches the three-agent performance in Fig. 2d, demonstrating Theorem 2.

Consider the example in Fig. 3b. From (10), each region gives one equation:

$$\begin{aligned} \mathcal{R}'_1: &\quad x_1 + x_3 + x_5 = 3a'_1, &(11a)\\ \mathcal{R}'_2: &\quad x_2 + x_3 + x_5 = 3a'_2, &(11b)\\ \mathcal{R}'_3: &\quad x_2 + x_4 + x_5 = 3a'_3, &(11c)\\ \mathcal{R}'_4: &\quad x_2 + x_4 + x_6 = 3a'_4. &(11d) \end{aligned}$$

There are six unknowns but only four equations; thus, there is no unique solution. Instead, we obtain several conditions about the representation points. First, representation points should lie in $[0, 1]$ because they represent prior probabilities. Second, if regular quantizers are desired, then each representation point should lie in the region that it represents.[1] This gives the following conditions:

$$0 < x_1 < b'_1,\ b'_1 < x_2 < 1,\ 0 < x_3 < b'_2,\ b'_2 < x_4 < 1,\ 0 < x_5 < b'_3,\ b'_3 < x_6 < 1. \quad (12)$$

From (11), x_2, x_3, x_4, and x_5 can be expressed in terms of x_1 and x_6. Hence, after finding a valid pair of (x_1, x_6) that satisfies (12), we can compute other variables, which will give us the full description of (q_1, q_2, q_3). Note that there are infinitely many pairs that satisfy (12). Choosing any pair will result in the same performance with respect to the mean Bayes risk because all of the resulting quantizers are associated with the same $(3K - 2)$-level quantizer.

Fig. 4 provides a set of numerical examples. To enable comparison with Fig. 2, the distribution of P_0 and other parameters are unchanged from before. Quantizers (q_1, q_2, q_3) were designed by first optimizing an associated q_{S} using the Lloyd-Max algorithm and then determining regular quantizers that would yield the same performance. The examples demonstrate the advantage of diversity in quantization.

5 Conclusion

This paper has introduced consideration of quantized prior probabilities in distributed hypothesis testing. The main results are equivalences that simplify optimal quantizer

[1]The K-level quantizers need not be regular.

design. Theorem 1 demonstrates the equivalence in performance between $2n+1$ agents with identical decision thresholds whose observations are corrupted by additive noises $\{W_i\}_{i=1}^{2n+1}$ and a single agent whose observation is corrupted by additive noise $V = \text{median}(\{W_i\}_{i=1}^{2n+1})$; this equivalence enables the design of prior probability quantizers for teams that use the same quantizer. However, using the same quantizer is far from optimal. The equivalence in Theorem 2 shows that a diverse team of three agents is able to nearly triple its effective number of quantization levels through joint optimization toward a common minimum mean Bayes risk goal.

References

[1] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psych. Rev.*, vol. 63, no. 2, pp. 81–97, Mar. 1956.

[2] K. R. Varshney, "Frugal hypothesis testing and classification," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, Jun. 2010.

[3] J. B. Rhim, L. R. Varshney, and V. K. Goyal, "Conflict in distributed hypothesis testing with quantized prior probabilities," in *IEEE Data Compression Conf.*, Mar. 2011, to appear.

[4] 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.

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

[6] 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.

[7] 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.

[8] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, pp. 1383–1400, Mar. 2010.

[9] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI Press, 1993, pp. 297–344.

[10] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[11] J. Marschak and R. Radner, *Economic Theory of Teams*. New Haven: Yale University Press, 1972.

[12] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2003.

[13] G. D. Lin and J. S. Huang, "Variances of sample medians," *Stat. Prob. Lett.*, vol. 8, no. 2, pp. 143–146, Jun. 1989.

[14] 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. 52, no. 4, pp. 1029–1050, Jul. 1984.