

Benefits of Collaboration and Diversity in Teams of Categorically-Thinking Decision Makers

Joong Bum Rhim^{*}, Lav R. Varshney[†], and Vivek K Goyal^{*}

^{*} Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[†] IBM Thomas J. Watson Research Center, Hawthorne, NY 10532, USA

Abstract—Certain information-processing limitations in hypothesis testing can be modeled as quantization of prior probabilities. While quantization hurts performance, a team of decision makers can minimize their performance loss by adopting diverse quantizers and collaborating on the design of their decision rules. In this paper, the benefits of diversity and collaboration in binary hypothesis testing are discussed. A set of N diverse K -level quantizers used by a team of N collaborating decision makers is as powerful as a single $(N(K-1)+1)$ -level quantizer used by them all. If the decision makers do not collaborate, a set of diverse quantizers is less powerful, but it is still better than a set of identical quantizers.

I. INTRODUCTION

Quantization of prior probabilities is a novel restriction on decision makers (DMs) first introduced in [1]. The need for quantization arises from a model for an information-processing limitation. When faced with an ensemble of decision making (hypothesis testing) problems with different prior probability distributions, it is ideal to vary the decision rule—perhaps even having a different decision rule for each problem in the ensemble. This may be impractical, especially when the ensemble is large; for example, human DMs reportedly think with around seven categories without getting confused [2], [3]. *Categorically-thinking* DMs apply one decision rule for each category of problems. For binary decision problems, categories are formed by quantization of a scalar prior probability.

Quantization of prior probabilities leads to suboptimal decision making, but quantizers for prior probabilities can be designed to minimize the cost of wrong decisions due to quantization. Here we evaluate decision making on the ensemble with the mean Bayes risk and the effect of quantization by the mean Bayes risk error (MBRE), which is the mean excess of the Bayes risk caused by quantization. Other criteria are possible, such as minimax Bayes risk [4]

Here we study categorically-thinking DMs in a distributed decision-making system described by voting and integrating. Every DM has an equal vote on binary candidates, and all the votes are integrated by a fixed fusion rule such as MAJORITY rule. The quantization of prior probabilities raises two interesting issues: collaboration and diversity.

DMs that share one cost function can *collaborate* to optimize everyone's decision rules for global objectives [5], [6], whereas DMs that have individual cost functions would

have *conflict* in optimizing decision rules for individual objectives [7]. Even if DMs share one cost function, they may not collaborate on decision making due to lack of communication [8]. Optimal quantizers have different properties in each of these cases, and consequently different design algorithms and performance.

Diversity is defined as DMs using different categorization (quantization of prior probabilities). Diverse quantization is compared to identical quantization in [6], [8]. It is perhaps immediate that DMs with differing objectives would optimally employ different quantizers. Even when DMs have a common cost function, diversity is useful in the sense that a set of diverse quantizers is tantamount to having finer quantization cells than a set of identical quantizers.

A group of DMs produces decisions of different quality depending on whether they collaborate and whether they have diverse quantizers. This paper summarizes extant results in this area. Section II describes the basic decision making model. Section III presents several quantization models and discusses optimal decision rules and quantizers. Section IV shows an example to compare them, and Section V concludes.

II. BASIC DECISION MAKING MODEL

Consider an ensemble of objects that are independently in state h_0 or h_1 with individual prior probabilities. One object is randomly chosen to be a subject of decision making, and its state is denoted by $H \in \{h_0, h_1\}$. The object is distinguishable from others by its prior probability of being in state h_0 , denoted by $p_0 = \mathbb{P}\{H = h_0\}$. Because of the random selection from the ensemble of objects, we consider p_0 to be a realization of a random variable P_0 with a density f_{P_0} .

The distributed decision making system under consideration consists of N DMs and one fusion center. Each DM has two kinds of imperfect information. DM i receives a realization y_i of a random variable Y_i , where Y_1, Y_2, \dots, Y_N are conditionally independent given the state h_m , with identical likelihoods $f_{Y_i|H}(y_i | h_m)$. In addition, DM i knows quantized prior probability $q_i(p_0)$. Using y_i and $q_i(p_0)$, DM i makes a hard decision $\hat{H}_i \in \{h_0, h_1\}$ and sends it to the fusion center. The fusion center makes a global decision based only on the local decisions: $\hat{H} = h_1$ if it receives h_1 from L or more DMs and $\hat{H} = h_0$ otherwise (L -out-of- N fusion rule).

The system incurs cost c_{10} for a false alarm ($\hat{H} = h_1$ when $H = h_0$) and c_{01} for a missed detection ($\hat{H} = h_0$ when $H = h_1$). For simplicity, correct global decisions incur zero cost. Note

This material is based upon work supported by the National Science Foundation under Grants 0729069 and 1101147 and the Korea Foundation for Advanced Studies (KFAS) Fellowship.

that there is a single common cost for the system decision; there are no individual costs for the DMs.

Given $P_0 = p_0$, the system has the Bayes risk

$$R(p_0) = c_{10}p_0P_E^I(p_0) + c_{01}(1-p_0)P_E^{II}(p_0),$$

where $P_E^I(p_0)$ and $P_E^{II}(p_0)$ denote the probabilities of false alarms and missed detections of the system. They are determined by each DM's probabilities of false alarms and missed detections, $P_{e,i}^I(p_0)$ and $P_{e,i}^{II}(p_0)$, in an inclusion-exclusion manner:

$$P_E^I = \sum_{n=L}^N \sum_{\substack{I \subseteq [N] \\ |I|=n}} \prod_{i \in I} P_{e,i}^I \prod_{j \in [N] \setminus I} (1 - P_{e,j}^I), \quad (1)$$

$$P_E^{II} = \sum_{n=N-L+1}^N \sum_{\substack{I \subseteq [N] \\ |I|=n}} \prod_{i \in I} P_{e,i}^{II} \prod_{j \in [N] \setminus I} (1 - P_{e,j}^{II}), \quad (2)$$

where $[N]$ denotes the set $\{1, 2, \dots, N\}$.

The average performance of the decision making system is measured by the mean Bayes risk

$$\mathbb{E}[R] = \int_0^1 R(p_0) f_{P_0}(p_0) dp_0. \quad (3)$$

The quantizers of interest are functional scalar quantizers [9]; the distortion measure is not the error of the prior probability but the error of the mean Bayes risk.

III. QUANTIZATION MODELS

In this section, we introduce several variations on the basic decision making model. These are compared quantitatively in Section IV.

Purely to serve as a baseline for comparison—i.e., not falling within the model of the previous section—suppose the DMs communicate their observations $\{y_i\}$ (rather than only hard decisions $\{\hat{H}_i\}$) perfectly to the fusion center, and the fusion center has perfect knowledge of the prior probability p_0 . Then the Bayes risk is minimized by the following likelihood ratio test:

$$\frac{f_{Y_1, \dots, Y_N | H}(y_1, \dots, y_N | 1)}{f_{Y_1, \dots, Y_N | H}(y_1, \dots, y_N | 0)} \underset{\hat{H}(y)=0}{\overset{\hat{H}(y)=1}{>}} \frac{c_{10}p_0}{c_{01}(1-p_0)}. \quad (4)$$

In the following, we consider first in Section III-A the restriction of DM-to-fusion center communication to hard decisions; then, in Section III-B, the restriction of the fusion center to use a quantized prior probability. Sections III-C–III-E study variations on the model in Section II.

A. Communication Constraints: Hard Decisions

DMs are connected to the fusion center with a channel of 1-bit capacity. Since the fusion center can only receive hard decisions from the DMs, the fusion center has a simple fixed fusion rule: the L -out-of- N rule. For example, it uses the MAJORITY rule ($\lceil \frac{N+1}{2} \rceil$ -out-of- N rule) or the OR rule (1-out-of- N rule). The DMs are informed of the fusion rule and adjust

their local decision rules to optimize the global objective:

$$\frac{f_{Y_a | H}(y_a | h_1)}{f_{Y_a | H}(y_a | h_0)} \underset{\hat{H}_a(y_a)=h_0}{\overset{\hat{H}_a(y_a)=h_1}{>}} \quad (5)$$

$$\frac{p_0 c_{10} \sum_{\substack{I \subseteq [N] \setminus \{a\} \\ |I|=L-1}} \prod_{i \in I} P_{e,i}^I \prod_{j \in [N] \setminus (I \cup \{a\})} (1 - P_{e,j}^I)}{(1-p_0) c_{01} \sum_{\substack{I \subseteq [N] \setminus \{a\} \\ |I|=N-L}} \prod_{i \in I} P_{e,i}^{II} \prod_{j \in [N] \setminus (I \cup \{a\})} (1 - P_{e,j}^{II})}.$$

The details are explained in [10].

B. Memory Constraints: Quantized Priors

The optimal decision rule at the fusion center is (4), but the DMs know inaccurate prior probabilities $q_i(p_0)$. A fusion center that does not know p_0 must replace p_0 by $\sum_i q_i(p_0)/N$ in (4), which makes the decision rule suboptimal.

C. Hard Decisions and Identical Local Quantizers

Now consider DMs that have to make hard decisions before communicating with a fusion center, as in Section III-A, and in addition use identically quantized prior probabilities. Even though they know the fusion rule, their decision rules that replace p_0 in (5) by $q_a(p_0)$ are worse than (5) because of inexact prior probabilities.

All DMs have the same decision rule because they have identical K -level quantizers $q_1(p_0) = \dots = q_N(p_0) = q(p_0)$, for all p_0 , whose quantization cells are $[0, b_1), [b_1, b_2), \dots, [b_{K-1}, 1]$ and representation points are a_1, \dots, a_K . The optimal K -level quantizer is

$$q^* = \arg \min_q \int_0^1 (c_{10}p_0 P_E^I(q(p_0)) + c_{01}(1-p_0) P_E^{II}(q(p_0))) \times f_{P_0}(p_0) dp_0$$

$$= \arg \min_{a_1^K, b_1^{K-1}} \sum_{k=1}^K \left(c_{10} P_E^I(a_k) \int_{b_{k-1}}^{b_k} p_0 f_{P_0}(p_0) dp_0 + c_{01} P_E^{II}(a_k) \int_{b_{k-1}}^{b_k} (1-p_0) f_{P_0}(p_0) dp_0 \right).$$

Since the total distortion is the sum of the distortion in each cell, the optimization of representation points in each cell can be performed separately.

In the case that the observations are of the state corrupted by conditionally-independent additive noise, this model is equivalent with respect to Bayes risk to a single-DM model with additive noise V , which is the L th largest noise $Y_{(L)}$ among Y_1, \dots, Y_N [6, Theorem 3]. This equivalence guarantees us that the optimization of the quantizer performed in the single-DM model leads to the same result as that performed in the N -DM model [6, Theorem 4]. The nearest neighbor and centroid conditions, which are necessary for optimal quantizers in the single-DM model, are given in [1], and the optimal quantizer can be obtained by the iterative Lloyd–Max algorithm.

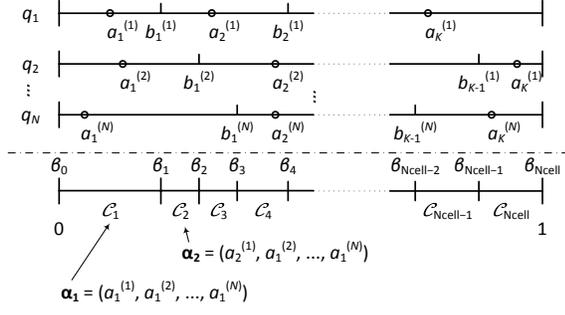


Fig. 1. A set of quantizers q_1, \dots, q_N categorizes cells $\mathcal{C}_1, \dots, \mathcal{C}_{N_{\text{cell}}}$.

D. Hard Decisions and Diverse Local Quantizers

The system has the same constraints as in Section III-C but DMs quantize prior probabilities differently so as to take advantage of diversity. Intuitively, a set of diverse quantizers should be better than a set of identical quantizers. Fig. 1 depicts an example of N diverse quantizers q_1, \dots, q_N , which have K cells $[0, b_1^{(i)}], [b_1^{(i)}, b_2^{(i)}], \dots, [b_{K-1}^{(i)}, 1]$ with corresponding representation points $a_1^{(i)}, a_2^{(i)}, \dots, a_K^{(i)}$, $i = 1, \dots, N$. Then the entire system has N_{cell} cells $\mathcal{C}_1, \dots, \mathcal{C}_{N_{\text{cell}}}$ to quantize prior probabilities, where a set of endpoints $\{0, \beta_1, \beta_2, \dots, \beta_{N_{\text{cell}}-1}, 1\}$, $0 < \beta_1 < \beta_2 < \dots < \beta_{N_{\text{cell}}-1} < 1$, is the union of endpoints of all quantizers q_1, \dots, q_N and $\mathcal{C}_k = [\beta_{k-1}, \beta_k)$. Since the maximum number of N_{cell} is $N(K-1)+1$, which is bigger than K for $N > 1$, this system can quantize prior probabilities with less errors than that in Section III-C.

In this model, the decision rules are based on differently quantized prior probabilities. The optimal quantizers are

$$\begin{aligned}
 & q_1^*, \dots, q_N^* \\
 &= \arg \min_{q_1, \dots, q_N} \int_0^1 (c_{10} p_0 P_E^I(q_1(p_0), \dots, q_N(p_0)) \\
 &\quad + c_{01} (1-p_0) P_E^{II}(q_1(p_0), \dots, q_N(p_0))) f_{P_0}(p_0) dp_0 \\
 &= \arg \min_{a_1^k, b_1^{K-1}} \sum_{k=1}^K \left(c_{10} P_E^I(\alpha_k) \int_{b_{k-1}}^{b_k} p_0 f_{P_0}(p_0) dp_0 \right. \\
 &\quad \left. + c_{01} P_E^{II}(\alpha_k) \int_{b_{k-1}}^{b_k} (1-p_0) f_{P_0}(p_0) dp_0 \right).
 \end{aligned}$$

Unlike in Section III-C, we cannot separately perform the optimization in each cell because α_k , the vectors of representation points of the quantizers within the cell \mathcal{C}_k , are not independent. In Fig. 1, for example, the second elements of α_1 and α_2 should be the same because they both reflect $a_1^{(2)}$. Thus, we extend the Lloyd–Max algorithm to optimize N diverse quantizers. The nearest neighbor and centroid conditions for optimal quantizers and the algorithm are described in [8]. Because of iterations within iterations, the algorithm is more complex than the Lloyd–Max algorithm.

E. Hard Decisions, Diverse Local Quantizers, and Collaboration

To make better use of the diversity in Section III-D, DMs can share their quantized values of prior probabilities. In Section III-D, using the set of diverse K -level quantizers is not as good as using a set of identical N_{cell} -level quantizers because the decision makers do not know others' quantization of prior probabilities. If they know others' quantized values of prior probability, then all decision makers can use better decision rules, which replace p_0 in (5) by $\frac{1}{N} \sum_{i=1}^N q_i(p_0)$. Sharing the quantized prior probabilities for better decision rules is what we call *collaboration*.

Under collaboration, there is an equivalence between diverse quantization and identical quantization. Using a set of diverse quantizers q_1, \dots, q_N is as good as using identical quantizer $q_S(\cdot) = \frac{1}{N} \sum_{i=1}^N q_i(\cdot)$ [6, Theorem 5]. Hence, an optimal set of diverse K -level quantizers is equivalent to an optimal $(N(K-1)+1)$ -level quantizer identically used by N decision makers [6, Theorem 6].

Even though decision makers use diverse quantizers, the design of optimal quantizers in this model is much easier than in the model of Section III-D. The equivalence allows us to design an optimal $(N(K-1)+1)$ -level quantizer q_S first and then disassemble the quantizer into N diverse K -level quantizers that satisfy $q_S(\cdot) = \frac{1}{N} \sum_{i=1}^N q_i(\cdot)$. The design of q_S follows the Lloyd–Max algorithm like in Section III-C, and the disassembly algorithm is introduced in [5], [6].

IV. BENEFITS

As an example, let us consider the following measurement model for $N = 3$ DMs:

$$Y_i = s_m + W_i, \quad i = 1, \dots, N, \quad m \in \{0, 1\}, \quad (6)$$

where $s_0 = 0$ when $H = h_0$, $s_1 = 1$ when $H = h_1$, and W_i are iid zero-mean Gaussian random variable with unit variance. The prior probability p_0 is uniformly distributed within $[0, 1]$. The Bayes costs of the decision are $c_{10} = c_{01} = 1$. Local decisions are fused by MAJORITY.

Fig. 2 depicts the performance of the decision making systems as p_0 is varied with $K = 3$ levels for quantizers. Performance is shown as excess Bayes risk with respect to the baseline system using (4). Comparing the three kinds of quantizers for prior probabilities, which are shown in Fig. 3, diverse quantizers yield smaller Bayes risk than identical quantizers. Also, the diverse quantizers are more useful when the decision makers collaborate.

Fig. 4 shows the mean Bayes risk depending on K , the number of quantizer levels. The difference between * and + or the difference between \square and \circ shows the performance loss due to the quantization of observations and the MAJORITY fusion rule. In Fig. 5, the curves depict the performance loss induced by identical quantization and diverse quantization. It also shows that collaboration improves the performance by the difference between \times and \circ .

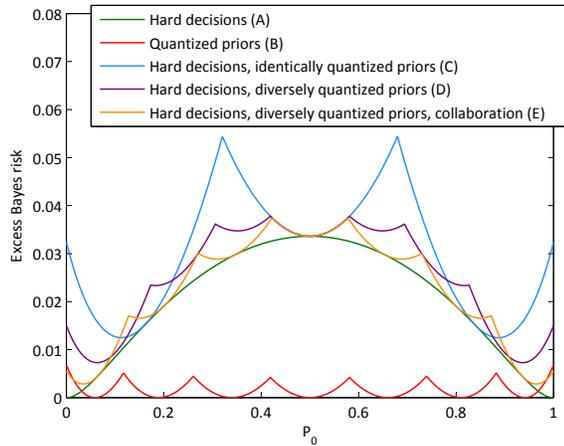


Fig. 2. Excess Bayes risk (compared to the use of perfect observations and prior probabilities in (4)) for 3-level quantization.

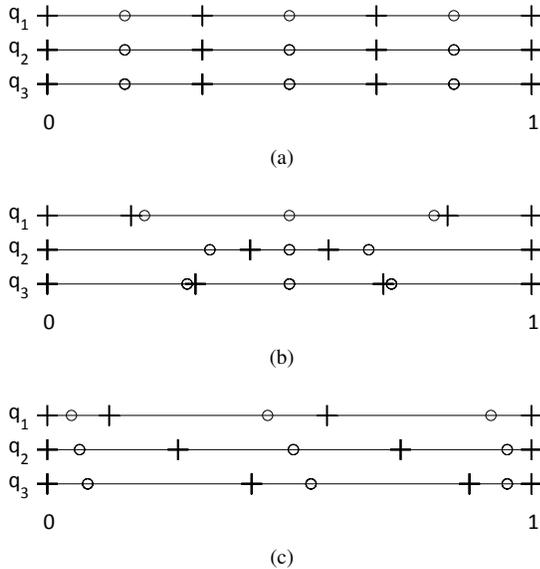


Fig. 3. Optimal quantizers (cell endpoints as '+'s and representation points as 'o's). (a) Identical quantization (Section III-C). (b) Diverse quantization (Section III-D). (c) Diverse quantization and collaboration (Section III-E).

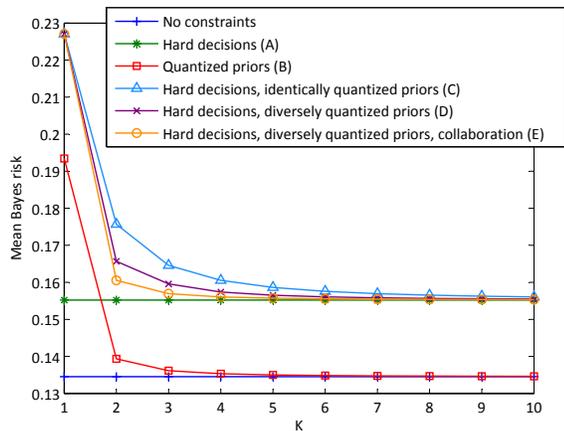


Fig. 4. Mean Bayes risk.

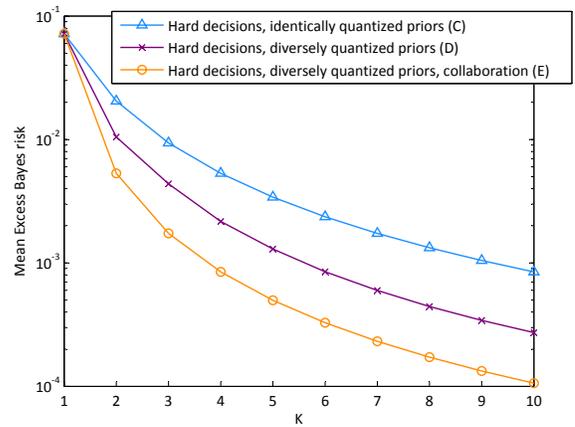


Fig. 5. Mean Excess Bayes risk (compared to the hard-decision-but-perfect- p_0 model (Section III-A)).

V. CONCLUSION

We have discussed the effect of quantization of prior probabilities on hypothesis testing. To minimize the negative influence due to the quantization, we have defined the mean Bayes risk error as the optimization criterion for prior-probability quantizers. Then we have compared the identical quantization to the diverse quantization.

Diverse quantization is the better option than identical quantization. Diverse quantizers require more computational cost to be designed, but once they are designed, they keep yielding better decisions without any additional cost.

Collaboration through sharing of quantizers among DMs is a way to amplify the benefit of diverse quantization. Collaboration requires communication among DMs to share their quantized values of prior probability at each decision making instance, but it minimizes the effect of quantization and yields the smallest mean Bayes risk.

REFERENCES

- [1] 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.
- [2] G. A. Miller, "The magical number seven, plus or minus two," *Psych. Rev.*, vol. 63, no. 2, pp. 81–97, Mar. 1956.
- [3] R. Fryer and M. O. Jackson, "A categorical model of cognition and biased decision making," *B.E. J. Theor. Econ.*, vol. 8, no. 1, Feb. 2008.
- [4] K. R. Varshney and L. R. Varshney, "Multilevel minimax hypothesis testing," in *Proc. IEEE/SP Workshop Stat. Signal Process.*, Nice, France, Jun. 2011, pp. 109–112.
- [5] J. B. Rhim, L. R. Varshney, and V. K. Goyal, "Collaboration in distributed hypothesis testing with quantized prior probabilities," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2011, pp. 303–312.
- [6] —, "Quantization of prior probabilities for collaborative distributed hypothesis testing," *arXiv:1109.2567*, 2011.
- [7] —, "Conflict in distributed hypothesis testing with quantized prior probabilities," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2011, pp. 313–322.
- [8] —, "Distributed decision making by categorically-thinking agents," in *Proc. NIPS 2011 Workshop on Decision Making with Multiple Imperfect Decision Makers*, Sierra Nevada, Spain, Dec. 2011.
- [9] V. Misra, V. K. Goyal, and L. R. Varshney, "Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5298–5325, Aug. 2011.
- [10] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.