Quantization of Prior Probabilities in Bayesian Group Decision-Making

by

Joong Bum Rhim

B.S. Electrical Engineering
KAIST, 2008

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author ....
Department of Electrical Engineering and Computer Science
August 30, 2010

Certified by ....
Vivek K Goyal
Esther and Harold Edgerton Career Development
Associate Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Quantization of Prior Probabilities in Bayesian Group Decision-Making

by

Joong Bum Rhim

Submitted to the Department of Electrical Engineering and Computer Science
on August 30, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

In Bayesian hypothesis testing, a decision is made based on a prior probability distribution over the hypotheses, an observation with a known conditional distribution given the true hypothesis, and an assignment of costs to different types of errors. In a setting with multiple agents and the principle of "one person, one vote", the decisions of agents are typically combined by the majority rule. This thesis considers collections of group hypothesis testing problems over which the prior itself varies. Motivated by constraints on memory or computational resources of the agents, quantization of the prior probabilities is introduced, leading to novel analysis and design problems.

Two hypotheses and three agents are sufficient to reveal various intricacies of the setting. This could arise with a team of three referees deciding by majority rule on whether a foul was committed. The referees face a collection of problems with different prior probabilities, varying by player. This scenario illustrates that even as all referees share the goal of making correct foul calls, opinions on the relative importance of missed detections and false alarms can vary. Whether cost functions are identical and whether referees use identical quantizers create variants of the problem.

When referees are identical in both their cost functions and their quantizers for the prior probabilities, it is optimal for the referees to use the same decision rules. The homogeneity of the referees simplifies the problem to an equivalent single-referee problem with a lower-variance effective noise. Then the quantizer optimization problem is reduced to a problem previously solved by Varshney and Varshney (2008). Centroid and nearest-neighbor conditions that are necessary for quantizer optimality are provided.

On the contrary, the problem becomes complicated when variations in cost functions or quantizers are allowed. In this case, decision-making and quantization problems create strategic form games; the decision-making game does always have a Nash equilibrium. The analysis shows that conflict between referees, in the form of variation in cost functions, makes overall team performance worse. Two ways to optimize quantizers are introduced and compared to each other.

In the setting that referees purely collaborate, in the form of having equal cost functions, the effect of variations between their quantizers is analyzed. It is shown that the referees have incentive to use different quantizers rather than identical quantizers even though their cost functions are identical. In conclusion, a diverse team with a common goal performs best.

Thesis Supervisor: Vivek K Goyal
Title: Esther and Harold Edgerton Career Development
Associate Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to thank my advisor Vivek Goyal for his every support. Our discussions helped me organize and develop my ideas. He always encouraged me so that I could go through difficulties.

I also thank the members of the STIR group; this group is small, but every member is close to each other and helps me easily settle down in this group. Especially, it was Lav Varshney who came up with my thesis topic and gave me useful feedback. John Sun showed me surprising intuitions for my work. Daniel Weller, Ahmed Kirmani, and Michael Snella provided comments that improved my work.

I thank my other mentors, Alan Oppenheim, who was always concerned about me and gave me counsel so that I could become comfortable here, and Muriel Médard, who I first learned how to research from.

I especially appreciate my family and friends' support. My parents and sister gave me strength to keep moving forward. Keun Sup Shim, Jae-Byum Chang, Donghyun Jin are my friends who have been known for twelve years and came to MIT with me. They helped me in personal issues instead of my family and we shared every happiness and difficulty. I am lucky to have them with me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Decision theory has been developed in various ways because there are numerous different situations of decision-making in practice. One of the simplest situations is that a single decision-maker chooses between two alternatives, such as an alarm detecting fire by monitoring heat and smoke or a referee judging a player's tackle to be fair or foul. They are called binary hypothesis testing problems, which have been widely studied in wireless communications.

The simplicity of binary hypothesis testing problems helps us understand fundamentals of decision-making: there is a beautifully simple decision rule and a useful operating characteristic that the probability of false alarm is convex in that of missed detection [1].

Another situation of decision-making is a distributed hypothesis testing problem. The interest in the problem originated with the requirements of military surveillance systems with distributed sensors [2, 3]. One type of distributed detection system is what consists of several decision-makers who vote for one candidate hypothesis and a fusion center that uses the majority rule. We are interested in this type of distributed detection system because it is widely used in real life as democratic decision-making. In such a system, a decision-maker may compete or cooperate with others in order to induce the global decision to be what he wants when all decision-makers have different preferences. In traditional distributed detection systems, however, detectors or decision-makers are supposed to share the same cost function and cooperate with others [4,5]. We are not aware of any previous work that has looked at the distributed detection system in which each detector has its own cost function and one vote.

We discuss the criterion of optimality of decision rules in this system, and compute equilibria of decision rules by defining a proper strategic form game.

In addition, practical decision-makers may have physical limitations. Let us consider decision-makers that perform a series of hypothesis testing on a population of objects. The decision-makers need to know exact prior probability of each object for each Bayesian hypothesis testing. However, they may have limited memory or computational capability. We assume that they use quantized versions of prior probabilities because of the limitation, which is a feasible assumption. Then each decision-maker needs an optimal quantizer of prior probabilities to minimize the error due to the quantization. Designing an optimal quantizer for prior probabilities for a single decision-maker using the Lloyd-Max algorithm has been studied in [6]. In group decision-making, the Lloyd-Max algorithm cannot be applied because of dependency among decision makers' quantization rules. We discuss difficulty in optimizing quantization rules and how the decision-makers' preferences affect optimal quantization rules.

When a distributed detection system consists of decision-makers that have the same preference, they can cooperate to make the best decision. In this case, game-theoretic issues do not occur in optimizing decision and quantization rules. Therefore, the analysis of quantization is similar to that in [6] except a diversity issue. Especially, decision-makers may be able to perform more accurate hypothesis testing by using different quantizers for prior probabilities than using the same quantizers. It is because the diverse quantizers can categorize the objects into more detailed groups than the same quantizers can. We investigate the extent of the benefit of diversity in quantizers and how to design the optimal diverse quantization rules.

1.1 Thesis Outline

This thesis explores the group decision-making by imperfect referees. The term referee is motivated by applications in sports. We consider a group of three referees, which is the smallest number of referees without controversy in majority vote. They observe the same object that has two possible states, but their observations are distorted by independent and identically distributed additive Gaussian noises. Each referee makes a local binary decision and all local decisions are sent to a fusion center, where a global decision is determined as the majority of the local decisions. The referees follow Bayesian hypothesis testing rule, i.e., each referee attempts to make decisions that minimize his Bayes risk. Since the referees are supposed to detect an object with arbitrary prior probability, they need to know the prior probability of what they observe in order to make such decisions.

Due to their limited processing capability, however, they can distinguish an object as one of K categories. In other words, an object belongs to one of K categories according to its prior probability and a referee's classification rule, and the referee recognizes the object has the prior probability that represents the category it belongs to. Consequently, due to his limitation, the referee uses the quantized prior probability in Bayesian hypothesis testing.

This work deals with two main issues: decision rules and quantization rules. Each referee wants the final decision to minimize his Bayes risk. According to others' decisions, however, the final decision may become different from what he wants, especially when the referees have different local cost functions. The discord is the reason that each referee encounters conflicts of interest and has to consider others' cost functions as well as his own cost function in order to determine an optimal decision rule. In addition, we assume that each referee has a proper quantization rule for prior probabilities according to his cost function. Even though it is assumed that all referees know the same prior probabilities in traditional distributed detection problems [5], the referees in our problem no longer have the same prior probabilities if they have different quantization schemes. Since different quantization schemes come from the referees' different cost functions, they face conflicts of interest not only because of different cost functions but also due to differently quantized prior probabilities. Therefore, we investigate methods for referees to develop optimal decision rules and quantization rules so that each referee can minimize his Bayes risk while he considers others' decision and quantization rules.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Cost function</td></tr>
<tr><td>Same</td><td>Different</td></tr>
<tr><td rowspan="2">Quantizer</td><td>Same</td><td>Identical referees
Chapter 3</td><td rowspan="2">Conflicting referees
Chapter 4
(Section 4.2)</td></tr>
<tr><td>Different</td><td>Collaborating referees
Chapter 4
(Section 4.3)</td></tr>
</table>

Table 1.1. Classification of teams of referees.

In this thesis, we discuss behaviors of each referee in a group Bayesian decision-making system. We propose several strategies by which a referee can conflict or cooperate with others in order to minimize his Bayes risk. In addition, we analyze how behaviors of other referees affect the design of an individual's quantization rule as well as his decision rule.

The rest of the thesis is organized as follows. Chapter 2 covers some relevant background on decision theory, quantization theory, and game theory.

Chapter 3 looks at the group decision-making when referees are identical. The term *identical* means that the referees have the same preference (same Bayes costs) and the same quantizers, Table 1.1. Since each referee's Bayes risk coincides with every other referee's Bayes risk, there are no conflicts of interest or game-theoretic issues. We derive optimal decision rules and quantization rules for the referees. Their performance is compared to a single referee's performance and it is discussed how they have an advantage over the single referee.

Chapter 4 discusses a more general case when referees are not identical. Conflicts of interest among the referees occur in this case because each referee's Bayes risk has a different formulation from the others'. Thus, they should compete with one another in order to achieve a preferable global decision. We analyze how the competition has an effect on optimal decision and quantization rules from game-theoretic point of view. In addition, we look at the referees who may have different quantization rules but collaborate for a common goal. This case is similar to the identical-referee case except that the referees can take advantage of diverse quantizers. We analyze how they benefit by diversity of quantization rules and present an algorithm to design optimal quantization schemes.

Finally, Chapter 5 summarizes the contributions of this thesis.

Chapter 2

Background

This thesis lies in the intersection of decision theory, quantization theory, and game theory. This chapter gives an overview of relevant concepts and terminology from these fields. Also, the results of a related previous work [6] are summarized.

■ 2.1 Bayesian Hypothesis Testing

Hypothesis testing is making a decision among a set of discrete possibilities. Hypothesis testing is used to a variety of fields including radar detection, speech recognition [7], and clinical trial investigation [8] as well as digital communication. Hypothesis testing is based on observations, which are distorted or incomplete due to noise, obstacles, or limitation of equipments. The goal in hypothesis testing is to make the best decision with imperfect observed data.

■ 2.1.1 Description of Model

The basic model in hypothesis testing problems is shown in Figure 2-1, where H denotes a hypothesis and Y denotes observed data. A hypothesis is a discrete random variable drawn from a set of M states, $\mathcal{H} = \{h_0, \dots, h_{M-1}\}$, which is called an alphabet. We observe a set of data Y which is jointly distributed with the hypothesis H. In general, we are given two kinds of information. First, the hypothesis has prior probability $P_H(h_m)$ with which the hypothesis is in state h_m, where $\sum_{m=0}^{M-1} P_H(h_m) = 1$ and $P_H(h_m) \geq 0$, for $m = 0, 1, \dots, M-1$. Second, the observation Y is characterized

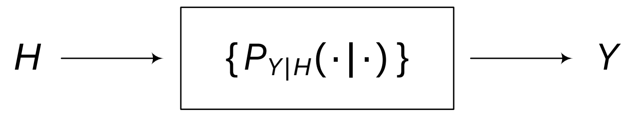


Figure 2-1. The observation model in Bayesian hypothesis testing problems.

by conditional probability distributions

$$P_{Y|H}(\cdot|h_m), \qquad m = 0, 1, \ldots, M-1$$

under each hypothesis h_m. The system between H and Y in Figure 2-1 can be fully described by the set of transition probabilities $\{P_{Y|H}(\cdot|h_0), P_{Y|H}(\cdot|h_1), \ldots, P_{Y|H}(\cdot|h_{M-1})\}$, and the system has the special name "channel" in communication.

The two kinds of information – the prior probabilities and transition probabilities – are sufficient to characterize the observed data Y. The observed data have the density function

$$P_Y(y) = \sum_{m=0}^{M-1} P_{H,Y}(h_m, y) = \sum_{m=0}^{M-1} P_{Y|H}(y|h_m) P_H(h_m).$$

Then it is possible to update our belief (or distribution) of the hypothesis based on the observed data by using Bayes' theorem. The likelihood of each hypothesis h_m when we observe $Y = y$ is given by

$$P_{H|Y}(h_m|y) = \frac{P_{Y|H}(y|h_m) P_H(h_m)}{P_Y(y)} = \frac{P_{Y|H}(y|h_m) P_H(h_m)}{\sum_{m=0}^{M-1} P_{Y|H}(y|h_m) P_H(h_m)}.$$

This probability is called posterior probability, which means that it is computed after $Y = y$ is observed. Hypothesis testing using Bayes' theorem is called Bayesian hypothesis testing.

■ 2.1.2 Criterion of Decision Rule

In general, it is impossible to recover the correct hypothesis every time because of incompleteness of observed data. Hence, the goal in Bayesian hypothesis testing problems is to make the best decision (and find the best decision rule) about the hypothesis which incurs minimum cost. We use $C(h_i, h_j)$ to denote the cost of deciding that the hypothesis is h_i when the correct hypothesis is $H = h_j$, and we call $C(\cdot, \cdot) : \mathcal{H}^2 \mapsto \mathbb{R}^+$ a cost function.

The solution of a hypothesis testing problem is given by a decision rule, which is a function $\hat{H} : \mathbb{R} \mapsto \mathcal{H}$ that maps each possible observation to one of the hypotheses. The criterion for performance of a decision rule $\hat{H}$ is the expected cost, which is referred to as Bayes risk and is computed by

$$\rho = \mathbb{E}[C(H, \hat{H}(Y))] = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} C(h_n, h_m)\mathbb{P}[\hat{H} = h_n | H = h_m] P_H(h_m).$$

In other words, the best decision rule in a Bayesian hypothesis testing problem is the decision rule that minimizes Bayes risk.

2.1.3 Binary Hypothesis

In a binary hypothesis testing problem, the hypothesis is in one of two values $\{h_0, h_1\}$, whose prior probabilities are $p_0 = \mathbb{P}[H = h_0]$ and $p_1 = \mathbb{P}[H = h_1] = 1 - p_0$. Our observation Y is a random variable whose distribution conditioned on $H = h_0$ or $H = h_1$ is respectively given by $P_{Y|H}(y|h_0)$ or $P_{Y|H}(y|h_1)$. Then a decision rule $\hat{H}(\cdot)$ leads to Bayes risk

$$\begin{aligned}
\rho &= c_{00} p_0 \mathbb{P}[\hat{H} = h_0 | H = h_0] + c_{10} p_0 \mathbb{P}[\hat{H} = h_1 | H = h_0] \\
&\quad + c_{11} p_1 \mathbb{P}[\hat{H} = h_1 | H = h_1] + c_{01} p_1 \mathbb{P}[\hat{H} = h_0 | H = h_1] \\
&= \int_{\mathcal{Y}_1} p_0 (c_{10} - c_{00}) P_{Y|H}(y|h_0)\, dy + \int_{\mathcal{Y}_0} p_1 (c_{01} - c_{11}) P_{Y|H}(y|h_1)\, dy + c_{00} p_0 + c_{11} p_1,
\end{aligned}$$

where $c_{ij} \triangleq C(h_i, h_j)$ and $\mathcal{Y}_i \triangleq \{y : \hat{H}(y) = h_i\}$. Valid cost functions have to satisfy

$$c_{ij} > c_{jj}, \qquad \forall i \neq j, \qquad \forall j = 1, \ldots, |\mathcal{H}|,$$

because it is reasonable that incorrect decisions are more costly than correct decisions. The decision rule $\hat{H}(\cdot)$ that minimizes ρ has a form of a likelihood ratio test:

$$\frac{P_{Y|H}(y|h_1)}{P_{Y|H}(y|h_0)} \underset{\hat{H}(y)=h_0}{\overset{\hat{H}(y)=h_1}{\gtrless}} \frac{p_0(c_{10}-c_{00})}{p_1(c_{01}-c_{11})}. \tag{2.1}$$

In the remainder of this paper, we assume that $c_{00} = c_{11} = 0$, which simplifies the likelihood ratio test to

$$\frac{P_{Y|H}(y|h_1)}{P_{Y|H}(y|h_0)} \underset{\hat{H}(y)=h_0}{\overset{\hat{H}(y)=h_1}{\gtrless}} \frac{p_0 c_{10}}{p_1 c_{01}}. \tag{2.2}$$

There are two kinds of errors in a binary hypothesis testing:

$$P_{e1} = \mathbb{P}[\hat{H} = h_1 | H = h_0] = \int_{\mathcal{Y}_1} P_{Y|H}(y|h_0)\, dy, \tag{2.3}$$

$$P_{e2} = \mathbb{P}[\hat{H} = h_0 | H = h_1] = \int_{\mathcal{Y}_0} P_{Y|H}(y|h_1)\, dy. \tag{2.4}$$

P_{e1} is called the probability of error of the first kind or probability of a false alarm, and P_{e2} is called the probability of error of the second kind or probability of a missed detection. One characteristic of the probabilities of errors is that P_{e2} is convex in P_{e1} if decision rule (2.2) that minimizes Bayes risk is used [1].

■ 2.2 Quantization

Quantization is the process of mapping from a continuous range of values to a set of discrete values. A quantizer consists of a set of regions $\mathcal{R} = \{\mathcal{R}_k; k \in \mathcal{K}\}$ and a set of representation points $\mathcal{C} = \{y_k; k \in \mathcal{K}\}$, where $\mathcal{K}$ is a countable index set [9], so that the quantizer is defined by

$$q(x) = y_k \text{ for } x \in \mathcal{R}_k.$$

The function $q(x)$ is called the quantization rule. A simple example of quantization is rounding off, which is defined by $\mathcal{R}_k = [k - 0.5, k + 0.5)$ and $y_k = k$ with $\mathcal{K} = \mathbb{Z}$, or $q(x) = \lfloor x + 0.5 \rfloor$.

The quality of a quantizer can be measured by comparing the resulting reproduction to its original value. Having a distortion measure $d(x, \hat{x})$ that specifies the cost or distortion of recovering x as $\hat{x}$, we can measure the quality of a quantization scheme by the average distortion. When the data is considered as a random variable whose probability density function is $f_X(x)$, the average distortion becomes

$$D(q) = \mathbb{E}[d(X, q(X))] = \sum_k \int_{\mathcal{R}_k} d(x, y_k) f_X(x)\, dx. \tag{2.5}$$

Having smaller average distortion means higher quality. One of the most common distortion measures is squared error $d(x, \hat{x}) = |x - \hat{x}|^2$ and $D(q)$ is then called the mean squared error (MSE).

Quantization is coding an input x to one of K binary codewords, where $K = |\mathcal{K}|$. Then since it requires $\log_2 K$ bits to describe each codeword, the rate of this quantization scheme is defined as $\log_2 K$ bits per sample. A quantizer with fixed-length codewords is said to have *fixed rate*. The goal of quantization is to encode data with as few bits as possible and to recover them with as small average distortion as possible. Thus, there is a trade-off between average distortion and rate.

A quantizer is called a vector quantizer if the dimension of its input is more than one, and it is called a scalar quantizer if the dimension is equal to one. In addition, a memoryless quantizer does not change sets of regions and representation points depending on the past. Fixed-rate memoryless scalar quantization is used in this work.

2.2.1 Optimality Conditions

A fixed-rate memoryless scalar quantizer consists of two components: a lossy encoder $\alpha : A \mapsto \mathcal{K}$, where A is an alphabet of input symbols, and a reproduction decoder $\beta : \mathcal{K} \mapsto \mathcal{C}$. Lloyd's conditions on (α, β) in order for the quantizer to be optimal are as follows: if any component of the code (α, β) is fixed, then the other component must have a specific form, which is described below.

- For a fixed lossy encoder α, the optimal decoder β is given by

  $$\beta(k) = \arg\min_{y} \mathbb{E}[d(X, y)|\alpha(X) = k].$$

  In other words, $\beta(k)$ is given by the value minimizing the expectation of the distortion between the value and the input x conditioned on that the encoder generates k for x. The values $\{\beta(k); k \in \mathcal{K}\}$ are called centroids.

- For a fixed reproduction decoder β, the optimal lossy encoder is a minimum-distortion (or nearest neighbor) encoder

  $$\alpha(x) = \arg\min_{k \in \mathcal{K}} d(x, \beta(k)).$$

The partition that satisfies both conditions is called a *Voronoi partition*.

2.2.2 Functional Quantization

In general, limitation of storage space requires quantization of data. The quantization may occur in the middle of a whole process and the quantized data can be used in remaining process. In this work, for example, a prior probability is quantized and the quantized value is used in Bayesian hypothesis testing. Then it is reasonable that we are concerned about the values we will get after the whole process rather than those

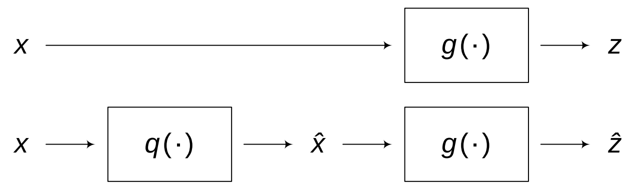


Figure 2-2. Fixed-rate scalar functional quantization.

right after quantization. Figure 2-2 depicts two systems: one without a quantizer and one with a quantizer in front of it. Since the latter system uses quantized value $\hat{x}$, the result becomes $\hat{z}$ instead of the desired value z.

Even if we use the optimal quantizer that minimizes the average of distortion $d(x, \hat{x})$, it does not guarantee that we can achieve the minimum distortion between z and $\hat{z}$. Therefore, the quantization rule $q(\cdot)$ should be determined so that it minimizes the average distortion

$$D(q) = \mathbb{E}_X[d(z, \hat{z})] = \sum_k \int_{\mathcal{R}_k} d(g(x), g(y_k)) f_X(x)\, dx,$$

which is different from (2.5).

Quantization of prior probabilities

Quantization of prior probabilities for Bayesian hypothesis testing is introduced in [6]. There is a population of objects, and each object has its own prior probability drawn from a density function $f_{P_0}(p_0)$. However, a referee has finite memory or limited information processing resources so that he can only work with at most K different prior probabilities. Thus, when he makes a decision about an arbitrary object, he maps its true prior probability to one of the K available values and then performs the Bayesian hypothesis test. In other words, he quantizes the prior probability before performs hypothesis test.

The objective of the paper is to find an optimal K-point quantizer $v_K(\cdot)$ for prior probabilities. Since a Bayesian referee pursues the minimum Bayes risk, $v_K(\cdot)$ is a functional quantizer that should minimize Bayes risk error $d(p_0, v_K(p_0))$ due to the quantization. Mean Bayes risk error

$$\int_0^1 d(p_0, v_K(p_0)) f_{P_0}(p_0)\, dp_0$$

is defined as the criterion for optimality of the quantizer, and the nearest neighbor condition and the centroid condition for optimal quantizers are derived. It is shown that if $f_{P_0}(p_0)$ is positive and continuous in $(0, 1)$ and $\int_0^1 d(p_0, a) f_{P_0}(p_0)\, dp_0$ is finite for all $a \in [0, 1]$, then the Lloyd-Max algorithm alternating the centroid and nearest neighbor conditions iteratively will converge to an optimal quantizer. High-rate approximation of distortion-rate function is also obtained in the paper.

The paper applies its results to human decision-making and derives an interesting conclusion about discrimination against minority. Consider two populations – a majority and a minority populations – and extend the definition of mean Bayes risk error to

$$D^{(2)} = \frac{M}{M+m} \mathbb{E}[d(P_0, v_{K_M}(P_0))] + \frac{m}{M+m} \mathbb{E}[d(P_0, v_{K_m}(P_0))],$$

where M is the number of the majority population, m is the number of the minority population, K_M is the number of points in the quantizer for the majority, and K_m is the number of points in the quantizer for the minority. If a referee has the total quota of representation points $K_t = K_M + K_m$, then his optimal allocation of the points will result in $K_M > K_m$ in order to minimize $D^{(2)}$. Therefore, even though the referee does not intend to, he will make more accurate decisions on the majority population than on the minority population.

However, the accuracy of the decisions is not enough to explain the discrimination that the referee calls more fouls on minority than on majority. The paper defines discrimination quantity

$$\Delta = \mathbb{E}\left[\mathbb{P}[\hat{H}_{K_m} = h_1] - \mathbb{P}[\hat{H}_{K_M} = h_1] \right],$$

where $\mathbb{P}[\hat{H}_K = h_1]$ is the probability of calling a foul when the referee uses a quantizer

v_K:

$$\mathbb{P}[\hat{H}_K = h_1] = p_0 P_{e1}(v_K(p_0)) + (1 - p_0)(1 - P_{e2}(v_K(p_0))).$$

The discrimination quantity may be written as

$$\Delta = \mathbb{E}\left[p_0 P_{e1}(v_{K_m}(p_0)) - (1 - p_0)P_{e2}(v_{K_m}(p_0))\right] - \mathbb{E}\left[p_0 P_{e1}(v_{K_M}(p_0)) - (1 - p_0)P_{e2}(v_{K_M}(p_0))\right].$$

If this discrimination quantity Δ is greater than zero, then the referee calls more fouls on minority; otherwise, he calls more fouls on majority. It is found out that Δ depends on Bayes costs c_{10} and c_{01} and the distribution $f_{P_0}(p_0)$ as well as the quantizers. For example, the discrimination against minority occurs if $c_{01} > c_{10}$ for a uniform prior probability. Analyzing various data about decisions by police, human resource professionals, and National Basketball Association referees, the paper concludes that all of them follow what is called the precautionary principle.

2.3 Game Theory

Game theory provides useful mathematical methods to analyze a system that consists of multiple agents whose actions have effect on the entire system. Game theory has developed methodologies that understand how an individual makes a decision when the individual's outcome depends on others' decisions. Game theory was introduced by von Neumann and Morgenstern in 1944 [10]. Initial studies considered two players' competitions in zero-sum games in which a player's gain or loss is exactly balanced by the loss or gain of the other player. Game theory has been expanded since 1950s so that it has been applied to political science [11], biology [12], computer science as well as economics [13].

The basic assumption in game theory is that all agents are rational in the sense that each agent attempts to maximize his payoff. Technically speaking, each agent is supposed to know his set of strategies and be capable of thinking through all possible outcomes. He chooses the option that gives him higher utility or payoff by computing expected payoff over unknown parameters and solving an optimization problem. Game theory explains many equilibrium concepts of players' strategies based on the rational decision-making process.

2.3.1 Strategic Form Game

A strategic or normal form game is a model of interactive decision-making in which all agents simultaneously make their decisions while they do not have any information about others' decisions. The game is defined by $(\mathcal{I}, (S_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}})$: the finite set of players $i \in \mathcal{I} = \{1, \ldots, I\}$, the set of available pure strategies (or actions) $s_i \in S_i$ for each player i, and the payoff (or utility) function $u_i : \prod_{i\in\mathcal{I}} S_i \mapsto \mathbb{R}$ for each player i. In addition, $S_{-i} = \prod_{j\in\mathcal{I}, j\neq i} S_i$ denotes the set of strategy profiles of all players other than player i, which are referred to as player i's opponents. The vector of strategies of player i's opponents is denoted by $s_{-i} \in S_{-i}$, and $(s_i, s_{-i}) \in S = \prod_{i\in\mathcal{I}} S_i$ is called strategy profile or outcome. The payoff of player i depends on (s_i, s_{-i}), and payoff functions describe the influence among players.

A *mixed strategy* is a probability distribution over pure strategies, which represents a player's probability of playing each pure strategy. The payoffs to a profile of mixed strategies are the expected values of the corresponding pure strategy payoffs. A game is said to be a *finite* game if the cardinality of S is finite; otherwise, it is an *infinite* game.

It is assumed in a strategic form game that all players are rational and have full knowledge about the structure of the game, i.e., $(\mathcal{I}, (S_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}})$. Player i cares not only about his strategy but also strategies taken by his opponents. The word opponents does not mean they attempt to beat player i. Rather, player i tries to maximizes his payoff, which may help or hurt his opponents. The central objective of game theory is to find equilibria of strategy profiles.

2.3.2 Dominant or Dominated Strategy

An easy way of anticipating which strategy a player would or would not choose is to find a strategy that always leads him to the largest or smallest payoff. A strategy $s_i \in S_i$ is *dominant* if for $\forall s_i' \in S_i$ and $\forall s_{-i} \in S_{-i}$,

$$u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}).$$

On the contrary, a strategy $s_i \in S_i$ is *strictly dominated* (by strategy s_i') if there exists some $s_i' \in S_i$ such that

$$u_i(s_i', s_{-i}) > u_i(s_i, s_{-i}), \forall s_{-i} \in S_{-i}.$$

A dominant or dominated strategy is a strong concept of decision-making because choosing or discarding the strategy does not depend on other players' choices.

If player i has a dominant strategy s_i, then it is reasonable to think that the player will choose s_i no matter how other players play. A dominant strategy equilibrium is a strategy profile $s^* = (s_1^*, \ldots, s_I^*)$ such that s_i^* is a dominant strategy for each player i, $\forall I \in \mathcal{I}$.

On the other hand, if player i has a strictly dominated strategy s_i, the player will never choose s_i. Thus, s_i can be discarded from strategy space S_i of player i and payoff function be redefined. Iterated elimination of strictly dominated strategies denotes the algorithm to remove strictly dominated strategy repeatedly and to save feasible strategy profiles.

Algorithm I: Iterative elimination of strictly dominated strategies

---

1) Define $S_i^0 = S_i, \forall i \in \mathcal{I}$.

2) Iterative the following process: at n-th iteration, for $i \in \mathcal{I}$,

   (i) $\bar{S}_i^n = \{s_i \in S_i^{n-1} : \exists s_i' \in S_i^{n-1} s.t. u_i(s_i', s_{-i}) > u_i(s_i, s_{-i}), \forall s_{-i} \in S_{-i}^{n-1}\}$.

   (ii) $S_i^n = S_i^{n-1} / \bar{S}_i^n$.

   (iii) $S_{-i}^n = \prod_{j \neq i} S_j^n$.

3) Define $S_i^\infty = \bigcap_{n=0}^{\infty} S_i^n$.

A problem is said to be solvable by iterative (strict) dominance if, for each player $i \in \mathcal{I}$, S_i^∞ is a singleton. S_i^∞ is nonempty and contains at least one pure strategy for each player i [14, 15].

■ 2.3.3 Nash Equilibrium

Even though iterated elimination of strictly dominated strategy is a very intuitive way to find an equilibrium, many games are not solvable by iterative strict dominance. Thus, we need a more robust equilibrium notion than dominant or dominated strategy.

A strategy profile $s^* = (s_1^*, \ldots, s_I^*)$ is a *Nash equilibrium* of a strategic game $(\mathcal{I}, \{S_i\}_{i=1}^I, \{u_i\}_{i=1}^I)$ if, for every player i,

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \forall s_i \in S_i.$$

In addition, a Nash equilibrium s_i^* is strict if, for every player i,

$$u_i(s_i^*, s_{-i}^*) > u_i(s_i, s_{-i}^*), \forall s_i \neq s_i^*.$$

Nash equilibrium is a reasonable equilibrium notion because no player has incentive to change his strategy in a Nash equilibrium.

There are several theorems about the existence of Nash equilibria:

Theorem 2.1 ([16]). *Every finite strategic-form game has a mixed strategy Nash equilibrium.*

Theorem 2.2 ([17]). *An infinite game has a mixed strategy Nash equilibrium if*

- *its strategy spaces S_i are nonempty compact sets and*
- *its payoff functions $u_i(s_i, s_{-i})$ are continuous in s.*

Theorem 2.3 ([17–19]). *An infinite game has a pure strategy Nash equilibrium if*

- *its strategy spaces S_i are nonempty compact convex sets,*
- *its payoff functions $u_i(s_i, s_{-i})$ are continuous in s_{-i}, and*
- *$u_i(s_i, s_{-i})$ are quasi-concave in s_i.*

A Nash equilibrium is a meaningful prediction of how the game will be played in the sense that if all players predict that a Nash equilibrium will occur, then it is the best choice for them to play it. Therefore, a Nash equilibrium has the property that players can predict it and predict that their opponents can predict it. This is why many applications of game theory, including this thesis, pay attention to Nash equilibria.

Chapter 3

Identical Referees

In general distributed detection and data fusion problems, one cost function is defined and shared by detectors. These problems do not contain game-theoretic issues such as conflicts among detectors. In the language of this thesis, the three referees share one cost function. This means that they have identical preferences in terms of how important it is to avoid missed detections and false alarms. We analyze how they make local decisions and quantize prior probabilities optimally in terms of Bayes risk. In this chapter, we additionally assume that if prior probabilities are quantized, the referees quantize in the same way; the referees are thus called identical.

The operating characteristic of the team of identical referees shows that using the same decision rules is optimal for them. A single-referee model equivalent to a three-referee model is introduced in order for us to compare the performance of a three-referee team to that of a single referee. The equivalent single-referee model is also useful to easily derive nearest neighbor and centroid conditions for optimal quantizers. Using the Lloyd-Max algorithm, we optimize quantization rules for several cases and show the results.

3.1 Problem Model

Figure 3-1 depicts the distributed detection and data fusion model under consideration. The object that referees want to detect is denoted by H. It has two possible states h_0 and h_1, whose prior probabilities are $p_0 = \mathbb{P}[H = h_0]$ and $p_1 = \mathbb{P}[H = h_1] = 1 - p_0$. It is assumed that $h_0, h_1 \in \mathbb{R}$ and $h_0 < h_1$. Referees observe Y_i, which are versions of H distorted by independent and identically distributed additive noises W_i

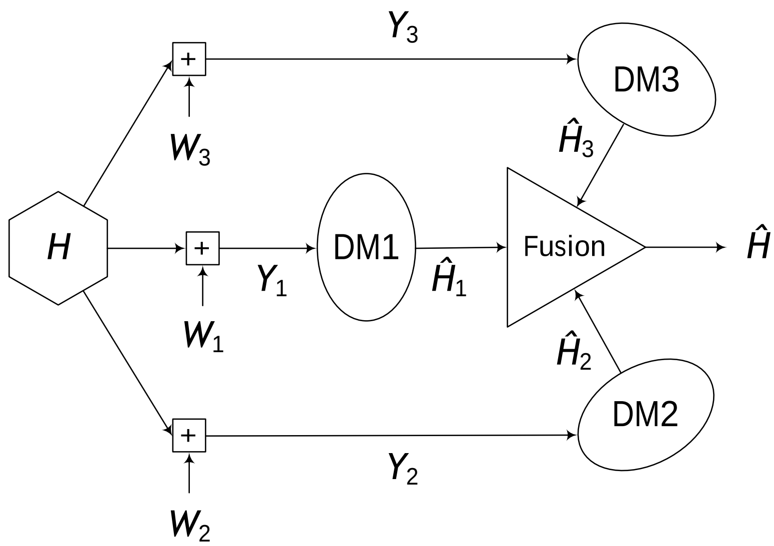


Figure 3-1. The distributed detection and fusion model explored in this work.

drawn from the Gaussian distribution with zero mean and variance σ^2. They make local decisions $\hat{H}_i(Y_i)$ according to their own decision rules $\hat{H}_i : \mathbb{R} \mapsto \{h_0, h_1\}$. Local decisions $\hat{H}_i$ are sent to a fusion center, where fusion of the decision obeys the majority rule. All referees have the same cost function $C(\hat{H}, H)$, which depends on a global decision rather than local decisions. Thus, global decisions matter to referees no matter what local decisions are. We assume that $C(h_i, h_i) = 0$ and $C(h_i, h_j) > 0$ for all $i = 0, 1$ and all $j \neq i$.

3.2 Decision Rule

We investigate the identical referees' Bayesian optimal decision rules. All referees determine decision rules that lead to the minimum Bayes risk. Because all referees have the same cost function, they have the same Bayes risk:

$$R = c_{10} p_0 \mathbb{P}[\hat{H} = h_1 | H = h_0] + c_{01}(1 - p_0)\mathbb{P}[\hat{H} = h_0 | H = h_1], \tag{3.1}$$

where $c_{ij} \triangleq C(h_i, h_j)$. Let $P_{e1}^{(i)} \triangleq \mathbb{P}[\hat{H}_i = h_1 | H = h_0]$ and $P_{E1} \triangleq \mathbb{P}[\hat{H} = h_1 | H = h_0]$. According to the majority fusion rule, $\hat{H} = h_1$ if at least two referees declare h_1, which means a false alarm occurs at the fusion center if at least two referees give false alarms. Thus, we can compute the probability of a global false alarm from the probabilities of local false alarms in an inclusion-exclusion manner:

$$P_{E1} = P_{e1}^{(1)} P_{e1}^{(2)} + P_{e1}^{(2)} P_{e1}^{(3)} + P_{e1}^{(3)} P_{e1}^{(1)} - 2P_{e1}^{(1)} P_{e1}^{(2)} P_{e1}^{(3)}. \tag{3.2}$$

Likewise,

$$P_{E2} = P_{e2}^{(1)} P_{e2}^{(2)} + P_{e2}^{(2)} P_{e2}^{(3)} + P_{e2}^{(3)} P_{e2}^{(1)} - 2P_{e2}^{(1)} P_{e2}^{(2)} P_{e2}^{(3)}, \tag{3.3}$$

where $P_{e2}^{(i)} \triangleq \mathbb{P}[\hat{H}_i = h_0 | H = h_1]$ and $P_{E2} \triangleq \mathbb{P}[\hat{H} = h_0 | H = h_1]$.

We can rewrite Bayes risk of referee 1 as follows:

$$\begin{aligned}
R &= c_{10} p_0 P_{E1} + c_{01}(1-p_0) P_{E2} \\
&= c_{10} p_0 (P_{e1}^{(2)} + P_{e1}^{(3)} - 2P_{e1}^{(2)} P_{e1}^{(3)}) P_{e1}^{(1)} + c_{10} p_0 P_{e1}^{(2)} P_{e1}^{(3)} \\
&\quad + c_{01}(1-p_0)(P_{e2}^{(2)} + P_{e2}^{(3)} - 2P_{e2}^{(2)} P_{e2}^{(3)}) P_{e2}^{(1)} + c_{01}(1-p_0) P_{e2}^{(2)} P_{e2}^{(3)} \\
&= c_{10} p_0 (P_{e1}^{(2)} + P_{e1}^{(3)} - 2P_{e1}^{(2)} P_{e1}^{(3)}) \int_{\mathcal{Y}_1^{(1)}} f_{Y_1|H}(y_1|h_0)\, dy_1 + c_{10} p_0 P_{e1}^{(2)} P_{e1}^{(3)} \\
&\quad + c_{01}(1-p_0)(P_{e2}^{(2)} + P_{e2}^{(3)} - 2P_{e2}^{(2)} P_{e2}^{(3)}) \int_{\mathcal{Y}_0^{(1)}} f_{Y_1|H}(y_1|h_1)\, dy_1 + c_{01}(1-p_0) P_{e2}^{(2)} P_{e2}^{(3)},
\end{aligned}$$

where $\mathcal{Y}_k^{(i)} \triangleq \{y_i : \hat{H}_i(y_i) = h_k\}$. In order to minimize Bayes risk, referee 1 should assign $\mathcal{Y}_0^{(1)}$ and $\mathcal{Y}_1^{(1)}$ such that $y_1 \in \mathcal{Y}_1^{(1)}$ if $c_{10} p_0 (P_{e1}^{(2)} + P_{e1}^{(3)} - 2P_{e1}^{(2)} P_{e1}^{(3)}) f_{Y_1|H}(y_1|h_0) < c_{01}(1-p_0)(P_{e2}^{(2)} + P_{e2}^{(3)} - 2P_{e2}^{(2)} P_{e2}^{(3)}) f_{Y_1|H}(y_1|h_1)$, and $y_1 \in \mathcal{Y}_0^{(1)}$ otherwise. Thus, his decision rule should be

$$\frac{f_{Y_1|H}(y_1|h_1)}{f_{Y_1|H}(y_1|h_0)} \underset{\hat{H}_1(y_1)=h_0}{\overset{\hat{H}_1(y_1)=h_1}{\gtrless}} \frac{c_{10} p_0 (P_{e1}^{(2)} + P_{e1}^{(3)} - 2P_{e1}^{(2)} P_{e1}^{(3)})}{c_{01}(1-p_0)(P_{e2}^{(2)} + P_{e2}^{(3)} - 2P_{e2}^{(2)} P_{e2}^{(3)})} \triangleq \eta_1.$$

Using the fact that noise W_1 is drawn from $\mathcal{N}(0, \sigma^2)$, we obtain

$$\frac{\exp[-\frac{(y_1-h_1)^2}{2\sigma^2}]}{\exp[-\frac{(y_1-h_0)^2}{2\sigma^2}]} \underset{\hat{H}_1(y_1)=h_0}{\overset{\hat{H}_1(y_1)=h_1}{\gtrless}} \eta_1,$$

or

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

Due to the symmetry among referees, the other referees have decision rules of similar form:

$$y_2 \underset{\hat{H}_2(y_2)=h_0}{\overset{\hat{H}_2(y_2)=h_1}{\gtrless}} \frac{h_1 - h_0}{2} + \frac{\sigma^2}{h_1 - h_0} \ln \left(\frac{c_{10} p_0 (P_{e1}^{(3)} + P_{e1}^{(1)} - 2P_{e1}^{(3)} P_{e1}^{(1)})}{c_{01}(1-p_0)(P_{e2}^{(3)} + P_{e2}^{(1)} - 2P_{e2}^{(3)} P_{e2}^{(1)})} \right) \triangleq \lambda_2,$$
$$y_3 \underset{\hat{H}_3(y_3)=h_0}{\overset{\hat{H}_3(y_3)=h_1}{\gtrless}} \frac{h_1 - h_0}{2} + \frac{\sigma^2}{h_1 - h_0} \ln \left(\frac{c_{10} p_0 (P_{e1}^{(1)} + P_{e1}^{(2)} - 2P_{e1}^{(1)} P_{e1}^{(2)})}{c_{01}(1-p_0)(P_{e2}^{(1)} + P_{e2}^{(2)} - 2P_{e2}^{(1)} P_{e2}^{(2)})} \right) \triangleq \lambda_3.$$

Thus, determining the referees' optimal decision rules is equivalent to finding the optimal values of decision thresholds λ_1, λ_2, and λ_3.

Probabilities of local errors can be determined from the decision thresholds:

$$P_{e1}^{(i)} = \mathbb{P}[Y_i \geq \lambda_i | H = h_0] = Q\left(\frac{\lambda_i - h_0}{\sigma}\right), \tag{3.4}$$
$$P_{e2}^{(i)} = \mathbb{P}[Y_i < \lambda_i | H = h_1] = Q\left(\frac{h_1 - \lambda_i}{\sigma}\right), \tag{3.5}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp[-\frac{t^2}{2}]\, dt$ is the Q-function. Bayes risk is described in terms of decision thresholds by substituting (3.4) and (3.5) into (3.2) and (3.3), and substituting them into (3.1):

$$\begin{aligned}
R = c_{10} p_0 \Big\{ & Q\left(\frac{\lambda_1 - h_0}{\sigma}\right) Q\left(\frac{\lambda_2 - h_0}{\sigma}\right) + Q\left(\frac{\lambda_2 - h_0}{\sigma}\right) Q\left(\frac{\lambda_3 - h_0}{\sigma}\right) \\
& + Q\left(\frac{\lambda_3 - h_0}{\sigma}\right) Q\left(\frac{\lambda_1 - h_0}{\sigma}\right) - 2Q\left(\frac{\lambda_1 - h_0}{\sigma}\right) Q\left(\frac{\lambda_2 - h_0}{\sigma}\right) Q\left(\frac{\lambda_3 - h_0}{\sigma}\right) \Big\} \\
& + c_{01}(1-p_0) \Big\{ Q\left(\frac{h_1 - \lambda_1}{\sigma}\right) Q\left(\frac{h_1 - \lambda_2}{\sigma}\right) + Q\left(\frac{h_1 - \lambda_2}{\sigma}\right) Q\left(\frac{h_1 - \lambda_3}{\sigma}\right) \\
& + Q\left(\frac{h_1 - \lambda_3}{\sigma}\right) Q\left(\frac{h_1 - \lambda_1}{\sigma}\right) - 2Q\left(\frac{h_1 - \lambda_1}{\sigma}\right) Q\left(\frac{h_1 - \lambda_2}{\sigma}\right) Q\left(\frac{h_1 - \lambda_3}{\sigma}\right) \Big\} \\
\triangleq\ & r(\lambda_1, \lambda_2, \lambda_3)
\end{aligned} \tag{3.6}$$

Conjecture 3.1. *Identical referees have a triplet of optimal decision thresholds* $(\lambda_1^*, \lambda_2^*, \lambda_3^*)$

such that $(\lambda_1^*, \lambda_2^*, \lambda_3^*)$ *is a global minimum of* $r(x, y, z)$ *and* $\lambda_1^* = \lambda_2^* = \lambda_3^*$. *The triplet leads to the minimum Bayes risk. In other words, it is optimal for identical referees to use the identical decision rules.*

In Figure 3-2, the gray region depicts the achievable region of (P_{E1}, P_{E2}) in the three-referee system. The region is lower bounded by the performance of referees who use the same decision rules (the red solid curve), and upper bounded by the performance of referees of whom the first two referees use the fixed decision thresholds ∞ and $-\infty$ and the last referee uses an arbitrary decision threshold (the blue dashed curve). Note that the performance of the latter referees is equal to that of a single referee. By rewriting (3.1), we obtain

$$P_{E2} = -\frac{c_{10}p_0}{c_{01}(1-p_0)} P_{E1} + \frac{1}{c_{01}(1-p_0)} R. \tag{3.7}$$

In order for the team of referees to achieve the smallest Bayes risk R, the line (3.7) should be tangent to the lower bound of the operating region. Thus, for any c_{10}, c_{01}, and p_0, the optimal P_{E1} and P_{E2} are always on the operating characteristic of referees who use the same decision rules, which means using the same optimal decision rules leads to the minimum Bayes risk.

From Conjecture 3.1, we can assume that all referees use λ as their decision thresholds and simplify (3.6) to

$$\begin{aligned} R &= r(\lambda, \lambda, \lambda) \\ &= c_{10}p_0 \left\{ 3Q^2\left(\frac{\lambda - h_0}{\sigma}\right) - 2Q^3\left(\frac{\lambda - h_0}{\sigma}\right) \right\} \\ &\quad + c_{01}(1-p_0) \left\{ 3Q^2\left(\frac{h_1 - \lambda}{\sigma}\right) - 2Q^3\left(\frac{h_1 - \lambda}{\sigma}\right) \right\}. \end{aligned}$$

Since P_{E2} is strictly convex in P_{E1} when all referees use the same decision thresholds, $r(\lambda, \lambda, \lambda)$ has exactly one stationary point, which is the global minimum. Thus, we can determine the optimal decision threshold λ^* by computing the solution of

$$\left. \frac{dr(\lambda, \lambda, \lambda)}{d\lambda} \right|_{\lambda = \lambda^*} = 0.$$

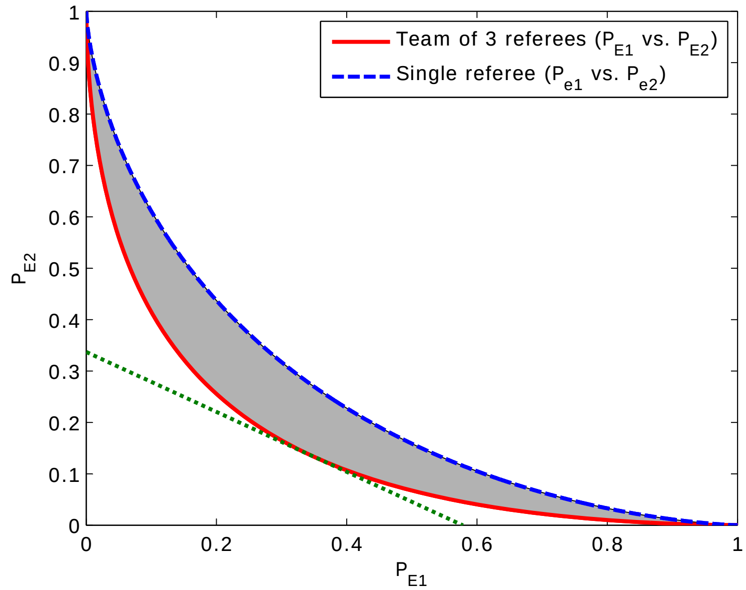
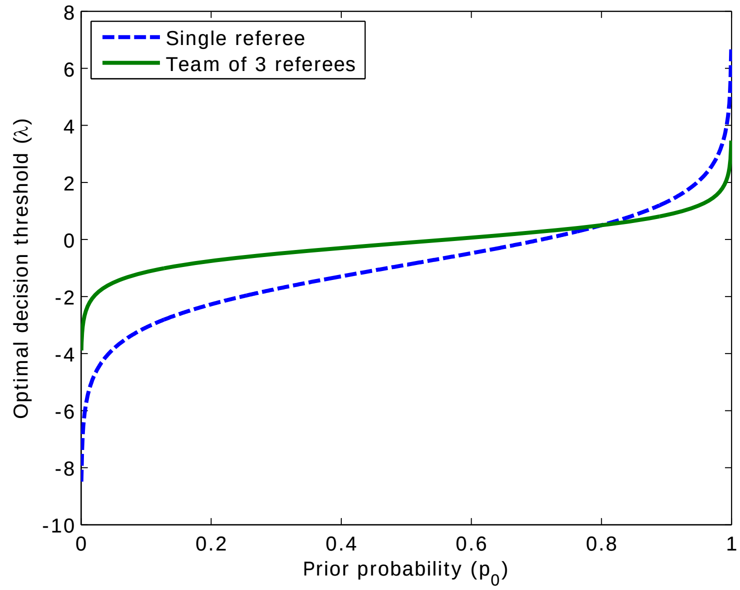


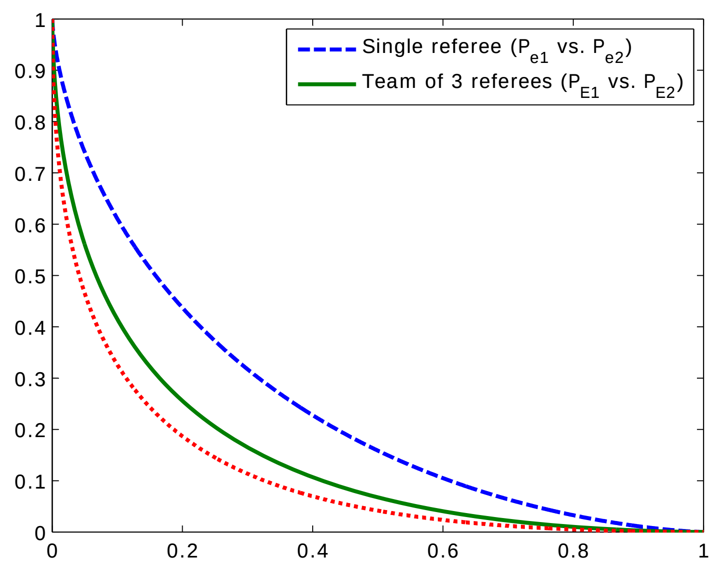
Figure 3-2. The operating region of the three-referee model for $h_0 = 0$, $h_1 = 1$, and $\sigma = 1$. The green dotted line depicts (3.7) for $c_{10} = 1$, $c_{01} = 4$, and $p_0 = 0.7$.

Figure 3-3a depicts change of the optimal decision threshold as a function of prior probability p_0 for a single referee and a team of three referees. In both cases, the optimal thresholds tend to be smaller than $(h_1 - h_0)/2$. Since $c_{10} > c_{01}$, referees think that not missing h_1 is more important than detecting h_0. The two curves meet at the p_0 such that $p_0/(1 - p_0) = c_{01}/c_{10}$, where the optimal threshold is $(h_1 - h_0)/2$. Compared to the optimal decision rule of the single referee, however, the team of referees uses decision thresholds that vary less as a function of p_0. The more observations referees have, the better decisions they can make. As referees have more observations, the referees' dependency on observations increases and their dependency on prior probability decreases. Thus, the optimal decision rule of the team depends on p_0 less than that of the single referee does.

Flipped versions of the operating characteristic curves in Figure 3-3b show that the team of referees can achieve smaller probabilities of errors than the single referee does, which means that the team's Bayes risk is smaller than the Bayes risk of a single referee for any p_0. This is pretty obvious because the team of referees have more information than the single referee. In order to prove this precisely and analyze



(a)



(b)

Figure 3-3. Differences between a single referee and a team of three identical referees for $h_0 = 0$, $h_1 = 1$, and $\sigma = 1$. All referees have Bayes costs $c_{10} = 1$ and $c_{01} = 4$. (a) Optimal decision threshold for prior probability p_0. (b) Flipped versions of the operating characteristic curves (redrawn from Figure 3-2). For comparison, the flipped operating characteristic curve of soft decision-making referees is also drawn in red dotted curve.

how much improvement in performance the team of referees can make, we introduce an equivalent single-referee model.

Corollary 3.2. *There exists a single-referee model that is equivalent to an identical-three-referee model. Let $f_v(v)$ and $f_W(w)$ denote the probability density functions of additive noises in the single-referee model and three-referee model, respectively. Then, they satisfy that*

$$f_V(v) = 6(F_W(v) - F_W^2(v))f_W(v),$$

where $F_W(v) = \int_{-\infty}^{v} f_W(w)\,dw$.

Proof. From Conjecture 3.1, we know that the best decision rule is the same for all referees. From this, we can think about an equivalent single-referee model, in which, for given p_0 and cost function, the referee uses the same decision rule and has the same Bayes risk as the referees in the three-referee model. Consider a single-referee model where the referee has the same cost function as the referees in the original model but he experiences different additive random noise V. Let $\tilde{P}_{e1}$ and $\tilde{P}_{e2}$ denote the probabilities of each error in the single-referee model. They are determined by the single referee's decision threshold $\tilde{\lambda}$:

$$\tilde{P}_{e1} = \int_{\tilde{\lambda}}^{\infty} f_V(v)\,dv,$$
$$\tilde{P}_{e2} = \int_{-\infty}^{\tilde{\lambda}} f_V(v)\,dv,$$

where $f_V(v)$ denotes the density function of noise V. When the prior probability of the object is p_0, the single referee has the Bayes risk $\tilde{R} = c_{10}p_0\tilde{P}_{e1} + c_{01}(1-p_0)\tilde{P}_{e2}$. If $\tilde{P}_{e1} = P_{E1}$ and $\tilde{P}_{e2} = P_{E2}$ for any $\tilde{\lambda} = \lambda_1 = \lambda_2 = \lambda_3$, then the single referee uses the same decision rule as the best decision rule of the team of referees in the three-referee model. We want to find the distribution of V such that it leads to the same probabilities of errors.

$$\begin{aligned} -f_V(\lambda - h_0) &= \frac{d\tilde{P}_{e1}}{d\lambda} = \frac{dP_{E1}}{d\lambda} = \frac{d}{d\lambda}(3P_{e1}^2 - 2P_{e1}^3) \\ &= 6(P_{e1} - P_{e1}^2)\frac{dP_{e1}}{d\lambda} \\ &= -6(F_W(\lambda - h_0) - F_W^2(\lambda - h_0))f_W(\lambda - h_0), \end{aligned}$$

where $F_W(w)$ denotes the cumulative distribution function of noise W in the three-referee model and we use that $P_{e1} = 1 - F_W(\lambda - h_0)$. In a similar way,

$$\begin{aligned} f_V(\lambda - h_1) &= \frac{d\tilde{P}_{e2}}{d\lambda} = \frac{dP_{E2}}{d\lambda} \\ &= 6(F_W(\lambda - h_1) - F_W^2(\lambda - h_1))f_W(\lambda - h_1). \end{aligned}$$

Thus, both the referees in the single-referee model and the team of referees in the original model use the same decision rules and have the same Bayes risks for any p_0 if the density functions of the noises in the two models satisfy

$$f_V(v) = 6(F_W(v) - F_W^2(v))f_W(v). \tag{3.8}$$

Note that $f_V(v)$ is a valid probability density function: $f_V(v) \geq 0$ for all v since $F_W(v) - F_W^2(v) \geq 0$ and $f_W(v) \geq 0$. Also, from that

$$\begin{aligned} \int_{-\infty}^{\infty} F_W(v)f_W(v)\,dv &= F_W(v)F_W(v)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f_W(v)F_W(v)\,dv \\ &= \frac{1}{2}F_W^2(v)\Big|_{-\infty}^{\infty} = \frac{1}{2}, \end{aligned} \tag{3.9}$$

and that

$$\begin{aligned} \int_{-\infty}^{\infty} F_W^2(v)f_W(v)\,dv &= F_W^2(v)F_W(v)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 2F_W(v)f_W(v)F_W(v)\,dv \\ &= \frac{1}{3}F_W^3(v)\Big|_{-\infty}^{\infty} = \frac{1}{3}, \end{aligned} \tag{3.10}$$

it is derived that

$$\int_{-\infty}^{\infty} f_V(v)\,dv = \int_{-\infty}^{\infty} 6(F_W(v) - F_W^2(v))f_W(v)\,dv = 1.$$

We refer to the model as the equivalent single-referee model of the three-referee model.

□

Consider the right-hand side of (3.8) and define a function $t(w) \triangleq 6(F_W(w) -$

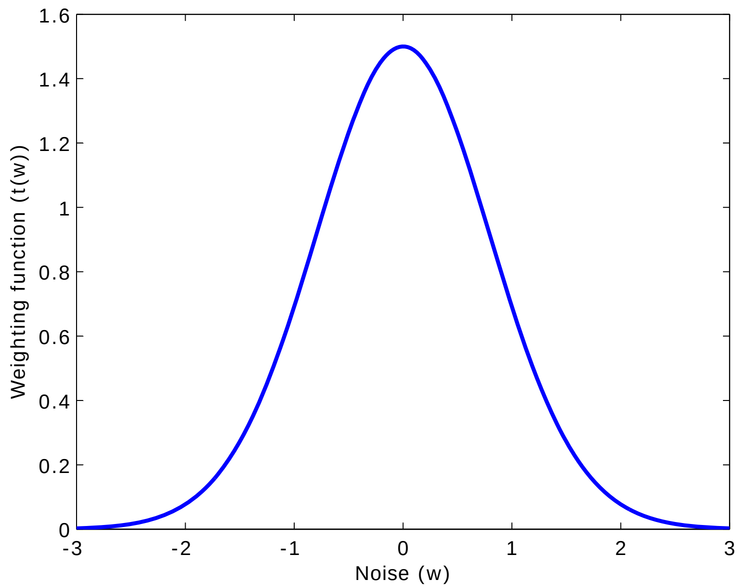


Figure 3-4. Weighting function $t(w)$ for the realization of noise W, which has the Gaussian density of $\mathcal{N}(0,1)$.

$F_W^2(w))$. The noise V in the equivalent single-referee model can be interpreted as a weighted version of noise W in the original model, where weighting function is $t(w)$. $t(w)$ reflects the effect of having three referees instead one referee. Figure 3-4 shows $t(w)$ for the realization of noise W. Since the cumulative distribution function $F_W(w)$ is a monotonically increasing function from 0 to 1 and $x - x^2$ is a concave function that has a global maximum at $x = 0.5$, $t(w)$ is greatest at the median of W and much smaller than 1 at both tails of W. Thus, $t(w)$ makes the tails of V thinner than those of W, and the variance of V is smaller than that of W. Figure 3-5 compares the density functions of noises in the three-referee model and its equivalent single-referee model.

The following lemma also shows that V has smaller variance than W does in a Gaussian-noise case:

Lemma 3.3. *The variance of V is proportional to the variance of W if W is normally distributed.*

Proof. Let $V^{(1)}$ denote the noise of the equivalent single-referee model when the noise of the three-referee model is $W^{(1)}$, whose distribution is $\mathcal{N}(0,1)$. The variance of V

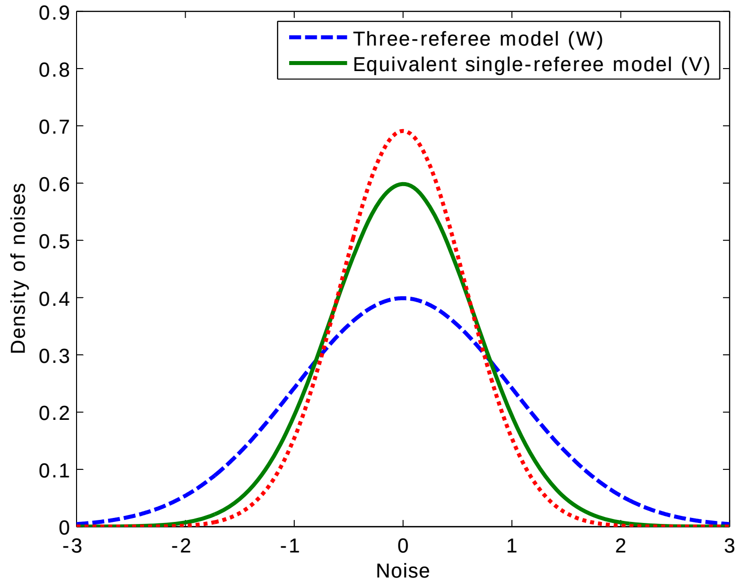


Figure 3-5. The density of noise V in a single-referee model that is equivalent to the three-referee model with noise W, which has the Gaussian density of $\mathcal{N}(0,1)$. For comparison, the density of $\frac{1}{3}W$, the effective noise of soft decision-making referees, is drawn in red dotted curve.

is

$$
\begin{aligned}
Var(V) &= \int_{-\infty}^{\infty} v^2 f_V(v)\,dv \\
&= \int_{-\infty}^{\infty} v^2 6(F_W(v) - F_W(v))f_W(v)\,dv \\
&= \int_{-\infty}^{\infty} v^2 6\left(\int_{-\infty}^{v} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{w^2}{2\sigma^2}\right] dw - \left(\int_{-\infty}^{v} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{w^2}{2\sigma^2}\right] dw\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{v^2}{2\sigma^2}\right] dv \\
&= \int_{-\infty}^{\infty} v^2 6\left(\int_{-\infty}^{v/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w'^2}{2}\right] dw' - \left(\int_{-\infty}^{v/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w'^2}{2}\right] dw'\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{v^2}{2\sigma^2}\right] dv \\
&= \int_{-\infty}^{\infty} \sigma^2 v'^2 6\left(\int_{-\infty}^{v'} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w'^2}{2}\right] dw' - \left(\int_{-\infty}^{v'} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w'^2}{2}\right] dw'\right)^2\right) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{v'^2}{2}\right] dv' \\
&= \sigma^2 Var(V^{(1)}),
\end{aligned} \tag{3.11}
$$

where $w' = w/\sigma$ and $v' = v/\sigma$. Therefore, the variance of V is proportional to σ^2, which is equal to the variance of W. Numerical calculation yields that $Var(V^{(1)}) \approx 0.449$. Thus, $Var(V) \approx 0.449\sigma^2 < Var(W) = \sigma^2$. $\square$

Now compare the three-referee model to a single-referee model with noise W, which has the same distribution as the noises in the three-referee model. Since the

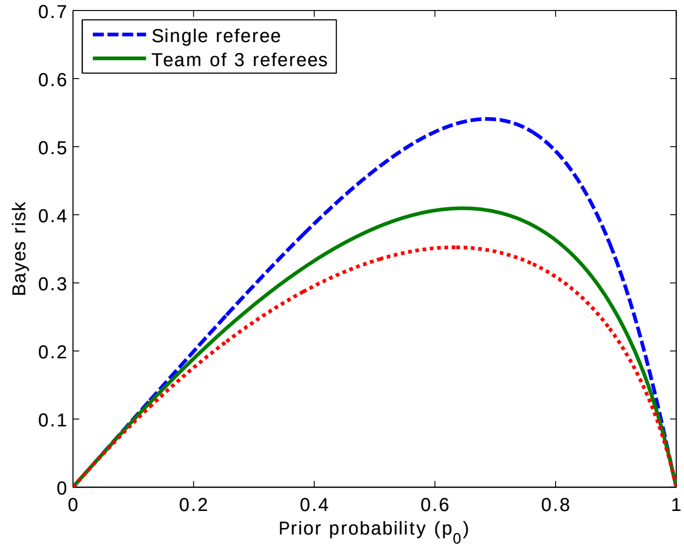


Figure 3-6. Bayes risks of a single referee and a team of three identical referees for $h_0 = 0$, $h_1 = 1$, and $\sigma = 1$. All referees have Bayes costs $c_{10} = 1$ and $c_{01} = 4$. For comparison, the Bayes risk of soft decision-making referees is also drawn in red dotted curve.

effective variance of noises in three-referee model, which is equal to the variance of V, is smaller than the variance of noise W in the single-referee model, the team of referees can achieve the probabilities of errors $P_{E1} = \tilde{P}_{e1}$ and $P_{E2} = \tilde{P}_{e2}$, which are respectively smaller than the single referee's probabilities of errors P_{e1} and P_{e2}. Therefore, the flipped operating characteristic curve of the team of three referees is always lower than that of the single referee as in Figure 3-3b, and consequently, the team of referees can achieve smaller Bayes risk than a single referee can, Figure 3-6.

Lemma 3.3 is also used to compare the performances of hard decision-making and soft decision-making. Consider a team of three referees under soft decision-making who transfer their exact observations Y_i to a fusion center so that it can make a soft decision. In this case, the fusion center can make the best decision based on the information $\frac{1}{3}(Y_1+Y_2+Y_3) = H + \frac{1}{3}(W_1+W_2+W_3)$ because it is a sufficient statistic of H. The effective variance of the noises of the referees is $\frac{1}{3}\sigma^2$, which is smaller than the variance of V, $0.449\sigma^2$. This explains the gap between the flipped operating characteristic curves of soft decision-making and hard decision-making in Figure 3-3b. The performance loss results from allowing each referee only one bit to represent his observation, which may be considered as quantization of the observation.

3.3 Quantization of Prior Probabilities

We aim to optimize the quantization rules for prior probabilities. Consider the situation when there is a population of objects. Each object has its own prior probability p_0 of being h_0, which is drawn from a probability density function $f_{P_0}(p_0)$. Referees in our model are supposed to observe any of the objects and make decisions. However, they have a constraint: they can distinguish objects only into K different categories with respect to their prior probabilities. The categorization is equivalent to quantization of prior probabilities. For example, for a given quantization rule $q(\cdot)$, two objects H_1 and H_2 belong to the same category if and only if their prior probabilities satisfy

$$q(\mathbb{P}[H_1 = h_0]) = q(\mathbb{P}[H_2 = h_0]).$$

We investigate which categorization scheme lets referees pay the minimum cost.

Each referee has his own quantizer $q_i(\cdot)$ as in Figure 3-7. When he observes an object H, he knows the quantized version of its prior probability $q_i(p_0)$. Thus, he makes an optimal decision $\hat{H}_i$ based on $q_i(p_0)$ along with his observation Y_i. In this section, we restrict referees so that they use the same quantizers for prior probabilities, i.e., $q_1(\cdot) = q_2(\cdot) = q_3(\cdot)$. Without the restriction, they may use differently quantized prior probabilities for hypothesis testing. Then they have different Bayes risks to minimize, but we did not deal with this case in Section 3.2. We will consider this case in Section 4.3.

The prior probability of an object p_0 has a value between $[0, 1]$. Since we consider a population of objects and each object has own prior probability, we regard p_0 as a realization of a random variable P_0 whose density function $f_{P_0}(p_0)$ is defined for $p_0 \in [0, 1]$. We consider a K-point quantizer, which partitions the whole interval into K regions $\mathcal{R}_1, \ldots, \mathcal{R}_K$ and has K points $a_1, \ldots, a_K$ that represent the regions. It is reasonable to consider the quantizer as a regular quantizer so that each region is contiguous (i.e., $\mathcal{R}_1 = [0, b_1], \mathcal{R}_2 = (b_1, b_2], \ldots, \mathcal{R}_K = (b_{K-1}, 1]$, where $0 < b_1 < b_2 < \cdots < b_K < 1$) and the representation point a_k belongs to the region $\mathcal{R}_k$.

Let $P_{E1}(p)$ and $P_{E2}(p)$ denote probabilities of errors when all referees make de-

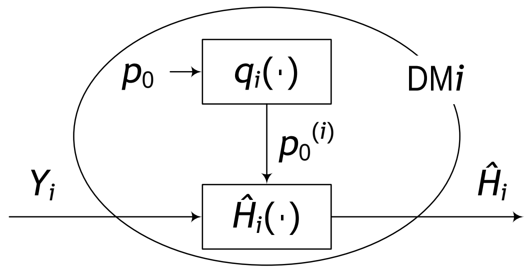


Figure 3-7. The model for referee i with a decision rule $\hat{H}_i(\cdot)$ and a quantization rule $q_i(\cdot)$.

cisions by using p, the quantized version of prior probability. If the referees use a quantizer $q(\cdot)$ and $q(p_0) = a_k$, then the probabilities of a global false alarm and a global missed detection are $P_{E1}(a_k)$ and $P_{E2}(a_k)$, respectively. Note that $P_{E1}(a_k)$ and $P_{E2}(a_k)$ are determined by the decision rule which minimizes

$$\bar{R} = c_{10} a_k P_{E1}(a_k) + c_{01}(1 - a_k) P_{E2}(a_k),$$

but $\bar{R}$ is not the actual Bayes risk that referees should take. Their actual Bayes risk[1] $\tilde{R}$ is computed by

$$\begin{aligned}\tilde{R} &= c_{10}\mathbb{P}[H = h_0] P_{E1}(a_k) + c_{01}\mathbb{P}[H = h_1] P_{E2}(a_k) \\ &= c_{10} p_0 P_{E1}(a_k) + c_{01}(1 - p_0) P_{E2}(a_k).\end{aligned}$$

The Bayes risk averaged over P_0 is

$$\begin{aligned}\mathbb{E}[\tilde{R}] &= \int_0^1 (c_{10} p_0 P_{E1}(q(p_0)) + c_{01}(1 - p_0) P_{E2}(q(p_0))) f_{P_0}(p_0)\, dp_0 \\ &= \sum_{k=1}^{K} \int_{R_k} (c_{10} p_0 P_{E1}(a_k) + c_{01}(1 - p_0) P_{E2}(a_k)) f_{P_0}(p_0)\, dp_0 \\ &= \sum_{k=1}^{K} \int_{b_{k-1}}^{b_k} (c_{10} p_0 P_{E1}(a_k) + c_{01}(1 - p_0) P_{E2}(a_k)) f_{P_0}(p_0)\, dp_0. \qquad (3.12)\end{aligned}$$

It is mean Bayes risk (MBR) that is the criterion for performance of a quantizer. There is a useful property in (3.12): MBR for each region is able to be computed

[1] We call $\tilde{R}$ mismatched Bayes risk because it is different from $R = c_{10} p_0 P_{E1}(p_0) + c_{01}(1 - p_0) P_{E2}(p_0)$, the Bayes risk when the referees know the true value of the prior probability.

independently of MBR for the other regions. Thus, representation points of the optimal quantizer should satisfy

$$a_k = \underset{a \in (b_{k-1}, b_k]}{\arg\min} \int_{b_{k-1}}^{b_k} (c_{10} p_0 P_{E1}(a) + c_{01}(1-p_0) P_{E2}(a)) f_{P_0}(p_0)\, dp_0, \tag{3.13}$$

which is called centroid condition.

In addition, there is another type of condition for boundaries of regions in the sense that p_0 should be mapped to a_k such that

$$k = \underset{k'}{\arg\min} \{c_{10} p_0 P_{E1}(a_{k'}) + c_{01}(1-p_0) P_{E2}(a_{k'})\}.$$

Then for $p_0 \in [a_k, a_{k+1}]$,

$$c_{10} p_0 P_{E1}(a_k) + c_{01}(1-p_0) P_{E2}(a_k) \underset{p_0 \in \mathcal{R}_k}{\overset{p_0 \in \mathcal{R}_{k+1}}{\gtrless}} c_{10} p_0 P_{E1}(a_{k+1}) + c_{01}(1-p_0) P_{E2}(a_{k+1}), \tag{3.14}$$

which is called nearest neighbor condition.

Since any identical-three-referee model has the equivalent single-referee model, we are able to take advantage of the results in [6]. Because

$$\left(\int_{b_{k-1}}^{b_k} c_{10} p_0 f_{P_0}(p_0)\, dp_0 \right) P_{E1}(a) + \left(\int_{b_{k-1}}^{b_k} c_{01}(1-p_0) f_{P_0}(p_0)\, dp_0 \right) P_{E2}(a)$$

has only one stationary point that is a minimum extremum [6, Theorem 2], a_k is the unique solution to

$$\left(\int_{b_{k-1}}^{b_k} c_{10} p_0 f_{P_0}(p_0)\, dp_0 \right) \left. \frac{P_{E1}(a)}{da} \right|_{a_k} + \left(\int_{b_{k-1}}^{b_k} c_{01}(1-p_0) f_{P_0}(p_0)\, dp_0 \right) \left. \frac{P_{E2}(a)}{da} \right|_{a_k} = 0. \tag{3.15}$$

In addition, the left-hand side of (3.14) is the line tangent to Bayes risk $c_{10} p_0 P_{E1}(p_0) + c_{01}(1-p_0) P_{E2}(p_0)$ at $p_0 = a_k$, and so is the right-hand side of (3.14) at $p_0 = a_{k+1}$. Thus, by [6, Theorem 1], the boundary between $\mathcal{R}_k$ and $\mathcal{R}_{k+1}$ is b_k such that the two expressions are equal at $p_0 = b_k$:

$$b_k = \frac{c_{01}\left(P_{E2}(a_{k+1}) - P_{E2}(a_k)\right)}{c_{01}\left(P_{E2}(a_{k+1}) - P_{E2}(a_k)\right) - c_{10}\left(P_{E1}(a_{k+1}) - P_{E1}(a_k)\right)}. \tag{3.16}$$

The strict convexity of $\tilde{R}$ in a_k shown in [6, Theorem 1] also implies that the quantizers that satisfy the centoid and nearest neighbor conditions are regular [20, Lemma 6.2.1].

The Lloyd-Max algorithm is an algorithm to find a quantizer that meets the centroid condition and the nearest neighbor condition. The algorithm alternates between optimizing representation points for a given set of endpoints through (3.15) and optimizing endpoints for the new representation points through (3.16). As given in [6, 21], if $f_{P_0}(p_0)$ is positive and continuous in $(0, 1)$ and

$$\int_0^1 \left(c_{10} p_0 P_{E1}(a) + c_{01}(1 - p_0) P_{E2}(a)\right) f_{P_0}(p_0)\, dp_0$$

is finite for all a, then the algorithm converges to an optimal quantizer.

The plots in Figure 3-8 depict Bayes risks due to the minimum-MBR quantizers as blue solid lines; the circle markers are representation points. The green solid curves are Bayes risk without quantization of prior probabilities, which is the same as in Figure 3-6. It is obvious that the mean error between the mismatched Bayes risk and the true Bayes risk decreases as K increases.

For comparison, Figure 3-8 also shows Bayes risks of a single-referee model with the same Bayes costs as dashed lines: the green dashed curves are unquantized Bayes risk and the blue dashed lines are mismatched Bayes risk. The results show that, for some p_0 which is closer to 0 or 1, the mismatched Bayes risk of the team of referees is greater than that of the single referee. Mean mismatched Bayes risk of the team of referees, however, is always smaller than that of the single referee.

Consider a single-referee model and a three-referee model where all referees have the same Bayes costs and use the same quantizer, which is optimized for the single referee. Let $a_1, \ldots a_K$ denote the quantizer's representation points and $b_1, \ldots, b_{K-1}$

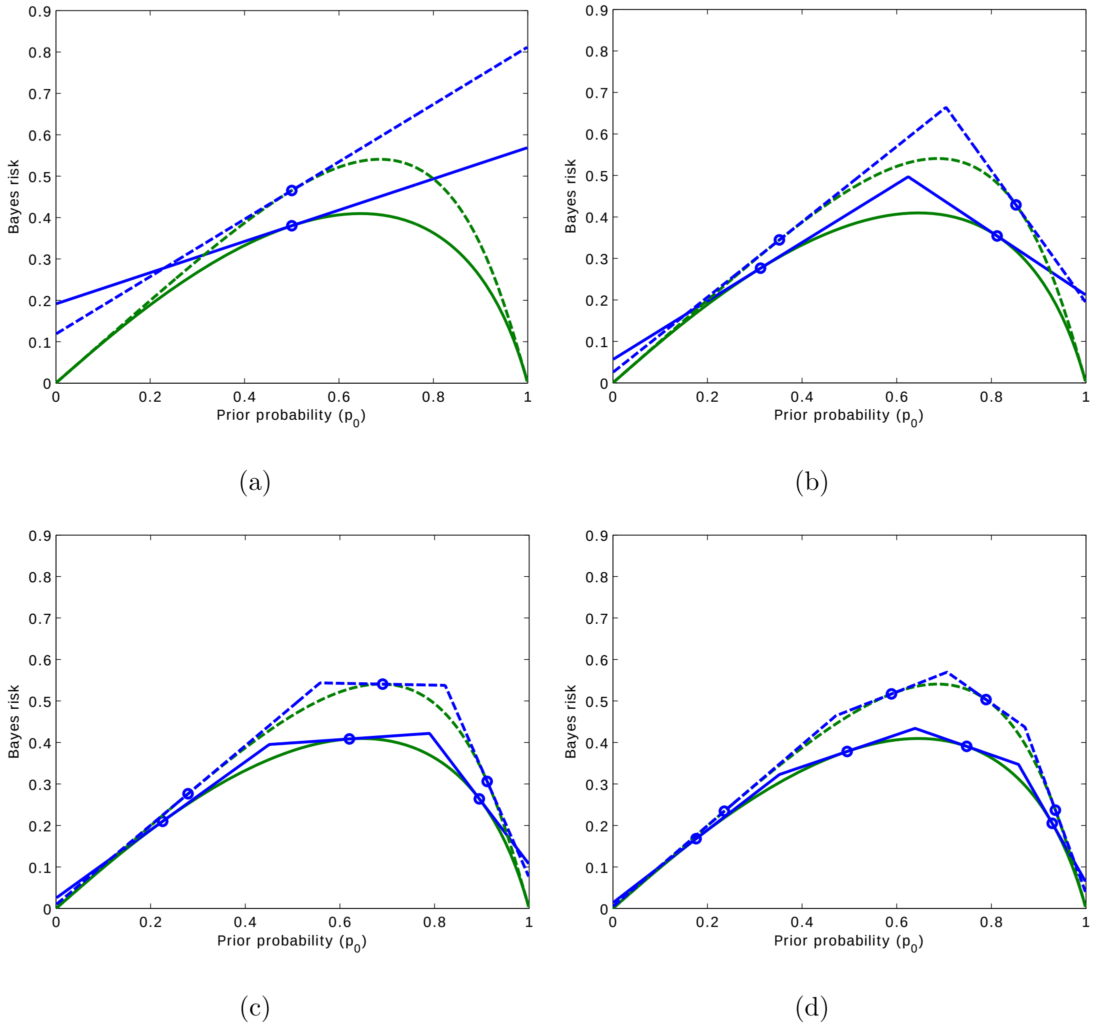


Figure 3-8. Quantizers for uniformly distributed P_0, $h_0 = 0$, $h_1 = 1$, $\sigma = 1$, and Bayes costs $c_{10} = 1$ and $c_{01} = 4$. Mismatched Bayes risk and unquantized Bayes risk are plotted for (a) $K = 1$, (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$ in three-referee model (as solid line) and single-referee model (as dashed line).

denote its endpoints. In Section 3.2, we get

$$P_{e1}(p_0) \geq P_{E1}(p_0),$$

$$P_{e2}(p_0) \geq P_{E2}(p_0),$$

where both inequalities hold with equality only for $p_0 = 0$ or $p_0 = 1$. Thus,

$$\begin{aligned}&\int_{b_{k-1}}^{b_k} (c_{10} p_0 f_{P_0}(p_0)\, dp_0)\, P_{e1}(a_k) + \int_{b_{k-1}}^{b_k} (c_{01}(1-p_0) f_{P_0}(p_0)\, dp_0)\, P_{e2}(a_k)\\ &> \int_{b_{k-1}}^{b_k} (c_{10} p_0 f_{P_0}(p_0)\, dp_0)\, P_{E1}(a_k) + \int_{b_{k-1}}^{b_k} (c_{01}(1-p_0) f_{P_0}(p_0)\, dp_0)\, P_{E2}(a_k)\end{aligned} \quad (3.17)$$

for any $a_k \in (0, 1)$. The left-hand side of (3.17) is the MBR of the single referee in region $\mathcal{R}_k$ and the right-hand side is that of the team of referees in the same region. Hence, the MBR of the single referee is greater than that of the team of referees. Even though the quantizer is optimal for the single referee, however, it may not be for the team of referees; they can achieve an even smaller MBR by optimizing their quantizers. Therefore, a team of three identical referees always makes better performance on average than a single referee can do even if they quantize prior probabilities.

Chapter 4

Non-Identical Referees

We considered the case when a team of referees share one cost function and collaborate in order to make the best global decision with regard to Bayes risk in the previous chapter. In general human group decision-making situations, however, each referee may have a cost function that is different from the other referees' cost functions. For example, each voter has his or her own political inclination; the individual's vote in a presidential election depends on his or her political inclination as well as evaluation of each candidate. Also in a business decision-making, an executive who pursues high profit has a different cost function from that of his partner who wants safe investment.

In this chapter, we analyze how referees make decisions and categorize objects when they are allowed to have their own cost functions. We define the decision-making and quantization problems in strategic form and apply a game-theoretic approach to analyze optimal decision and quantization rules. We discuss how a referee's decision rule is affected by the others' decision rules. It is shown that a Nash equilibrium of decision thresholds always exists. Designing an optimal set of quantization rules is difficult in this case because of dependency among the referees. Under the restriction of using the same endpoints, two ways to optimize quantization rules are introduced and compared to each other.

Furthermore, within this chapter, we consider referees who share a common cost function but may categorize differently. The referees behave like identical referees except that collaborating referees can take advantage of diverse quantization rules. It is shown that the collaborating referees have incentive to use diverse quantization rules rather than identical quantization rules. We investigate to what extent the diversity in quantizers makes the performance better and how to design the optimal diverse quantization rules.

4.1 Problem Model

The model of the decision-making problem in this chapter (Figure 4-1) is the same as the model used in Chapter 3 except that referees have their own cost functions. Referee i has a cost function $C_i(\hat{H}, H)$, or Bayes costs $c_{10}^{(i)} = C_i(h_1, h_0)$ and $c_{01}^{(i)} = C_i(h_0, h_1)$. His Bayes risk is

$$R_i = c_{10}^{(i)} p_0 P_{E1} + c_{01}^{(i)} (1 - p_0) P_{E2}$$

for an object whose prior probability is p_0. Note that his cost still depends on the global decision rather than his own decision. We assume $c_{10}^{(i)} \neq c_{10}^{(j)}$ and $c_{01}^{(i)} \neq c_{01}^{(j)}$ for $i \neq j$. Even though the referees have different Bayes risks, each referee still attempts to minimize his own Bayes risk, which is the reason that the referees face conflicts of interest. For example, suppose that referees 1 and 2 pay much bigger cost for missed detections than for false alarms and referee 3 pays much bigger cost for false alarms than for missed detections. Then, referees 1 and 2 tend to declare h_1 so that they can decrease the probability of a missed detection, and consequently, the global decision is highly likely to be h_1 regardless of referee 3's decision by the majority rule. However, they would make referee 3 unhappy because their decisions increase the probability of a false alarm.

Referees need to consider each others' decision-making due to the conflict of interests. Game theory provides useful methods to analyze the referees' strategies to pursue their goals under competition. Thus, we use a game-theoretic approach, especially investigating Nash equilibria, to analyze how referees make decisions and categorize objects so that each can achieve as small a Bayes risk as possible.

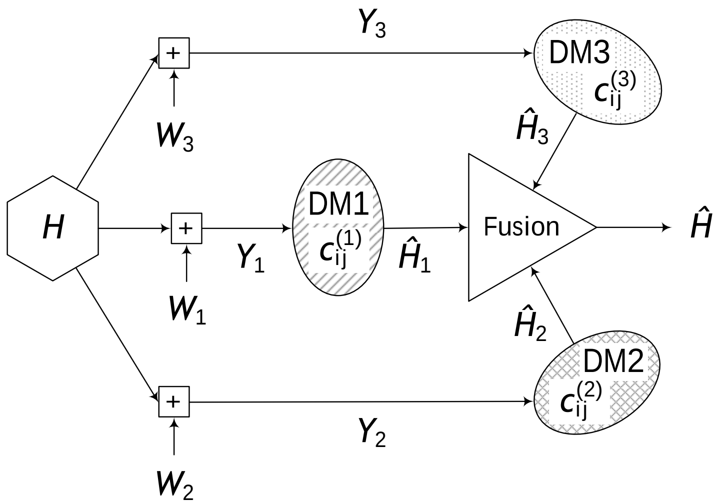


Figure 4-1. The model is the same as that in Chapter 3 except that referees may have different inclinations.

4.1.1 Problem Description in the Game-Theoretic Point of View

We define the decision-making and quantization problems in strategic form to apply game-theoretic approach to them. Each referee is assumed to know the other referees' cost functions but not to know their decisions when he makes his decision. First of all, we need to rewrite our model in strategic form $(\mathcal{I}, (S_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}})$, which consists of the finite set of players $i \in \mathcal{I} = \{1, \ldots, I\}$, the set of available strategies (or actions) $s_i \in S_i$ for each player i, and the payoff (or utility) function $u_i : \prod_{i\in\mathcal{I}} S_i \mapsto \mathbb{R}$ for each player i. In the situation where three referees make decisions based on differently quantized prior probabilities $p_0^{(1)}$, $p_0^{(2)}$, and $p_0^{(3)}$, respectively, the game can be described as follows:

Game I: Determination of decision rules

- $I = 3$ and $\mathcal{I} = \{1, 2, 3\}$ represent three referees,
- $S_i = \mathbb{R}, \forall i \in \mathcal{I}$ is a set of possible decision thresholds s_i for referee i,
- $u_i = -R_i = -c_{10}^{(i)} p_0^{(i)} P_{E1} - c_{01}^{(i)}(1 - p_0^{(i)}) P_{E2}, \forall i \in \mathcal{I}$ is the payoff function for referee i.

We define each referee's payoff function as the negative Bayes risk so that referees are able to minimize their Bayes risks by maximizing their payoff functions.

The game of determining optimal quantizers for prior probabilities can be defined in a similar way:

Game II: Determination of quantizers for prior probabilities

- $I = 3$ and $\mathcal{I} = \{1, 2, 3\}$ represent three referees,
- $s_i = (a_1^{(i)}, \ldots, a_K^{(i)}, b_1^{(i)}, \ldots, b_{K-1}^{(i)}), \forall i \in \mathcal{I}$, i.e., strategy is a quantizer for prior probabilities, where $a_k^{(i)}$ denotes the representation point of k-th region $[b_{k-1}^{(i)}, b_k^{(i)}) \subset [0, 1]$.
- $v_i = -\mathbb{E}[R_i] = -\int R_i f_{P_0}(p_0)\, dp_0, \forall i \in \mathcal{I}$ is the payoff function for referee i.

Note that there are $2K-1$ degrees of freedom in a strategy when referees use K-point quantizers. We define each player's payoff function as the negative of his mean Bayes risk so that they are able to determine minimum MBR quantizers.

4.2 Conflicting Referees

We investigate how conflicting referees will determine their optimal decision and quantization rules. When referees have different cost functions, each referee's optimal decision rule depends on others' decision rules. Conflicts among referees arise from this dependency. Figure 4-2 shows the conflict between referees 1 and 2. Initially, referees 1, 2, and 3 use -0.886, 1.886. and 0.5 as their decision thresholds, respectively. Because referee 1 notices that he can do better by changing his decision threshold, he changes his decision threshold from -0.886 to -1.5470 while referees 2 and 3 fix their decision rules. The change, however, also affects the performance of the decision rule of referee 2: not only is his minimum Bayes risk increased but also his optimal decision rule is changed. The change also affects the performance of the decision rule of referee 3. We say that referees are conflicting if their only goal is minimizing their

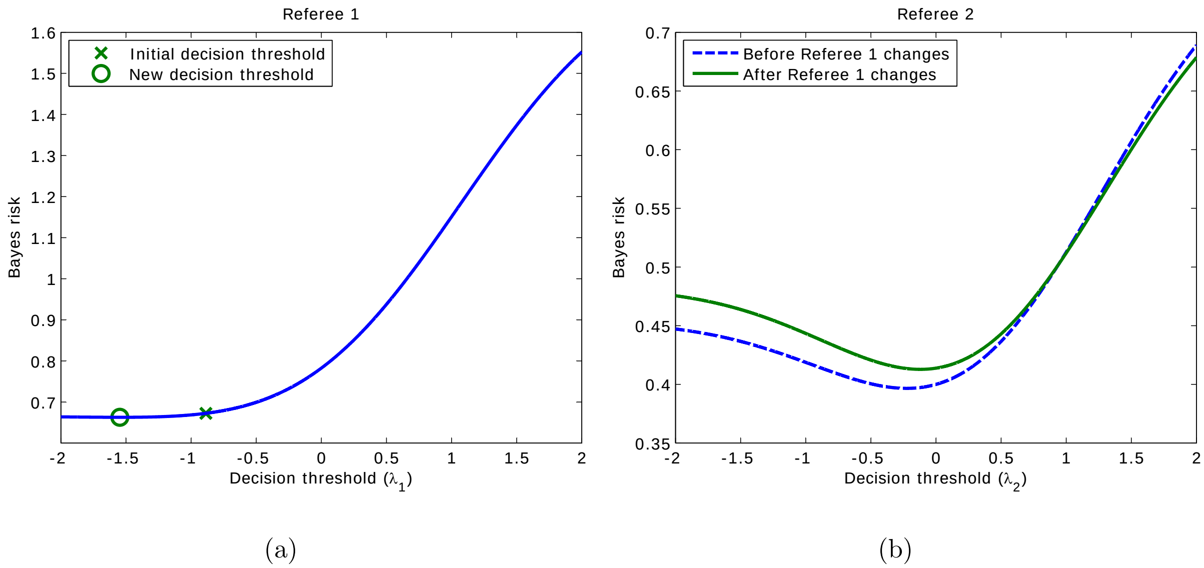


Figure 4-2. Change of the Bayes risks that shows how referees conflict. Referee 1 has Bayes costs $c_{10}^{(1)} = 1$ and $c_{01}^{(1)} = 4$, referee 2 has $c_{10}^{(2)} = 4$ and $c_{01}^{(2)} = 1$, and referee 3 has $c_{10}^{(3)} = 4$ and $c_{01}^{(3)} = 4$. $p_0 = 0.5$. (a) The Bayes risk of Referee 1 for his decision threshold. (b) The Bayes risks of Referee 2 before and after Referee 1 changes his decision threshold.

own Bayes risk. We will also consider the case when they are collaborating in Section 4.3.

4.2.1 Decision-Making Strategy

We need to find out how the referees fix their decision rules. The rules are easy to fix if there exists a dominant strategy in Game I, i.e., if there exists a decision rule that is optimal regardless of other referees' decision rules.

Theorem 4.1. *If the density function of noises is continuous and always greater than zero, then dominant strategies do not exist for any cost functions and* $p_0^{(i)}, i = 1, 2, 3$.

Proof. It is sufficient to consider dominant strategies of referee 1 due to the symmetry among the referees. By definition, s_1^* is dominant if, for $\forall s_1 \in S_1$ and $\forall (s_2, s_3) \in S_2 \times S_3$,

$$u_1(s_1^*, s_2, s_3) \geq u_1(s_1, s_2, s_3),$$

which is equivalent to

$$
\begin{aligned}
&-c_{10}^{(1)}p_0^{(1)}P_{E1}(s_1^*, s_2, s_3) - c_{01}^{(1)}(1-p_0^{(1)})P_{E2}(s_1^*, s_2, s_3) \\
&\geq -c_{10}^{(1)}p_0^{(1)}P_{E1}(s_1, s_2, s_3) - c_{01}^{(1)}(1-p_0^{(1)})P_{E2}(s_1, s_2, s_3),
\end{aligned} \tag{4.1}
$$

where $P_{E1}(s_1, s_2, s_3)$ and $P_{E2}(s_1, s_2, s_3)$ respectively denote probabilities of a global false alarm and a global missed detection when the decision threshold of referee i is s_i. According to our fusion rule,

$$
\begin{aligned}
P_{E1}(s_1, s_2, s_3) &= P_{e1}^{(1)}(s_1)P_{e1}^{(2)}(s_2) + P_{e1}^{(2)}(s_2)P_{e1}^{(3)}(s_3) + P_{e1}^{(3)}(s_3)P_{e1}^{(1)}(s_1) \\
&\quad -2P_{e1}^{(1)}(s_1)P_{e1}^{(2)}(s_2)P_{e1}^{(3)}(s_3), \qquad (4.2) \\
P_{E2}(s_1, s_2, s_3) &= P_{e2}^{(1)}(s_1)P_{e2}^{(2)}(s_2) + P_{e2}^{(2)}(s_2)P_{e2}^{(3)}(s_3) + P_{e2}^{(3)}(s_3)P_{e2}^{(1)}(s_1) \\
&\quad -2P_{e2}^{(1)}(s_1)P_{e2}^{(2)}(s_2)P_{e2}^{(3)}(s_3). \qquad (4.3)
\end{aligned}
$$

By defining $f_1(s_2, s_3) = P_{e1}^{(2)}(s_2)+P_{e1}^{(3)}(s_3)-2P_{e1}^{(2)}(s_2)P_{e1}^{(3)}(s_3)$ and $f_2(s_2, s_3) = P_{e2}^{(2)}(s_2)+P_{e2}^{(3)}(s_3) - 2P_{e2}^{(2)}(s_2)P_{e2}^{(3)}(s_3)$, (4.1) is equivalent to

$$
\begin{aligned}
&c_{10}^{(1)}p_0^{(1)}P_{e1}^{(1)}(s_1^*)f_1(s_2, s_3) + c_{01}^{(1)}(1-p_0^{(1)})P_{e2}^{(1)}(s_1^*)f_2(s_2, s_3) \\
&\leq c_{10}^{(1)}p_0^{(1)}P_{e1}^{(1)}(s_1)f_1(s_2, s_3) + c_{01}^{(1)}(1-p_0^{(1)})P_{e2}^{(1)}(s_1)f_2(s_2, s_3).
\end{aligned} \tag{4.4}
$$

Consider a variable $t = P_{e1}^{(1)}(s_1)$ and a function $g(t) = P_{e2}^{(1)}(s_1)$ such that $g(t)$ is referee 1's probability of a missed detection when his probability of a false alarm is t. If the density of noise W_1 is continuous and always greater than zero, then the function $P_{e1}^{(1)} : \mathbb{R} \mapsto [0, 1]$ is one-to-one and onto. Hence, it is possible to define an inverse function $(P_{e1}^{(1)})^{-1}$ of it and thus define $g(t) = P_{e2}^{(1)} \circ (P_{e1}^{(1)})^{-1}(t)$.

Subsituting t and $g(t)$ into (4.4), we get

$$
\begin{aligned}
&c_{10}^{(1)}p_0^{(1)}f_1(s_2, s_3)t^* + c_{01}^{(1)}(1-p_0^{(1)})f_2(s_2, s_3)g(t^*) \\
&\leq c_{10}^{(1)}p_0^{(1)}f_1(s_2, s_3)t + c_{01}^{(1)}(1-p_0^{(1)})f_2(s_2, s_3)g(t),
\end{aligned} \tag{4.5}
$$

where $t^* = P_{e1}^{(1)}(s_1^*)$. In order for s_1^* to be dominant, $h(t) \triangleq c_{10}^{(1)} p_0^{(1)} f_1(s_2,s_3)t + c_{01}^{(1)}(1-p_0^{(1)}) f_2(s_2,s_3)g(t)$ should have a global minimum point at t^*. Since $g(t)$ is monotonically decreasing and convex in t, the slope of $g(t)$ is negative and monotonically increasing in t. Thus, the location of the minimal extreme of $h(t)$ depends on $f_1(s_2,s_3)/f_2(s_2,s_3)$: the minimal extreme moves to the left as $f_1(s_2,s_3)/f_2(s_2,s_3)$ increases. Therefore, no t^* exists such that $h(t)$ is minimum at $t = t^*$ for all $(s_2,s_3) \in S_2 \times S_3$. □

On the other hand, there exist dominated strategies. By definition, s_1^* is a dominated strategy if, for all $(s_2,s_3) \in S_2 \times S_3$, there exists some $s_1 \in S_1$ such that

$$\begin{aligned}
&-c_{10}^{(1)} p_0^{(1)} P_{E1}(s_1^*, s_2, s_3) - c_{01}^{(1)}(1-p_0^{(1)}) P_{E2}(s_1^*, s_2, s_3) \\
&\leq -c_{10}^{(1)} p_0^{(1)} P_{E1}(s_1, s_2, s_3) - c_{01}^{(1)}(1-p_0^{(1)}) P_{E2}(s_1, s_2, s_3),
\end{aligned}$$

which is equivalent to

$$c_{10}^{(1)} p_0^{(1)} f_1(s_2,s_3)t^* + c_{01}^{(1)}(1-p_0^{(1)}) f_2(s_2,s_3)g(t^*) \geq c_{10}^{(1)} p_0^{(1)} f_1(s_2,s_3)t + c_{01}^{(1)}(1-p_0^{(1)}) f_2(s_2,s_3)g(t). \tag{4.6}$$

The left-hand side of (4.6), which is defined as $h(t^*)$, has local maximum points at 0 and 1, which do not depend on $f_1(s_2,s_3)/f_2(s_2,s_3)$. Thus, we can find some point s_1 such that $u_1(s_1^*, s_2, s_3) < u_1(s_1, s_2, s_3)$ for all $(s_2,s_3) \in S_2 \times S_3$, when $P_{e1}^{(1)}(s_1^*) = 0$ or $P_{e1}^{(1)}(s_1^*) = 1$. However, such s_1^* is ∞ or $-\infty$, and any other s_1 cannot be dominated because it can be dominant for some s_2 and s_3, which is shown in the proof of Theorem 4.1. Therefore, the problem is not solvable by iterative dominance.

Since no referee has a dominant strategy, it may seem arbitrary how to determine decision rules. We propose computing a Nash equilibrium as a reasonable way to determine them because any player does not benefit by changing his own strategy unilaterally in a Nash equilibrium. However, it does not mean that the Nash equilibrium is an optimal strategy profile: there may exist a strategy profile that leads to bigger benefit than the Nash equilibrium does. One famous example is shown in Figure 4-3, which is called *prisoner's dilemma*. The only Nash equilbrium in the game is

A \ B	Cooperate	Defect
Cooperate	3, 3	0, 5
Defect	5, 0	1, 1

Figure 4-3. A classical payoff matrix in prisoner's dilemma.

(defect, defect) but, in fact, playing (cooperate, cooperate) gives both players higher payoffs than playing the Nash equilibrium. This example shows that optimality of decision rules in this conflicting-referee case is difficult to be defined compared to that in the identical-referee case.

Nevertheless, following Nash equilibrium is one of the safest strategies for all players under the limitation that they should simultaneously make their own decisions without knowledge about each others' decisions, especially when they do not have a dominant strategy. Thus, we assume that referees in the model always follow Nash equilibria. The assertion requires existence of Nash equilibria. The game of decision-making is an infinite game because each player has an infinite strategy space, and we can show existence of Nash equilibria in the game by Theorem 2.3 [17–19].

Theorem 2.3 cannot be applied to Game I because strategy sets of Game I are convex but not compact. Hence we need to define another game:

Game I′: Determination of decision rules in terms of the probability of error

---

- $I = 3$ and $\mathcal{I} = \{1, 2, 3\}$ represent three referees,
- $T_i = [0, 1], \forall i \in \mathcal{I}$ is a set of possible probabilities of a false alarm t_i for referee i,
- $u'_i(t_1, t_2, t_3) = u_i(s_1, s_2, s_3) = -R_i = -c_{10}^{(i)} p_0^{(i)} P_{E1} - c_{01}^{(i)}(1 - p_0^{(i)}) P_{E2}, \forall i \in \mathcal{I}$ is the payoff function for referee i.

Lemma 4.2. *Game I and Game I′ are equivalent for additive Gaussian noises.*

Proof. Strategies s_i in Game I and t_i in Game I′ have the following relation:

$$t_i = P_{e1}^{(i)}(s_i).$$

Since a Gaussian distribution is continuous and always greater than zero, the functions $P_{e1}^{(i)} : \mathbb{R} \mapsto [0,1]$ are one-to-one and onto. Thus, there exist inverse functions $(P_{e1}^{(i)})^{-1}$, which means that t_i are uniquely determined by s_i and vice versa. Since choosing either t_i or s_i does not affect the players' payoff functions, Game I and Game I′ are equivalent. □

Theorem 4.3. *A pure Nash equilibrium always exists in Game I for additive Gaussian noises.*

Proof. We can prove Theorem 4.3 by showing that there always exists a pure Nash equilibrium in Game I′ for additive Gaussian noises. Note that strategy sets T_i are compact and convex.

Let $g_i(t_i) \triangleq P_{e2}^{(i)}(s_i)$, where s_i is the decision threshold such that $t_i \triangleq P_{e1}^{(i)}(s_i)$. We can rewrite the payoff function for referee 1 as follows:

$$\begin{aligned}
u_1 &= -c_{10}^{(1)}p_0^{(1)}[t_1t_2 + t_2t_3 + t_3t_1 - 2t_1t_2t_3] \\
&\quad -c_{01}^{(1)}(1-p_0^{(1)})[g_1(t_1)g_2(t_2) + g_2(t_2)g_3(t_3) + g_3(t_3)g_1(t_1) - 2g_1(t_1)g_2(t_2)g_3(t_3)] \\
&= -c_{10}^{(1)}p_0^{(1)}[t_2 + t_3 - 2t_2t_3]t_1 - c_{01}^{(1)}(1-p_0^{(1)})[g_2(t_2) + g_3(t_3) - 2g_2(t_2)g_3(t_3)]g_1(t_1) \\
&\quad -[c_{10}^{(1)}p_0^{(1)}t_2t_3 + c_{01}^{(1)}(1-p_0^{(1)})g_2(t_2)g_3(t_3)] \\
&= A_1t_1 + A_2g_1(t_1) - A_3,
\end{aligned}$$

where $A_1 \triangleq -c_{10}^{(1)}p_0^{(1)}[t_2+t_3-2t_2t_3]$, $A_2 \triangleq -c_{01}^{(1)}(1-p_0^{(1)})[g_2(t_2)+g_3(t_3)-2g_2(t_2)g_3(t_3)]$, and $A_3 \triangleq c_{10}^{(1)}p_0^{(1)}t_2t_3 + c_{01}^{(1)}(1-p_0^{(1)})g_2(t_2)g_3(t_3)$. A_3 is a constant with respect to t_1. $A_1 \leq 0$ because

$$t_2 + t_3 - 2t_2t_3 = t_2(1-t_3) + (1-t_2)t_3 \geq 0,$$

and likewise, $A_2 \leq 0$. Since $g_1(t_1)$ is convex in t_1 by the characteristic of probabilities of errors [1], $u_1'(t_1, t_2, t_3)$ is concave in t_1. By the symmetries among players,

$u'_i(t_1, t_2, t_3)$ is concave in t_i. Furthermore, $g_i(t_i)$ is continuous in t_i, so $u'_i(t_1, t_2, t_3)$ is continuous in t_{-i} as well as t_i.

Thus, the newly defined game satisfies all conditions for Theorem 2.3, which tells that Game I′ has a pure Nash equilibrium $(t_1^{NE}, t_2^{NE}, t_3^{NE})$. Then we can determine a strategy profile (s_1^*, s_2^*, s_3^*) that leads to $(t_1^{NE}, t_2^{NE}, t_3^{NE})$. Since the two games are equivalent, (s_1^*, s_2^*, s_3^*) is a pure Nash equilibrium in Game I. $\square$

A Nash equilibrium (s_1^*, s_2^*, s_3^*) satisfies

$$\begin{aligned}
\left.\frac{\partial u_1(s_1,s_2,s_3)}{\partial s_1}\right|_{(s_1^*,s_2^*,s_3^*)} &= \left. -c_{10}^{(1)} p_0^{(1)} \frac{\partial P_{E1}}{\partial s_1} - c_{01}^{(1)}(1-p_0^{(1)})\frac{\partial P_{E2}}{\partial s_1}\right|_{(s_1^*,s_2^*,s_3^*)} \\
&= -c_{10}^{(1)} p_0^{(1)} \left(P_{e1}^{(2)}(s_2^*) + P_{e1}^{(3)}(s_3^*) - 2P_{e1}^{(2)}(s_2^*)P_{e1}^{(3)}(s_3^*)\right) \left.\frac{dP_{e1}^{(1)}(s_1)}{ds_1}\right|_{s_1=s_1^*} \\
&\quad -c_{01}^{(1)}(1-p_0^{(1)}) \left(P_{e2}^{(2)}(s_2^*) + P_{e2}^{(3)}(s_3^*) - 2P_{e2}^{(2)}(s_2^*)P_{e2}^{(3)}(s_3^*)\right) \left.\frac{dP_{e2}^{(1)}(s_1)}{ds_1}\right|_{s_1=s_1^*} \\
&= 0.
\end{aligned} \tag{4.7}$$

for referee 1's payoff function. Likewise, the Nash equilibrium satisfies

$$\begin{aligned}
\left.\frac{\partial u_2(s_1,s_2,s_3)}{\partial s_2}\right|_{(s_1^*,s_2^*,s_3^*)} &= -c_{10}^{(2)} p_0^{(2)} \left(P_{e1}^{(3)}(s_3^*) + P_{e1}^{(1)}(s_1^*) - 2P_{e1}^{(3)}(s_3^*)P_{e1}^{(1)}(s_1^*)\right) \left.\frac{dP_{e1}^{(2)}(s_2)}{ds_2}\right|_{s_2=s_2^*} \\
&\quad -c_{01}^{(2)}(1-p_0^{(2)}) \left(P_{e2}^{(3)}(s_3^*) + P_{e2}^{(1)}(s_1^*) - 2P_{e2}^{(3)}(s_3^*)P_{e2}^{(1)}(s_1^*)\right) \left.\frac{dP_{e2}^{(2)}(s_2)}{ds_2}\right|_{s_2=s_2^*} \\
&= 0,
\end{aligned} \tag{4.8}$$

and

$$\begin{aligned}
\left.\frac{\partial u_3(s_1,s_2,s_3)}{\partial s_3}\right|_{(s_1^*,s_2^*,s_3^*)} &= -c_{10}^{(3)} p_0^{(3)} \left(P_{e1}^{(1)}(s_1^*) + P_{e1}^{(2)}(s_2^*) - 2P_{e1}^{(1)}(s_1^*)P_{e1}^{(2)}(s_2^*)\right) \left.\frac{dP_{e1}^{(3)}(s_3)}{ds_3}\right|_{s_3=s_3^*} \\
&\quad -c_{01}^{(3)}(1-p_0^{(3)}) \left(P_{e2}^{(1)}(s_1^*) + P_{e2}^{(2)}(s_2^*) - 2P_{e2}^{(1)}(s_1^*)P_{e2}^{(2)}(s_2^*)\right) \left.\frac{dP_{e2}^{(3)}(s_3)}{ds_3}\right|_{s_3=s_3^*} \\
&= 0.
\end{aligned} \tag{4.9}$$

A Nash equilibrium can be computed by solving (4.7)-(4.9). Note that the optimal decision threshold in the identical-referee case also satisfies (4.7)-(4.9) and, thus, is also a Nash equilibrium for the identical referees. Therefore, this method that follows a Nash equilibrium can be applied to the identical-referee case as well.

In general, a Nash equilibrium of conflicting referees does not satisfy $s_1^* = s_2^* = s_3^*$ unless

$$\frac{c_{10}^{(1)} p_0^{(1)}}{c_{01}^{(1)}(1 - p_0^{(1)})} = \frac{c_{10}^{(2)} p_0^{(2)}}{c_{01}^{(2)}(1 - p_0^{(2)})} = \frac{c_{10}^{(3)} p_0^{(3)}}{c_{01}^{(3)}(1 - p_0^{(3)})}.$$

Then their operating point will be located at somewhere middle of the operating region in Figure 3-2. However, the point is not the best choice for any referee because the referees can reduce either P_{E2} by moving their operating point vertically or P_{E1} by moving it horizontally, which will give all referees smaller Bayes risks than the Nash equilibrium does. This result shows that their performance suffers when they do not agree on Bayes costs and prior probabilities.

4.2.2 Quantization Strategy

It is reasonable that referees have different quantizers for prior probabilities when referees have different cost functions. However, it is much more complicated to determine optimal quantizers in the conflicting-referee case than in the identical-referee case. Below we will discuss the reason.

Proposition 4.4. *Game II does not always have a dominant strategy.*

Proof. It is simple to show. Consider 1-point quantizers, and each referee needs to determine one representation point $a_1^{(i)}$, $i = 1, 2, 3$. Then referee 1's payoff function is

$$\begin{aligned} v_i(a_1^{(1)}, a_1^{(2)}, a_1^{(3)}) &= \int_0^1 \left(-c_{10}^{(i)} p_0 P_{E1} - c_{01}^{(i)} (1 - p_0) P_{E2} \right) f_{P_0}(p_0)\, dp_0 \\ &= -c_{10}^{(i)} \mathbb{E}[P_0] P_{E1} - c_{01}^{(i)} \left(1 - \mathbb{E}[P_0]\right) P_{E2}. \end{aligned} \tag{4.10}$$

Comparing (4.10) to the payoff function of Game I

$$u_i = -c_{10}^{(i)} p_0^{(i)} P_{E1} - c_{01}^{(i)} \left(1 - p_0^{(i)}\right) P_{E2},$$

we can see that the two equations are the same if $p_0^{(i)} = \mathbb{E}[P_0], \forall i \in \mathcal{I}$. Since a dominant decision rule does not exist according to Theorem 4.1, neither does a dominant

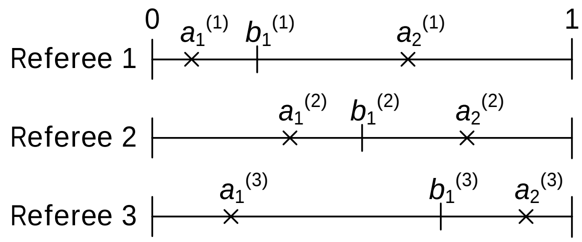


Figure 4-4. An example of possible quantizers that referees use.

representation point. □

Since it is not guaranteed that there exists a dominant strategy in Game I, we need to consider a Nash equilibrium. However, not only may a Nash equilibrium not exist in the game, but also it may be too complicated to find one if any. Because all referees are supposed to use the same quantizers in the identical-referee case, representation points for different regions are able to be independently chosen. In this conflicting-referee case, however, we need to consider dependency between different regions. Figure 4-4 depicts an example of triplets of 2-point quantizers. In the example, referee 1's mean Bayes risks in regions 1 and 2 are as follows:

$$\mathbb{E}[R_1]_{\mathcal{R}_1} = \int_0^{b_1^{(1)}} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_1^{(1)}, a_1^{(2)}, a_1^{(3)}) + c_{01}^{(1)} (1-p_0) \bar{P}_{E2}(a_1^{(1)}, a_1^{(2)}, a_1^{(3)}) \right) f_{P_0}(p_0)\, dp_0, \tag{4.11}$$

$$\begin{aligned}\mathbb{E}[R_1]_{\mathcal{R}_2} = &\int_{b_1^{(1)}}^{b_1^{(2)}} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_2^{(1)}, a_1^{(2)}, a_1^{(3)}) + c_{01}^{(1)} (1-p_0) \bar{P}_{E2}(a_2^{(1)}, a_1^{(2)}, a_1^{(3)}) \right) f_{P_0}(p_0)\, dp_0 \\ &+ \int_{b_1^{(2)}}^{b_1^{(3)}} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_2^{(1)}, a_2^{(2)}, a_1^{(3)}) + c_{01}^{(1)} (1-p_0) \bar{P}_{E2}(a_2^{(1)}, a_2^{(2)}, a_1^{(3)}) \right) f_{P_0}(p_0)\, dp_0 \\ &+ \int_{b_1^{(3)}}^{1} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_2^{(1)}, a_2^{(2)}, a_2^{(3)}) + c_{01}^{(1)} (1-p_0) \bar{P}_{E2}(a_2^{(1)}, a_2^{(2)}, a_2^{(3)}) \right) f_{P_0}(p_0)\, dp_0,\end{aligned} \tag{4.12}$$

where $\bar{P}_{E1}(x,y,z)$ and $\bar{P}_{E2}(x,y,z)$ denote the probabilities of a global false alarm and a global missed detection when referees 1, 2, and 3 respectively use quantized prior probabilities x, y, and z for decision-making. $a_1^{(2)}$ and $a_1^{(3)}$ are involved in (4.12) as well as in (4.11). Thus, via $a_1^{(2)}$ and $a_1^{(3)}$ that are affected by $a_1^{(1)}$, choice of $a_1^{(1)}$ affects on choice of $a_2^{(1)}$ and vice versa.

What is even worse is that we do not know how the variables are related. Figure 4-5 depicts a different example of triplets of 2-point quantizers. Note the structure of quantizers: $b_1^{(1)} < b_1^{(2)} < b_1^{(3)}$ in Figure 4-4, but $b_1^{(2)} < b_1^{(1)} < b_1^{(3)}$ in Figure 4-5. In this

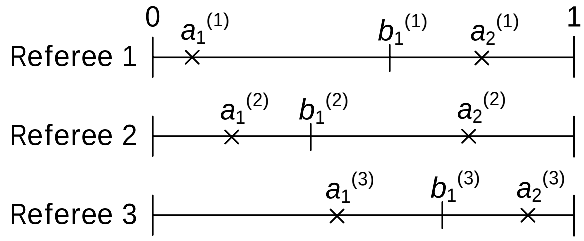


Figure 4-5. Another example of possible quantizers that referees use.

example, referee 1's mean Bayes risks in regions 1 and 2 are as follows:

$$\begin{aligned}
\mathbb{E}[R_1]_{\mathcal{R}_1} &= \int_0^{b_1^{(2)}} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_1^{(1)}, a_1^{(2)}, a_1^{(3)}) + c_{01}^{(1)}(1-p_0)\bar{P}_{E2}(a_1^{(1)}, a_1^{(2)}, a_1^{(3)})\right) f_{P_0}(p_0)\, dp_0 \\
&\quad + \int_{b_1^{(2)}}^{b_1^{(1)}} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_1^{(1)}, a_2^{(2)}, a_1^{(3)}) + c_{01}^{(1)}(1-p_0)\bar{P}_{E2}(a_1^{(1)}, a_2^{(2)}, a_1^{(3)})\right) f_{P_0}(p_0)\, dp_0, \\
\mathbb{E}[R_1]_{\mathcal{R}_2} &= \int_{b_1^{(1)}}^{b_1^{(3)}} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_2^{(1)}, a_2^{(2)}, a_1^{(3)}) + c_{01}^{(1)}(1-p_0)\bar{P}_{E2}(a_2^{(1)}, a_2^{(2)}, a_1^{(3)})\right) f_{P_0}(p_0)\, dp_0 \\
&\quad + \int_{b_1^{(3)}}^{1} \left(c_{10}^{(1)} p_0 \bar{P}_{E1}(a_2^{(1)}, a_2^{(2)}, a_2^{(3)}) + c_{01}^{(1)}(1-p_0)\bar{P}_{E2}(a_2^{(1)}, a_2^{(2)}, a_2^{(3)})\right) f_{P_0}(p_0)\, dp_0.
\end{aligned}$$

Due to the difference of the structure, how $a_1^{(1)}$ depends on the other variables is not the same in Figure 4-4 as in Figure 4-5. Since we do not know which structure is better, however, we have to consider all possible scenarios. The number of possible scenarios is $\frac{(3(K-1))!}{(K-1)!(K-1)!(K-1)!}$ for K-point quantizers, which means computational complexity is $O(3^K)$.

Quantization Using the Same Categorization

It makes the problem of quantization simpler to assume that referees use the same categorization (i.e., the same endpoints) for their quantizers. Under this assumption, all referees are allowed to optimize K representation points. Since choosing a representation point for a region is independent of representation points for other regions, the referees just need to consider the dependency among them within individual regions. Thus, the game of quantization can be split into K subgames, in which each players' strategy is defined as selecting one representation point.

Suppose that all referees use the set of fixed endpoints $\{b_0, b_1, \ldots, b_{K-1}, b_K\}$, where

$b_0 = 0$ and $b_K = 1$. The kth subgame is described as follows:

Game III: Determination of representation points for fixed categories

---

For $\mathcal{R}_k = (b_{k-1}, b_k]$,

- $I = 3$ and $\mathcal{I} = \{1, 2, 3\}$ represent three referees,
- $a_k^{(i)} \in \mathcal{R}_k = (b_{k-1}, b_k)$ is a representation point of $\mathcal{R}_k$ for referee i,
- $v_{ik}(a_k^{(1)}, a_k^{(2)}, a_k^{(3)}) = -\int_{\mathcal{R}_k} \left(c_{10}^{(i)} p_0 P_{E1} - c_{01}^{(i)} (1 - p_0) P_{E2} \right) f_{P_0}(p_0)\, dp_0$ is the payoff function for referee i.

The following equations hold at a Nash equilibrium $(a_k^{(1)*}, a_k^{(2)*}, a_k^{(3)*})$:

$$
\begin{aligned}
\frac{\partial v_{1k}(a_k^{(1)}, a_k^{(2)}, a_k^{(3)})}{\partial a_k^{(1)}} &= -c_{10}^{(1)} \epsilon_k^I \frac{\partial P_{E1}}{\partial a_k^{(1)}} - c_{01}^{(1)} \epsilon_k^{II} \frac{\partial P_{E2}}{\partial a_k^{(1)}} = 0, \\
\frac{\partial v_{2k}(a_k^{(1)}, a_k^{(2)}, a_k^{(3)})}{\partial a_k^{(2)}} &= -c_{10}^{(2)} \epsilon_k^I \frac{\partial P_{E1}}{\partial a_k^{(2)}} - c_{01}^{(2)} \epsilon_k^{II} \frac{\partial P_{E2}}{\partial a_k^{(2)}} = 0, \\
\frac{\partial v_{3k}(a_k^{(1)}, a_k^{(2)}, a_k^{(3)})}{\partial a_k^{(3)}} &= -c_{10}^{(3)} \epsilon_k^I \frac{\partial P_{E1}}{\partial a_k^{(3)}} - c_{01}^{(3)} \epsilon_k^{II} \frac{\partial P_{E2}}{\partial a_k^{(3)}} = 0,
\end{aligned}
\tag{4.13}
$$

where $\epsilon_k^I = \int_{\mathcal{R}_k} p_0 f_{P_0}(p_0)\, dp_0$ and $\epsilon_k^{II} = \int_{\mathcal{R}_k} (1 - p_0) f_{P_0}(p_0)\, dp_0$. We can find a Nash equilibrium by solving them.

In addition, we have another interesting way to determine representation points. Consider the first region $\mathcal{R}_1$ of the referees' quantizers. For any object whose prior probability is $p_0 \in \mathcal{R}_1$, the referees think that its prior probability is $a_1^{(1)}$, $a_1^{(2)}$, and $a_1^{(3)}$, respectively, and make their decisions based on the quantized versions of prior probability. Hence, they apply the same decision rules to different objects as long as they belong to the same category. Thus, it makes sense that the referees directly optimize their own decision rules for each category rather than representation points.

Game IV: Determination of decision rules for fixed categories

For $\mathcal{R}_k = (b_{k-1}, b_k]$,

- $I = 3$ and $\mathcal{I} = \{1, 2, 3\}$ represent three referees,
- $\lambda_k^{(i)}$ is a decision threshold for $p_0 \in \mathcal{R}_k$ for referee i,
- $v'_{ik}(\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)}) = -\int_{\mathcal{R}_k} \left(c_{10}^{(i)} p_0 P_{E1} - c_{01}^{(i)} (1 - p_0) P_{E2} \right) f_{P_0}(p_0)\, dp_0$ is the payoff function for referee i.

The payoff function in Game IV becomes

$$\begin{aligned} v'_{ik}(\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)}) &= -c_{10}^{(i)} \epsilon_k^I P_{E1} - c_{01}^{(i)} \epsilon_k^{II} P_{E2} \\ &= -(\epsilon_k^I + \epsilon_k^{II}) \left(c_{10}^{(i)} \left(\frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}} \right) P_{E1} + c_{01}^{(i)} \left(1 - \frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}} \right) P_{E2} \right). \end{aligned} \tag{4.14}$$

Since the objective of Game IV is to find a strategy that maximizes the payoff function, scalar multiplication of the payoff function does not change the result; we can use the following instead of (4.14):

$$v'_{ik}(\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)}) = -c_{10}^{(i)} \left(\frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}} \right) P_{E1} - c_{01}^{(i)} \left(1 - \frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}} \right) P_{E2}. \tag{4.15}$$

Comparison of (4.15) to the payoff function in Game I tells us that the two games are equivalent if $p_0 = \frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}}$. Therefore, direct optimization of decision rules is equivalent to quantizing $p_0 \in \mathcal{R}_k$ to $\frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}}$. Note that

$$\begin{aligned} \epsilon_k^I &= \int_{\mathcal{R}_k} p_0 f_{P_0}(p_0)\, dp_0 \\ &= \mathbb{P}[P_0 \in \mathcal{R}_k] \int_{\mathcal{R}_k} p_0 \frac{f_{P_0}(p_0)}{\mathbb{P}[P_0 \in \mathcal{R}_k]}\, dp_0 \\ &= \mathbb{P}[P_0 \in \mathcal{R}_k] \int_{\mathcal{R}_k} p_0 \frac{\mathbb{P}[P_0 \in \mathcal{R}_k | P_0 = p_0] f_{P_0}(p_0)}{\mathbb{P}[P_0 \in \mathcal{R}_k]}\, dp_0 \\ &= \mathbb{P}[P_0 \in \mathcal{R}_k] \mathbb{E}[P_0 | P_0 \in \mathcal{R}_k], \end{aligned}$$

where the third equality holds because $\mathbb{P}[P_0 \in \mathcal{R}_k | P_0 = p_0] = 1$ for any $p_0 \in \mathcal{R}_k$.

Since

$$\begin{aligned}\epsilon_k^I + \epsilon_k^{II} &= \int_{\mathcal{R}_k} (p_0 + (1 - p_0)) f_{P_0}(p_0)\, dp_0 \\ &= \mathbb{P}[P_0 \in \mathcal{R}_k],\end{aligned}$$

we obtain the simple expression:

$$\frac{\epsilon_k^I}{\epsilon_k^I + \epsilon_k^{II}} = \mathbb{E}[P_0 | P_0 \in \mathcal{R}_k].$$

Thus, direct optimization of decision rules is equivalent to quantizing $\forall p_0 \in \mathcal{R}_k$ to $\mathbb{E}[P_0|P_0 \in \mathcal{R}_k]$ no matter what their cost functions are. Note that $\mathbb{E}[P_0|P_0 \in \mathcal{R}_k]$ is the centroid of the region $\mathcal{R}_k$.

Figure 4-6 shows results of Games III and IV for several different sets of referees. While using Game IV is almost as good as using Game III for the referees in Figure 4-6a, the referees in Figure 4-6b and in Figure 4-6c had better use Game III. In many cases, Game III gives the better strategy that leads to lower mean Bayes risk than Game IV does.

P_{E1} and P_{E2} can be determined if either referees' quantized prior probabilities or their decision rules are known. Whereas the latter gives P_{E1} and P_{E2} directly from (4.2) and (4.3), the former does not: referees' decision rules should be determined by finding the Nash equilibrium in Game I, then P_{E1} and P_{E2} can be determined. At the Nash equilibrium, not only referee 1's decision rule but also referee 2 and 3's decision rules depend on referee 1's representation point. In other words, for payoffs which are functions of $\lambda_k^{(1)}$, $\lambda_k^{(2)}$, and $\lambda_k^{(3)}$, referee i searches for the optimal strategy along certain curve that is defined by the dependency between his representation points and $(\lambda_k^{(1)}, \lambda_k^{(2)}, \lambda_k^{(3)})$ in Game III, but he searches for the optimal strategy only along $\lambda_k^{(i)}$-axis in Game IV. Note that at a Nash equilibrium, any player does not benefit by changing his own strategy unilaterally. Since each referee in Game III consequently adjusts all of $\lambda_k^{(1)}$, $\lambda_k^{(2)}$, and $\lambda_k^{(3)}$ by changing his representation point, referees in Game III have more chances to find a better strategy profile than those in Game IV.

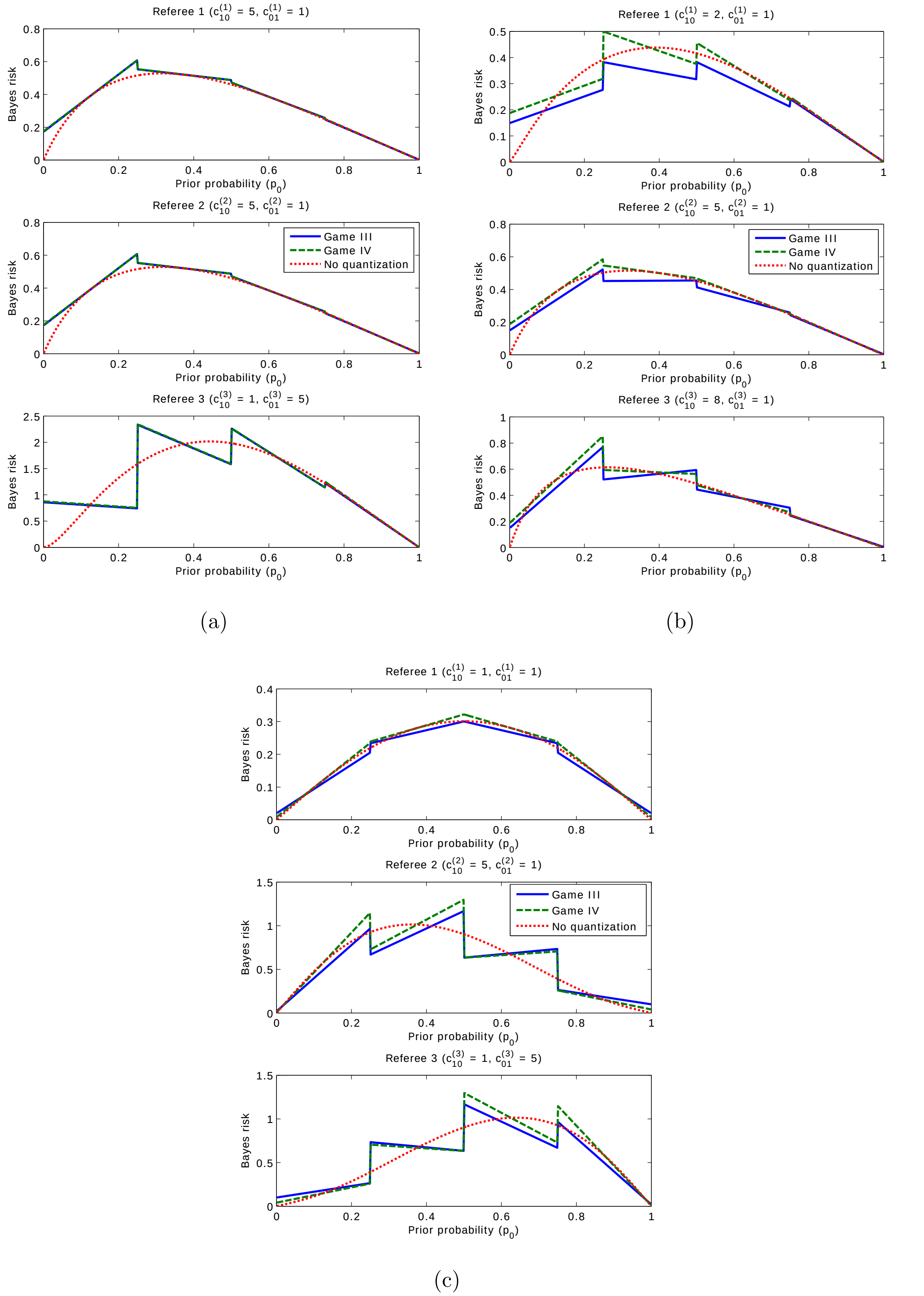


Figure 4-6. Comparison of conflicting referees' Bayes risks when they use the Nash equilibrium of representation points to when they use the Nash equilibrium of decision rules for $h_0 = 0$, $h_1 = 1$, and $\sigma = 1$.

4.3 Collaborating Referees

Collaborating referees is a generalized version of identical referees. We compare the performance of collaboration to that of conflict. We also explore the advantage of diversity in quantization rules for collaborating referees.

We say that referees are collaborating when they are attempting to a minimize Bayes risk function determined by mutual agreement. This function is formed by summing each referee's Bayes risk multiplied by weight w_i according to his power. In other words, the common risk is

$$\bar{R} = \sum_{i=1}^{3} w_i R_i,$$

where $w_i > 0$ for all $i \in \mathcal{I}$ and $\sum_{i=1}^{3} w_i = 1$. For example, $w_i = 1/3$ when all referees are of the same rank, and $\bar{R}$ becomes the average of Bayes risks of the referees. Even though each referee cannot minimize his own Bayes risk, minimizing the common risk has an effect on reducing each referee's Bayes risk because it is a part of the common risk.

The identical referees in Chapter 3 collaborate in the sense that they share one cost function, but here we do not constrain the categorization used by the referees to be the same. This makes it possible for the referees who share a common cost function to maximize their performance by using optimal diverse quantization rules.

4.3.1 Decision-Making Strategy

Consider referee i who has Bayes costs $c_{10}^{(i)}$ and $c_{01}^{(i)}$. He makes a decision on an object whose true prior proability is p_0 and quantized prior probability is $p_0^{(i)}$. His true Bayes risk is

$$c_{10}^{(i)} p_0 P_{E1} + c_{01}^{(i)} (1 - p_0) P_{E2},$$

but he thinks his Bayes risk is

$$R_i = c_{10}^{(i)} p_0^{(i)} P_{E1} + c_{01}^{(i)} (1 - p_0^{(i)}) P_{E2}.$$

Then the team of referees has a common risk

$$
\begin{aligned}
\bar{R} &= \sum_{i=1}^{3} w_i R_i \\
&= \left[\sum_{i=1}^{3} w_i c_{10}^{(i)} p_0^{(i)}\right] P_{E1} + \left[\sum_{i=1}^{3} w_i c_{01}^{(i)} (1 - p_0^{(i)})\right] P_{E2}.
\end{aligned}
$$

We can find the optimal decision rule that minimizes $\bar{R}$ in the same way as in the identical-referee case. This is possible because P_{E1} and P_{E2} depend not on local decisions but on their global decision and are the same for all referees even though they are not identical. According to the result in Section 3.2, all referees have the same optimal decision rule. Assuming an independent and identically distributed additive Gaussian noise $\mathcal{N}(0, \sigma^2)$, the optimal decision rule is

$$
y_i \underset{\hat{H}_i(y_i)=h_0}{\overset{\hat{H}_i(y_i)=h_1}{\gtrless}} \lambda,
$$

where λ is the unique solution of a nonlinear equation

$$
\lambda = \frac{h_1 + h_0}{2} + \frac{\sigma^2}{h_1 - h_0} \ln \frac{\left[\sum_{i=1}^{3} w_i c_{10}^{(i)} p_0^{(i)}\right] \left[Q(\frac{\lambda - h_0}{\sigma}) - Q(\frac{\lambda - h_0}{\sigma})^2\right]}{\left[\sum_{i=1}^{3} w_i c_{01}^{(i)} (1 - p_0^{(i)})\right] \left[Q(\frac{h_1 - \lambda}{\sigma}) - Q(\frac{h_1 - \lambda}{\sigma})^2\right]}.
$$

We compare results of collaboration to those of conflict in several examples in Figure 4-7. In the case of Figure 4-7a, the first two referees have the same Bayes costs; global decisions are highly likely to be determined by what referees 1 and 2 want since referee 3 rarely affects the global decisions when the referees conflict because the global decisions require only two referees' agreement. When they collaborate, on the contrary, since the characteristic (i.e., cost function) of referee 3 is considered in the common risk, he benefits by collaborating while the others make a loss with regard to Bayes risk.

When referees of similar characteristics make a team such as referees in the case of Figure 4-7b, collaboration is better than conflict because the characteristic of their common risk agrees with their own characteristics. Furthermore, in some cases like the case of Figure 4-7c, it depends on the prior probability of an observed object whether referees make a profit or loss by collaboration.

Note that cost function of referee 2 does not change in the three examples. For him, however, collaboration is always better than conflict in the first example, always worse in the second example, and sometimes better and elsewhere worse in the last example. This implies that the question of which one is better does not have a single answer; it depends on situations. Thus, we do not intend to argue that collaboration is better but want to see what happens when referees conflict or collaborate.

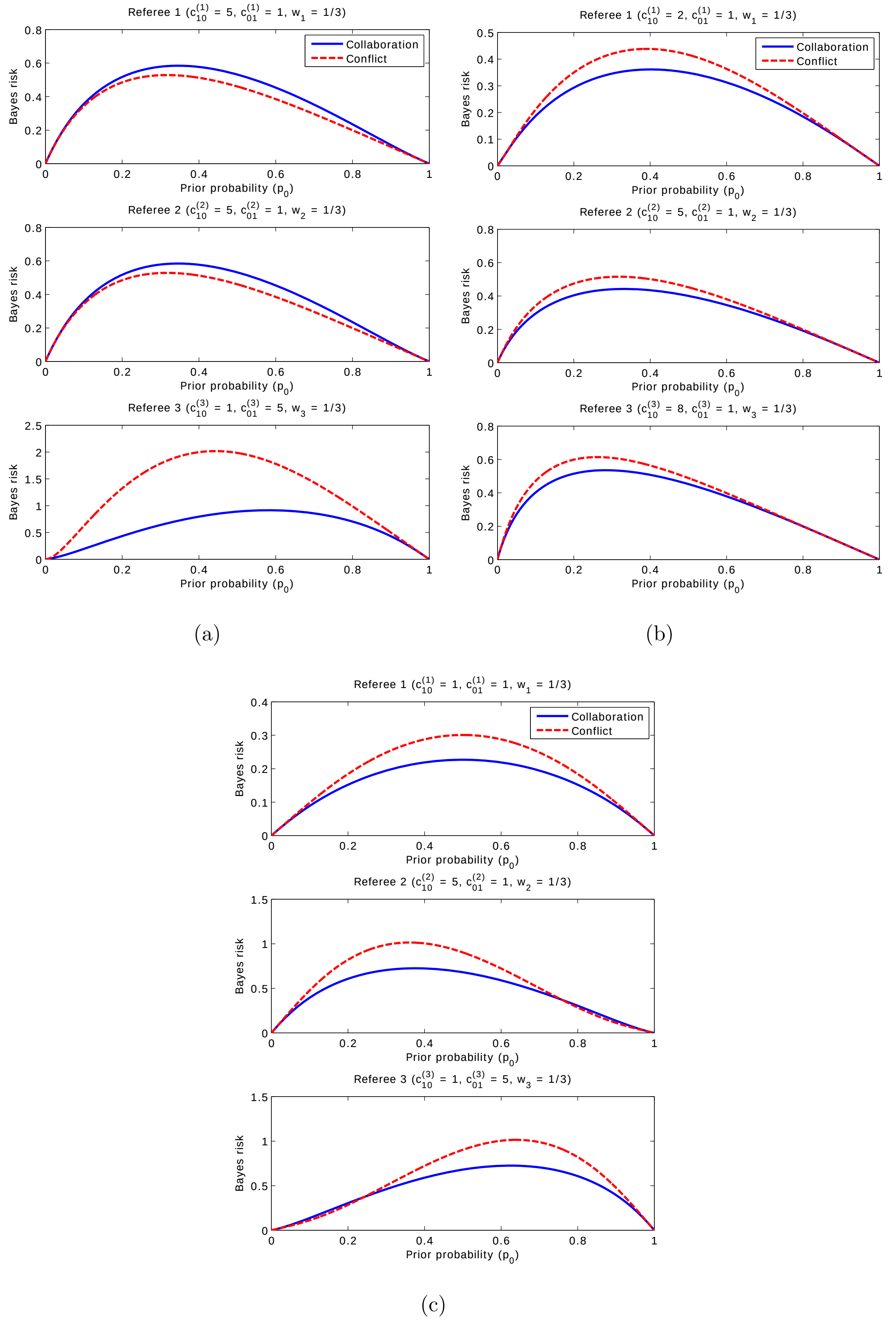


Figure 4-7. Comparison of Bayes risks in collaborating-referee cases to those in conflicting-referee cases. Referees use true prior probabilities.

4.3.2 Quantization Strategy - Using Diverse Quantizers

We know how referees collaborate to make a decision. Now we need to know how referees collaborate to categorize objects optimally. One method is using identical quantizers like in Section 3.2 since all referees share one common risk. We showed how to design such an optimal quantizer in Section 3.2. However, referees are able to collaborate even if they quantize prior probabilities differently, and thus we are free from the limitation that referees should use identical quantizers.

Consider three different quantizers for prior probabilities. Figure 4-8 shows one example of them; from the top, they are 2-point quantizers for referees 1, 2, and 3, respectively. In Figure 4-8, $a_k^{(i)}$ and $b_k^{(i)}$ respectively denote representation point and right endpoint of k-th region of the quantizer of referee i. Each quantizer divides the interval $[0, 1]$ into two partitions, but the whole quantization system divides it into four partitions: $[0, b_1^{(1)}]$, $(b_1^{(1)}, b_1^{(2)}]$, $(b_1^{(2)}, b_1^{(3)}]$, and $(b_1^{(3)}, 1]$. In general, three regular K-point quantizers can split the entire interval into at most $(3K - 2)$ partitions. Thus, it is possible for us to consider a virtual $(3K - 2)$-point quantizer that behaves exactly the same as the set of real quantizers like in Figure 4-9. Then we are faced with two problems: Does such a virtual quantizer exist? If it exists, how can we find the quantizer?

In order to answer the first question, let us introduce a set of virtual identical referees[1] who use the virtual quantizer in Figure 4-9. Suppose that their Bayes costs are c'_{10} and c'_{01}. In the first region $\mathcal{R}'_1$, mean Bayes risk of real referee i is

$$\begin{aligned}
&\int_{\mathcal{R}'_1} [c_{10}^{(i)} p_0 P_{E1} + c_{01}^{(i)} (1 - p_0) P_{E2}] f_{P_0}(p_0)\, dp_0 \\
&= \left[\int_{\mathcal{R}'_1} p_0 f_{P_0}(p_0)\, dp_0 \right] c_{10}^{(i)} P_{E1} + \left[\int_{\mathcal{R}'_1} (1 - p_0) f_{P_0}(p_0)\, dp_0 \right] c_{01}^{(i)} P_{E2}.
\end{aligned}$$

When the real referees collaborate with weight w_i, their mean common risk in the

[1] We call them virtual referees in order to distinguish them from real referees 1, 2, and 3.

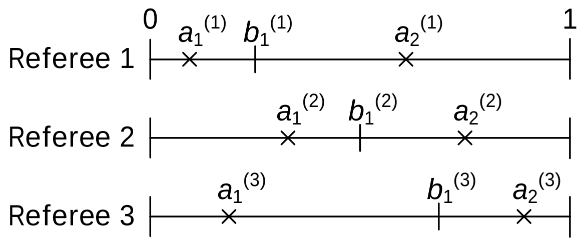


Figure 4-8. An example of diverse quantizers for prior probabilities.

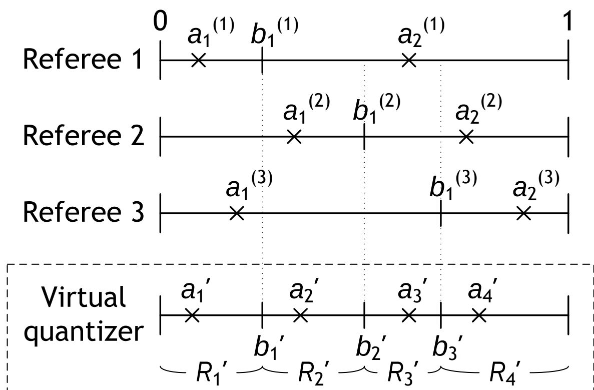


Figure 4-9. A virtual 4-point quantizer which is identical for all referees such that using it leads to the same results as using the real 2-point quantizers.

region becomes

$$\left[\int_{\mathcal{R}_1'} p_0 f_{P_0}(p_0)\,dp_0\right]\left[\sum_{i=1}^{3} w_i c_{10}^{(i)}\right] P_{E1} + \left[\int_{\mathcal{R}_1'} (1-p_0) f_{P_0}(p_0)\,dp_0\right]\left[\sum_{i=1}^{3} w_i c_{01}^{(i)}\right] P_{E2}. \quad (4.16)$$

In addition, mean Bayes risk of virtual referees in the region is

$$\left[\int_{\mathcal{R}_1'} p_0 f_{P_0}(p_0)\,dp_0\right] c_{10}' P_{E1} + \left[\int_{\mathcal{R}_1'} (1-p_0) f_{P_0}(p_0)\,dp_0\right] c_{01}' P_{E2}. \quad (4.17)$$

If $c_{10}' = \sum_{i=1}^{3} w_i c_{10}^{(i)}$ and $c_{01}' = \sum_{i=1}^{3} w_i c_{01}^{(i)}$, then (4.16) and (4.17) are the same for the same probabilities of errors, which means that the real referees' optimal decision rule is equal to the virtual referees'.[2] This argument is true for any regions $\mathcal{R}_k'$.

The next step is to investigate how to determine representation points for such

[2]Note that the real referees use the same decision rules, and so do the virtual referees. Even though the real referees are not identical, they are collaborating and their optimal decision rules are the same. Thus, we do not need to consider the real referees' using different decision rules.

decision rules. In the first region $\mathcal{R}'_1$, the Bayes risk of real referee i is

$$c_{10}^{(i)} a_1^{(i)} P_{E1} + c_{01}^{(i)} (1 - a_1^{(i)}) P_{E2},$$

and the common risk of real referees is

$$\left[\sum_{i=1}^{3} w_i c_{10}^{(i)} a_1^{(i)}\right] P_{E1} + \left[\sum_{i=1}^{3} w_i c_{01}^{(i)} (1 - a_1^{(i)})\right] P_{E2}. \tag{4.18}$$

By defining the quantizer of referee i as a function $q_i(p_0)$, we can write (4.18) as

$$\left[\sum_{i=1}^{3} w_i c_{10}^{(i)} q_i(p_0)\right] P_{E1} + \left[\sum_{i=1}^{3} w_i c_{01}^{(i)} (1 - q_i(p_0))\right] P_{E2},$$

where $p_0 \in \mathcal{R}'_1$.

The Bayes risk of the virtual referees is given by

$$c'_{10} a'_1 P_{E1} + c'_{01} (1 - a'_1) P_{E2}. \tag{4.19}$$

By comparing (4.18) to (4.19), we realize that the both real and virtual referees would use the same decision rules in the first region if their quantizers satisfy $\left[\sum_{i=1}^{3} w_i c_{10}^{(i)} a_1^{(i)}\right] = c'_{10} a'_1$ and $\left[\sum_{i=1}^{3} w_i c_{01}^{(i)} (1 - a_1^{(i)})\right] = c'_{01}(1 - a'_1)$. To summarize the above results, if a set of identical referees whose cost function is defined by $c'_{10} = \sum_{i=1}^{3} w_i c_{10}^{(i)}$ and $c'_{01} = \sum_{i=1}^{3} w_i c_{01}^{(i)}$ use the virtual $(3K-2)$-point quantizers whose representation point for the kth region is determined by the equation

$$\frac{\sum_{i=1}^{3} w_i c_{10}^{(i)} q_i(p_0)}{\sum_{i=1}^{3} w_i c_{10}^{(i)} (1 - q_i(p_0))} = \frac{c'_{10} a'_k}{c'_{01}(1 - a'_k)}, \tag{4.20}$$

then the real referees and virtual referees use the same decision rules for any $p_0 \in [0, 1]$. This result gives us the answers to the two questions: Yes, there exists such a virtual quantizer. We can design the virtual quantizer by using the same categorizations as the set of real quantizers and solving a set of linear equations about representation points.

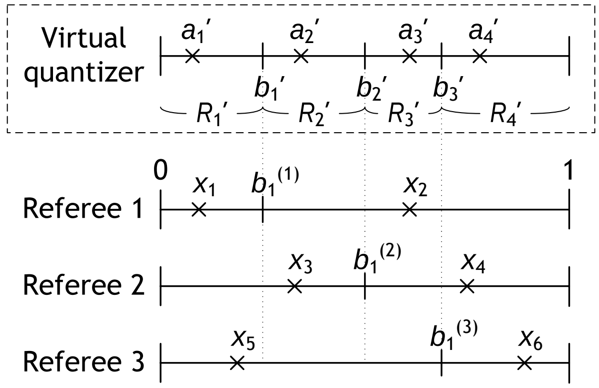


Figure 4-10. An example of diverse 2-point quantizers that are equivalent to the identical 4-point quantizer.

In the reverse direction, we are also able to find three real diverse quantizers which are equivalent to a virtual quantizer if we know the virtual quantizer. For convenience, we use $x_1, \ldots, x_6$ to denote the representation points of the real quantizers that we need to determine in Figure 4-10. In Figure 4-10, each region gives one equation like (4.20):

$$\alpha_{12}[\bar{c}_{10}^{(1)}x_1 + \bar{c}_{10}^{(2)}x_3 + \bar{c}_{10}^{(3)}x_5] = \alpha_{11}[\bar{c}_{01}^{(1)}(1-x_1) + \bar{c}_{01}^{(2)}(1-x_3) + \bar{c}_{01}^{(3)}(1-x_5)], \quad (4.21)$$

$$\alpha_{22}[\bar{c}_{10}^{(1)}x_2 + \bar{c}_{10}^{(2)}x_3 + \bar{c}_{10}^{(3)}x_5] = \alpha_{21}[\bar{c}_{01}^{(1)}(1-x_2) + \bar{c}_{01}^{(2)}(1-x_3) + \bar{c}_{01}^{(3)}(1-x_5)], \quad (4.22)$$

$$\alpha_{32}[\bar{c}_{10}^{(1)}x_2 + \bar{c}_{10}^{(2)}x_4 + \bar{c}_{10}^{(3)}x_5] = \alpha_{31}[\bar{c}_{01}^{(1)}(1-x_2) + \bar{c}_{01}^{(2)}(1-x_4) + \bar{c}_{01}^{(3)}(1-x_5)], \quad (4.23)$$

$$\alpha_{42}[\bar{c}_{10}^{(1)}x_2 + \bar{c}_{10}^{(2)}x_4 + \bar{c}_{10}^{(3)}x_6] = \alpha_{41}[\bar{c}_{01}^{(1)}(1-x_2) + \bar{c}_{01}^{(2)}(1-x_4) + \bar{c}_{01}^{(3)}(1-x_6)], \quad (4.24)$$

where $\alpha_{k1} \triangleq c'_{10}a'_k$, $\alpha_{k2} \triangleq c'_{01}(1-a'_k)$, $\bar{c}_{10}^{(i)} \triangleq w_i c_{10}^{(i)}$, and $\bar{c}_{01}^{(i)} \triangleq w_i c_{01}^{(i)}$. We can simplify (4.21)-(4.24) as the following nice matrix form:

$$\begin{bmatrix} \beta_1^{(1)} & 0 & \beta_1^{(2)} & 0 & \beta_1^{(3)} & 0 \\ 0 & \beta_2^{(1)} & \beta_2^{(2)} & 0 & \beta_2^{(3)} & 0 \\ 0 & \beta_3^{(1)} & 0 & \beta_3^{(2)} & \beta_3^{(3)} & 0 \\ 0 & \beta_4^{(1)} & 0 & \beta_4^{(2)} & 0 & \beta_4^{(3)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^3 \alpha_{11}\bar{c}_{01}^{(i)} \\ \sum_{i=1}^3 \alpha_{21}\bar{c}_{01}^{(i)} \\ \sum_{i=1}^3 \alpha_{31}\bar{c}_{01}^{(i)} \\ \sum_{i=1}^3 \alpha_{41}\bar{c}_{01}^{(i)} \end{bmatrix}, \quad (4.25)$$

where $\beta_k^{(i)} \triangleq \alpha_{k2}\bar{c}_{10}^{(i)} + \alpha_{k1}\bar{c}_{01}^{(i)}$.

However, there can be infinitely many solutions that satisfy (4.25) because the first matrix does not have full rank. Since the number of equations is generally less than the number of variables (i.e., the total number of representation points of real quantizers) by at least two, there is no unique solution. Instead, we obtain several conditions about the representation points: First, representation points are quantized prior probabilities and they should be lying within $[0, 1]$. Second, we want to design regular quantizers so each representation point should be lying within the region that is represented by the point. In the example of Figure 4-10, the conditions give us the following:

$$\begin{bmatrix} 0 \\ b'_1 \\ 0 \\ b'_2 \\ 0 \\ b'_3 \end{bmatrix} \prec \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \prec \begin{bmatrix} b'_1 \\ 1 \\ b'_2 \\ 1 \\ b'_3 \\ 1 \end{bmatrix}, \tag{4.26}$$

where $A \prec B$ means that A is smaller than B elementwise. We can write (4.25) as

$$\begin{bmatrix} 0 & \beta_1^{(2)} & 0 & \beta_1^{(3)} \\ \beta_2^{(1)} & \beta_2^{(2)} & 0 & \beta_2^{(3)} \\ \beta_3^{(1)} & 0 & \beta_3^{(2)} & \beta_3^{(3)} \\ \beta_4^{(1)} & 0 & \beta_4^{(2)} & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -\beta_1^{(1)} & 0 & \sum_{i=1}^3 \alpha_{11}\bar{c}_{01}^{(i)} \\ 0 & 0 & \sum_{i=1}^3 \alpha_{21}\bar{c}_{01}^{(i)} \\ 0 & 0 & \sum_{i=1}^3 \alpha_{31}\bar{c}_{01}^{(i)} \\ 0 & -\beta_4^{(3)} & \sum_{i=1}^3 \alpha_{41}\bar{c}_{01}^{(i)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_6 \\ 1 \end{bmatrix}, \tag{4.27}$$

or

$$\begin{bmatrix} x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 & \beta_1^{(2)} & 0 & \beta_1^{(3)} \\ \beta_2^{(1)} & \beta_2^{(2)} & 0 & \beta_2^{(3)} \\ \beta_3^{(1)} & 0 & \beta_3^{(2)} & \beta_3^{(3)} \\ \beta_4^{(1)} & 0 & \beta_4^{(2)} & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\beta_1^{(1)} & 0 & \sum_{i=1}^3 \alpha_{11}\bar{c}_{01}^{(i)} \\ 0 & 0 & \sum_{i=1}^3 \alpha_{21}\bar{c}_{01}^{(i)} \\ 0 & 0 & \sum_{i=1}^3 \alpha_{31}\bar{c}_{01}^{(i)} \\ 0 & -\beta_4^{(3)} & \sum_{i=1}^3 \alpha_{41}\bar{c}_{01}^{(i)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_6 \\ 1 \end{bmatrix}. \tag{4.28}$$

Thus, we have the following inequalities of x_1 and x_6 besides $0 < x_1 < b'_1$ and

$b'_3 < x_6 < 1$:

$$\begin{bmatrix} b'_1 \\ 0 \\ b'_2 \\ 0 \end{bmatrix} \prec \begin{bmatrix} 0 & \beta_1^{(2)} & 0 & \beta_1^{(3)} \\ \beta_2^{(1)} & \beta_2^{(2)} & 0 & \beta_2^{(3)} \\ \beta_3^{(1)} & 0 & \beta_3^{(2)} & \beta_3^{(3)} \\ \beta_4^{(1)} & 0 & \beta_4^{(2)} & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\beta_1^{(1)} & 0 & \sum_{i=1}^3 \alpha_{11}\bar{c}_{01}^{(i)} \\ 0 & 0 & \sum_{i=1}^3 \alpha_{21}\bar{c}_{01}^{(i)} \\ 0 & 0 & \sum_{i=1}^3 \alpha_{31}\bar{c}_{01}^{(i)} \\ 0 & -\beta_4^{(3)} & \sum_{i=1}^3 \alpha_{41}\bar{c}_{01}^{(i)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_6 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 1 \\ b'_2 \\ 1 \\ b'_3 \end{bmatrix} \tag{4.29}$$

All we have to do is to find a valid pair of (x_1, x_6) that satisfies (4.29) and compute other variables from (4.28). There are still infinitely many solutions but any of them makes the perfect three real quantizers.

Up to now, we discussed the relation and transformation between real diverse K-point quantizers and virtual identical $(3K-2)$-point quantizers. From now on, we show how the discussion helps us optimize real diverse quantizers when referees are collaborating. We showed that there exists a system consisting of virtual identical referees and a virtual $(3K-2)$-point quantizer such that for any p_0, the referees' decision rules are the same as those of the real collaborating referees using diverse K-point quantizers in a true system. Since their decision rules are the same, the mean Bayes risk of the virtual referees is equal to the mean common risk of the real referees. Thus, the diverse K-point quantizers can achieve as good performance as the virtual identical $(3K-2)$-point quantizers can and vice versa. In other words, if the virtual identical $(3K-2)$-point quantizers are optimal for the virtual referees then the real diverse K-point quantizers are also optimal for the real referees.

We propose to design the best diverse K-point quantizers from optimized identical $(3K-2)$-point quantizers. Optimizing diverse quantizers directly is very complicated but determining the best identical quantizers is easy, and we already know how to do that from Section 3.3.

Algorithm II: Design of optimal diverse quantizers

1) We are given three referees with Bayes costs $c_{10}^{(i)}$ and $c_{01}^{(i)}$ and weights w_i, $i = 1, 2, 3$.

2) Consider three virtual identical referees with cost functions $c'_{10} = \sum_{i=1}^{3} w_i c_{10}^{(i)}$ and $c'_{01} = \sum_{i=1}^{3} w_i c_{01}^{(i)}$.

3) Design the best $(3K-2)$-point quantizer for the virtual referees using the Lloyd-Max algorithm in Section 3.2.

4) Determine the optimal endpoints of three diverse K-point quantizers.

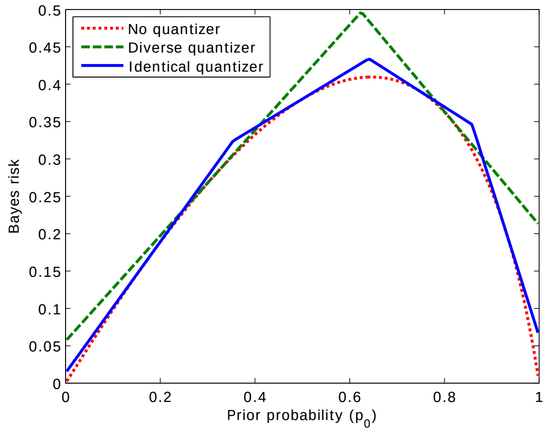
   (i) The quantizer for the virtual referees has $3K - 3$ endpoints except 0 and 1. Distribute them into three sets $B_1, \ldots, B_3$ so that the cardinality of each set becomes $K - 1$. The elements of B_i become the endpoints of K-point quantizer for referee i.

   (ii) Find the valid $B_1, \ldots, B_3$ in the sense that there exists a pair of variables that satisfies inequalities like (4.29), and determine values of the variables.

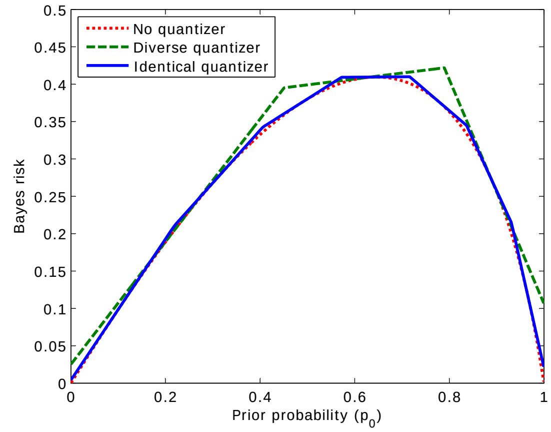
5) Compute the remaining variables (or representation points) from their relationships like (4.28).

Note that there are infinitely many pairs that satisfy (4.29) at Step 4)-(ii). However, choosing any pair will result in the same performance with respect to the common risk because all of the resulting quantizers are mapped to the same virtual quantizer.

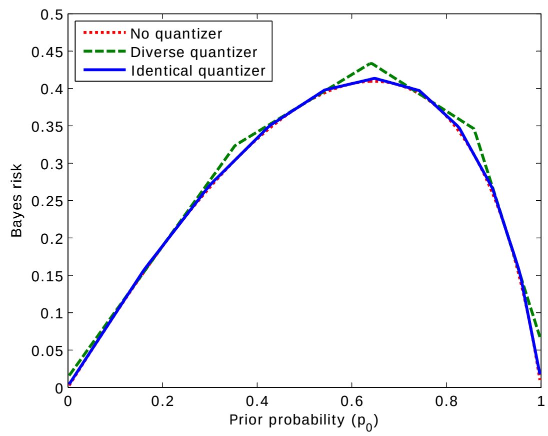
Figures 4-11 and 4-12 show resulting Bayes risks when referees use optimal diverse quantizers. Note that we make no assumptions about referees' Bayes costs. We can apply Algorithm II to design optimal quantization rules for referees having the same cost function. By introducing the concept of collaboration, we make it possible for such referees to use different quantization rules, which was not possible in Section 3.3. Figure 4-11 shows that using diverse quantizers is the better choice even for



(a)



(b)



(c)

Figure 4-11. Quantizers for uniformly distributed P_0, $h_0 = 0$, $h_1 = 1$, $\sigma = 1$, and collaborating referees who have the same Bayes costs $c_{10}^{(i)} = 1$ and $c_{01}^{(i)} = 4$. Bayes risks when collaborating referees use identical quantization rules and diverse quantization rules are plotted for (a) $K = 2$, (b) $K = 3$, and (c) $K = 4$.

referees with identical cost functions. Note that the set of diverse 2-point quantizers (represented by the solid line in Figure 4-11a) is as good as identical 4-point quantizers (represented by the dashed line in Figure 4-11c) for any p_0, which supports our discussion about relation between real diverse K-point quantizers and virtual identical $(3K-2)$-point quantizers. It is a very positive result because we always have incentive to use diverse quantizers. It is also an interesting result because optimal decision rules are identical but optimal categorization rules are not identical for a team of referees sharing the same cost function.

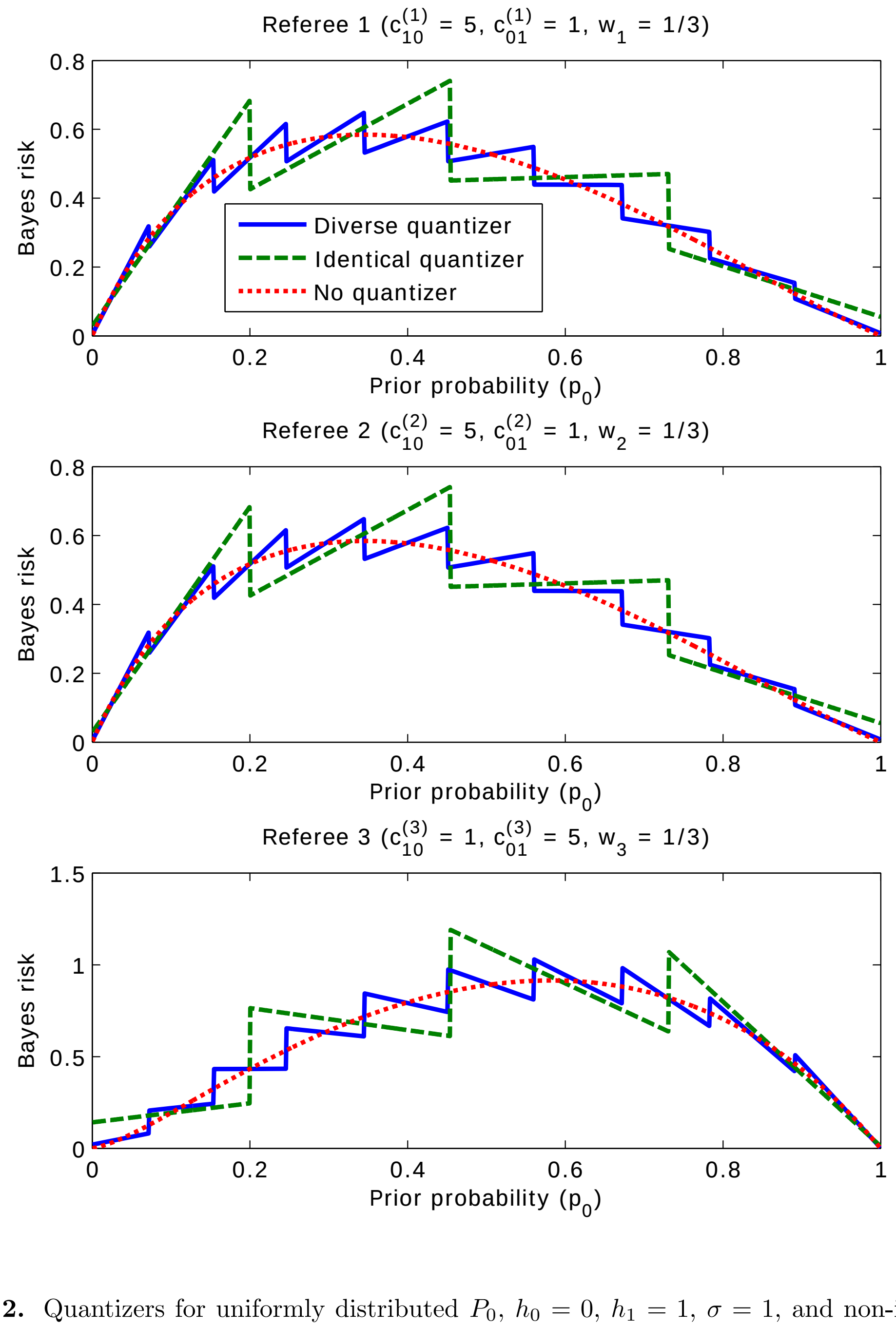


Figure 4-12. Quantizers for uniformly distributed P_0, $h_0 = 0$, $h_1 = 1$, $\sigma = 1$, and non-identical referees.

Chapter 5

Conclusion

In this thesis, we explored the questions regarding quantization of prior probabilities in Bayesian group decision-making. In the single-referee case [6], the main problem was to optimize the referee's quantization rule for the minimum mean Bayes risk. On the other hand, in the three-referee case, various issues arise not only from optimizing quantization rules but also from determining decision rules. This is because the referees mutually affect one another.

First of all, in Chapter 3, we consider the identical-referee case. Operation region of the three-referee model shows that identical referees cannot do better than using the same decision rules and cannot do worse than one referee does in terms of probabilities of errors. It is a reasonable result that using the same decision rules is optimal for identical referees in the sense that they are under the same circumstances for everything, such as cost functions, quantized prior probabilities, and density functions of noises. Using the same decision rules makes the problem simple: the Bayes risk can be defined as a function of one variable, and an optimal decision threshold is a global minimum of the function.

We assume that identical referees use the identical quantization rules to keep them identical. From the fact that the identical-three-referee model can be converted into an equivalent single-referee model, the results obtained for the single-referee model in [6] can be applied to the identical-three-referee model, and the centroid and nearest neighbor conditions are derived. Identical referees' quantization rules are optimized by the Lloyd-Max algorithm that alternates the conditions.

In Chapter 4, we consider non-identical referees to make a team. A referee's decision rule may help or hurt the other referees because all referees may have different cost functions. Game-theoretic methodologies are useful to analyze their behaviors. It is impossible to find the best decision threshold that minimizes all referees' Bayes risks since no referee has a dominant strategy. On the other hand, a Nash equilibrium turns out to exist for any cost functions and quantized prior probabilities. Furthermore, following a Nash equilibrium is one of the safest strategies for all referees in the sense that they can predict it and predict that their opponents can predict it. Thus, we assume that they determine their decision threshold as the Nash equilibrium. Note that the optimal decision threshold in the identical-referee case is also a Nash equilibrium for the identical referees.

The quantizer optimization problem has an issue about complexity. For non-identical referees, it is not possible to derive centroid and nearest neighbor conditions similar to the identical-referee case because of the dependency among referees. Thus optimization of three K-point quantizers has $3(2K-1)$ degrees of freedom. Furthermore, we need to consider the structure of the quantizers, which has $\frac{(3(K-1))!}{(K-1)!(K-1)!(K-1)!}$ possible scenarios. In order to decrease the computational complexity, we optimize the quantizers under the assumption that the referees use the same fixed categorization for their quantizers. Two methods for optimization are introduced: adjusting representation points of each category and finding optimal decision thresholds for each category. The results show that the former method leads to a better set of quantization rules.

In addition, we allow non-identical referees to collaborate with each other, which is a generalized case of the identical-referee case. Virtual identical referees and their virtual identical quantizers are derived from the similarity between the collaborating-referee case and the identical-referee case. By investigating the virtual referees and quantizers, we discover not only that collaborating referees' K-point diverse quantizers can have as good performance as $(3K-2)$-point identical quantizers but also how we design such diverse quantizers. We can apply this result to the identical-referee model in Chapter 3. Our main finding is that using the identical decision rules is optimal for the identical referees, but using the identical quantization rules is not optimal for them.

In summary, decision-makers may make a better decision when they have different categorizations for an object than when they have the same categorizations even though they have the same preference. We have shown that there is a definite improvement in using diverse categorizations. This thesis also presents the formulation of quantization of prior probabilities in Bayesian group decision-making in the game-theoretic point of view. This formulation helps us understand the problems and enables us to find the ways for decision-makers to determine their decision and quantization rules.

Bibliography

[1] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: John Wiley & Sons, 1968. 13, 22, 59

[2] J. Lawson, "An elementary theory of surveillance," Naval Electronics System Command, Tech. Rep., Aug. 1978. 13

[3] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 2, pp. 187–198, Mar. 1982. 13

[4] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 995–1000, Sep. 1989. 13

[5] P. K. Varshney, *Distributed Detection and Data Fusion.* New York: Springer-Verlag, 1997. 13, 15

[6] 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. 14, 19, 25, 47, 48, 81

[7] R. C. Rose, B. H. Juang, and C. H. Lee, "A training procedure for verifying string hypotheses in continuous speech recognition," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, May 1995, pp. 281–284. 19

[8] G. Guyatt, R. Jaeschke, N. Heddle, D. Cook, H. Shannon, and S. Walter, "Basic statistics for clinicians: 1. Hypothesis testing," *Canadian Medical Association Journal*, vol. 152, no. 1, pp. 27–32, Jan. 1995. 19

[9] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998. 23

[10] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press, 1944. 28

[11] A. Downs, *An Economic Theory of Democracy.* New York: Harper, 1992. 28

[12] M. Smith, *Evolution and the Theory of Games.* New York: Cambridge University Press, 1982. 28

[13] R. Gibbons, *Game Theory for Applied Economists.* Princeton, NJ: Princeton University Press, 1992. 28

[14] B. D. Bernheim, "Rationalizable strategic behavior," *Econometrica*, vol. 52, no. 4, pp. 1007–1028, Jul. 1984. 30

[15] D. G. Pearce, "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, vol. 52, no. 4, pp. 1029–1050, Jul. 1984. 30

[16] J. F. Nash, "Equilibrium points in n-person games," *Proc. Natl. Acad. Sci. USA*, vol. 36, no. 1, pp. 48–49, Jan. 1950. 31

[17] I. L. Glicksberg, "A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points," *Proceedings of the American Mathematical Society*, vol. 3, no. 1, pp. 170–174, Feb. 1952. 31, 58

[18] K. Fan, "Fixed-point and minimax theorems in locally convex topological linear spaces," *Proc. Natl. Acad. Sci. USA*, vol. 38, no. 2, pp. 121–126, Feb. 1952. 31, 58

[19] G. Debreu, "A social equilibrium existence theorem," *Proc. Natl. Acad. Sci. USA*, vol. 38, no. 10, pp. 886–893, Oct. 1952. 31, 58

[20] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Norwell, MA: Kluwer Academic Publishers, 1992. 48

[21] 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. 48