Large Deviations Approaches to Performance Analysis of Distributed Detection Systems

by

Po-Ning Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland at College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1994

Advisory Committee:

Professor Adrian Papamarcou, Chairman/Advisor
Professor Thomas E. Fuja
Professor Abram K. Kagan
Professor Armand M. Makowski
Professor Prakash Narayan

Abstract

Title of Dissertation: Large Deviations Approaches to Performance Analysis of Distributed Detection Systems

Po-Ning Chen, Doctor of Philosophy, 1994

Dissertation directed by: Professor Adrian Papamarcou
Department of Electrical Engineering

This dissertation studies the performance of the distributed detection systems by means of the large deviation techniques. Two distinct models are considered.

In the first model, the error performance is investigated as the number of sensors tends to infinity. It is assumed that the i.i.d. sensor data are quantized locally into m-ary messages and transmitted to the fusion center for binary hypothesis testing. The boundedness of the second moment of the post-quantization log-likelihood ratio is examined in relation to the asymptotic error exponent. It is found that when that second moment is unbounded, the Neyman-Pearson error exponent can become a function of the test level; whereas the Bayes error exponent remains, as previously conjectured by Tsitsiklis, unaffected. Large deviations techniques are also employed to show that in Bayes testing, the equivalence of absolutely optimal and best identical-quantizer systems is not limited to error exponents, but extends to the actual Bayes error probabilities up to a

multiplicative constant. Under a fairly general assumption, the same conclusion can be drawn for the Neyman-Pearson testing.

In the second model, a distributed detection system is considered in which two sensors and a fusion center jointly process the output of a random data source. It is assumed that the null and alternative distributions are spatially correlated Gaussian, differing in the mean; thus the random source is either noise only or a deterministic signal plus noise.

Two issues are considered. The first is whether contiguous marginal likelihood ratio quantizers are optimal. It is shown that this is not true in general, and a sufficient condition is obtained under in the case of a single observation per sensor. It is also demonstrated that in the case of larger samples, marginal likelihood ratios may not be sufficient statistics for local quantization.

The second issue is the performance gap between a centralized and a distributed system for detecting a signal based on N samples per sensor, each sample being independently corrupted by additive Gaussian noise. It is shown that if the number of quantization levels grows with sample size at a suitable polynomial rate, the ratio of error probabilities of the two systems is bounded by a multiplicative constant.

Large Deviations Approaches to Performance Analysis of Distributed Detection Systems

by

Po-Ning Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland at College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1994

Advisory Committee:

Professor Adrian Papamarcou, Chairman/Advisor
Professor Thomas E. Fuja
Professor Abram K. Kagan
Professor Armand M. Makowski
Professor Prakash Narayan

Acknowledgements

My dissertation advisor, Dr. Adrian Papamarcou, provided immeasurable assistance in the development of this research. Thanks are also to my family for their full support during my study.

Table of Contents

List of Tables

List of Figures

Large Deviations Approaches to Performance Analysis of Distributed Detection Systems

Po-Ning Chen

April 25, 2008

This comment page is not part of the dissertation.

Typeset by LaTeX using the `dissertation` class by Pablo A. Straub, University of Maryland.

Chapter 1

Introduction

A distributed or decentralized detection system consists of a number of observers (or data collecting units) and one or more data fusion centers. Each observer is coupled with a local data processor and communicates with the data fusion center through network links. The fusion center combines all information received from the observers and attempts to classify the source of the observations into one of finitely many categories.

A distributed detection system offers significant benefits compared to a centralized, or non-distributed, one in which the observer and the data fusion center are combined in one unit. This is because by employing several observers and local data processors, one can distribute the burden of information processing from the decision-making center to the local observers, thereby reducing the operational complexity of the decision-making system. Furthermore, the presence of local data processors enables the system to make decisions locally, and thus partial reliability is ensured in the event of a malfunction in the fusion center. Most importantly, the amount (and cost) of information exchange between the local observers and the fusion centers is reduced significantly. For these reasons, distributed detection systems are attractive alternatives to conventional centralized signal detection systems for processing large quantities of data collected by sensor networks. It should be noted, however, that the advantages of distributed detection are partially offset by a degradation in detection performance.

Structurally, distributed detection systems are more complex than centralized ones. Accordingly, the optimal design of a distributed detection system is significantly harder: it entails the joint optimization of local processors and fusion centers. The literature on distributed detection has focused on different facets of this optimization problem; some of the contributions related to the theme of this dissertation are listed below.

Tenney and Sandell [28] introduced distributed detection using a fixed processor (rule) at the fusion center, and addressed the optimization of the local processors. Chair and Varshney [37] treated the local processors as fixed, and optimized the fusion processor. Reibman and Nolte [1] combined these two approaches in a study of the overall optimization problem. Willett and Warren [34] considered sensor networks with i.i.d. observations and discussed the existence of symmetric solutions (characterized by identical local processors). They also explored necessary conditions for deterministic (non-randomized) local processors [35] to be optimal. The possibility of existence of non-symmetric solutions to sensor networks with i.i.d. observations has also been studied by other researchers [30] [12]. Tsitsiklis [30] investigated the error performance of such networks when the number of sensors is large, and showed that symmetric solutions are asymptotically optimal in the sense of error exponent. For a sensor network of fixed size with correlated local observations, Ahlswede and Csiszar [25] considered the error exponent as the number of local observations tends to infinity. Variations on this setup were also considered in [26, 27, 22].

In this dissertation, we use techniques drawn primarily from the theory of large deviations in order to study the error performance of optimal distributed detection systems. We consider two different models.

In the first model, i.i.d. sensor data are quantized locally into m-ary messages and transmitted to the fusion center for binary hypothesis testing. Our aim is to compare the error performance of the (absolutely) optimal solution to that of the best symmetric (identical-quantizer) solution as the number of sensors tends to infinity. Since the latter solution is much easier to obtain than the former, it is natural to seek an estimate of the performance loss resulting from using identical sensors. By using a refinement (due to Esseen) of the central limit theorem for independent but not identically distributed summands, we show that the (Bayes) error probability of the best identical-quantizer system is at most a constant multiple of that achieved by the absolutely optimal system. This is true as long as the two hypotheses are mutually absolutely continuous. A similar result is true for Neyman-Pearson testing at a fixed test level, although further assumptions, such as boundedness of second moments of the post-quantization log-likelihood ratio, are needed in that case.

The second model entails a two-sensor distributed system for detecting a known signal in additive Gaussian noise exhibiting spatial correlation. In the setup where each sensor draws one local observation and transmits a quantized message to the fusion center, a sufficient condition is found under which the optimal system can be implemented using marginal likelihood ratio tests. Under the same condition, we show that a symmetric sensor network (characterized by data statistics that are spatially permutation-invariant) admits a symmetric optimal solution. We also investigate the error performance of the system as the number N of local observations per sensor goes to infinity. We find that if the number of quantization levels increases with N at a suitable polynomial rate, the error probability of the system is at most a fixed multiple of that achieved by the centralized system.

Accordingly, the material of this dissertation is arranged into two parts, the first part consisting of Chapters 1, 2, and 3, and the second consisting of Chapter 4. Some general facts on distributional distance, error exponents and quantization are covered in Section 1.2. Chapter 2 contains a brief analysis of Neyman-Pearson testing under the assumption that the second moment of the post-quantization log-likelihood is bounded, followed by an examination of models that violate this assumption. The condition under which the Neyman-Pearson error performance of the best identical-quantizer system is at most a fixed multiple of the absolutely optimal system is also established in this chapter. Chapter 3 is devoted to the discussion of Bayes testing, leading to the result that the ratio of the Bayes error probabilities of the two aforementioned systems is confined between two constants. Results for the second model (a two-sensor distributed system for a known signal in additive Gaussian noise) are covered in Chapter 4. Some final comments appear in Chapter 5. Appendix A shows some simulation results on the ratio of the Bayesian error probabilities of the two systems. Appendix B contains proofs of two lemmas from Chapter 3, while in Appendix C it is demonstrated that the ratio of the Bayesian error performances of the two systems need not tend to unity. Some supplements to Chapter 4 are placed in Appendix D.

The remainder of this chapter will serve as an introduction to the first part of the dissertation.

1.1 Introduction

Perhaps the most common architecture in distributed (or decentralized) detection is the parallel feedforward system $\mathcal{S}_n$ depicted in Fig. 1.1. It consists of n geographically dispersed sensors, noiseless one-way communication links, and a fusion center. Each sensor makes an observation (denoted by Y_i) of a random source, quantizes Y_i into an m-ary message $U_i = g_i(Y_i)$, and then transmits U_i to the fusion center. Upon receipt of $(U_1, \ldots, U_n)$, the fusion center makes a global decision $\mathcal{D}(U_1, \ldots, U_n)$ about the nature of the random source.

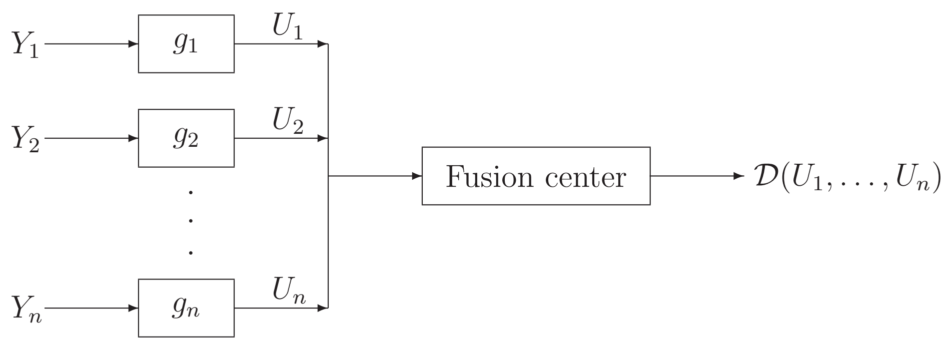

Figure 1.1: Distributed detection in $\mathcal{S}_n$.

The optimal design of $\mathcal{S}_n$ entails choosing quantizers $g_1, \ldots, g_n$ and a global decision rule $\mathcal{D}$ so as to optimize a given performance index. In this dissertation, we consider binary hypothesis testing under the (classical) Neyman-Pearson and Bayesian formulations. The first formulation dictates minimization of the type II error probability subject to an upper bound on the type I error probability; while the second stipulates minimization of the Bayes error probability, computed according to the prior probabilities of the two hypotheses.

The joint optimization of entities $g_1, \ldots, g_n$ and $\mathcal{D}$ in $\mathcal{S}_n$ is a hard computational task [32], except in trivial cases (such as when the observations Y_i lie in a set of size no greater than m). The complexity of the problem can only be reduced by introducing additional statistical structure in the observations. For example, it has been shown that whenever $Y_1, \ldots, Y_n$ are independent given each hypothesis, an optimal solution can be found in which $g_1, \ldots, g_n$ are threshold-type functions of the local likelihood ratio (possibly with some randomization for Neyman-Pearson testing). These results, which were developed in—among others—[28, 33, 31, 35] are directly relevant to the discussion in this dissertation. Still, we should note that optimization of $g_1, \ldots, g_n$ over the class of threshold-type likelihood-ratio quantizers is prohibitively complex when n is large.

Of equal importance are situations where the statistical model exhibits spatial symmetry in the form of permutation invariance with respect to the sensors. A natural question to ask in such cases is whether a symmetric optimal solution exists in which the quantizers g_i are identical; if so, then the optimal system design is considerably simplified. The answer is clearly negative for cases where sensor observations are highly dependent; as an extreme example, take $Y_1 = \ldots = Y_n = Y$ with probability 1 under each hypothesis, and note that any two identical quantizers lead to a redundancy. If one heuristically interprets dependence as the existence of a common "core" in all observations, then the previous example would suggest that this core is not best handled by identical quantizers. Yet as the core shrinks and the data become independent (and identically distributed), it also becomes conceivable that a symmetric solution consisting of identical quantizers may exist. Surprisingly, this is not always true; counterexamples to this symmetry conjecture were first given in [30], and were also explored in [34] and [12]. We now know that this asymmetry is by no means a pathology, and arises quite frequently in numerical simulations.

The initial goal of our inquiry was the characterization of those i.i.d. source models for which the optimal system $\mathcal{S}_n$ consists of identical quantizers. Early results—both analytical and numerical— showed that the presence of nonidentical quantizers in the optimal $\mathcal{S}_n$ is neither sporadic nor predictable. As a result, we found the determination of necessary and sufficient conditions for the existence of symmetric solutions quite intractable (notwithstanding the contributions of [34], where certain necessary conditions were derived). We subsequently focused on the study of asymptotically (large n) optimal systems, thereby continuing the inquiry initiated by Tsitsiklis in [30].

The general problem is as follows. System $\mathcal{S}_n$ is used for testing $H_0 : P$ versus $H_1 : Q$, where P and Q are one-dimensional marginals of the i.i.d. data $Y_1, \ldots, Y_n$. As n tends to infinity, both the minimum type II error probability $\beta_n^*(\alpha)$ (as function of the type I error probability bound α) and the Bayes error probability $\gamma_n^*(\pi)$ (as function of the prior probability π of H_0) vanish at an exponential rate. It thus becomes legitimate to adopt a measure of asymptotic performance based on the error exponents

$$\begin{aligned} e_{\mathrm{NP}}^*(\alpha) &\triangleq \lim_{n\to\infty} -\frac{1}{n}\log \beta_n^*(\alpha) \\ e_{\mathrm{B}}^*(\pi) &\triangleq \lim_{n\to\infty} -\frac{1}{n}\log \gamma_n^*(\pi) \ . \end{aligned}$$

It was shown in [30] that, under certain assumptions on the hypotheses P and Q, it is possible to achieve the same error exponents using identical quantizers.

Thus if $\beta_n^\diamond(\alpha)$, $\gamma_n^\diamond(\pi)$, $e_{\rm NP}^\diamond(\alpha)$ and $e_{\rm B}^\diamond(\pi)$ are the counterparts of $\beta_n^*(\alpha)$, $\gamma_n^*(\pi)$, $e_{\rm NP}^*(\alpha)$ and $e_{\rm B}^*(\pi)$ under the constraint that the quantizers $g_1, \ldots, g_n$ are identical, then

$$(\forall \alpha \in (0,1)) \qquad e_{\rm NP}^\diamond(\alpha) = e_{\rm NP}^*(\alpha)$$

and

$$(\forall \pi \in (0,1)) \qquad e_{\rm B}^\diamond(\pi) = e_{\rm B}^*(\pi) .$$

(Of course, for all n, $\beta_n^\diamond(\alpha) \geq \beta_n^*(\alpha)$ and $\gamma_n^\diamond(\pi) \geq \gamma_n^*(\pi)$.) This result provides some justification for restricting attention to identical quantizers when designing a system consisting of a large number of sensors.

Our work revolves around two issues that remained open in [30]. The first issue is the exact asymptotics of the minimum error probabilities achieved by the absolutely optimal and best identical-quantizer systems. Note that equality in the error exponents of $\gamma_n^*(\pi)$ and $\gamma_n^\diamond(\pi)$ does not in itself guarantee that for any given n, the values of $\gamma_n^*(\pi)$ and $\gamma_n^\diamond(\pi)$ are in any sense close. In particular, the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ may vanish at a subexponential rate, and thus the best identical-quantizer system may be vastly inferior to the absolutely optimal system. (The same argument can be made for $\beta_n^*(\alpha)/\beta_n^\diamond(\alpha)$).

In examining this problem, we first performed numerical simulations of Bayes testing in $\mathcal{S}_n$ (See Appendix A). These showed that the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ is (apparently) bounded from below by a positive constant which is, in many cases, reasonably close to unity. In this dissertation we substantiate these findings by using large deviations techniques to prove that $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ is, indeed, always bounded from below (it is, of course, upper bounded by unity). For Neyman-Pearson testing, we find that an additional regularity condition is required in order for the ratio $\beta_n^*(\alpha)/\beta_n^\diamond(\alpha)$ to be lower-bounded in n. In either case, we conclude that the optimal system essentially consists of "almost identical" quantizers, and is thus only marginally different from the best identical-quantizer system.

The second issue stemming from [30] concerns the assumption under which identical-quantizer systems were shown to attain the optimal error exponent. This assumption stipulates that the variance of the log-likelihood ratio at the quantizer output be bounded over all quantizer mappings g; where this variance is computed w.r.t. P in the case of Neyman-Pearson testing, and w.r.t. both P and Q in the case of Bayes testing.

It was conjectured in [29] that the aforementioned boundedness assumption may be superfluous in the case of Bayes testing. We are able to give proof to this conjecture. On the other hand, we have found that the boundedness assumption is significant in Neyman-Pearson testing. In its absence, the finite-sample exponent $(-1/n)\log\beta_n^*(\alpha)$ exhibits three distinct modes, including one in which the lim inf and lim sup are dependent on the type I error probability α. This is somewhat surprising, considering that the asymptotic Neyman-Pearson error exponent is independent of α in most detection problems of interest, including the present one when the boundedness assumption holds.

1.2 General background

A) Observational model Each sensor observation $Y = Y_i$ takes values in the measurable space $(\mathcal{Y}, \mathcal{B})$. The distribution of Y under the null (H_0) and alternative (H_1) hypotheses is denoted by P and Q, respectively. For simplicity, the same notation is used for the n-fold product of these measures defined on $(\mathcal{Y}^n, \mathcal{B}^n)$.

Assumption 1.1. *P and Q are mutually absolutely continuous, i.e., $P \equiv Q$.*

Although our analysis can be tailored to situations where P and Q are not mutually absolutely continuous, we feel that such problems are of somewhat secondary interest and that their inclusion in this dissertation would detract from its clarity and cohesiveness.

Under Assumption 1.1, the (pre-quantization) log-likelihood ratio

$$X(y) \triangleq \log \frac{dP}{dQ}(y)$$

is well-defined for $y \in \mathcal{Y}$ and is a.s. finite. (Since $P \equiv Q$, "almost surely" and "almost everywhere" are understood as under both P and Q.) In $\mathcal{S}_n$, the variable $X(Y_i)$ will also be denoted as X_i.

It is well-known that in a centralized system where the sensor observations $Y_1, \ldots, Y_n$ are directly available to the fusion center, the optimal hypothesis test (Neyman Pearson or Bayes) involves $(Y_1, \ldots, Y_n)$ only through the sufficient statistic $X_1 + \cdots + X_n$. If each X_i happens to be an m-ary random variable (so that it can be obtained from Y_i via an m-ary quantizer), then the optimal centralized system can be realized by $\mathcal{S}_n$. This trivial case is excluded by the following assumption.

Assumption 1.2. *Every measurable m-ary partition of $\mathcal{Y}$ contains an atom over which $X = \log(dP/dQ)$ is not almost everywhere constant.*

B) Distance functionals and error exponents The error exponents in centralized detection are given by the functionals $D(P\|Q)$ and $\rho(P, Q)$ defined below. Both entities are derived from the class of f-divergences [15] or Ali-Silvey distances [2].

The (Kullback-Leibler, informational) divergence, or relative entropy, of P relative to Q is defined by

$$D(P\|Q) \stackrel{\triangle}{=} E_P[X] \;=\; \int \log \frac{dP}{dQ}(y)\, dP(y) \; .$$

On the convex domain $\mathcal{M}_e(\mathcal{Y})$ consisting of distribution pairs (P, Q) with the property $P \equiv Q$, the functional $D(P\|Q)$ is nonnegative and convex. It is also equal [13] to the optimal Neyman-Pearson error exponent in testing P versus Q at any level $\alpha \in (0, 1)$ based on the i.i.d. observations $Y_1, \ldots, Y_n$.

The Chernoff exponent $\rho(P, Q)$ is derived from the moment generating function $\Psi(\theta)$ of X under Q:

$$\Psi(\theta) \stackrel{\triangle}{=} E_Q[\exp\{\theta X\}] \;=\; \int \left(\frac{dP}{dQ}(y)\right)^{\theta} dQ(y) \; .$$

For fixed $\theta \in [0, 1]$, both $\Psi(\theta)$ and $\log \Psi(\theta)$ (the cumulant generating function) are finite-valued concave functionals of the pair (P, Q) in $\mathcal{M}_e(\mathcal{Y})$; while for fixed $(P, Q) \in \mathcal{M}_e(\mathcal{Y})$, both $\Psi(\theta)$ and $\log \Psi(\theta)$ are finite and convex in $\theta \in [0, 1]$. This last property, together with the fact that $\Psi(0) = \Psi(1) = 1$, guarantees that $\Psi(\theta)$ has a minimum value which is less than or equal to unity, achieved by some $\theta^* \in (0, 1)$. We define

$$\rho(P, Q) \stackrel{\triangle}{=} -\log \Psi(\theta^*) \;=\; -\log \left[\min_{\theta \in (0,1)} \Psi(\theta) \right] \; ,$$

and note [13] that the Chernoff exponent coincides with the Bayes error exponent in centralized hypothesis testing.

The functionals $D(\cdot\|\cdot)$ and $\rho(\cdot, \cdot)$ are also associated with error exponents in $\mathcal{S}_n$. As was shown in [30] under a boundedness assumption, the error exponents $e^*_{\mathrm{NP}}(\alpha)$ and $e^*_{\mathrm{B}}(\pi)$ of the absolutely optimal system coincide with those achieved by the best identical-quantizer system, i.e., $e^\diamond_{\mathrm{NP}}(\alpha)$ and $e^\diamond_{\mathrm{B}}(\pi)$, respectively. Now, in an n-sensor identical-quantizer system, the quantizer outputs $U_i = g(Y_i)$ are clearly i.i.d. Thus, by the discussion of the previous paragraphs, a fixed (as n varies) quantizer mapping g yields error exponents $D(P_g\|Q_g)$ for Neyman-Pearson testing and $\rho(P_g, Q_g)$ for Bayes testing, where P_g and Q_g are the distributions of U_i under the null and alternative hypotheses, respectively. The optimal error exponent is then obtained by choosing the mapping g so as to maximize the appropriate functional.

C) Deterministic and randomized quantization; LRQ's We now consider quantization for $\mathcal{S}_n$ in somewhat greater detail.

A deterministic m-ary quantizer is a measurable mapping g from the observation space $(\mathcal{Y}, \mathcal{B})$ to the message space $\mathcal{U}_m \stackrel{\triangle}{=} \{1, \ldots, m\}$.

For completeness, it is also necessary to introduce randomized quantizers in $\mathcal{S}_n$. Each such quantizer g (w.l.o.g. we use the same symbol) is specified by a finite vector (g_k) of deterministic quantizers together with a pmf vector (λ_k). Before making an observation, each randomized quantizer in $\mathcal{S}_n$ independently selects a deterministic mapping according to its individual (λ_k), and then proceeds to encode its observation using the chosen mapping. Independent (across sensors) randomization is needed in order to preserve independence of the messages U_i (the possibility of cooperative, or synchronized, randomization has also been mentioned in [31] and [35]).

The distributions of the message $U \in \mathcal{U}_m$ produced by g are denoted by P_g and Q_g, and are readily obtainable from P, Q and g. The post-quantization log-likelihood ratio is defined on $\mathcal{U}_m$ by

$$X_g(u) \triangleq \log \frac{P_g(u)}{Q_g(u)} \ .$$

Clearly, if g is deterministic, then both the output message U and the log-likelihood ratio $X_g(U)$ are measurable functions of the observation Y. In that case, we obtain the smoothing property

$$\begin{aligned} \exp\{X_g(U)\} &= \frac{1}{Q\{g(Y)=U\}} \int_{\{g(Y)=U\}} \frac{dP}{dQ}(y)dQ(y) \\ &= E_Q[\exp\{X\}|g(Y)=U] \ . \end{aligned} \tag{1.2.1}$$

Throughout the dissertation, $\mathcal{S}_n$ will employ n distinct quantizers $g_{n1},\ldots,g_{nn}$. We will suppress the common "n" from most subscripts, and we will also abbreviate $X_{g_i}(U_i)$ as X_{g_i}. The joint distributions of the independent messages $U_1,\ldots,U_n$ will be denoted by P_g and Q_g, where $P_g \triangleq P_{g_1} \times \cdots \times P_{g_n}$ and $Q_g \triangleq P_{g_1} \times \cdots \times P_{g_n}$.

Of special importance to our analysis are the so-called (log-) likelihood ratio quantizers (LRQ's). The definitions and properties given below are adapted from [33] and [31].

A deterministic m-ary LRQ is specified by an m-ary partition

$$\tau \ = \ (I_1,\ldots,I_m)$$

of the real line, where the I_u's are consecutive intervals (with I_1 being the left-most). We call such τ a likelihood ratio partition (LRP), and emphasize that I_u can be a singleton or even empty. The output of the quantizer is given by the rule

$$U(Y) = u \quad \text{iff} \quad X(Y) \in I_u \ .$$

A randomized m-ary LRQ is obtained by randomly selecting, according to a pmf (p_k), one of finitely many deterministic m-ary LRP's τ_k, where the τ_k's may differ only in the endpoints of the constituent intervals. For example, if $[a, b]$ is the uth interval in τ_k, then the corresponding interval in $\tau_{k'}$ must be one of $[a, b]$, $[a, b)$, $(a, b]$ or (a, b). As a consequence of this definition, if the observed log-likelihood ratio X falls in the interior of such an interval, then the output of the randomized LRQ is, with probability one, the index of that interval. Thus randomized LRQ's are only marginally different from deterministic ones, and they become indistinguishable—in terms of the resulting output distributions (P_g, Q_g)—from the latter when the distribution of X (under either hypothesis) has no point masses.

The class of randomized m-ary quantizers will be denoted by $\mathcal{G}_m$, and the deterministic m-ary LRP's will be denoted by $\mathcal{T}_m$. A deterministic LRQ g will commonly be designated by its corresponding LRP τ, and P_g, Q_g and X_g will also appear as P_τ, Q_τ and X_τ, respectively. Note that the random variable X_τ is a measurable function of both Y and $X(Y)$.

D) Optimality of LRQ's The optimality of LRQ's in $\mathcal{S}_n$ was established in a series of results outlined below.

1. For Neyman-Pearson testing, it was shown in [33] that if the observations $Y_1, \ldots, Y_n$ are independent (but not necessarily identically distributed) under both hypotheses, then an optimal solution $(g_1, \ldots, g_n, \mathcal{D})$ for $\mathcal{S}_n$ exists in which all g_i's are randomized LRQ's. In general, the optimal Neyman-Pearson fusion rule $\mathcal{D}$ operating on the discrete messages $U_1, \ldots, U_n$ will also be randomized. However, for the special case where the distribution (under either hypothesis) of each pre-quantization log-likelihood X_i has no point masses, it can be shown

[35] that the fusion rule $\mathcal{D}$ must be deterministic. (We already know from the discussion of the previous subsection that the optimal g_i's are also deterministic in this special case).

2. For Bayesian testing with independent (but, again, not necessarily identically distributed) observations, it was shown in [28] that an optimal solution for $\mathcal{S}_n$ exists in which all g_i's are deterministic LRQ's, and $\mathcal{D}$ is also deterministic.

3. The error exponents achievable by the optimal $\mathcal{S}_n$ (assuming i.i.d. observations and the boundedness condition) are given [30] by

$$(\forall \alpha \in (0,1)) \qquad e^*_{\mathrm{NP}}(\alpha) = \sup_{g \in \mathcal{G}_m} D(P_g \| Q_g)$$

and

$$(\forall \pi \in (0,1)) \qquad e^*_{\mathrm{B}}(\pi) = \sup_{g \in \mathcal{G}_m} \rho(P_g, Q_g)$$

As g varies over $\mathcal{G}_m$, the resulting range of output distribution pairs (P_g, Q_g) is a closed convex subset $\mathcal{L}_m(P,Q)$ of $[0,1]^{2m}$. Using the convexity properties of the functionals $D(P_g\|Q_g)$ and $\Psi(\theta)$ defined earlier, one can show that both suprema will be achieved by distribution pairs that are extremal points of $\mathcal{L}_m(P,Q)$. These extremal points can, in turn, be generated by deterministic LRQ's, and it thus follows that

$$\sup_{g \in \mathcal{G}_m} D(P_g \| Q_g) = \max_{\tau \in \mathcal{T}_m} D(P_\tau \| Q_\tau) \triangleq D_m \,, \tag{1.2.2}$$

$$\sup_{g \in \mathcal{G}_m} \rho(P_g, Q_g) = \max_{\tau \in \mathcal{T}_m} \rho(P_\tau, Q_\tau) \triangleq \rho_m \,. \tag{1.2.3}$$

This argument is developed in detail in [31], and similar results have appeared in [33].

In the light of the above discussion, all problems studied in this dissertation have solutions which are randomized LRQ's, and the generic quantizer g can be restricted to this class.

Notation Throughout the dissertation, $a \wedge b \triangleq \min\{a, b\}$ and $a \vee b \triangleq \max\{a, b\}$.

Chapter 2

Neyman-Pearson testing in parallel distributed detection systems

We continue the discussion of the asymptotic performance of the system introduced in Section 1.1, with emphasis on Neyman-Pearson testing.

2.1 The boundedness assumption

The equality (and finiteness) of $e^*_{\mathrm{NP}}(\alpha)$ and $e^\diamond_{\mathrm{NP}}(\alpha)$ was established in Theorem 2 of [30] under the assumption that $E_P[X^2] < \infty$. Actually, the proof only utilized the following weaker condition for $\delta = 0$ on the post-quantization log-likelihood ratio.

Assumption 2.1. *There exists $\delta \geq 0$ for which*

$$\sup_{g \in \mathcal{G}_m} E_P[|X_g|^{2+\delta}] < \infty. \tag{2.1.1}$$

Let us briefly examine the above assumption. For an arbitrary randomized quantizer g, let $p_u = P_g(u)$ and $q_u = Q_g(u)$. Then

$$E_P[|X_g|^{2+\delta}] \;=\; \sum_{u=1}^{m} p_u \left|\log \frac{p_u}{q_u}\right|^{2+\delta} \;.$$

Our first observation is that the negative part X_g^- of X_g has bounded $(2+\delta)$-th moment under P, and thus Assumption 2.1 is equivalent to $\sup_g E_P[(X_g^+)^{2+\delta}] < \infty$. Indeed, we have

$$E_P[|X_g^-|^{2+\delta}] \;=\; \sum_{u:\, p_u<q_u} p_u \left|\log \frac{p_u}{q_u}\right|^{2+\delta} \;=\; \sum_{u:\, p_u<q_u} p_u \left|\log \frac{q_u}{p_u}\right|^{2+\delta} \;,$$

and using the inequality $|\log x^{1/(2+\delta)}| \le x^{1/(2+\delta)} - 1$ for $x \ge 1$, we obtain

$$\begin{aligned} E_P[(X_g^-)^{2+\delta}] \;&\le\; (2+\delta)^{2+\delta} \sum_{u:\, p_u<q_u} p_u \left(\left(\frac{q_u}{p_u}\right)^{1/(2+\delta)} - 1 \right)^{2+\delta} \\ &\le\; (2+\delta)^{2+\delta} \sum_{u=1}^{m} q_u \;\le\; (2+\delta)^{2+\delta} \;. \end{aligned}$$

Our second observation is as follows.

Theorem 2.1. *Assumption 2.1 is equivalent to the condition*

$$\sup_{\tau\in\mathcal{T}_2} E_P[|X_\tau|^{2+\delta}] \;<\; \infty \;. \tag{2.1.2}$$

Proof: Assumption 2.1 clearly implies (2.1.2). To prove the converse, let τ_t be the LRP in $\mathcal{T}_2$ defined by

$$\tau_t \stackrel{\triangle}{=} \big((-\infty, t]\,,\,(t,\infty)\big) \;, \tag{2.1.3}$$

and let $p(t) = P\{X > t\}$, $q(t) = Q\{X > t\}$. By (2.1.2), there exists $b < \infty$ such that for all $t \in \boldsymbol{R}$,

$$E_P[|X_{\tau_t}|^{2+\delta}] \;=\; (1-p(t)) \left|\log \frac{1-p(t)}{1-q(t)}\right|^{2+\delta} + p(t) \left|\log \frac{p(t)}{q(t)}\right|^{2+\delta} \;\le\; b \;. \tag{2.1.4}$$

Consider now an arbitrary deterministic m-ary quantizer g with output pmf's

$(p_1, \ldots, p_m)$ and $(q_1, \ldots, q_m)$, and let the maximum of $p_u|\log(p_u/q_u)|^{2+\delta}$ subject to $p_u \geq q_u$ be achieved at $u = u_*$. Then, from the first observation in this section, it follows that

$$E_P[|X_g|^{2+\delta}] \leq mp_* \left|\log \frac{p_*}{q_*}\right|^{2+\delta} + (2+\delta)^{2+\delta} .$$

To see that $p_* \left|\log(p_*/q_*)\right|^{2+\delta}$ is bounded from above if (2.1.4) holds, note that for a given $p_* = P\{Y \in C^*\}$, the value of $q_* = Q\{Y \in C^*\}$ can be bounded from below using the Neyman-Pearson lemma. In particular, there exist $t \in \boldsymbol{R}$ and $\mu \in [0,1]$ such that

$$\begin{aligned} p_* &= \mu P\{X = t\} + p(t) \ , \\ q_* &\geq \mu Q\{X = t\} + q(t) \ . \end{aligned}$$

If $P\{X = t\} = 0$, then

$$p_* \left|\log \frac{p_*}{q_*}\right|^{2+\delta} \leq p(t) \left|\log \frac{p(t)}{q(t)}\right|^{2+\delta} ,$$

where by virtue of (2.1.4), the r.h.s. is upper-bounded by b. Otherwise, the r.h.s. is of the form

$$\Big(p(t) + \mu P\{X = t\}\Big) \left|\log \frac{p(t) + \mu P\{X = t\}}{q(t) + \mu Q\{X = t\}}\right|^{2+\delta} ,$$

for some $\mu \in (0,1]$. For $\mu = 1$, this can again be bounded using (2.1.2): take τ_t' consisting of intervals $(-\infty, t)$ and $[t, \infty)$, or use τ_t and a simple continuity argument. Then the log-sum inequality [14, Theorem 2.7.1] can be applied together with the concavity of $f(t) = t^{1/(2+\delta)}$ to show that the same bound b is valid for $\mu \in (0,1)$ (details are omitted).

If g is a randomized quantizer, then the probabilities p_* and q_* defined previously will be expressible as $\sum_k \lambda_k p^{(k)}$ and $\sum_k \lambda_k q^{(k)}$. Here k ranges over a finite index set, and each pair $(p^{(k)}, q^{(k)})$ is derived from a deterministic quantizer. Again, using the log-sum inequality and the concavity of $f(t) = t^{1/(2+\delta)}$, one can obtain $p_* \left|\log(p_*/q_*)\right|^{2+\delta} \leq b$. $\square$

Remark It is easy to show that (2.1.2) is in fact equivalent to

$$\limsup_{t\to\infty} E_P[|X_{\tau_t}|^{2+\delta}] \ < \ \infty \ , \tag{2.1.5}$$

where τ_t is defined in (2.1.3).

Our next result is a refinement on the asymptotic equivalence of the best identical-quantizer system ($^\diamond$) and the absolutely optimal system (*). The proof is a variant of a standard argument for Stein's lemma (see, e.g., [14, Theorem 12.8.1]) and also parallels the condensed proof of Theorem 2 in [30]; it can be found in [10, Theorem 1].

Theorem 2.2. *If Assumption 2.1 holds, then for all* $\alpha \in (0,1)$*,*

$$-\frac{1}{n}\log\beta_n^*(\alpha) \ = \ D_m + O(n^{-1/2})$$

and

$$-\frac{1}{n}\log\beta_n^\diamond(\alpha) \ = \ D_m + O(n^{-1/2}) \ . \qquad \square$$

As an immediate corollary, we have $e^*_{\mathrm{NP}}(\alpha) = e^\diamond_{\mathrm{NP}}(\alpha) = D_m$, which is Theorem 2 in [30]. The stated result sharpens this equality by demonstrating that the finite-sample error exponents of the absolutely optimal and best identical-quantizer systems converge at a rate $O(n^{-1/2})$. It also motivates the following observations.

The first observation concerns the accuracy, or tightness, of the $O(n^{-1/2})$ convergence factor. Although the upper and lower bounds on $\beta_n^*(\alpha)$ in Theorem 2.2 are based on a suboptimal null acceptance region, it seems that in general, the

$O(n^{-1/2})$ rate cannot be improved on. (We have found examples to support this conjecture in the context of centralized testing.) At the same time, it is rather unlikely that the ratio $\beta_n^*(\alpha)/\beta_n^\diamond(\alpha)$ decays as fast as $\exp\{-c'\sqrt{n}\}$, i.e., there probably exists a lower bound which is tighter than what is implied by Theorem 2.2. As we shall see in the next section, under some regularity assumptions on the mean and variance of the post-quantization log-likelihood ratio computed w.r.t. the null distribution, the ratio $\beta_n^*(\alpha)/\beta_n^\diamond(\alpha)$ is indeed bounded from below.

The second observation is about the composition of an optimal quantizer set $(g_1, \ldots, g_n)$ for $\mathcal{S}_n$. The proof of Theorem 2.2 also yields an upper bound on the number of quantizers that are at least ε–distant from the deterministic LRQ that achieves $e_{\rm NP}^\diamond(\alpha)$. Specifically, if $K_n(\varepsilon)$ is the number of indices i for which

$$D(P_{g_i}\|Q_{g_i}) \;<\; D_m - \varepsilon$$

(where $\varepsilon > 0$), then

$$\frac{K_n(\varepsilon)}{n} \;=\; O(n^{-1/2}) \;.$$

Thus in an optimal system, most quantizers will be "essentially identical" to the one that achieves $e_{\rm NP}^\diamond(\alpha)$. As we shall see in the next section, this conclusion can be significantly strengthened if additional assumption is made in the case of Neyman-Pearson testing.

In the remainder of this section we discuss the asymptotics of Neyman-Pearson testing in situations where Assumption 2.1 does not hold. By the remark following the proof of Theorem 2.1, this condition is violated if and only if

$$\limsup_{t\to\infty} E_P[X_{\tau_t}^2] \;=\; \infty \;, \tag{2.1.6}$$

where τ_t is the binary LRP defined by

$$\tau_t \;=\; \Big((-\infty, t]\,,\,(t,\infty)\Big)\;.$$

We distinguish between three cases.

Case A. $$\limsup_{t\to\infty} E_P[X_{\tau_t}] \;=\; \infty\;.$$

Case B. $$0 \;<\; \limsup_{t\to\infty} E_P[X_{\tau_t}] \;<\; \infty\;.$$

Case C. $$\limsup_{t\to\infty} E_P[X_{\tau_t}] \;=\; 0 \quad \text{and} \quad \limsup_{t\to\infty} E_P[X_{\tau_t}^2] \;=\; \infty\;.$$

Example Let the observation space be the unit interval $(0,1]$ with its Borel field. For $a>0$, define the distributions P and Q by

$$P\{Y\le y\} = y\;, \qquad\qquad Q\{Y\le y\} = \exp\left\{\frac{a+1}{a}\left(1-\frac{1}{y^a}\right)\right\}\;.$$

The pdf of Q is strictly increasing in y, and thus the likelihood ratio $(dP/dQ)(y)$ is strictly decreasing in y. Hence the event $\{X>t\}$ can also be written as $\{Y<y_t\}$, where $y_t\to 0$ as $t\to\infty$. Using this equivalence, we can examine the limiting behavior of $E_P[X_{\tau_t}]$ and $E_P^2[X_{\tau_t}]$ to obtain:

a. $a>1$: $\lim_{t\to\infty} E_P[X_{\tau_t}] = \lim_{t\to\infty} E_P[X_{\tau_t}^2] = \infty$ (Case A)

b. $a=1$: $\lim_{t\to\infty} E_P[X_{\tau_t}] = 2,\ \lim_{t\to\infty} E_P[X_{\tau_t}^2] = \infty$ (Case B)

c. $1/2<a<1$: $\lim_{t\to\infty} E_P[X_{\tau_t}] = 0,\ \lim_{t\to\infty} E_P[X_{\tau_t}^2] = \infty$ (Case C)

d. $a\le 1/2$: $\lim_{t\to\infty} E_P[X_{\tau_t}^2] < \infty$ (Assumption 2.1 is satisfied) .

In Case A, the error exponents $e^*_{\rm NP}(\alpha)$ and $e^\diamond_{\rm NP}(\alpha)$ are both infinite. This result is neither difficult to prove nor surprising, considering that $E_P[X_g] =$

$D(P_g \| Q_g)$ can be made arbitrarily large by choice of the quantizer g.

Theorem 2.3. *If* $\limsup_{t\to\infty} E_P[X_{\tau_t}] = \infty$*, then for all* $m \geq 2$ *and* $\alpha \in (0,1)$*,*

$$e^*_{\mathrm{NP}}(\alpha) = e^{\diamond}_{\mathrm{NP}}(\alpha) = \infty . \qquad \square$$

We now turn to case B, which is more interesting. Using the notation $p(t) = P\{X > t\}$ and $q(t) = Q\{X > t\}$ introduced earlier, we have

$$E_P[X_{\tau_t}] = (1 - p(t)) \log \frac{1 - p(t)}{1 - q(t)} + p(t) \log \frac{p(t)}{q(t)} .$$

The first summand on the r.h.s. clearly tends to zero (as $t \to \infty$, which is understood throughout), hence the lim sup of the second summand $p(t)\log[p(t)/q(t)]$ is greater than zero. Since $p(t)$ tends to zero, both $\log[p(t)/q(t)]$ and $p(t)\log^2[p(t)/q(t)]$ have lim sup equal to infinity. Thus in particular, (2.1.6) always holds in Case B.

A separate argument (which we omit) reveals that the centralized error exponent $D(P\|Q) = E_P[X]$ is also infinite in Case B. Yet unlike Case A, the decentralized error exponent $e^*_{\mathrm{NP}}(\alpha)$ obtained here is not infinite. Quite surprisingly, if this exponent exists, then it must depend on the test level α. This is stated in the following theorem.

Theorem 2.4. *Consider hypothesis testing with* m*-ary quantization, where* $m \geq 2$*. If*

$$0 < \limsup_{t\to\infty} E_P[X_{\tau_t}] < \infty , \qquad (2.1.7)$$

then there exist:

(i) an increasing sequence of integers $\{n_k, k \in \mathbf{N}\}$ *and a function* $L : (0,1) \mapsto$

$(0,\infty)$ *which is monotonically increasing to infinity, such that*

$$\liminf_{k\to\infty} -\frac{1}{n_k} \log \beta^{\diamond}_{n_k}(\alpha) \;\geq\; L(\alpha) \vee D_m \;;$$

(ii) a function $M : (0,1) \mapsto (0,\infty)$ *which is monotonically increasing to infinity and is such that*

$$\limsup_{n\to\infty} -\frac{1}{n} \log \beta^{*}_{n}(\alpha) \;\leq\; M(\alpha) \;.$$

Proof: (i) Lower bound. As was argued in the proof of Theorem 2.2, an error exponent equal to D_m can be achieved using identical quantizers; hence one part of the bound follows immediately. In what follows, we construct a sequence of identical-quantizer detection schemes with finite sample-error exponent almost always exceeding $L(\alpha)$, where $L(\alpha)$ increases to infinity as α tends to unity.

Let $\nu \stackrel{\triangle}{=} \limsup_{t\to\infty} E_P[X_{\tau_t}]$, so that $0 < \nu < \infty$ by (2.1.7). By subsequence selection, we obtain a sequence of LRP's τ_{t_k} with the property that $\nu_k \stackrel{\triangle}{=} E_P[X_{\tau_k}]$ converges to ν as $k \to \infty$. (We eliminate t from all subscripts to simplify the notation.) Letting $p_k = p(t_k)$, $q_k = q(t_k)$, $\varepsilon_k = \log[(1-p_k)/(1-q_k)]$ and $\zeta_k = \log(p_k/q_k)$, we can write

$$\nu_k \;=\; (1-p_k)\varepsilon_k + p_k\zeta_k \;,$$

where by the discussion preceding this theorem, p_k and ε_k tend to zero and ζ_k increases to infinity.

Fix $\omega > 0$ and assume w.l.o.g. that $\zeta_k > \zeta_{k-1} + (1/\omega)$. Consider a system consisting of n_k sensors, where $n_k = \lfloor \omega\zeta_k \rfloor$, and let each sensor employ the same binary LRQ with LRP τ_k. This choice is clearly suboptimal (since $m \geq 2$), but it suffices for our purposes.

Define the set $\mathcal{A}_k \subset \{1,2\}^{n_k}$ by

$$\mathcal{A}_k \;=\; \{(u_1,\ldots,u_{n_k}) : \text{ at least one } u_j \text{ equals } 2\}\ .$$

Recalling that $U_j = 2$ iff the observed log-likelihood ratio is larger than t_k, we have

$$P_\tau(\mathcal{A}_k) \;=\; 1-(1-p_k)^{n_k}\ ,$$

and thus

$$\begin{aligned}
\lim_{k\to\infty}(1-P_\tau(\mathcal{A}_k)) &= \lim_{k\to\infty}(1-p_k)^{n_k} \\
&= \lim_{k\to\infty}(1-p_k)^{\omega\zeta_k} \\
&= \left(\lim_{k\to\infty}(1-p_k)^{\zeta_k}\right)^{\omega} \;=\; \exp\{-\nu\omega\}\ ,
\end{aligned}$$

where the last equality follows from the fact that $\zeta_k \to \infty$ and $p_k\zeta_k \to \nu$ as $k\to\infty$.

Thus given any $\eta > 0$, for all sufficiently large k the set $\mathcal{A}_k$ is admissible (albeit not necessarily optimal) as a null acceptance region for testing at level $\alpha = \exp\{-\nu\omega+\eta\}$. For this value of α, we have

$$\begin{aligned}
\beta^{\diamond}_{n_k}(\alpha) &\le Q_\tau(\mathcal{A}_k) \\
&= 1-(1-q_k)^{n_k} \\
&= 1-(1-p_k\exp\{-\zeta_k\})^{n_k} \\
&= n_kp_k\exp\{-\zeta_k\} - \frac{n_k(n_k-1)}{2!}p_k^2\exp\{-2\zeta_k\} \\
&\qquad + \cdots - (-1)^{n_k}p_k^{n_k}\exp\{-n_k\zeta_k\} \\
&\le n_kp_k\exp\{-\zeta_k\} + n_k^2p_k^2\exp\{-2\zeta_k\} + \cdots
\end{aligned}$$

Summing the geometric series, we obtain

$$\beta^{\diamond}_{n_k}(\alpha) \ \le \ \frac{n_k p_k \exp\{-\zeta_k\}}{1 - n_k p_k \exp\{-\zeta_k\}} \ .$$

The r.h.s. denominator tends to unity because $\zeta_k \to \infty$ and $n_k p_k \to \omega\nu$ as $k \to \infty$. Since $\zeta_k / n_k \to 1/\omega$, we conclude that

$$\liminf_{k\to\infty} -\frac{1}{n_k} \log \beta^{\diamond}_{n_k}(\alpha) \ \ge \ \frac{1}{\omega} \ = \ \frac{\nu}{\log(1/\alpha) + \eta} \ .$$

As $\omega > 0$ and $\eta > 0$ were chosen arbitrarily, it follows that

$$\liminf_{k\to\infty} -\frac{1}{n_k} \log \beta^{\diamond}_{n_k}(\alpha) \ \ge \ \frac{\nu}{\log(1/\alpha)}$$

for all $\alpha \in (0, 1)$. The lower bound in statement (i) of the theorem is obtained by taking $L(\alpha) \stackrel{\triangle}{=} \nu / \log(1/\alpha)$.

(ii) Upper bound. Consider an optimal detection scheme for $\mathcal{S}_n$, with the same setup as in the proof of Theorem 2.2. Recall in particular that the fusion center employs a randomized test with log-likelihood threshold η_n and randomization constant μ_n.

For θ to be an upper bound on the error exponent of $\beta^*_n(\alpha)$, it suffices that $n\theta$ be greater than η_n and such that the events $\left\{\sum_{i=1}^n X_{g_i} \le n\theta\right\}$ and $\left\{\sum_{i=1}^n X_{g_i} \ge \eta_n\right\}$ have significant overlap under P_g. Indeed, if $\theta > \eta_n / n$ is such that for all sufficiently large n,

$$\mu_n P_g \left\{\sum\nolimits_{i=1}^n X_{g_i} = \eta_n\right\} + P_g \left\{n\theta \ge \sum\nolimits_{i=1}^n X_{g_i} > \eta_n\right\} \ \ge \ \varepsilon \ > \ 0 \ , \qquad (2.1.8)$$

then $\beta^*_n(\alpha) > \varepsilon \exp\{-n\theta\}$, as required.

The threshold η_n is rather difficult to determine, so we use an indirect method for finding θ. We have

$$P_g\left\{\sum_{i=1}^n X_{g_i} > n\theta\right\} \;\le\; P_g\left\{\sum_{i=1}^n |X_{g_i}| > n\theta\right\} \;\le\; \frac{1}{\theta}\sup_{g\in\mathcal{G}_m} E_P[|X_g|]\ ,$$

where the last bound follows from the Markov inequality. We claim that the supremum in this relationship is finite. This is because the negative part of X_g has bounded expectation under P (see the discussion following Assumption 2.1), and the proof of Theorem 2.1 can be easily modified to show that $\nu' \triangleq \sup_{g\in\mathcal{G}_m} E_P[|X_g|]$ is finite iff $\nu = \limsup_{t\to\infty} E_P[X_{\tau_t}]$ is (which is our current hypothesis). Thus

$$P_g\left\{\sum_{i=1}^n X_{g_i} > n\theta\right\} \;\le\; \frac{\nu'}{\theta}\ .$$

Now let $\varepsilon > 0$ and $\theta = \nu'/(1-\alpha-\varepsilon)$, so that $P_g\{\sum_{i=1}^n X_{g_i} > n\theta\} \ge 1-\alpha-\varepsilon$. The Neyman-Pearson lemma immediately yields $n\theta > \eta_n$. Also, using

$$\mu_n P_g\left\{\sum_{i=1}^n X_{g_i} = \eta_n\right\} + P_g\left\{\sum_{i=1}^n X_{g_i} > \eta_n\right\} \;=\; 1-\alpha$$

and a simple contradiction, we obtain (2.1.8). Thus the chosen value of θ is an upper bound on $\limsup_n(-1/n)\log\beta_n^*(\alpha)$. Since $\varepsilon > 0$ can be made arbitrarily small, we have that

$$M(\alpha) \triangleq \frac{\nu'}{1-\alpha}$$

is also an upper bound. □

From Theorem 2.4 we conclude that in Case B, the error exponent $e^*_{\mathrm{NP}}(\alpha)$ must lie between the bounds $L(\alpha)$ and $M(\alpha)$ whenever it exists as a limit. (Since $\nu' \ge D_m \ge \nu$ and $1-\alpha \le \log(1/\alpha)$, the inequality $M(\alpha) \ge L(\alpha)$ is indeed true.) These bounds are shown in Figure 2.1. We note that an earlier result (Theorem 3 in [10]) employed Poisson's theorem to yield a weaker bound $L(\alpha)$.

Finally, we briefly argue that the error exponents in Case C are identical to those obtained under Assumption 2.1 (boundedness of $E_P[|X_g|^{2+\delta}]$). Indeed, let $g_1, \ldots, g_n$ be optimal randomized LRQ's for $\mathcal{S}_n$. It is easy to show, by techniques similar to those employed in the proof of Theorem 2.1, that the condition $\limsup_{t\to\infty} E_P[X_{\tau_t}] = 0$ implies uniform integrability in the following sense: given $\varepsilon \in (0,1)$, there exists $b = b(\varepsilon, m)$ such that $E_P[|X_g|\mathbf{1}_{\{|X_g|>b\}}] < \varepsilon^2$ for all m-ary LRQ's g. Letting $X_g' = X_g \mathbf{1}_{\{X_g \le b\}}$ and $X_g'' = X_g \mathbf{1}_{\{X_g > b\}}$, we have

$$P_g\left\{\sum\nolimits_{i=1}^n X_{g_i} > n(D_m + 2\varepsilon)\right\} \;\le\; P_g\left\{\sum\nolimits_{i=1}^n X'_{g_i} > n(D_m + \varepsilon)\right\} + P_g\left\{\sum\nolimits_{i=1}^n X''_{g_i} > n\varepsilon\right\} \quad (2.1.9)$$

Now $|X_g'| \le b + X_g^-$, hence by the observation preceding Theorem 2.1, $\mathrm{Var}_P[X_g']$ is bounded. Since $E_P[X_g'] \le D_m + \varepsilon^2 < D_m + \varepsilon$, we can apply the Chebyshev inequality to conclude that first summand on the r.h.s. of (2.1.9) tends to zero as n tends to infinity. As for the second summand, it is less than ε by the Markov inequality and the bound $E_P[|X_g''|] < \varepsilon^2$. Retracing the proof of the converse (upper bound) part of Theorem 2.4, we obtain $\limsup_n (-1/n) \log \beta_n^*(\alpha) \le D_m + 2\varepsilon$ for any $\varepsilon \in (0, \varepsilon(\alpha))$, and thus finally, $e^*_{\mathrm{NP}}(\alpha) = e^{\diamond}_{\mathrm{NP}}(\alpha) = D_m$.

2.2 Error probability asymptotics under the boundedness assumption

In this section, we discuss the asymptotics of Neyman-Pearson testing in situations where Assumption 2.1 holds.

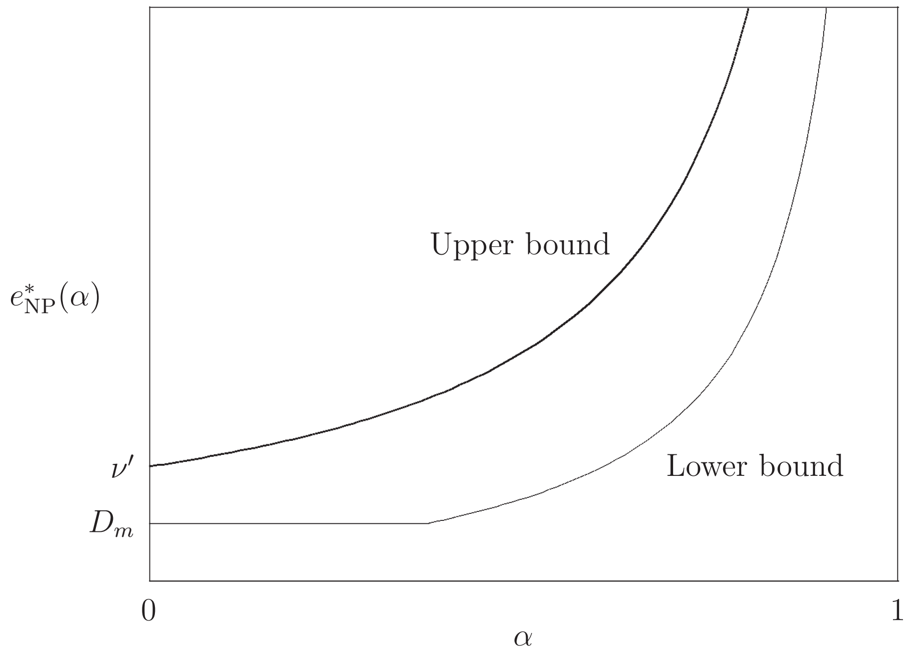

Figure 2.1: Upper and lower bounds on $e^*_{\mathrm{NP}}(\alpha)$ in Case B.

2.2.1 Properties of the post-quantization distribution

For the problem in hand, it is known that optimal solutions can be found in the subclass of deterministic and randomized likelihood ratio threshold quantizers [33]. This property will be of no particular value in what follows; this is because quantizers will be characterized almost entirely in terms of the induced post-quantization distribution pair. For a quantizer $g \in \mathcal{G}_m$, this is defined as the pair of vectors $\boldsymbol{p} \stackrel{\triangle}{=} (p(1), \ldots, p(m))$ and $\boldsymbol{q} \stackrel{\triangle}{=} (q(1), \ldots, q(m))$, where

$$p(u) \;=\; P\{U = u\} \qquad \text{and} \qquad q(u) \;=\; Q\{U = u\}$$

for $1 \leq u \leq m$. Write $\boldsymbol{f} = (\boldsymbol{p}, \boldsymbol{q})$, and define the post-quantization log-likelihood ratio of g by

$$X_{\boldsymbol{f}}(u) \stackrel{\triangle}{=} \log \frac{p(u)}{q(u)} \;.$$

Finally, denote by $\mathcal{F}_m \subset [0,1]^{2m}$ the class of all $\boldsymbol{f}$'s induced by quantizers in $\mathcal{G}_m$.

The moments of $X_{\boldsymbol{f}}$ under P will play an important role in the subsequent discussion. The following notations will be used:

$$\begin{aligned} \mu(\boldsymbol{f}) &\triangleq E_P[X_{\boldsymbol{f}}] \\ \sigma^2(\boldsymbol{f}) &\triangleq E_P[|X_{\boldsymbol{f}} - \mu(\boldsymbol{f})|^2] \\ \tau(\boldsymbol{f}, \delta) &\triangleq E_P[|X_{\boldsymbol{f}} - \mu(\boldsymbol{f})|^{2+\delta}] \end{aligned}$$

Note that $\mu(\boldsymbol{f})$ is just a Kullback-Leibler divergence, i.e.,

$$\mu(\boldsymbol{f}) \;=\; D(\boldsymbol{p}\|\boldsymbol{q}) \;=\; \sum_{u=1}^{m} p(u) \log \frac{p(u)}{q(u)} \;.$$

By Stein's lemma [13], $D(\boldsymbol{p}\|\boldsymbol{q})$ is the asymptotic type II error exponent obtained when all sensors employ a common quantizer with post-quantization distribution pair $\boldsymbol{f}$. Consequently,

$$D_m \triangleq \sup_{\boldsymbol{f} \in \mathcal{F}_m} \mu(\boldsymbol{f})$$

equals the error exponent of the best identical-quantizer system and also—by the results in [31]—that of the optimal system. Under Assumption 1.2, $D_m > D_{m-1}$ for all $m \geq 2$.

Recall that Assumption 2.1 requires that the $(2+\delta)$-th moment of the post-quantization likelihood ratio evaluated under null distribution be uniformly bounded over all possible quantizers. Although our result in this section remains true even if Assumption 2.1 holds in the weakest sense, i.e., with (2.1.1) valid only for $\delta = 0$, inclusion of such extreme case would detract from the clarity and cohesiveness of the proof. Hence, in what follows, δ will be understood as positive and satisfying (2.1.1). As a consequence of Assumption 2.1, $\mu(\boldsymbol{f})$, $\sigma^2(\boldsymbol{f})$

and $\tau(\boldsymbol{f}, \delta)$ will always be bounded above. Note that (2.1.1) is implied by the stronger (and easier to verify) condition that the pre-quantization log-likelihood X satisfies

$$E_P\left[|X|^{2+\delta}\right] \;<\; \infty \; .$$

By the discussion in Section 2.1, (2.1.1) is also equivalent to the condition (2.1.5).

For convenience, $\boldsymbol{f}_{ni} = (\boldsymbol{p}_{ni}, \boldsymbol{q}_{ni})$ will denote the post-quantization distribution pair of the ith quantizer in $\mathcal{S}_n$. The corresponding log-likelihood ratio $X_{\boldsymbol{f}_{ni}}(U_{ni})$ will be abbreviated as X_{ni}. The quantities $\mu(\boldsymbol{f}_{ni})$, $\sigma^2(\boldsymbol{f}_{ni})$ and $\tau(\boldsymbol{f}_{ni}, \delta)$ will be abbreviated as μ_{ni}, σ^2_{ni} and $\tau_{ni}(\delta)$, respectively. Then

$$m_n \triangleq \sum_{i=1}^{n} \mu_{ni} \; , \qquad s_n^2 \triangleq \sum_{i=1}^{n} \sigma^2_{ni} \; , \quad \text{and} \quad r_n(\delta) \triangleq \sum_{i=1}^{n} \tau_{ni}(\delta) \; .$$

The functions $\varphi(\cdot)$ and $\Phi(\cdot)$ will denote the unit Gaussian pdf and cdf, respectively. Also, it is assumed throughout that $m \geq 2$.

With the notation defined above, we proceed to derive some properties of the functions $\mu(\boldsymbol{f})$, $\sigma^2(\boldsymbol{f})$ and $\tau(\boldsymbol{f}, \delta)$, and to introduce an additional assumption which is needed for our main results in this chapter.

It is well-known (see, e.g., [31]) that the set $\mathcal{F}_m$ of post-quantization distribution pairs is a closed convex subset of

$$\mathcal{X}_m \triangleq \left\{ (\boldsymbol{p}, \boldsymbol{q}) \in [0,1]^{2m} : \; \sum_{u=1}^{m} p(u) = \sum_{u=1}^{m} q(u) = 1 \right\} \; .$$

Assumption 1.1 implies that for any $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{F}_m$, $p(u) = 0$ iff $q(u) = 0$; while Assumption 1.2 guarantees that $\mathcal{F}_m$ and $\mathcal{X}_m$ have the same dimension, namely $2m - 2$. From [31], all distribution pairs $(\boldsymbol{p}, \boldsymbol{q})$ on the relative boundary of $\mathcal{F}_m$ can be generated by randomized likelihood-ratio threshold quantizers. The set $\mathcal{F}_2$ is depicted in Figure 2.2.

The functions $\mu(\boldsymbol{f})$, $\sigma^2(\boldsymbol{f})$ and $\tau(\boldsymbol{f},\delta)$ are continuous at any point $\boldsymbol{f} \in \mathcal{F}_m$ with strictly positive coordinates. Under Assumption 2.1, $\mu(\boldsymbol{f})$ and $\sigma^2(\boldsymbol{f})$ are also continuous on the whole domain $\mathcal{F}_m$. Indeed, for any $\boldsymbol{f}' \in \mathcal{F}_m$,

$$p'(u)\left|\log\frac{p'(u)}{q'(u)}\right|^2 \leq B^{\frac{2}{2+\delta}}(p'(u))^{\frac{\delta}{2+\delta}} , \tag{2.2.10}$$

where B is the value of the supremum in (2.1.1). Since the r.h.s. tends to zero together with $p'(u)$, both $\mu(\boldsymbol{f})$ and $\sigma^2(\boldsymbol{f})$ are also continuous at points $\boldsymbol{f} = (\boldsymbol{p}, \boldsymbol{q})$ such that $p(u) = q(u) = 0$ for some u.

Consider now the set

$$\mathcal{O}_m \stackrel{\triangle}{=} \{\boldsymbol{f} \in \mathcal{F}_m : \mu(\boldsymbol{f}) = D_m\} ,$$

where D_m is the maximum of $\mu(\boldsymbol{f})$ over $\mathcal{F}_m$. Continuity of $\mu(\boldsymbol{f})$ on $\mathcal{F}_m$ implies that $\mathcal{O}_m$ is a nonempty closed subset of $\mathcal{F}_m$. Convexity of $\mu(\boldsymbol{f})$—which is of the strict variety on any convex subdomain that excludes points with $\boldsymbol{p} = \boldsymbol{q}$ [14, Thm. 2.7.2]—implies that $\mathcal{O}_m$ consists of extreme points of $\mathcal{F}_m$ (see Figure 2.2). From [31], it is also known that the distribution pairs in $\mathcal{O}_m$ can be generated by deterministic likelihood-ratio threshold quantizers.

It is also useful to note that every point in $\mathcal{O}_m$ has strictly positive entries. If this were not true, i.e., if $p(u) = q(u) = 0$ for some $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{O}_m$ and some u, then $(\boldsymbol{p}, \boldsymbol{q})$ would also correspond to an $(m-1)$-ary quantizer. Using Assumption 1.2 and standard properties of the Kullback-Leibler divergence, it would then be possible to refine this $(m-1)$-ary quantizer to an m-ary one yielding a higher value of $\mu(\cdot)$ and thus also a contradiction.

Let $\Delta(\boldsymbol{f}, \mathcal{O}_m)$ denote the minimum Euclidean distance of $\boldsymbol{f}$ from the closed

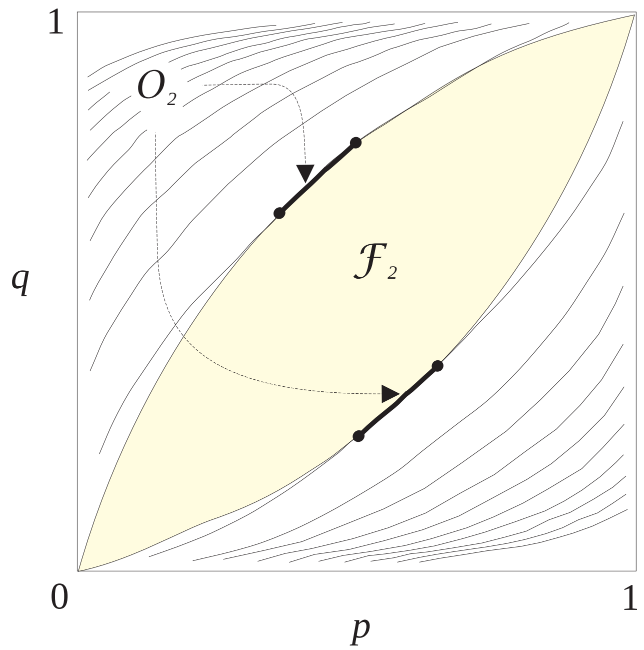

Figure 2.2: $\mathcal{F}_2$ and $\mathcal{O}_2$ as subsets of the two-dimensional set $\mathcal{X}_2$. Shown also are contours of $\mu(\cdot)$.

set $\mathcal{O}_m$, and define

$$\mathcal{N}_a(\mathcal{O}_m) \stackrel{\triangle}{=} \{\boldsymbol{f} \in \mathcal{F}_m : \ \Delta(\boldsymbol{f}, \mathcal{O}_m) \leq a\} \ .$$

By the foregoing discussion, $a > 0$ can be chosen such that

$$\mathcal{N}_a(\mathcal{O}_m) \subset [a, 1-a]^{2m} \ . \tag{2.2.11}$$

From (2.2.11) it follows that on the closed set $\mathcal{N}_a(\mathcal{O}_m)$, all partial derivatives (of any order) of $\sigma^2(\cdot)$ are continuous and hence bounded. It is then straightforward to show that exists a finite A_1 such that

$$(\forall \boldsymbol{f}_1 \in \mathcal{N}_a(\mathcal{O}_m), \ \boldsymbol{f}_2 \in \mathcal{F}_m) \qquad |\sigma^2(\boldsymbol{f}_1) - \sigma^2(\boldsymbol{f}_2)| \ \leq \ A_1 \cdot ||\boldsymbol{f}_1 - \boldsymbol{f}_2|| \ . \tag{2.2.12}$$

Also from (2.2.11), it follows that both $\sigma^2(\boldsymbol{f})$ and $\tau(\boldsymbol{f}, \delta)$ are bounded away from zero on $\mathcal{N}_a(\mathcal{O}_m)$.

The final assumption is as follows.

Assumption 2.2. *There exists $A_2 > 0$ such that for all $\boldsymbol{f} \in \mathcal{F}_m$,*

$$D_m - \mu(\boldsymbol{f}) \;\geq\; A_2 \cdot \Delta^2(\boldsymbol{f}, \mathcal{O}_m) \ .$$

A few comments on the above condition are in order here:

1. Although it was possible to construct examples of sets $\mathcal{F}_m$ that violate Assumption 2.2, these examples were not derived from common parametric families of distributions.

2. Assumption 2.2 is satisfied in the case of a finite observation space $\mathcal{Y}$. This is because $\mathcal{F}_m$ is polygonal, hence $\mathcal{O}_m$ consists solely of vertices of $\mathcal{F}_m$. Strict convexity of $\mu(\boldsymbol{f})$ along the edges of $\mathcal{O}_m$ then leads to the required inequality.

3. In cases where the distribution of the pre-quantization log-likelihood ratio X is absolutely continuous, each distribution pair $\boldsymbol{f}$ on the relative boundary of $\mathcal{F}_m$ can be parametrized by the likelihood-ratio threshold quantizer that generates $\boldsymbol{f}$, i.e., $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{t})$, where $\boldsymbol{t}$ is a $(m-1)$-dimensional vector of (increasing) thresholds. It can then be shown that Assumption 2.2 is satisfied if $\mu(\boldsymbol{f}(\boldsymbol{t}))$ is twice continuously differentiable with respect to $\boldsymbol{t}$ and achieves its global maxima at points $\boldsymbol{t}$ where its Hessian is (strictly) negative-definite. For $m = 2$ and $m = 3$, this condition was not violated in any of the parametric tests investigated with the aid of MATHEMATICA, including testing for a fixed signal in Gaussian noise.

2.2.2 Error probability bounds

This section establishes upper and lower bounds on the error probabilities of hypothesis tests based on the messages $U_{n1}, \ldots, U_{nn}$.

Lemma 2.1. *(Lower bound on type II error probability) Let* $\lim_{n\to\infty} n^{-1} m_n = D_m$ *and* $\delta \leq 1$. *Then for* n *sufficiently large, the minimum type II error probability* $\beta_n(\alpha)$ *attainable subject to an upper bound* α *on the type I probability satisfies*

$$\beta_n(\alpha) \;\geq\; \frac{\varphi(\Phi^{-1}(\alpha))}{2\, s_n\, e} \exp\left\{ -\frac{36}{\varphi(\Phi^{-1}t(\alpha))} \frac{r_n(\delta)}{s_n^{1+\delta}} \right\} \exp\left\{ -s_n\, \Phi^{-1}(\alpha) - m_n \right\} \; .$$

Proof: Since the messages $U_{n1}, \ldots, U_{nn}$ are independent, a sufficient statistic for the optimal global decision rule is the sum of the post-quantization log-likelihoods. Here, the normalized form $s_n^{-1} \sum_{i=1}^n (X_{ni} - \mu_{ni})$ is used, which has zero mean and unit variance under P. The cdf of $s_n^{-1} \sum_{i=1}^n (X_{ni} - \mu_{ni})$ under P is denoted by $H_n(\cdot)$.

From Assumption 2.1 and the condition $\lim_{n\to\infty} n^{-1} m_n = D_m$, it is easily deduced (see the proof of Lemma 2.3) that

$$\lim_{n\to\infty} \frac{r_n(\delta)}{s_n^{2+\delta}} \;=\; 0, \tag{2.2.13}$$

$$\lim_{n\to\infty} s_n \;=\; \infty \; . \tag{2.2.14}$$

The first equality is just Lyapounov's condition under which $s_n^{-1} \sum_{i=1}^n (X_{ni} - \mu_{ni})$ converges in distribution to a unit Gaussian variable [5, Thm. 27.3]. The Berry-Esseen Theorem [21, Sec. 2.5.1, Thm. 2.6] yields that

$$(\forall\, n) \qquad \sup_{t\in\Re} |H_n(t) - \Phi(t)| \;\leq\; 6\, \frac{r_n(\delta)}{s_n^{2+\delta}} \tag{2.2.15}$$

provided $\delta \leq 1$.

Let $\varepsilon_n \stackrel{\triangle}{=} \Phi^{-1}\left(\alpha + 6r_n(\delta)s_n^{-2-\delta}\right)$. Using a first-order Taylor series expansion of $\Phi(\cdot)$ and (2.2.13), is easy to show that

$$0 \;\leq\; \varepsilon_n - \Phi^{-1}(\alpha) \;\leq\; \frac{12}{\varphi(\Phi^{-1}(\alpha))}\frac{r_n(\delta)}{s_n^{2+\delta}} \;\to\; 0\;. \tag{2.2.16}$$

If the acceptance region for the alternative hypothesis is chosen as

$$\mathcal{U}_n \stackrel{\triangle}{=} \left\{\frac{1}{s_n}\sum_{i=1}^{n}(X_{ni}-\mu_{ni}) \;\leq\; \varepsilon_n\right\},$$

then by (2.2.15),

$$P(\mathcal{U}_n) \;=\; H_n(\varepsilon_n) \;\geq\; \Phi(\varepsilon_n) - 6\frac{r_n(\delta)}{s_n^{2+\delta}} = \alpha.$$

Although this choice violates the type I error constraint, it is nevertheless useful for lower-bounding $\beta_n(\alpha)$. Indeed, $\mathcal{U}_n$ is defined via a likelihood ratio partition and hence by the Neyman-Pearson lemma, $P(\mathcal{U}_n) \geq \alpha$ implies $Q(\mathcal{U}_n^c) \leq \beta_n(\alpha)$. Therefore, for $\eta_n > 0$,

$$\begin{aligned}
& \beta_n(\alpha) \\
\geq\;& Q\left(\mathcal{U}_n^c\right) \\
\geq\;& Q\left\{\varepsilon_n + \eta_n \;\geq\; \frac{1}{s_n}\sum_{i=1}^{n}(X_{ni}-\mu_{ni}) \;>\; \varepsilon_n\right\} \\
\geq\;& P\left\{\varepsilon_n + \eta_n \;\geq\; \frac{1}{s_n}\sum_{i=1}^{n}(X_{ni}-\mu_{ni}) \;>\; \varepsilon_n\right\}\exp\left\{-s_n\left(\varepsilon_n+\eta_n\right) - m_n\right\} \\
\geq\;& h_n(\eta_n)\exp\left\{-s_n\,\eta_n\right\}\exp\{-s_n\,\varepsilon_n - m_n\}\;,
\end{aligned} \tag{2.2.17}$$

where $h_n(\eta_n) \stackrel{\triangle}{=} \Phi\left(\varepsilon_n + \eta_n\right) - \Phi(\varepsilon_n) - 12r_n(\delta)s_n^{-2-\delta}$ and the last inequality follows from (2.2.15). If η_n is chosen so as to approach zero as n tends to infinity, then

$$\lim_{n\to\infty}\frac{\Phi\left(\varepsilon_n+\eta_n\right)-\Phi(\varepsilon_n)}{\eta_n} \;=\; \varphi(\lim_{n\to\infty}\varepsilon_n) \;=\; \varphi(\Phi^{-1}(\alpha))\;.$$

By (2.2.13) and (2.2.14), an appropriate choice for η_n is

$$\eta_n \triangleq \frac{24}{\varphi(\Phi^{-1}(\alpha))} \frac{r_n(\delta)}{s_n^{2+\delta}} + \frac{1}{s_n} \; .$$

Hence, for all sufficiently large n,

$$h_n(\eta_n) \; \geq \; \frac{1}{2}\,\varphi(\Phi^{-1}(\alpha))\,\eta_n - 12\,\frac{r_n(\delta)}{s_n^{2+\delta}} \; = \; \frac{\varphi(\Phi^{-1}(\alpha))}{2\,s_n}. \tag{2.2.18}$$

The combination of (2.2.17), (2.2.18), and (2.2.16) yields

$$\begin{aligned} \beta_n(\alpha) \quad &\geq \quad \frac{\varphi(\Phi^{-1}(\alpha))}{2\,s_n} \exp\left\{ -\frac{24}{\varphi(\Phi^{-1}(\alpha))} \frac{r_n(\delta)}{s_n^{1+\delta}} - 1 \right\} \\ &\qquad \cdot \exp\left\{ -s_n \left(\Phi^{-1}(\alpha) + \frac{12}{\varphi(\Phi^{-1}(\alpha))} \frac{r_n(\delta)}{s_n^{2+\delta}} \right) - m_n \right\} \\ &= \quad \frac{\varphi(\Phi^{-1}(\alpha))}{2\,s_n\,e} \exp\left\{ -\frac{36}{\varphi(\Phi^{-1}(\alpha))} \frac{r_n(\delta)}{s_n^{1+\delta}} \right\} \exp\left\{ -s_n\,\Phi^{-1}(\alpha) - m_n \right\}, \end{aligned}$$

thereby completing the proof. □

Remark. Although the assumption $\lim_{n\to\infty} n^{-1} m_n = D_m$ in the hypothesis of Lemma 2.1 may appear somewhat contrived, it is, as will be seen in the following section, satisfied by both the optimal and best-identical quantizer designs.

The derivation of the upper bound on the type II error probability employs a basic large deviations technique in addition to the Berry-Esseen theorem.

Lemma 2.2. *(Upper bound on type II error probability) Let* $\lim_{n\to\infty} n^{-1} m_n = D_m$ *and* $\delta \leq 1$. *Then for* n *sufficiently large, the minimum type II error probability* $\beta_n(\alpha)$ *attainable subject to an upper bound* α *on the type I probability* *satisfies*

$$\beta_n(\alpha) \;\leq\; \frac{\varphi(\Phi^{-1}(\alpha))}{s_n}\left(2+\frac{12}{\varphi(\Phi^{-1}(\alpha))}\frac{r_n(\delta)}{s_n^{1+\delta}}\right)$$
$$\cdot \exp\left\{\frac{12}{\varphi(\Phi^{-1}(\alpha))}\frac{r_n(\delta)}{s_n^{1+\delta}}\right\}\exp\left\{-s_n\,\Phi^{-1}(\alpha)-m_n\right\}.$$

Proof: The large deviations notation used here is the same as in [8]. Let $Z_n \triangleq \sum_{i=1}^{n}(X_{ni}-\mu_{ni})$ and denote the distribution of Z_n under Q by $F_n(\cdot)$. The tilted distribution with parameter θ corresponding to $F_n(\cdot)$ is defined by

$$dF_n^{(\theta)}(x) \;=\; \frac{\exp(\theta x)dF_n(x)}{\int \exp(\theta x')dF_n(x')} \;=\; \frac{\exp(\theta x)dF_n(x)}{M_n(\theta)}, \tag{2.2.19}$$

where $M_n(\theta)$ is the moment generating function of the distribution $F_n(\cdot)$. Note that

$$M_n(1) \;=\; \exp\{-m_n\}$$

and

$$F_n^{(1)}(s_n x) \;=\; H_n(x)\;,$$

where $H_n(\cdot)$ is the cdf of $s_n^{-1}Z_n$ under P (also used in the proof of Lemma 2.1).

As in the proof of Lemma 2.1, consider an acceptance region for the alternative hypothesis given by

$$\mathcal{V}_n \triangleq \left\{\frac{1}{s_n}\sum_{i=1}^{n}(X_{ni}-\mu_{ni}) \;\leq\; \zeta_n\right\},$$

where $\zeta_n \triangleq \Phi^{-1}\left(\alpha - 6r_n(\delta)s_n^{-2-\delta}\right)$. By analogy to (2.2.16), it is now true that

$$0 \;\leq\; \Phi^{-1}(\alpha)-\zeta_n \;\leq\; \frac{12}{\varphi(\Phi^{-1}(\alpha))}\frac{r_n(\delta)}{s_n^{2+\delta}} \;\to\; 0\;. \tag{2.2.20}$$

Also by (2.2.15),

$$P(\mathcal{V}_n) \;=\; H_n(\zeta_n) \;\leq\; \Phi(\zeta_n)+6r_n(\delta)s_n^{-2-\delta} \;=\; \alpha.$$

Thus $\mathcal{V}_n$ satisfies the type I error constraint and

$$\begin{aligned}
\beta_n(\alpha) &\leq Q(\mathcal{V}_n^c) \\
&= \int_{s_n \zeta_n}^{\infty} dF_n(x) \\
&= M_n(1) \int_{s_n \zeta_n}^{\infty} \exp(-x)\, dF_n^{(1)}(x) \\
&= \exp\{-m_n\} \int_{\zeta_n}^{\infty} \exp(-s_n\, x)\, dH_n(x) \\
&= \exp\{-s_n\, \zeta_n - m_n\} \cdot I_n \;, \qquad (2.2.21)
\end{aligned}$$

where

$$I_n = \int_0^{\infty} \exp\{-s_n\, x\}\; dH_n(x + \zeta_n).$$

Note that $\lambda_n(dx) \stackrel{\triangle}{=} s_n \exp\{-s_n\, x\} dx$ defines an exponential distribution on $[0, \infty)$. Integration by parts and (2.2.15) then yield

$$\begin{aligned}
I_n &= \int_0^{\infty} [\, H_n(x + \zeta_n) - H_n(\zeta_n)\,]\ \lambda_n(dx) \\
&\leq \int_0^{\infty} \left[\Phi(x + \zeta_n) - \Phi(\zeta_n) + 12\, \frac{r_n(\delta)}{s_n^{2+\delta}} \right] \lambda_n(dx) \\
&= \frac{1}{s_n} \int_0^{\infty} \varphi(x + \zeta_n) \lambda_n(dx) \ + \ 12\, \frac{r_n(\delta)}{s_n^{2+\delta}}. \qquad (2.2.22)
\end{aligned}$$

From (2.2.20), it follows that

$$\begin{aligned}
\left|\, \varphi(x + \zeta_n) - \varphi(x + \Phi^{-1}(\alpha)) \,\right| &\leq \sup_{x \in \Re} |\varphi'(x)| \cdot \left|\, \zeta_n - \Phi^{-1}(\alpha) \,\right| \\
&\leq \frac{12}{\sqrt{2\pi e}\, \varphi(\Phi^{-1}(\alpha))}\, \frac{r_n(\delta)}{s_n^{2+\delta}} \ ,
\end{aligned}$$

which in the light of (2.2.13) implies that $\varphi(x + \zeta_n)$ converges uniformly to $\varphi(x + \Phi^{-1}(\alpha))$. By (2.2.14), $\lambda_n(\cdot)$ converges to a degenerate distribution with unit mass at the origin. The first integral in (2.2.22) will therefore converge to

$\varphi(\Phi^{-1}(\alpha))$ (see [8, Lemma 1]). Thus, if n is sufficiently large,

$$I_n \;\le\; \frac{\varphi(\Phi^{-1}(\alpha))}{s_n}\left(2+\frac{12}{\varphi(\Phi^{-1}(\alpha))}\,\frac{r_n(\delta)}{s_n^{1+\delta}}\right). \tag{2.2.23}$$

From (2.2.20), (2.2.21) and (2.2.23) it follows that

$$\begin{aligned}\beta_n(\alpha) \quad &\le \quad \frac{\varphi(\Phi^{-1}(\alpha))}{s_n}\left(2+\frac{12}{\varphi(\Phi^{-1}(\alpha))}\,\frac{r_n(\delta)}{s_n^{1+\delta}}\right)\\ &\qquad \cdot \exp\left\{\frac{12}{\varphi(\Phi^{-1}(\alpha))}\,\frac{r_n(\delta)}{s_n^{1+\delta}}\right\}\exp\left\{-s_n\,\Phi^{-1}(\alpha)-m_n\right\},\end{aligned}$$

as required. □

Clearly, the validity of Lemmas 2.1 and 2.2 extends to centralized Neyman-Pearson testing with independent—but not necessarily identically distributed—observations under appropriate assumptions. In the special case of i.i.d. observations for which the third moment of the log-likelihood is finite, the type II error probability of the centralized system is given by

$$\beta_n(\alpha) \;=\; C_1(n)\frac{\varphi(\Phi^{-1}(\alpha))}{\sqrt{n}\sigma}\exp\left\{-\sqrt{n}\sigma\Phi^{-1}(\alpha)-n\mu\right\},$$

where $C_1(n)$ is a bounded constant satisfying

$$\begin{aligned}\exp\left\{-1-\frac{36\tau(1)}{\varphi(\Phi^{-1}(\alpha))\sigma^2}\right\} \;&\le\; C_1(n)\\ &\le\; \left(2+\frac{12\tau(1)}{\varphi(\Phi^{-1}(\alpha))\sigma^2}\right)\exp\left\{\frac{12\tau(1)}{\varphi(\Phi^{-1}(\alpha))\sigma^2}\right\}\end{aligned}$$

for all sufficiently large n.

2.2.3 Main results

This section compares the error performance of the optimal and best identical quantizer systems $\mathcal{S}_n^*$ and $\mathcal{S}_n^\diamond$. Basic properties of $\mathcal{S}_n^*$ are given in the following lemma.

Lemma 2.3. $\mathcal{S}_n^*$ *satisfies*

(i) $\lim_{n\to\infty} n^{-1} m_n = D_m$;

(ii) $r_n(\delta) s_n^{-1-\delta} = O(n^{(1-\delta)/2})$.

Proof: As is shown in [10, Theorem 1], Assumption 2.1 implies

$$\beta_n(\alpha) \geq \exp\left\{-C\sqrt{n} - m_n\right\}$$

for some $C > 0$. Using this lower bound, it is straightforward to show that if statement (i) is not true, then the error exponent of $\beta_n(\alpha)$ is less than D_m, which is impossible for $\mathcal{S}_n^*$.

From (i) it follows that in $\mathcal{S}_n^*$, the fraction of quantizers with post-quantization distribution pairs that lie outside $\mathcal{N}_a(\mathcal{O}_m)$ (defined in subsection 2.2.1) is $o(1)$. This, in conjunction with the fact that both $\sigma^2(\boldsymbol{f})$ and $\tau(\boldsymbol{f}, \delta)$ are bounded away from zero and infinity on $\mathcal{N}_a(\mathcal{O}_m)$, readily yields statement (ii). A similar argument proves (2.2.13) and (2.2.14). □

In upper-bounding the type II error probability of $\mathcal{S}_n^\diamond$, it is clear that any identical-quantizer system $\mathcal{S}_n^\circ$ with

$$\mu_{ni} \;=\; \mu_n^\circ \qquad \text{and} \qquad \sigma_{ni}^2 \;=\; (\sigma_n^\circ)^2$$

can be used. Statement (ii) of Lemma 2.3 is trivially true for $\mathcal{S}_n^\circ$. Under the assumption

$$\lim_{n\to\infty} \mu_n^\circ \;=\; D_m \;, \tag{2.2.24}$$

statement (i) will also be true. Both $\mathcal{S}_n^*$ and $\mathcal{S}_n^\circ$ will then satisfy the hypotheses of Lemmas 2.1 and 2.2, and

$$\frac{\beta_n^*(\alpha)}{\beta_n^\diamond(\alpha)} \;\geq\; \frac{\beta_n^*(\alpha)}{\beta_n^\circ(\alpha)} \;\geq\; C_2(n) \exp\left\{\Phi^{-1}(\alpha)\,\left(\sqrt{n}\sigma^\circ - s_n^*\right) + (n\mu_n^\circ - m_n^*)\right\} \;,$$

where

$$C_2(n) \;=\; \frac{1}{2e\left(2+\frac{12}{\varphi(\Phi^{-1}(\alpha))}\frac{r_n^\circ(\delta)}{(s_n^\circ)^{1+\delta}}\right)}\exp\left\{-\frac{12}{\varphi(\Phi^{-1}(\alpha))}\left(3\frac{r_n^*(\delta)}{(s_n^*)^{1+\delta}}+\frac{r_n^\circ(\delta)}{(s_n^\circ)^{1+\delta}}\right)\right\} .$$

Note that by Lemma 2.3(ii), $C_2(n)$ is of the form $\exp\{-c(\delta,\alpha)n^{(1-\delta)/2}\}$. In particular, $C_2(n)$ is bounded below if $\delta = 1$.

The main result is as follows.

Theorem 2.5. *Let $\delta \le 1$ satisfy (2.1.1). If $\alpha \le 1/2$, or if $\alpha > 1/2$ and Assumption 2.2 holds, then*

$$\frac{\beta_n^*(\alpha)}{\beta_n^\diamond(\alpha)} \;\ge\; \exp\{-c'(\delta,\alpha)n^{\frac{1-\delta}{2}}\} .$$

In particular, if the supremum of $E_P\left[|X_{\boldsymbol{f}}|^3\right]$ over $\mathcal{F}_m$ is finite, then the ratio $\beta_n^*(\alpha)/\beta_n^\diamond(\alpha)$ is bounded from below.

Proof: By the foregoing discussion, it suffices to prove that there is a choice of μ_n° which satisfies (2.2.24) and yields

$$\liminf_{n\to\infty}\left[\Phi^{-1}(\alpha)\,\left(\sqrt{n}\sigma_n^\circ - s_n^*\right) + (n\mu_n^\circ - m_n^*)\right] \;>\; -\infty.$$

The following arguments will show that there exist such choices of μ_n° that yield the above bound for arbitrary parameters $m_n = \sum_{i=1}^n \mu_{ni}$ and $s_n^2 = \sum_{i=1}^n \sigma_{ni}^2$ replacing m_n^* and $(s_n^*)^2$, respectively. For simplicity, the superscript "$\circ$" will be dropped from μ_n° and σ_n°.

Case 1: $\alpha \leq 1/2$. In this case, $\Phi^{-1}(\alpha) = -\gamma^2 \leq 0$ and

$$\begin{aligned}
& \Phi^{-1}(\alpha)\,\left(\sqrt{n}\sigma_n - s_n\right) + (n\mu_n - m_n) \\
= \;& \gamma^2\left(s_n - \sqrt{n}\sigma_n\right) + (n\mu_n - m_n) \\
\geq \;& \gamma^2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_{ni} - \sqrt{n}\sigma_n\right) + (n\mu_n - m_n) \\
= \;& \sum_{i=1}^{n}\left[\left(\mu_n - \frac{\gamma^2}{\sqrt{n}}\sigma_n\right) - \left(\mu_{ni} - \frac{\gamma^2}{\sqrt{n}}\sigma_{ni}\right)\right], \qquad (2.2.25)
\end{aligned}$$

where the Cauchy-Schwarz inequality $n \cdot \sum_{i=1}^{n} z_i^2 \geq (\sum_{i=1}^{n} z_i)^2$ was used.

The function $\mu(\boldsymbol{f}) - \gamma^2 n^{-1/2}\sigma(\boldsymbol{f})$ is continuous on $\mathcal{F}_m$, hence it achieves a global maximum at (say) $\boldsymbol{f} = \boldsymbol{f}_n$. By the fact that $\sigma^2(\boldsymbol{f})$ is bounded, this maximum value approaches D_m as n tends to infinity. Therefore the choice $(\mu_n, \sigma_n^2) = (\mu(\boldsymbol{f}_n), \sigma^2(\boldsymbol{f}_n))$ satisfies (2.2.24) and yields a nonnegative value for the lower bound in (2.2.25). The required result follows.

Case 2: $\alpha > 1/2$. In this case $\Phi^{-1}(\alpha) = \gamma^2 > 0$. Let $\boldsymbol{f} = \boldsymbol{f}_0$ be the post-quantization distribution pair in $\mathcal{O}_m$ with the highest $\sigma^2(\boldsymbol{f})$, and for all n take

$$(\mu_n, \sigma_n^2) \;=\; (\mu(\boldsymbol{f}_0), \sigma^2(\boldsymbol{f}_0)) \;=\; (D_m, \sigma^2)\;.$$

Note that this choice trivially satisfies (2.2.24). Also, denote by J_n the set of indices i for which $\sigma_{ni}^2 \geq \sigma^2$. Then

$$\begin{aligned}
s_n - \sqrt{n}\sigma \quad &= \quad \frac{s_n^2 - n\sigma^2}{s_n + \sqrt{n}\sigma} \\
&\leq \quad \frac{s_n^2 - n\sigma^2}{\sqrt{n}\sigma} \\
&\leq \quad \frac{1}{\sqrt{n}\sigma}\sum_{i \in J_n}\left[\sigma_{ni}^2 - \sigma^2\right]\;.
\end{aligned}$$

Let $(\mu_{ni}, \sigma_{ni}^2)$ correspond to $\boldsymbol{f}_{ni}$ and consider for each i a point $\boldsymbol{f}'_{ni}$ in $\mathcal{O}_m$

such that $||\boldsymbol{f}_{ni} - \boldsymbol{f}'_{ni}|| = \Delta(\boldsymbol{f}, \mathcal{O}_m)$. Since σ^2 is the maximum variance attainable on $\mathcal{O}_m$, it follows with the aid of (2.2.12) that

$$\begin{aligned}\sum_{i \in J_n} \left[\sigma_{ni}^2 - \sigma^2\right] &\leq \sum_{i \in J_n} \sigma_{ni}^2 - (\sigma'_{ni})^2 \\ &\leq A_1 \cdot \sum_{i \in J_n} \Delta(\boldsymbol{f}_{ni}, \mathcal{O}_m) \\ &\leq A_1 \cdot \sum_{i=1}^{n} \Delta(\boldsymbol{f}_{ni}, \mathcal{O}_m)\end{aligned}$$

and therefore that

$$s_n - \sqrt{n}\sigma \;\leq\; \frac{A_1}{\sqrt{n}\sigma} \sum_{i=1}^{n} \Delta(\boldsymbol{f}_{ni}, \mathcal{O}_m) \; . \tag{2.2.26}$$

Assumption 2.2 and the Cauchy-Schwarz inequality yield

$$\begin{aligned}nD_m - m_n &= \sum_{i=1}^{n} [D_m - \mu_{ni}] \\ &\geq A_2 \cdot \sum_{i=1}^{n} \Delta^2(\boldsymbol{f}_{ni}, \mathcal{O}_m) \\ &\geq \frac{A_2}{n} \left(\sum_{i=1}^{n} \Delta(\boldsymbol{f}_{ni}, \mathcal{O}_m) \right)^2 \; .\end{aligned} \tag{2.2.27}$$

From (2.2.26) and (2.2.27) it follows that

$$\begin{aligned}&\Phi^{-1}(\alpha) \; \left(\sqrt{n}\sigma_n - s_n\right) + (n\mu_n - m_n) \\ =\;& \gamma^2 \left(\sqrt{n}\sigma - s_n\right) + (nD_m - m_n) \\ \geq\;& -\frac{\gamma^2 A_1}{\sqrt{n}\sigma} \sum_{i=1}^{n} \Delta(\boldsymbol{f}_{ni}, \mathcal{O}_m) + \frac{A_2}{n} \left(\sum_{i=1}^{n} \Delta(\boldsymbol{f}_{ni}, \mathcal{O}_m) \right)^2 \; .\end{aligned}$$

The required result holds because the quadratic $A_2 z^2 - \gamma^2 \sigma^{-1} A_1 z$ has a finite minimum as z ranges over $\Re$. □

Chapter 3

Bayes testing in parallel distributed detection systems

3.1 Preliminaries

We now turn to the asymptotic study of optimal Bayes detection in $\mathcal{S}_n$. The prior probabilities of H_0 and H_1 are denoted by π and $1-\pi$, respectively, the probability of error of the absolutely optimal system is denoted by $\gamma_n^*(\pi)$, and the probability of error of the best identical-quantizer system is denoted by $\gamma_n^\diamond(\pi)$.

In our analysis, we will always assume that $\mathcal{S}_n$ employs deterministic m-ary LRQ's represented by LRP's $\tau_1, \ldots, \tau_n$. This is clearly sufficient by the discussion in Section 1.2.D.

Upon receiving the messages $U_1, \ldots, U_n$ produced by the quantizers, the fusion center uses the MAP rule to decide in favor of H_0 iff

$$\frac{P\left(U_1, \ldots, U_n\right)}{Q\left(U_1, \ldots, U_n\right)} > \frac{1-\pi}{\pi} ,$$

or equivalently, iff

$$X_{\tau_1} + \cdots + X_{\tau_n} > \log \frac{1-\pi}{\pi} \ .$$

The resulting probability of error is

$$\gamma_n(\pi) \ = \ \pi P\{X_{\tau_1} + \cdots + X_{\tau_n} \leq \eta\} + (1-\pi)Q\{X_{\tau_1} + \cdots + X_{\tau_n} > \eta\} \ , \quad (3.1.1)$$

where $\eta \stackrel{\triangle}{=} \log[(1-\pi)/\pi]$. When the LRP's are optimal, then $\gamma_n(\pi) = \gamma_n^*(\pi)$. (Note that the X_{τ_i}'s are measurable functions of the respective observations Y_i, and thus it is legitimate to use the product measure P instead of P_τ in expressions such as (3.1.1)).

Our aim is to show eventually that the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ is bounded away from zero as long as $P \equiv Q$. The groundwork for this result is developed in this section, and is divided into five units for convenience.

A) Basic large deviations results The values of $\gamma_n^*(\pi)$ and $\gamma_n^\diamond(\pi)$ can be approximated using a large deviations technique which originated in [6] and [3]. Our exposition here is fairly complete; for a general reference, see [7, Chapter VII, A] or [20, Section 5.4 and Appendix 5A].

For fixed n, let

$$\Psi_{\tau_i}(\theta) \stackrel{\triangle}{=} E_Q[\exp\{\theta X_{\tau_i}\}]$$

and

$$M_n(\theta) \stackrel{\triangle}{=} E_Q[\exp\{\theta(X_{\tau_1} + \cdots + X_{\tau_n})\}] \ .$$

Then by independence of the X_{τ_i}'s we have $M_n(\theta) = \prod_{i=1}^n \Psi_{\tau_i}(\theta)$.

We recall from Section 1.2.B that for each i, $\log \Psi_{\tau_i}(\theta)$ is convex in θ. This implies that $\log M_n(\theta) = \sum_{i=1}^n \log \Psi_{\tau_i}(\theta)$ is convex, and it is also strictly so unless all X_{τ_i}'s are trivial (which can be safely excluded). Since $\log M_n(0) =$

$\log M_n(1) = 0$, both $\log M_n(\theta)$ and $M_n(\theta)$ have a unique minimum achieved by $\theta_n \in (0,1)$. We have

$$\begin{aligned} M_n(\theta_n) &\geq \prod_{i=1}^{n} \min_{\theta\in(0,1)} \Psi_{\tau_i}(\theta) \\ &= \prod_{i=1}^{n} \exp\{-\rho(P_{\tau_i}, Q_{\tau_i})\} \;\geq\; \exp\{-n\rho_m\} \;, \end{aligned} \tag{3.1.2}$$

where ρ_m is defined in (1.2.3). Equality holds throughout iff every τ_i is an m-ary LRP that achieves ρ_m.

Let $\mathcal{X}_i$ be the (m–point) range of X_{τ_i}, and denote the distributions of X_{τ_i} under H_0 and H_1 by $\mathcal{P}_i$ and $\mathcal{Q}_i$, respectively. The tilted distribution of X_{τ_i} is the measure $\tilde{Q}_i$ on $\mathcal{X}_i$ given by

$$\tilde{Q}_i(x) \;=\; \frac{\exp\{\theta_n x\}\mathcal{Q}_i(x)}{\Psi_{\tau_i}(\theta_n)} \tag{3.1.3}$$

(cf. (2.2.19)), where θ_n is as defined in the previous paragraph, i.e., it satisfies $M_n'(\theta_n) = 0$.

By extension, the tilted distribution of the vector $(X_{\tau_1}, \ldots, X_{\tau_n})$ is defined as the product measure $\tilde{Q}_1 \times \cdots \times \tilde{Q}_n$, which clearly preserves independence of the X_{τ_i}'s. Denoting expectation w.r.t. the tilted distribution by $E_{\tilde{Q}}$, we have the identity

$$E_{\tilde{Q}}[X_{\tau_1} + \cdots + X_{\tau_n}] \;=\; 0 \;. \tag{3.1.4}$$

The Bayes error probability is given by (3.1.1). This can be expressed in terms of the tilted distribution by noting that

$$Q\{X_{\tau_1}+\cdots+X_{\tau_n}>\eta\} = \sum_{x_1+\cdots+x_n>\eta} \mathcal{Q}_1(x_1)\cdots\mathcal{Q}_n(x_n)$$
$$= M_n(\theta_n) \sum_{x_1+\cdots+x_n>\eta} \exp\{-\theta_n(x_1+\cdots+x_n)\}\cdot\tilde{Q}_1(x_1)\cdots\tilde{Q}_n(x_n) \; ; \tag{3.1.5}$$

and also since $\mathcal{P}_i(x)=\exp\{x\}\mathcal{Q}_i(x)$,

$$P\{X_{\tau_1}+\cdots+X_{\tau_n}\le\eta\}$$
$$= M_n(\theta_n) \sum_{x_1+\cdots+x_n\le\eta} \exp\{(1-\theta_n)(x_1+\cdots+x_n)\}\cdot\tilde{Q}_1(x_1)\cdots\tilde{Q}_n(x_n) \; . \tag{3.1.6}$$

Expressing the sums in (3.1.5) and (3.1.6) in terms of the tilted cdf $\tilde{F}_n(\cdot)$ of $X_{\tau_1}+\cdots+X_{\tau_n}$, we can rewrite (3.1.1) as

$$\gamma_n(\pi)$$
$$= M_n(\theta_n)\left[\pi\int_{x\le\eta}\exp\{(1-\theta_n)x\}d\tilde{F}_n(x)+(1-\pi)\int_{x>\eta}\exp\{-\theta_n x\}d\tilde{F}_n(x)\right] \; . \tag{3.1.7}$$

The Chernoff (upper) bound on $\gamma_n(\pi)$ can be easily derived from (3.1.7). More useful for our purposes is the lower bound

$$\gamma_n(\pi) \ge [\pi\wedge(1-\pi)]M_n(\theta_n)\int_{\boldsymbol{R}}\exp\{(1-\theta_n)x\wedge(-\theta_n x)\}d\tilde{F}_n(x)$$
$$\ge [\pi\wedge(1-\pi)]M_n(\theta_n)\int_{\boldsymbol{R}}\exp\{-|x|\}d\tilde{F}_n(x) \; , \tag{3.1.8}$$

which is independent of η and holds for all $\theta_n\in(0,1)$. We note that (3.1.8) also yields the lower bound appearing in [9, relationship (3.42)], which—together with the Chernoff bound— was used in [30] to prove the result $e^*_{\mathrm{B}}(\pi)=e^{\diamond}_{\mathrm{B}}(\pi)$ under the boundedness assumption $\sup_{\tau\in\mathcal{T}_m}\Psi''_\tau(\theta)<\infty$ or its equivalent form (see also [10, Theorem 5])

$$\sup_{\tau \in \mathcal{T}_m} \sup_{\tilde{Q}} E_{\tilde{Q}}[X_\tau^2] \; < \; \infty \; . \tag{3.1.9}$$

B) CLT approximations Tighter bounds on $\gamma_n(\pi)$ can be derived from (3.1.7) by a central limit theorem approximation. Since the independent sum $(X_{\tau_1} + \cdots + X_{\tau_n})/s_n$ has zero mean and unit variance under the tilted distribution, it may—under certain conditions—converge weakly to a Gaussian $\mathcal{N}(0,1)$ variable. If so, then $\tilde{F}_n$ can be approximated by a Gaussian cdf. In the (i.i.d.) case of identical LRP's $\tau_{ni} = \tau$, this technique yields (see, e.g., [3, 17, 18])

$$\liminf_{n\to\infty} \gamma_n(\pi)\sqrt{n}\exp\{n\rho(P_\tau, Q_\tau)\} \; = \; c_{\min} \; > \; 0 \tag{3.1.10}$$

and

$$\limsup_{n\to\infty} \gamma_n(\pi)\sqrt{n}\exp\{n\rho(P_\tau, Q_\tau)\} \; = \; c_{\max} \; < \; \infty \; , \tag{3.1.11}$$

where $c_{\min}$ and $c_{\max}$ depend on P_τ, Q_τ, π and η. Furthermore, $c_{\min} = c_{\max}$ except in certain cases where the variable X_τ has a lattice form.

From (3.1.10) we immediately obtain

$$\limsup_{n\to\infty} \gamma_n^\diamond(\pi)\sqrt{n}\exp\{n\rho_m\} \; < \; \infty \; . \tag{3.1.12}$$

In Section 3.2 we will show implicitly that

$$\liminf_{n\to\infty} \gamma_n^*(\pi)\sqrt{n}\exp\{n\rho_m\} \; > \; 0 \; , \tag{3.1.13}$$

and that therefore $\liminf_{n\to\infty}(\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)) > 0$. The following lemma and its corollary will be essential in proving that result.

Lemma 3.1. *If there are uniform (in n and $i \le n$) bounds $a > 0$ and $b < \infty$ such that*

C1. $|X_{\tau_{ni}}| \le b$ *a.s.; and*

C2. $\operatorname{Var}_{\tilde{Q}}[X_{\tau_{ni}}] \geq a^2$,

then

$$\liminf_{n\to\infty} \sqrt{n} \int_{\boldsymbol{R}} \exp\{-|x|\} d\tilde{F}_n(x) \;>\; 0 \;. \tag{3.1.14}$$

Proof: Let $\tilde{H}_n(\cdot)$ be the tilted cdf of the normalized sum $(X_{\tau_1}+\ldots+X_{\tau_n})/s_n$, and $\Phi(\cdot)$ be the cdf of a $\mathcal{N}(0,1)$ distribution. A CLT approximation due to Esseen [16, Section XVI.5, Theorem 2] gives, for all x and n,

$$|\tilde{H}_n(x) - \Phi(x)| \;\leq\; \frac{6}{s_n^3} \sum_{i=1}^{n} E_{\tilde{Q}}[|X_{\tau_i} - E[X_{\tau_i}]|^3] \;.$$

By an elementary argument, C1 can be shown to imply both $\operatorname{Var}_{\tilde{Q}}[X_{\tau_i}] \leq b^2$ and $E_{\tilde{Q}}[|X_{\tau_i} - E[X_{\tau_i}]|^3] \leq b^3$. As $s_n \geq a\sqrt{n}$ by C2, we obtain

$$|\tilde{H}_n(x) - \Phi(x)| \;\leq\; 6(b/a)^3 n^{-1/2} \;. \tag{3.1.15}$$

We have

$$\int_{\boldsymbol{R}} \exp\{-|x|\} d\tilde{F}_n(x) \;=\; \int_{\boldsymbol{R}} \exp\{-s_n|x|\} d\tilde{H}_n(x) \;.$$

Restricting the range of the right-hand integral to $(-tn^{-1/2},\, tn^{-1/2}]$, where $t > 0$ will be specified later, we obtain

$$\begin{aligned}\int_{\boldsymbol{R}} \exp\{-|x|\} d\tilde{F}_n(x) \quad &\geq \quad [\tilde{H}_n(tn^{-1/2}) - \tilde{H}_n(-tn^{-1/2})] \exp\{-ts_n n^{-1/2}\} \\ &\geq \quad [\tilde{H}_n(tn^{-1/2}) - \tilde{H}_n(-tn^{-1/2})] \exp\{-tb\} \;.\end{aligned}$$

It remains to show that for all sufficiently large n, the difference $\tilde{H}_n(tn^{-1/2}) - \tilde{H}_n(-tn^{-1/2})$ can be made greater than $cn^{-1/2}$ for $c > 0$. From the CLT approximation (3.1.15), we have

$$\begin{aligned}\tilde{H}_n(tn^{-1/2}) - \tilde{H}_n(-tn^{-1/2}) &\geq \Phi(tn^{-1/2}) - \Phi(-tn^{-1/2}) - 12(b/a)^3 n^{-1/2} \\ &\geq 2tn^{-1/2}\Phi'(tn^{-1/2}) - 12(b/a)^3 n^{-1/2} \ . \\ &= 2n^{-1/2}[t\Phi'(tn^{-1/2}) - 6(b/a)^3] \ .\end{aligned}$$

Since for fixed t, $\Phi'(tn^{-1/2}) \to \Phi'(0) = (2\pi)^{-1/2}$ as $n \to \infty$, any choice $t > 6\sqrt{2\pi}(b/a)^3$ will suffice. □

Corollary 3.1. *Under conditions C1 and C2 in the hypothesis of Lemma 3.1,*

$$\liminf_{n\to\infty} \gamma_n(\pi)\sqrt{n}\exp\{n\rho_m\} \ > \ 0.$$

Proof: This follows immediately from (3.1.2), (3.1.8) and (3.1.14). □

C) Inequalities for Chernoff exponents The following fact is well-known (see, e.g., [24] or [19]), and is established using Jensen's inequality together with (1.2.1).

Lemma 3.2. *Let g be a deterministic m-ary quantizer. Then for all $\theta \in (0,1)$,*

$$E_Q[\exp\{\theta X_g\}] \ \geq \ E_Q[\exp\{\theta X\}] \ ,$$

or equivalently, $\Psi_g(\theta) \geq \Psi(\theta)$, with equality if and only if $X_g = X$ a.s. □

Using Assumption 1.2 of Chapter 1 and a simple argument, we obtain the following corollary to Lemma 3.2.

Lemma 3.3. *If $l' < l \leq m$, then*

$$\rho_{l'} \ < \ \rho_l \ < \ \rho(P,Q) \ .$$ □

D) A classification of LRP's We denote by $\mathcal{T}_m(\delta)$ the class of all τ's in $\mathcal{T}_m$

with the property that

$$(\forall u \in \mathcal{U}_m) \qquad\qquad P_\tau(u) \wedge Q_\tau(u) \;\geq\; \delta \; .$$

Clearly for $\delta > \delta' > 0$,

$$\mathcal{T}_m(\delta) \;\subset\; \mathcal{T}_m(\delta') \;\subset\; \mathcal{T}_m(0) \;=\; \mathcal{T}_m \; .$$

The complement of $\mathcal{T}_m(\delta)$ w.r.t. $\mathcal{T}$ is denoted by $\mathcal{R}_m(\delta)$.

The next two lemmas together imply that the class $\mathcal{T}_m(\delta)$ satisfies conditions C1 and C2 in the hypothesis of Lemma 3.1. The first result follows immediately from the definition of $\mathcal{T}_m(\delta)$, while the second lemma is proved in Appendix B.

Lemma 3.4. *Let $\delta > 0$. If $\mathcal{T}_m(\delta)$ is nonempty, then*

$$\sup_{\tau\in\mathcal{T}_m(\delta)} \sup_{u\in\mathcal{U}_m} |X_\tau(u)| \;<\; \log(1/\delta). \qquad \square$$

Lemma 3.5. *Let $m \geq 2$ and $\delta > 0$. If $\mathcal{T}_m(\delta)$ is nonempty, then*

$$\inf_{\tau\in\mathcal{T}_m(\delta)} \inf_{\tilde{Q}} \mathrm{Var}_{\tilde{Q}}[X_\tau] \;>\; 0 \; ,$$

where $\tilde{Q}$ varies over all tilted distributions of X_τ. $\square$

Next we consider the class $\mathcal{R}_m(\delta) \stackrel{\triangle}{=} \mathcal{T}_m \setminus \mathcal{T}_m(\delta)$ for $\delta > 0$. It is reasonable to expect that as δ decreases to zero, the performance of any LRP τ in $\mathcal{R}_m(\delta)$ (measured in terms of $\rho(P_\tau, Q_\tau)$) approaches that of a LRP in $\mathcal{T}_{m-1}$. This is effectively stated in the following lemma, which is proved in Appendix B.

Lemma 3.6. *For every $2 \leq l \leq m$,*

$$\lim_{\delta\downarrow 0} \sup_{\tau\in\mathcal{R}_l(\delta)} \rho(P_\tau, Q_\tau) \;=\; \rho_{l-1} \; . \qquad \square$$

E) An inequality for Bayes testing The following lemma completes the preparation for the results of Section 3.2.

Lemma 3.7. *Consider two independent observations Y_A and Y_B such that*

$$H_0: \quad Y_A \sim P_A \text{ and } Y_B \sim P_B \ ;$$

$$H_1: \quad Y_A \sim Q_A \text{ and } Y_B \sim Q_B \ .$$

Let $\gamma_A(\pi)$, $\gamma_B(\pi)$ and $\gamma_{AB}(\pi)$ be the the minimum probability of error attainable in testing H_0 vs. H_1 on the basis of Y_A, Y_B and (Y_A, Y_B), respectively. Then for all $\pi \in [0,1]$,

$$\gamma_{AB}(\pi) \ \geq \ 2\gamma_A(\pi)\gamma_B(1/2) \ .$$

Proof: Without loss of generality, assume that Y_A and Y_B are discrete observations with alphabets $\mathcal{Y}_A$ and $\mathcal{Y}_B$, respectively. Then

$$\begin{aligned}
&\gamma_{AB}(\pi) \\
= \ & \sum_{y\in\mathcal{Y}_A}\sum_{y'\in\mathcal{Y}_B} \pi P_A(y)P_B(y') \wedge (1-\pi)Q_A(y)Q_B(y') \\
\geq \ & \sum_{y\in\mathcal{Y}_A}\sum_{y'\in\mathcal{Y}_B} \{\pi P_A(y)[P_B(y') \wedge Q_B(y')]\} \wedge \{(1-\pi)Q_A(y)[P_B(y') \wedge Q_B(y')]\} \\
= \ & \sum_{y\in\mathcal{Y}_A}\sum_{y'\in\mathcal{Y}_B} [\pi P_A(y) \wedge (1-\pi)Q_A(y)]\,[P_B(y') \wedge Q_B(y')] \\
= \ & 2\gamma_A(\pi)\gamma_B(1/2) \ ,
\end{aligned}$$

which completes the proof. □

3.2 Main results

We now use the tools developed in Section 3.1 to prove the following theorem.

Theorem 3.1. *In Bayes testing with m-ary quantization,*

$$\liminf_{n\to\infty} \frac{\gamma_n^*(\pi)}{\gamma_n^\diamond(\pi)} > 0 \tag{3.2.16}$$

for all $\pi \in (0,1)$.

Proof: An upper bound on $\gamma_n^\diamond(\pi)$ is immediately obtained from (3.1.12). It remains to find a good lower bound on $\gamma_n^*(\pi)$.

Step 1. Two subsystems. Fix n, and consider an absolutely optimal system employing LRP's $\tau_1, \dots, \tau_n$. Also fix $\delta > 0$ (to be specified later), and assume w.l.o.g. that

$$(\forall i \le n_A) \quad \tau_i \in \mathcal{T}_m(\delta) \qquad \text{and} \qquad (\forall i > n_A) \quad \tau_i \in \mathcal{R}_m(\delta) \ .$$

Let $n_B = n - n_A$, and define two systems $\mathcal{S}_{n_A}$ and $\mathcal{S}_{n_B}$ as follows. System $\mathcal{S}_{n_A}$ observes $Y_1, \dots, Y_{n_A}$, compresses these observations using the LRP's $\tau_1, \dots, \tau_{n_A}$, and then performs an optimal test based on $U_1, \dots, U_{n_A}$; while $\mathcal{S}_{n_B}$ observes $Y_{n_A+1}, \dots, Y_n$, quantizes using $\tau_{n_A+1}, \dots, \tau_n$, and decides based on $U_{n_A+1}, \dots, U_n$. The Bayes error probabilities for these two systems are denoted by $\gamma_{n_A}(\pi)$ and $\gamma_{n_B}(\pi)$. Since the vectors $(U_1, \dots, U_{n_A})$ and $(U_{n_A+1}, \dots, U_n)$ are independent, we can apply Lemma 3.7 to obtain

$$\gamma_n^*(\pi) \ \ge \ 2\gamma_{n_A}(\pi)\gamma_{n_B}(1/2) \ . \tag{3.2.17}$$

We examine $\gamma_{n_A}(\pi)$ and $\gamma_{n_B}(\pi)$ separately. By Lemmas 3.4 and 3.5, every LRP in $\mathcal{T}_m(\delta)$ satisfies conditions C1 and C2 in the hypothesis of Lemma 3.1, hence by Corollary 3.1,

$$\liminf_{n\to\infty} \gamma_{n_A}(\pi)\sqrt{n_A} \exp\{n_A \rho_m\} \ > \ 0. \tag{3.2.18}$$

Step 2. Conditioning in $\mathcal{S}_{n_B}$. To lower-bound $\gamma_{n_B}(\pi)$, we need to consider

LRP's in $\mathcal{R}_m(\delta)$. In this case, Lemma 3.1 does not necessarily apply, nor is the weaker boundedness assumption (3.1.9) (which was used in the lower bound in [30]) valid in general. To remove these obstacles, we condition on the event Δ specified as follows. For each $i \geq n_A + 1$, we let

$$J_i \;=\; \{u \in \mathcal{U}_m : P_{\tau_i}(u) \;\wedge\; Q_{\tau_i}(u) \geq \delta\} \;.$$

Furthermore, we let Δ_i be the event that X_{τ_i} lies in J_i, and define

$$\Delta \;=\; \bigcap_{i=n_A+1}^{n} \Delta_i \;.$$

From (3.1.1) we obtain the lower bound

$$\begin{aligned}\gamma_{n_B}(\pi) \;\geq\; & \left[\pi P\left\{\sum\nolimits_{i=n_A+1}^{n} X_{\tau_i} \leq \eta | \Delta\right\}\right. \\ & \left. + (1-\pi) Q\left\{\sum\nolimits_{i=n_A+1}^{n} X_{\tau_i} > \eta | \Delta\right\}\right] (P(\Delta) \;\wedge\; Q(\Delta)) \;,\end{aligned}$$

where P and Q represent product measures by the convention of Section 1.2. To estimate $P(\Delta) \wedge Q(\Delta)$, we recall that since the measures P and Q on the marginal space $(\mathcal{Y}, \mathcal{B})$ are mutually absolutely continuous, the inequality $P(C) \wedge Q(C) < \delta$ also implies $P(C) \vee Q(C) < \xi(\delta)$, where $\xi(\delta) \to 0$ as $\delta \to 0$. Therefore

$$P(\Delta_i) \;\wedge\; Q(\Delta_i) \;\geq\; 1 - m\xi(\delta) \;,$$

and by independence of the Δ_i's,

$$P(\Delta) \;\wedge\; Q(\Delta) \;\geq\; (1 - m\xi(\delta))^{n_B} \;.$$

We conclude that

$$\begin{aligned}\gamma_{n_B}(\pi) \;\geq\; & \left[\pi P\left\{\sum\nolimits_{i=n_A+1}^{n} X_{\tau_i} \leq \eta | \Delta\right\}\right. \\ & \left. + (1-\pi) Q\left\{\sum\nolimits_{i=n_A+1}^{n} X_{\tau_i} > \eta | \Delta\right\}\right] \exp\{-n_B \varepsilon_1(\delta)\} \quad (3.2.19)\end{aligned}$$

where $\varepsilon_1(\delta) \to 0$ as $\delta \to 0$. Our aim is to show that the first factor on the r.h.s. of (3.2.19) decreases at a rate which bounded away from the optimal ρ_m.

Step 3. Restoration of $\mathcal{S}_{n_B}$. Consider next a system $\mathcal{S}^{\Delta}_{n_B}$ with independent observations $Y^{\Delta}_{n_A+1}, \ldots, Y^{\Delta}_n$. Each Y^{Δ}_i takes values in the restriction $(\Delta_i, \mathcal{B}_{\Delta_i})$ of $(\mathcal{Y}, \mathcal{B})$ on Δ_i, with distributions given, for $C \in \mathcal{B}_{\Delta_i}$, by

$$P^{\Delta}_i(C) \;=\; \frac{P(C)}{P(\Delta_i)} \qquad \text{and} \qquad Q^{\Delta}_i(C) \;=\; \frac{Q(C)}{Q(\Delta_i)} \; .$$

For all $y \in \Delta_i$, (a version of) the likelihood ratio $dP^{\Delta}_i/dQ^{\Delta}_i$ is given by

$$\frac{dP^{\Delta}_i}{dQ^{\Delta}_i}(y) \;=\; \frac{Q(\Delta_i)}{P(\Delta_i)} \frac{dP}{dQ}(y) \; ,$$

and thus the log-likelihood ratio X^{Δ}_i satisfies

$$X^{\Delta}_i(y) \;=\; X(y) + \mu_i \; ,$$

where $\mu_i \stackrel{\triangle}{=} \log[Q(\Delta_i)/P(\Delta_i)]$.

The last relationship implies that if $g : \mathcal{Y} \mapsto \mathcal{U}_m$ is a deterministic LRQ, then so is its restriction g^{Δ}_i to Δ_i; and that if

$$\tau \;=\; (I_1, \ldots, I_m)$$

is a LRP corresponding to g, then

$$\tau^{\Delta}_i \;=\; (I_1 + \mu_i, \ldots, I_m + \mu_i)$$

is a LRP corresponding to g^{Δ}_i. In the special case $\tau = \tau_i$ (the only case of interest), the definition of Δ_i implies that the output of the quantizer g^{Δ}_i will a.s. lie in the set J_i; where, by the assumption $\tau_i \in \mathcal{R}_m(\delta)$, the size of J_i is at most $m-1$. Thus w.l.o.g., the partition τ^{Δ}_i can be coarsened by absorbing all intervals with indices not in J_i into intervals with indices in J_i, i.e., τ^{Δ}_i can be taken as a $|J_i|$–ary LRP. Writing $P^{\Delta}_{\tau_i}$ and $Q^{\Delta}_{\tau_i}$ for the distributions induced on J_i

by this trimmed version of τ_i^Δ, we have

$$(\forall u \in J_i) P_{\tau_i}^\Delta(u) \;=\; \frac{P_{\tau_i}(u)}{P(\Delta_i)} \qquad \text{and} \qquad Q_{\tau_i}^\Delta(u) \;=\; \frac{Q_{\tau_i}(u)}{Q(\Delta_i)}$$

and thus for all $u \in J_i$,

$$P_{\tau_i}^\Delta(u) \wedge Q_{\tau_i}^\Delta(u) \;\geq\; \delta \; . \tag{3.2.20}$$

(One could say that τ_i^Δ lies in $\mathcal{T}_{J_i}(\delta)$, where the latter class is defined w.r.t. the observation space $(\Delta_i, \mathcal{B}_{\Delta_i})$ and the base measures P_i^Δ and Q_i^Δ.) Furthermore, if $X_{\tau_i}^\Delta$ is the post-quantization log-likelihood ratio corresponding to τ_i^Δ, then for all $x \in \boldsymbol{R}$,

$$P_i^\Delta\{X_{\tau_i}^\Delta = x\} \;=\; P\{X_{\tau_i} = x - \mu_i | \Delta_i\} \tag{3.2.21}$$

and

$$Q_i^\Delta\{X_{\tau_i}^\Delta = x\} \;=\; Q\{X_{\tau_i} = x - \mu_i | \Delta_i\} \tag{3.2.22}$$

Step 4. Relevance of $\mathcal{S}_{n_B}^\Delta$. In $\mathcal{S}_{n_B}$, the events $\{X_{\tau_i} = x_i\} \cap \Delta_i$ are independent and thus

$$P\left(\bigcap\nolimits_{i=n_A+1}^n \{X_{\tau_i} = x_i\} \Big| \bigcap\nolimits_{i=n_A+1}^n \Delta_i\right) \;=\; \prod_{i=n_A+1}^n P\{X_{\tau_i} = x_i | \Delta_i\} \; .$$

The r.h.s. can be rewritten in terms of (3.2.21), hence

$$P\left(\bigcap\nolimits_{i=n_A+1}^n \{X_{\tau_i} = x_n\} \Big| \Delta\right) \;=\; \prod_{i=n_A+1}^n P_i^\Delta\{X_{\tau_i}^\Delta = x_i + \mu_i\} \; ,$$

and the same holds (by (3.2.22)) with Q replacing P. Thus letting $\eta_n = \eta + \sum_{i=n_A+1}^n \mu_i$ and defining product measures $P^\Delta = P_{n_A+1}^\Delta \times \cdots \times P_n^\Delta$ and $Q^\Delta = Q_{n_A+1}^\Delta \times \cdots \times Q_n^\Delta$, we obtain

$$\pi P\left\{\sum\nolimits_{i=n_A+1}^n X_{\tau_i} \leq \eta | \Delta\right\} \;+\; (1-\pi) Q\left\{\sum\nolimits_{i=n_A+1}^n X_{\tau_i} > \eta | \Delta\right\} =$$
$$\pi P^\Delta\left\{\sum\nolimits_{i=n_A+1}^n X_{\tau_i}^\Delta \leq \eta_n\right\} \;+\; (1-\pi) Q^\Delta\left\{\sum\nolimits_{i=n_A+1}^n X_{\tau_i}^\Delta > \eta_n\right\} \; .$$

The r.h.s. of this equation is the overall probability of error in $\mathcal{S}_{n_B}^{\Delta}$ when the fusion center performs a log-likelihood ratio threshold test with threshold η_n. This is clearly no less than the Bayes error probability $\gamma_{n_B}^{\Delta}(\pi)$, which is obtained by resetting η_n to $\eta = \log[(1-\pi)/\pi]$. Thus in conjunction with (3.2.19), we conclude that

$$\gamma_{n_B}(\pi) \;\geq\; \exp\{-n_B\varepsilon_1(\delta)\}\gamma_{n_B}^{\Delta}(\pi)\ . \tag{3.2.23}$$

Step 5. Error performance of $\mathcal{S}_{n_B}^{\Delta}$. We will use the results of Section 3.1 to obtain a lower bound for $\gamma_{n_B}(\pi)$. In what follows, the entities $\Psi_\tau^{\Delta}(\cdot)$, $M_\tau^{\Delta}(\cdot)$, $\theta_{n_B}^{\Delta}$ and $\tilde{F}_{n_B}^{\Delta}$ will be defined—with obvious modifications—as in Section 3.1.A.

The idea is to adapt Lemma 3.1 and its corollary to $\mathcal{S}_{n_B}^{\Delta}$. By virtue of (3.2.20), each $X_{\tau_i}^{\Delta}$ is absolutely bounded by $\log(1/\delta)$, and thus condition C1 in the hypothesis of Lemma 3.1 is satisfied with $X_{\tau_i}^{\Delta}$ replacing X_{τ_i}. Condition C2 can be checked as follows.

Let i be such that $|J_i| \geq 2$, and take $r > 0$ as in the proof of Lemma 3.5:

$$Q\{X \leq x\} \;\geq\; r \qquad \text{and} \qquad Q\{X \geq x'\} \;\geq\; r\ .$$

Without loss of generality, we can take δ sufficiently small so that $m\xi(\delta) \leq r/2$. Then

$$Q_i^{\Delta}\{X_i^{\Delta} \leq x + \mu_i\} \;=\; \frac{Q(\{X_i \leq x\} \cap \Delta_i)}{Q(\Delta_i)} \;\geq\; \frac{r}{2Q(\Delta_i)} \;\geq\; \frac{r}{2}$$

uniformly in i, and similarly $Q_i^{\Delta}\{X_i^{\Delta} \geq x'\} \geq r/2$. Retracing the proof of the lemma with $r/2$ replacing r, we obtain $|X_{\tau_i}^{\Delta}(|J_i|) - X_{\tau_i}^{\Delta}(1)| \;\geq\; b(x+\mu_i, x'+\mu_i, r/2) > 0$. Continuing the argument using (3.2.20), we conclude that

$$\mathrm{Var}_{\tilde{Q}}[X_{\tau_i}^{\Delta}] \;\geq\; b^2(x+\mu_i, x'+\mu_i, r/2)\frac{\delta}{2} \;>\; 0\ .$$

Thus condition C2 is satisfied by $X^{\Delta}_{\tau_i}$ provided i is such that $|J_i| \geq 2$. Let $n_{B'}$ be the number of such indices i, and observe that the remaining $n_B - n_{B'}$ indices are of no interest because $X^{\Delta}_{\tau_i} = 0$ a.s. when $|J_i| = 1$. We can thus apply Lemma 3.1 to conclude that

$$\liminf_{n\to\infty} \sqrt{n_B} \int_{\boldsymbol{R}} \exp\{-|x|\} d\tilde{F}^{\Delta}_{n_B}(x) \geq \liminf_{n\to\infty} \sqrt{n_{B'}} \int_{\boldsymbol{R}} \exp\{-|x|\} d\tilde{F}^{\Delta}_{n_{B'}}(x) > 0 \,, \tag{3.2.24}$$

where $\tilde{F}^{\Delta}_{n_B}(\cdot) = \tilde{F}^{\Delta}_{n_{B'}}(\cdot)$ by the earlier remark.

In order to combine (3.2.24) with (3.1.8), we need a lower bound on $M^{\Delta}_{n_B}(\theta^{\Delta}_{n_B})$. We obtain this by writing, as in (3.1.2),

$$M^{\Delta}_{n_B}(\theta^{\Delta}_{n_B}) \;\geq\; \prod_{i=n_A+1}^{n} \min_{\theta\in(0,1)} \Psi^{\Delta}_{\tau_i}(\theta) \;, \tag{3.2.25}$$

and then by examining each $\Psi^{\Delta}_{\tau_i}(\theta)$ separately. We have, for $\theta \in (0,1)$,

$$\begin{aligned}
\Psi^{\Delta}_{\tau_i}(\theta) &= \sum_{u\in J_i} [P^{\Delta}_{\tau_i}(u)]^{\theta} [Q^{\Delta}_{\tau_i}(u)]^{1-\theta} \\
&= [P(\Delta_i)]^{-\theta} [Q(\Delta_i)]^{\theta-1} \sum_{u\in J_i} [P_{\tau_i}(u)]^{\theta} [Q_{\tau_i}(u)]^{1-\theta} \\
&\geq \sum_{u\in J_i} [P_{\tau_i}(u)]^{\theta} [Q_{\tau_i}(u)]^{1-\theta} \\
&= \Psi_{\tau_i}(\theta) - \sum_{u\notin J_i} [P_{\tau_i}(u)]^{\theta} [Q_{\tau_i}(u)]^{1-\theta} \\
&\geq \Psi_{\tau_i}(\theta) - \sum_{u\notin J_i} P_{\tau_i}(u) \,\vee\, Q_{\tau_i}(u) \;.
\end{aligned}$$

Now, $\Psi_{\tau_i}(\theta) \geq \exp\{-\rho(P_{\tau_i}, Q_{\tau_i})\}$ (which is bounded away from zero), and $P_{\tau_i}(u) \vee Q_{\tau_i}(u) \leq \xi(\delta)$ for $u \notin J_i$. Thus

$$\Psi^{\Delta}_{\tau_i}(\theta) \;\geq\; \exp\{-\rho(P_{\tau_i}, Q_{\tau_i}) - \varepsilon_2(\delta)\} \;,$$

where $\varepsilon_2(\delta) \to 0$ as $\delta \to 0$. Since $\tau_i \in \mathcal{R}_m(\delta)$, we can invoke Lemma 3.6 to obtain

$$\Psi^{\Delta}_{\tau_i}(\theta) \;\geq\; \exp\{-\rho_{m-1} - \varepsilon_3(\delta)\} \;,$$

where again $\varepsilon_3(\delta) \to 0$ as $\delta \to 0$. This holds for all $s \in (0,1)$, and consequently (3.2.23) yields

$$M^{\Delta}_{n_B}(\theta^{\Delta}_{n_B}) \;\geq\; \exp\{-n_B(\rho_{m-1} + \varepsilon_3(\delta))\} \;.$$

Using this lower bound on $M^{\Delta}_{n_B}(\theta^{\Delta}_{n_B})$ in conjunction with (3.2.24) and (3.1.8), we obtain

$$\liminf_{n\to\infty} \gamma^{\Delta}_{n_B}(\pi)\sqrt{n_B}\exp\{n_B(\rho_{m-1} + \varepsilon_3(\delta))\} \;>\; 0 \;. \tag{3.2.26}$$

Step 6. Conclusion. In the light of (3.2.23) and (3.2.26), we have

$$\liminf_{n\to\infty} \gamma_{n_B}(\pi)\sqrt{n_B}\exp\{n_B(\rho_{m-1} + \varepsilon_4(\delta))\} \;>\; 0 \tag{3.2.27}$$

with $\varepsilon_4(\delta) \to 0$ as $\delta \to 0$. Since the error exponent obtained by $\mathcal{S}_{n_B}$ is at best only marginally greater than ρ_{m-1}, system $\mathcal{S}_{n_B}$ is markedly inferior to the best identical-quantizer system of the same size. Hence it is natural to expect that the size n_B of $\mathcal{S}_{n_B}$ is bounded in n. As we will see, this is indeed true and implies that $\gamma^*_n(\pi)/\gamma^\diamond_n(\pi)$ is bounded from below.

Combining (3.2.17) with (3.2.18) and (3.2.27), we obtain

$$\liminf_{n\to\infty} \gamma^*_n(\pi)\sqrt{n_A n_B}\exp\{n_A\rho_m + n_B(\rho_{m-1} + \varepsilon_4(\delta))\} \;>\; 0 \;,$$

where we assume that $n_B \geq 1$. By (3.1.12), the quantity $\gamma^\diamond_n(\pi)\sqrt{n}\exp\{n\rho_m\}$ is bounded from above, hence there exists $c > 0$ such that

$$\frac{\gamma^*_n(\pi)}{\gamma^\diamond_n(\pi)} \;\geq\; c\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\exp\{n_B(\rho_m - \rho_{m-1} - \varepsilon_4(\delta))\}$$

for all sufficiently large n. By Lemma 3.3, the difference $\rho_m - \rho_{m-1} - \varepsilon_4(\delta)$ can be made larger than $(\rho_m - \rho_{m-1})/2 > 0$ by choosing δ suitably small, and thus

$$\frac{\gamma_n^*(\pi)}{\gamma_n^\diamond(\pi)} \geq \frac{c}{\sqrt{n_B}} \exp\{n_B(\rho_m - \rho_{m-1})/2\} \tag{3.2.28}$$

for all sufficiently large n. As the r.h.s. of the previous inequality is bounded away from zero for all $n_B \geq 1$, the theorem is proved. Observe further that although the r.h.s. of (3.2.28) increases (in n_B) to infinity, we always have $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi) \leq 1$; thus n_B must also be bounded from above.

The case $n_B = 0$ is straightforward. Indeed, (3.2.16) follows directly from (3.2.18) and (3.1.12). □

Theorem 3.1 refines the equality of error exponents $e_{\mathrm{B}}^*(\pi)$ and $e_{\mathrm{B}}^\diamond(\pi)$ by showing that the actual ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ is bounded away from zero provided $P \equiv Q$. No further conditions, such as the boundedness assumption (3.1.9) used in [30], are needed. The latter fact was indeed conjectured in [29], together with the possibility that the optimal LRQ's in $\mathcal{S}_n$ employ thresholds that are confined in a fixed (as $n \to \infty$) finite interval. We do not have proof of the second—and quite stronger—conjecture; our proof only establishes that the number of LRQ's with one or more thresholds outside a certain fixed interval $[-T(\delta), T(\delta)]$ must be bounded in n.

Also related to the composition of an optimal set $(\tau_1, \ldots, \tau_n)$ of LRP's is the following observation.

Corollary 3.2. *For $\varepsilon > 0$, the number $K_n(\varepsilon)$ of optimal LRP's τ_i such that*

$$\rho(P_{\tau_i}, Q_{\tau_i}) < \rho_m - \varepsilon \tag{3.2.29}$$

is bounded.

Proof: Suppose that, in the proof of Theorem 3.1, n_ε of the first n_A quantizers satisfy (3.2.29). Then (3.2.18) can be strengthened into

$$\liminf_{n\to\infty} \gamma_{n_A}(\pi)\sqrt{n_A}\exp\{n_A\rho_m - n_\varepsilon\varepsilon\} \;>\; 0 \;,$$

in which case (3.2.28) becomes

$$\liminf_{n\to\infty} \frac{\gamma_n^*(\pi)}{\gamma_n^\diamond(\pi)} \;\geq\; \frac{c}{\sqrt{n_B}}\exp\{n_B[(\rho_m - \rho_{m-1})/2] + n_\varepsilon\varepsilon\} \;.$$

From $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi) \leq 1$ it follows again that n_ε, n_B, and $K_n(\varepsilon)$ (which is at most $n_\varepsilon + n_B$) are bounded. The argument is easily adapted to the case $n_B = 0$. □

The significance of this result becomes transparent when the sample space $\mathcal{Y}$ is finite. In that case, there are at most $\binom{|\mathcal{Y}|+m-1}{m-1}$ deterministic LRQ's, and thus only finitely many pairs (P_τ, Q_τ) of output distributions. Clearly, we can choose ε small enough so that $n - K_n(\varepsilon)$ is the exact number of quantizers in the system that achieve ρ_m. Then by Corollary 3.2, the number of remaining quantizers (the ones that do not achieve ρ_m) must be bounded. If, in particular, there is a unique output distribution pair that achieves ρ_m, then an optimal system exists in which the same quantizer is used by all but a bounded number of sensors.

As we pointed out in Section 1.1, our numerical results have often indicated that the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ is close to unity. It is thus possible that under certain conditions as yet unknown to us, the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ tends to unity as n approaches infinity. Nevertheless, we have a counterexample (appearing in Appendix C) which shows that this cannot be true in general. In that example, the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$ is smaller than a constant $r < 1$ infinitely often in n.

Chapter 4

Distributed detection in additive correlated Gaussian noise

In the second part of this dissertation, we consider the two-sensor distributed detection system $\mathcal{S}_2$ depicted in Figure 4.1, where the null and alternative distributions are Gaussian, differing in the mean. This is the standard hypothesis testing model for detecting a known signal in Gaussian noise.

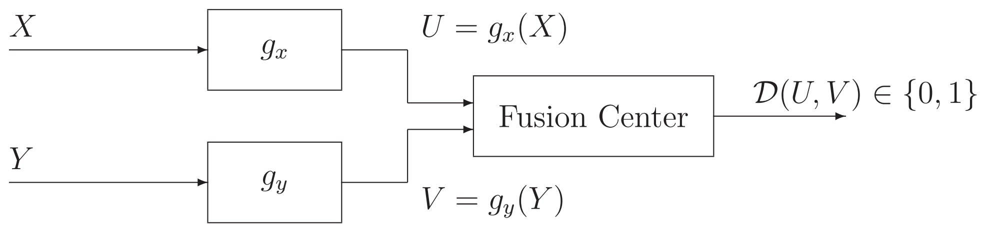

Figure 4.1: Distributed detection in two-sensor system $\mathcal{S}_2$.

4.1 Introduction

When the observations are correlated across sensors, the optimal design of a distributed detection system is in general very complicated. Early results [28] have shown that the joint optimization of local quantizers g_x, g_y, and global decision $\mathcal{D}$ may yield solutions in which g_x and g_y are not based on marginal likelihood ratio tests. This is one instance where distributed detection departs from the traditional statistical framework where likelihood ratios are sufficient for most purposes.

As in the first part of this dissertation, the complexity of the optimal design problem can be reduced by introducing additional statistical structure in the observations. For an extreme example, let the local observations be equal with probability 1 under each hypothesis. Then there is no loss of optimality in making a final decision based on only one local observation, which can be obviously handled by a classical detection technique. A more interesting (and nontrivial) case is when the null and alternative distributions are spatially correlated Gaussian, differing in the mean. This is the standard model used for the detection of a deterministic signal in Gaussian noise.

Due to the special structure of the additive Gaussian noise model, one might expect that the optimal distributed system exhibits a simple form in which g_x and g_y can be represented as threshold-type functions of the local (i.e., marginal) likelihood ratio. Unfortunately, this is not in general true; a counterexample in which each sensor draws two local observations is constructed in Section 4.4. Thus the optimal design problem is still prohibitively complex even if the null and alternative hypothesis distributions are spatially correlated Gaussian, differing in the mean.

The initial goal of our investigation of the optimal design problem was the characterization of noise models for which the optimal system employs marginal likelihood ratio tests. In the setup where each sensor draws one local observation, we succeeded in obtaining a sufficient condition on the noise mean and covariance under which the optimal binary quantizers are contiguous partitions of the marginal observation space. Since the marginal likelihood ratio is a linear function of the local observation (X or Y), this result implies that g_x and g_y are threshold-type functions of the marginal likelihood ratio. It also reduces the optimization to identifying break points (thresholds) in the marginal observation space.

We also examined whether the aforementioned sufficient condition is also necessary, and found that violation of this condition may in certain—but not all—cases render the contiguous marginal likelihood ratio partition suboptimal. We reached this conclusion by examining the special case where the noise marginals are the same for both sensors; the sufficient condition is then equivalent to negative correlation between X and Y. We found that for values of the correlation coefficient $\rho(X, Y)$ close to -1, local quantizers based on non-contiguous likelihood ratio partitions outperform those based on contiguous likelihood ratio partitions. We were not able to establish the same for $\rho(X, Y)$ close to 0^-.

The above discussion motivates the following question. Assuming that the aforementioned sufficient condition is valid, does symmetry in the signal and noise models (same marginal for both sensors) imply symmetry in the optimal solution, with g_x and g_y being identical contiguous binary partitions of the real line? We found that this is indeed true, and in such case, optimal design is further simplified.

Next, we considered a large-sample variant in which the sensors observe a sequence of temporally i.i.d observations $(X_1, Y_1), \ldots, (X_N, Y_N)$ and locally compress their observations using a quantizer of variable (in N) size, i.e., $U = g_{x,N}(X_1, \ldots, X_N)$ and $V = g_{y,N}(Y_1, \ldots, Y_N)$. After transmitting the local decisions to the fusion center, a final decision about the nature of the source is made.

In the case of binary local quantizers $g_{x,N}$ and $g_{y,N}$, earlier results had shown that the Bayes error exponent is smaller than that achieved by the centralized system; while the Neyman-Pearson error exponent remains unaffected [23]. However, by employing asymptotically zero-rate compression in each sensor, defined by $\limsup_{N\to\infty}(1/N)\log(\|g_{x,N}\| \vee \|g_{y,N}) = 0$ and $\lim_{N\to\infty} \|g_{x,N}\| \wedge \|g_{y,N}\| = \infty$, the Bayes exponent (in N) of the distributed system becomes equivalent to that of the centralized system. We were able to prove that by employing a suitable polynomial (in N) number of quantization levels, the performance gap between the distributed and centralized systems can be reduced to the extent that the ratio between the actual error probabilities is bounded.

The material in this chapter is arranged as follows. The sufficient condition under which the contiguous LRP is optimal is derived in Section 4.2, followed by a counterexample where violation of this condition renders the contiguous LRP suboptimal. The optimality of a symmetric solution in the case of symmetric signal and noise models is proved in Section 4.3. In Section 4.4, an example is constructed where the marginal LRP is not optimal. Section 4.5 is devoted to the discussion of the temporal asymptotics, and contains the result that under suitable asymptotically zero rate compression, the ratio of the error probabilities of the centralized and distributed systems can be confined between two constants. Some supplementary results are placed in Appendix D.

Throughout this chapter, the observation statistics are denoted by

$$H_0 : P_{xy} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right) \tag{4.1.1}$$

$$H_1 : Q_{xy} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \eta \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right), \tag{4.1.2}$$

with the obvious i.i.d. extension in the asymptotic case.

4.2 Optimality of contiguous marginal likelihood ratio partition

In this section, a condition on the mean vector and covariance matrix of the Gaussian noise is derived which guarantees the optimality of marginal scalar quantizers with contiguous partitions.

Lemma 4.1. *It suffices to employ quantizers of X and Y with marginal partitions, provided*

$$\sigma_{xy}(\eta\sigma_x^2 - \mu\sigma_{xy})(\mu\sigma_y^2 - \eta\sigma_{xy}) \geq 0. \tag{4.2.3}$$

Proof: Let $\gamma_{\text{AND}}(\pi, \mathcal{A} \times \mathcal{B})$ be the Bayes error under AND fusion, where π is the prior probability of the null hypothesis and $\mathcal{A} \times \mathcal{B}$ is the acceptance region for the alternative hypothesis. Then

$$\gamma_{\text{AND}}(\pi, \mathcal{A} \times \mathcal{B}) = (1-\pi) + \int_{\mathcal{A}} \left[\pi \int_{\mathcal{B}} P_{xy}(x,y)dy - (1-\pi)\int_{\mathcal{B}} Q_{xy}(x,y)dy\right] dx.$$

Hence, if for any given $\mathcal{B}$,

$$f_{\mathcal{B}}(x) \triangleq \log \frac{\int_{\mathcal{B}} P_{xy}(x,y)dy}{\int_{\mathcal{B}} Q_{xy}(x,y)dy}$$

is a monotone function of x, then the optimal partition $\mathcal{A}$ w.r.t. the given $\mathcal{B}$ should be of the desired form, i.e., a contiguous binary partition of the real line. By symmetry, the optimal $\mathcal{B}$ should also be a contiguous marginal likelihood ratio partition. (A similar argument can be applied to the cases of OR and XOR fusion.)

The monotonicity of the function $f_{\mathcal{B}}(x)$ can be checked by taking the derivative of $f_{\mathcal{B}}(x)$. In differentiating each of the integrals in the expression for $f_{\mathcal{B}}(x)$ w.r.t. x, the order of integration and differentiation can be interchanged. Let $\Delta = \sigma_x^2\sigma_y^2 - \sigma_{xy}^2$; we then have

$$\begin{aligned}
f'_{\mathcal{B}}(x) &= \frac{\int_B \left(-\frac{\sigma_y^2}{\Delta}x + \frac{\sigma_{xy}}{\Delta}y\right) e^{-\frac{\sigma_y^2}{2\Delta}x^2 - \frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy} dy}{\int_B e^{-\frac{\sigma_y^2}{2\Delta}x^2 - \frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy} dy} \\
&- \frac{\int_B \left(-\frac{\sigma_y^2}{\Delta}(x-\mu) + \frac{\sigma_{xy}}{\Delta}(y-\eta)\right) e^{-\frac{\sigma_y^2}{2\Delta}(x-\mu)^2 - \frac{\sigma_x^2}{2\Delta}(y-\eta)^2 + \frac{\sigma_{xy}}{\Delta}(x-\mu)(y-\eta)} dy}{\int_B e^{-\frac{\sigma_y^2}{2\Delta}(x-\mu)^2 - \frac{\sigma_x^2}{2\Delta}(y-\eta)^2 + \frac{\sigma_{xy}}{\Delta}(x-\mu)(y-\eta)} dy} \\
&= \frac{\int_B \left(-\frac{\sigma_y^2}{\Delta}x + \frac{\sigma_{xy}}{\Delta}y\right) e^{-\frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy} dy}{\int_B e^{-\frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy} dy} \\
&- \frac{\int_B \left(-\frac{\sigma_y^2}{\Delta}x + \frac{\sigma_{xy}}{\Delta}y + \frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\Delta}\right) e^{-\frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy + \frac{\eta\sigma_x^2 - \mu\sigma_{xy}}{\Delta}y} dy}{\int_B e^{-\frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy + \frac{\eta\sigma_x^2 - \mu\sigma_{xy}}{\Delta}y} dy}.
\end{aligned}$$

Let

$$\widetilde{P}(y) \triangleq \frac{\exp\left\{-\frac{\sigma_y^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy\right\}}{\int_{\mathcal{B}} \exp\left\{-\frac{\sigma_y^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy\right\} dy},$$

and

$$\tilde{Q}(y) \triangleq \frac{\exp\left\{-\frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy + \frac{\eta\sigma_x^2-\mu\sigma_{xy}}{\Delta}y\right\} dy}{\int_{\mathcal{B}} \exp\left\{-\frac{\sigma_x^2}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy + \frac{\eta\sigma_x^2-\mu\sigma_{xy}}{\Delta}y\right\} dy}.$$

Then

$$f'_{\mathcal{B}}(x) \;=\; \frac{\sigma_{xy}}{\Delta}(E_{\widetilde{P}}[Y] - E_{\tilde{Q}}[Y]) - \frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\Delta}.$$

Note that $\tilde{Q}(y)$ is actually a tilted distribution of $\widetilde{P}(y)$, i.e.

$$\tilde{Q}(y) \;=\; \widetilde{P}^{(\theta)}(y),$$

for $\theta \triangleq (\eta\sigma_x^2 - \mu\sigma_{xy})/\Delta$ (cf. (2.2.19)). Therefore,

$$f'_{\mathcal{B}}(x) \;=\; \frac{\sigma_{xy}}{\Delta}(E_{\widetilde{P}}[Y] - E_{\widetilde{P}^{(\theta)}}[Y]) - \frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\Delta}.$$

Finally, in light of the following inequalities:

$$\theta > 0 \;\Rightarrow\; E_{\widetilde{P}}[Y] \le E_{\widetilde{P}^{(\theta)}}[Y]$$

$$\theta < 0 \;\Rightarrow\; E_{\widetilde{P}}[Y] \ge E_{\widetilde{P}^{(\theta)}}[Y]$$

we obtain that the condition (4.2.3) implies the monotonicity of $f_{\mathcal{B}}(x)$. □

It is worth mentioning that (4.2.3) is also a necessary condition for $f_{\mathcal{B}}(x)$ (defined in the above proof) to be a monotone function of x. Necessity is shown in Lemma D.1.

As we mentioned before, if condition (4.2.3) holds, the system design will be much simpler: one only needs to consider scalar quantizers with contiguous partition regions. However, the violation of (4.2.3) may result in non-contiguous optimal partitions. This becomes apparent in the following (degenerate) example.

Example. Assume equal priors and $\|g_x\| = \|g_y\| = 2$. Let $\sigma_x^2 = \sigma_y^2 = \mu = \eta = 1$ and $\sigma_{xy} = -1$. Observe that the entire probability mass is accumulated on a straight line under each hypothesis.

There are three nontrivial possibilities for the fusion rule: AND fusion, OR fusion and XOR fusion. These are shown schematically in Figures 4.2(a), 4.3(b) and 4.3(a) (or 4.4(a)), respectively. In each figure, x and y coordinate axes are suppressed, horizontal and vertical lines correspond to (noncontiguous) partition boundaries, slanted lines are supports of P_{xy} and Q_{xy} (as marked), and shaded rectangles constitute the acceptance region for the alternative hypothesis H_1.

For Bayes testing, the performance analysis under AND and OR fusion is exactly the same. Consequently, if a contiguous partition is outperformed by a contiguous one under AND fusion, the same is true under OR fusion, as well.

We will thus consider AND fusion only, shown in Figure 4.2(a). The Bayes error probability is the average of the probabilities carried by the bold segments of the support lines. It can be seen that the noncontiguous partition of Figure 4.2(b) (with AND fusion) yields a lower Bayes error probability.

Under XOR fusion, we need to consider two types of contiguous partitions, shown in Figures 4.3(a) and 4.4(a). For the case shown in Figure 4.3(a), it can be seen that the Bayes error probability is exactly the same as that obtained from the acceptance region of Figure 4.3(b). The latter figure represents OR fusion, which has been already considered.

This leaves us with the case shown in Figure 4.4(a). Again, we can construct a noncontiguous partition such as that in Figure 4.4(b) which yields a lower

Bayes error probability under XOR fusion. □

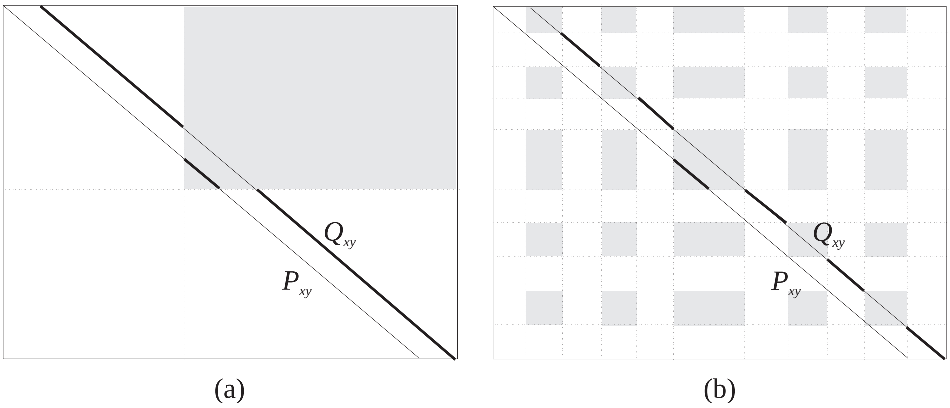

Figure 4.2: (a) Optimal contiguous LRP under AND fusion. (b) A non-contiguous improvement. The shaded area is the acceptance region for H_1.

Remark. The above example clearly represents an extreme case where either of the local observations is a sufficient statistic for centralized testing. The same effect, however, can be obtained by choosing $\sigma_{xy} \approx -1$ and applying a continuity argument. A nondegenerate counterexample is constructed in Appendix D.2.

4.3 Optimal partition thresholds for symmetric noise model

The question of whether a symmetric optimal solution exists for a sensor network with symmetric statistics has received a fair amount of attention in the literature [12, 10, 11, 30, 34, 36]. As discussed in Section 1.1, such a symmetric optimal solution does not always exist (even in the independent case), which is contrary to intuition. One important exception is the case of the additive Gaussian noise

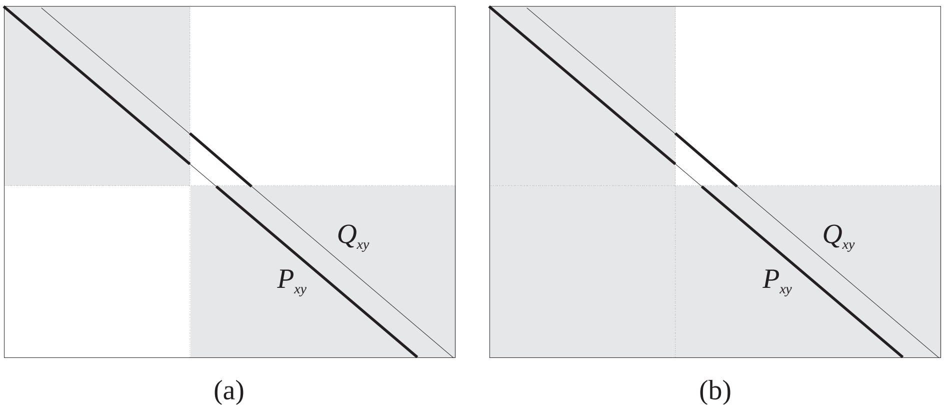

Figure 4.3: (a) Contiguous LRP under XOR fusion. (b) Equivalent contiguous LRP under OR fusion. The shaded area is the acceptance region for H_1.

model with spatially uncorrelated components, for which the existence of symmetric optimal solutions was substantiated in [36]. By using a similar technique, we will extend this result to the case of spatially correlated noise components, under the design constraint that the (binary) local quantizers are contiguous partitions of the observation space.

Theorem 4.1. *Let the signal and noise models be symmetric, and consider binary quantization. If each local quantizer uses a contiguous partition of the observation space, then the partition thresholds are equal.*

***Proof*:** The symmetry assumption implies $\mu = \eta$ and $\sigma_x^2 = \sigma_y^2$. Assume w.l.o.g. $\sigma_x^2 = \sigma_y^2 = 1$. We consider the the case of OR fusion in detail; the analysis for AND and XOR fusion is virtually identical.

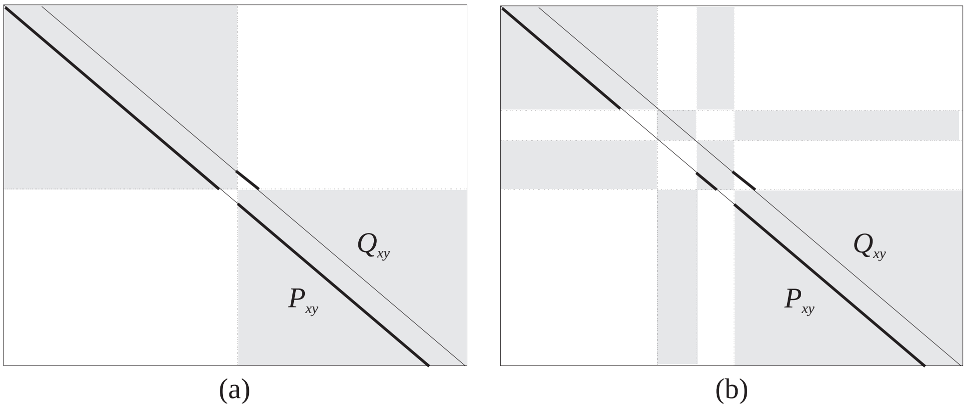

Figure 4.4: (a) Contiguous LRP under XOR fusion. (b) A non-contiguous improvement. The shaded area is the acceptance region for H_1.

Let $(-\infty, s) \times (-\infty, t)$ be the acceptance region for null hypothesis. Define

$$P(x,y) \triangleq \frac{1}{2\pi\Delta}\exp\left\{-\frac{1}{2\Delta}x^2 - \frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}xy\right\}$$

$$Q(x,y) \triangleq \frac{1}{2\pi\Delta}\exp\left\{-\frac{1}{2\Delta}(x-\mu)^2 - \frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(x-\mu)(y-\mu)\right\}$$

where $\Delta \triangleq 1 - \sigma_{xy}^2$. Then the corresponding Bayes error is

$$\gamma(s,t) = \pi_0\left(1 - \int_{-\infty}^{t}\int_{-\infty}^{s} P(x,y)dxdy\right) + (1-\pi_0)\int_{-\infty}^{t}\int_{-\infty}^{s} Q(x,y)dxdy, \tag{4.3.4}$$

where π_0 is the prior probability of null hypothesis. By taking the derivatives of (4.3.4) w.r.t. s, we obtain

$$\begin{aligned} s &= f(s,t) \\ &\triangleq \frac{\mu}{2} + \frac{\Delta}{\mu}\log\frac{\pi_0}{1-\pi_0} - \frac{\Delta}{\mu}\log\int_{-\infty}^{t} e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)}dy \\ &\quad + \frac{\Delta}{\mu}\log\int_{-\infty}^{t} e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta}sy}dy \end{aligned}$$

and similarly, $t = f(t, s)$. Therefore, the optimal thresholds should satisfy

$$s = f(s, t) \tag{4.3.5}$$

$$t = f(t, s) \tag{4.3.6}$$

To prove the theorem, it suffices to show the solutions of the above equation pairs should be located on the line $s = t$.

Let $h_c(s) \triangleq f(s, c-s)$ for some constant c. Then

$$\begin{aligned}
\frac{\partial h_c(s)}{\partial s} &= \left[-\frac{\Delta}{\mu} \frac{\int_{-\infty}^{c-s} \frac{\sigma_{xy}}{\Delta}(y-\mu) e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)} dy}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)} dy} \right. \\
&\quad \left. + \frac{\Delta}{\mu} \frac{\int_{-\infty}^{c-s} \frac{\sigma_{xy}}{\Delta} y e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta} sy} dy}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta} sy} dy} \right] \\
&\quad - \left[-\frac{\Delta}{\mu} \frac{e^{-\frac{1}{2\Delta}(c-s-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(c-s-\mu)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)} dy} \right. \\
&\quad \left. + \frac{\Delta}{\mu} \frac{e^{-\frac{1}{2\Delta}(c-s)^2 + \frac{\sigma_{xy}}{\Delta} s(c-s)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta} sy} dy} \right] \\
&= \left[-\frac{\Delta}{\mu} \left(\frac{-\sigma_{xy} e^{-\frac{1}{2\Delta}(c-s-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(c-s-\mu)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)} dy} + \frac{\sigma_{xy}^2}{\Delta}(s-\mu) \right) \right. \\
&\quad \left. + \frac{\Delta}{\mu} \left(\frac{-\sigma_{xy} e^{-\frac{1}{2\Delta}(c-s)^2 + \frac{\sigma_{xy}}{\Delta} s(c-s)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta} sy} dy} + \frac{\sigma_{xy}^2}{\Delta} s \right) \right] \\
&\quad - \left[-\frac{\Delta}{\mu} \frac{e^{-\frac{1}{2\Delta}(c-s-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(c-s-\mu)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)} dy} \right. \\
&\quad \left. + \frac{\Delta}{\mu} \frac{e^{-\frac{1}{2\Delta}(c-s)^2 + \frac{\sigma_{xy}}{\Delta} s(c-s)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta} sy} dy} \right] \\
&= \frac{\Delta}{\mu}(1+\sigma_{xy}) \frac{e^{-\frac{1}{2\Delta}(c-s-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(c-s-\mu)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}(y-\mu)^2 + \frac{\sigma_{xy}}{\Delta}(s-\mu)(y-\mu)} dy} \\
&\quad - \frac{\Delta}{\mu}(1+\sigma_{xy}) \frac{e^{-\frac{1}{2\Delta}(c-s)^2 + \frac{\sigma_{xy}}{\Delta} s(c-s)}}{\int_{-\infty}^{c-s} e^{-\frac{1}{2\Delta}y^2 + \frac{\sigma_{xy}}{\Delta} sy} dy} + \sigma_{xy}^2
\end{aligned}$$

Let $g(u) \triangleq \dfrac{\exp\{-u^2/2\}}{\Phi(u)}$. From [36, inequality (12)], $g(t-s) - g(t) < s\sqrt{2\pi}$. Then

$$\begin{aligned}
\frac{\partial h_c(s)}{\partial s} &= \frac{\Delta(1+\sigma_{xy})}{\sqrt{2\pi\Delta}\mu}\left[g\left(\frac{(c-s-\mu)-\sigma_{xy}(s-\mu)}{\sqrt{\Delta}}\right) - g\left(\frac{(c-s)-\sigma_{xy}s}{\sqrt{\Delta}}\right)\right] \\
&\quad +\sigma_{xy}^2 \\
&< \frac{\Delta(1+\sigma_{xy})}{\sqrt{2\pi\Delta}\mu}\left[\frac{\mu(1-\sigma_{xy})}{\sqrt{\Delta}}\sqrt{2\pi}\right] + \sigma_{xy}^2 \\
&= 1 - \sigma_{xy}^2 + \sigma_{xy}^2 \\
&= 1 \qquad (4.3.7)
\end{aligned}$$

Suppose there exists (s', t') with $s' > t'$ satisfying (4.3.5) and (4.3.6). Let $c' = s' + t'$. Then

$$\begin{aligned}
1 = \frac{f(s',t') - f(t',s')}{s'-t'} &= \frac{f(s', c'-s') - f(c'-s', s')}{s' - (c'-s')} \\
&= \frac{h_{c'}(s') - h_{c'}(c'-s')}{2s'-c'} \\
&= h'_{c'}(\tilde{s}), \quad \text{for some } \tilde{s} \in (c'-s', s') \\
&< 1, \quad \text{from (4.3.7)}
\end{aligned}$$

which is a contradiction. □

In conjunction with Lemma 4.1, the above theorem implies the following corollary.

Corollary 4.1. *Let the signal and noise models be symmetric, and consider binary quantization. If $\sigma_{xy} \geq 0$, then an optimal solution exists in which both quantizers use the same contiguous partition of the observation space.*

4.4 Optimality of marginal likelihood ratio partitions in higher dimensions

The discussion in Section 4.2 revealed the possibility that non-contiguous marginal likelihood ratio partitions can outperform contiguous ones if (4.2.3) is violated. Still, marginal likelihood ratios remain sufficient statistics for quantization, since they are trivially equivalent to the data themselves.

In the case of two or more observations corrupted by additive noise, marginal likelihood ratios are no longer equivalent to the data sets, and their optimality is at issue. Indeed, it appears that in higher-dimensional additive Gaussian noise models exhibiting spatial dependence and temporal independence, suboptimality of marginal likelihood ratio partitions is the rule rather than the exception. We substantiate this for the case of two observations per sensor ($N = 2$). using techniques borrowed from the proof of Lemma 4.1

Observation 4.1. Suppose $\mathcal{A}\times\mathcal{B}$ is the optimal acceptance region for alternative hypothesis under AND fusion, where $\mathcal{A} \subset \mathcal{X}^2$ and $\mathcal{B} \subset \mathcal{Y}^2$. Also, let $p(x, \bar{x}, y, \bar{y})$ and $q(x, \bar{x}, y, \bar{y})$ be the null and alternative densities respectively. Then the Bayes error probability for equal priors is given by

$$\begin{aligned}
\gamma_2^*(1/2) &= \frac{1}{2}\int_{\mathcal{A}}\int_{\mathcal{B}} P(x, \bar{x}, y, \bar{y})(dyd\bar{y})(dxd\bar{x}) \\
&\quad +\frac{1}{2} - \frac{1}{2}\int_{\mathcal{A}}\int_{\mathcal{B}} Q(x, \bar{x}, y, \bar{y})(dyd\bar{y})(dxd\bar{x}) \\
&= \frac{1}{2} + \frac{1}{2}\int_{\mathcal{A}} \left(P(x, \bar{x}, \mathcal{B}) - Q(x, \bar{x}, \mathcal{B})\right) dxd\bar{x}\,.
\end{aligned}$$

Therefore, given a partition $\mathcal{B}$ on $\mathcal{Y}^2$, the optimal partition on $\mathcal{X}$ should be of the form:

$$f_{\mathcal{B}}(x, \bar{x}) \triangleq \frac{P(x, \bar{x}, \mathcal{B})}{Q(x, \bar{x}, \mathcal{B})} \gtrless \tau.$$

□

Counterexample. Given random vectors $(X, \overline{X}, Y, \overline{Y})$ in which (X, Y) and $(\overline{X}, \overline{Y})$ are independent, and where

$$(X, Y) \quad : \quad \begin{cases} H_0: \ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{xy} \\ \sigma_{xy} & 1 \end{pmatrix}\right) \\ H_1: \ \mathcal{N}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{xy} \\ \sigma_{xy} & 1 \end{pmatrix}\right) \end{cases}$$

and

$$(\overline{X}, \overline{Y}) \quad : \quad \begin{cases} H_0: \ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \bar{\sigma}_{xy} \\ \bar{\sigma}_{xy} & 1 \end{pmatrix}\right) \\ H_1: \ \mathcal{N}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & \bar{\sigma}_{xy} \\ \bar{\sigma}_{xy} & 1 \end{pmatrix}\right) \end{cases},$$

the marginal likelihood ratio test can be written as

$$x + \bar{x} \gtrless \tau_x, \qquad \text{and} \qquad y + \bar{y} \gtrless \tau_y.$$

Based on Observation 4.1, if the marginal likelihood ratio test is optimal, then there exists a function $g(\cdot)$ such that $f_{\mathcal{B}}(x, s - x) = g(s)$ under the (assumed) optimal partition $\mathcal{B}$, which should also be a marginal likelihood ratio partition.

Since $\mathcal{B}$ is a marginal likelihood ratio partition, it can be expressed in terms of the sum $t = y + \bar{y}$. Using the fact that X and $\bar{X}$ are independent and that Y and $\bar{Y}$ are conditionally independent given $(X, \bar{X})$, we obtain the following conditional distributions:

$$(Y|X=x) \ \sim \ \begin{cases} H_0: & \mathcal{N}(\sigma_{xy}x, 1-\sigma_{xy}^2) \\ H_1: & \mathcal{N}(1+\sigma_{xy}(x-1), 1-\sigma_{xy}^2) \end{cases},$$

$$(\overline{Y}|\overline{X}=\bar{x}) \ \sim \ \begin{cases} H_0: & \mathcal{N}(\bar{\sigma}_{xy}\bar{x}, 1-\bar{\sigma}_{xy}^2) \\ H_1: & \mathcal{N}(1+\bar{\sigma}_{xy}(\bar{x}-1), 1-\bar{\sigma}_{xy}^2) \end{cases}$$

and

$$(T=Y+\overline{Y}|X=x,\overline{X}=\bar{x}) \ \sim \ \begin{cases} H_0: & \mathcal{N}(\sigma_{xy}x+\bar{\sigma}_{xy}\bar{x}, \Delta) \\ H_1: & \mathcal{N}(\sigma_{xy}x+\bar{\sigma}_{xy}\bar{x}-\sigma_{xy}-\bar{\sigma}_{xy}+2, \Delta) \end{cases},$$

where $\Delta \triangleq 2-\sigma_{xy}^2-\bar{\sigma}_{xy}^2$. Therefore,

$$\begin{aligned} & f_{\mathcal{B}}(x,\bar{x}) \\ = \ & \frac{\int_{\mathcal{B}} P(x,\bar{x},t)dt}{\int_{\mathcal{B}} Q(x,\bar{x},t)dt} \\ = \ & \left(\int_{\mathcal{B}} e^{-\frac{1}{2\Delta}(t-\sigma_{xy}x-\bar{\sigma}_{xy}\bar{x})^2}dt\right) \Big/ \left(\int_{\mathcal{B}} e^{x+\bar{x}-1}e^{-\frac{1}{2\Delta}(t-\sigma_{xy}x-\bar{\sigma}_{xy}\bar{x}+\sigma_{xy}+\bar{\sigma}_{xy}-2)^2}dt\right) \\ = \ & \left(\int_{\mathcal{B}} e^{-\frac{1}{2\Delta}t^2+\frac{\sigma_{xy}x+\bar{\sigma}_{xy}\bar{x}}{\Delta}t}dt\right) \Big/ \left(\int_{\mathcal{B}} e^{-\frac{1}{2\Delta}t^2+\frac{\sigma_{xy}x+\bar{\sigma}_{xy}\bar{x}-\sigma_{xy}-\bar{\sigma}_{xy}+2}{\Delta}t+C(x,\bar{x})}dt\right), \end{aligned}$$

where

$$C(x,\bar{x}) \triangleq -\frac{(\sigma_{xy}+\bar{\sigma}_{xy}-2)^2}{2\Delta} + \frac{(\sigma_{xy}+\bar{\sigma}_{xy}-2)(\sigma_{xy}x+\bar{\sigma}_{xy}\bar{x})}{\Delta} + (x+\bar{x}-1).$$

Thus, if $\sigma_{xy}=\bar{\sigma}_{xy}=\sigma$, then $f_{\mathcal{B}}(x,s-x)=g(s)$ where

$$g(s) \triangleq \frac{\int_{\mathcal{B}} \exp\left\{-\frac{1}{2\Delta}t^2+\frac{\sigma s}{\Delta}t\right\}dt}{\int_{\mathcal{B}} \exp\left\{-\frac{1}{2\Delta}t^2+\frac{\sigma s}{\Delta}t\right\}\exp\left\{\frac{s-2}{1+\delta}\right\}dt}.$$

If, however, $\sigma_{xy} \neq \bar{\sigma}_{xy}$, then there does not appear to exist $g(s)$ satisfying

$f_{\mathcal{B}}(x, s-x) = g(s)$. $\square$

It is therefore possible that stationarity of signal and noise parameters is a necessary condition for the optimality of marginal LRP's in situations where two or more observations are taken per sensor and the noise is temporally independent. We doubt that stationarity alone is not sufficient for the optimality of marginal LRP's, but we have not succeeded in producing a counterexample.

4.5 Temporal asymptotics of error probabilities

In this section, we study the temporal asymptotics of two-sensor distributed detection used for testing a null hypothesis $H_0 : P_{\underline{xy}}$ against an alternative hypothesis $H_1 : Q_{\underline{xy}}$ based on a discrete data source $(\underline{X}, \underline{Y}) = \{(X_1, Y_1), \ldots, (X_N, Y_N)\}$, which is assumed to be i.i.d. in time. Again, at each time instant i, (X_i, Y_i) is bivariate Gaussian with different mean under each hypothesis. The final decision is based on the messages $U = g_{x,N}(X_1, \ldots, X_N)$ and $V = g_{y,N}(Y_1, \ldots, Y_N)$.

For the problem in hand, the centralized asymptotics can be easily computed. By adopting parameters in (4.1.1) and (4.1.2) for the temporal marginal distributions of (X_i, Y_i), we find that the centralized Neyman-Pearson type II error probability is

$$\begin{aligned} & \beta_2^*(\alpha, N, +\infty) \\ = \ & 1 - \Phi\left(\Phi^{-1}(\alpha) + \sqrt{2ND}\right) \\ \asymp \ & \frac{1}{\sqrt{N}V} \exp\left\{-ND - \sqrt{N}\Phi^{-1}(\alpha)V\right\}, \end{aligned} \quad (4.5.8)$$

where $V^2 = 2D = (\eta^2\sigma_x^2 + \mu^2\sigma_y^2 - 2\mu\eta\sigma_{xy})/(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)$, and $\asymp$ is defined by

$$a_n \asymp b_n \iff (\exists\ A \text{ and } B) \ni: A \geq \limsup_{n\to\infty} \frac{a_n}{b_n} \geq \liminf_{n\to\infty} \frac{a_n}{b_n} \geq B.$$

(Note: In subsequent notation, the centralized Neyman-Person type II error probability $\beta_2^*(\alpha, N, +\infty)$ is a function of the type I error bound α, number of local observations N, and the rate constraint $R = +\infty$; while the subscript "$_2$" represents the number of local sensors in the distributed system.)

We wish to investigate whether the error performance of the distributed system is at most a fixed multiple of that achieved by the centralized system in situations where $g_{x,N}$ and $g_{y,N}$ satisfy

$$\lim_{N\to\infty} \frac{1}{N} \log\left(\|g_{x,N}\| \vee \|g_{y,N}\|\right) = 0 \tag{4.5.9}$$

and

$$\lim_{N\to\infty} \|g_{x,N}\| \wedge \|g_{y,N}\| = \infty\,. \tag{4.5.10}$$

Conditions (4.5.9) and (4.5.10) represent (asymptotic) zero rate compression, which is denoted by $R = 0^+$ [27]. We succeed in proving this result under a slightly stronger assumption (than (4.5.10)), namely

$$\liminf_{N\to\infty} \frac{\|g_{x,N}\| \wedge \|g_{y,N}\|}{N^{1/2+b}} > 0\,, \tag{4.5.11}$$

for some $b > 0$.

By (4.5.8), it suffices to show that

$$\beta_2^*(\alpha, N, 0^+) \ \leq\ \frac{C_{NP}}{\sqrt{N}} \exp\left\{-ND - \sqrt{N}\Phi^{-1}(\alpha)V\right\},$$

for some constant C_{NP}.

Lemma 4.2. *Suppose (4.5.11) is true. Then*

$$\beta_2^*(\alpha, N, 0^+) \;\le\; \frac{C_{NP}}{\sqrt{N}} \exp\left\{-ND - \sqrt{N}\Phi^{-1}(\alpha)V\right\}$$

for some positive constant C_{NP}.

***Proof*:** We present the details under the assumption that $(\mu\sigma_y^2 - \eta\sigma_{xy}) > 0$ and $(\eta\sigma_x^2 - \mu\sigma_{xy}) > 0$. The proofs for the other cases are similar. Note that if $\mu\sigma_y^2 - \eta\sigma_{xy} = 0$ or $\eta\sigma_x^2 - \mu\sigma_{xy} = 0$, the statement of the lemma follows trivially because

$$\log\frac{P_{xy}(x,y)}{Q_{xy}(x,y)} = \frac{\sigma_x^2(\mu\sigma_y^2 - \eta\sigma_{xy})}{\mu(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)}\log\frac{P_x(x)}{Q_x(x)} + \frac{\sigma_y^2(\eta\sigma_x^2 - \mu\sigma_{xy})}{\eta(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)}\log\frac{P_y(y)}{Q_y(y)}\,,$$

and the joint log-likelihood ratio only depends one marginal log-likelihood ratio.

By (4.5.11), there exist $a > 0$ and N_0 such that for all $N > N_0$,

$$\|g_{x,N}\| \wedge \|g_{y,N}\| > N^{1/2+b}/a\ .$$

The marginal likelihood ratio pair $(\log[P_x(X)/Q_x(X)], \log[P_y(Y)/Q_y(Y)])$ is distributed as follows:

$$H_0: \quad \mathcal{N}\left(\begin{pmatrix} \dfrac{\mu^2}{2\sigma_x^2} \\ \dfrac{\eta^2}{2\sigma_y^2} \end{pmatrix}, \begin{pmatrix} \dfrac{\mu^2}{\sigma_x^2} & \dfrac{\sigma_{xy}}{\sigma_x^2\sigma_y^2}\mu\eta \\ \dfrac{\sigma_{xy}}{\sigma_x^2\sigma_y^2}\mu\eta & \dfrac{\eta^2}{\sigma_y^2} \end{pmatrix}\right)$$

$$H_1: \quad \mathcal{N}\left(\begin{pmatrix} -\dfrac{\mu^2}{2\sigma_x^2} \\ -\dfrac{\eta^2}{2\sigma_y^2} \end{pmatrix}, \begin{pmatrix} \dfrac{\mu^2}{\sigma_x^2} & \dfrac{\sigma_{xy}}{\sigma_x^2\sigma_y^2}\mu\eta \\ \dfrac{\sigma_{xy}}{\sigma_x^2\sigma_y^2}\mu\eta & \dfrac{\eta^2}{\sigma_y^2} \end{pmatrix}\right)$$

Let

$$Z \triangleq \frac{\log\dfrac{P_x^N(\underline{X})}{Q_x^N(\underline{X})} - N\dfrac{\mu^2}{2\sigma_x^2}}{\dfrac{\mu}{\sigma_x}\sqrt{N}}, \qquad \text{and} \qquad W \triangleq \frac{\log\dfrac{P_y^N(\underline{Y})}{Q_y^N(\underline{Y})} - N\dfrac{\eta^2}{2\sigma_y^2}}{\dfrac{\eta}{\sigma_y}\sqrt{N}}.$$

We plot W versus Z in Figure 4.5.

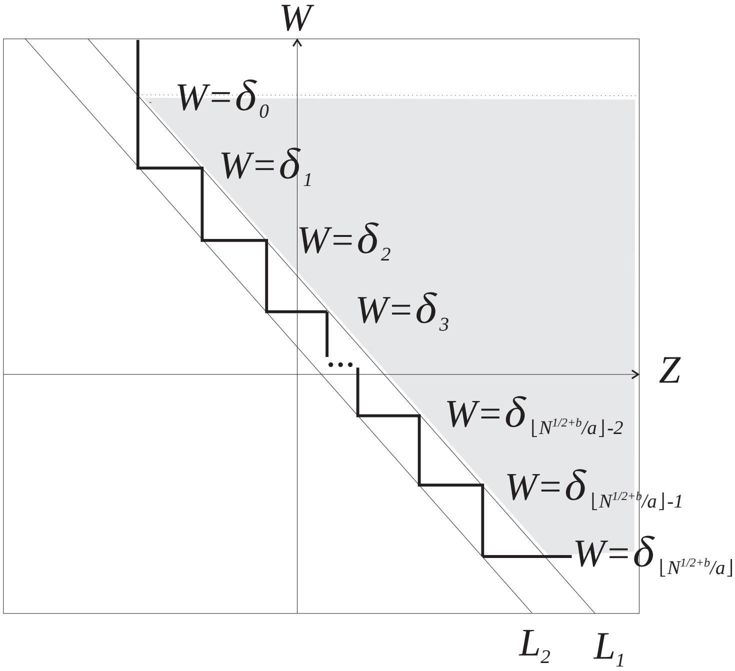

Figure 4.5: A suitable partition for asymptotic zero rate compression in $\mathcal{S}_2$ under Neyman-Pearson testing.

The equation of the straight line L_1 in Figure 4.5 is

$$\frac{\sigma_x(\mu\sigma_y^2 - \eta\sigma_{xy})}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} Z + \frac{\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} W \;=\; V\varepsilon_N,$$

where $\varepsilon_N \triangleq \Phi^{-1}\left(\alpha - 2a/N^{1/2+b}\right)$.

Hence, the probability of the region to the right of line L_1 under the null hypothesis is

$$P_{xy}^N \left\{ \frac{\log \frac{P_{xy}^N(\underline{x},\underline{y})}{Q_{xy}^N(\underline{x},\underline{y})} - ND}{V\sqrt{N}} \;>\; \varepsilon_N \right\} \;=\; 1 - \alpha + \frac{2a}{N^{1/2+b}}.$$

Define $\delta_{\lfloor N^{1/2+b}/a \rfloor} \triangleq \Phi^{-1}(a/N^{1/2+b})$, $\delta_0 \triangleq -\delta_N$, and

$$\delta_j \triangleq \delta_0 - j\frac{\delta_0 - \delta_{\lfloor N^{1/2+b}/a \rfloor}}{\lfloor N^{1/2+b}/a \rfloor}.$$

(Note that $\delta_{\lfloor N^{1/2+b}/a \rfloor} < 0$ if N is large enough.) Consider the piecewise-linear boundary shown in bold in Figure 4.5. The break points of this boundary lie on the the straight line L_2 which is parallel to L_1. The equation of L_2 is

$$\begin{aligned} &\frac{\sigma_x(\mu\sigma_y^2 - \eta\sigma_{xy})}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} Z + \frac{\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} W \\ = \quad & V\varepsilon_N - \frac{\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})(\delta_0 - \delta_{\lfloor N^{1/2+b}/a \rfloor})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)\lfloor N^{1/2+b}/a \rfloor} \\ = \quad & V\varepsilon_N + \frac{2\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})\Phi^{-1}(a/N^{1/2+b})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)\lfloor N^{1/2+b}/a \rfloor} . \end{aligned}$$

Now consider the region $\mathcal{U}$ to the right of the piecewise-linear boundary. Ignoring the type I error constraint for the moment, $\mathcal{U}$ is feasible as acceptance region for the null hypothesis if the message alphabet sizes $\|g_{x,N}\|$ and $\|g_{y,N}\|$ satisfy (4.5.11). We have

$$\begin{aligned} P_{xy}^N(\mathcal{U}) \quad \geq \quad & P_{xy}^N(\text{shaded area in Figure 4.5}) \\ = \quad & P_{xy}^N(\text{region to the right of line } L_1) - P_{xy}^N(W > \delta_0) \\ & -P_{xy}^N(W < \delta_{\lfloor N^{1/2+b}/a \rfloor}) \\ = \quad & 1 - \alpha + \frac{2a}{N^{1/2+b}} - \frac{a}{N^{1/2+b}} - \frac{a}{N^{1/2+b}} \\ = \quad & 1 - \alpha \end{aligned}$$

and

$$
\begin{aligned}
& Q_{xy}^N(\mathcal{U}) \\
\leq\ & Q_{xy}^N(\text{region to the right of line } L_2) \\
=\ & Q_{xy}^N \left\{ \frac{\log \frac{P_{xy}^N(\underline{x},\underline{y})}{Q_{xy}^N(\underline{x},\underline{y})} - ND}{V\sqrt{N}} \ \geq\ \varepsilon_N + \frac{2\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})\Phi^{-1}(a/N^{1/2+b})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)V\lfloor N^{1/2+b}/a \rfloor} \right\} \\
=\ & 1 - \Phi\left(\varepsilon_N + \frac{2\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})\Phi^{-1}(a/N^{1/2+b})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)V\lfloor N^{1/2+b}/a \rfloor} + \sqrt{2ND} \right) \\
\leq\ & \frac{1}{\varepsilon_N + \frac{2\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})\Phi^{-1}(a/N^{1/2+b})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)V\lfloor N^{1/2+b}/a \rfloor} + \sqrt{2ND}} \\
& \exp\left\{ -\frac{1}{2}\left(\varepsilon_N + \frac{2\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})\Phi^{-1}(a/N^{1/2+b})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)V\lfloor N^{1/2+b}/a \rfloor} \right)^2 \right\} \\
& \exp\left\{ -ND - \sqrt{2ND}\left(\varepsilon_N + \frac{2\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})\Phi^{-1}(a/N^{1/2+b})}{(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)V\lfloor N^{1/2+b}/a \rfloor} \right) \right\} \\
\leq\ & \frac{C_{NP}}{\sqrt{N}} \exp\left\{ -ND - \sqrt{N}\Phi^{-1}(\alpha)V \right\}
\end{aligned}
$$

for some $C_{NP} > 0$. The last step follows because

$$
\varepsilon_N = \Phi^{-1}(\alpha) + O(1/N^{1/2+b}), \qquad \text{and} \qquad \frac{\Phi^{-1}(a/N^{1/2+b})}{N^b} = o(1)\,.
$$

□

In what follows, we use a parallel argument to prove a a similar result for Bayes testing under a stronger assumption (than (4.5.11)), namely

$$
\liminf_{N\to\infty} \frac{\|g_{x,N}\| \wedge \|g_{y,N}\|}{N} > 0. \tag{4.5.12}
$$

Again, it is easy to verify that the Bayes error of the centralized system satisfies

$$
\gamma_2^*(\pi, N, +\infty) \ \asymp\ \frac{1}{\sqrt{N}} \exp\left\{-\rho(P_{xy}\|Q_{xy})\,N\right\},
$$

where $\rho(P_{xy}\|Q_{xy}) = D/4$ is the Chernoff exponent of the temporal marginal distributions P_{xy} and Q_{xy} (cf. Section 1.2). Hence, it suffices to show that

$$\gamma_2^*(\alpha, N, 0^+) \;\le\; \frac{C_B}{\sqrt{N}} \exp\left\{-\frac{D}{4}N\right\},$$

for some positive constant C_B, where π is the prior probability of the null hypothesis.

Lemma 4.3. *Suppose (4.5.12) holds. Then*

$$\gamma_2^*(\pi, N, 0^+) \;\le\; \frac{C_B}{\sqrt{N}} \exp\left\{-\frac{D}{4}N\right\}$$

for some $C_B > 0$.

***Proof*:** Again, we assume $\mu\sigma_y^2 - \eta\sigma_{xy} > 0$ and $\eta\sigma_x^2 - \mu\sigma_{xy} > 0$. The other cases are proved similarly.

By assumption, we can find $a > 0$ and N_0 such that $\|g_{x,N}\| \wedge \|g_{y,N}\| > N/a$ for $N > N_0$.

Define

$$Z = \frac{1}{\sigma_x\sqrt{N}}\sum_{i=1}^{n} X_i - \frac{\mu\sqrt{N}}{2\sigma_x} \qquad \text{and} \qquad W = \frac{1}{\sigma_y\sqrt{N}}\sum_{i=1}^{n} Y_i - \frac{\eta\sqrt{N}}{2\sigma_y}.$$

Then the statistics of (Z, W) are

$$H_0: \quad \mathcal{N}\left(\begin{pmatrix} -\dfrac{\mu\sqrt{N}}{2\sigma_x} \\ -\dfrac{\eta\sqrt{N}}{2\sigma_y} \end{pmatrix}, \begin{pmatrix} 1 & \dfrac{\sigma_{xy}}{\sigma_x\sigma_y} \\ \dfrac{\sigma_{xy}}{\sigma_x\sigma_y} & 1 \end{pmatrix}\right)$$

$$H_1: \quad \mathcal{N}\left(\begin{pmatrix} \dfrac{\mu\sqrt{N}}{2\sigma_x} \\ \dfrac{\eta\sqrt{N}}{2\sigma_y} \end{pmatrix}, \begin{pmatrix} 1 & \dfrac{\sigma_{xy}}{\sigma_x\sigma_y} \\ \dfrac{\sigma_{xy}}{\sigma_x\sigma_y} & 1 \end{pmatrix}\right)$$

We plot Z versus W in Figure 4.6. The equation of the straight line L_1 corresponding to $\log[P^N_{xy}(\underline{X},\underline{Y})/Q^N_{xy}(\underline{X},\underline{Y})] = 0$ is

$$\sigma_x(\mu\sigma_y^2 - \eta\sigma_{xy})Z + \sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})W = 0.$$

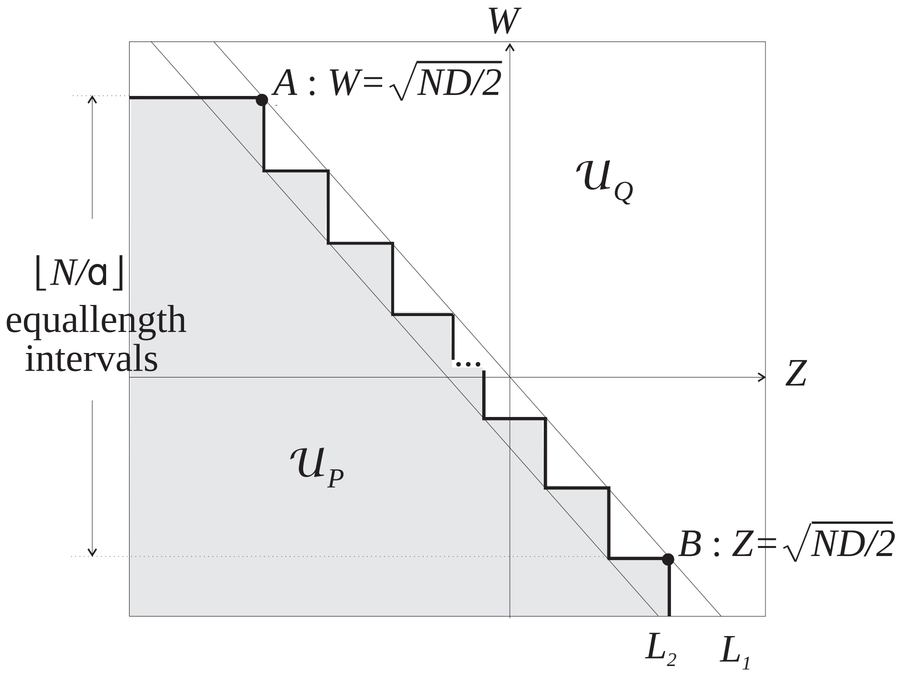

Figure 4.6: A suitable partition for asymptotic zero-rate compression in $\mathcal{S}_2$ under Bayes testing.

In Figure 4.6, we choose points A and B on L_1 such that the W-coordinate of point A is $\sqrt{ND/2}$ and the Z-coordinate of point B is $\sqrt{ND/2}$. We then divide the line segment AB into $\lfloor N/a \rfloor$ equal intervals, and construct the piecewise linear boundary shown in bold. As in the proof of the previous lemma, this boundary can be realized when the message alphabets satisfy (4.5.12).

The equation of line L_2 is

$$\begin{aligned}\sigma_x(\mu\sigma_y^2 - \eta\sigma_{xy})Z + \sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})W &= -\frac{\sqrt{ND/2}}{\lfloor N/a \rfloor}\left(1 + \frac{\sigma_x(\mu\sigma_y^2 - \eta\sigma_{xy})}{\sigma_y(\eta\sigma_x^2 - \mu\sigma_{xy})}\right) \\ &\triangleq -\frac{b_N}{\sqrt{N}}.\end{aligned}$$

Note that $b_N = O(1)$. We then have

$$\begin{aligned}
Q_{xy}^N(\mathcal{U}_P) &\leq Q_{xy}^N\left(\log\frac{P_{xy}^N(\underline{X},\underline{Y})}{Q_{xy}^N(\underline{X},\underline{Y})} \geq 0\right) \\
&= 1 - \Phi\left(\sqrt{\frac{ND}{2}}\right) \\
&\leq \frac{1}{\sqrt{ND/2}}\exp\left\{-\frac{D}{4}N\right\}
\end{aligned}$$

and

$$\begin{aligned}
P_{xy}^N(\mathcal{U}_Q) &\leq P_{xy}^N\left[\log\frac{P_{xy}^N(\underline{X},\underline{Y})}{Q_{xy}^N(\underline{X},\underline{Y})} \leq b_N\right] \\
&\quad + P_{xy}^N\left[W > \sqrt{ND/2}\right] + P_{xy}^N\left[Z > \sqrt{ND/2}\right] \\
&= \Phi\left(\frac{b_N - ND}{\sqrt{N}V}\right) + \Phi\left(-\sqrt{ND/2} - \frac{\eta}{2\sigma_y}\sqrt{N}\right) \\
&\quad + \Phi\left(-\sqrt{ND/2} - \frac{\mu}{2\sigma_x}\sqrt{N}\right) \\
&= \Phi\left(\frac{b_N - ND}{\sqrt{N}V}\right) + 2\Phi\left(-\sqrt{ND/2}\right) \\
&\leq \frac{1}{\sqrt{ND/2} - b_N/(V\sqrt{N})}\exp\left\{-\frac{D}{4}N - \frac{b_N^2}{4ND} + b_N\right\} \\
&\quad + \frac{1}{\sqrt{ND/8}}\exp\left\{-\frac{D}{4}N\right\} \\
&\leq \frac{C}{\sqrt{N}}\exp\left\{-\frac{D}{4}N\right\}
\end{aligned}$$

for some positive constant C. Hence

$$\gamma_2^*(\pi, N, 0^+) \leq \frac{C_B}{\sqrt{N}}\exp\left\{-\frac{D}{4}N\right\},$$

where $C_B \triangleq (\pi\sqrt{2}/\sqrt{D}) \vee ((1-\pi)C)$. $\square$

Chapter 5

Conclusions

5.1 Spatial asymptotics in parallel distributed detection

Our investigation of error exponents in the absence of boundedness conditions (such as Assumption 2.1 and (3.1.9)) has yielded interesting and illuminating results: notably Theorem 2.4 on the dependence of the Neyman-Pearson error exponent on the test level α; and Theorem 3.1, which affirms the previously conjectured redundancy of the boundedness assumption in Bayes testing.

The main conclusion of Theorem 3.1 is that the performance ratio between the absolutely optimal and best identical-quantizer systems is bounded. This result, in conjunction with Corollary 3.2, shows that the degree of asymptotic equivalence of these two systems is far greater than what is implied by the equality of error exponents $e^*_{\mathrm{B}}(\pi)$ and $e^{\diamond}_{\mathrm{B}}(\pi)$. A stronger version of Theorem 3.1 could include a lower bound on $\gamma^*_n(\pi)/\gamma^{\diamond}_n(\pi)$; although this was not attempted in this dissertation (the proof of Theorem 3.1 is already too technical), this refinement would be quite welcome.

We have ascertained that Theorem 3.1 also holds for L-ary hypothesis testing, where $L > 2$. In that case, it is known from [30, Theorem 1] that there exists an asymptotically exponentially optimal system $\mathcal{S}_n^\diamond$ employing—in fixed proportions as n varies—at most $L(L-1)/2$ fixed quantizers. We have obtained a variant of Lemma 3.1 based on a multidimensional Berry-Esseen theorem [4, Corollary 17.2], and proceeded to modify the proof of Theorem 3.1 so as to consider the $L(L-1)/2$ binary likelihood-ratio comparisons involved in L-ary testing (details omitted).

In designing a large parallel distributed system for Neyman-Pearson detection based on i.i.d. observations, the use of identical quantizers is marginally suboptimal. While previous results had shown that the best identical-quantizer system $\mathcal{S}_n^\diamond$ achieves the same error exponent as the absolutely optimal system $\mathcal{S}_n^*$, a much stronger result (Theorem 2.1) was obtained in this dissertation, namely that the ratio between the actual type II error probabilities is in many cases bounded.

It is noteworthy that Theorem 2.1 differentiates between the cases $\alpha \leq 1/2$ and $\alpha > 1/2$, requiring an additional regularity condition (Assumption 2.2) for the latter case. If Assumption 2.2 is violated, then it is possible to have $\beta_n^*(\alpha)/\beta_n^\diamond(\alpha) \to 0$ even when the third moment of the post-quantization log-likelihood is bounded, i.e., $\delta = 1$ in (2.1.1). Boundedness of third moments is not an unreasonable requirement, considering that boundedness of second moments ($\delta = 0$ in (2.1.1)) is needed in order to show equality of error exponents in $\mathcal{S}_n^\diamond$ and $\mathcal{S}_n^*$.

Finally, it should be noted that the lower bound of Theorem 2.1 also holds for the limiting case $\delta = 0$, but that the proof requires a more elaborate central limit theorem argument.

5.2 Distributed detection of a signal in additive Gaussian noise

We considered the problem of optimal distributed detection of a known signal in additive Gaussian noise; specifically, some of the difficulties and curiosities due to correlation in the noise.

The tractable nature of the Gaussian model enabled us to derive a sufficient condition under which contiguous marginal LRQ's are optimal. Whether this condition is necessary is still unknown to us. We did, however, show by counterexample that contiguous marginal LRQ's are not always optimal.

We also demonstrated that for large samples corrupted by additive Gaussian noise, marginal LRQ's may not be optimal at all. Still, in terms of asymptotic performance, we showed that it is possible for a distributed system to achieve an error probability similar to (at most a fixed multiple of) that achieved by a centralized system, provided the number of local quantization levels increases with sample size at a suitable polynomial rate.

Appendix A

Numerical results on the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$

Finding $\gamma_n^*(\pi)$ is a hard computational task: as discussed in [32], the complexity of the problem is exponential in n. If the observation space is continuous, then the complexity of the problem explodes as discretization becomes finer. In what follows, numerical solutions are obtained for the simplest case of ternary local observations and binary quantization per sensor. The results are given in Table A.1 and A.2, where "$\diamond$" indicates that the identical quantizer system is optimal, while "$*$" indicates that the identical quantizer system is suboptimal.

Equal priors for the two hypotheses are assumed. The pre-quantization distribution pairs in the two examples considered are given by:

Example A.1.

P	:	0.4	0.3	0.3
Q	:	0.3	0.4	0.4

and

Example A.2.

P	:	0.8	0.15	0.05
Q	:	1/3	1/3	1/3

Further simulations on Example A.2 show that even for very large values of n (in the hundreds), it is possible for the optimal system to contain nonidentical quantizers. In addition, the ratio $\gamma_n^*(1/2)/\gamma_n^\diamond(1/2)$ does not seem to converge. This feature is also observed in the example of Appendix C.

Number of local sensors	$\gamma_n^*(1/2)$	$\gamma_n^\diamond(1/2)$	$\frac{\gamma_n^*(1/2)}{\gamma_n^\diamond(1/2)}$	optimal system
2	0.435000	0.435000	1	$\diamond$
3	0.423000	0.432000	0.9791667	$*$
4	0.411750	0.411750	1	$\diamond$
5	0.403380	0.404370	0.9975517	$*$
6	0.394470	0.400005	0.9861626	$*$
7	0.386417	0.386417	1	$\diamond$
8	0.379368	0.381810	0.993605	$*$
9	0.372790	0.376476	0.990209	$*$
10	0.366336	0.366336	1	$\diamond$
11	0.360618	0.363362	0.992448	$*$
12	0.354808	0.357262	0.993131	$*$
13	0.349351	0.349351	1	$\diamond$
14	0.344263	0.347517	0.990636	$*$
15	0.339222	0.340816	0.995323	$*$
16	0.334390	0.334390	1	$\diamond$
17	0.329889	0.333272	0.989849	$*$
18	0.325241	0.325918	0.997923	$*$
19	0.320632	0.321830	0.996278	$*$
20	0.316633	0.320893	0.986725	$*$
21	0.311862	0.313803	0.993815	$*$
22	0.304137	0.305419	0.995802	$*$
23	0.293484	0.293484	1	$\diamond$

Table A.1: Numerical results on the ratio $\gamma_n^*(1/2)/\gamma_n^\diamond(1/2)$ in Example A.1

Number of local sensors	$\gamma_n^*(1/2)$	$\gamma_n^\diamond(1/2)$	$\frac{\gamma_n^*(1/2)}{\gamma_n^\diamond(1/2)}$	optimal system
2	0.230550	0.234450	0.983365	$*$
3	0.179413	0.179413	1	$\diamond$
4	0.144485	0.144485	1	$\diamond$
5	0.131442	0.131442	1	$\diamond$
6	0.097879	0.097879	1	$\diamond$
7	0.090847	0.095704	0.9492497	$*$
8	0.070427	0.070427	1	$\diamond$
9	0.063026	0.063026	1	$\diamond$
10	0.052997	0.052997	1	$\diamond$
11	0.043497	0.043497	1	$\diamond$
12	0.041287	0.041287	1	$\diamond$
13	0.031335	0.031335	1	$\diamond$
14	0.029030	0.030057	0.965832	$*$
15	0.023460	0.023460	1	$\diamond$
16	0.020772	0.020772	1	$\diamond$
17	0.018160	0.018160	1	$\diamond$
18	0.014796	0.014796	1	$\diamond$
19	0.014050	0.014461	0.9715787	$*$
20	0.010907	0.010907	1	$\diamond$

Table A.2: Numerical results on the ratio $\gamma_n^*(1/2)/\gamma_n^\diamond(1/2)$ in Example A.2

Appendix B

Proof of Lemmas 3.5 and 3.6

Proof of Lemma 3.5 Assumption 1.2 of Section 1.2 implies that the pre-quantization log-likelihood ratio X is not almost surely constant. Thus there exist $x < 0$, $x' > 0$ and $r > 0$ such that

$$Q\{X \leq x\} \ \geq \ r \qquad \text{and} \qquad Q\{X \geq x'\} \ \geq \ r \ . \tag{B.1}$$

Now let $m \geq 2$, $\delta > 0$, and consider an arbitrary LRP $\tau = (I_1, \ldots, I_m)$ in $\mathcal{T}_m(\delta)$. The values of P_τ and Q_τ here are all greater than zero, so the equation $X_\tau(u) = \log[P_\tau(u)/Q_\tau(u)]$ is meaningful. We always have

$$X_\tau(1) \ \leq \ 0 \ , \qquad X_\tau(m) \ \geq \ 0 \ . \tag{B.2}$$

In what follows we will find a nontrivial lower bound to $X_\tau(m) - X_\tau(1)$ in terms of x, x', and r. For that purpose, we decompose I_1 into $A = I_1 \cap (-\infty, x]$ and $B = I_1 \setminus (-\infty, x]$, and denote the (P, Q)-probabilities of these intervals by (p_A, q_A) and (p_B, q_B), respectively. Also, we fix a point t outside the interior of $I_1 \cup I_m$.

Clearly $p_A \leq q_A \exp x$ and $p_B \leq q_B \exp t$, and thus

$$\begin{aligned} \exp\{X_\tau(1)\} &= \frac{p_A + p_B}{q_A + q_B} \\ &\leq \frac{q_A \exp x + q_B \exp t}{q_A + q_B} \\ &= \frac{q_A}{q_A + q_B} \exp x + \frac{q_B}{q_A + q_B} \exp t \ . \end{aligned}$$

If $x \notin I_1$ (so that $B = \emptyset$), then the r.h.s. bound equals e^x, and by virtue of (B.2), we have

$$X_\tau(m) - X_\tau(1) \ \geq \ -X_\tau(1) \ \geq \ -x \ > \ 0 \ . \tag{B.3}$$

Otherwise, if $x \in I_1$, the bound is a proper mixture of $\exp x$ and $\exp t$ (greater than $\exp x$), whose value increases in q_B/q_A. Noting that $q_B/q_A < (1-r)/r$ by (B.1), we obtain

$$X_\tau(1) \ \leq \ \log[re^x + (1-r)e^t] \ . \tag{B.4}$$

An identical argument for I_m yields, for $x' \notin I_m$,

$$X_\tau(m) - X_\tau(1) \ \geq \ X_\tau(m) \ \geq \ x' \ > \ 0 \ ; \tag{B.5}$$

and for $x' \in I_m$,

$$X_\tau(m) \ \geq \ \log[re^{x'} + (1-r)e^t] \ . \tag{B.6}$$

It remains to find a lower bound to $X_\tau(m) - X_\tau(1)$ when both $x \in I_1$ and $x' \in I_n$. Using (B.4) and (B.6), we obtain for this case

$$X_\tau(m) - X_\tau(1) \ \geq \ \log \frac{re^{x'} + (1-r)e^t}{re^x + (1-r)e^t} \ . \tag{B.7}$$

Since $t \in [x, x']$ and the r.h.s. is decreasing in t, we can set $t = x'$ to obtain

$$X_\tau(m) - X_\tau(1) \ \geq \ -\log[re^{x-x'} + (1-r)] \ > \ 0 \ . \tag{B.8}$$

From (B.3), (B.5) and (B.8) we conclude that $|X_\tau(m) - X_\tau(1)| \geq b(x, x', r) > 0$, and that the bound is independent of δ (we do, however, need $\delta > 0$). The remainder is straightforward. We write for simplicity $p_u = P_\tau(u)$ and $q_u = Q_\tau(u)$ and note that the tilted distribution with parameter $\theta \in (0,1)$ satisfies

$$\tilde{Q}(X_\tau(u)) \;=\; \frac{p_u^\theta q_u^{1-\theta}}{\sum_{v=1}^m p_v^\theta q_v^{1-\theta}} \;,$$

where the denominator of the rightmost term is just $\Psi_\tau(\theta)$. Since for $\theta \in (0,1)$, $\Psi_\tau(\theta) \leq 1$ and $p_u^\theta q_u^{1-\theta} \geq \delta$, we conclude that both $\tilde{Q}(X_\tau(1)) \geq \delta$ and $\tilde{Q}(X_\tau(m)) \geq \delta$. Therefore $\mathrm{Var}_{\tilde{Q}}[X_\tau] \geq \delta b^2(x, x', r)/2 > 0$. □

Proof of Lemma 3.6 The quantity

$$r(\delta) \stackrel{\triangle}{=} \sup_{\tau \in \mathcal{R}_l(\delta)} \rho(P_\tau, Q_\tau)$$

decreases together with δ, and thus the limit $r \stackrel{\triangle}{=} \lim_{\delta \downarrow 0} r(\delta)$ is well-defined.

First we observe that $r(\delta) \geq \rho_{l-1}$ for every $\delta > 0$. Indeed, any LRP τ_{l-1} in $\mathcal{T}_{l-1}$ that achieves ρ_{l-1} can be refined to a LRP τ_l in $\mathcal{R}_l(\delta)$ by splitting the rightmost interval of τ_{l-1} at a point t such that $P\{X \geq t\} < \delta$. It follows easily from Lemma 3.2 that $\rho(P_{\tau_l}, Q_{\tau_l}) \geq \rho(P_{\tau_{l-1}}, Q_{\tau_{l-1}})$, and thus also $r(\delta) \geq \rho_{l-1}$.

It remains to show that $\rho_{l-1} \geq r$. For every $\delta > 0$, we choose a $\tau_l = \tau_l(\delta)$ in $\mathcal{R}_l(\delta)$ such that

$$\rho(P_{\tau_l}, Q_{\tau_l}) \;=\; -\log \Psi_{\tau_l}(\theta_l) \;\geq\; r(\delta) - \varepsilon \;, \tag{B.9}$$

where $\varepsilon > 0$. Without loss of generality, we can assume that $\theta_l = \theta_l(\delta)$ lies in a fixed (as δ varies) closed interval $[a, b]$, where $0 < a < b < 1$. This is because by Lemma 3.2 and the convexity of $\Psi(\theta)$, we can find $a = a(\varepsilon) > 0$ and $b = b(\varepsilon) < 1$

such that for all $\theta \in [0, a(\varepsilon)] \cup [b(\varepsilon), 1]$,

$$1 \;\geq\; \Psi_{\tau_l}(\theta) \;\geq\; \Psi(\theta) \;\geq\; \exp\{-\varepsilon\} \;.$$

Therefore, if θ_l lies in $[0, a] \cup [b, 1]$, it must satisfy

$$\exp \varepsilon \;\geq\; \frac{\Psi_{\tau_l}(\theta_l)}{\Psi_{\tau_l}(a)} \;\geq\; \exp\{-\varepsilon\}$$

and thus also by (B.9),

$$-\log \Psi_{\tau_l}(\theta_l) \;\geq\; r(\delta) - 2\varepsilon \;. \tag{B.10}$$

We can thus replace θ_l by a if $\theta_l \notin [a, b]$.

In the usual notation, let $P_\tau = (p_1, \ldots, p_l)$ and $Q_\tau = (q_1, \ldots, q_l)$ be the output probability distributions generated by $\tau_l = (I_1, \ldots, I_l)$. Then there exists at least one index u for which $p_u \wedge q_u < \delta$. Since P and Q are mutually absolutely continuous, $P(A) \wedge Q(A) < \delta$ implies $P(A) \vee Q(A) < \xi(\delta)$, where $\xi(\delta) \to 0$ as $\delta \to 0$. Thus we also have $p_u \vee q_u < \xi(\delta)$.

Taking w.l.o.g. the said index as $u = 1$, we can merge I_1 with I_2 to obtain a LRP $\tau_{l-1} \in \mathcal{T}_{l-1}$. This new $(l-1)$-ary LRP is equivalent to the l-ary LRP

$$\tau'_l \;=\; (I_1 \cup I_2,\, \emptyset,\, I_3,\, \cdots,\, I_l) \;,$$

and thus $\Psi_{\tau_{l-1}}(\theta) = \Psi_{\tau'_l}(\theta)$ for all θ. Also, observe that the pairs (P_{τ_l}, Q_{τ_l}) and $(P_{\tau'_l}, Q_{\tau'_l})$ differ by no more than $\xi(\delta)$ in any given coordinate.

Now when $\tau \in \mathcal{T}_l$, the function $\Psi_\tau(\theta)$ is obtained from

$$\Theta(\theta; \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq \sum_{u=1}^{l} \lambda_u^\theta \mu_u^{1-\theta}$$

by setting $\boldsymbol{\lambda} = P_\tau$ and $\boldsymbol{\mu} = Q_\tau$. It is straightforward to show that $\Theta(\theta; \boldsymbol{\lambda}, \boldsymbol{\mu})$ is continuous on $[a, b] \times [0, 1]^{2l}$ provided $0 < a < b < 1$. Recalling that $\Psi_\tau(\theta)$ is bounded from below by $\exp\{-\rho(P, Q)\} > 0$, we conclude that

$$|\log \Psi_{\tau_{l-1}}(\theta) - \log \Psi_{\tau_l}(\theta)|$$

can be made smaller than ε uniformly in $\theta \in [a, b]$ by taking δ (and thus also $\xi(\delta)$) sufficiently small.

By virtue of (B.10), we now have $\rho_{l-1} > r(\delta) - 3\varepsilon \geq r - 3\varepsilon$, and since ε was arbitrary, we conclude that $\rho_{l-1} \geq r$. □

Appendix C

A counterexample to $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi) \to$ 1

Consider a ternary observation space $\mathcal{Y} = \{a_1, a_2, a_3\}$ with binary quantization. The two hypotheses are assumed equally likely, with

y	a_1	a_2	a_3
$P(y)$	$1/12$	$1/4$	$2/3$
$Q(y)$	$1/3$	$1/3$	$1/3$
$(dP/dQ)(y)$	$1/4$	$3/4$	2

There are only two nontrivial deterministic LRQ's: $\hat{g}$, which partitions $\mathcal{Y}$ into $\{a_1\}$ and $\{a_2, a_3\}$; and $\bar{g}$, which partitions $\mathcal{Y}$ into $\{a_1, a_2\}$ and $\{a_3\}$. The corresponding output distributions and log-likelihood ratios are given by

$$\begin{array}{ccc} u & 1 & 2 \\ P_{\hat{\tau}}(u) & 1/12 & 11/12 \\ Q_{\hat{\tau}}(u) & 1/3 & 2/3 \\ X_{\hat{\tau}}(u) & -\log 4 & \log(11/8) \end{array} \quad \text{and} \quad \begin{array}{ccc} u & 1 & 2 \\ P_{\bar{\tau}}(u) & 1/3 & 2/3 \\ Q_{\bar{\tau}}(u) & 2/3 & 1/3 \\ X_{\bar{\tau}}(u) & -\log 2 & \log 2 \end{array} \tag{C.1}$$

Direct computation shows that

$$\rho(P_{\bar{\tau}}, Q_{\bar{\tau}}) \;=\; \frac{1}{2}\log\frac{9}{8} \;=\; 0.0589 \;>\; 0.0534 \;=\; \rho(P_{\hat{\tau}}, Q_{\hat{\tau}}) \;,$$

and thus for n sufficiently large, the best identical-quantizer system $\mathcal{S}_n^\diamond$ employs $\bar{g}$ on all n sensors (the other choice yields a suboptimal error exponent and thus eventually a higher value of $\gamma_n(\pi)$).

We will now show by contradiction that if $\mathcal{S}_n^*$ is an absolutely optimal system consisting of deterministic LRQ's, then for all even values of n, at least one of the quantizers in $\mathcal{S}_n^*$ must be $\hat{g}$.

Assume the contrary, i.e., $\mathcal{S}_n^*$ is such that for all $i \le n$, $g_i = \bar{g}$. Now consider the problem of optimizing the quantizer g_n in $\mathcal{S}_n$ subject to the constraint that each of the remaining quantizers $g_1, \ldots, g_{n-1}$ equals $\bar{g}$. From the discussion in Section 1.2.D, we know that either $g_n = \hat{g}$ or $g_n = \bar{g}$ is optimal. Our assumption about $\mathcal{S}_n^*$ then implies that $g_n = \bar{g}$ is, in fact, optimal.

To see why this cannot be so if n is even, consider the Bayes error probability for this problem. Writing u_1^n for $(u_1, \ldots, u_n)$, we have

$$\begin{aligned}
&\gamma_n\left(\frac{1}{2}\right)\\
&= \frac{1}{2}\sum_{u_1^n \in \{1,2\}^n} [P_{\bar{\tau}}(u_1^{n-1})P_{g_n}(u_n)] \wedge [Q_{\bar{\tau}}(u_1^{n-1})Q_{g_n}(u_n)]\\
&= \frac{1}{2}\sum_{u_1^{n-1} \in \{1,2\}^{n-1}} \sum_{u_n \in \{1,2\}} [P_{\bar{\tau}}(u_1^{n-1})P_{g_n}(u_n)] \wedge [Q_{\bar{\tau}}(u_1^{n-1})Q_{g_n}(u_n)]\\
&= \sum_{u_1^{n-1} \in \{1,2\}^{n-1}} \left[\frac{P_{\bar{\tau}}(u_1^{n-1}) + Q_{\bar{\tau}}(u_1^{n-1})}{2}\right] \gamma\left(\frac{P_{\bar{\tau}}(u_1^{n-1})}{P_{\bar{\tau}}(u_1^{n-1}) + Q_{\bar{\tau}}(u_1^{n-1})}\right) \qquad \text{(C.2)}
\end{aligned}$$

where $\gamma(\cdot)$ represents the Bayes error probability function of the nth sensor/quantizer pair (note that in this equation, the argument of $\gamma(\cdot)$ is just the posterior probability of H_0 given u_1^{n-1}).

We note from (C.1) that the log-likelihood ratio

$$\log \frac{P_{\bar{\tau}}(u_1^{n-1})}{Q_{\bar{\tau}}(u_1^{n-1})} = X_{\bar{\tau}}(u_1) + \cdots + X_{\bar{\tau}}(u_{n-1})$$

can be also expressed as $(2l_{n-1}(u) - n + 1)(\log 2)$, where $l_{n-1}(u)$ is the number of 2's in u_1^{n-1}. Now $l_{n-1}(U)$ is a binomial variable under either hypothesis, and we can rewrite (C.2) as

$$\gamma_n\left(\frac{1}{2}\right) = \sum_{l=0}^{n-1} \binom{n-1}{l} \left[\frac{2^l + 2^{n-l-1}}{2 \cdot 3^{n-1}}\right] \gamma\left(\frac{2^{2l-n+1}}{2^{2l-n+1}+1}\right) . \tag{C.3}$$

The two candidates for γ are $\hat{\gamma}$ and $\bar{\gamma}$, given by

$$\hat{\gamma}(\pi) = \left[\frac{1}{12}\pi \wedge \frac{1}{3}(1-\pi)\right] + \left[\frac{11}{12}\pi \wedge \frac{2}{3}(1-\pi)\right]$$

$$\bar{\gamma}(\pi) = \left[\frac{1}{3}\pi \wedge \frac{2}{3}(1-\pi)\right] + \left[\frac{2}{3}\pi \wedge \frac{1}{3}(1-\pi)\right]$$

and shown in Figure C.1. Note that $\hat{\gamma}(\pi) = \bar{\gamma}(\pi)$ for $\pi \le 1/3$, $\pi = 4/7$ and $\pi \ge 4/5$. Thus the critical values of l in (C.3) are those for which $2^{2l-n+1}/(2^{2l-n+1}+1)$ lies in the union of (1/3, 4/7) and (4/7, 4/5).

If n is odd, then the range of $2l - n + 1$ in (C.3) comprises the even integers between $-n+1$ and $n-1$ inclusive. The only critical value of l is $(n-1)/2$, for which the posterior probability of H_0 is 1/2. Since $\hat{\gamma}(1/2) = 3/8 > 1/3 = \bar{\gamma}(1/2)$, the optimal choice is $\bar{g}$.

If n is even, then $2l - n + 1$ ranges over all odd integers between $-n+1$ and $n-1$ inclusive. Here the only critical value of l is $n/2$, which makes the posterior probability of H_0 equal to 2/3. Since $\hat{\gamma}(2/3) = 5/18 < 1/3 = \bar{\gamma}(2/3)$, $\hat{g}$ is optimal.

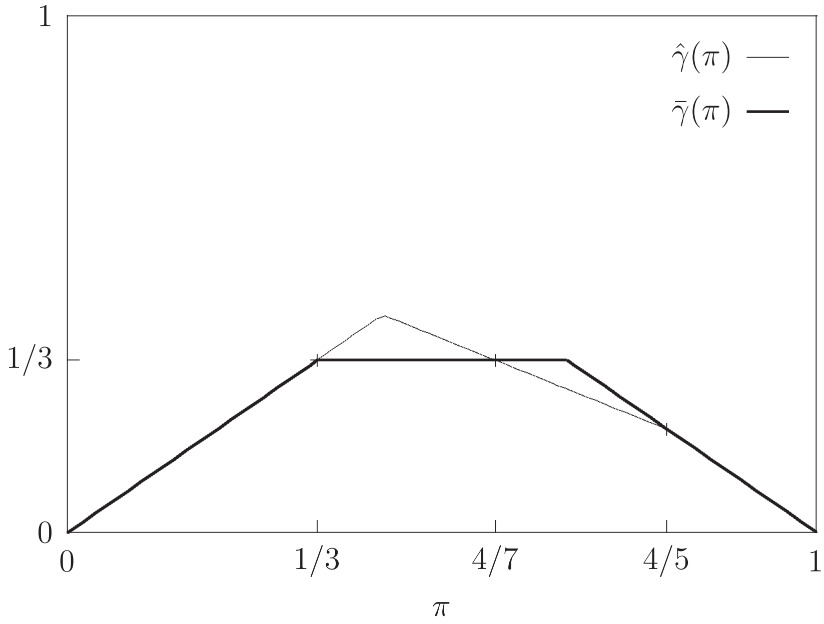

Figure C.1: Bayes error probabilities associated with $\hat{g}$ and $\bar{g}$.

We thus obtain the required contradiction, together with the inequality

$$\begin{aligned}\gamma^{\diamond}_{2k}\left(\frac{1}{2}\right) - \gamma^{*}_{2k}\left(\frac{1}{2}\right) \quad &\geq \quad \binom{2k-1}{k}\left[\frac{2^k + 2^{k-1}}{2 \cdot 3^{2k-1}}\right]\left[\bar{\gamma}\left(\frac{2}{3}\right) - \hat{\gamma}\left(\frac{2}{3}\right)\right] \\ &= \quad \frac{1}{8}\binom{2k-1}{k}\left(\frac{2}{9}\right)^k .\end{aligned}$$

Stirling's formula gives $(4\pi k)^{-1/2}\exp\{-k\log 4\}(1+o(1))$ for the binomial coefficient, and thus

$$\liminf_{k\to\infty}\left(\gamma^{\diamond}_{2k}\left(\frac{1}{2}\right) - \gamma^{*}_{2k}\left(\frac{1}{2}\right)\right)\sqrt{\pi k}\left(\frac{9}{8}\right)^k \;\geq\; \frac{1}{16} . \tag{C.4}$$

Since $(9/8)^k = \exp\{2k\rho(P_{\bar{\tau}}, Q_{\bar{\tau}})\} = \exp\{2k\rho_2\}$, we immediately deduce from (C.4) and (3.1.12) that

$$\limsup_{k\to\infty}\frac{\gamma^{*}_{2k}(1/2)}{\gamma^{\diamond}_{2k}(1/2)} \;<\; 1 .$$

A finer approximation to

$$\gamma^{\diamond}_{2k}\left(\frac{1}{2}\right) \;=\; Q\{X_{\bar{\tau}}(U_1)+\cdots+X_{\bar{\tau}}(U_{2k}) > 0\} + \frac{1}{2}Q\{X_{\bar{\tau}}(U_1)+\cdots+X_{\bar{\tau}}(U_{2k}) = 0\}$$

using [18, Theorem 1] and Stirling's formula for the first and second summands, respectively, yields

$$\lim_{k\to\infty} \gamma^{\diamond}_{2k}\left(\frac{1}{2}\right)\sqrt{\pi k}\left(\frac{9}{8}\right)^{k} \;=\; \frac{3}{2}\;.$$

From (C.4) we then obtain

$$\limsup_{k\to\infty} \frac{\gamma^{*}_{2k}(1/2)}{\gamma^{\diamond}_{2k}(1/2)} \;\leq\; \frac{23}{24}\;.$$

Appendix D

Supplements to Chapter 4

D.1 Proof of necessity of (4.2.1) for the monotonity of $f_{\mathcal{B}}(x)$

Lemma D.1. *If* $\sigma_{xy}(\eta\sigma_x^2 - \mu\sigma_{xy})(\mu\sigma_x^2 - \eta\sigma_{xy}) \;<\; 0$, *then* $\exists\ \mathcal{B} \in \mathcal{Y}$ *such that* $\log(P_{xy}(x, \mathcal{B}))/(Q_{xy}(x, \mathcal{B}))$ *is not a monotone function of* x.

***Proof*:** Let $\mathcal{B}(b) \triangleq (-b - 1/b, -b + 1/b) \cup (b - 1/b, b + 1/b)$ for some $b > 1$, and let

$$R(y) \triangleq \frac{\exp\left\{-\frac{\sigma_y^2}{2\Delta}y^2\right\}}{\displaystyle\int_{\mathcal{B}} \exp\left\{-\frac{\sigma_y^2}{2\Delta}y^2\right\} dy}.$$

Then from the proof of Lemma 4.1,

$$\widetilde{P}(y) \;=\; R^{(s)}(y), \qquad \text{and} \qquad \tilde{Q}(y) \;=\; R^{(s+t)}(y),$$

where $s = \sigma_{xy}x/\Delta$, and $t = (\eta\sigma_x^2 - \mu\sigma_{xy})/\Delta$. Also

$$f'_{\mathcal{B}}(x) \;=\; \frac{\sigma_{xy}}{\Delta}\left(E_{R^{(s)}}[Y] - E_{R^{(s+t)}}[Y]\right) - \frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\Delta}.$$

Since $\lim_{s\to\infty} E_{R^{(s)}}[Y] = b + 1/b$ and $\lim_{s\to-\infty} E_{R^{(s)}}[Y] = -b - 1/b$,

$$\lim_{|x|\to\infty} f'_{\mathcal{B}}(x) = -\frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\Delta}.$$

Observe that $E_{R^{(0)}}[Y] = 0$. Hence,

$$f'_{\mathcal{B}}(0) = -\frac{\sigma_{xy}}{\Delta} E_{R^{(t)}}[Y] - \frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\Delta}.$$

Since

$$\exp\left\{-\frac{2\sigma_y^2}{\Delta} + 2tb\right\} \le \frac{R^{(t)}[(b-1/b, b+1/b)]}{R^{(t)}[(-b-1/b, -b+1/b)]} \le \exp\left\{\frac{2\sigma_y^2}{\Delta} + 2tb\right\},$$

we obtain

$$E_{R^{(t)}}[Y] \ge \frac{e^{2tb} - e^{2\sigma_y^2/\Delta}}{e^{2tb} + e^{2\sigma_y^2/\Delta}} b - \frac{1}{b} \ge \frac{e^{2tb} - e^{2\sigma_y^2/\Delta}}{e^{2tb} + e^{2\sigma_y^2/\Delta}} b - 1$$

$$E_{R^{(t)}}[Y] \le \frac{e^{2tb} - e^{-2\sigma_y^2/\Delta}}{e^{2tb} + e^{-2\sigma_y^2/\Delta}} b + \frac{1}{b} \le \frac{e^{2tb} - e^{-2\sigma_y^2/\Delta}}{e^{2tb} + e^{-2\sigma_y^2/\Delta}} b + 1.$$

So by choosing

$$b > \frac{\Delta + 2\sigma_y^2}{2\Delta|\eta\sigma_x^2 - \mu\sigma_{xy}|} \vee \frac{e+1}{e-1}\left(1 + \left|\frac{\mu\sigma_y^2 - \eta\sigma_{xy}}{\sigma_{xy}}\right|\right),$$

we have $f'_{\mathcal{B}}(0) f'(+\infty) < 0$ and $f'_{\mathcal{B}}(0) f'(-\infty) < 0$. □

D.2 Example of suboptimality of contiguous marginal LRP's

AND fusion is considered only. An example for OR/XOR fusion rule can be created in a similar way.

Example Equal priors and complete compression are assumed. Let $\sigma_x^2 = \sigma_y^2 = 1$, $\mu = \eta = 4$ and $\sigma_{xy} = -1 + \varepsilon/2$ for $\varepsilon \in (0, 2)$.

Claim D.1. *For the above example, there exists ε_0 in $(0, 2)$ such that for all $0 < \varepsilon < \varepsilon_0$, contiguous marginal LRP's are suboptimal in Bayes testing.*

***Proof*:** Let γ_2^* represent the optimal Bayes error.

1.

$$\gamma_2^* \le Pr\{\text{use only one local observation}\} = \Phi(-2) = 0.0227. \quad \text{(D.1)}$$

2. Suppose $(c, \infty) \times (d, \infty)$ is the optimal acceptance region for the alternative hypothesis, obtained from a contiguous marginal LRP. From Theorem 4.1, $c = d$. In addition, $c \ge -4$; for if $c < -4$, then

$$\gamma_2^* \ge \frac{1}{2} P\left[(-4, \infty) \times (-4, \infty)\right] \quad \text{(D.2)}$$

$$\ge \frac{1}{2} P\left[|X + Y| < \sqrt{\varepsilon} \text{ and } |X - Y| < \sqrt{4 - \varepsilon}\,\right] \quad \text{(D.3)}$$

(See Figure D.1)

$$= \frac{1}{2}\left[1 - 2\Phi(-1)\right]^2$$

$$= 0.2329713 > 0.0227,$$

which contradicts (D.1). Note that r.h.s.(D.2)$\ge$r.h.s.(D.3) holds for all $-1 \le \sigma_{xy} \le 1$. Hence $c \ge -4$ is a universal bound for the best contiguous likelihood ratio partition.

3. [*Main argument*] Define

$$\mathcal{L}_1 \triangleq \{|x + y| \le \varepsilon^{1/4}\}, \quad \text{and} \quad \mathcal{L}_2 \triangleq \{|x + y - 8| \le \varepsilon^{1/4}\}.$$

Then

$$\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset \quad \text{and} \quad P[\mathcal{L}_1^c] = Q[\mathcal{L}_2^c] = 2\Phi\left(-\varepsilon^{-1/4}\right).$$

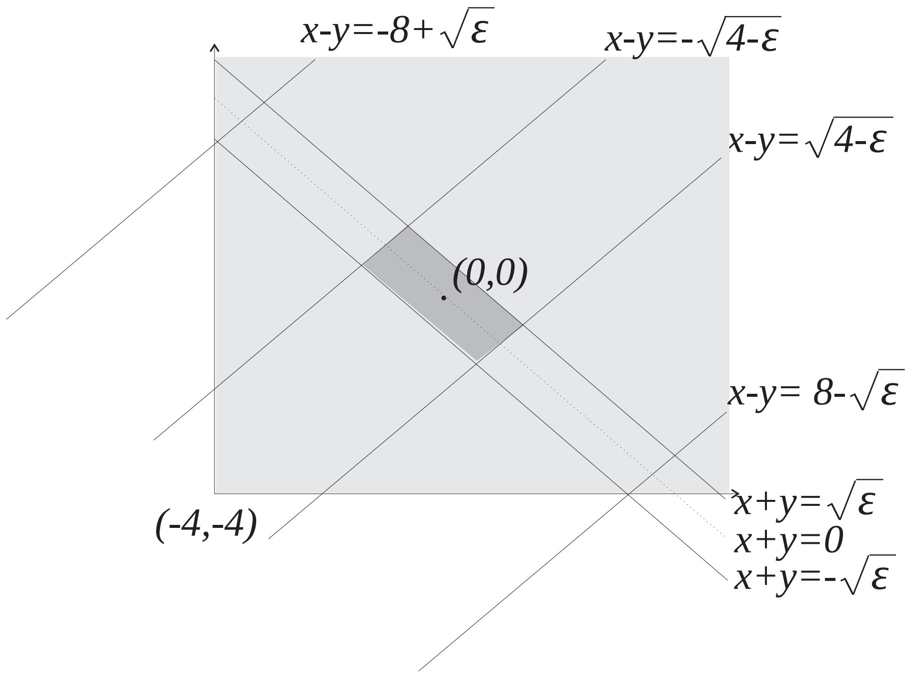

Figure D.1: Lower-bounding the Bayes error probability.

Therefore

$$0 \;\leq\; \gamma_2^* - \gamma_2^*(\mathcal{L}_1 \cup \mathcal{L}_2) \;\leq\; 2\Phi\left(-\varepsilon^{-1/4}\right),$$

where $\gamma_2^*(\mathcal{L}_1 \cup \mathcal{L}_2)$ is defined as the contribution to the Bayes error of the region $\mathcal{L}_1 \cup \mathcal{L}_2$ only (See Figure D.2). Actually, for any partition,

$$0 \;\leq\; \gamma_2 - \gamma_2(\mathcal{L}_1 \cup \mathcal{L}_2) \;\leq\; 2\Phi\left(-\varepsilon^{-1/4}\right)$$

is true.

For the new non-contiguous partition shown in Figure D.3, the new Bayes error should satisfy

$$\gamma_2^*(\mathcal{L}_1 \cup \mathcal{L}_2) - \gamma_2(\mathcal{L}_1 \cup \mathcal{L}_2) \;=\; \frac{1}{2}Q(\mathcal{B}),$$

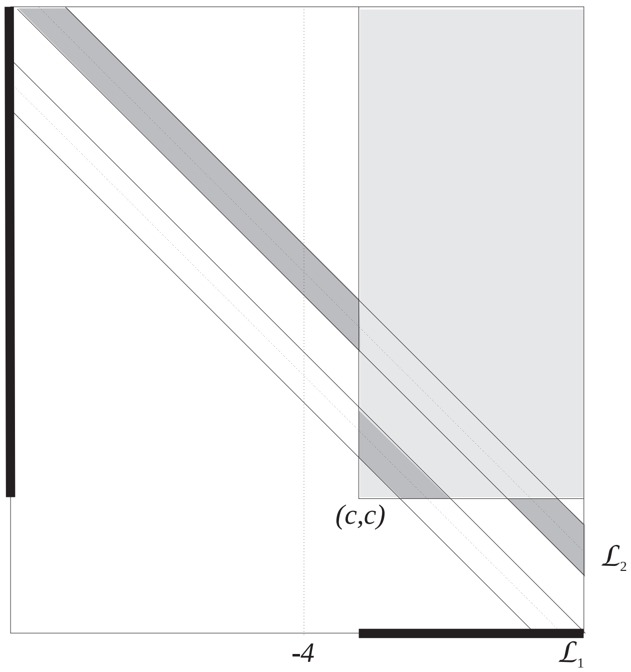

Figure D.2: The Bayes error restricted to $\mathcal{L}_1 \cup \mathcal{L}_2$.

where

$$
\begin{aligned}
Q(\mathcal{B}) \;\; &\geq \;\; Q\big(|X+Y-8| \;\leq\; \varepsilon^{1/4} \\
&\qquad\qquad \text{and} \quad -32-3\varepsilon^{1/4} \;\geq\; X-Y \;>\; -48+\varepsilon^{1/4}\big)
\end{aligned}
$$

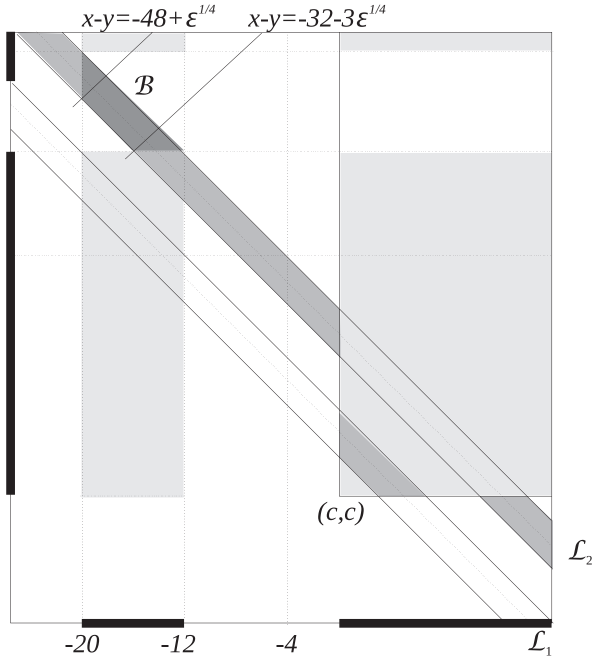

Figure D.3: A non-contiguous improvement.

Hence

$$\begin{aligned}
\gamma_2^* - \gamma_2 &\geq \gamma_2^*(\mathcal{L}_1 \cup \mathcal{L}_2) - \left(\gamma_2(\mathcal{L}_1 \cup \mathcal{L}_2) + 2\Phi\left(-\varepsilon^{-1/4}\right)\right) \\
&= \frac{1}{2}Q(\mathcal{B}) - 2\Phi\left(-\varepsilon^{-1/4}\right) \\
&\geq \frac{1}{2}\left[1 - 2\Phi\left(-\varepsilon^{-1/4}\right)\right] \\
&\quad \times \left[\Phi\left(\frac{-32 - 3\varepsilon^{1/4}}{\sqrt{4-\varepsilon}}\right) - \Phi\left(\frac{-48 + \varepsilon^{1/4}}{\sqrt{4-\varepsilon}}\right)\right] - 2\Phi\left(-\varepsilon^{-1/4}\right) \\
&> \frac{1}{4}\left[\Phi(-16) - \Phi(-24)\right] \;>\; 0, \quad \text{for } 0 < \varepsilon < \varepsilon_0, \text{ for some } \varepsilon_0.
\end{aligned}$$

Consequently, a better non-contiguous partition can always be found if $0 < \varepsilon < \varepsilon_0$. □

Bibliography

[1] A. R. Reibman and L. W. Nolte. Optimal detection and performance of distributed detection systemes. *IEEE Trans. on Aerospace and Electronic Systems*, AES–23(1):24–30, Jan. 1987.

[2] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Royal Stat. Soc., series B*, 28:131–142, 1966.

[3] R. Bahadur and R. R. Rao. On deviations of the sample mean. *Ann. Math. Statist.*, 31:1015–1027, 1960.

[4] R. N. Bhattacharya and R. R. Rao. *Normal Approximation and Asymptotic Expansions*. Wiley, New York, 1976.

[5] P. Billingsley. *Probability and Measure*. Wiley, New York, 2nd edition, 1986.

[6] D. Blackwell and J. L. Hodges, Jr. The probability in the extreme tail of a convolution. *Ann. Math. Stat.*, 30:1113–1120, 1959.

[7] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation*. Wiley, New York, 1991.

[8] J. A. Bucklew and J. S. Sadowsky. A contribution to the theory of Chernoff bounds. *IEEE Trans. Info. Theory*, IT–39(1):249–254, 1993.

[9] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels—i. *Inform. and Control*, 10:65–103, 1967.

[10] P.-N. Chen and A. Papamarcou. Asymptotic results on the symmetry of distributed detection. In *Proc. 26th Conference on Information Sciences and Systems*, pages 180–184, Princeton, New Jersey, Mar. 1992.

[11] P.-N. Chen and A. Papamarcou. New asymptotic results in parallel distributed detection. *IEEE Trans. Info. Theory*, 39(6):1847–1863, Nov. 1993.

[12] M. Cherikh and P. B. Cantor. Counterexamples in distributed detection. *IEEE Trans. Info. Theory*, IT–38(1):162–165, 1992.

[13] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Math. Stat.*, 23:493–507, 1952.

[14] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[15] I. Csiszár. Information-type measure of difference of probability distributions and indirect observations. *Stud. Scientarium Mathematicerium Hungarica*, 2:229–318, 1967.

[16] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.

[17] J. A. Fill. Asymptotic expansions for large deviation probabilities in the strong law of large numbers. *Probab. Th. Rel. Fields*, 81:213–233, 1989.

[18] J. A. Fill and M. J. Wichura. The convergence rate for the strong law of large numbers: General lattice distributions. *Probab. Th. Rel. Fields*, 81:189–212, 1989.

[19] T. J. Flynn and R. M. Gray. Encoding of correlated observations. *IEEE Trans. Info. Theory*, IT–33(6):773–787, Nov. 1987.

[20] R. G. Gallager. *Information Theory and Reliable Communication.* Wiley, New York, 1968.

[21] P. Hall. *Rates of Convergence in the Central Limit Theorem.* Pitman, Boston, 1982.

[22] Hossam M. H. Shalaby, Adrian Papamarcou. Multiterminal detection with zero-rate data compression. *IEEE Trans. on Information Theory*, 38(2):254–267, Mar. 1992.

[23] Jenn-Sen Leu. Topics in noise parameter estimation and decentralized signal detection. *Dissertation*, 1993.

[24] H. V. Poor. Fine quantization in signal detection and estimation. *IEEE Trans. Info. Theory*, IT–34(5):960–972, Sep. 1988.

[25] Rudolf Ahlswede and I. Csiszar. Hypothesis testing with communication constraints. *IEEE Trans. on Information Theory*, IT–32(4):533–542, Jul. 1986.

[26] Te Sun Han. Hypothesis testing with multiterminal data compression. *IEEE Trans. on Information Theory*, IT–33(6):759–772, Nov. 1987.

[27] Te Sun Han, and Kingo Kobayashi. Exponential-type error probabilities for multiterminal hypothesis testing. *IEEE Trans. on Information Theory*, 35(1):2–14, Jan. 1989.

[28] R. R. Tenney and N. R. Sandell, Jr. Detection with distributed sensors. *IEEE Trans. Aerospace Electron. Syst.*, AES-17(4):501–510, July 1981.

[29] J. N. Tsitsiklis. On threshold rules in decentralized detection. In *Proc. 25th Conference on Decision and Control*, pages 232–236, Athens, Greece, 1986.

[30] J. N. Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals and Systems*, 1(2):167–182, 1988.

[31] J. N. Tsitsiklis. Extremal properties of likelihood-ratio quantizers. In *Proc. 29th Conference on Decision and Control*, pages 2680–2685, Honolulu, Hawaii, 1990.

[32] J. N. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. *IEEE Trans. Automat. Contr.*, AC–30(5):440–446, 1985.

[33] D. Warren and P. Willett. Optimal decentralized detection for conditionally independent sensors. In *Proc. American Control Conference*, pages 1326–1329, 1989.

[34] P. Willett and D. Warren. Decentralized detection: When are identical sensors identical? In *Proc. Conference on Information Sciences and Systems*, pages 287–292, Baltimore, Maryland, Mar. 1991.

[35] P. Willett and D. Warren. The suboptimality of randomized tests in distributed and quantized detection systems. *IEEE Trans. Info. Theory*, IT–38(2):355–361, 1992.

[36] William W. Irving and John N. Tsitsiklis. Some properties of optimal thresholds in decentralized detection. *IEEE Trans. on Automatic Control*, 39(4):835–838, Apr. 1994.

[37] Z. Chair and P. K. Varshney. Optimal data fusion in multiple sensor detection system. *IEEE Trans. on Aerospace and Electronic Systems*, AES–22(1):98–101, Jan. 1988.