Multiterminal Detection with Zero-Rate Data Compression

Hossam M. H. Shalaby, *Member, IEEE,* and Adrian Papamarcou, *Member, IEEE*

Abstract—**The asymptotic performance of a multiterminal detection system comprising a central detector and two remote sensors that have access to discrete, spatially dependent, and temporally memoryless observations is discussed. It is assumed that prior to transmitting information to the central detector, each sensor compresses its observations at a rate which approaches zero as the sample size tends to infinity; and that on the basis of the compressed data from both sensors, the central detector seeks to determine whether the true distribution of the observations belongs to a null class Π or an alternative class Ξ. Under the criterion that stipulates minimization of the type II error rate subject to an upper bound ϵ on the type I error rate, we show that in the case of simple distribution classes ($|\Pi| = |\Xi| = 1$), the error exponent achievable by such a system has a simple characterization, is independent of the value of ϵ, and is insensitive to changes in compression rate as long as the asymptotic rate on one of the sensors is zero. By considering four different settings, it is then demonstrated that these conclusions do not hold in the case of composite distribution classes.**

Key Words: **hypothesis testing, distributed detection, data compression, quantization, error exponents, blowing-up lemma.**

I. Introduction

We consider the problem of testing a null hypothesis H_0 against an alternative H_1 on the basis of compressed data from a discrete-time, discrete-alphabet, memoryless multiple source. In its simplest form, our setup comprises two remote sensors S_X and S_Y that are linked to a *central detector*. The sensors S_X and S_Y observe the respective components of the random sequence $\{(X_i, Y_i)\}_{i=1}^n$, and encode their observations into a maximum of M_n and N_n messages, respectively. Upon receipt of the two codewords, the central detector accepts or rejects the null hypothesis in conformity with the classical criterion that stipulates minimization of the probability of falsely accepting H_0 (*type II error*) subject to a fixed upper bound ϵ on the probability of falsely rejecting H_0 (*type I error*).

Distributed detection systems of the above type have been widely studied in the recent literature. The models most frequently encountered [9]–[17] employ fixed codebook sizes $M_n = M$ and $N_n = N$, where M and N are often equal to 2. In such cases, the central detector receives from each sensor what amounts to a local decision, possibly accompanied by an assessment (on a fixed finite-valued scale) of the sensor's confidence in that decision. Of course, it is also possible to design distributed detection systems employing varying codebook sizes M_n and N_n, as is the case with certain models discussed in the information-theoretic literature [1]–[6] and in this paper.

It is worth noting that for one particular model, namely that in which M_n and N_n are large enough so that no compression is needed, the analysis is well known. In that case, the central detector knows the observed sequence $\{(X_i, Y_i)\}_{i=1}^n$ precisely, and the optimal decision rule for testing $H_0 : P_{XY}$ versus $H_1 : Q_{XY}$ at any level ϵ is specified by the Neyman–Pearson lemma. Furthermore, the resulting minimum type II error probability $\beta_n(\epsilon)$ satisfies the asymptotic identity

$$-\lim_n \frac{1}{n}\log \beta_n(\epsilon) = D(P_{XY}\|Q_{XY}).$$

The quantity appearing on the left-hand side of this equation (which is due to Stein [7]) is termed the *error exponent* for the hypothesis testing problem. On the right-hand side, $D(\cdot\|\cdot)$ denotes informational divergence.

Unfortunately, in cases where data compression is mandatory, the determination of the optimal system is a highly complex task that involves the joint optimization of the local data encoders and the central detector. This difficulty has motivated the study of tractable compression/decision schemes which are *asymptotically* optimal, i.e., achieve the same error exponents as their optimal counterparts. The investigations in [2]–[5] are examples of such studies.

In [2], Ahlswede and Csiszár discussed the problem of hypothesis testing under fixed-rate compression on one sensor. In the special case of testing against independence (i.e., $Q_{XY} = Q_X \times Q_Y$), they obtained a single-letter characterization of the error exponent by recourse to entropy characterization techniques. Also, in the general case where $Q_{XY} > 0$, they showed that the error exponent is independent of the level ϵ. Yet the problem of single-letter characterization of the error exponent in the case $Q_{XY} \neq Q_X \times Q_Y$ remained unsolved; single-letter lower bounds to that exponent were obtained in both [2] and [3] using compression/decision schemes whose asymptotic optimality was not established. In a somewhat different model involving exponentially decaying

Manuscript received October 4, 1990; revised August 11, 1991. This work was supported by the Systems Research Center (a National Science Foundation Engineering Research Center) at the University of Maryland. This work was presented in part at the 1990 IEEE International Symposium on Information Theory, San Diego, CA, Jan. 14–19, 1990.

H. M. H. Shalaby was with the Electrical Engineering Department and the Systems Research Center at the University of Maryland, College Park, MD 20742. He is now with the Department of Electrical Engineering, Faculty of Engineering, University of Alexandria, Alexandria 21544, Egypt.

A. Papamarcou is with the Electrical Engineering Department and the Systems Research Center at the University of Maryland, College Park, MD 20742.

0018-9448/92$03.00

bounds on the type I error rate, Han and Kobayashi [4] developed good upper bounds on the error exponent for fixed-rate compression on both sensors.

In this paper, we consider the hypothesis testing problem under data compression at (asymptotically) *zero rate*. In other words, we assume that the codebook sizes satisfy constraints of the type

$$R_X(n) = \frac{1}{n}\log M_n \to 0, \qquad R_Y(n) = \frac{1}{n}\log N_n \to 0.$$

Our inquiry was motivated by the study in [3] of hypothesis testing under two-sided one-bit ($M_n = 2,\ N_n = 2$), and one-sided one-bit ($M_n = 2,\ N_n = \infty$), compression. For those systems, Han proposed a simple scheme that compressed both S_X and S_Y to one bit, was independent of the level ϵ, and yielded a simply characterized lower bound on the error exponent. He then proved by converse theorems the tightness of the lower bound

1) for all values of ϵ in the case of two-sided one-bit compression;
2) for a range $(0, \epsilon_0)$ of values of ϵ, where $\epsilon_0 < 1$, in the case of one-sided one-bit compression.

Related work has appeared in the pioneering paper of Amari and Han [5], where differential-geometric arguments were used to establish the error exponent in certain cases of two-sided zero-rate compression under the additional constraint of symmetric encoding. We should also note that under the assumption of exponential decay in the level ϵ, Han and Kobayashi [4] derived the error exponent under two-sided one-bit compression and provided lower bounds for that exponent under one-sided one-bit, and two-sided zero-rate, compression.

We complement and extend these results as follows. First, for fixed-level simple hypothesis testing under the positivity constraint $Q_{XY} > 0$, we prove that the two-sided one-bit compression/decision scheme proposed by Han in [3] is, for *all* $\epsilon \in (0, 1)$, asymptotically optimal in the broader class of one-sided zero-rate compression/decision schemes. Thus, an optimal distributed detection system employing two sensors, of which one transmits data at a vanishing rate while the other supplies complete information about its observations, is asymptotically no better than optimal system in which each sensor transmits a single binary digit. It also follows as a special case that optimal systems for fixed codebook-size compression ($M_n = M,\ N_n = N$) have the same asymptotic performance regardless of the values M and N. In other words, no gain in asymptotic performance can result by allowing each sensor to transmit a quantized, or *soft*, decision [16] instead of a binary, or *hard*, decision.

Our second body of contributions is in the area of composite hypothesis testing under zero-rate compression. By considering four different problems in this area, we demonstrate that the error exponent here is not only a function of the null and alternative distribution classes, but also depends on the level ϵ and the sequences of codebook sizes M_n and N_n. This conclusion is in sharp antithesis to our findings in the area of simple hypothesis testing.

The formulation of the general problem is given in Section II, together with pertinent notation. The converse theorem for simple hypothesis testing appears in Section III, followed in Section IV by an extension to the multivariate case (r sensors, where $r > 2$). Section V serves as introduction to the problems in composite hypothesis testing discussed in Sections VI and VII.

II. Problem Statement and Preliminaries

A. General Notation

The observations of S_X and S_Y are denoted by the sequences $X^n = (X_1, \cdots, X_n) \in \mathscr{X}^n$ and $Y^n = (Y_1, \cdots, Y_n) \in \mathscr{Y}^n$, respectively, and the alphabets $\mathscr{X}$ and $\mathscr{Y}$ are assumed finite. Since the multiple source is memoryless, the sequence of pairs $((X_1, Y_1), \cdots, (X_n, Y_n)) \in (\mathscr{X} \times \mathscr{Y})^n$ is i.i.d. under both hypotheses. In what follows, it will be convenient to deal with the product space $\mathscr{X}^n \times \mathscr{Y}^n$ instead of $(\mathscr{X} \times \mathscr{Y})^n$, and thus, the observations will be collectively represented by the pair $(X^n, Y^n) \in \mathscr{X}^n \times \mathscr{Y}^n$.

By virtue of the aforementioned i.i.d. assumption, all distributions of interest can be specified through bivariate distributions on $\mathscr{X} \times \mathscr{Y}$. Under the null hypothesis, the distribution of any pair (X_i, Y_i) is usually denoted by P_{XY}, and its respective marginals by P_X and P_Y. The distributions of X^n, Y^n, and (X^n, Y^n) under the same hypothesis are denoted by P_X^n, P_Y^n and P_{XY}^n, respectively. The i.i.d. assumption then implies that for all (x^n, y^n) in $\mathscr{X}^n \times \mathscr{Y}^n$,

$$P_{XY}^n(x^n, y^n) = \prod_{i=1}^{n} P_{XY}(x_i, y_i).$$

Analogous notation is employed for the alternative hypothesis, with Q replacing P. We will also have occasion to use distributions $\bar{P}_{XY}$, $\tilde{P}_{XY}$ and $\hat{P}_{XY}$ on $\mathscr{X} \times \mathscr{Y}$, which will yield marginals and higher order distributions in the same manner as P_{XY} and Q_{XY}.

The spaces of all distributions on $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{X} \times \mathscr{Y}$ will be denoted by $\mathscr{P}(\mathscr{X})$, $\mathscr{P}(\mathscr{Y})$ and $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$, respectively.

The compression of X^n and Y^n is effected by encoders f_n, and g_n, respectively, where

$$f_n : \mathscr{X}^n \mapsto \{1, \cdots, M_n\}, \quad \text{and} \quad g_n : \mathscr{Y}^n \mapsto \{1, \cdots, N_n\}.$$

For one-sided zero-rate compression of X^n we assume that $N_n \geq |\mathscr{Y}|^n$ and

$$M_n \geq 2, \qquad \lim_n \frac{1}{n}\log M_n = 0, \tag{2.1}$$

and similarly for one-sided zero-rate compression of Y^n, we have $M_n \geq |\mathscr{X}|^n$ and

$$N_n \geq 2, \qquad \lim_n \frac{1}{n}\log N_n = 0. \tag{2.2}$$

For two-sided zero-rate compression, both (2.1) and (2.2) are assumed.

The central detector is represented by the function

$$\phi_n : \{1, \cdots, M_n\} \times \{1, \cdots, N_n\} \mapsto \{0, 1\},$$

where the output 0 signifies the acceptance of the null hypothesis H_0, and 1 its rejection. This induces a partition of the original (i.e., noncompressed) sample space $\mathscr{X}^n \times \mathscr{Y}^n$ into an *acceptance* region

$$\mathscr{A}_n \stackrel{\text{def}}{=} \{(x^n, y^n) \in \mathscr{X}^n \times \mathscr{Y}^n : \phi_n(f_n(x^n), g_n(y^n)) = 0\},$$

and a *critical* (or rejection) region $\mathscr{A}_n^c$.

By nature of the encoding process, the acceptance region can be decomposed into M_n rectangles $C_i \times F_i$ in $\mathscr{X}^n \times \mathscr{Y}^n$ that possess disjoint projections C_i on $\mathscr{X}^n$. More precisely, if for every $1 \le i \le M_n$ we define

$$C_i = \{x^n \in \mathscr{X}^n : f_n(x^n) = i\}$$

$$\text{and} \quad F_i = \{y^n \in \mathscr{Y}^n : \phi_n(i, g(y^n)) = 0\},$$

then we can write

$$\mathscr{A}_n = \bigcup_{i=1}^{M_n} C_i \times F_i, \tag{2.3}$$

where $(\forall i \neq j)$ $C_i \cap C_j = \emptyset$. We can obtain an alternative representation for $\mathscr{A}_n$ by partitioning $\mathscr{Y}^n$ into N_n sets:

$$\mathscr{A}_n = \bigcup_{i=1}^{N_n} D_i \times G_i, \tag{2.4}$$

where $(\forall i \neq j)$ $D_i \cap D_j = \emptyset$. Note that (2.3) and (2.4) *jointly* characterize all admissible acceptance regions under *two-sided* compression with codebook sizes M_n (for X^n) and N_n (for Y^n). Taken *separately*, the above conditions characterize the admissible acceptance regions under *one-sided* compression of X^n and Y^n, respectively.

B. Simple Hypothesis Testing

The optimal acceptance region for testing $H_0 : P$ versus $H_1 : Q$ at a given level $\epsilon \in (0, 1)$ is one that minimizes $Q^n_{XY}(\mathscr{A}_n)$ over all acceptance regions $\mathscr{A}_n$ that

C1) yield a value of $P^n_{XY}(\mathscr{A}_n^c)$ less than or equal to ϵ; and

C2) satisfy the appropriate compression constraints; namely,

- (2.1) and (2.3) for one-sided compression of X^n;
- (2.2) and (2.4) for one-sided compression of Y^n;
- (2.1), (2.2), (2.3), and (2.4) for two-sided compression.

The resulting *minimum* probability of type II error is denoted by $\beta_n(M_n, N_n, \epsilon)$, and the associated error exponent is given by

$$\theta(M, N, \epsilon) \stackrel{\text{def}}{=} -\lim_n \frac{1}{n} \log \beta_n(M_n, N_n, \epsilon),$$

provided the limit on the right-hand side exists.

C. Composite Hypothesis Testing

Let Π and Ξ be disjoint subsets of $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$. For testing $H_0 : P_{XY} \in \Pi$ versus $H_1 : Q_{XY} \in \Xi$ at a given level ϵ, we employ the *uniformly most powerful* (UMP) test. Thus for a given level $\epsilon \in (0, 1)$, we seek to minimize the quantity

$$\sup_{Q \in \Xi} Q^n_{XY}(\mathscr{A}_n)$$

over all acceptance regions $\mathscr{A}_n$ that meet the constraints

C1') $P^n_{XY}(\mathscr{A}_n^c) \le \epsilon$ for all P_{XY} in Π; and

C2) as before.

We use the notation

$$\beta_n(M_n, N_n, \epsilon) \stackrel{\text{def}}{=} \min_{\mathscr{A}_n} \sup_{Q \in \Xi} Q^n_{XY}(\mathscr{A}_n),$$

and define the error exponent $\theta(M, N, \epsilon)$ as before.

D. Typical Sequences

Our proofs rely on the concept of a typical sequence, as developed in [8]. We cite some basic definitions and facts on typical sequences.

The type of a sequence $x^n \in \mathscr{X}^n$ is the distribution λ_x on $\mathscr{X}$ defined by the relationship

$$(\forall a \in \mathscr{X}) \qquad \lambda_x(a) \stackrel{\text{def}}{=} \frac{1}{n} N(a \mid x^n),$$

where $N(a \mid x^n)$ is the number of terms in x^n equal to a. The set of all types of sequences in $\mathscr{X}^n$, namely $\{\lambda_x : x^n \in \mathscr{X}^n\}$, will be denoted by $\mathscr{P}_n(\mathscr{X})$.

Given a type $\hat{P}_X \in \mathscr{P}_n(\mathscr{X})$, we will denote by $\hat{T}^n_X$ the set of sequences $x^n \in \mathscr{X}^n$ of type $\hat{P}_X$:

$$\hat{T}^n_X \stackrel{\text{def}}{=} \{x^n \in \mathscr{X}^n : \lambda_x = \hat{P}_X\}.$$

Also, for an arbitrary distribution $\tilde{P}_X$ on $\mathscr{X}$ and a constant $\eta > 0$, we will denote by $\tilde{T}^n_{X,\eta}$ the set of $(\tilde{P}_X, \eta)$-typical sequences in $\mathscr{X}^n$. A sequence x^n is $(\tilde{P}_X, \eta)$-typical if $|\lambda_x(a) - \tilde{P}_X(a)| \le \eta$ for every letter $a \in \mathscr{X}$ and, in addition, $\lambda_x(a) = 0$ for every a such that $\tilde{P}_X(a) = 0$. Thus, if $\|\cdot\|$ denotes the sup norm and $\ll$ denotes absolute continuity, we have

$$\tilde{T}^n_{X,\eta} \stackrel{\text{def}}{=} \{x^n \in \mathscr{X}^n : \|\lambda_x - \tilde{P}_X\| \le \eta, \lambda_x \ll \tilde{P}_X\}.$$

In the same manner, we will denote by $T^n_{X,\eta}$ and $\bar{T}^n_{X,\eta}$ the sets of (P_X, η)- and $(\bar{P}_X, \eta)$- (respectively) typical sequences in $\mathscr{X}^n$. We will have no need to consider sequences with exact or approximate type Q_X.

The proofs of the following lemmas appear in [8]. As usual, $|\mathscr{A}|$ denotes the size of $\mathscr{A}$.

Lemma 1: The size of $\mathscr{P}_n(\mathscr{X})$ is at most $(n+1)^{|\mathscr{X}|}$. For any $\hat{P}_X$ in $\mathscr{P}_n(\mathscr{X})$ and Q_X in $\mathscr{P}(\mathscr{X})$,

$$(n+1)^{-|\mathscr{X}|} \exp[nH(\hat{P}_X)] \le |\hat{T}^n_X| \le \exp[nH(\hat{P}_X)],$$

and

$$(n+1)^{-|\mathscr{X}|} \exp[-nD(\hat{P}_X \| Q_X)] \le Q^n_X(\hat{T}^n_X) \le \exp[-nD(\hat{P}_X \| Q_X)].$$

Lemma 2: For any distribution P_X on $\mathscr{X}$ and $\eta > 0$,

$$P_X^n(T_{X,\eta}^n) \geq 1 - \frac{|\mathscr{X}|}{4n\eta^2}.$$

One can easily modify this exposition to accommodate pairs $(x^n, y^n) \in \mathscr{X}^n \times \mathscr{Y}^n$ by reverting to their representation in $(\mathscr{X} \times \mathscr{Y})^n$. Thus the type of (x^n, y^n) is the distribution λ_{xy} on $\mathscr{X} \times \mathscr{Y}$ such that

$$\lambda_{xy}(a, b) = \frac{1}{n}|\{i : (x_i, y_i) = (a, b)\}|,$$

and the class $\mathscr{P}_n(\mathscr{X} \times \mathscr{Y})$, as well as the sets $\hat{T}_{XY}^n \subset \mathscr{X}^n \times \mathscr{Y}^n$ and $\tilde{T}_{XY,\eta}^n \subset \mathscr{X}^n \times \mathscr{Y}^n$, are defined accordingly.

In this and the following sections, we will omit the superscript n from T^n, as n will be essentially constant.

III. A Converse Theorem for Simple Hypothesis Testing

In this section, we derive the error exponent for simple hypothesis testing under the positivity condition $Q_{XY} > 0$. We show that the error exponent $\theta(M, N, \epsilon)$ is independent of ϵ and the compression scheme used (one-sided or two-sided), provided the asymptotic zero-rate constraints (2.1) and/or (2.2) are met. Furthermore, its value is given by the minimum of the quantity

$$D(\tilde{P}_{XY} \| Q_{XY})$$

over all bivariate distributions $\tilde{P}_{XY}$ on $\mathscr{X} \times \mathscr{Y}$ whose marginals on $\mathscr{X}$ and $\mathscr{Y}$ agree with those of P_{XY}.

The positive result, namely the existence of a sequence of acceptance regions that achieve the above value, was shown in [3]. The acceptance regions used in that work had the simple rectangular form

$$T_{X,\eta} \times T_{Y,\eta},$$

and were thus admissible under the most stringent of two-sided compression schemes, namely $M_n = N_n = 2$. Our result here is a strong converse for one-sided compression of X^n, i.e., we show that for every value of $\epsilon \in (0, 1)$ and every sequence of acceptance regions $\mathscr{A}_n$ satisfying (C1), (2.1) and (2.3), the following is true:

$$-\liminf_n \frac{1}{n} \log Q_{XY}^n(\mathscr{A}_n) \leq \min_{\substack{\tilde{P}_{XY}: \\ \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y}} D(\tilde{P}_{XY} \| Q_{XY}).$$

By symmetry, the same is true for one-sided compression of Y^n, and *a fortiori*, for two-sided compression.

Theorem 1: Let P_{XY} be arbitrary, and $Q_{XY} > 0$. For all $\epsilon \in (0, 1)$ and sequences M_n and $\mathscr{A}_n$ satisfying (2.1) and (2.3), the following is true: if for every n,

$$P_{XY}^n(\mathscr{A}_n^c) \leq \epsilon,$$

then

$$-\liminf_n \frac{1}{n} \log Q_{XY}^n(\mathscr{A}_n) \leq \min_{\substack{\tilde{P}_{XY}: \\ \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y}} D(\tilde{P}_{XY} \| Q_{XY}).$$

Proof: By (2.3), we have

$$\mathscr{A}_n = \bigcup_{i=1}^{M_n} C_i \times F_i,$$

where the C_i's are pairwise disjoint. Assume that $P_{XY}^n(\mathscr{A}^c) \leq \epsilon$, or equivalently, $P_{XY}^n(\mathscr{A}) \geq 1 - \epsilon$. Then there exists an index i_0 such that

$$P_{XY}^n(C_{i_0} \times F_{i_0}) \geq \frac{1-\epsilon}{M_n}.$$

Letting $C = C_{i_0}$ and $F = F_{i_0}$, we can rewrite it as

$$P_{XY}^n(C \times F) \geq \exp(-n\delta_n), \tag{3.1}$$

where

$$\delta_n = \delta_n(M_n, \epsilon) = -\frac{1}{n}\log(1-\epsilon) + \frac{1}{n}\log M_n,$$

and $\delta_n \to 0$ by (2.1). Equation (3.1) clearly implies that

$$P_X^n(C) \geq \exp(-n\delta_n) \quad \text{and} \quad P_Y^n(F) \geq \exp(-n\delta_n). \tag{3.2}$$

Thus, asymptotically, neither C nor F has "exponentially small" probability. By the blowing-up lemma [8, Theorem 5.4], this fact implies that both sets possess Hamming k_n-neighborhoods which are asymptotically "as thin" as the sets themselves (i.e., $k_n/n \to 0$), and whose probabilities approach unity as n tends to infinity. Specifically, let $d(\cdot, \cdot)$ denote Hamming distance, and define the Hamming k-neighborhood $\Gamma^k C$ of C by

$$\Gamma^k C \stackrel{\text{def}}{=} \{u^n \in \mathscr{X}^n : (\exists x^n \in C)\; d(x^n, u^n) \leq k\}.$$

The blowing-up lemma asserts that under condition (3.2), there exist sequences k_n and γ_n satisfying

$$k_n/n \to 0 \quad \text{and} \quad \gamma_n \to 0,$$

and such that

$$P_X^n(\Gamma^{k_n} C) \geq 1 - \gamma_n \quad \text{and} \quad P_Y^n(\Gamma^{k_n} F) \geq 1 - \gamma_n. \tag{3.3}$$

Furthermore, k_n and γ_n depend only on $|\mathscr{X}|$, $|\mathscr{Y}|$ and δ_n, and *not* on P_{XY}.

In what follows, we will use k instead of k_n in all superscripts.

Equation (3.3) clearly holds true if we replace P by $\tilde{P}$, where $\tilde{P}_{XY}$ satisfies the marginal constraints

$$\tilde{P}_X = P_X \quad \text{and} \quad \tilde{P}_Y = P_Y.$$

Using the elementary property $\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1$, we then obtain

$$\tilde{P}_{XY}^n(\Gamma^k C \times \Gamma^k F) \geq \tilde{P}_X^n(\Gamma^k C) + \tilde{P}_Y^n(\Gamma^k F) - 1,$$

and hence,

$$\tilde{P}_{XY}^n(\Gamma^k C \times \Gamma^k F) \geq 1 - 2\gamma_n. \tag{3.4}$$

Thus, under the n-fold product of $\tilde{P}_{XY}$, the probability of the rectangle $\Gamma^k C \times \Gamma^k F$ approaches unity as n tends to infinity. By Lemma 2, the same is true of the set of $(\tilde{P}_{XY}, \eta)$-typical elements in $\mathscr{X}^n \times \mathscr{Y}^n$, where $\eta = \eta_n = n^{-1/3}$. Indeed,

$$\tilde{P}^n_{XY}(\tilde{T}_{XY,\eta}) \geq 1 - \frac{|\mathscr{X}||\mathscr{Y}|}{4n\eta_n^2} = 1 - \frac{|\mathscr{X}||\mathscr{Y}|}{4n^{1/3}}.$$

Hence, for all sufficiently large n, we obtain

$$\tilde{P}^n_{XY}((\Gamma^k C \times \Gamma^k F) \cap \tilde{T}_{XY,\eta}) \geq \tfrac{1}{2}. \tag{3.5}$$

By definition of $\tilde{T}_{XY,\eta}$, we have the following decomposition:

$$\tilde{T}_{XY,\eta} = \bigcup_{\substack{\hat{P}_{XY} \in \mathscr{P}_n(\mathscr{X} \times \mathscr{Y}): \\ \|\hat{P}_{XY} - \tilde{P}_{XY}\| \leq \eta, \\ \hat{P}_{XY} \ll \tilde{P}_{XY}}} \hat{T}_{XY}.$$

Thus, observing that the elements of a given $\hat{T}_{XY}$ are equiprobable under any i.i.d. measure, we can rewrite (3.5) as

$$\sum_{\substack{\hat{P}_{XY} \in \mathscr{P}_n(\mathscr{X} \times \mathscr{Y}): \\ \|\hat{P}_{XY} - \tilde{P}_{XY}\| \leq \eta, \\ \hat{P}_{XY} \ll \tilde{P}_{XY}}} \tilde{P}^n_{XY}(\hat{T}_{XY}) \frac{|(\Gamma^k C \times \Gamma^k F) \cap \hat{T}_{XY}|}{|\hat{T}_{XY}|} \geq \frac{1}{2}.$$

At least one of the fractions in the above sum must be greater than or equal to 1/2; hence, there exists a type $\hat{P}_{XY} \in \mathscr{P}_n(\mathscr{X} \times \mathscr{Y})$ satisfying

$$\|\hat{P}_{XY} - \tilde{P}_{XY}\| \leq \eta \quad \text{and} \quad \hat{P}_{XY} \ll \tilde{P}_{XY},$$

and such that

$$\frac{|(\Gamma^k C \times \Gamma^k F) \cap \hat{T}_{XY}|}{|\hat{T}_{XY}|} \geq \frac{1}{2}.$$

Since pairs (x^n, y^n) of the same type are also equiprobable under Q^n_{XY}, we conclude that for the previous type $\hat{P}_{XY}$,

$$\begin{aligned} Q^n_{XY}&(\Gamma^k C \times \Gamma^k F) \\ &\geq Q^n_{XY}((\Gamma^k C \times \Gamma^k F) \cap \hat{T}_{XY}) \\ &= Q^n_{XY}(\hat{T}_{XY}) \frac{|(\Gamma^k C \times \Gamma^k F) \cap \hat{T}_{XY}|}{|\hat{T}_{XY}|} \\ &\geq \frac{1}{2} Q^n_{XY}(\hat{T}_{XY}). \end{aligned} \tag{3.6}$$

We have, thus, established that the probabilities of the sets $\Gamma^k C \times \Gamma^k F$ and $\hat{T}_{XY}$ are of the same exponential order under Q^n_{XY}. We now show that the same is true of the pair $\Gamma^k C \times \Gamma^k F$ and $C \times F$. The argument is similar to that given in [2, Section IV].

Consider an arbitrary element (u^n, v^n) of $\Gamma^k C \times \Gamma^k F$. By definition of Γ^k, there exists at least one element $(x^n, y^n) \in C \times F$ such that (x_i, y_i) differs from (u_i, v_i) for at most $2k_n$ values of i. We, thus, have

$$\begin{aligned} Q^n_{XY}(u^n, v^n) &= \prod_{i=1}^n Q_{XY}(u_i, v_i) \\ &\leq \rho^{-2k} \prod_{i=1}^n Q_{XY}(x_i, y_i) \\ &= \rho^{-2k} Q^n_{XY}(x^n, y^n), \end{aligned} \tag{3.7}$$

where

$$\rho \stackrel{\text{def}}{=} \min_{x \in \mathscr{X},\, y \in \mathscr{Y}} Q_{XY}(x, y) > 0.$$

As (u^n, v^n) ranges over $\Gamma^k C \times \Gamma^k F$, each element (x^n, y^n) of $C \times F$ will be selected at most $|\Gamma^k(x^n)| \cdot |\Gamma^k(y^n)|$ times. By virtue of this, (3.7) yields

$$Q^n_{XY}(\Gamma^k C \times \Gamma^k F) \leq \rho^{-2k} |\Gamma^k(x^n)||\Gamma^k(y^n)| Q^n_{XY}(C \times F).$$

From [8], we have the upper bound

$$|\Gamma^k(x^n)| \leq \exp\left[n\left(h\left(\frac{k_n}{n}\right) + \frac{k_n}{n}\log|\mathscr{X}|\right)\right],$$

where $h(\cdot)$ denotes the binary entropy function. Thus, we may write

$$Q^n_{XY}(\Gamma^k C \times \Gamma^k F) \leq \exp(n\xi_n) Q^n_{XY}(C \times F), \tag{3.8}$$

where

$$\xi_n = 2h\left(\frac{k_n}{n}\right) + \frac{k_n}{n}\log(|\mathscr{X}||\mathscr{Y}|) - \frac{2k_n}{n}\log\rho \to 0.$$

As a final step, we combine (3.6) and (3.8) with the upper bound on $Q^n_{XY}(\hat{T}_{XY})$ provided by Lemma 1. Thus,

$$\begin{aligned} Q^n_{XY}&(C \times F) \\ &\geq \tfrac{1}{2}\exp(-n\xi_n) Q^n_{XY}(\hat{T}_{XY}) \\ &\geq \frac{(n+1)^{-|\mathscr{X}||\mathscr{Y}|}}{2} \exp[-n(D(\hat{P}_{XY}\|Q_{XY}) + \xi_n)] \\ &\geq \exp[-n(D(\hat{P}_{XY}\|Q_{XY}) + \zeta_n)], \end{aligned}$$

where

$$\zeta_n = \zeta_n(\rho, \epsilon, M_n, |\mathscr{X}|, |\mathscr{Y}|) \to 0.$$

Over the range of pairs $(\tilde{P}_{XY}, \tilde{Q}_{XY})$ such that and $\tilde{Q}_{XY} \geq \rho$, the divergence functional $D(\tilde{P}_{XY}\|\tilde{Q}_{XY})$ is convex and bounded, and thus also uniformly continuous. It follows that we can find a sequence

$$\mu_n = \mu_n(\rho, |\mathscr{X}|, |\mathscr{Y}|) \to 0$$

such that

$$\|\hat{P}_{XY} - \tilde{P}_{XY}\| \leq \eta_n = n^{-1/3} \Rightarrow |D(\hat{P}_{XY}\|Q_{XY}) - D(\tilde{P}_{XY}\|Q_{XY})| \leq \mu_n.$$

Hence,

$$Q^n_{XY}(C \times F) \geq \exp[-n(D(\tilde{P}_{XY}\|Q_{XY}) + \zeta_n + \mu_n)], \tag{3.9}$$

and consequently

$$-\liminf_{n} \frac{1}{n} \log Q_{XY}^{n}(\mathscr{A}_n) \le D(\tilde{P}_{XY} \| Q_{XY}).$$

Since $\tilde{P}_{XY}$ satisfies the appropriate marginal constraints, the proof is complete. □

This result, in conjunction with the positive part of [3, Theorem 5] yields the following theorem.

Theorem 2: If $Q_{XY} > 0$, the error exponent for $H_0 : P_{XY}$ versus $H_1 : Q_{XY}$ under one-sided or two-sided zero-rate compression is given by

$$\theta(M, N, \epsilon) = \min_{\substack{\tilde{P}_{XY}: \\ \tilde{P}_X = P_X,\ \tilde{P}_Y = P_Y}} D(\tilde{P}_{XY} \| Q_{XY}).$$

Remark: In the proof of the converse theorem, the constants ζ_n and μ_n appearing on the right-hand side of (3.9) are independent of the distributions P_{XY}, $\tilde{P}_{XY}$, and depend on Q_{XY} only through the lower bound ρ. With this in mind, we state without proof the following variant of Theorem 1, which will be useful in establishing converse results in the sections that follow.

Theorem 3: Fix $\rho > 0$ and $\epsilon \in (0, 1)$, and let M_n be a sequence of integers satisfying (2.1). Then there exists a sequence

$$\nu_n = \nu_n(\rho, \epsilon, M_n, |\mathscr{X}|, |\mathscr{Y}|) \to 0$$

such that for every $\tilde{Q}_{XY} \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ that satisfies $\tilde{Q}_{XY} \ge \rho$, and every $\tilde{P}_{XY} \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})$, $C \in \mathscr{X}^n$, $F \in \mathscr{Y}^n$ that satisfy either

$$(\exists P_{XY} : P_X = \tilde{P}_X,\ P_Y = \tilde{P}_Y) \qquad P_{XY}^n(C \times F) \ge \frac{1-\epsilon}{M_n}$$

or, more generally,

$$\tilde{P}_X^n(C) \ge \frac{1-\epsilon}{M_n}, \qquad \tilde{P}_Y^n(F) \ge \frac{1-\epsilon}{M_n},$$

the following is true:

$$\tilde{Q}_{XY}^n(C \times F) \ge \exp\left[-n\left(D(\tilde{P}_{XY} \| \tilde{Q}_{XY}) + \nu_n\right)\right].$$

IV. Arbitrary Numbers of Sensors

The results of the previous section can be extended to multiterminal detection systems employing r sensors, where $r > 2$. Here the problem is that of testing $H_0 : P$ versus $H_1 : Q$, where P and Q are r-variate distributions. As in the case $r = 2$, we assume $Q > 0$.

Without going into detail, we give the following statements that can be established by suitably modifying the arguments in the proofs of [3, Theorem 5] and Theorem 1.

1) If at least $r - 1$ sensors employ zero-rate compression, then the error exponent is given by the minimum of $D(\tilde{P} \| Q)$ over all r-variate distributions $\tilde{P}$ whose univariate marginals agree with those of P. As in the bivariate case, the value of the exponent does not depend on the level ϵ and the codebook sizes. Thus, in particular, systems employing one-bit compression per source component can attain the same asymptotic performance as more complex systems employing zero-rate compression on $r - 1$ source components, and no compression at all on the remaining component.
2) More generally, if r_c out of r source components are compressed at asymptotically zero rate and the remaining $r_u = r - r_c$ are not compressed, the error exponent is given by the minimum of $D(\tilde{P} \| Q)$ over all distributions $\tilde{P}$ that agree with P on
   a) the univariate marginals corresponding to the compressed source components, and
   b) the r_u-variate marginal corresponding to the r_u components that are not compressed.

Thus, the latter r_u components are essentially treated as one. It also follows for $r_u \ge 2$ that if we impose zero-rate compression on any one of these r_u components, then the error exponent will (in general) decrease.

V. Composite Hypothesis Testing

In the remainder of this work, we consider issues of optimal zero-rate compression for *composite* hypothesis testing. For disjoint classes Π and Ξ of bivariate distributions on $\mathscr{X} \times \mathscr{Y}$, we wish to test

$$H_0 : P_{XY} \in \Pi \quad \text{against} \quad H_1 : Q_{XY} \in \Xi$$

subject to the compression rate constraints introduced earlier in this paper. The precise formulation of the problem appears in Section II-C.

The key result of our inquiry into the simple hypothesis testing problem with zero-rate data compression was that under a positivity assumption on the alternative distribution, the error exponent $\theta(M, N, \epsilon)$ exists and is independent of the sequences M, N and the level ϵ. Furthermore, as the proof of the positive part (given in [3]) indicates, it is possible to specify a sequence of asymptotically optimal acceptance regions *solely* in terms of the null distribution P, and thus the alternative distribution enters the picture only in the computation of the error exponent $\theta(M, N, \epsilon)$.

In examining the composite hypothesis testing problem, we have found that these conclusions are of limited validity in this case. That is, given two classes of distributions Π and Ξ, the error exponent for testing the hypothesis $H_0 : \Pi$ versus $H_1 : \Xi$ depends in general on the sequences M, N and the level ϵ. Furthermore, the choice of optimal acceptance regions is influenced by *both* Π and Ξ.

Our results, which are presented in the following two sections, highlight similarities and differences between the simple and composite hypothesis settings by reference to both general and specific problems parametrized by Π, Ξ, M and N. Some of the more specialized results admit generalizations, albeit at the expense of compactness in the characterization of the error exponent. It seems to us that the general problem of determining error exponents for arbitrary Π, Ξ, M and N resists coherent treatment, and is, thus, placed outside the scope of this work.

The following notation will be used in Sections VI and VII.

1) For a class Π of distributions on $\mathscr{X} \times \mathscr{Y}$, the corresponding classes of marginals are denoted by

$$\Pi_X = \{P_X \in \mathscr{P}(\mathscr{X}) : \exists P_{XY} \in \Pi\}$$
$$\text{and} \quad \Pi_Y = \{P_Y \in \mathscr{P}(\mathscr{Y}) : \exists P_{XY} \in \Pi\}.$$

2) If P_X, P_Y, Q_{XY} are distributions on $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{X} \times \mathscr{Y}$, respectively, we let

$$d(P_X, P_Y \| Q) \stackrel{\text{def}}{=} \min_{\substack{\tilde{P}_{XY}: \\ \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y}} D(\tilde{P}_{XY} \| Q_{XY}).$$

More generally, if Δ, Λ and Ξ are classes of distributions on the same spaces (respectively), then

$$d(\Delta, \Lambda \| \Xi) \stackrel{\text{def}}{=} \inf_{\substack{Q_{XY} \in \Xi, \\ \tilde{P}_{XY} : \tilde{P}_X \in \Delta,\, \tilde{P}_Y \in \Lambda}} D(\tilde{P}_{XY} \| Q_{XY}).$$

VI. Adequate Codebooks

Our main observation here is that if the codebook sizes M_n and N_n allow the codes to distinguish between distributions in the marginal classes Π_X and Π_Y derived from Π, then the error exponent has no further dependence on M, N, and ϵ, and is achieved by a sequence of acceptance regions specified solely in terms of Π.

We first consider the case in which M_n and N_n increase to infinity while satisfying the zero rate compression constraint, i.e.,

$$\lim_n M_n = \lim_n N_n = \infty$$

and

$$\lim_n \frac{1}{n} \log M_n = \lim_n \frac{1}{n} \log N_n = 0. \qquad (6.1)$$

In this case it is possible to encode the types of the observed sequences x^n and y^n with arbitrary accuracy and thus construct acceptance regions which are similar to those given in [3] for the simple hypothesis testing problem. To prove the converse, we employ Theorem 3, assuming the uniform positivity constraint

$$\rho_{\inf} \stackrel{\text{def}}{=} \inf_{Q \in \Xi} \min_{(x, y) \in \mathscr{X} \times \mathscr{Y}} Q_{XY}(x, y) > 0. \qquad (6.2)$$

This ensures that the convex function $D(\cdot \| \cdot)$ is bounded on $\mathscr{P}(\mathscr{X} \times \mathscr{Y}) \times \Xi$ and is thus uniformly continuous. The result is as follows.

Theorem 4: If $\Pi \subset \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ is arbitrary, and (6.2) on Ξ and (6.1) on M, N are satisfied, then

$$\theta(M, N, \epsilon) = \inf_{P_{XY} \in \Pi,\, Q_{XY} \in \Xi} d(P_X, P_Y \| Q_{XY}).$$

Proof: See Appendix A.

Next we assume that the codebook sizes are fixed in n, i.e., $M_n = M$, $N_n = N$. In this case, it is no longer possible to encode the type of the observed sequences with arbitrary accuracy, and the conclusion of Theorem 4 does not hold in general. However, in the special case

$$M \geq |\Pi_X| + 1, \qquad N \geq |\Pi_Y| + 1, \qquad (6.3)$$

it is still possible for the codes to distinguish between distributions in Π_X and Π_Y, and by a straightforward adaptation of the proof given for the positive part of Theorem 4, we readily obtain the following.

Theorem 5: If $\Pi \subset \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ is finite and (6.2) on Ξ and (6.3) on (M, N) are satisfied, then

$$\theta(M, N, \epsilon) = \inf_{P_{XY} \in \Pi,\, Q_{XY} \in \Xi} d(P_X, P_Y \| Q_{XY}). \qquad \square$$

VII. Inadequate Codebooks

In this section, we consider situations in which the prescribed codebook sizes do not allow the codes to distinguish between the distributions in Π_X and Π_Y. A consequence of this inadequacy is that the optimal system design will depend on both the null and the alternative distribution classes, as well as on the actual codebook sizes and the value of the level ϵ.

For simplicity we will assume that the class Π is finite. As we pointed out earlier, some of our proofs admit cumbersome but straightforward generalizations to situations in which Π is infinite. However, since our aim is to highlight salient differences from the simple hypothesis testing problem, we choose to restrict our attention to the simplest possible case. For this reason we will also assume that that the alternative hypothesis is simple, i.e., $\Xi = \{Q_{XY}\}$.

For finite Π, we already know from Theorem 5 that the codebooks will be adequate if they have fixed sizes M and N such that

$$M \geq |\Pi_X| + 1, \qquad N \geq |\Pi_Y| + 1.$$

Thus, we investigate what happens when at least one of the inequalities is not satisfied. Specifically, we consider three cases:

Problem a) $|\Pi| = |\Pi_X| = |\Pi_Y| = 2, M = 2, N = 3;$
Problem b) $|\Pi| = |\Pi_X| = |\Pi_Y| = 2, M = 2, N = 2;$
Problem c) $|\Pi| > |\Pi_X|,\ M = 2,\ N \geq |\Pi_Y|.$

In Problem a), if $\Pi = \{P_{XY}, \overline{P}_{XY}\}$, then the two codewords available to the S_X encoder do not suffice in order to classify the observed x^n sequence as (approximate) type P_X, (approximate) type $\overline{P}_X$, or neither. At first glance, the logical choice is to pair the two distributions in Π_X together and thus encode x^n as either lying in the set $T_{X,\eta} \cup \overline{T}_{X,\eta}$ or its complement. Yet, this encoder is not always optimal and may be outperformed by one that *separates* the distributions P_X and $\overline{P}_X$ by placing them in two disjoint and mutually exhaustive classes. The complete result is given by the following theorem.

Theorem 6: Let $\Pi = \{P_{XY}, \overline{P}_{XY}\}$, where $P_X \neq \overline{P}_X$ and $P_Y \neq \overline{P}_Y$. If $Q_{XY} > 0$, then for $0 < \epsilon < 1$,

$$\theta(2, 3, \epsilon) = \theta^{(1)} \vee \theta^{(2)},$$

where

$$\theta^{(1)} \stackrel{\text{def}}{=} d(\Pi_X, \Pi_Y \| Q)$$

and

$$\theta^{(2)} \stackrel{\text{def}}{=} d(P_X, P_Y \| Q) \wedge d(\bar{P}_X, \bar{P}_Y \| Q) \wedge \min_{\tilde{P}_X \in \mathscr{P}(\mathscr{X})} \{ d(\tilde{P}_X, P_Y \| Q) \vee d(\tilde{P}_X, \bar{P}_Y \| Q) \}.$$

Proof: See Appendix B.

In Problem b), where $M = N = 2$, *neither* encoder can provide a ternary classification of the received sequence. As in case a), either of two encoding schemes can be asymptotically optimal. The first classifies each of the two observed sequences as "type Π" or "not type Π," while the second uses partitions of $\mathscr{P}(\mathscr{X})$ and $\mathscr{P}(\mathscr{Y})$ that separate the marginals of P_{XY} and $\bar{P}_{XY}$. Specifically, we have

Theorem 7: Let $\Pi = \{P_{XY}, \bar{P}_{XY}\}$, where $P_X \neq \bar{P}_X$ and $P_Y \neq \bar{P}_Y$. If $Q_{XY} > 0$, then for $0 < \epsilon < 1$,

$$\theta(2, 2, \epsilon) = \theta^{(1)} \vee \theta^{(3)},$$

where $\theta^{(1)}$ is as defined in Theorem 6, and $\theta^{(3)}$ is the supremum, over all partitions $\{\Phi, \bar{\Phi}\}$ of $\mathscr{P}(\mathscr{X})$ and $\{\Psi, \bar{\Psi}\}$ of $\mathscr{P}(\mathscr{Y})$, of the quantity

$$d(\Phi \cup \{P_X\}, \Psi \cup \{P_Y\} \| Q) \wedge d(\bar{\Phi} \cup \{\bar{P}_X\}, \bar{\Psi} \cup \{\bar{P}_Y\} \| Q).$$

Proof: See Appendix C.

It should be noted that although the definition of $\theta^{(3)}$ given in the statement of Theorem 7 is seemingly more involved than the one given for $\theta^{(2)}$ in Theorem 6, it is analytically possible to reduce $\theta^{(3)}$ to an expression that involves no maximization, namely

$$\theta^{(3)} = \{ D(P_X \| Q_X) \wedge D(\bar{P}_Y \| Q_Y) \} \vee \{ D(\bar{P}_X \| Q_X) \wedge D(P_Y \| Q_Y) \}. \quad (7.1)$$

This characterization, which is derived in Appendix D, is also useful for determining the maximizing classes Φ and Ψ in the original definition of $\theta^{(3)}$.

We finally consider Problem c), which would be a straightforward generalization of Problem a) if it were not for the possibility that $|\Pi| > |\Pi_X|$. When this arises, i.e., two or more distributions in Π have the same X marginal P_X, it is possible to improve upon the second encoding scheme (the one which induces a nonempty partition of Π_X) by separating sequences that have approximate type P_X. The resulting scheme, however, is only admissible for values of the level ϵ greater than $1/2$, and thus the error exponent depends on ϵ. This is somewhat of a surprise, considering the chain of strong converse theorems which have been derived in [2]–[4], and in this work. The statement of our result is given below, where 1_X denotes the set of degenerate distributions of $\mathscr{P}(\mathscr{X})$.

Theorem 8: Let $\Pi < \infty$, $M = 2$, and $N \geq |\Pi_Y| + 1$. Also, let $\{\Delta, \bar{\Delta}\}$ denote a partition of Π. If $Q_{XY} > 0$, then for $\epsilon \in (0, 1/2) \cup (1/2, 1)$, the following is true:

$$\theta(2, N, \epsilon) = \theta^{(1)} \vee \theta^{(4)}(\epsilon),$$

where

$$\theta^{(1)} = d(\Pi_X, \Pi_Y \| Q),$$

$$\theta^{(4)}(\epsilon) = \begin{cases} \max_{\Delta, \bar{\Delta}: \Delta_X \cap \bar{\Delta}_X = \emptyset} \tau(\Delta, \bar{\Delta}), & \text{if } 0 < \epsilon < \frac{1}{2}, \\ \max_{\substack{\Delta, \bar{\Delta}: \\ \Delta_X \cap \bar{\Delta}_X \cap 1_X = \emptyset}} \tau(\Delta, \bar{\Delta}), & \text{if } \frac{1}{2} < \epsilon < 1, \end{cases}$$

and

$$\tau(\Delta, \bar{\Delta}) = d(\Delta_X, \Delta_Y \| Q) \wedge d(\bar{\Delta}_X, \bar{\Delta}_Y \| Q) \wedge \min_{\tilde{P}_X} \{ d(\tilde{P}_X, \Delta_Y \| Q) \vee d(\tilde{P}_X, \bar{\Delta}_Y \| Q) \}.$$

Remark: We have been unable to evaluate $\theta(2, N, 1/2)$.

Proof: See Appendix E.

The results of this section illustrate that the error exponent in distributed hypothesis testing with zero-rate compression depends on the codebook sizes M, N, as well as on the level ϵ; and that the choice of asymptotically optimal acceptance regions (those that achieve $\theta^{(2)}$, $\theta^{(3)}$ and $\theta^{(4)}(\epsilon)$ in Theorems 6, 7, and 8) is also affected by the alternative distribution Q. We should also add that the dominant error exponent in each of the above three theorems is not trivially determined, e.g., $\theta^{(1)}$ does not always dominate $\theta^{(2)}$. Furthermore, in regard to Theorem 8, we have examples in which the dominant error exponent $\theta^{(4)}(\epsilon)$ actually decreases as ϵ drops below $1/2$.

VIII. Concluding Remarks

The positivity assumption on the alternative hypothesis was essential for the derivation of the converse results in this paper. Without this assumption, we could not have applied the blowing-up lemma in the proof of the pivotal Theorems 2 and 3. The same difficulty was encountered in the proof of the converse result in [2, Theorem 6], which also employed the blowing-up lemma. We hope that this obstacle will eventually be removed.

In the meantime, we should note that in the case of simple hypothesis testing, there are instances where $Q \not> 0$ and $D(\tilde{P} \| Q)$ is trivially minimized by $\tilde{P} = P$. In such cases, the resulting minimum is equal to the error exponent under no data compression (cf. Stein's lemma [7]), and the converse result follows immediately.

We must also emphasize that Theorem 2 does not subsume its counterpart in [3]. Although the converse theorem appearing in that work was valid for one-bit compression of S_X and for $\epsilon \in (0, \epsilon_0)$ only, the hypothesis of that theorem did not impose any constraints on Q_{XY} other than $D(P_{XY} \| Q_{XY}) < \infty$.

Appendix A

Proof of Theorem 4

Direct Part: Let $a_n = \lfloor M_n^{1/|\mathscr{X}|} \rfloor$. Then by an elementary geometrical construction we can partition $\mathscr{P}(\mathscr{X})$ into at most $a_n^{|\mathscr{X}|} \leq M_n$ cells $\mathscr{C}_i^n$ of maximum dimension (measured by sup norm) not exceeding a_n^{-1}; clearly $a_n^{-1} \to 0$ since $M_n \to \infty$. The same is true for $\mathscr{P}(\mathscr{Y})$ with b_n replacing a_n, i.e., $b_n = \lfloor N_n^{1/|\mathscr{Y}|} \rfloor$.

We denote the $\mathscr{P}(\mathscr{Y})$-counterpart of $\mathscr{C}_i^n$ by $\mathscr{F}_j^n$, and we write

$$C_i^n = \bigcup_{\hat{P}_X \in \mathscr{C}_i^n} \hat{T}_X, \qquad F_j^n = \bigcup_{\hat{P}_Y \in \mathscr{F}_j^n} \hat{T}_Y.$$

Based on the above partition, we devise a compression/decision scheme as follows. First, we require that each encoder transmit the cell index corresponding to the observed type, i.e.,

$$f_n(x^n) = i, \qquad \text{iff } x^n \in C_i^n,$$
$$g_n(y^n) = j, \qquad \text{iff } y^n \in F_j^n.$$

Next, we seek an acceptance region $\mathscr{A}_n \subset \mathscr{X}^n \times \mathscr{Y}^n$ such that

$$\mathscr{A}_n \supset \bigcup_{P_{XY} \in \Pi} T^n_{XY,\eta}, \tag{A.1}$$

for some fixed $\eta > 0$. This is because the above set has P^n_{XY}-probability that *uniformly* approaches unity for all $P_{XY} \in \Pi$ (by Lemma 2), and this automatically ensures that the type I error bound is met for every $\epsilon \in (0,1)$. We define $\mathscr{A}_n$ as the smallest union of rectangles $C_i^n \times F_j^n$ that contains

$$\bigcup_{P_{XY} \in \Pi} T_{X,\xi} \times T_{Y,\xi},$$

where ξ is a multiple of η chosen so as to ensure that (A.1) holds.

Since ξ is fixed and the dimension of each C_i^n and F_j^n shrinks to zero as n approaches infinity, it is also true that for n sufficiently large,

$$\mathscr{A}_n \subset \bigcup_{P_{XY} \in \Pi} T_{X,2\xi} \times T_{Y,2\xi}.$$

By a standard argument based on the definition of typicality, we also have

$$T_{X,2\xi} \times T_{Y,2\xi} \subset \bigcup_{\substack{\tilde{P}_{XY}: \\ \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y}} \tilde{T}_{XY,\zeta},$$

where ζ is a fixed multiple of ξ and η. We conclude that

$$\mathscr{A}_n \subset \bigcup_{\substack{\tilde{P}_{XY}: \\ (\exists P_{XY} \in \Pi)\, \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y}} \tilde{T}_{XY,\zeta}.$$

A union bound on $Q^n(\mathscr{A}_n)$ for $Q \in \Xi$ can now be established using Lemma 1 and the fact that $D(\cdot \| \cdot)$ is uniformly continuous on $\mathscr{P}(\mathscr{X} \times \mathscr{Y}) \times \Xi$:

$$Q(\mathscr{A}_n) \le |\mathscr{P}_n(\mathscr{X} \times \mathscr{Y})| \exp\left[-n \inf_{\substack{\tilde{P}_{XY}: \\ (\exists P_{XY} \in \Pi)\, \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y}} \cdot \left(D(\tilde{P}_{XY} \| Q_{XY}) - \mu'(\zeta) \right) \right]$$
$$\le \exp\left[-n \inf_{P_{XY} \in \Pi} \left(d(P_X, P_Y \| Q_{XY}) - \mu(\zeta) \right) \right],$$

where $\mu(\zeta)$ goes to zero together with ζ (and, hence, also η). We, therefore, have

$$\beta_n(M_n, N_n, \epsilon) \le \exp\left[-n \inf_{P_{XY} \in \Pi,\, Q_{XY} \in \Xi} \cdot \left(d(P_X, P_Y \| Q_{XY}) - \mu(\zeta) \right) \right].$$

Since $\mu(\zeta)$ can be made arbitrarily small by choice of η, we conclude that

$$\theta(M, N, \epsilon) \ge \inf_{P_{XY} \in \Pi,\, Q_{XY} \in \Xi} d(P_X, P_Y \| Q_{XY}).$$

Converse Part: Let $\mathscr{A}_n$ be an admissible acceptance region. By (2.3), for every distribution P_{XY} in Π, we can find a rectangle $C \times F \subset \mathscr{X}^n \times \mathscr{Y}^n$ such that

$$P^n_{XY}(C \times F) \ge (1 - \epsilon)/M_n.$$

Applying Theorem 3 with $\rho = \rho_{\inf}$, we obtain a universal sequence $\nu_n \to 0$ with the property that for every $Q_{XY} \in \Xi$, $P_{XY} \in \Pi$ and $\tilde{P}_{XY} \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ such that $\tilde{P}_X = P_X$, $\tilde{P}_Y = P_Y$, the following is true:

$$Q^n_{XY}(\mathscr{A}_n) \ge \exp\left[-n\left(D(\tilde{P}_{XY} \| Q_{XY}) + \nu_n \right) \right].$$

We conclude that

$$\beta_n(M_n, N_n, \epsilon) \ge \exp\left[-n \inf_{\substack{\tilde{P}_{XY}: (\exists P_{XY} \in \Pi)\, \tilde{P}_X = P_X,\, \tilde{P}_Y = P_Y, \\ Q_{XY} \in \Xi}} \cdot \left(D(\tilde{P}_{XY} \| Q_{XY}) + \nu_n \right) \right]$$

and hence,

$$\theta(M, N, \epsilon) \le \inf_{P_{XY} \in \Pi,\, Q_{XY} \in \Xi} d(P_X, P_Y \| Q_{XY}). \qquad \square$$

Appendix B

Proof of Theorem 6

Direct Part: We restrict our attention to encoders that group sequences of the same type together. The sensible choice for the S_Y encoder is one that specifies whether the sequence y^n lies in $T_{Y,\eta}$, $\bar{T}_{Y,\eta}$ or $(T_{Y,\eta} \cup \bar{T}_{Y,\eta})^c$.

For the S_X encoder, the first choice is one that specifies whether the type of the observed sequence x^n is close to *either one or none* of the distributions P_X, $\bar{P}_X$, i.e.,

$$C_1 = T_{X,\eta} \cup \bar{T}_{X,\eta}, \qquad C_2 = \left(T_{X,\eta} \cup \bar{T}_{X,\eta} \right)^c.$$

With this choice of encoders, the smallest acceptance region that satisfies the type I error constraint under both P_{XY} and $\bar{P}_{XY}$ is

$$\mathscr{A}_n^{(1)} = \left(T_{X,\eta} \cup \bar{T}_{X,\eta} \right) \times \left(T_{Y,\eta} \cup \bar{T}_{Y,\eta} \right).$$

The Q^n-probability of the above set can be upper-bounded in the standard fashion (viz. Appendix A):

$$Q^n_{XY}(\mathscr{A}_n) \le \exp\left[-n \left(\min_{\tilde{P}_X \in \{P_X, \bar{P}_X\},\, \tilde{P}_Y \in \{P_Y, \bar{P}_Y\}} \cdot D(\tilde{P}_{XY} \| Q_{XY}) - \mu(\eta) \right) \right]$$

where $\mu(\eta) \to 0$ as $\eta \to 0$. This yields, since η is arbitrary small,

$$\theta(2, 3, \epsilon) \ge \theta^{(1)} = d(\Pi_X, \Pi_Y | Q). \tag{B.1}$$

The second candidate for the S_X encoder is one that *separates* sequences of approximate type P_X from ones of approximate type $\bar{P}_X$. Formally, we define a ball of radius η centered at P_X by

$$\mathscr{B}_\eta(P_X) \stackrel{\text{def}}{=} \left\{ \tilde{P}_X \in \mathscr{P}(\mathscr{X}) : \| \tilde{P}_X - P_X \| \le \eta,\, \tilde{P}_X \ll P_X \right\}. \tag{B.2}$$

We then consider two sets Φ and $\bar{\Phi}$ of distributions such that

$$\Phi \subset \mathscr{P}(\mathscr{X}) - \mathscr{B}_\eta(P_X) - \mathscr{B}_\eta(\bar{P}_X)$$
$$\text{and} \quad \bar{\Phi} = \mathscr{P}(\mathscr{X}) - \mathscr{B}_\eta(P_X) - \mathscr{B}_\eta(\bar{P}_X) - \Phi,$$

and we let the encoder partition $\mathscr{X}^n$ into

$$C_1' = T_{X,\eta} \cup \bigcup_{\tilde{P}_X \in \Phi} \tilde{T}_X$$

$$\text{and} \quad C_2' = (C_1')^c = \overline{T}_{X,\eta} \cup \bigcup_{\tilde{P}_X \in \overline{\Phi}} \tilde{T}_X. \tag{B.3}$$

This expression contains a slight abuse of notation, since not every $\tilde{P}_X$ in Φ or $\overline{\Phi}$ will be in $\mathscr{P}_n(\mathscr{X})$. In what follows, however, it will be convenient to write

$$\bigcup_{P_X \in \Phi} T_X \quad \text{and} \quad \bigcup_{P_X \in \Phi} T_{X,\eta},$$

for

$$\bigcup_{\hat{P}_X \in \Phi \cap \mathscr{P}_n(\mathscr{X})} \hat{T}_X \quad \text{and} \quad \bigcup_{\hat{P}_X \in \mathscr{B}_\eta(\Phi) \cap \mathscr{P}_n(\mathscr{X})} \hat{T}_X,$$

respectively.

The second choice of S_X encoder (together with the S_Y encoder introduced in the beginning of the proof) yields the acceptance region

$$\mathscr{A}_n^{(2)} = (C_1' \times T_{Y,\eta}) \cup (C_2' \times \overline{T}_{Y,\eta}).$$

Note that unlike $\mathscr{A}_n^{(1)}$, $\mathscr{A}_n^{(2)}$ does not contain $T_{X,\eta} \times \overline{T}_{Y,\eta}$ or $\overline{T}_{X,\eta} \times T_{Y,\eta}$. It does, however, contain pairs (x^n, y^n) whose marginal type λ_x is close to neither P_X nor $\overline{P}_X$.

To estimate $Q^n(\mathscr{A}_n^{(2)})$, we decompose each of C_1' and C_2' into two sets as in definition (B.3). We then treat $\mathscr{A}_n^{(2)}$ as a union of four disjoint sets, and upper-bound their Q^n-probabilities in the usual way:

$$Q_{XY}^n(T_{X,\eta} \times T_{Y,\eta}) \le \exp\left[-n(d(P_X, P_Y \| Q) - \mu(\eta))\right],$$

$$Q_{XY}^n(\overline{T}_{X,\eta} \times \overline{T}_{Y,\eta}) \le \exp\left[-n(d(\overline{P}_X, \overline{P}_Y \| Q) - \mu(\eta))\right],$$

$$Q_{XY}^n\left(\bigcup_{\tilde{P}_X \in \Phi} \tilde{T}_X \times T_{Y,\eta}\right) \le \exp\left[-n\left(\inf_{\tilde{P}_X \in \Phi} d(\tilde{P}_X, P_Y \| Q) - \mu(\eta)\right)\right],$$

$$Q_{XY}^n\left(\bigcup_{\tilde{P}_X \in \overline{\Phi}} \tilde{T}_X \times \overline{T}_{Y,\eta}\right) \le \exp\left[-n\left(\inf_{\tilde{P}_X \in \overline{\Phi}} d(\tilde{P}_X, \overline{P}_Y \| Q) - \mu(\eta)\right)\right],$$

where $\mu(\eta) \to 0$ as $\eta \to 0$.

Thus, the error exponent associated with this choice of acceptance region is greater than or equal to the minimum of the four exponents appearing in the above bounds, namely the quantity

$$d(P_X, P_Y \| Q) \wedge d(\overline{P}_X, \overline{P}_Y \| Q) \wedge \inf_{\tilde{P}_X \in \Phi} d(\tilde{P}_X, P_Y \| Q) \wedge \inf_{\tilde{P}_X \in \overline{\Phi}} d(\tilde{P}_X, \overline{P}_Y \| Q).$$

At this point, we should note that by letting η shrink to zero, we have expanded the classes Φ and $\overline{\Phi}$ in the vicinity P_X and $\overline{P}_X$ so that $\Phi \cup \overline{\Phi} = \mathscr{P}(\mathscr{X}) - \{P_X\} - \{\overline{P}_X\}$. This is justified by continuity of $d(\cdot\,, \cdot \| Q)$, which further allows us to treat Φ and $\overline{\Phi}$ in the previous expression as constituting a partition of $\mathscr{P}(\mathscr{X})$.

It remains to find that partition $\{\Phi, \overline{\Phi}\}$ of $\mathscr{P}(\mathscr{X})$ which maximizes

$$v(P_X) \wedge \bar{v}(\overline{P}_X) \wedge \inf_{\tilde{P}_X \in \Phi} v(\tilde{P}_X) \wedge \inf_{\tilde{P}_X \in \overline{\Phi}} \bar{v}(\tilde{P}_X),$$

where $v(\cdot) \stackrel{\text{def}}{=} d(\cdot\,, P_Y \| Q)$ and $\bar{v}(\cdot) \stackrel{\text{def}}{=} d(\cdot\,, \overline{P}_Y \| Q)$. This is easily accomplished by noting that

$$\begin{aligned}
&\inf_{\Phi} v(\tilde{P}_X) \wedge \inf_{\overline{\Phi}} \bar{v}(\tilde{P}_X) \\
&\le \inf_{\Phi}\left[v(\tilde{P}_X) \vee \bar{v}(\tilde{P}_X)\right] \wedge \inf_{\overline{\Phi}}\left[v(\tilde{P}_X) \vee \bar{v}(\tilde{P}_X)\right] \\
&= \inf_{\mathscr{P}(\mathscr{X})}\left[v(\tilde{P}_X) \vee \bar{v}(\tilde{P}_X)\right] \\
&= \inf_{\tilde{P}_X : v(\tilde{P}_X) \ge \bar{v}(\tilde{P}_X)} v(\tilde{P}_X) \wedge \inf_{\tilde{P}_X : v(\tilde{P}_X) < \bar{v}(\tilde{P}_X)} \bar{v}(\tilde{P}_X).
\end{aligned}$$

Thus, an optimal partition consists of the sets

$$\Phi = \{\tilde{P}_X : v(\tilde{P}_X) \ge \bar{v}(\tilde{P}_X)\}$$

$$\text{and} \quad \tilde{\Phi} = \{\tilde{P}_X : v(\tilde{P}_X) < \bar{v}(\tilde{P}_X)\},$$

and the error exponent associated with the resulting $\mathscr{A}^{(2)}$ is given by

$$\theta^{(2)} = d(P_X, P_Y \| Q) \wedge d(\overline{P}_X, \overline{P}_Y \| Q) \wedge \min_{\tilde{P}_X \in \mathscr{P}(\mathscr{X})} \{d(\tilde{P}_X, P_Y \| Q) \vee d(\tilde{P}_X, \overline{P}_Y \| Q)\}.$$

We conclude that $\theta(2, 3, \epsilon) \ge \theta^{(2)}$, and in light of (B.1),

$$\theta(2, 3, \epsilon) \ge \theta^{(1)} \vee \theta^{(2)}.$$

Converse Part: For fixed n, consider an admissible acceptance region $\mathscr{A}_n$. By nature of the encoding, $\mathscr{A}_n$ can be written as

$$\mathscr{A}_n = (C_1 \times F_1) \cup (C_2 \times F_2),$$

where C_1 and C_2 form a partition of $\mathscr{X}^n$, and at most one of F_1, F_2 may be empty. From the type I error constraint

$$P_{XY}(\mathscr{A}_n) \ge 1 - \epsilon \quad \text{and} \quad \overline{P}_{XY}(\mathscr{A}_n) \ge 1 - \epsilon,$$

it follows that two cases may arise.

Case 1: For i and j distinct, we have

$$P_{XY}^n(C_i \times F_i) \ge P_{XY}^n(C_j \times F_j)$$

$$\text{and} \quad \overline{P}_{XY}^n(C_i \times F_i) \ge \overline{P}_{XY}^n(C_j \times F_j).$$

This clearly implies that

$$\tilde{P}_X^n(C_i) \ge \frac{1-\epsilon}{2} \quad \text{and} \quad \tilde{P}_Y^n(F_i) \ge \frac{1-\epsilon}{2},$$

for any $\tilde{P}_X \in \Pi_X$, $\tilde{P}_Y \in \Pi_Y$. From Theorem 3, we obtain

$$-\frac{1}{n}\log Q_{XY}^n(C_i \times F_i) \le d(\Pi_X, \Pi_Y \| Q) + \nu_n = \theta^{(1)} + \nu_n,$$

where $\nu_n \to 0$ as $n \to \infty$, and, thus, also

$$-\frac{1}{n}\log Q^n(\mathscr{A}_n) \le \theta^{(1)} + \nu_n. \tag{B.4}$$

Case 2: For i and j distinct, we have

$$P_{XY}^n(C_i \times F_i) \ge P_{XY}^n(C_j \times F_j)$$

$$\text{and} \quad \overline{P}_{XY}^n(C_i \times F_i) < \overline{P}_{XY}^n(C_j \times F_j). \tag{B.5}$$

Using Theorem 3 once again, we obtain respectively

$$-\frac{1}{n}\log Q^n_{XY}(C_i \times F_i) \le d(P_X, P_Y \| Q) + \nu_n$$

and

$$-\frac{1}{n}\log Q^n_{XY}(C_j \times F_j) \le d(\bar{P}_X, \bar{P}_Y \| Q) + \nu_n.$$

Hence,

$$-\frac{1}{n}\log Q^n(\mathscr{A}_n) \le d(P_X, P_Y \| Q) \wedge d(\bar{P}_X, \bar{P}_Y \| Q) + \nu_n. \tag{B.6}$$

Relationship (B.5) also implies that

$$P^n_Y(F_i) \ge \frac{1-\epsilon}{2} \quad \text{and} \quad \bar{P}^n_Y(F_j) \ge \frac{1-\epsilon}{2}.$$

By virtue of Theorem 3, these inequalities can lead to a further upper bound on $Q^n(\mathscr{A}_n)$ provided there exists a distribution $\tilde{P}_X \in \mathscr{P}(\mathscr{X})$ for which either $\tilde{P}^n_X(C_i)$ or $\tilde{P}^n_X(C_j)$ exceeds a fixed value independent of n. But the last disjunction is true for every $\tilde{P}_X$, since C_i and C_j are complementary events. We, thus, obtain the upper bound

$$-\frac{1}{n}\log Q^n_{XY}(\mathscr{A}_n) \le \min_{\tilde{P}_X \in \mathscr{P}(\mathscr{X})} \cdot \{ d(\tilde{P}_X, P_Y \| Q) \vee d(\tilde{P}_X, \bar{P}_Y \| Q)\} + \nu_n,$$

which, together with (B.6), yields

$$-\frac{1}{n}\log Q^n_{XY}(\mathscr{A}_n) \le \theta^{(2)} + \nu_n.$$

Finally, by combining the bound for Case 1 (B.4) with the above bound for Case 2, we obtain the converse statement

$$\theta(2,3,\epsilon) \le \theta^{(1)} \vee \theta^{(2)}. \qquad \square$$

Appendix C

Proof of Theorem 7

Direct Part. Since $M = 2$ as in Theorem 6, we consider the same two candidates for the S_X encoder:

$$f: C_1 = T_{X,\eta} \cup \bar{T}_{X,\eta}, \qquad C_2 = (T_{X,\eta} \cup \bar{T}_{X,\eta})^c$$

and

$$f': C_1' = T_{X,\eta} \cup \bigcup_{\tilde{P}_X \in \Phi} \tilde{T}_X, \qquad C_2' = \bar{T}_{X,\eta} \cup \bigcup_{\tilde{P}_X \in \bar{\Phi}} \tilde{T}_X,$$

where $(\Phi, \bar{\Phi})$ form a partition of $\mathscr{P}(\mathscr{X}) - \mathscr{B}_\eta(P_X) - \mathscr{B}_\eta(\bar{P}_X)$. Observe that, in this case, $N = 2$ also, and thus, it is no longer possible for the S_Y encoder to specify whether y^n lies in $T_{Y,\eta}$, $\bar{T}_{Y,\eta}$ or $(T_{Y,\eta} \cup \bar{T}_{Y,\eta})^c$. Proceeding as for S_X, we propose the following two encoders for S_Y:

$$g: F_1 = T_{Y,\eta} \cup \bar{T}_{Y,\eta}, \qquad F_2 = (T_{Y,\eta} \cup \bar{T}_{Y,\eta})^c$$

and

$$g': F_1' = T_{Y,\eta} \cup \bigcup_{\tilde{P}_Y \in \Psi} \tilde{T}_Y, \qquad F_2' = \bar{T}_{Y,\eta} \cup \bigcup_{\tilde{P}_Y \in \bar{\Psi}} \tilde{T}_Y,$$

where $(\Psi, \bar{\Psi})$ are defined in a similar manner.

Given these possibilities for encoding S_X and S_Y, there are only two reasonable choices for the acceptance region $\mathscr{A}_n$:

$$\mathscr{A}_n^{(1)} = C_1 \times F_1 \quad \text{and} \quad \mathscr{A}_n^{(3)} = (C_1' \times F_1') \cup (C_2' \times F_2').$$

Noting that the region $\mathscr{A}_n^{(1)}$ is identical to the one used in the proof of Theorem 6, we obtain

$$\theta(2,2,\epsilon) \ge \theta^{(1)} = d(\Pi_X, \Pi_Y \| Q).$$

To evaluate the error exponent associated with $\mathscr{A}_n^{(3)}$, we follow the corresponding procedure for $\mathscr{A}_n^{(2)}$ in the proof of Theorem 6. Since

$$\mathscr{A}_n^{(3)} = \left(\bigcup_{\Phi \cup \mathscr{B}_\eta(P_X)} \tilde{T}_X \times \bigcup_{\Psi \cup \mathscr{B}_\eta(P_Y)} \tilde{T}_Y \right) \cup \left(\bigcup_{\bar{\Phi} \cup \mathscr{B}_\eta(\bar{P}_X)} \tilde{T}_X \times \bigcup_{\bar{\Psi} \cup \mathscr{B}_\eta(\bar{P}_Y)} \tilde{T}_Y \right),$$

we obtain

$$-\lim_n \frac{1}{n}\log Q^n(\mathscr{A}_n^{(3)}) = d(\Phi \cup \{P_X\}, \Psi \cup \{P_Y\} \| Q) \wedge d(\bar{\Phi} \cup \{\bar{P}_X\}, \bar{\Psi} \cup \{\bar{P}_Y\} \| Q).$$

Once again, it is legitimate to assume that in the above equation, $\{\Phi, \bar{\Phi}\}$, $\{\Psi, \bar{\Psi}\}$ constitute partitions of the entire spaces $\mathscr{P}(\mathscr{X})$ and $\mathscr{P}(\mathscr{Y})$, respectively. The best error exponent attainable by a sequence of acceptance regions of the form $\mathscr{A}_n^{(3)}$ is therefore

$$\theta^{(3)} = \sup_{\Phi, \Psi} \{ d(\Phi \cup \{P_X\}, \Psi \cup \{P_Y\} \| Q) \wedge d(\bar{\Phi} \cup \{\bar{P}_X\}, \bar{\Psi} \cup \{\bar{P}_Y\} \| Q) \}.$$

We conclude that

$$\theta(2,2,\epsilon) \ge \theta^{(1)} \vee \theta^{(3)}.$$

Converse Part: In this case every admissible acceptance region $\mathscr{A}_n$ can be written as

$$\mathscr{A}_n = (C_1 \times F_1) \cup (C_2 \times F_2),$$

where C_1, C_2 are complementary, while F_1, F_2 are constrained by $F_2 \in \{\emptyset, \mathscr{Y}^n, F_1^c\}$. As in the proof of Theorem 6, two cases may arise.

Case 1: For i and j distinct, we have

$$P^n_{XY}(C_i \times F_i) \ge P^n_{XY}(C_j \times F_j)$$
$$\text{and} \quad \bar{P}^n_{XY}(C_i \times F_i) \ge \bar{P}^n_{XY}(C_j \times F_j).$$

This is same as Case 1 in the proof of Theorem 6, whence we obtain

$$-\frac{1}{n}\log Q^n(\mathscr{A}_n) \le \theta^{(1)} + \nu_n.$$

Note that this case subsumes the situation in which F_2 is empty.

Case 2: For i and j distinct, we have

$$P^n_{XY}(C_i \times F_i) \ge P^n_{XY}(C_j \times F_j)$$
$$\text{and} \quad \bar{P}^n_{XY}(C_i \times F_i) < \bar{P}^n_{XY}(C_j \times F_j).$$

We easily deduce that

$$P^n_X(C_i) \ge \frac{1-\epsilon}{2}, \qquad P^n_Y(F_i) \ge \frac{1-\epsilon}{2},$$

and

$$\bar{P}_X^n(C_j) \geq \frac{1-\epsilon}{2}, \qquad \bar{P}_Y^n(F_j) \geq \frac{1-\epsilon}{2}.$$

Let us define the classes

$$\Phi_n = \{\tilde{P}_X : \tilde{P}_X^n(C_i) \geq \tfrac{1}{2}\}, \qquad \Psi_n = \{\tilde{P}_Y : \tilde{P}_Y^n(F_i) \geq \tfrac{1}{2}\},$$

and

$$\Phi_n^* = \{\tilde{P}_X : \tilde{P}_X^n(C_j) > \tfrac{1}{2}\}, \qquad \Psi_n^* = \{\tilde{P}_Y : \tilde{P}_Y^n(F_j) > \tfrac{1}{2}\}.$$

Since C_1 and C_2 are complementary, $\Phi_n^* = \Phi_n^c$. For F_1 and F_2, we have either $F_2 = F_1^c$ or $F_2 = \mathscr{Y}^n$. In the former case we have again $\Psi_n^* = \Psi_n^c$, while in the latter, either Ψ_n or Ψ_n^* is equal to $\mathscr{P}(\mathscr{Y})$.

By the foregoing discussion, all marginal distributions $\tilde{P}_X \in \Phi_n \cup \{P_X\}$, $\tilde{P}_Y \in \Psi_n \cup \{P_Y\}$, satisfy

$$\tilde{P}_X^n(C_i) \geq \frac{1-\epsilon}{2} \quad \text{and} \quad \tilde{P}_Y^n(F_i) \geq \frac{1-\epsilon}{2}.$$

Applying Theorem 3, we obtain

$$-\frac{1}{n}\log Q_{XY}^n(C_i \times F_i) \leq d(\Phi_n \cup \{P_X\}, \Psi_n \cup \{P_Y\} \| Q) + \nu_n. \quad \text{(C.1)}$$

Similarly, for $C_j \times F_j$, we have

$$-\frac{1}{n}\log Q_{XY}^n(C_j \times F_j) \leq d(\Phi_n^* \cup \{\bar{P}_X\}, \Psi_n^* \cup \{\bar{P}_Y\} \| Q) + \nu_n. \quad \text{(C.2)}$$

We must show that the smaller of the two bounds appearing in (C.1) and (C.2) is less than or equal to $\theta^{(3)}$ as defined in the statement of the theorem. This is certainly true if $\Psi_n^* = \Psi_n^c$, since we can then take

$$\{\Phi, \bar{\Phi}\} = \{\Phi_n, \Phi_n^*\} \quad \text{and} \quad \{\Psi, \bar{\Psi}\} = \{\Psi_n, \Psi_n^*\}$$

in the definition of $\theta^{(3)}$. Otherwise, if w.l.o.g. $\Psi_n^* = \mathscr{P}(\mathscr{Y})$, the same conclusion can be reached by taking

$$\{\Phi, \bar{\Phi}\} = \{\Phi_n, \Phi_n^*\} \quad \text{and} \quad \{\Psi, \bar{\Psi}\} = \{\Psi_n, \Psi_n^c\}.$$

Thus, we have obtained

$$-\frac{1}{n}\log Q_{XY}^n(\mathscr{A}_n) \leq \theta^{(3)} + \nu_n.$$

This, together with our result for Case 1, yields the converse statement

$$\theta(2,2,\epsilon) \leq \theta^{(1)} \vee \theta^{(3)}. \qquad \square$$

Appendix D

Derivation of (7.1)

If we let, for all $\Phi \subset \mathscr{P}(\mathscr{X})$ and $\Psi \subset \mathscr{P}(\mathscr{Y})$,

$$\alpha(\Phi, \Psi) \stackrel{\text{def}}{=} \inf_{(\tilde{P}_X, \tilde{P}_Y) \in (\Phi \times \Psi) \cup (\Phi^c \times \Psi^c)} d(\tilde{P}_X, \tilde{P}_Y \| Q),$$

then the definition of $\theta^{(3)}$ becomes

$$\theta^{(3)} = \sup_{\substack{\Phi, \Psi: \\ (P_X, P_Y) \in \Phi \times \Psi, \\ (\bar{P}_X, \bar{P}_Y) \in \Phi^c \times \Psi^c}} \alpha(\Phi, \Psi). \quad \text{(D.1)}$$

We must show that $\theta^{(3)}$ can be expressed as in (7.1), or equivalently, that $\theta^{(3)} = \theta'$, where

$$\theta' \stackrel{\text{def}}{=} \{D(P_X \| Q_X) \wedge D(\bar{P}_Y \| Q_Y)\} \vee \{D(\bar{P}_X \| Q_X) \wedge D(P_Y \| Q_Y)\}.$$

1) To show that $\theta^{(3)} \geq \theta'$, let $\Phi = \{P_X\}$ and $\Psi = \{\bar{P}_Y\}^c$. Then

$$\theta^{(3)} \geq \alpha(\Phi, \Psi) = d(P_X, \{\bar{P}_Y\}^c \| Q) \wedge d(\{P_X\}^c, \bar{P}_Y \| Q) = D(P_X \| Q_X) \wedge D(\bar{P}_Y \| Q_Y), \quad \text{(D.2)}$$

where the last equality follows by continuity of divergence. Similarly, if $\Phi = \{\bar{P}_X\}^c$ and $\Psi = \{P_Y\}$, we have

$$\theta^{(3)} \geq D(P_Y \| Q_Y) \wedge D(\bar{P}_X \| Q_X). \quad \text{(D.3)}$$

Combining (D.2) with (D.3) we obtain $\theta^{(3)} \leq \theta'$.

2) To show the reverse inequality $\theta^{(3)} \leq \theta'$, let

$$A = \text{cl}\,\Phi, \quad \bar{A} = \text{cl}\,\Phi^c, \quad B = \text{cl}\,\Psi, \quad \bar{B} = \text{cl}\,\Psi^c,$$

where cl denotes closure under sup norm. Then by continuity of divergence,

$$\alpha(\Phi, \Psi) = \min_{(\tilde{P}_X, \tilde{P}_Y) \in (A \times B) \cup (\bar{A} \times \bar{B})} d(\tilde{P}_X, \tilde{P}_Y \| Q).$$

We must show that $\alpha(\Phi, \Psi) \leq \theta'$ for every Φ and Ψ. This is trivially true if $(Q_X, Q_Y) \in (A \times B) \cup (\bar{A} \times \bar{B})$, in which case we have

$$\alpha(\Phi, \Psi) = d(Q_X, Q_Y \| Q) = 0.$$

Hence, we may assume that

$$(Q_X, Q_Y) \notin (A \times B) \cup (\bar{A} \times \bar{B}). \quad \text{(D.4)}$$

We provide an upper bound on $\alpha(\Phi, \Psi)$ as follows. First we note that

$$(A \cap \bar{A}) \times \mathscr{P}(\mathscr{Y}) \subset (A \times B) \cup (\bar{A} \times \bar{B}),$$

so that

$$\alpha(\Phi, \Psi) \leq \min_{(\tilde{P}_X, \tilde{P}_Y) \in (A \cap \bar{A}) \times \mathscr{P}(\mathscr{Y})} d(\tilde{P}_X, \tilde{P}_Y \| Q) = \min_{(\tilde{P}_X, \tilde{P}_Y):\, \tilde{P}_X \in A \cap \bar{A}} D(\tilde{P}_{XY} \| Q_{XY}).$$

Using the log-sum inequality, we can show that above minimum is equal to

$$\min_{\tilde{P}_X \in A \cap \bar{A}} D(\tilde{P}_X \| Q_X).$$

By symmetry we conclude that

$$\alpha(\Phi, \Psi) \leq \min_{\tilde{P}_X \in A \cap \bar{A}} D(\tilde{P}_X \| Q_X) \wedge \min_{\tilde{P}_Y \in B \cap \bar{B}} D(\tilde{P}_Y \| Q_Y). \quad \text{(D.5)}$$

Two cases may arise, according to whether Q_X lies in A or $\bar{A}$ (note that it cannot lie in $A \cap \bar{A}$ by (D.4)).

Case 1: $Q_X \in A$: Since $\bar{P}_X \in \bar{A}$, there exists $\lambda \in (0, 1]$ such that

$$\hat{P}_X = \lambda \bar{P}_X + (1-\lambda) Q_X \in A \cap \bar{A}.$$

This yields

$$\min_{\tilde{P}_X \in A \cap \bar{A}} D(\tilde{P}_X \| Q_X) \le D(\hat{P}_X \| Q_X)$$
$$\le \lambda D(\bar{P}_X \| Q_X) + (1-\lambda) D(Q_X \| Q_X)$$
$$\le D(\bar{P}_X \| Q_X),$$

where the last inequality follows by convexity of divergence.

From (D.4), we also have the $Q_Y \in \bar{B}$. An analogous argument for $Q_Y \in \bar{B}$ and $P_Y \in B$ yields

$$\min_{\tilde{P}_Y \in B \cap \bar{B}} D(\tilde{P}_Y \| Q_Y) \le D(P_Y \| Q_Y).$$

From (D.5), we conclude that

$$\alpha(\Phi, \Psi) \le D(\bar{P}_X \| Q_X) \wedge D(P_Y \| Q_Y). \qquad (D.6)$$

Case 2: $Q_X \in \bar{A}$: Again (D.4) implies that $Q_Y \in B$. As in Case 1, we obtain

$$\alpha(\Phi, \Psi) \le D(P_X \| Q_X) \wedge D(\bar{P}_Y \| Q_Y). \qquad (D.7)$$

From (D.6) and (D.7) we conclude that $\alpha(\Phi, \Psi) \le \theta'$, and hence also $\theta^{(3)} \le \theta'$. □

Appendix E

Proof of Theorem 8

Direct Part: Once again we take $\mathscr{A}_n^{(1)}$ as in the proof of Theorem 6, whence we obtain $\theta(2, N, \epsilon) \ge \theta^{(1)}$.

To construct $\mathscr{A}_n^{(2)}$ by analogy to Theorem 6, we partition the space Π_X into Λ, $\bar{\Lambda}$, and the space $\mathscr{P}(\mathscr{X}) - \mathscr{B}_\eta(\Pi_X)$ into Φ, $\bar{\Phi}$. We then have

$$\mathscr{A}_n^{(2)} = \Bigg(\bigcup_{\tilde{P}_X \in \Phi \cup \mathscr{B}_\eta(\Lambda)} \tilde{T}_X \times \bigcup_{P_{XY} \in \Pi:\, P_X \in \Lambda} \tilde{T}_{Y,\eta} \Bigg) \cup \Bigg(\bigcup_{\tilde{P}_X \in \bar{\Phi} \cup \mathscr{B}_\eta(\bar{\Lambda})} \tilde{T}_X \times \bigcup_{P_{XY} \in \Pi:\, P_X \in \bar{\Lambda}} \tilde{T}_{Y,\eta} \Bigg),$$

which is readily seen to satisfy the type I error constraint for every ϵ and every distribution in Π.

Note that instead of partitioning Π_X into Λ and $\bar{\Lambda}$, one can begin by partitioning Π itself into Δ and $\bar{\Delta}$ such that $\Delta_X \cap \bar{\Delta}_X = \emptyset$. Then one can write equivalently

$$\mathscr{A}_n^{(2)} = \Bigg(\bigcup_{\tilde{P}_X \in \Phi \cup \mathscr{B}_\eta(\Delta_X)} \tilde{T}_X \times \bigcup_{\tilde{P}_Y \in \Delta_Y} \tilde{T}_{Y,\eta} \Bigg) \cup \Bigg(\bigcup_{\tilde{P}_X \in \bar{\Phi} \cup \mathscr{B}_\eta(\bar{\Delta}_X)} \tilde{T}_X \times \bigcup_{\tilde{P}_Y \in \bar{\Delta}_Y} \tilde{T}_{Y,\eta} \Bigg),$$

and by the argument given in the proof of Theorem 6,

$$\theta(2, N, \epsilon) \ge \tau(\Delta, \bar{\Delta}) = d(\Delta_X, \Delta_Y \| Q) \wedge d(\bar{\Delta}_X, \bar{\Delta}_Y \| Q) \wedge \min_{\tilde{P}_X} \{ d(\tilde{P}_X, \Delta_Y \| Q) \vee d(\tilde{P}_X, \bar{\Delta}_Y \| Q) \}.$$

Taking the maximum over all partitions $\{\Delta, \bar{\Delta}\}$ of Π satisfying $\Delta_X \cap \bar{\Delta}_X = \emptyset$, we obtain for all $\epsilon \in (0,1)$,

$$\theta(2, N, \epsilon) \ge \max_{\Delta, \bar{\Delta}:\, \Delta_X \cap \bar{\Delta}_X = \emptyset} \tau(\Delta, \bar{\Delta}).$$

The constraint $\Delta_X \cap \bar{\Delta}_X = \emptyset$ is essential in the above construction of $\mathscr{A}_n^{(2)}$; its removal would allow

$$C_1' = \bigcup_{\tilde{P}_X \in \Phi \cup \mathscr{B}_\eta(\Delta_X)} \tilde{T}_X \quad \text{and} \quad C_2' = \bigcup_{\tilde{P}_X \in \bar{\Phi} \cup \mathscr{B}_\eta(\bar{\Delta}_X)} \tilde{T}_X$$

to have nonempty intersection and, hence, be inadmissible under the given compression scheme. If, however, $1/2 < \epsilon < 1$, then it is possible to relax the said constraint to

$$\Delta_X \cap \bar{\Delta}_X \cap 1_X = \emptyset$$

in the following manner. For every $\tilde{P}_X$ that lies in $\mathscr{B}_\eta(\Delta_X \cap \bar{\Delta}_X)$ (and hence not in 1_X if η is properly chosen), we can partition $\tilde{T}_X$ into two sets $\tilde{T}_X^+$ and $\tilde{T}_X^-$ of sizes that differ by at most 1, and redefine C_1' and C_2' by

$$C_1' = \bigcup_{\tilde{P}_X \in \Phi \cup \mathscr{B}_\eta(\Delta_X - \bar{\Delta}_X)} \tilde{T}_X \cup \bigcup_{\tilde{P}_X \in \mathscr{B}_\eta(\Delta_X \cap \bar{\Delta}_X)} \tilde{T}_X^+$$

and

$$C_2' = \bigcup_{\tilde{P}_X \in \bar{\Phi} \cup \mathscr{B}_\eta(\bar{\Delta}_X - \Delta_X)} \tilde{T}_X \cup \bigcup_{\tilde{P}_X \in \mathscr{B}_\eta(\Delta_X \cap \bar{\Delta}_X)} \tilde{T}_X^-.$$

We can then complete the construction of $\mathscr{A}_n^{(2)}$ in the usual manner.

It is easily seen that for every $P_{XY} \in \Pi$ such that $P_X \notin \Delta_X \cap \bar{\Delta}_X$, and every $\epsilon \in (0,1)$,

$$P_{XY}^n(\mathscr{A}_n^{(2)}) \ge 1 - \epsilon$$

for n sufficiently large. The same is true for every $P_{XY} \in \Pi$ such that $P_X \in \Delta_X \cap \bar{\Delta}_X$, if $\epsilon \in (1/2, 1)$. To see this, let w.l.o.g. $P_{XY} \in \Delta$. Then

$$P_{XY}^n(\mathscr{A}_n^{(2)}) \ge P_{XY}^n \Bigg(\bigcup_{\tilde{P}_X \in \mathscr{B}_\eta(P_X)} \tilde{T}_X^+ \times T_{Y,\eta} \Bigg)$$
$$\ge P_X^n \Bigg(\bigcup_{\tilde{P}_X \in \mathscr{B}_\eta(P_X)} \tilde{T}_X^+ \Bigg) + P_Y^n(T_{Y,\eta}) - 1$$
$$\ge \frac{1}{2} - \lambda_n + 1 - \frac{|\mathscr{Y}|}{4n\eta^2} - 1,$$

where $\lambda_n \to 0$ since $\mathscr{B}_\eta(P_X)$ contains no degenerate distributions. We conclude that for n sufficiently large,

$$P_{XY}^n(\mathscr{A}_n^{(2)}) \ge 1 - \epsilon.$$

By computing the error exponent as before, we obtain for $1/2 < \epsilon < 1$,

$$\theta(2, N, \epsilon) \ge \max_{\substack{\Delta, \bar{\Delta}: \\ \Delta_X \cap \bar{\Delta}_X \cap 1_X = \emptyset}} \tau(\Delta, \bar{\Delta}).$$

This concludes the proof of the direct part.

Converse Part: As in the proof of the converse part of Theorem 6, we express $\mathscr{A}_n$ as

$$\mathscr{A}_n = (C_1 \times F_1) \cup (C_2 \times F_2),$$

where C_1 and C_2 form a partition of $\mathscr{X}^n$, and at most one of F_1, F_2 may be empty. Once again, two cases may arise.

Case 1: For i and j distinct, we have

$$(\forall P_{XY} \in \Pi) \qquad P_{XY}^n(C_i \times F_i) \ge P_{XY}^n(C_j \times F_j).$$

This implies that

$$-\frac{1}{n} \log Q^n(\mathscr{A}_n) \le \theta^{(1)} + \nu_n.$$

Case 2: The sets Δ and $\bar{\Delta}$ defined next form a nontrivial partition of Π:

$$\Delta = \{P_{XY} \in \Pi: P_{XY}^n(C_1 \times F_1) \geq P_{XY}^n(C_2 \times F_2)\},$$
$$\bar{\Delta} = \{P_{XY} \in \Pi: P_{XY}^n(C_1 \times F_1) < P_{XY}^n(C_2 \times F_2)\}.$$

We claim further that $\Delta_X \cap \bar{\Delta}_X \cap 1_X = \emptyset$. Indeed, if there exist $P_{XY} \in \Delta$ and $\tilde{P}_{XY} \in \bar{\Delta}$ such that $P_X = \tilde{P}_X$, then

$$P_X^n(C_1) \geq \frac{1-\epsilon}{2}, \qquad \tilde{P}_X^n(C_2) = P_X^n(C_2) > \frac{1-\epsilon}{2}.$$

Since C_1 and C_2 are complementary and have positive probability under P_X^n, P_X cannot be degenerate.

As in Case 2 in the proof of the converse part of Theorem 6, we obtain for all $\epsilon \in (0, 1)$,

$$-\frac{1}{n}\log Q^n(\mathscr{A}_n) \leq \tau(\Delta, \bar{\Delta}) + \nu_n.$$

It remains to show that if $\epsilon \in (0, 1/2)$, the above bound is also valid for a partition $\{\Omega, \bar{\Omega}\}$ of Π such that $\Omega_X \cap \bar{\Omega}_X = \emptyset$. To construct such a partition, we argue as follows.

For $P_X \in \Pi_X$, we consider the set $\mathscr{H}(P_X)$ of distributions in Π that have P_X as X-marginal:

$$\mathscr{H}(P_X) \overset{\text{def}}{=} \{\tilde{P}_{XY} \in \Pi: \tilde{P}_X = P_X\}.$$

We let $\lambda > 0$ be independent of n, and we assume for the moment that for every $P_X \in \Pi_X$, we can find $i \in \{1, 2\}$ such that

$$(\forall \tilde{P}_{XY} \in \mathscr{H}(P_X)) \qquad \tilde{P}_{XY}^n(C_i \times F_i) \geq \lambda. \tag{E.1}$$

If so, then we can partition Π_X into Λ_1 and Λ_2 by placing each of the members P_X of Π_X in Λ_i iff i is the smallest index for which the previous relationship holds. This, in turn, yields a partition Ω, $\bar{\Omega}$ of Π through

$$\Omega = \bigcup_{P_X \in \Lambda_1} \mathscr{H}(P_X) \quad \text{and} \quad \bar{\Omega} = \bigcup_{P_X \in \Lambda_2} \mathscr{H}(P_X).$$

Clearly $\Omega_X = \Lambda_1$, $\bar{\Omega}_X = \Lambda_2$, and from the definition of Λ_i and relationship (E.1), we obtain the desired bound

$$-\frac{1}{n}\log Q^n(\mathscr{A}_n) \leq \tau(\Omega, \bar{\Omega}) + \nu_n.$$

Thus, the issue is to prove that for suitable $\lambda > 0$, every $P_X \in \Pi_X$ is such that (E.1) holds for $i = 1$ or $i = 2$. By definition of the classes Δ and $\bar{\Delta}$, this is true for $P_X \in \Delta_X - \bar{\Delta}_X$ and $P_X \in \bar{\Delta}_X - \Delta_X$. To show that it is also true for $P_X \in \Delta_X \cap \bar{\Delta}_X$, assume the contrary, namely that there exists $P_{XY} \in \Delta$ and $\tilde{P}_{XY} \in \bar{\Delta}$ with $\tilde{P}_X = P_X$ and

$$P_{XY}^n(C_1 \times F_1) < \lambda, \qquad \tilde{P}_{XY}^n(C_2 \times F_2) < \lambda.$$

This implies that

$$P_X^n(C_1) \geq \tilde{P}_{XY}^n(C_1 \times F_1) > 1 - \epsilon - \lambda,$$
$$P_X^n(C_2) \geq P_{XY}^n(C_2 \times F_2) > 1 - \epsilon - \lambda,$$

and hence,

$$P_X^n(C_1) + P_X^n(C_2) > 2 - 2\epsilon - 2\lambda.$$

Thus, if $\epsilon < 1/2$, we can set $\lambda = (1 - 2\epsilon)/3 > 0$ to obtain the desired contradiction:

$$P_X^n(C_1) + P_X^n(C_2) > 1 + \lambda. \qquad \square$$

References

[1] T. Berger, "Decentralized estimation and decision theory," presented at the *IEEE Seven Springs Workshop Inform. Theory,* Mt. Kisco, NY, Sept. 1979.

[2] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inform. Theory,* vol. IT-32, pp. 533-542, July 1986.

[3] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inform. Theory,* vol. IT-33, pp. 759-772, Nov. 1987.

[4] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inform. Theory,* vol. 35, pp. 2-14, Jan. 1989.

[5] S. I. Amari and T. S. Han, "Statistical inference under multiterminal rate restrictions: A differential geometric approach," *IEEE Trans. Inform. Theory,* vol. 35, pp. 217-227, Mar. 1989.

[6] Z. Zhang and T. Berger, "Estimation via compressed information," *IEEE Trans. Inform. Theory,* vol. 34, pp. 198-211, Mar. 1988.

[7] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations," *Ann. Math. Statist.,* vol. 23, pp. 493-507, 1952.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1982.

[9] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerospace Electron. Syst.,* vol. AES-17, pp. 501-510, July 1981.

[10] D. Teneketzis and P. Varaiya, "The decentralized quickest decision problem," *IEEE Trans. Automat. Contr.,* vol. AC-29, pp. 641-644, July 1984.

[11] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Automat. Contr.,* vol. AC-30, pp. 440-446, May 1985.

[12] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace Electron. Syst.,* vol. AES-22, pp. 98-101, Jan. 1986.

[13] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerospace Electron. Syst.,* vol. AES-23, pp. 24-30, Jan. 1987.

[14] S. C. Thomopoulos, R. Viswanathan, and D. C. Bougoulias, "Optimal decision fusion in multiple sensor systems," *IEEE Trans. Aerospace and Electron. Syst.,* vol. AES-23, pp. 644-653, Sept. 1987.

[15] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Cont., Signals, and Syst.,* vol. 1,2, pp. 167-182, 1988.

[16] J. J. Chao and C. C. Lee, "Optimal local decision space partitioning for distributed detection," *IEEE Trans. Aerospace Electron. Syst.,* vol. 25, pp. 536-543, July 1989.

[17] E. Geraniotis and Y. A. Chau, "Robust data fusion for multi-sensor detection systems," *IEEE Trans. Inform. Theory,* vol. 36, pp. 1265-1279, Nov. 1990.