

Multiple hypothesis testing for arbitrarily varying sources

NAIRA M. GRIGORYAN AND ASHOT N. HARUTYUNYAN

Highly unstable environments can be modeled by arbitrarily varying (information) sources (AVS). We conduct a study of multiple hypothesis testing (HT) for those sources within two approaches existing in an information-theoretic area of statistical analysis. First we characterize the attainable exponent trade-offs for all kind of error probabilities and indicate the corresponding decision schemes or testing strategies. Then we treat the same problem from an optimality achieving perspectives.

Moreover, Chernoff bounds for both the binary and M -ary HT are specified via indication of a Sanov theorem for AVS's. Additional geometric interpretations help to digest the structure of HT in derived solutions.

KEYWORDS AND PHRASES: Arbitrarily varying sources, multiple hypothesis testing, error exponents, Chernoff bounds, Sanov theorem.

1. Introduction

Unstable systems and environments can be characterized by multiple probabilistic laws (distributions) depending on “states” they stay at rather than a single such law. Moreover, how those states commute with each other has no statistical description and is arbitrary. The corresponding model of information source, arbitrarily varying source (AVS), was introduced by Berger [2] in a source coding game context as a rate-distortion problem. An extended analysis of this problem in terms of error probability exponents can be found in [12].

This paper was presented in part [9] at IEEE Information Theory Workshop 2010.

Beyond natural phenomena which can be modeled by AVS's, examples of these kind of environments range from biometric verification systems (biometric templates are taken at arbitrary states of the enroller device or individual, see [13]) to cloud computing networks with state variability of virtual devices migrating over all the cloud landscape. Many problems of decision making in those transient environments relate to hypothesis testing (HT) and especially to multiple HT. That motivates the current study of multiple HT bounds within the classical framework of statistics and information theory initially started with the paper [9].

The problem of testing of hypotheses for AVS's is not new and is formulated by Fu and Shen [8] in binary case as extension of simpler discrete memoryless source (DMS) scenario. So the HT problem of [8] targets the identification of the distribution family of observations instead of predicting the generic distribution of a DMS without option to reject both the hypothetical laws about it. Furthermore, the work [1] considered the problem by Fu and Shen [8] in case of decision making with known to statistician states of the source. This side information simplifies the decision making.

The prior research on HT topic starts with the groundbreaking papers by Hoeffding [14], Csiszár and Longo [7], and Blahut [4]. The latter characterized the achievable error probability exponents in binary HT for DMS's and established also the strong tie between information theory and HT. Further advancement in the area covers the extension of binary HT to any M alternative hypotheses. Haroutunian's [10] generalization of HT for DMS was devoted to the maximum (best, optimal) error probability exponents for a collection of error kinds while having a "core" set of them within certain levels. This approach is known as log-asymptotic optimal (LAO) testing having roots in Birgé [3]. In particular, the work [10] specifies the LAO interdependencies between error exponents/reliabilities for all possible pairs of hypotheses. The same HT problem with an alternative setting is the subject of [16]. Being interested in attainable interconnections between all sort of error probabilities it specifies the corresponding achievable region of those tradeoffs, thus introducing a broader and unified description to the problem solution.

In this paper we extend the results by [10], [8], and [16], simultaneously, considering the multiple HT problem for AVS's targeting the specification of both the achievability region of error exponents and the LAO tradeoffs of those exponents. Moreover, we deepen the analysis to derive the Chernoff error bounds in HT for AVS's, thus extending the result by Leang and Johnson [15] for DMS's. For that purpose we prove a generalized Sanov theorem. In all cases the decision schemes supporting the error probability

bounds are constructed. We attach also geometric interpretations to our analysis.

The rest of the paper is organized as follows. Section 2 introduces the problem formally with necessary definitions. Section 3 is a summary on the employed in the paper methodological tool of empirical distributions in AVS context. In Section 4 the achievable reliability region we are looking for is characterized. The optimal error bounds within that region is the subject of Section 5. In Section 6 a Sanov theorem for AVS's is formulated and proved. In Section 7 the Chernoff bound for M -ary HT problem is presented and specialized to the binary case. Finally, Section 8 addresses geometric interpretations of HT mechanisms for AVS's.

2. Models of information source and HT

An AVS is a varying state-based DMS with finite alphabet \mathcal{X} . Those states alter within another finite set \mathcal{S} in a non-probabilistic and arbitrary manner. It means that an AVS $\{X, \mathcal{S}\}$ can be defined by a family of conditional probability distributions (PD) G_s^* (subset of all distributions $\mathcal{P}(\mathcal{X})$ on \mathcal{X}) depending on source state $s \in \mathcal{S}$:

$$(1) \quad \mathcal{G}^* \triangleq \{G_s^*, s \in \mathcal{S}\}$$

with $G_s^* \triangleq \{G^*(x|s), x \in \mathcal{X}\}$. From the condition of the state-based DMS, the probability of a vector $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$ emitted by the AVS while sequentially stayed at components of a vector $\mathbf{s} \in \mathcal{S}^N$ is determined by the following product

$$G^*(\mathbf{x}|\mathbf{s}) \triangleq \prod_{n=1}^N G^*(x_n|s_n).$$

The probability of a subset $\mathcal{A}_N \subset \mathcal{X}^N$ subject to a $\mathbf{s} \in \mathcal{S}^N$ is the sum

$$G^*(\mathcal{A}_N|\mathbf{s}) \triangleq \sum_{\mathbf{x} \in \mathcal{A}_N} G^*(\mathbf{x}|\mathbf{s}).$$

The multiple HT problem for this source admits $M > 2$ different hypotheses about the distribution (1):

$$(2) \quad H_m : \mathcal{G}^* = \mathcal{G}_m, \quad m = \overline{1, M}$$

with

$$(3) \quad \mathcal{G}_m \triangleq \{G_{m,s}, s \in \mathcal{S}\},$$

where

$$G_{m,s} \triangleq \{G_m(x|s), x \in \mathcal{X}\}, s \in \mathcal{S}.$$

Let G_m be the stochastic matrix defined by (3). Based on N observations of the source, namely on \mathbf{x} , one should make decision in favor of one (of M) hypotheses (2). A test φ_N by the statistician is an algorithm partitioning \mathcal{X}^N into M disjoint subsets \mathcal{A}_N^m , $m = \overline{1, M}$:

$$\mathcal{A}_N^m \cap \mathcal{A}_N^l = \emptyset, \quad m \neq l, \quad \bigcup_m \mathcal{A}_N^m = \mathcal{X}^N.$$

If $\mathbf{x} \in \mathcal{A}_N^m$ then the test adopts the hypothesis H_m . Inevitably, in decision making in favor of one of M alternatives he/she may commit $M(M-1)$ different kinds of errors, which are denoted by $\alpha_{l,m}(\varphi_N)$ – the probability (maximized over \mathcal{S}^N) of the erroneous acceptance by the test φ_N of the hypothesis H_l provided that H_m is true,

$$(4) \quad \alpha_{l,m}(\varphi_N) \triangleq \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathcal{A}_N^l | \mathbf{s}), \quad 1 \leq m \neq l \leq M.$$

Another type of error can be observed related to wrong decision in case of true H_m with the probability

$$(5) \quad \alpha_m(\varphi_N) \triangleq \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\overline{\mathcal{A}_N^m} | \mathbf{s}) \leq \sum_{l \neq m} \alpha_{l,m}(\varphi_N), \quad m = \overline{1, M}.$$

From the perspectives of study of achievable reliabilities in M -ary HT [16], consider the $M(M-1)$ -dimensional vector $\mathbf{E} \triangleq \{E_{l,m}\}_{l,m=\overline{1, M}, l \neq m}$, with each index (l, m) corresponding to error exponent $-\frac{1}{N} \log \alpha_{l,m}(\varphi_N)$.

Definition 1. \mathbf{E} is an achievable collection of error exponents (reliabilities) if for all $\varepsilon > 0$ there exists a decision scheme $\{\mathcal{A}_N^m\}_{m=1}^M$ or test φ_N with

$$-\frac{1}{N} \log \alpha_{l,m}(\varphi_N) > E_{l,m} - \varepsilon$$

for all $m \neq l$ and large enough N .

Let \mathcal{R}_{AVS} denotes the set of all achievable reliability vectors. One of our goals is to characterize this region.

Another goal is to look for optimal tests and trade-offs between error exponents under certain conditions. For that purpose we introduce the following definitions:

$$(6) \quad \bar{E}_{l,m} \triangleq \bar{E}_{l,m}(\varphi) \triangleq \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{l,m}(\varphi_N), \quad l, m = \overline{1, M}, l \neq m,$$

$$(7) \quad \bar{E}_m \triangleq \bar{E}_m(\varphi) \triangleq \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_m(\varphi_N), \quad m = \overline{1, M},$$

where $\varphi \triangleq \{\varphi_N\}_{N=1}^\infty$. Note that the equations (5), (6) and (7) imply

$$(8) \quad \bar{E}_m(\varphi) \leq \min_{l=\overline{1, M}, l \neq m} [\bar{E}_{l,m}(\varphi)], \quad m = \overline{1, M}.$$

Naturally, the asymptotic exponents in (6) and (7) are in cross trade-offs for any decision scheme. Therefore, the quality of a test scheme can be measured by the maximum exponents it provides for selected error kinds while keeping the rest of them within acceptable degrees. This concept of optimality is historically preceding (see Birge [3], Haroutunian [10]) the generic notion of achievability region for error exponent trade-offs by Tuncel [16]. In [10] the optimal test is defined by $M - 1$ fixed error exponents which indicate the “degrees of freedom” in the HT problem formulation. Therefore, in our case, the definition of a test optimality can be formulated in the following way.

Definition 2. A test φ^* is optimal if for every other test φ' achieving equal-degree errors for wrong decisions on every true hypotheses:

$$\bar{E}_m^* \triangleq \bar{E}_m^*(\varphi^*) = \bar{E}'_m \triangleq \bar{E}'_m(\varphi'), \quad m = \overline{1, M - 1},$$

it holds that all the pair-wise error exponents of φ^* are superior to those achieved by the other test:

$$\begin{aligned} \bar{E}_{l,m}^* \triangleq \bar{E}_{l,m}(\varphi^*) &\geq \bar{E}'_{l,m} \triangleq \bar{E}_{l,m}(\varphi'), \quad l, m = \overline{1, M}, l \neq m, \\ \bar{E}_M(\varphi^*) &\geq \bar{E}_M(\varphi'). \end{aligned}$$

So by looking for this kind of optimal tests and relevant error exponents they achieve, we are interested in special/optimal “trajectories” in

the achievability region \mathcal{R}_{AVS} that provide the best error trade-offs for pairs of hypotheses with the same degrees of overall error probabilities.

3. Empirical distributions in AVS context

The language of typical sequences is a convenient tool to apply in our analysis. In the sequel we'll use important properties of typical sequences (empirical distributions) tuned for the AVS context. In this section we make a short survey of some basic concepts and facts needed for further expansion of the subject.

Let

$$\mathcal{P}(\mathcal{S}) \triangleq \{P(s), s \in \mathcal{S}\}$$

be the collection of all distributions on \mathcal{S} and let PG be the marginal distribution on \mathcal{X} defined by

$$PG(x) \triangleq \sum_{s \in \mathcal{S}} P(s)G(x|s), x \in \mathcal{X}.$$

The type of the vector $\mathbf{s} \in \mathcal{S}^N$ is the empirical distribution defined by $P_{\mathbf{s}}(s) \triangleq \frac{1}{N}N(s|\mathbf{s})$, where $N(s|\mathbf{s})$ is the number of occurrences of s in \mathbf{s} . Let's denote the set of all types of N -size state vectors by $\mathcal{P}^N(\mathcal{S})$. For a pair of sequences $\mathbf{x} \in \mathcal{X}^N$ and $\mathbf{s} \in \mathcal{S}^N$ let $N(x, s|\mathbf{x}, \mathbf{s})$ be the number of occurrences of (x, s) in $\{x_n, s_n\}_{n=1}^N$. The conditional type $G_{\mathbf{x}, \mathbf{s}}$ of the vector \mathbf{x} with respect to the vector \mathbf{s} is defined by

$$(9) \quad G_{\mathbf{x}, \mathbf{s}}(x|s) \triangleq N(x, s|\mathbf{x}, \mathbf{s})/N(s|\mathbf{s}), \quad x \in \mathcal{X}, s \in \mathcal{S}.$$

The joint type of vectors \mathbf{x} and \mathbf{s} is the empirical distribution $P_{\mathbf{s}} \circ G_{\mathbf{x}, \mathbf{s}}$ defined by

$$P_{\mathbf{s}} \circ G_{\mathbf{x}, \mathbf{s}} = \{P_{\mathbf{s}}(s)G_{\mathbf{x}, \mathbf{s}}(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}.$$

For brevity we may omit the indices for type notations hereafter, however keep the unambiguity of formulations.

Let $\mathcal{G}^N(\mathcal{X}|\mathcal{S})$ be the set of all conditional types (9). Denote by $\mathcal{T}_G^N(\mathcal{X}|\mathbf{s})$ the set of vectors \mathbf{x} which have the conditional type G for given \mathbf{s} having type P . Let the conditional entropy of G given type P be $H(G|P) \triangleq - \sum_x P(s)G(x|s) \log G(x|s)$.

Denote by $D(G \| G_m|P)$ the KL divergence between G and G_m given type P and by $D(PG \| PG_m)$ the one between PG and PG_m .

For every $G_m \in \mathcal{G}_m$, $m = \overline{1, M}$, the following inequality holds

$$(10) \quad D(G \parallel G_m | P) \geq D(PG \parallel PG_m).$$

The upper bound for the cardinality of $\mathcal{G}^N(\mathcal{X}|\mathcal{S})$ is

$$(11) \quad |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| < (N+1)^{|\mathcal{X}||\mathcal{S}|}.$$

We need some properties for types. The size of type class $G \in \mathcal{G}^N(\mathcal{X}|\mathcal{S})$, where $P \in \mathcal{P}^N(\mathcal{S})$, has the following estimate

$$(12) \quad |\mathcal{T}_G^N(\mathcal{X}|\mathbf{s})| \leq \exp\{NH(G|P)\}.$$

For a distribution $G_m \in \mathcal{G}(\mathcal{X}|\mathcal{S})$ the sequence $\mathbf{x} \in \mathcal{T}_G^N(\mathcal{X}|\mathbf{s})$ has the probability

$$(13) \quad G_m^N(\mathbf{x}|\mathbf{s}) = \exp\{-N[H(G|P) + D(G \parallel G_m|P)]\}.$$

Note that (12) and (13) give the estimate for conditional type class probability

$$(14) \quad G_m^N(\mathcal{T}_G^N(\mathcal{X}|\mathbf{s})|\mathbf{s}) \geq (N+1)^{-|\mathcal{X}||\mathcal{S}|} \exp\{-ND(G \parallel G_m|P)\},$$

$$(15) \quad G_m^N(\mathcal{T}_G^N(\mathcal{X}|\mathbf{s})|\mathbf{s}) \leq \exp\{-ND(G \parallel G_m|P)\}.$$

4. Characterizing the achievable reliabilities

The achievability region characterization below extends the result by [16]. In other words, it describes the tradeoff between all error exponents for M -ary AVS HT problem. To demonstrate the result, we'll need the convex hull of the family of distributions \mathcal{G}_m for each $m = \overline{1, M}$,

$$(16) \quad \mathcal{W}_m \triangleq \left\{ W_m(x) \triangleq \sum_{s \in \mathcal{S}} \lambda_s G_{m,s}(x|s), x \in \mathcal{X}, 0 \leq \lambda_s \leq 1, \sum_{s \in \mathcal{S}} \lambda_s = 1 \right\}.$$

Then we introduce the region \mathcal{E}_{AVS} of $M(M-1)$ -dimensional vectors

$$(17) \quad \mathcal{E}_{\text{AVS}} \triangleq \left\{ \mathbf{E} : \forall W \exists l, \right. \\ \left. \text{such that } \min_{W_m \in \mathcal{W}_m} D(W \parallel W_m) > E_{l,m} \text{ for } \forall m \neq l \right\}.$$

We want to show in the next theorem that (17) completely characterizes the desired \mathcal{R}_{AVS} .

Theorem 1. \mathcal{E}_{AVS} is an achievable region of reliabilities: $\mathcal{E}_{AVS} \subset \mathcal{R}_{AVS}$. Conversely, if $\mathbf{E} \in \mathcal{R}_{AVS}$, then for any $\delta > 0$, $\mathbf{E}_\delta \in \mathcal{E}_{AVS}$, where $\mathbf{E}_\delta \triangleq \{E_{l,m} - \delta\}_{l \neq m}$.

Remark 1. Setting $|\mathcal{S}| = 1$ in Theorem 1 we'll readily get the region of achievable exponents in M -ary hypothesis testing for DMS [16]. Denote it by \mathcal{R}_{DMS} .

Theorem 2. $\mathcal{E}_{DMS} \subset \mathcal{R}_{DMS}$. Conversely, if $\mathbf{E} = \{E_{l,m}\} \in \mathcal{R}_{DMS}$, then for any $\delta > 0$, $\mathbf{E}_\delta \in \mathcal{E}_{DMS}$, where $\mathbf{E}_\delta \triangleq \{E_{l,m} - \delta\}$,

$$\mathcal{E}_{DMS} \triangleq \{\mathbf{E} : \forall W \exists l, \text{ such that } D(W|W_m) > E_{l,m} \text{ for } \forall m \neq l\}.$$

Remark 2. For $M = 2$ we have the binary AVS HT model by Fu and Shen [8] and the corresponding results. It can be easily seen by representing the region \mathcal{E}_{AVS} in the following way:

$$(18) \quad \mathcal{E}_{AVS}^* \triangleq \left\{ \mathbf{E} : \forall W \text{ either } \min_{W_1 \in \mathcal{W}_1} D(W|W_1) > E_{2,1}, \right. \\ \left. \text{or } \min_{W_2 \in \mathcal{W}_2} D(W|W_2) > E_{1,2} \right\}.$$

The alternative characterization of (18) is the following set

$$\mathcal{E}_{AVS}^* \triangleq \{\{E_{1,2}, E_{2,1}\} : E_{1,2} < E_{1,2}(E_{2,1})\},$$

where

$$E_{1,2}(E_{2,1}) = \min_{W_2 \in \mathcal{W}_2} \min_{W_1 \in \mathcal{W}_1} \min_{W: D(W||W_1) < E_{2,1}} D(W || W_2),$$

which is the binary AVS HT exponent function [8].

Proof. Let us prove the first part of Theorem 1. If $\mathbf{E} \in \mathcal{E}_{AVS}$, then from (12), (13) and (14), for any $\mathbf{s} \in \mathcal{S}^N$ with type $P_{\mathbf{s}} = P$ we have

$$(19) \quad G_m(\mathcal{A}_N^l | \mathbf{s}) = \sum_{\mathbf{x} \in \mathcal{A}_N^l} G_m(\mathbf{x} | \mathbf{s}) \\ \leq \sum_{\mathcal{T}_G^N(\mathcal{X} | \mathbf{s}) \subset \mathcal{A}_N^l} \exp\{-ND(G || G_m | P)\} \\ \leq |\mathcal{G}^N(\mathcal{X} | \mathcal{S})| \exp\{-ND(PG || PG_{m,s})\},$$

where G is a type $G \in \mathcal{G}^N(\mathcal{X} | \mathcal{S})$.

For every $W_m \in \mathcal{W}_m$ there exists $\mathbf{s} \in \mathcal{S}^N$, such that $W_m = P_{\mathbf{s}}G_m$. Therefore, from (19) and the polynomial order (11) of $\mathcal{G}^N(\mathcal{X}|\mathcal{S})$ we have

$$\begin{aligned}\alpha_{l,m}(\varphi_N) &\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-N \min_{W_m} D(W \| W_m)\} \\ &\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-N E_{l,m}\} \\ &\leq \exp\{-N(E_{l,m} - \delta)\}.\end{aligned}$$

This signals about a proven direct part of the theorem.

To proceed with the converse reasoning we should assume that a $\mathbf{E} \in \mathcal{R}_{\text{AVS}}$. This assures that for every $\varepsilon > 0$ there exists a decision scheme $\{\mathcal{A}_N^m\}_{m=1}^M$ that starting from an $N > N_0(\varepsilon)$ validates the inequality

$$(20) \quad -\frac{1}{N} \alpha_{l,m}(\varphi_N) > E_{l,m} - \varepsilon$$

for all $m \neq l$. For a picked $\delta > 0$ we need to show that

$$(21) \quad \forall W \exists l, \text{ s. t. } \min_{W_m \in \mathcal{W}_m} D(W|W_m) > E_{l,m} - \delta, \forall m \neq l.$$

To come to this statement we apply an estimate. For each $W_m \in \mathcal{W}_m$ and $\mathcal{A}_N^l \subseteq \mathcal{X}^N$ the following inequality holds for N large enough:

$$(22) \quad W_m(\mathcal{A}_N^l) \leq \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathcal{A}_N^l|\mathbf{s}) = \alpha_{l,m}(\varphi_N).$$

To show (22), first note that for any $W_m \in \mathcal{W}_m$ there exist a collection of λ_s , $0 \leq \lambda_s \leq 1$, $s \in \mathcal{S}$, with the property $\sum_{s \in \mathcal{S}} \lambda_s = 1$, such that $W_m =$

$\sum_{s \in \mathcal{S}} \lambda_s G_{m,s}$. Whence, for $\lambda_{\mathbf{s}} \triangleq \prod_{n=1}^N \lambda_{s_n}$ and any $\mathcal{A}_N^l \in \mathcal{X}^N$, $\mathbf{x} \in \mathcal{A}_N^l$, at which the test φ_N adopts the hypothesis H_l , the following can be performed:

$$\begin{aligned}W_m(\mathbf{x}) &= \prod_{n=1}^N W_m(x_n) = \prod_{n=1}^N \sum_{s \in \mathcal{S}} \lambda_s G_m(x_n|s) = \sum_{\mathbf{s} \in \mathcal{S}^N} \lambda_{\mathbf{s}} \prod_{n=1}^N G_m(x_n|s_n) \\ &\leq \max_{\mathbf{s} \in \mathcal{S}^N} \prod_{n=1}^N G_m(x_n|s_n) \leq \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathbf{x}|\mathbf{s}).\end{aligned}$$

Thus (22) follows.

Now turning to the proof for (21), by the continuity of $D(\cdot \| W_m)$ in the first argument, there exists a type $Q \in \mathcal{P}^N(\mathcal{X})$ that for $N > N_1(\varepsilon)$ satisfies

$$(23) \quad D(Q \| W_m) \leq D(W \| W_m) + \delta/2$$

for a fixed m .

Let $\overline{W}_{m^*} \triangleq \arg[\max_m \min_{W_m \in \mathcal{W}_m} D(W \| W_m)]$, then in light of (22) and (12) we have

$$\begin{aligned} \alpha_{m^*}(\varphi_N) &\geq \overline{W}_{m^*}(\overline{\mathcal{A}}_N^{m^*}) \\ &\geq \overline{W}_{m^*}(\overline{\mathcal{A}}_N^{m^*} \cap \mathcal{T}_Q^N(\mathcal{X})) \\ &= \sum_{\overline{\mathcal{A}}_N^{m^*} \cap \mathcal{T}_Q^N(\mathcal{X})} \exp\{-N[H(Q) + D(Q \| \overline{W}_{m^*})]\} \\ &\geq |\overline{\mathcal{A}}_N^{m^*} \cap \mathcal{T}_Q^N(\mathcal{X})| \exp\{-NH(Q)\} \exp\{-ND(Q \| \overline{W}_{m^*})\}, \end{aligned}$$

where Q is a type-approximation of W defined by (23) for $\overline{W}_{m^*}^N$. Note that $|\overline{\mathcal{A}}_N^{m^*} \cap \mathcal{T}_Q^N(\mathcal{X})| \exp\{-NH(Q)\} \geq \exp\{-N\delta/4\}$ for $N > N_2(\delta)$. It follows from the inequality $|\overline{\mathcal{A}}_N^{m^*} \cap \mathcal{T}_Q^N(\mathcal{X})| \geq \frac{|\mathcal{T}_Q^N(\mathcal{X})|}{M}$:

$$(24) \quad \begin{aligned} &|\overline{\mathcal{A}}_N^{m^*} \cap \mathcal{T}_Q^N(\mathcal{X})| \exp\{-NH(Q)\} \\ &\geq |\mathcal{T}_Q^N(\mathcal{X})| \exp\{-NH(Q)\} \exp\left\{-N \frac{\log M}{N}\right\} \\ &\geq \exp\{-N\delta/4\}. \end{aligned}$$

Whence, for $N > \max\{N_1(\delta), N_2(\delta)\}$ we have

$$\begin{aligned} \alpha_{m^*}(\varphi_N) &\geq \exp\{-N[D(Q \| \overline{W}_{m^*}) - \delta/4]\} \\ &\geq \exp\{-N[D(W \| \overline{W}_{m^*}) + \delta/4]\} \end{aligned}$$

that with (20) and $\varepsilon = 3\delta/4$ gives

$$E_{m^*} - \delta < -\frac{1}{N} \log \alpha_{m^*}(\varphi_N) < D(W \| \overline{W}_{m^*})$$

for $N > \max\{N_0(\varepsilon), N_1(\delta), N_2(\delta)\}$ and for every W . □

5. Error exponents in optimal tradeoffs

Now we proceed with the goal of characterization of optimal tradeoffs for error probabilities in terms of Definition 2 for fixed exponents \overline{E}_m , $m =$

$\overline{1, M-1}$. First we introduce the following sets:

$$(25) \quad \mathcal{B}_m(\mathcal{W}_m) \triangleq \left\{ W : \min_{W_m \in \mathcal{W}_m} D(W \parallel W_m) < \overline{E}_m \right\}, \quad m = \overline{1, M-1},$$

$$(26) \quad \mathcal{B}_M(\mathcal{W}_M) \triangleq \left\{ W : \min_{W_m \in \mathcal{W}_m} D(W \parallel W_m) > \overline{E}_m, \quad m = \overline{1, M-1} \right\},$$

and also define

$$(27) \quad \overline{E}_{l,m}^* \triangleq \min_{W \in \mathcal{B}_l(\mathcal{W}_l)} \min_{W_m \in \mathcal{W}_m} D(W \parallel W_m), \\ l = \overline{1, M-1}, \quad m = \overline{1, M}, \quad l \neq m,$$

$$(28) \quad \overline{E}_{M,m}^* \triangleq \min_{W \in \mathcal{B}_M(\mathcal{W}_M)} \min_{W_m \in \mathcal{W}_m} D(W \parallel W_m), \quad m = \overline{1, M-1},$$

$$(29) \quad \overline{E}_M^* \triangleq \min_{W \in \mathcal{B}_M(\mathcal{W}_M)} \min_{W_M \in \mathcal{W}_M} D(W \parallel W_M).$$

Theorem 3. *The optimal test φ^* is characterized by the reliabilities defined in (27)–(29) with the following restrictions for \overline{E}_m^* , $m = \overline{1, M-1}$:*

$$(30) \quad \overline{E}_1^* < \min_{m=2, M} \left\{ \min_{\substack{W_1 \in \mathcal{W}_1, \\ W_m \in \mathcal{W}_m}} D(W_m \parallel W_1) \right\},$$

$$(31) \quad \overline{E}_m^* < \min_{l \neq m} \left\{ \min_{l=1, m-1} \overline{E}_{l,m}^*, \min_{l=m+1, M} \min_{\substack{W_l \in \mathcal{W}_l, \\ W_m \in \mathcal{W}_m}} D(W_l \parallel W_m) \right\}, \\ m = \overline{2, M-1}.$$

Remark 3. Setting $|\mathcal{S}| = 1$ in Theorem 3 we'll readily get the optimal exponents in M -ary hypothesis testing for DMS [10, 11].

Observe that the sets (25)–(26) specialize to

$$\mathcal{B}_m(W_m) \triangleq \{W : D(W \parallel W_m) < \overline{E}_m\}, \quad m = \overline{1, M-1}, \\ \mathcal{B}_M(W_M) \triangleq \{W : D(W \parallel W_m) > \overline{E}_m, \quad m = \overline{1, M-1}\},$$

and the definitions (27)–(29) specialize to

$$(32) \quad \overline{E}_{l,m}^* \triangleq \min_{W \in \mathcal{B}_l(W_l)} D(W \parallel W_m), \quad l = \overline{1, M-1}, \quad m = \overline{1, M}, \quad l \neq m,$$

$$(33) \quad \overline{E}_{M,m}^* \triangleq \min_{W \in \mathcal{B}_M(W_M)} D(W \parallel W_m), \quad m = \overline{1, M-1},$$

$$(34) \quad \overline{E}_M^* \triangleq \min_{W \in \mathcal{B}_M(W_M)} D(W \parallel W_m).$$

Theorem 4. *The optimal test φ^* is characterized by the reliabilities defined in (32)–(34) with the following restrictions for \bar{E}_m^* , $m = \overline{1, M-1}$:*

$$(35) \quad \bar{E}_1^* < \min_{m=2, \overline{M}} \{D(W_m \parallel W_1)\},$$

$$(36) \quad \bar{E}_m^* < \min_{l \neq m} \left\{ \min_{l=1, m-1} \bar{E}_{l, m}^*, \min_{l=m+1, \overline{M}} D(W_l \parallel W_m) \right\}, \quad m = \overline{2, M-1}.$$

Proof. (of Theorem 3). Let the decision on m be made based on the partition:

$$\mathcal{D}_m \triangleq \bigcup_{W \in \mathcal{B}_m(\mathcal{W}_m)} \mathcal{T}_W^N(\mathcal{X})$$

for $m = \overline{1, M}$. Note that $\mathcal{D}_m \cap \mathcal{D}_l \neq \emptyset$ and $\mathcal{D}_m \cap \mathcal{D}_M \neq \emptyset$. We are going to show that this is a partition that enables the optimal test sequence $\varphi^* = \{\varphi_N^*\}_{N=1}^\infty$.

For $\bar{W}_m \triangleq \arg \min_{W_m \in \mathcal{W}_m} D(W \parallel W_m)$, $m = \overline{1, M}$, and φ^* we perform the following analysis applying the unconditional version of (14):

$$\begin{aligned} \alpha_{M, m}(\varphi_N^*) &\geq \bar{W}_m(\mathcal{D}_M) \\ &\geq \bar{W}_m \left(\bigcup_{W \in \mathcal{B}_M(\mathcal{W}_M)} \mathcal{T}_W^N(\mathcal{X}) \right) \\ &\geq \max_{W \in \mathcal{B}_M(\mathcal{W}_M)} \exp\{-N[D(W \parallel \bar{W}_m) + o_N(1)]\} \\ &= \exp \left\{ -N \left[\min_{W \in \mathcal{B}_M(\mathcal{W}_M)} D(W \parallel \bar{W}_m) + o_N(1) \right] \right\}. \end{aligned}$$

In a similar way we can obtain the inequalities

$$(37) \quad \alpha_{l, m}(\varphi_N^*) \geq \exp \left\{ -N \left[\min_{W_m \in \mathcal{W}_m} \min_{W \in \mathcal{B}_l(\mathcal{W}_l)} D(W \parallel W_m) + o_N(1) \right] \right\},$$

$$\begin{aligned} (38) \quad \alpha_M(\varphi_N^*) &\geq \bar{W}_M(\bar{\mathcal{D}}_M) \\ &\geq \bar{W}_M \left(\bigcup_{W \in \bar{\mathcal{B}}_M(\mathcal{W}_M)} \mathcal{T}_W^N(\mathcal{X}) \right) \\ &\geq \max_{W \in \bar{\mathcal{B}}_M(\mathcal{W}_M)} \exp\{-N[D(W \parallel \bar{W}_M) + o_N(1)]\} \\ &= \exp \left\{ -N \left[\min_{W \in \bar{\mathcal{B}}_M(\mathcal{W}_M)} D(W \parallel \bar{W}_M) + o_N(1) \right] \right\}. \end{aligned}$$

The proof of the converse inequalities

$$(39) \quad \alpha_{l,m}(\varphi_N^*) \leq \exp \left\{ -N \left[\min_{W_m \in \mathcal{W}_m} \min_{W \in \mathcal{B}_l(\mathcal{W}_l)} D(W \parallel W_m) + o_N(1) \right] \right\}$$

$$(40) \quad \alpha_M(\varphi_N^*) \leq \exp \left\{ -N \left[\min_{W_M \in \mathcal{W}_M} \min_{W \in \bar{\mathcal{B}}_M(\mathcal{W}_M)} D(W \parallel W_M) + o_N(1) \right] \right\}$$

are presented in the sequel.

From (10), (12), (13) and (15) and $\mathbf{s} \in \mathcal{S}^N$ with type $P_{\mathbf{s}} = P$ we have

$$(41) \quad \begin{aligned} G_m(\mathcal{B}_l(\mathcal{W}_l)|\mathbf{s}) &= \sum_{\mathbf{x} \in \mathcal{B}_l(\mathcal{W}_l)} G_m(\mathbf{x}|\mathbf{s}) \\ &\leq \sum_{\mathcal{T}_G(\mathcal{X}|\mathbf{s}) \subset \mathcal{B}_l} \exp\{-ND(G \parallel G_m|P)\} \\ &\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-ND(PG \parallel PG_m)\} \end{aligned}$$

$$(42) \quad \begin{aligned} G_M(\bar{\mathcal{B}}_M(\mathcal{W}_M)|\mathbf{s}) &= \sum_{\mathbf{x} \in \bar{\mathcal{B}}_M(\mathcal{W}_M)} G_M(\mathbf{x}|\mathbf{s}) \\ &\leq \sum_{\mathcal{T}_G(\mathcal{X}|\mathbf{s}) \subset \bar{\mathcal{B}}_M} \exp\{-ND(G \parallel G_M|P)\} \\ &\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-ND(PG \parallel PG_M)\} \end{aligned}$$

where $G \in \mathcal{G}^N(\mathcal{X}|\mathcal{S})$ is a type.

For every $W_m \in \mathcal{W}_m$ there exists $\mathbf{s} \in \mathcal{S}^N$, such that $W_m = P_{\mathbf{s}}G_m$. Hence, from (41) for $l \neq m, l, m = \overline{1, M}$, we come to

$$\alpha_{l,m}(\varphi_N^*) \leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp \left\{ -N \min_{W_m \in \mathcal{W}_m} \min_{W \in \mathcal{B}_l(\mathcal{W}_l)} D(W \parallel W_m) \right\},$$

and from (42) we get

$$\alpha_M(\varphi_N^*) \leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp \left\{ -N \min_{W_M \in \mathcal{W}_M} \min_{W \in \bar{\mathcal{B}}_l(\mathcal{W}_l)} D(W \parallel W_M) \right\}.$$

Taking into account (37), (39), (38), (40) and the continuity of the functional $D(W \parallel W_m)$ we obtain that the limit $\lim_{N \rightarrow \infty} \{\sup -N^{-1} \log \alpha_{l,m}(\varphi_N^*)\}$ exists and equals to $\bar{E}_{l,m}^*$. The same continuity argument applies to the exponent of $\alpha_M(\varphi_N^*)$.

To accomplish the proof by contradictory assumption let φ' be a test defined by the sets (\mathcal{D}'_m) such that

$$\overline{E}'_{l,m} > \overline{E}^*_{l,m}, \quad l \neq m = \overline{1, M}. \quad \overline{E}'_M > \overline{E}^*_M.$$

It yields for N large enough that

$$\alpha_{l,m}(\varphi'_N) < \alpha_{l,m}(\varphi^*_N), \quad \alpha_{M,m}(\varphi'_N) < \alpha_{M,m}(\varphi^*_N).$$

Below we examine the relation between $(\mathcal{D}_m, \mathcal{D}_M)$ and $(\mathcal{D}'_m, \mathcal{D}'_M)$. Four cases are possible:

- 1) $\mathcal{D}_m \cap \mathcal{D}'_m = \emptyset$,
- 2) $\mathcal{D}_m \subset \mathcal{D}'_m$,
- 3) $\mathcal{D}'_m \subset \mathcal{D}_m$,
- 4) $\mathcal{D}_m \cap \mathcal{D}'_m \neq \emptyset$.

The same cases exist also for \mathcal{D}_M and \mathcal{D}'_M . Consider the case of $\mathcal{D}_m \cap \mathcal{D}'_m = \emptyset$. It follows that there exists $l \neq m$ such that $\mathcal{D}_m \cap \mathcal{D}'_l \neq \emptyset$. That is there exists a W such that $D(W \parallel \overline{W}_m) < \overline{E}'_m$, so $\mathcal{T}_W^N(\mathcal{X}) \subset \mathcal{D}'_l$. Compute

$$\begin{aligned} \alpha_{l,m}(\varphi'_N) &= \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathcal{D}'_l | \mathbf{s}) \\ &\geq \overline{W}_m^N(\mathcal{T}_W^N(\mathcal{X})) \\ &\geq \exp\{-N[D(W \parallel \overline{W}_m) + o_N(1)]\} \\ &> \exp\{-N[\overline{E}'_m + o_N(1)]\}. \end{aligned}$$

Thus $\overline{E}'_{l,m} < \overline{E}'_m = \overline{E}^*_m$ which contradicts to (8). □

6. Generalized sanov theorem

Here we are interested in large deviations framework for AVS's. Let a family of conditional probabilities \mathcal{G} be defined according to (3) and let $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ be a subset of distributions on \mathcal{X} . As in (16) we define and denote by \mathcal{W} the convex hull of \mathcal{G} . Note that

$$G(\mathcal{C} | \mathbf{s}) \triangleq G(\mathcal{C} \cap \mathcal{P}^N(\mathcal{X}) | \mathbf{s}) \triangleq \sum_{\mathbf{x}: G_{\mathbf{x}} \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} G(\mathbf{x} | \mathbf{s}).$$

Let us define the following set for each \mathbf{s}

$$\mathcal{F}_{\mathbf{s}} \triangleq \{G_{\mathbf{x},\mathbf{s}} \in \mathcal{P}^N(\mathcal{X}|\mathcal{S}) : \text{ such that } G_{\mathbf{x}} \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})\}.$$

AVS version of Sanov's theorem states:

Theorem 5. *For given set \mathcal{C} ,*

$$(43) \quad \max_{\mathbf{s}} G(\mathcal{C}|\mathbf{s}) \leq (N + 1)^{|\mathcal{X}||\mathcal{S}|} \exp \left\{ -N \min_{W \in \mathcal{W}} D(V^* \| W) \right\},$$

where $V^* \triangleq \arg \min_{V \in \mathcal{C}} D(V \| W)$. If, in addition, the set \mathcal{C} is the closure of its interior, then

$$(44) \quad \max_{\mathbf{s}} \frac{1}{N} \log G(\mathcal{C}|\mathbf{s}) \rightarrow - \min_{W \in \mathcal{W}} D(V^* \| W).$$

With $|\mathcal{S}| = 1$ we get Sanov's theorem for DMS. It is formulated in the next claim.

Theorem 6. *For given set $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$,*

$$G(\mathcal{C} \cap \mathcal{P}^N(\mathcal{X})) \leq (N + 1)^{|\mathcal{X}|} \exp\{-ND(G^* \| G)\},$$

where $G^* \triangleq \arg \min_{G' \in \mathcal{C}} D(G' \| G)$. If, in addition, \mathcal{C} is the closure of its interior, then $\frac{1}{N} \log G(\mathcal{C}) \rightarrow -D(G^* \| G)$.

Proof. We first prove the upper bound (43). For every $\mathbf{s} \in \mathcal{S}^N$ we have the following chain of estimates which is enough to be sure in (43).

$$\begin{aligned} G(\mathcal{C}|\mathbf{s}) &= G(\mathcal{C} \cap \mathcal{P}^N(\mathcal{X})|\mathbf{s}) \\ &= \sum_{\mathbf{x}: G_{\mathbf{x}} \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} G(\mathbf{x}|\mathbf{s}) \\ &= \sum_{\mathbf{x}: G_{\mathbf{x}} \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} \exp\{-N[H(G_{\mathbf{x},\mathbf{s}}|P_{\mathbf{s}}) + D(G_{\mathbf{x},\mathbf{s}} \| G|P_{\mathbf{s}})]\} \\ &= \sum_{G_{\mathbf{x},\mathbf{s}} \in \mathcal{F}_{\mathbf{s}}} \sum_{\mathbf{x} \in T_{G_{\mathbf{x},\mathbf{s}}}(X|\mathbf{s})} \exp\{-N[H(G_{\mathbf{x},\mathbf{s}}|P_{\mathbf{s}}) + D(G_{\mathbf{x},\mathbf{s}} \| G|P_{\mathbf{s}})]\} \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \exp \left\{ -N \left[\min_{G_{\mathbf{x},s} \in \mathcal{F}_s} D(G_{\mathbf{x},s} \parallel G|P_s) - \epsilon \right] \right\} \\
&\stackrel{(b)}{\leq} \exp \left\{ -N \left[\min_{G_{\mathbf{x},s} \in \mathcal{F}_s} D(P_s G_{\mathbf{x},s} \parallel P_s G) - \epsilon \right] \right\} \\
&\stackrel{(c)}{=} \exp \{ -N [D(G_{\mathbf{x}}^* \parallel W) - \epsilon] \} \\
&\stackrel{(d)}{\leq} \exp \left\{ -N \left[\min_{V \in \mathcal{C}} D(V \parallel W) - \epsilon \right] \right\} \\
&= \exp \{ -N [D(V^* \parallel W) - \epsilon] \} \\
&\leq \exp \left\{ -N \left[\min_{W \in \mathcal{W}} D(V^* \parallel W) - \epsilon \right] \right\},
\end{aligned}$$

where

- (a) follows from the fact that there are only polynomially many types in $\mathcal{P}^N(\mathcal{X})$ and from (12),
- (b) follows from the property for divergence (10),
- (c) $G_{\mathbf{x}}^* \triangleq P_s G_{\mathbf{x},s}^*$, $G_{\mathbf{x},s}^* \triangleq \arg \min_{G_{\mathbf{x},s} \in \mathcal{F}_s} D(P_s G_{\mathbf{x},s} \parallel P_s G)$,
- (d) follows from the argument that $G_{\mathbf{x}}^* \in \mathcal{C}$.

Now we turn to the lower bound. Assume that \mathcal{C} is the closure of its interior. Since $\mathcal{P}^N(\mathcal{X})$ is dense in the set of all distributions then $\mathcal{C} \cap \mathcal{P}^N(\mathcal{X})$ is non-empty for all $N > N_0$, for some N_0 . Furthermore, the assumption implies that for every $W \in \mathcal{W}$ we can find a sequence of distributions Z_n such that $Z_n \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})$ and $D(Z_n \parallel W) \rightarrow D(V^* \parallel W)$.

To show the limit (44) we make the following estimate for any $W \in \mathcal{W}$ employing the inequality (22):

$$\begin{aligned}
\max_{\mathbf{s} \in \mathcal{S}^N} G(\mathcal{C}|\mathbf{s}) &\geq W(\mathcal{C}) = \sum_{Z \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} W(\mathcal{T}_V^N(\mathcal{X})) \\
&\geq \sum_{Z_n \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} W(\mathcal{T}_{Z_n}^N(\mathcal{X})) \\
&\geq (N+1)^{|\mathcal{X}|} \max_{Z_n \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} W(\mathcal{T}_{Z_n}^N(\mathcal{X})) \\
&\stackrel{(a)}{\geq} \exp \left\{ -N \min_{Z_n \in \mathcal{C} \cap \mathcal{P}^N(\mathcal{X})} D(Z_n \parallel W) \right\},
\end{aligned}$$

where (a) follows from the unconditional version of (14). Consequently,

$$\begin{aligned} & \liminf \frac{1}{N} \log \max_{\mathbf{s} \in \mathcal{S}^N} G(\mathcal{C}|\mathbf{s}) \\ & \geq \liminf \left(-\frac{1}{N} - \min_{W \in \mathcal{W}} D(Z_n \| W) \right) \\ & = - \min_{W \in \mathcal{W}} D(V^* \| W). \end{aligned}$$

Combining this with the upper bound establishes the theorem. \square

7. Chernoff error bounds

Here the idea is to minimize the overall probability of error given by the weighted sum of the individual probabilities of error. Those weights are obtained via Bayesian inference. Therefore, in our setup we can assume that there are M hypotheses

$$H_m : \mathcal{G}_m = \{G_m(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}, \quad m = \overline{1, M},$$

with associated prior probabilities π_m . Denoting the optimal decision regions by \mathcal{A}_m and their complements by \mathcal{A}_m^c , $m = \overline{1, M}$, Theorem 6 yields

$$\max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathcal{A}_m^c|\mathbf{s}) \leq \exp \left\{ -N \left[\min_{W_m \in \mathcal{W}_m} D(V_m^* \| W_m) - \epsilon \right] \right\}$$

for the error probability under each hypothesis, where

$$V_m^* \triangleq \arg \min_{V \in \mathcal{A}_m^c} D(V \| W_m).$$

The overall probability of error is

$$P_e(\varphi_N) \triangleq \sum_{m=1}^M \pi_m \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathcal{A}_m^c|\mathbf{s}).$$

Let

$$D_{\text{AVS}}^*(M) = \lim_{N \rightarrow \infty} \min_{\varphi_N} -\frac{1}{N} \log P_e(\varphi_N).$$

Theorem 7. $D_{AVS}^*(M) = C_{AVS}(\mathcal{W}_1, \dots, \mathcal{W}_M)$, where $C_{AVS}(\mathcal{W}_1, \dots, \mathcal{W}_M)$ is called the Chernoff distance for AVS and defined as

$$C_{AVS}(\mathcal{W}_1, \dots, \mathcal{W}_M) \triangleq \min_m \left\{ \min_{W_m \in \mathcal{W}_m} D(V_m^* \parallel W_m) \right\}.$$

Proof.

$$\begin{aligned} P_e(\varphi_N) &= \sum_{m=1}^M \pi_m \max_{\mathbf{s} \in \mathcal{S}^N} G_m(\mathcal{A}_m^c | \mathbf{s}) \\ &\leq \sum_{m=1}^M \pi_m \exp \left\{ -N \left[\min_{W_m \in \mathcal{W}_m} D(V_m^* \parallel W_m) - \epsilon \right] \right\} \\ &\leq \exp \left\{ -N \left[\min_m \min_{W_m \in \mathcal{W}_m} D(V_m^* \parallel W_m) - \epsilon \right] \right\} \\ &= \exp \left\{ -N [C_{AVS}(\mathcal{W}_1, \dots, \mathcal{W}_M) - \epsilon] \right\}. \end{aligned}$$

□

Remark 4. For $|\mathcal{S}| = 1$ we get the result by Leang and Johnson [15] for the Chernoff bound on DMS setting of HT.

For the binary case of $M = 2$, Theorem 7 specializes to the following lemma.

Lemma 8. *The Chernoff information between two families of distributions can be specified by the following representation:*

$$\begin{aligned} C_{AVS}(\mathcal{W}_1, \mathcal{W}_2) &= \min \left\{ \min_{W_1 \in \mathcal{W}_1} D(V_1^* \parallel W_1), \min_{W_2 \in \mathcal{W}_2} D(V_2^* \parallel W_2) \right\} \\ &= \min \{ D(V_1^* \parallel W_1^*), D(V_2^* \parallel W_2^*) \} \\ &= C_{DMS}(W_1^*, W_2^*), \end{aligned}$$

where $W_i^* = \arg \min_{W_i \in \mathcal{W}_i} D(V_i^* \parallel W_i)$, $i = 1, 2$, and $C_{DMS}(W_1^*, W_2^*)$ is the Chernoff information for distributions W_1^* and W_2^* which according to [6] [Chapter 12] satisfy the equalities

$$D(V_1^* \parallel W_1^*) = D(V_2^* \parallel W_2^*), \quad V_1^* = V_2^* = V^*.$$

With $|\mathcal{S}| = 1$ the lemma easily specifies the binary Chernoff bound D^* for DMS that according to [6] is determined by the following theorem.

Theorem 9.

$$D^* = D(W_{\lambda^*} \parallel W_1) = D(W_{\lambda^*} \parallel W_2).$$

where

$$W_\lambda = \frac{W_1^\lambda(x)W_2(x)^{1-\lambda}}{\sum_{a \in \mathcal{X}} W_1^\lambda(a)W_2(a)^{1-\lambda}}$$

and λ^* is the value of λ such that $D(W_\lambda^* \parallel W_1) = D(W_\lambda^* \parallel W_2)$.

8. Geometry of decision schemes and bounds

Fig. 1 presents a geometric interpretation for the decision scheme in Theorem 3. Fig. 2 and 3 illustrate the geometry of the Chernoff bounds derived in Section 7 for binary and M -ary HT, respectively, (c.f. [5] for DMS's).

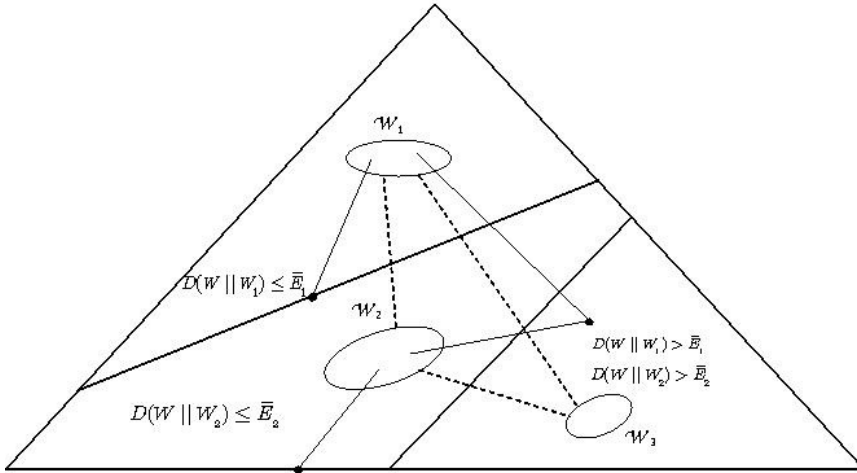


Figure 1: Multiple HT for AVS. Dash-lines represent distances

$$\min_{\substack{W_l \in \mathcal{W}_l, \\ W_m \in \mathcal{W}_m}} D(W_m \parallel W_l)$$

between convex hulls $(\mathcal{W}_l, \mathcal{W}_m)$, where $l, m = \bar{1}, \bar{3}$, $l \neq m$.

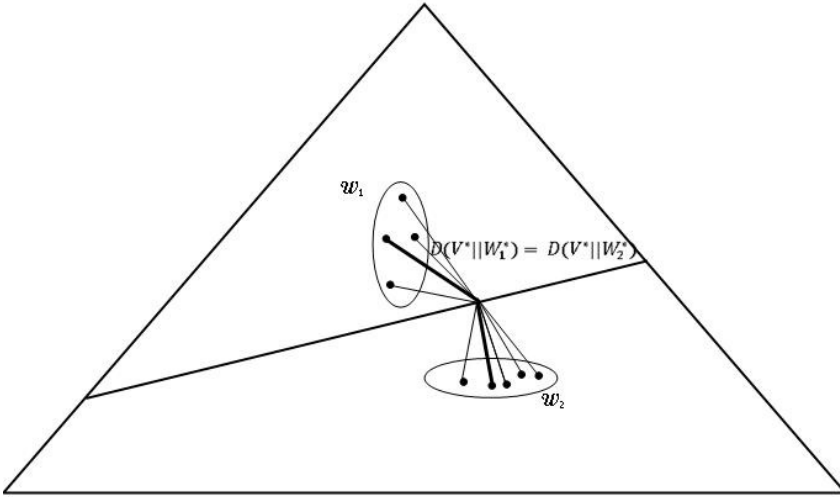


Figure 2: Chernoff bound for AVS in binary HT.

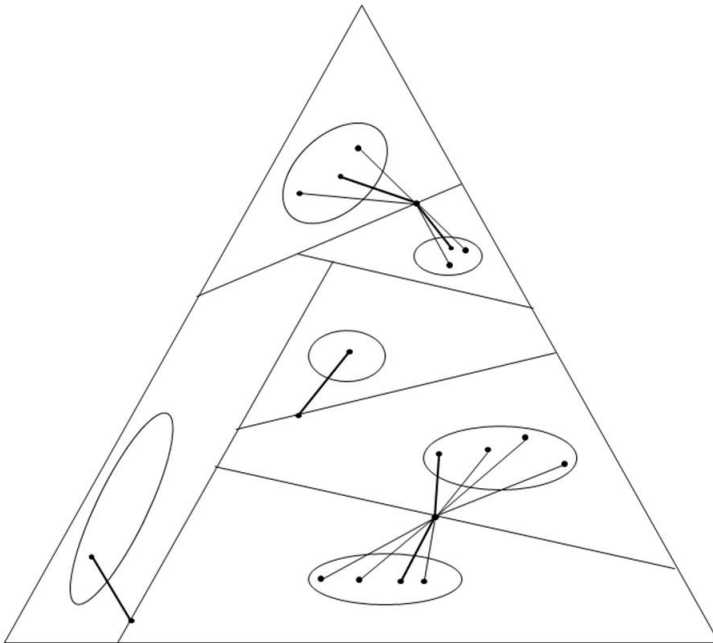


Figure 3: Chernoff bound for AVS in multiple HT.

Acknowledgement

Authors are thankful to the anonymous reviewers for highly useful comments which helped to substantially improve the paper.

References

- [1] R. Ahlswede, E. Aloyan and E. Haroutunian, *On logarithmically asymptotically optimal hypothesis testing for arbitrarily varying sources with side information*. Lecture Notes in Computer Science, 4123, pp. 547–552, Springer Verlag, 2006.
- [2] T. Berger, *The source coding game*. IEEE Trans. Inform. Theory, vol. IT-17, no. 1, pp. 71–76, January 1971.
- [3] L. Birgé, *Vitesse maximale de décroissance des erreurs et tests optimaux associés*. Z. Wahrsch. Verw. Gebiete, vol. 55, pp. 261–173, 1981.
- [4] R. E. Blahut, *Hypothesis testing and information theory*. IEEE Trans. Inform. Theory, Vol. IT-20, no. 4, pp. 405–417, 1974.
- [5] M. Brandon Westover, *Asymptotic geometry of multiple hypothesis testing*. IEEE Trans. Inform. Theory, vol. 54, no. 7, pp. 3327–3329, 2008.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, Wiley, 1991.
- [7] I. Csiszár and G. Longo, *On the error exponent for source coding and for testing simple statistical hypotheses*. Studia Scientiarum Mathem. Hung., vol. 6, pp. 181–191, 1971.
- [8] F.-W. Fu and S.-Y. Shen, *Hypothesis testing for arbitrarily varying source with exponential-type constraint*. IEEE Trans. Inform. Theory, vol. 44, no. 2, pp. 892–895, 1998.
- [9] N. M. Grigoryan and A. N. Harutyunyan, *Error exponents in multiple hypothesis testing for arbitrarily varying sources*. Proc. IEEE Workshop Inform. Theory, Dublin, Ireland, Aug. 30 – Sept. 3, 5 p., 2010.
- [10] E. A. Haroutunian, *Logarithmically asymptotically optimal testing of multiple statistical hypotheses*. Problems of Control and Information Theory, vol. 19, no. 5–6, pp. 413–421, 1990.

- [11] E. A. Haroutunian, M. E. Haroutunian and A. N. Harutyunyan, *Reliability Criteria in Information Theory and in Statistical Hypotheses Testing*. Foundation and Trends in Communications and Information Theory, vol. 4, no. 2–3, pp. 97–263, 2008.
- [12] A. N. Harutyunyan and A. J. Han Vinck, *Error exponent in AVS coding*. Proc. IEEE Intern. Symp. Inform. Theory, pp. 2166–2170, Seattle, WA, July 9–14, 2006.
- [13] A. Harutyunyan, N. Grigoryan, S. Voloshynovsky and O. Koval, *A new biometric identification model and the multiple hypothesis testing for arbitrarily varying objects*. Proc. Int. Conf. Biometric Special Interest Group (BIOSIG), Lecture Notes in Informatics (LNI), vol. P-191, Darmstadt, Germany, Sept. 8–9, pp. 305–312, 2011.
- [14] W. Hoeffding, *Asymptotically optimal tests for multinomial distributions*. Ann. Math. Statist., vol. 36, pp. 369–401, 1965.
- [15] C. Leang and D. H. Johnson, *On the asymptotics of M-hypothesis bayesian detection*. IEEE Trans. Inform. Theory, vol. 43, no. 1, pp. 280–282, 1997.
- [16] E. Tuncel, *On error exponents in hypothesis testing*. IEEE Trans. Inform. Theory, vol. 51, no. 8, pp. 2945–2950, 2005.

VMWARE, INC.

E-mail address: ngrigoryan@vmware.com

VMWARE, INC.

E-mail address: aharutyunyan@vmware.com

RECEIVED NOVEMBER 4, 2013

ACCEPTED MAY 19, 2015