On Multiple Hypothesis Testing with Rejection Option

Grigoryan, N.; Harutyunyan, A.; Voloshynovskyy, Svyatoslav; Koval, Oleksiy

On Multiple Hypothesis Testing with Rejection Option

Naira Grigoryan[*] IIAP NAS RA Email: nar.gri@gmail.com

Ashot Harutyunyan[*] IIAP NAS RA Email: ashot@iem.uni-due.de

Svyatoslav Voloshynovskiy University of Geneva Email: svolos@unige.ch

Oleksiy Koval University of Geneva Email: oleksiy.koval@unige.ch

***Abstract*—We study the problem of multiple hypothesis testing (HT) in view of a rejection option. That model of HT has many different applications. Errors in testing of M hypotheses regarding the source distribution with an option of rejecting all those hypotheses are considered. The source is discrete and arbitrarily varying (AVS). The tradeoffs among error probability exponents/reliabilities associated with *false acceptance of rejection decision* and *false rejection of true distribution* are investigated, the optimal decision strategies are outlined. The special case of discrete memoryless source (DMS) is also discussed. An interesting insight that the analysis implies is the phenomenon (comprehensible in terms of supervised/unsupervised learning) that in optimal discrimination within M hypothetical distributions one permits always lower error than in deciding to decline the set of hypotheses. Geometric interpretations of the optimal decision schemes and bounds in multi-HT for AVS's are given.**

I. Introduction

Recent impetuous progress in computer and public network infrastructure as well as in multimedia data manipulating software created an unprecedented yet often uncontrolled possibilities for multimedia content modification and redistribution over various public services and networks including Flickr and YouTube. Since in multiple cases these actions concern privacy sensitive data, a significant research effort was made targeting efficient means of their identification as well as related performance analysis [11], [12], [17]. While early reported results [13] were mostly dedicated to the capacity analysis of identification systems, more recent considerations are based on multiple HT framework with a rejection option. Possible examples for binary data statistics are presented in [20] and [22]. Motivated by the prior art, we extend the problem of content identification as multiple HT with rejection to a broader class of source priors including AVS's. In this regard, the current study implies a solution to a new biometric identification problem [24]. Our analysis lies within the frames of the works by Hoeffding [1], Csiszár and Longo [2], Blahut [3], Birgé [4], Haroutunian [6], Fu and Shen [10], Tuncel [14], Grigoryan and Harutyunyan [21] with the aim of specifying the asymptotic bounds for error probabilities. Those papers do not treat an option of rejection. In particular, [3] characterizes the optimum relation between two error exponents in binary HT and [6] (see also [15], [19]) and [14] study the multiple ($M > 2$) HT for DMS's in terms of logarithmically asymptotic optimality (LAO) and error exponent achievability, respectively. Later advances in the binary and M-ary HT for a more general class of sources – AVS's (see also its coding framework [16]), are the subjects of [10] and [21], respectively. The latter derives also Chernoff bounds for HT on AVS's and extends the finding by Leang and Johnson [9] for DMS's. Our work is a further extension of M-ary HT for discrete sources in terms of errors occurring with respect to an additional rejection decision. The focus is on the attainable region of error exponents which trade-off between the *false acceptance of rejection decision* and *false rejection of true distribution*. A similar model of HT with empirically observed statistics for Markov sources has been explored by Gutman in [5]. Compared to [5] we make a new look into the compromises among error events. Another relevant prior work [7] investigates the exponential rate in binary HT of sources with known and unknown statistics.

II. Models of source and HT

Let $\mathcal{X}$ and $\mathcal{S}$ be finite alphabets of an information source and its states, respectively. Let $\mathcal{P}(\mathcal{X})$ be the set of all probability distributions (PD) on $\mathcal{X}$. The source in our focus is defined by the following family of conditional PD's G_s^* depending on arbitrarily and not probabilistically varying source state $s \in \mathcal{S}$:

$$\mathcal{G}^* \triangleq \{G_s^*,\, s \in \mathcal{S}\} \tag{1}$$

with $G_s^* \triangleq \{G^*(x|s), \quad x \in \mathcal{X}\}$. An output source vector $\mathbf{x} \triangleq (x_1, ..., x_N) \in \mathcal{X}^N$ will have the following probability if dictated by a state vector $\mathbf{s} \in \mathcal{S}^N$: $G^{*N}(\mathbf{x}|\mathbf{s}) \triangleq G_{\mathbf{s}}^*(\mathbf{x}) \triangleq \prod_{n=1}^N G^*(x_n|s_n)$. Furthermore, the probability of a subset $\mathcal{A}_N \subset \mathcal{X}^N$ subject to $\mathbf{s} \in \mathcal{S}^N$ is measured by the sum $G^{*N}(\mathcal{A}_N|\mathbf{s}) \triangleq G_{\mathbf{s}}^*(\mathcal{A}_N) \triangleq \sum\limits_{\mathbf{x}\in\mathcal{A}_N} G_{\mathbf{s}}^*(\mathbf{x})$.

Our model of HT is determined by $M+1$ hypotheses about the source distribution (1):

$$H_m : \mathcal{G}^* = \mathcal{G}_m, \quad H_R : \text{none of } H_m\text{'s is true}$$

with

$$\mathcal{G}_m \triangleq \{G_{m,s},\, s \in \mathcal{S}\}, \tag{2}$$

where $G_{m,s} \triangleq \{G_m(x|s),\ x \in \mathcal{X}\}$, $s \in \mathcal{S}$, $m = \overline{1, M}$. Let G_m be the stochastic matrix defined by (2). Based on N observations of the source one should make a decision in favor of one of those hypotheses. Typically it can be performed by

[*]The authors research was supported by national grant 11-1b255.

a decision maker/detector applying a test φ_N as a partition of $\mathcal{X}^N$ into $M+1$ disjoint subsets $\mathcal{A}_N^m$, $m=\overline{1,M}$, and $\mathcal{A}_N^R$. If $\mathbf{x}\in\mathcal{A}_N^m$ then the test adopts the hypothesis H_m. If $\mathbf{x}\in\mathcal{A}_N^R$, the test rejects all the hypotheses H_m, $m=\overline{1,M}$. The test design aims at achieving certain level of errors during the process of decision making. $(M+1)M$ different kinds of errors are possible. The probability of an erroneous acceptance of hypothesis H_l, when H_m was true is

$$\alpha_{l,m}(\varphi_N)\stackrel{\triangle}{=}\max_{\mathbf{s}\in\mathcal{S}^N} G_m^N(\mathcal{A}_N^l|\mathbf{s}),\quad 1\le l\ne m\le M. \tag{3}$$

And the error probability of false rejection, when H_m was true is defined by

$$\alpha_{R,m}(\varphi_N)\stackrel{\triangle}{=}\max_{\mathbf{s}\in\mathcal{S}^N} G_m^N(\mathcal{A}_N^R|\mathbf{s}),\quad m=\overline{1,M}. \tag{4}$$

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

$$\begin{aligned}\alpha_m(\varphi_N) &\stackrel{\triangle}{=} \max_{\mathbf{s}\in\mathcal{S}^N} G_m^N(\overline{\mathcal{A}}_N^m|\mathbf{s})\\ &= \sum_{l\ne m}^{M}\alpha_{l,m}(\varphi_N)+\alpha_{R,m}(\varphi_N),\ m=\overline{1,M}.\end{aligned}\tag{5}$$

So, we study the following error probability exponents/reliabilities (log-s and exp-s being to the base 2) by (3) and (4):

$$E_{l|m}(\varphi)\stackrel{\triangle}{=}\limsup_{N\to\infty}-\frac{1}{N}\log\alpha_{l|m}^N(\varphi_N),\quad l\ne m=\overline{1,M}, \tag{6}$$

$$E_{R,m}(\varphi)\stackrel{\triangle}{=}\limsup_{N\to\infty}-\frac{1}{N}\log\alpha_{R,m}^N(\varphi_N),\quad m=\overline{1,M}, \tag{7}$$

where $\varphi\stackrel{\triangle}{=}\{\varphi_N\}_{N=1}^{\infty}$. From (5), (6), and (7) it follows that

$$E_m(\varphi)=\min_{l\ne m}\left[E_{l|m}(\varphi),E_{R,m}(\varphi)\right]. \tag{8}$$

In view of achievability concept [14] for reliabilities in M-ary HT, consider the $M(M+1)$-dimensional point $\mathbf{E}\stackrel{\triangle}{=}\{E_{R,m},E_m\}_{m=\overline{1,M}}$ with respect to the error exponent pairs $(-\frac{1}{N}\log\alpha_{R,m}(\varphi_N),-\frac{1}{N}\log\alpha_m(\varphi_N))$, where the decision regions $\mathcal{A}_N^m$ ($m=\overline{1,M}$) and $\mathcal{A}_N^R$ satisfy $\mathcal{A}_N^m\cap\mathcal{A}_N^l=\emptyset$ for $m\ne l$, $\mathcal{A}_N^m\cap\mathcal{A}_N^R=\emptyset$ and $\bigcup\limits_m\mathcal{A}_N^m=\mathcal{X}^N/\mathcal{A}_N^R$.

Definition 1. $\mathbf{E}$ is called achievable if for all $\varepsilon>0$ there exists a decision scheme $\{\mathcal{A}_N^m,\mathcal{A}_N^R\}_{m=1}^M$ such that

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

for N large enough. Let $\mathcal{R}_{\text{AVS}}(M)$ denotes the set of all achievable reliabilities.

III. Basic Properties

Here we resume some necessary material on the typical sequences [8]. Let $\mathcal{P}(\mathcal{S})\stackrel{\triangle}{=}\{P(s),\,s\in\mathcal{S}\}$ be the collection of all PD's on $\mathcal{S}$ and let PG be a marginal PD on $\mathcal{X}$ defined by $PG(x)\stackrel{\triangle}{=}\sum\limits_{s\in\mathcal{S}}P(s)G(x|s)$, $x\in\mathcal{X}$.

The type of the vector $\mathbf{s}\in\mathcal{S}^N$ is the empirical PD $P_{\mathbf{s}}(s)\stackrel{\triangle}{=}\frac{1}{N}N(s|\mathbf{s})$, 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-length 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

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

The joint type of vectors $\mathbf{x}$ and $\mathbf{s}$ is the PD $P_{\mathbf{s}}\circ G_{\mathbf{x},\mathbf{s}}\stackrel{\triangle}{=}\{P_{\mathbf{s}}(s)G_{\mathbf{x},\mathbf{s}}(x|s),\ x\in\mathcal{X},\ s\in\mathcal{S}\}$. For brevity the type notations can be used without indices. Let $\mathcal{G}^N(\mathcal{X}|\mathcal{S})$ be the set of all conditional types (9) and $\mathcal{G}(\mathcal{X})$ be the set of all distributions defined on $\mathcal{X}$. Denote by $\mathcal{T}_G^N(X|\mathbf{s})$ 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 $H(G|P)$. The notation $H(Q)$ will stand for the unconditional entropy of $Q\in\mathcal{P}(\mathcal{X})$. Denote by $D(G\parallel G_m|P)$ the KL divergence between G and G_m given type P and by $D(PG\parallel PG_m)$ the one between marginals PG and PG_m. The following inequality holds for every $G_m\in\mathcal{G}_m$:

$$D(G\parallel G_m|P)\ge D(PG\parallel PG_m). \tag{10}$$

We need the next properties:

$$|\mathcal{G}^N(\mathcal{X}|\mathcal{S})|\le(N+1)^{|\mathcal{X}||\mathcal{S}|}, \tag{11}$$

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

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

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

(12) and (13) give an estimate for conditional type class probability

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

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

IV. Region of Acheivable Reliabilities

Introduce the following convex hulls for each $m=\overline{1,M}$

$$\mathcal{W}_m\stackrel{\triangle}{=}\{W_m(x)\stackrel{\triangle}{=}\sum_{s\in\mathcal{S}}\lambda_s G_{m,s}(x|s)\}, \tag{16}$$

where $x\in\mathcal{X}$, $0\le\lambda_s\le1$, $\sum\limits_{s\in\mathcal{S}}\lambda_s=1$, and the region

$$\begin{aligned}\mathcal{E}_{\text{AVS}}(M) \stackrel{\triangle}{=} \{\mathbf{E}:\ &\forall W\ \exists\, m\ (m=\overline{1,M}),\ \text{s. t.}\\ &\min_{W_m\in\mathcal{W}_m}D(W\parallel W_m)>E_m \text{ and } \exists\, W \text{ s. t.}\\ &\min_{W_m\in\mathcal{W}_m}D(W\parallel W_m)>E_{R,m}\text{ for all } m\}.\end{aligned}\tag{17}$$

Our main result shows that (17) completely characterizes $\mathcal{R}_{\text{AVS}}(M)$.

Theorem 1: $\mathcal{E}_{\text{AVS}}(M)$ is an achievable region of reliabilities $\mathcal{E}_{\text{AVS}}(M)\subset\mathcal{R}_{\text{AVS}}(M)$. Moreover, if $\mathbf{E}\in\mathcal{R}_{\text{AVS}}(M)$, then for any $\delta>0$, $\mathbf{E}_\delta\in\mathcal{E}_{\text{AVS}}(M)$, where $\mathbf{E}_\delta\stackrel{\triangle}{=}\{E_{R,m}-\delta,E_m-\delta\}_{m=\overline{1,M}}$.

Proof: For the direct part, if $\mathbf{E} \in \mathcal{E}_{\text{AVS}}(M)$, then from (10), (12), (13) and (15) for any type $G \in \mathcal{G}^N(\mathcal{X}|\mathcal{S})$ and $\mathbf{s} \in \mathcal{S}^N$ with type $P_{\mathbf{s}} = P$ we have

$$G^N_{m,s}(\overline{\mathcal{A}}^m_N|\mathbf{s}) = \sum_{\mathbf{x}\in\overline{\mathcal{A}}^m_N} G^N_{m,s}(\mathbf{x}|\mathbf{s})$$
$$\leq \sum_{\mathcal{T}^N_G(X|\mathbf{s})\subset\overline{\mathcal{A}}^m_N} \exp\{-ND(G \parallel G_{m,s}|P)\}$$
$$\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-ND(PG \parallel PG_{m,s})\}. \quad (18)$$

For every $W_m \in \mathcal{W}_m$ there exists $\mathbf{s} \in \mathcal{S}^N$, such that $W_m = P_{\mathbf{s}} G_{m,s}$. Hence, from (18) and (11) we come to

$$\begin{aligned} \alpha_m(\varphi_N) &\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-N \min_{W_m} D(W \parallel W_m)\} \\ &\leq |\mathcal{G}^N(\mathcal{X}|\mathcal{S})| \exp\{-N E_m\} \\ &\leq \exp\{-N(E_m - \delta)\}. \end{aligned}$$

In the same way we could get the necessary inequality for $\alpha_{R,m}(\varphi_N)$, that is

$$\alpha_{R,m}(\varphi_N) \leq \exp\{-N(E_{R,m} - \delta)\}. \quad (19)$$

This closes the proof of the direct part.

For the converse we assume that $\mathbf{E} \in \mathcal{R}_{\text{AVS}}(M)$. This provides that for every $\varepsilon > 0$ there exists a decision scheme $\{\mathcal{A}^m_N, \mathcal{A}^R_N\}^M_{m=1}$ that makes the following inequalities true as soon as $N > N_0(\varepsilon)$:

$$-\frac{1}{N}\log \alpha_{R,m}(\varphi_N) > E_{R,m} - \varepsilon,\ -\frac{1}{N}\log \alpha_m(\varphi_N) > E_m - \varepsilon, \quad (20)$$

for all m's. Pick a $\delta > 0$ and show that

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

$$\exists W \text{ s. t. } \min_{W_m\in\mathcal{W}_m} D(W \parallel W_m) > E_{R,m} - \delta \text{ for all } m. \quad (22)$$

For that we prove the next fact. For every $W_m \in \mathcal{W}_m$ and $\mathcal{A}^m_N \subseteq \mathcal{X}^N$ the inequality holds:

$$W^N_m(\mathcal{A}^m_N) \leq \max_{\mathbf{s}\in\mathcal{S}^N} G^N_m(\mathcal{A}^m_N|\mathbf{s}). \quad (23)$$

To show (23), first note that for $W_m \in \mathcal{W}_m$ there exists a collection of λ_s's (by (16)) s.t. $W_m = \sum_{s\in\mathcal{S}} \lambda_s G_{m,s}$. Whence, for $\lambda_{\mathbf{s}} \triangleq \prod_{n=1}^N \lambda_{s_n}$ and any $\mathcal{A}^m_N \subset \mathcal{X}^N$, $\mathbf{x} \in \mathcal{A}^m_N$, the following estimate implies

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

Therefore we get (23).

Turning to (21), by the continuity of $D(\cdot \parallel W_m)$ there exists a type $Q \in \mathcal{P}^N(\mathcal{X})$ that for $N > N_1(\varepsilon)$ and a fixed m satisfies

$$D(Q \parallel W_m) \leq D(W \parallel W_m) + \delta/2. \quad (24)$$

Let $\overline{W}_{m^*} \triangleq \arg[\max_m \min_{W_m\in\mathcal{W}_m} D(W \parallel W_m)]$, then in light of (23) and (12) we have

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

where Q is a type-approximation of W defined by (24) for $\overline{W}^N_{m^*}$. Note that $|\overline{\mathcal{A}}^{m^*}_N \cap \mathcal{T}^N_Q(X)| \exp\{-NH(Q)\} \geq \exp\{-N\delta/4\}$ for $N > N_2(\delta)$. It follows from the inequality $|\overline{\mathcal{A}}^{m^*}_N \cap \mathcal{T}^N_Q(X)| \geq \frac{|\mathcal{T}^N_Q(X)|}{M}$:

$$\begin{aligned} |\overline{\mathcal{A}}^{m^*}_N &\cap \mathcal{T}^N_Q(X)| \exp\{-NH(Q)\} \\ &\geq |\mathcal{T}^N_Q(X)| \exp\{-NH(Q)\} \exp\{-N\frac{\log M}{N}\} \\ &\geq \exp\{-N\delta/4\}. \end{aligned} \quad (25)$$

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

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

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

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

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

Now we have to proceed with the proof of (22). Suppose $\overline{W}_m \triangleq \arg \min_{W_m\in\mathcal{W}_m} D(W \parallel W_m)$. For a picked $\delta > 0$, if $\mathbf{E}_\delta \notin \mathcal{E}_{\text{AVS}}(M)$ then $\forall W\, \exists m$ satisfying $D(W \parallel \overline{W}_m) \leq E_{R,m} - \delta$.

According to (23), (12), (24) and (25) we have

$$\begin{aligned} \alpha_{R,m}(\varphi_N) &\geq \overline{W}^N_m(\mathcal{A}^R_N) \\ &\geq \overline{W}^N_m(\mathcal{A}^R_N \cap \mathcal{T}^N_Q(X)) \\ &= \sum_{\mathcal{A}^R_N\cap\mathcal{T}^N_Q(X)} \exp\{-N[H(Q) \\ &\quad + D(Q \parallel \overline{W}_m)]\} \\ &\geq |\mathcal{A}^R_N \cap \mathcal{T}^N_Q(X)| \exp\{-NH(Q)\} \times \\ &\quad \times \exp\{-ND(Q \parallel \overline{W}_m)\} \\ &\geq \exp\{-N[D(W \parallel \overline{W}_m) - \delta/4]\} \\ &\geq \exp\{-N[E_{R,m} - \delta/4]\}. \end{aligned}$$

However the last inequality is in conflict with (20) for $\varepsilon < \delta/4$ and $N > \max\{N_0(\varepsilon), N_1(\delta), N_2(\delta)\}$. ■

V. Optimal decision schemes

Here we look for optimal decision schemes and the corresponding best error exponents in the following sense (similar to LAO test [6], [15]). Let $E_m,\ m = \overline{1,M}$, be fixed: what are the "maximum" values for $\{E^*_{l,m}, E^*_{R,m}\}_{l\neq m=\overline{1,M}}$ such that there is no other $\{E'_{l,m}, E'_{R,m}\}_{l\neq m=\overline{1,M}}$ satisfying $E'_{l,m} > E^*_{l,m}$ and $E'_{R,m} > E^*_{R,m}$ for all $l \neq m = \overline{1,M}$? Consider the following test sequence φ^* in terms of the sets

$$\mathcal{B}_R \triangleq \{W : \min_{W_m\in\mathcal{W}_m} D(W \parallel W_m) > E_m \text{ for all } m\},$$

$$\mathcal{B}_m \triangleq \{W : \min_{W_m\in\mathcal{W}_m} D(W \parallel W_m) < E_m\}, \quad m = \overline{1,M}.$$

Define ($l \neq m = \overline{1,M}$):

$$E_{R,m}(\varphi^*) \triangleq E^*_{R,m} \triangleq \min_{W\in\mathcal{B}_R}\min_{W_m\in\mathcal{W}_m} D(W \parallel W_m), \quad (26)$$

$$E_{l,m}(\varphi^*) \triangleq E^*_{l,m} \triangleq \min_{W\in\mathcal{B}_l}\min_{W_m\in\mathcal{W}_m} D(W \parallel W_m). \quad (27)$$

Theorem 2: Let the following inequalities hold:

$$E_1 < \min_m\{\min_{W_m\in\mathcal{W}_m, W_1\in\mathcal{W}_1} D(W_m \parallel W_1)\},$$

$$E_m < \min_{l\neq m}\{\min_{l=\overline{1,m-1}} E_{l,m}, \min_{l=\overline{m+1,M}} \min_{\substack{W_l\in\mathcal{W}_l,\\ W_m\in\mathcal{W}_m}} D(W_l \parallel W_m)\},$$

then there exist optimal sequence of tests and the corresponding optimal vector of reliabilities are defined by (26)-(27).

Proof: Let the decision on R or an m be made based on the partition: $\mathcal{D}_m \triangleq \bigcup_{W\in\mathcal{B}_m}\mathcal{T}^N_W(X)$, $\mathcal{D}_R \triangleq \bigcup_{W\in\mathcal{B}_R}\mathcal{T}^N_W(X)$. Note that $\mathcal{D}_m\cap\mathcal{D}_l \neq \emptyset$ and $\mathcal{D}_m\cap\mathcal{D}_R \neq \emptyset$, $m \neq l = \overline{1,M}$.

For $\overline{W}_m \triangleq arg\min_{W_m\in\mathcal{W}_m} D(W \parallel W_m)$, $m = \overline{1,M}$, and $\varphi \triangleq \{\varphi^*_N\}^\infty_{N=1}$ perform (applying unconditional verion of (14))

$$\begin{aligned}
\alpha_{R,m}(\varphi_N) &\geq \overline{W}^N_m(\mathcal{D}_R)\\
&\geq \overline{W}^N_m(\bigcup_{W\in\mathcal{B}_R}\mathcal{T}^N_W(X))\\
&\geq \max_{W\in\mathcal{B}_R}\exp\{-N[D(W \parallel \overline{W}_m) + o_N(1)]\}\\
&= \exp\{-N[\min_{W\in\mathcal{B}_R} D(W \parallel \overline{W}_m) + o_N(1)]\}.
\end{aligned}$$

In a similar way we can obtain the inequality

$$\alpha_{l,m}(\varphi_N) \geq \exp\{-N[\min_{W_m\in\mathcal{W}_m}\min_{W\in\mathcal{D}_l} D(W \parallel W_m) + o_N(1)]\}. \quad (28)$$

The proof of the converse inequalities

$$\alpha_{R,m}(\varphi_N) \leq \exp\{-N[\min_{W_m\in\mathcal{W}_m}\min_{W\in\mathcal{D}_R} D(W \parallel W_m) + o_N(1)]\} \quad (29)$$

$$\alpha_{l,m}(\varphi_N) \leq \exp\{-N[\min_{W_m\in\mathcal{W}_m}\min_{W\in\mathcal{D}_l} D(W \parallel W_m) + o_N(1)]\} \quad (30)$$

are omitted here because of space restrictions.

Taking into account (28), (29), (30) and the continuity of the functional $D(W \parallel W_m)$ we obtain that the limit $\lim_{N\to\infty}\{\sup -N^{-1}\log\alpha^N_{l,m}(\varphi^*_N)\}$ exists and equals to $E^*_{l,m}$.

The proof will be accomplished if we demonstrate that φ^* is optimal. Let φ' be a test defined by the sets $(\mathcal{D}'_m, \mathcal{D}'_R)$ s.t.

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

It yields for N large enough that

$$\alpha^N_{l,m}(\varphi'_N) < \alpha^N_{l,m}(\varphi^*_N), \quad \alpha^N_{R,m}(\varphi'_N) < \alpha^N_{R,m}(\varphi^*_N).$$

Below we examine the relation between $(\mathcal{D}_m, \mathcal{D}_R)$ and $(\mathcal{D}'_m, \mathcal{D}'_R)$. 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 $\mathcal{D}_R$ and $\mathcal{D}'_R$. Consider $\mathcal{D}_m\cap\mathcal{D}'_m = \emptyset$ case. It follows that there exists $l \neq m$ such that $\mathcal{D}_m\cap\mathcal{D}'_l \neq \emptyset$. That is $\exists W$ such that $D(W \parallel \overline{W}_m) < E'_m$, so $\mathcal{T}^N_W(X) \subset \mathcal{D}'_l$. Compute

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

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

Remark 1: It can be proved also that

$$\min\left[\min_{l=\overline{1,M}, l\neq m} E^*_{l,m}, E^*_{R,m}\right] = E^*_{R,m} \text{ for all } m = \overline{1,M}.$$

This means that *discrimination is always easier than rejection*.

VI. Geometric interpretations

Fig. 1 presents a geometric interpretation for the decision scheme in Theorem 1. Relevantly, Fig. 2 and 3 illustrate the geometry of the Chernoff bounds derived in [21] for the multi-HT where the rejection is not an alternative (c.f. [18] for DMS's). Those interpretations are comprehensible with conceptual details given in [21].

VII. Results for DMS

With assumption of $\mathcal{S} = 1$ we get the model of multi-HT with rejection for DMS:

$$H_m : G^* = G_m,\ H_R : \text{none of } H_m\text{'s is true},$$

with $G_m \triangleq \{G_m(x), x \in \mathcal{X}\}$, $m = \overline{1,M}$. The problem here is to make a decision regarding the generic G^* among M alternative PD's $G_m,\ m = \overline{1,M}$, and the rejection. Let

$$\begin{aligned}
\mathcal{E}(M) \triangleq\ &\{\mathbf{E} : \quad \forall Q\ \exists\, m\ (m = \overline{1,M}),\ \text{s. t.}\\
&D(Q \parallel G_m) > E_m \text{ and } \exists\, Q \text{ s. t.}\\
&D(Q \parallel G_m) > E_{R,m} \text{ for all } m\}
\end{aligned}$$

and let $\mathcal{R}(M)$ be the DMS version of $\mathcal{R}_{\text{AVS}}(M)$.

Theorem 3: Theorem 1 implies that $\mathcal{E}(M) \subset \mathcal{R}(M)$. Conversely, if $\mathbf{E} \in \mathcal{R}(M)$, then for any $\delta > 0$, $\mathbf{E}_\delta \in \mathcal{E}(M)$, where $\mathbf{E}_\delta \triangleq \{E_{R,m} - \delta, E_m - \delta\}_{m=\overline{1,M}}$.

To formulate the DMS counterpart of Theorem 2 define the sets:

$$\mathcal{B}_R(\text{DMS}) \triangleq \{Q :\ D(Q \parallel G_m) > E_m, \text{ for all } m = \overline{1,M}\},$$

$$\mathcal{B}_m(\text{DMS}) \stackrel{\triangle}{=} \{Q:\ D(Q \parallel G_m) < E_m\},\ m = \overline{1, M}.$$

Furthermore

$$E^*_{R,m} \stackrel{\triangle}{=} \min_{Q \in \mathcal{B}_R(\text{DMS})} D(Q \parallel G_m), \quad m = \overline{1, M}, \tag{31}$$

$$E^*_{l,m} \stackrel{\triangle}{=} \min_{Q \in \mathcal{B}_l(\text{DMS})} D(Q \parallel G_m), \quad l \neq m = \overline{1, M}. \tag{32}$$

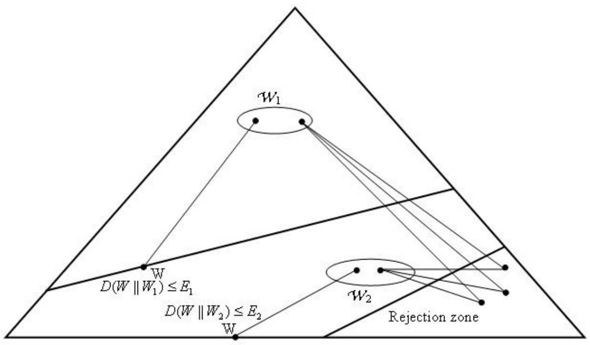

Fig. 1: Multiple HT with rejection.

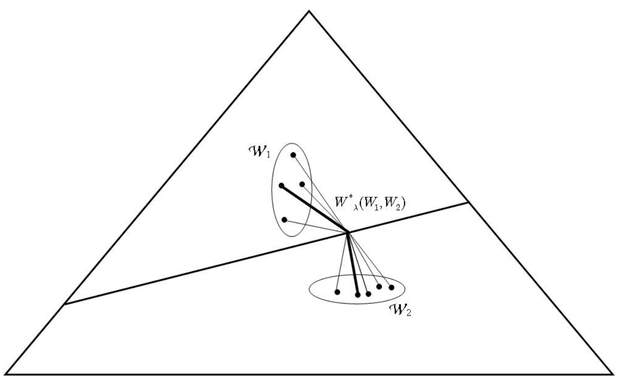

Fig. 2: Chernoff bounds: binary HT: AVS.

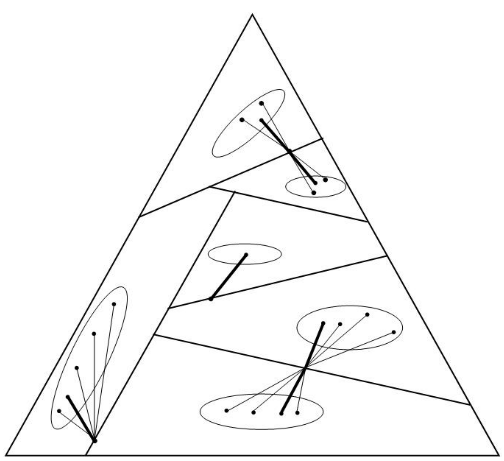

Fig. 3: Chernoff bounds: multiple HT: AVS.

Theorem 4: If $D(G_m \parallel G_l) > 0$, $m \neq l = \overline{1, M}$, and

$$E_1 < \min_m \{D(G_m \parallel G_1)\},$$

$$E_m < \min_{l \neq m} \{ \min_{l = \overline{1, m-1}} E_{l,m}, \min_{l = \overline{m+1, M}} D(G_l \parallel G_m)\},$$

then the optimal vector of reliabilities are defined by (31)-(32).

According to [23] the authors claim to have obtained Theorem 4 independently.

References

[1] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions", *Ann. Math. Statist.*, vol. 36, pp. 369–401, 1965.

[2] 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.

[3] R.E. Blahut, "Hypothesis testing and information theory", *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.

[4] L. Birgé, "Vitess maximals de déroissance des erreurs et tests optimaux associes", *Z. Wahrsch. Verw. Gebiete*, vol. 55, pp. 261–173, 1981.

[5] M. Gutman, "Asymptotically optimal classification for multiple tests with empirically observed statistics", *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 401–408, 1989.

[6] E.A. Haroutunian, "Logarithmically asymptotically optimal testing of multiple statistical hypotheses", *Problems of Control and Inform. Theory,* vol. 19, no. 5-6, pp. 413–421, 1990.

[7] O. Zeitouni and M. Gutman, "On universal hypotheses testing via large deviations", *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 285–290, March 1991.

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

[9] 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.

[10] 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.

[11] J. Haitsma, T. Kalker, and J. Oostveen, "Robust audio hashing for content identification", *Proc. Int. Workshop Content-Based Multimedia Indexing*, pp. 117–125, Brescia, Italy, Sept. 2001.

[12] F. Lefebvre and B. Macq, "Rash : RAdon Soft Hash algorithm", *Proc. EUSIPCO*, Toulouse, France, Sept. 3–6, 2002.

[13] F. Willems, T. Kalker, J. Goseling, and J.P. Linnartz, "On the capacity of a biometrical identification system", *Proc. IEEE Intern. Symp. Inf. Theory*, p. 82, Yokohama, Japan, June 29 – July 4, 2003.

[14] E. Tuncel, "On error exponents in hypothesis testing", *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2945–2950, 2005.

[15] R.F. Ahlswede and E.A. Haroutunian, "On statistical hypothesis optimal testing and identification", *Lecture Notes in Computer Science* 4123, pp. 553–571, Springer Verlag, 2006.

[16] A.N. Harutyunyan and A.J. Han Vinck, "Error exponents in AVS coding", *Proc. IEEE Intern. Symp. Inform. Theory*, pp. 2166–2170, Seattle, WA, July 9–14, 2006.

[17] O. Koval, S. Voloshynovskiy, F. Beekhof, and T. Pun, "Decision-theoretic consideration of robust perceptual hashing: link to practical algorithms", *WaCha2007*, Saint Malo, France, June 15, 2007.

[18] M. Brandon Westover, "Asymptotic geometry of multiple hypothesis testing", *IEEE Trans. Inform. Theory*, vol. 54, no. 7, pp. 3327–3329, 2008.

[19] E.A. Haroutunian, M.E. Haroutunian, and A.N. Harutyunyan, *Reliability Criteria in Information Theory and in Statistical Hypothesis Testing*, FnT in Communic. and Inform. Theory, vol. 4, no 2–3, pp 97–263, 2008.

[20] A.L. Varna, A. Swaminathan, and M. Wu, "A decision theoretic framework for analyzing hash-based content identification systems", *Proc. ACM Digital Rights Management Workshop*, pp. 67–76, Oct. 27, 2008.

[21] 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.

[22] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification", *Proc. IEEE Workshop Inform. Theory*, Dublin, Ireland, Aug. 30 – Sept. 3, 5 p., 2010.

[23] E. Haroutunian and P. Hakobyan, Personal communication, 2011.

[24] A. Harutyunyan, N. Grigoryan, S. Voloshynovsky, and O. Koval, "A new biometric identification model and the multiple hypothesis testing for arbitrarily varying objects", accepted, *BIOSIG*, Darmstadt, Germany, Sept. 8–9, 8 p., 2011.