Arimoto-Rényi Conditional Entropy and Bayesian Hypothesis Testing

Igal Sason
Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion-Israel Institute of Technology
Haifa 32000, Israel
E-mail: sason@ee.technion.ac.il

Sergio Verdú
Department of Electrical Engineering
Princeton University
New Jersey 08544, USA
E-mail: verdu@princeton.edu

***Abstract*—This paper gives upper and lower bounds on the minimum error probability of Bayesian M-ary hypothesis testing in terms of the Arimoto-Rényi conditional entropy of an arbitrary order α. The improved tightness of these bounds over their specialized versions with the Shannon conditional entropy ($\alpha = 1$) is demonstrated. In particular, in the case where M is finite, we show how to generalize Fano's inequality under both the conventional and list-decision settings. As a counterpart to the generalized Fano's inequality, allowing M to be infinite, a lower bound on the Arimoto-Rényi conditional entropy is derived as a function of the minimum error probability. Explicit upper and lower bounds on the minimum error probability are obtained as a function of the Arimoto-Rényi conditional entropy.**

Index Terms – Information measures, hypothesis testing, Arimoto-Rényi conditional entropy, Rényi divergence, Fano's inequality, minimum probability of error.

I. Introduction

In Bayesian M-ary hypothesis testing, we have:

- M possible explanations, hypotheses or models for the $\mathcal{Y}$-valued data $\{P_{Y|X=m}, m \in \mathcal{X}\}$ where $|\mathcal{X}| = M$; and
- a prior distribution P_X on $\mathcal{X}$, the set of model indices.

The minimum probability of error of X given Y, denoted by $\varepsilon_{X|Y}$, is achieved by the *maximum-a-posteriori* decision rule. A number of bounds on $\varepsilon_{X|Y}$ involving information measures have been obtained in the literature, most notably:

1) Fano's inequality [10] gives an upper bound on the conditional entropy $H(X|Y)$ as a function of $\varepsilon_{X|Y}$ when M is finite.
2) Shannon's inequality [28] (see also [34]) gives an explicit lower bound on $\varepsilon_{X|Y}$ as a function of $H(X|Y)$, also when M is finite.
3) Tightening another bound by Shannon [27], Poor and Verdú [23] gave a lower bound on $\varepsilon_{X|Y}$ (generalized in [5]) as a function of the distribution of the conditional information (whose expected value is $H(X|Y)$).
4) Baladová [3], Chu and Chueh [6, (12)] and Hellman and Raviv [14, (41)] showed that
$$\varepsilon_{X|Y} \leq \tfrac{1}{2}\, H(X|Y) \text{ bits} \tag{1}$$
for finitely valued random variables. It is also easy to show that (see, e.g., [11, (21)])
$$\varepsilon_{X|Y} \leq 1 - \exp\bigl(-H(X|Y)\bigr). \tag{2}$$
Tighter and generalized upper bounds on $\varepsilon_{X|Y}$ were obtained by Kovalevsky [19], Tebbe and Dwyer [30], and Ho and Verdú [15, (109)].
5) Based on the fundamental tradeoff of an auxiliary binary hypothesis test, Polyanskiy *et al.* [21] gave the *meta-converse* implicit lower bound on $\varepsilon_{X|Y}$.
6) In the case $M = 2$, Hellman and Raviv [14] gave an upper bound on $\varepsilon_{X|Y}$ as a function of the prior probabilities and the Rényi divergence of order $\alpha \in [0, 1]$ between the two models. The special case of $\alpha = \frac{1}{2}$ yields the *Bhattacharyya bound* [16].
7) In [9] and [33], Devijver and Vajda derived upper and lower bounds on $\varepsilon_{X|Y}$ as a function of the quadratic Arimoto-Rényi conditional entropy $H_2(X|Y)$.
8) Building up on [14], Kanaya and Han [17] showed that in the case of independent identically distributed (i.i.d.) observations, $\varepsilon_{X|Y^n}$ and $H(X|Y^n)$ vanish exponentially at the same speed, which is governed by the Chernoff information between the closest hypothesis pair.
9) Generalizing Fano's inequality, Han and Verdú [13] gave lower bounds on the mutual information $I(X;Y)$ as a function of $\varepsilon_{X|Y}$, one of which was generalized by Polyanskiy and Verdú [22] to give a lower bound on the α-mutual information.
10) In [29], Shayevitz gave a lower bound, in terms of the Rényi divergence, on the maximal worst-case miss-detection exponent for a binary composite hypothesis testing problem when the false-alarm probability decays to zero with the number of i.i.d. observations.
11) Tomamichel and Hayashi [31], [32] studied optimal exponents of binary composite hypothesis testing, expressed in terms of Rényi's information measures. A measure of dependence was studied in [32] (see also Lapidoth and Pfister [20]) along with its role in composite hypothesis testing.

This paper (whose extended version is [25]) gives upper and lower bounds on $\varepsilon_{X|Y}$ not in terms of $H(X|Y)$ but in terms of the *Arimoto-Rényi* conditional entropy $H_\alpha(X|Y)$ of an arbitrary order α. Indeed, in this paper we find pleasing counterparts to the bounds in Items 1), 4), 6), 7), 8) and 9), resulting in generally tighter bounds. In addition, we enlarge the scope of the problem to consider not only $\varepsilon_{X|Y}$ but the probability that a list decision rule (which is allowed to

978-1-5090-4096-4/17/$31.00

output a set of L hypotheses) does not include the true one. Previous work on extending Fano's inequality to the setup of list decision rules includes [1, Section 5] and [18, Lemma 1].

Section II introduces the basic notation and definitions of Rényi information measures. Section III contains the main results in the paper on the interplay between $\varepsilon_{X|Y}$ and $H_\alpha(X|Y)$, giving counterparts to a number of those existing results mentioned above. In particular:

1) an upper bound on $H_\alpha(X|Y)$ as a function of $\varepsilon_{X|Y}$ is derived for positive α; it provides an implicit lower bound on $\varepsilon_{X|Y}$ as a function of $H_\alpha(X|Y)$;
2) explicit lower bounds on $\varepsilon_{X|Y}$ are given as a function of $H_\alpha(X|Y)$ for both positive and negative α;
3) the lower bounds are extended to the list-decoding setting;
4) as a counterpart to the generalized Fano's inequality, we derive a lower bound on $H_\alpha(X|Y)$ as a function of $\varepsilon_{X|Y}$ capitalizing on the Schur concavity of Rényi entropy.

Due to space limitations, all proofs are provided in [25], which, in addition, gives upper bounds on the minimum error probability as a function of the Rényi divergence and the Chernoff information. In the setup of discrete memoryless channels, we analyze in [25] the exponentially vanishing decay of the Arimoto-Rényi conditional entropy of the transmitted codeword given the channel output when averaged over a code ensemble.

II. Rényi Information Measures

Definition 1: [24] Let P_X be a probability distribution on a discrete set $\mathcal{X}$. The *Rényi entropy of order* $\alpha \in (0,1)\cup(1,\infty)$ of X is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x\in\mathcal{X}} P_X^\alpha(x). \tag{3}$$

By its continuous extension,

$$H_0(X) = \log\bigl|\{x \in \mathcal{X} \colon P_X(x) > 0\}\bigr|, \tag{4}$$

$$H_1(X) = H(X), \tag{5}$$

$$H_\infty(X) = \log \frac{1}{p_{\max}} \tag{6}$$

where $p_{\max}$ is the largest of the masses of X.

Definition 2: For $\alpha \in (0,1)\cup(1,\infty)$, the binary Rényi entropy is defined, for $p \in [0,1]$, as

$$h_\alpha(p) = H_\alpha(X) = \frac{1}{1-\alpha} \log\bigl(p^\alpha + (1-p)^\alpha\bigr), \tag{7}$$

where X is a binary random variable with probabilities p and $1-p$. The continuous extension of the binary Rényi entropy at $\alpha = 1$ yields the binary entropy function:

$$h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}. \tag{8}$$

In order to put forth generalizations of Fano's inequality and bounds on the error probability, we consider Arimoto's proposal for the conditional Rényi entropy (named, for short, the Arimoto-Rényi conditional entropy).

Definition 3: [2] Let P_{XY} be defined on $\mathcal{X}\times\mathcal{Y}$, where X is a discrete random variable. The *Arimoto-Rényi conditional entropy of order* $\alpha \in [0,\infty]$ of X given Y is defined as follows:

- If $\alpha \in (0,1)\cup(1,\infty)$, then

$$\begin{aligned} & H_\alpha(X|Y) \\ &= \frac{\alpha}{1-\alpha} \log \mathbb{E}\left[\left(\sum_{x\in\mathcal{X}} P_{X|Y}^\alpha(x|Y)\right)^{\frac{1}{\alpha}}\right] \qquad (9)\\ &= \frac{\alpha}{1-\alpha} \log \sum_{y\in\mathcal{Y}} P_Y(y) \exp\left(\frac{1-\alpha}{\alpha}\, H_\alpha(X|Y=y)\right), \qquad (10) \end{aligned}$$

  where (10) applies if Y is a discrete random variable.
- By its continuous extension, the Arimoto-Rényi conditional entropy of orders 0, 1 and ∞ is defined as

$$\begin{aligned} H_0(X|Y) &= \operatorname{ess\,sup} H_0\left(P_{X|Y}(\cdot|Y)\right) && (11)\\ &= \log \max_{y\in\mathcal{Y}} \bigl|\mathrm{supp} P_{X|Y}(\cdot|y)\bigr| && (12)\\ &= \max_{y\in\mathcal{Y}} H_0(X\,|\,Y=y), && (13)\\ H_1(X|Y) &= H(X|Y), && (14)\\ H_\infty(X|Y) &= -\log \mathbb{E}\Bigl[\max_{x\in\mathcal{X}} P_{X|Y}(x|Y)\Bigr] && (15) \end{aligned}$$

  where (12) and (13) apply if Y is a discrete random variable.

Although not nearly as important, sometimes in the context of finitely valued random variables, it is useful to consider the unconditional and conditional Rényi entropies of negative orders $\alpha \in (-\infty, 0)$ in (3) and (9) respectively. Basic properties of $H_\alpha(X|Y)$ appear in [12] and [25].

The third Rényi information measure used in this paper is the binary Rényi divergence.

Definition 4: For $\alpha \in (0,1)\cup(1,\infty)$, the binary Rényi divergence is defined as the continuous extension to $[0,1]^2$ of

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1} \log\Bigl(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\Bigr). \tag{16}$$

By analytic continuation in α, for $(p,q) \in (0,1)^2$,

$$d_0(p\|q) = 0, \tag{17}$$

$$d_1(p\|q) = d(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}, \tag{18}$$

$$d_\infty(p\|q) = \log \max\left\{\frac{p}{q}, \frac{1-p}{1-q}\right\} \tag{19}$$

where $d(\cdot\|\cdot)$ in (18) denotes the binary relative entropy.

III. Arimoto-Rényi Conditional Entropy and Error Probability

A. Upper bound on the Arimoto-Rényi conditional entropy: Generalized Fano's inequality

The minimum error probability $\varepsilon_{X|Y}$ can be achieved by *maximum-a-posteriori* decision rule $\mathcal{L}^*\colon \mathcal{Y} \to \mathcal{X}$:

$$\begin{aligned}
\varepsilon_{X|Y} &= \min_{\mathcal{L}\colon \mathcal{Y}\to\mathcal{X}} \mathbb{P}[X \neq \mathcal{L}(Y)] && (20)\\
&= \mathbb{P}[X \neq \mathcal{L}^*(Y)] && (21)\\
&= 1 - \mathbb{E}\left[\max_{x\in\mathcal{X}} P_{X|Y}(x|Y)\right] && (22)\\
&\leq 1 - p_{\max} && (23)
\end{aligned}$$

where (23) is the minimum error probability achievable among blind decision rules that disregard the observations.

Fano's inequality links the decision theoretic uncertainty $\varepsilon_{X|Y}$ and the information theoretic uncertainty $H(X|Y)$ through

$$\begin{aligned}
H(X|Y) &\leq \log M - d\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big) && (24)\\
&= h(\varepsilon_{X|Y}) + \varepsilon_{X|Y} \log(M-1). && (25)
\end{aligned}$$

Although the form of Fanos inequality in (24) is not nearly as popular as (25), it turns out to be the version that admits an elegant generalization to the Arimoto-Renyi conditional entropy. It is straightforward to obtain (25) by averaging a conditional version with respect to the observation. This simple route to the desired result is not viable in the case of $H_\alpha(X|Y)$ since it is not an average of Rényi entropies of conditional distributions. The conventional proof of Fano's inequality in [7, pp. 38–39] based on the use of the chain rule for entropy is also doomed to failure for the Arimoto-Rényi conditional entropy of order $\alpha \neq 1$ since it does not satisfy the chain rule.

Before we generalize Fano's inequality by linking $\varepsilon_{X|Y}$ with $H_\alpha(X|Y)$ for $\alpha \in [0,\infty)$, note that for $\alpha = \infty$, the following identity holds in view of (22):

$$\varepsilon_{X|Y} = 1 - \exp\big(-H_\infty(X|Y)\big). \qquad (26)$$

Theorem 1: Let P_{XY} be a probability measure defined on $\mathcal{X} \times \mathcal{Y}$ with $|\mathcal{X}| = M < \infty$. For all $\alpha \in (0,\infty)$,

$$H_\alpha(X|Y) \leq \log M - d_\alpha\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big). \qquad (27)$$

Equality holds in (27) if and only if

$$P_{X|Y}(x|y) = \begin{cases} \frac{\varepsilon_{X|Y}}{M-1}, & x \neq \mathcal{L}^*(y)\\ 1 - \varepsilon_{X|Y}, & x = \mathcal{L}^*(y). \end{cases} \qquad (28)$$

for all $y \in \mathcal{S}$ such that $P_Y(\mathcal{S}) = 1$, where $\mathcal{L}^*$ is a deterministic MAP decision rule (see (21)).

Proof: See [25, Theorem 3]. ■

In information theoretic problems, it is common to encounter the case in which X and Y are actually vectors of dimension n. Fano's inequality ensures that vanishing error probability implies vanishing normalized conditional entropy as $n \to \infty$. As we see next, the picture with the Arimoto-Rényi conditional entropy is more nuanced.

Theorem 2: Let $\{X_n\}$ be a sequence of random variables, with X_n taking values on $\mathcal{X}_n$ for $n \in \mathbb{N}$ and assume that there exists an integer $M \geq 2$ such that $|\mathcal{X}_n| \leq M^n$ for all n.[1] Let $\{Y_n\}$ be an arbitrary sequence of random variables, for which $\varepsilon_{X_n|Y_n} \to 0$ as $n \to \infty$. The following results hold for $H_\alpha(X_n|Y_n)$:

a) if $\alpha \in (1,\infty]$, then $H_\alpha(X_n|Y_n) \to 0$;
b) if $\alpha = 1$, then $\frac{1}{n} H(X_n|Y_n) \to 0$;
c) if $\alpha \in [0,1)$, then $\frac{1}{n} H_\alpha(X_n|Y_n)$ is upper bounded by $\log M$; nevertheless, it does not necessarily tend to 0.

Proof: See [25, Theorem 4]. ■

B. List decoding

In this section we consider the case where the decision rule outputs a list of choices. The extension of Fano's inequality to list decoding was initiated in [1, Section 5]. It is useful for proving converse results in conjunction with the blowing-up lemma ([8, Lemma 1.5.4]). The main idea of the successful combination of these two tools is that, given an arbitrary code, one can blow-up the decoding sets in such a way that the probability of decoding error can be as small as desired for sufficiently large blocklength; since the blown-up decoding sets are no longer disjoint, the resulting setup is a list decoder with subexponential list size.

A generalization of Fano's inequality for list decoding of size L is[2]

$$H(X|Y) \leq \log M - d\big(P_{\mathcal{L}} \| 1 - \tfrac{L}{M}\big), \qquad (29)$$

where $P_{\mathcal{L}}$ denotes the probability of X not being in the list. As we noted before, averaging a conditional version with respect to the observation is not viable in the case of $H_\alpha(X|Y)$ with $\alpha \neq 1$. A pleasing generalization of (29) to the Arimoto-Rényi conditional entropy does indeed hold as the following result shows.

Theorem 3: Let P_{XY} be a probability measure defined on $\mathcal{X} \times \mathcal{Y}$ where $|\mathcal{X}| = M$. Consider a decision rule[3] $\mathcal{L}\colon \mathcal{Y} \to \binom{\mathcal{X}}{L}$, and denote the decoding error probability by

$$P_{\mathcal{L}} = \mathbb{P}\big[X \notin \mathcal{L}(Y)\big]. \qquad (30)$$

Then, for all $\alpha \in (0,1) \cup (1,\infty)$,

$$\begin{aligned}
H_\alpha(X|Y) &\leq \log M - d_\alpha\big(P_{\mathcal{L}} \| 1 - \tfrac{L}{M}\big) && (31)\\
&= \frac{1}{1-\alpha} \log\Big(L^{1-\alpha}\left(1-P_{\mathcal{L}}\right)^\alpha + (M-L)^{1-\alpha} P_{\mathcal{L}}^\alpha\Big) && (32)
\end{aligned}$$

with equality in (31) if and only if

$$P_{X|Y}(x|y) = \begin{cases} \frac{P_{\mathcal{L}}}{M-L}, & x \notin \mathcal{L}(y)\\ \frac{1-P_{\mathcal{L}}}{L}, & x \in \mathcal{L}(y). \end{cases} \qquad (33)$$

[1] Note that this encompasses the conventional setting in which $\mathcal{X}_n = \mathcal{A}^n$.

[2] See [18, Lemma 1] for a weaker version of (29).

[3] $\binom{\mathcal{X}}{L}$ stands for the set of all the subsets of $\mathcal{X}$ with cardinality L, with $L \leq |\mathcal{X}|$.

Proof: See [25, Theorem 8]. ∎

Theorem 4: Let P_{XY} be a probability measure defined on $\mathcal{X} \times \mathcal{Y}$ with $|\mathcal{X}| = M < \infty$, which satisfies

$$P_{X|Y}(x|y) > 0, \quad (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{34}$$

and let $\mathcal{L}\colon \mathcal{Y} \to \binom{\mathcal{X}}{L}$. Then, for all $\alpha \in (-\infty, 0)$, the probability that the decoding list does not include the correct decision satisfies

$$P_{\mathcal{L}} \geq \exp\left(\frac{1-\alpha}{\alpha}\Big[H_\alpha(X|Y) - \log(M-L)\Big]\right). \tag{35}$$

Proof: See [25, Theorem 9]. ∎

Theorem 5: Let P_{XY} be a probability measure defined on $\mathcal{X} \times \mathcal{Y}$ which satisfies (34), with $\mathcal{X}$ being finite or countably infinite, and let $\mathcal{L}\colon \mathcal{Y} \to \binom{\mathcal{X}}{L}$. Then, for all $\alpha \in (1, \infty)$,

$$P_{\mathcal{L}} \geq 1 - \exp\left(\frac{1-\alpha}{\alpha}\Big[H_\alpha(X|Y) - \log L\Big]\right). \tag{36}$$

Proof: See [25, Theorem 10]. ∎

Remark 1: The implicit lower bound on $\varepsilon_{X|Y}$ given by the generalized Fano's inequality in (31) is tighter than the explicit lower bound in (36).

C. Lower bounds on the Arimoto-Rényi conditional entropy

The major existing lower bounds on the Shannon conditional entropy $H(X|Y)$ as a function of the minimum error probability $\varepsilon_{X|Y}$ are:

1) In view of [15, Theorem 11], (1) (shown in [3, Theorem 1], [6, (12)] and [14, (41)] for finite alphabets) holds for a general discrete random variable X. As an example where (1) holds with equality, let X and Y be random variables defined on $\{0, 1\}$ with $P_X(0) = \eta \in \left(0, \frac{1}{2}\right]$, $P_{Y|X}(1|0) = 1$, and $P_{Y|X}(1|1) = \frac{\eta}{1-\eta}$. Then, $\varepsilon_{X|Y} = \eta$ and $H(X|Y) = 2\eta$ bits.
2) Due to Kovalevsky [19], Tebbe and Dwyer [30] (see also [11]) in the finite alphabet case, and to Ho and Verdú [15, (109)] in the general case,

$$\phi(\varepsilon_{X|Y}) \leq H(X|Y) \tag{37}$$

where $\phi\colon [0, 1) \to [0, \infty)$ is the piecewise linear function that is defined on the interval $t \in \left[1 - \frac{1}{k}, 1 - \frac{1}{k+1}\right)$ as

$$\phi(t) = t\,k(k+1)\log\left(\frac{k+1}{k}\right) + (1-k^2)\log(k+1) + k^2\log k \tag{38}$$

where k is an arbitrary positive integer. Note that (37) is tighter than (1) since $\phi(t) \geq 2t\log 2$.

In view of (26), since $H_\alpha(X|Y)$ is monotonically decreasing in α, one can readily obtain the following bound:

$$H_\alpha(X|Y) \geq \log\frac{1}{1-\varepsilon_{X|Y}} \tag{39}$$

for $\alpha \in [0, \infty]$ with equality if $\alpha = \infty$.

The next result gives a counterpart to Theorem 1, and a generalization of (37).

Theorem 6: If $\alpha \in (0, 1) \cup (1, \infty)$, then

$$\frac{\alpha}{1-\alpha}\log g_\alpha(\varepsilon_{X|Y}) \leq H_\alpha(X|Y), \tag{40}$$

where the piecewise linear function $g_\alpha\colon [0, 1) \to \mathcal{D}_\alpha$, with $\mathcal{D}_\alpha = [1, \infty)$ for $\alpha \in (0, 1)$ and $\mathcal{D}_\alpha = (0, 1]$ for $\alpha \in (1, \infty)$, is defined by

$$g_\alpha(t) = \Big(k(k+1)^{\frac{1}{\alpha}} - k^{\frac{1}{\alpha}}(k+1)\Big)t + k^{\frac{1}{\alpha}+1} - (k-1)(k+1)^{\frac{1}{\alpha}} \tag{41}$$

on the interval $t \in \left[1 - \frac{1}{k}, 1 - \frac{1}{k+1}\right)$ for an arbitrary positive integer k.

Proof: See [25, Theorem 11]. ∎

Remark 2: The most useful domain of applicability of Theorem 6 is $\varepsilon_{X|Y} \in [0, \frac{1}{2}]$, in which case the lower bound specializes to ($k = 1$)

$$\frac{\alpha}{1-\alpha}\log\Big(1 + \big(2^{\frac{1}{\alpha}} - 2\big)\varepsilon_{X|Y}\Big) \leq H_\alpha(X|Y) \tag{42}$$

which yields (1) as $\alpha \to 1$.

Remark 3: Theorem 6 generalizes (37) since

$$\lim_{\alpha\to 1}\frac{\alpha}{1-\alpha}\,\log\,g_\alpha(\tau) = \phi(\tau), \tag{43}$$

for all $\tau \in [0, 1]$ with ϕ defined in (38).

Remark 4: As $\alpha \to \infty$, (40) is asymptotically tight.

Remark 5: Theorem 6 gives a tighter bound than (39), unless $\varepsilon_{X|Y} \in \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots \frac{M-1}{M}\}$ (M is allowed to be ∞ here) in which case they are identical, and independent of α (see Figure 1).

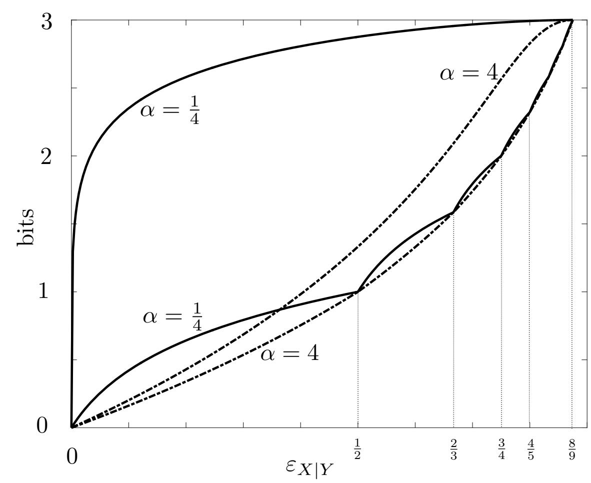

Fig. 1. Upper and lower bounds on $H_\alpha(X|Y)$ in Theorems 1 and 6, respectively, as a function of $\varepsilon_{X|Y} \in [0, 1 - \frac{1}{M}]$ for $\alpha = \frac{1}{4}$ (solid lines) and $\alpha = 4$ (dash-dotted lines) with $M = 8$.

The following result is illustrated by Figure 1.

Theorem 7: Let $M \in \{2, 3, \ldots\}$ be finite, and let the upper and lower bounds on $H_\alpha(X|Y)$ as a function of $\varepsilon_{X|Y}$, as given in Theorems 1 and 6, be denoted by $u_{\alpha,M}(\cdot)$ and $l_\alpha(\cdot)$, respectively. Then,

a) these bounds coincide if and only if X is a deterministic function of the observation Y or X is equiprobable on the set $\mathcal{X}$ and independent of Y;

b) the limit of the ratio of the upper-to-lower bounds when $\varepsilon_{X|Y} \to 0$ is given by

$$\lim_{\varepsilon_{X|Y}\to 0} \frac{u_{\alpha,M}(\varepsilon_{X|Y})}{l_\alpha(\varepsilon_{X|Y})} = \begin{cases} \infty, & \alpha \in (0,1) \\ \frac{1}{2-2^{\frac{1}{\alpha}}}, & \alpha \in (1,\infty). \end{cases} \tag{44}$$

Proof: See [25, Appendix B]. ∎

The following result is a consequence of Theorem 6:

Theorem 8: Let $k \in \mathbb{N}$, and $\alpha \in (0,1)\cup(1,\infty)$. If $\log k \le H_\alpha(X|Y) < \log(k+1)$, then

$$\varepsilon_{X|Y} \le \frac{\exp\left(\frac{1-\alpha}{\alpha}\, H_\alpha(X|Y)\right) - k^{\frac{1}{\alpha}+1} + (k-1)(k+1)^{\frac{1}{\alpha}}}{k(k+1)^{\frac{1}{\alpha}} - k^{\frac{1}{\alpha}}(k+1)}. \tag{45}$$

Furthermore, the upper bound on $\varepsilon_{X|Y}$ as a function of $H_\alpha(X|Y)$ is asymptotically tight in the limit where $\alpha \to \infty$.

Proof: See [25, Theorem 12]. ∎

Remark 6: By letting $\alpha \to 1$ in the right side of (45), the bound by Ho and Verdú [15, (109)] is recovered:

$$\varepsilon_{X|Y} \le \frac{H(X|Y) + (k^2-1)\log(k+1) - k^2 \log k}{k(k+1)\log\left(\frac{k+1}{k}\right)} \tag{46}$$

if $\log k \le H(X|Y) < \log(k+1)$ for an arbitrary $k \in \mathbb{N}$.

ACKNOWLEDGMENT

This work has been supported by the Israeli Science Foundation (ISF) under Grant 12/12, by ARO-MURI contract number W911NF-15-1-0479 and in part by the Center for Science of Information, an NSF Science and Technology Center under Grant CCF-0939370.

REFERENCES

[1] R. Ahlswede, P. Gács and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, no. 2, pp. 157–177, 1976 (correction in vol. 39, no. 4, pp. 353–354, 1977).

[2] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory - 2nd Colloquium*, Keszthely, Hungary, 1975, Colloquia Mathematica Societatis Janós Bolyai (I. Csiszár and P. Elias editors), Amsterdam, the Netherlands: North Holland, vol. 16, pp. 41–52, 1977.

[3] L. Baladová, "Minimum of average conditional entropy for given mimimum probability of error," *Kybernetika*, vol. 2, no. 5, pp. 416–422, 1966.

[4] M. Ben-Bassat and J. Raviv, "Rényi's entropy and probability of error," *IEEE Trans. on Information Theory*, vol. 24, no. 3, pp. 324–331, May 1978.

[5] P. N. Chen and F. Alajaji, "A generalized Poor-Verdú error bound for multihypothesis testing," *IEEE Trans. on Information Theory*, vol. 58, no. 1, pp. 311–316, January 2012.

[6] J. Chu and J. Chueh, "Inequalities between information measures and error probability," *Journal of the Franklin Institute*, vol. 282, no. 2, pp. 121–125, August 1966.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, Second edition, 2006.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Second edition, Cambridge University Press, 2011.

[9] P. A. Devijver, "On a new class of bounds on Bayes risk in multihypothesis pattern recognition," *IEEE Trans. on Computers*, vol. C-23, no. 1, pp. 70–80, January 1974.

[10] R. M. Fano, *Class Notes for Course 6.574: Transmission of Information*, MIT, Cambridge, MA, USA, 1952.

[11] M. Feder and N. Merhav, "Relations between entropy and error probability," *IEEE Trans. on Information Theory*, vol. 40, no. 1, pp. 259–266, January 1994.

[12] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Trans. on Information Theory*, vol. 60, no. 11, pp. 6801–6810, November 2014.

[13] T. S. Han and S. Verdú, "Generalizing the Fano inequality," *IEEE Trans. on Information Theory*, vol. 40, no. 4, pp. 1247–1251, July 1994.

[14] M. E. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. on Information Theory*, vol. 16, no. 4, pp. 368–372, July 1970.

[15] S. W. Ho and S. Verdú, "On the interplay between conditional entropy and error probability," *IEEE Trans. on Information Theory*, vol. 56, no. 12, pp. 5930–5942, December 2010.

[16] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. on Communication Technology*, vol. 15, no. 1, pp. 52–60, February 1967.

[17] F. Kanaya and T. S. Han, "The asymptotics of posterior entropy and error probability for Bayesian estimation," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1988–1992, November 1995.

[18] Y. H. Kim, A. Sutivong and T. M. Cover, "State amplification," *IEEE Trans. on Information Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.

[19] V. A. Kovalevsky, "The problem of character recognition from the point of view of mathematical statistics," in *Reading Automata and Pattern Recognition* (in Russian), Naukova Dumka, Kyjev, ed. 1965.

[20] A. Lapidoth and C. Pfister, "Two measures of dependence," *Proceedings of the 2016 IEEE International Conference on the Science of Electrical Engineering*, Eilat, Israel, November 16–18, 2016.

[21] Y. Polyanskiy, H. V. Poor and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[22] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," *Proceedings of the Forty-Eighth Annual Allerton Conference on Communication, Control and Computing*, pp. 1327–1333, Urbana-Champaign, Illinois, USA, October 2010.

[23] H. V. Poor and S. Verdú, "A lower bound on the probability of error in multihypothesis testing," *IEEE Trans. on Information Theory*, vol. 41, no. 6, pp. 1992–1994, November 1995.

[24] A. Rényi, "On measures of entropy and information," *Proceedings of the Fourth Berkeley Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berkeley, California, USA, 1961.

[25] I. Sason and S. Verdú, "Arimoto-Rényi conditional entropy and Bayesian M-ary hypothesis testing," submitted to the *IEEE Trans. on Information Theory*, September 2016. Available: https://arxiv.org/abs/1701.01974.

[26] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July-October 1948.

[27] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, vol. 1, pp. 6–25, September 1957.

[28] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal on Research and Development*, vol. 2, no. 4, pp. 289–293, October 1958.

[29] O. Shayevitz, "On Rényi measures and hypothesis testing," *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pp. 800–804, Saint Petersburg, Russia, August 2011.

[30] D. Tebbe and S. Dwyer, "Uncertainty and the probability of error," *IEEE Trans. on Information Theory*, vol. 14, no. 3, pp. 516–518, May 1968.

[31] M. Tomamichel and M. Hayashi, "Correlation detection and an operational interpretation of the Rényi mutual information," *Proceedings of the 2015 IEEE International Symposium on Information Theory*, pp. 1447–1451, Hong Kong, China, June 2015.

[32] M. Tomamichel and M. Hayashi, "Operational interpretation of Rényi conditional mutual information via composite hypothesis testing against Markov distributions," *Proc. of the 2016 IEEE International Symposium on Information Theory*, pp. 585–589, Barcelona, Spain, July 2016.

[33] I. Vajda, "Bounds on the minimal error probability on checking a finite or countable number of hypotheses," (in Russian), *Problems of Information Transmission*, vol. 4, no. 1, pp. 9–19, 1968.

[34] S. Verdú, "Shannon's inequality," *2011 Workshop on Information Theory and Applications*, San Diego, California, USA, February 2011.