Improved Bounds on Guessing Moments via Rényi Measures

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 provides upper and lower bounds on the optimal guessing moments of a random variable taking values on a finite set when side information may be available. These moments quantify the number of guesses required for correctly identifying the unknown object and, similarly to Arikan's bounds, they are expressed in terms of the Arimoto-Rényi conditional entropy. Although Arikan's bounds are asymptotically tight, the improvement of the bounds in this paper is significant in the non-asymptotic regime. Relationships between moments of the optimal guessing function and the MAP error probability are provided, characterizing the exact locus of their attainable values.**

Index Terms – Guessing moments, MAP decision rules, M-ary hypothesis testing, error probability, Rényi information measures.

I. Introduction

The problem of guessing discrete random variables has found a variety of applications in information theory, coding theory, cryptography, and searching and sorting algorithms. The central object of interest is the distribution of the number of guesses required to identify a realization of a random variable X, taking values on a finite or countably infinite set $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$, by asking questions of the form "Is X equal to x?". A *guessing function* is a one-to-one function $g\colon \mathcal{X} \to \mathcal{X}$, which can be viewed as a permutation of the elements of $\mathcal{X}$ in the order in which they are guessed. We can envision a generic algorithm that outputs $g^{-1}(1)$; a supervisor checks whether $X = g^{-1}(1)$, if so then the algorithm halts; otherwise, the algorithm outputs $g^{-1}(2)$ and the process repeats until the value of X is guessed correctly. Therefore, the number of guesses is $g(x)$ when the true outcome is $x \in \mathcal{X}$.

Lower and upper bounds on the minimal expected number of required guesses for correctly identifying the realization of X, expressed as a function of the Shannon entropy $H(X)$, have been respectively derived by Massey [6] and by McEliece and Yu [7]. More generally, given a probability mass function P_X on $\mathcal{X}$, it is of interest to minimize the generalized guessing moment

$$\mathbb{E}[g^\rho(X)] = \sum_{x \in \mathcal{X}} P_X(x) g^\rho(x), \quad \rho > 0. \tag{1}$$

For an arbitrary positive ρ, the ρ-th moment of the number of guesses is minimized by selecting the guessing function to be a *ranking function* g_X, for which $g_X(x) = k$ if $P_X(x)$ is the k-th largest mass. Upper and lower bounds on the ρ-th moment of ranking functions, expressed in terms of the Rényi entropy $H_\alpha(X)$ of order $\alpha = \frac{1}{1+\rho}$, were derived by Arikan [1], followed by a refined upper bound by Boztaş [3]. Although if $|\mathcal{X}|$ is small, it is straightforward to evaluate numerically the guessing moments, the benefit of bounds expressed in terms of Rényi entropies is particularly relevant when dealing with a random vector $X^n = (X_1, \ldots, X_n)$ whose letters belong to a finite alphabet $\mathcal{A}$; computing all the probabilities of the mass function P_{X^n} over the set $\mathcal{A}^n$, and then sorting them in decreasing order for the calculation of the ρ-th moment of the optimal guessing function for the elements of $\mathcal{A}^n$ has exponential complexity in n. Therefore, it becomes infeasible even for moderate values of n. In contrast, regardless of the value of n, bounds on guessing moments which depend on the Rényi entropy are readily computable if for example $\{X_i\}_{i=1}^n$ are independent; in which case, the Rényi entropy of the vector is equal to the sum of the Rényi entropies of its components (hence, the exponential complexity is reduced to linear complexity in n; furthermore, in the i.i.d. case, the complexity of computing the Rényi entropy of X^n is independent of n). Arikan's bounds are asymptotically tight for random vectors of length n as $n \to \infty$, so another benefit of these bounds is that they provide the correct exponential growth rate of the guessing moments for sufficiently large n. In [1], Arikan generalized his bounds to allow side information, leading to asymptotically tight bounds which are expressed in terms of the Arimoto-Rényi conditional entropy [2].

Section II defines the Rényi information measures. Section III provides upper and lower bounds on the minimal guessing moments of a random variable taking a finite number of values where side information on its value may be available. In the non-asymptotic regime, these bounds improve earlier results by Arikan [1] and Boztaş [3]. Section III also provides tight lower and upper bounds which establish relationships between the MAP error probability in M-ary hypothesis testing, and the moments of the optimal guessing function for correctly identifying X when side information Y is available. Due to space limitations, all proofs appear in the full paper version [11].

II. Preliminaries

The information measures used in this paper apply to discrete random variables.

Definition 1: [8] 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

$$H_\alpha(X) = \frac{1}{1-\alpha}\,\log \sum_{x\in\mathcal{X}} P_X^\alpha(x). \tag{2}$$

By its continuous extension,

$$H_0(X) = \log\big|\{x\in\mathcal{X}\colon P_X(x)>0\}\big|, \tag{3}$$
$$H_1(X) = H(X), \tag{4}$$
$$H_\infty(X) = \log\frac{1}{p_{\max}} \tag{5}$$

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

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

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1}\,\log\Big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\Big). \tag{6}$$

Furthermore, for $\alpha=1$, $d_\alpha(p\|q)$ is equal to the binary relative entropy:

$$d_1(p\|q) = d(p\|q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}. \tag{7}$$

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

$$H_\alpha(X|Y)$$
$$= \frac{\alpha}{1-\alpha}\,\log\,\mathbb{E}\left[\left(\sum_{x\in\mathcal{X}} P_{X|Y}^\alpha(x|Y)\right)^{\frac{1}{\alpha}}\right] \tag{8}$$
$$= \frac{\alpha}{1-\alpha}\,\log\sum_{y\in\mathcal{Y}} P_Y(y)\,\exp\left(\frac{1-\alpha}{\alpha}\,H_\alpha(X|Y=y)\right), \tag{9}$$

  where (9) 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

$$H_0(X|Y) = \sup_{y\in\mathcal{Y}} H_0(X\,|\,Y=y), \tag{10}$$
$$H_1(X|Y) = H(X|Y), \tag{11}$$
$$H_\infty(X|Y) = -\log\,\mathbb{E}\Big[\max_{x\in\mathcal{X}} P_{X|Y}(x|Y)\Big] \tag{12}$$

  where (10) applies if Y is a discrete random variable.

Properties of the Arimoto-Rényi conditional entropy were studied in [5], [9] and [10].

As in [10, Section 4], we find several useful results satisfied by the Arimoto-Rényi conditional entropy of negative orders.

III. Improved Bounds on Guessing Moments

This section provides improved upper and lower bounds on the guessing moments of a discrete random variable. The upper bounds correspond to the case where the guessing function is a ranking function. These bounds are proved in [11, Section 3].

A. Key result

Theorem 1: Given a discrete random variable X taking values on a set $\mathcal{X}$, a function $g\colon\mathcal{X}\to(0,\infty)$, and a scalar $\rho\neq 0$, then

1)

$$\sup_{\beta\in(-\rho,+\infty)\setminus\{0\}} \frac{1}{\beta}\left[H_{\frac{\beta}{\beta+\rho}}(X) - \log\sum_{x\in\mathcal{X}} g^{-\beta}(x)\right]$$
$$\le \frac{1}{\rho}\log\mathbb{E}[g^\rho(X)] \tag{13}$$
$$\le \inf_{\beta\in(-\infty,-\rho)\setminus\{0\}} \frac{1}{\beta}\left[H_{\frac{\beta}{\beta+\rho}}(X) - \log\sum_{x\in\mathcal{X}} g^{-\beta}(x)\right]. \tag{14}$$

2) For $\tau\in\mathbb{R}$, define the probability mass function

$$Q_\tau(x) = \frac{g^{-\tau}(x)}{\sum\limits_{a\in\mathcal{X}} g^{-\tau}(a)}, \quad x\in\mathcal{X}, \tag{15}$$

provided that the sum in the right side of (15) is finite. The following results hold:

a) If $P_X = Q_\rho$ and $\mathcal{X}$ is a finite set, then

$$\frac{1}{\rho}\log\mathbb{E}[g^\rho(X)] = -\frac{1}{\rho}\log\left(\frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} g^{-\rho}(x)\right). \tag{16}$$

b) If $P_X = Q_\nu$ with $\nu>0$ and $\nu\neq\rho$, then the supremum in the left side of (13) is attained at $\beta=\nu-\rho$, and the inequality in (13) is an identity. Conversely, if (13) is an identity and the supremum is attained at $\beta^*\in(-\rho,+\infty)\setminus\{0\}$, then $P_X = Q_{\rho+\beta^*}$.

c) If $P_X = Q_\nu$ with $\nu<0$ and $\nu\neq\rho$, then the infimum in the right side of (14) is attained at $\beta=\nu-\rho$, and the inequality in (14) is an identity. Conversely, if (14) is an identity and the infimum is attained at $\beta^*\in(-\infty,-\rho)\setminus\{0\}$, then $P_X = Q_{\rho+\beta^*}$.

Remark 1: For $\rho>0$, the supremum over β in the right side of (13) involves negative orders of the Rényi entropy whenever $\beta\in(-\rho,0)$. The optimal value of $\beta\in(-\rho,\infty)\setminus\{0\}$ can be negative; furthermore, for every such β, Theorem 1 asserts the existence of a probability mass function for which (13) is achieved with equality. Allowing Rényi entropy of negative orders in Theorem 1 is therefore beneficial.

The particularization of Item 1) in Theorem 1 to $\beta=1$ yields the following, generally looser, bound:

Corollary 1: [4, Lemma 2] Let X and g be as in Theorem 1, and $\rho\in(-1,0)\cup(0,\infty)$. Then,

$$\frac{1}{\rho}\,\log\mathbb{E}\big[g^\rho(X)\big] \ge H_{\frac{1}{1+\rho}}(X) - \log\sum_{x\in\mathcal{X}}\frac{1}{g(x)}. \tag{17}$$

B. Lower bounds

Theorem 1 enables to derive lower bounds on guessing moments with or without side information, improving the bounds in [1].

Theorem 2: Let X be a random variable taking values on the finite set $\mathcal{X} = \{1, \dots, M\}$, and let $g\colon \mathcal{X} \to \mathcal{X}$ be an arbitrary guessing function. Then, for every $\rho \neq 0$,

$$\frac{1}{\rho} \log \mathbb{E}\big[g^\rho(X)\big] \geq \sup_{\beta \in (-\rho, \infty) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log u_M(\beta) \right] \tag{18}$$

with

$$u_M(\beta) = \begin{cases} \log_{\mathrm{e}} M + \gamma + \dfrac{1}{2M} - \dfrac{5}{6(10M^2+1)} & \beta = 1, \\ \min\left\{ \zeta(\beta) - \frac{(M+1)^{1-\beta}}{\beta-1} - \frac{(M+1)^{-\beta}}{2}, \, u_M(1) \right\} & \beta > 1, \\ 1 + \dfrac{1}{1-\beta} \left[\left(M + \frac{1}{2}\right)^{1-\beta} - \left(\frac{3}{2}\right)^{1-\beta} \right] & |\beta| < 1, \\ \dfrac{M^{1-\beta} - 1}{1-\beta} + \frac{1}{2}\left(1 + M^{-\beta}\right) & \beta \leq -1 \end{cases} \tag{19}$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant, and $\zeta(\beta) = \sum_{n=1}^{\infty} \frac{1}{n^\beta}$ is the Riemann zeta function for $\beta > 1$.

Remark 2: Specializing Theorem 2 to $\beta = 1$ and using $u_M(1) \leq 1 + \log_{\mathrm{e}} M$ for $M \geq 2$, we obtain

$$\frac{1}{\rho} \log \mathbb{E}\big[g^\rho(X)\big] \geq H_{\frac{1}{1+\rho}}(X) - \log\big(1 + \log_{\mathrm{e}} M\big) \tag{20}$$

for $\rho \in (-1, \infty)$. This bound was obtained in the range $\rho > 0$ by Arikan [1, (1)].

The following remark justifies the utility of Theorem 2.

Remark 3: Since Theorem 1 also applies to guessing functions, it gives a lower bound on $\frac{1}{\rho} \log \mathbb{E}\big[g^\rho(X)\big]$ where $u_M(\beta)$ in (18)–(19) is replaced by $\sum_{j=1}^{M} \frac{1}{j^\beta}$ for $\beta \in (-\rho, \infty) \setminus \{0\}$. While numerical evidence shows that it is slightly better than the bound in Theorem 2 for large M, the latter bound is much easier to compute if M is large.

C. Upper bounds

The average number of guesses is minimized by taking the guessing function to be the ranking function g_X, for which $g_X(x) = k$ if $P_X(x)$ is the k-th largest mass [6]. Although the tie breaking affects the choice of g_X, the distribution of $g_X(X)$ does not depend on how ties are resolved. Not only does this strategy minimize the average number of guesses, but it also minimizes the ρ-th moment of the number of guesses for every $\rho > 0$.

Theorem 3: [1, Proposition 4] Let X be a discrete random variable taking values on a set $\mathcal{X}$, and let g_X be the ranking function according to P_X. Then, for all $\rho > 0$,

$$\mathbb{E}[g_X^\rho(X)] \leq \exp\Big(\rho H_{\frac{1}{1+\rho}}(X)\Big). \tag{21}$$

The following results tighten Theorem 3.

Theorem 4: Under the assumptions in Theorem 3, for all $\rho \geq 0$,

$$\mathbb{E}[g_X^\rho(X)] \leq \frac{1}{1+\rho} \left[\exp\Big(\rho H_{\frac{1}{1+\rho}}(X)\Big) - 1 \right] + \exp\Big((\rho-1)^+ H_{\frac{1}{\rho}}(X)\Big) \tag{22}$$

where $(x)^+ \triangleq \max\{x, 0\}$ for $x \in \mathbb{R}$.

In the range $\rho \in [0, 2]$, we can tighten (22) according to the following result.

Theorem 5: Under the assumptions in Theorem 3,

a) For $\rho \in [0, 1]$

$$\mathbb{E}[g_X^\rho(X)] \leq \frac{1}{1+\rho} \exp\Big(\rho H_{\frac{1}{1+\rho}}(X)\Big) + \frac{\rho - (1-\rho)(2^\rho - 1)(1 - p_{\max})}{1+\rho}. \tag{23}$$

b) For $\rho \in [1, 2]$

$$\mathbb{E}[g_X^\rho(X)] \leq \frac{1}{1+\rho} \exp\Big(\rho H_{\frac{1}{1+\rho}}(X)\Big) + \frac{1}{\rho} \exp\Big((\rho-1) H_{\frac{1}{\rho}}(X)\Big) + \frac{\rho^2 - \rho - 1}{\rho(1+\rho)}. \tag{24}$$

Furthermore, both (23) and (24) hold with equality if X is deterministic.

Remark 4: Particularizing (24) to $\rho = 1$ and $\rho = 2$, we recover the bounds on the first and second moments in [3, Theorem 3]. Furthermore, the bounds in (23) and (24) provide a continuous transition at $\rho = 1$.

Theorem 6: Under the setting in Theorem 3, for $\rho \geq 2$,

$$\mathbb{E}[g_X^\rho(X)] \leq 1 + \sum_{j=0}^{\lfloor \rho \rfloor} c_j(\rho) \left[\exp\Big((\rho - j)\, H_{\frac{1}{1+\rho-j}}(X) \Big) - 1 \right], \tag{25}$$

where $\{c_j(\rho)\}$ is given by

$$c_j(\rho) = \begin{cases} \dfrac{1}{1+\rho} & j = 0 \\ \frac{1}{2} & j = 1 \\ \dfrac{\rho \dots (\rho - j + 2)}{2^j} & j \in \{2, \dots, \lfloor \rho \rfloor - 1\} \\ \dfrac{\rho \dots (\rho - j + 2)}{2^{j-1}\,(\rho - j + 1)} & j = \lfloor \rho \rfloor \end{cases} \tag{26}$$

and $\lfloor x \rfloor$ denotes the largest integer that is smaller than or equal to x.

Remark 5: In contrast to [3, Theorem 3], Theorems 5 and 6 provide an explicit upper bound on $\mathbb{E}[g_X^\rho(X)]$ for $\rho > 0$ as a function of Rényi entropies of X. Note also that the upper bounds in (24) and (25) coincide at $\rho = 2$.

Remark 6: Numerical evidence shows that none of the bounds in (22) and (25) supersedes the other for $\rho > 2$.

Example 1: Let X be geometrically distributed restricted to $\{1, \ldots, M\}$ with the probability mass function

$$P_X(k) = \frac{(1-a)\, a^{k-1}}{1-a^M}, \quad k \in \{1, \ldots, M\} \tag{27}$$

where $a = 0.9$ and $M = 32$. Table I compares $\frac{1}{3}\log_{\mathrm{e}} \mathbb{E}[g_X^3(X)]$ to its various lower and upper bounds (LBs and UBs, respectively). Notice that in this example, the upper bound in (25) improves the bound in (22).

TABLE I
COMPARISON OF $\frac{1}{3}\log_{\mathrm{e}} \mathbb{E}[g_X^3(X)]$ AND BOUNDS IN EXAMPLE 1.

(20) LB	Theorem 2 LB	$\frac{1}{3}\log_{\mathrm{e}} \mathbb{E}[g_X^3(X)]$ exact value	(25) UB	(22) UB	(21) UB
1.864	2.593	2.609	2.920	2.939	3.360

D. Bounds on guessing moments with side information

This subsection extends the lower and upper bounds in Sections III-B and III-C to allow side information Y for guessing the value of X. These bounds tighten the results in [1, Theorem 1] and [1, Proposition 4] for all $\rho > 0$.

Theorem 7: Let X and Y be discrete random variables taking values on the sets $\mathcal{X} = \{1, \ldots, M\}$ and $\mathcal{Y}$, respectively. For all $y \in \mathcal{Y}$, let $g(\cdot|y)$ be a guessing function of X given that $Y = y$. Then, for $\rho \in (0, \infty)$,

$$\begin{aligned} &\frac{1}{\rho}\, \log \mathbb{E}\big[g^\rho(X|Y)\big] \\ &\geq \sup_{\beta \in (-\rho, 0) \cup (0, \infty)} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X|Y) - \log u_M(\beta) \right] \end{aligned} \tag{28}$$

with $u_M(\cdot)$ as defined in (19).

Theorem 8: Let X and Y be discrete random variables taking values on sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. For all $y \in \mathcal{Y}$, let $g_{X|Y}(\cdot|y)$ be a ranking function of X given that $Y = y$. Then, for $\rho \in (0, \infty)$, the upper bounds in Theorems 4, 5 and 6 hold for $\mathbb{E}[g_{X|Y}^\rho(X|Y)]$ by replacing the Rényi entropy $H_\alpha(X)$ with the conditional Rényi entropy $H_\alpha(X|Y)$ for $\alpha > 0$.

E. Guessing moments and the minimum probability of error

Let X and Y be discrete random variables,[1] taking values on the sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. The minimum probability of error of X given Y, denoted by $\varepsilon_{X|Y}$, is achieved by the maximum-a-posteriori (MAP) decision rule. Hence,

$$\varepsilon_{X|Y} = \sum_{y \in \mathcal{Y}} P_Y(y) \left[1 - \max_{x \in \mathcal{X}} P_{X|Y}(x|y) \right]. \tag{29}$$

[1]The assumption that Y is a discrete random variable can be easily dispensed with.

In contrast, the moments of the ranking function $\mathbb{E}[g_{X|Y}^\rho(X|Y)]$ quantify the number of guesses required for correctly identifying the unknown object X on the basis of Y. It is therefore natural to establish relationships between both quantities. First note that by definition,

$$1 - \varepsilon_{X|Y} = \mathbb{P}[g_{X|Y}(X|Y) = 1]. \tag{30}$$

In [10], we derived lower and upper bounds on $\varepsilon_{X|Y}$ as a function of $H_\alpha(X|Y)$ for an arbitrary order α. In this section, Theorems 7–8 provide lower and upper bounds on guessing moments of a ranking function $g_{X|Y}(X|Y)$ as a function of Arimoto-Rényi conditional entropies. As a natural continuation to these studies, we derive tight lower and upper bounds on $\mathbb{E}[g_{X|Y}^\rho(X|Y)]$ as a function of $\varepsilon_{X|Y}$.

Theorem 9: Let X and Y be discrete random variables taking values on sets $\mathcal{X} = \{1, \ldots, M\}$ and $\mathcal{Y}$, respectively. Then, for $\rho > 0$,

$$f_\rho(\varepsilon_{X|Y}) \leq \mathbb{E}[g_{X|Y}^\rho(X|Y)] \tag{31}$$

$$\leq 1 + \left(\tfrac{1}{M-1} \sum_{j=2}^{M} j^\rho - 1 \right) \varepsilon_{X|Y} \tag{32}$$

where the function $f_\rho \colon [0,1) \to [0, \infty)$ is given by

$$f_\rho(u) = (1-u) \sum_{j=1}^{k_u} j^\rho + [1 - (1-u)k_u](k_u+1)^\rho, \tag{33}$$

$$k_u = \left\lfloor \frac{1}{1-u} \right\rfloor. \tag{34}$$

The lower and upper bounds in (31) and (32) are tight:

- Let $p_{\max}(y) = \max\limits_{x \in \mathcal{X}} P_{X|Y}(x|y)$ for $y \in \mathcal{Y}$. The lower bound is attained if and only if $p_{\max}(y) = p_{\max}$ is fixed for all $y \in \mathcal{Y}$, and conditioned on $Y = y$, X has $\left\lfloor \frac{1}{p_{\max}} \right\rfloor$ masses equal to $p_{\max}$, and an additional mass equal to $1 - p_{\max} \left\lfloor \frac{1}{p_{\max}} \right\rfloor$ whenever $\frac{1}{p_{\max}}$ is not an integer.
- The upper bound is attained if and only if regardless of $y \in \mathcal{Y}$, conditioned on $Y = y$, X is equiprobable among its $M-1$ conditionally least likely values on $\mathcal{X}$.

Remark 7: The lower and upper bounds in (31) and (32) coincide in each of the extreme cases $\varepsilon_{X|Y} = 0$ or $1 - \frac{1}{M}$.

In view of Theorems 2 and 9, the next result provides an explicit lower bound on $\varepsilon_{X|Y}$ as a function of $H_\alpha(X|Y)$ for any non-zero $\alpha < 1$.

Theorem 10: Let X and Y be discrete random variables taking values on sets $\mathcal{X} = \{1, \ldots, M\}$ and $\mathcal{Y}$, respectively. Then, for all $\alpha \in (-\infty, 0) \cup (0, 1)$,

$$\varepsilon_{X|Y} \geq \sup_{\rho > 0} \left\{ \frac{\exp\left(\left(\frac{1}{\alpha} - 1\right) \left[H_\alpha(X|Y) - \log u_M\left(\frac{\alpha\rho}{1-\alpha}\right) \right] \right) - 1}{\frac{1}{M-1} \sum\limits_{j=2}^{M} j^\rho - 1} \right\} \tag{35}$$

with $u_M(\cdot)$ as defined in (19).

In [10], we derived the following lower bounds on $\varepsilon_{X|Y}$ as a function of $H_\alpha(X|Y)$:

1) $\alpha > 0$: A generalization of Fano's inequality in [10, Theorem 3] is given by

$$H_\alpha(X|Y) \leq \log M - d_\alpha\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big) \tag{36}$$

with $d_\alpha(\cdot\|\cdot)$ as defined in (6).

2) $\alpha < 0$: An explicit lower bound in [10, Theorem 6] is given by

$$\varepsilon_{X|Y} \geq \exp\left(\frac{1-\alpha}{\alpha}\Big[H_\alpha(X|Y) - \log(M-1)\Big]\right). \tag{37}$$

Remark 8: Shannon's inequality [12] (see also [13]) gives an explicit lower bound on $\varepsilon_{X|Y}$ as a function of $H(X|Y)$ when M is finite:

$$\varepsilon_{X|Y} \geq \frac{1}{6}\,\frac{H(X|Y)}{\log M + \log\log M - \log H(X|Y)}, \tag{38}$$

where, in the right side of (38), the base of the logarithm and the units of the conditional entropy must be equal (and can be arbitrary). The bound in (35) becomes trivial in the limit where $\alpha \uparrow 1$ since, for any fixed $\rho > 0$, (19) implies that $u_M\Big(\frac{\alpha\rho}{1-\alpha}\Big) \to 1$, and therefore the lower bound on $\varepsilon_{X|Y}$ tends to zero in this case. Nevertheless, numerical experimentation shows that the convergence of this bound in (35) to zero is only affected by values of α very close to 1, as it is illustrated in Example 2 with a comparison to Shannon's bound in (38).

Example 2: Let X and Y be random variables taking values on $\mathcal{X} = \{1,2,3,4\}$, and let

$$\big[P_{XY}(x,y)\big]_{(x,y)\in\mathcal{X}^2} = \frac{1}{100}\begin{pmatrix} 9 & 3 & 4 & 9 \\ 9 & 9 & 3 & 4 \\ 4 & 9 & 9 & 3 \\ 3 & 4 & 9 & 9 \end{pmatrix}. \tag{39}$$

It can be verified from (29) that $\varepsilon_{X|Y} = \frac{16}{25} = 0.640$. Table II

TABLE II
EXAMPLE 2: LOWER BOUNDS ON $\varepsilon_{X|Y}$.

α	(35)	(36)	(37)
-1	0.463	–	0.447
$-\frac{1}{2}$	0.475	–	0.355
$-\frac{1}{4}$	0.482	–	0.206
$\frac{1}{5}$	0.494	0.523	–
$\frac{1}{2}$	0.502	0.530	–
$\frac{4}{5}$	0.510	0.536	–

shows a slight advantage of the lower bound in (36) over (35) for $\alpha \in (0,1)$, and a superiority of the lower bound in (35) over (37) for some negative values of α.

In view of Remark 8, the lower bound in (35) for α close to 1 is compared with Shannon's lower bound in (38). For $\alpha = 0.99$, the lower bound in (35) is equal to 0.515 (note that it is slightly looser than (36), which is equal to 0.540); on the other hand, the lower bound in (38) is equal to 0.146.

Theorem 9 establishes relationships between the ρ-th moment of the optimal guessing function, for fixed $\rho > 0$, and the MAP error probability. This characterizes the exact locus of their attainable values. The following result suggests an easy way to determine the MAP error probability on the basis of the knowledge of these ρ-th moments at an arbitrarily small right neighborhood of $\rho = 0$.

Theorem 11: Let X and Y be discrete random variables taking values on sets $\mathcal{X} = \{1,\ldots,M\}$ and $\mathcal{Y}$, respectively. For an integer $k \geq 0$, let $z_k = \frac{\mathrm{d}^k}{\mathrm{d}\rho^k}\,\mathbb{E}[g^\rho_{X|Y}(X|Y)]\Big|_{\rho=0}$. Then,

$$\varepsilon_{X|Y} = 1 - \frac{1}{c_M}\begin{vmatrix} z_0 & 1 & \cdots & 1 \\ z_1 & \log_{\mathrm{e}} 2 & \cdots & \log_{\mathrm{e}} M \\ \vdots & \vdots & \vdots & \vdots \\ z_{M-1} & \log_{\mathrm{e}}^{M-1} 2 & \cdots & \log_{\mathrm{e}}^{M-1} M \end{vmatrix}$$

with

$$c_M = \begin{cases} \log_{\mathrm{e}} 2, & M = 2, \\ \displaystyle\prod_{k=2}^{M} \log_{\mathrm{e}} k \prod_{2\leq i<j\leq M} \log_{\mathrm{e}}\Big(\frac{j}{i}\Big), & M \geq 3. \end{cases} \tag{40}$$

REFERENCES

[1] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 99–105, January 1996.

[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] S. Boztaş, "Comments on "An inequality on guessing and its application to sequential decoding"," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 2062–2063, November 1997.

[4] T. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2494–2498, Honolulu, Hawaii, USA, July 2014.

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

[6] J. L. Massey, "Guessing and entropy," *Proceedings of the 1994 IEEE International Symposium on Information Theory*, p. 204, Trondheim, Norway, June 1994.

[7] R. J. McEliece and Z. Yu, "An inequality on entropy," *Proceedings of the 1995 IEEE International Symposium on Information Theory*, p. 329, Whistler, Canada, September 1995.

[8] A. Rényi, "On measures of entropy and information," *Proceedings of the 4th Berkeley Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berkeley, California, USA, 1961.

[9] Y. Sakai and K. Iwata, "Sharp bounds on Arimoto's conditional Rényi entropies between two distinct orders," *Proceedings of the 2017 IEEE International Symposium on Information Theory*, pp. 2985–2989, Aachen, Germany, June 2017.

[10] I. Sason and S. Verdú, "Arimoto-Rényi conditional entropy and Bayesian M-Ary hypothesis testing," *IEEE Trans. on Information Theory*, vol. 64, no. 1, pp. 4–25, January 2018.

[11] I. Sason and S. Verdú, "Improved bounds on lossless source coding and guessing moments via Rényi measures," *IEEE Trans. on Information Theory*, vol. 64, no. 6, pp. 4323–4346, June 2018.

[12] 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.

[13] S. Verdú, "Shannon's inequality," *2011 Workshop on Information Theory and Applications*, San Diego, California, USA, February 2011. [Online]. Available: http://ita.ucsd.edu/workshop/11/files/paper/paper_374.pdf.