Arimoto-Rényi Conditional Entropy and Bayesian M-ary Hypothesis Testing

Igal Sason (Technion) Sergio Verdú (Princeton)

Department of Electrical Engineering
Technion, Israel
November 23rd, 2017

Hypothesis Testing

- Bayesian M-ary hypothesis testing:
  - X is a random variable taking values on $\mathcal{X}$ with $|\mathcal{X}| = M$;
  - a prior distribution P_X on $\mathcal{X}$;
  - M hypotheses for the $\mathcal{Y}$-valued data $\{P_{Y|X=m}, m \in \mathcal{X}\}$.

Hypothesis Testing

- Bayesian M-ary hypothesis testing:
  - X is a random variable taking values on $\mathcal{X}$ with $|\mathcal{X}| = M$;
  - a prior distribution P_X on $\mathcal{X}$;
  - M hypotheses for the $\mathcal{Y}$-valued data $\{P_{Y|X=m}, m \in \mathcal{X}\}$.
- $\varepsilon_{X|Y}$: the minimum probability of error of X given Y
  - achieved by the *maximum-a-posteriori* (MAP) decision rule. Hence,

$$\varepsilon_{X|Y} = \mathbb{E}\left[1 - \max_{x \in \mathcal{X}} P_{X|Y}(x|Y)\right] \tag{1}$$

$$= 1 - \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{X,Y}(x, y). \tag{2}$$

where (2) holds when Y is discrete.

Example

Let X and Y be random variables defined on the set $\mathcal{A} = \{1, 2, 3\}$, and let

$$\left[P_{XY}(x,y)\right]_{(x,y)\in\mathcal{A}^2} = \frac{1}{45}\begin{pmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{pmatrix}. \tag{3}$$

Then,

$$\varepsilon_{X|Y} = 1 - \left(\tfrac{8}{45} + \tfrac{9}{45} + \tfrac{7}{45}\right) = \tfrac{7}{15}. \tag{4}$$

Interplay $\varepsilon_{X|Y} \longleftrightarrow$ information measures

- Bounds on $\varepsilon_{X|Y}$ involving information measures exist in the literature. Those works attest that there is a considerable motivation for studying the relationships between $\varepsilon_{X|Y}$ and information measures.
- $\varepsilon_{X|Y}$ is rarely directly computable, and the best bounds are information theoretic.

Interplay $\varepsilon_{X|Y} \longleftrightarrow$ information measures

- Bounds on $\varepsilon_{X|Y}$ involving information measures exist in the literature. Those works attest that there is a considerable motivation for studying the relationships between $\varepsilon_{X|Y}$ and information measures.
- $\varepsilon_{X|Y}$ is rarely directly computable, and the best bounds are information theoretic.
- Useful for
  - the analysis of M-ary hypothesis testing
  - proofs of coding theorems.

Interplay $\varepsilon_{X|Y} \longleftrightarrow$ information measures

- Bounds on $\varepsilon_{X|Y}$ involving information measures exist in the literature. Those works attest that there is a considerable motivation for studying the relationships between $\varepsilon_{X|Y}$ and information measures.
- $\varepsilon_{X|Y}$ is rarely directly computable, and the best bounds are information theoretic.
- Useful for
  - the analysis of M-ary hypothesis testing
  - proofs of coding theorems.
- In this talk, we introduce:

upper and lower bounds on $\varepsilon_{X|Y}$ in terms of the *Arimoto-Rényi* conditional entropy $H_\alpha(X|Y)$ of any order α, and apply them in coding.

The Rényi Entropy

Definition

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

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X^\alpha(x) \tag{5}$$

By its continuous extension,

$$H_0(X) = \log \bigl| \{x \in \mathcal{X} \colon P_X(x) > 0\} \bigr|, \tag{6}$$

$$H_1(X) = H(X), \tag{7}$$

$$H_\infty(X) = \log \tfrac{1}{p_{\max}} \tag{8}$$

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

The Binary Rényi Divergence

Definition

For $\alpha \in (0,1) \cup (1,\infty)$, the binary Rényi divergence of order α is given by

$$d_\alpha(p\|q) = \frac{1}{\alpha - 1}\,\log\Big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\Big). \tag{9}$$

The Binary Rényi Divergence

Definition

For $\alpha \in (0,1) \cup (1,\infty)$, the binary Rényi divergence of order α is given by

$$d_\alpha(p\|q) = \frac{1}{\alpha - 1} \log\left(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\right). \tag{9}$$

$$\lim_{\alpha \uparrow 1} d_\alpha(p\|q) = d(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \tag{10}$$

Rényi Conditional Entropy ?

- If we mimic the definition of $H(X|Y)$ and define conditional Rényi entropy as
$$\sum_{y\in\mathcal{Y}} P_Y(y)\, H_\alpha(X|Y=y),$$
we find that, for $\alpha \neq 1$, the conditional version may be larger than $H_\alpha(X)$!

Rényi Conditional Entropy ?

- If we mimic the definition of $H(X|Y)$ and define conditional Rényi entropy as

$$\sum_{y \in \mathcal{Y}} P_Y(y)\, H_\alpha(X|Y=y),$$

  we find that, for $\alpha \neq 1$, the conditional version may be larger than $H_\alpha(X)$!

- To remedy this situation, Arimoto introduced a notion of conditional Rényi entropy, $H_\alpha(X|Y)$ (named Arimoto-Rényi conditional entropy), which is upper bounded by $H_\alpha(X)$.

The Arimoto-Rényi Conditional Entropy (cont.)

Definition

Let P_{XY} be defined on $\mathcal{X} \times \mathcal{Y}$, where X is a discrete random variable.

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

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}\left[\left(\sum_{x \in \mathcal{X}} P^{\alpha}_{X|Y}(x|Y)\right)^{\frac{1}{\alpha}}\right] \tag{11}$$

The Arimoto-Rényi Conditional Entropy (cont.)

Definition

Let P_{XY} be defined on $\mathcal{X} \times \mathcal{Y}$, where X is a discrete random variable.

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

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}\left[\left(\sum_{x\in\mathcal{X}} P^{\alpha}_{X|Y}(x|Y)\right)^{\frac{1}{\alpha}}\right] \tag{11}$$

$$= \frac{\alpha}{1-\alpha} \log \sum_{y\in\mathcal{Y}} P_Y(y) \exp\left(\frac{1-\alpha}{\alpha} H_\alpha(X|Y=y)\right), \tag{12}$$

where (12) applies if Y is a discrete random variable.

The Arimoto-Rényi Conditional Entropy (cont.)

- By its continuous extension,

$$H_0(X|Y) = \operatorname{ess\,sup} H_0\left(P_{X|Y}(\cdot|Y)\right) \tag{13}$$
$$= \max_{y \in \mathcal{Y}} H_0(X \,|\, Y = y), \tag{14}$$

$$H_1(X|Y) = H(X|Y), \tag{15}$$

$$H_\infty(X|Y) = \log \frac{1}{\mathbb{E}\left[\max_{x \in \mathcal{X}} P_{X|Y}(x|Y)\right]} \tag{16}$$

where (14) applies if Y is a discrete random variable.

The Arimoto-Rényi Conditional Entropy (cont.)

- By its continuous extension,

$$H_0(X|Y) = \operatorname{ess\,sup} H_0\left(P_{X|Y}(\cdot|Y)\right) \tag{13}$$

$$= \max_{y \in \mathcal{Y}} H_0(X \,|\, Y = y), \tag{14}$$

$$H_1(X|Y) = H(X|Y), \tag{15}$$

$$H_\infty(X|Y) = \log \frac{1}{\mathbb{E}\left[\max_{x \in \mathcal{X}} P_{X|Y}(x|Y)\right]} \tag{16}$$

where (14) applies if Y is a discrete random variable.

Monotonicity Properties

- $H_\alpha(X|Y)$ is monotonically decreasing in α throughout the real line.
- $\frac{\alpha-1}{\alpha} H_\alpha(X|Y)$ is monotonically increasing in α on $(0,\infty)$ & $(-\infty,0)$.

Fano's Inequality

Let X take values in $|\mathcal{X}| = M$, then

$$H(X|Y) \leq h(\varepsilon_{X|Y}) + \varepsilon_{X|Y} \log(M-1) \tag{17}$$

Fano's Inequality

Let X take values in $|\mathcal{X}| = M$, then

$$H(X|Y) \le h(\varepsilon_{X|Y}) + \varepsilon_{X|Y} \log(M-1) \tag{17}$$

$$= \log M - d\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big) \tag{18}$$

Fano's Inequality

Let X take values in $|\mathcal{X}| = M$, then

$$H(X|Y) \leq h(\varepsilon_{X|Y}) + \varepsilon_{X|Y} \log(M-1) \tag{17}$$

$$= \log M - d\left(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\right) \tag{18}$$

- (18) is not nearly as popular as (17);
- (18) turns out to be the version that admits an elegant (although not immediate) generalization to the Arimoto-Rényi conditional entropy.

Generalization of Fano's Inequality

- It is easy to get Fano's inequality by averaging $H(X|Y=y)$ with respect to the observation y:

$$H(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y)\, H(X|Y=y).$$

Generalization of Fano's Inequality

- It is easy to get Fano's inequality by averaging $H(X|Y=y)$ with respect to the observation y:

$$H(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y)\, H(X|Y=y).$$

- This simple route is not viable in the case of $H_\alpha(X|Y)$ since it is not an average of Rényi entropies of conditional distributions:

$$H_\alpha(X|Y) \neq \sum_{y \in \mathcal{Y}} P_Y(y)\, H_\alpha(X|Y=y), \quad \alpha \neq 1. \tag{19}$$

Generalization of Fano's Inequality

- It is easy to get Fano's inequality by averaging $H(X|Y=y)$ with respect to the observation y:

$$H(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y)\, H(X|Y=y).$$

- This simple route is not viable in the case of $H_\alpha(X|Y)$ since it is not an average of Rényi entropies of conditional distributions:

$$H_\alpha(X|Y) \neq \sum_{y \in \mathcal{Y}} P_Y(y)\, H_\alpha(X|Y=y), \quad \alpha \neq 1. \tag{19}$$

- The standard proof of Fano's inequality, also fails for $H_\alpha(X|Y)$ of order $\alpha \neq 1$ since it does not satisfy the chain rule.

Generalization of Fano's Inequality (cont.)

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 equality holds:

$$\varepsilon_{X|Y} = 1 - \exp\bigl(-H_\infty(X|Y)\bigr). \qquad (20)$$

Generalization of Fano's Inequality (cont.)

Lemma

Let $\alpha \in (0,1) \cup (1,\infty)$ *and* $(\beta,\gamma) \in (0,\infty)^2$. *Then,*

$$f_{\alpha,\beta,\gamma}(u) = \left(\gamma(1-u)^\alpha + \beta u^\alpha\right)^{\frac{1}{\alpha}}, \quad u \in [0,1] \tag{21}$$

is

- *strictly convex for* $\alpha \in (1,\infty)$;
- *strictly concave for* $\alpha \in (0,1)$.

$$f''_{\alpha,\beta,\gamma}(u) = (\alpha-1)\beta\gamma\Big(\gamma(1-u)^\alpha + \beta u^\alpha\Big)^{\frac{1}{\alpha}-2}\big(u(1-u)\big)^{\alpha-2} \tag{22}$$

which is strictly negative if $\alpha \in (0,1)$, and strictly positive if $\alpha \in (1,\infty)$.

Generalization of Fano's Inequality (cont.)

Theorem

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

$$H_\alpha(X|Y) \leq \log M - d_\alpha\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big). \tag{23}$$

Equality holds in (23) if and only if, for all y,

$$P_{X|Y}(x|y) = \begin{cases} \dfrac{\varepsilon_{X|Y}}{M-1}, & x \neq \mathcal{L}^*(y) \\ 1 - \varepsilon_{X|Y}, & x = \mathcal{L}^*(y) \end{cases} \tag{24}$$

where $\mathcal{L}^* \colon \mathcal{Y} \to \mathcal{X}$ is a deterministic MAP decision rule.

Generalization of Fano's Inequality (cont.)

If X, Y are vectors of dimension n, then $\varepsilon_{X|Y} \to 0 \;\Rightarrow\; \frac{1}{n} H(X|Y) \to 0$.
However, the picture with $H_\alpha(X|Y)$ is more nuanced !

Generalization of Fano's Inequality (cont.)

If X, Y are vectors of dimension n, then $\varepsilon_{X|Y} \to 0 \ \Rightarrow \ \frac{1}{n} H(X|Y) \to 0$.
However, the picture with $H_\alpha(X|Y)$ is more nuanced !

Theorem

Assume

- $\{X_n\}$ *is a sequence of random variables;*
- X_n *takes values on* $\mathcal{X}_n$ *such that* $|\mathcal{X}_n| \leq M^n$ *for* $M \geq 2$ *and all* n*;*
- $\{Y_n\}$ *is a sequence of random variables, for which* $\varepsilon_{X_n|Y_n} \to 0$.

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.*

Lower Bound on $H_\alpha(X|Y)$

Theorem

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

$$\frac{\alpha}{1-\alpha} \log g_\alpha(\varepsilon_{X|Y}) \leq H_\alpha(X|Y), \tag{25}$$

with the piecewise linear function

$$g_\alpha(t) = \left(k(k+1)^{\frac{1}{\alpha}} - k^{\frac{1}{\alpha}}(k+1)\right)t + k^{\frac{1}{\alpha}+1} - (k-1)(k+1)^{\frac{1}{\alpha}} \tag{26}$$

on the interval $t \in \left[1-\frac{1}{k}, 1-\frac{1}{k+1}\right)$ *for* $k \in \{1,2,\ldots\}$.

- Not restricted to finite M.

Proof Outline

Lemma

Let X be a discrete random variable attaining maximal mass $p_{\max}$. Then, for $\alpha \in (0,1) \cup (1,\infty)$,

$$H_\alpha(X) \geq s_\alpha(\varepsilon_X) \tag{27}$$

where $\varepsilon_X = 1 - p_{\max}$ is the minimum error probability of guessing X, and $s_\alpha \colon [0,1) \to [0,\infty)$ is given by

$$s_\alpha(x) := \frac{1}{1-\alpha} \log \left(\left\lfloor \frac{1}{1-x} \right\rfloor (1-x)^\alpha + \left(1 - (1-x) \left\lfloor \frac{1}{1-x} \right\rfloor \right)^\alpha \right).$$

Equality holds in (27) if and only if P_X has $\left\lfloor \frac{1}{p_{\max}} \right\rfloor$ masses equal to $p_{\max}$.

The proof relies on the Schur-concavity of $H_\alpha(\cdot)$.

Proof Outline (cont.)

For every $y \in \mathcal{Y}$, the lemma yields $H_\alpha(X \,|\, Y = y) \geq s_\alpha\big(\varepsilon_{X|Y}(y)\big)$.

Proof Outline (cont.)

For every $y \in \mathcal{Y}$, the lemma yields $H_\alpha(X \,|\, Y = y) \geq s_\alpha\big(\varepsilon_{X|Y}(y)\big)$.

For $\alpha \in (0,1)$, let $f_\alpha \colon [0,1) \to [1,\infty)$ be defined as

$$f_\alpha(x) = \exp\left(\tfrac{1-\alpha}{\alpha}\, s_\alpha(x)\right)$$

- g_α is the piecewise linear function which coincides with f_α at all points $1 - \frac{1}{k}$ for $k \in \mathbb{N}$;
- g_α is the lower convex envelope of f_α;

$$\begin{aligned} H_\alpha(X|Y) &\geq \tfrac{\alpha}{1-\alpha}\, \log \mathbb{E}\left[f_\alpha\big(\varepsilon_{X|Y}(Y)\big)\right] \ \text{(Lemma; } f_\alpha \text{ increasing)} \\ &\geq \tfrac{\alpha}{1-\alpha}\, \log \mathbb{E}\left[g_\alpha\big(\varepsilon_{X|Y}(Y)\big)\right] \ (g_\alpha \leq f_\alpha) \\ &\geq \tfrac{\alpha}{1-\alpha}\, \log\, g_\alpha(\varepsilon_{X|Y}) \ \text{(Jensen)} \end{aligned}$$

Proof Outline (cont.)

For every $y \in \mathcal{Y}$, the lemma yields $H_\alpha(X \,|\, Y = y) \geq s_\alpha\big(\varepsilon_{X|Y}(y)\big)$.

For $\alpha \in (0, 1)$, let $f_\alpha \colon [0, 1) \to [1, \infty)$ be defined as

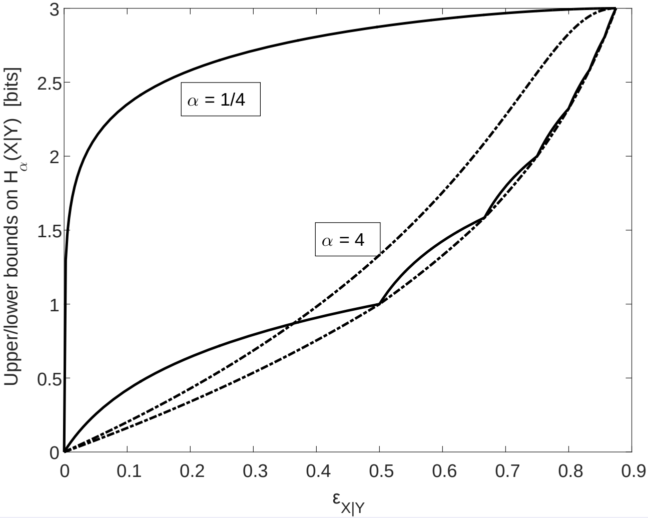
$$f_\alpha(x) = \exp\left(\tfrac{1-\alpha}{\alpha}\; s_\alpha(x)\right)$$

- g_α is the piecewise linear function which coincides with f_α at all points $1 - \frac{1}{k}$ for $k \in \mathbb{N}$;
- g_α is the lower convex envelope of f_α;

$$\begin{aligned} H_\alpha(X|Y) &\geq \tfrac{\alpha}{1-\alpha}\, \log \mathbb{E}\left[f_\alpha\big(\varepsilon_{X|Y}(Y)\big)\right] \ \text{(Lemma; } f_\alpha \text{ increasing)} \\ &\geq \tfrac{\alpha}{1-\alpha}\, \log \mathbb{E}\left[g_\alpha\big(\varepsilon_{X|Y}(Y)\big)\right] \ (g_\alpha \leq f_\alpha) \\ &\geq \tfrac{\alpha}{1-\alpha}\, \log\, g_\alpha(\varepsilon_{X|Y}) \ \text{(Jensen)} \end{aligned}$$

For $\alpha \in (1, \infty)$, $-g_\alpha$ is the lower convex envelope of $-f_\alpha$, and f_α is monotonically decreasing. Proof is similar.

$H_\alpha(X|Y) \longleftrightarrow \varepsilon_{X|Y}$



Asymptotic Tightness

Both upper and lower bounds on $\varepsilon_{X|Y}$ are asymptotically tight as $\alpha \to \infty$.

Asymptotic Tightness

Both upper and lower bounds on $\varepsilon_{X|Y}$ are asymptotically tight as $\alpha \to \infty$.

Special cases

As $\alpha \to 1$, we get existing bounds as special cases:

- Fano's inequality,
- Its counterpart by Kovalevsky ('68), and Tebbe and Dwyer ('68).

Asymptotic Tightness

Both upper and lower bounds on $\varepsilon_{X|Y}$ are asymptotically tight as $\alpha \to \infty$.

Special cases

As $\alpha \to 1$, we get existing bounds as special cases:

- Fano's inequality,
- Its counterpart by Kovalevsky ('68), and Tebbe and Dwyer ('68).

Upper bound on $\varepsilon_{X|Y}$

The most useful domain of applicability of the counterpart to the generalization of Fano's inequality is $\varepsilon_{X|Y} \in [0, \frac{1}{2}]$, in which case the lower bound specializes to $(k = 1)$

$$\frac{\alpha}{1-\alpha} \log\left(1 + \left(2^{\frac{1}{\alpha}} - 2\right)\varepsilon_{X|Y}\right) \leq H_\alpha(X|Y). \tag{28}$$

List Decoding

- Decision rule outputs a list of choices.
- The extension of Fano's inequality to list decoding, expressed in terms of the conditional Shannon entropy, was initiated by Ahlswede, Gacs and Körner ('66).
- Useful for proving converse results.

Generalization of Fano's Inequality for List Decoding

- A generalization of Fano's inequality for list decoding of size L is

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

  where $P_{\mathcal{L}}$ denotes the probability of X not being in the list.
- Averaging a conditional version of $H_\alpha(X|Y=y)$ with respect to the observation is not viable in the case of $H_\alpha(X|Y)$ with $\alpha \neq 1$.

Generalization of Fano's Inequality for List Decoding (cont.)

Theorem (Fixed List Size)

Let P_{XY} *be a probability measure defined on* $\mathcal{X} \times \mathcal{Y}$ *where* $|\mathcal{X}| = M$. *Consider a decision rule*[a] $\mathcal{L} \colon \mathcal{Y} \to \binom{\mathcal{X}}{L}$, *and denote the decoding error probability by* $P_{\mathcal{L}} = \mathbb{P}\big[X \notin \mathcal{L}(Y)\big]$. *Then, for all* $\alpha \in (0,1) \cup (1,\infty)$,

$$H_\alpha(X|Y) \leq \log M - d_\alpha\big(P_{\mathcal{L}} \| 1 - \tfrac{L}{M}\big) \tag{30}$$

with equality in (30) *if and only if*

$$P_{X|Y}(x|y) = \begin{cases} \dfrac{P_{\mathcal{L}}}{M-L}, & x \notin \mathcal{L}(y) \\[2ex] \dfrac{1-P_{\mathcal{L}}}{L}, & x \in \mathcal{L}(y). \end{cases} \tag{31}$$

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

Arimoto-Rényi Conditional Entropy Averaged over Codebook Ensembles

- Consider the channel coding setup with a code ensemble $\mathcal{C}$, over which we are interested in averaging the Arimoto-Rényi conditional entropy of the channel input given the channel output.
- Denote such averaged quantity by

$$\mathbb{E}_{\mathcal{C}}\left[H_\alpha(X^n|Y^n)\right]$$

  where $X^n = (X_1, \dots, X_n)$ and $Y^n = (Y_1, \dots, Y_n)$.
- Some motivation for this study:
  - The normalized equivocation $\frac{1}{n}H(X^n|Y^n)$ was used by Shannon to prove that reliable communication is impossible at rates above capacity;
  - The asymptotic convergence to zero of the equivocation $H(X^n|Y^n)$ at rates below capacity was studied by Feinstein.

Coding Theorem 1 (Feder and Merhav, 1994)

For a DMC with transition probability matrix $P_{Y|X}$, the conditional entropy of the transmitted codeword given the channel output, averaged over a random coding selection with per-letter distribution P_X such that $I(P_X, P_{Y|X}) > 0$, is bounded (in nats) by

$$\mathbb{E}_{\mathcal{C}}\left[H(X^n|Y^n)\right] \leq \left(1 + \frac{1}{\rho^*(R, P_X)}\right) \exp\bigl(-nE_{\mathrm{r}}(R, P_X)\bigr)$$

with

- $R = \frac{\log M}{n} \leq I(P_X, P_{Y|X})$;
- E_{r} is the random-coding error exponent, given by

$$E_{\mathrm{r}}(R, P_X) = \max_{\rho \in [0,1]} \rho \left(I_{\frac{1}{1+\rho}}(P_X, P_{Y|X}) - R \right); \qquad (32)$$

- the argument that maximizes (32) is denoted by $\rho^*(R, P_X)$.

Coding Theorem 2 (ISSV, 2017)

The following results hold under the setting in the previous theorem:

- For all $\alpha > 0$, and rates R below the channel capacity C,

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mathbb{E}_{\mathcal{C}}\left[H_\alpha(X^n|Y^n)\right] \leq E_{\mathsf{sp}}(R), \tag{33}$$

where $E_{\mathsf{sp}}(\cdot)$ denotes the sphere-packing error exponent

$$E_{\mathsf{sp}}(R) = \sup_{\rho\geq 0} \rho \left(\max_{Q_X} I_{\frac{1}{1+\rho}}(Q_X, P_{Y|X}) - R \right) \tag{34}$$

with the maximization in the right side of (34) over all single-letter distributions Q_X defined on the input alphabet.

Coding Theorem 2 (ISSV '17, cont.)

- For all $\alpha \in (0,1)$,

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{E}_{\mathcal{C}}\big[H_\alpha(X^n|Y^n)\big] \geq \alpha E_{\mathrm{r}}(R, P_X) - (1-\alpha)R, \tag{35}$$

provided that

$$R < R_\alpha(P_X, P_{Y|X}) \tag{36}$$

where $R_\alpha(P_X, P_{Y|X})$ is the unique solution $r \in (0, I(P_X, P_{Y|X}))$ to

$$E_{\mathrm{r}}(r, P_X) = \left(\frac{1}{\alpha} - 1\right) r. \tag{37}$$

Coding Theorem 2 (ISSV '17, cont.)

- The rate $R_\alpha(P_X, P_{Y|X})$ is monotonically increasing and continuous in $\alpha \in (0,1)$, and

$$\lim_{\alpha \downarrow 0} R_\alpha(P_X, P_{Y|X}) = 0, \tag{38}$$

$$\lim_{\alpha \uparrow 1} R_\alpha(P_X, P_{Y|X}) = I(P_X, P_{Y|X}). \tag{39}$$

Coding Theorem 3 (ISSV '17, cont.)

Let $P_{Y|X}$ be the transition probability matrix of a memoryless binary-input output-symmetric channel, and let $P_X^* = \left[\frac{1}{2}\ \frac{1}{2}\right]$. Let R_c, R_0, and C denote the critical and cutoff rates and the channel capacity, respectively, and let

$$\alpha_c = \frac{R_c}{R_0} \in (0, 1). \tag{40}$$

The rate $R_\alpha = R_\alpha(P_X^*, P_{Y|X})$, with the symmetric input distribution P_X^*, can be expressed as follows:

a) for $\alpha \in (0, \alpha_c]$, $R_\alpha = \alpha R_0$;

b) for $\alpha \in (\alpha_c, 1)$, $R_\alpha \in (R_c, C)$ is the solution to $E_{sp}(r) = \left(\frac{1}{\alpha} - 1\right) r$;

c) R_α is continuous, monotonically increasing in $\alpha \in [\alpha_c, 1)$ from R_c to C.

Example: BSC(δ)

- Consider a BSC with crossover probability δ, and let $P_X = \left[\frac{1}{2}\ \frac{1}{2}\right]$.
- the cutoff rate, critical rate and capacity (in bits) are given by

$$R_0 = 1 - \log\left(1 + \sqrt{4\delta(1-\delta)}\right), \tag{41}$$

$$R_{\mathrm{c}} = 1 - h\left(\frac{\sqrt{\delta}}{\sqrt{\delta} + \sqrt{1-\delta}}\right), \tag{42}$$

$$C = I(P_X, P_{Y|X}) = 1 - h(\delta). \tag{43}$$

- The sphere-packing error exponent is given by

$$E_{\mathrm{sp}}(R) = d\left(\delta_{\mathrm{GV}}(R) \,\|\, \delta\right) \tag{44}$$

where the normalized Gilbert-Varshamov distance is denoted by

$$\delta_{\mathrm{GV}}(R) = h^{-1}(1-R). \tag{45}$$

Example: BSC(δ) (cont.)

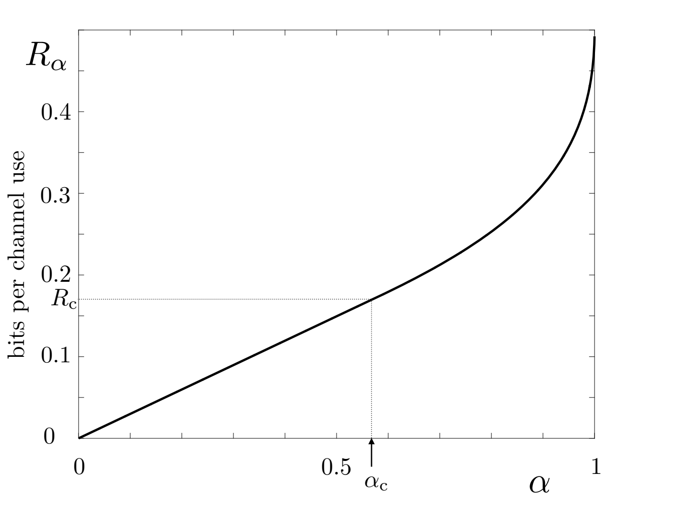


Figure: The rate R_α for $\alpha \in (0, 1)$ for BSC(δ) with crossover prob. $\delta = 0.110$.

Conclusions

- We have shown new bounds on the minimum Bayesian error prob. $\varepsilon_{X|Y}$ of M-ary hypothesis testing.
- Our major focus has been the Arimoto-Rényi conditional entropy of the hypothesis index given the observation.

Conclusions

- We have shown new bounds on the minimum Bayesian error prob. $\varepsilon_{X|Y}$ of M-ary hypothesis testing.
- Our major focus has been the Arimoto-Rényi conditional entropy of the hypothesis index given the observation.
- Changing the conventional form of Fano's inequality from

$$H(X|Y) \leq h(\varepsilon_{X|Y}) + \varepsilon_{X|Y} \log(M-1) \tag{46}$$
$$= \log M - d\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big) \tag{47}$$

to the right side of (47), where $d(\cdot\|\cdot)$ is the binary relative entropy, allows a natural generalization where the Arimoto-Rényi conditional entropy of an arbitrary positive order α is upper bounded by

$$H_\alpha(X|Y) \leq \log M - d_\alpha\big(\varepsilon_{X|Y} \| 1 - \tfrac{1}{M}\big) \tag{48}$$

with $d_\alpha(\cdot\|\cdot)$ denoting the binary Rényi divergence.

Conclusions (Cont.)

- The Schur-concavity of the Rényi entropy yields a lower bound on $H_\alpha(X|Y)$ in terms of $\varepsilon_{X|Y}$, which holds even if $M = \infty$. It recovers existing bounds by letting $\alpha \to 1$.

Conclusions (Cont.)

- The Schur-concavity of the Rényi entropy yields a lower bound on $H_\alpha(X|Y)$ in terms of $\varepsilon_{X|Y}$, which holds even if $M = \infty$. It recovers existing bounds by letting $\alpha \to 1$.
- Our techniques were extended to list decoding with a fixed list size, generalizing all the $H_\alpha(X|Y)$–$\varepsilon_{X|Y}$ bounds to that setting.

Conclusions (Cont.)

- The Schur-concavity of the Rényi entropy yields a lower bound on $H_\alpha(X|Y)$ in terms of $\varepsilon_{X|Y}$, which holds even if $M = \infty$. It recovers existing bounds by letting $\alpha \to 1$.
- Our techniques were extended to list decoding with a fixed list size, generalizing all the $H_\alpha(X|Y)$–$\varepsilon_{X|Y}$ bounds to that setting.
- Application: We analyzed the exponentially vanishing decay of the Arimoto-Rényi conditional entropy of the transmitted codeword given the channel output for DMCs and random coding ensembles.

Further Results in This Work

- Explicit lower bounds on $\varepsilon_{X|Y}$ as a function of $H_\alpha(X|Y)$ for an arbitrary α (also, for $\alpha < 0$).
- Explicit lower bounds on the list decoding error probability for fixed list size as a function of $H_\alpha(X|Y)$ for an arbitrary α (also, for $\alpha < 0$).
- We also explored some facets of the role of binary hypothesis testing in analyzing M-ary Bayesian hypothesis testing problems, and have shown new bounds in terms of Rényi divergence.

Journal Paper

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