Improved Bounds on Guessing Moments via Rényi Measures

Igal Sason (Technion)

Joint work with Sergio Verdú (Princeton)

ISIT 2018

Vail, Colorado, USA
June 17-22, 2018

Guessing

The problem of guessing discrete random variables has found a variety of applications in

- Shannon theory,
- coding theory,
- cryptography,
- searching and sorting algorithms,

etc.

The central object of interest:
The distribution of the number of guesses required to identify a realization of a random variable, taking values on a finite or countably infinite set.

Guessing and Ranking functions

- X is a discrete random variable taking values on $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$.
- One wishes to guess the value of X by repeatedly asking questions of the form "Is X equal to x ?" until X is guessed correctly.
- A guessing function is a 1-to-1 function $g \colon \mathcal{X} \to \mathcal{X}$ where the number of guesses is equal to $g(x)$ if $X = x \in \mathcal{X}$.
- For $\rho > 0$, $\mathbb{E}[g^\rho(X)]$ is minimized by selecting g to be a ranking function g_X, for which $g_X(x) = k$ if $P_X(x)$ is the k-th largest mass.
- Having side information $Y = y$ on X, we refer to the conditional ranking function $g_{X|Y}(\cdot|y)$.
- $\mathbb{E}[g^\rho_{X|Y}(X|Y)]$ is the ρ-th moment of the number of guesses required for correctly identifying the unknown object X on the basis of Y.

The Rényi Entropy

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

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

By its continuous extension, $H_1(X) = H(X)$.

The Arimoto-Rényi Conditional Entropy

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

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

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}\left[\left(\sum_{x\in\mathcal{X}} P^\alpha_{X|Y}(x|Y)\right)^{\frac{1}{\alpha}}\right] \quad (2)$$

$$= \frac{\alpha}{1-\alpha} \log \sum_{y\in\mathcal{Y}} P_Y(y) \exp\left(\frac{1-\alpha}{\alpha} H_\alpha(X|Y=y)\right), \quad (3)$$

  where (3) applies if Y is a discrete random variable.
- Continuous extension at $\alpha = 0, 1, \infty$ with $H_1(X|Y) = H(X|Y)$.

$H_\alpha(X|Y)$ and Guessing Moments

Theorem (Arikan '96)

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 $g_{X|Y}(\cdot|y)$ be a ranking function of X given that $Y = y$. Then, for $\rho > 0$,

$$\frac{1}{\rho} \log \mathbb{E}\left[g^\rho_{X|Y}(X|Y)\right] \geq H_{\frac{1}{1+\rho}}(X|Y) - \log(1 + \log_{\mathrm{e}} M), \tag{4}$$

$$\frac{1}{\rho} \log \mathbb{E}\left[g^\rho_{X|Y}(X|Y)\right] \leq H_{\frac{1}{1+\rho}}(X|Y). \tag{5}$$

$H_\alpha(X|Y)$ and Guessing Moments

Theorem (Arikan '96)

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 $g_{X|Y}(\cdot|y)$ be a ranking function of X given that $Y = y$. Then, for $\rho > 0$,

$$\frac{1}{\rho} \log \mathbb{E}\big[g^{\rho}_{X|Y}(X|Y)\big] \geq H_{\frac{1}{1+\rho}}(X|Y) - \log(1 + \log_{\mathrm{e}} M), \tag{4}$$

$$\frac{1}{\rho} \log \mathbb{E}\big[g^{\rho}_{X|Y}(X|Y)\big] \leq H_{\frac{1}{1+\rho}}(X|Y). \tag{5}$$

Arikan's result yields an asymptotically tight error exponent:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\big[g^{\rho}_{X^n|Y^n}(X^n|Y^n)\big] = \rho H_{\frac{1}{1+\rho}}(X|Y)$$

when $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. $\quad [X^n := (X_1, \ldots, X_n)]$.

Key Result

Theorem

Given a discrete random variable X taking values on a set $\mathcal{X}$, an arbitrary non-negative function $g \colon \mathcal{X} \to (0, \infty)$, and a scalar $\rho \neq 0$, then

$$\sup_{\beta \in (-\rho, +\infty) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log \sum_{x \in \mathcal{X}} g^{-\beta}(x) \right]$$

$$\leq \frac{1}{\rho} \log \mathbb{E}[g^\rho(X)] \tag{6}$$

$$\leq \inf_{\beta \in (-\infty, -\rho) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log \sum_{x \in \mathcal{X}} g^{-\beta}(x) \right]. \tag{7}$$

Letting $\beta = 1$ yields the lower bound by Courtade and Verdú (ISIT '14).

Theorem: Consequence of Key Result

Let $g\colon \mathcal{X} \to \mathcal{X}$ be an arbitrary guessing function. Then, for every $\rho \neq 0$,

$$\frac{1}{\rho} \log \mathbb{E}\big[g^\rho(X)\big] \geq \sup_{\beta \in (-\rho,\infty)\setminus\{0\}} \frac{1}{\beta} \Big[H_{\frac{\beta}{\beta+\rho}}(X) - \log u_M(\beta)\Big] \tag{8}$$

with

$$u_M(\beta) = \begin{cases} \log_{\mathrm{e}} M + \gamma + \frac{1}{2M} - \frac{5}{6(10M^2+1)} & \beta = 1, \\ \min\Big\{\zeta(\beta) - \frac{(M+1)^{1-\beta}}{\beta-1} - \frac{(M+1)^{-\beta}}{2},\, u_M(1)\Big\} & \beta > 1, \\ 1 + \dfrac{1}{1-\beta}\, \Big[\big(M+\frac{1}{2}\big)^{1-\beta} - \big(\frac{3}{2}\big)^{1-\beta}\Big] & |\beta| < 1, \\ \dfrac{M^{1-\beta}-1}{1-\beta} + \frac{1}{2}\big(1 + M^{-\beta}\big) & \beta \leq -1. \end{cases} \tag{9}$$

- $u_M(\beta)$ is an upper/ lower bound on $\sum_{n=1}^{M} \frac{1}{n^\beta}$ for $\beta > 0$ or $\beta < 0$, resp.;
- $\gamma \approx 0.5772$ is Euler's constant;
- $\zeta(\beta) = \sum_{n=1}^{\infty} \frac{1}{n^\beta}$ is Riemann's zeta function for $\beta > 1$.

Lower Bound: Special Case

Specializing to $\beta = 1$, and using

$$u_M(1) = \sum_{j=1}^{M} \tfrac{1}{j} \leq 1 + \log_e M, \quad M \geq 2, \tag{10}$$

we obtain

$$\tfrac{1}{\rho} \, \log \mathbb{E}\big[g^\rho(X)\big] \geq H_{\frac{1}{1+\rho}}(X) - \log\big(1 + \log_e M\big) \tag{11}$$

for $\rho \in (-1, \infty)$. Bound (11) was obtained for $\rho > 0$ by Arikan.

Upper Bounds on Optimal Guessing Moments

- We also derive upper bounds on the ρ-th moment of optimal guessing (i.e., if $g = g_X$);
- In the non-asymptotic regime (finite M), they improve
  - the asymptotically tight bound by Arikan (1996);
  - its refinement by Boztaș (1997).

Upper Bounds on Optimal Guessing Moments

- We also derive upper bounds on the ρ-th moment of optimal guessing (i.e., if $g = g_X$);
- In the non-asymptotic regime (finite M), they improve
  - the asymptotically tight bound by Arikan (1996);
  - its refinement by Boztaș (1997).

1st Upper Bound on Optimal Guessing Moments

For $\rho > 0$

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

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

2nd Upper Bound on Optimal Guessing Moments

1. For $\rho \in [0, 1]$

$$\mathbb{E}[g_X^\rho(X)] \le \frac{1}{1+\rho} \exp\Big(\rho H_{\frac{1}{1+\rho}}(X)\Big) + \frac{\rho - (1-\rho)(2^\rho - 1)(1 - p_{\max})}{1+\rho}. \tag{12}$$

2. For $\rho \in [1, 2]$

$$\mathbb{E}[g_X^\rho(X)] \le \frac{1}{1+\rho} \exp\Big(\rho H_{\frac{1}{1+\rho}}(X)\Big) + \frac{1}{\rho} \exp\Big((\rho-1) H_{\frac{1}{\rho}}(X)\Big) + \frac{\rho^2 - \rho - 1}{\rho(1+\rho)}. \tag{13}$$

Furthermore, both (12) and (13) hold with equality if X is deterministic.

3rd Upper Bound on Optimal Guessing Moments

$$\mathbb{E}[g_X^\rho(X)] \leq 1 + \sum_{j=0}^{\lfloor \rho \rfloor} c_j(\rho) \left[\exp\left((\rho - j)\, H_{\frac{1}{1+\rho-j}}(X) \right) - 1 \right], \tag{14}$$

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

$$c_j(\rho) = \begin{cases} \dfrac{1}{1+\rho} & j = 0 \\ \frac{1}{2} & j = 1 \\ \dfrac{\rho \ldots (\rho - j + 2)}{2^j} & j \in \{2, \ldots, \lfloor \rho \rfloor - 1\} \\ \dfrac{\rho \ldots (\rho - j + 2)}{2^{j-1}\,(\rho - j + 1)} & j = \lfloor \rho \rfloor \end{cases} \tag{15}$$

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

Numerical Results

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

$$P_X(k) = \frac{(1-a)\, a^{k-1}}{1-a^M}, \quad k \in \{1, \dots, M\} \tag{16}$$

where $a = 0.9$ and $M = 32$. Table 1 compares $\frac{1}{3} \log_{\mathrm{e}} \mathbb{E}[g_X^3(X)]$ to its various lower and upper bounds (LBs and UBs, respectively).

Table: Comparison of $\frac{1}{3} \log_{\mathrm{e}} \mathbb{E}[g_X^3(X)]$ and bounds.

Arikan's LB	Improved LB	$\frac{1}{3} \log_{\mathrm{e}} \mathbb{E}[g_X^3(X)]$ exact value	Improved UB	Arikan's UB
1.864	2.593	2.609	2.920	3.360

Bounds on Guessing Moments with Side Information

- Our lower and upper bounds extend to allow side information Y for guessing the value of X.
- These bounds tighten the results by Arikan for all $\rho > 0$.
- With side information Y, all bounds stay valid by the replacement of $H_\alpha(X)$ with the Arimoto-Rényi conditional entropy $H_\alpha(X|Y)$.

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,

$$\varepsilon_{X|Y} = \mathbb{E}\left[1 - \max_{x \in \mathcal{X}} P_{X|Y}(x|Y)\right]. \tag{17}$$

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,

$$\varepsilon_{X|Y} = \mathbb{E}\left[1 - \max_{x \in \mathcal{X}} P_{X|Y}(x|Y)\right]. \tag{17}$$

- Identity:

$$\varepsilon_{X|Y} = 1 - \mathbb{P}[g_{X|Y}(X|Y) = 1]. \tag{18}$$

Exact Locus of $(\varepsilon_{X|Y}, \mathbb{E}[g^\rho_{X|Y}(X|Y)])$

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

$$f_\rho(\varepsilon_{X|Y}) \le \mathbb{E}[g^\rho_{X|Y}(X|Y)] \le 1 + \left(\frac{2^\rho + \ldots + M^\rho}{M-1} - 1 \right) \varepsilon_{X|Y} \tag{19}$$

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

$$f_\rho(u) = (1-u) \sum_{j=1}^{k_u} j^\rho + [1 - (1-u)k_u](k_u+1)^\rho, \tag{20}$$

$$k_u = \left\lfloor \frac{1}{1-u} \right\rfloor. \tag{21}$$

The Upper and Lower Bounds Are Tight

- Let

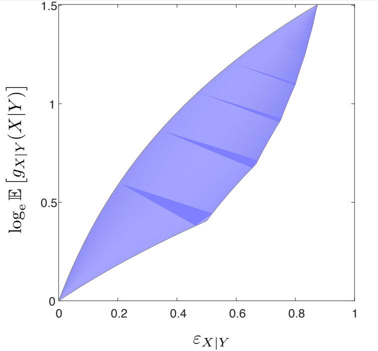
$$p_{\max}(y) = \max_{x \in \mathcal{X}} P_{X|Y}(x|y)$$

for $y \in \mathcal{Y}$. The lower bound is attained if and only if

1. $p_{\max}(y) = p_{\max}$ is fixed for all $y \in \mathcal{Y}$;
2. 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 $1 - p_{\max} \left\lfloor \frac{1}{p_{\max}} \right\rfloor$ if $\frac{1}{p_{\max}}$ 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 $\mathcal{X}$.

$\varepsilon_{X|Y} \longleftrightarrow \mathbb{E}[g^{\rho}_{X|Y}(X|Y)]$



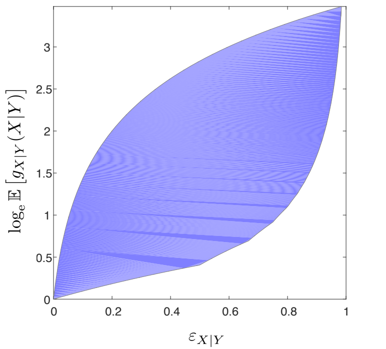


Figure: locus of attainable values of $(\varepsilon_{X|Y}, \log_e \mathbb{E}[g_{X|Y}(X|Y)])$. The random variable X takes $M = 8$ (left plot) or $M = 64$ (right plot) possible values.

$\varepsilon_{X|Y} \longleftrightarrow \mathbb{E}[g^{\rho}_{X|Y}(X|Y)]$

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,

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

with

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

Summary

- Derivation of new 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.
- Similarly to Arikan's bounds, they are expressed in terms of the Arimoto-Rényi conditional entropy.
- Arikan's bounds are asymptotically tight. However, the improvement of the new bounds is significant in the non-asymptotic regime.
- Application: improved non-asymptotic bounds for fixed-to-variable optimal lossless source coding without the prefix constraint (my ISIT talk to be given on Friday at 9:50 AM).
- Relationships between moments of the optimal guessing function and the MAP error probability are provided, characterizing the exact locus of the attainable values of $(\varepsilon_{X|Y}, H_\alpha(X|Y))$.

Journal Paper

I.S. 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.