Entropy and Guessing: Old and New Results

Igal Sason
(EE Department, Technion)

Workshop on Mathematical Data Science (MDS) 2019

Dürnstein, Austria

October 13–15, 2019

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 a finite or countably infinite set $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$.

Guessing and Ranking functions

- X is a discrete random variable taking values on a finite or countably infinite set $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$.
- 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.

Guessing and Ranking functions

- X is a discrete random variable taking values on a finite or countably infinite set $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$.
- 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 $X = x \in \mathcal{X}$.

Guessing and Ranking functions

- X is a discrete random variable taking values on a finite or countably infinite set $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$.
- 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.

Guessing and Shannon Entropy (Massey, ISIT '94)

Average number of successive guesses with an optimal strategy satisfies

$$\mathbb{E}[g_X(X)] \geq \tfrac{1}{4} \exp\bigl(H(X)\bigr) + 1$$

provided $H(X) \geq 2$ bits. It is tight within a factor of $\frac{4}{\mathrm{e}}$ when X is geometrically distributed.

Guessing and Shannon Entropy (McEliece and Yu, ISIT '95)

If X takes no more than $M < \infty$ possible values, then

$$\mathbb{E}[g_X(X)] \leq \left(\frac{M-1}{2\log M}\right) H(X)$$

This upper bound on the number of guesses is tight if and only if X is equiprobable with $P_X(x) = \frac{1}{M}$ for each x, or if X is deterministic.

Can we Get Bounds on the ρ-th Moments ($\rho > 0$) by Using a Generalized Information-Theoretic Measure of Shannon's Entropy ?

The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

(a) $H(p_1, p_2, \cdots, p_n)$ *is a symmetric function of its variables for* $n = 2, 3, \cdots$.
(b) $H(p, 1-p)$ *is a continuous function of* p *for* $0 \leqq p \leqq 1$.
(c) $H(1/2, 1/2) = 1$.
(d) $H[tp_1, (1-t)p_1, p_2, \cdots, p_n] = H(p_1, p_2, \cdots, p_n) + p_1 H(t, 1-t)$
for any distribution $\mathcal{P} = (p_1, p_2, \cdots, p_n)$ *and for* $0 \leqq t \leqq 1$.

1961

ON MEASURES OF ENTROPY AND INFORMATION

ALFRÉD RÉNYI

(a) $H(p_1, p_2, \cdots, p_n)$ *is a symmetric function of its variables for* $n = 2, 3, \cdots$.
(b) $H(p, 1 - p)$ *is a continuous function of* p *for* $0 \leqq p \leqq 1$.
(c) $H(1/2, 1/2) = 1$.

(d') If X and Y are independent, then $H(X, Y) = H(X) + H(Y)$

Rényi entropy

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{A}} P_X^\alpha(x), \quad \alpha \in (0,1) \cup (1,\infty)$$

$$\tfrac{\alpha}{1-\alpha} \log \|P_X\|_\alpha$$

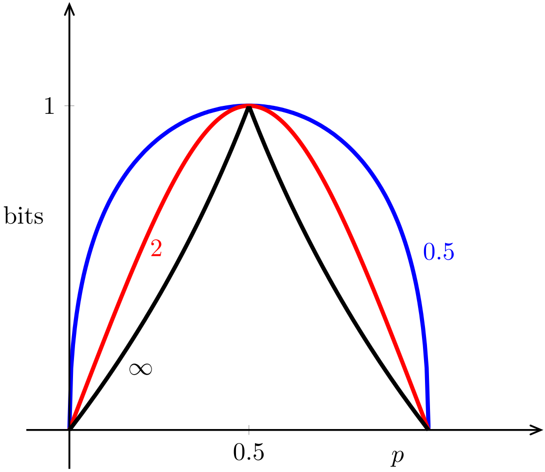
$$H_1(X) = H(X),$$

$$H_\infty(X) = \min_{x \in \mathcal{A}} \log \frac{1}{P_X(x)}$$

$$H_0(X) = \log \left| \{x \in \mathcal{A} \colon P_X(x) > 0\} \right|$$

$$H_2(X) = \log \frac{1}{\sum_{x \in \mathcal{A}} P_X^2(x)}$$

Rényi entropy: binary distribution



Applications of Rényi Entropy

- Random search (Rényi, 1965).
- Statistical physics (Tsallis, 1988).
- Secret-key generation (Renner-Wolf, 2005).
- Data compression (Campbell, 1965).
- Hypothesis testing and coding theorems (Csiszár, 1995).
- Guessing (Arikan, 1996).

$H_\alpha(X)$ and Guessing Moments

Theorem (Arikan '96)

Let X be a discrete random variable taking values on $\mathcal{X} = \{1, \ldots, M\}$. Let $g_X(\cdot)$ be a ranking function of X. Then, for $\rho > 0$,

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

$$\frac{1}{\rho} \log \mathbb{E}\big[g_X^\rho(X)\big] \leq H_{\frac{1}{1+\rho}}(X).$$

$H_\alpha(X)$ and Guessing Moments

Theorem (Arikan '96)

Let X be a discrete random variable taking values on $\mathcal{X} = \{1, \ldots, M\}$. Let $g_X(\cdot)$ be a ranking function of X. Then, for $\rho > 0$,

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

$$\frac{1}{\rho} \log \mathbb{E}\big[g_X^\rho(X)\big] \leq H_{\frac{1}{1+\rho}}(X).$$

Arikan's result yields an asymptotically tight error exponent:

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}\big[g_{X^n}^\rho(X^n)\big] = \rho H_{\frac{1}{1+\rho}}(X), \quad \forall\, \rho > 0$$

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

Bounds on Guessing Moments with Side Information

- 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.

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 $\alpha \neq 1$, it 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 $\alpha \neq 1$, it 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), where $H_\alpha(X|Y) \leq H_\alpha(X)$.

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.

The Arimoto-Rényi Conditional Entropy

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

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

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

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

  where the last equality 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 $\rho > 0$,

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

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

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

Theorem (Arikan '96)

Let X and Y be discrete random variables taking values on the sets $\mathcal{X} = \{1, \dots, M\}$ 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 $\rho > 0$,

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

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

Arikan's result yields an asymptotically tight error exponent:

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

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

Some Recent Results on Guessing

Guessing with an Encoder

- Prior to guessing X, the guesser is provided some side information about X in the form of $f(X)$, where $f\colon \mathcal{X} \to \{1, \ldots, M\}$.
- A guessing function $G(\cdot|\cdot)\colon \mathcal{X} \times \{1, \ldots, M\} \to \{1, \ldots, |\mathcal{X}|\}$: $G(\cdot|m)\colon \mathcal{X} \to \{1, \ldots, |\mathcal{X}|\}$ is a bijective function $\forall\, m \in \{1, \ldots, M\}$.

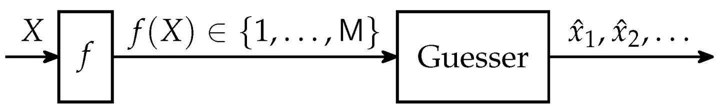


Figure: Guessing with an encoders f.

A. Bracher, E. Hof and A. Lapidoth, "Guessing attacks on distributed storage systems," 2017. Accepted for publication in the IEEE T-IT, 2019.

Guessing with an Encoder (Cont.)

If among all encoders, f^* minimizes the ρ-th moment of the number of guesses required by an optimal guesser to guess X after observing $f(X)$, then

$$\begin{aligned} &\frac{1}{(1+\ln|\mathcal{X}|)^\rho} \; \exp\left(\rho\Big[H_{\frac{1}{1+\rho}}(X) - \log M\Big]\right) \\ &\leq \mathbb{E}[G^*(X|f^*(X))^\rho] \\ &\leq 1 + \exp\left(\rho\Big[H_{\frac{1}{1+\rho}}(X) - \log M + 1\Big]\right). \end{aligned}$$

Guessing with an Encoder (Cont.)

In guessing a sequence of i.i.d. random variables

$$X^n = (X_1, \dots, X_n),$$

a description rate of $H_{\frac{1}{1+\rho}}(X) + \varepsilon$ bits per symbol (with $\varepsilon > 0$ arbitrarily small) is needed to drive the ρ-th moment of the number of guesses to 1 as the length of the sequence (n) tends to infinity.

$$\lim_{n\to\infty} \mathbb{E}[G^*(X^n \,|\, f_n^*(X^n))^{\rho}] = 1.$$

Guessing with Distributed Encoders

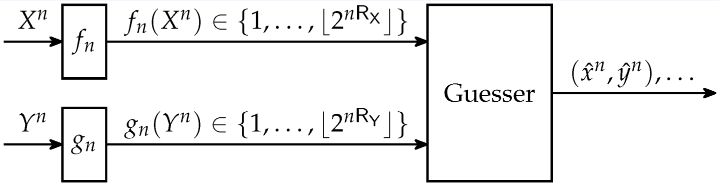


Figure: Guessing with distributed encoders f_n and g_n.

A. Bracher, A. Lapidoth and C. Pfister, "Guessing with distributed encoders," *Entropy*, March 2019.

Guessing with Distributed Encoders

Analog of Slepian-Wolf coding (distributed lossless source coding).

- Two dependent sources generate a pair of sequences $X^n := (X_1, \ldots, X_n)$ and $Y^n := (Y_1, \ldots, Y_n)$
- The pairs $\{(X_i, Y_i)\}_{i=1}^n$ are taken from a finite alphabet $\mathcal{X} \times \mathcal{Y}$.

Guessing with Distributed Encoders

Analog of Slepian-Wolf coding (distributed lossless source coding).

- Two dependent sources generate a pair of sequences $X^n := (X_1, \ldots, X_n)$ and $Y^n := (Y_1, \ldots, Y_n)$
- The pairs $\{(X_i, Y_i)\}_{i=1}^n$ are taken from a finite alphabet $\mathcal{X} \times \mathcal{Y}$.
- Each of the two sequences is observed by a different encoder, which produces a rate-limited description to the sequence it observes:
  - The sequence X^n is described by one of $\lfloor \exp(nR_X) \rfloor$ labels.
  - The sequence Y^n is described by one of $\lfloor \exp(nR_Y) \rfloor$ labels.

$$f_n \colon \mathcal{X}^n \to \{1, \ldots, \lfloor \exp(nR_X) \rfloor\}, \quad R_X \geq 0,$$
$$g_n \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor \exp(nR_Y) \rfloor\}, \quad R_Y \geq 0.$$

Guessing with Distributed Encoders

Analog of Slepian-Wolf coding (distributed lossless source coding).

- Two dependent sources generate a pair of sequences $X^n := (X_1, \ldots, X_n)$ and $Y^n := (Y_1, \ldots, Y_n)$
- The pairs $\{(X_i, Y_i)\}_{i=1}^n$ are taken from a finite alphabet $\mathcal{X} \times \mathcal{Y}$.
- Each of the two sequences is observed by a different encoder, which produces a rate-limited description to the sequence it observes:
  - The sequence X^n is described by one of $\lfloor \exp(nR_X) \rfloor$ labels.
  - The sequence Y^n is described by one of $\lfloor \exp(nR_Y) \rfloor$ labels.

$$f_n \colon \mathcal{X}^n \to \{1, \ldots, \lfloor \exp(nR_X) \rfloor\}, \quad R_X \geq 0,$$
$$g_n \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor \exp(nR_Y) \rfloor\}, \quad R_Y \geq 0.$$

- The two rate-limited descriptions are provided to a guessing device, which produces guesses of the form $(\hat{x}^n, \hat{y}^n)$ until $(\hat{x}^n, \hat{y}^n) = (x^n, y^n)$.

Achievable Rate Pairs

For a fixed $\rho > 0$, a rate pair $(R_X, R_Y) \in \mathbb{R}_+^2$ is called achievable if there exists a sequence of distributed encoders and guessing functions $\{f_n, g_n, G_n\}$ such that the ρ-th moment of the number of guesses tends to 1 as we let n tend to infinity.

$$\lim_{n\to\infty} \mathbb{E}\big[G_n\big(X^n, Y^n \,|\, f_n(X^n), g_n(Y^n)\big)^{\rho}\big] = 1.$$

Exact Characterization of the Rate Region

Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be i.i.d. according to P_{XY}. Consider the rate region $\mathcal{R}(\rho)$ which is defined to be the set of rate tuples (R_X, R_Y) such that

$$R_X \geq H_{\frac{1}{1+\rho}}(X|Y),$$

$$R_Y \geq H_{\frac{1}{1+\rho}}(Y|X),$$

$$R_X + R_Y \geq H_{\frac{1}{1+\rho}}(X, Y).$$

Then, all rate pairs in the interior of $\mathcal{R}(\rho)$ are achievable, while those outside $\mathcal{R}(\rho)$ are not achievable.

An Important Difference From Slepian-Wolf Coding

- Slepian-Wolf coding allows separate encoding with the same sum-rate as with joint encoding, $H(X, Y)$.
- This is not necessarily true in the setting of guessing with distributed encoders.
- Specifically, for $\rho > 0$, if

$$H_{\frac{1}{1+\rho}}(X|Y) + H_{\frac{1}{1+\rho}}(Y|X) > H_{\frac{1}{1+\rho}}(X, Y),$$

  then the single-rate constraints on R_X and R_Y together impose a stronger constraint on the sum-rate than the third constraint on $R_X + R_Y$. It then requires a larger sum-rate than joint encoding.

Improving Arikan's Bounds in the Non-Asymptotic Setting

Result (I.S. & S. Verdú, IEEE T-IT, June 2018)

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

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

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

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

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

Theorem: Consequence of the Result

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

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

where $u_M(\beta)$ is an upper/ lower bound on $\sum_{n=1}^{M} \frac{1}{n^{\beta}}$ for $\beta > 0$ or $\beta < 0$, respectively.

Theorem: Consequence of the Result

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

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

with

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

- $\gamma \approx 0.5772$ is Euler's constant;
- $\zeta(\beta) = \sum\limits_{n=1}^{\infty} \frac{1}{n^\beta}$ is Riemann's zeta function for $\beta > 1$.

Lower Bound: Special Case

Specializing to $\beta = 1$, and using an upper bound on the harmonic sum:

$$u_M(1) = \sum_{j=1}^{M} \tfrac{1}{j} \leq 1 + \log_{\mathrm{e}} M, \quad M \geq 2,$$

we obtain

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

for $\rho \in (-1, \infty)$. The latter bound 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 $\rho > 0$

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

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

2nd Upper Bound on Optimal Guessing Moments

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

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

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

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

Furthermore, both bounds hold with equality if X is deterministic.

3rd Upper Bound on Optimal Guessing Moments

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

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

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

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

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

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

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

Arikan's LB	Improved LB	$\mathbb{E}[g_X^3(X)]$ exact value	Improved UB	Arikan's UB
268	$2,390$	$2,507$	$6,374$	$23,861$

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 $H_\alpha(X|Y)$.

Guessing, Rényi Entropy and Majorization

- Majorization: a simple & productive concept in the theory of inequalities.

Guessing, Rényi Entropy and Majorization

- Majorization: a simple & productive concept in the theory of inequalities.
- Cicalese *et al.* (IEEE T-IT, April '18):

  If X is a RV taking n possible values, and the size of the support of $f(X)$ equals m with $m < n$, how close $H\big(f(X)\big)$ can be to $H(X)$?

Guessing, Rényi Entropy and Majorization

- Majorization: a simple & productive concept in the theory of inequalities.
- Cicalese *et al.* (IEEE T-IT, April '18):

  If X is a RV taking n possible values, and the size of the support of $f(X)$ equals m with $m < n$, how close $H(f(X))$ can be to $H(X)$?
- Their goal: computing

$$\max_f H\big(f(X)\big) = \max_f \Big\{ H\big(f(X)\big) - H\big(f(X)|X\big) \Big\} = \max_f I\big(X; f(X)\big)$$

with max. over all functions mapping a set of cardinality n to a set of cardinality $m < n$.

Guessing, Rényi Entropy and Majorization

- Majorization: a simple & productive concept in the theory of inequalities.
- Cicalese *et al.* (IEEE T-IT, April '18):

  If X is a RV taking n possible values, and the size of the support of $f(X)$ equals m with $m < n$, how close $H\big(f(X)\big)$ can be to $H(X)$?
- Their goal: computing

$$\max_f H\big(f(X)\big) = \max_f \Big\{ H\big(f(X)\big) - H\big(f(X)|X\big) \Big\} = \max_f I\big(X; f(X)\big)$$

  with max. over all functions mapping a set of cardinality n to a set of cardinality $m < n$.
- Useful in the context of data clustering.

Motivation

- Rényi measures are very useful, so how about ...

  Generalizing this question to $H_\alpha\big(f(X)\big)$ for any $\alpha > 0$ (not trivial).

- Possible Applications:
  - Guessing (Arikan '96);
  - Lossless compression problems (Campbell '65).

Journal Paper

I. S., "Tight bounds on the Rényi entropy via majorization with applications to guessing and compression," *Entropy*, vol. 20, no. 12, paper 896, pp. 1–25, November 2018.

Setting

Let

- $\alpha > 0$;
- $\mathcal{X}$ and $\mathcal{Y}$ be finite sets of cardinalities
$$|\mathcal{X}| = n, \quad |\mathcal{Y}| = m, \quad n > m \geq 2;$$
without any loss of generality, let
$$\mathcal{X} = \{1, \dots, n\}, \quad \mathcal{Y} = \{1, \dots, m\};$$
- $\mathcal{P}_n$ $(n \geq 2)$ be the set of probability mass functions (pmf) on $\mathcal{X}$;
- X be a RV taking values on $\mathcal{X}$ with a pmf $P_X \in \mathcal{P}_n$;
- $\mathcal{F}_{n,m}$ be the set of deterministic functions $f \colon \mathcal{X} \to \mathcal{Y}$;
- $f \in \mathcal{F}_{n,m}$ is not one-to-one since $m < n$.

Bad News

For an arbitrary $\alpha > 0$, the maximization problem

$$\max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big) \qquad (2 \leq m < n)$$

is strongly NP-hard.

- Unless $\mathsf{P} = \mathsf{NP}$, there is no poly. time algorithm which, for any $\varepsilon > 0$, computes an admissible deterministic function $f_\varepsilon \in \mathcal{F}_{n,m}$ such that

$$H_\alpha\big(f_\varepsilon(X)\big) \geq (1 - \varepsilon) \max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big).$$

Good News

We can efficiently construct (by the use of Huffman algorithm) an admissible function $f^* \in \mathcal{F}_{n,m}$ s.t.

$$H_\alpha\bigl(f^*(X)\bigr) \geq \max_{f \in \mathcal{F}_{n,m}} H_\alpha\bigl(f(X)\bigr) - v(\alpha), \quad \alpha > 0$$

where

$$v(\alpha) := \begin{cases} \log\left(\dfrac{\alpha-1}{2^\alpha-2}\right) - \dfrac{\alpha}{\alpha-1}\,\log\left(\dfrac{\alpha}{2^\alpha-1}\right), & \alpha \neq 1, \\ \log\left(\dfrac{2}{e\,\ln 2}\right) \approx 0.08607\,\text{bits}, & \alpha = 1. \end{cases}$$

$v\colon (0,\infty) \to (0, \log 2)$ is monotonically increasing, continuous, and

$$\lim_{\alpha \downarrow 0} v(\alpha) = 0, \qquad \lim_{\alpha \to \infty} v(\alpha) = \log 2 \text{ (1 bit)}.$$

Plot

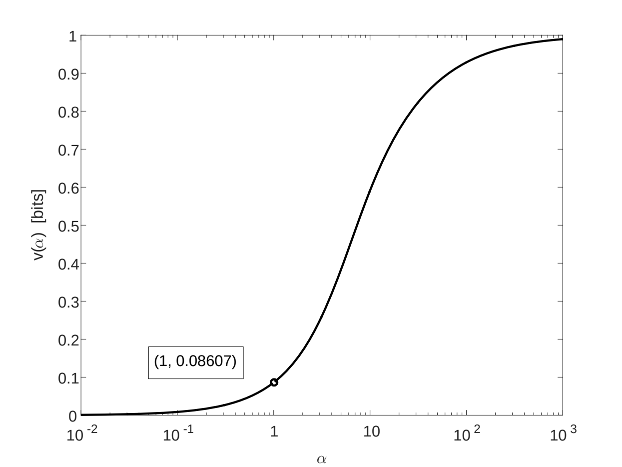


Figure: A plot of $v(\alpha)$ as a function of $\alpha > 0$.

Upper Bounding $\max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big)$

Let

- X be a discrete RV with pmf P_X, which takes n possible values, and assume that
$$P_X(1) \geq P_X(2) \geq \ldots \geq P_X(n).$$
- $f \in \mathcal{F}_{n,m}$;
- Q_X be the pmf of $f(X)$; assume that
$$\begin{aligned} &Q_X(1) \geq Q_X(2) \geq \ldots \geq Q_X(m), \\ &Q_X(m+1) = \ldots = Q_X(n) = 0. \end{aligned}$$

Then, P_X is majorized by Q_X:

$$P_X \prec Q_X \quad \left(\sum_{i=1}^{k} P_X(i) \leq \sum_{i=1}^{k} Q_X(i), \; \forall\, k \in \{1, \ldots, n\} \right).$$

Upper Bounding $\max\limits_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big)$ (Cont.)

Since the Rényi entropy is Schur-concave, it follows that

$$H_\alpha(X) = H_\alpha(P_X) \geq H_\alpha(Q_X) = H_\alpha\big(f(X)\big).$$

Upper Bounding $\max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big)$ (Cont.)

Since the Rényi entropy is Schur-concave, it follows that

$$H_\alpha(X) = H_\alpha(P_X) \geq H_\alpha(Q_X) = H_\alpha\big(f(X)\big).$$

Let $\mathcal{P}_n$ $(n \geq 2)$ be the set of prob. mass functions (pmfs) on $\mathcal{X}$ $(|\mathcal{X}| = n)$. Since the pmf of $f(X)$ majorizes the pmf of X, then

$$\max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big) \leq \max_{Q \in \mathcal{P}_m : P_X \prec Q} H_\alpha(Q).$$

Upper Bounding $\max\limits_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big)$ (Cont.)

Since the Rényi entropy is Schur-concave, it follows that

$$H_\alpha(X) = H_\alpha(P_X) \geq H_\alpha(Q_X) = H_\alpha\big(f(X)\big).$$

Let $\mathcal{P}_n$ $(n \geq 2)$ be the set of prob. mass functions (pmfs) on $\mathcal{X}$ $(|\mathcal{X}| = n)$. Since the pmf of $f(X)$ majorizes the pmf of X, then

$$\max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big) \leq \max_{Q \in \mathcal{P}_m : P_X \prec Q} H_\alpha(Q).$$

- The max. in the right side of the last ineq. is explicitly calculated.
- A function $f^* \in \mathcal{F}_{n,m}$ is efficiently constructed such that

$$H_\alpha\big(f^*(X)\big) \geq \max_{Q \in \mathcal{P}_m : P_X \prec Q} H_\alpha(Q) - v(\alpha) \geq \max_{f \in \mathcal{F}_{n,m}} H_\alpha\big(f(X)\big) - v(\alpha).$$

Solving the Maximum Rényi Entropy Problem

$$\max_{Q \in \mathcal{P}_m : P_X \prec Q} H_\alpha(Q)$$

with $X \in \{1, \ldots, n\}$, $m < n$, and $\alpha > 0$.

Solution: $Q = R_m(P_X)$

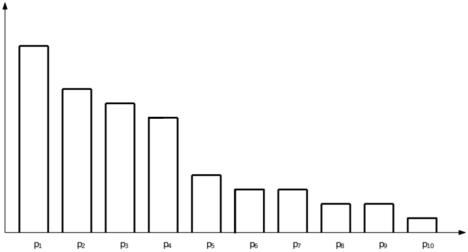
- If $P_X(1) < \frac{1}{m}$, then $R_m(P_X)$ is the equiprobable dist. on $\{1, \ldots, m\}$;
- Otherwise, $R_m(P_X) := Q_X \in \mathcal{P}_m$ with

$$Q_X(i) = \begin{cases} P_X(i), & i \in \{1, \ldots, n^*\}, \\ \dfrac{1}{m - n^*} \displaystyle\sum_{j=n^*+1}^{n} P_X(j), & i \in \{n^* + 1, \ldots, m\}, \end{cases}$$

where n^* is the max. integer i s.t. $P_X(i) \geq \frac{1}{m-i} \sum_{j=i+1}^{n} P_X(j)$.

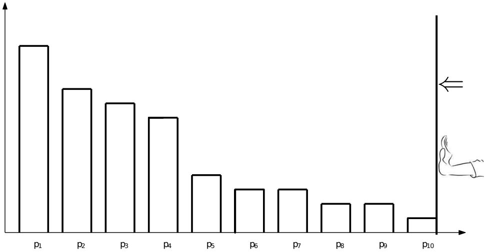
Intuitively

p



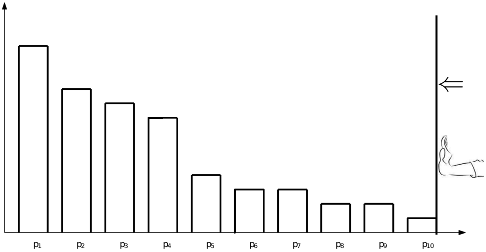
Intuitively

p

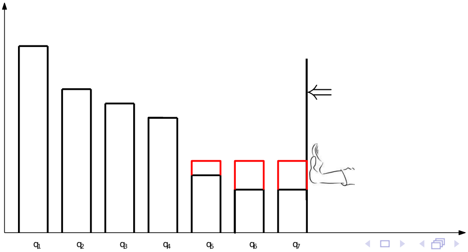


Intuitively

p



$R(\mathbf{p})$



The Algorithm by Huffman Coding

1. Start from the pmf $P_X \in \mathcal{P}_n$ with $P_X(1) \geq \ldots \geq P_X(n)$;
2. Merge successively pairs of probability masses by applying the Huffman algorithm;
3. Stop the process in Step 2 when a probability mass function $Q \in \mathcal{P}_m$ is obtained (with $Q(1) \geq \ldots \geq Q(m)$);
4. Construct the deterministic function $f^* \in \mathcal{F}_{n,m}$ by setting $f^*(k) = j \in \{1, \ldots, m\}$ for all probability masses $P_X(k)$, with $k \in \{1, \ldots, n\}$, being merged in Steps 2–3 into the node of $Q(j)$.

Theorem: Guessing Moments

Let

- $\{X_i\}_{i=1}^k$ be i.i.d. with $X_1 \sim P_X$ taking values on a set $\mathcal{X}$, $|\mathcal{X}| = n$;
- $Y_i = f(X_i)$, for every $i \in \{1, \ldots, k\}$, where $f \in \mathcal{F}_{n,m}$ is a deterministic function with $m < n$;
- $$g_{X^k} \colon \mathcal{X}^k \to \{1, \ldots, n^k\}, \quad g_{Y^k} \colon \mathcal{Y}^k \to \{1, \ldots, m^k\}$$

  be, respectively, ranking functions of the random vectors

  $$X^k := (X_1, \ldots, X_k), \quad Y^k := (Y_1, \ldots, Y_k).$$

Notation

For $m \in \{2, \ldots, n\}$, let

$$\widetilde{X}_m \sim R_m(P_X).$$

Theorem: Guessing Moments (Cont.)

1. For every deterministic function $f \in \mathcal{F}_{n,m}$, and for all $\rho > 0$,

$$\frac{1}{k} \log \frac{\mathbb{E}\big[g_{X^k}^{\rho}(X^k)\big]}{\mathbb{E}\big[g_{Y^k}^{\rho}(Y^k)\big]} \geq \rho \left[H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(\widetilde{X}_m) \right] - \frac{\rho \log(1 + k \ln n)}{k}.$$

Theorem: Guessing Moments (Cont.)

1. For every deterministic function $f \in \mathcal{F}_{n,m}$, and for all $\rho > 0$,

$$\frac{1}{k} \log \frac{\mathbb{E}\big[g_{X^k}^{\rho}(X^k)\big]}{\mathbb{E}\big[g_{Y^k}^{\rho}(Y^k)\big]} \geq \rho \left[H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(\widetilde{X}_m) \right] - \frac{\rho \log(1 + k \ln n)}{k}.$$

2. For the deterministic function $f^* \in \mathcal{F}_{n,m}$, constructed by Huffman algorithm, with $Y_i = f^*(X_i)$ for all $i \in \{1, \ldots, k\}$, we have

$$I\big(X;\, f^*(X)\big) \geq \max_{f \in \mathcal{F}_{n,m}} I\big(X;\, f(X)\big) - 0.08607 \text{ bits},$$

and, for all $\rho > 0$,

$$\begin{aligned} &\frac{1}{k} \log \frac{\mathbb{E}\big[g_{X^k}^{\rho}(X^k)\big]}{\mathbb{E}\big[g_{Y^k}^{\rho}(Y^k)\big]} \\ &\leq \rho \left[H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(\widetilde{X}_m) + v\left(\frac{1}{1+\rho}\right) \right] + \frac{\rho \log(1 + k \ln m)}{k}. \end{aligned}$$

Theorem: Guessing Moments (Cont.)

For every $\rho > 0$,

3. The gap between the universal lower bound on $\frac{1}{k} \log \frac{\mathbb{E}\left[g^{\rho}_{X^k}(X^k)\right]}{\mathbb{E}\left[g^{\rho}_{Y^k}(Y^k)\right]}$, for all $f \in \mathcal{F}_{n,m}$ (with $Y_i = f(X_i)$), and the upper bound with the specific function $f = f^* \in \mathcal{F}_{n,m}$, is at most

$$\rho\, v\left(\frac{1}{1+\rho}\right) + \frac{2\rho \log(1 + k \log_{\mathrm{e}} n)}{k} \approx \frac{0.08607\, \rho}{1+\rho} + O\left(\frac{\log k}{k}\right) \text{ bits}$$

while $f = f^*$ also almost achieves the maximal mutual information of $I\big(X; f(X)\big)$ up to a difference of 0.08607 bits.

Letting $k \to \infty$, the gap in the normalized ratio of the ρ-th guessing moments is less than 0.08607 bits for all $\rho > 0$, and the construction of the function $f^* \in \mathcal{F}_{n,m}$ does not depend on ρ.

Ongoing Activity in Guessing Problems

The topic of guessing from an IT perspective is very active these days.

- Noisy guesses (N. Merhav, arXiv:1910.00215).
- Asymptotic analysis of card guessing with feedback (P. Liu, arXiv:1908.07718).
- A unified framework for problems on guessing, source coding and task partitioning (A. Kumar et al., arXiv:1907.06889).
- Guessing individual sequences using finite-state machines (N. Merhav, arXiv:1906.10857).
- Optimal guessing under non-extensive framework and associated moment bounds (A. Ghosh, arXiv:1905.07729).
- Guessing probability in quantum key distribution (X. Wang et al., arXiv:1904.12075).
- Guessing random additive noise decoding with soft detection symbol reliability information (K. Duffey and M. Medard, arXiv:1902.03796).

Thanks for this Wonderful Workshop