Adder MAC and estimates for Rényi entropy

Ganesh Ajjanagadde, Yury Polyanskiy

Department of EECS, MIT, Cambridge, MA, 02139
email: {gajjanag,yp}@mit.edu

September 30, 2015

Overview

1. Definition of the binary adder MAC and motivation
2. The Rényi entropy conjecture
3. The difficulty of "single-letterization" of Rényi entropy
4. Approaches towards "single-letterization" of Rényi entropy
5. Proof
6. Future work

The binary adder MAC

- The binary adder MAC over n channel uses:

Definition

$$Y^n = A^n + B^n \quad \textit{where}$$

$$A^n \perp\!\!\!\perp B^n \in \{0,1\}^n, \quad Y^n \in \{0,1,2\}^n$$

The binary adder MAC

- The binary adder MAC over n channel uses:

Definition

$$Y^n = A^n + B^n \quad \textit{where}$$

$$A^n \perp\!\!\!\perp B^n \in \{0,1\}^n, \quad Y^n \in \{0,1,2\}^n$$

- Capacity region [Ahlswede, 1971, Liao, 1972]:

$$R_1 \leq \log 2,\ R_2 \leq \log 2,\ R_1 + R_2 \leq 1.5 \log 2\,.$$

The binary adder MAC

- The binary adder MAC over n channel uses:

Definition

$$Y^n = A^n + B^n \quad where$$

$$A^n \perp\!\!\!\perp B^n \in \{0,1\}^n, \quad Y^n \in \{0,1,2\}^n$$

- Capacity region [Ahlswede, 1971, Liao, 1972]:

$$R_1 \leq \log 2,\ R_2 \leq \log 2,\ R_1 + R_2 \leq 1.5 \log 2\,.$$

- $R_1 + R_2 \leq 1.5 \log 2$ is really

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log M^*(n, \epsilon) = 1.5 \log 2\,,$$

where ϵ is the "average probability of error" and $M^*(n, \epsilon)$ is the cardinality of the maximum codebook.

More refined capacity estimates

- Improved converse [Dueck, 1981, Ahlswede, 1982]:

$$\log M^*(n, \epsilon) \leq \frac{3n}{2} \log 2 + K_\epsilon \sqrt{n} \log n \, .$$

More refined capacity estimates

- Improved converse [Dueck, 1981, Ahlswede, 1982]:

$$\log M^*(n, \epsilon) \leq \frac{3n}{2} \log 2 + K_\epsilon \sqrt{n} \log n \, .$$

- Random coding achievability bound via [Polyanskiy et al., 2010]:

$$\log M^*(n, \epsilon) \geq \frac{3n}{2} \log 2 - \frac{\sqrt{n}}{2} Q^{-1}(\epsilon) + O(\log n) \, ,$$

where $Q^{-1}(\epsilon)$ is the normal quantile function.

More refined capacity estimates

- Improved converse [Dueck, 1981, Ahlswede, 1982]:

$$\log M^*(n, \epsilon) \leq \frac{3n}{2} \log 2 + K_\epsilon \sqrt{n} \log n \,.$$

- Random coding achievability bound via [Polyanskiy et al., 2010]:

$$\log M^*(n, \epsilon) \geq \frac{3n}{2} \log 2 - \frac{\sqrt{n}}{2} Q^{-1}(\epsilon) + O(\log n) \,,$$

  where $Q^{-1}(\epsilon)$ is the normal quantile function.

- Mismatch of $\log n$ between the achievability and converse.

Our motivation

- Suppose the random coding achievability bound has the right second order term. So far, $\sqrt{n}$ second order terms are attributed to the i.i.d nature of the noise cf. [Polyanskiy et al., 2010, Strassen, 1962]. The adder MAC lacks channel noise, demonstrating the fundamental importance of random coding for communication.

Our motivation

- Suppose the random coding achievability bound has the right second order term. So far, $\sqrt{n}$ second order terms are attributed to the i.i.d nature of the noise cf. [Polyanskiy et al., 2010, Strassen, 1962]. The adder MAC lacks channel noise, demonstrating the fundamental importance of random coding for communication.
- Can we tighten the converse bound from $\sqrt{n} \log n$ to $\sqrt{n}$?

Outline

Our approach to tightening the converse

- In order to tighten the converse, it suffices to maximize Rényi mutual information K_α (defined in [Csiszár, 1995]).
- As adder MAC is noiseless, K_α coincides with Rényi entropy H_α:

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha} \log \sum_x [P_X(x)]^\alpha .$$

Our approach to tightening the converse

- In order to tighten the converse, it suffices to maximize Rényi mutual information K_α (defined in [Csiszár, 1995]).
- As adder MAC is noiseless, K_α coincides with Rényi entropy H_α:

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha} \log \sum_x [P_X(x)]^\alpha .$$

- We thus propose:

Conjecture

For any $A^n \perp\!\!\!\perp B^n$ taking values in $\{0,1\}^n$

$$H_\alpha(A^n + B^n) \le nH_\alpha(Y^*) \quad \forall \alpha \in [0,1] \tag{1}$$

where $P_{Y^*} = [\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$.

Our approach to tightening the converse

Intuition

H_α for $\alpha \approx 1$ gives us the second order variations of entropy. $\alpha = 1$ gives the capacity region, $\alpha \approx 1$ the second order terms.

- Assuming Conjecture 1, data processing for Rényi mutual information K_α ([Polyanskiy and Verdú, 2010, (32),(60)]) at $\alpha = 1 - \frac{1}{\sqrt{n}}$ gives us:

$$\log M^*(n, \epsilon) \leq \frac{3n}{2} \log 2 + O(\sqrt{n}),$$

an improvement over the existing $O(\sqrt{n} \log n)$ second order term.

Evidence for the Rényi entropy conjecture

- Analytical: proof for $n = 1$, $n = 2, \alpha \leq 0.5$. Proofs make heavy use of majorization.
- Numerical: tested up to $n = 7$ using various optimization toolboxes. Problem is non-convex, so no guarantees at the moment.

Outline

The difficulty of using Rényi entropy

- Shannon entropy satisfies:

$$H(P_{XY}) \le H(P_X) + H(P_Y)\,\text{(subadditivity)}$$

- Rényi entropy ($\forall \alpha \notin \{0, 1\}$) does not, cf. [Aczél and Daróczy, 1975]:

$$H_\alpha(P_{XY}) \not\le H_\alpha(P_X) + H_\alpha(P_Y).$$

Even worse, $\forall \alpha \in (0, 1)$, one can fix $H_\alpha(P_X)$ and $H_\alpha(P_Y)$ and make $H_\alpha(P_{XY}) \nearrow \infty$. [Kovacevic et al., 2013]

The difficulty of using Rényi entropy

- Shannon entropy satisfies:

$$H(P_{XY}) \le H(P_X) + H(P_Y) \,(\text{subadditivity})$$

- Rényi entropy ($\forall \alpha \notin \{0, 1\}$) does not, cf. [Aczél and Daróczy, 1975]:

$$H_\alpha(P_{XY}) \not\le H_\alpha(P_X) + H_\alpha(P_Y).$$

  Even worse, $\forall \alpha \in (0, 1)$, one can fix $H_\alpha(P_X)$ and $H_\alpha(P_Y)$ and make $H_\alpha(P_{XY}) \nearrow \infty$. [Kovacevic et al., 2013]
- Subadditivity is the most natural induction ("single-letterization") tool we know, so how can we circumvent this?

First attempt

An estimate resembling sub-additivity obtained via Minkowski's inequality:

Theorem

$$H_\alpha(P_{XY}) \le H_\alpha(P_X) + H_{\frac{1}{\alpha}}(P_{Y_\alpha}) \, (\forall \alpha \ge 0) \, ,$$

where

$$\mathbb{P}[Y_\alpha = y] = \sum_x P_{XY}(x, y)^\alpha \exp[(\alpha - 1) H_\alpha(X, Y)] \, .$$

- P_{Y_α} is the **marginal of tilted joint distribution**.

First attempt

An estimate resembling sub-additivity obtained via Minkowski's inequality:

Theorem

$$H_\alpha(P_{XY}) \leq H_\alpha(P_X) + H_{\frac{1}{\alpha}}(P_{Y_\alpha})\,(\forall \alpha \geq 0)\,,$$

where

$$\mathbb{P}[Y_\alpha = y] = \sum_x P_{XY}(x,y)^\alpha \exp[(\alpha - 1)H_\alpha(X,Y)]\,.$$

- P_{Y_α} is the **marginal of tilted joint distribution**.
- Unfortunately, what we really need is **tilted marginal of joint distribution** for induction ☹.

A more sophisticated attempt

- Tilting towards a uniform distribution increases the entropy.

A more sophisticated attempt

- Tilting towards a uniform distribution increases the entropy.
- (roughly) How much tilt on the marginals do we need in order to restore subadditivity?

A more sophisticated attempt

- Tilting towards a uniform distribution increases the entropy.
- (roughly) How much tilt on the marginals do we need in order to restore subadditivity?

Definition

Let $P = (p_1, p_2, \ldots, p_N)$ *denote a probability vector on* N *atoms. Define* $P^\beta = \frac{1}{Z}\left(p_1^\beta, p_2^\beta, \ldots, p_N^\beta\right)$. *Call this the "*$\beta$*-tilt of* P*".*

Definition

Define the set of allowable tilts:

$$T_{\alpha,n} = \{\beta : \forall X^n \; H_\alpha(X^n) \leq \sum_{i=1}^{n} H_\alpha({P_{X_i}}^\beta)\}\,.$$

Properties of $T_{\alpha,n}$

- $0 \in T_{\alpha,n}$ since $P^{\beta=0}$ is a uniform distribution.

Properties of $T_{\alpha,n}$

- $0 \in T_{\alpha,n}$ since $P^{\beta=0}$ is a uniform distribution.
- $T_{\alpha,n} = [0, t)$ or $[0, t]$. This follows from:

Lemma

For all $\alpha \geq 0$, $H_\alpha(P^\beta)$ is non-increasing with β. Moreover, it is strictly decreasing unless P is uniform.

Properties of $T_{\alpha,n}$

- $0 \in T_{\alpha,n}$ since $P^{\beta=0}$ is a uniform distribution.
- $T_{\alpha,n} = [0,t)$ or $[0,t]$. This follows from:

Lemma

For all $\alpha \geq 0$, $H_\alpha(P^\beta)$ is non-increasing with β. Moreover, it is strictly decreasing unless P is uniform.

Theorem

$$\forall \alpha \in (0,1) \quad \sup(T_{\alpha,n}) \in \left[0, \frac{1}{n-(n-1)\alpha}\right]. \tag{2}$$

$$\forall \alpha \in (1,\infty) \quad \sup(T_{\alpha,n}) \in \left[\frac{1}{\alpha}, \frac{1}{n} + \frac{n-1}{n\alpha}\right]. \tag{3}$$

Properties of $T_{\alpha,n}$

- $0 \in T_{\alpha,n}$ since $P^{\beta=0}$ is a uniform distribution.
- $T_{\alpha,n} = [0,t)$ or $[0,t]$. This follows from:

Lemma

For all $\alpha \geq 0$, $H_\alpha(P^\beta)$ is non-increasing with β. Moreover, it is strictly decreasing unless P is uniform.

Theorem

$$\forall \alpha \in (0,1) \quad \sup(T_{\alpha,n}) \in \left[0, \frac{1}{n-(n-1)\alpha}\right]. \tag{2}$$

$$\forall \alpha \in (1,\infty) \quad \sup(T_{\alpha,n}) \in \left[\frac{1}{\alpha}, \frac{1}{n} + \frac{n-1}{n\alpha}\right]. \tag{3}$$

- Essentially a negative result

Proof strategy

- Lower bounds of the intervals are easy: the one for $0 < \alpha < 1$ corresponds to the tilted distribution at $\beta = 0$ becoming uniform.

Proof strategy

- Lower bounds of the intervals are easy: the one for $0 < \alpha < 1$ corresponds to the tilted distribution at $\beta = 0$ becoming uniform.
- Strategy for upper bounds: fix the marginals, and try to find a "maximal Rényi entropy coupling (joint)". Also pick suitable marginals, and study the asymptotics as alphabet size $N \to \infty$.

Definition

$\mathcal{C}(P_{X_1}, P_{X_2}, \ldots, P_{X_n})$ is called the set of couplings with marginals $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$. It consists of all joint distributions P_{X^n} whose marginals are $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$.

Definition

$P^*_{X^n} = \underset{P_{X^n} \in \mathcal{C}(P_{X_1}, P_{X_2}, \ldots, P_{X_n})}{\arg\max} H_\alpha(P_{X^n})$ is called a maximal Rényi entropy coupling of $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$ and order α.

Some special couplings

- For $\alpha = 1$, $P^*_{X^n} = P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_n}$.

Some special couplings

- For $\alpha = 1$, $P^*_{X^n} = P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_n}$.
- For $\alpha = 2$ (collision entropy), we have (roughly):

$$\begin{aligned} P^* = {} & U_1 \otimes P_{X_2} \otimes P_{X_3} \otimes \cdots \otimes P_{X_n} \\ & + P_{X_1} \otimes U_2 \otimes P_{X_3} \otimes \cdots \otimes P_{X_n} \\ & + \\ & \vdots \\ & + P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_{n-1}} \otimes U_n \\ & - (n-1) U_1 \otimes U_2 \otimes \cdots \otimes U_n. \end{aligned} \tag{4}$$

Here, U_i denote uniform random variables over the respective alphabets of the P_{X_i}. Proof follows from KKT conditions.

Some special couplings

- For $\alpha = 1$, $P^*_{X^n} = P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_n}$.
- For $\alpha = 2$ (collision entropy), we have (roughly):

$$\begin{aligned} P^* &= U_1 \otimes P_{X_2} \otimes P_{X_3} \otimes \cdots \otimes P_{X_n} \\ &\quad + P_{X_1} \otimes U_2 \otimes P_{X_3} \otimes \cdots \otimes P_{X_n} \\ &\quad + \\ &\quad \vdots \\ &\quad + P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_{n-1}} \otimes U_n \\ &\quad - (n-1) U_1 \otimes U_2 \otimes \cdots \otimes U_n. \end{aligned} \tag{4}$$

  Here, U_i denote uniform random variables over the respective alphabets of the P_{X_i}. Proof follows from KKT conditions.
- We use this coupling in characterizing $T_{\alpha,n}$ for $\alpha \in (1, \infty)$.

Some special marginals I

- Let $P = (p_1, p_1, \ldots, p_1, \frac{p_1}{N}, \frac{p_1}{N}, \ldots, \frac{p_1}{N})$ where p_1 occurs $M = N^{1-\beta}$ times, and normalization is ensured by taking $p_1 = \frac{N}{MN+N-M}$. Then,

$$H_\alpha(P) = (1-\beta)\log(N) + O(1) \quad \forall \alpha \in [\beta, \infty]\,.$$

Some special marginals I

- Let $P = (p_1, p_1, \ldots, p_1, \frac{p_1}{N}, \frac{p_1}{N}, \ldots, \frac{p_1}{N})$ where p_1 occurs $M = N^{1-\beta}$ times, and normalization is ensured by taking $p_1 = \frac{N}{MN+N-M}$. Then,

$$H_\alpha(P) = (1-\beta)\log(N) + O(1) \quad \forall \alpha \in [\beta, \infty] \,.$$

- Used as the marginal for proving $T_{\infty,n} = \{0\}$.

Some special marginals II

- Let $P = (p_1, p_1, \ldots, p_1, \frac{n-1}{nN}, \frac{n-1}{nN}, \ldots, \frac{n-1}{nN})$ where n is held fixed, where p_1 occurs $M = N^\gamma$ times, and normalization is ensured by taking $p_1 = \frac{1}{nM} + \frac{n-1}{nN}$. Then,

$$H_\alpha(P) = \gamma \log(N) + O(1) \quad \alpha \in (1, \infty].$$
$$H_\alpha(P) = \log(N) + O(1) \quad \alpha \in [0, 1).$$

Some special marginals II

- Let $P = (p_1, p_1, \ldots, p_1, \frac{n-1}{nN}, \frac{n-1}{nN}, \ldots, \frac{n-1}{nN})$ where n is held fixed, where p_1 occurs $M = N^\gamma$ times, and normalization is ensured by taking $p_1 = \frac{1}{nM} + \frac{n-1}{nN}$. Then,

$$H_\alpha(P) = \gamma \log(N) + O(1) \quad \alpha \in (1, \infty].$$
$$H_\alpha(P) = \log(N) + O(1) \quad \alpha \in [0, 1).$$

- Although $H_\alpha(P)$ is continuous in α, this tells us that there can be an arbitrarily sharp transition at $\alpha = 1$.

Some special marginals II

- Let $P = (p_1, p_1, \ldots, p_1, \frac{n-1}{nN}, \frac{n-1}{nN}, \ldots, \frac{n-1}{nN})$ where n is held fixed, where p_1 occurs $M = N^\gamma$ times, and normalization is ensured by taking $p_1 = \frac{1}{nM} + \frac{n-1}{nN}$. Then,

$$H_\alpha(P) = \gamma \log(N) + O(1) \quad \alpha \in (1, \infty].$$
$$H_\alpha(P) = \log(N) + O(1) \quad \alpha \in [0, 1).$$

- Although $H_\alpha(P)$ is continuous in α, this tells us that there can be an arbitrarily sharp transition at $\alpha = 1$.
- Used as the marginal for characterizing $T_{\alpha,n}$ for $\alpha \in (1, \infty)$.

Proof wrapup

- The special marginals and couplings shown here are at the core of the proof.
- Using them and letting alphabet size $N \to \infty$, we obtain upper bounds on the allowable tilts that are asymptotically tight as $n \to \infty$.

Proof wrapup

- The special marginals and couplings shown here are at the core of the proof.
- Using them and letting alphabet size $N \to \infty$, we obtain upper bounds on the allowable tilts that are asymptotically tight as $n \to \infty$.
- For simplicity, above focused on $\alpha \in (1, \infty]$ case. Turns out a slight generalization of couplings from [Kovacevic et al., 2013] works for $\alpha \in (0, 1)$.

Outline

Future work

- The Rényi entropy conjecture remains open.

Future work

- The Rényi entropy conjecture remains open.
- Alternative approaches to single letterization of Rényi entropy: tilting is clearly not enough.

Future work

- The Rényi entropy conjecture remains open.
- Alternative approaches to single letterization of Rényi entropy: tilting is clearly not enough.
- Maximal Rényi entropy couplings can be used as uninformative priors instead of the standard product distribution. Evaluation of these in statistical settings is unexplored.

Summary

1. Definition of the binary adder MAC and motivation
2. The Rényi entropy conjecture
3. The difficulty of "single-letterization" of Rényi entropy
4. Approaches towards "single-letterization" of Rényi entropy
5. Proof
6. Future work

Thanks

This material is based upon work supported by the National Science Foundation CAREER award under grant agreement CCF-12-53205 and by the SuperUROP program of the Dept. of EECS at MIT.

References I

Aczél, J. and Daróczy, Z. (1975).
On measures of information and their characterizations.
New York.

Ahlswede, R. (1971).
Multi-way communication channels.
In *Proc. 1971 IEEE Int. Symp. Inf. Theory (ISIT)*, pages 23–52, Tsahkadsor, Armenia, USSR.

Ahlswede, R. (1982).
An elementary proof of the strong converse theorem for the multiple-access channel.
J. Combinatorics, Information and System Sciences, 7(3).

Csiszár, I. (1995).
Generalized cutoff rates and Renyi's information measures.
IEEE Trans. Inf. Theory, 41(1):26 –34.

References II

Dueck, G. (1981).
The strong converse to the coding theorem for the multiple–access channel.
J. Comb. Inform. Syst. Sci, 6(3):187–196.

Kovacevic, M., Stanojevic, I., and Senk, V. (2013).
On the entropy of couplings.
CoRR, abs/1303.3235.

Liao, H. (1972).
Multiple access channels.
PhD thesis, Dept. of Elect. Eng., U. of Hawaii, Honolulu, HI.

Polyanskiy, Y., Poor, H. V., and Verdú, S. (2010).
Channel coding rate in the finite blocklength regime.
IEEE Trans. Inf. Theory, 56(5):2307–2359.

References III

Polyanskiy, Y. and Verdú, S. (2010).
Arimoto channel coding converse and rényi divergence.
In *Proc. 2010 48th Allerton Conference*, pages 1327–1333, Allerton Retreat Center, Monticello, IL, USA.

Strassen, V. (1962).
Asymptotische Abschätzungen in Shannon's Informationstheorie.
In *Trans. 3d Prague Conf. Inf. Theory*, pages 689–723, Prague.