Bounding Techniques for the Intrinsic Uncertainty of Channels

Or Ordentlich
Joint work with Ofer Shayevitz

July 4th, 2014
ISIT, Honolulu, HI, USA

Motivation

DMC

- For DMCs $C = \max_{P(X)} I(X;Y)$
- Calculating $I(X;Y)$ is easy

Motivation

DMC

- For DMCs $C = \max_{P(X)} I(X;Y)$
- Calculating $I(X;Y)$ is easy

Channels with memory

- Assuming information stability [Dobrushin 1973]

$$C = \lim_{n\to\infty} \max_{P_{\mathbf{X}}} \frac{1}{n} I(\mathbf{X};\mathbf{Y})$$

- Calculating $I(\mathbf{X};\mathbf{Y})$ may be difficult

Motivation

DMC

- For DMCs $C = \max_{P(X)} I(X;Y)$
- Calculating $I(X;Y)$ is easy

Channels with memory

- Assuming information stability [Dobrushin 1973]

$$C = \lim_{n\to\infty} \max_{P_{\mathbf{X}}} \frac{1}{n} I(\mathbf{X};\mathbf{Y})$$

- Calculating $I(\mathbf{X};\mathbf{Y})$ may be difficult

- For many interesting channels $P_{\mathbf{XY}}$ has sparse support:
  -deletion, insertion, trapdoor,...

Motivation

DMC

- For DMCs $C = \max_{P(X)} I(X;Y)$
- Calculating $I(X;Y)$ is easy

Channels with memory

- Assuming information stability [Dobrushin 1973]

$$C = \lim_{n\to\infty} \max_{P_{\mathbf{X}}} \frac{1}{n} I(\mathbf{X};\mathbf{Y})$$

- Calculating $I(\mathbf{X};\mathbf{Y})$ may be difficult

- For many interesting channels $P_{\mathbf{XY}}$ has sparse support:
  -deletion, insertion, trapdoor,...

Want lower bounds on $I(\mathbf{X};\mathbf{Y})$ that are useful for such channels

Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\bar{\mathbf{X}} \sim P_{\mathbf{X}}$, $\bar{\mathbf{Y}} \sim P_{\mathbf{Y}}$, $\bar{\mathbf{X}} \perp\!\!\!\perp \bar{\mathbf{Y}}$

Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\bar{\mathbf{X}} \sim P_{\mathbf{X}}$, $\bar{\mathbf{Y}} \sim P_{\mathbf{Y}}$, $\bar{\mathbf{X}} \perp\!\!\!\perp \bar{\mathbf{Y}}$
- AEP:

$$I(\mathbf{X};\mathbf{Y}) \approx -\log\left(\Pr\left((\bar{\mathbf{X}},\bar{\mathbf{Y}}) \in \mathcal{T}\right)\right)$$

Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\bar{\mathbf{X}} \sim P_{\mathbf{X}}$, $\bar{\mathbf{Y}} \sim P_{\mathbf{Y}}$, $\bar{\mathbf{X}} \perp\!\!\!\perp \bar{\mathbf{Y}}$
- AEP:

$$I(\mathbf{X};\mathbf{Y}) \approx -\log\left(\Pr\left((\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \in \mathcal{T}\right)\right)$$

- Computing $\Pr\left((\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \in \mathcal{T}\right)$ may be difficult
- Lower bound by replacing $\mathcal{T}$ with some $\mathcal{S} \supseteq \mathcal{T}$

Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\bar{\mathbf{X}} \sim P_{\mathbf{X}}$, $\bar{\mathbf{Y}} \sim P_{\mathbf{Y}}$, $\bar{\mathbf{X}} \perp\!\!\!\perp \bar{\mathbf{Y}}$
- ~~AEP:~~

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\Pr\left((\bar{\mathbf{X}},\bar{\mathbf{Y}}) \in \mathcal{S}\right)\right)$$

- Computing $\Pr\left((\bar{\mathbf{X}},\bar{\mathbf{Y}}) \in \mathcal{T}\right)$ may be difficult
- Lower bound by replacing $\mathcal{T}$ with some $\mathcal{S} \supseteq \mathcal{T}$

Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\bar{\mathbf{X}} \sim P_{\mathbf{X}}$, $\bar{\mathbf{Y}} \sim P_{\mathbf{Y}}$, $\bar{\mathbf{X}} \perp\!\!\!\perp \bar{\mathbf{Y}}$
- ~~AEP:~~

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{Y}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

- Computing $\Pr\left((\bar{\mathbf{X}},\bar{\mathbf{Y}}) \in \mathcal{T}\right)$ may be difficult
- Lower bound by replacing $\mathcal{T}$ with some $\mathcal{S} \supseteq \mathcal{T}$

Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\bar{\mathbf{X}} \sim P_{\mathbf{X}}$, $\bar{\mathbf{Y}} \sim P_{\mathbf{Y}}$, $\bar{\mathbf{X}} \perp\!\!\!\perp \bar{\mathbf{Y}}$

- ~~AEP~~:

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{Y}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

- Computing $\Pr\left((\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \in \mathcal{T}\right)$ may be difficult
- Lower bound by replacing $\mathcal{T}$ with some $\mathcal{S} \supseteq \mathcal{T}$

- A simple choice is the **support** $\mathcal{S} \triangleq \{(\mathbf{x}, \mathbf{y}) \ : \ P_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) > 0\}$

Motivation

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{Y}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

- Bound relatively easy to compute: involves only marginals and support
- Gives reasonable results for certain "sparse" distributions

Motivation

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{Y}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

- Bound relatively easy to compute: involves only marginals and support
- Gives reasonable results for certain "sparse" distributions

Can we find better bounds that only involve marginals and support?

- Yes - replace $\mathcal{S}$ with $\bar{\mathcal{S}} \triangleq \{\mathbf{x}\in\mathcal{T}_{\mathbf{X}}, \mathbf{y}\in\mathcal{T}_{\mathbf{Y}} \;:\; P_{\mathbf{XY}}(\mathbf{x},\mathbf{y})>0\}$

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}|\mathcal{T}_{\mathbf{X}}}\mathbb{E}_{\mathbf{Y}|\mathcal{T}_{\mathbf{Y}}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

[*Diggavi & Grossglauser '01*] [*Drinea & Mitzenmacher '07*]...

Motivation

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{Y}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

- Bound relatively easy to compute: involves only marginals and support
- Gives reasonable results for certain "sparse" distributions

Can we find better bounds that only involve marginals and support?

- Yes - replace $\mathcal{S}$ with $\bar{\mathcal{S}} \triangleq \{\mathbf{x}\in\mathcal{T}_{\mathbf{X}}, \mathbf{y}\in\mathcal{T}_{\mathbf{Y}} \ : \ P_{\mathbf{XY}}(\mathbf{x},\mathbf{y})>0\}$

$$I(\mathbf{X};\mathbf{Y}) \geq -\log\left(\mathbb{E}_{\mathbf{X}|\mathcal{T}_{\mathbf{X}}}\mathbb{E}_{\mathbf{Y}|\mathcal{T}_{\mathbf{Y}}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}\right)$$

[*Diggavi & Grossglauser '01*] [*Drinea & Mitzenmacher '07*]...

Main Result

$$I(\mathbf{X};\mathbf{Y}) \geq -\mathbb{E}_{\mathbf{Y}}\log\mathbb{E}_{\mathbf{X}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}} - \mathbb{E}_{\mathbf{X}}\log\mathbb{E}_{\mathbf{Y}}\frac{\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}{\mathbb{E}_{\mathbf{X}}\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}$$

Examples

$$I(\mathbf{X};\mathbf{Y}) \geq -\mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}} - \mathbb{E}_{\mathbf{X}} \log \mathbb{E}_{\mathbf{Y}} \frac{\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}{\mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}$$

Examples

$$I(\mathbf{X};\mathbf{Y}) \geq -\mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}} - \mathbb{E}_{\mathbf{X}} \log \mathbb{E}_{\mathbf{Y}} \frac{\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}{\mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}$$

When is this bound useful?

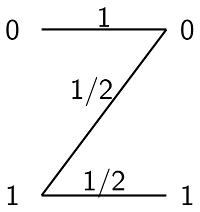
Examples

$$I(\mathbf{X};\mathbf{Y}) \geq -\mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}} - \mathbb{E}_{\mathbf{X}} \log \mathbb{E}_{\mathbf{Y}} \frac{\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}{\mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}$$

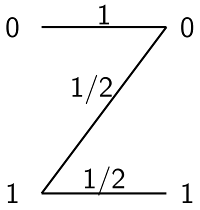
When is this bound useful?

- Not for "fully-connected" channels:
  All pairs $(\mathbf{x},\mathbf{y}) \in \mathcal{S}$ - the bound gives $I(\mathbf{X};\mathbf{Y}) \geq 0$
- Can be pretty good for channels with "low-connectivity"

Example: Z-Channel



Example: Z-Channel



Bounds for IID $\mathrm{Ber}(p)$ Input

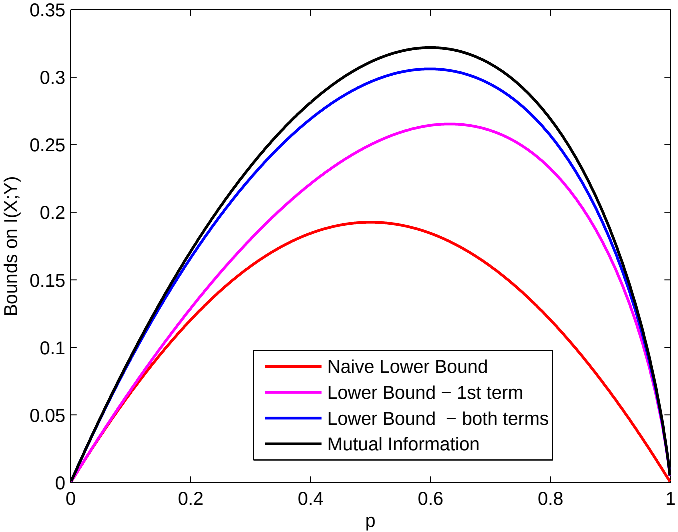
- Mutual information: $I(X;Y) = H\left(\frac{1}{2}(1+p)\right) - (1-p)$
- Naive bound:

$$I(X,Y) \geq \log\left(\mathbb{E}_X \mathbb{E}_Y \mathbb{1}_{\{(X,Y)\in\mathcal{S}\}}\right) = -\frac{1}{2}\log\left(1 - \frac{p}{2}(1-p)\right)$$

- Our bound:

$$I(X,Y) \geq -\frac{1}{2}(1-p)\log(1-p) - p\log\left(\frac{1}{2}(1+p)\right) - (1-p)\log\left(\frac{1}{2}(2+p)\right)$$

Example: Z-Channel



Preliminaries

Channels via Conditional Probability

- Channel $\Leftrightarrow$ Conditional distribution $P_{\mathbf{Y}|\mathbf{X}}$
- Input alphabet $\mathcal{X}^n$
- Output alphabet $\mathcal{Y}^*$

Preliminaries

Channels via Conditional Probability

- Channel $\Leftrightarrow$ Conditional distribution $P_{\mathbf{Y}|\mathbf{X}}$
- Input alphabet $\mathcal{X}^n$
- Output alphabet $\mathcal{Y}^*$

Channels via Actions (Functional Representation Lemma)

- $P_{\mathbf{A}}$ - a distribution over mappings $\mathcal{X}^n \mapsto \mathcal{Y}^*$
- Channel $\Leftrightarrow$ Action $\mathbf{A} \sim P_{\mathbf{A}}$, $\mathbf{A} \perp\!\!\!\perp \mathbf{X}$

$$\mathbf{Y} = \mathbf{A}(\mathbf{X})$$

- The choice of $P_{\mathbf{A}}$ is not unique

Preliminaries

The Intrinsic Uncertainty

- Input distribution $P_{\mathbf{X}}$
- $H(\mathbf{A}|\mathbf{X},\mathbf{Y})$ is the *intrinsic uncertainty*

Preliminaries

The Intrinsic Uncertainty

- Input distribution $P_{\mathbf{X}}$
- $H(\mathbf{A}|\mathbf{X},\mathbf{Y})$ is the *intrinsic uncertainty*

Capacity

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \\ &= H(\mathbf{Y}) - (H(\mathbf{Y},\mathbf{A}|\mathbf{X}) - H(\mathbf{A}|\mathbf{X},\mathbf{Y})) \\ &= H(\mathbf{Y}) - H(\mathbf{A}|\mathbf{X}) - H(\mathbf{Y}|\mathbf{A},\mathbf{X}) + H(\mathbf{A}|\mathbf{X},\mathbf{Y}) \\ &= H(\mathbf{Y}) - H(\mathbf{A}) + H(\mathbf{A}|\mathbf{X},\mathbf{Y}) \end{aligned}$$

- Lower bounding the intrinsic uncertainty = lower bounding MI

Examples for Action Sets and Intrinsic Uncertainty

The Binary Symmetric Channel

- Action $\Leftrightarrow$ IID Noise sequence $\mathbf{W} \sim \mathrm{Ber}(p)$
- $\mathbf{Y} = \mathbf{A}(\mathbf{X}) = \mathbf{X} \oplus \mathbf{W}$
- $H(\mathbf{A}|\mathbf{X}, \mathbf{Y}) = 0$

Examples for Action Sets and Intrinsic Uncertainty

The Binary Symmetric Channel

- Action $\Leftrightarrow$ IID Noise sequence $\mathbf{W} \sim \mathrm{Ber}(p)$
- $\mathbf{Y} = \mathbf{A}(\mathbf{X}) = \mathbf{X} \oplus \mathbf{W}$
- $H(\mathbf{A}|\mathbf{X}, \mathbf{Y}) = 0$

The Z-Channel

- Action $\Leftrightarrow$ IID Noise sequence $\mathbf{W} \sim \mathrm{Ber}(\frac{1}{2})$
- $Y_i = \{\mathbf{A}(\mathbf{X})\}_i = \begin{cases} X_i & X_i = 0 \\ X_i \oplus W_i & X_i = 1 \end{cases}$
- Action masked when $X_i = 0 \quad \Rightarrow \quad H(\mathbf{A}|\mathbf{X}, \mathbf{Y}) > 0$

Examples for Action Sets and Intrinsic Uncertainty

The Binary Deletion Channel

- Deletes bits independently with probability d
- Action $\Leftrightarrow$ IID deletion pattern $\mathbf{W} \sim \mathrm{Ber}(d)$
- $\mathbf{X} \mapsto \mathbf{Y}$ via many different actions $\Rightarrow$ $H(\mathbf{A}|\mathbf{X},\mathbf{Y}) > 0$
- For example: $\mathbf{x} = 01100$ and $\mathbf{y} = 110$

Examples for Action Sets and Intrinsic Uncertainty

The Binary Deletion Channel

- Deletes bits independently with probability d
- Action $\Leftrightarrow$ IID deletion pattern $\mathbf{W} \sim \mathrm{Ber}(d)$
- $\mathbf{X} \mapsto \mathbf{Y}$ via many different actions $\Rightarrow$ $H(\mathbf{A}|\mathbf{X}, \mathbf{Y}) > 0$
- For example: $\mathbf{x} = \not{0}11\not{0}0$ and $\mathbf{y} = 110$

Examples for Action Sets and Intrinsic Uncertainty

The Binary Deletion Channel

- Deletes bits independently with probability d
- Action $\Leftrightarrow$ IID deletion pattern $\mathbf{W} \sim \mathrm{Ber}(d)$
- $\mathbf{X} \mapsto \mathbf{Y}$ via many different actions $\Rightarrow$ $H(\mathbf{A}|\mathbf{X},\mathbf{Y}) > 0$
- For example: $\mathbf{x} = \not{0}110\not{0}$ and $\mathbf{y} = 110$

Examples for Action Sets and Intrinsic Uncertainty

The Binary Deletion Channel

- Deletes bits independently with probability d
- Action $\Leftrightarrow$ IID deletion pattern $\mathbf{W} \sim \mathrm{Ber}(d)$
- $\mathbf{X} \mapsto \mathbf{Y}$ via many different actions $\Rightarrow$ $H(\mathbf{A}|\mathbf{X}, \mathbf{Y}) > 0$
- For example: $\mathbf{x} = \not{0}110\not{0}$ and $\mathbf{y} = 110$

Other Channels with memory and positive intrinsic uncertainty

- Insertion channel
- Trapdoor channel
- Permutation channels
- ....

Main Tool

We would like to lower bound the intrinsic uncertainty

$$H(\mathbf{A}|\mathbf{X},\mathbf{Y}) = \mathbb{E}\log\left(\frac{1}{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}\right)$$

Main Tool

We would like to lower bound the intrinsic uncertainty

$$H(\mathbf{A}|\mathbf{X},\mathbf{Y}) = \mathbb{E}\log\left(\frac{1}{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}\right)$$

Variational Principle [Dupuis & Ellis]

For any distribution P and function $f(x)$ s.t. $|\mathbb{E}_P \log f(X)| < \infty$,

$$\mathbb{E}_P \log f(X) = \min_Q \left(\log \mathbb{E}_Q f(X) + D(P||Q)\right)$$

The minimum is uniquely attained by

$$Q^*(x) = \frac{P(x)/f(x)}{\mathbb{E}_P(1/f(x))}$$

Main Tool

We would like to lower bound the intrinsic uncertainty

$$H(\mathbf{A}|\mathbf{X},\mathbf{Y}) = \mathbb{E}\log\left(\frac{1}{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}\right)$$

Variational Principle [Dupuis & Ellis]

For any distribution P and function $f(x)$ s.t. $|\mathbb{E}_P \log f(X)| < \infty$,

$$\mathbb{E}_P \log f(X) = \min_Q \left(\log \mathbb{E}_Q f(X) + D(P||Q)\right)$$

The minimum is uniquely attained by

$$Q^*(x) = \frac{P(x)/f(x)}{\mathbb{E}_P(1/f(x))}$$

In our case $f = 1/P(\mathbf{A}|\mathbf{X},\mathbf{Y})$

A General Bound

Using the variational principle + chain rule of relative entropy + convexity of relative entropy

Theorem

The intrinsic uncertainty is lower bounded by

$$H(\mathbf{A}|\mathbf{X},\mathbf{Y}) \geq -H(\mathbf{Y}) - \mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y}) - \mathbb{E}_{\mathbf{X},\mathbf{A}} \log \mathbb{E}_{\mathbf{Y}} \frac{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}{E_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y})}$$

A General Bound

Using the variational principle + chain rule of relative entropy + convexity of relative entropy

Theorem

The intrinsic uncertainty is lower bounded by

$$I(\mathbf{X};\mathbf{Y}) \geq -H(\mathbf{A}) - \mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y}) \\ - \mathbb{E}_{\mathbf{X},\mathbf{A}} \log \mathbb{E}_{\mathbf{Y}} \frac{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}{E_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y})}$$

A General Bound

Using the variational principle + chain rule of relative entropy + convexity of relative entropy

Theorem

The intrinsic uncertainty is lower bounded by

$$I(\mathbf{X};\mathbf{Y}) \geq -H(\mathbf{A}) - \mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y}) - \mathbb{E}_{\mathbf{X},\mathbf{A}} \log \mathbb{E}_{\mathbf{Y}} \frac{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}{E_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y})}$$

- Bound's tightness depends on the choice of $P_{\mathbf{A}}$
- For BSC certain choices of $P_{\mathbf{A}}$ yield tight bounds and other choices yield $I(\mathbf{X};\mathbf{Y}) \geq 0$

A General Bound

Using the variational principle + chain rule of relative entropy + convexity of relative entropy

Theorem

The intrinsic uncertainty is lower bounded by

$$I(\mathbf{X};\mathbf{Y}) \geq -H(\mathbf{A}) - \mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y})$$
$$- \mathbb{E}_{\mathbf{X},\mathbf{A}} \log \mathbb{E}_{\mathbf{Y}} \frac{P(\mathbf{A}|\mathbf{X},\mathbf{Y})}{E_{\mathbf{X},\mathbf{A}} P(\mathbf{A}|\mathbf{X},\mathbf{Y})}$$

- Bound's tightness depends on the choice of $P_{\mathbf{A}}$
- For BSC certain choices of $P_{\mathbf{A}}$ yield tight bounds and other choices yield $I(\mathbf{X};\mathbf{Y}) \geq 0$

For a certain choice of $P_{\mathbf{A}}$ the bound becomes much simpler...

Uniform Action Set

Definition

A channel has a *uniform action set* if

$$\mathbf{A} \sim \mathrm{Uniform}(\mathcal{A})$$

Uniform Action Set

Definition

A channel has a *uniform action set* if

$$\mathbf{A} \sim \mathrm{Uniform}(\mathcal{A})$$

Theorem

For channels with uniform action set:

$$I(\mathbf{X};\mathbf{Y}) \geq -\mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}} - \mathbb{E}_{\mathbf{X}} \log \mathbb{E}_{\mathbf{Y}} \frac{\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}{\mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}$$

where

$$\mathcal{S} \triangleq \{(\mathbf{x},\mathbf{y}) \ : \ \exists \mathbf{a} \in \mathcal{A} \text{ s.t. } \mathbf{a}(\mathbf{x}) = \mathbf{y}\} = \{(\mathbf{x},\mathbf{y}) \ : \ P_{\mathbf{XY}}(\mathbf{x},\mathbf{y}) > 0\}$$

Uniform Action Set

Proposition

For each channel $P_{\mathbf{Y}|\mathbf{X}}$ there exist a uniform action set

Uniform Action Set

Proposition

For each channel $P_{\mathbf{Y}|\mathbf{X}}$ there exist a uniform action set

Proof

- Let $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{|\mathcal{A}|}\}$ be some action set
- $P_{\mathbf{A}}$ is a probability assignment on $\mathcal{A}$ consistent with $P_{\mathbf{Y}|\mathbf{X}}$
- Duplicate each action $\mathbf{a}_i$ to M_i identical actions with equal probabilities $\frac{P_{\mathbf{A}}(\mathbf{a}_i)}{M_i}$
- Choose the M_is such that all actions in the extended set are equiprobable

Uniform Action Set

Proposition

For each channel $P_{\mathbf{Y}|\mathbf{X}}$ there exist a uniform action set

Corollary (Our Main Result)

For any joint distribution $P_{\mathbf{XY}}$

$$I(\mathbf{X};\mathbf{Y}) \geq -\mathbb{E}_{\mathbf{Y}} \log \mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}} - \mathbb{E}_{\mathbf{X}} \log \mathbb{E}_{\mathbf{Y}} \frac{\mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}{\mathbb{E}_{\mathbf{X}} \mathbb{1}_{\{(\mathbf{X},\mathbf{Y})\in\mathcal{S}\}}}$$

where $\mathcal{S} \triangleq \{(\mathbf{x},\mathbf{y}) \ : \ P_{\mathbf{XY}}(\mathbf{x},\mathbf{y}) > 0\}$.

Example: Binary Deletion Channel

Capacity

- Only bounds are known
- Best lower bounds use input with memory [*Diggavi & Grossglauser '01*] [*Drinea & Mitzenmacher '07*] [*Kirsch & Drinea '10*] ...
- Some implicitly analyze the first summand in our bound

Example: Binary Deletion Channel

Capacity

- Only bounds are known
- Best lower bounds use input with memory [*Diggavi & Grossglauser '01*] [*Drinea & Mitzenmacher '07*] [*Kirsch & Drinea '10*] ...
- Some implicitly analyze the first summand in our bound

Rates for a memoryless input

- Only bounds are known
- $1 - H_2(d)$ achievable for $d \in [0, \frac{1}{2})$ [*Gallager '61*]
- Recently improved for
  - Small d [*Rahmati & Duman '13*]
  - $d \to 0$ [*Kanoria & Montanari '13*] [*Drmota et al '12*]
- And our bound?

Example: Binary Deletion Channel

New Bound (Memoryless Input)

$$\boxed{\lim_{n\to\infty} \frac{1}{n} I(\mathbf{X};\mathbf{Y}) \geq 1 - H_2(d) + g(d)}$$

where $g(d) > 0$ for all $d \in (0, \frac{1}{2})$, and is given by

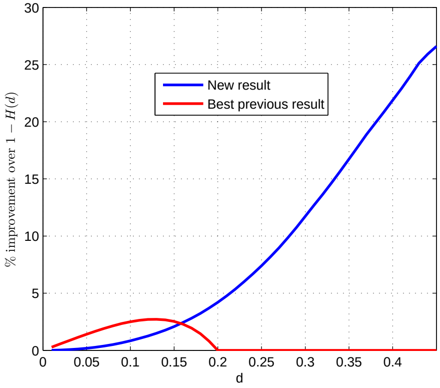
$$g(d) = \min_{\alpha\in[0,1]} \left(D_2(\alpha||1-d) - (1 - H_2(\langle\alpha\rangle)) + \Lambda^*(\alpha)\right)$$

$$\Lambda^*(\alpha) = \max_{t>0} \left(\alpha t - \frac{1}{5} \sum_{k_1} \sum_{k_2} 2^{-(k_1+k_2-1)} \log \lambda_{Z_{k_1,k_2}}(t)\right)$$

$$\lambda_{Z_i}^{k_1,k_2}(t) = 2^{k_1(t-1)} + 2^{t-1} \frac{1-2^{k_1(t-1)}}{1-2^{t-1}} \left(2^{t-1} \frac{1-2^{k_2(t-1)}}{1-2^{t-1}} + 2^{k_2(t-1)-t}\right)$$

Example: Binary Deletion Channel

% improvement over Gallager's bound $1 - H_2(d)$ (IID input):



Concluding Remarks

Summary

- Novel lower bound on $I(\mathbf{X};\mathbf{Y})$ that depends only on $P_{\mathbf{X}}, P_{\mathbf{Y}}$ and the support of $P_{\mathbf{XY}}$
- Bound is useful for channels with memory and low-connectivity
- Main tool: The Variational Principle
- For the deletion channel with IID input our bound improves best existing bounds (for some regime of d)

Concluding Remarks

Summary

- Novel lower bound on $I(\mathbf{X};\mathbf{Y})$ that depends only on $P_{\mathbf{X}}, P_{\mathbf{Y}}$ and the support of $P_{\mathbf{XY}}$
- Bound is useful for channels with memory and low-connectivity
- Main tool: The Variational Principle
- For the deletion channel with IID input our bound improves best existing bounds (for some regime of d)

Future Research

- Evaluate bound for different inputs and different channels, e.g., deletion with Markov input, trapdoor channels, etc...
- Can improve the bound to better trade-off complexity and accuracy: Replace $\mathcal{S}$ with a subset of the support whose probability approaches 1

Thanks for your attention!