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]
  \[
  C = \lim_{n \to \infty} \max_{P_X} \frac{1}{n} I(X; Y)
  \]
- Calculating $I(X; Y)$ 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]
  $$C = \lim_{n \to \infty} \max_{P_X} \frac{1}{n} I(X; Y)$$
- Calculating $I(X; Y)$ may be difficult
  - For many interesting channels $P_{XY}$ 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]
  \[ C = \lim_{n \to \infty} \max_{P_X} \frac{1}{n} I(X; Y) \]
- Calculating $I(X; Y)$ may be difficult

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

Want lower bounds on $I(X; Y)$ that are useful for such channels
Motivation

- $\mathbf{X}, \mathbf{Y}$ are random vectors with joint distribution $P_{\mathbf{XY}}$
- $\tilde{\mathbf{X}} \sim P_{\mathbf{X}}, \tilde{\mathbf{Y}} \sim P_{\mathbf{Y}}, \tilde{\mathbf{X}} \parallel \tilde{\mathbf{Y}}$
Motivation

- $X, Y$ are random vectors with joint distribution $P_{XY}$
- $\bar{X} \sim P_X$, $\bar{Y} \sim P_Y$, $\bar{X} \parallel \bar{Y}$
- AEP:

\begin{equation}
I(X; Y) \approx -\log \left( \Pr \left( (\bar{X}, \bar{Y}) \in T \right) \right)
\end{equation}
Motivation

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

AEP:

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

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

- $X, Y$ are random vectors with joint distribution $P_{XY}$
- $\bar{X} \sim P_X$, $\bar{Y} \sim P_Y$, $\bar{X} \parallel \bar{Y}$
- AEP:

\[
I(X; Y) \geq -\log \left( \Pr \left( (\bar{X}, \bar{Y}) \in S \right) \right)
\]

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

- \( \mathbf{X}, \mathbf{Y} \) are random vectors with joint distribution \( P_{XY} \)
- \( \bar{X} \sim P_X, \bar{Y} \sim P_Y, \bar{X} \parallel \bar{Y} \)
- AEP:

\[
I(\mathbf{X}; \mathbf{Y}) \geq -\log \left( \mathbb{E}_X \mathbb{E}_Y 1\{ (\mathbf{X}, \mathbf{Y}) \in S \} \right)
\]

Computing \( \Pr((\bar{X}, \bar{Y}) \in \mathcal{T}) \) may be difficult

Lower bound by replacing \( \mathcal{T} \) with some \( S \supseteq \mathcal{T} \)
Motivation

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

AEP:

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

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

- A simple choice is the support $S \triangleq \{(x, y) : P_{\mathbf{X}\mathbf{Y}}(x, y) > 0\}$
Motivation

\[ I(X; Y) \geq - \log \left( \mathbb{E}_X \mathbb{E}_Y 1_{\{ (X,Y) \in S \}} \right) \]

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

\[ I(X; Y) \geq - \log \left( \mathbb{E}_X \mathbb{E}_Y 1_{\{(X,Y) \in S\}} \right) \]

- 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 \( S \) with \( \bar{S} \triangleq \{ x \in \mathcal{T}_X, y \in \mathcal{T}_Y : P_{X|Y}(x, y) > 0 \} \)

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

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

\[ I(X; Y) \geq -\log \left( \mathbb{E}_X \mathbb{E}_Y 1_{\{(X,Y)\in S\}} \right) \]

- 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 \( S \) with \( \bar{S} \triangleq \{ x \in \mathcal{T}_X, y \in \mathcal{T}_Y : P_{XY}(x,y) > 0 \} \)

\[ I(X; Y) \geq -\log \left( \mathbb{E}_{X|\mathcal{T}_X} \mathbb{E}_{Y|\mathcal{T}_Y} 1_{\{(X,Y)\in \bar{S}\}} \right) \]

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

Main Result

\[ I(X; Y) \geq -\mathbb{E}_Y \log \mathbb{E}_X 1_{\{(X,Y)\in S\}} - \mathbb{E}_X \log \mathbb{E}_Y \frac{1_{\{(X,Y)\in S\}}}{\mathbb{E}_X 1_{\{(X,Y)\in S\}}} \]
Examples

\[
l(X; Y) \geq -\mathbb{E}_Y \log \mathbb{E}_X 1\{ (X, Y) \in S \} - \mathbb{E}_X \log \mathbb{E}_Y \frac{1\{ (X, Y) \in S \}}{\mathbb{E}_X 1\{ (X, Y) \in S \}}\]

Or Ordentlich and Ofer Shayevitz
Bounding Techniques for the Intrinsic Uncertainty of Channels
Examples

\[ I(X; Y) \geq -\mathbb{E}_Y \log \mathbb{E}_X 1\{(X,Y)\in S\} - \mathbb{E}_X \log \mathbb{E}_Y \frac{1\{(X,Y)\in S\}}{\mathbb{E}_X 1\{(X,Y)\in S\}} \]

When is this bound useful?
Examples

\[ I(X; Y) \geq -\mathbb{E}_Y \log \mathbb{E}_X 1_{\{(X,Y)\in S\}} - \mathbb{E}_X \log \mathbb{E}_Y \frac{1_{\{(X,Y)\in S\}}}{\mathbb{E}_X 1_{\{(X,Y)\in S\}}} \]

When is this bound useful?

- **Not for “fully-connected” channels:**
  All pairs \((x, y) \in S\) - the bound gives \(I(X; Y) \geq 0\)

- **Can be pretty good for channels with “low-connectivity”**
Example: Z-Channel

\[
\begin{array}{c}
0 \quad 1 \quad 0 \\
\downarrow \quad 1/2 \\
1 \quad 1/2 \quad 1
\end{array}
\]
Example: Z-Channel

Bounds for IID \(\text{Ber}(p)\) Input

- Mutual information: \(I(X; Y) = H\left(\frac{1}{2}(1 + p)\right) - (1 - p)\)
- Naive bound:
  \[
  I(X, Y) \geq \log\left(\mathbb{E}_X \mathbb{E}_Y \mathbb{1}_{\{(X, Y) \in S\}}\right) = -\frac{1}{2} \log\left(1 - \frac{p}{2}(1 - p)\right)
  \]
- 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

Bounds on $I(X;Y)$

- Naive Lower Bound
- Lower Bound – 1st term
- Lower Bound – both terms
- Mutual Information
Channels via Conditional Probability

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

Channels via Conditional Probability

- Channel \iff Conditional distribution \( P_{Y|X} \)
- Input alphabet \( \mathcal{X}^n \)
- Output alphabet \( \mathcal{Y}^* \)

Channels via Actions (Functional Representation Lemma)

- \( P_A \) - a distribution over mappings \( \mathcal{X}^n \rightarrow \mathcal{Y}^* \)
- Channel \iff Action \( A \sim P_A, A \| X \)

\[ Y = A(X) \]

- The choice of \( P_A \) is not unique
Preliminaries

The Intrinsic Uncertainty

- Input distribution $P_X$
- $H(A|X,Y)$ is the *intrinsic uncertainty*
Preliminaries

The Intrinsic Uncertainty

- Input distribution $P_X$
- $H(A|X, Y)$ is the intrinsic uncertainty

Capacity

$$I(X; Y) = H(Y) - H(Y|X)$$
$$= H(Y) - (H(Y, A|X) - H(A|X, Y))$$
$$= H(Y) - H(A|X) - H(Y|A, X) + H(A|X, Y)$$
$$= H(Y) - H(A) + H(A|X, Y)$$

- Lower bounding the intrinsic uncertainty $= $ lower bounding MI
Examples for Action Sets and Intrinsic Uncertainty

The Binary Symmetric Channel

- Action $\iff$ IID Noise sequence $W \sim \text{Ber}(p)$
- $Y = A(X) = X \oplus W$
- $H(A|X, Y) = 0$
Examples for Action Sets and Intrinsic Uncertainty

The Binary Symmetric Channel

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

The Z-Channel

- Action $\leftrightarrow$ IID Noise sequence $\mathbf{W} \sim \text{Ber}(\frac{1}{2})$
- $Y_i = \{A(X)\}_i = \begin{cases} X_i & X_i = 0 \\ X_i \oplus W_i & X_i = 1 \end{cases}$
- Action masked when $X_i = 0 \implies H(A|X, Y) > 0$
Examples for Action Sets and Intrinsic Uncertainty

The Binary Deletion Channel

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

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

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

Other Channels with memory and positive intrinsic uncertainty

- Insertion channel
- Trapdoor channel
- Permutation channels
- ....
We would like to lower bound the intrinsic uncertainty

\[ H(A|X, Y) = \mathbb{E} \log \left( \frac{1}{P(A|X, Y)} \right) \]
Main Tool

We would like to lower bound the intrinsic uncertainty

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

Variational Principle [Dupuis & Ellis]

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

\[ \mathbb{E}_P \log f(X) = \min_Q (\log \mathbb{E}_Q f(X) + D(P||Q)) \]

The minimum is uniquely attained by

\[ Q^*(x) = \frac{P(x)/f(x)}{\mathbb{E}_P(1/f(x))} \]
We would like to lower bound the intrinsic uncertainty

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

**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 \( f = 1/P(A|X,Y) \)
A General Bound

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

**Theorem**

The intrinsic uncertainty is lower bounded by

\[
H(A|X, Y) \geq -H(Y) - \mathbb{E}_Y \log \mathbb{E}_{X,A} P(A|X, Y) \\
- \mathbb{E}_{X,A} \log \mathbb{E}_Y \frac{P(A|X, Y)}{\mathbb{E}_{X,A} P(A|X, Y)}
\]
A General Bound

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

The intrinsic uncertainty is lower bounded by

\[ I(X; Y) \geq -H(A) - \mathbb{E}_Y \log \mathbb{E}_{X,A} P(A|X, Y) \]

\[ - \mathbb{E}_{X,A} \log \mathbb{E}_Y \frac{P(A|X, Y)}{P(A|X, Y)} \]
A General Bound

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

Theorem

The intrinsic uncertainty is lower bounded by

\[ I(X; Y) \geq -H(A) - \mathbb{E}_Y \log \mathbb{E}_{X,A} P(A|X, Y) \]

\[ - \mathbb{E}_{X,A} \log \mathbb{E}_Y \frac{P(A|X, Y)}{\mathbb{E}_{X,A} P(A|X, Y)} \]

- Bound’s tightness depends on the choice of \( P_A \)
- For BSC certain choices of \( P_A \) yield tight bounds and other choices yield \( I(X; Y) \geq 0 \)
A General Bound

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

Theorem

The intrinsic uncertainty is lower bounded by

\[ I(X; Y) \geq -H(A) - \mathbb{E}_Y \log \mathbb{E}_{X,A} P(A|X,Y) \]

\[ - \mathbb{E}_{X,A} \log \mathbb{E}_Y \frac{P(A|X,Y)}{\mathbb{E}_{X,A} P(A|X,Y)} \]

- Bound’s tightness depends on the choice of \( P_A \)
- For BSC certain choices of \( P_A \) yield tight bounds and other choices yield \( I(X; Y) \geq 0 \)

For a certain choice of \( P_A \) the bound becomes much simpler...
Definition

A channel has a *uniform action set* if

\[ A \sim \text{Uniform}(A) \]
Uniform Action Set

Definition

A channel has a *uniform action set* if

\[ A \sim \text{Uniform}(A) \]

Theorem

For channels with uniform action set:

\[
I(X; Y) \geq - \mathbb{E}_Y \log \mathbb{E}_X \mathbb{1}_\{(X,Y) \in S\} - \mathbb{E}_X \log \mathbb{E}_Y \mathbb{1}_\{(X,Y) \in S\}
\]

where

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

For each channel $P_{Y|X}$ there exist a uniform action set.
Uniform Action Set

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

**Proof**
- Let $\mathcal{A} = \{a_1, \ldots, a_{|A|}\}$ be some action set
- $P_\mathcal{A}$ is a probability assignment on $\mathcal{A}$ consistent with $P_{Y|X}$
- Duplicate each action $a_i$ to $M_i$ identical actions with equal probabilities $\frac{P_\mathcal{A}(a_i)}{M_i}$
- Choose the $M_i$s such that all actions in the extended set are equiprobable
Uniform Action Set

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

Corollary (Our Main Result)
For any joint distribution $P_{XY}$

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

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

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

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

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

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

\[
\lambda_{Z_{k_1,k_2}}(t) = 2^{k_1(t-1)} + 2^{t-1} \frac{1 - 2^{k_1(t-1)}}{1 - 2^{t-1}} \left( 2^{t-1} \frac{1 - 2^{k_2(t-1)}}{1 - 2^{t-1}} + 2^{k_2(t-1)-t} \right)
\]
Example: Binary Deletion Channel

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

![Graph showing % improvement over Gallager's bound for the Binary Deletion Channel]

- Blue line: New result
- Red line: Best previous result

$d$ values range from 0 to 0.4, and % improvement values range from 0 to 30.
Concluding Remarks

Summary

- Novel lower bound on $I(X; Y)$ that depends only on $P_X, P_Y$ and the support of $P_{XY}$
- 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 $I(X; Y)$ that depends only on $P_X, P_Y$ and the support of $P_{XY}$
- 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 $S$ with a subset of the support whose probability approaches 1
Thanks for your attention!