SHARED INFORMATION

Prakash Narayan

with

Imre Csiszár, Sirin Nitinawarat, Himanshu Tyagi, Shun Watanabe
Acknowledgement

Praneeth Boda

Himanshu Tyagi ♦ Shun Watanabe
Outline

Two-terminal model: Mutual information
  Operational meaning in:
  ▶ Channel coding: channel capacity
  ▶ Lossy source coding: rate distortion function
  ▶ Binary hypothesis testing: Stein’s lemma

Interactive communication and common randomness
  ▶ Two-terminal model: Mutual information
  ▶ Multiterminal model: Shared information

Applications
Outline

Two-terminal model: Mutual information

Operational meaning in:

- Channel coding: channel capacity
- Lossy source coding: rate distortion function
- Binary hypothesis testing: Stein’s lemma

Interactive communication and common randomness

Applications
Mutual Information

*Mutual information is a measure of mutual dependence between two rvs.*
**Mutual Information**

*Mutual information is a measure of mutual dependence between two rvs.*

Let $X_1$ and $X_2$ be $\mathbb{R}$-valued rvs with joint probability distribution $P_{X_1 X_2}$.

The **mutual information** between $X_1$ and $X_2$ is

\[
I(X_1 \wedge X_2) = \begin{cases} 
\mathbb{E}_{P_{X_1 X_2}} \left[ \log \frac{dP_{X_1 X_2}}{dP_{X_1} \times P_{X_2}} (X_1, X_2) \right], & \text{if } P_{X_1 X_2} \prec P_{X_1} \times P_{X_2} \\
\infty, & \text{if } P_{X_1 X_2} \nprec P_{X_1} \times P_{X_2}
\end{cases}
\]

\[
= D \left( P_{X_1 X_2} \mid\mid P_{X_1} \times P_{X_2} \right). \quad (\text{Kullback–Leibler divergence})
\]

When $X_1$ and $X_2$ are **finite-valued**, 

\[
I(X_1 \wedge X_2) = H(X_1) + H(X_2) - H(X_1, X_2)
\]

\[
= H(X_1) - H(X_1 \mid X_2) = H(X_2) - H(X_2 \mid X_1)
\]

\[
= H(X_1, X_2) - \left[ H(X_1 \mid X_2) + H(X_2 \mid X_1) \right].
\]
Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be finite alphabets, and $W : \mathcal{X}_1 \to \mathcal{X}_2$ be a stochastic matrix.

Discrete memoryless channel (DMC):

$$W^{(n)}(x_{21}, \ldots, x_{2n} \mid x_{11}, \ldots, x_{1n}) = \prod_{i=1}^{n} W(x_{2i} \mid x_{1i}).$$
Channel Capacity

Goal: Make code rate $\frac{1}{n} \log M$ as large as possible while keeping
\[
\max_m P(\phi(X_{21}, \ldots, X_{2n}) \neq m \mid f(m))
\]
to be small, in the asymptotic sense as $n \to \infty$.

[C.E. Shannon, 1948]

Channel capacity $C = \max_{P_{X_1}:P_{X_2 \mid X_1 = W}} I(X_1 \land X_2)$. 
Let $\{X_1^t\}_{t=1}^{\infty}$ be an $\mathcal{X}_1$-valued i.i.d. source.

$$d((x_{11}, \ldots, x_{1n}), (x_{21}, \ldots, x_{2n})) = \frac{1}{n} \sum_{i=1}^{n} d(x_{1i}, x_{2i}).$$
Rate Distortion Function

\[ \text{Goal: Make (compression) code rate } \frac{1}{n} \log J \text{ as small as possible while keeping} \]

\[ P \left( \frac{1}{n} \sum_{i=1}^{n} d(X_{1i}, X_{2i}) \leq \Delta \right) \]

to be large, in the asymptotic sense as \( n \to \infty \).

[Shannon, 1948, 1959]

Rate distortion function \( R(\Delta) = \min_{P_{X_2|X_1}: \mathbb{E}[d(X_1,X_2)] \leq \Delta} I(X_1 \wedge X_2) \).
Simple Binary Hypothesis Testing

Let \( \{(X_{1t}, X_{2t})\}_{t=1}^\infty \) be an \( X_1 \times X_2 \)-valued i.i.d. process generated according to

\[
H_0 : P_{X_1 X_2} \quad \text{or} \quad H_1 : P_{X_1} \times P_{X_2}.
\]

Test:

Decides \( H_0 \) w.p. \( T(0 \mid x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}) \),

\( H_1 \) w.p. \( T(1 \mid x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}) = 1 - T(0 \mid \ldots) \).

Stein’s lemma [H. Chernoff, 1956]: For every \( 0 < \epsilon < 1 \),

\[
\lim_{n \to \infty} \frac{-1}{n} \log \inf_{T: P_{H_0}(T \text{ says } H_0) \geq 1 - \epsilon} P_{H_1}(T \text{ says } H_0)
\]

\[
= D(P_{X_1 X_2} \parallel P_{X_1} \times P_{X_2}) = I(X_1 \land X_2).
\]
Outline

Two-terminal model: Mutual information

Interactive communication and common randomness

- Two-terminal model: Mutual information
- Multiterminal model: Shared information

Applications
Set of terminals $\mathcal{M} = \{1, \ldots, m\}$.

$X_1, \ldots, X_m$ are finite-valued rvs with known joint distribution $P_{X_1 \ldots X_m}$ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m$.

Terminal $i \in \mathcal{M}$ observes data $X_i$.

Multiple rounds of interactive communication on a noiseless channel of unlimited capacity; all terminals hear all communication.
Interactive Communication

Interactive communication

- Assume: Communication occurs in consecutive time slots in \( r \) rounds.

- Communication is described in terms of the mappings

\[
f_{11}, \ldots, f_{1m}, f_{21}, \ldots, f_{2m}, \ldots, f_{r1}, \ldots, f_{rm}
\]

  - \( f_{ji} \): message in round \( j \) from terminal \( i \), \( 1 \leq j \leq r, 1 \leq i \leq m \)
  - \( f_{ji} \) is any function of \( X_i \) and of all previous communication.
Interactive Communication

Interactive communication

- Assume: Communication occurs in consecutive time slots in $r$ rounds.

- Communication is described in terms of the mappings

$$f_{11}, \ldots, f_{1m}, f_{21}, \ldots, f_{2m}, \ldots, f_{r1}, \ldots, f_{rm}$$

  - $f_{ji}$: message in round $j$ from terminal $i$, $1 \leq j \leq r$, $1 \leq i \leq m$
  - $f_{ji}$ is any function of $X_i$ and of all previous communication.

- The corresponding rvs representing the communication are

$$F = F(X_1, \ldots, X_m) = (F_{11}, \ldots, F_{1m}, F_{21}, \ldots, F_{2m}, \ldots, F_{r1}, \ldots, F_{rm})$$

  - $F_{11} = f_{11}(X_1)$, $F_{12} = f_{12}(X_2, F_{11})$, \ldots
  - $F_{ji} = f_{ji}(X_i; \text{ all previous communication})$.

Simple communication: $F = (F_1, \ldots, F_m)$, $F_i = f_i(X_i)$, $1 \leq i \leq m$. 
Applications

Data exchange: Omniscience.

Signal recovery: Data compression.

Function computation.

Cryptography: Secret key generation.

¿Applications in control?
Example: Function Computation

\[ X_1 = \begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix} \xrightarrow{F_1} \xleftarrow{F_2} X_2 = \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix} \]

[S. Watanabe]

- \( X_{11}, X_{12}, X_{21}, X_{22} \) are mutually independent \((0.5, 0.5)\) bits.
- Terminals 1 and 2 wish to compute:

\[ G = g(X_1, X_2) = \mathbb{1}\left((X_{11}, X_{12}) = (X_{21}, X_{22})\right). \]

- Simple communication: \( F = \left(F_1 = (X_{11}, X_{12}), F_2 = (X_{21}, X_{22})\right) \).
  - Communication complexity: \( H(F) = 4 \) bits.
  - No privacy: Terminal 1 or 2, or an observer of \( F \), learns all the data \( X_1, X_2 \).
Example: Function Computation

\[
X_1 = \begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix} \quad \xrightarrow{F_{11}} \quad \begin{pmatrix} F_{11} \\ F_{12} \end{pmatrix} \quad \xrightarrow{F_{21}} \quad \begin{pmatrix} F_{21} \\ F_{22} \end{pmatrix} \quad \xleftarrow{F_{21}} \quad X_2 = \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix}
\]

- **Interactive communication 1:**
  - \( F_{11} = X_{11} \oplus X_{12}, \ F_{12} = X_{21} \oplus X_{22} \).
  - If \( F_{11} \neq F_{12} \), protocol over.
  - If \( F_{11} = F_{12} \), then \( F_{21} = X_{11}, \ F_{22} = X_{21} \).
  - Complexity: \( H(\mathbf{F}) = 3 \) bits.
  - Some privacy: W.p. 0.5 both terminals, or an observer of \( \mathbf{F} \), learn that \( X_1 \neq X_2 \); and w.p. 0.5 everyone learns \( X_1, X_2 \).
Example: Function Computation

\[
X_1 = \begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix} \quad F_{11} \quad X_2 = \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix}
\]

▶ Interactive communication 2:

- \( F = \left( F_{11} = (X_{11}, X_{12}), \; F_{12} = G \right) \).
- Complexity: \( H(F) = 2.81 \) bits.
- Some privacy: Terminal 2, or an observer of \( F \), learns \( X_1 \); terminal 1, or an observer of \( F \), either learns \( X_2 \) w.p. 0.25 or w.p. 0.75 that \( X_2 \) differs from \( X_1 \).
Related Work

▶ Exact function computation
- Yao '79: Communication complexity.
- Gallager '88: Algorithm for parity computation in a network.
- Giridhar-Kumar '05: Algorithms for computing functions over sensor networks.
- Orlitsky-El Gamal '84: Communication complexity with secrecy.

▶ Information theoretic function computation
- Körner-Marton '79: Minimum rate for computing parity.
- Orlitsky-Roche '01: Two terminal function computation.
- Nazer-Gastpar '07: Computation over noisy channels.
- Ma-Ishwar '08: Distributed source coding for interactive computing.
- Ma-Ishwar-Gupta '09: Multi-round function computation in colocated networks.
- Tyagi-Watanabe '13, '14 Secrecy generation, secure computing.

▶ Compressing interactive communication
- Schulman '92: Coding for interactive communication.
- Braverman-Rao '10: Information complexity of communication.
- Kol-Raz '13, Heupler '14: Interactive communication over noisy channels.
For $0 \leq \epsilon < 1$, given interactive communication $\mathbf{F}$, an rv $L = L(X_1, \ldots, X_m)$ is $\epsilon$-CR for the terminals in $\mathcal{M}$ using $\mathbf{F}$, if there exist local estimates

$$L_i = L_i(X_i, \mathbf{F}), \ i \in \mathcal{M},$$

of $L$ satisfying

$$P\left(L_i = L, \ i \in \mathcal{M}\right) \geq 1 - \epsilon.$$
Common Randomness

\[
\begin{align*}
\text{COMMUNICATION NETWORK} & \quad \sim = L \\
X_1 & \quad X_2 \quad X_m \\
L_1 & \quad L_2 \quad L_m \quad \simeq L
\end{align*}
\]

**Examples:**

- **Omniscience:** \( L = (X_1, \ldots, X_m) \).
- **Single signal:** \( L = X_{i^*} \), for some fixed \( i^* \in \mathcal{M} \).
- **Function computation:** \( L = g(X_1, \ldots, X_m) \) for a given \( g \).
- **Secret CR, i.e., secret key:** \( L \) with \( I(L \land F) \simeq 0 \).
¿What is the maximal CR, as measured by $H(L|F)$, that can be generated by a given interactive communication $F$ for a distributed processing task?
A Basic Question

What is the maximal CR, as measured by \( H(L|F) \), that can be generated by a given interactive communication \( F \) for a distributed processing task?

Answer in two steps:

- Fundamental property of interactive communication
- Upper bound on amount of CR achievable with interactive communication.

Shall start with the case of \( m = 2 \) terminals.
**Lemma:** [U. Maurer], [R. Ahlswede - I. Csiszár]

For interactive communication $F$ of the terminals $i \in \mathcal{M} = \{1, 2\}$, with terminal $i$ possessing “initial” data $X_i$, 

$$I(X_1 \wedge X_2 | F) \leq I(X_1 \wedge X_2).$$

In particular, independent rvs $X_1, X_2$ remain so upon conditioning on an interactive communication.
**Fundamental Property of Interactive Communication**

**Lemma:** [U. Maurer], [R. Ahlswede - I. Csiszár]

For interactive communication $F$ of the terminals $i \in \mathcal{M} = \{1, 2\}$, with terminal $i$ possessing "initial" data $X_i$,

$$I(X_1 \land X_2|F) \leq I(X_1 \land X_2).$$

In particular, independent rvs $X_1, X_2$ remain so upon conditioning on an interactive communication.

**Proof:** For interactive communication $F = (F_{11}, F_{12}, \ldots, F_{r1}, F_{r2})$,

$$I(X_1 \land X_2) = I(X_1, F_{11} \land X_2)$$
$$\geq I(X_1 \land X_2|F_{11})$$
$$= I(X_1 \land X_2, F_{12}|F_{11})$$
$$\geq I(X_1 \land X_2|F_{11}, F_{12}),$$

followed by iteration. □
An Equivalent Form

For interactive communication $F$ of terminals 1 and 2:

$$I(X_1 \land X_2|F) \leq I(X_1 \land X_2)$$

$\Updownarrow$

$$H(F) \geq H(F|X_1) + H(F|X_2).$$
Upper Bound on CR for Two Terminals

Using
- $L$ is $\epsilon$-CR for $\mathcal{M} = \{1, 2\}$ with interactive $F$; and
- $H(F) \geq H(F|X_1) + H(F|X_2)$,

we get

$$H(L|F) \leq H(X_1, X_2) - \left[ H(X_1|X_2) + H(X_2|X_1) \right] + 2\nu(\epsilon),$$

where $\lim_{\epsilon \to 0} \nu(\epsilon) = 0$. 
Lemma: [I. Csiszár - P. Narayan] Let $L$ be any $\epsilon$-CR for the terminals $i \in \mathcal{M} = \{1, 2\}$ with terminal $i$ possessing “initial” data $X_i$, achievable with interactive communication $F$. Then

$$H(L|F) \leq I(X_1 \wedge X_2) + 2\nu, \quad \lim_{\epsilon \to 0} \nu(\epsilon) = 0.$$ 

Remark: When $\{(X_{1t}, X_{2t})\}_{t=1}^{\infty}$ is an $\mathcal{X}_1 \times \mathcal{X}_2$-valued i.i.d. process, the upper bound is attained.
Interactive Communication for $m \geq 2$ Terminals

**Theorem 1:** [I. Csiszár-P. Narayan]
For interactive communication $F$ of the terminals $i \in \mathcal{M} = \{1, \ldots, m\}$, with terminal $i$ possessing “initial” data $X_i$,

$$H(F) \geq \sum_{B \in \mathcal{B}} \lambda_B H(F|X_{B^c})$$

for every family $\mathcal{B} = \{B \subsetneq \mathcal{M}, B \neq \emptyset\}$ and set of weights ("fractional partition")

$$\lambda \triangleq \left\{ 0 \leq \lambda_B \leq 1, B \in \mathcal{B}, \text{satisfying } \sum_{B \in \mathcal{B}: B \ni i} \lambda_B = 1 \forall i \in \mathcal{M} \right\}.$$  

Equality holds if $X_1, \ldots, X_m$ are mutually independent.

Special case of:
CR for \( m \geq 2 \) Terminals: A Suggestive Analogy

[S. Nitinawarat-P. Narayan]

For interactive communication \( F \) of the terminals \( i \in M = \{1, \ldots, m\} \), with terminal \( i \) possessing "initial" data \( X_i \),

\[
\begin{align*}
\left( m = 2 : H(F) & \geq H(F|X_1) + H(F|X_2) \iff I(X_1 \wedge X_2|F) \leq I(X_1 \wedge X_2) \right)
\end{align*}
\]

\[
H(F) \geq \sum_{B \in B} \lambda_B H(F|X_{B^c})
\]

\[
\uparrow
\]

\[
H(X_1, \ldots, X_m|F) - \sum_{B \in B} \lambda_B H(X_B|X_{B^c}, F)
\]

\[
\leq H(X_1, \ldots, X_m) - \sum_{B \in B} \lambda_B H(X_B|X_{B^c}).
\]
An Analogy

[S. Nitinawarat-P. Narayan]

For interactive communication $F$ of the terminals $i \in M = \{1, \ldots, m\}$, with terminal $i$ possessing “initial” data $X_i$,

$$H(F) \geq \sum_{B \in \mathcal{B}} \lambda_B H(F|X_{B^c})$$

$$\Downarrow$$

$$H(X_1, \ldots, X_m|F) - \sum_{B \in \mathcal{B}} \lambda_B H(X_B|X_{B^c}, F) \leq H(X_1, \ldots, X_m) - \sum_{B \in \mathcal{B}} \lambda_B H(X_B|X_{B^c}).$$

Does the RHS suggest a measure of mutual dependence among the rvs $X_1, \ldots, X_m$?
CR for $m \geq 2$ Terminals

**Theorem 2:** [I. Csiszár-P. Narayan]

Given $0 \leq \epsilon < 1$, for an $\epsilon$-CR $L$ for $\mathcal{M}$ achieved with interactive communication $\mathbf{F}$,

$$H(L|\mathbf{F}) \leq H(X_1, \ldots, X_m) - \sum_{B \in \mathcal{B}} \lambda_B H(X_B|X_B^c) + m\nu$$

for every fractional partition $\lambda$ of $\mathcal{M}$, with $\nu = \nu(\epsilon) = \epsilon \log|\mathcal{L}| + h(\epsilon)$.

**Remarks:**

- The proof of Theorem 2 relies on Theorem 1.
- When $\{(X_{1t}, \ldots, X_{mt})\}_{t=1}^{\infty}$ is an i.i.d. process, the upper bound is attained.
Theorem 2: [I. Csiszár-P. Narayan]

\[ H(L|F) \leq H(X_1, \ldots, X_m) - \max_{\lambda} \sum_{B \in B} \lambda_B H(X_B|X_{B^c}) + m \nu \]

\[ \Delta \triangleq SI(X_1, \ldots, X_m) + m \nu \]
Extensions

Theorems 1 and 2 extend to:

- random variables with densities [S. Nitinawarat-P. Narayan]
- a larger class of probability measures [H. Tyagi-P. Narayan].
Shared Information and Kullback-Leibler Divergence


\[
SI(X_1, \ldots, X_m) = H(X_1, \ldots, X_m) - \max_\lambda \sum_{B \in B} \lambda_B H(X_B | X_{B^c})
\]

\[
(m = 2) = H(X_1, X_2) - \left[ H(X_1 | X_2) + H(X_2 | X_1) \right] = I(X_1 \land X_2)
\]

\[
(m = 2) = D(P_{X_1X_2} \| P_{X_1} \times P_{X_2})
\]
Shared Information and Kullback-Leibler Divergence


\[ SI(X_1, \ldots, X_m) = H(X_1, \ldots, X_m) - \max_{B \in \mathcal{B}} \lambda B H(X_B | X_{B^c}) \]

\[ (m = 2) = H(X_1, X_2) - \left[ H(X_1|X_2) + H(X_2|X_1) \right] = I(X_1 \land X_2) \]

\[ (m = 2) = D(P_{X_1X_2}||P_{X_1 \times X_2}) \]

\[ (m \geq 2) = \min_{2 \leq k \leq m} \min_{A_k = (A_1, \ldots, A_k)} \frac{1}{k - 1} D\left(P_{X_1 \ldots X_m}|| \prod_{i=1}^{k} P_{X_{A_i}} \right) \]

and equals 0 iff \( P_{X_1 \ldots X_m} = P_{X_A} P_{X_{Ac}} \) for some \( A \subset M \).

Does shared information have an operational significance as a measure of the mutual dependence among the rvs \( X_1, \ldots, X_m \)?
Outline

Two-terminal model: Mutual information

Interactive communication and common randomness

Applications
For $L = (X_1, \ldots, X_m)$, Theorem 2 gives

$$H(F) \geq H(X_1, \ldots, X_m) - SI(X_1, \ldots, X_m) - m\nu,$$

which, for $m = 2$, is

$$H(F) \geq H(X_1|X_2) + H(X_2|X_1) - 2\nu.$$

[Slepian – Wolf]
Recovery of a Single Signal

\[ L \]

\[ X_1 \]
\[ X_2 \]
\[ X_m \]
\[ L_1 \]
\[ L_2 \]
\[ L_m \approx L \]

[COMMUNICATION NETWORK]
\sim \ = \ L
\[ F \]

[S. Nitinawat-P. Narayan]

With \( L = X_1 \), by Theorem 2

\[ H(F) \geq H(X_1) - SI(X_1, \ldots, X_m) - m\nu, \]

which, for \( m = 2 \), gives

\[ H(F) \geq I(X_1 \wedge F) \geq H(X_1|X_2) - 2\nu. \]

[Slepian-Wolf]
Terminals \(1, \ldots, m\) generate CR \(L\) satisfying the *secrecy condition*

\[ I(L \land F) \approx 0. \]

By Theorem 2,

\[ H(L) \approx H(L|F) \leq SI(X_1, \ldots, X_m) + m\nu. \]

- Secret key generation [I. Csiszár-P. Narayan]
- Secure function computation [H. Tyagi-P. Narayan]
Shared information and a Hypothesis Testing Problem

\[ SI(X_1, \ldots, X_m) = \min_{2 \leq k \leq m} \min_{A_k=(A_1,\ldots,A_k)} \frac{1}{k-1} D\left( P_{X_1\ldots X_m} \parallel \prod_{i=1}^{k} P_{X_{A_i}} \right) \]

- Related to exponent of "\( P_e \)-second kind" for an appropriate binary composite hypothesis testing problem, involving restricted CR \( L \) and communication \( F \).

In Closing ...

¿ How useful is the concept of *shared information* ?

**A:** Operational meaning in specific cases of distributed processing ...
In Closing ... 

¿ How useful is the concept of *shared information*? 

**A:** Operational meaning in specific cases of distributed processing ...

For instance

- Consider $n$ i.i.d. repetitions (say, in time) of the rvs $X_1, \ldots, X_m$.
- Data at time instant $t$ is $X_{1t}, \ldots, X_{mt}$, $t = 1, \ldots, n$.
- Terminal $i$ observes the i.i.d. data $(X_{1i}, \ldots, X_{ni})$, $i \in \mathcal{M}$.
- Shared information-based results are asymptotically tight (in $n$):
  - *Minimum rate* of communication for omniscience.
  - *Maximum rate* of a secret key.
  - *Necessary condition* for secure function computation.
  - Several problems in information theoretic cryptography.
Shared Information: Many Many Open Questions ...

- Significance in network source and channel coding?
- Interactive communication over noisy channels?
- Communication links described by (an undirected) graph?
- Continuous-time models?

...