Trading Information Complexity for Error

Hamed Hatami

joint work with Yuval Dagan, Yuval Filmus, Yaqiao Li

School of Computer Science
McGill University

April 10, 2017

- Setting: the standard two-party communication model.
- Main question: How much one can save in information complexity by allowing an error of ϵ?

- Setting: the standard two-party communication model.
- Main question: How much one can save in information complexity by allowing an error of ϵ?
- How does $\mathrm{IC}_\mu(f, 0) - \mathrm{IC}_\mu(f, \epsilon)$ behave?

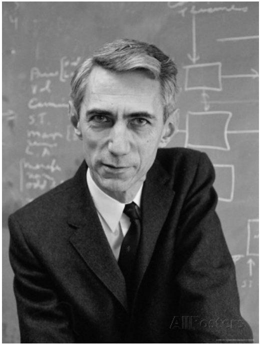
- Setting: the standard two-party communication model.
- Main question: How much one can save in information complexity by allowing an error of ϵ?
- How does $\mathrm{IC}_\mu(f, 0) - \mathrm{IC}_\mu(f, \epsilon)$ behave?
- How does $\mathrm{IC}(\mathrm{AND}, 0) - \mathrm{IC}(\mathrm{AND}, \epsilon)$ behave?

  [BGPW13]: We know $\mathrm{IC}(\mathrm{AND}, 0) \approx 1.4923$

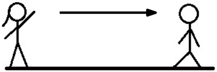
Information complexity

Continuation of Shannon's information theory.

Shannon (1916-2001)

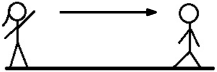


Shannon's setting



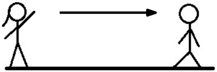
- One-way channel: Alice receives independent samples $X_1, X_2, \ldots, X_n$ of a random variable X.
- She wants to transmit them to Bob.

Shannon's setting



- One-way channel: Alice receives independent samples $X_1, X_2, \ldots, X_n$ of a random variable X.
- She wants to transmit them to Bob.
- On average how many bits does she need to send?

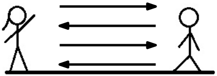
Shannon's setting



- One-way channel: Alice receives independent samples $X_1, X_2, \ldots, X_n$ of a random variable X.
- She wants to transmit them to Bob.
- On average how many bits does she need to send?
- $$\lim_{n\to\infty} \frac{C_n(X)}{n} = \mathbb{H}(X),$$

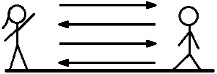
  where $\mathbb{H}(X)$ is the entropy of X and it captures the amount of information in X.

Communication complexity



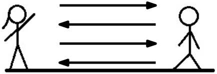
- Consider a pair of random variables $(X, Y) \sim \mu$.
- Alice receives X and Bob receives Y.

Communication complexity



- Consider a pair of random variables $(X, Y) \sim \mu$.
- Alice receives X and Bob receives Y.
- They want to compute $f(X, Y)$ collaboratively.

Communication complexity



- Consider a pair of random variables $(X, Y) \sim \mu$.
- Alice receives X and Bob receives Y.
- They want to compute $f(X, Y)$ collaboratively.
- How many bits of communication is necessary?

Communication Tasks

Let π be the communication protocol (has access to randomness).

Communication Tasks

Let π be the communication protocol (has access to randomness).

- $[f, \epsilon]$: For every x, y,

$$\Pr[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

Communication Tasks

Let π be the communication protocol (has access to randomness).

- $[f, \epsilon]$: For every x, y,

$$\Pr[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

- $[f, \mu, \epsilon]$:

$$\Pr_{xy \sim \mu}[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

Communication Tasks

Let π be the communication protocol (has access to randomness).

- $[f, \epsilon]$: For every x, y,

$$\Pr[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

- $[f, \mu, \epsilon]$:

$$\Pr_{xy \sim \mu}[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

- common notation: $R_\epsilon(f) = CC[f, \epsilon]$ and $D_\mu^\epsilon(f) = CC[f, \mu, \epsilon]$.

Communication Tasks

Let π be the communication protocol (has access to randomness).

- $[f, \epsilon]$: For every x, y,

$$\Pr[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

- $[f, \mu, \epsilon]$:

$$\Pr_{xy \sim \mu}[f(x,y) \neq \pi(x,y)] \leq \epsilon.$$

- common notation: $R_\epsilon(f) = CC[f, \epsilon]$ and $D^\epsilon_\mu(f) = CC[f, \mu, \epsilon]$.

Theorem (Yao's minimax theorem)

$$\mathrm{CC}[f, \epsilon] = \max_\mu \mathrm{CC}[f, \mu, \epsilon].$$

Information complexity

What is information complexity?

- It extends Shannon's information theory.

Information complexity

What is information complexity?

- It extends Shannon's information theory.
- Same setting as communication complexity:
  - Alice and Bob want to compute $f(X, Y)$ collaboratively.

Information complexity

What is information complexity?

- It extends Shannon's information theory.
- Same setting as communication complexity:
    - Alice and Bob want to compute $f(X, Y)$ collaboratively.
    - Now they want to minimize the information cost.

Information complexity

What is information complexity?

- It extends Shannon's information theory.
- Same setting as communication complexity:
  - Alice and Bob want to compute $f(X, Y)$ collaboratively.
  - Now they want to minimize the information cost.

Information cost of a protocol is the amount of information Alice and Bob learn from the communicated bits Π about each other's inputs.

$$\mathrm{IC}_\mu(\pi) = I(X;\Pi|Y) + I(Y;\Pi|X)$$

Information complexity of a communication task T:

$$\mathrm{IC}_\mu(T) := \inf_\pi \mathrm{IC}_\mu(\pi) \qquad \text{where } \pi \text{ performs } T$$

Information complexity of a communication task T:

$$\mathrm{IC}_\mu(T) := \inf_\pi \mathrm{IC}_\mu(\pi) \qquad \text{where } \pi \text{ performs } T$$

Examples

$$\mathrm{IC}_\mu[f,\epsilon] \qquad \text{and} \qquad \mathrm{IC}_\mu[f,\nu,\epsilon].$$

Information complexity of a communication task T:

$$\mathrm{IC}_\mu(T) := \inf_\pi \mathrm{IC}_\mu(\pi) \qquad \text{where } \pi \text{ performs } T$$

Examples

$$\mathrm{IC}_\mu[f,\epsilon] \qquad \text{and} \qquad \mathrm{IC}_\mu[f,\nu,\epsilon].$$

Theorem (Braverman-Rao'10: Amortized communication = Information complexity)

For $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{\mathrm{CC}([f,\mu,\epsilon]^n)}{n} = \mathrm{IC}_\mu[f,\mu,\epsilon].$$

Information complexity of a communication task T:

$$\mathrm{IC}_\mu(T) := \inf_\pi \mathrm{IC}_\mu(\pi) \qquad \text{where } \pi \text{ performs } T$$

Examples

$$\mathrm{IC}_\mu[f,\epsilon] \qquad \text{and} \qquad \mathrm{IC}_\mu[f,\nu,\epsilon].$$

Theorem (Braverman-Rao'10: Amortized communication = Information complexity)

For $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{\mathrm{CC}([f,\mu,\epsilon]^n)}{n} = \mathrm{IC}_\mu[f,\mu,\epsilon].$$

Theorem (Braverman'12: Amortized communication = Information complexity)

For $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{\mathrm{CC}([f,\epsilon]^n)}{n} = \max_\mu \mathrm{IC}_\mu[f,\epsilon] = \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon].$$

$\lim_{n\to\infty} \mathrm{CC}([f,\mu,\epsilon]^n)/n = \mathrm{IC}_\mu[f,\mu,\epsilon]$.
$\lim_{n\to\infty} \mathrm{CC}([f,\epsilon]^n)/n = \max_\mu \mathrm{IC}_\mu[f,\epsilon] = \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$.

$\lim_{n\to\infty} \mathrm{CC}([f,\mu,\epsilon]^n)/n = \mathrm{IC}_\mu[f,\mu,\epsilon]$.
$\lim_{n\to\infty} \mathrm{CC}([f,\epsilon]^n)/n = \max_\mu \mathrm{IC}_\mu[f,\epsilon] = \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$.

Definition (Prior-free information complexity)

$$\mathrm{IC}(f,\epsilon) := \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$$

and

$$\mathrm{IC}^D(f,\epsilon) := \max_{\mu} \mathrm{IC}_\mu[f,\mu,\epsilon]$$

$\lim_{n\to\infty} \mathrm{CC}([f,\mu,\epsilon]^n)/n = \mathrm{IC}_\mu[f,\mu,\epsilon]$.
$\lim_{n\to\infty} \mathrm{CC}([f,\epsilon]^n)/n = \max_\mu \mathrm{IC}_\mu[f,\epsilon] = \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$.

Definition (Prior-free information complexity)

$$\mathrm{IC}(f,\epsilon) := \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$$

and

$$\mathrm{IC}^D(f,\epsilon) := \max_{\mu} \mathrm{IC}_\mu[f,\mu,\epsilon]$$

Theorem (Braverman'12)

$$\mathrm{IC}(f,0) = \mathrm{IC}^D(f,0),$$

and for every $0 < \alpha < 1$,

$$\mathrm{IC}^D(f,\epsilon) \le \mathrm{IC}(f,\epsilon) \le \frac{1}{1-\alpha}\mathrm{IC}^D(f,\epsilon\alpha),$$

$\lim_{n\to\infty} \mathrm{CC}([f,\mu,\epsilon]^n)/n = \mathrm{IC}_\mu[f,\mu,\epsilon]$.
$\lim_{n\to\infty} \mathrm{CC}([f,\epsilon]^n)/n = \max_\mu \mathrm{IC}_\mu[f,\epsilon] = \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$.

Definition (Prior-free information complexity)

$$\mathrm{IC}(f,\epsilon) := \max_{\mu,\nu} \mathrm{IC}_\mu[f,\nu,\epsilon]$$

and

$$\mathrm{IC}^D(f,\epsilon) := \max_{\mu} \mathrm{IC}_\mu[f,\mu,\epsilon]$$

Theorem (Braverman'12)

$$\mathrm{IC}(f,0) = \mathrm{IC}^D(f,0),$$

and for every $0 < \alpha < 1$*,*

$$\mathrm{IC}^D(f,\epsilon) \leq \mathrm{IC}(f,\epsilon) \leq \frac{1}{1-\alpha}\mathrm{IC}^D(f,\epsilon\alpha),$$

[Braverman'12]: $\mathrm{IC}^D(f,\epsilon) = \mathrm{IC}(f,\epsilon)$?

Set Disjointness

$$S, T \subseteq \{1, \ldots, n\}, \qquad \mathrm{DISJ}_n(S, T) = \begin{cases} 1 & S \cap T = \emptyset \\ 0 & S \cap T \neq \emptyset \end{cases}$$

Set Disjointness

$$S, T \subseteq \{1, \ldots, n\}, \qquad \mathrm{DISJ}_n(S, T) = \begin{cases} 1 & S \cap T = \emptyset \\ 0 & S \cap T \neq \emptyset \end{cases}$$

Theorem (KS92, Raz92)

For $\epsilon < 1/2$,

$$\mathrm{CC}[\mathrm{DISJ}_n, \epsilon] = \Theta(n).$$

Set Disjointness

$$S, T \subseteq \{1, \ldots, n\}, \qquad \mathrm{DISJ}_n(S, T) = \begin{cases} 1 & S \cap T = \emptyset \\ 0 & S \cap T \neq \emptyset \end{cases}$$

Theorem (KS92, Raz92)

For $\epsilon < 1/2$,

$$\mathrm{CC}[\mathrm{DISJ}_n, \epsilon] = \Theta(n).$$

Disjointness as OR of AND's

$$\mathrm{DISJ}_n(X, Y) = \neg \vee_{i=1}^{n} (X_i \wedge Y_i).$$

Set Disjointness

$$S, T \subseteq \{1, \ldots, n\}, \qquad \mathrm{DISJ}_n(S, T) = \begin{cases} 1 & S \cap T = \emptyset \\ 0 & S \cap T \neq \emptyset \end{cases}$$

Theorem (KS92, Raz92)

For $\epsilon < 1/2$,

$$\mathrm{CC}[\mathrm{DISJ}_n, \epsilon] = \Theta(n).$$

Disjointness as OR of AND's

$$\mathrm{DISJ}_n(X, Y) = \neg \vee_{i=1}^{n} (X_i \wedge Y_i).$$

An idea: Computing $\mathrm{DISJ}_n(X, Y)$ is essentially equivalent to computing all $X_i \wedge Y_i$ (amortized).

Set Disjointness

Theorem (BGPW13)

$$\lim_{\epsilon \to 0} \mathrm{CC}[\mathrm{DISJ}_n, \epsilon] \approx \mathrm{IC}(\mathrm{DISJ}_n, 0) \approx 0.4827n + o(n)$$

where

$$0.4827... = \mathrm{IC}^0(\mathrm{AND}, 0) := \max_{\mu} \{\mathrm{IC}_{\mu}(\mathrm{AND}, 0) : \mu(1,1) = 0\}.$$

Set Disjointness

Theorem (BGPW13)

$$\lim_{\epsilon \to 0} \mathrm{CC}[\mathrm{DISJ}_n, \epsilon] \approx \mathrm{IC}(\mathrm{DISJ}_n, 0) \approx 0.4827n + o(n)$$

where

$$0.4827... = \mathrm{IC}^0(\mathrm{AND}, 0) := \max_{\mu} \{\mathrm{IC}_\mu(\mathrm{AND}, 0) : \mu(1,1) = 0\}.$$

[BGPW13]'s questions

- What is the behaviour of $\mathrm{IC}(\mathrm{AND}, 0) - \mathrm{IC}(\mathrm{AND}, \epsilon)$ as $\epsilon \to 0$?
- What is the behaviour of $\mathrm{IC}[\mathrm{DISJ}_n, 0] - \mathrm{IC}[\mathrm{DISJ}_n, \epsilon]$ as $\epsilon \to 0$?

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}(\mathrm{AND}, 0) - \mathrm{IC}(\mathrm{AND}, \epsilon) = \Theta(h(\epsilon)),$$

$$\mathrm{IC}[\mathrm{DISJ}_n, \epsilon] \text{ and } \mathrm{CC}[\mathrm{DISJ}_n, \epsilon] = 0.4827n - \Theta(h(\epsilon)n).$$

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}(\mathrm{AND}, 0) - \mathrm{IC}(\mathrm{AND}, \epsilon) = \Theta(h(\epsilon)),$$

$$\mathrm{IC}[\mathrm{DISJ}_n, \epsilon] \;\; \textit{and} \;\; \mathrm{CC}[\mathrm{DISJ}_n, \epsilon] = 0.4827n - \Theta(h(\epsilon)n).$$

- The lower bound $\mathrm{IC}_\mu(\mathrm{AND}, \epsilon) \geq \mathrm{IC}_\mu(\mathrm{AND}, 0) - \Theta(h(\epsilon))$ is difficult. We do not know what the optimal protocol is for $\mathrm{IC}_\mu(\mathrm{AND}, \epsilon)$.

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}(\mathrm{AND}, 0) - \mathrm{IC}(\mathrm{AND}, \epsilon) = \Theta(h(\epsilon)),$$

$$\mathrm{IC}[\mathrm{DISJ}_n, \epsilon] \text{ and } \mathrm{CC}[\mathrm{DISJ}_n, \epsilon] = 0.4827n - \Theta(h(\epsilon)n).$$

- The lower bound $\mathrm{IC}_\mu(\mathrm{AND}, \epsilon) \geq \mathrm{IC}_\mu(\mathrm{AND}, 0) - \Theta(h(\epsilon))$ is difficult. We do not know what the optimal protocol is for $\mathrm{IC}_\mu(\mathrm{AND}, \epsilon)$.
- The approach of [BGPW13] based on verifying some local convexity conditions does not seem to work.

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}^D[\mathrm{DISJ}_n, \epsilon] := \max_\mu \mathrm{IC}_\mu[\mathrm{DISJ}_n, \mu, \epsilon] =$$

$$n[\mathrm{IC}^0(\mathrm{AND}) - \Theta(\sqrt{h(\epsilon)})] + O(\log n).$$

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}^D[\mathrm{DISJ}_n, \epsilon] := \max_\mu \mathrm{IC}_\mu[\mathrm{DISJ}_n, \mu, \epsilon] =$$

$$n[\mathrm{IC}^0(\mathrm{AND}) - \Theta(\sqrt{h(\epsilon)})] + O(\log n).$$

This with the previous theorem separates distributional prior-free and non-distributional prior-free information complexity and shows that Braverman's analysis is tight

$$\mathrm{IC}^D(\mathrm{DISJ}_n, \epsilon) \neq \mathrm{IC}(\mathrm{DISJ}_n, \epsilon).$$

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}_\mu[\mathrm{DISJ}_n, \mu, \epsilon] \le n[\mathrm{IC}^0(\mathrm{AND}) - \Theta(\sqrt{h(\epsilon)})] + O(\log n).$$

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}_{\mu}[\mathrm{DISJ}_n, \mu, \epsilon] \le n[\mathrm{IC}^0(\mathrm{AND}) - \Theta(\sqrt{h(\epsilon)})] + O(\log n).$$

- Alice and Bob try to compute $(X_i \wedge Y_i)$ for all $i = \sigma_1, \sigma_2, \ldots$ for a random permutation σ of $\{1, 2, \ldots, n\}$.

Theorem (Dagan-Filmus-H-Li)

$$\mathrm{IC}_\mu[\mathrm{DISJ}_n, \mu, \epsilon] \leq n[\mathrm{IC}^0(\mathrm{AND}) - \Theta(\sqrt{h(\epsilon)})] + O(\log n).$$

- Alice and Bob try to compute $(X_i \wedge Y_i)$ for all $i = \sigma_1, \sigma_2, \ldots$ for a random permutation σ of $\{1, 2, \ldots, n\}$.
- To compute $(X_i \wedge Y_i)$, they run an (almost) optimal for $\mathrm{IC}_{\nu_{\sigma_i}}(\mathrm{AND}, \epsilon/2p, 1 \to 0)$, where
    - $1 \to 0$ means one-sided error.
    - $p = \Pr_\mu[\mathrm{DISJ}_n(X, Y) = 1]$.
    - ν_{σ_i} is the corresponding marginal conditioned on the event that the protocol has not yet terminated.

Open Problems

We proved

$$\mathrm{IC}^0(\mathrm{AND}, \epsilon) \leq \frac{\mathrm{IC}(\mathrm{DISJ}_n, \epsilon)}{n} \leq \frac{\mathrm{IC}(\mathrm{DISJ}_n, \epsilon, 1 \to 0)}{n} \leq \mathrm{IC}^0(\mathrm{AND}, \epsilon, 1 \to 0)$$

- Here $1 \to 0$ denote one-sided error (if output is 0 we are always correct).
- IC^0 means we only consider μ with $\mu(1,1) = 0$.

Open Problems

We proved

$$\mathrm{IC}^0(\mathrm{AND},\epsilon) \le \frac{\mathrm{IC}(\mathrm{DISJ}_n,\epsilon)}{n} \le \frac{\mathrm{IC}(\mathrm{DISJ}_n,\epsilon,1\to 0)}{n} \le \mathrm{IC}^0(\mathrm{AND},\epsilon,1\to 0)$$

- Here $1 \to 0$ denote one-sided error (if output is 0 we are always correct).
- IC^0 means we only consider μ with $\mu(1,1)=0$.

Conjecture

$$\frac{\mathrm{IC}(\mathrm{DISJ}_n,\epsilon)}{n} = \mathrm{IC}^0(\mathrm{AND},\epsilon,1\to 0) \pm o(1).$$

- We prove that for every function
$\mathrm{IC}_\mu(f,0) - \Theta(\sqrt{h(\epsilon)}) \leq \mathrm{IC}_\mu(f,\epsilon) \leq \mathrm{IC}_\mu(f,0) - \Theta(h(\epsilon))$.

  Is the upper-bound always the truth?