The Communication Complexity of Correlation

Prahladh Harsha
Rahul Jain
David McAllester
Jaikumar Radhakrishnan

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

Alice | Bob

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

$x \longrightarrow \boxed{\text{Alice}} \qquad\qquad \boxed{\text{Bob}}$

Input (to Alice): $x \leftarrow_R X$

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

$$x \longrightarrow \boxed{\text{Alice}} \text{------} z \longrightarrow \boxed{\text{Bob}}$$

Input (to Alice): $x \leftarrow_R X$

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

$$x \longrightarrow \boxed{\text{Alice}} \text{------} z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

$$x \longrightarrow \boxed{\text{Alice}} \text{------} z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Question: What is the minimum "expected" number of bits (i.e, $|Z|$) Alice needs to send Bob? (say, $T(X : Y)$)

Transmitting Correlated Variables

(X, Y) – pair of correlated random variables

$$x \longrightarrow \boxed{\text{Alice}} \text{——-} z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Question: What is the minimum "expected" number of bits (i.e, $|Z|$) Alice needs to send Bob? (say, $T(X : Y)$)

Easy to check: $T(X : Y) \geq I[X : Y]$ (mutual information)

Correlated variables – an example

- $W = (i, b)$ — random variable uniformly distributed over $[n] \times \{0, 1\}$.
- X and Y – two n bit strings such that
  - $X[i] = Y[i] = b$
  - remaining $2n - 1$ bits independently and uniformly chosen.

Correlated variables – an example

- $W = (i, b)$ — random variable uniformly distributed over $[n] \times \{0, 1\}$.
- X and Y – two n bit strings such that
  - $X[i] = Y[i] = b$
  - remaining $2n - 1$ bits independently and uniformly chosen.

Exercise: $I[X : Y] = o(1)$ but $T(X : Y) = \Theta(\log n)$

Information Theory – Preliminaries

(X, Y) - pair of random variables

1. Entropy:

$$H[X] \doteqdot \sum_x p(x) \log \frac{1}{p(x)},$$

   where $p(x) = \Pr[X = x]$.
   $\approx$ minimum expected number of bits to encode X (upto ± 1)

2. Conditional Entropy:

$$H[Y|X] \doteqdot \sum_x \Pr[X = x] \cdot H[Y|_{X=x}]$$

3. Joint Entropy: $H[XY] = H[X] + H[Y|X]$

4. Mutual Information:

$$\begin{aligned} I[X:Y] &\doteqdot H[X] + H[Y] - H[XY] \\ &= H[Y] - H[Y|X] \end{aligned}$$

Information Theory – Preliminaries (Contd)

1. Independence: X and Y are independent if

$$\Pr[X = x, Y = y] = \Pr[X = x] \cdot \Pr[Y = y], \text{for all } x, y$$

   Equivalently, $I[X : Y] = 0$.
2. Markov Chain: X—-Z—-Y is called a Markov chain if X and Y are conditionally independent given Z (ie., $I[X : Y|Z] = 0$)
3. Data Processing Inequality:

$$X\text{—-}Z\text{—-}Y \implies I[X : Z] \geq I[X : Y]$$

Transmitting Correlated Variables

$$x \longrightarrow \boxed{\text{Alice}} \text{------} z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Question: What is the minimum "expected" number of bits (i.e, $|z|$) Alice needs to send Bob? (say, $T(X:Y)$)

Transmitting Correlated Variables

$$x \longrightarrow \boxed{\text{Alice}} \text{———-} z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Question: What is the minimum "expected" number of bits (i.e, $|z|$) Alice needs to send Bob? (say, $T(X : Y)$)

$$T(X : Y) = \min_Z H[Z]$$

where the minimum is over all Markov chains X—-Z—-Y (i.e., Z such that X and Y are independent given Z)

Asymptotic Version (with error)

$$\left.\begin{matrix} x_1\\ x_2\\ \vdots\\ x_m\end{matrix}\right\} \longrightarrow \boxed{\text{Alice}} \text{---} z \longrightarrow \boxed{\text{Bob}} \longrightarrow \left\{\begin{matrix}\widetilde{y}_1\\ \widetilde{y}_2\\ \vdots\\ \widetilde{y}_m\end{matrix}\right.$$

Input: $x_1, x_2, \ldots, x_m$ —- i.i.d. samples from X

Asymptotic Version (with error)

$$\left.\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix}\right\} \longrightarrow \boxed{\text{Alice}} \;\text{---}z \longrightarrow \boxed{\text{Bob}} \longrightarrow \left\{\begin{matrix} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_m \end{matrix}\right.$$

Input: $x_1, x_2, \ldots, x_m$ —- i.i.d. samples from X

Output: $\widetilde{y_1}, \widetilde{y_2}, \ldots, \widetilde{y_m}$ such that

$$\left\| ((X_1, Y_1), \ldots, (X_m, Y_m)) - \Big((X_1, \widetilde{Y}_1), \ldots, (X_m, \widetilde{Y}_m)\Big) \right\|_1 \leq \lambda$$

Asymptotic Version (with error)

$$\left.\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array}\right\} \longrightarrow \boxed{\text{Alice}} \text{---} z \longrightarrow \boxed{\text{Bob}} \longrightarrow \left\{\begin{array}{c} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_m \end{array}\right.$$

Input: $x_1, x_2, \ldots, x_m$ — i.i.d. samples from X

Output: $\widetilde{y_1}, \widetilde{y_2}, \ldots, \widetilde{y_m}$ such that

$$\left\| ((X_1, Y_1), \ldots, (X_m, Y_m)) - \left((X_1, \widetilde{Y}_1), \ldots, (X_m, \widetilde{Y}_m)\right) \right\|_1 \leq \lambda$$

- $T_\lambda(X^m, Y^m) = \min E[|Z|]$
- Common Information:

$$C(X:Y) \doteqdot \liminf_{\lambda \to 0} \left[\lim_{m \to \infty} \frac{T_\lambda(X^m : Y^m)}{m} \right]$$

Asymptotic Version (with error)

Theorem (Wyner 1975)

$$C(X : Y) = \min_{Z} I[XY : Z],$$

where the minimum is taken over all Z *such that* X—-Z—-Y.

Asymptotic Version (with error)

Theorem (Wyner 1975)

$$C(X:Y) = \min_Z I[XY:Z],$$

where the minimum is taken over all Z such that X—-Z—-Y.

For instance in the example, $C(X:Y) = 2 - o(1)$
i.e., Can send significantly less in the asymptotic case (2 bits on average) compared to the one-shot case ($\Theta(\log n)$ bits).

Asymptotic Version

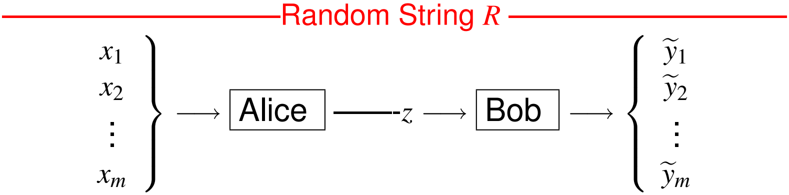
$$\left.\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array}\right\} \longrightarrow \boxed{\text{Alice}} \text{------}z \longrightarrow \boxed{\text{Bob}} \longrightarrow \left\{\begin{array}{c} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_m \end{array}\right.$$

- $\widetilde{y_1}, \widetilde{y_2}, \ldots, \widetilde{y_m}$ such that

$$\left\| ((X_1, Y_1), \ldots, (X_m, Y_m)) - \left((X_1, \widetilde{Y}_1), \ldots, (X_m, \widetilde{Y}_m)\right) \right\|_1 \leq \lambda$$

- $T_\lambda(X^m, Y^m) = \min E[|Z|]$

Asymptotic Version with Shared Randomness



- $\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y_m}$ such that
$$\left\| ((X_1, Y_1), \ldots, (X_m, Y_m)) - \left((X_1, \widetilde{Y}_1), \ldots, (X_m, \widetilde{Y}_m)\right) \right\|_1 \leq \lambda$$
- $T_\lambda^R(X^m, Y^m) = \min E[|Z|]$

Asymptotic Version with Shared Randomness

Random String R

$$\left.\begin{array}{c} x_1\\ x_2\\ \vdots\\ x_m\end{array}\right\} \longrightarrow \boxed{\text{Alice}} \longrightarrow z \longrightarrow \boxed{\text{Bob}} \longrightarrow \left\{\begin{array}{c}\widetilde{y}_1\\ \widetilde{y}_2\\ \vdots\\ \widetilde{y}_m\end{array}\right.$$

- $\widetilde{y_1}, \widetilde{y_2}, \ldots, \widetilde{y_m}$ such that

$$\left\| ((X_1, Y_1), \ldots, (X_m, Y_m)) - \left((X_1, \widetilde{Y}_1), \ldots, (X_m, \widetilde{Y}_m)\right) \right\|_1 \leq \lambda$$

- $T_\lambda^R(X^m, Y^m) = \min E[|Z|]$

Theorem (Winter 2002)

$$\liminf_{\lambda \to 0} \left[\lim_{m \to \infty} \frac{T_\lambda^R(X^m : Y^m)}{m} \right] = I[X : Y]$$

Main Result: One-shot Version

$$\overbrace{\hspace{6cm}}^{R}$$

$$x \longrightarrow \boxed{\text{Alice}} \;\text{——}\; z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Communication: $T^{R}(X : Y) = E[|Z|]$

Main Result: One-shot Version

$$\overset{R}{\overline{\hspace{20em}}}$$

$$x \longrightarrow \boxed{\text{Alice}} \;\text{---}\; z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Communication: $T^R(X:Y) = E[|Z|]$

Theorem (Main Result)

$$I[X:Y] \leq T^R(X:Y) \leq I[X:Y] + 2\log I[X:Y] + O(1)$$

Main Result: One-shot Version

$$\overline{\qquad\qquad\qquad} \; R \; \overline{\qquad\qquad\qquad}$$

$$x \longrightarrow \boxed{\text{Alice}} \text{ ------ } z \longrightarrow \boxed{\text{Bob}} \longrightarrow y$$

Input (to Alice): $x \leftarrow_R X$

Output (from Bob): $y \leftarrow_R Y|_{X=x}$ (ie., conditional distribution)

Communication: $T^R(X : Y) = E[|Z|]$

Theorem (Main Result)

$$I[X : Y] \leq T^R(X : Y) \leq I[X : Y] + 2\log I[X : Y] + O(1)$$

Characterization of Mutual Information (upto lower order logarithmic terms)

One-shot vs. Asymptotic

One-shot vs. Asymptotic

Typical Sets

- For large n, n i.i.d samples of X fall in "typical sets"
- Typical sets – all elements equally probable and number of elements $\approx 2^{nH[X]}$.
- Asymptotic statements – arguably properties of typical sets as opposed to the underlying distributions.

One-shot vs. Asymptotic

Typical Sets

- For large n, n i.i.d samples of X fall in "typical sets"
- Typical sets – all elements equally probable and number of elements $\approx 2^{nH[X]}$.
- Asymptotic statements – arguably properties of typical sets as opposed to the underlying distributions.

Applications

- Asymptotic versions not strong enough for applications.

Outline

Generating one distribution from another

P, Q — two distributions such that $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$.
Rejection Sampling Procedure:

$$\underline{q_1}\ \underline{q_2}\ \underline{q_3}\ \underline{q_4}\ \underline{q_5}\ \cdots\cdots\cdots\ \underline{q_i}\ \cdots\cdots\cdots$$

$$\Downarrow$$

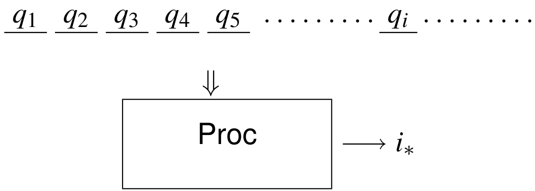
$$\boxed{\text{Proc}} \longrightarrow i_*$$

Input: An infinite stream of independently drawn samples from Q

Output: Index i_* such that q_{i_*} is a sample from P

Generating one distribution from another

P, Q — two distributions such that $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$.
Rejection Sampling Procedure:



Input: An infinite stream of independently drawn samples from Q

Output: Index i_* such that q_{i_*} is a sample from P

Question: What is the minimum expected length of the index (i.e., $E[l(i_*)]$) over all such procedures?

Naive procedure

- Sample according to P to obtain item x
- Wait till item x appears in the stream and output corresponding index

Naive procedure

- Sample according to P to obtain item x
- Wait till item x appears in the stream and output corresponding index

$$E[l(i_*)] \approx \sum_x p(x) \log \frac{1}{q(x)}$$

Relative Entropy

P, Q — two distributions

$$S(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Relative Entropy

P, Q — two distributions

$$S(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Properties:

- Asymmetric

Relative Entropy

P, Q — two distributions

$$S(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Properties:

- Asymmetric
- $S(P\|Q) < \infty \Leftrightarrow \mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$

Relative Entropy

P, Q — two distributions

$$S(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Properties:

- Asymmetric
- $S(P\|Q) < \infty \Leftrightarrow \mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$
- $S(P\|Q) = 0 \Leftrightarrow P \equiv Q$

Relative Entropy

P, Q — two distributions

$$S(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Properties:

- Asymmetric
- $S(P\|Q) < \infty \Leftrightarrow \mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$
- $S(P\|Q) = 0 \Leftrightarrow P \equiv Q$
- $S(P\|Q) \geq 0$

Rejection Sampling Lemma

Lemma (Rejection Sampling Lemma)

There exists a rejection sampling procedure that generates P from Q such that

$$E[l(i_*)] \leq S(P\|Q) + 2\log S(P\|Q) + O(1).$$

Proof of Main Result

Fact: $I[X:Y] = E_{x \leftarrow X}\left[S\left(Y|_{X=x} \| Y\right)\right]$

Proof of Main Result

Fact: $I[X:Y] = E_{x \leftarrow X}\left[S\left(Y|_{X=x} \| Y\right)\right]$

Proof.

- Common random string: sequence of samples from marginal Y
- On input x, Alice performs rejection sampling procedure to generate $Y|_{X=x}$ from Y

$$
\begin{aligned}
E[|Z|] &= E_{x \leftarrow X}\left[S(Y|_{X=x} \| Y) + 2\log S(Y|_{X=x} \| Y) + O(1)\right] \\
&\leq I[X:Y] + 2\log I[X:Y] + O(1)
\end{aligned}
$$

□

Greedy Approach to Rejection Sampling

$$\underline{q_1}\ \underline{q_2}\ \underline{q_3}\ \underline{q_4}\ \underline{q_5}\ \cdots\cdots\cdots\ \underline{q_i}\ \cdots\cdots\cdots$$

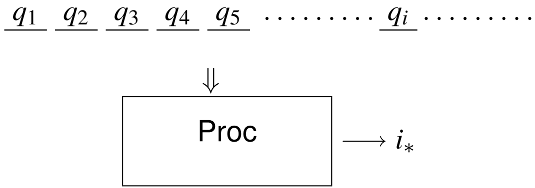
$$\Downarrow$$

$$\boxed{\text{Proc}} \longrightarrow i_*$$

Input: An infinite stream of independently drawn samples from Q

Output: Index i_* such that q_{i_*} is a sample from P

Greedy Approach to Rejection Sampling



Input: An infinite stream of independently drawn samples from Q

Output: Index i_* such that q_{i_*} is a sample from P

Greedy Approach: At each iteration, fill distribution P with best possible sub-distribution of Q, while maintaining that sum is always less then P

Greedy Approach

1. Set $p_1(x) \leftarrow p(x)$
   [$p_i(x) =$ Probability for item x that still needs to be satisfied]
2. Set $s_0 \leftarrow 0$

   [$s_i = \Pr$[Greedy stops before examining $(i+1)$th sample]]
3. For $i \leftarrow 1$ to ∞
   3.1 Examine sample q_i
   3.2 If $q_i = x$,
      - With probability $\min\left\{\frac{p_i(x)}{(1-s_{i-1})\cdot q(x)}, 1\right\}$ output i

   3.3 Updates:
      - For all x, set $p_{i+1}(x) \leftarrow$ $p_i(x) - (1-s_{i-1})q(x) \cdot \min\left\{\frac{p_i(x)}{(1-s_{i-1})q(x)}, 1\right\} = p_i(x) - \alpha_{i,x}$
      - Set $s_i \leftarrow s_{i-1} + \sum_x \alpha_{i,x}$

Greedy Approach

1. Set $p_1(x) \leftarrow p(x)$
   [$p_i(x)$ = Probability for item x that still needs to be satisfied]
2. Set $s_0 \leftarrow 0$

   [$s_i = \Pr$[Greedy stops before examining $(i+1)$th sample]]
3. For $i \leftarrow 1$ to ∞
   3.1 Examine sample q_i
   3.2 If $q_i = x$,
      - With probability $\min\left\{\frac{p_i(x)}{(1-s_{i-1})\cdot q(x)}, 1\right\}$ output i

   3.3 Updates:
      - For all x, set $p_{i+1}(x) \leftarrow$ $p_i(x) - (1-s_{i-1})q(x) \cdot \min\left\{\frac{p_i(x)}{(1-s_{i-1})q(x)}, 1\right\} = p_i(x) - \alpha_{i,x}$
      - Set $s_i \leftarrow s_{i-1} + \sum_x \alpha_{i,x}$

Clearly, distribution generated is P.

Expected Index Length

Expected Index Length

$$
\begin{aligned}
s_i &= \Pr[\text{ Greedy stops before examining sample } q_{i+1}] \\
\alpha_{i,x} &= \Pr[\text{ Greedy outputs } x \text{ in iteration } i].
\end{aligned}
$$

Expected Index Length

$$\begin{aligned} s_i &= \Pr[\text{ Greedy stops before examining sample } q_{i+1}] \\ \alpha_{i,x} &= \Pr[\text{ Greedy outputs } x \text{ in iteration } i]. \end{aligned}$$

$$\text{Note } p(x) = \sum_i \alpha_{i,x}$$

Expected Index Length

$$\begin{aligned} s_i &= \Pr[\text{ Greedy stops before examining sample } q_{i+1}] \\ \alpha_{i,x} &= \Pr[\text{ Greedy outputs } x \text{ in iteration } i]. \end{aligned}$$

$$\text{Note } p(x) = \sum_i \alpha_{i,x}$$

Suppose $\alpha_{i+1,x} > 0$,
i.e., previous iterations not sufficient to provide the required probability $p(x)$ to item x. Hence,

Expected Index Length

$$
\begin{aligned}
s_i &= \Pr[\text{ Greedy stops before examining sample } q_{i+1}] \\
\alpha_{i,x} &= \Pr[\text{ Greedy outputs } x \text{ in iteration } i].
\end{aligned}
$$

$$\text{Note } p(x) = \sum_i \alpha_{i,x}$$

Suppose $\alpha_{i+1,x} > 0$,
i.e., previous iterations not sufficient to provide the required probability $p(x)$ to item x. Hence,

$$
\begin{aligned}
\sum_{j=1}^{i} (1 - s_{j-1}) \cdot q(x) &< p(x) \\
i(1 - s_i) q(x) &< p(x)
\end{aligned}
$$

.

Expected Index Length

$$\begin{aligned} s_i &= \Pr[\text{ Greedy stops before examining sample } q_{i+1}] \\ \alpha_{i,x} &= \Pr[\text{ Greedy outputs } x \text{ in iteration } i]. \end{aligned}$$

$$\text{Note } p(x) = \sum_i \alpha_{i,x}$$

Suppose $\alpha_{i+1,x} > 0$,
i.e., previous iterations not sufficient to provide the required probability $p(x)$ to item x. Hence,

$$\begin{aligned} \sum_{j=1}^{i} (1 - s_{j-1}) \cdot q(x) &< p(x) \\ i(1 - s_i)q(x) &< p(x) \\ i &< \frac{1}{1 - s_i} \cdot \frac{p(x)}{q(x)}. \end{aligned}$$

Expected Index Length

$$
\begin{aligned}
E[l(i)] &\approx E[\log i] = \sum_x \sum_i \alpha_{i,x} \cdot \log i \\
&\leq \sum_x \sum_i \alpha_{i,x} \cdot \log \left(\frac{1}{1 - s_{i-1}} \cdot \frac{p(x)}{q(x)} \right) \\
&= \sum_x \sum_i \alpha_{i,x} \left(\log \frac{1}{1 - s_{i-1}} + \log \frac{p(x)}{q(x)} \right) \\
&= \sum_x p(x) \log \frac{p(x)}{q(x)} + \sum_i \left(\sum_x \alpha_{i,x} \right) \log \frac{1}{1 - s_{i-1}} \\
&= S(P \| Q) + \sum_i \alpha_i \log \frac{1}{1 - s_{i-1}} \leq S(P \| Q) + \int_0^1 \log \frac{1}{1 - s} \mathrm{d}s \\
&= S(P \| Q) + O(1)
\end{aligned}
$$

Expected Index Length

$$
\begin{aligned}
E[l(i)] &\approx E[\log i] = \sum_x \sum_i \alpha_{i,x} \cdot \log i \\
&\leq \sum_x \sum_i \alpha_{i,x} \cdot \log \left(\frac{1}{1 - s_{i-1}} \cdot \frac{p(x)}{q(x)} \right) \\
&= \sum_x \sum_i \alpha_{i,x} \left(\log \frac{1}{1 - s_{i-1}} + \log \frac{p(x)}{q(x)} \right) \\
&= \sum_x p(x) \log \frac{p(x)}{q(x)} + \sum_i \left(\sum_x \alpha_{i,x} \right) \log \frac{1}{1 - s_{i-1}} \\
&= S(P \| Q) + \sum_i \alpha_i \log \frac{1}{1 - s_{i-1}} \leq S(P \| Q) + \int_0^1 \log \frac{1}{1 - s} \mathrm{d}s \\
&= S(P \| Q) + O(1)
\end{aligned}
$$

Actually, $l(i) = \log i + 2 \log \log i$, hence extra log term in final result

Greedy Approach (Contd)

- Clearly, the greedy approach generates the target distribution P
- Furthermore, it can be shown that the expected index length is at most

$$E[l(i)] \leq S(P\|Q) + 2\log S(P\|Q) + O(1).$$

Outline

Two Party Communication Complexity Model [Yao]

$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$

Alice

Bob

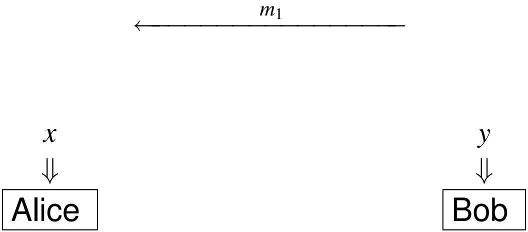
Two Party Communication Complexity Model [Yao]

$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$

x	y
$\Downarrow$	$\Downarrow$
$\boxed{\text{Alice}}$	$\boxed{\text{Bob}}$

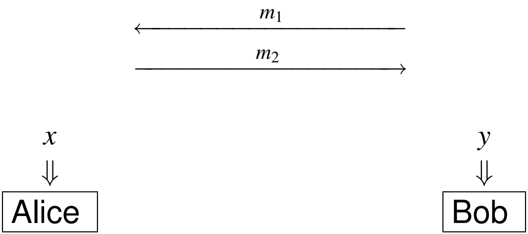
Two Party Communication Complexity Model [Yao]

$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$



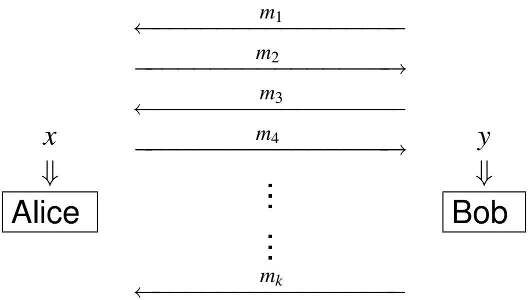
Two Party Communication Complexity Model [Yao]

$$f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$$



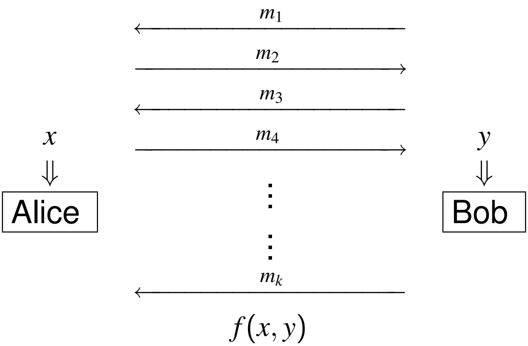
Two Party Communication Complexity Model [Yao]

$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$



Two Party Communication Complexity Model [Yao]

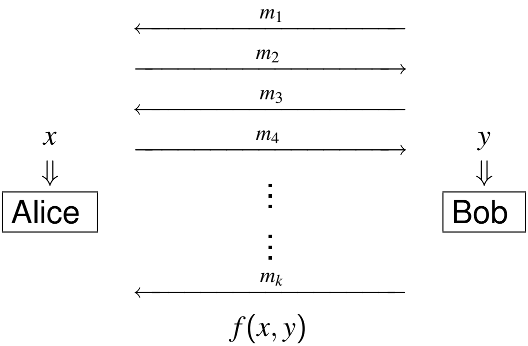
$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$



k-round protocol computing f

Two Party Communication Complexity Model [Yao]

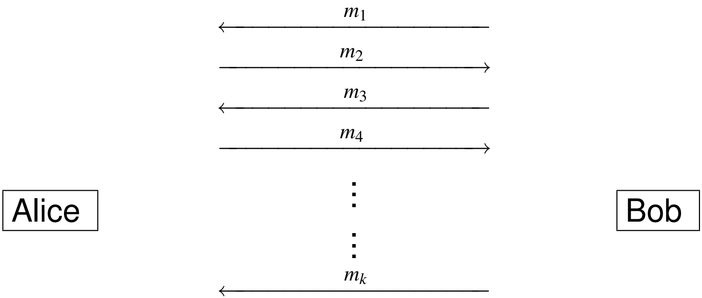
$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$



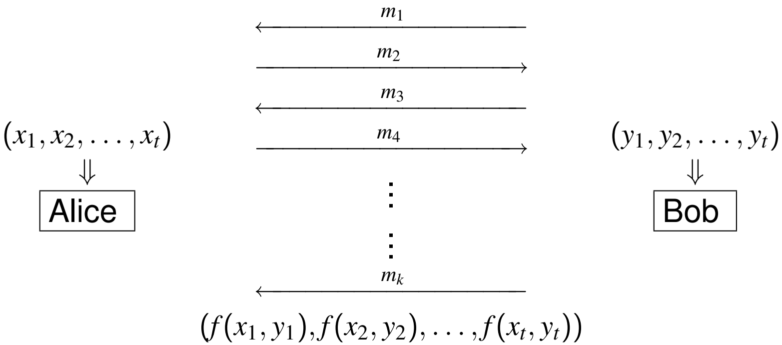
k-round protocol computing f

Question: How many bits must Alice and Bob exchange to compute f?

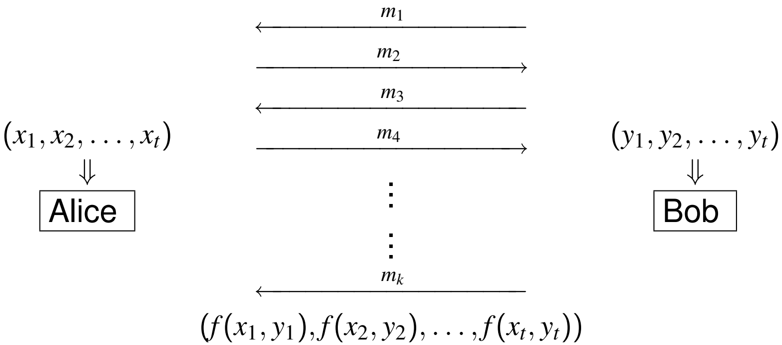
Direct Sum Question



Direct Sum Question

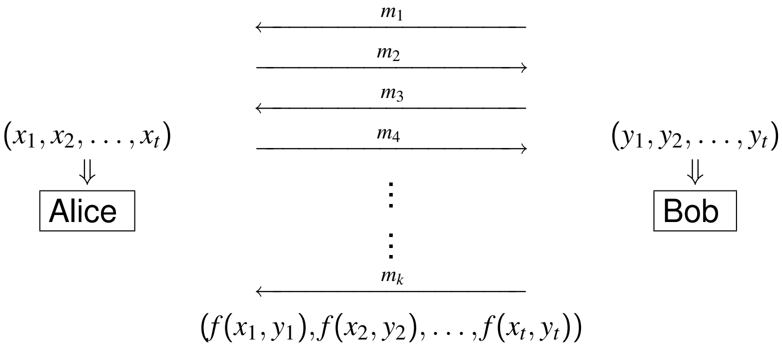


Direct Sum Question



Direct Sum Question: Does the number of bits communicated increase t fold?

Direct Sum Question



Direct Sum Question: Does the number of bits communicated increase t fold?

Direct Product Question: Keeping number of bits communicated fixed, does success probability fall exponentially in t?

Communication Complexity Measures

- randomized communication complexity:

$$R_\epsilon^k(f) = \min_{\Pi} \max_{(x,y)} \text{(number of bits communicated)}$$

Π —- k-round public-coins randomized protocol, that computes f correctly with probability at least $1-\epsilon$ on each input (x,y).

Communication Complexity Measures

- randomized communication complexity:

$$R_\epsilon^k(f) = \min_{\Pi} \max_{(x,y)} \text{(number of bits communicated)}$$

  Π —- k-round public-coins randomized protocol, that computes f correctly with probability at least $1-\epsilon$ on each input (x,y).

- distributional communication complexity: For a distribution μ on the inputs $\mathcal{X} \times \mathcal{Y}$,

$$D_\epsilon^{\mu,k}(f) = \min_{\Pi} \max_{(x,y)} \text{(number of bits communicated)}$$

  where Π —- "deterministic" k-round protocol for f with average error at most ϵ under μ.

Communication Complexity Measures

- randomized communication complexity:

$$R^k_\epsilon(f) = \min_{\Pi} \max_{(x,y)} \text{(number of bits communicated)}$$

  Π —- k-round public-coins randomized protocol, that computes f correctly with probability at least $1 - \epsilon$ on each input (x, y).
- distributional communication complexity: For a distribution μ on the inputs $\mathcal{X} \times \mathcal{Y}$,

$$D^{\mu,k}_\epsilon(f) = \min_{\Pi} \max_{(x,y)} \text{(number of bits communicated)}$$

  where Π —- "deterministic" k-round protocol for f with average error at most ϵ under μ.

Theorem (Yao's minmax principle)

$$R^k_\epsilon(f) = \max_{\mu} D^{\mu,k}_\epsilon(f)$$

Direct Sum Results

$R_\epsilon^k(f)$ vs. $R_\epsilon^k(f^{\oplus t})$ and $D_\epsilon^{\mu,k}(f)$ vs. $D_\epsilon^{\mu^t,k}(f^{\oplus t})$

Direct Sum Results

$R^k_\epsilon(f)$ vs. $R^k_\epsilon(f^{\oplus t})$ and $D^{\mu,k}_\epsilon(f)$ vs. $D^{\mu^t,k}_\epsilon(f^{\oplus t})$

1. [Chakrabarti, Shi, Wirth and Yao 2001]
   Direct Sum result for "Equality function" in *Simultaneous message passing* model.

Direct Sum Results

$R^k_\epsilon(f)$ vs. $R^k_\epsilon(f^{\oplus t})$ and $D^{\mu,k}_\epsilon(f)$ vs. $D^{\mu^t,k}_\epsilon(f^{\oplus t})$

1. [Chakrabarti, Shi, Wirth and Yao 2001]
   Direct Sum result for "Equality function" in *Simultaneous message passing* model.
2. [Jain, Radhakrishnan and Sen 2005]
   Extended above result to all functions in simultaneous message passing model and one-way communication model.

Direct Sum Results

$R^k_\epsilon(f)$ vs. $R^k_\epsilon(f^{\oplus t})$ and $D^{\mu,k}_\epsilon(f)$ vs. $D^{\mu^t,k}_\epsilon(f^{\oplus t})$

1. [Chakrabarti, Shi, Wirth and Yao 2001]
   Direct Sum result for "Equality function" in *Simultaneous message passing* model.
2. [Jain, Radhakrishnan and Sen 2005]
   Extended above result to all functions in simultaneous message passing model and one-way communication model.
3. [Jain, Radhakrishnan and Sen 2003]
   For bounded round communication models, for any f and any product distribution μ,

$$D^{\mu^t,k}_\epsilon(f^{\oplus t}) \geq t\left(\frac{\delta^2}{2k} \cdot D^{\mu,k}_{\epsilon+2\delta}(f) - 2\right)$$

Improved Direct Sum Result

Theorem

For any function f and any product distribution μ,

$$D_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq \frac{t}{2}\left(\delta D_{\epsilon+\delta}^{\mu,k}(f) - O(k)\right).$$

Improved Direct Sum Result

Theorem

For any function f and any product distribution μ,

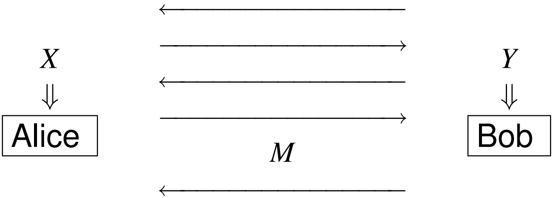
$$D_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq \frac{t}{2}\left(\delta D_{\epsilon+\delta}^{\mu,k}(f) - O(k)\right).$$

Applying Yao's minmax principle,

$$R_{\epsilon}^{k}(f^{\oplus t}) \geq \max_{\text{product } \mu}\left(\frac{t}{2}\left(\delta D_{\epsilon+\delta}^{\mu,k}(f) - O(k)\right)\right).$$

Information Cost

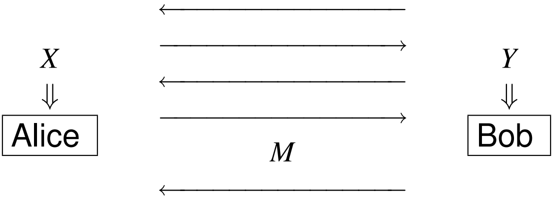
Private Coins Protocol Π



M— transcript

Information Cost

Private Coins Protocol Π



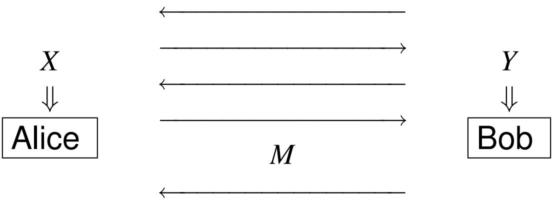
M— transcript

Information Cost of Π wrt μ:

$$\mathrm{IC}^{\mu}(\Pi) = I[XY : M]$$

Information Cost

Private Coins Protocol Π



M— transcript

Information Cost of Π wrt μ:

$$\mathrm{IC}^{\mu}(\Pi) = I[XY : M]$$

For a function f, let

$$\mathrm{IC}_{\epsilon}^{\mu,k}(f) = \min_{\Pi} \mathrm{IC}^{\mu}(\Pi),$$

where Π — k-round private-coins protocols for f with error at most ϵ under μ.

Three Lemmata

Lemma (Direct Sum for Information Cost)

For μ – product distribution,

$$\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f).$$

Three Lemmata

Lemma (Direct Sum for Information Cost)

For μ – product distribution,

$$\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f).$$

Lemma (IC upper bounds distributional complexity)

$$\mathsf{IC}_{\epsilon}^{\mu,k}(f) \leq D_{\epsilon}^{\mu,k}(f).$$

Three Lemmata

Lemma (Direct Sum for Information Cost)

For μ – product distribution,

$$\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f).$$

Lemma (IC upper bounds distributional complexity)

$$\mathsf{IC}_{\epsilon}^{\mu,k}(f) \leq D_{\epsilon}^{\mu,k}(f).$$

Lemma ((Improved) Message Compression)

$$D_{\epsilon+\delta}^{\mu,k}(f) \leq \frac{1}{\delta}\left[2 \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f) + O(k)\right].$$

Three Lemmata

Lemma (Direct Sum for Information Cost)

For μ – product distribution,

$$\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f).$$

Lemma (IC upper bounds distributional complexity)

$$\mathsf{IC}_{\epsilon}^{\mu,k}(f) \leq D_{\epsilon}^{\mu,k}(f).$$

Lemma ((Improved) Message Compression)

$$D_{\epsilon+\delta}^{\mu,k}(f) \leq \frac{1}{\delta}\left[2 \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f) + O(k)\right].$$

Three Lemmata imply improved direct sum result.

Direct Sum for Information Cost

For μ – product distribution, $\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f)$.

Direct Sum for Information Cost

For μ – product distribution, $\mathrm{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathrm{IC}_{\epsilon}^{\mu,k}(f)$.

Private Coins Protocol Π achieves $I = IC_{\epsilon}^{\mu,k}(f^{\oplus t})$

$$\begin{array}{ccc} & \longleftarrow & \\ X_1, X_2, \ldots, X_t & \longrightarrow & Y_1, Y_2, \ldots, Y_t \\ \Downarrow & M & \Downarrow \\ \boxed{\text{Alice}} & \longleftarrow & \boxed{\text{Bob}} \end{array}$$

M— transcript

Direct Sum for Information Cost

For μ – product distribution, $\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f)$.

Private Coins Protocol Π achieves $I = IC_{\epsilon}^{\mu,k}(f^{\oplus t})$

$$\longleftarrow$$

$X_1, X_2, \ldots, X_t \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_t$

$\Downarrow \qquad M \qquad \Downarrow$

Alice $\quad \longleftarrow \quad$ Bob

M— transcript

Chain rule: $I[XY : M] \geq \sum_{i=1}^{t} I[X_iY_i : M]$

Direct Sum for Information Cost

For μ – product distribution, $\mathsf{IC}_{\epsilon}^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathsf{IC}_{\epsilon}^{\mu,k}(f)$.

Private Coins Protocol Π achieves $I = IC_{\epsilon}^{\mu,k}(f^{\oplus t})$

$$\begin{array}{ccc}
 & \longleftarrow & \\
X_1, X_2, \ldots, X_t & \longrightarrow & Y_1, Y_2, \ldots, Y_t \\
\Downarrow & M & \Downarrow \\
\boxed{\text{Alice}} & \longleftarrow & \boxed{\text{Bob}}
\end{array}$$

M— transcript

Chain rule: $I[XY : M] \geq \sum_{i=1}^{t} I[X_iY_i : M]$

Claim: For each i, $I[X_iY_i : M] \geq \mathsf{IC}_{\epsilon}^{\mu,k}(f)$

Direct Sum for Information Cost

For μ – product distribution, $\mathrm{IC}_\epsilon^{\mu^t,k}(f^{\oplus t}) \geq t \cdot \mathrm{IC}_\epsilon^{\mu,k}(f)$.

Private Coins Protocol Π achieves $I = IC_\epsilon^{\mu,k}(f^{\oplus t})$

$$
\begin{array}{ccc}
 & \longleftarrow & \\
X_1, X_2, \ldots, X_t & \longrightarrow & Y_1, Y_2, \ldots, Y_t \\
\Downarrow & M & \Downarrow \\
\boxed{\text{Alice}} & \longleftarrow & \boxed{\text{Bob}}
\end{array}
$$

M— transcript

Chain rule: $I[XY : M] \geq \sum_{i=1}^t I[X_i Y_i : M]$

Claim: For each i, $I[X_i Y_i : M] \geq \mathrm{IC}_\epsilon^{\mu,k}(f)$

Proof: On input X_i and Y_i, Alice and Bob fill in other components (based on product distribution μ) and perform above protocol

$\square$

IC upper bounds distributional complexity

$$\mathsf{IC}_{\epsilon}^{\mu,k}(f) \le D_{\epsilon}^{\mu,k}(f).$$

IC upper bounds distributional complexity

$$\mathsf{IC}_{\epsilon}^{\mu,k}(f) \leq D_{\epsilon}^{\mu,k}(f).$$

Proof.

Let Π be a protocol that achieves $D_{\epsilon}^{\mu,k}(f)$ and M be its transcript. Then,

$$\begin{aligned} D_{\epsilon}^{\mu,k}(f) &\geq E[|M|] \\ &\geq H[M] \\ &\geq I[XY : M] \\ &\geq \mathsf{IC}_{\epsilon}^{\mu,k}(f) \end{aligned}$$

□

Message Compression

To prove: $D^{\mu,k}_{\epsilon+\delta}(f) \leq \frac{1}{\delta}\left[2 \cdot \mathsf{IC}^{\mu,k}_{\epsilon}(f) + O(k)\right]$

Private Coins Protocol Π

$$\xrightarrow{\qquad\qquad m_1 \qquad\qquad}$$

$$\xleftarrow{\qquad\qquad m_2 \qquad\qquad}$$

$$\xrightarrow{\qquad\qquad m_3 \qquad\qquad}$$

$$\xleftarrow{\qquad\qquad m_4 \qquad\qquad}$$

$$\vdots$$

$$\xrightarrow{\qquad\qquad m_k \qquad\qquad}$$

Information Cost I

Message Compression

To prove: $D^{\mu,k}_{\epsilon+\delta}(f) \leq \frac{1}{\delta}\left[2 \cdot \mathrm{IC}^{\mu,k}_{\epsilon}(f) + O(k)\right]$

$$
\begin{array}{c}
\text{Private Coins Protocol } \Pi \\
\xrightarrow{\qquad m_1 \qquad} \\
\xleftarrow{\qquad m_2 \qquad} \\
\xrightarrow{\qquad m_3 \qquad} \\
\xleftarrow{\qquad m_4 \qquad} \\
\vdots \\
\xrightarrow{\qquad m_k \qquad} \\
\text{Information Cost } I
\end{array}
\quad \longrightarrow \quad
\begin{array}{ccc}
\multicolumn{3}{c}{\text{Public Coins Protocol } \Pi'} \\
 & \xrightarrow{\quad z_1 \quad} & m_1 \\
m_2 & \xleftarrow{\quad z_2 \quad} & \\
 & \xrightarrow{\quad z_3 \quad} & m_3 \\
m_4 & \xleftarrow{\quad z_4 \quad} & \\
 & \vdots & \\
 & \xrightarrow{\quad z_k \quad} & m_k \\
\multicolumn{3}{c}{\sum_{i=1}^{k} E\left[Z_i\right] \leq 2I + O\left(k\right)}
\end{array}
$$

Message Compression

To prove: $D^{\mu,k}_{\epsilon+\delta}(f) \leq \frac{1}{\delta}\left[2 \cdot \mathrm{IC}^{\mu,k}_{\epsilon}(f) + O(k)\right]$

Private Coins Protocol Π		Public Coins Protocol Π'
$\xrightarrow{\quad m_1 \quad}$		$\xrightarrow{\ z_1\ }\ m_1$
$\xleftarrow{\quad m_2 \quad}$		$m_2\ \xleftarrow{\ z_2\ }$
$\xrightarrow{\quad m_3 \quad}$	$\longrightarrow$	$\xrightarrow{\ z_3\ }\ m_3$
$\xleftarrow{\quad m_4 \quad}$		$m_4\ \xleftarrow{\ z_4\ }$
$\vdots$		$\vdots$
$\xrightarrow{\quad m_k \quad}$		$\xrightarrow{\ z_k\ }\ m_k$
Information Cost I		$\sum_{i=1}^{k} E\left[Z_i\right] \leq 2I + O\left(k\right)$

$$I = I[XY : M] = I[XY : M_1 M_2 \ldots M_k] = \sum_{i=1}^{k} I[XY : M_i | M_1 M_2 \ldots M_{i-1}]$$

Message Compression

To prove: $D^{\mu,k}_{\epsilon+\delta}(f) \leq \frac{1}{\delta}\left[2 \cdot \mathrm{IC}^{\mu,k}_{\epsilon}(f) + O(k)\right]$

Private Coins Protocol Π $\longrightarrow$ Public Coins Protocol Π'

$$\begin{array}{c}
\xrightarrow{\qquad\quad m_1 \qquad\quad} \\
\xleftarrow{\qquad\quad m_2 \qquad\quad} \\
\xrightarrow{\qquad\quad m_3 \qquad\quad} \\
\xleftarrow{\qquad\quad m_4 \qquad\quad} \\
\vdots \\
\xrightarrow{\qquad\quad m_k \qquad\quad}
\end{array}
\qquad \longrightarrow \qquad
\begin{array}{rcl}
 & \xrightarrow{\quad z_1 \quad} & m_1 \\
m_2 & \xleftarrow{\quad z_2 \quad} & \\
 & \xrightarrow{\quad z_3 \quad} & m_3 \\
m_4 & \xleftarrow{\quad z_4 \quad} & \\
 & \vdots & \\
 & \xrightarrow{\quad z_k \quad} & m_k
\end{array}$$

Information Cost I $\qquad$ $\sum_{i=1}^{k} E\left[Z_i\right] \leq 2I + O\left(k\right)$

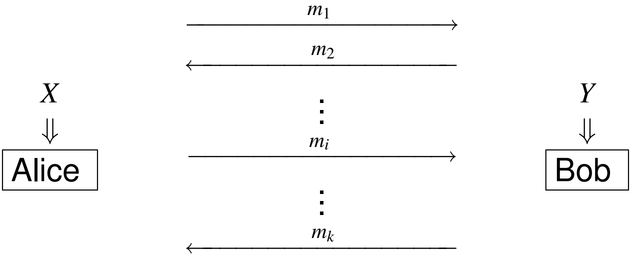
$$I = I[XY : M] = I[XY : M_1 M_2 \ldots M_k] = \sum_{i=1}^{k} I[XY : M_i | M_1 M_2 \ldots M_{i-1}]$$

Sufficient to prove,

For all $i, E[Z_i] \leq 2I[XY : M_i | M_1 M_2 \ldots M_{i-1}] + O(1)$

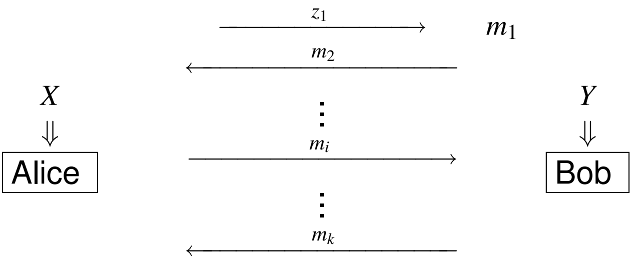
Message Compression (Contd)

To prove: For all $i, E[Z_i] \leq 2I[XY : M_i | M_1 M_2 \dots M_{i-1}] + O(1)$



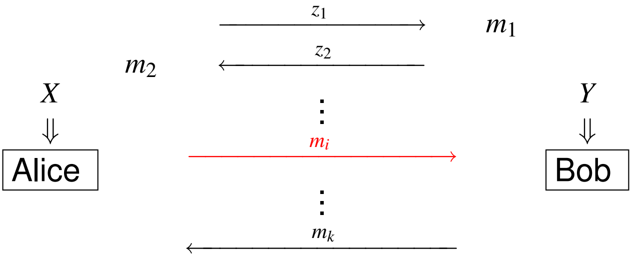
Message Compression (Contd)

To prove: For all $i, E[Z_i] \leq 2I[XY : M_i | M_1 M_2 \ldots M_{i-1}] + O(1)$



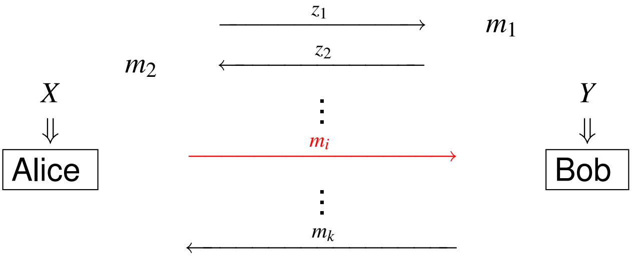
Message Compression (Contd)

To prove: For all $i, E[Z_i] \leq 2I[XY : M_i | M_1 M_2 \ldots M_{i-1}] + O(1)$



Message Compression (Contd)

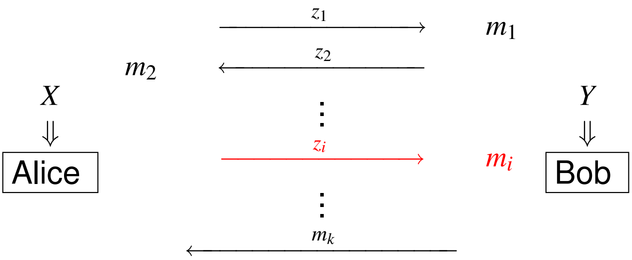
To prove: For all $i, E[Z_i] \leq 2I[XY : M_i | M_1 M_2 \ldots M_{i-1}] + O(1)$



Conditioned on all earlier messages, message M_i is independent of Bob's Input Y.
Hence, $I_i = I[XY : M_i | m_1 \ldots m_{i-1}] = I[X : M_i | m_1 \ldots m_{i-1}]$

Message Compression (Contd)

To prove: For all $i, E[Z_i] \leq 2I[XY : M_i|M_1M_2 \ldots M_{i-1}] + O(1)$



Conditioned on all earlier messages, message M_i is independent of Bob's Input Y.

Hence, $I_i = I[XY : M_i|m_1 \ldots m_{i-1}] = I[X : M_i|m_1 \ldots m_{i-1}]$

Main Result implies M_i can be generated on Bob's side sending only $I_i + 2\log I_i + O(1) \leq 2I_i + O(1)$ bits

Summarizing Results

- A characterization of mutual information and relative entropy in terms of communication complexity (modulo lower order log terms)
- An improved direct sum result for communication complexity

Thank You