CSCE 658: Randomized Algorithms

Lecture 13

Samson Zhou

Class Logistics

- March 5: Lecture canceled, i.e., do NOT show up to HRBB 126 (unless you want to see an empty classroom)

Information Theory

- Suppose X is a random variable taking on values $[n] \coloneqq \{1,2,\ldots,n\}$ and let $p_i \coloneqq \Pr[X = i]$ for all $i \in [n]$
- Concepts generalize to continuous domains

Entropy

- Suppose X is a random variable taking on values $[n] := \{1,2,\ldots,n\}$ and let $p_i := \Pr[X = i]$ for all $i \in [n]$

- The entropy $H(X) = \sum_i p_i \log_2 \frac{1}{p_i}$ of X measures its uncertainty
- We have $H(X) \leq \log_2 n$ with equality at $p_i = \frac{1}{n}$ for all $i \in [n]$

Entropy

- Suppose X is the outcome of a fair coin flip. What is $H(X)$?

- Suppose X is the outcome of a flip of a coin that is HEADS with probability $\frac{1}{2}$. What is $H(X)$?

- Suppose X is the outcome of a flip of a coin that is HEADS with probability $\frac{1}{4}$. What is $H(X)$?

Entropy

- Suppose X is the outcome of a fair coin flip. What is $H(X)$?

- Suppose X is the outcome of a flip of a coin that is HEADS with probability $\frac{1}{2}$. What is $H(X)$? $\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$

- Suppose X is the outcome of a flip of a coin that is HEADS with probability $\frac{1}{4}$. What is $H(X)$? $\frac{1}{4}\log_2 4 + \frac{3}{4}\log_2 \frac{4}{3} \approx 0.811$

Entropy

- Suppose X is the outcome of a fair coin flip. What is $H(X)$?
- Suppose X is the outcome of a flip of a coin that is HEADS with probability p. What is $H(X)$?
- Suppose X is the outcome of a flip of a coin that is HEADS with probability $1 - p$. What is $H(X)$?

Entropy

- Suppose X is the outcome of a fair coin flip. What is $H(X)$? 1
- Suppose X is the outcome of a flip of a coin that is HEADS with probability p. What is $H(X)$? $p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$
- Suppose X is the outcome of a flip of a coin that is HEADS with probability $1-p$. What is $H(X)$? $p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$

Conditional and Joint Entropy

- Let X and Y be random variables

- Conditional entropy $H(X|Y) = \sum_y H(X|Y = y) \cdot \Pr[Y = y]$
- Conditioning can only decrease entropy: $H(X|Y) \leq H(X)$

- Proof is by concavity of the log function and Jensen's inequality

Joint Entropy

- Joint entropy:

$$H(X,Y) = \sum_{x,y} \Pr[(X,Y) = (x,y)] \cdot \log_2 \frac{1}{\Pr[(X,Y)=(x,y)]}$$

Chain Rule for Entropy

- $H(X,Y) = H(X) + H(Y|X)$

$$H(X,Y) = \sum_{x,y} \Pr[(X,Y) = (x,y)] \cdot \log_2 \frac{1}{\Pr[(X,Y) = (x,y)]}$$

$$= \sum_{x,y} \Pr[X = x] \cdot \Pr[Y = y|X = x] \cdot \log_2 \frac{1}{\Pr[(X,Y) = (x,y)]}$$

$$= \sum_{x,y} \Pr[X = x] \Pr[Y = y|X = x] \left(\log_2 \frac{1}{\Pr[X = x]} \cdot \frac{1}{\Pr[Y = y|X = x]} \right)$$

Mutual Information

- Mutual information between X and Y is $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X)$

- "Amount of information" obtained about one random variable from observing the other random variable

- We have $I(X;X) = H(X) - H(X|X) = H(X)$

Trivia Question #9 (Conditional Mutual Information)

- For the conditional mutual information between X and Y given Z, $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$, which of the following is always true?

- $I(X;Y|Z) \geq I(X;Y)$
- $I(X;Y|Z) = I(X;Y)$
- $I(X;Y|Z) \leq I(X;Y)$
- None of the above

Conditional Mutual Information

- Suppose $X = Y = Z$

- $I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|Z) - H(X|Z) = 0$
- Y does not reveal anything about X that Z has not already revealed

- $I(X;Y) = H(X) - H(X|Y) = H(X) - 0 = H(X)$
- In this case, $I(X;Y|Z) \leq I(X;Y)$

Conditional Mutual Information

- Suppose $X, Y \in \{0,1\}$ uniformly at random and $X \equiv Y + Z \ (mod\ 2)$

- $I(X; Y|Z) = H(X|Z) - H(X|Y, Z) = H(X) - 0 = H(X)$
- X is completely determined by Y once Z is fixed

- $I(X; Y) = H(X) - H(X|Y) = H(X) - H(X) = 0$
- In this case, $I(X; Y|Z) \geq I(X; Y)$

Chain Rule for Mutual Information

- $I(X,Y;Z) = I(X;Z) + I(Y;Z|X)$
- By induction, $I(X_1, \dots X_n; Z) = \sum_i I(X_i; Z|X_1, \dots, X_{i-1})$

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

(Chain Rule for Entropy)

$$= H(X) + H(Y|X) - H(X|Z) - H(Y|X,Z)$$

$$= I(X;Z) + I(Y;Z|X)$$

Markov Chain

- A Markov chain $X \rightarrow Y \rightarrow Z$ is a sequence of random variables where the outcome of each random variable only depends on the value of the previous random variable

- In other words, the distribution of Z depends solely on the realization of Y, regardless of the value of X

Data Processing Inequality

- Suppose $X \to Y \to Z$ is a Markov chain. Then

$$I(X;Z) \leq I(X;Y)$$

- In other words, any post-processing function applied to Y to obtain Z can only lose information about X

- Consequently, we also have

$$H(X|Y) \leq H(X|Z)$$

Data Processing Inequality

- Suppose $X \to Y \to Z$ is a Markov chain. Then

$$I(X;Z) \leq I(X;Y)$$

- By the chain rule for mutual information,

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z) = I(X;Y) + I(X;Z|Y)$$

- By definition, we have $I(X;Z|Y) = H(X|Y) - H(X|Y,Z)$
- Since Z is independent of X conditioned on Y, then $H(X|Y,Z) = H(X|Y)$ so that $I(X;Z|Y) = 0$
- Then we have $I(X;Z) + I(X;Y|Z) = I(X;Y)$

Fano's Inequality

- Suppose $X \to Y \to Z$ is a Markov chain and $P_e = \Pr[X \neq Z]$. Suppose X is a random variable taking on values $[n]$. Then
$$H(X|Y) \leq H(P_e) + P_e \cdot \log_2(n-1)$$

- Average information lost in a noisy channel

Fano's Inequality

- Suppose $X \to Y \to Z$ is a Markov chain and $P_e = \Pr[X \neq Z]$. Suppose X is a random variable taking on values $[n]$. Then
$$H(X|Y) \leq H(P_e) + P_e \cdot \log_2(n-1)$$
- By data processing inequality, $H(X|Y) \leq H(X|Z)$
- Let $E = 1$ if there is an error, i.e., $X \neq Z$ and $E = 0$ otherwise

- $H(X|Z) = H(X|Z) + H(E|X,Z) = H(E,X|Z)$, by chain rule of entropy and because E is fixed conditioned on X, Z

Fano's Inequality

- Putting these together, Fano's inequality will hold if

$$H(E, X|Z) \leq H(P_e) + P_e \cdot \log_2(n-1)$$

- By chain rule of entropy, $H(E, X|Z) = H(E|Z) + H(X|E, Z)$
- By definition of P_e, we have $H(E|Z) \leq H(P_e)$
- By conditional entropy,

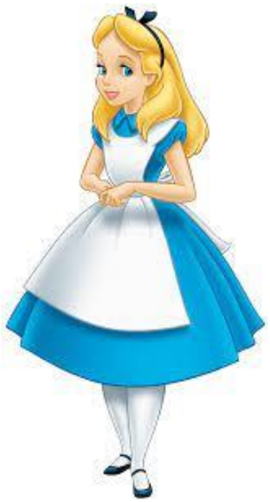
$$\begin{aligned} H(X|E, Z) &= \Pr[E=0]\, H(X|X', E=0) + \Pr[E=1]\, H(X|X', E=1) \\ &= (1-P_e) \cdot 0 + P_e \cdot H(X|X', E=1) \\ &\leq P_e \cdot \log_2(n-1) \end{aligned}$$

Communication Complexity

- Multiple players each hold an input and are trying to solve a problem on the collection of their inputs
- Multiple models: blackboard setting, number-on-forehead

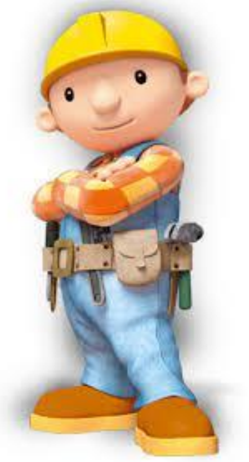
Communication Complexity

- Two-player communication problem
- Alice holds some input A and Bob holds some input B
- One-way communication or total communication



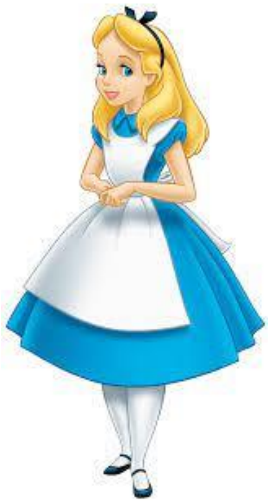
A

B



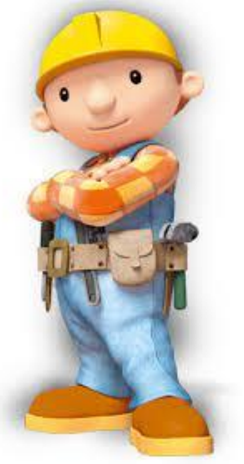
Index Problem

- Alice holds some input $A \in \{0,1\}^n$ and Bob holds some input $B := i \in [n]$
- Goal: Alice sends a message to Bob so that with probability at least $\frac{2}{3}$ (over the protocol's randomness), Bob can determine A_i



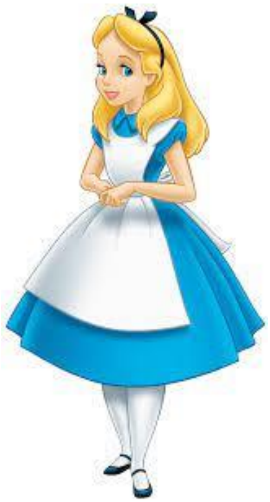
$A \in \{0,1\}^n$ $\longrightarrow$ $i \in [n]$

M



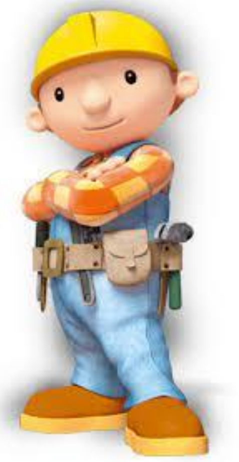
Index Problem

- Suppose $A \in \{0,1\}^n$ is drawn uniformly at random
- Alice sends M to Bob, so that for all $i \in [n]$, $\Pr\left[\widehat{A_i} = A_i\right] \geq \frac{2}{3}$
- By Fano's inequality, $H(A_i|M) \leq H\left(\frac{2}{3}\right) + \frac{1}{3}(\log_2 2 - 1) = H\left(\frac{2}{3}\right)$



$A \in \{0,1\}^n$ $\longrightarrow$ $i \in [n]$

Π



Index Problem

- By the chain rule for mutual information,

$$I(A;M) = \sum_{i\in[n]} I(A_i;M,A_1,\dots,A_{i-1})$$
$$= \sum_{i\in[n]} H(A_i|\,A_1,\dots,A_{i-1}) - H(A|M,A_1,\dots,A_{i-1})$$

- Since the bits of A are independent, $H(A_i|\,A_1,\dots,A_{i-1}) = 1$.
- Since conditioning can only decrease entropy, $H(A|M,A_1,\dots,A_{i-1}) \le H(A|M) \le H\left(\frac{2}{3}\right)$

Index Problem

- By the chain rule for mutual information,

$$I(A;M) = \sum_{i\in[n]} I(A_i; M, A_1, \ldots, A_{i-1})$$

$$= \sum_{i\in[n]} H(A_i| A_1, \ldots, A_{i-1}) - H(A|M, A_1, \ldots, A_{i-1})$$

$$= \sum_{i\in[n]} 1 - H\left(\frac{1}{3}\right) = \Omega(n)$$

- Thus, we have that $|M| \geq H(M) \geq I(A;M) = \Omega(n)$

Streaming Lower Bounds

- Alice creates a stream A and runs streaming algorithm S on A
- Send the state $S(A)$ of the algorithm to Bob
- Bob takes $S(A)$ and updates the state of the algorithm on a second part of the stream B

- If Bob solves INDEX (or some other communication problem), then the space required by streaming algorithm S is at least the one-way communication complexity of INDEX (or the other communication problem)

Streaming Lower Bounds, Example 1

- Given a stream of length m on a universe of size n, how many unique items appear in the stream?

- Alice takes A from INDEX and sends the coordinates of A
- Bob computes the number of unique items in A
- Bob then adds the number i to the stream and again computes the number of unique items in the new dataset
- If the numbers differ, then $A_i = 0$

Streaming Lower Bounds, Example 1

- Given a stream of length m on a universe of size n, how many unique items appear in the stream?

- This algorithm solves INDEX with input $\{0,1\}^n$ and thus requires space $\Omega(n)$

Streaming Lower Bounds, Example 2

- Given a stream of length m on a universe of size n inducing a frequency vector f, can we determine whether $f_i = f_j$ for a query pair i, j given after the stream?

- Alice takes A from INDEX with universe size $n - 1$ and sends the coordinates of A
- Bob asks whether $f_i = f_n$ (observe n never appears in the stream)
- If $f_i = f_n$, then $A_i = 0$. Otherwise $A_i = 1$.

Streaming Lower Bounds, Example 2

- Given a stream of length m on a universe of size n, how many unique items appear in the stream?

- This algorithm solves INDEX with input $\{0,1\}^{n-1}$ and thus requires space $\Omega(n-1) = \Omega(n)$