EC5.102: Information and Communication

Module: Information theory

Arti D. Yardi

Email address: arti.yardi@iiit.ac.in

Reference books

- Thomas M. Cover and Joy A. Thomas, "Elements of Information Theory", Wiley India press, Edition 2.

What is information?

How to quantitatively measure and represent information?

- Let us first look at how we assess the amount of information in our daily lives using common sense.
- How to quantitatively measure?

Type 1: Sachin Tendulkar retired from Professional Cricket :Known fact
Narendra Modi is Prime Minister of India :Known fact
↳ Zero information

Type 2: It will rain in Hyderabad in the month of August. Since August is known to be monsoon time: Not much uncertainty
↳ Little information

Type 3: There is one more surprise quiz today!
Are you sure? An unlikely event
↳ Large information

- Information: A numerical measure of the uncertainty of an experimental outcome.

How to quantify information?

- Consider the following three sentences.

Sun rises in the East

It will rain in Hyderabad in August month

An earthquake is going to hit Hyderabad tomorrow

Which sentence gives you more information and why?

- Consider the following three *digital* sentences.

1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1

1 1 0 1 0 1 1 0 1 0
1 0 1 0 1 1 1 0 1 1

1 0 1 1 0 0 1 0 0 1
1 0 1 0 0 0 1 0 1 1

Which sentense gives you more information and why?

- How to quantify **information**?

Information theory related questions

- How we can measure the amount of information?
- How we can ensure the correctness of information?
- What to do if information gets corrupted by errors?
- How much memory does it require to store or transmit information?
- Can we reduce the time taken to transfer the information?

Information theory

- Information theory was invented by Shannon in 1948.
- Basis of today's modern digital communication systems were laid by Shannon.
- One of the basic postulates of information theory is that information can be treated like a measurable physical quantity, such as density or mass.
- Shannon's answer: The information content of a message consists simply of the number of 1s and 0s it takes to transmit it. Information can thus be measured in bits.
- Hence, the elementary unit of information is a binary unit: a bit, which can be either 1 or 0; "true" or "false"; "yes" or "no", etc.

How to quantitatively measure and represent information?

- How to quantitatively measure?

Type 1: Sachin Tendulkar retired from Professional Cricket :Known fact
Narendra Modi is Prime Minister of India :Known fact
↳ Zero information

Type 2: It will rain in Hyderabad in the month of August. Since August is known to be monsoon time: Not much uncertainty
↳ Little information

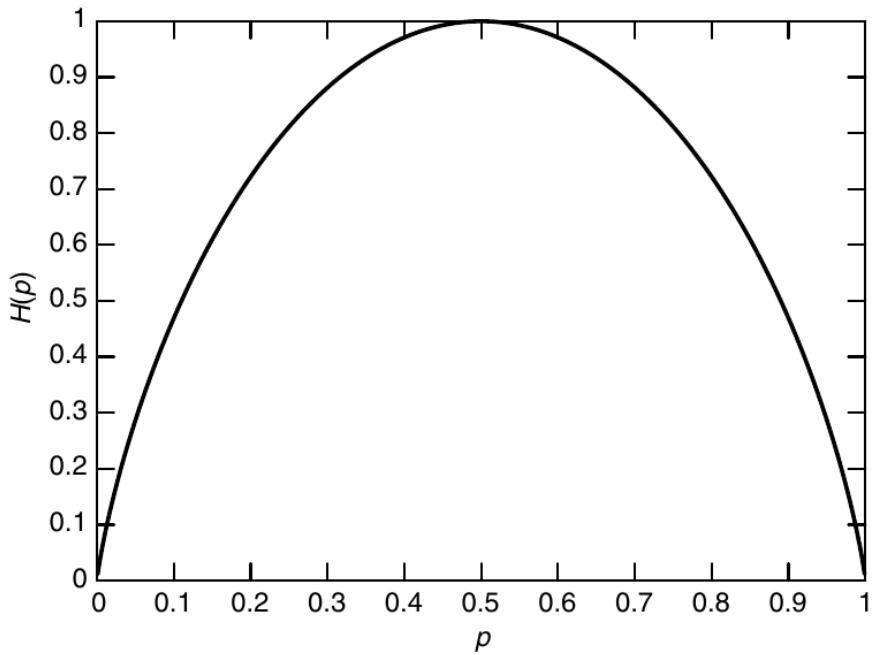
Type 3: There is one more surprise quiz today!
Are you sure? An unlikely event
↳ Large information

Information is a numerical measure of the uncertainty of an experimental outcome

Entropy: Examples

Example: Entropy of a Bernoulli random variable

- Entropy $H(X)$ of r.v. X: $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$
- Bernoulli r.v. with parameter p
  - How do you define a Bernoulli r.v. with parameter p?
  - Entropy of a Bernoulli random variable $= -p \log_2 p - (1-p) \log_2 (1-p) =: H(p)$
  - Plot of $H(p)$ versus p and its interpretation:



  - Entropy of a fair coin $= 1$ bit

Example: Entropy of a discrete uniform random variable

- Consider a discrete uniform random variable with support set $\{1, 2, \ldots, n\}$. Find its entropy.

- Example:
  - Consider a fair dice (each outcome is equally likely). Compute the entropy of this source.
  - If this dice was loaded such that outcomes 6 and 5 are more likely than others $p(X = 5) = 0.5$ and $p(X = 6) = 1/3$. The rest of the outcomes occur with equally probability. Compute the entropy in this scenario.
  - Can you observe something from the above two examples?
  - Hint: Recall the shape of entropy of Bernoulli RV ($H(p)$ vs p curve)

- $H(X) \leq log|\mathcal{X}|$, where $|\mathcal{X}|$ denotes the number of elements in the range of X, with equality if and only X has a uniform distribution over $\mathcal{X}$.

  (Proof: Not discussed)

Recap of the previous class

Recap

- Joint and conditional RV
- Information: Measure of uncertainty
- Shannon information content $h(A)$ corresponding to event A with probability $P(A)$ is defined as,

$$h(A) = \log_2 \left(\frac{1}{\mathbb{P}(A)} \right) \quad \text{bits}$$

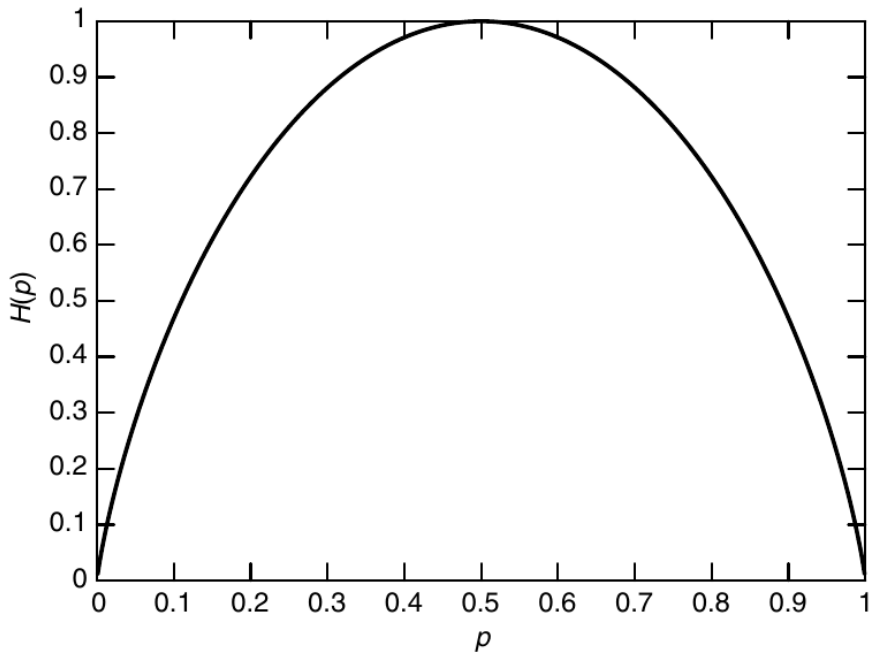
- Entropy $H(X)$ of a discrete RV X with the support set $\mathcal{X}$ is defined as

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{1}{p(x)} \right) \quad \text{bits}$$

Entropy: Examples

Example: Entropy of a Bernoulli random variable

- Entropy $H(X)$ of r.v. X: $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$
- Bernoulli r.v. with parameter p
  - How do you define a Bernoulli r.v. with parameter p?
  - Entropy of a Bernoulli random variable $= -p \log_2 p - (1-p) \log_2 (1-p) =: H(p)$
  - Plot of $H(p)$ versus p and its interpretation:



  - Entropy of a fair coin $= 1$ bit

Example: Entropy of a discrete uniform random variable

- Consider a discrete uniform random variable with support set $\{1, 2, \ldots, n\}$. Find its entropy.

- Example:
  - Consider a fair dice (each outcome is equally likely). Compute the entropy of this source.
  - If this dice was loaded such that outcomes 6 and 5 are more likely than others $p(X = 5) = 0.5$ and $p(X = 6) = 1/3$. The rest of the outcomes occur with equally probability. Compute the entropy in this scenario.
  - Can you observe something from the above two examples?
  - Hint: Recall the shape of entropy of Bernoulli RV ($H(p)$ vs p curve)

- $H(X) \leq log|\mathcal{X}|$, where $|\mathcal{X}|$ denotes the number of elements in the range of X, with equality if and only X has a uniform distribution over $\mathcal{X}$.

  (Proof: Not discussed)

Joint entropy and conditional entropy

Joint entropy and conditional entropy

- Joint entropy $H(X, Y)$ of a pair of discrete random variables (X, Y) is defined as

$$\begin{aligned} H(X, Y) &= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \\ &= -\mathbb{E} \log p(X, Y) \end{aligned}$$

- If $(X, Y) \sim p(x, y)$, the conditional entropy $H(Y|X)$ is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \\ &= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= -\mathbb{E} \log p(Y|X) \end{aligned}$$

Joint entropy and conditional entropy

- Definition of joint and conditional entropy:

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y)$$

$$H(Y|X) = \sum_{x\in\mathcal{X}} p(x)H(Y|X=x)$$

- What is $H(X,Y)$ if X and Y are independent?
- What is $H(X|Y)$ if X and Y are independent?
- Is $H(X|Y) = H(Y|X)$?
- **Is there any relation between $H(X,Y)$ and $H(X|Y)$?** Chain rule

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

(Proof: In class)

Joint entropy and conditional entropy: Examples

Joint entropy and conditional entropy: Example-1

- $H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$

  $H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y)$

- Example:

  Let (X, Y) have the following joint distribution:

X \ Y	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

- Find the following

  $H(X|Y) =$? bits

  $H(Y|X) =$? bits

  $H(X, Y) =$? bits

Joint entropy and conditional entropy: Example-2

- $H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$

  $H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y)$

- Example:

  Let (X, Y) have the following joint distribution:

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

- Find the following

  $H(X|Y)$

  $H(Y|X)$

  $H(X, Y)$

Joint entropy and conditional entropy: Example-3

Assume that the probability of being male is $p(M) = 0.5$ and so likewise for being female $p(F) = 0.5$. Suppose that 20% of males are T (i.e. tall): $p(T|M) = 0.2$; and that 6% of females are tall: $p(T|F) = 0.06$.

- Calculate the probability that if somebody is "tall" (meaning taller than 6 ft or whatever), that person must be male (Calculate $p(M|T)$).
- If you know that somebody is male, how much information do you gain (in bits) by learning that he is also tall?
- How much do you gain by learning that a female is tall?
- Finally, how much information do you gain from learning that a tall person is female?

(Homework)

Recap of the previous class

Recap

- Entropy = Information = Uncertainty
- Shannon information content $h(A)$ is information associated with event A is

$$h(A) = -\log \mathbb{P}(A)$$

- Entropy $H(X)$ of X is uncertainty associated with RV X, defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

- Joint entropy of X and Y is the combined information of X and Y, defined as

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$

- Conditional entropy $H(X|Y)$ is uncertainty about X given Y, defined as

$$H(X|Y) = \sum_{y \in \mathcal{Y}} H(X|Y = y) p(y)$$

$$H(X|Y = y) = -\sum_{x \in \mathcal{X}} p(x|y) \log p(x|y)$$

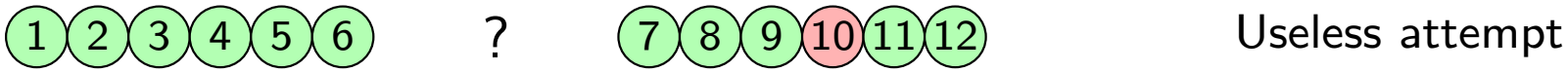
Weighing Puzzle

Weighing puzzle

- You are given 12 balls, all equal in weight except for one that is either heavier or lighter.
- You are also given a two-pan balance and in each use of the balance you may put any number balls in the pan. Outcome of the pan can be:
  - Left and right pan weigh the same.
  - Left pan is heavier.
  - Left pan is lighter.
- Your task is to design a strategy to determine which is the odd ball and whether it is heavier or lighter than the others in as few uses of the balance as possible.
- Attempt to find a solution before you check the answer.

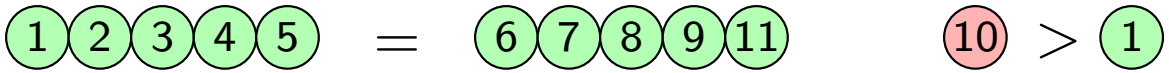
Weighing puzzle: Solution

- Suppose the balls are numbered as $1, 2, \ldots, 12$.
- Attempt 1:



Useless attempt

- Attempt 2: Pan is used two times.

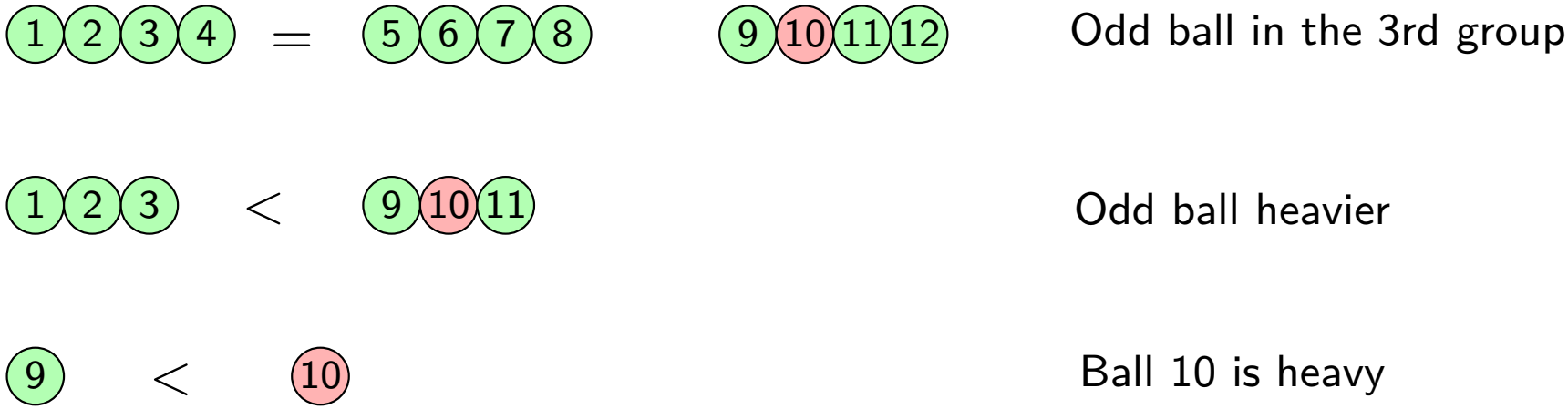


"Lucky outcome"!

  - You might not be always lucky!
  - Can you always find the odd ball by using pan at most two times?
  - If yes, can you prove it?
  - If no, can you provide a counter example?
  - Answer: Using pan at most two times is not sufficient for all possible arrangements!

Weighing puzzle: Solution

- Attempt 3: Pan is used three times.

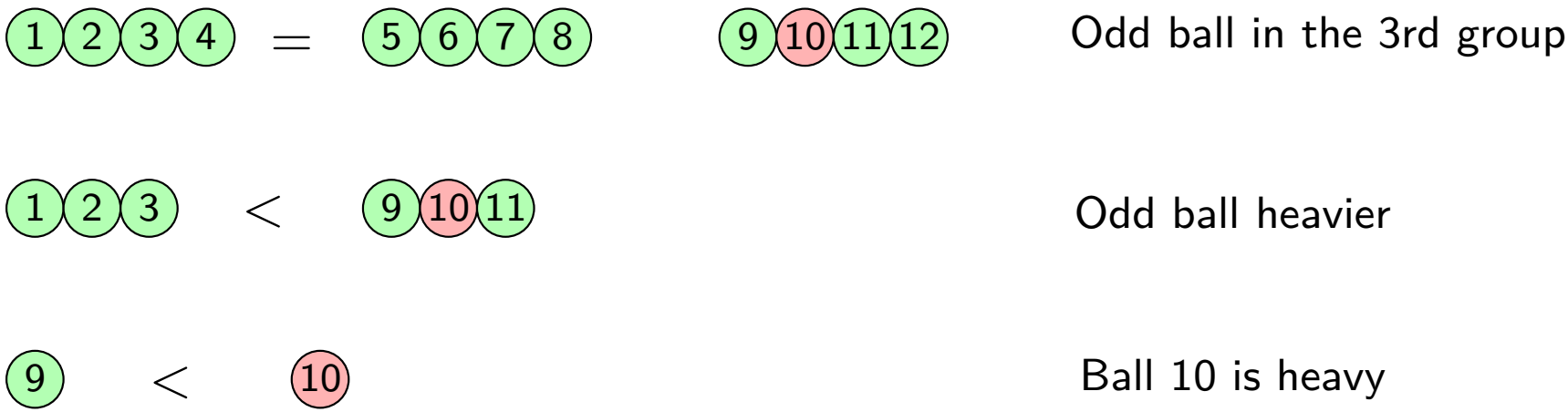


  - You might not be always lucky!
  - Can you always find the odd ball by using pan at most three times?
  - If yes, can you prove it?
  - If no, can you provide a counter example?
  - Answer: Using pan at most three times is sufficient for all possible arrangements!

Information theory provide answers to such puzzles!

Information theoretic insights to weighing puzzle

- An optimal solution consists of using pan three times



- Why is this solution optimal?
- Hint: At each step of an optimal procedure, the three outcomes of the pan ("left heavier", "right heavier", and "balanced") are as close as possible to equiprobable.
- Can you see a connection to entropy of a discrete RV?

Relative entropy

Relative entropy

- Entropy: Measure of uncertainty of a random variable
- Relative entropy: Measure of distance between two distributions
- Consider two pmfs $p(x)$ and $q(x)$
  - If $p(x) = q(x)$, then we want relative entropy between them to be equal to 0
  - If $p(x)$ and $q(x)$ are too different from each other, we want relative entropy $= \infty$
- Relative entropy $D(p||q)$ between $p(x)$ and $q(x)$ is defined as

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

  - When $D(p||q) = 0$?
  - When $D(p||q) = \infty$?
- $D(p||q)$ is also called as Kullback-Liebler (KL) distance/divergence

Mutual information

Mutual information

- Consider two random variables X and Y with a joint pmf $p(x, y)$.
- Product distribution definition: $p(x, y) = p(x)p(y)$
- Recall: Relative entropy $D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$
- Mutual information $I(X; Y)$ is defined as relative entropy between joint distribution $p(x, y)$ and the product distribution $p(x)p(y)$

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

- Relationship between entropy and mutual information:

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

(Proof: In class)

- $I(X; Y) = H(Y) - H(Y|X)$? Yes/No?

Mutual information: Example

- $I(X;Y) = H(X) - H(X|Y)$
- Example:

  Let (X, Y) have the following joint distribution. Find $I(X;Y)$

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

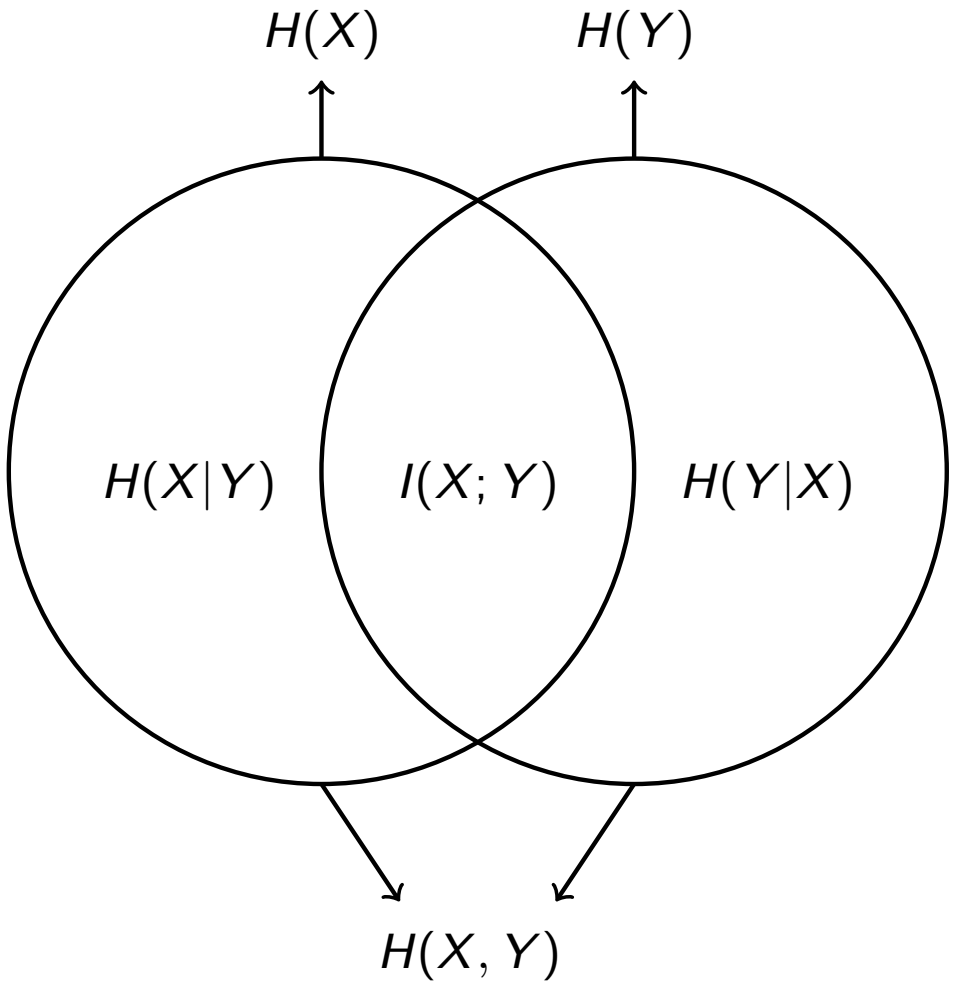
  - From Example-3 we have:

    $H(X) = 7/4$ bits

    $H(X|Y) = 11/8$ bits

  - $I(X;Y) = 7/4 - 11/8 = 3/8$ bits

Mutual information and entropy



$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= H(X) + H(Y) - H(X,Y) \\
&= I(Y;X)
\end{aligned}
$$

Chain rules

- Chain rule of probability
  - $P(A_1, A_2) = P(A_1)P(A_2|A_1)$
  - $P(A_1, A_2, A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)$
- Chain rule for entropy:
  - $H(X, Y) = H(X) + H(Y|X)$
  - For $n = 3$: $H(X_1, X_2, X_3) = H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1)$
- Chain rule for mutual information:
  - $I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1)$
  - $I(X_1, X_2, X_3; Y) = I(X_1; Y) + I(X_2; Y|X_1) + I(X_3; Y|X_2, X_1)$
- We will not prove these.
- Can you write an expression for n RVs?