Basics of Information Theory

Lecture 1

Samuel Brody

February 2009

Background

Developed by Claude Shannon
at Bell labs in the 1940's

"A Mathematical Theory of Communication",
Bell System Technical Journal, 27, 1948

- intended to describe the process of conveying information through a noisy channel
- deals with the nature of information and uncertainty
- huge impact in many fields

Entropy

- $p(x)$ is a probability mass function of variable X , over a discrete set of symbols (or alphabet) $X = \{x_1, x_2, ..., x_n\}$:

$$p(x_i) = Prob(X = x_i), \quad x_i \in X$$

- The entropy (or self-information) is the average uncertainty of a single random variable:

$$H(p) = H(X) = -\sum_{x \in X} p(x) log_2 p(x)$$

Entropy - cont.

Definition:

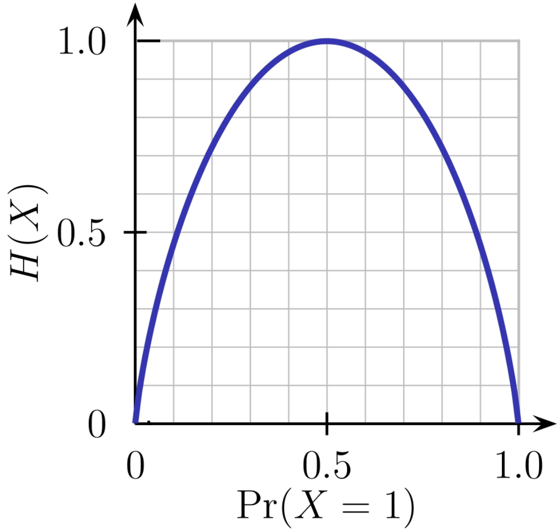
- $H(X) = -\sum_{x \in X} p(x) log_2 p(x)$
- (weighted) average uncertainty
- the number of *yes / no* questions
- complete certainty*= 0, complete uncertainty = $log_2|X|$

Examples:

- single fair coin toss: $H(X) = -\frac{1}{2} log_2 \frac{1}{2} - \frac{1}{2} log_2 \frac{1}{2} = 1$
- number of heads in two coin tosses:
$$H(X) = -\frac{1}{4} log_2 \frac{1}{4} - \frac{1}{2} log_2 \frac{1}{2} - \frac{1}{4} log_2 \frac{1}{4} = \frac{2}{4} + \frac{1}{2} + \frac{2}{4} = 1\frac{1}{2}$$
- random letter in a book:
If all letters were equally likely, it would be $log_2 26$. In practice, much less.

*by definition, $0 log 0 = 0$

Entropy of Biased Coin Toss

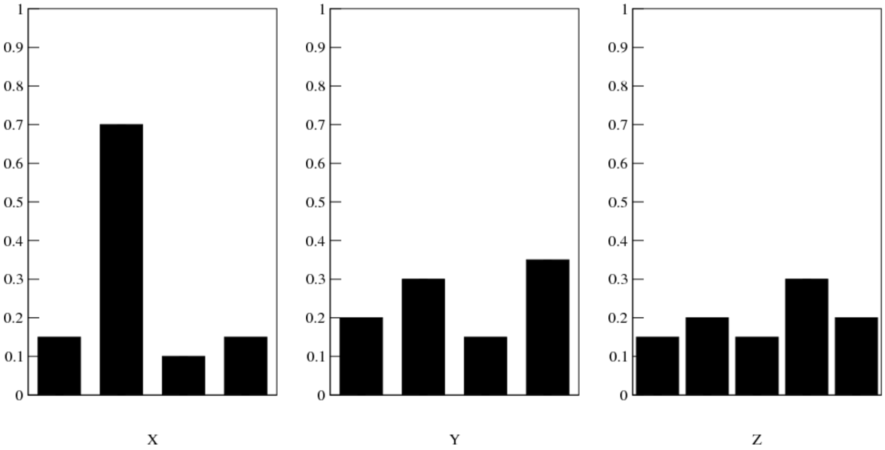


The amount of uncertainty depends on:

- the number of choices

- the distribution of probabilities

Some Intuition



$$H(X) < H(Y) < H(Z)$$

Joint Entropy

- The joint entropy of a pair of discrete random variables $X, Y \sim p(x, y)$ is the amount of information needed on average to specify both their values.

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) log_2 p(x, y)$$

Joint Entropy - Examples

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) log_2 p(x, y)$$

Examples:

1. $X \sim$ a fair throw of a die, $Y \sim$ a fair coin toss

	X = 1	X = 2	X = 3	X = 4	X = 5	X = 6
Y = H	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
Y = T	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

2. $X \sim$ a fair throw of a die, if the result is odd, $Y \sim$ a fair coin toss, otherwise, Y is heads.

	X = 1	X = 2	X = 3	X = 4	X = 5	X = 6
Y = H	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$
Y = T	$\frac{1}{12}$	0	$\frac{1}{12}$	0	$\frac{1}{12}$	0

- in the first case, $H(X, Y) = log_2 12 = log_2 6 + log_2 2 = H(X) + H(Y)$
- in the second case, $H(X, Y) < log_2 12$, because knowledge of X gives us some knowledge about Y.

Conditional Entropy

The conditional entropy of a discrete random variable Y given another variable X is the remaining uncertainty of Y given that the other party knows the value of X :

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) log_2 p(y|x)$$

Chain Rule

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

$$H(X_1, ..., X_n) = H(X_1) + H(X_2|X_1) + ... + H(X_n|X_1, ..., X_{n-1})$$

Dice and coin examples:

- in the first case, the variables were independent, so ...

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

- in the second example, X provides information about Y, so ...

$$H(Y|X) < H(Y)$$

Mutual Information

By the chain rule for entropy,

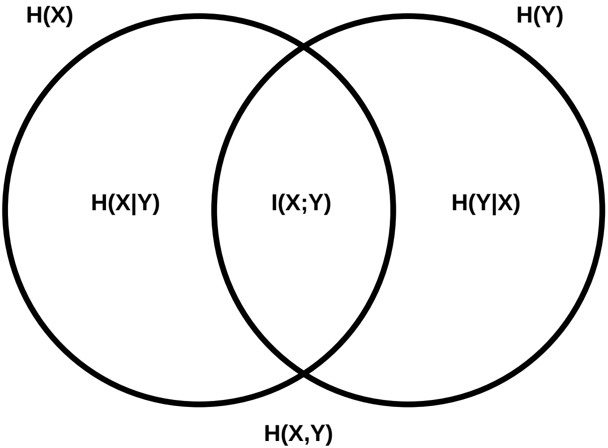
$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

Therefore,

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

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

Mutual Information



Characteristics of Mutual Information

- symmetric
- non-negative
- zero if variables are independent

- Also, since $H(X|X) = 0$ it follows that:

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

Common Cases of Mutual Information

- Correlated Phenomena
  - height and weight
  - obesity and heart disease
- Observable features indicating unobserved state
  - symptoms and diseases
  - protein expression and cell cycle
  - words used and emotions/opinion of the author
  - SAT scores and intelligence (?)
- Sequences (*note: MI does not depend on order!*)
  - weather patterns
  - medical history
  - stock prices

Summary

Entropy

- expresses the amount of uncertainty about the state of a variable
- a function of a probability distribution over the possible states

Mutual Information

- the amount of information shared by two different variables

or

- how much one variable tells us about the other