Lab 9

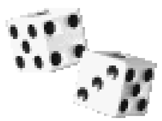
Dr. Weitao Xu

Information theory

- Entropy
- Joint distribution
- Conditional distribution

Measurement of Information

“How to measure information in terms of bits?”



= ? bits

= ? bits

Shannon’s Information Theory

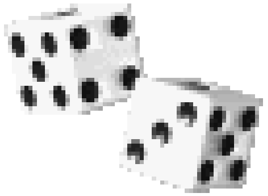
Shannon’s measure of information is the number of bits to represent the amount of uncertainty (randomness) in a data source, and is defined as entropy

$$H = -\sum_{i=1}^{n} p_i \log(p_i)$$

Where there are *n* symbols *1, 2, … n*, each with probability of occurrence of *pi*

For example

- Tossing a dice:
  - Outcomes are 1,2,3,4,5,6
  - Each occurs at probability 1/6
  - Information provided by tossing a dice is



$$H = -\sum_{i=1}^{6} p(i)\log_2 p(i) = -\sum_{i=1}^{6} p(i)\log_2 p(i)$$

$$= -\sum_{i=1}^{6} \frac{1}{6}\log_2 \frac{1}{6} = \log_2 6 = 2.585\ \text{bits}$$

- Entropy is greatest when the probabilities of the outcomes are equal

Let's consider a fair coin experiment

- The entropy $H = -\frac{1}{2} \log 0.5 - \frac{1}{2} \log 0.5 = 1$
- Consider a biased coin, $P(H) = 0.98$, $P(T) = 0.02$
- $H = -0.98 * \log 0.98 - 0.02 * \log 0.02 =$

  $= 0.98 * 0.029 + 0.02 * 5.643 = 0.0285 + 0.1129 = 0.1414$

Joint distribution

- If *X and Y are two random variables, the probability distribution that defines* their simultaneous behavior is called a joint probability distribution.

For example:

- *Suppose Person A is rolling a dice, and person b is flipping a coin. Every time A rolls a dice, B flips a coin. They repeat this process by 10 times*
- *X is a random variable representing the outcomes of rolling a dice* (X can be 0,1,2,3,4,5,6)
- *Y is a random variable representing the outcomes of flipping a coin (Y can be H, T)*

Experiment No	1	2	3	4	5	6	7	8	9	10
X	6	4	2	1	3	4	5	2	5	6
Y	H	T	H	T	T	H	T	H	T	H

(2,H) happens 2 times, so P(2,H)=2/10

Pi,j	1	2	3	4	5	6
H	0	0.2	0	0.1	0	0.2
T	0.1	0	0.1	0.1	0.2	0

For joint distribution, we have $\sum_x \sum_y p_{X,Y}(x,y) = 1$

Conditional distribution

- If *X and Y are two random variables,* conditional distribution P(Y=y|X=x) means the probability that Y=y happens given X=x.

For example:

- *Suppose Person A is rolling a dice, and person b is flipping a coin. Every time A rolls a dice, B flips a coin. They repeat this process by 10 times*
- *X is a random variable representing the outcomes of rolling a dice* (X can be 0,1,2,3,4,5,6)
- *Y is a random variable representing the outcomes of flipping a coin (Y can be H, T)*

Experiment No	1	2	3	4	5	6	7	8	9	10
X	6	4	2	1	3	4	5	2	5	6
Y	H	T	H	T	T	H	T	H	T	H

1 happens 1 times, but (1,H) never happens in the above table, so P(H|1)=0/1=0
(1,T) happens 1 time, so P(T|1)=1/1=1

Py|x	1	2	3	4	5	6
H	0	1	0	0.5	0	1
T	1	0	1	0.5	1	0

For conditional distribution, we have the sum of Py|x is 1 for every x.

How to calculate each conditional distribution?

	1	2	3	4	5	6	7	8	9	10
Dice	6	4	2	1	3	4	5	2	5	6
coin	H	T	H	T	T	H	T	H	T	H

For Py|x: Calculate the number of events (x,y)happens divided by the number of events x happens.
For example, 4 happens 2 times, and (4,T) happens 1 time, (4,H) happens 1 time, so PT|4 is ½=0.5, PH|4 is ½=0.5

Py|x	1	2	3	4	5	6
H	0/1	1/1	0/1	1/2	0/1	1/1
T	1/1	0/1	1/1	1/2	1/1	0/1