PART C: Measure of Information

✓ Quantities OF information

Information theory is based on probability theory and statistics. It's often concerns with measures of information of the distributions associated with random variables. Important quantities of information are **Entropy** which is *a measure of information in a single random variable*, and **Mutual Information** which is a measure of information in common between two random variables. The choice of *logarithm base* formulae determines the unit of information entropy that is used. A common unit of information is **Bit**, based on the binary logarithm. Other units include the **Nat**, which is based on the natural logarithm, and the **Hartley**, which is based on the common logarithm. In what follows, an expression of the form $p \log p$ is considered by convention to be equal to zero whenever P approximate to 0.

✓ Measure of Information

The occurrence of an event of low probability causing a big surprise and it reaches more information than others resulting from the occurrence of an event of high probability. In other words, the information associated with the element of surprise. Since, information concerned with the amount of information, not their quality, which is calculated by using a probability. The measure of information is concerns of an event occurs and neglecting any other details.

Let P : Probability of the occurrences of event.

I : The information gained.

$$I \propto \frac{1}{p}$$

$$I = f(p)$$

When $P \to 1, \quad I \to 0$

$P \to 0, \quad I \to 1 \qquad (0 \le P \le 1)$

In general small P gives large I and vice versa.

✓ Self – Information

Shannon was able to derive a method for measuring information called **self-information** of a message m:

$$I(m) = \text{Log}\,(1/P(m))$$

$$I(m) = -\,\text{Log}\,(P(m))$$

Where, **P(m)** is the probability that message ***m*** is chosen from all possible choices in the message space **M**. The base of logarithm affects a scaling factor and, consequently, the unit of measured information is expressed. In other words, logarithm of **base 2**, will expressed in unit of **bit**.

When using the logarithm of **base** e, the unit will be the nat. For the log of **base 10**, the unit will be **hartley.**

Note

You can use the following equation to convert from logarithm in base 2 to logarithm in base 10.

$$\mathbf{Log_2(v) = 3.322*log_{10}(v)}$$

If an event *C* represents **intersection** of two independent events *A* and *B*, then the amount of information which show that *C* has happened, equals to the **sum** of the amounts of information for each of *A* and *B* respectively.

$$I(A \cap B)=I(A)+I(B)$$

Example 1**:** On tossing a coin, the chance of 'tail' is 0.5. When it is proclaimed that indeed 'tail' occurred, this amounts to

I('tail') = $\log_2$ (1/0.5) = $\log_2$ 2 = 1 bits of information.

OR

I('tail') = $\log_2$ (1/0.5) = 3.322*$\log_{10}$(2) = 1 bits of information.

Example 2**:** When throwing a fair dice, the probability of 'four' is 1/6. When it is proclaimed that 'four' has been thrown, the amount of self-information is

I('four') = $\log_2$ (1/(1/6)) = $\log_2$ (6) = 2.585 bits.

OR

I('four') = $\log_2$ (1/(1/6)) = 3.322*$\log_{10}$ (6) = 2.585 bits.

Example 3**:** When, independently, two dice are thrown, the amount of information associated with {throw 1 = 'two' & throw 2 = 'four'} equals

I('throw 1 is two & throw 2 is four') = $\log_2$ (1/P(throw 1 = 'two')+ $\log_2$ (1/P(throw 2 = 'four'))

= $\log_2$ (1/(1/6)) + $\log_2$(1/(1/6))

= 3.322*$\log_{10}$ (6) + 3.322*$\log_{10}$ (6)

= 2.585 + 2.585 = 5.170 bits.

Example 4**:** In the same two dice situation we can also consider the information present in the statement "The sum of the two dice is five".

Sample space = 36

The total chances to get five = 4 ways (1+4, 4+1, 2+3, 3+2)

The probability = 4/36 = 1/9

I('The sum of throws 1 and 2 is five') = $\log_2$ (1/P('throw 1 and 2 sum to five'))

= $\log_2 (1/(4/36))$ Or $\log_2 (1/(1/9))$

= $\log_2 (9)$

= $3.322 * \log_{10} (9) = 3.17$ bits.

✓ Properties of Information

1. Information is positive quantity $I(x) > 0$
2. If $P(x) = 1$ then $I(x) = 0$ that's mean there is no information.
3. For any tow events we are using Joint Probability.

$$I(X_1 * X_2) = I(X_1) + I(X_2)$$

$$I(X)^2 = I(X*X) = I(X) * I(X) = 2\, I(X)$$

So …

$$I(X^n) = n * I(X)$$

4. The information joint with probability. So any changes in the probability causes changes in the information.

Example 1**:** Find the amount of information for tossing a coin:

On tossing a coins the chance of a "tail" or " head " is equal. So the probability P ("tail") or P (" head") = 0.5. That's mean

❖ Tossing a coins one times gives the information as following:

$I = -\log_2 (P) = \log_2 (1/P)$

$= -\log_2 (1/2) = \log_2 (2)$

$= 1$ bit

❖ Tossing a coins **n** times gives the information as following:

$I = -\log_2 (P^n) = \log_2 ((1/P)^n)$

$= -\log_2 ((1/2)^n) = \log_2 (2^n)$

$= n * \log_2 (2)$

$= n$ bits

***Example 2*:** Let A, B, C and D be for characters have the following probabilities 1/2 , 1/4, 1/8 and 1/8. Find the amount of self-information for each character, then find the amount of information for a message X=BDA. Suppose that all characters are independent.

$I_A = \log_2(1/P_A)$

$= \log_2(2) = 1$ bit

$I_B = \log_2(1/P_B)$

$= \log_2(4) = 2$ bit

$I_C = \log_2(1/P_C)$

$= \log_2(8) = 3$ bit

$I_D = \log_2(1/P_D)$

$= \log_2(8) = 3$ bit

The amount of information in the message X=BDA is :

$I_X = I_B + I_D + I_A$

$= 2 + 3 + 1$

$= 6$ bits