✓ Conditional entropy

The **conditional entropy** or **conditional uncertainty** (also called the **equivocation** of X about Y) *is a measure to determine the uncertainty of a random variable Y depending on the value of another variable X*. This condition occurs in a succession of symbols so that each symbol which affected the symbol next to it. The conditional entropy is calculated according to the following formula:

$$H(Y|X) = \sum_{x,y} P(x,y) \log2(1/P(y|x)) \quad \text{unit of information}$$

$$H(Y|X) = H(x, y) – H(x)$$

$$H(X|Y) = H(x, y) – H(y)$$

✓ Redundancy

Redundancy: Is the concept in information theory and occurs due to correlation letters with each other, this means existence of unimportant or incorrect symbols in the message more than is necessary. The redundancy is calculated according to the following relationship:

$$R = 1 – \text{Actual Entropy}$$

***Example 1*:** Simple language consists of A, B symbols and produces the following chain (AAABBAAAABBB). Calculate:

1. Probability for each of A & B.
2. Conditional entropy for the chain.
3. Entropy in the absence of correlation.
4. Redundancy of this language.

1) P(A) = 7/12 = 0.583 P(B) = 5/12 = 0.417

2) $H = \sum_{i=1}^{n} P_i \log_2(1/P_i)$

$= P_A * \log_2 (1/P_A) + P_B * \log_2 (1/P_B)$

$= 7/12*3.322*\log_{10}(12/7) + 5/12*3.322*\log_{10}(12/5)$

$= 0.583*3.322*0.234 + 0.417*3.322*0.38$

$= 0.453 + 0.526$

$= 0.979$ bit

Note : The Conditional Entropy occurs in a succession of symbols so that each symbol which affected the symbol next to it.

3) P(A, A) = 5/12 $P(A|A) = P_{(A,A)}/P_{(A)} = 5/12 * 12/7 = 5/7$

P(A, B) = 2/12 $P(B|A) = P_{(A,B)}/P_{(A)} = 2/12 * 12/7 = 2/7$

P(B, A) = 1/12 $P(A|B) = P_{(A,B)}/P_{(B)} = 1/12 * 12/5 = 1/5$

P(B, B) = 3/12 $P(B|B) = P_{(B,B)}/P_{(B)} = 3/12 * 12/5 = 3/5$

$H(Y|X) = \sum_{x,y} P(x,y) \log_2(1/P(x|y))$
$= P(_{A, A}) * \log_2(1/P(_{A|A})) + P(_{A, B}) * \log_2(1/P(_{B|A}))$
$+ P(_{B, A}) * \log_2(1/P(_{A|B})) + P(_{B, B}) * \log_2(1/P(_{B|B}))$
$= 5/12 * 3.322 * \log_{10}(7/5) + 2/12 * 3.322 * \log_{10}(7/2)$
$+ 1/12 * 3.322 * \log_{10}(5/1) + 3/12 * 3.322 * \log_{10}(5/3)$
$= 0.417 * 3.322 * 0.699 + 0.167 * 3.322 * 0.544$
$+ 0.083 * 3.322 * 0.193 + 0.25 * 3.322 * 0.222$
$= 0.202 + 0.302 + 0.193 + 0.184 =$ **0.881 bit**

4) R = 1 – actual entropy
= 1 – 0.881
= 0.119 bits with correlation

***Example2*:** Simple language consists of A, B symbols and produces the following chain (ABABBBBAAAABBABAB). Calculate:

1. Probability for each of A & B.
2. Entropy in the absence of correlation.
3. Conditional entropy for the chain.
4. Redundancy of this language.

1) P(A) = 8/17 = 0.471 P(B) = 9/17 = 0.529
2) $H = \sum_{i=1}^{2} P_i \log_2(1/P_i)$

$= P_A * \log_2(1/P_A) + P_B * \log_2(1/P_B)$
$= 8/17*3.322*\log_{10}(17/8) + 9/17*3.322*\log_{10}(17/9)$
$= 0.471*3.322*0.327 + 0.529*3.322*0.276$
$= 0.512 + 0.485 =$ **0.997 bit**

3) P(A, A) = 3/17 P(A|A) = 3/8
P(A, B) = 5/17 P(B|A) = 5/8
P(B, A) = 4/17 P(A|B) = 4/9
P(B, B) = 4/17 P(B|B) = 4/9

$H(Y|X) = \sum_{x,y} P(x, y) \log_2(1/P(x|y))$

$= P(_{A, A}) * \log_2(1/P(_{A|A})) + P(_{A, B}) * \log_2(1/P(_{B|A}))$
$+ P(_{B, A}) * \log_2(1/P(_{A|B})) + P(_{B, B}) * \log_2(1/P(_{B|B}))$
$= 3/17 * 3.322 * \log_{10}(8/3) + 5/17 * 3.322 * \log_{10}(8/5)$
$+ 4/9 * 3.322 * \log_{10}(9/4) + 4/9 * 3.322 * \log_{10}(9/4)$
$= 0.176 * 3.322 * 0.426 + 0.294 * 3.322 * 0.204$
$+ 0.444 * 3.322 * 0.352 + 0.444 * 3.322 * 0.352$
$= 0.249 + 0.199 + 0.519 + 0.519$
= 1.486 bits

4) R = 1 – actual entropy
   = 1 – 1.486
   = - 0.486 bits with correlation

✓ Mutual Information

Mutual Information *is the term that refers to the measure of information transmission across channels by measuring the information shared between X and Y*. In other words it is the amount of information that can be obtained by the recipient after receiving the message from the sender.

Assume that X represents a series of sent symbols, Y represents a series of received symbols, for each X is sent, there is y received. So the mutual information I(Xi, Yj) can be calculated according to the following law:

Mutual Information = information received - information lost

$$I(X_i, Y_j) = I(X_i) - I(X_i|Y_j)$$
$$= -\log(P(X_i)) - (-\log(P(X_i|Y_j)))$$

$$I(X_i, Y_j) = \log\frac{P(Xi|Yj)}{P(Xi)} \quad \text{unit of information}$$

Prove that: $I(X_i, Y_j) = I(Y_j, X_i) = \log\frac{P(Xi|Yj)}{P(Xi)}$

<u>Proof:</u>

$$P(X_i|Y_j) = \frac{P(Xi,Yj)}{P(Yj)}$$

$$I(X_i, Y_j) = \log\frac{\frac{P(Xi,Yj)}{P(Yj)}}{P(Xi)}$$

$$P(Xi, Yj) = P(Yj, Xi) = P(X_i) * P(Y_j)$$
$$P(Yj, Xi) = P(X_i) * P(Y_j | X_i)$$

$$I(X_i, Y_j) = \log\frac{\frac{P(Xi)*P(Yj|Xi)}{P(Yj)}}{P(Xi)}$$

$$I(X_i, Y_j) = \log\frac{P(Yj|Xi)}{P(Yi)}$$

✓ Limits Of Mutual Information

Let X being a series of symbols sent, and Y is series of symbols received, the mutual information will calculated as follows:

$$I(X_i, Y_j) = \log\frac{P(Xi|Yj)}{P(Xi)}$$

- If X sent through a noiseless channel, it will calculated as follows:

$$P(X_i|Y_j) = 1$$

And

$$I(X_i,Y_j)_{MAX} = \log(\frac{1}{P(Xi)})$$

$$I(X_i,Y_j)_{MAX} = I(X_i)$$

- If X sent through a noisy channel, it will calculated as follows:

$$P(X_i|Y_j) = P(X_i)$$

And

$$I(X_i,Y_j)_{MIN} = \log(\frac{P(Xi)}{P(Xi)})$$

$$= \log(1)$$

$$= 0$$ ***No Information is transferred***

So the limit of mutual information is:

When $I(X_i,Y_j)_{MIN} = 0 \Longrightarrow I(X_i,Y_j)_{Max} = I(X_i)$

$\Downarrow$ $P(X_i|Y_j) = P(X_i)$ $\qquad$ $\Downarrow$ $P(X_i|Y_j) = 1$