Conditional Entropy

- $H(Y|X)$ is the expected value of $H(Y|X=x)$

$$H(Y|X) = \Sigma_{x\in X}\, p(X=x)\, H(Y|X=x)$$

$$= \Sigma_{x\in X}\, p(X=x)\, [\, -\Sigma_{y\in Y}\, p(Y=y|X=x) \log p(Y=y|X=x)]$$

$$= -\Sigma_{x\in X}\, \Sigma_{y\in Y}\, p(X=x, Y=y) \log p(Y=y|X=x)$$

$$= E_{p(X,Y)}\, (-\log p(Y=y|X=x))$$

- $H(Y|X)$ is the expected value of $-\log p(Y=y|X=x)$

In the example

P(XY)	Y=0	Y=1
X=0	1/3	1/6
X=1	1/6	1/3

- Find $H(Y|X)$

- $H(Y|X) = p(X=0)\, H(Y|X=0) + p(X=1)\, H(Y|X=1)$
- $p(X = 1) = p(X = 0) = 1/2$

- $H(Y|X=0) = - \Sigma_{y \in \{0,1\}} p(Y=y|\, X=0) \log p(Y=y|X=0)$

- $P(Y=0|X=0)= p(Y=0, X=0)/ p(X=0) = (2/6)/(1/2)=2/3$
- $P(Y=1|X=0)= p(Y=1, X=0)/ p(X=0) = (1/6)/(1/2)=1/3$

- $H(Y|X=1) = - \Sigma_{y \in \{0,1\}} p(Y=y|\, X=1) \log p(Y=y|X=1)$

- $P(Y=0|X=1)= p(Y=0, X=1)/ p(X=1) = (1/6)/(1/2)=1/3$
- $P(Y=1|X=1)= p(Y=1, X=1)/ p(X=1) = (2/6)/(1/2)=2/3$

H(Y|X)

$$H(Y \mid X=0) = -[p(Y=1 \mid X=0)\log p(Y=1 \mid X=0) +$$

$$p(Y=0 \mid X=0)\log p(Y=0 \mid X=0)] = -(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}) = 0.918$$

$$H(Y \mid X=1) = -[p(Y=1 \mid X=1)\log p(Y=1 \mid X=1) +$$

$$p(Y=0 \mid X=1)\log p(Y=0 \mid X=1)] = -(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}) = 0.918$$

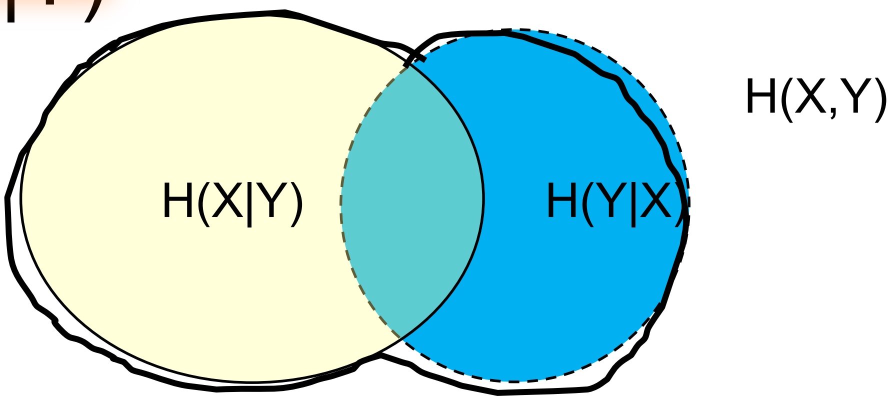
$$H(Y \mid X) = p(X=0)H(Y \mid X=0) + p(X=1)H(Y \mid X=1)$$

$$= (\frac{1}{2} \times 0.918) + (\frac{1}{2} \times 0.918) = 0.918 \qquad bit$$

- For this example: H(X,Y)= H(X) + H(Y|X)
- 1.918 = 1 + 0.918

Theorem

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



- (Theorem 2.2.1 - CT)
- (Example 2.2.1 - CT)

- Note that $H(Y|X) \neq H(X|Y)$

Mutual Information

- X and Y are two random variables with joint probability distribution $p(x,y)$.
- The information that observing Y gives about X is
  - $I(X;Y) = H(X) - H(X|Y)$
- This is the expected reduction in uncertainty about X when Y is known.
  - Similarly the information that X gives about Y is
  - $I(Y;X) = H(Y) - H(Y|X)$

Mutual Information

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

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

  $\quad - (-\Sigma_{x \in X} \Sigma_{y \in Y} p(X=x,Y=y) \log p(X=x|Y=y)$

  $= \Sigma_{x \in X} \Sigma_{y \in Y} p(X=x,Y=y) \log [p(X=x|Y=y) / p(X=x)]$

  $= \Sigma_{x \in X} \Sigma_{y \in Y} p(X=x,Y=y) \log [p(X=x, Y=y) / p(X=x)p(Y=y)]$

  $= \Sigma_{x \in X} \Sigma_{y \in Y} p(X=x,Y=y) \log [p(Y=y| X=x) / p(Y=y)] = I(Y;X)$

  That is: $I(X;Y) = I(Y;X)$

  – X says as much about Y as Y says about X.

Example: Erasure Channel

An information source:

p(X=0)= 2/3 ,   p(X=1)= 1/3

H(X) =.918 bits

X

01011

Y

?10?1

Example: Erasure Channel

An information source:
p(X=0)= 2/3 ,   p(X=1)= 1/3
H(X) =.918 bits

- How much is H(Y)?

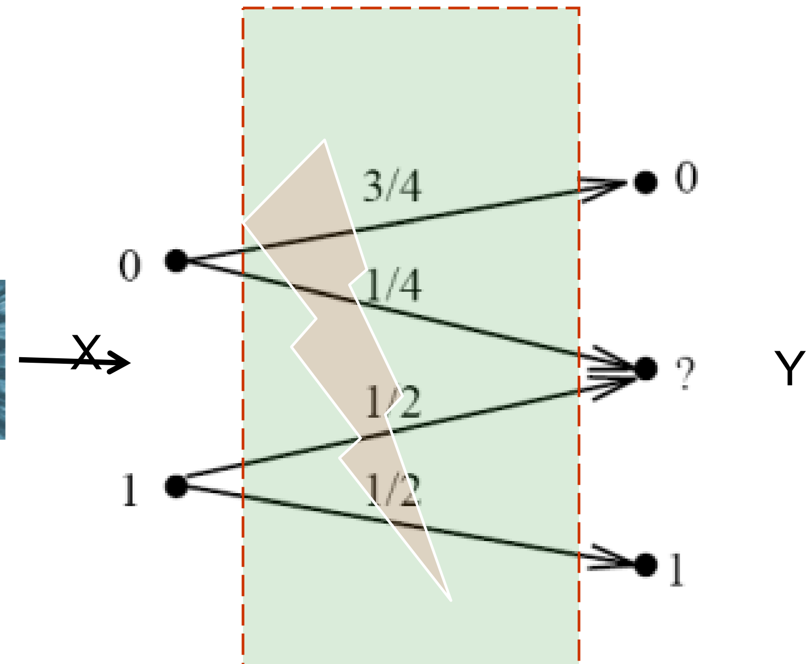
Find p(y):

- p(Y=0) = p(X=0)p(Y=0|X=0)
  =2/3 x 3/4 =1/2

- p(Y=1) = p(X=1)p(Y=1|X=1)
  =1/3 x 1/2 =1/6

- p(Y=?) = p(X=1)p(Y=?|X=1) +
  p(X=0)p(Y=?|X=0)

  =1/3 x 1/2 + 2/3 x 1/4 = 1/3



Entropy of Y

- Find $H(Y\ ) = -\sum_{y \in Y} p(y) \log p(y)$

  $= 1/2 \times 1 + 1/6 \times 2.58 + 1/3 \times 1.58$

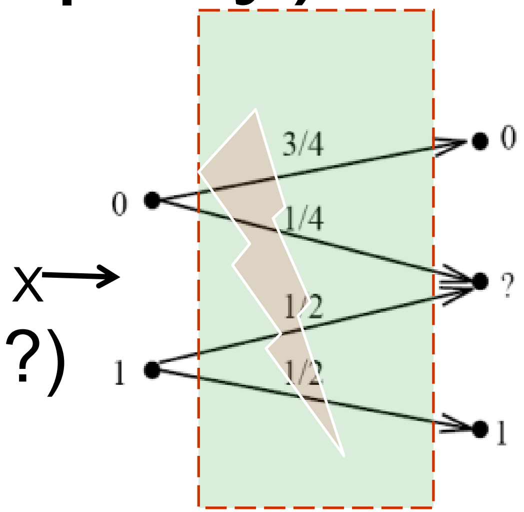
  $= 1.95$ bits

- $H(Y) > H(X)$. Has information increased?

- $I(X;Y) = H(X) - H(X|Y)$ is the information about X that Bob receives.

$H(X|Y)= \Sigma_{y \in Y}\, p(Y=y)H(X|Y=y)$

- $H(X|Y=0) = 0$
- $H(X|Y=1) = 0$
- $H(X|Y=?) = -\Sigma_{x \in X}\, p(x|Y=?) \log p(x|Y=?)$



- $p(X = 0|Y =?) =$ p (X = 0, Y =?)/p(Y=?)

  $= p(X=0)p(Y=? \mid X=0) / p(Y=?)$

  $= (2/3 \times 1/4)/(1/3) = 1/2$

- $p(X = 1|Y =?) =$ p (X = 1, Y =?)/p(Y=?)

  $= p(X=1)p(Y=? \mid X=1) / p(Y=?)$

  $= (1/3 \times 1/2)/(1/3) = 1/2$

Entropies

- $H(X|Y) = \Sigma_{y \in Y=\{0,?,1\}} p(y)H(X|Y=y)$

  $= 1/3 \times 1 = 0.33$ bits
- $I(X;Y) = H(X) - H(X|Y)$

  $= 0.9183 - 0.3333 = 0.5850$ bits

- Note that $H(X|Y=?) = 1$ bit
- $H(X|Y=?) > H(X)$ (= 0.9183 bits)

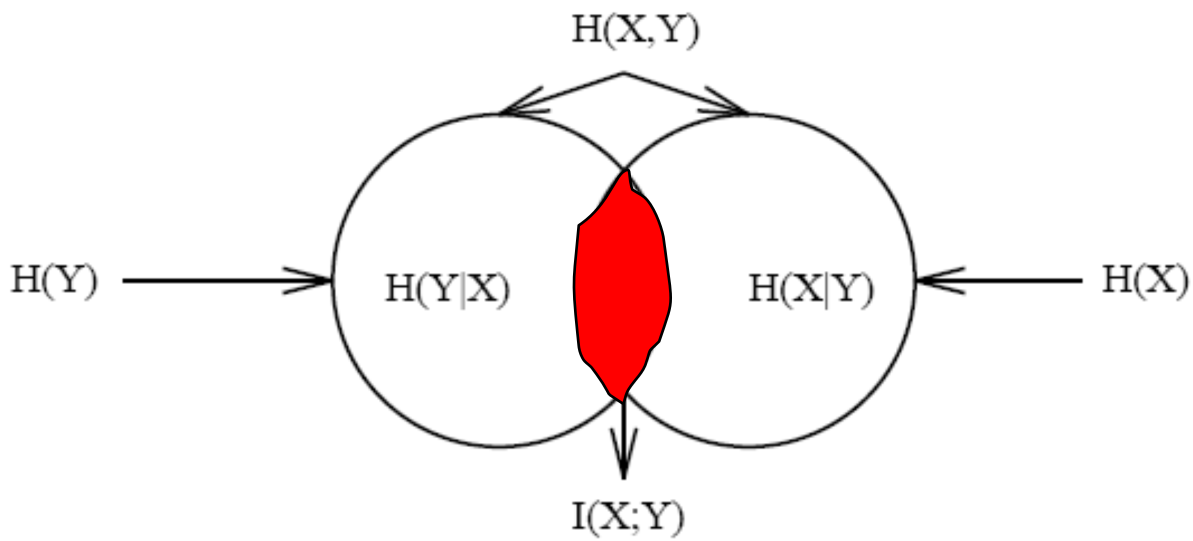
$I(X;Y)$

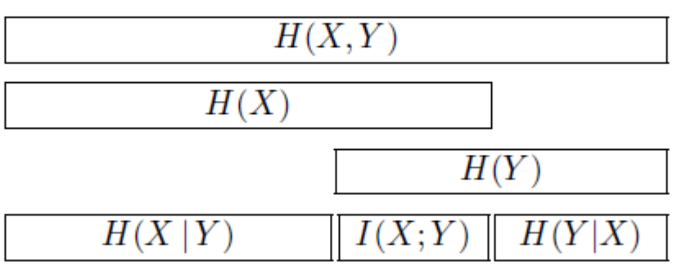
- Symmetry: The symmetry of this function implies:
  - information theory is not concerned with the cause/effect type of relation. It merely looks at statistical dependence.

- If X and Y are independent, we have
  - $H(X|Y) = H(X)$ and $I(X;Y) = 0$
  - Y does not give any information about X.

I(X;Y)

- $I(X;Y) = H(X) + H(Y) - H(X,Y)$





DM Info Theory..