$\boxed{\text{Entropy}}$

① **Entropy** $H(X) = -\sum_{x} p(x) \log p(x)$ $X \to$ Random Variable

a) $H(X) \geq 0$

b) $H(X) = 0 \Leftrightarrow$ random variable X is not 'random'.

c) Interpretation: X is more random, $H(X)$ is bigger (entropy)

$\downarrow$

achieve maximum when X is uniformly distributed

② **Jointly Entropy** $\boxed{H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)}$

a) $H(X,Y) \geq 0$

b) $H(X,Y) = 0 \Leftrightarrow (X,Y)$ is not "random"

c) (X,Y) is more random than X, in the sense that X is the marginal of (X,Y)

Therefore, in principle, $H(X,Y) \geq H(X)$ generally holds!

$H(X|Y) = \sum_{y} p(y) \cdot H(X|Y=y) = -\sum_{y} p(y) \sum_{x} P_{X|Y}(x|y) \cdot \log P_{X|Y}(x|y)$

③ **Conditional Entropy** $H(X|Y) = H(X,Y) - H(Y) \geq 0$ $= -\sum_{x,y} P_{X,Y}(x,y) \log P_{X|Y}(x|y)$

$\downarrow$

$H(X|Y)$

$= \mathbb{E}[H(X|Y)]$

$$= -\sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_Y(y)}$$

$$= -\sum_{x,y} P_{X,Y}(x,y) \log P_{X|Y}(x|y)$$

a) $H(X|Y) \geq 0$

b) measure the uncertainty of X given Y

④ **Mutual info** $I(X;Y) = I(Y;X)$

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

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

$= \cancel{-}\sum P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{\ }$

joint dist

x,y $P_X(x)\,P_Y(y)$ $\boxed{\text{independent } X, Y}$

$$= D\left(P_{X,Y}(\cdot,\cdot) \,\|\, P_{\tilde{X},\tilde{Y}}(\cdot,\cdot)\right) \quad \text{where} \quad P_{\tilde{X},\tilde{Y}}(x,y) = P_X(x)\,P_Y(y)$$

a) $I(X;Y) = I(Y;X) \geq 0$

b) $H(X|Y) \leq H(X) \leq H(X,Y)$

mutual info condifonal entropy $H(X|Y)$

⑤ $\boxed{\text{KL- divergence}}$ $D\left(P_X(\cdot) \,\|\, Q_X(\cdot)\right)$

$$= \sum_x P_X(x) \log \frac{P_X(x)}{Q_X(x)} \geq 0$$

$$I(X,Y) := D\left(P_{X,Y}(\cdot,\cdot) \,|\, Q_{X,Y}(\cdot,\cdot)\right)$$

$\boxed{Q_{X,Y} \text{ is the independently joint distribution of } X \,\&\, Y}$

Summary

$\boxed{X \rightsquigarrow (X,Y)}$

$\downarrow$

$H(X,Y) - H(X) = \boxed{H(Y|X) \longrightarrow H(Y)}$

$\downarrow$

$\boxed{H(Y) - H(Y|X)} \rightarrow I(X;Y) = I(Y;X)$

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

Interpretation:

① $H(X) \longrightarrow$ the measure of randomness of X

$= -\sum p(x) \log_2 p(x) \in [0, \log_2 K]$

② $H(X,Y) \longrightarrow$ similar to $H(X)$

③ $H(X|Y) = E_Y[f(Y)]$

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

$\downarrow$

measure the randomness of

$f(Y) = H(\tilde{X}_y)$

$\hookrightarrow$ give a $Y=y$, we can construct a r.v.

$\tilde{X}_y = X|Y=y \rightarrow \boxed{H(\tilde{X}_y)}$

$X|Y$, in the sense that average all choices of $Y=y$

→ measure the randomness of $X|Y=y$

→ average over all choices of Y

③ $I(X;Y) = H(X) - H(X|Y)$ → $\boxed{\text{mutual info}}$

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

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