ex

anna.basetti

February 2020

1a

$H((X+Y)\mid X) = H(Y\mid X)$ follows from the fact that

$$p(x+y\mid x) := \frac{P(X=x, X+Y=x+y)}{P(X=x)} = \frac{P(X=x, Y=y)}{P(X=x)} = p(y\mid x),$$

because the two conditions $(X=x, X+Y=x+y)$ and $(X=x, Y=y)$ are equivalent:

$$\begin{cases} X=x \\ Y=y \end{cases} \iff \begin{cases} X=x \\ X+Y=x+y \end{cases}.$$

This yields

$$\begin{aligned} H((X+Y)\mid X) &= -E\log p((X+Y)\mid X) \\ &= -E\log p(Y\mid X) \\ &= H(Y\mid X). \end{aligned}$$

$H(X+Y)\geq H(Y)$ for independent X, Y follows from

$$\begin{aligned} &0 \leq I(X+Y;X) \\ &\quad= H(X+Y) - H(X+Y\mid X) \\ &\quad\overset{1a}{=} H(X+Y) - H(Y\mid X) \\ &\quad\overset{ind.}{=} H(X+Y) - H(Y) \\ &\quad\iff H(X+Y)\geq H(Y). \end{aligned}$$

1b

In previous courses a statistic $T(X)$ was said to be sufficient for θ if the conditional distribution of X given T was independent of θ, meaning $p(X\mid T(X),\theta = p(X\mid T(X)$ (we will call this "def 1" when we use this to justify an equality below). Multiple applications of Bayes' rule give

$$p(\theta\mid X, T(X)) = \frac{p(X\mid\theta, T(X))p(\theta\mid T(X))}{p(X\mid T(X))} \overset{def1}{=} \frac{p(X\mid T(X))p(\theta\mid T(X))}{p(X\mid T(X))} = p(\theta\mid T(X)).$$

Also, because $T(X)$ is a function of X, we have $P(T(X) = t, X = x) = P(X = x)$ for all values of $(x, t(x))$, meaning $p(X, T(X)) = x(X)$, and analogously $p(\theta, X, T(X)) = p(\theta, X)$. This yields

$$p(\theta \mid X, T(X)) = \frac{p(\theta, X, T(X))}{p(X, T(X))} = p(\theta \mid X).$$

Since Shannon entropy only depends on the probability function, this means that

$$H(\theta \mid X, T(X)) = H(\theta \mid T(X)) = H(\theta \mid X).$$

Once we have shown this, it is very easy to show that the definition of sufficiency given in previous courses is equivalent to the information-theoretical one. Namely,

$$I(\theta; X, T(X)) = H(\theta) - H(\theta \mid X, T(X)) = H(\theta) - H(\theta \mid T(X)) = I(\theta; T(X)),$$

$$I(\theta; X, T(X)) = H(\theta) - H(\theta \mid X, T(X)) = H(\theta) - H(\theta \mid X) = I(\theta; X)$$

meaning $I(\theta; T(X)) = I(\theta; X)$, which is precisely the information-theoretical definition of sufficiency ("def 2").

We shall now show that def 1 $\Longrightarrow$ def 2. This follows from a few identities which are easily seen by drawing out the Venn-diagram visualization of entropies and informations for three variables. We have

$I(\theta; X \mid T(X)) = I(\theta; X) - I(\theta; X; T(X))$

OBS! Mutual information only defined for 2 RVs! Maybe skip this step

$= I(\theta; X) - (I(\theta; X) - H(\theta \mid X) + H(\theta \mid X, T(X)))$

$= H(\theta \mid X) - H(\theta) + I(\theta; X) + I(\theta, T(X)) - I(\theta; X, T(X)))$

Well... assuming that $I(\theta; X, T(X))$ motsvarar det jag trodde hette $I(\theta; X; T(X))$

alltså intersektionen av alla tre regioner i venn-diagrammet

$= I(\theta; T(X)) - I(\theta; X, T(X))$

$\overset{def2}{=} I(\theta; X) - I(\theta; X, T(X)).$

Again, being a function of X, $T(X)$ doesn't bring any new information in the system than X already contains. Thus we have $I(\theta; X, T(X)) = I(\theta; X)$, giving

$$I(\theta; X \mid T(X)) = I(\theta; X) - I(\theta; X) = 0$$

which can only be the case if $X \mid T(X)$ is independent of θ, i.e. def 1.

We have therefore shown that the two definitions are equivalent.