The Chain Rule

The chain rule expresses the relation between the conditional entropy and the joint/maringal entropies of the variables involved. We first state and prove the chain rule for two random variables, and then generalize it to n variables.

Proposition: Chain Rule

Let X and Y be random variables. Then

$$H(XY) = H(X) + H(Y|X)\,.$$

Proof hint

We encourage you to try to prove this for yourself. As a starting point, write out the definition of $H(XY)$, and rewrite the terms $P_{XY}(x,y)$ into a conditional form in order to relate it to $H(Y|X)$.

Show full proof

The chain rule is a matter of rewriting:

$$\begin{aligned} H(XY) &= -\sum_{x,y} P_{XY}(x,y) \log P_{XY}(x,y) \\ &= -\sum_{x,y} P_{XY}(x,y) \log\big(P_X(x) P_{Y|X}(y|x)\big) \\ &= -\sum_{x,y} P_{XY}(x,y) \log P_X(x) - \sum_{x,y} P_{XY}(x,y) \log P_{Y|X}(y|x) \\ &= -\sum_{x} P_X(x) \log P_X(x) - \sum_{x,y} P_{XY}(x,y) \log P_{Y|X}(y|x) \\ &= -\sum_{x} P_X(x) \log P_X(x) - \sum_{x} P_X(x) \sum_{y} P_{Y|X}(y|x) \log P_{Y|X}(y|x) \\ &= H(X) + H(Y|X)\,. \end{aligned}$$

This was to be shown.

The chain rule immediately results in the so-called 'independence bound':

Corollary: Subadditivity (independence bound)

$$H(XY) \le H(X) + H(Y)\,.$$

Equality holds iff X and Y are independent.

Proof

$$H(XY) = H(X) + H(Y|X) \leq H(X) + H(Y),$$

where the equality is due to the chain rule and the inequality is due to **the upper bound on the conditional entropy** that $H(Y|X) \leq H(Y)$ (and equal iff X and Y are independent).

We can naturally generalize the definition of conditional entropy by applying it to the conditional distribution $P_{XY|\mathcal{A}}$; this results in $H(X|Y, \mathcal{A})$, the entropy of X given Y and conditioned on the event $\mathcal{A}$. Since the entropy is a function of the *distribution* of a random variable, the chain rule also holds when conditioning on an event $\mathcal{A}$. Furthermore, it holds that

$$H(X|YZ) = \sum_z P_Z(z) H(X|Y, Z{=}z)\,,$$

which is straightforward to verify. With this observation, we see that the chain rule generalizes as follows.

Corollary: generalized chain rule

Let X, Y and Z be random variables. Then

$$H(XY|Z) = H(X|Z) + H(Y|XZ)\,.$$

Inductively applying the (generalized) chain rule implies that for any sequence $X_1, \ldots, X_n$ of random variables:

$$H(X_1 \cdots X_n) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_{n-1} \cdots X_1)\,.$$

Combining this with the upper bound on the conditional entropy, we see that subadditivity generalizes to

$$H(X_1 \cdots X_n) \leq \sum_{i=1}^{n} H(X_i).$$