Bounds on the Conditional Entropy

On this page, we show that the conditional entropy $H(X|Y)$ is lower bounded by zero, and upper bounded by the entropy $H(X)$. That is, *on average* any additional information (i.e., knowing Y) can only *decrease* the uncertainty about X.

Proposition

Let X and Y be random variables with respective images $\mathcal{X}$ and $\mathcal{Y}$. Then

$$0 \leq H(X|Y) \leq H(X)$$

Equality on the left-hand side holds iff X is determined by Y, i.e., for all $y \in \mathcal{Y}$, there is an $x \in \mathcal{X}$ such that $P_{X|Y}(x|y) = 1$. Equality on the right-hand side holds iff X and Y are independent.

Proof

The lower bound follows trivially from the definition and from the positivity of entropy, and so does the characterization of when $H(X|Y) = 0$.

For the upper bound, note that

$$H(X|Y) = \sum_y P_Y(y) \sum_x P_{X|Y}(x|y) \log \frac{1}{P_{X|Y}(x|y)} = \sum_{x,y} P_{XY}(x,y) \log \frac{P_Y(y)}{P_{XY}(x,y)}$$

and

$$H(X) = \sum_x P_X(x) \log \frac{1}{P_X(x)} = \sum_{x,y} P_{XY}(x,y) \log \frac{1}{P_X(x)}$$

where the last equality is derived by marginalization. In both expressions, we may restrict the sum to those pairs (x,y) with $P_{XY}(x,y) > 0$.

Using Jensen's inequality, it follows that

$$H(X|Y) - H(X) = \sum_{x,y:P_{XY}(x,y)>0} P_{XY}(x,y) \log \frac{P_X(x)P_Y(y)}{P_{XY}(x,y)}$$

$$\leq \log\left(\sum_{x,y:P_{XY}(x,y)>0} P_X(x)P_Y(y) \right)$$

$$\leq \log\left(\sum_{x,y} P_X(x)P_Y(y) \right)$$

$$= \log\left(\left(\sum_{x\in\mathcal{X}} P_X(x)\right)\left(\sum_{y\in\mathcal{Y}} P_Y(y)\right) \right)$$

$$= \log 1 = 0.$$

The second inequality follows by the monotonicity of the logarithm function.

It remains to show that these inequalities are equalities if and only if X and Y are independent. Try it yourself first; you can use the equality condition of Jensen's inequality to characterize the first inequality in the derivation above.

Show solution

We start with the if-direction. Suppose that X and Y are independent, i.e., $P_X(x)P_Y(y) = P_{XY}(x,y)$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. Then for all $(x,y),(x',y') \in \mathcal{X} \times \mathcal{Y}$,

$$\frac{P_X(x)P_Y(y)}{P_{XY}(x,y)} = 1 = \frac{P_X(x')P_Y(y')}{P_{XY}(x',y')}.$$

By the equality condition in Jensen's inequality, the first inequality in the derivation above is an equality. The second inequality in the derivation above is easily seen to be equality as well: the second term is bigger only because we add those summands $P_X(x)P_Y(y)$ for which $P_{XY}(x,y) = 0$, but for those x and y, $P_X(x)P_Y(y) = 0$ as well. For the only-if-direction, suppose that $H(X|Y) = H(X)$, that is, both inequalities in the above derivation are equalities. Then (from the fact that the second inequality is an equality), we know that whenever $P_{XY}(x,y) = 0$, also $P_X(x)P_Y(y) = 0$. When $P_{XY}(x,y) > 0$, we know from the equality condition in Jensen's inequality that for all $y' \in \mathcal{Y}$ for which $P_{XY}(x,y') > 0$,

$$\frac{P_X(x)P_Y(y)}{P_{XY}(x,y)} = \frac{P_X(x)P_Y(y')}{P_{XY}(x,y')}.$$

Working out the term $P_{XY}(x,y)$ as $P_{X|Y}(x|y)P_Y(y)$ (and similarly for $P_{XY}(x,y')$), and cancelling/rearranging terms, we get that

$$P_{X|Y}(x|y) = P_{X|Y}(x|y')$$

for all y' for which $P_{XY}(x,y') > 0$. From this, we may conclude that

$$P_X(x) = \sum_{y''\in\mathcal{Y}} P_Y(y'')P_{X|Y}(x|y'') = \sum_{y''\in\mathcal{Y}} P_Y(y'')P_{X|Y}(x|y) = P_{X|Y}(x|y),$$

where the second inequality follows from the fact that $P_{X|Y}(x|y) = P_{X|Y}(x|y'')$ for all x, y, y''. Finally we conclude that

$P_{XY}(x,y) = P_{X|Y}(x|y)P_Y(y) = P_X(x)P_Y(y)$. This equality now holds for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, so X and Y are independent.