Properties of Shannon Entropy

In this section, we list a few properties of the Shannon entropy, and show a trick to compute it more easily by hand.

Proposition: Positivity of entropy

Let X be a random variable with image $\mathcal{X}$. Then

$$0 \leq H(X).$$

Equality holds iff there exists $x \in \mathcal{X}$ with $P_X(x) = 1$ (and thus $P_X(x') = 0$ for all $x' \neq x$).

Proof

For all $x \in \{X\}$, we have $0 \leq P_X(x) \leq 1$, and hence $-P_X(x) \log P_X(x) \geq 0$. So $H(X)$, which is the sum of those terms, is always nonnegative. To characterize the condition for equality, note that by definition of Shannon entropy, $H(X) = 0$ when $P_X(x) = 1$ for some x. On the other hand, if $H(X) = 0$ then for any x with $P_X(x) > 0$ it must be that $\log(1/P_X(x)) = 0$ and hence $P_X(x) = 1$.

Proposition: Upper bound on entropy

Let X be a random variable with image $\mathcal{X}$. Then

$$H(X) \leq \log(|\mathcal{X}|).$$

Equality holds iff $P_X(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$.

Proof hint

We encourage you to try to find the proof for this proposition yourself. As a first step, you may want to write out the definition of $H(X)$ apply Jensen's inequality.

Show full proof

The function $f : \mathbb{R}_{>0} \to \mathbb{R}$ defined by $y \mapsto \log y$ is strictly concave on $\mathbb{R}_{>0}$. Thus, by Jensen's inequality:

$$H(X) = \sum_{x \in \mathcal{X}} P_X(x) \cdot \log \frac{1}{P_X(x)} \leq \log\Big(\sum_{x \in \mathcal{X}} P_X(x) \cdot \frac{1}{P_X(x)}\Big) = \log\Big(\sum_{x \in \mathcal{X}} 1\Big) = \log(|\mathcal{X}|).$$

Furthermore, since we may restrict the sum to all x with $P_X(x) > 0$, equality holds if and only if $\log(1/P_X(x)) = \log(1/P_X(x'))$, and thus $P_X(x) = P_X(x')$, for all $x, x' \in \mathcal{X}$.

When working with explicit distributions, one can always compute the entropy of a random variable by filling in all the probabilities in the definition of entropy. However, in some cases there is some structure to the distribution. In those cases, the entropy can be computed in a smarter and faster way. This is especially useful when you are computing the entropy by hand, but can also help when analysing the entropy of a more complex distribution containing some unknown variables.

Video-2018-06-27-12-10-50_Smart ways to compute entropy.MP4

In general, the entropy of a random variable with probabilities $p_1, \ldots, p_n$ can be expressed as

$$H(p_1, \ldots, p_k, p_{k+1}, \ldots, p_n) = h\left(\sum_{i=1}^{k} p_i\right) + \left(\sum_{i=1}^{k} p_i\right) \cdot H\left(\frac{p_1}{\sum_{i=1}^{k} p_i}, \ldots, \frac{p_k}{\sum_{i=1}^{k} p_i}\right) + \left(\sum_{i=k+1}^{n} p_i\right)$$

You can of course use this trick multiple times in a row to break down the entropies on the right-hand side of this equation even further.

Exercise

Consider a random variable X with $\mathcal{X} = a, b, c$ and $P_X(a) = \frac{1}{2}$, $P_X(b) = P_X(c) = \frac{1}{4}$. Compute the entropy of X using the techniques shown in the video.

Show solution

We can think of this distribution as the result of two fair coin tosses: if the first coin comes out heads, the outcome is a. If it comes out tails, we toss another fair coin to determine whether the outcome is b or c.

An appropriate underlying probability space (Ω, P) could be $\Omega = \mathsf{hh}, \mathsf{ht}, \mathsf{th}, \mathsf{tt}$ and $P(\omega) = \frac{1}{4}$ for all $\omega \in \Omega$. Then we define the function $X : \Omega \to \mathcal{X}$ as

$$X(\mathsf{hh}) = X(\mathsf{ht}) = a, \qquad X(\mathsf{th}) = b, \qquad X(\mathsf{tt}) = c.$$

This yields the correct distribution P_X.

The following computation now leads to the entropy of X:

$$H(X) = h\left(\frac{1}{2}\right) + \frac{1}{2}h(0) + \frac{1}{2}h\left(\frac{1}{2}\right) = \frac{3}{2}.$$

The first coin toss determines whether the outcome is a (on heads h) or something else (on tails t). On heads, the second coin toss does not give any more information, whereas on tails, the second coin toss still decides between outcome b and outcome c.

created: 2018-12-12