Definition: Shannon Entropy

In this section, we explore a measure for the amount of uncertainty of random variables. Watch the video below for an introduction to this topic:

Consider some probabilistic event $\mathcal{A}$ that occurs with probability $P[\mathcal{A}]$ for some probability measure P. The **surprisal value** $\log \frac{1}{P[\mathcal{A}]}$ indicates how surprised we should be when the event $\mathcal{A}$ occurs: events with small probabilities yield high surprisal values, and vice versa. An event that occurs with certainty ($P_X(\mathcal{A}) = 1$) yields a surprisal value of 0.

For a random variable X, we consider the *expected* surprisal value to be an indicator of how much uncertainty is contained in the variable, or how much information is gained by revealing the outcome. This expected surprisal value is more commonly known as the (Shannon) entropy of a random variable:

Definition: Entropy

Let X be a random variable with image $\mathcal{X}$. The (Shannon) entropy $H(X)$ of X is defined as

$$H(X) := \sum_{x \in \mathcal{X}} P_X(x) \cdot \log \frac{1}{P_X(x)} = -\sum_{x \in \mathcal{X}} P_X(x) \cdot \log P_X(x)\,,$$

with the convention that the $\log$ function represents the *binary* logarithm $\log_2$.

There are three things to note about this definition:

- The entropy of X is a function (solely) of the *distribution* P_X of X, and the elements of $\mathcal{X}$ (i.e., the values that X can take on) are completely irrelevant for the entropy. Therefore, it would be formally more correct to write $H(P_X)$. However, it is customary (and more convenient) to write $H(X)$ whenever there is no ambiguity about the underlying distribution.
- The summand $P_X(x) \cdot \log P_X(x)$ is technically undefined whenever $P_X(x) = 0$. As a convention, we set $P_X(x) \cdot \log P_X(x) = 0$ in this case. This choice is justified by taking the limit where $P_X(x)$ goes to 0 (see the exercise below).
- The entropy of X can also be expressed as the expectation of the random variable $\log\big(1/P_X(X)\big)$:

$$H(X) = \mathbb{E}_X\left[\log \frac{1}{P_X(X)}\right].$$

Exercise

Prove that $\lim_{p\to 0^+} p\log(p) = 0$. This is a justification for the convention that we set $P_X(x)\cdot\log P_X(x) = 0$ whenever $P_X(x) = 0$.

Show solution

First, note that

$$\lim_{p\to 0^+} \log p = -\infty$$

and

$$\lim_{p\to 0^+} \frac{1}{p} = \infty.$$

So, by l'Hopital's rule,

$$\lim_{p\to 0^+} p\log p = \lim_{p\to 0^+} \frac{\log p}{1/p} = \lim_{p\to 0^+} \frac{[\log p]'}{[1/p]'} = \lim_{p\to 0^+} -\ln(2)\frac{1/p}{1/p^2} = \lim_{p\to 0^+} -\ln(2)p = 0.$$

Example

Consider a random variable X with $\mathcal{X} = a, b, c, d$ and $P_X(a) = \frac{1}{2}$, $P_X(b) = P_X(c) = \frac{1}{4}$, $P_X(d) = 0$. The entropy of X is

$$H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 + 0 = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 0 = \frac{3}{2}.$$