Entropy Rate

Intuitively, some of the stochastic processes we have seen in the previous sections are more predictable than others. The periodic Markov process from **Example 3** is not so surprising anymore as soon as the first b is observed. In this section, we will study a measure for the unpredictability of a stochastic process: the entropy rate.

Definition: Entropy rate

The entropy rate $H(\{X_i\})$ of a stochastic process $\{X_i\}$ is

$$H(\{X_i\}) := \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n),$$

if the limit exists, and undefined otherwise.

In the literature, the entropy rate is often denoted $H(\mathcal{X})$, referring to the common support of the variables in the stochastic process. The notation $H(X)$ is also sometimes used, but this can be ambiguous and confusing. The entropy rate reflects the way in which the entropy of the sequence (observed so far) grows as n grows large.

Example

Consider a process $\{X_i\}$ where the X_i are i.i.d. sampled from P_X. Then

$$\begin{aligned} H(\{X_i\}) &= \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots X_n) \\ &= \lim_{n\to\infty} \frac{1}{n} (H(X_1) + H(X_2) + \ldots + H(X_n)) \\ &= \lim_{n\to\infty} \frac{n}{n} H(X) \\ &= H(X). \end{aligned}$$

So, every new coin toss increases the entropy of the entire observed sequence by $H(X)$.

We can also define an alternative measure of the unpredictability of a stochastic process, where we focus not on the amount of entropy in the entire sequence observed so far, but on the amount of entropy present in the current random variable, given the past sequence.

Definition: Entropy rate given the past

The entropy rate given the past $H'(\{X_i\})$ of a stochastic process $\{X_i\}$ is

$$H'(\{X_i\}) := \lim_{n\to\infty} H(X_n | X_1, \ldots, X_{n-1}),$$

if the limit exists, and undefined otherwise.

For all stationary processes, this alternative definition turns out to coincide with the original definition of entropy rate. In order to show this, we need an analytic statement about the convergence of sums.

Theorem: Cesàro mean

If $\lim\limits_{n\to\infty} a_n = a$ and $b_n = \frac{1}{n} \sum_{i=1}^{n} a_i$, then $\lim\limits_{n\to\infty} b_n = a$.

Proof

05 Cesàro mean.mp4

Theorem

For a stationary process $\{X_i\}$, it holds that $H(\{X_i\}) = H'(\{X_i\})$ (and both limits exist).

Proof

We first show that $H(X_n \mid X_1, \dots, X_{n-1})$ is a non-increasing function of n:

$$\begin{aligned} H(X_n|X_1,\dots,X_{n-1}) &= H(X_{n+1}|X_2,\dots,X_n) && \text{(stationary)} \\ &\geq H(X_{n+1}|X_1,X_2,\dots,X_n) && \text{(Bounds on the Conditional Entropy)}. \end{aligned}$$

Combined with the fact that $H(X_n \mid X_1, \dots, X_{n-1})$ is lower bounded by 0, this implies that the limit $\lim_{n\to\infty} H(X_n \mid X_1, \dots, X_{n-1})$ must exist. It is $H'(\{X_i\})$. It remains to show that $H(\{X_i\}) = H'(\{X_i\})$:

$$\begin{aligned} H(\{X_i\}) &= \lim_{n\to\infty} \frac{1}{n} H(X_1,\dots,X_n) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i \mid X_1,\dots,X_{i-1}) \\ &= H'(\{X_i\}). \end{aligned}$$

The final equality follows from the Cesaro mean.

Example 2: A finite-state time-invariant Markov process, continued

For a Markov process with transition matrix

$$R = \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix},$$

we have **previously computed** the stationary distribution to be $(\mu_a, \mu_b) = (5/8, 3/8)$. Hence, if we start in this stationary distribution, the entropy rate for this time-invariant stationary Markov process is

$$H(\{X_i\}) = H'(\{X_i\}) = \lim_{n\to\infty} H(X_n|X^{n-1}) = H(X_2|X_1),$$

where P_{X_1} is the stationary distribution, i.e. $P_{X_1}(\mathtt{a}) = 5/8, P_{X_1}(\mathtt{b}) = 3/8$. Therefore,

$$\begin{aligned} H(\{X_i\}) = H(X_2|X_1) &= P_{X_1}(\mathtt{a}) \cdot H(X_2|X_1 = \mathtt{a}) + P_{X_1}(\mathtt{b}) \cdot H(X_2|X_1 = \mathtt{b}) \\ &= 5/8 \cdot h(0.3) + 3/8 \cdot h(0.5) \\ &\approx 0.926\,. \end{aligned}$$

Note that the entropy rate $H(\{X_i\}$ is not equal to the entropy of the stationary distribution (which is $h(5/8) \approx 0.792$ in this case)!