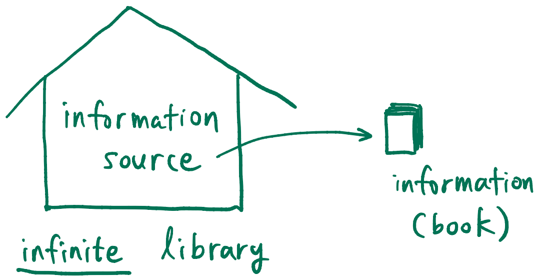
1-2 Shannon Entropy

Shannon theory

- noiseless channel coding theorem – data compression
- noisy channel coding theorem – channel capacity

You'll learn why entropy is defined so.

Model



why infinite?

set of alphabets $\chi = \{a, b, c, \cdots\}$

random variable $X = x$ with probability $p(x)$, $x \in \chi$

information source → a sequence of random variables of length n

$X^n \equiv X_1 X_2 \cdots X_n$ (n: length)

X_i's are i.i.d. (independent and identically distributed)

$X^n = x^n \equiv x_1 x_2 \cdots x_n$ with probability $p(x^n)$

ex) $\chi = \{a, b\}$, $p(a) = 2/3$, $p(b) = 1/3$, $n = 2$

Result

As $n \to \infty$, you can encode the message into $\boxed{nH(X)}$ bits.

$$\boxed{H(X) \equiv -\sum_x p(x) \log p(x)}$$ Shannon entropy

ex) $\chi = \{a, b, c, d\}$

$p(a) = 1/2$, $p(b) = 1/8$, $p(c) = 1/4$, $p(d) = 1/8$

method 1: $a \to 00$, $b \to 01$, $c \to 10$, $d \to 11$

You need $2n$ bits.

method 2: $a \to 0$, $c \to 10$, $b \to 110$, $d \to 111$

$abcd \to 011010111$ (there is no ambiguity!)

\# of bits per character $= \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 \cdot 2 = \frac{7}{4}$

You need $\frac{7}{4} n$ bits.

Sketch of the proof

$x^n = x_1 x_2 \cdots x_n, \quad n \to \infty$

$x_i \in \{c_1, c_2, ..., c_\alpha, ...\}, \quad p_\alpha \equiv p(c_\alpha)$

typical vs atypical sequences

typical x^n: # of $c_1 \approx np_1$, # of $c_2 \approx np_2$, ···

ex) coin tossing

asymptotic equipartition theorem

Almost all x^ns are typical! (as $n \to \infty$)

$p(\text{typical } x^n) \approx p_1^{np_1} p_2^{np_2} \cdots$ (all identical)

# of typical sequences $\approx \dfrac{1}{p(\text{typical } x^n)}$

You encode only typical sequences into nR bits. $\boxed{R: \text{ compression rate}}$

$p(\text{error}) \to 0$ as $n \to \infty$

$$
\begin{aligned}
nR &\approx \log(\text{\# of typical sequences}) \\
&= -\log\left(p_1^{np_1} p_2^{np_2} \cdots\right) \\
&\approx -np_1 \log p_1 - np_2 \log p_2 - \cdots \\
&= n\left(-\sum_\alpha p_\alpha \log p_\alpha\right) \\
&= nH(X)
\end{aligned}
$$

Q) Is this optimal? converse theorem