Theorem: Shannon's Source-Coding Theorem (Optimal Codes)

We now know that prefix-free codes can achieve the same minimal code lengths for a source P_X as the more general class of uniquely decodable codes. How small is this minimal code length in general? In this section we explore the following relation between the minimal code length and the entropy of the source:

Theorem: Shannon's source-coding theorem (for symbol codes)

For any source P_X, we have the following bounds:

$$H(X) \leq \ell_{\min}(P_X) \leq H(X) + 1.$$

$H(X) \leq \ell_{\min}(P_X)$

The proof relies on Kraft's inequality. Let C be a code, and write ℓ_x for $\ell(C(x))$ as a notational convenience. For the lower bound, we have that

$$\begin{aligned} H(X) - \ell_C(P_X) &= -\sum_{x \in \mathcal{X}} P_X(x) \log(P_X(x)) - \sum_{x \in X} P_X(x) \ell_x \\ &= \sum_{x \in \mathcal{X}} P_X(x) \left(-\log(P_X(x)) - \log\left(2^{\ell_x}\right)\right) \\ &= \sum_{x \in \mathcal{X}} P_X(x) \log\left(\frac{1}{P_X(x) \cdot 2^{\ell_x}}\right) \\ &\leq \log\left(\sum_{x \in \mathcal{X}} \frac{1}{2^{\ell_x}}\right) && \text{(by Jensen's inequality)} \\ &\leq \log(1) = 0 && \text{(by Kraft's inequality)} \end{aligned}$$

$\ell_{\min}(P_X) \leq H(X) + 1$

For the upper bound, let us denote by ℓ_x the surprisal value in bits rounded up to the next integer, i.e. for any $x \in \mathcal{X}$,

$$\ell_x := \left\lceil \log \frac{1}{P_X(x)} \right\rceil,$$

and note that

$$\sum_{x \in \mathcal{X}} 2^{-\ell_x} \leq \sum_{x \in \mathcal{X}} 2^{-\log \frac{1}{P_X(x)}} = \sum_{x \in \mathcal{X}} P_X(x) = 1.$$

Therefore, by Kraft's inequality, there exists a prefix-free code C such that $\ell(C(x)) = \ell_x$ for all $x \in \mathcal{X}$. This code satisfies

$$\begin{aligned}
\ell_C(P_X) &= \sum_{x \in \mathcal{X}} P_X(x) \ell_x \\
&\leq \sum_{x \in \mathcal{X}} P_X(x) \left(\log \frac{1}{P_X(x)} + 1 \right) \\
&= -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x) + \sum_{x \in \mathcal{X}} P_X(x) \\
&= H(X) + 1.
\end{aligned}$$

We have thus constructed a code C with $\ell_C(P_X) \leq H(X) + 1$, so $\ell_{\min}(P_X) \leq H(X) + 1$.