Math 5251 Entropy and Shannon's Noiseless Coding Theorem (§3.3)

Given a source $W = \{w_1, \ldots, w_m\}$ with probabilities $p_1, \ldots, p_m$

the entropy $H(W) := -\sum_{i=1}^{m} p_i \log_2(p_i)$

gives a surprisingly precise upper and lower bound on the minimum possible value of

$$\text{avglength}(f) = \sum_{i=1}^{m} p_i \, \ell(f(w_i))$$

for all uniquely decipherable n-ary encodings

$$f : W \longrightarrow \mathcal{C} \subset \Sigma^* \text{ with } n = |\Sigma| :$$

THEOREM (Shannon 1948) The above minimum satisfies

$$\frac{H(W)}{\log_2(n)} \leq \text{avg length}(f) < 1 + \frac{H(W)}{\log_2(n)}$$

To prove the lower bound, need a basic inequality:

LEMMA: Given probabilities $p_1, \dots, p_m$
(so $p_i \in [0,1]$, $p_1 + \dots + p_m = 1$)
and real numbers $q_1, \dots, q_m \geq 0$ with
$q_1 + \dots + q_m \leq 1$ (so maybe not probabilities!),
one has $\sum_{i=1}^{m} p_i \log_2\left(\frac{1}{p_i}\right) \leq \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{q_i}\right)$.

This will follow easily from some calculus in a bit.
But first let's see how to use it.

COROLLARY 1 (of LEMMA): Among all sample spaces $\Omega = \{\omega_1, \dots, \omega_m\}$ (prob $p_1, \dots, p_m$) of size m, uniform distribution has highest entropy:

$$H(\Omega) \leq H\left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right)$$

Proof:

$$H(\Omega) = \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{p_i}\right) \leq \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{1/m}\right)$$

$$= \log_2(m) \cdot \sum_{i=1}^{m} p_i = \log_2(m)$$

$$= H\left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right) \blacksquare$$

Take $q_i := \frac{1}{m}\ \forall i$ in LEMMA

COROLLARY 2: (of LEMMA) For any u.d. n-ary encoding $W \xrightarrow{f} \mathcal{C}$,

one has $$\frac{H(W)}{\log_2(n)} \leq \text{avglength}(f)$$

Shannon's lower bound

proof: Given the u.d. n-ary encoding $W \xrightarrow{f} \mathcal{C}$ with codeword lengths $l_i := l(f(w_i))$,

we know from McMillan that $\sum_{i=1}^{m} \frac{1}{n^{l_i}} \leq 1$.

Hence if we take $q_i := \frac{1}{n^{l_i}}$ then $q_1 + \ldots + q_m \leq 1$,

and we can apply the LEMMA to conclude

$$H(W) = \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{p_i}\right) \leq \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{q_i}\right)$$

$$= \sum_{i=1}^{m} p_i \log_2\left(n^{l_i}\right)$$

$$= \sum_{i=1}^{m} p_i l_i \log_2(n)$$

$$= \log_2(n)\, \text{avglength}(f)$$

i.e. $$\frac{H(W)}{\log_2(n)} \leq \text{avglength}(f) \quad \blacksquare$$

The upperbound in Shannon's Theorem says
$\exists$ an n-ary u.d. encoding $W \xrightarrow{f} \mathcal{C} \subset \Sigma^*$
with $\text{avglength}(f) < 1 + \frac{H(W)}{\log_2(n)}$.

---

proof of upper bound:

Pick positive integers $l_1, \dots, l_m$ uniquely via
$l_i \in [\alpha_i, 1+\alpha_i)$ where $\alpha_i := \log_n\left(\frac{1}{p_i}\right)$ (>0)

i.e. $\log_n\left(\frac{1}{p_i}\right) \leq l_i \overset{(*)}{<} 1 + \log_n\left(\frac{1}{p_i}\right)$ for $i = 1, 2, \dots, m$.

$\Downarrow$

Then $\frac{1}{p_i} \leq n^{l_i}$

$\Downarrow$

$p_i \geq \frac{1}{n^{l_i}} \Rightarrow 1 = \sum_{i=1}^{m} p_i \geq \sum_{i=1}^{m} \frac{1}{n^{l_i}}$

$\overset{\text{Kraft}}{\Longrightarrow}$ $\exists$ a u.d. encoding $W \xrightarrow{f} \mathcal{C}$
with codelengths $l_i = l(f(w_i))$

use (*)

But then

$$\text{avglength}(f) = \sum_{i=1}^{m} p_i l_i < \sum_{i=1}^{m} p_i \left(1 + \log_n\left(\frac{1}{p_i}\right)\right)$$
$$= \sum_{i=1}^{n} p_i + \sum_{i=1}^{m} p_i \log_n\left(\frac{1}{p_i}\right)$$
$$= 1 + \frac{H(W)}{\log_2(n)} \quad \blacksquare$$

use $\log_n(x) = \frac{\log_2(x)}{\log_2(n)}$

Let's return to prove ...

LEMMA: For $p_1, \ldots, p_m$ probabilities and $q_1, \ldots, q_m \geq 0$ with $q_1 + \ldots + q_m \leq 1$,

$$\sum_{i=1}^{m} p_i \log_2\left(\frac{1}{p_i}\right) \leq \sum p_i \log_2\left(\frac{1}{q_i}\right).$$

proof: Want to show

$$\sum_{i=1}^{m} p_i \log_2\left(\frac{1}{p_i}\right) - \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{q_i}\right) \overset{?}{\leq} 0$$

$$= \sum_{i=1}^{m} p_i \left(\log_2\left(\frac{1}{p_i}\right) - \log_2\left(\frac{1}{q_i}\right)\right)$$

$$= \sum_{i=1}^{m} p_i \log_2\left(\frac{1/p_i}{1/q_i}\right)$$

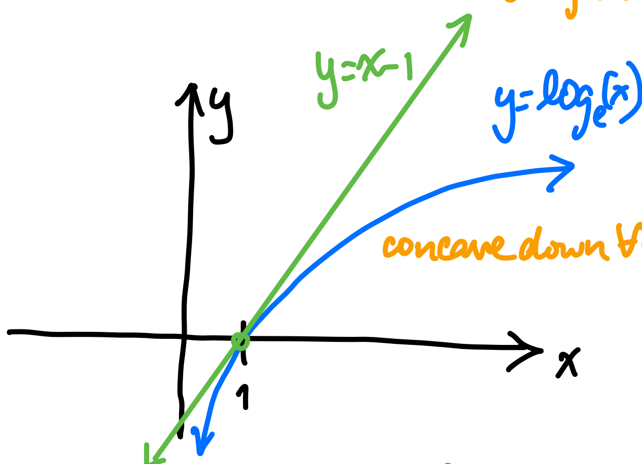
$$= \sum_{i=1}^{m} p_i \log_2\left(\frac{q_i}{p_i}\right)$$

So we want $\sum_{i=1}^{m} p_i \log_2\left(\frac{q_i}{p_i}\right) \overset{?}{\leq} 0$

some positive reals $x = \frac{q_i}{p_i} > 0$

We claim $\log_2(x) \overset{(**)}{\leq} \frac{x-1}{\log_e(2)} \quad \forall x > 0$

or equivalently $\log_e(x) \leq x - 1$:



tangent line has slope $\left[\frac{d}{dx}\log_e(x)\right]_{x=1} = \left[\frac{1}{x}\right]_{x=1} = 1$

concave down $\forall x > 0$ since $\frac{d^2}{dx^2}\log_e(x) = \frac{-1}{x^2}$

Hence $\sum_{i=1}^{m} p_i \log_2\left(\frac{q_i}{p_i}\right) \underset{\text{use }(**)}{\leq} \sum_{i=1}^{m} p_i \left(\frac{q_i}{p_i} - 1\right) \Big/ \log_e(2)$

$$= \frac{1}{\log_e(2)}\left(\sum_{i=1}^{m} q_i - \sum_{i=1}^{m} p_i\right) = \frac{1}{\log_e(2)}\left(\sum_{i=1}^{m} q_i - 1\right) \leq 0$$

since $\sum_{i=1}^{m} q_i \leq 1$ by our hypotheses $\blacksquare$