Entropy is a lower bound on average code length

Last lecture, we derived an upper bound on average code length of a Huffman code. Today we derive the lower bound.

Before doing so, we note that if $p(A_i)$ is a power of 2 for all $i = 1 \ldots N$, then the average length of the Huffman code is less than or equal to the entropy. The reason is that

$$\lceil \log(\frac{1}{p(A_i)}) \rceil = \log(\frac{1}{p(A_i)})$$

and so

$$\lambda_{Huff} \;\leq\; \sum_{i=1}^{N} p(A_i)\ \lceil \log(\frac{1}{p(A_i)}) \rceil \;=\; \sum_{i=1}^{N} p(A_i)\ \log(\frac{1}{p(A_i)}) \;=\; H$$

The inequality in the previous line was proven last lecture.

You might next ask whether there is a situation in which the average codelength of a Huffman code is strictly less than the entropy. The answer is no.

Theorem 5.1 *The average code length of a prefix code is greater than or equal to the entropy H.*

Proof Take any prefix code. Let λ_i be the codeword lengths. We show that $H \leq \overline{\lambda}$.

$$\begin{aligned} H - \overline{\lambda} &= \sum_{i=1}^{N} (\log(\frac{1}{p(A_i)}) - \lambda_i) p(A_i) \\ &= \sum_{i=1}^{N} (\log\ (\frac{2^{-\lambda_i}}{p(A_i)})\ p(A_i)) \end{aligned}$$

We apply Jensen's inequality (see below) where $a_i = \frac{2^{-\lambda_i}}{p(A_i)}, \quad p(A_i) = p_i$

$$\begin{aligned} &\leq \log\ (\ \sum_{i=1}^{N} \frac{2^{-\lambda_i}}{p(A_i)} p(A_i)\ ) \\ &= \log\ (\ \sum_{i=1}^{N} 2^{-\lambda_i}\ ) \\ &\leq \log 1, \quad \text{by Kraft inequality} \\ &= 0 \qquad \square \end{aligned}$$

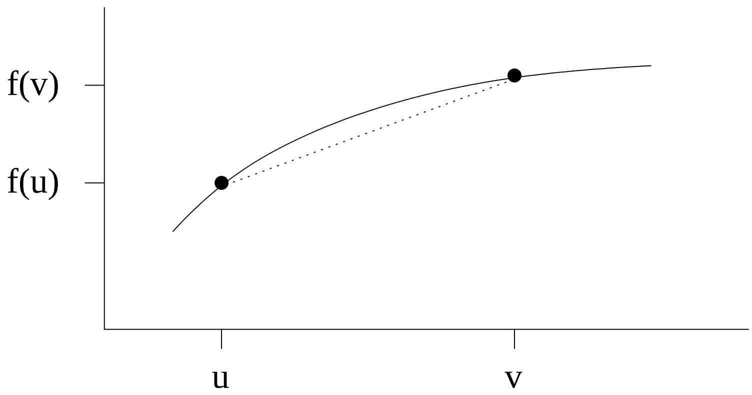
The following result is normally stated for any real valued convex function $f(x)$, that is, any function for which

$$(1-t)f(u) + tf(v) \;\leq\; f(\ (1-t)\ u \;+\; t\ v\ )\ .$$

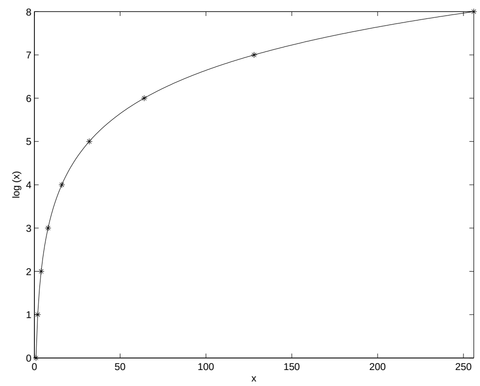
whenever $0 \leq t \leq 1$ and $u, v \in \Re$. I will state the result for the special case $f(x) = \log x$, since that is all we will need for COMP 423.

Theorem 5.2 (Jenson's Inequality) *Suppose we have a set of positive numbers $\{a_1, a_2, \ldots a_N\}$ and corresponding probabilities $p_1,\ p_2, \ldots, p_N$. Then,*

$$\sum_{i=1}^{N} p_i\ \log(a_i) \;\leq\; \log(\ \sum_{i=1}^{N} p_i\ a_i)$$



convex function



$\log x$

Proof We prove it by induction.

The theorem is true for $N = 2$, since $\log(x)$ is a convex function, namely

$$p_0 \log a_1 + (1 - p_0) \log a_2 \leq \log(p_0 a_1 + (1 - p_0) a_2)$$

[See Appendix to these notes.]

We assume the theorem is true for $N = K$ and prove for $N = K + 1$.

The trick is to define a new probability function, p'_i on the first K values $a_1, \ldots, a_K$ only, i.e. for $i \in 1, \ldots, K$, we define

$$p'_i \equiv \frac{p_i}{1 - p_{K+1}}$$

$$\sum_{i=1}^{K+1} \log(a_i)\ p_i \quad = \quad \sum_{i=1}^{K} \log(a_i)\ p_i \ + \ \log(a_{K+1})\ p_{K+1}$$

We wish to apply the induction step, using the p'_i probabilities. Multiplying and diving by $(1 - p_{K+1})$ gives

$$\sum_{i=1}^{K+1} \log(a_i)\ p_i \quad = \quad (1 - p_{K+1}) \sum_{i=1}^{K} \log(a_i) \frac{p_i}{1 - p_{K+1}} + \log(a_{K+1}) p_{K+1}$$

and applying induction hypothesis gives

$$\leq \quad (1 - p_{K+1})\ \log(\sum_{i=1}^{K} a_i\ p'_i\ ) \ + \ p_{K+1}\ \log(a_{K+1})$$

This has the form $(1 - t) \log(u) + t \log(v)$ where $0 \leq t \leq 1$, so we can apply the $N = 2$ case (see Appendix) to get:

$$\sum_{i=1}^{K+1} \log(a_i)\ p_i \ \leq \ \log(\sum_{i=1}^{K+1} a_i\ p_i)$$

which is what we wanted to prove. □

Claim 5.1 *Given an alphabet of N symbols, the uniform probability $p(A_i) = \frac{1}{N}$ yields maximum entropy.*

Proof We first apply Jensen's inequality to the definition of entropy.

$$H = \sum_{i=1}^{N} p(A_i) \log \frac{1}{p(A_i)} \leq \log(\sum_{i=1}^{N} \frac{p(A_i)}{p(A_i)}) \;=\; \log N$$

Now notice that this upper bound on H is achieved by the uniform probability function. i.e.

$$H = \sum_{i=1}^{N} \frac{1}{N} \log N = \log N$$

Example

Use Jensen's inequality to derive an upper bound on

$$\sum_{i=1}^{N} \log \log i \ .$$

Solution: Multiplying by $\frac{N}{N} = 1$,

$$\begin{aligned}
\sum_{i=1}^{N} \log \log i &= \frac{N}{N} \sum_{i=1}^{N} \log \log i \\
&= N \sum_{i=1}^{N} \frac{1}{N} \log(\log i) \\
&\leq N \log(\sum_{i=1}^{N} \frac{1}{N} (\log i)) \\
&\leq N \log(\log(\sum_{i=1}^{N} \frac{i}{N}) \\
&= N \log(\log(\frac{N(N+1)}{2N}) \\
&= N \log(\log(\frac{N+1}{2})) \\
&= N \log(\log(N+1) - 1)
\end{aligned}$$

Appendix (not covered in class – you are NOT responsible for this.)

For completely, I want to show the base case of Jensen's inequality, namely:

$$p_0 \log a_1 + (1-p_0)\log a_2 \le \log(p_0 a_1 + (1-p_0)a_2)$$

where $0 \le p_0 \le 1$. First, note that

$$\ln x = \ln(2^{\log x}) \;= \log x \;\ln 2.$$

Taking the derivative, we get

$$\frac{d \log x}{dx} \;=\; \frac{1}{\ln 2}\frac{d\;\ln x}{dx} \;=\; \frac{1}{x \ln 2}$$

and taking the derivative again, we get

$$\frac{d^2 \log x}{dx^2} \;=\; -\frac{1}{x^2 \ln 2} < 0.$$

The second derivative is always negative. (Note: $\log x$ is only defined on $x > 0$.)

Rather than continuing with $\log x$, let's prove the result for any function $f(x)$ which has the property: $f''(x) < 0$ for some range of x. That is, I want to prove that

$$p_0 f(a_1) + (1-p_0) f(a_2) \le f(p_0 a_1 + (1-p_0)a_2) \tag{1}$$

Here goes: Take a Taylor expansion of $f(x)$ about some point x_0.

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \ldots$$

Treat x and x_0 as fixed. Define the function $g(u) = f'(u)$. It is easy to show, using the intermediate value theorem of Calculus, that there exists some x^* between x_0 and x such that

$$g(x) = g(x_0) + g'(x^*)(x-x_0)$$

namely there must be some x^* such that $g'(x^*) = \frac{g(x)-g(x_0)}{x-x_0}$. Integrating $g(u) = f'(u)$ from $u = x_0$ to $u = x$, we get

$$f(x) - f(x_0) = f'(x_0)(x-x_0) + \frac{f''(x^*)}{2}(x-x_0)^2 \;.$$

But $f''(x^*) < 0$ (since the second derivative is assumed to be negative everywhere). Thus,

$$f(x) < f(x_0) + f'(x_0)(x-x_0).$$

The next move is to define

$$x_0 = p_0 a_1 + (1-p_0)a_2$$

and to create two inequalities by letting $x = a_1$ or a_2, respectively. This gives:

$$f(a_1) < f(p_0 a_1 + (1-p_0)a_2) + f'(x_0)(a_1 - (p_0 a_1 + (1-p_0)a_2))$$

$$f(a_2) < f(p_0 a_1 + (1-p_0)a_2) + f'(x_0)(a_2 - (p_0 a_1 + (1-p_0)a_2))$$

where I have left the $f'(x_0)$ expression as is, rather than substituting for x_0, since this expression will disappear below.

We're almost done. Multiply the first inequality by p_0 and the second inequality by $(1-p_0)$, and add the two inequalities. This gives the result, namely Eq. (1).