Proof of optimality of uniform distribution of information content

We define the problem as follows: given an utterance u to be expressed in n units, suppose that the difficulty incurred by each unit w_n is proportional to some power k of its negative conditional log-probability:

$$\mathit{diff}(w_i) \propto [-\log P(w_i|w_{1\cdots i-1})]^k$$

and that the total difficulty of u is the sum of the difficulties of all its units.

Theorem. *For any given joint probability p_u for u, setting the conditional probability of each w_i equal at $p_u^{\frac{1}{n}}$ minimizes the total difficulty of u when $k > 1$, and maximizes it when $k < 1$.*

Proof. The proof follows from a simple application of Jensen's inequality, which states that for any random variable X and any convex function f,

$$E[f(X)] \geq f(E[x])$$

and the reverse inequality for any concave function f. Define $p_i \equiv P(w_i|w_{1\cdots i-1})$ (note that by definition, the p_i are constrained such that $\prod_{i=1}^{n} p_i = p_u$). Let X be the random variable

$$P(X = -\log p_i) = \frac{1}{n}$$

and f be the function $f(x) = x^k$. We have

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} \frac{1}{n}[-\log p_i] \\ &= -\frac{1}{n} \log \prod_{i=1}^{n} p_i \\ &= -\frac{1}{n} \log p_u \\ &= -\log p_u^{\frac{1}{n}} \end{aligned}$$

When $k > 1$, f is convex, so by Jensen's inequality we have

$$\sum_{i=1}^{n} \frac{1}{n}[-\log p_i]^k \geq -\log p_u^{\frac{1}{n}}$$

and multiplying through by n gives us the desired result. When $k < 1$, f is concave, and the desired result follows by identical logic.

□