Last lecture, we defined the arithmetic code $C(\vec{x})$ of a given sequence $\vec{x} = (i_1, i_2, \ldots, i_n)$, in terms of the cumulutive probabilities $F(\vec{x})$ and the probabilility $p(\vec{x})$. Today we will look at: Given a way of computing probabilities of individual sequences, how can the cumulatives $F(\vec{x})$ be computed?

How to compute the cumulative distribution?

We show how to compute $F(i_1, i_2, \ldots, i_n)$ inductively. It is easy to compute $F(i_1)$, namely the marginal cumulative on the first variable X_1, evaluated at the particular event $X_1 = i_1$. We then show how to compute $F(i_1, i_2, \ldots, i_k, i_{k+1})$, given $F(i_1, i_2, \ldots, i_k)$. (Notice that k here is not the k we used for the order of the Markov model. We use k because our method is "inductive" and k is typically the variable we use.)

First, consider the base case $\vec{x} = (i_1)$. One considers the marginal $p(X_1)$ and applies the definition of a cumulative distribution function $F(X_1)$ at $X_1 = i_1$:

$$F(\ i_1\ ) \ = \ \sum_{j=1}^{i_1} p(X_1 = j)$$

Now let's do the induction step. Consider the cumulative distribution function defined on the first $k+1$ symbols in the sequence, where $k \geq 1$, namely

$$F(i_1, i_2, \ldots, i_{k+1}) \ \equiv \ \sum_{(j_1 \ldots j_{k+1}) \ \leq \ (i_1 \ldots i_{k+1})} p(j_1, j_2, \ldots, j_{k+1})$$

We break $F(i_1, i_2, \ldots, i_{k+1})$ into two terms as:

$$F(i_1, i_2, \ldots, i_{k+1}) \ = \ [\sum_{(j_1 \ldots j_k) \leq pred(i_1 \ldots i_k)} \ \sum_{j=1}^{N} p(j_1, j_2, \ldots, j_k, \ j) \ ] \ + \ \sum_{j=1}^{i_{k+1}} p(i_1, i_2, \ldots, i_k, j)$$

or equivalently

$$F(i_1, i_2, \ldots, i_{k+1}) \ = \ [\sum_{(j_1 \ldots j_k) \ \leq \ pred(i_1 \ldots i_k)} p(j_1, j_2, \ldots, j_k) \ ] \ + \ \sum_{j=1}^{i_{k+1}} p(i_1, i_2, \ldots, i_k, j)$$

The first term is just $F(\ pred(i_1, i_2, \ldots, i_k)\ )$, that is, the cumulative of the marginal of the first k elements of the sequence, evaluated at $pred(i_1, \ldots, i_k)$.

We can rewrite the second term by taking each term within the sum and applying the definition of conditional probability:

$$p(i_1, i_2, \ldots, i_k, j) \ = \ p(i_1, i_2, \ldots, i_k) \ p(j \mid i_1, i_2, \ldots, i_k) \tag{1}$$

Substituting into the two terms of the sum above gives:

$$F(i_1, i_2, \ldots, i_{k+1}) \ = \ F(\ pred(i_1, i_2, \ldots, i_k)) \ + \ p(i_1, i_2, \ldots, i_k) \sum_{j=1}^{i_{k+1}} p(j \mid i_1, i_2, \ldots, i_k) \tag{2}$$

Aha! This gives us an inductive method for calculating the cumulative distribution function. Here's how:

Suppose we can compute $F(i_1, \ldots, i_k)$ and $p(i_1, \ldots, i_k)$. We immediately get $F(pred(i_1, \ldots, i_k))$, since

$$F(pred(i_1, \ldots, i_k)) \;=\; F(i_1, \ldots, i_k) - p(i_1, \ldots, i_k).$$

To compute $p(i_1, \ldots, i_k, i_{k+1})$ and $F(i_1, i_2, \ldots, i_{k+1})$ we just need to compute $p(j \mid i_1, \ldots, i_k)$ for all $j \leq i_{k+1}$ and plug into the above equation. Assuming we can compute each of the conditional probabilities in constant time (see upcoming lectures), this requires $O(N)$ operations for any k, since $i_{k+1} \leq N$.

Since we require $O(N)$ operations for each $k \leq n$, we require $O(Nn)$ operations in total. This is significantly less than the $O(N^n)$ operations that would be required by Huffman coding!

l_k, u_k notation

Given $\vec{x} = (i_1, i_2, \ldots, i_n)$, define two sequences l_k and u_k such that:

$$\begin{aligned} l_k &= F(pred(\, i_1, i_2, \ldots, i_k)) \\ u_k &= F(\, i_1, i_2, \ldots, i_k) \end{aligned}$$

Note

$$\begin{aligned} u_k - l_k &= F(\, i_1, i_2, \ldots, i_k) - F(pred(\, i_1, i_2, \ldots, i_k)) \\ &= p(i_1, \ldots, i_k) \end{aligned}$$

l stands for "lower" and u stands for "upper". In particular,

$$\begin{aligned} l_n &= F(pred(\vec{x})) \\ u_n &= F(\, \vec{x}\,) \end{aligned}$$

Using Eq. (2) and the above relations, we can rewrite l_{k+1} and u_{k+1} as follows:

$$l_{k+1} = l_k \;+\; (u_k - l_k) \sum_{j < i_{k+1}} p(j \mid i_1, i_2, \ldots, i_k) \tag{3}$$

$$u_{k+1} = l_k \;+\; (u_k - l_k) \sum_{j \leq i_{k+1}} p(j \mid i_1, i_2, \ldots, i_k) \tag{4}$$

Note that there is one case where the definition of l_{k+1} and u_{k+1} is awkward, namely if $i_{k+1} = 1$. In this case, the summation in Eq. (3) has no terms and the summation is 0.

A key property of the sequences l_k and u_k is that the $[l_k, u_k]$ intervals are nested, that is, for all k,

$$[l_{k+1}, u_{k+1}] \;\subseteq\; [l_k, u_k]$$

or equivalent, that l_k is non-decreasing and u_k is non-increasing. To show this, we use the fact that $l_k < u_k$, which is obvious from the definition, since $u_k - l_k = p(i_1, \ldots, i_k)$.

To show $l_k \leq l_{k+1}$,

$$\begin{aligned} l_{k+1} &= l_k + (u_k - l_k) \sum_{j < i_{k+1}} p(j \mid i_1, i_2, \ldots, i_k) \\ &\geq l_k \end{aligned}$$

since $u_k - l_k > 0$. We get a strict inequality if and only if $i_{k+1} > 1$, since otherwise the summation vanishes.

To show that $u_k \geq u_{k+1}$,

$$\begin{aligned} u_{k+1} &= l_k + (u_k - l_k) \sum_{j \leq i_{k+1}} p(j \mid i_1, i_2, \ldots, i_k) \\ &\leq l_k + (u_k - l_k) \cdot 1 \\ &= u_k \end{aligned}$$

By inspection, we get a strict inequality if and only if $i_{k+1} < N$.

F notation

In upcoming lectures, I will simplify the notation even further, by getting rid of the summation in the l_k, u_k update equations. Instead, I will replace the summation with the condition cumulative distribution, so that:

$$l_{k+1} = l_k + (u_k - l_k)\ F(\ pred(i_{k+1}) \mid i_1, i_2, \ldots, i_k)$$

$$u_{k+1} = l_k + (u_k - l_k)\ F(\ i_{k+1} \mid i_1, i_2, \ldots, i_k)$$