CSC 310, Fall 2011 — Theory Assignment #2

Due in class on November 14. Worth 7% of the course grade.

Note that this assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own.

Question 1 (50 marks): Suppose that the designer of a data compression system uses a model for a source of binary symbols, X_1, X_2, $\ldots$, X_n, in which these symbols are independent given the probability, p_1, of symbol 1. (Of course, the probability of symbol 0 is $p_0 = 1-p_1$.) The designer doesn't know p_1, but uses a prior model in which p_1 is either 0, 1, or 1/2, with $P(p_1 = 0) = P(p_1 = 1) = 1/4$ and $P(p_1 = 1/2) = 1/2$.

Consider two ways of encoding such a sequence of length n. (We assume that n is fixed, and known to both the encoder and decoder.) These are:

> **Method A:** Estimate p_1 by examining the whole sequence of n symbols (with the estimate being either 0, 1, or 1/2), then transmit the estimated p_1 using an optimal code based on the prior probabilities, and finally transmit X_1, $\ldots$, X_n using an optimal code based on p_1 (and the assumption that the symbols are independent).
>
> **Method B:** Transmit each of X_1, $\ldots$, X_n using a method such as arithmetic coding that compresses nearly down to the entropy, basing the transmission of X_i on the predictive distribution of X_i given X_1, $\ldots$, X_{i-1}. These predictive probabilities are found by summing the predictive probabilties based on each possible value of p_1 times the posterior probability of that value of p_1 given X_1, $\ldots$, X_{i-1} (which is the prior probability when $i = 1$).

Answer the following, justifying your answers:

a) When using Method A, what is the best way to estimate p_1 — that is, the way that results in the smallest number of bits being transmitted on average?

b) When using Method A, and the estimation method from part (a) above, how many bits (total) are transmitted when $X_1,\ldots,X_n$ has n_0 0s and n_1 1s (with $n_0 + n_1 = n$)?

c) Derive a simple way of finding the predictive distribution for X_i given X_1, $\ldots$, X_{i-1} that is needed to implement Method B.

d) How many bits are transmitted using Method B when X_1, $\ldots$, X_n are all symbol 1? Express your answer in as simple a form as possible. In this and later parts of this question, assume that with the encoding method used, the number of bits that are transmitted as a result of encoding a symbol having probability p is $\log_2(1/p)$ (which need not be an integer).

e) How many bits are transmitted using Method B when X_1, $\ldots$, X_{n-1} are all symbol 1 and X_n is symbol 0? Express your answer in as simple a form as possible.

f) How many bits are transmitted using Method B when X_1 is symbol 0 and X_2, $\ldots$, X_n are all symbol 1? Express your answer in as simple a form as possible.

g) Discuss which of Method A and Method B is better in terms of compression.

Question 2 (50 marks): Recall that for each context the PPM method keeps track of which symbols have been seen before, along with the count of how many times each such symbol has been seen, and transmits a symbol that has been seen before using a probability for it that is proportional to its count. A fixed count of 1 is allocated to an "escape" symbol, that is used when a symbol not seen before needs to be transmitted, using a lower-order context. In this question, we will consider using this scheme when the maximum order is 0, so that symbols are modeled as being independent (not depending on any preceding symbols). In this case, the order 0 context will contain all symbols that have been seen before, and the escape from this context is to the order -1 context, in which all symbols have fixed counts 1, and hence are all equally likely.

Assume that an encoding method such as arithmetic coding is used, so that the number of bits that are transmitted as a result of encoding a symbol having probability p is $\log_2(1/p)$ (which need not be an integer).

a) Describe two deficiencies in this scheme that lead to a code tree in which some nodes have only one child — equivalently, these deficiencies result in the arithmetic coding interval from 0 to 1 having portions that can never contain the final transmitted message.

b) Show that this PPM scheme of maximum order 0 produces exchangeable codes — that is, show that the number of bits in which a message is encoded remains the same if the order of symbols in the message is permuted.

c) Consider a modified scheme in which the count for the escape symbol is not fixed at 1, but instead is equal to $(I-m)/I$, where I is the size of the alphabet and m is the number of distinct symbols seen previously (initially 0), and also, after a symbol is encoded in the order -1 context, its count in the order -1 context is decreased to 0 (so that the probability in the order -1 context of a symbol not previously seen is $1/(I-m)$, rather than $1/I$). Show that this modified scheme never encodes a message in more bits than the original scheme, and characterize the set of messages that the modified scheme encodes in fewer bits than the original scheme.

d) Show that the modified scheme of part (c) does *not* produce an exchangeable code.

e) [5 Mark Bonus] Find a further modification of the modified scheme that does produce an exchangeable code, while still retaining whatever you consider to be its advantages. Do you think being exchangeable makes it better?