Work for Information Theory Supervision I

Supervision 1 covers entropy, the source-coding theorem, Huffman codes and maximum entropy priors. Questions are drawn from a variety of sources: past Tripos questions, MacKay, and some of my own.

Please submit work by 18:30 the day before the supervision.

Questions

1. Get ready to talk in the supervision about entropy, the source-coding theorem and its proof. [There's no need to submit anything in advance for this question.]

2. Suppose that female pandas who live beyond the age of 20 outnumber male pandas in the same age group by three to one. How much information, in bits, is gained by learning that a panda that lives beyond 20 is male?

3. What is the maximum possible entropy H of an alphabet consisting of N different letters? In such a maximum entropy alphabet, what is the probability of its most likely letter? What is the probability of its least likely letter?

4. If discrete symbols from an alphabet S having entropy $H(S)$ are encoded into blocks of length n symbols, we derive a new alphabet of symbols blocks S^n. If the occurrence of symbols is independent, derive the entropy $H(S^n)$ of this new alphabet of symbols blocks.

5. Why are fixed length codes inefficient for alphabets whose letters are not equiprobable? Discuss this in relation to Morse Code.

6. A fair coin is secretly flipped until the first head occurs. Let X denote the number of flips required. The flipper will truthfully answer any "yes-no" questions about his experiment, and we wish to discover thereby the value of X as efficiently as possible.

   (a) What is the most efficient possible sequence of such questions?

   (b) On average, how many questions should we need to ask?

   (c) Relate the sequence of questions to the bits in a uniquely decodable prefix code for X.

7. Is it possible to construct a prefix code in which the codewords have the following lengths: 1, 2, 3, 3, 4, 4?

8. Consider three variable-length codes for a four-symbol alphabet {A, B, C, D } having probabilities $p(x)$ as shown:

x	$p(x)$	Code 1	Code 2	Code 3
A	0.25	00	10	01
B	0.5	1	0	0
C	0.125	01	110	011
D	0.125	10	111	111

   Compare the average codeword length[1] of each code to the entropy of the alphabet, and for each code give all possible decodings of the bit sequence '1001' as a complete message. Which codes are uniquely decodable; which have the prefix (instantaneous) property; which code is best, and why?

[1]Note: don't forget to use the probabilities when computing the average!

9. Find a probability distribution $\{p_1, p_2, p_3, p_4\}$ such that there are two optimal codes that assign different lengths $\{l_i\}$ to the four symbols.

10. Construct an ensemble where the difference between the entropy and the expected length of the Huffman code is as large as you can make it.

11. You are tasked with investigating a funky random number generator, which generates integers i, where $1 \leq i \leq n$. The true distribution $(p_1, p_2, ..., p_n)$ is unknown, but the average μ is known. In order to perform inference and estimate the posterior distribution using Bayes theorem, we need a prior distribution.

    (a) Show, using the method of Lagrange multipliers, that for the maximum entropy prior p_i can be written as $C r^i$ for some C and r.

    (b) Use normalisation to find C.

    (c) Use the known average of the distribution to obtain the following equation:

    $$n r^{n+1} - (n+1) r^n + 1 = \mu(r^n - 1)(r - 1)$$

    (d) This equation is difficult to solve analytically[2], so solve it numerically (e.g. using Python, Mathematica or Matlab). For $n = 6$, plot the distribution of p_i for $\mu = 3.5$, $\mu = 2$ and $\mu = 5$.

[2] Possible? Not sure. Challenge?