

1. Show that two sets  $A, B \subset \Omega$  are independent if and only if

$$I(A \cap B) = I(A) + I(B) .$$

The *mutual information* of two random variables  $X, Y$  is

$$I(Y, X) = H(Y) - H(Y|X) .$$

Show that  $I(X, Y) = I(Y, X)$  and that  $I(Y, X) \geq 0$ . When is there equality?

2. Show that  $H(Y|X) \geq 0$  with equality if and only if  $Y = f(X)$  almost surely for some function  $f$ . (That is, there is a function  $f$  with  $\mathbb{P}(f(X) = Y) = 1$ .)

Prove that  $H(Y|X) \leq H(Y)$  but give an example to show that  $H(Y|X = x)$  may exceed  $H(Y)$  for some value  $x$ .

3. Two players  $A$  and  $B$  play a best of 5 set tennis match. Let  $X$  be the number of sets won by  $A$  and  $Y$  the total number of sets played. Assuming that the players are equally matched and the outcome of the sets are independent, compute the conditional entropies  $H(X|Y)$ ,  $H(Y|X)$  and the mutual information  $I(X; Y)$ .
4. (a) Give an example of a decodable code which is not prefix-free. (Hint: Reverse the code words in a prefix-free code.)  
 (b) Give an example of a non-decodable code which satisfies Kraft's inequality.  
 (c) A *comma code* is one where a special letter — comma — occurs at the end of each code word and nowhere else. Show that a comma code is prefix-free and check directly that comma codes satisfy the Kraft inequality.
5. The product of two codes  $c_j : \mathcal{A}_j \rightarrow \mathcal{B}_j^*$  is

$$g : \mathcal{A}_1 \times \mathcal{A}_2 \rightarrow (\mathcal{B}_1 \times \mathcal{B}_2)^* ; \quad g : (a_1, a_2) \mapsto c_1(a_1)c_2(a_2) .$$

Show that the product of two prefix-free codes is prefix-free but that the product of a decodable code and a prefix-free code need not even be decodable.

6. Jensen's inequality states that if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a convex function and  $p_1, \dots, p_n$  is a probability distribution then  $f(\sum p_i x_i) \leq \sum p_i f(x_i)$  for any  $x_1, \dots, x_n \in \mathbb{R}$ . Prove Gibbs' inequality by applying Jensen's inequality to the convex function  $f(x) = -\log x$ .
7. Show that the function  $f(t) = t \ln t$  is convex on the unit interval.

Let  $\mathbf{p} = (p_j)$  and  $\mathbf{q} = (q_j)$  be two probability distributions. Show that, for any number  $\lambda$  with  $0 < \lambda < 1$ , the entropy satisfies:

$$H((1 - \lambda)\mathbf{p} + \lambda\mathbf{q}) \geq (1 - \lambda)H(\mathbf{p}) + \lambda H(\mathbf{q}) .$$

Let  $p_1, p_2, p_3$  be a probability distribution. Show that  $H(p_1, p_2, p_3) \leq H(p_1, 1 - p_1) + (1 - p_1)$  and determine when equality occurs.

8. Use the methods of Shannon–Fano and Huffman to construct prefix-free binary codes for the alphabet  $a_1, a_2, \dots, a_5$  emitted either (a) with equal probabilities, or (b) with probabilities 0.3, 0.3, 0.2, 0.15, 0.05. Compare the expected word lengths in each case to the entropy.
9. Letters  $a_1, a_2, \dots, a_5$  are emitted with probabilities 0.4, 0.2, 0.2, 0.1, 0.1. Find an optimal binary code. Determine whether there are optimal binary codes with either (a) all but one codeword of the same length, or (b) each codeword a different length.
10. A gastric infection is known to originate in exactly one of  $m$  restaurants; the probability it originates in the  $j$ th being  $p_j$ . A health inspector has samples from all of the  $m$  restaurants and by testing the pooled samples from a set  $S$  of them can determine with certainty whether the infection originates in  $S$  or its complement. Let  $N(p_1, \dots, p_m)$  denote the minimum expected number of such tests needed to locate the infection. Show that  $H(p_1, \dots, p_m) \leq N(p_1, \dots, p_m) \leq H(p_1, \dots, p_m) + 1$ , and determine when the lower bound is attained.

- 
11. Let  $X, Y$  be two random variables that each take values in an alphabet  $\mathcal{A}$  of size  $K$ . Let  $I$  be the indicator random variable

$$I = \begin{cases} 1 & \text{if } Y = X; \\ 0 & \text{otherwise.} \end{cases}$$

- (a) Show that

$$H(Y, I|X) = H(I|Y, X) + H(Y|X) = H(Y|I, X) + H(I|X) .$$

[Hint: Recall that  $H(A, B) = H(A|B) + H(B)$ .]

- (b) Show that

$$\begin{aligned} H(I|Y, X) &= 0 \\ H(Y|I, X) &= \mathbb{P}(I = 0)H(Y|I = 0, X) + \mathbb{P}(I = 1)H(Y|I = 1, X) \leq \mathbb{P}(I = 0) \log_2(K - 1) \\ H(I|X) &\leq H(I) . \end{aligned}$$

Deduce Fano's inequality:

$$H(Y|X) \leq H(p, 1 - p) + p \log_2(K - 1) \leq 1 + p \log_2(K - 1)$$

where  $p = \mathbb{P}(Y \neq X)$ .

12. Use a Lagrange multiplier to solve the following constrained optimisation problem: Given  $p_i > 0$  with  $\sum_{i=1}^m p_i = 1$  find real numbers  $l_1, \dots, l_m$  to minimise  $\sum_{i=1}^m p_i l_i$  subject to  $\sum_{i=1}^m D^{-l_i} \leq 1$ .
13. Extend the definition of entropy to a random variable taking values in the non-negative integers. Compute the expected value  $E(X)$  and entropy  $H(X)$  of a random variable  $X$  with  $P(X = k) = p(1 - p)^k$ . Show that among non-negative integer valued random variables with the same expected value,  $X$  achieves the maximum possible entropy.
14. In a horse race with  $m$  horses the probability that the  $i$ th horse wins is  $p_i$ . The odds offered on each horse are  $a_i$ -for-1, i.e. a bet of  $\mathcal{L}x$  on the  $i$ th horse will yield  $\mathcal{L}a_i x$  pounds if the horse wins, and nothing otherwise. A gambler bets a proportion  $b_i$  of his wealth on horse  $i$ , with  $\sum_{i=1}^m b_i = 1$ . He seeks to maximise  $W = \sum_{i=1}^m p_i \log(a_i b_i)$ . Suggest a motivation for this choice. Solve to find the  $b_i$  that maximize  $W$ . Show that in the case where all of the odds are equal this maximum and the entropy  $H(p_1, \dots, p_m)$  sum to a constant.
15. (Huffman encoding for non-binary codes.) In the binary case the Huffman code is defined by combining the two letters with smallest probabilities. In general, for a code  $h : \mathcal{A} \rightarrow \mathcal{B}^*$  into an alphabet  $\mathcal{B}$  with size  $D$ , we combine the  $D$  letters with smallest probabilities. In order to be able to do this, first add extra letters to  $\mathcal{A}$ , each with probability 0, so that the size of  $\mathcal{A}$  is congruent to 1 modulo  $D - 1$ .

Carry this out for a ternary coding of an alphabet with probabilities 0.2, 0.2, 0.15, 0.15, 0.1, 0.1, 0.05, 0.05.

16. You are given  $N$  apparently identical coins, one of which may be a forgery. Forged coins are either too light or too heavy. You have a balance, on which you may place any of the coins you like and determine whether the coins in one pan are together lighter, heavier or the same weight as those in the other. Using the balance you wish to detect whether there is a forgery and, if so, which coin it is and whether it is lighter or heavier.

Prove that at least  $\log_3(2N + 1)$  weighings are required.

\* Show that for  $N = 12$  three weighings suffice.

---

*An annotated version of this example sheet is available for supervisors from DPMMS.*

*Please send any comments or corrections to me at: t.k.carne@dpmms.cam.ac.uk .*