MATHEMATICAL TRIPOS PART II 2013
CODING AND CRYPTOGRAPHY
EXAMPLE SHEET 2

The first 12 examples are minimal to cover the course; you are also encouraged to try questions 13–15.

1 In a binary symmetric channel we usually take the probability p of error to be strictly less than $1/2$. Why do we not consider the case $1 \geq p > 1/2$? What if $p = 1/2$?

2 In an examination each candidate is asked to write down a Candidate Number of the form $2234A$, $2235B$, $2236C$,... (the eleven possible letters are repeated cyclically) and a Desk Number. (Thus candidate 0004 sitting at desk 425 writes down $0004D - -425$.) The first four numbers in the Candidate Number identify the candidate uniquely. Show that if the candidate makes one error in the Candidate Number then that error can be detected without using the Desk Number. Would this be true if there were nine possible letters repeated cyclically? Would this be true if there were twelve possible letters repeated cyclically? Give reasons.

Show that if we combine the Candidate Number and the Desk Number the combined code is 1-error correcting.

3 Into the Stuart's Handicap at Royal Basket are entered m horses, the probability that the jth horse wins being p_j. The odds offered on each horse are a_j-for-1 (meaning a wager of £x on the jth horse will yield £$a_j x$ if the horse wins, and nothing otherwise). Chevalier de Méré[1] bets a proportion b_j of his bankroll on horse j, with $\sum_{j=1}^m b_j = 1$. He seeks to maximise $W = \sum_{j=1}^m p_j \log(a_j b_j)$. Suggest a motivation for this choice. Solve to find the b_j that maximise W. Show that in the case where all of the odds are equal this maximum and the entropy $H(p_1, \ldots, p_m)$ sum to a constant.

[1] The fortunes and misfortunes of a famous gambler, the Chevalier de Méré, were the origin of an algebraic approach to probability. A noted rake and *bon vivant*, the Chevalier had made his pile by always betting small favourable odds on getting at least one six in four tosses of a die, then lost it by always betting small odds on getting at least one double six in twenty-four double tosses. 'Il est très bon ésprit,' wrote Pascal to Fermat about the Chevalier, 'mais quel dommage, il n'est pas géomètre.'

4 If you look at the inner title page of almost any book published between 1974 and 2007, you will find its International Standard Book Number (ISBN-10). The ISBN-10 uses single digits selected from 0, 1, …, 8, 9 and X representing 10. Each ISBN-10 consists of nine such digits a_1, a_2, …, a_9 followed by a single check digit a_{10} chosen so that

$$(*)\qquad 10a_1 + 9a_2 + \cdots + 2a_9 + a_{10} \equiv 0 \pmod{11}.$$

(In more sophisticated language, our code C consists of those elements $\mathbf{a} \in \mathbb{F}_{11}^{10}$ such that $\sum_{j=1}^{10}(11-j)a_j = 0$.)

(i) Find a couple of books[2] and check that $(*)$ holds for their ISBNs.

(ii) Show that $(*)$ will not work if you make a mistake in writing down one digit of an ISBN.

(iii) Show that $(*)$ may fail to detect two errors.

(iv) Show that $(*)$ will not work if you interchange two distinct adjacent digits (a transposition error).

(v) Does (iv) remain true if we remove the word 'adjacent' ? Errors of type (ii) and (iv) are the most common in typing.

In communication between publishers and booksellers, both sides are anxious that errors should be detected but would prefer the other side to query errors rather than to guess what the error might have been.

(vi) Since the ISBN contained information such as the name of the publisher, only a small proportion of possible ISBNs could be used[3] and the system described above started to 'run out of numbers'. A new system was introduced which is compatible with the system used to label most consumer goods. After January 2007, the appropriate code became a 13 digit ISBN-13 number $x_1x_2\dots x_{13}$ with each digit selected from 0, 1, …, 8, 9 and the check digit x_{13} computed by using the formula

$$x_{13} \equiv -(x_1 + 3x_2 + x_3 + 3x_4 + \cdots + x_{11} + 3x_{12}) \pmod{10}.$$

Show that we can detect single errors. Give an example to show that we cannot detect all transpositions.

5 Suppose we use eight hole tape with the standard paper tape code (i.e. the simple parity check code of length 8) and the probability that an error occurs at a particular place on the tape (i.e. a hole occurs where it should not or fails to occur where it should) is 10^{-4}. A program requires about 10 000 lines of tape (each line containing eight places) using the paper tape code. Using the Poisson approximation, direct calculation (possible with a hand calculator but really no advance on the Poisson method), or otherwise, show that the probability that the tape will be accepted as error free by the decoder is less than .04%.

Suppose now that we use the Hamming scheme (making no use of the last place in each line). Explain why the program requires about 17 500 lines of tape but that any particular line will be correctly decoded with probability about $1 - (21 \times 10^{-8})$ and the probability that the entire program will be correctly decoded is better than 99.6%.

6 Determine the set of integers n for which the repetition code of length n is perfect for the binary alphabet $\{0, 1\}$.

7 Two codewords are chosen independently at random from $\mathbb{F}_2^n$ with each string equally likely. What is the expected Hamming distance between them?

[2] try a place called the 'College Library' (ask the Porters where it is).

[3] The same problem occurs with telephone numbers. If we use the Continent, Country, Town, Subscriber system we will need longer numbers than if we just numbered each member of the human race.

8 Let C be the binary $[11, 12]$-code consisting of the word 10111000100 and its ten cyclic shifts (that is 01011100010, 00101110001 and so on) together with the zero codeword. Show that C has minimum distance 5.

9 The original Hamming code was a 7-bit code used in an 8-bit system (paper tape).

(i) Consider the code $c : \{0,1\}^4 \to \{0,1\}^8$ obtained by using the Hamming code for the first seven bits $x_1, \dots, x_7$ and then a check digit x_8 chosen such that

$$x_1 + x_2 + \cdots + x_8 \equiv 0 \pmod{2}.$$

Find the minimum distance for this code. How many errors can it detect? How many can it correct?

(ii) Given a code of length n which corrects e errors can you always construct a code of length $n+1$ which detects $2e+1$ errors?

10 We usually work under the assumption that all messages sent through our noisy channel are equally likely. In this question we drop this assumption.

Suppose that each bit sent through a binary symmetric channel has probability $p = 1/3$ of being mistransmitted. There are four codewords 1100, 0110, 0001, 1111 sent with probabilities $1/4$, $1/2$, $1/12$, $1/6$. If you receive 1001 what will you decode it as, using each of the following rules?

(i) The ideal observer rule: find $\mathbf{b} \in C$ so as to maximise

$$\mathbb{P}(\mathbf{b} \text{ sent} \mid \mathbf{u} \text{ received}).$$

(ii) The maximum likelihood rule: find $\mathbf{b} \in C$ so as to maximise

$$\mathbb{P}(\mathbf{u} \text{ received} \mid \mathbf{b} \text{ sent}).$$

11 Define $A(n, \delta)$ as the maximum size of a binary code of length n with minimum distance δ. Write down the values of $A(n, 1)$, $A(n, 2)$, $A(n, n-1)$ and $A(n, n)$. Prove that

$$\frac{2^n}{V(n, \delta - 1)} \leqslant A(n, \delta) \leqslant \frac{2^n}{V(n, \frac{1}{2}(\delta - 1))}.$$

12 (i) Construct a $(7, 8, 4)$-code from Hamming's code.

(ii) Prove that if $\delta < n$ then $A(n, \delta) \leqslant 2A(n-1, \delta)$.

(iii) Prove that if δ is even then $A(n-1, \delta-1) = A(n, \delta)$.

(iv) Hence compute $A(6, 4)$.

13 Let C be an (n, m, d)-code. Show that

$$m(m-1)d \leqslant \sum \sum d(\mathbf{c}_i, \mathbf{c}_j) \leqslant \frac{1}{2} n m^2$$

where the sum is over all codewords $\mathbf{c}_i$ and $\mathbf{c}_j$ of C. Use this to give an upper bound on $A(n, d)$ in the case $n < 2d$.

14 Your employer announces that he is abandoning the old-fashioned paternalistic scheme under which he guarantees you a fixed sum Kx (where, of course, K, $x > 0$) when you retire. Instead, he will empower you by giving you a fixed sum x now, to invest as you wish. In order to help you and the rest of the staff, your employer arranges that you should obtain advice from a financial whizz-kid with a top degree from Cambridge. After a long lecture in which the whizz-kid manages to be simultaneously condescending, boring and incomprehensible, you come away with the following information.

When you retire, the world will be in exactly one of n states. By means of a piece of financial wizardry called hedging, the whizz-kid can offer you a pension plan which for the cost of x_i will return $Kx_iq_i^{-1}$ if the world is in state i, but nothing otherwise. (Here $q_i > 0$ and $\sum_{i=1}^n q_i = 1$.) The probability that the world will be in state i is p_i. You must invest the entire fixed sum. (Formally, $\sum_{i=1}^n x_i = x$. You must also take $x_i \geq 0$.) On philosophical grounds you decide to maximise the expected value S of the logarithm of the sum received on retirement. Assuming that you will have to live off this sum for the rest of your life, explain why this choice is reasonable or why it is unreasonable.

Find the appropriate choices of x_i. Do they depend on the q_i?

Suppose that K is fixed, but the whizz-kid can choose q_i. We may suppose that what is good for you is bad for him so he will seek to minimise S for your best choices. Show that he will choose $q_i = p_i$. Show that, with these choices,

$$S = \log Kx.$$

15 (i) Show that $-t \geq \log(1-t)$ for $0 \leq t < 1$.

(ii) Show that, if $\delta_N > 0$, $1 - N\delta_N > 0$ and $N^2\delta_N \to \infty$, then

$$\prod_{m=1}^{N-1}(1 - m\delta_N) \to 0.$$

(iii) Let $V(n,r)$ be the number of points in a Hamming ball of radius r in $\mathbb{F}_2^n$ and let $p(n,N,r)$ be the probability that N such balls chosen at random do not intersect. By observing that if m non-intersecting balls are already placed, then an $m+1$st ball which does not intersect them must certainly not have its centre in one of the balls already placed, show that, if $N_n^2 2^{-n} V(n, r_n) \to \infty$, then $p(n, N_n, r_n) \to 0$.

(iv) Assuming $\alpha \leqslant \frac{1}{2}$, show that, if $2\beta + H(\alpha) > 1$, then $p(n, 2^{\beta n}, \alpha n) \to 0$.

Thus simply throwing balls down at random will not give very good systems of balls with empty intersections.

SM, Lent Term 2013
Comments on and corrections to this sheet may be emailed to `sm@dpmms.cam.ac.uk`