Solution Booklet

DMM 2021

1 Power Round

The theme is *Error Correction Codes*. The are a total of 50 points for this round.

1.1 Check digit

Consider the typical credit card number, which has 16 digits. Let d_i denote the ith digit. The first 15 digits is the account number, and the last digit is the *check digit*, which is given by the remainder of the sum of the first 15 digits, i.e.

$$d_{16} = d_1 + d_2 + \ldots + d_{15} \pmod{10}.$$

For example, a valid credit card number could be 1234 1234 1234 1236. If we have made a mistake in a digit and instead typed 7234 1234 1234 1236, the computer can easily check that the sum of the first 15-digits is $2 \not\equiv 6 \pmod{10}$ and detect an error.

Problem 1: (7 points total)

(a) (2 points) Describe all the valid card numbers of the form 2021 1106 _ _42 1337.

By computation we obtain $d_9 + d_{10} \equiv 1 \pmod{10}$. Hence there are 10 valid combinations where $(d_9, d_{10}) = (1, 0), (2, 9), (3, 8), \ldots, (0, 1)$.

(b) (2 points) Suppose we are given an invalid credit card number where we know that exactly one of the sixteen digits is wrong. Are we able to recover the correct credit card number? Prove your answer.

No. consider $A = 0000\,0000\,0000\,0000$ and $B = 1000\,0000\,0000\,0001$. If we are given the number $0000\,0000\,0000\,0001$, we cannot tell if it was originally A or B.

(c) (3 points) We can also consider a general formula for the check digit. Let a_i be an integer from 1 to 10. We can choose the following formula for the check digit:

$$d_{16} \equiv a_1 d_1 + a_2 d_2 + \ldots + a_{15} d_{15} \pmod{10}.$$

Find the number of tuples $(a_1, a_2, \ldots, a_{15})$ that we can choose such that an error is always detected if one of the digits is wrong.

4^{15}. We must choose a_i that is coprime to 2 or 5. There are 4 choices for each a_i.

1.2 Error Correction Codes

Consider a set-up that involve binary strings or bitstrings, which are finite sequences of the digits 0 and 1. Bob and Dylan wants to send messages to each other, but they can only do so via bitstrings.

Suppose Bob wants to send the letter A or B to Dylan, via the encoding

$$A = 0, B = 1.$$

During transmission, each bit has a probability $p = 0.1$ of changing due to random noise, and this is independent for each bit. If Bob wants to send A to Dylan using the string 0, there is a 10% chance that Dylan receives the string 1 and misinterprets the message as B.

Instead, Bob and Dylan can agree on an encoding scheme H that allows them to send a letter A or B via the following bitstrings:

$$A = 000, B = 111.$$

Bob sends the bitstring 000 or 111 to Dylan. When Dylan receives the bitstring, he checks whether it is more likely to be A or B.

For example if Dylan receives the bitstring 010, he can tell that Bob has most likely sent the letter A. This is because it will take two errors to modify 111 into 010, which is less likely to happen. Hence we have designed a code that is safe if there is at most one bit of error during transmission.

Problem 2: (2 points) Using H, what is the probability that the correct message is received by Dylan? Give the numeric answer.

The message is correct if there is at most one error. This happens with probability $(1-p)^3 + 3p(1-p)^2 = 0.972$

For any encoding scheme, the set of bitstrings that can be sent are called *codewords*. For H, the codewords are $\{000, 111\}$. Define the *distance* between two strings X and Y to be $d(X, Y) =$ total number of positions where the bit differs. Eg. $d(00, 01) = 1$.

For any encoding scheme, if we receive S, we choose the codeword that has the minimum distance from S. If there are two codewords with the same minimum distance from S, then there is no valid decoding.

Consider the encoding scheme H' given by

$$A = 00011, B = 11111, C = 00000.$$

Problem 3: (6 points total)

(a) (1 point) Under H', suppose that the bitstring X is decoded as C. What are all the possible values of $d(00000, X)$? Prove you answer.

d can be $0, 1, 2$, for $X = 00000, 10000, 11000$. If $d \geq 3$ then X will be decoded as A or B.

(b) (2 points) Find all the bitstrings X where we cannot decide on a valid decoding of X.

There are 8 such bitstrings given by

$$d(X, A) = d(X, C) : (00001, 00010, 10001, 10010, 01001, 01010, 00101, 00110)$$

A tie cannot occur between B and C, nor can it occur between B and A.

(c) (3 points) What is the probability that Bob sends A, but Dylan decodes the bitstring as B?

Changing any of the last two digits of 00011 increases the distance to A and B by 1. Hence at least two of the first three digits must be modified.

If two of the first three digits are changed to 1, then we can have at most one change in the last two digits. The probability in this case is $3p^2(1-p)(1-p^2)$

If all three digits are changed to 1, then this string will always be decoded as B. This has probability p^3. In total, we have $p^3 + 3p^2(1-p)(1-p^2) = 0.02773$.

1.3 A Larger Code

Consider a general encoding scheme. The bitstrings all have length N, and there are a total of R codewords. The minimum distance between any two codewords is denoted by D.

(For H, we have $N = 3, R = 2, D = 3$.)

Problem 4: (7 points total)

(a) (2 points) Prove the following triangle inequality: For any three codewords X, Y, Z we have

$$d(X, Y) \leq d(X, Z) + d(Y, Z)$$

If the ith digit of X and Y are different, then the ith digit of Z must be different from that of X or Y.

(b) (2 points) Define $M = \lfloor \frac{D-1}{2} \rfloor$. Conclude from earlier that if X is a codeword and $d(X, S) \leq M$, then S must be decoded as X.

Suppose otherwise, then for some other codeword $Y \neq X$ we also have $d(Y, S) \leq M$. Then $d(X, Y) \leq 2M \leq D - 1$, a contradiction.

(c) (3 points) Prove that for any encoding scheme, we have

$$R \leq \frac{2^N}{\sum_{t=0}^{M} \binom{N}{t}}.$$

By 4(b), for each codeword X there are at least $\sum_{t=0}^{M} \binom{N}{t}$ bitstrings that decodes to X. Hence there are at least $R \cdot (\sum_{t=0}^{M} \binom{N}{t})$ bitstrings with a valid decoding, and this is bounded by the total number of possible bitstrings 2^N.

We will now consider an encoding scheme J that allows us to send any 4-letter word made up of A and B. Convert the word w to a bitstring $\overline{x_1x_2x_3x_4}$ by changing $A \to 0, B \to 1$. To this bitstring, we add 3 more digits x_5, x_6, x_7 defined by

$$x_5 = x_1 + x_2 + x_4 \pmod 2, x_6 = x_1 + x_3 + x_4 \pmod 2, x_7 = x_2 + x_3 + x_4 \pmod 2.$$

Thus J is a scheme where $N = 7$ and $R = 16$.

Problem 5: (7 points total)

(a) (3 points) Prove that the distance between any two codewords in J cannot be 1 or 2.

For any two codewords, there must be at least one difference in the first 4 digits. If the number of differences in the first 4 digits is exactly 1, then two or three of the last 3 digits must be different. If the number of differences is exactly 2, then at least one of x_5, x_6, x_7 changes by exactly 1.

(b) (4 points) Show that $D = 3$, and prove that every bitstring of length 7 has a decoding.

$D = 3$ can be shown by the codewords $0000000, 1000110$. Hence for each codeword X, there are at least $1 + 7 = 8$ bitstrings that decode to X. Hence there are at least $8 * 16 = 128$ bitstrings with valid decodings, which must include all bitstrings of length 7.

If there is at most one erroneous bit in X, we can find the correct decoding of the bitstring X. This can be done in a tedious way by comparing X with each codeword. However, we would like to consider a more efficient method to decode X. Consider the three numbers $y_1, y_2, y_3 \in \{0, 1\}$ given by

$$y_1 \equiv x_1 + x_2 + x_4 + x_5 \pmod 2,$$
$$y_2 \equiv x_1 + x_3 + x_4 + x_6 \pmod 2,$$
$$y_3 \equiv x_2 + x_3 + x_4 + x_7 \pmod 2.$$

Problem 6: (4 points total)

(a) (1 point) Compute (y_1, y_2, y_3) for the codeword representing $BAAA$ if there is no bit error, and also if we change the 3rd bit.

The codeword for $BAAA$ is 1000110. We have $(y_1, y_2, y_3) = (0, 0, 0)$. If the $3rd$ bit is wrong, the bitstring becomes 1010110 and $(y_1, y_2, y_3) = (0, 1, 1)$

(b) (3 points) Suppose that there is at most one erroneous bit in some random bitstring X. Show that we can deduce which bit has been changed (or none at all) given only the values of y_1, y_2, y_3.

Note that by plugging in the formulas for x_5, x_6, x_7, we can show that $(y_1, y_2, y_3) = (0, 0, 0)$ when there is no errors. Next, if exactly one of $y_i = 1$, then the error must be in bit x_{i+4}. If exactly one of $y_i = 0$, we have $i = 1 \implies x_3$ is wrong, $i = 2 \implies x_2$ is wrong, $i = 3 \implies x_1$ is wrong. Finally if $y_1 = y_2 = y_3$ then x_4 is wrong.

1.4 Generic Codes

In real life, we may need to encode very long messages, and hence we have to construct larger encoding schemes. For this section you will explore the limits of what kind of encodings we can construct.

Recall the inequality from Problem 4(c). We want to find positive integers R, M, N such that the equality can be achieved. In other words, we want

$$R = \frac{2^N}{\sum_{t=0}^{M} \binom{N}{t}}. \tag{1}$$

Problem 7: (3 points) Suppose that $D = 3$ throughout this problem. Find all combinations of (N, R) such that the equation in (1) is satisfied.

We have $R(N+1) = 2^N$. Hence $N = 2^k - 1$ for some positive integer k, which implies that $R = 2^{2^k-k-1}$. Thus $(N, R) = (2^k - 1, 2^{2^k-k-1})$.

Problem 8: (8 points total) Suppose that $D = 7$ throughout this problem.

(a) (3 points) Show that

$$(N^2 - N + 6)(N + 1) = 3 \cdot 2^k$$

for some positive integer $k \geq 2$. Conclude that $N + 1 = 2^l$ or $3 \cdot 2^l$ for some non-negative integer l.

The equation in (1) rewrites as $R(1 + \binom{N}{1} + \binom{N}{2} + \binom{N}{3}) = 2^N$. Hence R is a power of 2.

Multiplying 6 on both sides, we can obtain the equation $(N^2 - N + 6)(N + 1) = 3 \cdot 2^k$ for $k \geq 2$.

(b) (3 points) Show that if $l \geq 4$, then the equation in (1) has no solutions.

Suppose that $N+1 = 2^l$. Then substituting this into $N^2 - N + 6 = 3 \cdot 2^{k-l}$, we have $2^{2l} - 3 \cdot 2^l + 8 = 3 \cdot 2^{k-l}$. If $l \geq 4$, then $3 \cdot 2^{k-l} \equiv 8 \pmod{16}$, which implies that $k - l = 3$. Thus we have $2^{2l} - 3 \cdot 2^l - 16 = 0$, which has no solutions when $l \geq 4$. A similar argument shows that $N+1 = 3 \cdot 2^l$ also has no solutions.

(c) (2 points) Find all the pairs (N, R) that satisfy the equation in (1).

Checking $l = 1, 2, 3$, we see that the only solutions are $(N, R) = (7, 2)$ and $(23, 4096)$

The last problem concerns finding encodings that attain equality in (1).

Problem 9: (6 points) Construct an encoding for each combination (N, R) in Problem 7.

This is the Hamming code, and the idea is to extend the construction used in the scheme J.

For the bitstring of length $2^k - 1$, we designate bit $1, 2, 4, \ldots, 2^{k-1}$ to be the parity bits. The remaining $2^k - k - 1$ are used for the message.

For bit 2^i, we define it to be the sum of all the bits at position j, where the $(i+1)$th digit in the binary representation of j is 1 and $j \neq 2^i$.