

# Assignment 1

Joshua Hwang (44302650)

7 March

## Question 2

We transform the question into formal mathematical notation.

$$P(\text{Error uncorrected}|\text{Error occurred}) = \frac{P(\text{Error uncorrected} \cap \text{Error occurred})}{P(\text{Error occurred})}$$

All errors less than  $e$  can be solved thus all uncorrected errors (excluding the case where there's no error) can be represented as a summation from  $e + 1$  to  $n$ .

$$= \frac{\sum_{i=e+1}^n {}^nC_e p^{n-i} q^i}{P(\text{Error occurred})}$$

The chance an error occurs is the opposite of no error occurring (which is  ${}^nC_0 p^n q^0$ ).

$$\begin{aligned} &= \frac{\sum_{i=e+1}^n {}^nC_e p^{n-i} q^i}{1 - {}^nC_0 p^n q^0} \\ &= \frac{\sum_{i=e+1}^n {}^nC_e p^{n-i} q^i}{1 - p^n} \end{aligned}$$

## Question 5

Find probability of undetected error with  $p = 0.9$  and  $q = 0.1$ .

**a**

Using a (5,4) parity check binary code. It is guaranteed to check a single error but can also detect an odd number of errors. Thus,

$${}^5C_1 0.9^4 0.1^1 = 0.32805$$

$${}^5C_3 0.9^2 0.1^3 = 0.0081$$

$${}^5C_5 0.9^0 0.1^5 = 0.00001$$

$$\text{Sum of probabilities} = 0.33616$$

The chance that an error occurs at all (will be used later) is,

$$\begin{aligned} P(\text{Error occurred}) &= 1 - P(\text{No error}) \\ &= 1 - 0.9^5 \\ &= 0.40591 \end{aligned}$$

This is the probability of detecting an error even if an error isn't present,  $P(\text{Error uncorrected} \cap \text{Error occurred})$ . We actually want the probability with respect to mutated code. Thus we use the previous questions formula to properly convert it,

$$\begin{aligned} \frac{P(\text{Error uncorrected} \cap \text{Error occurred})}{P(\text{Error occurred})} &= \frac{0.33616}{0.40591} \\ &= 0.8282 \quad \text{To 4 decimal places} \end{aligned}$$

Now the undetected error is  $1 - 0.8282 = 0.1718$ .

**b**

Using a (12, 4) triple repetition binary code. It is guaranteed to detect any 2 errors. For 3 errors we can detect all errors excluding when the error occurs in the same place in each triplet. i.e. abcd efgh ijkl errors occurring at a,e,i will be undetected, there are  ${}^4C_1$  cases where this happens. For 4 and 5 errors it will again detect an error since there is no way the triplets could each have the same errors. At 6 errors it is possible each triplet has two errors in the same locations, there are  ${}^4C_2$  ways this could happen. Up until 12 errors can also be detected from the same reasoning as 4 and 5 errors. If every single bit in the code is mutated there is no way we can detect such an error.

$$\begin{aligned}
{}^1_2C_1 0.9^{11} 0.1^1 &= 0.3766 \\
{}^1_2C_2 0.9^{10} 0.1^2 &= 0.2301 \\
({}^1_2C_3 - {}^4C_1) 0.9^9 0.1^3 &= 0.0837 \\
{}^1_2C_4 0.9^8 0.1^4 &= 0.0213 \\
{}^1_2C_5 0.9^7 0.1^5 &= 0.0038 \\
({}^1_2C_6 - {}^4C_2) 0.9^6 0.1^6 &= 0.0005 \\
{}^1_2C_7 0.9^5 0.1^7 &= 0.0000 \\
{}^1_2C_8 0.9^4 0.1^8 &= 0.0000 \\
{}^1_2C_9 0.9^3 0.1^9 &= 0.0000 \\
{}^1_2C_{10} 0.9^2 0.1^{10} &= 0.0000 \\
{}^1_2C_{11} 0.9^1 0.1^{11} &= 0.0000 \\
\text{Sum of probabilities} &= 0.7160
\end{aligned}$$

Now we calculate the chance of an error occurring,

$$\begin{aligned}
P(\text{Error occurred}) &= 1 - P(\text{No error}) \\
&= 1 - 0.9^{12} \\
&= 0.7176
\end{aligned}$$

Now we calculate our final answer just like the previous question.

$$\begin{aligned}\frac{P(\text{Error uncorrected} \cap \text{Error occurred})}{P(\text{Error occurred})} &= \frac{0.716}{0.7176} \\ &= 0.9978 \quad \text{To 4 decimal places}\end{aligned}$$

Now the undetected error is  $1 - 0.9978 = 0.0022$ .

## Question 6

**a**

Without loss of generality we split our set of possible words into four.

$$S = S_{00} \cup S_{01} \cup S_{10} \cup S_{11}$$

Where  $S_{ij}$  begins with either i then j. Since we need 9 sequences one of the subsets has at least 3 sequences. Without loss of generality we'll set  $S_{01}$  as this set. We'll denote them as,

$$01a_1b_1c_1d_1$$

$$01a_2b_2c_2d_2$$

$$01a_3b_3c_3d_3$$

Where each letter represents a digit in each word. Now each word must be distance 3 apart. Since the order of the digits doesn't matter we'll arbitrarily set  $a_1 \neq a_2$ ,  $b_1 \neq b_2$ ,  $c_1 \neq c_2$  without losing generality. Now at least 2 of  $a_3$ ,  $b_3$ ,  $c_3$  must not be equal to their  $a_2$ ,  $b_2$ ,  $c_2$  counterparts to maintain a distance of 3. Without loss of generality we will take  $a_3$  and  $b_3$  as these values. But this will mean  $a_3 = a_1$  and  $b_3 = b_1$ . If  $a_3 = a_1$  and  $b_3 = b_1$  then there aren't enough digits to make the third sequence different enough from the first. Since it cannot be done in the most general case it means it cannot be done at all.

**b**

Much like the previous question, without loss of generality we split our set of possible words into 8. With each subset containing the sequences that share a common beginning.

$$S = S_{000} \cup S_{001} \cup S_{010} \cup S_{011} \dots$$

Since there are 17 code words one of the subsets has at least 3 sequences. Without loss of generality we'll set  $S_{101}$  as this set. We'll denote them as,

$$101a_1b_1c_1d_1$$

$$101a_2b_2c_2d_2$$

$$101a_3b_3c_3d_3$$

Where each letter represents a digit in each word. Now each word must be distance 3 apart. Since the order of the digits doesn't matter we'll arbitrarily set  $a_1 \neq a_2$ ,  $b_1 \neq b_2$ ,  $c_1 \neq c_2$  without losing generality. Now at least 2 of  $a_3$ ,  $b_3$ ,  $c_3$  must not be equal to their  $a_2$ ,  $b_2$ ,  $c_2$  counterparts to maintain a distance of 3. Without loss of generality we will take  $a_3$  and  $b_3$  as these values. But this will mean  $a_3 = a_1$  and  $b_3 = b_1$ . If  $a_3 = a_1$  and  $b_3 = b_1$  then there aren't enough digits to make the third sequence different enough from the first. Since it cannot be done in the most general case it means it cannot be done at all.

**c**

Let the number of words be  $N$  where  $N > \frac{2^n}{n+1}$ . Each word will contain  $n$  binary letters meaning at most we can have  $2^n$  possible words.

Let each sequence in our code words be represented as  $A_i$  where  $i$  goes from 1 to  $N$ . Now consider all sequences a distance 1 from an  $A_i$ . Create a set for  $A_i$  with itself and all points a distance 1 away surrounding it,  $\beta_i = A_i \cup$

{All other sequences a distance 1 away}. This set contains  $1 + n$  elements because an error can occur in any of the  $n$  bits in our sequences and we're also adding the  $A_i$  as an element.

Since each of the  $A_i$  are a distance 3 away from each other it makes sense that  $B_i \cap B_j = \emptyset$  by the axioms of a metric. To prove this, consider a point  $x \in B_i$  but  $x \neq A_i$ . By definition this point is 1 away from  $A_i$ . If it was in another set  $B_j$  then it would also be at most a distance 1 from  $A_j$ . But that would mean

$$\begin{aligned} |x - A_i| + |x - A_j| &\geq |A_i - A_j| \\ 1 + 1 &\geq |A_i - A_j| \\ 2 &\geq |A_i - A_j| \end{aligned}$$

Which immediately breaks the initial requirement for a distance 3 apart. Thus we have  $N$  non intersecting sets with  $n + 1$  elements each. Thus our elements come up to  $N \times (n + 1)$ .

$$\begin{aligned} N &> \frac{2^n}{n + 1} \\ N \times (n + 1) &> 2^n \end{aligned}$$

But the set of all possible  $n$  length binary sequences can only come up to  $2^n$  thus we have a contradiction. It is not possible to have a code with more than  $\frac{2^n}{n+1}$  codes with a distance 3.