Name: ______________________________

UW Math Circle

Week 23 – Information Theory

1 Error correction

Below is a list of eight words that we will use.

ban	bat	ben	bet	pan	pat	pen	pet

1. Secretly choose one of the eight words above, then write a message on a note to communicate this word to your other group members. However, you must first pass your note to your instructor who will copy your message, possibly changing one letter, then your group members will only see this changed note.

   (a) Discuss a strategy with your group on how to accurately communicate any of the above words. Try out your strategy!

   (b) What if you are only allowed to use the letters "b,p,a,e,n,t" and no spaces in your message?

2. Next we will code the eight words above into 0's and 1's using the following table.

Letter	b	p	a	e	n	t
Bit	0	1	0	1	0	1

For example, "bet" would be "011". A sequence of 0's and 1's is called a *binary string*.

(a) Code the following words into binary strings: "bat", "pen", "pet".

(b) Decode the following binary strings into words: "001", "010", "100".

3. Again you must communicate a word from above to your group members as in Problem 1. However, this time you may only write a binary string of some fixed length. If your instructor is allowed to possibly *flip a bit* (change a 0 to a 1 or change a 1 to a 0), how can you communicate your word to your group? Try to make the length of your binary string as small as possible.

4. Codes that still allow you to communicate information even if a bit is flipped are called *error correcting codes*. One method is a *repetition code* in which you simply repeat each 0 and 1 three times. In this case, 0101 becomes 000111000111 and 0010 becomes 000000111000. Then you send this longer string.

   (a) If you want to communicate the message 1100, what longer string should you send?

   (b) If you receive the message 111000011000 and you know at most one bit was flipped, what was the original message?

   (c) Is it always possible to recover the message if you know at most one bit was flipped? What if up to two bits might be flipped? Why or why not?

5. Problem 4 provides an error correcting code that communicates a length 4 binary string by sending a length 12 binary string. Can you find an error correcting code that sends messages of a length less than 12 in which you can still recover the message if a bit is flipped? How small can you make this length?

2 Data compression

Computers store and send information in 0's and 1's, so we often need to convert words and other data into a binary string. Data compression is the study of systems that do this while keeping the binary string as short as possible.

6. You have a list of words that only use the letters "a,b,c,d" that you need to convert into binary strings using the following conversion rule.

Letter	a	b	c	d
Binary string	0	10	110	111

For example, "add" becomes "0111111".

(a) Convert the following words into binary strings "cab", "dad", "bad'.

(b) Decode the following binary strings into words "0111", "01010", "111010".

(c) In order to shorten the coded binary strings, you try the following new rule:
a $\mapsto$ 0, b $\mapsto$ 10, c $\mapsto$ 11, d $\mapsto$ 01.
Can you code and decode every word without ambiguity? Why or why not?

(d) Can you code and decode every word without ambiguity when using the original conversion rule? Why or why not?

7. We say the binary string 1011 is a *prefix* of the longer string 1011011 because it appears at the beginning of the longer string. The previous coding method is called a *prefix code* because this method only works when no binary string in the conversion table is a prefix of another (challenge: why?).

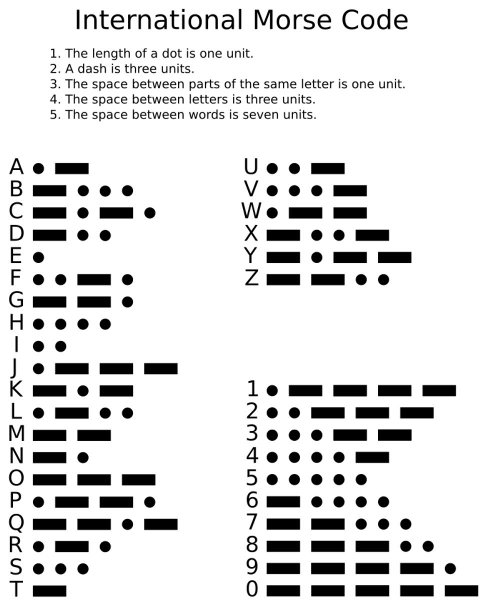
   (a) Suppose you come across an alien language that uses only the characters "◯, △, □". How can you fill out the below table to make a prefix code for words in this language?

Character	◯	△	□
Binary string			

   (b) Add up the lengths of all the strings in your table. Is there a coding scheme with a smaller sum of lengths? If so, what is it? If not, why not?

   (c) You look at a few alien words such as "△△ ◯ △△" and "□△ ◯ △△△◯" and you notice the character "△" appears more than the other two. How should you match your solution from (7b) to these characters so that when you convert these words into binary strings, the strings are as short as possible?

8. Morse code is a way of sending messages with a sequence of dots "•" and dashes "-" using the below chart. To send Morse code with a flashlight, you could leave the flashlight



ON for 1 second to communicate a "•", and leave the flashlight ON for 3 seconds to communicate a "-". In between a "•" and a "-" in the same letter, you would leave the flashlight OFF for 1 second. In between letters, you stay OFF for 3 seconds (so you know when a letter ends) and in between words, you stay OFF for 7 seconds (so you know when a word ends). Decode the following:

- -   • -   -   • • • •   • -   -   • • -   • - -

Debate with your group: is Morse code an example of a prefix code? If so, what are the characters? If not, why?

9. In problem 6 you studied a prefix code using strings of length $\{1, 2, 3, 3\}$ and in problem 7 you found a coding method using strings of length $\{1, 2, 2\}$.

   (a) Is it possible to find a prefix code using strings of length $\{1, 2, 2, 2\}$ like we tried to do in problem 6c?

   (b) In general, what sets of lengths are possible to find prefix codes for?

3 Information

10. Let's play the following number guessing game:

    Step 1: Your instructor will choose a secret number from 1 to 16.

    Step 2: Your group will guess a number and your instructor will tell you if the secret number is "higher" or the secret number is "not higher" (so even if you guess the exact secret number, your instructor will say "not higher").

    Step 3: Continue guessing numbers until you are sure what the secret number is.

    Discuss with your group what you think the best strategy is. What do you think makes a strategy "good" or the "best"?

11. Let's study the same guessing game, but now the secret number can only be from 1 to 4.

    (a) If you guess "2", how many possibilities remain if the instructor says "higher"? What if the instructor says "not higher"? In each case, how many more guesses do you need to know the secret number?

    (b) If you guess "1", how many possibilities remain if the instructor says "higher" and if they say "not higher"? In each case, how many more guesses do you need to know the secret number? What is the average number of guesses you need over all possible secret numbers?

12. When your instructor says "higher" or "not higher" you get information. Each time the possibilities are halved, we say you get *1 bit* of information. For example:

    - If in the 1-16 game you guess 8 and your instructor says "higher", the number of possibilities is multiplied by 1/2 and so you get 1 bit of information.
    - If in the 1-16 game you guess 4 and your instructor says "not higher", the number of possibilities is multiplied by 1/4 and you get 2 bits of information.

(a) How many bits of information do you get if you guess 14 in the 1-16 game and your instructor says “higher”?

(b) How many bits of information do you get if you guess 1 in the 1-16 game and your instructor says “not higher”?

13. In general, if the number of possibilities is multiplied by the fraction p, then you get I bits of information if $2^I = 1/p$. The solution to this equation is written as $I = \log_2(1/p)$. For example: $\log_2(8) = 3$ because $2^3 = 8$.

    (a) What is $\log_2(16)$?

    (b) What is $\log_2(2)$?

14. Now let’s study the 1-4 guessing game again.

    (a) How many bits of information do you get from your first guess if you guess “2”?

    (b) How many bits of information do you get from your first guess if you guess “1” for each possible secret number?

    (c) How many bits of information do you get from a first guess of “1” if you average over all possible secret numbers? (use the approximation $\log_2(1/(3/4)) \approx 0.4$)). Note: this average information is often called *entropy* and making this higher makes for a good guessing strategy.

15. What strategy in the number guessing game gives the highest entropy for each guess?