**How Hard Can It Be?**

The FBI intercepted a message containing 10 numbers they think is a phone number from the Pacific Time zone. Don asks Charlie if it is feasible to break a phone number code using a computer. Charlie replies that it will take time but suggests that expected entropy can give them some indication of the difficulty involved. The expected entropy of a code is its level of “difficulty” to break. More characters or numbers in a code give it more possibilities requiring more time to break it.

**1.** To calculate the **expected entropy**, you must know the probability of each digit appearing. For the last four digits, the probability for selecting any of the digits 0–9 is equally likely. What is the probability of selecting 1 digit from 0–9?

**2.** The **entropy** uses logarithms with a base of 2. In general the expected entropy for each number appearing is

****

where *P*(*n*) is the probability of the digit appearing.

What expression would represent the expected entropy for a 5 appearing?

**3.** What is the expected entropy for a 5 appearing?

**4.** To find the expected entropy for the encryption used on the last four digits of the phone number, add the expected entropies for each of the numbers 0–9. What is the expected entropy for the last four digits?

**5.** The expected entropy for the area code (first three digits in the phone number) will be different than that of the last four because not all the numbers 0–9 appear in area code with the same probability. For example, of the 36 area codes in the Pacific time zone, the probability of a 4 appearing is about 0.06. Based on this probability, what is the expected entropy for a 4 appearing in the area code?

**6.** The table below shows the probability for each number appearing in the 36 area codes. Determine the expected entropy for the encryption of the area code.

 

**7.** If you want to break the code for a seven-digit telephone number, would you start with the area code or the last four digits? Explain.

The FBI agents intercepted a second message they think is indicating where the serial poisoner and the “insider” are going to meet. Unfortunately, the message is encoded and there is no indication what the key to the code might be.

**8.** Assuming all the letters are equally likely to appear in the message, what is the expected entropy for the code used to create the written messages?

**9.** Because of the structure of words and sentences in English, the probability of each letter appearing is not equal. The table below shows the approximate probability for each letter and a space appearing in the English language. Determine the expected entropy of the letters based on these probabilities. (Hint: Use the list feature on your calculator to do the calculations more efficiently.)



**10.** If punctuation marks were encoded and added to the table above, would you expect the new entropy to be greater or less than the entropy found for 9? Explain.

**11.** Based on the entropy from Parts 8 and 9, which code would be the more difficult to decipher?

You may wish to answer all the questions in a **spreadsheet tab**; this will make it easy for you to make the calculations (especially the log to base 2 calculations).

**You may hand in a printout of the spreadsheet tab.**