ECE563 Fall 2016 **Homework 1**

Date Assigned: 23 August 2016.
Date Due: 1 September 2016 in class.
Suggested Reading: Chapter 2.1 of Cover-Thomas and Part I of Claude E. Shannon's seminal paper "A mathematical theory of communication", Bell Systems Technical Journal, October 1948.
Thanks: Prof. Sergio Verdú.

1. One card is drawn from each of k decks of 52 cards.
   (a) Show that the maximum information provided by the unordered collection of k drawn cards is equal to $k \log 52$.
   (b) Suppose that k is a multiple of 52. Using Stirling's approximation show that the information of an outcome where each card appears the same number of times is $\frac{51}{2} \log k + O(1)$ as $k \to \infty$.

2. Each year some 35 million cows, 150 million pigs, 275 million turkeys and 9 billion chickens are slaughtered in the United States. A numerical answer is expected for the following questions; simplify as much as you can.
   (a) Assuming that the slaughtering of each of the species is not seasonal, estimate the information provided by the fact that the first animal species (among those four) to be slaughtered on April 30th is a pig.
   (b) How much information is provided by the sequence in which the animals are slaughtered?

3. A randomization process was used in ancient Greece to select a jury of 500 citizens out of a pool of 5500 candidates. The citizens were randomly divided in 10 groups of 550, and each group followed the following procedure. Tokens identifying each citizen were inserted in the next available slot in a 50×11 stone matrix (the kleroterion). Balls were extracted from a tube containing ten white balls and one black ball in randomized order. If the black ball appeared in the kth draw, the citizens in the kth column were selected for jury duty and the other 500 were dismissed.
   (a) Find the information in bits provided by the composition of the jury.
   (b) Find the probability that the answer to a) is strictly lower than what it would have been had each candidate drawn a ball from the tube independently. Note that this alternative selection process is not only much more painstaking but it results in a random jury size.

4. In the social network Twitter, messages can have up to 140 characters drawn from the alphabet of 94 visible characters (47 two-character keys in the standard US keyboard) plus return and space. What is the base of the logarithm that would justify calling the corresponding information unit a twit?

5. Let $L = \sum_{i=1}^{m} p_i \ell_i^{100}$ be the expected value of the 100$^{\text{th}}$ power of the word lengths associated with an encoding of the random variable X. Let $L_1 = \min L$ over all instantaneously decodable codes and let $L_2 = \min L$ over all uniquely decodable codes. What inequality relationship exists between L_1 and L_2?

6. Consider a random variable X with pmf $\mathbf{p} = (p_1, \ldots, p_n)$. In this question we state three axioms that capture what one might expect from a measure of information gained from knowing X. Then we derive an expression for the information measure H from these axioms. The three axioms are the following.

   (a) H should be continuous in $\mathbf{p}$.

   (b) If all the p_i are equal, then H should be monotonic increasing function of n.

   (c) If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. For instance $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + 0.5H(\frac{2}{3}, \frac{1}{3})$.

   Show that the only function H satisfying the above three properties is of the form $H = -K \sum_{i=1}^{n} p_i \log p_i$ for some positive constant K.

   Hint: Show that the only strictly monotonically increasing real-valued function with the property $\phi(k^m) = m\phi(k)$ for all positive integers m and k is the function $\phi(m) = \log m$ with an arbitrary base for the logarithm.

   References: You can see how Claude Shannon proved this theorem in his seminal paper, cited above. In Shannon's words, these axioms are "in no way necessary for the theory" but "lend a certain plausibility" to the definition of entropy and related information measures. A more recent, but also classical, reference is: I. Csiszár, Axiomatic characterizations of information measures, *Entropy*, 2008.

7. **Optional**: Consider two probability vectors $\mathbf{p}$ and $\mathbf{q}$ of length n. We say that the vector $\mathbf{p}$ *majorizes* the vector $\mathbf{q}$ if $\sum_{i=1}^{k} p_{[i]} \geq \sum_{i=1}^{k} q_{[i]}$ for every $k = 1 \ldots n-1$. Here $p_{[i]}$ is the i^{th} largest element of the vector $\mathbf{p}$. The intuition is that the vector $\mathbf{q}$ is "more random" than the vector $\mathbf{p}$. Show that $H(\mathbf{p}) \leq H(\mathbf{q})$.

   References: Majorization is covered in depth in a classic book by Marshall and Olkin as well as much less comprehensively in Chapter 3.2 of Horn and Johnson's book titled Topics in Matrix Analysis.