COSC 6222 :: Test #1

Instructions:

1. Attempt all questions. All questions have equal weight.
2. Consulting with other people – whether students of the class or not – is **prohibited**.
3. All resources are allowed (so long as they satisfy point 2). However, for each question, **you must declare all resources you consulted.** If you consulted books (including the text), give the titles; if you consulted websites, give the URLs.
4. **Due date:** Monday, March 1, 11:59 PM. Submit by email to: aeckford@yorku.ca

1. Let p(x,y) represent a joint probability of the binary random variables x and y, where p(0,0) = 2/5, p(1,0) = 2/5, p(0,1) = 1/5, and p(x,y) = 0 otherwise. Calculate:
   a. H(X)
   b. H(X|Y)
   c. H(X,Y)
   d. The Huffman code for the pair (X,Y), and its average length.

2. Let f(x,y) and g(x,y) represent joint distributions on x and y. Let f(y|x), f(x|y), and f(y) represent the appropriate conditional and marginal densities obtained from f(x,y); and similarly for g(y|x),g(x|y), and g(y). Show that:

   $$D(f(y \mid x) \,\|\, g(y \mid x)) \le D(f(x \mid y) \,\|\, g(x \mid y)) + D(f(y) \,\|\, g(y)),$$

   and give the condition under which equality is achieved.

3. Let {A, B} represent a set of symbols for a source X, where Pr(A) = 0.8 and Pr(B) = 0.2.
   a. For this source, what is the minimum average length of any uniquely decodable code using an encoding alphabet with k symbols?
   b. Calculate the quantity in part a for k = 2 (i.e., binary) and k = 3 (i.e., ternary).
   c. Give the binary arithmetic code for AAAABA.

4. Let $X = (X_1, X_2, \ldots, X_n)$ represent a sequence of iid random variables. Let C represent the set of all sequences X such that the probability of X is greater than p^*, i.e.,

   $$C = \{X : p(X) \ge p^*\}.$$

   a. Show that $|C| \le 1/p^*$.
   b. For what values of p^* is it true that $\Pr(X \in C) \to 1$ as $n \to \infty$?

5. Consider the random variable X, where X = 0 if the sun does not rise in the morning, and X = 1 if the sun does rise in the morning. In all of recorded history, which encompasses 2.2×10^6 mornings, the sun has risen every morning.
   a. Give the universal probability $q(x_{i+1}|x^i)$ of the sun rising tomorrow (or not), conditioned on the observation that the sun has already risen every morning in recorded history.
   b. Given your answer for part a, what is the optimal encoding length for the observation that the sun rises tomorrow? What is the optimal length if the sun does not rise?