201700080 Information Theory and Statistics
11 April 2019, 8:45 - 11:45

This test consists of 6 problems for a total of 31 points. All answers need to be justified. The use of a non-programmable calculator (not a "GR") is allowed and advised. No books, notes, or other materials may be used.

Formulas you may find useful:

$$D(P \| Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$$
$$H(X, Y \mid Z) = H(X \mid Z) + H(Y \mid X, Z)$$
$$I(X ; Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z)$$

1. [2 pt] Let X, Y and Z be jointly distributed random variables. Prove the following inequality and find conditions for equality:
$$H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X).$$

2. Consider a source with alphabet $\mathcal{X} = \{a, b, c, d, e, f, g\}$ and $p(a) = p(b) = p(c) = 0.1$, $p(d) = p(e) = p(f) = 0.15$, $p(g) = 0.25$. You want to compress this source and store the compressed data onto a device that is using ternary storage. This means that you will compress into the alphabet $\{0, 1, 2\}$.
   a. [1 pt] Explain the property 'instantaneous decodability'.
   b. [1 pt] Consider the code C for which $C(a) = \mathbf{0}$, $C(b) = \mathbf{1}$, $C(c) = \mathbf{20}$, $C(d) = \mathbf{21}$, $C(e) = \mathbf{22}$ and $C(f) = \mathbf{200}$ and $C(g) = \mathbf{201}$. Is C instantaneously decodable?
   c. [1 pt] Compute the entropy $H(X)$ of the source.
   d. [2 pt] Construct a ternary Huffman code for the source. Denote this code by C' and let L' denote its codewords lengths. Compute $\mathbb{E}[L']$.
   e. [2 pt] Give two different arguments for why it is not possible to construct a prefix code for this source for which the expected codeword length is 1.7. Formulate as precisely as possible the theory that you are using.

3. A fountain code is used to communicate bits m_1, m_2, m_3. You receive the encoded symbols $m_1 + m_2 = 0$, $m_1 = 0$, $m_1 + m_2 + m_3 = 1$ and $m_2 + m_3 = 1$.
   a. [1 pt] Construct a bipartite graph that represents the information in the encoded symbols.

b. [3 pt] Use belief propagation to decode m_1, m_2, m_3. Explicitly show all steps involved. *(You do not receive points for the values of m_1, m_2, m_3, only for the correct procedure.)*

4. Consider the channel with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, with $P(Y = 1 \mid X = 1) = P(Y = 0 \mid X = 0) = 1 - e$.

   a. [2 pt] Give the definition of channel capacity and state the noisy channel coding theorem.

   b. [3 pt] Find the channel capacity for this channel as a function of e. Express your answer using the binary entropy function.

5. You receive a sample of n observations $x_1, x_2, \ldots, x_n$, that are independent and identically distributed according to a geometric distribution for which the success probability p is unknown. It is known that either $p = p_1$ or $p = p_2$, i.e. the distribution for a single observation is either $P_1(X = x) = (1 - p_1)^{x-1} p_1$ or $P_2(X = x) = (1 - p_2)^{x-1} p_2$, $x = 1, 2, \ldots$.

   a. [1 pt] Give the general formulation of a binary hypothesis testing problem. *(Specify the hypotheses, the decision rule and the possible errors.)*

   b. [2 pt] Formulate a hypothesis test and a Neyman-Pearson optimal decision rule for deciding on the value of p. *(Hint: Express the decision rule in terms of the log-likelihood ratio and write it as explicitly as possible in terms of the observed values $x_1, x_2, \ldots, x_n$.)*

   c. [2 pt] Compute $D(P_1 \| P_2)$. *(Hint: $\mathbb{E}[X] = 1/p$)*

   d. [2 pt] State the Chernoff-Stein result and discuss the implications for this problem.

6. Let X be continuous random variable for which the probability density function $f_\theta(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}$ for $x \geq 0$ and $f_\theta(x) = 0$ for $x < 0$. Here θ is an unknown parameter. For this random variable it is known that $\mathbb{E}[X^2] = 2\theta$ and $\mathrm{Var}[X^2] = 4\theta^2$.

   After observing n independent realizations of this random variable you want to estimate θ. One possible way to estimate θ is to use $\tilde{\theta} = \frac{1}{2n} \sum_{i=1}^{n} x_i^2$.

   a. [2 pt] Compute the bias and the variance of $\tilde{\theta}$.

   b. [2 pt] Assume that θ is unknown and compute the Fisher information $J(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial^2 \theta} \ln f_\theta(X)\right]$ for a single observation.

   c. [2 pt] State the Cramer-Rao bound and discuss in your own words the implications of the results from a) and b). *(If you did not solve a) you may assume that $\tilde{\theta}$ is unbiased and that $\mathrm{Var}[\tilde{\theta}] = 2\theta^2/n$. If you did not solve b) you may assume $J(\theta) = 1/\theta^2$)*