201700080 Information Theory and Statistics
9 April 2020, 8:45 - 11:45

This test consists of 7 problems for a total of 27 points. All answers need to be justified: Show details of derivations and state which existing results you are using. During the exam you are allowed to use:

- Book: *"Information Theory, Inference and Learning Algorithms"*, by Mackay,
- Your own notes,
- All information and linked material that is on the Canvas site of the course,
- A non-graphical, non-programmable, calculator.

1. Please read the following paragraph carefully, and copy the text below it verbatim to your answer sheet. To find more information, please consult the remote written assessment rules. By testing you remotely in this fashion, we express our trust that you will adhere to the ethical standard of behaviour expected of you. This means that we trust you to answer the questions and perform the assignments in this test to the best of your own ability, without seeking or accepting the help of any source that is not explicitly allowed by the conditions of this test.

   Text to be copied:

   I will make this test to the best of my own ability, without seeking or accepting the help of any source not explicitly allowed by the conditions of the test.

2. Let X and Y be random variables that take values $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$, respectively. Let $Z = X + Y$.

   a. [2 pt] Prove that $H(Z \mid X) = H(Y \mid X)$.

   b. [2 pt] Prove that, if X and Y are independent, that $H(Z) \geq H(X)$ and $H(Z) \geq H(Y)$.

3. Let (X_i, Y_i) be i.i.d. $\sim p(x, y)$. We form the log likelihood ratio of the hypothesis that X and Y are independent vs. the hypothesis that X and Y are dependent as

$$\frac{1}{n} \log \frac{p(X^n)p(Y^n)}{p(X^n, Y^n)}$$

   In this exercise we consider the limit $n \to \infty$.

   a. [1 pt] Which result from probability theory can be used to analyze $\lim_{n\to\infty} \frac{1}{n} \log \frac{p(X^n)p(Y^n)}{p(X^n,Y^n)}$? Why?

   b. [1 pt] Express this limit in terms of (an) information measure(s) that we have seen in class.

4. Consider data compression of a source with alphabet $\mathcal{X} = \{a, b, c, d\}$ and $p(a) = 0.4$ and $p(b) = p(c) = p(d) = 0.2$.

   a. [1 pt] Explain the property 'instantaneous decodability' and why it is important.

   b. [2 pt] Use the dictionary-based Lempel-Ziv algorithm to encode the sequence

   $$c\ a\ a\ a\ a\ d\ c\ a\ d\ a\ b\ a\ c\ c\ a$$

   Give the encoded representation as a list of pairs (i, x), with i an integer and $x \in \mathcal{X}$ (there is no need to describe the bit-level representation).

   e. [2 pt] If we encode a sequence of length n and consider $n \to \infty$, how many bits per source symbol will the Lempel-Ziv algorithm produce?

5. Consider the channel $Y = XZ$, where X and Z are independent random variables with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $P(Z = 1) = q$.

   a. [2 pt] Let $P(X = x) = p$. Find $I(X\,;Y)$.

   b. [3 pt] Prove that *(Hint: $\frac{d}{dx}H_2(x) = \log\left(\frac{1-x}{x}\right)$.)*

   $$C = \log\left(2^{\frac{H_2(q)}{q}} + 1\right) - \frac{H_2(q)}{q}.$$

   c. [3 pt] Describe in at most 50 words (but as precisely as possible), what the role of C is in a communication system.

6. A manufacturer of dice has announced that there have been production problems with some of its dice. The dice with errors have a 1/5 probability of throwing 6 and 4/25 probability of another number. You have a dice of this manufacturer at home and you want to test if it has this production error. You observe n rolls of the dice for this test.

   a. [1 pt] Formulate the corresponding binary hypothesis testing problem and a class of optimal decision rules. *(Specify the hypotheses, the decision rule and the possible errors.)*

   b. [3 pt] Use Le Cam's theorem and Pinsker's inequality to establish a lower bound on the total error probability $\alpha + \beta$.

7. Let $X_1, X_2, \ldots, X_n$ be i.i.d. normally distributed random variables with mean zero and variance θ.

   a. [3 pt] Compute the Fisher information $J_n(\theta)$. *(Hint: For $X \sim \mathcal{N}(0, \theta)$, $\mathbb{E}[X^4] = 3\theta^2$.)*

   b. [1 pt] Give a lower bound on the variance of an unbiased estimator of θ. (Explain)