201700080 Information Theory and Statistics
19 April 2022, 13:45 – 16:45

This test consists of 5 problems for a total of 36 points. All answers need to be justified. The use of a non-programmable calculator (not a "GR") is allowed. A handwritten single side A4 cheat sheet is allowed. No additional books or notes may be used.

1. The centered Laplace distribution is a continuous distribution with pdf

$$f_b(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right),$$

where $b > 0$ is a scale parameter and $x \in \mathbb{R}$.

You receive a sample of n observations $x_1, x_2, \dots, x_n$, that are independent and identically distributed according to a Laplace distribution for which the scale b is unknown. It is known that either $b = b_1$ or $b = b_2$, with $b_1 > b_2$.

In this exercise you are going to work on the binary hypothesis testing problem for choosing between b_1 and b_2. Therefore, let P_1 and P_2 be the continuous probability distributions with densities f_{b_1} and f_{b_2}, respectively.

a. [4 pt] Specify a binary hypothesis testing problem for choosing between b_1 and b_2. Derive an optimal decision rule.

b. [2 pt] Compute $D(P_1 \| P_2)$. *(Hint: If $X \sim \mathrm{Laplace}(b)$ then $|X| \sim \mathrm{Exp}(b^{-1})$.)*

c. [2 pt] State the Chernoff-Stein result. Make sure to define and explain the variables and quantities that are involved.

2. Let X_1 and X_2 be identically distributed discrete random variables. They are not necessarily independent. Let

$$\rho = \frac{I(X_1; X_2)}{H(X_1)}.$$

a. [2 pt] Show that $\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}$. *(Hint: You need to use the fact that X_1 and X_2 are identically distributed.)*

b. [2 pt] Show that $0 \leq \rho \leq 1$.

c. [2 pt] When is $\rho = 0$? In addition to giving a mathematical expression, explain in words as simple as possible.

d. [2 pt] When is $\rho = 1$? In addition to giving a mathematical expression, explain in words as simple as possible.

P.T.O. (Please turn over)

3. 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. [3 pt] Explain the property 'instantaneous decodability' and why it is important.

   b. [3 pt] Use a Lempel-Ziv algorithm to encode the sequence

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

   You may:

   1. use a dictionary-based Lempel-Ziv algorithm (as in the book of Mackay) and give the encoded representation as a list of pairs (i, x), with i an integer and $x \in \mathcal{X}$, *or*
   2. use a sliding-window Lempel-Ziv algorithm (as shown in, for instance, the lecture slides) and give the encoded representation as a list of triple (i, k, x), with i and k integers and $x \in \mathcal{X}$.

4. Consider the channel with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $P(Y = 0|X = 0) = 1$ and $P(Y = 0|X = 1) = \frac{1}{2}$. In this exercise, use the natural logarithm in all expressions.

   a. [3 pt] Let $P(X = 1) = p$. Give expressions for $H(Y)$ and $H(Y \mid X)$.

   b. [3 pt] Compute the capacity of the the channel. Give a numerical answer with two decimals precision. *(Hint: the derivative of* $-\frac{1}{2}x\log(\frac{1}{2}x) - (1 - \frac{1}{2}x)\log(1 - \frac{1}{2}x)$ *is equal to* $\frac{1}{2}\log(\frac{2-x}{x})$*)*

5. Let X be an exponentially distributed random variable with rate parameter $\theta > 0$, i.e. $f_\theta(x) = \theta e^{-\theta x}$, $x \geq 0$. Consider the estimator $\hat{\theta}_n = \frac{n}{\sum_{j=1}^n x_j}$, for estimating θ based on observations $x_1, \ldots, x_n$. It is known (and you can use these facts in your solutions) that this estimator is unbiased and efficient.

   a. [2 pt] Compute $\mathbb{E}[\hat{\theta}_n]$.

   b. [3 pt] Compute the Fisher information $J(\theta)$.

   c. [3 pt] Compute $\mathrm{var}(\hat{\theta}_n)$. State theoretical results that you are using and explain.