Solutions to sample exam

201700080 Information Theory and Statistics, 2024–2025
Jasper Goseling & Maike de Jongh

---

1. Consider a Pareto distribution with density function

$$f_\theta(x) = \theta x_0^\theta x^{-\theta-1}, \quad x \geq x_0, \quad \theta > 1.$$

Assume that $x_0 > 0$ is given, but θ is unknown.

a. [3 pt] Compute the Fisher information $J(\theta)$.
Solution:

$$\begin{aligned}
\ln f_\theta(x) &= \ln\theta + \theta \ln x_0 - (\theta+1)\ln x, \\
\frac{\partial}{\partial\theta}\ln f_\theta(x) &= \frac{1}{\theta} + \ln x_0 - \ln x \\
\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) &= -\frac{1}{\theta^2} \\
J(\theta) &= \frac{1}{\theta^2}
\end{aligned}$$

b. [2pt] State the Cramer-Rao bound and discuss in your own words the implications of the result from a) on unbiased estimators of θ based on n observations.
Solution:
Cramer-Rao bound: if $\hat{\theta}(X_1, ..., X_n)$ is an unbiased estimator of θ, then

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{J_n(\theta)}.$$

Using the result from a), we obtain

$$J_n(\theta) = nJ(\theta) = \frac{n}{\theta^2},$$

Thus, any unbiased estimator $\hat{\theta}(X_1, ..., X_n)$ satisfies

$$\mathrm{var}(\hat{\theta}) \geq \frac{\theta^2}{n}.$$

2. Every evening, the University canteen serves either pasta, curry or fish & chips. The cook claims he picks one of the three recipes uniformly at random each night. However, students have long suspected that he is actually selecting recipes according to the distribution $P(\text{pasta}) = \frac{2}{3}$, $P(\text{curry}) = \frac{1}{6}$ and $P(\text{fish \& chips}) = \frac{1}{6}$.

a. [2pt] Give the log-likelihood function of the presumed distribution and the distribution claimed by the cook.

Solution:

Let P_1 and P_2 denote the distribution claimed by the cook and the presumed distribution respectively, i.e. $P_1(\text{pasta}) = P_1(\text{curry}) = P_1(\text{fish \& chips}) = \frac{1}{3}$ and $P_2(\text{pasta}) = \frac{2}{3}$, $P_2(\text{curry}) = P_2(\text{fish \& chips}) = \frac{1}{6}$. The log-likelihood function is given by

$$L(x) = \log \Lambda(x) = \log \frac{P_1(x)}{P_2(x)} = \begin{cases} \log(\frac{1}{2}) = -1 & \text{if } x = \text{pasta}, \\ \log(2) = 1 & \text{if } x = \text{curry}, \\ \log(2) = 1 & \text{if } x = \text{fish \& chips}. \end{cases}$$

b. [2pt] Philip decides to test the hypothesis by eating in the canteen for n evenings in a row and recording the served meals. Formulate a binary hypothesis testing problem for choosing between the two distributions.

Solution:

Let $a(x)$ denote the total number of times Philip was served meal x in the canteen, where $x = \text{pasta}, \text{curry}, \text{fish \& chips}$. Let $H_1 : Q = P_1$ and $H_2 : Q = P_2$. We have

$$L(x) = \frac{1}{n}\sum_{i=1}^{n} L(x_i) = \frac{a(\text{curry}) + a(\text{fish\& chips}) - a(\text{pasta})}{n}.$$

Now, define the following decision rule for choosing between the two distributions:

$$\Psi_t^* = \begin{cases} 1 & \text{if } L(x) > t \\ 2 & \text{if } L(x) \leq t, \end{cases}$$

with $t \in \mathbb{R}$.

c. [3pt] Verify Pinsker's inequality for the two distributions.

Solution:

Pinsker's inequality states:

$$||P_1 - P_2||_{\text{TV}} \leq \sqrt{\frac{1}{2\log e} D(P_1||P_2)}.$$

We have

$$||P_1 - P_2||_{\text{TV}} = \max_{A \in \mathcal{X}} |P(A) - Q(A)| = \frac{1}{3},$$

where $\mathcal{X} = \{\text{pasta}, \text{curry}, \text{fish \& chips}\}$. Also,

$$D(P_1||P_2) = \sum_{x \in \mathcal{X}} P_1(x) \log \frac{P_1(x)}{P_2(x)} = \frac{1}{3}.$$

Since

$$\sqrt{\frac{1}{2\log e}D(P_1||P_2)} \approx 0.3399 \geq \frac{1}{3} = ||P_1 - P_2||_{\mathrm{TV}},$$

Pinsker's inequality is satisfied for these two distributions.

3. [3pt] Let X, Y and Z be jointly distributed random variables. Prove the following inequality and find conditions for equality:

$$H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X).$$

Solution:

$$\begin{aligned} H(X,Y,Z) - H(X,Y) &= H(X,Z) + H(Y|X,Z) - H(X) - H(Y|X) \\ &= H(X,Z) - H(X) + H(Y|X,Z) - H(Y|X) \\ &\leq H(X,Z) - H(X), \end{aligned} \tag{1}$$

Since $H(Y|X) \geq H(Y|X,Z)$. Condition for equality is $H(Y|X) = H(Y|X,Z)$, or Y conditioned on X is independent of Z (i.e. $Z \to X \to Y$ is a Markov chain).

4. Consider a source with alphabet $\mathcal{X} = \{a,b,c,d,e\}$ and $p(a) = 0.3$, $p(b) = 0.2$, $p(c) = 0.15$, $p(d) = 0.2$ and $p(e) = 0.15$. Consider the code C for which $C(a) = \texttt{00}$, $C(b) = \texttt{01}$, $C(c) = \texttt{101}$, $C(d) = \texttt{11}$ and $C(e) = \texttt{1001}$.

   a. [1 pt] Use code C to encode the message *aacade* and give the resulting bitstring.
   Solution:
   `000010100111001`.

   b. [2 pt] Is C a prefix code? Is it a uniquely decodable code? Justify your answers.
   Solution:
   Yes and yes.

   c. [1 pt] Compute the expected codeword length $\mathbb{E}[L]$ for C.
   Solution:
   $\mathbb{E}[L] = 2.45$.

   d. [1 pt] Compute the entropy $H(X)$ of the source.
   Solution:
   $H(X) \approx 2.27$.

   e. [3 pt] Construct a Huffman code for the source. Denote this code by C' and let L' denote its codeword length. Compute $\mathbb{E}[L']$.
   Solution:
   $C'(a) = \texttt{10}$, $C'(b) = \texttt{00}$, $C'(c) = \texttt{110}$, $C'(d) = \texttt{01}$, $C'(e) = \texttt{111}$. We have $\mathbb{E}[L'] = 2.3$.

f. [1 pt] Is it possible to construct a prefix code for the source for which the expected codword length is 2.2? Explain.

Solution:
No.

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

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

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

Solution:
$C = \max_{P_X} I(X; Y)$, where $I(X; Y) = H(Y) - H(Y|X)$. Let $P(X = 1) = p$. We have

$$H(Y|X) = pH(Y|X = 1) + (1 - p)H(Y|X = 0) = H_2(e).$$

Hence,

$$C = 1 - H_2(e).$$

6. A message from a source with $\mathcal{X} = \{\texttt{a}, \texttt{b}, ..., \texttt{z}\}$ is encoded using the sliding-window based Lempel-Ziv code. The message is

`aabacabracadaaa`

a. [2 pt] Encode the message and give the resulting sequence of (o, ℓ, c) triples (no need to give the bit-level representation).

Solution:
$(-, -, \texttt{a})$, $(1, 1, \texttt{b})$, $(2, 1, \texttt{c})$, $(4, 2, \texttt{r})$, $(5, 3, \texttt{d})$, $(12, 2, \texttt{a})$.

b. [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?