201700080 Information Theory and Statistics
15 April 2025, 8:45 – 11:45

This test consists of 5 problems for a total of 33 points. All answers need to be justified. Hints can be used without proof. The use of a non-programmable calculator (not a "GR") is allowed. A single-sided A4 cheat sheet is allowed. No additional books or notes may be used.

1. Gamma distribution. The gamma distribution with parameters $\theta > 0$ and $k \in \{1, 2, \dots\}$ has the following properties:

- We write $X \sim \text{Gamma}(k, \theta)$ for X that is gamma distributed with parameters k and θ. We have
  - $f(x) = \frac{1}{(k-1)!\theta^k} x^{k-1} e^{-x/\theta}$,
  - $\mathbb{E}[X] = k\theta$,
  - $\mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{(k-1)\theta}$,
- If $X_1, X_2, \dots, X_n$ are independent and identically distributed (i.i.d.) random variables with $X_i \sim \text{Gamma}(k, \theta)$, then the sum $\sum_{i=1}^n X_i \sim \text{Gamma}(nk, \theta)$.
- If $X \sim \text{Gamma}(k, \theta)$, then $cX \sim \text{Gamma}(k, c\theta)$ for $c > 0$.

You receive an independent sample $x_1, x_2, \dots, x_n$ of values from the random variable $X \sim \text{Gamma}(k, \theta)$, where k is known and θ is unknown. You want to estimate θ using the estimator $\hat{\theta} = \frac{C}{\sum_{i=1}^n x_i}$, where C is a constant.

(a) [3 pt] Find the value of C such that $\hat{\theta}$ is an unbiased estimator of θ.

(b) [3 pt] Give a lower bound on the variance of $\hat{\theta}$.

2. [5 pt] **Bottleneck.** Let X_1, X_2 and X_3 be discrete random variables that form a Markov chain X_1—X_2—X_3, i.e.,

$$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2).$$

Also, $x_1 \in \{1, 2, ..., n\}$, $x_2 \in \{1, 2, ..., k\}$, $x_3 \in \{1, 2, ..., n\}$ and $k \leq n$.

Show that the dependence between X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$.

P.T.O. (Please turn over)

3. **Hypothesis test.** Consider a binary hypothesis testing problem in which $P_1(X = k) = (1-p_1)^k p_1$ and $P_2(X = k) = (1-p_2)^k p_2$, $k = 0, 1, 2, \ldots$ We have $p_1 = 1/2 + a$ and $p_2 = 1/2 - a$, with $0 < a < 1/2$. We observe n independent samples $x_1, x_2, \ldots, x_n$ from the random variable X. *(Hint: $\mathbb{E}_{P_1}[X] = (1-p_1)/p_1$ and $\mathbb{E}_{P_2}[X] = (1-p_2)/p_2$.)*

   (a) [2 pt] State an optimal decision rule for choosing between the two distributions based on n observations. There's no need to simplify/compute the decision rule.

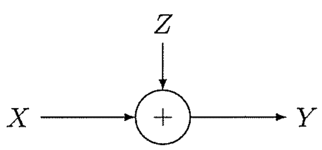
   (b) [3 pt] Show that $\|P_1 - P_2\|_{\mathrm{TV}} = 2a$. *(Hints:*
   - $\|P_1 - P_2\|_{\mathrm{TV}} = \frac{1}{2}\sum_{k=0}^{\infty} |P_1(X = k) - P_2(X = k)|$,
   - *Split the sum into two parts:* $k = 0$ *and* $k \geq 1$.
   - $\sum_{k=1}^{\infty} r^k = \frac{r}{1-r}$ *for* $0 < r < 1$.*)*

   (c) [1 pt] Give a lower bound on $\alpha + \beta$.

   (d) [3 pt] Suppose that P_1 is the actual distribution. Compute

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{P_1(x_i)}{P_2(x_i)}.$$

4. [5 pt] **Sum channel.** Find the channel capacity of the following discrete memoryless channel:



   where $P(Z = 0) = P(Z = a) = 1/2$, with $a \in \mathbb{R}$, and $X \in \{0, 1\}$. The channel is defined as $Y = X + Z$. Give the channel capacity for all possible values of a. *(Hint: The erasure channel has capacity $1 - \epsilon$, where ϵ is the erasure probability.)*

5. **Shannon-Fano code.** Consider a source with alphabet $\mathcal{X} = \{a, b, c, d, e\}$ and $p(a) = 0.3$, $p(b) = 0.2$, $p(c) = 0.2$, $p(d) = 0.15$ and $p(e) = 0.15$. The Shannon-Fano code is a compression algorithm that assigns codewords to symbols in such a way that the length of the codeword of symbol x is $\lceil -\log_2 p(x) \rceil$. We use the notation $\lceil \cdot \rceil$ to denote the ceiling function, which rounds a number up to the nearest integer.

   (a) [3 pt] Is it possible to construct a prefix code for the source for which the expected codeword length is 2.2? Explain.

   (b) [2 pt] Construct a (prefix) Shannon-Fano code for the source.

   (c) [3 pt] Is the expected codeword length of a Shannon-Fano code optimal? If so, explain. If not, construct a better code.