1 Reading (optional)

1. Read [1, Chapters 11,13]

2 Exercises

NOTE: Each exercise is 10 points. Only 3 exercises per assignment will be graded. If you submit more than 3 solved exercises please indicate which ones you want to be graded.

1 Consider a probability measure $\mathbb{P}$ and a measure-preserving transformation $\tau : \Omega \to \Omega$. Prove:

$\tau$-ergodic iff for any measurable $A, B$ we have

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}[A \cap \tau^{-k} B] \to \mathbb{P}[A] \mathbb{P}[B].$$

Comment: Thus ergodicity is a weaker condition than mixing: $\mathbb{P}[A \cap \tau^{-n} B] \to \mathbb{P}[A] \mathbb{P}[B]$.

2 Consider a three-state Markov chain $S_1, S_2, \ldots$ with the following transition probability matrix

$$P = \begin{bmatrix}
\frac{1}{2} & 0 & 0 \\
0 & \frac{1}{2} & \frac{1}{2} \\
1 & 0 & 0
\end{bmatrix}. $$

Compute the limit of $\frac{1}{n} \mathbb{E}[l(f(S^n))]$ when $n \to \infty$. Does your answer depend on the distribution of the initial state $S_1$?

3 Enumerative Codes. Consider the following simple universal compressor for binary sequences: Given $x^n \in \{0, 1\}^n$, denote by $n_1 = \sum_{i=1}^{n} x_i$ and $n_0 = n - n_1$ the number of ones and zeros in $x^n$. First encode $n_1 \in \{0, 1, \ldots, n\}$ using $\lceil \log_2(n + 1) \rceil$ bits, then encode the index of $x^n$ in the set of all strings with $n_1$ number of ones using using $\lceil \log_2 \left( \frac{n}{n_1} \right) \rceil$ bits. Concatenating two binary strings, we obtain the codeword of $x^n$. This defines a lossless compressor $f : \{0, 1\}^n \to \{0, 1\}^*$. 

1. Verify that $f$ is a prefix code.

2. Let $S^n_{\theta} \overset{i.i.d.}{\sim} \text{Bern} (\theta)$. Show that for any $\theta \in [0, 1],$

$$\mathbb{E} [l(f(S^n_{\theta}))] \leq nh(\theta) + \log n + O(1),$$

where $h(\cdot)$ is the binary entropy function. Conclude that the average code length $\frac{1}{n} \mathbb{E} [l(f(S^n_{\theta}))]$ achieves the entropy simultaneously for all $\theta$, as $n \to \infty$.

3. Show that

$$\sup_{0 \leq \theta \leq 1} \{\mathbb{E} [l(f(S^n_{\theta}))] - nh(\theta)\} \geq \log n + O(1).$$

Compare with the performance of the optimal universal codes.

[Optional: Explain why enumerative coding fails to achieve the optimal redundancy.]
Hint: The following non-asymptotic version of Stirling approximation might be useful

\[ 1 \leq \frac{n!}{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n} \leq \frac{e}{\sqrt{2\pi}}, \quad \forall n \in \mathbb{N}. \]

4 Let \( P_0 \) and \( P_1 \) be distributions on \( \mathcal{X} \). Recall that the region of achievable pairs \( (P_0[Z = 0], P_1[Z = 0]) \) via randomized tests \( P_{Z|X} : \mathcal{X} \to \{0, 1\} \) is denoted

\[ \mathcal{R}(P_0, P_1) \triangleq \bigcup_{P_{Z|X}} (P_0[Z = 0], P_1[Z = 0]) \subseteq [0, 1]^2. \]

Let also \( P_{Y|X} : \mathcal{X} \to \mathcal{Y} \) be a random transformation, which carries \( P_j \) to \( Q_j \) according to \( P_j \xrightarrow{P_{Y|X}} Q_j, \ j = 0, 1 \). Compare the regions \( \mathcal{R}(P_0, P_1) \) and \( \mathcal{R}(Q_0, Q_1) \). What does this say about \( \beta_\alpha(P_0, P_1) \) vs. \( \beta_\alpha(Q_0, Q_1) \)?

Comment: This is the most general form of data-processing, all the other ones (divergence, mutual information, \( f \)-divergence, total-variation, Rényi-divergence, etc) are corollaries.

5 Let \( P_0 \) and \( P_1 \) be two distributions on a finite alphabet \( \mathcal{X} \) such that \( P_0 \sim P_1 \) (that is, \( P_0(x) > 0 \iff P_1(x) > 0 \)). Denote the loglikelihood ratio by

\[ F = \log \frac{P_0(X)}{P_1(X)}. \]

Denote by \( P_{F_0} \) and \( P_{F_1} \) the distribution of \( F \) under \( P_0 \) and \( P_1 \), resp. (That is, \( P_{F_0}, P_{F_1} \) are distributions on \( \mathbb{R} \)).

1. Can distribution \( P_{F_1} \) be recovered from \( P_{F_0} \)?
2. What are the general properties of \( P_{F_0} \)? (list as many as possible)
3. Given a distribution \( Q \) on \( \mathbb{R} \) with such properties can you define \( P_0 \) and \( P_1 \) such that \( P_{F_0} = Q \)?

6 Consider distribution \( P \) and \( Q \) with the density in Fig. 1.

1. Compute the expression of \( \beta_\alpha(P, Q) \).
2. Plot the region \( \mathcal{R}(P, Q) \).
3. Specify the tests achieving \( \beta_\alpha \) for \( \alpha = 5/6 \) and \( \alpha = 1/2 \), respectively.
References
