Homework 1

Due: 16:30, 09/21, 2023 (in class)

Homework Policy: (Read before You Start to Work)

- Copying from other students' solution is not allowed. If caught, all involved students get 0 point on that particular homework. Caught twice, you will be asked to drop the course.
- Collaboration is welcome. You can work together with **at most one partner** on the homework problems which you find difficult. However, you should write down your own solution, not just copying from your partner's.
- Your partner should be the same for the entire homework.
- Put your collaborator's name beside the problems that you collaborate on.
- When citing known results from the assigned references, be as clear as possible.

1. (Another kind of typical sequences) [18]

In this problem, let us consider another kind of typical sequences defined as follows.

Definition. For $\gamma \in (0,1)$, a sequence s^n is called γ-typical with respect to a DMS $S \sim \mathsf{P}_S$ if

$$|\pi(a|s^n) - \mathsf{P}_S(a)| \le \gamma \mathsf{P}_S(a), \ \forall\, a \in \mathcal{S},$$

where $\pi(a|s^n) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{s_i = a\}$. The γ-typical set

$$\mathcal{T}_\gamma^{(n)}(S) := \{s^n \in \mathcal{S}^n \mid s^n \text{ is } \gamma\text{-typical with respect to } S\}.$$

a) Show that the typical sequence and typical set defined above also satisfy AEP (Proposition 1 of Unit 1) with $\mathcal{A}_\delta^{(n)}(S)$ replace by $\mathcal{T}_\gamma^{(n)}(S)$, and the δ in properties 1, 3, and 4 replaced by something, denoted by $\xi(\gamma)$, depending on γ. Specify this $\xi(\gamma)$. [6]

b) Show that $\mathcal{T}_\gamma^{(n)} \subseteq \mathcal{A}_\delta^{(n)}$ where $\delta = \xi(\gamma)$ found in a). [6]

c) Find an alphabet $\mathcal{S}$, a reference probability mass function P_S, and γ such that $\forall\, \delta' > 0, n \in \mathbb{N}$, $\mathcal{A}_{\delta'}^{(n)} \not\subseteq \mathcal{T}_\gamma^{(n)}$. [6]

Remark. From b) and c) we see that the typicality defined in this problem is *stronger* than that defined in the lecture. Hence, they are called strong typicality and weak typicality respectively in the literature.

2. (Finer asymptote for lossless source coding achievability) [20]

Consider a discrete memoryless source $S_i \overset{\text{i.i.d.}}{\sim} \mathsf{P}_S$, $i = 1, 2, \ldots$, where P_S is the PMF of the source S. Let $\mathrm{R} > \mathrm{H}(S)$, that is,

$$\mathrm{R} = \mathrm{H}(S) + \delta,$$

where $\delta > 0$ denotes a constant. Then, it can be shown as a corollary of the lossless source coding theorem in our lecture that there exists a sequence of $(n, \lfloor n\mathrm{R} \rfloor)$ codes such that $\forall\, \epsilon > 0$,

$$\mathsf{P}_{\mathsf{e}}^{(n)} \le \epsilon \quad \text{for } n \text{ sufficiently large.} \tag{†}$$

Notably, the gap to the fundamental limit, δ, is a constant not depending on ϵ.

Suppose we would like to achieve (†) for a given $\epsilon \in (0, 1/2)$. Is it possible to derive a finer asymptote for $\mathrm{R} - \mathrm{H}(S)$, the gap to the fundamental limit? In this problem, we are going to show that $\mathrm{R} - \mathrm{H}(S) = \Theta(n^{-1/2})$ suffices.

a) (Warm-up) Let $\varsigma(S) > 0$ denote the standard deviation of $\log \frac{1}{\mathsf{P}_S(S)}$ when $S \sim \mathsf{P}_S$ and $\Phi(\cdot)$ denote the CDF of a standard normal RV. Use the central limit theorem to prove the following

$$\lim_{n\to\infty} \mathsf{Pr}\left\{ \prod_{i=1}^n \mathsf{P}_S(S_i) \ge 2^{-n\left(\mathrm{H}(S) + n^{-1/2}\delta\,\varsigma(S)\right)} \right\} = \Phi(\delta) \qquad [8]$$

With the above, if we define a set of length-n source sequences $\mathcal{B}_\delta^{(n)}(S)$ as follows:

$$\mathcal{B}_\delta^{(n)}(S) := \left\{ s^n \,\middle|\, \prod_{i=1}^n \mathsf{P}_S(s_i) \ge 2^{-n\left(\mathrm{H}(S) + n^{-1/2}\delta\,\varsigma(S)\right)} \right\},$$

one can control the probability of each sequence in $\mathcal{B}_\delta^{(n)}(S)$ from below and hence can control the cardinality of this set from above. Also, we know that $\mathsf{Pr}\{S^n \in \mathcal{B}_\delta^{(n)}(S)\} \to \Phi(\delta)$ as $n \to \infty$ from Part a). It is hence tempting to use label all the sequences in $\mathcal{B}_\delta^{(n)}$ and give up the rest as a source encoding scheme. However, to upper bound the error probability, knowing its limit as $n \to \infty$ is not enough. Berry-Esseen theorem is a standard refinement of the CLT.

b) Show that (†) can be attained using the aforementioned scheme if the rate approaches $\mathrm{H}(S)$ from above as $n \to \infty$ in the following manner:

$$\mathrm{R}_n = \mathrm{H}(S) - n^{-1/2}\varsigma(S)\Phi^{-1}(\epsilon) + \zeta_n$$

where $\zeta_n = O(n^{-1})$ denotes a positive sequence tends to zero not slower than n^{-1}. [12]

Remark. The above is not optimal – the optimal asymptote of the rate (when $\varsigma(S) > 0$) is

$$\mathrm{R}_n = \mathrm{H}(S) - n^{-1/2}\varsigma(S)\Phi^{-1}(\epsilon) - \frac{\log n}{2n} + O(1/n).$$

3. (An alternative lossless source coding theorem) [12]

For a discrete memoryless source $\{S_i \mid i \in \mathbb{N}\}$, consider a sequence of $(n, \lfloor n\mathrm{R} \rfloor)$ source codes indexed by $n = 1, 2, \ldots$ with compression rate $\mathrm{R} > 0$.

Prove the following statements.

a) If $\mathrm{R} > \mathrm{H}(S)$, there exist a sequence of $(n, \lfloor n\mathrm{R} \rfloor)$ codes with

$$\lim_{n \to \infty} \mathsf{P}_{\mathsf{e}}^{(n)} = 0.$$

In other words, the probability of error can be driven to zero as $n \to \infty$. [6]

b) If $\mathrm{R} < \mathrm{H}(S)$, for any sequence of $(n, \lfloor n\mathrm{R} \rfloor)$ codes, the sequence of error probabilities must converge to 1, that is,

$$\lim_{n \to \infty} \mathsf{P}_{\mathsf{e}}^{(n)} = 1. \quad [6]$$