Homework 1 Solution

TA: Heng-Chien Liou [*]

r11942067@ntu.edu.tw

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

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

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

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

In the following, please assume $|\mathcal{S}| < \infty$.

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 $\forall\, \delta' > 0, n \in \mathbb{N}$, $\mathcal{A}_{\delta'}^{(n)} \nsubseteq \mathcal{T}_\gamma^{(n)}$. [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.

Solution:

a) **Property 1** $\forall s^n \in \mathcal{T}_\gamma^{(n)}$, we have $\forall a \in \mathcal{S}$, $|\pi(a|s^n) - \mathsf{P}_S(a)| \le \gamma \mathsf{P}_S(a)$

$$\frac{1}{n}\log \mathsf{Pr}\{S^n = s^n\} = \sum_{a\in\mathcal{S}} \pi(a|s^n)\log \mathsf{P}_S(a)$$

[*] With contributions by Wen-Shao Ho and Chen-Hao Hsiao.

$$\begin{cases} \le \sum_{a\in\mathcal{S}} (1-\gamma)\mathsf{P}_S(a)\log \mathsf{P}_S(a) = -(1-\gamma)\mathrm{H}(S) \\ \ge \sum_{a\in\mathcal{S}} (1+\gamma)\mathsf{P}_S(a)\log \mathsf{P}_S(a) = -(1+\gamma)\mathrm{H}(S)\,. \end{cases}$$

Hence, with $\xi(\gamma) = \gamma \mathrm{H}(S)$

$$\begin{aligned} \mathsf{Pr}\{S^n = s^n\} &= \sum_{a\in\mathcal{S}} \pi(a|s^n)\log \mathsf{P}_S(a) \\ & \begin{cases} \le 2^{-n(1-\gamma)\mathrm{H}(S)} = 2^{-n(\mathrm{H}(S)-\xi(\gamma))} \\ \ge 2^{-n(1+\gamma)\mathrm{H}(S)} = 2^{-n(\mathrm{H}(S)+\xi(\gamma))}. \end{cases} \end{aligned}$$

Property 2 Consider any $\epsilon > 0$. By basic definition on set and probability,

$$\begin{aligned} \mathsf{Pr}\left\{s^n \in \mathcal{T}_\gamma^{(n)}\right\} &= \mathsf{Pr}\left\{\forall a \in \mathcal{S}, |\pi(a|s^n) - \mathsf{P}_S(a)| \le \gamma \mathsf{P}_S(a)\right\} \\ &= \mathsf{Pr} \bigcap_{a\in\mathcal{S}} \{|\pi(a|s^n) - \mathsf{P}_S(a)| \le \gamma \mathsf{P}_S(a)\} \\ &= 1 - \mathsf{Pr} \bigcup_{a\in\mathcal{S}} \{|\pi(a|s^n) - \mathsf{P}_S(a)| > \gamma \mathsf{P}_S(a)\} \end{aligned}$$

For given $a \in \mathcal{S}$, by the weak law of large number, $\exists n_a \in \mathcal{N}$ such that $\forall n \ge n_a$

$$\mathsf{Pr}\left\{|\pi(a|s^n) - \mathsf{P}_S(a)| > \gamma \mathsf{P}_S(a)\right\} \le \frac{\epsilon}{|\mathcal{S}|}.$$

Select $n_0 = \max\{n_a : a \in \mathcal{S}\}$, then $\forall n \ge n_0$, by union bound

$$\begin{aligned} \mathsf{Pr} \bigcup_{a\in\mathcal{S}} \{|\pi(a|s^n) - \mathsf{P}_S(a)| > \gamma \mathsf{P}_S(a)\} &\le \sum_{a\in\mathcal{S}} \mathsf{Pr}\left\{|\pi(a|s^n) - \mathsf{P}_S(a)| > \gamma \mathsf{P}_S(a)\right\} \\ &\le \sum_{a\in\mathcal{S}} \frac{\epsilon}{|\mathcal{S}|} = \epsilon \end{aligned}$$

Property 3 by Property 1,

$$1 \ge \mathsf{Pr}\left\{s^n \in \mathcal{T}_\gamma^{(n)}\right\} = \sum_{s^n \in \mathcal{T}_\gamma^{(n)}} \mathsf{Pr}\{S^n = s^n\} \ge |\mathcal{T}_\gamma^{(n)}| 2^{-n(\mathrm{H}(S)+\xi(\gamma))}$$

Property 4 by Property 1 and 2, $\forall n \ge n_0$

$$1 - \epsilon \le \mathsf{Pr}\left\{s^n \in \mathcal{T}_\gamma^{(n)}\right\} = \sum_{s^n \in \mathcal{T}_\gamma^{(n)}} \mathsf{Pr}\{S^n = s^n\} \le |\mathcal{T}_\gamma^{(n)}| 2^{-n(\mathrm{H}(S)-\xi(\gamma))}$$

b) $\forall s^n = (s_1, s_2, ..., s_n) \in \mathcal{T}_\gamma^{(n)}$, by the previous argument and the property of memorylessness:

$$-(1+\gamma)\mathrm{H}(S) \le \frac{1}{n}\sum_{i=1}^{n} \log \mathsf{P}_S(s_i) = \frac{1}{n}\log \mathsf{Pr}\{S^n = s^n\} \le -(1-\gamma)\mathrm{H}(S)\,.$$

$\mathcal{T}_\gamma^{(n)} \subseteq \mathcal{A}_\delta^n$ with $\delta = \xi(\gamma) = \gamma \mathrm{H}(S)$

c) For some $\gamma < 1$, consider $\mathcal{S} = \{0, 1\}$ and $\mathsf{P}_S(0) = \mathsf{P}_S(1) = \frac{1}{2}$. Consider $s^n = 0^n$. $\mathrm{H}(S) = 1 \Rightarrow 0^n \in \mathcal{A}_{\delta'}^n, \forall \delta' > 0$. But $0^n \notin \mathcal{T}_\gamma^{(n)}$.

Grading Policy:

a) Property 1 worths [2] points, Property 2 worths [3] points, and Property 3, 4 worth [1] point in total.

b) Correct arguments about set relations [4] and correct arithemtic details [2].

c) A correct example [3], and justification [3].

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 $\forall\, \epsilon > 0$,

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

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

Suppose we would like to achieve ($\dagger$) 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 ($\dagger$) 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

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

Solution:

a) Central limit theorem states that for i.i.d random variables $X_1, \cdots, X_n$ with finite mean μ, variance σ^2. As $n \to \infty$, the random variable $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$, where $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$.
Namely, for any real number z,

$$\lim_{n\to\infty} \mathsf{Pr}\left\{\sqrt{n}(\bar{X}_n - \mu) \le z\right\} = \Phi\left(\frac{z}{\sigma}\right).$$

Choose $X_i = \log\frac{1}{P_S(S_i)}$ and $z = \delta\varsigma(S)$ in the Central limit theorem:

$$\begin{aligned}
&\lim_{n\to\infty} \mathsf{Pr}\left\{\sqrt{n}\left(-\frac{1}{n}\log\prod_{i=1}^n P_S(S_i) - \mathsf{E}\left[\log\frac{1}{P_S(S)}\right]\right) \le \delta\varsigma(S)\right\} = \Phi(\delta)\\
\Leftrightarrow &\lim_{n\to\infty} \mathsf{Pr}\left\{\left(-\frac{1}{n}\log\prod_{i=1}^n P_S(S_i)\right) \le H(S) + n^{-1/2}\delta\varsigma(S)\right\} = \Phi(\delta)\\
\Leftrightarrow &\lim_{n\to\infty} \mathsf{Pr}\left\{\prod_{i=1}^n P_S(S_i) \ge 2^{-n\left(H(S)+n^{-1/2}\delta\varsigma(S)\right)}\right\} = \Phi(\delta).
\end{aligned}$$

b) Note that we can now label all the sequences in $\mathcal{B}_\delta^{(n)}(S)$ and give up the rest, we want to find $\delta(\epsilon)$ such that when n is sufficiently large,

$$\mathsf{Pr}\left\{S^n \in \mathcal{B}_{\delta(\epsilon)}^{(n)}(S)\right\} \ge 1 - \epsilon.$$

Combine Berry-Esseen theorem and subproblem a), we know that

$$\mathsf{Pr}\left\{S^n \in \mathcal{B}_{\delta(\epsilon)}^{(n)}(S)\right\} \ge \Phi(\delta(\epsilon)) - c\frac{\rho^3}{\varsigma(S)^3}n^{-1/2}.$$

Therefore, it suffices to find $\delta(\epsilon)$ such that

$$\Phi(\delta(\epsilon)) - c\frac{\rho^3}{\varsigma(S)^3}n^{-1/2} \ge 1 - \epsilon \Leftrightarrow \delta(\epsilon) \ge \Phi^{-1}\left(1 - \epsilon + c\frac{\rho^3}{\varsigma(S)^3}n^{-1/2}\right).$$

The order of $\Phi^{-1}\left(1 - \epsilon + c\frac{\rho^3}{\varsigma(S)^3}n^{-1/2}\right)$ can be further characterized using Taylor expansion at $1-\epsilon$ on the function $\Phi^{-1}(x)$:

$$\Phi^{-1}(x) = \Phi^{-1}(1-\epsilon) + \frac{\Phi^{-1}(1-\epsilon)}{1!}(x - (1-\epsilon)) + O(x^2).$$

Plug in $x = 1 - \epsilon + c\frac{\rho^3}{\varsigma(S)^3} n^{-1/2}$ we have

$$\Phi^{-1}\left(1 - \epsilon + c\frac{\rho^3}{\varsigma(S)^3} n^{-1/2}\right) = \Phi^{-1}(1-\epsilon) + O(n^{-1/2}) = -\Phi^{-1}(\epsilon) + O(n^{-1/2}).$$

The result above indicates that if we choose $\delta(\epsilon) = -\Phi^{-1}(\epsilon) + O(n^{-1/2})$, we can guarantee

$$\mathsf{Pr}\left\{S^n \in \mathcal{B}^{(n)}_{\delta(\epsilon)}(S)\right\} \geq 1 - \epsilon.$$

Similar to the proof of the properties of the typical set, for n sufficiently large:

$$\begin{aligned} 1 \geq \mathsf{Pr}\left\{S^n \in \mathcal{B}^{(n)}_{\delta(\epsilon)}(S)\right\} &\geq |\mathcal{B}^{(n)}_{\delta(\epsilon)}(S)| 2^{-n\left(H(S) + n^{-1/2}\delta(\epsilon)\varsigma(S)\right)} \\ \Rightarrow |\mathcal{B}^{(n)}_{\delta(\epsilon)}(S)| \leq 2^{n\left(H(S) + n^{-1/2}\delta(\epsilon)\varsigma(S)\right)} &= 2^{n\left(H(S) - n^{-1/2}\varsigma(S)\Phi^{-1}(\epsilon) + O(n^{-1})\right)}. \end{aligned}$$

Therefore, the choice

$$R_n = H(S) - n^{-1/2}\varsigma(S)\Phi^{-1}(\epsilon) + O(n^{-1}),$$

makes it possible to encode all the sequences in $\mathcal{B}^{(n)}_{\delta(\epsilon)}(S)$ and also satisfies the requirement that $\mathsf{Pr}\left\{S^n \in \mathcal{B}^{(n)}_{\delta(\epsilon)}(S)\right\} \geq 1 - \epsilon$.

Grading Policy:

a) Justification for your application of central limit theorem [4], other details [4].

b) Overall logic in scheme and analysis [4], application of Berry-Esseen theorem [4], other details such as mean value theorem or Taylor's theorem [4].

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

For a discrete memoryless source $\{S_i \,|\, i \in \mathbb{N}\}$, 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

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

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,

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

Solution:

a) Let $\delta = \mathrm{R} - \mathrm{H}(S)$. A simple scheme is to use $\mathcal{A}_{\delta/2}^{(n)}$ and standard aymptotic equipartition property for coding. Here we try to prove it using the lossless source coding theorem in the lecture.

Consider a sequence $\{\epsilon_m\}_{m=1}^{\infty} > 0$ such that $\lim_{m\to\infty} \epsilon_m = 0$. For any $n, \epsilon_m > 0$, consider $\exists k^*$ such that for any $k \geq k^*(n, \epsilon_m)$, there exists a (n, k) code with $\mathsf{P}_{\mathsf{e}}^{(n)} < \epsilon_m$. By the lossless source coding theorem in the lecture and the Archimedean property, $\exists n_m > \max_{i=1}^{m-1} \{n_i\}$ such that $\forall n \geq n_m, k^*(n, \epsilon_m) < n(\mathrm{H}(S) + \delta/2) < \lfloor n\mathrm{R} \rfloor$. $\{n_m\}_{m}^{\infty}$ then is a strictly increasing sequence of integers.

Therefore, we can construct a sequence of $(n, \lfloor n\mathrm{R} \rfloor)$ codes as follows: for n such that $n_m \leq n < n_{m+1}$ select $k = k^*(n, \epsilon_m)$ and one of the correpsonding (n, k) code $(\mathrm{enc}_n, \mathrm{dec}_n)$ such that $\mathsf{P}_{\mathsf{e}}^{(n)} < \epsilon_m$. By the above justification, $k < \lfloor n\mathrm{R} \rfloor$.

$\{(\mathrm{enc}_n, \mathrm{dec}_n)\}_n$ is then a desirable code with the property.

b) Let $\delta = \mathrm{H}(S) - \mathrm{R}$. Again, a simple strategy is to use $\mathcal{A}_{\delta/2}^{(n)}$ and standard aymptotic equipartition property to characterize the error probability for any sequence of codes. Here we try to prove it using the lossless source coding theorem in the lecture.

We attempt to prove by contradiction. Suppose there exist a sequence of $(n, \lfloor n\mathrm{R} \rfloor)$ codes with error probabilitties not converging to 1, namely, for some $\epsilon < 0$, $\forall m \in \mathbb{N}, \exists n'(m) \geq m$ such that $\mathsf{P}_{\mathsf{e}}^{(n'(m))} < 1 - \epsilon$.

$\forall m \in \mathbb{N}$, the $(n'(m), \lfloor n'(m)R \rfloor)$ code from the above sequences, by definition, is then a $(n'(m), \lfloor n'(m)R \rfloor, 1 - \epsilon)$ code.

$$\mathsf{P}_{\mathsf{e}}^{(n'(m))} < 1 - \epsilon, \ \forall m \in \mathbb{N} \tag{1}$$

On the other hand, for such ϵ, by the lossless source coding theorem, we know that $\exists n_{\delta,\epsilon} \in \mathbb{N}$ such that $\forall n \geq n_{\delta,\epsilon}$,

$$k^*(n, 1 - \epsilon) > n(\mathrm{H}(S) - \delta/2) > n\mathrm{R} \tag{2}$$

From (1), for $m \geq n_{\delta,\epsilon}$, $\exists n'(m) \geq m \geq n_{\delta,\epsilon}$ and a corresponding $(n'(m), \lfloor n'(m)R \rfloor)$ code with error smaller than $1 - \epsilon$. But from (2), for such $n'(m) \geq n_{\delta,\epsilon}$, the smallest length required is $k^*(n'(m), 1 - \epsilon) > n'(m)\mathrm{R}$. This leads to contradiction.

Grading Policy:

a) A correct scheme [3] and the justification [3]; alternatively, a correct proof of existence [6].

b) A proof that applies for all codes [4] and other details [2].