Information Theory HW1

許博翔

September 21, 2023

Problem 1. I'll prove (b) first, and then use (b) to prove (a) for convenience.

(b) Suppose that $s^n = (s_1, s_2, \ldots, s_n) \in \mathcal{T}_\gamma^{(n)}(S)$.

By the definition of $\mathcal{T}_\gamma^{(n)}(S)$, $\forall a \in \mathbf{S}$, $\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{s_i = a\} - P_S(a)\right| \le \gamma P_S(a)$.

$\Rightarrow \forall a \in \mathbf{S}$, $\left|\frac{1}{n}\sum_{i=1}^{n}\log(P_S(a))\mathbb{I}\{s_i = a\} - P_S(a)\log(P_S(a))\right| \le \gamma P_S(a)\log(P_S(a))$.

$\Rightarrow \sum_{a\in\mathbf{S}}\left|\frac{1}{n}\sum_{i=1}^{n}\log(P_S(a))\mathbb{I}\{s_i = a\} - P_S(a)\log(P_S(a))\right| \le \sum_{a\in\mathbf{S}}\gamma P_S(a)\log(P_S(a))$.

By triangular inequality,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\log(P_S(s_i)) + H(S)\right|$$
$$= \left|\sum_{a\in\mathbf{S}}\frac{1}{n}\sum_{i=1}^{n}\log(P_S(a))\mathbb{I}\{s_i = a\} - \sum_{a\in\mathbf{S}}P_S(a)\log(P_S(a))\right|$$
$$\le \sum_{a\in\mathbf{S}}\left|\frac{1}{n}\sum_{i=1}^{n}\log(P_S(a))\mathbb{I}\{s_i = a\} - P_S(a)\log(P_S(a))\right|$$
$$\le \sum_{a\in\mathbf{S}}\gamma P_S(a)\log(P_S(a)) = -\gamma H(S).$$

Taking $\delta = \xi(\gamma) := -\gamma H(S)$, and we get $\left|\frac{1}{n}\sum_{i=1}^{n}\log(P_S(s_i)) + H(S)\right| \le \delta$, which means $s^n \in \mathcal{A}_\delta^{(n)}(S)$.

$\therefore \mathcal{T}_\gamma^{(n)}(S) \subseteq \mathcal{A}_\delta^{(n)}(S)$.

(a) Recall from (b), we take $\delta = \xi(\gamma) := -\gamma H(S)$.

The 4 properties in the proposition are:

(1) The original property is: $\forall s^n \in \mathcal{A}_\delta^{(n)}(S)$, $2^{-n(H(S)+\delta)} \le \Pr\{S^n = s^n\} \le 2^{-n(H(S)-\delta)}$.

$\because$ from (b) we know that $\forall s^n \in \mathcal{T}_\gamma^{(n)}(S),\ s^n \in \mathcal{A}_\delta^{(n)}(S)$.

$\therefore 2^{-n(H(S)+\delta)} \leq \Pr\{S^n = s^n\} \leq 2^{-n(H(S)-\delta)}$.

(2) Let $A_n(a) := \{s^n \in \mathbf{S}^n : |\pi(a|s^n) - P_S(a)| > \gamma P_S(a)\}$.

Since $S \sim P_S$ is a DMS, the random variables $\{X_i\}_{i=1}^\infty$ where $X_i := \mathbb{I}\{S_i = a\}$ are i.i.d.

The average of X_i, denote as μ, $= \Pr\{S_i = a\} = P_S(a)$.

$\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i = \dfrac{1}{n}\sum_{i=1}^n \mathbb{I}\{S_i = a\} = \pi(a|S^n)$.

Take $\epsilon > \gamma P_S(a)$.

By the weak law of large numbers, $\lim\limits_{n\to\infty} \Pr\{S^n \in A_n(a)\} = \lim\limits_{n\to\infty} \Pr\{|\pi(a|S^n) - P_S(a)| > \gamma P_S(a)\} = \lim\limits_{n\to\infty} \Pr\{|\bar{X}_n - \mu| > \gamma P_S(a)\} \leq \lim\limits_{n\to\infty} \Pr\{|\bar{X}_n - \mu| \geq \epsilon\} = 0$.

$\because \mathcal{T}_\gamma^{(n)}(S) = \mathbf{S}^n \setminus \displaystyle\bigcup_{a\in\mathbf{S}} A_n(a)$.

$\therefore \lim\limits_{n\to\infty} \Pr\{S^n \in \mathcal{T}_\gamma^{(n)}(S)\} = 1 - \lim\limits_{n\to\infty} \Pr\{S^n \in \displaystyle\bigcup_{a\in\mathbf{S}} A_n(a)\} \geq 1 - \lim\limits_{n\to\infty} \sum_{a\in\mathbf{S}} \Pr\{S^n \in A_n(a)\} = 1$.

$\therefore \forall\epsilon > 0$, by the definition of limits, $\Pr\{S^n \in \mathcal{T}_\gamma^{(n)}(S)\} \geq 1 - \epsilon$ for n large enough.

(3) $\because \mathcal{T}_\gamma^{(n)}(S) \subseteq \mathcal{A}_\delta^{(n)}(S)$.

$\therefore |\mathcal{T}_\gamma^{(n)}(S)| \leq |\mathcal{A}_\delta^{(n)}(S)| \leq 2^{n(H(S)+\delta)}$.

(4) By (2), $\forall\epsilon > 0$, for n large enough, there is $1 - \epsilon \leq \Pr\{S^n \in \mathcal{T}_\gamma^{(n)}(S)\} = \displaystyle\sum_{s^n \in \mathcal{T}_\gamma^{(n)}(S)} \Pr\{S^n = s^n\} \overset{(1)}{\leq} \sum_{s^n \in \mathcal{T}_\gamma^{(n)}(S)} 2^{-n(H(S)-\delta)} = |\mathcal{T}_\gamma^{(n)}(S)| 2^{-n(H(S)-\delta)}$.

$\therefore \forall\epsilon > 0$, for n large enough, there is $|\mathcal{T}_\gamma^{(n)}(S)| \geq (1-\epsilon)2^{n(H(S)-\delta)}$.

(c) Consider $\mathbf{S} = \{0, 1\},\ P_S(0) = P_S(1) = \dfrac{1}{2}, \gamma = 0.1$.

For the sequence $s^n = 0^n$, $|\pi(0|s^n) - P_S(0)| = \dfrac{1}{2} \nleq 0.05 = \gamma P_S(0)$.

$\Rightarrow 0^n \notin \mathcal{T}_\gamma^{(n)}(S)$.

However, $\forall\delta' > 0$, $\left|\dfrac{1}{n}\displaystyle\sum_{i=1}^n \log P_S(s_i) + H(S)\right| = \left|\dfrac{1}{n}\sum_{i=1}^n \log\dfrac{1}{2} - \log\dfrac{1}{2}\right| = 0 \leq \delta'$.

$\Rightarrow 0^n \in \mathcal{A}_{\delta'}^{(n)}$.

$\therefore \mathcal{A}_{\delta'}^{(n)} \nsubseteq \mathcal{T}_\gamma^{(n)}(S)$.

Problem 2.

(a) Define $X_i = \log \frac{1}{P_S(S_i)}$. Since S_i are i.i.d, X_i are also i.i.d.

Since $P_S(S_i) \le 1$, we get that $\log \frac{1}{P_S(S_i)} \ge 0$.

$\Rightarrow \mathrm{E}[|X_i|] = \mathrm{E}[X_i] = \mathrm{E}[\log \frac{1}{P_S(S_i)}] = H(S) < \infty$.

$$\prod_{i=1}^{n} P_S(S_i) \ge 2^{-n(H(S)+n^{-1/2}\delta\varsigma(S))}$$

$$\iff \prod_{i=1}^{n} \frac{1}{P_S(S_i)} \le 2^{n(H(S)+n^{-1/2}\delta\varsigma(S))}$$

$$\iff \sum_{i=1}^{n} \log \frac{1}{P_S(S_i)} \le n(H(S)+n^{-1/2}\delta\varsigma(S))$$

$$\iff \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - H(S) \le n^{-1/2}\delta\varsigma(S)$$

$$\iff \frac{\sqrt{n}(\overline{X_n} - \mathrm{E}[X_i])}{\varsigma(S)} \le \delta.$$

By central limit theorem, $\frac{\sqrt{n}(\overline{X_n} - \mathrm{E}[X_i])}{\varsigma(S)} \xrightarrow{d} Z \sim N(0,1)$ as $n \to \infty$.

$$\Rightarrow \Pr\left\{\prod_{i=1}^{n} P_S(S_i) \ge 2^{-n(H(S)+n^{-1/2}\delta\varsigma(S))}\right\} = \Pr\left\{\frac{\sqrt{n}(\overline{X_n} - \mathrm{E}[X_i])}{\varsigma(S)} \le \delta\right\}$$

$\to \Pr\{Z \le \delta\} = \Phi(\delta)$ as $n \to \infty$.

(b) Let $Z \sim N(0,1)$, by Berry-Esseen theorem, $|\Pr\{S^n \in \mathcal{B}_\delta^{(n)}(S)\} - \Pr\{Z \le \delta\}| = \left|\Pr\left\{\frac{\sqrt{n}(\overline{X_n} - E[X_i])}{\varsigma(S)} \le \delta\right\} - \Pr\{Z \le \delta\}\right| \le cn^{-1/2}$ for some constant $c > 0$.

$\Rightarrow \Pr\{S^n \in \mathcal{B}_\delta^{(n)}(S)\} \ge \Pr\{Z \le \delta\} - cn^{-1/2} = \Phi(\delta) - cn^{-1/2}$.

Take $\delta = \Phi^{-1}(1 - \epsilon + cn^{-1/2}) = -\Phi^{-1}(\epsilon - cn^{-1/2})$, we get that $\Pr\{S^n \in \mathcal{B}_\delta^{(n)}(S)\} \ge 1 - \epsilon$.

$\Rightarrow \Pr\{S^n \notin \mathcal{B}_\delta^{(n)}(S)\} \le \epsilon$.

Since $\frac{d\Phi^{-1}(x)}{dx} = \left.\frac{1}{\frac{d\Phi(y)}{dy}}\right|_{y=\Phi^{-1}(x)} = \left.\sqrt{2\pi}e^{y^2/2}\right|_{y=\Phi^{-1}(x)} = \sqrt{2\pi}e^{(\Phi^{-1}(x))^2/2}$, there is $\Phi^{-1}(\epsilon - cn^{-1/2}) \approx \Phi^{-1}(\epsilon) - \sqrt{2\pi}e^{(\Phi^{-1}(\epsilon))^2/2}cn^{-1/2} = \Phi^{-1}(\epsilon) - O(n^{-1/2})$ for n sufficiently large.

$\Rightarrow \delta = -\Phi^{-1}(\epsilon) + \zeta_n'$, where $\zeta_n' = O(n^{-1/2})$.

Lemma 2.1. $\exists \zeta_n = O(n^{-1})$ s.t. $nk \le \lfloor n(k + \zeta_n)\rfloor$.

Proof. Consider $\zeta_n = \frac{1}{n}$, we get that $\lfloor n(k+\zeta_n)\rfloor = \lfloor n(k+\frac{1}{n})\rfloor = \lfloor nk\rfloor + 1 \ge$

nk. ■

Since $\displaystyle\sum_{s^n \in \mathcal{B}_\delta^{(n)}(S)} P_{S^n}(s^n) \leq \sum_{s^n} P_{S^n}(s^n) = 1$,

and if $s^n \in B$, then $P_{S^n}(s^n) = \displaystyle\prod_{i=1}^{n} P_S(s_i) \geq 2^{-n(H(S)+n^{-1/2}\delta\varsigma(S))}$.

$\therefore |\mathcal{B}_\delta^{(n)}(S)| 2^{-n(H(S)+n^{-1/2}\delta\varsigma(S))} = \displaystyle\sum_{s^n \in \mathcal{B}_\delta^{(n)}(S)} 2^{-n(H(S)+n^{-1/2}\delta\varsigma(S))} \leq \sum_{s^n \in \mathcal{B}_\delta^{(n)}(S)} P_{S^n}(s^n) \leq 1$.

$\Rightarrow |\mathcal{B}_\delta^{(n)}(S)| \leq 2^{n(H(S)+n^{-1/2}\delta\varsigma(S))}$.

By **Lemma (2.1)**, there exists $\zeta_n'' \in O(n^{-1})$ s.t. $n(H(S)+n^{-1/2}\delta\varsigma(S)) \leq \lfloor n(H(S)+n^{-1/2}\delta\varsigma(S)+\zeta_n'') \rfloor$.

Take $R = H(S) + n^{1/2}\varsigma(S)\delta + \zeta_n'' = H(S) - n^{-1/2}\varsigma(S)\Phi^{-1}(\epsilon) + n^{-1/2}\varsigma(S)\zeta_n' + \zeta_n''$.

Since $n^{-1/2}\varsigma(S)\zeta_n' = O(n^{-1})$, we get that $R = H(S) - n^{-1/2}\varsigma(S)\Phi^{-1}(\epsilon) + \zeta_n$ for some $\zeta_n = O(n^{-1})$.

Therefore, $\mathcal{B}_\delta^{(n)}(S)$ is an $(n, \lfloor nR \rfloor)$ code with $P_e^{(n)} \leq \epsilon$.

Problem 3.

(a) Let $\delta \in (0, R - H(S))$, and $\mathcal{A}_\delta^{(n)}(S)$ be the δ-typical set defined in Definition 1.

By the third property of Proposition 1, we know that $|\mathcal{A}_\delta^{(n)}(S)| \leq 2^{n(H(S)+\delta)} \overset{H(S)+\delta<R \Rightarrow n(H(S)+\delta)<nR-1 \text{ for } n \text{ large enough}}{<} 2^{\lfloor nR \rfloor}$ for n large enough.

$\Rightarrow \mathcal{A}_\delta^{(n)}(S)$ is an $(n, \lfloor nR \rfloor)$ code.

By the second property of Proposition 1, we know that $\forall \epsilon > 0,\ \exists N$ s.t. $\forall n \geq N,\ P_e^{(n)} = \Pr\{S^n \notin \mathcal{A}_\delta^{(n)}(S)\} \leq \epsilon$.

Since $P_e^{(n)} \geq 0$, therefore by the definition of limits, $\displaystyle\lim_{n\to\infty} P_e^{(n)} = 0$.

$\therefore$ such sequence exists, and it is $\mathcal{A}_\delta^{(n)}(S)$.

(b) For a given $(n, \lfloor nR \rfloor)$ code, let $\mathcal{B}^{(n)}$ denote the range of the decoding function.

Let $\delta \in (0, H(S) - R)$, and $\mathcal{A}_\delta^{(n)}(S)$ be the δ-typical set defined in Definition 1.

By the first property of Proposition 1, we know that $\forall s^n \in \mathcal{A}_\delta^{(n)}(S),\ \Pr\{S^n = s^n\} \leq 2^{-n(H(S)-\delta)}$.

$\Rightarrow \Pr\{S^n \in \mathcal{A}_\delta^{(n)}(S) \cap \mathcal{B}^{(n)}\} = \displaystyle\sum_{s^n \in \mathcal{A}_\delta^{(n)}(S) \cap \mathcal{B}^{(n)}} \Pr\{S^n = s^n\}$

$$\leq \sum_{s^n \in \mathcal{A}_\delta^{(n)}(S) \cap \mathcal{B}^{(n)}} 2^{-n(H(S)-\delta)} \leq \sum_{s^n \in \mathcal{B}^{(n)}} 2^{-n(H(S)-\delta)}$$

$$= |\mathcal{B}^{(n)}| 2^{-n(H(S)-\delta)} \leq 2^{\lfloor nR \rfloor - n(H(S)-\delta)} \leq 2^{-n(H(S)-R-\delta)}.$$

Since $H(S) - R - \delta > 0$ by definition of δ, we get that

$\lim_{n\to\infty} P_e^{(n)} = \lim_{n\to\infty} \Pr\{S^n \in \mathcal{A}_\delta^{(n)}(S) \cap \mathcal{B}^{(n)}\} \geq \lim_{n\to\infty} (1 - 2^{-n(H(S)-R-\delta)}) = 1.$

On the other hand, $P_e^{(n)} \leq 1$, so there is $\lim_{n\to\infty} P_e^{(n)} = 1$.