Solutions: More Exercises

Problem 1

We know that $X_i \sim \text{Bernoulli}(\theta)$ so $\mathbb{P}[X_i = 1] = \theta$ and $\mathbb{P}[X_i = 0] = 1 - \theta$.

(a) $\theta = 1/2$ so $H(X_i) = 1$ for $i = 1, \dots, n$.
By definition $(X_1, \dots, X_n) = (0, \dots, 0) \in A_\epsilon^{(n)}$ if and only if

$$2^{-n[H(X)+\epsilon]} \leq \mathbb{P}[(X_1, \dots, X_n) = (0, \dots, 0)] \leq 2^{-n[H(X)-\epsilon]},$$

or

$$2^{-n[H(X)+\epsilon]} \leq \prod_{i=1}^{n} \overbrace{\mathbb{P}[X_i = 0]}^{1/2} \leq 2^{-n[H(X)-\epsilon]}.$$

Then we can take the $\log(\cdot)$ of both side of the above inequalities. Because that the $\log(\cdot)$ function is an increasing function the order of the inequalities do not change and we have

$$-n[1+\epsilon] \leq -n \leq -n[1-\epsilon],$$

$$1 + \epsilon \geq 1 - \epsilon.$$

So $(0, \dots, 0) \in A_\epsilon^{(n)}$ if and only if $\epsilon \geq 0$ which means that it is true for every ϵ. In fact for $\theta = 1/2$ all of the 2^n possible sequences are belong to $A_\epsilon^{(n)}$.

(b) Again by definition $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$ if and only if

$$2^{-n[H(X)+\epsilon]} \leq \overbrace{\mathbb{P}[(X_1, \dots, X_n) = (x_1, \dots, x_n)]}^{\mathbb{P}[X_1=x_1]\cdots\mathbb{P}[X_n=x_n]} \leq 2^{-n[H(X)-\epsilon]},$$

if and only if

$$-n[H(X)+\epsilon] \leq \log \prod_{i=1}^{n} \mathbb{P}[X_i = x_i] \leq -n[H(X)-\epsilon],$$

if and only if

$$-n[H(X)+\epsilon] \leq \log\left(\theta^{L(x^n)}.(1-\theta)^{n-L(x^n)}\right) \leq -n[H(X)-\epsilon],$$

which only depends on $L(x^n)$.

(c) For the probability of observing a sequence $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$, by definition we have the following bounds

$$2^{-n[H(X)+\epsilon]} \leq \mathbb{P}[(X_1, \dots, X_n) = (x_1, \dots, x_n)] \leq 2^{-n[H(X)-\epsilon]},$$

so we can write

$$\frac{\mathbb{P}[\text{most probable sequence}]}{\mathbb{P}[\text{least probable sequence}]} = \frac{2^{-n[H(X)-\epsilon]}}{2^{-n[H(X)+\epsilon]}} = 2^{2n\epsilon} \xrightarrow{n\to\infty} \infty,$$

which shows that the typical sequences are not "approximately equiprobable".

(d) From (b) we can write

$$-n\left[H(X)+\epsilon\right] \leq L(x^n)\log(\theta) + (n - L(x^n))\log(1-\theta) \leq -n\left[H(X)-\epsilon\right],$$

$$-n\left[H(X)+\epsilon\right] \leq L(x^n)\log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta) \leq -n\left[H(X)-\epsilon\right],$$

and

$$-n\left[H(X)+\log(1-\theta)+\epsilon\right] \leq L(x^n)\log\left(\frac{\theta}{1-\theta}\right) \leq -n\left[H(X)+\log(1-\theta)-\epsilon\right].$$

We know that

$$\begin{cases} \log\left(\frac{\theta}{1-\theta}\right) > 0 \ : \ \theta > 1/2, \\ \log\left(\frac{\theta}{1-\theta}\right) < 0 \ : \ \theta < 1/2, \end{cases}$$

so for $\theta > 1/2$ we have

$$\frac{-n\left[H(\theta)+\log(1-\theta)+\epsilon\right]}{\log\left(\frac{\theta}{1-\theta}\right)} \leq L(x^n) \leq \frac{-n\left[H(\theta)+\log(1-\theta)-\epsilon\right]}{\log\left(\frac{\theta}{1-\theta}\right)},$$

where $H(\theta) = -\theta\log(\theta) - (1-\theta)\log(1-\theta)$. Then we have

$$n\cdot\theta - n\cdot\frac{\epsilon}{\log\left(\frac{\theta}{1-\theta}\right)} \leq L(x^n) \leq n\cdot\theta + n\cdot\frac{\epsilon}{\log\left(\frac{\theta}{1-\theta}\right)}.$$

So by choosing $p=\theta$ and $\alpha = \frac{\epsilon}{\log\left(\frac{\theta}{1-\theta}\right)}$ we have $C^{(n)}(\alpha,p) = A_\epsilon^{(n)}$.

Problem 2

Every set S_i is chosen uniformly at random from all 2^m possible subsets of the set $\{1,\ldots,m\}$. Then at step i we ask the question: is $X \in S_i$? Based on the answers to these questions we want to find the value of X.

(a) Let us assume that $X=1$. Then we can write the event that the question S_i at ith step has the same answer for object 1 and 2 as follows

$$\text{same answer at } i\text{th step} = \overbrace{[1\in S_i, 2\in S_i]}^{\text{event} A_i} \cup \overbrace{[1\notin S_i, 2\notin S_i]}^{\text{event} B_i},$$

because $A_i \cap B_i = \emptyset$. So for the probabilities we have

$$\mathbb{P}\left[\text{same answer at } i\text{th step}\right] = \mathbb{P}\left[1\in S_i, 2\in S_i\right] + \mathbb{P}\left[1\notin S_i, 2\notin S_i\right] = \mathbb{P}\left[A_i\right] + \mathbb{P}\left[B_i\right]$$

Now we can use the chain rule to write

$$\mathbb{P}\left[1\in S_i, 2\in S_i\right] = \mathbb{P}\left[2\in S_i \mid 1\in S_i\right]\mathbb{P}\left[1\in S_i\right].$$

Then it can be easily observed that

$$\mathbb{P}\left[1\in S_i\right] = \frac{2^{m-1}}{2^m} = 1/2,$$

and

$$\mathbb{P}[2 \in S_i \mid 1 \in S_i] = \frac{2^{m-2}}{2^{m-1}} = 1/2,$$

so

$$\mathbb{P}[A_i] = \mathbb{P}[1 \in S_i, 2 \in S_i] = \frac{1}{4}.$$

Using a similar argument we can show that

$$\mathbb{P}[B_i] = \mathbb{P}[1 \notin S_i, 2 \notin S_i] = \frac{1}{4}.$$

So we have

$$\mathbb{P}[\text{same answer with object 2 after } k \text{ step}] = \prod_{i=1}^{k} \mathbb{P}[A_i \cup B_i] = \prod_{i=1}^{k} \mathbb{P}[A_i] + \mathbb{P}[B_i] = 1/2^k.$$

(b) We can use the argument of part (a) to write

$$\mathbb{P}[\text{same answer with object } i \text{ after } k \text{ step}] = 1/2^k,$$

so the expected number of objects in the set $\{2, \ldots, m\}$ that have the same answers to the k questions as does the correct object 1 is

$$\text{average number of wrong objects} = \frac{m-1}{2^k} = \frac{2^n - 1}{2^k} = 2^{n-k} - 2^{-k}.$$

(c) Choosing $k = n + \sqrt{n}$ we have

$$\text{average number of wrong objects} = 2^{n-n-\sqrt{n}} - 2^{-n-\sqrt{n}} = 2^{-\sqrt{n}} - 2^{-n-\sqrt{n}}.$$

(d) Let E be the number of wrong objects after asking k questions. This is a random variable that in part (b) and (c) we calculated its expected value. For $k = n + \sqrt{n}$ we found that $\mathbb{E}[E] = 2^{-\sqrt{n}} - 2^{-n-\sqrt{n}}$. Now we have

$$\lim_{n \to \infty} \mathbb{E}[E] \to 0.$$

But we want to show that the probability of wrong answers goes to zero as $n \to \infty$. To this end, we use the Markov's inequality as following

$$\mathbb{P}[E \geq 1] = \mathbb{P}\left[E \geq \underbrace{\frac{1}{2^{-\sqrt{n}} - 2^{-n+\sqrt{n}}}}_{t} \underbrace{(2^{-\sqrt{n}} - 2^{-n-\sqrt{n}})}_{\mu}\right] \overset{\text{(i)}}{\leq} 2^{-\sqrt{n}} - 2^{-n-\sqrt{n}},$$

where (i) follows from Markov's inequality. So for the probability of having wrong answer we have

$$\lim_{n \to \infty} \mathbb{P}[E \geq 1] \longrightarrow 0.$$

Because the random variable E takes its value from $\{0, 1, 2, \ldots\}$ we conclude that $E = 0$ with probability approaching 1 as $n \to \infty$.

Problem 3

$\mathcal{P}$ is a given set of participants and $\mathcal{A}$ is a collection of subsets of $\mathcal{P}$.

(a) We know that $A, B \subset \mathcal{P}$, $B \notin \mathcal{A}$ and $A \cup B \in \mathcal{A}$. Using chain rule we can write

$$H\left(X_A, S \mid X_B\right) = H\left(X_A \mid X_B\right) + \underbrace{H\left(S \mid X_A, X_B\right)}_{0},$$

so we have $H\left(X_A \mid X_B\right) = H\left(X_A, S \mid X_B\right)$. Again applying the chain rule we obtain

$$H\left(X_A \mid X_B\right) = H\left(X_A, S \mid X_B\right) = \underbrace{H\left(S \mid X_B\right)}_{H(S)} + H\left(X_A \mid X_B, S\right),$$

and we are done.

(b) In this part we have $B \in \mathcal{A}$. Using the chain rule we can write

$$\begin{aligned} H\left(X_A \mid X_B\right) &= H\left(X_A, S \mid X_B\right) - H\left(S \mid X_A, X_B\right) \\ &= H\left(X_A \mid X_B, S\right) + \underbrace{H\left(S \mid X_B\right)}_{0} - \underbrace{H\left(S \mid X_A, X_B\right)}_{0}. \end{aligned}$$

The last term is zero because we have $H\left(S \mid X_A, X_B\right) \le H\left(S \mid X_B\right) = 0$, since conditioning reduces the entropy.

(c) Here we assume that $A, B, C \subset \mathcal{P}$ where $A \cup C \in \mathcal{A}$, $B \cup C \in \mathcal{A}$, and $C \notin \mathcal{A}$. Then we can write

$$\begin{aligned} I\left(X_A; X_B \mid X_C\right) &= \overbrace{\underbrace{H\left(X_B \mid \underbrace{X_C}_{\notin A}\right)}_{H(S)+H(X_B \mid X_C, S)}}^{\text{from (a)}} - \overbrace{\underbrace{H\left(X_B \mid \underbrace{X_A, X_C}_{\in A}\right)}_{H(X_B \mid X_A, X_C, S)}}^{\text{from (b)}} \\ &= H(S) + \underbrace{I\left(X_B; X_A \mid X_C, S\right)}_{\ge 0}, \end{aligned}$$

so we have $I\left(X_A; X_B \mid X_C\right) \ge H(S)$.

Problem 4

Two random variables X and X' are i.i.d. with entropy $H(X)$.

(a) We want to show that $\mathbb{P}\left[X = X'\right] \ge 2^{-H(X)}$. Suppose that $X \sim P(x)$ and let us write

$$\begin{aligned} 2^{-H(X)} &= 2^{\mathbb{E}[\log P(X)]} \\ &\overset{\text{(i)}}{\le} \mathbb{E}\left[2^{\log P(X)}\right] \\ &= \sum_x P(x) 2^{\log P(x)} \\ &= \sum_x P(x)^2 \\ &= \mathbb{P}\left[X = X'\right], \end{aligned}$$

where (i) comes from the Jensen's inequality applying on the function $f(y) = 2^y$. Because the function f is convex we have $\mathbb{E}f(Y) \geq f(\mathbb{E}Y)$. So defining a new random variable $Y \triangleq \log P(X)$ we have

$$2^{\mathbb{E}\overbrace{\log P(X)}^{Y}} \leq \mathbb{E}2^{\overbrace{\log P(X)}^{Y}},$$

which results in (i).

(b) This part is also very similar to the previous part. Let us write

$$\begin{aligned}
2^{-H(P)-D(P\|Q)} &= 2^{\sum P(x)\log P(x)+\sum P(x)\log\frac{Q(x)}{P(x)}} = 2^{\sum P(x)\log Q(x)} \\
&= 2^{\mathbb{E}_p \overbrace{\log Q(x)}^{Y}} \leq \mathbb{E}_p 2^{\log Q(x)} = \sum P(x)\, 2^{\log Q(x)} \\
&= \sum P(x)\, Q(x) \\
&= \mathbb{P}\left[X = X'\right].
\end{aligned}$$

The same method applies for the other one.

Problem 5

(a) Suppose that we have two stochastic processes $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ such that $Y_i = \Phi(X_i)$ for $i = 1, 2, \ldots$, where $\Phi(\cdot)$ is some deterministic function. Then we can write

$$(Y_1, \ldots, Y_n) = (\Phi(X_1), \ldots, \Phi(X_n)) = F(X_1, \ldots, X_n),$$

or

$$(Y_1^n) = F(X_1^n),$$

for some deterministic multivariate function $F(\cdot)$. Now using the chain rule we have

$$H(X_1^n, Y_1^n) = \begin{cases} H(X_1^n) + \overbrace{H(Y_1^n \mid X_1^n)}^{=0}, \\ H(Y_1^n) + \overbrace{H(X_1^n \mid Y_1^n)}^{\geq 0}, \end{cases}$$

so we can conclude that $H(X_1^n) \geq H(Y_1^n)$. Then we take the limit as follows

$$\lim_{n\to\infty} \frac{1}{n} H(X_1^n) \geq \lim_{n\to\infty} \frac{1}{n} H(Y_1^n),$$

which results in

$$H(\mathcal{Y}) \leq H(\mathcal{X}).$$

(b) Refer to the proof of part (c).

(c) $Z_i = \Psi(X_i, \ldots, X_{i+l})$ for $i = 1, 2, \ldots$, and $i \leq l \leq n$ where l is a fixed number. Then we can write

$$(Z_1, \ldots, Z_{n-l}) = [\Psi(X_1, \ldots, X_{1+l}), \ldots, \Psi(X_{n-l}, \ldots, X_n)] = F(X_1, \ldots, X_n),$$

where $F(\cdot)$ is a deterministic multivariate function. Using the same argument given in part (a) we have

$$H\left(Z_1^{n-l}\right) \leq H(X_1^n).$$

Multiplying both side by $1/n$ and taking the limit we obtain

$$\lim_{n\to\infty} \frac{n-l}{n} \frac{1}{n-l} H\left(Z_1^{n-l}\right) \leq \lim_{n\to\infty} \frac{1}{n} H\left(X_1^n\right),$$

or we can write

$$\lim_{n\to\infty} \frac{1}{n-l} H\left(Z_1^{n-l}\right) \leq \lim_{n\to\infty} \frac{1}{n} H\left(X_1^n\right),$$

because for fixed values of l we have $\lim_{n\to\infty} \frac{n-l}{n} = 1$. Then we can conclude that

$$H(\mathcal{Z}) \leq H(\mathcal{Y}).$$

(d) We have a second order Markov process over alphabet $\{0.1\}$ which realized as follows

$$X_n = \begin{cases} \text{Bernoulli}\,(0.5) & \text{if} \quad X_{n-1} = X_{n-2}, \\ \\ \text{Bernoulli}\,(0.9) & \text{if} \quad X_{n-1} \neq X_{n-2}. \end{cases}$$

Let us define the Markov process $Z_n = (X_n, X_{n+1})$. Obviously Z_n is a first order Markov process or a Markov chain (it is a process that the current state only depends on the last state). From the definition of process Z_n we can write

$$H(Z_1, \ldots, Z_n) = H(X_1, \ldots, X_{n+1}).$$

Taking the limit we can write

$$\lim_{n\to\infty} \frac{1}{n} H(Z_1, \ldots, Z_n) = \lim_{n\to\infty} \frac{n+1}{n} \frac{1}{n+1} H(X_1, \ldots, X_{n+1}),$$

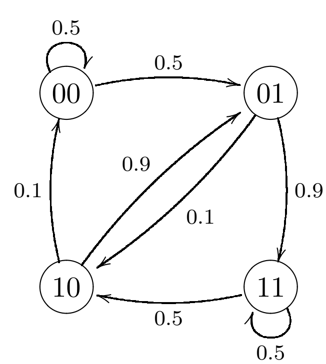
so we have

$$H(\mathcal{Z}) = H(\mathcal{X}).$$

To find the entropy rate of process X_n we can find the entropy rate of process Z_n that is a first order Markov chain with the states: $00, 01, 10, 11$. From the definition of the process X_n we can derive the transition matrix of the Markov chain Z_n which is as follows

$$P = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0.1 & 0.9 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix} \end{array},$$

which is equivalent to the following transition graph



To find the entropy rate of the Markov chain Z_n firstly we have to find it stationary distribution μ which is the solution of the following system of equations

$$\mu = \mu P.$$

Solving the above system of equations we obtain

$$\begin{array}{cccc} & 00 & 01 & 10 & 11 \\ \mu = & [0.087 & 0.4352 & 0.4352 & 0.7833] . \end{array}$$

Then for the entropy rates we can write

$$H(\mathcal{X}) = H\left(\mathcal{Z}\right) = H\left(Z_2 \mid Z_1\right) = \sum_{i \in \{00,01,10,11\}} \mu_i \sum_{j \in \{00,01,10,11\}} P_{ij} \log P_{ij}.$$