

Information Theory HW2

許博翔

October 5, 2023

Problem 1.

(a) Define $Q_X(x) = q_x$.

$$\begin{aligned} H(X) + \sum_{i=1}^{\infty} p_i \log q_i &= \sum_{i=1}^{\infty} \mathbb{E} \left[\log \frac{Q_X}{P_X} \right] \stackrel{\cdot \log \text{ is concave}}{\leq} \log \mathbb{E} \left[\frac{Q_X}{P_X} \right] = \log \left(\sum_{i=1}^{\infty} p_i \cdot \frac{q_i}{p_i} \right) = \\ &= \log \left(\sum_{i=1}^{\infty} q_i \right) = \log 1 = 0. \\ \therefore H(X) &\leq - \sum_{i=1}^{\infty} p_i \log q_i. \end{aligned}$$

(b) $-\log q_i$ is an arithmetic sequence $\Rightarrow q_i$ is a geometric sequence.

Suppose that $q_i = q_0 r^i$, where $1 < r < 1$ and $q_0 > 0$.

$$\begin{aligned} \because 1 &= \sum_{i=1}^{\infty} q_i = \frac{q_0 r}{1-r} \\ \Rightarrow q_0 &= \frac{1-r}{r}. \\ \because \mu_X &= \sum_{i=1}^{\infty} i q_i = \sum_{i=1}^{\infty} q_0 i r^i = \sum_{i=1}^{\infty} \sum_{j=1}^i q_0 r^i = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} q_0 r^i = \sum_{j=1}^{\infty} \frac{q_0 r^j}{1-r} = \frac{q_0 r}{(1-r)^2} \\ \Rightarrow \frac{1}{1-r} &= \mu_X \\ \therefore r &= 1 - \frac{1}{\mu_X} = \frac{\mu_X - 1}{\mu_X}, \quad q_0 = \frac{\frac{1}{\mu_X}}{1 - \frac{1}{\mu_X}} = \frac{1}{\mu_X - 1}. \\ -\log q_i &= -\log q_0 r^i = -\log q_0 - i \log r. \end{aligned}$$

Take $\alpha = -\log r = \log(\mu_X) - \log(\mu_X - 1)$, $\beta = -\log q_0 = \log(\mu_X - 1)$ satisfies the conditions.

$$\therefore \text{the answer is } q_i = \frac{(\mu_X - 1)^{i-1}}{(\mu_X)^i}, \quad \alpha = \log(\mu_X) - \log(\mu_X - 1), \quad \beta = \log(\mu_X - 1).$$

$$\begin{aligned} \text{(c)} \quad - \sum_{i=1}^{\infty} p_i \log q_i &= \sum_{i=1}^{\infty} p_i (\alpha i + \beta) = \alpha \mu_X + \beta = \log(\mu_X) \mu_X - \log(\mu_X - 1) \mu_X + \\ &+ \log(\mu_X - 1) = \mu_X (\log(\mu_X) - (1 - \frac{1}{\mu_X}) \log(\mu_X - 1)) = \mu_X (\log(\frac{\mu_X}{\mu_X - 1}) - \frac{1}{\mu_X} \log(\frac{\mu_X}{\mu_X - 1})) + \end{aligned}$$

$\frac{1}{\mu_X} \log(\mu_X) = \mu_X \left(-\left(1 - \frac{1}{\mu_X}\right) \log\left(1 - \frac{1}{\mu_X}\right) - \frac{1}{\mu_X} \log\left(\frac{1}{\mu_X}\right) \right) = \mu_X h_b(\mu_X^{-1})$.
 $\therefore H(X) \leq \mu_X h_b(\mu_X^{-1})$, and the equation holds when $p_i = q_i$ for all i , that is,
 $X \sim \text{Geo}\left(\frac{1}{\mu_X}\right)$ is the geometric distribution.

Problem 2.

$$\begin{aligned} \text{(a)} \quad & \int_2^\infty \frac{1}{x(\log x)^\alpha} dx = \int_{x=2}^\infty (\log x)^{-\alpha} d(\log x) \\ & = \begin{cases} \frac{1}{1-\alpha} (\log x)^{1-\alpha} \Big|_{x=2}^\infty & , \text{ if } \alpha \neq 1, \text{ which converges } \iff 1-\alpha < 0 \iff \alpha > 1, \\ \log \log x \Big|_{x=2}^\infty & , \text{ if } \alpha = 1, \text{ which does not converges} \end{cases} \\ & \quad \text{since } \lim_{y \rightarrow \infty} y^a = 0 \text{ for } a < 0, \text{ and } \lim_{y \rightarrow \infty} y^a \text{ does not exist for } a > 0. \\ & \therefore \sum_{n=2}^\infty \frac{1}{n(\log n)^\alpha} \text{ converges } \iff \alpha > 1. \end{aligned}$$

(b) First, we know that the series converges $\iff \alpha > 1$, so we only consider $\alpha > 1$.

$$\begin{aligned} H(X_\alpha) &= -E(\log P_{X_\alpha}) = \sum_{n=2}^\infty \frac{1}{s_\alpha n (\log n)^\alpha} \log(s_\alpha n (\log n)^\alpha) = \sum_{n=2}^\infty \frac{\log s_\alpha}{s_\alpha n (\log n)^\alpha} + \\ & \sum_{n=2}^\infty \frac{1}{s_\alpha n (\log n)^{\alpha-1}} + \sum_{n=2}^\infty \frac{\alpha \log \log n}{s_\alpha n (\log n)^\alpha}. \end{aligned}$$

For $\alpha \leq 2$, since $H(X_\alpha) > \sum_{n=2}^\infty \frac{1}{s_\alpha n (\log n)^{\alpha-1}} \rightarrow \infty$ from (a); therefore $H(X_\alpha)$ diverges to ∞ .

$$\begin{aligned} \text{For } \alpha > 2, \text{ since } H(X_\alpha) &< \sum_{n=2}^\infty \frac{\log s_\alpha}{s_\alpha n (\log n)^\alpha} + \sum_{n=2}^\infty \frac{1}{s_\alpha n (\log n)^{\alpha-1}} + \sum_{n=2}^\infty \frac{\alpha \log \log n}{s_\alpha n (\log n)^\alpha} \\ \log \log n &< \log n \text{ for } n \geq 2 \\ &< \sum_{n=2}^\infty \frac{\log s_\alpha}{s_\alpha n (\log n)^\alpha} + \sum_{n=2}^\infty \frac{1}{s_\alpha n (\log n)^{\alpha-1}} + \sum_{n=2}^\infty \frac{\alpha}{s_\alpha n (\log n)^{\alpha-1}} \\ &= \log s_\alpha + \frac{(1+\alpha)s_{\alpha-1}}{s_\alpha} < \infty, \end{aligned}$$

and $\sum_{n=2}^m \frac{1}{s_\alpha n (\log n)^\alpha} \log(s_\alpha n (\log n)^\alpha)$ is increasing as m increases.

$$\Rightarrow H(X_\alpha) = \sum_{n=2}^\infty \frac{1}{s_\alpha n (\log n)^\alpha} \log(s_\alpha n (\log n)^\alpha) \text{ converges.}$$

$\therefore H(X_\alpha)$ exists if $\alpha > 2$, and diverges to ∞ if $1 < \alpha \leq 2$.

Problem 3. Note that $P_{X_{\Theta_i}[i]}(\theta_i, x_i)$ is defined as $\Pr\{\Theta_i = \theta_i \wedge X_{\theta_i}[i] = x_i\}$, while $P_{X_{\theta_i}[i]}(x_i)$ is defined as $\Pr\{X_{\theta_i}[i] = x_i\}$.

Since $X_{\theta_i}[i]$ and Θ_i are independent, there is $P_{X_{\Theta_i}[i]}(\theta_i, x_i) = P_{\Theta_i}(\theta_i)P_{X_{\theta_i}[i]}(x_i)$.

- (a) $\because \forall l, n \in \mathbb{N}, P_{X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_n}[n]} = P_{X_{\Theta}[1], X_{\Theta}[2], \dots, X_{\Theta}[n]}$
 X_{Θ} is stationary no matter Θ is 0 or 1 $\stackrel{=}{=} P_{X_{\Theta}[l+1], X_{\Theta}[l+2], \dots, X_{\Theta}[l+n]} = P_{X_{\Theta_{l+1}}[l+1], X_{\Theta_{l+2}}[l+2], \dots, X_{\Theta_{l+n}}[l+n]}.$

$\therefore \{X_{\Theta_i}[i]\}$ is stationary.

By the definition of entropy rates,

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\log P_{X_k[1], X_k[2], \dots, X_k[n]}] &= \lim_{n \rightarrow \infty} \frac{1}{n} H(X_k[1], X_k[2], \dots, X_k[n]) = \mathcal{H}_k. \\ \Rightarrow \mathcal{H}(\{X_{\Theta_i}[i]\}) &= \lim_{n \rightarrow \infty} \frac{1}{n} H(X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_n}[n]) \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\log P_{X_{\Theta}[1], X_{\Theta}[2], \dots, X_{\Theta}[n]}] \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} (\Pr\{\Theta = 0\} \mathbb{E}[\log \Pr\{\Theta = 0\} P_{X_0[1], X_0[2], \dots, X_0[n]}] \\ &\quad + \Pr\{\Theta = 1\} \mathbb{E}[\log \Pr\{\Theta = 1\} P_{X_1[1], X_1[2], \dots, X_1[n]}]) \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} ((1-q) \mathbb{E}[\log(1-q) + \log P_{X_0[1], X_0[2], \dots, X_0[n]}] + q \mathbb{E}[\log q + \log P_{X_1[1], X_1[2], \dots, X_1[n]}]) \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} ((1-q) \log(1-q) + q \log q) + (1-q) \mathcal{H}_0 + q \mathcal{H}_1 = (1-q) \mathcal{H}_0 + q \mathcal{H}_1. \end{aligned}$$

- (b) Suppose $\Theta_1 \sim \text{Ber}(q)$.

$$\text{Since } \{\Theta_i\} \text{ is stationary, } \begin{pmatrix} 1-q & q \end{pmatrix} = \begin{pmatrix} 1-q & q \end{pmatrix} \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

$$\Rightarrow 1-q = (1-q)(1-\alpha) + q\beta$$

$$\Rightarrow \alpha(1-q) = q\beta$$

$$\Rightarrow q = \frac{\alpha}{\alpha + \beta}.$$

$$\because P_{X_{\Theta_{i+1}}[i+1]|X_{\Theta_i}[i]} \stackrel{X_j[i] \text{ is independent of } X_{j'}[i'] \text{ for any } (i', j') \neq (i, j)}{=} P_{\Theta_{i+1}|\Theta_i} P_{X_{\Theta_{i+1}}[i+1]}.$$

$$\therefore P_{X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_n}[n]} = \prod_{i=1}^n P_{X_{\Theta_i}[i]|X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_{i-1}}[i-1]} = P_{X_{\Theta_1}[1]} \prod_{i=2}^n P_{X_{\Theta_i}[i]|X_{\Theta_{i-1}}[i-1]}$$

$$= P_{\Theta_1} P_{X_{\Theta_1}[1]} \prod_{i=2}^n P_{\Theta_i|\Theta_{i-1}} P_{X_{\Theta_i}[i]} = \left(P_{\Theta_1} \prod_{i=2}^n P_{\Theta_i|\Theta_{i-1}} \right) \prod_{i=1}^n P_{X_{\Theta_i}[i]}$$

$$X_j[i] \text{ is independent of } X_{j'}[i'] \text{ for any } (i', j') \neq (i, j) \stackrel{=}{=} P_{\Theta_1, \Theta_2, \dots, \Theta_n} P_{X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_n}[n]}.$$

$$\Rightarrow \forall l, n \in \mathbb{N}, P_{X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_n}[n]} = P_{\Theta_1, \Theta_2, \dots, \Theta_n} P_{X_{\Theta_1}[1], X_{\Theta_2}[2], \dots, X_{\Theta_n}[n]} \stackrel{\{X_0[i]\}, \{X_1[i]\}, \{\Theta_i\} \text{ are stationary}}{=}$$

$$P_{\Theta_{l+1}, \Theta_{l+2}, \dots, \Theta_{l+n}} P_{X_{\Theta_{l+1}}[l+1], X_{\Theta_{l+2}}[l+2], \dots, X_{\Theta_{l+n}}[l+n]} = P_{X_{\Theta_{l+1}}[l+1], X_{\Theta_{l+2}}[l+2], \dots, X_{\Theta_{l+n}}[l+n]}.$$

$\therefore \{X_{\Theta_i}[i]\}$ is stationary.

By theorem 11, $\mathcal{H}(\{X_{\Theta_i}[i]\}) = H(X_{\Theta_2}[2]|X_{\Theta_1}[1])$

$$= - \sum_{\theta_1, \theta_2, x_1, x_2} P_{X_{\Theta_1}[1]}(\theta_1, x_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) P_{X_{\Theta_2}[2]}(x_2) (\log(P_{\Theta_2|\Theta_1}(\theta_2|\theta_1)) + \log(P_{X_{\Theta_2}[2]}(x_2)))$$

$$= - \sum_{\theta_1, \theta_2, x_1, x_2} P_{X_{\Theta_1}[1]}(\theta_1, x_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) P_{X_{\Theta_2}[2]}(x_2) \log(P_{\Theta_2|\Theta_1}(\theta_2|\theta_1))$$

$$\begin{aligned}
& - \sum_{\theta_1, \theta_2, x_1, x_2} P_{X_{\Theta_1}[1]}(\theta_1, x_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) P_{X_{\theta_2}[2]}(x_2) \log(P_{X_{\theta_2}[2]}(x_2)) \\
& = - \sum_{\theta_1, \theta_2, x_1} P_{X_{\Theta_1}[1]}(\theta_1, x_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) \log(P_{\Theta_2|\Theta_1}(\theta_2|\theta_1)) \\
& + \sum_{\theta_1, \theta_2, x_1} P_{X_{\Theta_1}[1]}(\theta_1, x_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) H(X_{\theta_2}[2]) \\
& = - \sum_{\theta_1, \theta_2} P_{\Theta_1}(\theta_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) \log(P_{\Theta_2|\Theta_1}(\theta_2|\theta_1)) \\
& + \sum_{\theta_1, \theta_2} P_{\Theta_1}(\theta_1) P_{\Theta_2|\Theta_1}(\theta_2|\theta_1) H(X_{\theta_2}[2]) \\
& = -(1-q)(1-\alpha) \log(1-\alpha) - (1-q)\alpha \log(\alpha) - q\beta \log(\beta) - q(1-\beta) \log(1-\beta) \\
& + H(X_0[2])((1-q)(1-\alpha) + q\beta) + H(X_1[2])((1-q)\alpha + q(1-\beta)) \\
& \{X_k[i]\} \text{ are i.i.d.} \Rightarrow \underline{\mathcal{H}_k} = H(\{X_k[i]\}) = H(X_k[i]) \quad (1-q)H_b(\alpha) + qH_b(\beta) + \mathcal{H}_0((1-q)(1-\alpha) + q\beta) + \mathcal{H}_1((1-q)\alpha + q(1-\beta)) \\
& = \frac{\beta}{\alpha + \beta} H_b(\alpha) + \frac{\alpha}{\alpha + \beta} H_b(\beta) + \mathcal{H}_0\left(\frac{\beta}{\alpha + \beta}(1-\alpha) + \frac{\alpha}{\alpha + \beta}\beta\right) + \mathcal{H}_1\left(\frac{\beta}{\alpha + \beta}\alpha + \frac{\alpha}{\alpha + \beta}(1-\beta)\right) \\
& = \frac{\beta}{\alpha + \beta} (H_b(\alpha) + \mathcal{H}_0) + \frac{\alpha}{\alpha + \beta} (H_b(\beta) + \mathcal{H}_1).
\end{aligned}$$