# Information Theory HW3

許博翔

October 17, 2023

**Problem 1.**

(a) Consider $\phi_{\tau,\gamma}(x) := \begin{cases} 1, \text{ if } LR(x) > \tau \\ \gamma, \text{ if } LR(x) = \tau \\ 0, \text{ if } LR(x) < \tau \end{cases}$.

$LR(0) = \dfrac{P_1(0)}{P_0(0)} = \dfrac{1-p_1}{1-p_0}$.

$LR(1) = \dfrac{P_1(1)}{P_0(1)} = \dfrac{p_1}{p_0}$.

$\because p_0 < p_1$.

$\therefore LR(1)\dfrac{p_1}{p_0} > 1 > \dfrac{1-p_1}{1-p_0} > LR(0)$.

By Neyman-Pearson theorem, $\phi_{\tau,\gamma}$ is optimal.

$\pi_{1|0}(\phi_{\tau,\gamma}) = P_0\{LR(X) > \tau\} + \gamma P_0\{LR(X) = \tau\}$.

$\pi_{0|1}(\phi_{\tau,\gamma}) = P_1\{LR(X) < \tau\} + (1-\gamma) P_1\{LR(X) = \tau\}$.

We only need to consider the cases $\tau = LR(x)$ for some $x$, since other cases can be reduced to these cases by setting $\gamma$ properly.

For $\tau = LR(0)$, $\pi_{1|0} = P_0(1) + \gamma P_0(0) = p_0 + \gamma(1-p_0)$; $\pi_{0|1} = 0 + (1-\gamma)P_1(0) = (1-\gamma)(1-p_1)$.

For $\tau = LR(1)$, $\pi_{1|0} = 0 + \gamma P_0(1) = \gamma p_0$; $\pi_{0|1} = P_1(0) + (1-\gamma)P_1(1) = 1 - p_1 + (1-\gamma)p_1$.

**Problem 2.**

(a) $\displaystyle\int_2^\infty \frac{1}{x(\log x)^\alpha}dx = \int_{x=2}^\infty (\log x)^{-\alpha} d(\log x)$

$$= \begin{cases} \dfrac{1}{1-\alpha}(\log x)^{1-\alpha}\Big|_{x=2}^{\infty} & \text{, if } \alpha \neq 1 \text{, which converges } \iff 1-\alpha < 0 \iff \alpha > 1, \\ & \text{since } \lim\limits_{y\to\infty} y^a = 0 \text{ for } a < 0 \text{, and } \lim\limits_{y\to\infty} y^a \text{ does not exist for } a > 0\cdot \\ \log\log x|_{x=2}^{\infty} & \text{, if } \alpha = 1 \text{, which does not converges} \end{cases}$$

$\therefore \displaystyle\sum_{n=2}^\infty \frac{1}{n(\log n)^\alpha}$ converges $\iff \alpha > 1$.

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

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

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

For $\alpha > 2$, since $H(X_\alpha) < \displaystyle\sum_{n=2}^\infty \frac{\log s_\alpha}{s_\alpha n(\log n)^\alpha} + \sum_{n=2}^\infty \frac{1}{s_\alpha n(\log n)^{\alpha-1}} + \sum_{n=2}^\infty \frac{\alpha\log\log n}{s_\alpha n(\log n)^\alpha}$

$\displaystyle\overset{\log\log n<\log n \text{ for } n\geq 2}{<} \sum_{n=2}^\infty \frac{\log s_\alpha}{s_\alpha n(\log n)^\alpha} + \sum_{n=2}^\infty \frac{1}{s_\alpha n(\log n)^{\alpha-1}} + \sum_{n=2}^\infty \frac{\alpha}{s_\alpha n(\log n)^{\alpha-1}}$

$= \log s_\alpha + \dfrac{(1+\alpha)s_{\alpha-1}}{s_\alpha} < \infty$,

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

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

$\therefore H(X_\alpha)$ exists if $\alpha > 2$, and diverges to $\infty$ 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 $\Theta_i$ 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) $\because \forall l, n \in \mathbb{N},\ P_{X_{\Theta_1}[1], X_{\Theta_2}[2], \ldots, X_{\Theta_n}[n]} = P_{X_\Theta[1], X_\Theta[2], \ldots, X_\Theta[n]}$

$\overset{X_\Theta \text{ is stationary no matter } \Theta \text{ is 0 or 1}}{=} P_{X_\Theta[l+1], X_\Theta[l+2], \ldots, X_\Theta[l+n]} = P_{X_{\Theta_{l+1}}[l+1], X_{\Theta_{l+2}}[l+2], \ldots, X_{\Theta_{l+n}}[l+n]}\cdot$

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

By the definition of entropy rates,

$\displaystyle\lim_{n\to\infty} -\frac{1}{n}\mathrm{E}[\log P_{X_k[1], X_k[2], \ldots, X_k[n]}] = \lim_{n\to\infty}\frac{1}{n}H(X_k[1], X_k[2], \ldots, X_k[n]) = \mathcal{H}_k$.

$$\Rightarrow \mathcal{H}(\{X_{\Theta_i}[i]\}) = \lim_{n\to\infty} \frac{1}{n} H(X_{\Theta_1}[1], X_{\Theta_2}[2], \ldots, X_{\Theta_n}[n])$$
$$= \lim_{n\to\infty} -\frac{1}{n} \mathrm{E}[\log P_{X_\Theta[1],X_\Theta[2],\ldots,X_\Theta[n]}]$$
$$= \lim_{n\to\infty} -\frac{1}{n} \big(\Pr\{\Theta=0\}\mathrm{E}[\log \Pr\{\Theta=0\} P_{X_0[1],X_0[2],\ldots,X_0[n]}]$$
$$+\Pr\{\Theta=1\}\mathrm{E}[\log \Pr\{\Theta=1\} P_{X_1[1],X_1[2],\ldots,X_1[n]}]\big)$$
$$= \lim_{n\to\infty} -\frac{1}{n} \big((1-q)\mathrm{E}[\log(1-q) + \log P_{X_0[1],X_0[2],\ldots,X_0[n]}] + q\mathrm{E}[\log q + \log P_{X_1[1],X_1[2],\ldots,X_1[n]}]\big)$$
$$= \lim_{n\to\infty} -\frac{1}{n}((1-q)\log(1-q) + q\log q) + (1-q)\mathcal{H}_0 + q\mathcal{H}_1 = (1-q)\mathcal{H}_0 + q\mathcal{H}_1.$$

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

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

$$\Rightarrow 1-q = (1-q)(1-\alpha) + q\beta$$
$$\Rightarrow \alpha(1-q) = q\beta$$
$$\Rightarrow q = \frac{\alpha}{\alpha+\beta}.$$

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

$$\therefore P_{X_{\Theta_1}[1],X_{\Theta_2}[2],\ldots,X_{\Theta_n}[n]} = \prod_{i=1}^{n} P_{X_{\Theta_i}[i]|X_{\Theta_1}[1],X_{\Theta_2}[2],\ldots,X_{\Theta_{i-1}}[i-1]} = P_{X_{\Theta_1}[1]} \prod_{i=2}^{n} P_{X_{\Theta_i}[i]|X_{\Theta_{i-1}}[i-1]}$$
$$= P_{\Theta_1} P_{X_{\theta_1}[1]} \prod_{i=2}^{n} P_{\Theta_i|\Theta_{i-1}} P_{X_{\theta_i}[i]} = \left(P_{\Theta_1} \prod_{i=2}^{n} P_{\Theta_i|\Theta_{i-1}}\right) \prod_{i=1}^{n} P_{X_{\theta_i}[i]}$$
$$\overset{X_j[i] \text{ is independent of } X_{j'}[i'] \text{ for any } (i',j')\neq(i,j)}{=} P_{\Theta_1,\Theta_2,\ldots,\Theta_n} P_{X_{\theta_1}[1],X_{\theta_2}[2],\ldots,X_{\theta_n}[n]}.$$

$$\Rightarrow \forall l,n \in \mathbb{N},\ P_{X_{\Theta_1}[1],X_{\Theta_2}[2],\ldots,X_{\Theta_n}[n]} = P_{\Theta_1,\Theta_2,\ldots,\Theta_n} P_{X_{\theta_1}[1],X_{\theta_2}[2],\ldots,X_{\theta_n}[n]} \overset{\{X_0[i]\},\{X_1[i]\},\{\Theta_i\} \text{ are stationary}}{=}$$
$$P_{\Theta_{l+1},\Theta_{l+2},\ldots,\Theta_{l+n}} P_{X_{\theta_1}[l+1],X_{\theta_2}[l+2],\ldots,X_{\theta_n}[l+n]} = P_{X_{\Theta_{l+1}}[l+1],X_{\Theta_{l+2}}[l+2],\ldots,X_{\Theta_{l+n}}[l+n]}.$$

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

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

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

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