Information Theory HW3

許博翔

October 19, 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.

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

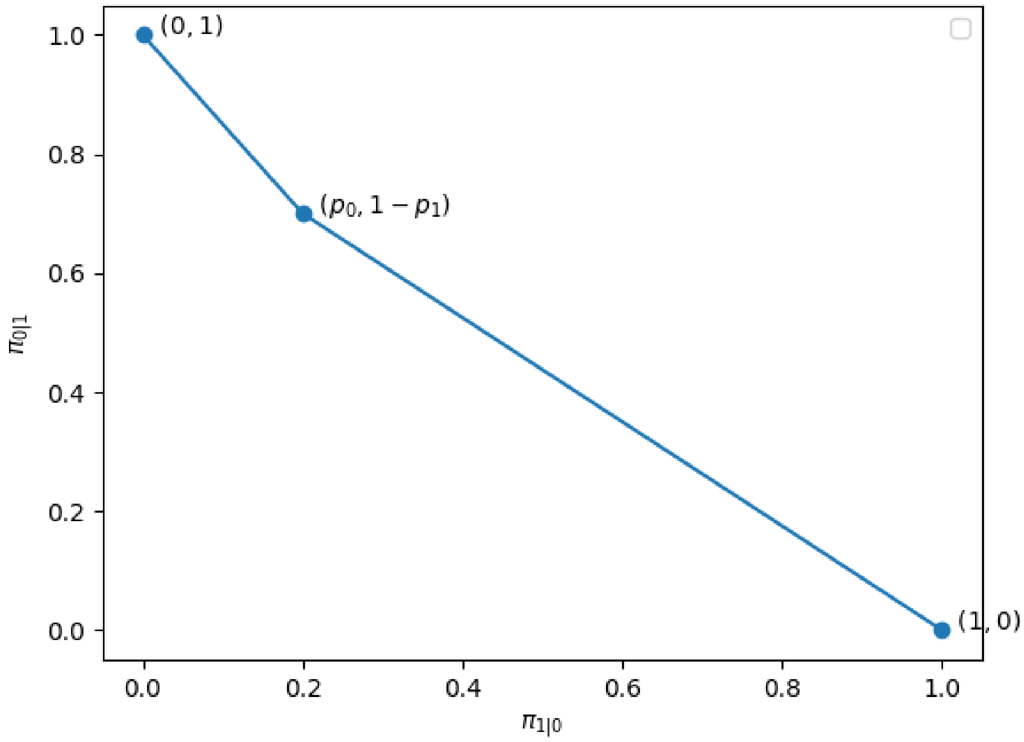
$\pi_{0|1}(\phi_{\tau,\gamma}) = P_1\{LR(X) < \tau\} + (1-\gamma) P_1\{LR(X) = \tau\}$.

We only need to consider the cases $\tau = LR(x)$ for some x, since other cases can be reduced to these cases by setting γ 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 $\tau = LR(1)$, $\pi_{1|0} = 0 + \gamma P_0(1) = \gamma p_0$; $\pi_{0|1} = P_1(0) + (1-\gamma)P_1(1) = 1 - p_1 + (1-\gamma)p_1$.

The above forms two segments, and their intersection is $(p_0, 1-p_1)$, which can be calculated by setting γ in the first segment to 0 or in the second segment to 1.



(b) Let Y be the random variable denoting the length of the observed sequence. We can see that $P_Y(y) = p(1-p)^{y-1}$.

$P\{Y > y\} = \sum_{z=y+1}^{\infty} p(1-p)^{z-1} = \frac{p(1-p)^y}{1-(1-p)} = (1-p)^y$.

$P\{Y < y\} = \sum_{z=1}^{y-1} p(1-p)^{z-1} = \frac{p(1-(1-p)^{y-1})}{1-(1-p)} = 1-(1-p)^{y-1}$.

$P_0(y) = p_0(1-p_0)^{y-1}, P_1(y) = p_1(1-p_1)^{y-1}$.

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

$LR(y) = \frac{P_1(y)}{P_0(y)} = \frac{p_1(1-p_1)^{y-1}}{p_0(1-p_0)^{y-1}}$.

Since $p_0 < p_1$, there is $\frac{1-p_1}{1-p_0} < 1$.

$\Rightarrow LR(y)$ is an decreasing function of y.

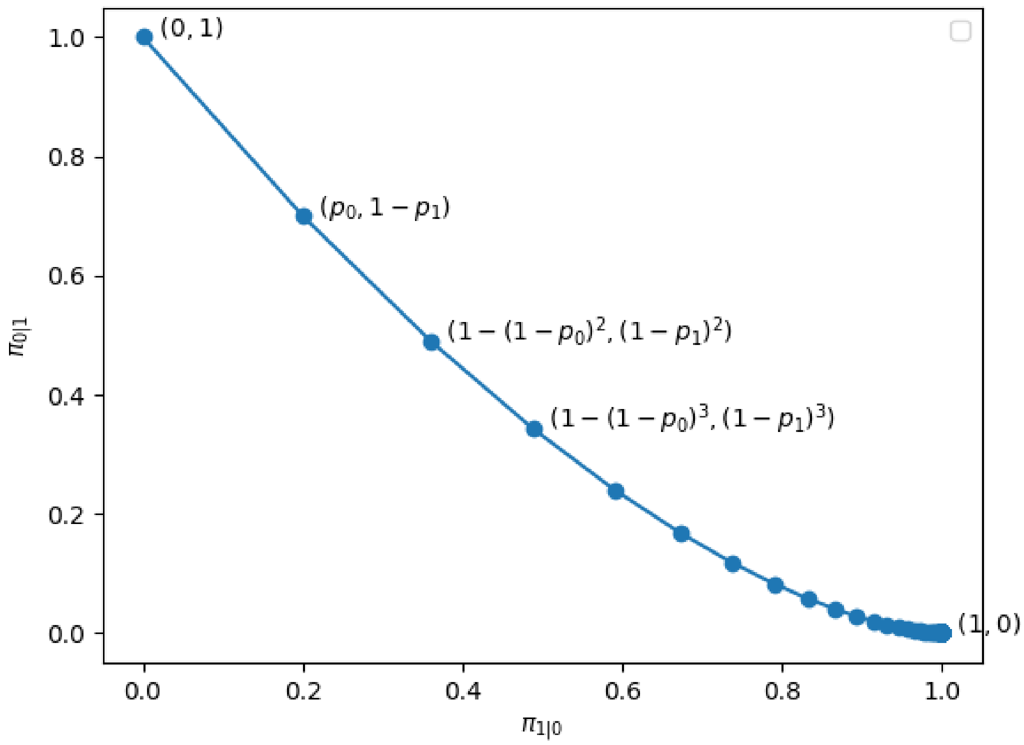
By Neyman-Pearson theorem, $\phi_{\tau,\gamma}$ is optimal.

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

Since $LR(y)$ is decreasing, for $\tau = LR(y)$, $\pi_{1|0}(\phi_{\tau,\gamma}) = P_0\{Y < y\} + \gamma P_0\{Y = y\} = 1 - (1-p_0)^{y-1} + \gamma p_0(1-p_0)^{y-1} = 1 - (1-p_0)^{y-1}(1-\gamma p_0)$.

$\pi_{0|1}(\phi_{\tau,\gamma}) = P_1\{Y > y\} + (1-\gamma)P_1\{Y = y\} = (1-p_1)^y + (1-\gamma)p_1(1-p_1)^{y-1} = (1-\gamma p_1)(1-p_1)^{y-1}$.

For each y, it forms a segment, where the intersection of the segments formed by y and $y+1$ is $(1-(1-p_0)^y, (1-p_1)^y)$, which can be calculated by setting γ in the segment formed by y to 1 or in the other segment to 0.



(c) Let Y be the random variable denoting the length of the observed sequence. The probability that a given sequence with length y and n 1s appears is $p^n(1-p)^{y-n}$, and there are $\binom{y}{n}$ sequences of this kind.

$\therefore P\{Y = y\} = \binom{y}{n} p^n (1-p)^{y-n}$.

Note that if $a < b$ or $b < 0$, then $\binom{a}{b}$ is defined as 0.

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

$LR(y) = \dfrac{P_1(y)}{P_0(y)} = \dfrac{\binom{y}{n} p_1^n (1-p_1)^{y-n}}{\binom{y}{n} p_0^n (1-p_0)^{y-n}} = \dfrac{p_1^n (1-p_1)^{y-n}}{p_0^n (1-p_0)^{y-n}}$.

Since $p_0 < p_1$, there is $\dfrac{1-p_1}{1-p_0} < 1$.

$\Rightarrow LR(y)$ is an decreasing function of y.

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

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

Since $LR(y)$ is decreasing, for $\tau = LR(y)$, $\pi_{1|0}(\phi_{\tau,\gamma}) = P_0\{Y < y\} + \gamma P_0\{Y = y\} = \displaystyle\sum_{z=0}^{y-1} \binom{z}{n} p_0^n (1-p_0)^{z-n} + \gamma \binom{y}{n} p_0^n (1-p_0)^{y-n}$.

$\pi_{0|1}(\phi_{\tau,\gamma}) = P_1\{Y > y\} + (1-\gamma) P_1\{Y = y\} = \displaystyle\sum_{z=y+1}^{\infty} \binom{z}{n} p_1^n (1-p_1)^{z-n} + (1-\gamma) \binom{y}{n} p_1^n (1-p_1)^{y-n}$.

The optimal solution is $\pi_{1|0}(\phi_{\tau,\gamma}) = \displaystyle\sum_{z=0}^{y-1} \binom{z}{n} p_0^n (1-p_0)^{z-n} + \gamma \binom{y}{n} p_0^n (1-p_0)^{y-n} = \epsilon$, where y is the minimum integer such that $\displaystyle\sum_{z=0}^{y} \binom{z}{n} p_0^n (1-p_0)^{z-n} \geq \epsilon$.

$\Rightarrow \displaystyle\sum_{z=0}^{y} \binom{z}{n} p_1 (1-p_1)^{z-n} \ \pi_{0|1}(\phi_{\tau,\gamma})$

Problem 2.

(a) $\pi_0^{(n)}(x^n) = \Pr\{\Theta = 0 | X^n = x^n\} = \Pr\{X_i \overset{\text{i.i.d.}}{\sim} P_0 | X^n = x^n\} = \dfrac{\Pr\{X_i \overset{\text{i.i.d.}}{\sim} P_0 \wedge X^n = x^n\}}{\Pr\{X^n = x^n\}} = \dfrac{\Pr\{X_i \overset{\text{i.i.d.}}{\sim} P_0 \wedge X^n = x^n\}}{\Pr\{(X_i \overset{\text{i.i.d.}}{\sim} P_0 \wedge X^n = x^n) \vee (X_i \overset{\text{i.i.d.}}{\sim} P_1 \wedge X^n = x^n)\}} = \dfrac{\pi_0^{(0)} \prod_{i=1}^{n} P_0(x_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(x_i) + \pi_1^{(0)} \prod_{i=1}^{n} P_1(x_i)}$.

Similarly, $\pi_1^{(n)}(x^n) = \dfrac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(x_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(x_i) + \pi_1^{(0)} \prod_{i=1}^{n} P_1(x_i)}$.

(b) $-\frac{1}{n}\log \pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i) = -\frac{1}{n}\left(\log \pi_0^{(0)} + \sum_{i=1}^{n} \log(P_0(X_i))\right) \overset{\text{SLLN}}{\to} -\frac{1}{n}\log \pi_0^{(0)} - \mathrm{E}[\log(P_0(X))] \overset{\log \pi_0^{(0)} \text{ is a constant}}{\to} -\mathrm{E}[\log(P_0(X))] = H(X)$ as $n \to \infty$.

From HW2 we know that $H(X) \leq -\sum_{i=1}^{\infty} P_0(i)\log P_1(i)$, with equality $\iff$ $P_1 \sim P_0$.

$-\frac{1}{n}\log \pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i) = -\frac{1}{n}\left(\log \pi_1^{(0)} + \sum_{i=1}^{n} \log(P_1(X_i))\right) \overset{\text{SLLN}}{\to} -\frac{1}{n}\log \pi_1^{(0)} - \mathrm{E}[\log(P_1(X))] \overset{\log \pi_1^{(0)} \text{ is a constant}}{\to} -\mathrm{E}[\log(P_1(X))] > H(X)$ as $n \to \infty$.

$\Rightarrow \dfrac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i)} \to \exp(nE[\log(P_1(X))] + nH(X)) = \exp(E[\log(P_1(X))] + H(X))^n \overset{\mathrm{E}[\log(P_1(X))]+H(X)<0}{\to} 0$ as $n \to \infty$.

$\therefore \pi_0^{(n)}(X^n) = \dfrac{\pi_0^{(0)} \prod_{i=1}^{n} P_0(x_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(x_i) + \pi_1^{(0)} \prod_{i=1}^{n} P_1(x_i)} = \dfrac{1}{1 + \frac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i)}} \to \dfrac{1}{1+0} = 1$ as $n \to \infty$.

As what we computed above, for any constant $c > 0$, $-\frac{1}{n}\log \dfrac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)}{c\pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i)} = H(X) + \mathrm{E}[\log(P_1(X))] + \frac{1}{n}\log c \overset{c \text{ is a constant}}{\to} H(X) + \mathrm{E}[\log(P_1(X))] = D(P_0\|P_1)$.

$\because$ log is an increasing function, and $\dfrac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i)} < \dfrac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)}{\pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i) + \pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)} = \pi_1^{(n)}(X^n) < \dfrac{\pi_1^{(0)} \prod_{i=1}^{n} P_1(X_i)}{2\pi_0^{(0)} \prod_{i=1}^{n} P_0(X_i)}$ when $n \to \infty$.

$\therefore$ by squeeze theorem, $-\frac{1}{n}\log \pi_1^{(n)}(X^n) \to D(P_0\|P_1)$ as $n \to \infty$.

Problem 3.

(a) Let $X \sim P$.

$\mathrm{D}(P\|G(p)) = \sum_{x=1}^{\infty} P(x)\log\frac{P(x)}{Q(x)} = \sum_{x=1}^{\infty} P(x)\log\frac{P(x)}{(1-p)p^{x-1}} = \mathrm{H}(X) - \mathrm{E}[\log((1-p)p^{X-1})] = \mathrm{H}(X) - \log(1-p) - \mathrm{E}[(X-1)\log(p)] = H(X) - \log(1-p) -$

$\log(p)E[X-1] = H(X) - \log(1-p) + \log p - \mu \log p$.
$\frac{d}{dp}\mathrm{D}(P\|G(p)) = \frac{1}{1-p} + \frac{1}{p} - \frac{1}{p}\mu = \frac{1-(1-p)\mu}{p(1-p)}$, which equals to 0 $\iff$ $\frac{1}{1-p} = \mu \iff p = 1 - \frac{1}{\mu}$.
One can also verify that if $p < 1 - \frac{1}{\mu}$, $\frac{d}{dp}\mathrm{D}(P\|G(p)) < 0$ and if $p > 1 - \frac{1}{\mu}$, $\frac{d}{dp}\mathrm{D}(P\|G(p)) > 0$.
$\therefore$ the minimum possible value of $\mathrm{D}(P\|G(p))$ occurs when $p = 1 - \frac{1}{\mu}$, that is, the distribution is $G(1-\frac{1}{\mu})$, and $\mathrm{D}(P\|G(p)) = H(X) - \log\mu + (1-\mu)\log(1-\mu)$.

(b) Let $X_i \sim P_i, Y \sim R$ where $R(y) := \frac{1}{m}\sum_{i=1}^{m} P_i(y)$.

From HW2 we know that $H(R) \leq -\sum_{j=1}^{\infty} R(j)\log Q(j)$, with equality $\iff$ $Q \sim R$.
$\Rightarrow \sum_{i=1}^{m}\mathrm{D}(P_i\|Q) = \sum_{i=1}^{m}\left(H(X_i) - \sum_{j=1}^{\infty} P_i(j)\log Q(j)\right) = \sum_{i=1}^{m} H(X_i) - \sum_{j=1}^{\infty}\left(\sum_{i=1}^{m} P_i(j)\right)\log Q(j) = \sum_{i=1}^{m} H(X_i) - m\sum_{j=1}^{\infty} R(j)\log Q(j) \geq \sum_{i=1}^{m} H(X_i) - mH(R)$.
$\therefore \min_{Q\in\mathcal{P}(X)} \sum_{i=1}^{m} D(P_i\|Q) = \sum_{i=1}^{m} H(X_i) - mH(R)$, with minimizer $Q = R$, that is, $Q(y) = \frac{1}{m}\sum_{i=1}^{m} P_i(y)$.