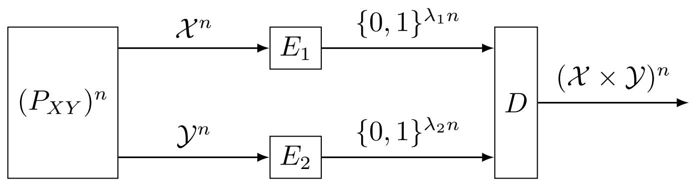
信息论

参考答案

1. (10 分) 本题中，信息熵 H 使用 2 作为底数. 令 P_{XY} 为一个支撑有限 $\mathcal{X}\times\mathcal{Y}$ 上的联合分布. 令常数 λ_1,λ_2 满足 $\lambda_1 > H(X|Y)$，$\lambda_2 > H(Y|X)$，$\lambda_1+\lambda_2 > H(X,Y)$.



(1) 请构造两个压缩函数 $E_1:\mathcal{X}^n\to\{0,1\}^{\lfloor\lambda_1 n\rfloor}$，$E_2:\mathcal{X}^n\to\{0,1\}^{\lfloor\lambda_2 n\rfloor}$，和一个解压缩函数 $D:\{0,1\}^{\lfloor\lambda_1 n\rfloor+\lfloor\lambda_2 n\rfloor}\to(\mathcal{X}\times\mathcal{Y})^n$，并证明

$$\Pr_{((X_1,Y_1),\ldots,(X_n,Y_n))\sim(P_{XY})^n}\Big[D(E_1(X_1,\ldots,X_n),E_2(Y_1,\ldots,Y_n))\neq((X_1,Y_1),\ldots,(X_n,Y_n))\Big]\le 2^{-\Theta(n)}.$$

(2) 如果改变参数, 满足如下条件之一: a) $\lambda_1 < H(X|Y)$, b) $\lambda_2 < H(Y|X)$, c) $\lambda_1+\lambda_2 < H(X,Y)$. 说明这时不能构造满足前一问要求的压缩函数和解压缩函数.

解

(1) 随机选取 E_1,E_2，然后我们证明 E_1,E_2 大概率是符合要求的压缩函数.

用 $X_{1:n},Y_{1:n}$ 分别表示 $(X_1,\ldots,X_n),(Y_1,\ldots,Y_n)$. 存在一个与 n 无关的实数 $\delta>0$ 使得 $\lambda_1 > H(X|Y)+2\delta$，$\lambda_2 > H(Y|X)+2\delta$，$\lambda_1+\lambda_2 > H(X,Y)+2\delta$. 根据 Chernoff bound（或 Sanov Theorem）

$$\begin{aligned}
\Pr[-\log(P_{XY}^n(X_{1:n},Y_{1:n}))\ge n(H(X,Y)+\delta)]&\le 2^{-\Theta(n)},\\
\Pr[-\log(P_{X|Y}^n(X_{1:n}|Y_{1:n}))\ge n(H(X|Y)+\delta)]&\le 2^{-\Theta(n)},\\
\Pr[-\log(P_{Y|X}^n(Y_{1:n}|X_{1:n}))\ge n(H(Y|X)+\delta)]&\le 2^{-\Theta(n)}.
\end{aligned}$$

据此我们定义集合 $\mathcal{D}_n$

$$\mathcal{D}_n := \left\{ \begin{matrix}(x_1,\ldots,x_n,\\ \quad y_1,\ldots,y_n)\end{matrix} \middle| \begin{matrix}\sum_{i=1}^n -\log(P_{XY}(x_i,y_i)) = -\log(P_{XY}^n(x_{1:n},y_{1:n}))\le n(H(X,Y)+\delta)\\ \sum_{i=1}^n -\log(P_{X|Y}(x_i|y_i)) = -\log(P_{X|Y}^n(x_{1:n}|y_{1:n}))\le n(H(X|Y)+\delta)\\ \sum_{i=1}^n -\log(P_{Y|X}(y_i|x_i)) = -\log(P_{Y|X}^n(y_{1:n}|x_{1:n}))\le n(H(Y|X)+\delta)\end{matrix} \right\}$$

从 P_{XY}^n 中采样得到的 $X_{1:n},Y_{1:n}$ 以 $1-2^{\Theta(n)}$ 的概率落在 $\mathcal{D}_n$ 当中. 我们定义解压缩函数 D 为

$$D(c_1,c_2)=(x_1,\ldots,x_n,y_1,\ldots,y_n)\text{ 如果存在唯一的}(x_1,\ldots,x_n,y_1,\ldots,y_n)\in\mathcal{D}_n$$
$$\text{满足}E_1(x_1,\ldots,x_n)=c_1\wedge E_2(y_1,\ldots,y_n)=c_2.$$

这个解压缩函数并不是最优的，但是方便分析.

对于任何 $(x_{1:n}, y_{1:n}) \in \mathcal{D}_n$，其可以正确被解压，即 $D(E_1((x_{1:n}), E_2(y_{1:n}))) = (x_{1:n}, y_{1:n})$，当且仅当没有另一个 $(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n$ 被编码到同样的 (c_1, c_2). 这个概率 (随机性来源于 E_1, E_2) 可以用 union bound 估计

$$\begin{aligned}
&\Pr[D(E_1((x_{1:n}), E_2(y_{1:n}))) \neq (x_{1:n}, y_{1:n})] \\
&\leq \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ (x'_{1:n}, y'_{1:n}) \neq (x_{1:n}, y_{1:n})}} \Pr[(E_1(x'_{1:n}), E_2(y'_{1:n})) = (E_1(x_{1:n}), E_2(y_{1:n}))] \\
&= \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} \neq x_{1:n} \wedge y'_{1:n} \neq y_{1:n}}} \Pr[(E_1(x'_{1:n}), E_2(y'_{1:n})) = (E_1(x_{1:n}), E_2(y_{1:n}))] \\
&\quad + \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} = x_{1:n} \wedge y'_{1:n} \neq y_{1:n}}} \Pr[E_2(y'_{1:n}) = E_2(y_{1:n})] + \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} \neq x_{1:n} \wedge y'_{1:n} = y_{1:n}}} \Pr[E_1(x'_{1:n}) = E_1(x_{1:n})] \\
&= \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} \neq x_{1:n} \wedge y'_{1:n} \neq y_{1:n}}} 2^{-(\lambda_1 + \lambda_2)n} + \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} = x_{1:n} \wedge y'_{1:n} \neq y_{1:n}}} 2^{-\lambda_2 n} + \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} \neq x_{1:n} \wedge y'_{1:n} = y_{1:n}}} 2^{-\lambda_1 n}.
\end{aligned}$$

注意到，$\mathcal{D}_n$ 的大小不是很大. 每一个 $(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n$ 都满足

$$P^n_{XY}(x'_{1:n}, y'_{1:n}) \geq 2^{-n(H(X,Y)+\delta)}.$$

它们的概率之和小于等于 1，因此 $|\mathcal{D}_n| \leq 2^{n(H(X,Y)+\delta)}$. 类似地，集合

$$\left\{ y'_{1:n} \;\middle|\; (x_{1:n}, y'_{1:n}) \in \mathcal{D}_n \right\}$$

的大小也不是很大. 其中的每一个 $y'_{1:n}$ 都满足

$$P^n_{Y|X}(y'_{1:n} | x_{1:n}) \geq 2^{-n(H(Y|X)+\delta)}.$$

因为概率只和不超过 1，所以满足条件的 $y'_{1:n}$ 不超过 $2^{n(H(Y|X)+\delta)}$ 个. 对称地，集合

$$\left\{ x'_{1:n} \;\middle|\; (x'_{1:n}, y_{1:n}) \in \mathcal{D}_n \right\}$$

的大小不超过 $2^{n(H(X|Y)+\delta)}$. 回到之前 union bound 的估计，

$$\begin{aligned}
&\Pr[D(E_1((x_{1:n}), E_2(y_{1:n}))) \neq (x_{1:n}, y_{1:n})] \\
&= \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} \neq x_{1:n} \wedge y'_{1:n} \neq y_{1:n}}} 2^{-(\lambda_1 + \lambda_2)n} + \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} = x_{1:n} \wedge y'_{1:n} \neq y_{1:n}}} 2^{-\lambda_2 n} + \sum_{\substack{(x'_{1:n}, y'_{1:n}) \in \mathcal{D}_n \\ x'_{1:n} \neq x_{1:n} \wedge y'_{1:n} = y_{1:n}}} 2^{-\lambda_1 n} \\
&\leq 2^{n(H(X,Y)+\delta) - (\lambda_1 + \lambda_2)n} + 2^{-n(H(Y|X)+\delta) - \lambda_2 n} + 2^{n(H(X|Y)+\delta) - \lambda_1 n} \\
&\leq 3 \cdot 2^{-\delta n}.
\end{aligned}$$

综合上面几条，正确解压缩的概率很高

$$\begin{aligned}&\Pr_{(X_{1:n},Y_{1:n})\sim(P_{XY})^n}\Big[D(E_1(X_{1:n}),E_2(Y_{1:n}))=(X_{1:n},Y_{1:n})\Big]\\=&\Pr_{(X_{1:n},Y_{1:n})\sim(P_{XY})^n}\Big[D(E_1(X_{1:n}),E_2(Y_{1:n}))=(X_{1:n},Y_{1:n})\;\Big|\;(X_{1:n},Y_{1:n})\in\mathcal{D}_n\Big]\Pr[(X_{1:n},Y_{1:n})\in\mathcal{D}_n]\\&\geq(1-2^{-\Theta(n)})(1-2^{-\Theta(n)})=1-2^{-\Theta(n)}.\end{aligned}$$

(2) 如果 $\lambda_1+\lambda_2<H(X,Y)$，一定不存在压缩函数 E_1,E_2 和解压缩函数 D 能大概率正确解压缩. 因为否则 E_1,E_2 可以合并成一个 P_{XY}^n 的压缩函数 $E:(\mathcal{X}\times\mathcal{Y})^n\to\{0,1\}^{(\lambda_1+\lambda_2)n}$，可以被 D 大概率正确解压. 而我们知道这样的压缩函数是不存在的.

如果 $\lambda_1<H(X|Y)$，一定不存在压缩函数 E_1,E_2 和解压缩函数 D 能大概率正确解压缩. 因为否则 E_1 就是一个压缩函数，可以被 E_2,D 合并成的一个利用旁信息的解压缩函数 $D':\mathcal{Y}^n\times\{0,1\}^{\lambda_1 n}\to\mathcal{X}^n$ 大概率正确解压缩. 而我们知道，即使利用旁信息，也不可能把 P_X^n 压缩到 $n(H(X|Y)-\delta)$ 比特以内.

2. (8 分) (n,m,k)-纠错码是一对映射 $E:\{0,1\}^n\to\{0,1\}^m$，$D:\{0,1\}^m\to\{0,1\}^n$ 满足

$$\forall x\in\{0,1\}^n,\forall c\in\{0,1\}^m,\Delta(c,E(x))\leq k\implies D(c)=x.$$

这里 Δ 表示汉明距离（Hamming distance）.

(1) 证明存在常数 α. 当 $m>4k$ 且 $m\geq 2n+\alpha k\ln(m/2k)$ 时，存在 (n,m,k)-纠错码.

提示：可以使用不等式 $\frac{\log p\cdot\log(1-p)}{\log e}\leq h(p)\leq\frac{\log p\cdot\log(1-p)}{\log 2}$. 其中 $h(p)$ 表示 $\mathrm{Bern}(p)$ 的熵.

(2) 证明存在常数 α. 当 $m>4k$ 且 $m\geq n+\alpha k\ln(m/2k)$ 时，存在 (n,m,k)-纠错码.

解 如果编码 $E:\{0,1\}^n\to\{0,1\}^m$ 满足

$$\forall\text{distinct } x,y\in\{0,1\}^n,\Delta(E(x),E(y))>2k$$

那么对任意 $c\in\{0,1\}^m$，存在至多一个 $x\in\{0,1\}^n$ 满足 $\Delta(c-E(x))\leq k$. 这时可以定义解码 D 使得 D,E 构成一个 (n,m,k)-纠错码.

(1) 随机选取一个映射 E. 定义 $C_x=E(x)$，这样 C_x（$x\in\{0,1\}^n$）是相互独立的随机变量. 对任意不同的 $x,y\in\{0,1\}^n$

$$\Pr\Big[\Delta(C_x-C_y)\leq 2k\Big]=\Pr_{Z\sim\mathrm{Binom}(m,\frac{1}{2})}\Big[Z\leq 2k\Big]\leq\exp\Big(-m\cdot D\Big(\frac{2k}{m}\;\Big\|\;\frac{1}{2}\Big)\Big).$$

根据提示中的不等式，

$$\begin{aligned}D\Big(\frac{2k}{m}\;\Big\|\;\frac{1}{2}\Big)&=\log 2-h\Big(\frac{2k}{m}\Big)\geq\log 2-\frac{\log(\frac{2k}{m})\cdot\log(1-\frac{2k}{m})}{\log 2}\\&=\log 2+\log_2(\tfrac{m}{2k})\cdot\log(1-\tfrac{2k}{m})\geq\log 2-2\log 2\cdot\frac{2k}{m}\log_2(\tfrac{m}{2k}).\end{aligned}$$

其中还利用了 $\log(1-p) \geq -2\log 2 \cdot p$ 对任何 $p \in [0, \frac{1}{2}]$. 因此

$$\Pr\Big[\Delta(C_x, C_y) \leq 2k\Big] \leq 2^{-m+4k\log_2(\frac{m}{2k})}.$$

根据 union bound,

$$\Pr\Big[\exists \text{distinct } x, y, \Delta(C_x, C_y) \leq 2k\Big] < 2^{2n} \cdot 2^{-m+4k\log_2(\frac{m}{2k})}.$$

只要 $m \geq 2n + 4k\log_2(\frac{m}{2k})$，这个概率便严格小于 1. 说明存在 (n, m, k)-纠错码.

(2) 随机选取一个线性映射 $E : \{0,1\}^n \to \{0,1\}^n$. 仍然定义 $C_x = E(x)$，这样 C_x （$x \in \{0,1\}^n$）是两两独立的随机变量. 仍有对任意不同的 $x, y \in \{0,1\}^n$

$$\Pr\Big[\Delta(C_x, C_y) \leq 2k\Big] \leq 2^{-m+4k\log_2(\frac{m}{2k})}.$$

对于任意 x, y，考虑 $z = x \oplus y$，因为 E 是线性映射

$$\Delta(C_x, C_y) = \|E(x) \oplus E(y)\|_1 = \|E(z)\|_1 = \|E(z) \oplus E(0)\|_1 = \Delta(C_z, C_0).$$

这说明, 任意两个编码见的汉明距离大, 等价于零的编码与任何非零的编码之间的汉明距离大.

$$\Pr\Big[\exists \text{distinct } x, y, \Delta(C_x, C_y) \leq 2k\Big] = \Pr\Big[\exists z \neq 0, \Delta(C_z, C_0) \leq 2k\Big] < 2^n \cdot 2^{-m+4k\log_2(\frac{m}{2k})}.$$

只要 $m \geq n + 4k\log_2(\frac{m}{2k})$，这个概率便严格小于 1. 说明存在 (n, m, k)-纠错码.

3. (10 分) 在有限空间 Ω 上有两个分布 P, Q. 区分器 $\mathcal{D}$ 是一个输入域为 Ω, 输出域为 $\{0,1\}$ 的算法 (更准确地说, 是 kernel). 我们希望让伪阳性概率 ε_{FP} 和伪阴性概率 ε_{FN} 尽量小.

$$\varepsilon_{\text{FP}} = \Pr_{X\sim P}[\mathcal{D}(X) \to 1], \qquad \varepsilon_{\text{FN}} = \Pr_{X\sim Q}[\mathcal{D}(X) \to 0].$$

(1) 定义 likelihood ratio 为 $L : \Omega \to [-\infty, +\infty]$, $L(x) = \log(\frac{Q(x)}{P(x)})$.
证明: 对任何区分器 $\mathcal{D}$, 存在算法 $\mathcal{D}' : [-\infty, +\infty] \to \{0,1\}$, 使得

$$\Pr_{X\sim P}[\mathcal{D}(X) \to 1] = \Pr_{X\sim P}[\mathcal{D}'(L(X)) \to 1], \qquad \Pr_{X\sim Q}[\mathcal{D}(X) \to 0] = \Pr_{X\sim Q}[\mathcal{D}'(L(X)) \to 0].$$

(2) 证明: 为了最小化 $\varepsilon_{\text{FP}}, \varepsilon_{\text{FN}}$, 只须考虑如下的 likelihood ratio test 区分器 $\mathcal{D}_{\tau,\theta}$

$$\mathcal{D}_{\tau,\theta}(x) = \begin{cases} 1, & \text{if } L(x) > \tau \\ \text{Bern}(\theta), & \text{if } L(x) = \tau \\ 0, & \text{if } L(x) < \tau \end{cases}$$

(3) 改为区分 P^n 和 Q^n. 这时区分器是输入域为 Ω^n, 输出域为 $\{0,1\}$ 的算法. 随着 n 的增长, 是否可以让 $\varepsilon_{\text{FP}}, \varepsilon_{\text{FN}}$ 分别以 $\exp(-n\alpha), \exp(-n\beta)$ 的速度趋近于 0?

具体来说, 请确定以下区域的边界

$$\left\{(\alpha,\beta)\in\mathbb{R}_+^2 \,\middle|\, \text{对任意充分大的 } n\text{, 存在区分器 } \mathcal{D}\text{, 同时满足} \begin{matrix}\Pr_{X\sim P}[\mathcal{D}(X)\to 1]\le\exp(-n\alpha)\\ \Pr_{X\sim Q}[\mathcal{D}(X)\to 0]\le\exp(-n\beta)\end{matrix}\right\}$$

为了统一记号, 对任意 $\lambda\in[0,1]$, 定义分布 P_λ 为 $P_\lambda(x)\propto(P(x))^{1-\lambda}(Q(x))^\lambda$.

提示: 上次作业第 3 题.

解

(1) 用随机变量 X 表示从 P 或 Q 的采样. 定义随机变量 $Y=L(X)$. 这样就得到了两个联合分布 P_{XY},Q_{XY}. 根据 Y 的定义, 有一个 (退化的) kernel $P_{Y|X}$ 满足 $P_{XY}=P_XP_{Y|X}$, $Q_{XY}=Q_XP_{Y|X}$. 同样根据 Y 的定义, 对任意 x s.t. $y=L(x)\in(-\infty,+\infty)$,

$$P_{x|y}(x|y)=\frac{P(x)}{\sum\limits_{x'\text{ s.t. }L(x')=y}P(x')}=\frac{e^yP(x)}{\sum\limits_{x'\text{ s.t. }L(x')=y}e^yP(x')}=\frac{Q(x)}{\sum\limits_{x'\text{ s.t. }L(x')=y}Q(x')}=Q_{x|y}(x|y)$$

因此可以定义 kernel $P_{X|Y}$ 使得 $P_{XY}=P_YP_{X|Y}$, $Q_{XY}=Q_YP_{X|Y}$.

令 $\mathcal{D}'(y)$ 首先从条件分布 $P_{X|Y=y}$ 采样 x, 再输出 $\mathcal{D}(x)$ 的结果. 这样当 y 采样自 P_Y (resp. Q_Y) 时, x 的分布服从 P_X (resp. Q_X). 便证明了题目.

区分从 P 或 Q 采样的随机变量 X 时, 能从 X 中计算出的量都被称为是统计量. 例如这里定义的 Y 就是一个统计量. 而额外满足 $P_{XY}=P_YP_{X|Y}$, $Q_{XY}=Q_YP_{X|Y}$ 的统计量被称为充分统计量 (sufficient statistic), 因为它已经包括了区分 P,Q 的所有有用信息.

因此, 我们不妨改为区分 P_Y,Q_Y. 等价地, 不妨假设 $x=L(x)=\log\frac{Q(x)}{P(x)}$.

(2) 假设 $\mathcal{D}$ 不是 LRT 区分器, 那么存在 $\alpha>\beta$ 使得

$$\begin{aligned}Q(\alpha)&>0, & P(\beta)&>0,\\ \Pr[\mathcal{D}(\alpha)\to 1]&<1, & \Pr[\mathcal{D}(\beta)\to 1]&>0.\end{aligned}$$

我们构造一个严格优于 $\mathcal{D}$ 的区分器 $\mathcal{D}'$. 选择一个充分小的 $\varepsilon>0$, 定义

$$\Pr[\mathcal{D}'(x)\to 1]=\begin{cases}\Pr[\mathcal{D}(\alpha)\to 1]+\varepsilon P(\beta), & \text{if } x=\alpha\\ \Pr[\mathcal{D}(\beta)\to 1]-\varepsilon P(\alpha), & \text{if } x=\beta\\ \Pr[\mathcal{D}(x)\to 1], & \text{otherwise}\end{cases}$$

于是

$$\begin{aligned}\Pr_{X\sim P}[\mathcal{D}'(x)\to 1]&=\Pr_{X\sim P}[\mathcal{D}(x)\to 1]+P(\alpha)\varepsilon P(\beta)-P(\beta)\varepsilon P(\alpha)=\Pr_{X\sim P}[\mathcal{D}(x)\to 1].\\ \Pr_{X\sim Q}[\mathcal{D}'(x)\to 1]&=\Pr_{X\sim Q}[\mathcal{D}(x)\to 1]+Q(\alpha)\varepsilon P(\beta)-Q(\beta)\varepsilon P(\alpha)\\ &=\Pr_{X\sim Q}[\mathcal{D}(x)\to 1]+Q(\alpha)\varepsilon P(\beta)-e^\beta P(\beta)\varepsilon e^{-\alpha}Q(\alpha)>\Pr_{X\sim Q}[\mathcal{D}(x)\to 1].\end{aligned}$$

也就是说, $\mathcal{D}'$ 在不改变 $\varepsilon_{\mathrm{FP}}$ 的同时改善了 $\varepsilon_{\mathrm{FN}}$.

(3) 因为已经证明 LRT 是最好的区分器, 只需考虑以下区分器的错误概率: $\mathcal{D}(x_1,\ldots,x_n)=1$ iff $\sum_i x_i > n\tau$, 其中 τ 是任意阈值. 这时

$$\varepsilon_{\mathrm{FP}} = \Pr_{X^n\sim P^n}\Big[\sum_i X_i > n\tau\Big], \qquad \varepsilon_{\mathrm{FN}} = \Pr_{X^n\sim Q^n}\Big[\sum_i X_i \le n\tau\Big].$$

为了让错误概率不逼近 1, 阈值 τ 必须满足

$$\mathop{\mathbb{E}}_{X\sim P}[X] \le \tau \le \mathop{\mathbb{E}}_{X\sim Q}[X]. \tag{*}$$

根据上次作业, 我们可以用 Chernoff bound 证明

$$\varepsilon_{\mathrm{FP}} = \Pr_{X^n\sim P^n}\Big[\sum_i X_i > n\tau\Big] \le \exp(-nD(P^*\|P)).$$

这里 $P^*(x)\propto \exp(\lambda x)P(x)$, 其中的参数 λ 由 $\mathbb{E}_{X\sim P^*}[X]=\tau$ 唯一确定. 注意到

$$P^*(x)\propto \exp(\lambda x)P(x) = \Big(\frac{Q(x)}{P(x)}\Big)^{\lambda} P(x) = Q(x)^{\lambda}P(x)^{1-\lambda}.$$

所以 $P^*=P_\lambda$, 其中 λ 由 $\mathbb{E}_{X\sim P_\lambda}[X]=\tau$ 确定. 因为 $\mathbb{E}_{X\sim P_\lambda}[X]$ 随 τ 单调增长, 且 τ 的取值在 (*) 中, 所以 $\lambda\in[0,1]$.

对称地, $\varepsilon_{\mathrm{FN}} \le \exp(-nD(P_\lambda\|Q))$, 其中 λ 由 $\mathbb{E}_{X\sim P_\lambda}[X]=\tau$ 确定.

对于任意 $\lambda\in[0,1]$, 区分器 $\mathcal{D}(x_1,\ldots,x_n)=1$ iff $\sum_i x_i > n\,\mathbb{E}_{X\sim P_\lambda}[X]$ 可以实现

$$\varepsilon_{\mathrm{FP}} = \exp(-nD(P_\lambda\|P)), \qquad \varepsilon_{\mathrm{FN}} = \exp(-nD(P_\lambda\|Q)).$$

因此题目所求区域一定包含了曲线 $(D(P_\lambda\|P), D(P_\lambda\|Q))_{\lambda\in(0,1)}$ 之下的部分.

同时, 根据前几问, 我们知道 LRT 是最优的. 根据上次作业, 我们知道 Chernoff bound 是足够紧的. 因此曲线 $(D(P_\lambda\|P), D(P_\lambda\|Q))_{\lambda\in(0,1)}$ 是题目所求区域的边界.

4. (8 分) 有两个相互独立的秘密，分别用随机变量 C_0, C_1 表示，满足 $H[C_0]=H[C_1]=n>0$. 根据 C_0, C_1，用一个随机算法生成 A_0, A_1, B_0, B_1. Alice 选择 $\alpha\in\{0,1\}$，并获得 A_α. Bob 选择 $\beta\in\{0,1\}$，并获得 B_β. 我们要求，无论 (α,β) 是多少:

- Alice 和 Bob 各自都没有得到 (C_0, C_1) 的任何信息;
- Alice 和 Bob 联合起来可以知道 $C_{\alpha\beta}$，但没有得到 $C_{1-\alpha\beta}$ 的任何信息.

请问 A_0, A_1, B_0, B_1 可以有多短.

(1) 将两条要求用信息量（熵、条件熵、互信息等）表示.

(2) 证明 $\max(H[A_0], H[A_1], H[B_0], H[B_1]) > n$.

(3) 证明 $\max(H[A_0], H[A_1], H[B_0], H[B_1]) \ge 1.5n$.

(4) (0 分) 证明上一问的界是紧的. 构造 C_0, C_1 的分布，以及生成 A_0, A_1, B_0, B_1 的随机算法.

解

(1) 条件一：对任意 $\alpha,\beta \in \{0,1\}$

$$I(C_0,C_1;A_\alpha)=0,\qquad I(C_0,C_1;B_\beta)=0.$$

条件二：对任意 $\alpha,\beta \in \{0,1\}$

$$H(C_{\alpha\beta}|A_\alpha,B_\beta)=0,\qquad I(C_{1-\alpha\beta};A_\alpha,B_\beta)=0.$$

(2) 因为 A_0 不包含 C_0 的信息 $H[C_0|A_0]=n$，而 (A_0,B_1) 包含 C_0 的所有信息 $H[C_0|A_0,B_1]=0$；所以 $I[C_0;B_1|A_0]=n$. 这说明 B_1 熵至少是 n，

$$H[B_1]\ge H[B_1|A_0]\ge I[C_0;B_1|A_0]=n.$$

同理，$H[A_0],H[A_1],H[B_0]\ge n$.

用反证法，假设 $H[A_0]=H[A_1]=H[B_0]=H[B_1]=n$，我们试图推出矛盾.

注意到，此时 $I[C_0;B_1|A_0]=H[C_0|A_0]=H[B_1|A_0]=n$. 这不仅说明可以由 A_0,B_1 确定 C_0，也对称地说明可以由 A_0,C_0 确定 B_1：

$$H[B_1|A_0,C_0]=H[B_0|A_0]-I[C_0;B_1|A_0]\le H[B_0]-I[C_0;B_1|A_0]=0.$$

同理，$H[A_1|B_0,C_0]=0$，可以由 B_0,C_0 确定 A_1. 整理一下，如果已知 A_0,B_0，可以确定 C_0，进而可以确定 A_1,B_1，再进一步可以确定 C_1. 这与 $I(C_1;A_0,B_0)=0$ 矛盾.

(3) 只需定量化上一问的论证. 令 $kn=\max(H[A_0],H[A_1],H[B_0],H[B_1])$.

$$H[B_1|A_0,C_0]=H[B_0|A_0]-I[C_0;B_1|A_0]\le H[B_0]-I[C_0;B_1|A_0]=(k-1)n.$$

对称地，$H[B_1|A_0,C_0]\le (k-1)n$.

$$\begin{aligned} n=H[C_1|A_0,B_0]&\le H[C_1,A_1,B_1|A_0,B_0]=H[A_1,B_1|A_0,B_0,C_0]\\ &\le H[A_1|A_0,B_0,C_0]+H[B_1|A_0,B_0,C_0]\le H[A_1|B_0,C_0]+H[B_1|A_0,C_0]\le(2k-2)n. \end{aligned}$$

得到 $k\ge 1.5$.

(4) 令 C_0,C_1 都服从 $\{0,1\}^2$ 上的均匀分布. 对于 $\{0,1\}^2$ 中的序对，我们用下标 L,R 表示其左分量和右分量.

使用如下算法生成 $A_0,A_1,B_0,B_1\in\{0,1\}^3$：采样随机的 C_0',C_0'' 满足 $C_0'\oplus C_0''=C_0$. 采样随机的 C_1',C_1'' 满足 $C_1'\oplus C_1''=C_1$. 输出 A_0,A_1,B_0,B_1

$$\begin{aligned} &A_0=(C_0',C_{0,R}''\oplus C_{1,R}''), && B_0=(C_0'',C_{0,L}'\oplus C_{1,L}''),\\ &A_1=(C_1',C_{0,R}'), && B_1=(C_1'',C_{0,L}''). \end{aligned}$$