信息论第六讲作业解答

中国科学技术大学《信息论 A》006125.01 班助教组

2024 年 5 月 27 日

第 1 题

Prove Corollary 1.

$$C(\Gamma) = \max_{P_X} I(X;Y), \tag{1}$$

$$s.t.\ \mathbf{E}[c(X)] \leq \Gamma. \tag{2}$$

a):For capacity-cost function $C(\Gamma)$, the optimization problem (1) is infeasible for any $\Gamma < \Gamma_{\min} := \min_{x\in\mathcal{X}} c(x)$ and we simply set $C(\Gamma) = 0$ therein.

b):There exists a threshold $\Gamma_{\max}$ such that for any $\Gamma \geq \Gamma_{\max}, C(\Gamma) = C(\infty)$ which is the capacity without cost constraint.

c):The capacity-cost function $C(\Gamma)$ is a non-decreasing concave function in Γ.

d):If $C(\infty) > 0$, then $C(\Gamma)$ is strictly increasing for $\Gamma \in [\Gamma_{\min}, \Gamma_{\max}]$, and in the optimization problem (1) the inequality constraint can be replaced by an equality constraint.

证明: a):对于任意的P_X, 我们均有：

$$\mathbf{E}[c(X)] = \sum_{x\in\mathcal{X}} p(x)c(x) \geq \Gamma_{\min} := \min_{x\in\mathcal{X}} c(x),$$

从而对于所有的$\Gamma < \Gamma_{\min}$, 这样的信源分布是不可能存在的, 认为此时无法按要求传递信息, 视$C(\Gamma) = 0$.

b):当我们在无约束下求解$\max_{P_X} I(X;Y)$得到一组$\mathcal{P}^*$, 从中取得$p^*(X)$使得$\mathbf{E}[c(X)]$最小, 此时便有：

$$\Gamma = \sum p^*(x)c(x) := \Gamma_{\max} \leq \max_{x\in\mathcal{X}} c(x),$$

那么对于任意$\Gamma \geq \Gamma_{\max}$, 我们仍然能在约束$\Gamma_{\max}$下取到$C(\Gamma)$的极值, 这也便意味着$C(\Gamma) = C(\infty)$, 等价于无约束.

c):对于任意的$\Gamma_{\min} \le \Gamma_1 < \Gamma_2$, 记满足(2)的分布集合为$\mathcal{P}_\Gamma$, 那么满足$\Gamma_1$约束时一定满足$\Gamma_2$约束, 即$\mathcal{P}_{\Gamma_1} \subseteq \mathcal{P}_{\Gamma_2}$, 那么$\mathcal{P}_{\Gamma_1}$下的最优解一定也属于$\mathcal{P}_{\Gamma_2}$, 因为我们求的是极大值, 所以容量代价函数是非减的.

我们希望进一步证明代价容量函数是凹的, 即对于$\Gamma_{\min} \le \Gamma_1 < \Gamma_2$以及$0 \le \lambda \le 1$有：

$$C((1-\lambda)\Gamma_1 + \lambda\Gamma_2) \ge (1-\lambda)C(\Gamma_1) + \lambda C(\Gamma_2),$$

为此, 我们取在Γ_1和Γ_2约束下达到极值的概率密度函数$P_{1,X}, P_{2,X}$, 定义$P_{\lambda,X} = (1-\lambda)P_{1,X} + \lambda P_{2,X}$, 我们有：

$$\begin{aligned}
\mathbf{E}_{X\sim P_{\lambda,X}}[c(X)] &= \sum_{x\in\mathcal{X}} p_\lambda(x)c(x) \\
&= (1-\lambda)\sum_{x\in\mathcal{X}} p_1(x)c(x) + \lambda \sum_{x\in\mathcal{X}} p_2(x)c(x) \\
&\le (1-\lambda)\Gamma_1 + \lambda\Gamma_2,
\end{aligned}$$

从而在约束$(1-\lambda)\Gamma_1 + \lambda\Gamma_2$下我们至少拥有了一个分布$P_{\lambda,X} = (1-\lambda)P_{1,X} + \lambda P_{2,X}$, 又因为我们求的是极大值：

$$C((1-\lambda)\Gamma_1 + \lambda\Gamma_2) \ge I(X;Y)|_{P_{\lambda,X}}. \tag{3}$$

进一步, 藉由互信息关于P_X的凹性(concave), 我们又有：

$$I(X;Y)|_{P_{\lambda,X}} \ge (1-\lambda)I(X;Y)|_{P_{1,X}} + \lambda I(X;Y)|_{P_{2,X}} = (1-\lambda)C(\Gamma_1) + \lambda C(\Gamma_2). \tag{4}$$

由(3)(4)：

$$C((1-\lambda)\Gamma_1 + \lambda\Gamma_2) \ge (1-\lambda)C(\Gamma_1) + \lambda C(\Gamma_2).$$

d):若不是严格增长的, 我们一定可以找到$\Gamma_{\min} \le \Gamma_1 < \Gamma_2 \le \Gamma_{\max}$, 使得$C(\Gamma_1) = C(\Gamma_2) = C(\Gamma)$, 其中$\Gamma_1 \le \Gamma \le \Gamma_2$, 由我们上面证明的凹性：

$$\Gamma_2 = \frac{\Gamma_{max} - \Gamma_2}{\Gamma_{max} - \Gamma_1}\Gamma_1 + \frac{\Gamma_2 - \Gamma_1}{\Gamma_{max} - \Gamma_1}\Gamma_{\max}$$
$$C(\Gamma_2) \ge \frac{\Gamma_{max} - \Gamma_2}{\Gamma_{max} - \Gamma_1}C(\Gamma_1) + \frac{\Gamma_2 - \Gamma_1}{\Gamma_{max} - \Gamma_1}C(\Gamma_{\max})$$

但是我们又假定$C(\Gamma_1) = C(\Gamma_2)$, 那么只有$C(\Gamma_1) = C(\Gamma_2) = C(\Gamma_{\max})$, 但是这又和我们对于$\Gamma_{\max}$的定义矛盾, 所以容量代价函数是严格递增的.

进一步, 若存在$\Gamma_{\min} \le \Gamma_1 \le \Gamma_{\max}$使得在(2)约束下(1)取到极值时, 有$\mathbf{E}[c(X)] = \Gamma' < \Gamma_1$, 那么我们可以断言$C(\Gamma') \ge C(\Gamma_1)$, 但是这又和严格递增矛盾, 命题得证. □

第 2 题

Find as many counterpart relationships between source representation and channel transmission as you can.

解: 如下表所示： □

表 1: Counterpart Relationships

	信源重建	信道传输
物理过程	$S \to W \to \hat{S}$	$W \to X \to Y \to \hat{W}$
实际意义	失真约束下最小化码率	代价约束下最大化传递信息量
操作性定义	$f_n^{(s)} : \mathcal{S}^n \to \{1, 2, ...M_n\}$ $g_n^{(s)} : \{1, 2, ...M_n\} \to \hat{\mathcal{S}}^n$ $R = \frac{\lceil \log_2 M_n \rceil}{n}$bits/source symbol $\lim_{n\to\infty} Ed(S, \hat{S}) \leq D$	$f_n^{(c)} : \{1, 2, ...M_n\} \to \mathcal{X}^n$ $g_n^{(c)} : \mathcal{Y}^n \to \{1, 2, ...M_n\}$ $R = \frac{\lceil \log_2 M_n \rceil}{n}$bits/channel use $\lim_{n\to\infty} P_{e,max} = 0$
数学表达	$\min_{p(\hat{S}\|S):Ed(S,\hat{S})\leq D} I(S; \hat{S})$	$\max_{p(X):Ec(X)\leq\Gamma} I(X; \hat{X})$
固定条件	$p(s)$:源分布 $d(s, \hat{s})$:失真形式	$p(\hat{x}\|x)$:信道分布（噪声） $c(x)$:代价形式
求解问题	最优重建表示的转移$p(\hat{s}\|s)$	最优可对抗噪声的分布$p(x)$
函数结构	递减凸函数	递增凹函数

第 3 题

Calculate the capacity for the following DMC

$$P_{Y|X} = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 \end{bmatrix},$$

where the i-th row j-th column element corresponds to $P_{Y|X}(j|i)$. (Hint: Symmetry is helpful, but you need to explain why your simplification using symmetry does not lose optimality.)

可以验证对所有 $a \in (0, 1) \cup (1, \infty)$ 和 $p \in (0, 1)$ 有

$$\frac{d}{dp}(-p \log_a(p) - (1 - p) \log_a(1 - p)) = \log_a \left(\frac{1 - p}{p}\right).$$

在求信道容量和率失真函数时我们可以直接用这个结论.

方法一: 设 $Y|X \sim P_{Y|X}$. 定义随机变量 Y_2 在 $Y=2$ 时取 1, 否则取 0.

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) + H(Y_2|Y) - H(Y|X) \\
&= H(Y_2) + H(Y|Y_2) - \log_2(3)P_X(2) \\
&= H(Y_2) + P_{Y_2}(0)H(Y|Y_2=0) - \log_2(3)P_X(2) \\
&\le H(Y_2) + P_{Y_2}(0) - \log_2(3)P_X(2) \\
&= h_2\left(\frac{P_X(2)}{3}\right) + 1 - \frac{P_X(2)}{3} - \log_2(3)P_X(2),
\end{aligned}$$

其中第二行的等号是因为 $H(Y_2|Y)=0$, 第四行的等号是因为如果 $P_{Y_2}(1)>0$ 则 $H(Y|Y_2=1)=0$, 最后一行的等号是因为 Y_2 以概率 $P_X(2)/3$ 取 1, 以概率 $1-P_X(2)/3$ 取 0. 用 $g(t)=h_2(t/3)+1-t/3-\log_2(3)t$ 定义 $g:[0,1]\to\mathbb{R}$. 对每个 $t\in(0,1)$ 有

$$g'(t) = \frac{1}{3}\log_2\left(\frac{3-t}{54t}\right),$$

所以 g 在 $[0,3/55]$ 上递增, 在 $[3/55,1]$ 上递减. 因此

$$I(X;Y) \le g\left(\frac{3}{55}\right) = \log_2\left(\frac{55}{27}\right).$$

如果 X 分别以概率 26/55, 3/55 和 26/55 取 1, 2 和 3, 则 $H(Y|Y_2=0)=H(Y|Y\neq 2)=1$,

$$h_2\left(\frac{P_X(2)}{3}\right) + 1 - \frac{P_X(2)}{3} - \log_2(3)P_X(2) = g\left(\frac{3}{55}\right),$$

所以 $I(X;Y)=\log_2(55/27)$. 作为 $Y|X\sim P_{Y|X}$ 的条件下 $I(X;Y)$ 的最大值, 信道容量等于 $\log_2(55/27)$ 比特. □

方法一不涉及 "simplification using symmetry". 方法二说明了 "simplification using symmetry does not lose optimality" 的意思.

方法二: 设 $Y|X\sim P_{Y|X}$. 定义随机变量 X_1 和 Y_1 满足 $P_{X_1}(1)=P_X(3)$, $P_{X_1}(2)=P_X(2)$, $P_{X_1}(3)=P_X(1)$ 和 $Y_1|X_1\sim P_{Y|X}$. 再定义随机变量 X_2 和 Y_2 满足 $P_{X_2}=(P_X+P_{X_1})/2$ 和 $Y_2|X_2\sim P_{Y|X}$. 可以看出 $I(X;Y)=I(X_1;Y_1)$. 根据互信息的凹凸性 (第 2 讲定理 7),

$$\frac{1}{2}I(X;Y) + \frac{1}{2}I(X_1;Y_1) \le I(X_2;Y_2).$$

所以 $I(X;Y)\le I(X_2;Y_2)$. 可以算出

$$P_{X_2}(1) = P_{X_2}(3) = \frac{1-P_X(2)}{2}, P_{X_2}(2) = P_X(2),$$

$$I(X_2;Y_2) = h_2\left(\frac{P_X(2)}{3}\right) + 1 - \frac{P_X(2)}{3} - \log_2(3)P_X(2) = g(P_X(2)).$$

这里 g 的定义和方法一中一样. 所以我们还是只需要求 g 在 $[0,1]$ 上的最大值. □

第 4 题

Consider a DMC with $P_{Y|X} = [p_{i,j}]_{i=1,...,|x|;j=1,...,|y|}$.

a) *Add an erasure output* e *as follows: for any input* x, *output* e *with probability* α, *otherwise output* $j \in \{1,\dots,|y|\}$ *according to* $P_{Y|X}$. *Prove that the capacity of such a channel is* $(1-\alpha)C$, *where* C *is the capacity of the original DMC* $P_{Y|X}$.

b) *Instead, if add an erasure output* e *as follows: first pass an input* x *through* $P_{Y|X}$, *and then change its output to* e *with probability* α. *What is the capacity of such a channel?*

证明: a):我们记原先的转移概率矩阵为$P_{Y|X}$, 之后的为$\tilde{P}_{Y|X}$.

$$\begin{aligned}
I(X;\tilde{Y}) &= H(X) - H(X|\tilde{Y}) \\
&= H(X) - P(\tilde{Y}=e)H(X|\tilde{Y}=e) - P(\tilde{Y}\neq e)H(X|\tilde{Y}\neq e) \\
&= H(X) - \alpha H(X) - (1-\alpha)\sum_{y\neq e} p(\tilde{Y}=y_j|\tilde{Y}\neq e)H(X|\tilde{Y}=y_j) \\
&= (1-\alpha)H(X) - (1-\alpha)H(X|Y) \\
&= (1-\alpha)I(X;Y).
\end{aligned}$$

我们提供使用另一种互信息计算方式的解法:

$$\begin{aligned}
I(X;\tilde{Y}) &= H(\tilde{Y}) - H(\tilde{Y}|X) \\
&= H(\alpha, 1-\alpha) + (1-\alpha)H(\frac{p(\tilde{y_1})}{1-\alpha}, ..., \frac{p(\tilde{y_m})}{1-\alpha}) \\
&\quad - \sum_x x_i[H(\alpha, 1-\alpha) + (1-\alpha)H(\frac{p(\tilde{y_1}|x_i)}{1-\alpha}, ..., \frac{p(\tilde{y_m}|x_i)}{1-\alpha})] \\
&= (1-\alpha)H(Y) - (1-\alpha)H(Y|X) \\
&= (1-\alpha)I(X;Y).
\end{aligned}$$

综上所述, 对于任意一个分布$P_{Y|X}$, 我们均有$I(X;\hat{Y}) = (1-\alpha)I(X;Y)$, 从而:

$$\max I(X;\tilde{Y}) = \max(1-\alpha)I(X;Y) = (1-\alpha)C.$$

b):我们只需说明两个信道的等价性, 或者说, 重点关注X到$\tilde{Y}$的转移概率矩阵是否一致. 此问中设计了两个连续的信道, $X \to Y \to \tilde{Y}$:

$$\begin{aligned}
p(\tilde{y_j}|x_i) &= \sum_{y_k} p(y_k|x_i)p(\tilde{y_j}|y_k) \\
&= \begin{cases} (1-\alpha)p(y_j|x_i), & \tilde{y_j} \neq e \\ \sum_{y_k} p(y_k|x_i)\alpha = \alpha, & \tilde{y_j} = e \end{cases}
\end{aligned}$$

很容易看出这和a)问中描述是一致的. $\square$

第 5 题

Consider a BEC with erasure probability $\alpha > 0$ and with noiseless feedback without delay. Explicitly describe a block coding method (i.e., all channel codewords having a prescribed length n that does not depend upon the feedback, and finally letting n go to infinity) that achieves the channel capacity $(1-\alpha)$ bits/channel use. Note that the described method should be explicit, not the random codebook as in Section IV.

解: 设 E_1, E_2, $\cdots$ 是一列独立同分布的随机变量, E_1 分别以 $1-\alpha$ 和 α 的概率取 0 和 1, $1/2 < \beta < 1$.

用 n 表示一个任意的正整数. 记

$$\nu(n) = \lceil \alpha n + n^{\beta} \rceil, \{0,1\}^{n-\nu(n)} = \{\underline{b}(1), \underline{b}(2), \cdots, \underline{b}(2^{n-\nu(n)})\}.$$

让一个编码器对每个正整数 $w \leq 2^{n-\nu(n)}$ 用 $\underline{b}(w)$ 传输消息 w. 如果 $\underline{b}(w)$ 的某个符号被信道擦除, 编码器应该重传这个符号, 直到符号被正确传输或已经使用了 n 次信道. 如果 $\underline{b}(w)$ 的最后一个符号被正确传输时编码器使用信道的次数小于 n, 编码器接下来可以传输 0 直到信道已被使用 n 次. 再设计以下译码器. 如果 $\underline{y} \in \{0,1,e\}^n$ 中 e 的个数小于等于 $\nu(n)$ 则存在唯一的正整数 $w \leq 2^{n-\nu(n)}$ 使 $\underline{b}(w)$ 是 $\underline{y}$ 中不等于 e 的部分的前缀, 译码器收到 $\underline{y}$ 后应该输出这个 w. 如果 $\underline{y} \in \{0,1,e\}^n$ 中 e 的个数大于 $\nu(n)$, 译码器收到 $\underline{y}$ 后应该输出 1. 这个码的码率为

$$\frac{\log_2(2^{n-\nu(n)})}{n} = \frac{n-\nu(n)}{n} > \frac{n-\alpha n - n^{\beta} - 1}{n} = 1 - \alpha - n^{\beta-1} - \frac{1}{n}.$$

对每个正整数 $w \leq 2^{n-\nu(n)}$, 可以看出消息 w 被传错的概率 $P_{e,w}^{(n)}$ 小于等于被擦除的符号多于 $\nu(n)$ 的概率, 所以

$$P_{e,w}^{(n)} \leq P\left[\sum_{i=1}^{n} E_i > \nu(n)\right] \leq P\left[\sum_{i=1}^{n} E_i > \alpha n + n^{\beta}\right].$$

可以看出 $\sum_{i=1}^{n} E_i$ 的均值为 αn, 方差为 $n\alpha(1-\alpha)$. 这样最大错误概率

$$P_{e,\max}^{(n)} \leq P\left[\sum_{i=1}^{n} E_i > \alpha n + n^{\beta}\right] \leq \frac{n\alpha(1-\alpha)}{(n^{\beta})^2} = \alpha(1-\alpha)n^{1-2\beta},$$

其中第二个大于等于号用了 Chebyshev 不等式.

因为 $n \to \infty$ 时码率的下界 $1-\alpha-n^{\beta-1}-1/n$ 趋于 $1-\alpha$, $P_{e,\max}^{(n)}$ 趋于 0, 所以上述编译码方法达到 $1-\alpha$ 的码率. □

第 6 题

Instead of the asymptotically vanishing error probability criterion, consider strictly zero-error channel transmission. Define a zero-error achievable rate as the code rate at which there exists a block code such that the probability of error is strictly zero, and define the zero-error capacity as the supremum of all zero-error achievable rates.

a) Argue that for a BSC, the zero-error capacity is zero whenever $\delta > 0$.

b) Find the zero-error capacity of the noisy typewritter channel with 26 letters. The noisy typewritter channel is as follows: $\mathcal{X} = \mathcal{Y} = \{a, b, c, \dots, z\}$, $P_{Y|X}(a|a) = P_{Y|X}(b|b) = \cdots = P_{Y|X}(y|y) = P_{Y|X}(z|z) = 1/2$, and $P_{Y|X}(b|a) = P_{Y|X}(c|b) = \cdots = P_{Y|X}(z|y) = P_{Y|X}(a|z) = 1/2$.

c) Show that the zero-error rate $\frac{1}{2}\log_2 5$ bits/channel use is achievable for a noisy typewritter channel with $\mathcal{X} = \mathcal{Y} = \{a, b, c, d, e\}$.

a) 证明: 设正整数 $M \geq 2$, f 是从 $\{1, 2, \cdots, M\}$ 到 $\{0,1\}^n$ 的函数, g 是从 $\{0,1\}^n$ 到 $\{1, 2, \cdots, M\}$ 的函数. 我们来证明分组码 (f, g) 的错误概率大于 0. 任取

$$\underline{y} = (y_1, y_2, \cdots, y_n) \in \{0,1\}^n.$$

因为 $M \geq 2$, 所以存在正整数 $m \leq M$ 满足 $m \neq g(\underline{y})$. 设

$$f(m) = (x_1, x_2, \cdots, x_n).$$

用 (f, g) 发送的 m 被误认为是 $g(\underline{y})$ 的概率大于等于 $f(m)$ 被信道传成 $\underline{y}$ 的概率, 也就是 $\prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$. 因为 $P_{Y|X}(y|x) > 0$ 对所有 $x, y \in \{0,1\}$ 成立, 所以 $\prod_{i=1}^{n} P_{Y|X}(y_i|x_i) > 0$. 因此 (f, g) 的错误概率大于 0.

换句话说, 所有码率大于 0 的分组码都有大于 0 的错误概率, 所有大于 0 的码率都不可达. 作为可达码率的上确界, 零错容量等于 0. □

[?] 的 7.1.3 小节也讲到了打字机信道. 我们可以把打字机信道形象地画成图 1 的左半部分.

b) 解: 用

$$f(1) = \mathrm{a}, f(2) = \mathrm{c}, \cdots, f(13) = \mathrm{y},$$
$$g(\mathrm{a}) = g(\mathrm{b}) = 1, g(\mathrm{c}) = g(\mathrm{d}) = 2, \cdots, g(\mathrm{y}) = g(\mathrm{z}) = 13$$

定义 $f: \{1, 2, \cdots, 13\} \to \mathcal{X}$ 和 $g: \mathcal{Y} \to \{1, 2, \cdots, 13\}$. (f, g) 是一个长度为 1 的分组码, 可以用图 1 表示. (f, g) 不会出错, 且有码率 $\log_2(13)/1 = \log_2(13)$, 所以 $\log_2(13)$ 是零错可达码率. 零错容量大于等于 $\log_2(13)$.

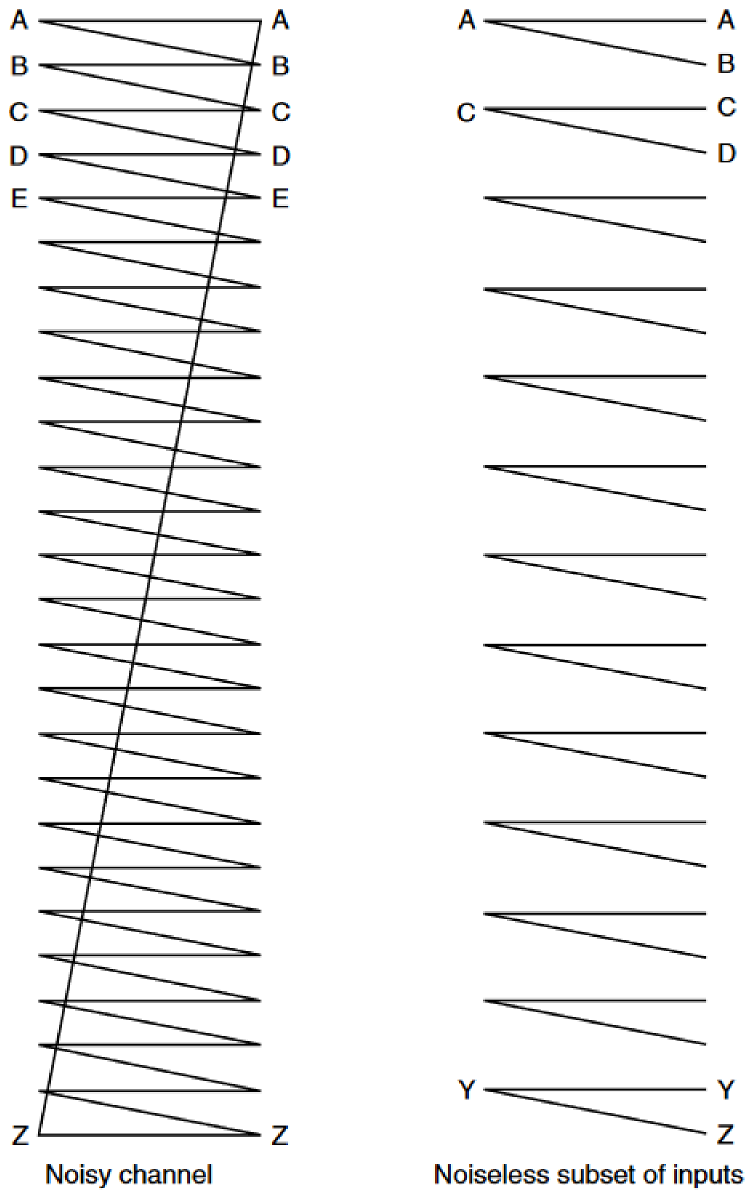

图 1: 打字机信道和它的一个错误概率为 0 的码. 图来自于 [?].

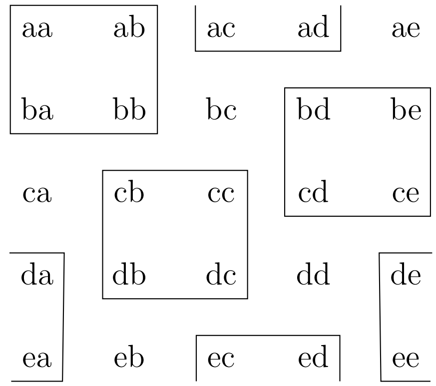

图 2: $\mathcal{X}^2 = \mathcal{Y}^2$ 中的 5 个不重叠的方块.

如果一列分组码的错误概率都等于 0, 则它们的错误概率当然也是趋于 0 的. 所以零错容量不会超过普通的信道容量即 $Y|X \sim P_{Y|X}$ 的条件下 $I(X;Y)$ 的最大值. 如果 $Y|X \sim P_{Y|X}$, 则 $I(X;Y) = H(Y) - H(Y|X) \le \log_2(26) - \log_2(2) = \log_2(13)$. 所以普通的信道容量不超过 $\log_2(13)$, 零错容量也不会超过 $\log_2(13)$.

综上所述, 零错容量等于 $\log_2(13)$. □

c) 证明: 我们可以把 $\mathcal{X}^2 = \mathcal{Y}^2$ 的所有元素排成图 2 所示的矩阵. $\mathcal{X}^2$ 中的字母对经信道传输可能不变, 也可能变成它在图 2 中右, 下或右下方的字母对. 这 4 个字母对组成一个 2×2 的方块. 这里认为矩阵是循环的, 如 ee 可能变成 ee, ea, ae 或 aa. 我们可以把图 2 中 5 个不重叠的 2×2 方块的左上角作为码字, 也就是用

$$f(1) = \mathrm{aa}, f(2) = \mathrm{cb}, f(3) = \mathrm{ec}, f(4) = \mathrm{bd}, f(5) = \mathrm{de}$$

定义 $f : \{1,2,3,4,5\} \to \mathcal{X}^2$. 这样存在 $g : \mathcal{Y}^2 \to \{1,2,3,4,5\}$ 使分组码 (f,g) 不会出错. (f,g) 的码率是 $\log_2(5)/2$, 所以 $\log_2(5)/2$ 是零错可达码率. □

第 7 题

We may combine several DMCs to create a sum DMC. Write a DMC $P_{Y|X}$ as a matrix whose i-th row j-th column element is $P_{Y|X}(j|i)$. Given DMCs $P_{Y|X,k}$ over input and output alphabets $\mathcal{X}_k$ and $\mathcal{Y}_k$, for $k = 1, \ldots, K$, the sum DMC has its matrix as $P_{Y|X,sum} = \operatorname{diag}\left\{P_{Y|X,1}, \ldots, P_{Y|X,K}\right\}$, i.e., a block-diagonal matrix whose diagonal submatrices $P_{Y|X,1}, \ldots, P_{Y|X,K}$. Calculate the capacity of $P_{Y|X,sum}$ and discuss the capacity-achieving input distribution in terms of the capacity-achieving input distributions for the component DMCs.

解: 我们记结合后X的分布为

$$(\underbrace{p_X(x_{1,1}),p_X(x_{1,2}),...,p_X(x_{1,n_1})}_{p_1},...,\underbrace{p_X(x_{k,1}),p_X(x_{k,2}),...,p_X(x_{k,n_k})}_{p_K}),$$

n_i为第i个子信道中的元素数量. 并记$p_X(x_{i,j})=p_ip_{X,i}(x_j), j=1,2,...,n_i$, 右侧第一项即为选中第$i$个DMC中元素的概率, 第二项为指定该信道下取到对应元素的概率, 我们用X_i表示这个概率分布下的随机变量, 也就是第i个信道的信源分布. 在这些符号辅助下, 我们试着展开互信息的表达式：

$$\begin{aligned}
I(X;Y) &= H(X)-H(X|Y)\\
&= H(p_1,p_2,...,p_K)+\sum_{i=1}^{K}p_iH(p_{X,i}(x_1),...,p_{X,i}(x_{n_i}))-H(X|Y)\\
&= H(p_1,p_2,...,p_K)+\sum_{i=1}^{K}p_iH(X_i)-\sum_{i=1}^{K}p_iH(X_i|Y_i)\\
&= H(p_1,p_2,...,p_K)+\sum_{i=1}^{K}p_iI(X_i;Y_i)
\end{aligned}\tag{5}$$

此处唯一需要注意的是(5)中使用$p(Y\in\mathcal{Y}_i)H(X|Y\in\mathcal{Y}_i)=p_iH(X_i|Y_i)$, 这是因为我们指定了$y\in\mathcal{Y}_i$之后, 发现$Y_i$只能由$X_i$转移得到, $p(Y\in\mathcal{Y}_i)=p_i$, 且：

$$\begin{aligned}
H(X|Y\in\mathcal{Y}_i) &= \sum_{j=1}^{m_i}p(Y=y_{i,j}|Y\in\mathcal{Y}_i)H(X|Y=y_{i,j},Y\in\mathcal{Y}_i)\\
&= \sum_{j=1}^{m_i}p_{Y,i}(y_j)H(X|Y_j=y_j)\\
&= \sum_{j=1}^{m_i}p_{Y,i}(y_j)H(X_i|Y_j=y_j)=H(X_i|Y_i)
\end{aligned}$$

从而对于原先的优化问题, 我们表示如下：

$$\begin{aligned}
\max_P I(X;Y) &= \max_P\max_{P_{X,i}}H(p_1,p_2,...,p_K)+\sum_{i=1}^{K}p_iI(X_i;Y_i)\\
&= \max_P H(p_1,p_2,...,p_K)+\sum_{i=1}^{K}p_iC_i\\
&= \max_P\sum_{i=1}^{K}p_i(C_i-\log_2(p_i)),\qquad \text{s.t.}\sum_{i=1}^{K}p_i=1.
\end{aligned}$$

对这个问题的求解, 我们使用Lagrange乘子法：

$$J(P)=\sum_{i=1}^{K}p_i(C_i-\log_2(p_i))+\lambda p_i,$$

$$J'(p_i) = C_i - \log_2(p_i) - 1 - \lambda = 0,$$

$$\Rightarrow p_i = 2^{\beta + C_i},\ \text{which } 2^{\beta} = \frac{1}{\sum_i 2^{C_i}}.\ \text{整理可得：} C = \log_2(\sum_{i=1}^{k} 2^{C_i}).$$

例：$P_{Y|X,1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,$P_{Y|X,2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$,$C_1 = 1, C_2 = \log_2(3)$, 带入结果可知信道容量为$\log_2(5)$. 也可以解出$(p_1, p_2) = (\frac{2}{5}, \frac{3}{5})$, 进一步根据各个信道本身达到信道容量的分布, 得知$P_{Y|X,sum} = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$, 同样可以检验上述结果. □