信息论第三讲作业解答

中国科学技术大学《信息论 A》006125.01 班助教组

2025 年 4 月 13 日

第 1 题

Consider a source with two components, (S_1, S_2), where S_1 and S_2 are independent, and let the distortion measure be of the form of $d((s_1, s_2), (\hat{s}_1, \hat{s}_2)) = d_1(s_1, \hat{s}_1) + d_2(s_2, \hat{s}_2)$. Denote the rate-distortion function of S_i under distortion d_i as $R_i(D), i = 1, 2$. Find the rate-distortion function $R(D)$, expressed in terms of $R_1(D)$ and $R_2(D)$.

解: 复合信源 (S_1, S_2) 的率失真函数为

$$
\begin{aligned}
R(D) &= \min_{\substack{P_{\hat{S}_1,\hat{S}_2|S_1,S_2} \\ \mathbf{E}[d((s_1,s_2),(\hat{s}_1,\hat{s}_2))]\le D}} I(S_1, S_2; \hat{S}_1, \hat{S}_2) \\
&\overset{(a)}{=} \min_{\substack{P_{\hat{S}_1,\hat{S}_2|S_1,S_2} \\ \mathbf{E}[d_1(s_1,\hat{s}_1)]+\mathbf{E}[d_2(s_2,\hat{s}_2)]\le D}} I(S_1, S_2; \hat{S}_1, \hat{S}_2) \\
&\overset{(b)}{\ge} \min_{\substack{P_{\hat{S}_1|S_1}, P_{\hat{S}_2|S_2} \\ \mathbf{E}[d_1(s_1,\hat{s}_1)]+\mathbf{E}[d_2(s_2,\hat{s}_2)]\le D}} I(S_1; \hat{S}_1) + I(S_2; \hat{S}_2) \\
&= \min_{D_1+D_2\le D} \min_{\substack{P_{\hat{S}_1|S_1}, P_{\hat{S}_2|S_2} \\ \mathbf{E}[d_1(s_1,\hat{s}_1)]=D_1, \mathbf{E}[d_2(s_2,\hat{s}_2)]=D_2}} I(S_1; \hat{S}_1) + I(S_2; \hat{S}_2) \\
&= \min_{D_1+D_2\le D} \left[\min_{\substack{P_{\hat{S}_1|S_1} \\ \mathbf{E}[d_1(s_1,\hat{s}_1)]=D_1}} I(S_1; \hat{S}_1) + \min_{\substack{P_{\hat{S}_2|S_2} \\ \mathbf{E}[d_2(s_2,\hat{s}_2)]=D_2}} I(S_2; \hat{S}_2) \right] \\
&= \min_{D_1+D_2\le D} R_1(D_1) + R_2(D_2),
\end{aligned}
$$

其中 (a) 基于题目中的失真度量形式, (b) 来源于

$$
\begin{aligned}
I(S_1, S_2; \hat{S}_1, \hat{S}_2) &= H(S_1, S_2) - H(S_1, S_2|\hat{S}_1, \hat{S}_2) \\
&= H(S_1) + H(S_2) - H(S_1|\hat{S}_1, \hat{S}_2) - H(S_2|S_1, \hat{S}_1, \hat{S}_2)
\end{aligned}
$$

$$\geq H(S_1) + H(S_2) - H(S_1|\hat{S}_1) - H(S_2|\hat{S}_2)$$
$$= I(S_1;\hat{S}_1) + I(S_2;\hat{S}_2).$$

当对S_1与S_2独立的编译码时, 即满足

$$P_{\hat{S}_1,\hat{S}_2,S_1,S_2}(\hat{s}_1,\hat{s}_2,s_1,s_2) = P_{S_1,\hat{S}_1}(s_1,\hat{s}_1)P_{S_2,\hat{S}_2}(s_2,\hat{s}_2)$$

时, 上述不等号的取等条件成立且满足失真约束. 因此$R(D)$满足

$$R(D) = \min_{D_1+D_2\leq D} R_1(D_1) + R_2(D_2).$$

□

第 2 题

If the rate-distortion function for a DMS S under distortion measure $d(s,\hat{s}), (s,\hat{s}) \in \mathcal{S}\times\hat{\mathcal{S}}$, is $R(D)$, what is the rate-distortion function when the distortion measure is changed to $d_{k,b}(s,\hat{s}) = kd(s,\hat{s}) + b$ for some $k > 0, b \geq 0$?

解:

$$\begin{aligned}
R_{k,b}(D) &= \min_{P_{\hat{S}|S}:\mathbf{E}\left[d_{k,b}(s,\hat{s})\right]\leq D} I(S;\hat{S}) \\
&= \min_{P_{\hat{S}|S}:\mathbf{E}[kd(s,\hat{s})+b]\leq D} I(S;\hat{S}) \\
&= \min_{P_{\hat{S}|S}:\mathbf{E}[d(s,\hat{s})]\leq \frac{D-b}{k}} I(S;\hat{S}) \\
&= R\left(\frac{D-b}{k}\right).
\end{aligned}$$

□

第 3 题

We may use a matrix $\mathbf{D}$ to collectively represent the distortion measure for source and reproduction with finite alphabets; that is, the i-th row j-th column of $\mathbf{D}$ corresponds to $d(s_i,\hat{s}_j)$. If $\mathbf{D}$ satisfies that all its columns are permutations of a certain vector $\left[d_1, d_2, \ldots, d_{|\mathcal{S}|}\right]$, prove the following lower bound of the rate-distortion function:

$$R(D) \geq H(S) - H(V) \tag{1}$$

where V is the random variable that attains the largest entropy among all random variables over $\{1, 2, \ldots, |\mathcal{S}|\}$ satisfying $\sum_{i=1}^{|\mathcal{S}|} P_V(i) d_i \leq D$. Furthermore, prove that if S is uniform and the rows of $\mathbf{D}$ are permutations of each other, then the lower bound (1) is tight, i.e., it is exactly the rate-distortion function $R(D)$.

证明:

$$\begin{aligned}
I(S;\hat{S}) =& H(S) - H(S|\hat{S}) \\
=& H(S) - \sum_j p(\hat{s}_j) H(S|\hat{S}=\hat{s}_j) \\
\geq& H(S) - \sum_j p(\hat{s}_j) H_{D_j}(V) \qquad (2) \\
\geq& H(S) - H_{\sum_j p(\hat{s}_j) D_j}(V) \qquad (3) \\
\geq& H(S) - H_D(V). \qquad (4)
\end{aligned}$$

从而 $R(D) \geq H(S) - H(V)$，其中, 我们定义 $H_{D_j}(V) = \max\limits_{V:\sum_{i=1}^{|\mathcal{S}|} P_V(i) d_i \leq D_j} H(V)$, $D_j = \sum_{i=1}^{|\mathcal{S}|} p(s_i|\hat{s}_j) d(s_i, \hat{s}_j)$. 关于 (3) 的证明, 如下：

$$\begin{aligned}
\lambda_1 H_{D_1}(V) + \lambda_2 H_{D_2}(V) &= \lambda_1 \max_{\sum_i P_V(i) d_i \leq D_1} H(V) + \lambda_2 \max_{\sum_i P_V(i) d_i \leq D_2} H(V) \\
&= \lambda_1 H(p^*) + \lambda_2 H(q^*) \qquad (5) \\
&\leq H(\lambda_1 p^* + \lambda_2 q^*) \qquad (6) \\
&\leq \max_{\sum_i P_V(i) d_i \leq \lambda_1 D_1 + \lambda_2 D_2} H(V) \qquad (7) \\
&= H_{\lambda_1 D_1 + \lambda_2 D_2}(V).
\end{aligned}$$

其中 $\lambda_1 + \lambda_2 = 1$, (5) 表明找到了两个最优的分布, (6) 使用了熵函数的凹性, (7) 因为这个分布满足优化问题的失真约束 $\sum_i (\lambda_1 p_i^* + \lambda_2 q_i^*) d_i = \sum_i \lambda_1 p_i^* d_i + \sum_i \lambda_2 q_i^* d_i \leq \lambda_1 D_1 + \lambda_2 D_2$.

关于 (4) 式的证明, 我们知道 $\sum_j p(\hat{s}_j) D_j = \sum_j p(\hat{s}_j) \sum_i p(s_i|\hat{s}_j) d(s_i, \hat{s}_j) \leq D$, 又 $H_{D_j}(V)$ 是关于 D_j 的单调增函数，所以有 $H_{\sum_j p(\hat{s}_j) D_j}(V) \leq H_D(V)$.

当 S 服从均匀分布并且行元素是可排列的, 我们可以构造出一个达到这个下界的概率转移矩阵. 令 $p^* = (p_1^*, p_2^*, ..., p_{|\mathcal{S}|}^*)$ 为 $H_D(V)$ 取到极大值时的 p. 我们说明可以通过 p 实现一个反向传播的信道, 我们取 $p(s=i \mid \hat{s}=j) = p_k^*$, 其中 k 满足 $d(s_i, \hat{s}_j) = d_k$, 在这种信道假设下, 我们取 $p(\hat{s}_j) = 1/|\mathcal{S}|$, 这样我们就可以得到

$$p(s_i) = \sum_j p(\hat{s}_j) p(s_i \mid \hat{s}_j) = \frac{1}{|\mathcal{S}|},$$

$$D_j = \sum_{i=1}^{|\mathcal{S}|} p(s_i \mid \hat{s}_j) d(s_i, \hat{s}_j) = \sum_{i=1}^{|\mathcal{S}|} p_k^* d_k = D, \text{ for all j}$$

所以对于平均失真，我们有

$$\sum_{i,j} p(\hat{s}_j)p(s_i|\hat{s}_j)d(s_i,\hat{s}_j) = \sum_j p(\hat{s}_j)D = D,$$

此时

$$\begin{aligned} I(S;\hat{S}) =& H(S) - \sum_j p(\hat{s}_j)H(S|\hat{S}=\hat{s}_j) \\ =& H(S) - \sum_j p(\hat{s}_j)H_D(V) \\ =& H(S) - H_D(V) \\ =& \log_2|\mathcal{S}| - H(V). \end{aligned}$$

又因为

$$\begin{aligned} R(D) &\le I(S;\hat{S}) \\ &= H(S) - H(V) \end{aligned}$$

从而对于 S 均匀分布且失真矩阵行列是重排的的信道模型, 率失真函数 $R(D)$ 达到下界，即 $R(D) = H(S) - H(V)$.

注 1. 本题中 S 均匀分布且失真矩阵行列是重排是率失真函数等于下界的一个充分条件, 但是仅仅从题目条件是不能推出达到下界时就有 S 均匀分布, 一个反例就是 $B(\delta)$ 的信道.

□

第 4 题

Calculate and plot the rate-distortion functions for the following models.
a) Source alphabet is $\mathcal{S} = \{-1,0,1\}$, reproduction alphabet is $\hat{\mathcal{S}} = \{-1,1\}$, source distribution is uniform $P_S(s) = \{1/3,1/3,1/3\}$, and distortion measure is

$$\mathbf{D} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

b) Source alphabet is $\mathcal{S} = \{-1,1\}$, reproduction alphabet is $\hat{\mathcal{S}} = \{-1,0,1\}$, source distribution is uniform $P_S(s) = \{1/2,1/2\}$, and distortion measure is

$$\mathbf{D} = \begin{bmatrix} 0 & 1/2 & 1 \\ 1 & 1/2 & 0 \end{bmatrix}.$$

解: a) 根据失真矩阵 $\mathbf{D}$ 的定义, 我们可以得到

$$D_{\min} = \sum_{s\in\mathcal{S}} P_S(s) \min_{\hat{s}\in\hat{\mathcal{S}}} d(s,\hat{s}) = 0, \qquad D_{\max} = \min_{\hat{s}\in\hat{\mathcal{S}}} \sum_{s\in\mathcal{S}} P_S(s) d(s,\hat{s}) = \frac{1}{3}.$$

方法一：

因此当 $D \geq 1/3$ 时, $R(D) = 0$. 当 $D \leq 1/3$ 时

$$\begin{aligned} I(S;\hat{S}) &= H(S) - H(S \mid \hat{S}) \\ &= \log_2 3 - H(S \mid \hat{S}). \end{aligned}$$

定义 $Y = S\hat{S}$, 则有 $H(S \mid \hat{S}) = H(S\hat{S} \mid \hat{S}) = H(Y \mid \hat{S})$, 第一个等号是因为在给定 $\hat{S}$ 时, S 到 $S\hat{S}$ 为双射. 因此有:

$$\begin{aligned} I(S;\hat{S}) &= \log_2 3 - H(S \mid \hat{S}) \\ &\geq \log_2 3 - H(Y). \end{aligned}$$

要最小化 $I(S;\hat{S})$, 则需最大化 $H(Y)$. 注意到当 $Y = -1$ 时, 失真 $d = 1$, 而当 $Y = 0$ 或 1 时, $d = 0$. 另外 $Y = 0$ 当且仅当 $S = 0$, 因此 $P(Y = 0) = 1/3$. 设 $P(Y = -1) = p$, 则 $P(Y = 1) = 2/3 - p$. 由失真约束, 有 $p \leq D \leq \frac{1}{3}$. 由熵的凹性可以说明, 当 $p = D$ 时, $H(Y)$ 达到最大. 此时, 有:

$$H(Y) = \frac{1}{3}\log_2 3 - D\log_2 D - \left(\frac{2}{3} - D\right)\log_2\left(\frac{2}{3} - D\right).$$

因此:

$$\begin{aligned} I(S;\hat{S}) &\geq \log_2 3 - H(Y) \\ &= \frac{2}{3}\log_2 3 + D\log_2 D + \left(\frac{2}{3} - D\right)\log_2\left(\frac{2}{3} - D\right). \end{aligned}$$

而为了满足上述不等式要求,我们构造反向传播信道：$P(\hat{S} = -1) = P(\hat{S} = 1) = 1/2, P(S = -1 \mid \hat{S} = -1) = P(S = 1 \mid \hat{S} = 1) = 2/3 - D, P(S = 0 \mid \hat{S} = -1) = P(S = 0 \mid \hat{S} = 1) = 1/3, P(S = -1 \mid \hat{S} = 1) = P(S = -1 \mid \hat{S} = 1) = D$, 即构造如下从$\hat{S}$到$S$的转移概率矩阵：

$$\begin{bmatrix} 2/3 - D & 1/3 & D \\ D & 1/3 & 2/3 - D \end{bmatrix}.$$

综上：

$$R(D) = \begin{cases} \frac{2}{3}\log_2 3 + D\log_2 D + \left(\frac{2}{3} - D\right)\log_2\left(\frac{2}{3} - D\right) & 0 \leq D < 1/3 \\ 0 & D \geq 1/3 \end{cases}.$$

方法二：上面所提到的方法沿用了我们在计算伯努利信道时, 使用反向传播信道的方法. 我们再给出一种相对容易理解的计算方法.

$$\begin{aligned} I(S;\hat{S}) &= \log_2 3 - H(S \mid \hat{S}) \\ &= H(\hat{S}) - H(\hat{S} \mid S). \end{aligned}$$

我们取两个正向的转移概率矩阵如下：

$$Q_1 = \begin{bmatrix} p_1 & 1-p_1 \\ p_3 & 1-p_3 \\ 1-p_2 & p_2 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} p_2 & 1-p_2 \\ 1-p_3 & p_3 \\ 1-p_1 & p_1 \end{bmatrix}.$$

我们可以发现, Q_1, Q_2 这两个转移概率矩阵对应的失真相同, 但是根据互信息关于转移概率矩阵的凸性, 取 $Q = \frac{Q_1+Q_2}{2}$ 在达到相同失真的同时, 具有更小的互信息. 从而我们可以假设转移概率矩阵服从如下分布：

$$Q = \begin{bmatrix} p & 1-p \\ 1/2 & 1/2 \\ 1-p & p \end{bmatrix}.$$

在此基础上, 我们进行失真的计算：

$$\mathbf{E}[d(S,\hat{S})] = 2 * \frac{1}{3}(1-p) \le D, \tag{8}$$

$$I(S;\hat{S}) = H(\hat{S}) - H(\hat{S} \mid S) = 1 - \frac{1}{3}(1 + 2h_2(p)). \tag{9}$$

同样由熵的凹性, 我们选择 $p = 1 - \frac{3D}{2}$, 恰好满足失真约束, 带入 9 式可得到答案.

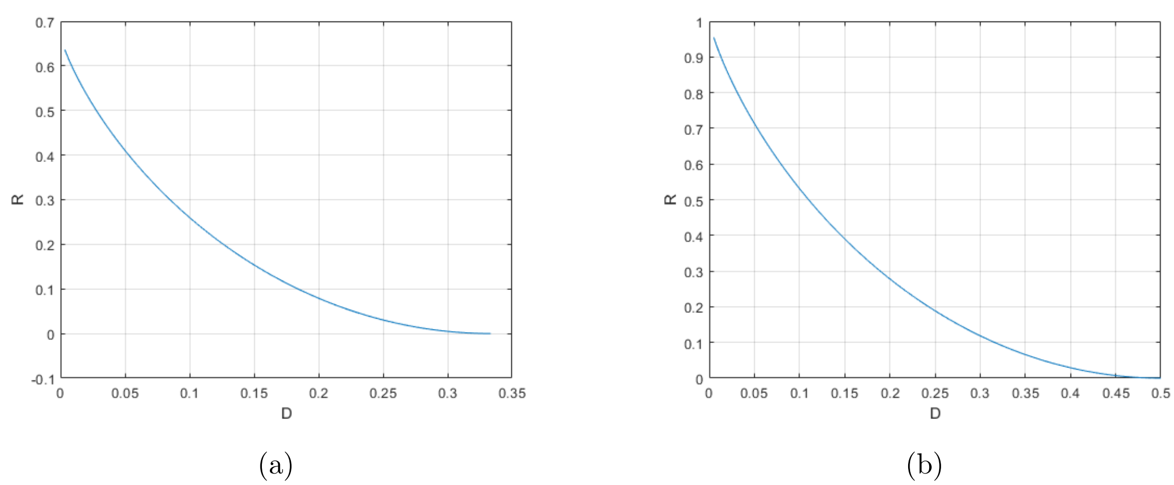

图 1: 第 4 题的 R-D 曲线

b) Source alphabet is $\mathcal{S}=\{-1,1\}$, reproduction alphabet is $\hat{\mathcal{S}}=\{-1,0,1\}$, source distribution is uniform $P_S(s)=\{1/2,1/2\}$, and distortion measure is

$$\mathcal{D}=\begin{bmatrix} 0 & 1/2 & 1 \\ 1 & 1/2 & 0 \end{bmatrix}.$$

解：当 $D\geq 1/2, R(D)=0$, 下面考虑 $D\leq 1/2$ 时的情况. 设转移概率矩阵 Q_1 和 Q_2, 及其对应的互信息为 $I(Q_1)$ 和 $I(Q_2)$, 设

$$Q_1=\begin{bmatrix} 1-q_1-q_2 & q_1 & q_2 \\ q_4 & q_3 & 1-q_3-q_4 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} 1-q_3-q_4 & q_3 & q_4 \\ q_2 & q_1 & 1-q_1-q_2 \end{bmatrix}.$$

可以发现 Q_1 和 Q_2 对应的期望失真相等, 令 $Q=\frac{Q_1+Q_2}{2}$, 则 Q 对应的失真同样与 Q_1 和 Q_2 的失真相等. 由于互信息对 Q 的凸性, 有 $I(Q)\leq I(Q_1)=I(Q_2)$. 但此时的优化问题求解较为复杂. 因此我们考虑对称的反向传播信道.

考虑如下反向测试信道, $P(\hat{S}=-1)=P(\hat{S}=1)=r/2, P(\hat{S}=0)=1-r, \hat{S}$ 到 S 的转移概率矩阵为

$$\begin{bmatrix} 1-p & p \\ 1/2 & 1/2 \\ p & 1-p \end{bmatrix},$$

$$\mathbb{E}[d]=rp+\frac{1-r}{2}\leq D, \tag{10}$$

$$\begin{aligned} I(S;\hat{S}) &= H(S)-H(S\mid\hat{S}) \\ &= 1-rh_2(p)-(1-r) \qquad (11) \\ &= r(1-h_2(p)). \end{aligned}$$

所以有:

$$\begin{aligned} I(S;\hat{S}) &= r(1-h_2(p)) \qquad (12) \\ &\geq (1-2D)\frac{1-h_2(p)}{1-2p} \qquad (13) \\ &\geq 1-h_2(D). \end{aligned}$$

其中 (12) 是因为由于 (10) 得到 $0\leq p\leq D$ 以及 $r\geq\frac{1-2D}{1-2p}$, (13) 是因为 $\frac{1-h_2(p)}{1-2p}$ 在 $[0,1/2]$ 上单调递减, 取 $p=D$ 有极小值. 从而率失真函数为：

$$R(D)=\begin{cases} 1-h_2(D) & 0\leq D<1/2 \\ 0 & D\geq 1/2 \end{cases}.$$

□

第 5 题

If a DMS S is Bernoulli with parameter δ and the distortion measure is Hamming, when $P_{\hat{S}|S}$ achieves $R(D)$, describe the encoding procedures (4.38) and (4.39) (4.40), and find out their differences.

解: 我们只考虑讲义例子中的情况即 $0 \le D \le \delta \le 1/2$. 对所有 s, $\hat{s} \in \{0,1\}$ 记

$$L_{s,\hat{s}} = \log_2\left(\frac{P_{S,\hat{S}}(s,\hat{s})}{P_S(s)P_{\hat{S}}(\hat{s})}\right).$$

在讲义的例子中我们已经求出 $P_{S|\hat{S}}(0|0) = P_{S|\hat{S}}(1|1) = 1-D$, $P_{S|\hat{S}}(1|0) = P_{S|\hat{S}}(0|1) = D$ (见讲义 (4.27) 式上面的一段话,即反向传播信道的构造), 所以

$$L_{0,0} = \log_2\left(\frac{1-D}{1-\delta}\right), L_{0,1} = \log_2\left(\frac{D}{1-\delta}\right), L_{1,0} = \log_2\left(\frac{D}{\delta}\right), L_{1,1} = \log_2\left(\frac{1-D}{\delta}\right).$$

设 $\underline{s} = (s_1, s_2, \cdots, s_n)$, $\underline{\hat{s}} = (\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_n)$. 对所有 s, $\hat{s} \in \{0,1\}$, 我们用 $N(s,\hat{s}|\underline{s},\underline{\hat{s}})$ 表示满足 $i \le n$, $s_i = s$ 和 $\hat{s}_i = \hat{s}$ 的正整数 i 的个数.

对于最优编码规则 (4.38), 我们先分析 $d(\underline{s},\underline{\hat{s}})$:

$$\begin{aligned} d(\underline{s},\underline{\hat{s}}) &= \frac{1}{n}\sum_{s=0}^{1}\sum_{\hat{s}=0}^{1} N(s,\hat{s}|\underline{s},\underline{\hat{s}})d(s,\hat{s}) \\ &= \frac{1}{n}(N(0,1|\underline{s},\underline{\hat{s}}) + N(1,0|\underline{s},\underline{\hat{s}})) \end{aligned} \tag{14}$$

对于编码规则 (4.39) (4.40), 我们有 $d(\underline{s},\underline{\hat{s}}) \le D + \epsilon$ 等价于

$$N(0,1|\underline{s},\underline{\hat{s}}) + N(1,0|\underline{s},\underline{\hat{s}}) \le n(D+\epsilon) \tag{15}$$

即 $\underline{s}$ 和 $\underline{\hat{s}}$ 不相同的符号个数不超过 $n(D+\epsilon)$. 由于

$$\begin{aligned} i(\underline{s},\underline{\hat{s}}) &= \frac{1}{n}\sum_{s=0}^{1}\sum_{\hat{s}=0}^{1} N(s,\hat{s}|\underline{s},\underline{\hat{s}})\log_2\left(\frac{P_{S,\hat{S}}(s,\hat{s})}{P_S(s)P_{\hat{S}}(\hat{s})}\right) \\ &= \frac{1}{n}\sum_{s=1}^{1}\sum_{\hat{s}=0}^{1} N(s,\hat{s}|\underline{s},\underline{\hat{s}})L_{s,\hat{s}} \end{aligned}$$

且 $I(S;\hat{S}) = R(D) = h_2(\delta) - h_2(D)$, $i(\underline{s},\underline{\hat{s}}) \le I(S;\hat{S}) + \epsilon$ 等价于

$$\sum_{s=0}^{1}\sum_{\hat{s}=0}^{1} N(s,\hat{s}|\underline{s},\underline{\hat{s}})L_{s,\hat{s}} \le n(h_2(\delta) - h_2(D) + \epsilon) \tag{16}$$

最后我们总结如下, 对于 $\epsilon > 0$, 给定 $\underline{S} = \underline{s}$,

- 最优编码规则 (4.38) 为找到满足下列优化问题的 w^*

$$w^* = \arg\min_{w\in\{1,\dots,M_n\}} \frac{1}{n}\left(N(0,1|\underline{s},\mathcal{C}(w)) + N(1,0|\underline{s},\mathcal{C}(w))\right)$$

- 编码规则 (4.39) (4.40) 为：
  如果存在 $w \in \{1,2,\cdots,M_n\}$ 满足

$$N(0,1|\underline{s},\mathcal{C}(w)) + N(1,0|\underline{s},\mathcal{C}(w)) \le n(D+\epsilon)$$

$$\sum_{s=0}^{1}\sum_{\hat{s}=0}^{1} N(s,\hat{s}|\underline{s},\mathcal{C}(w))L_{s,\hat{s}} \le n(h_2(\delta) - h_2(D) + \epsilon)$$

  则令 w^* 等于最小的这样的 w; 否则令 $w^* = 1$.

编码规则 (4.39) (4.40) 存在编码不成功的时候，最优编码规则 (4.38) 总是可以找到当前码本下的最优码字. □

第 6 题

Consider a Bernoulli(1/2) *DMS S. If we use the following codebook:*

$$\mathbf{C} = \{0000, 0101, 1010, 1111\},$$

calculate the expected Hamming distortion with the encoding procedure (4.38).

解: 我们知道 (4.38) 是最优编码准则，即在码本 $\mathbf{C}$ 中找到能使当前信源符号 $\underline{s}$ 失真最小的码字对应的索引 w^*。根据最优编码 $w^* = \arg\min_{w\in\{1,\dots,4\}} d(\underline{s},\mathbf{C}(w))$, 我们用下面的表格给出一种编码方法：

此时我们可以计算出期望失真为

$$\begin{aligned} D &= \mathbf{E}[d(\underline{S},\underline{\hat{S}})] \\ &= \sum_{\underline{s}} P_{\underline{S}}(\underline{s})d(\underline{s},\mathbf{C}(w^*)) \\ &= \frac{1}{4}\times\frac{1}{16}(0\times 4 + 1\times 8 + 2\times 4) \\ &= \frac{1}{4} \end{aligned}$$

□

表 1: 第 6 题一种最优编码方法

$\underline{s}$	w^*	$\mathbf{C}(w^*)$	$d(\underline{s}, \mathbf{C}(w^*))$
0000	1	0000	$\frac{1}{4} \times 0$
0001	1	0000	$\frac{1}{4} \times 1$
0010	1	0000	$\frac{1}{4} \times 1$
0011	1	0000	$\frac{1}{4} \times 2$
0100	1	0000	$\frac{1}{4} \times 1$
0101	2	0101	$\frac{1}{4} \times 0$
0110	1	0000	$\frac{1}{4} \times 2$
0111	2	0101	$\frac{1}{4} \times 1$
1000	1	0000	$\frac{1}{4} \times 1$
1001	1	0000	$\frac{1}{4} \times 2$
1010	3	1010	$\frac{1}{4} \times 0$
1011	4	1111	$\frac{1}{4} \times 1$
1100	1	0000	$\frac{1}{4} \times 2$
1101	4	1111	$\frac{1}{4} \times 1$
1110	4	1111	$\frac{1}{4} \times 1$
1111	4	1111	$\frac{1}{4} \times 0$

第 7 题

Suppose that there are two independent Bernoulli DMSs S_1 and S_2 with parameters $\delta_1 \le 1/2$ and $\delta_2 \le 1/2$, respectively. At each time we sample one of the DMSs in a memoryless fashion so as to get a new DMS, denoted by S. Let the probability of sampling S_1 (resp. S_2) be λ (resp. $1-\lambda$). What is the rate-distortion function of S under Hamming distortion?

解: 根据题意，我们有

$$\begin{aligned} P(S=1) &= \lambda\delta_1 + (1-\lambda)\delta_2, \\ P(S=0) &= \lambda(1-\delta_1) + (1-\lambda)(1-\delta_2). \end{aligned}$$

且 $\lambda\delta_1 + (1-\lambda)\delta_2 = \lambda(\delta_1 - \delta_2) + \delta_2 \le \delta_1 - \delta_2 + \delta_2 = \delta_1 \le \frac{1}{2}$, 从而依据讲义中 Example 4.2 的结果，即可得

$$R(D) = h_2(\lambda\delta_1 + (1-\lambda)\delta_2) - h_2(D) \text{ if } 0 \le D \le \lambda\delta_1 + (1-\lambda)\delta_2, \text{ and } R(D) = 0 \text{ otherwise.}$$

□

第 8 题

Consider a DMS S uniform over $\mathcal{S} = \{1, \ldots, m\}$, with reproduction alphabet $\hat{\mathcal{S}} = \mathcal{S}$, and Hamming distortion measure. Calculate the rate-distortion function $R(D)$. This also gives an example where the Fano's inequality (Theorem 3.9) attains equality.

解:

根据汉明失真，首先由定义得到 $D_{\min} = 0$, $D_{\max} = 1 - \frac{1}{m}$. 当 $m = 1$ 时, 显然有 $R(D) = 0$, 故下文考虑 $m \ge 2$.

由于失真度量与错误概率满足 $D = \Pr(S \ne \hat{S})$, 因此由 Fano's inequality 可得

$$H(S|\hat{S}) \le h_2(D) + D\log(m-1). \tag{17}$$

由于 S 服从均匀分布, 因此有

$$\begin{aligned} I(S;\hat{S}) &= H(S) - H(S|\hat{S}) \\ &\ge \log m - h_2(D) - D\log(m-1), \end{aligned} \tag{18}$$

其中不等号是由于 Fano's inequality 不等式 (17).

定义随机变量 E 满足当 $\hat{S} \ne S$ 时 $E = 1$, 否则 $E = 0$. 由 Fano's inequality 的取等条件可知, 当且仅当以下两个条件同时满足时 (可参考第三章中的练习 8), 不等式 (18) 取等.

1. $\hat{S}$ 与 E 独立;
2. $P_{S|\hat{S},E}(s|\hat{s},1)$ 满足均匀分布.

因此不妨设定 $\hat{S}$ 服从均匀分布, 反向测试信道分布为

$$P_{S|\hat{S}}(s|\hat{s}) = \begin{cases} 1-D & \text{if } \hat{s}=s \\ D/(m-1) & \text{if } \hat{s}\neq s \end{cases}.$$

可验证此时 S 服从均匀分布且满足失真约束. 综上, 此时率失真函数为

$$R(D) = \begin{cases} \log m - h_2(D) - D\log(m-1) & \text{if } 0 \le D \le 1-\frac{1}{m} \\ 0 & \text{if } D > 1-\frac{1}{m} \end{cases}.$$

□

第 9 题

In the problem formulation in Section 4.1, instead of assuming that the distortion measure d is a mapping from $\mathcal{S}\times\hat{\mathcal{S}}$ to $[0,\infty)$, let it be a random variable $U\in[0,\infty)$ generated according to some conditional probability distribution $P_{U|S,\hat{S}}(u|s,\hat{s})$. Generalize Theorem 4.1 to address this extended problem formulation, and point out necessary modifications in the proofs of converse and achievability.

解: 根据题意失真度量 d 定义为 $d(s,\hat{s}) = \mathbf{E}[U|S=s,\hat{S}=\hat{s}]$ 以及

$$\mathbf{E}[U|\underline{S}=\underline{s},\underline{\hat{S}}=\underline{\hat{s}}] = d(\underline{s},\underline{\hat{s}}) = \sum_{i=1}^{n} d(s_i,\hat{s}_i) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[U|S_i=s_i,\hat{S}_i=\hat{s}_i] \tag{19}$$

所以可达性定义 4.1 中的 (4.7) 式应该写为

$$\lim_{n\to\infty}\mathbf{E}[\mathbf{E}[U|\underline{S},\underline{\hat{S}}]] \le D \tag{20}$$

以及信息率失真函数的定义 4.3 应该写为

$$R_I(D) = \min_{P_{\hat{S}|S}:\mathbf{E}[\mathbf{E}[U|S,\hat{S}]]\le D} I(S;\hat{S}) \tag{21}$$

其余定义不变，我们依然可以证明 $R(D)=R_I(D)$.

首先关于逆定理的证明，我们只需将讲义中 (4.34)-(4.36) 式改为

$$\begin{aligned} R &\ge I(S_Q;\hat{S}_Q) \\ &\ge R_I(\mathbf{E}[\mathbf{E}[U|S_Q,\hat{S}_Q]]) \end{aligned}$$

$$
\begin{aligned}
&= R_I(\mathbf{E}[\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[U|S_i,\hat{S}_i]]) \\
&= R_I(\mathbf{E}[\mathbf{E}[U|\underline{S},\underline{\hat{S}}]]) \\
&\geq R_I(D)
\end{aligned}
$$

上式的最后一个等号基于 (19) 式.

对于可达性的证明，编码规则 (4.39) 重新写为

$$
\mathbf{E}[U|\underline{s},\mathbf{C}(w)] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[U|s_i,\mathbf{C}_i(w)] \leq D+\epsilon \tag{22}
$$

所以讲义中 (4.43) 式重新写为

$$
\begin{aligned}
\mathbf{E}_{\mathbf{C},\underline{S}}[\mathbf{E}[U|\underline{S},\underline{\hat{S}}]] = &\sum_{C,\underline{s}\in\mathcal{S}^n:E(\mathbf{C},\underline{s})=0} P_{\mathbf{C}}(C)P_{\underline{S}}(\underline{s})\mathbf{E}[U|\underline{s},\mathbf{C}(w)] \\
&+\sum_{C,\underline{s}\in\mathcal{S}^n:E(\mathbf{C},\underline{s})=1} P_{\mathbf{C}}(C)P_{\underline{S}}(\underline{s})\mathbf{E}[U|\underline{s},\mathbf{C}(1)]
\end{aligned}
$$

基于此我们依然可以得到第一个求和项的上界为 $D+\epsilon$，第二个求和项的上界为 $P_f d_{\max}$，对 P_f 的后续分析与讲义相同，即可达性证明依然成立。

□

第 10 题

Our statement of Shannon's fundamental theorem for source coding explicitly makes the assumption that the distortion measure $d(s,\hat{s})$ is bounded, i.e., $d_{\max} < \infty$.

a) *Explain why the proof of the achievability part of the theorem may break down if the distortion measure is unbounded, i.e., if there exists some $(s,\hat{s}) \in \mathcal{S}\times\hat{\mathcal{S}}$ such that $d(s,\hat{s}) = \infty$.*
b) *Consider the case where there exists $\hat{s}^* \in \hat{\mathcal{S}}$ such that $d(s,\hat{s}^*) < \infty$, $\forall s \in \mathcal{S}$. Argue that in this case, the rate-distortion function $R(D)$ is still given by Theorem 4.1, i.e., $\min_{P_{\hat{S}|S}} I(S;\hat{S})$ subject to $\mathbf{E}[d(S,\hat{S})] \leq D$.*
c) *Argue that there exist some cases of $(\mathcal{S},\hat{\mathcal{S}},d(s,\hat{s}))$ such that the rate-distortion function has to be as large as $\log|\mathcal{S}|$; that is, no efficient representation is possible and $R(D)$ may differ from $R_I(D)$.*
d) *Calculate the rate-distortion function of the following setup: $\mathcal{S}=\{0,1\}$, $\hat{\mathcal{S}}=\{0,1,e\}$, S is Bernoulli$(1/2)$, $d(s,\hat{s}) = 0$ if $\hat{s}=s$, 1 if $\hat{s}=e$, and ∞ if $\hat{s}\neq s$ and $\hat{s}\neq e$.*

a) 解: $d(s,\hat{s}) \le d_{\max}$ 对所有 $s \in \mathcal{S}$ 和 $\hat{s} \in \hat{\mathcal{S}}$ 成立的假设在讲义中被用来证明 (4.43) 式的第三行

$$\sum_{C,\underline{s} \in \mathcal{S}^n : E(\mathbf{C},\underline{s})=1} P_{\mathbf{C}}(C) P_{\underline{S}}(\underline{s}) d(\underline{s}, \mathcal{C}(1)) \tag{23}$$

不超过 $P_f d_{\max}$. 如果存在 $s \in \mathcal{S}$ 和 $\hat{s} \in \hat{\mathcal{S}}$ 使 $d(s,\hat{s}) = \infty$ 则(23)式可能等于 ∞. 这样就无法证明 n 很大时 $\mathbf{E}_{\mathbf{C},\underline{S}}[d(\underline{S},\underline{\hat{S}})] \le D + 2\epsilon$. □

b) 解: 讲义第三节的关于逆定理的证明没有用到 d 有界的假设, 所以这一节的结论依然成立, 即如果 $R < R_{\mathrm{I}}(D)$ 则 (R,D) 不可达.

d 无界给讲义第四节的证明带来的唯一问题是 (23) 式可能等于 ∞. 我们可以向码本添加由 n 个 $\hat{s}^*$ 组成的第 0 行 $\mathbf{C}(0)$ 并修改编码器来解决这个问题.

编码规则 (4.39) (4.40) 可以修改为: 如果存在 $w \in \{1,2,\cdots,M_n\}$ 满足 $d(\underline{s},\mathbf{C}(w)) \le D+\epsilon$ 和 $i(\underline{s},\mathbf{C}(w)) \le I(S;\hat{S}) + \epsilon$, 编码器还是输出最小的这样的 w. 但如果不存在这样的 w, 编码器应该输出 0. 这样讲义的 (4.43) 式变成

$$\mathbf{E}_{\mathbf{C},\underline{S}}[d(\underline{S},\underline{\hat{S}})] = \sum_{C,\underline{s} \in \mathcal{S}^n : E(\mathbf{C},\underline{s})=0} P_{\mathbf{C}}(C) P_{\underline{S}}(\underline{s}) d(\underline{s}, \mathcal{C}(w)) + \sum_{C,\underline{s} \in \mathcal{S}^n : E(\mathbf{C},\underline{s})=1} P_{\mathbf{C}}(C) P_{\underline{S}}(\underline{s}) d(\underline{s}, \mathcal{C}(0)).$$

对每个 $\underline{s} = (s_1, s_2, \cdots, s_n) \in \mathcal{S}^n$ 有

$$d(\underline{s}, \mathcal{C}(0)) = \frac{1}{n} \sum_{i=1}^{n} d(s_i, \hat{s}^*) \le \max_{s \in \mathcal{S}} d(s, \hat{s}^*) < \infty.$$

所以只要 P_f 很小, $\mathbf{E}_{\mathcal{C},\underline{S}}[d(\underline{S},\underline{\hat{S}})]$ 还是可以小于 $D+2\epsilon$. 因此若 $R > R_{\mathrm{I}}(D)$ 则 (R,D) 可达.

综上所述, 使 (R,D) 可达的所有 R 组成的集合有下确界 $R_{\mathrm{I}}(D)$, 率失真函数的值等于 $R_{\mathrm{I}}(D)$. □

c) 解: 用

$$d(s,\hat{s}) = \begin{cases} 0, & s = \hat{s} \\ \infty, & s \ne \hat{s} \end{cases}$$

定义 $\{0,1\}^2$ 上的失真度量 d. 设一个无记忆信源的符号 S_1, S_2, $\cdots$ 取值于 $\{0,1\}$, $P_{S_1}(0)$ 和 $P_{S_1}(1)$ 都大于 0. 用 R 表示这个信源在 d 下的率失真函数. 对每个正整数 n 定义 $\underline{S}_n = [S_1, S_2, \cdots, S_n]$. 由于 $d(0,1) = d(1,0) = \infty$, 我们无法用讲义中的定理 4.1 求 R. 接下来我们用率失真函数的定义求 R.

设 (R,D) 是可达率失真对. 根据讲义定义 4.1, 存在一列编译码器 $\{(f_n,g_n)\}_{n=1}^{\infty}$ 满足

$$\frac{\log_2(M_n)}{n} = R, \tag{24}$$

$$\lim_{n\to\infty} \mathbf{E}[d(\underline{S}_n, g_n(f_n(\underline{S}_n)))] \le D,$$

其中 M_n 表示 (f_n, g_n) 的指标 (indices) 的个数. 又因为 $D < \infty$, 所以存在正整数 n_1 使对每个正整数 $n > n_1$ 有 $\mathbf{E}[d(\underline{S}_n, g_n(f_n(\underline{S}_n)))] < \infty$. 如果 n 是正整数, $\underline{S}_n \neq g_n(f_n(\underline{S}_n))$ 的概率大于 0, 则 $\mathbf{E}[d(\underline{S}_n, g_n(f_n(\underline{S}_n)))] = \infty$. 所以对每个正整数 $n > n_1$, $\underline{S}_n = g_n(f_n(\underline{S}_n))$ 以概率 1 成立. 对每个离散随机变量 Z 用 $\alpha[Z]$ 表示 $\{z|P[Z = z] > 0\}$ 元素的个数. 对每个正整数 $n > n_1$, $g_n(f_n(\underline{S}_n))$ 只可能取 $g_n(1)$, $g_n(2)$, $\cdots$ 或 $g_n(M_n)$, 所以 $\alpha[g_n(f_n(\underline{S}_n))] \leq M_n$. 这样对每个正整数 $n > n_1$ 有 $M_n \geq \alpha[\underline{S}_n] = 2^n$, $\log_2(M_n)/n \geq 1$. 结合 24 式得 $R \geq 1$.

对每个 $D \geq 0$, 作为使 (R, D) 可达的所有 R 组成集合的下确界, $R(D) \geq 1$, 即 $R(D) \geq \log_2(2)$. □

注意在解 (c) 问时没有指定具体信源. 即使信源熵很小, 比如信源为 $Bernoulli(0.01)$ 分布, $R(D)$ 还是大于等于 $\log_2(2)$. 这表明失真度量有界对我们课上讲的结论而言是一个很重要的条件. 当然, 率失真函数等于 $\log_2(|\mathcal{S}|)$ 的例子还有很多. 比如在讲义 Bernoulli 信源的例子中, 如果 $\delta = 1/2$ 则 $R(0) = h_2(1/2) - h_2(0) = 1 = \log_2(2)$.

d) 解: 题意的失真矩阵为:

$$\mathbf{D} = \begin{bmatrix} 0 & \infty & 1 \\ \infty & 0 & 1 \end{bmatrix}$$

可以求得 $D_{\min} = 0$, 又

$$D_{\max} = \min_{\hat{s} \in \{0,1,e\}} \sum_{s=0}^{1} P_S(s) d(s, \hat{s}) = \min(\infty, \infty, 1) = 1,$$

所以对每个 $D \in [1, \infty)$ 有 $R(D) = 0$.

设 $0 \leq D < 1$, $\mathbf{E}[d(S, \hat{S})] \leq D$. 因为 $\mathbf{E}[d(S, \hat{S})] < \infty$, 所以 $P_{S,\hat{S}}(0, 1) = 0$, $P_{S,\hat{S}}(1, 0) = 0$. 如果 $P_{\hat{S}}(0) > 0$ 则 $P_{S|\hat{S}}(0|0) = 1$, $H(S|\hat{S} = 0) = 0$. 如果 $P_{\hat{S}}(1) > 0$ 则 $P_{S|\hat{S}}(1|1) = 1$, $H(S|\hat{S} = 1) = 0$. 如果 $P_{\hat{S}}(e) > 0$ 则 $H(S|\hat{S} = e) \leq \log_2(2) = 1$. 这样

$$\begin{aligned} H(S|\hat{S}) &= \sum_{\hat{s}, P_{\hat{S}}(\hat{s}) > 0} P_{\hat{S}}(\hat{s}) H(S|\hat{S} = \hat{s}) \\ &= P_{\hat{S}}(0) H(S|\hat{S} = 0) + P_{\hat{S}}(1) H(S|\hat{S} = 1) + P_{\hat{S}}(e) H(S|\hat{S} = e) \\ &\leq P_{\hat{S}}(e). \end{aligned}$$

又因为

$$\begin{gathered} \mathbf{E}[d(S, \hat{S})] = 1 P_{S,\hat{S}}(0, e) + 1 P_{S,\hat{S}}(1, e) = P_{\hat{S}}(e) \leq D, \\ I(S; \hat{S}) = H(S) - H(S|\hat{S}) = 1 - H(S|\hat{S}) \geq 1 - P_{\hat{S}}(e). \end{gathered}$$

所以 $I(S; \hat{S}) \geq 1 - D$.

如果 $0 \le D < 1$, 对所有 $s \in \{0,1\}$ 和 $\hat{s} \in \{0,1,e\}$, 如果有以下转移概率矩阵

$$P_{\hat{S}|S} = \begin{bmatrix} 1-D & 0 & D \\ 0 & 1-D & D \end{bmatrix},$$

则有 $\mathbf{E}[d(S,\hat{S})] = D$, $H(S|\hat{S}) = P_{\hat{S}}(e) = D$, $I(S;\hat{S}) = 1-D$. 对每个 $D \in [0,1)$, 作为 $\mathbf{E}[d(S,\hat{S})] \le D$ 的条件下 $I(S;\hat{S})$ 的最小值, 即有 $R(D) = 1-D$.

综上所述,

$$R(D) = \begin{cases} 1-D, & 0 \le D < 1 \\ 0, & D \ge 1 \end{cases}. \qquad \square$$