信息论 5 月 26 日第三次小测解答

中国科学技术大学《信息论 A》006125.01 班助教组

2025 年 5 月 27 日

第 1 题

A length-n binary repetition code has only two codewords: all 0_s *and* 1_s*. What is the code rate of this code? Find out its generator matrix and parity check matrix.*

解: 由于只有两个码字，即 $M=2$，故码率：

$$R=\frac{\log_2 M}{n}=\frac{1}{n}.$$

重复码是将二进制消息 0 或 1 重复 n 次，则生成矩阵：

$$\boldsymbol{G}=\begin{bmatrix}1 & 1 & 1 & \cdots & 1\end{bmatrix}_{1\times n}.$$

我们可以将生成矩阵直接视为系统型，第一个元素视为一个 1×1 的单位阵，即

$$\boldsymbol{G}=\begin{bmatrix}\boldsymbol{I}_{1\times 1} \mid \boldsymbol{P}_{1\times(n-1)}\end{bmatrix}_{1\times n},$$

其中 $\boldsymbol{P}=\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}_{1\times(n-1)}$.

那么对应的校验矩阵为：

$$\begin{aligned}\boldsymbol{H}&=\left[\boldsymbol{P}^{\mathrm{T}}_{(n-1)\times 1} \mid \boldsymbol{I}_{(n-1)\times(n-1)}\right]_{(n-1)\times n}\\&=\begin{bmatrix}1 & 1 & 0 & \cdots & 0\\1 & 0 & 1 & \cdots & 0\\\vdots & \vdots & \vdots & \ddots & 0\\1 & 0 & 0 & \cdots & 1\end{bmatrix}_{(n-1)\times n}.\end{aligned}$$

（注：校验矩阵答案不唯一，也可以把生成矩阵中的最后一个元素视为单位阵，这样两个子矩阵顺序前后调换，则对应校验矩阵内的子矩阵也要前后调换顺序。） □

第 2 题

Verify that for BSC with crossover probability $\delta < \frac{1}{2}$, the syndrome decoding rule is exactly the ML decoding rule in Sec. 6.4.

证明: 设码长为 n，对于翻转概率 $\delta < 1/2$ 的 BSC 信道来说，ML 译码等价于：

$$\begin{aligned}\hat{\underline{x}} &= \arg\max_{\underline{x}\in\mathcal{C}} P_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) \\ &= \arg\max_{\underline{x}\in\mathcal{C}} \delta^{d(\underline{y},\underline{x})}(1-\delta)^{(n-d(\underline{y},\underline{x}))} \\ &= \arg\min_{\underline{x}\in\mathcal{C}} d(\underline{y},\underline{x});\end{aligned}$$

其中 $d(\underline{y},\underline{x})$ 表示 $\underline{y}$ 与 $\underline{x}$ 的码字距离。

根据伴随式译码规则, $\underline{s} = \boldsymbol{H}\underline{y}^{\mathrm{T}} = \boldsymbol{H}(\underline{x}^{\mathrm{T}} \oplus \underline{e}^{\mathrm{T}}) = \boldsymbol{H}\underline{e}^{\mathrm{T}}$，一个 $\underline{s}$ 会对应 $|\mathcal{C}|$ 组 $(\underline{x}, \underline{e})$，然后我们是去找该 $\underline{s}$ 对应的具有最小重量的错误模式 e，即 $\min_{\underline{x}\in\mathcal{C}} w(e) = \min_{\underline{x}\in\mathcal{C}} d(\underline{y},\underline{x})$，故伴随式译码与 ML 译码等价。 □

第 3 题

For the $(7,4,3)$ Hamming code applied to a BSC with crossover probability $\delta < \frac{1}{2}$, calculate the average decoding error probability of the syndrome decoding rule; – you only need to give the coefficients for δ^2 and δ^3 terms.

解: 由于 $(7,4,3)$ 汉明码最多只能纠正一位错误，因此使用伴随式译码的错误概率

$$P_e = 1 - P(0\text{ 位错误}) - P(1\text{ 位错误}),$$

其中，对于该翻转概率为 δ 的 BSC 信道来说，码长为 7 的传输码字发生 0 位错误的概率 $P(0\text{ 位错误}) = (1-\delta)^7$，发生 1 位错误的概率 $P(1\text{ 位错误}) = \binom{7}{1}\delta(1-\delta)^6 = 7\delta(1-\delta)^6$，因此，译码错误概率

$$P_e = 1 - (1-\delta)^7 - 7\delta(1-\delta)^6.$$

进而，P_e 中 δ^2 和 δ^3 的系数为:

- δ^2 的系数：$-\binom{7}{2} + 7 \times \binom{6}{1} = -21 + 7 \times 6 = 21$;
- δ^3 的系数：$\binom{7}{3} - 7 \times \binom{6}{2} = 35 - 7 \times 15 = -70$.

（注：译码错误概率也可以写成 $P_e = P(2\text{ 位错误}) + P(3\text{ 位错误}) + \cdots + P(7\text{ 位错误}) = \binom{7}{2}\delta^2(1-\delta)^5 + \binom{7}{3}\delta^3(1-\delta)^3 + \cdots$，同样可以得出 δ^2 的系数：$\binom{7}{2} = 21$；δ^3 的系数：$-\binom{7}{2}\binom{5}{1} + \binom{7}{3} = -21 \times 5 + 35 = -70$.）

□

第 4 题

Apply the $(7,4,3)$ *Hamming code as a lossy source code for a Bernoulli*$(\frac{1}{2})$ *source with source block length* $n = 7$*. Specifically, for any length-7 source sequence* $(S_1, S_2, \ldots, S_7)$*, find the codeword with the smallest Hamming distance to it, as the reproduction. Calculate the expected Hamming distortion of this code, and compare it with the rate-distortion function.*

解: 首先，信源服从 $Bernoulli(\frac{1}{2})$ 分布，则长度为 7 的发送信源序列 $(S_1, S_2, \ldots, S_7) \in \mathbb{F}_2^7$，共有 $2^7 = 128$ 种等概可能。

使用 $(7,4,3)$ 汉明码作为有损信源编码，则重建序列 $(\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_7) \in \mathcal{C}_{(7,4,3)}$，共有 16 种重建码字。

容易知道如果发送的信源序列恰好就是这 16 种重建码字，则汉明失真为 0，如果发送的是其它，则选择与之汉明距离最小的码字作为重建，根据 $(7,4,3)$ 汉明码是完备码 * 可知，一定存在失真为 1 的重建码字，因此平均失真：

$$\mathbf{E}[d(\underline{S}, \underline{\hat{S}})] = (\frac{16}{128} \times 0 + \frac{112}{128} \times 1) \times \frac{1}{7} = \frac{1}{8}.$$

码率为 $R = \frac{\log_2 16}{n} = \frac{4}{7}$ (bits/source symbol)，根据 $Bernoulli(\frac{1}{2})$ 信源的率失真函数 $R(D) = h_2(1/2) - h_2(D) = 1 - h_2(D) = 1 - (-D\log_2 D - (1-D)\log_2(1-D))$; $0 \le D \le \frac{1}{2}$，得到失真：

$$D = h_2^{-1}(3/7) \approx 0.0877 < \frac{1}{8},$$

该结果体现了在相同码率下，$(7,4,3)$ 汉明码与无限码长编码方法之间的性能差距。

（* 注：完备码 (Perfect Code)：如果任意可能的 n 长 q 进制向量 $\underline{w_0} \in \mathbb{F}_q^n$，存在唯一的一个有效码字 $\underline{w} \in \mathcal{C}$ 最多有 e 个位置与 $\underline{w_0}$ 不同，其中 $e = (d_{min}-1)/2$，则称 $\mathcal{C}$ 是完备码。） □

第 5 题

Apply the $(7,4,3)$ *Hamming code to a BEC with erasure probability* α*. How should we decode it? How many erasures can this code tolerate?*

解: 如果经过 BEC 信道，那么接收到的 $\underline{y}$ 中有些位置就可能会包含 e，即被擦除，这些位置已知，且其他位置均被正确传输。

因此译码规则设计如下：将被擦除位置的 e 依次替换成 0/1，例如：

- 如果有 1 个 e：对应位置依次替换为 0 和 1，得到 $\mathcal{Y}' = \{\underline{y_0}', \underline{y_1}'\}$;
- 如果有 2 个 e：对应位置依次替换为 $00, 01, 10, 11$，得到 $\mathcal{Y}' = \{\underline{y_{00}}', \underline{y_{01}}', \underline{y_{10}}', \underline{y_{11}}'\}$;
- 如果有 3 个 e：对应位置依次替换为 $000, 001, 010, 011, 100, 101, 110, 111$，得到 $\mathcal{Y}' = \{\underline{y_{000}}', \underline{y_{001}}', \underline{y_{010}}', \underline{y_{011}}', \underline{y_{100}}', \underline{y_{101}}', \underline{y_{110}}', \underline{y_{111}}'\}$;

- $\cdots$

然后计算伴随式 $\underline{S}=\{\boldsymbol{H}\underline{y}'^{\mathrm{T}},\underline{y}'\in\mathcal{Y}'\}$，选择其中 $\underline{s}=000$ 对应的 $\underline{y}'$ 作为译码结果。

下面说明想要保证译码正确的充分条件是：e 的个数 $n_e\leq 2$.

首先，$(7,4,3)$ 汉明码的最小码距 $d_{min}=3$，对于任一有效码字，与之最近的其它有效码字距离为 3，即

$$\min_{\underline{c}\in\mathcal{C}_{(7,4,3)},\underline{c}\neq\underline{c_0}} d(\underline{c},\underline{c_0})=3,\forall\underline{c_0}\in\mathcal{C}_{(7,4,3)}.$$

其次，$\mathcal{Y}'$ 当中任一向量与其它向量的的最大距离等于 e 的个数，即

$$\max_{\underline{y}'\in\mathcal{Y}',\underline{y}'\neq\underline{y_0}'} d(\underline{y}',\underline{y_0}')=n_e,\forall\underline{y_0}'\in\mathcal{Y}'.$$

例如：当只有 1 个 e 时，$\underline{y_0}'$ 和 $\underline{y_1}'$ 的距离为 1；当有 2 个 e 时，$\underline{y_{00}}',\underline{y_{01}}',\underline{y_{10}}',\underline{y_{11}}'$ 当中任一向量与其它三个向量的距离最大为 2.

并且，$\mathcal{Y}'$ 当中一定存在发送码字 $\underline{x}$，即 $\underline{x}\in\mathcal{Y}'$.

因此，若 $n_e\leq 2$，则

$$\max_{\underline{y}'\in\mathcal{Y}',\underline{y}'\neq\underline{x}} d(\underline{y}',\underline{x})=n_e<3=\min_{\underline{c}\in\mathcal{C}_{(7,4,3)},\underline{c}\neq\underline{x}} d(\underline{c},\underline{x}),$$

进而 $\mathcal{Y}'$ 中除去发送码字 $\underline{x}$ 之外，一定不存在其它有效码字，即 $\underline{y}'\notin\mathcal{C}_{(7,4,3)},\forall\underline{y}'\in\mathcal{Y}'$ and $\underline{y}'\neq\underline{x}$，此时计算的伴随式 $\underline{S}$ 当中只存在一个 $\underline{s}=000$，对应的就是发送码字 $\underline{x}$，可以译码正确；而当 e 的个数 ≥ 3 时，$\underline{y}'$ 当中有可能出现不止一个有效码字，计算出来的 $\underline{s}$ 就不止一个 000，不能保证译码正确。

（需要说明的是：$n_e\leq 2$ 只是保证译码正确的充分条件，并非充要条件。因为当 $n_e\geq 3$ 时，如果被擦除是某些特定位置，也是可以译码正确的。

以 $n_e=3$ 为例，不妨假设发送码字 $\underline{x}=0000000$，如果是后三位被擦除，接收到 $\underline{y}=0000eee$，则 $\mathcal{Y}'=\{0000000,\ldots,0000111\}$，我们不难发现 $\mathcal{Y}'$ 当中只有一个有效码字，此时计算的伴随式 $\underline{S}$ 也就只有一个 000，可以保证译码正确。这是因为码本 $\mathcal{C}_{(7,4,3)}$ 当中与 0000000 距离为 3 的有效码字没有 0000111，事实上与 0000000 距离为 3 的有效码字仅有 7 个。所以当某些特定位置被擦除时也可以保证译码正确。）

$\square$