信息论与编码

马啸
maxiao@mail.sysu.edu.cn

计算机学院
中山大学

2021 年春季学期

引例一

- 信源：独立均匀硬币序列，每秒钟产生一个比特；
- 信道：离散无记忆的二元对称信道(Binary Symmetric Channel)，$p = 0.001$；
- 信道编码: “有效”、“可靠"地传输信源产生的序列。

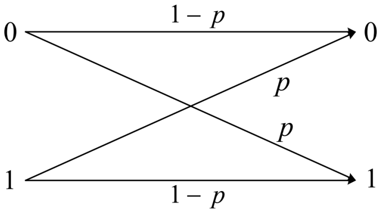


Figure: 二元对称信道(BSC)

我们考虑以下几种情形：

①信道每秒钟只能使用1次。

引例一

- 信源：独立均匀硬币序列，每秒钟产生一个比特；
- 信道：离散无记忆的二元对称信道(BSC)，$p = 0.001$；
- 信道编码: “有效”、“可靠"地传输信源产生的序列。

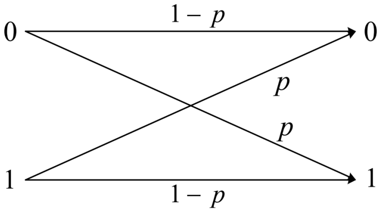


Figure: 二元对称信道(BSC)

我们考虑以下几种情形：

①信道每秒钟只能使用1次。此时没有任何余地。

引例一

②信道每秒钟可使用“无限多次”(“带宽”不受限)。

- 编码(N 长重复码):

$$\begin{array}{ccc} u=0 & \mapsto & \underline{x}=00\cdots 0 \\ u=1 & \mapsto & \underline{x}=11\cdots 1 \end{array}$$

- 传输: $\underline{x} \rightsquigarrow \underline{y}$
- 译码:

$$\hat{u}=\begin{cases} 0 & \underline{y}\in\mathcal{B}_0 \\ 1 & \underline{y}\in\mathcal{B}_1 \\ e & \underline{y}\in\mathcal{B}_2 \end{cases}$$

其中，$\mathcal{B}_i, i=0,1,2$ 构成一个接收空间的划分:

$$\begin{array}{l} \mathcal{B}_0=\{\text{ 所有Hamming重量小于N/2的接收序列}\} \\ \mathcal{B}_1=\{\text{ 所有Hamming重量大于N/2 的接收序列}\} \\ \mathcal{B}_2=\{\text{ 所有Hamming重量等于N/2的接收序列}\} \end{array}$$

- 错误概率: $P_e=\sum_{i\geq N/2}^{N}\begin{pmatrix} n \\ i \end{pmatrix} p^i(1-p)^{N-i}\to 0 \text{ as } N\to\infty.$
- 传输效率：$1/N\to 0$ as $N\to\infty$.

引例一

③信道每秒钟可使用两次。
方案1：重复2次.

- 编码: $0 \to 00, 1 \to 11$.
- 译码划分:

$$\mathcal{B}_0 = \{00\}$$
$$\mathcal{B}_1 = \{11\}$$
$$\mathcal{B}_2 = \{10, 01\}$$

- 正确概率:

$$P_c = (1-p)^2 = 0.998$$

- 传输效率: $1/2 = 0.5$.

引例一

③信道每秒钟可使用两次。
方案2：2个比特一起编码。

- 编码: $00 \to 0000, 01 \to 0111, 10 \to 1001, 11 \to 1110.$
- 译码划分:

$$\begin{aligned}\mathcal{B}_0 &= \{0000, 0010, 0100\} \\ \mathcal{B}_1 &= \{0011, 0101, 0110, 0111\} \\ \mathcal{B}_2 &= \{1001, 1011, 1101\} \\ \mathcal{B}_3 &= \{1010, 1100, 1110\} \\ \mathcal{B}_4 &= \{0001, 1000, 1111\}\end{aligned}$$

- 正确概率:

$$\begin{aligned}P_c = \frac{1}{4}\Big(&2(1-p)^3p + (1-p)^4 + 3(1-p)^3p + (1-p)^4 \\ &+2(1-p)^3p + (1-p)^4 + 2(1-p)^3p + (1-p)^4\Big) \\ &= 0.9982 > (1-p)^2 = 0.998\end{aligned}$$

- 传输效率: $2/4 = 0.5.$

引例一

③信道每秒钟可使用两次。

方案3：汉明码

信息：

$$u_0\ \ u_1\ \ u_2\ \ u_3$$

编码：

$$\begin{array}{lllll} & u_0 & u_1 & u_2 & u_3 \\ u_4=u_1+u_2+u_3 & \surd & \times & \times & \times \\ u_5=u_0+u_2+u_3 & \times & \surd & \times & \times \\ u_6=u_0+u_1+u_3 & \times & \times & \surd & \times \end{array}$$

码字：

$$u_0\ \ u_1\ \ u_2\ \ u_3\ \ u_4\ \ u_5\ \ u_6$$

接收：

$$c_0\ \ c_1\ \ c_2\ \ c_3\ \ c_4\ \ c_5\ \ c_6$$

正确概率：

$$P_c=(1-p)^7+\binom{7}{1}(1-p)^6p=0.999$$

传输效率：$4/7\approx 0.5714$ 如果没有错误，则三个方程都成立；
如果只有一个错误(共7种情况)，则至少有一个方程不成立。非常“神奇”（巧妙），一种错误模式，对应一个方程成立/不成立的模式。

引例二

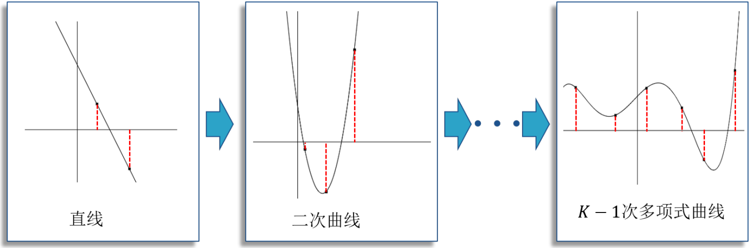
重复码也可以用来纠删

$$
\begin{aligned}
\text{学习进步} &\to \text{学？？？} \\
\text{学习进步} &\to \text{学习？？} \\
\text{学习进步} &\to \text{学习进？}
\end{aligned}
$$

引例二

纠删码构造

$y = a_0 + a_1 x$ $\quad$ $y = a_0 + a_1 x + a_2 x^2$ $\quad$ $y = a_0 + a_1 x + \cdots + a_{K-1} x^{K-1}$



引例二

4点确定一条3次多项式曲线

$$\text{学习进步} \rightarrow a_0 \ a_1 \ a_2 \ a_3 \rightarrow y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

从曲线上取12个点：$P_1 \ P_2 \ P_3 \ \cdots \ P_{12}$

广播这些点，任何人收到其中4个点，都可以恢复$a_0 \ a_1 \ a_2 \ a_3$。

引例二

3点确定一条2次多项式曲线

$$a_0\ a_1\ a_2 \to y = a_0 + a_1 x + a_2 x^2$$

从曲线上取n个点：$P_1\ P_2\ P_3\ \cdots\ P_n$

广播这些点，任何人收到其中3个点，都可以恢复$a_0\ a_1\ a_2$。

引例二

秘密共享

$$\text{秘密：}\quad a_0\ a_1\ a_2\ a_3 \rightarrow y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

从曲线上取n个点：$P_1\ P_2\ P_3\ \cdots\ P_n$，把这些点分给一群人。

少于4个点是没有办法解密的；而持有任何4个或以上的点的人群，都可以解密$a_0\ a_1\ a_2\ a_3$。

信道

信道

一个信道由输入集$\mathcal{X}$、输出集$\mathcal{Y}$以及信道概率转移函数

$$P(y^n \mid x^n), \quad x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$$

来刻画。

1. 给定x^n，　$P(* \mid x^n)$ 是$\mathcal{Y}^n$ 上的概率质量函数;
2. 给定x^n ，可能有许多序列y^n 与之“对应”；
3. 若重复发送一个给定的x^n 多次，在接收空间会形成一个“云团”。

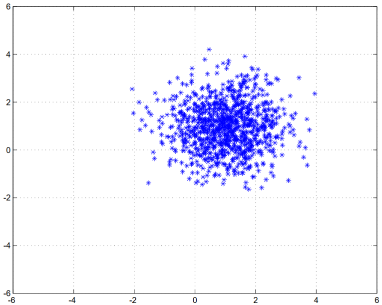


Figure: 在一个AWGN上多次发送(1，1)点形成的接收“云团”

信道

通常地，信道编译码方案（或算法）可以描述如下：

1. 消息集: $\mathcal{U} = \{1, 2, \cdots, M_n\}$。消息变量 U假设是 $\mathcal{U}$上的均匀随机变量。由此，相当于$\log M_n$ 比特。
2. 码书: $\mathcal{C} = \{x^n(1), x^n(2), \cdots, x^n(M_n)\} \subset \mathcal{X}^n$
3. 译码划分：把$\mathcal{Y}^n$ 划分成$M_n + 1$ 个不相交的区域$\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{M_n}$;
4. 编码: 若发送$u = i$ 时，发送码字$x^n(i)$;
5. 传输：$x^n \rightsquigarrow y^n$;
6. 译码：若接收向量y^n 落入$\mathcal{B}_i$, 则译成 $\hat{u} = i$。显然落入$\mathcal{B}_0$ 时，则译码一定出错;
7. 码率: $R = \frac{\log M_n}{n}$;
8. 错误率: $P(\hat{U} \neq U)$ 。

码字个数要多，错误率要低

信道编码定理

信道编码定理

Definition 1

二维离散随机变量(X, Y)的互信息定义为

$$I(X;Y) = E\left(\log\frac{P(Y\mid X)}{P(Y)}\right) = \sum_{x,y} P(x,y)\log\frac{P(y\mid x)}{P(y)}$$

Theorem 2 (信道编码定理)

离散无记忆信道$(\mathcal{X}, \mathcal{Y}, P(y\mid x))$ 的信道容量是$C = \max_{P(x)} I(X;Y)$。就是说，若$R < C$, 则存在编译码方案，使得误码率趋于0; 若$R > C$, 则不可能。

常见信道模型与信道容量

Noiseless Binary Channel

In this case, any transmitted bit is received without error.

Figure: Noiseless binary channel. $C = 1$ bit.

Hence, one error-free bit can be transmitted per use of the channel, and the capacity is 1 bit. We can also calculate the information capacity $C = \max I(X;Y) = 1$ bit, which is achieved by using $p(x) = (\frac{1}{2}, \frac{1}{2})$.

常见信道模型与信道容量

Noisy Channel with Nonoverlapping Outputs

This channel has two possible outputs corresponding to each of the two inputs. The channel appears to be noisy, but really is not. Even though the output of the channel is a random consequence of the input, the input can be determined from the output, and hence every transmitted bit can be recovered without error.

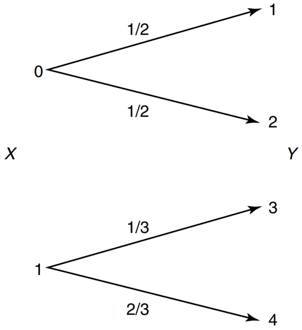


Figure: Noisy channel with nonoverlapping outputs. $C = 1$ bit.

常见信道模型与信道容量

Noisy Channel with Nonoverlapping Outputs

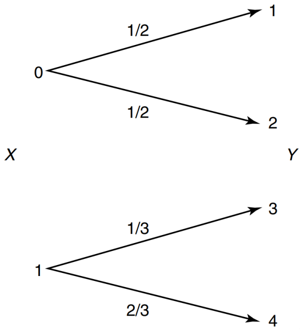


Figure: Noisy channel with nonoverlapping outputs. C = 1 bit.

The capacity of this channel is also 1 bit per transmission. We can also calculate the information capacity $C = \max I(X;Y) = 1$ bit, which is achieved by using $p(x) = (\frac{1}{2}, \frac{1}{2})$.

常见信道模型与信道容量

Noisy Typewriter

In this case the channel input is either received unchanged at the output with probability $\frac{1}{2}$ or is transformed into the next letter with probability $\frac{1}{2}$. If the input has 26 symbols and we use every alternate input symbol, we can transmit one of 13 symbols without error with each transmission.

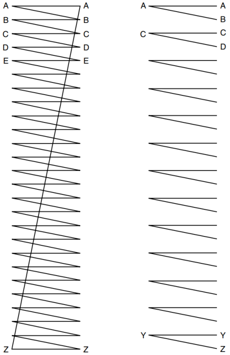


Figure: Noisy Typewriter. $C = \log 13$ bits.

常见信道模型与信道容量

Noisy Typewriter

Hence, the capacity of this channel is $\log 13$ bits per transmission. We can also calculate the information capacity

$$
\begin{aligned}
C &= \max I(X;Y) \\
&= \max(H(Y) - H(Y|X)) \\
&= \max H(Y) - 1 \\
&= \log 26 - 1 \\
&= \log 13
\end{aligned}
$$

achieved by using $p(x)$ distributed uniformly over all the inputs.

常见信道模型与信道容量

Binary Symmetric Channel

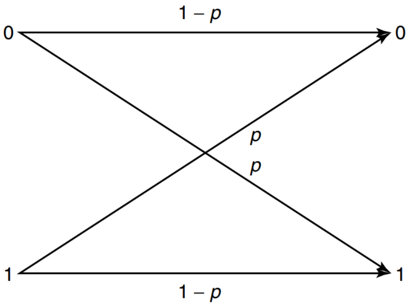


Figure: Binary symmetric channel. $C = 1 - H(p)$ bits.

This is a binary channel in which the input symbols are complemented with probability p. This is the simplest model of a channel with errors, yet it captures most of the complexity of the general problem.

常见信道模型与信道容量

Binary Symmetric Channel

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum p(x) H(Y|X=x) \\
&= H(Y) - \sum p(x) H(p) \\
&= H(Y) - H(p) \\
&\leq 1 - H(p)
\end{aligned}
$$

等号成立当且仅当 X 为均匀分布。所以，BSC信道的容量为

$$
C = 1 - H(p) \text{ bits}
$$

常见信道模型与信道容量

Binary Erasure Channel

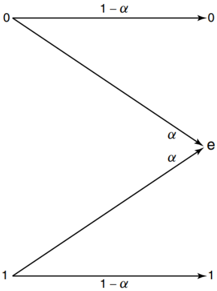


Figure: Binary erasure channel.

The analog of the binary symmetric channel in which some bits are lost (rather than corrupted) is the binary erasure channel. In this channel, a fraction α of the bits are erased. The receiver knows which bits have been erased. The binary erasure channel has two inputs and three outputs.

常见信道模型与信道容量

Binary Erasure Channel

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} (H(Y) - H(Y|X)) \\
&= \max_{p(x)} H(Y) - H(\alpha)
\end{aligned}
$$

The first guess for the maximum of $H(Y)$ would be $\log 3$, but we cannot achieve this by any choice of input distribution $p(x)$. Letting E be the event $\{Y = e\}$, using the expansion

$$H(Y) = H(Y, E) = H(E) + H(Y|E)$$

and letting $\Pr(X = 1) = \pi$, we have

$$H(Y) = H((1-\pi)(1-\alpha), \alpha, \pi(1-\alpha)) = H(\alpha) + (1-\alpha)H(\pi)$$

常见信道模型与信道容量

Binary Erasure Channel

Hence

$$
\begin{aligned}
C &= \max_{p(x)} H(Y) - H(\alpha) \\
&= \max_{\pi} (1-\alpha)H(\pi) + H(\alpha) - H(\alpha) \\
&= \max_{\pi} (1-\alpha)H(\pi) \\
&= 1-\alpha
\end{aligned}
$$

where capacity is achieved by $\pi = \frac{1}{2}$.

常见信道模型与信道容量

Symmetric Channel

Consider the channel with transition matrix:

$$p(y \mid x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Here the entry in the xth row and the yth column denotes the conditional probability $p(y|x)$ that y is received when x is sent. All the rows of the probability transition matrix are permutations of each other and so are the columns. Such a channel is said to be *symmetric*. Letting $\mathbf{r}$ be a row of the transition matrix, we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ &= H(Y) - H(\mathbf{r}) \\ &\leq \log|\mathcal{Y}| - H(\mathbf{r}) \end{aligned}$$

with equality if the output distribution is uniform.

常见信道模型与信道容量

Symmetric Channel

But $p(x) = 1/|\mathcal{X}|$ achieves a uniform distribution on Y, as seen from

$$p(y) = \sum_{x \in \mathcal{X}} p(y \mid x) p(x) = \frac{1}{|\mathcal{X}|} \sum p(y \mid x) = c \frac{1}{|\mathcal{X}|} = \frac{1}{|\mathcal{Y}|}$$

where c is the sum of the entries in one column of the probability transition matrix.

Thus, the channel has the capacity

$$C = \max_{p(x)} I(X;Y) = \log 3 - H(0.5, 0.3, 0.2)$$

and C is achieved by a uniform distribution on the input. The transition matrix of the symmetric channel defined above is doubly stochastic.

常见信道模型与信道容量

Symmetric Channel

In the computation of the capacity, we used the facts that the rows were permutations of one another and that all the column sums were equal. Considering these properties, we can define a generalization of the concept of a symmetric channel as follows:

Definition 3

A channel is said to be *symmetric* if the rows of the channel transition matrix $p(y|x)$ are permutations of each other and the columns are permutations of each other. A channel is said to be *weakly symmetric* if every row of the transition matrix $p(\cdot|x)$ is a permutation of every other row and all the column sums $\sum_x p(y|x)$ are equal.

常见信道模型与信道容量

Symmetric Channel

For example, the channel with transition matrix

$$p(y \mid x) = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}$$

is weakly symmetric but not symmetric.

The above derivation for symmetric channels carries over to weakly symmetric channels as well. We have the following theorem for weakly symmetric channels:

Theorem 4

For a weakly symmetric channel,

$$C = \log |\mathcal{Y}| - H(\textit{ row of transition matrix })$$

and this is achieved by a uniform distribution on the input alphabet.

常见信道模型与信道容量

下面定理给出$\{Q_k\}$达到DMC信道容量的充要条件。

Theorem 5

输入概率矢量$\mathbf{Q} = \{Q_1, Q_2, \cdots, Q_{K-1}\}$达到转移概率为$P(j|k)$的$DMC$的容量$C$的充要条件是

$$\begin{aligned} I(x=k;Y) &= C, \text{对所有} k, Q_k > 0 \\ I(x=k;Y) &\le C, \text{对所有} k, Q_k = 0 \end{aligned} \quad (1)$$

其中$I(x=k;Y)$是信道输入$x=k$时，信道输入出一个字母的平均互信息，即

$$I(x=k;Y) = \sum_j P(j|k) \log \frac{P(j|k)}{\sum_i Q_i P(j|i)}$$

常见信道模型与信道容量

证明: 由$I(X;Y)$是信道输入分布$\{Q_k\}$的concave函数。C 是函数$I(X;Y)$对所有可能分布求的极值。由K-T条件知，输入分布为最佳分布的充要条件是$I(X;Y)$ 对分布$\{Q_k\}$满足

$$\begin{aligned} \frac{\partial I(X;Y)}{\partial Q_k} &= \lambda, Q_k > 0 \\ \frac{\partial I(X;Y)}{\partial Q_k} &\le \lambda, Q_k = 0 \end{aligned} \tag{2}$$

其中λ是拉格朗日乘子，它由条件$\sum_{k=0}^{K-1} Q_k = 1$确定。因为

常见信道模型与信道容量

$$
\begin{aligned}
\frac{\partial I(X;Y)}{\partial Q_k} =& \frac{\partial}{\partial Q_k}\sum_j\sum_m Q_m P(j|m)\log\frac{P(j|m)}{\sum_i Q_i P(j|i)} \\
=& \sum_j\left[P(j|k)\log\frac{P(j|k)}{\sum_i Q_i P(j|i)} - (\log e)\sum_m Q_k P(j|m)\frac{P(j|k)}{\sum_i Q_i P(j|i)}\right] \\
=& \sum_j P(j|k)\log\frac{P(j|k)}{\sum_i Q_i P(j|i)} - (\log e)\sum_j P(j|k) \\
=& I(x=k;Y) - \log e
\end{aligned} \tag{3}
$$

将(3)代入(2),并令$C=\lambda+\log e$就得到(1)。以Q_k乘式(1)两边，并对所有$k\in X$求和，就可得到给定信道再求得的分布Q_k下，信道输入和输出之间的信息量

$$I(X;Y)=C$$

作业

Exercise 1.[王育民(2013)]

计算由下述转移概率矩阵给定的DMC的容量。

(a)

$$\begin{bmatrix} \text{1-P} & \text{P} & 0 \\ 0 & \text{1-P} & \text{P} \\ \text{P} & 0 & \text{1-P} \end{bmatrix}$$

(b)

$$\begin{bmatrix} \frac{1-P}{2} & \frac{1-P}{2} & \frac{P}{2} & \frac{P}{2} \\ \frac{P}{2} & \frac{P}{2} & \frac{1-P}{2} & \frac{1-P}{2} \end{bmatrix}$$

(c)

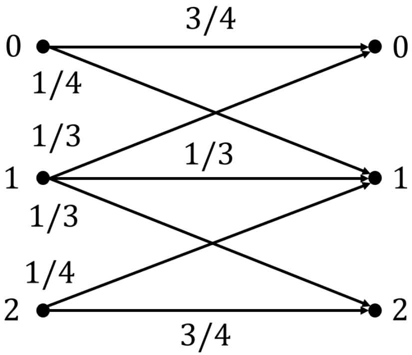
$$\begin{bmatrix} \text{1-P} & \text{P} & 0 \\ \text{P} & \text{1-P} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

作业

Exercise 2.[王育民(2013)]

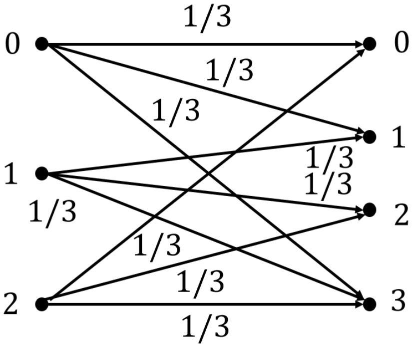
计算图中DMC的容量及最佳输入分布。

(a)



作业

(b)



谢谢！