XI'AN JIAOTONG UNIVERSITY

信息论与编码

第2章 信源

张建国

XI'AN JIAOTONG UNIVERSITY

主要内容与基本要求

- **主要内容**
  - 消息的数值编码；
  - 信源的种类及其信息率；
  - 各种信息熵。
- **基本要求**
  - 了解信源的数值编码过程；
  - 了解信源的种类；
  - 理解和掌握信息熵、信息率的概念和计算方法；
  - 理解连续无记忆信源的相对熵、绝对熵；
  - 理解离散记忆信源的条件熵。

XI'AN JIAOTONG UNIVERSITY

本章目录

XI'AN JIAOTONG UNIVERSITY

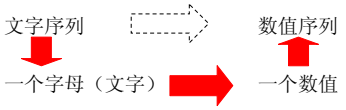
2.1 消息的数值编码

把不同的消息转换成统一的数值（二元或多元）序列。

消息——文字、声音、图像——数值序列。

消息是知识在传递过程中的表现形式，可以是文字的、声音的或图像的。信息论要研究的是消息在传输过程中的特性的度量，而不关心其具体形式和手段，通过数值编码可以将不同的消息形式转换成统一的形式，这有助于我们对信息传输过程的分析和讨论。我们要关注的是数值序列的统计特性。

2.1.1 文字消息



2.1 消息的数值编码

XI'AN JIAOTONG UNIVERSITY

一些文字消息的例子。

- 英文(ASCII)：26个大小写字母，数字0－9，标点符号，控制符等，共128个。每个符号用一个字节表示，最高位为0。
- 中文(GB2312-80)：二级汉字，共6763个，至少13位，用两个字节表示，两字节的最高位为1。（2^{12}=4096 < 6763 < 2^{13}=8192）
- 中文(GBK)：GB2312+BIG5，大于20000个，占用了15位，用两个字节表示，第1字节的最高位为1，第2个字节没有要求。（2^{15}=32768）
- 还有其它的一些编码方式，如UCS-2（16位），UCS-4（32位）等等。转换（存储、传输）机制有UTF-8，UTF-16，UTF-32等。

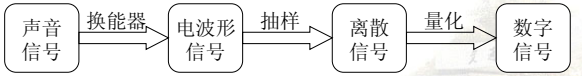
biáng

2.1 消息的数值编码

XI'AN JIAOTONG UNIVERSITY

2.1.2 声音信号

- 声音信号包括语音（Speech）、音乐（Music）。
- 电波形信号。换能器把声音信号转换成时间幅度均连续的电波形信号。
- 离散信号。对电波形信号抽样后形成时间离散信号。只要满足Nyquist采样定理，这一过程就是可恢复的。
- 数字信号。通过模数变换器对离散信号进行幅度上的量化后得到时间幅度均离散的数字信号。**量化过程是不可恢复的。**



2.1 消息的数值编码

XI'AN JIAOTONG UNIVERSITY

经过上述过程，我们将时间、幅度都连续的声音信号转换成了数值的序列。注意重要的两步：时间上的抽样，幅度上的量化。其中，抽样过程是可以恢复的，但量化过程是不可以恢复的，存在量化误差。

当信号幅度一定时，量化噪声随量化级数的增大而减小，所以为了减小量化误差，应取足够大的量化级数。但是级数的增大也意味着增加信源信息量，这要耗费信道容量。因而应在量化级数与接收消息的质量间采取合理的折中。

量化信号是不精确的连续信号，也可以说是足够精确的连续信号。

两个例子：

语音（电话）：3.4kHz，8k sample/s，分256级(2^8)，64kbps。

音乐：20Hz~20kHz，44k sample/s，分4096级(2^{12})，528kbps

2.1 消息的数值编码

XI'AN JIAOTONG UNIVERSITY

2.1.3 图像信号

- 电视
  - 二维 （扫描）
  - 运动的图像 （视觉暂存现象）
  - 彩电的制式:NTSC(美，日)，PAL(中，英，德)，SECAM（法）
  - 数据率：PAL制，带宽5.5M，采样后11M，量化后88M，实际上速率在110M左右。
- 计算机上的视频格式
  - MPEG1 VCD 1.5Mbps 352×288 1.5×60×45/8=506.25MB
  - MPEG2 DVD 12Mbps 720×480/640×480 4050MB/45min
  - RMVB 225K，350K，450K bps

2.1 消息的数值编码

XI'AN JIAOTONG UNIVERSITY

2.1.4 二进制编码

将不同种类信源所产生的消息转换成统一的数值序列，简化了对信息传输过程的分析和讨论。为了进一步简化，还可以将任一多进制数值序列归一化为基本的二进制序列。

每一个M进制数值 ⇨ 若干个二进制数值(N个)

M进制数值序列 ⇨ 二进制数值序列

N和M的关系 $N=\lceil \log_2 M \rceil$ 。

序列长度的关系：$L_2 = NL_M$ 。

$$x(t) \xrightarrow{\text{采样}} \sum_k x(kT_s)\delta(t-kT_s) \xrightarrow{\text{量化}} \sum_k a_m\delta(t-kT_s) \xrightarrow{\text{二进制编码}} \sum_n b_n\delta(t-nT_b)$$

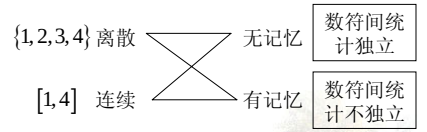
2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

- 从信源消息的取值集合角度看，可以分为**离散信源**和**连续信源**。
- 从信源消息的统计特性看，可以分为**无记忆**（Memoryless）信源、**有记忆**（Memory）信源（还有平稳信源、非平稳信源以及马尔科夫信源等）。无记忆是指发送的数符之间相互独立。

根据以上组合，信息论将要研究的信源分为以下四种。

- 离散无记忆信源
- 连续无记忆信源
- 离散有记忆信源
- 连续有记忆信源。



信源和信道在时间上都是离散的。

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

2.2.1 离散无记忆信源

Discrete Memoryless Source，DMS。

- 离散：数值的取值集合是离散的；
- 无记忆：发送的数符间相互统计独立。（放回与不放回的抽取）

离散无记忆信源信息熵的计算

设信源以速率 r产生数符 x，$x \in \{1,2,\cdots,M\}$，且有

$$P(x=i)=p_i,\ i=1,\cdots,M$$

则

$$H(X)=\sum_i p_i \log\frac{1}{p_i}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

信息率的定义

信源产生信息的速率，是信源在单位时间内发出的平均信息量。

信息率的计算

$$R=rH(X)$$

信息率是信息论中唯一一次出现具有时间意义的概念。以后的信道速率是指传一次需要传多少信息量。

2.2.2 连续无记忆信源

连续无记忆信源发出的每一数符的取值集合是连续的，而相继发出的数符间又是相互独立的。

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

关键是如何求**连续时的信息熵**。

连续取值要用到概率密度函数，具有如下性质：

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

解决思路：首先通过量化将连续型信源变换成离散型信源，用离散信源的分析方法进行分析；然后再将量化单位无限缩小，研究量化单位趋于0时离散信源熵的极限。

设信源以速率 r 产生数符 X，$X \in [0, A]$，且概率密度为 $p(x)$。

量化：将数符 X 的取值均匀量化为 M 级，量化台阶 $\Delta = A/M$ 。

得到量化取值集合

$$X_q = \left\{\frac{\Delta}{2}, \frac{3\Delta}{2}, \cdots (M - \frac{1}{2})\Delta\right\}$$

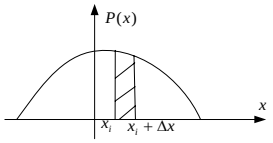
2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

经量化处理后，概率空间由概率密度 $p(x)$ 变为离散的概率值。每个样值发生的概率 $P(x_i)$ 为：

$$P(x_i) = p\left\{x_i - \frac{\Delta}{2} < x \le x_i + \frac{\Delta}{2}\right\}$$

$$= \int_{x_i - \frac{\Delta}{2}}^{x_i + \frac{\Delta}{2}} p(x)dx \approx p(x_i) \cdot \Delta$$



此时，每个样值的自信息量以及信源熵可以用离散信源的理论进行计算，所得离散信源的熵可近似作为此连续信源的熵。

$$H(X_q) = \sum_i P(x_i) \log \frac{1}{P(x_i)} \approx \sum_i p(x_i) \cdot \Delta \cdot \log \frac{1}{p(x_i) \cdot \Delta}$$

$$= \sum_i p(x_i) \cdot \Delta \cdot \log \frac{1}{p(x_i)} + \sum_i p(x_i) \cdot \Delta \cdot \log \frac{1}{\Delta}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

令量化级数无限大，即令量化台阶无穷小，当量化台阶趋于零时，离散信源就还原成连续信源。也就是说，如果令上式中的量化台阶趋于零，就可以得到连续信源的熵。

$$H_{abs}(X) = \lim_{\Delta \to 0} H(X_q) = \lim_{\Delta \to 0} \sum_i p(x_i) \log \frac{1}{p(x_i)} \Delta + \lim_{\Delta \to 0} \sum_i p(x_i) \Delta \cdot \log \frac{1}{\Delta}$$

$$= \int p(x) \log \frac{1}{p(x)} dx + \left[\int p(x)dx\right] \cdot \lim_{\Delta \to 0} \log \frac{1}{\Delta}$$

第一项由且仅由 $p(x)$ 决定，对给定信源为一确定值，称为**相对熵**，用 $H(X)$ 表示。 $H(X) \triangleq \int p(x) \log \frac{1}{p(x)} dx$

第二项，因为 $\Delta \to 0$，所以它趋于无限大。

故连续信源的**绝对熵**为：$H_{abs}(X) = H(X) + \infty$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

可见，连续变量所含的信息量是无穷大的。

简单理解，由于 $H(X) = E[I_x]$，连续消息每一样值**只有对应的概率密度**，其所占概率为0，根据自信息量的定义，连续消息每一样值的自信息量都是无限大，况且量化前样值集合的幅度连续，有无限多幅度值。

量化后，样值集合的幅度值为有限个，每个样值的概率也不再是零。所以，量化会损失信息量。

称其为绝对熵的原因：

- 当量化级数无穷大时，它趋于无限大；
- 连续信源的各种熵，包括条件熵、信宿熵、联合熵等，将都会有这一项，且都是当量化级数无穷大时其数值趋于无限大。

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

连续信源信息熵的特性：

- 两个信源的绝对熵不可比较。
- 无穷大无法传递，仅相对熵参加传递。

绝对熵虽然有明确的物理意义，但在分析信源的信息特性时并没有实际的意义。

相对熵不能代表连续信源的平均不确定性大小，因为连续信源的平均不确定性为无穷大，也不能代表连续信源输出的信息量。其意义在于：

- 这种定义可以与离散信源的熵在形式上统一起来，只不过是求和变为求积分，概率分布变为概率密度。
- 在讨论信息传输特性时，关心的是传输前后熵之间的差值（熵差）。相对熵在一定程度上可以用于度量连续信源的统计特性（混乱程度）。

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

相对熵仅与连续信源的概率密度有关，不同概率密度的信源具有不同的相对熵。因此**它表征了信源间平均信息量的差异，在一定程度上能够量度连续信源的信息特性（混乱程度）。**

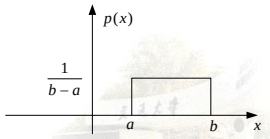
绝对熵非负，而相对熵可正可负。

$$H_{abs}(X)\geq 0 \qquad H(X)>,=,<0$$

例1、均匀分布情况

均匀分布的连续信源 X 的概率空间为：

$$[\mathbf{X}\cdot P]:\begin{cases}\mathbf{X}: & [a,b]\\ P(\mathbf{X}): p(x)=\begin{cases}\dfrac{1}{b-a}, & a\leq x\leq b\\ 0, & x>b, x<a\end{cases}\end{cases}$$



2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

可以算出，均匀分布连续信源的相对熵为：

$$H(X)=-\int_a^b p(x)\log p(x)\mathrm{d}x=\log(b-a)$$

这表明均匀分布的连续信源 X 的相对熵 $H(X)$，等于取值区间的上下限值之差 $(b-a)$ 的对数。

对于底大于1的对数，当 $(b-a)$ 小于1时，$H(X)<0$，说明均匀分布的连续信源 X 的相对熵将会出现负值。这说明**连续信源的相对熵不具有非负性**。

由于连续信源 X 的相对熵 $H(X)$ 只是绝对信息熵 $H_{abs}(X)$ 中的定值的部分，虽然 $H(X)$ 可能出现负值，但与无限大的绝对熵相加，仍为无限大的正数。

若连续无记忆信源每秒产生 r 个符号，则其信息率为 $R=rH(X)$。

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

2.2.3 离散记忆信源

记忆：相继数符间不独立。

记忆可减少信息量，即利用记忆可减少数符的不确定性。这说明记忆会带来冗余。

现有的自然信源大多是冗余的。（举例）

为了讨论信源的冗余，引入联合熵与条件熵。

联合熵（联合观察多个符号的平均信息量）

二个符号的联合熵可以用下式计算：

$$H(XY)=\sum_x\sum_y p(x,y)\log\frac{1}{p(x,y)}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

条件熵，观察记忆信源的另一种途径是通过条件熵。

离散记忆信源信息熵的计算

当信源发出的当前数符概率与之前的K个数符相关，则称信源具有K阶记忆。离散一阶记忆信源当前数符的条件信息熵：

$$H(X_n \mid x_{n-1}=j)=\sum_i p(x_n=i \mid x_{n-1}=j)\log\frac{1}{p(x_n=i \mid x_{n-1}=j)}$$

可以简写为：

$$H(X \mid x_j)=\sum_i p(x_i \mid x_j)\log\frac{1}{p(x_i \mid x_j)}$$

可以看到，这个熵是随机的，它随前一个符号的变化而变化。可以在此基础上进一步求平均：

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

$$H(X_n \mid X_{n-1})=\sum_j p(x_{n-1}=j)H(X_n \mid x_{n-1}=j)$$

$$=\sum_j p(x_{n-1}=j)\sum_i p(x_n=i \mid x_{n-1}=j)\log\frac{1}{p(x_n=i \mid x_{n-1}=j)}$$

$$=\sum_j\sum_i p(x_{n-1}=j)p(x_n=i \mid x_{n-1}=j)\log\frac{1}{p(x_n=i \mid x_{n-1}=j)}$$

$$=\sum_j\sum_i p(x_n=i,x_{n-1}=j)\log\frac{1}{p(x_n=i \mid x_{n-1}=j)}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

联合熵与条件熵之间的关系（以二维为例）

$$H(XY)=-\sum_x\sum_y p(x,y)\log p(x,y)$$

$$=-\sum_x\sum_y p(x)\,p(y \mid x)\log\left[p(x)\,p(y \mid x)\right]$$

$$=-\sum_x\sum_y p(x)\,p(y \mid x)\log p(x)-\sum_x\sum_y p(x)\,p(y \mid x)\log p(y \mid x)$$

$$=-\sum_x p(x)\log p(x)\sum_y p(y \mid x)-\sum_x\sum_y p(x,y)\log p(y \mid x)$$

$$=H(X)+H(Y \mid X)$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

例1：已知某单符号离散信源的概率空间为：

$\mathbf{X}$:	a_1,	a_2,	a_3
$P(\mathbf{X})$:	$\frac{11}{36}$	$\frac{4}{9}$	$\frac{1}{4}$

该信源发出的消息均为二重符号序列 $(a_i,a_j),(i,j=1,2,3)$，两个符号的关联性用条件概率 $P(a_j \mid a_i)$ 表示，如表2所示，求 $H(X^2)$。

表2 例2 给出的条件概率

a_i \ a_j	a_1	a_2	a_3
a_1	9/11	2/11	0
a_2	1/8	3/4	1/8
a_3	0	2/9	7/9

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

解：所求

$$H(X^2)=-\sum_{i=1}^{3}\sum_{j=1}^{3}P(a_i,a_j)\log P(a_i,a_j)$$

由$P(a_i,a_j)=P(a_j)P(a_i\mid a_j)$，可求出9个联合概率。

$$P(a_1,a_1)=P(a_1)P(a_1\mid a_1)=(11/36)\times(9/11)=1/4$$
$$P(a_1,a_2)=P(a_1)P(a_2\mid a_1)=(11/36)\times(2/11)=1/18$$
$$\vdots$$
$$P(a_3,a_3)=P(a_3)P(a_3\mid a_3)=(1/4)\times(7/9)=7/36$$

代入公式，可得：

$$H(X^2)=2.412\text{比特/符号序列}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

也可以通过原始信源熵和条件熵求扩展后的熵，有

$$H(X)=-\sum_{i=1}^{3}P(a_i)\log P(a_i)=1.542\ \text{比特/符号}$$

$$H(X_2\mid X_1)=-\sum_{i=1}^{3}\sum_{j=1}^{3}P(a_ia_j)\log P(a_j\mid a_i)=0.870\ \ \text{比特/符号}$$

$$H(X)+H(X_2\mid X_1)=2.412\text{比特/符号序列}$$

由此可见：

$$H(X^2)=H(X)+H(X_2\mid X_1)$$
$$H(X)\geq H(X_2\mid X_1)$$
$$H(X^2)\leq 2H(X)$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

例2：某箱子中有红色球和蓝色球各2个。在第k次抽取前把第k-2次抽出的球放回箱子，而第k-1次抽取的球留在外面。

求 $H(X)$ 、$H(X_k\mid x_{k-1}=red)$ 、$H(X_{k-1},X_k)$ 及 $H(X_k\mid X_{k-1})$

解：$\begin{bmatrix}X\\p(x)\end{bmatrix}=\begin{bmatrix}\text{red} & \text{blue}\\0.5 & 0.5\end{bmatrix}$　　$H(X)=\log 2$

$$H(X_k\mid x_{k-1}=red)=\frac{1}{3}\log 3+\frac{2}{3}\log\frac{3}{2}=0.92$$

$p(x_{k-1},x_k)$		X_k red	X_k blue
X_{k-1}	red	1/6	1/3
	blue	1/3	1/6

$H(X_{k-1},X_k)=1.92$

$H(X_k\mid X_{k-1})=0.92$

$p(x_k\mid x_{k-1})$		X_k red	X_k blue
X_{k-1}	red	1/3	2/3
	blue	2/3	1/3

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

上述结论说明符号间的关联性使信源输出的信息量减少。

根据熵的链接准则

- 对于一阶记忆信源，有 $H(X^2)=H(X)+H(X\mid X)$
- 对于K-1阶记忆信源，有

$$H(X^K)=H(X)+H(X\mid X)+\cdots+H(X\mid \underbrace{XX\cdots X}_{K-1\text{个}})$$

且 $H(X^K)\leq KH(X)$

信源的序列熵。若信源输出一个L长序列，则信源的序列熵为：

$$H(X^L)=H(X_1,X_2,\cdots,X_L)$$
$$=H(X_1)+H(X_2\mid X_1)+\cdots+H(X_L\mid X_1,X_2,\cdots,X_{L-1})$$

每个符号的平均熵为：$H_L(\mathbf{X})=H(X^L)/L$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

信源符号独立且等概时信息熵最大，记为 $H_0(X)=\log K$。

假设信源无记忆，但不等概，可计算单个符号的熵：$H_1(X)$。

若符号间有记忆，则通过计算联合熵，可得每个符号的平均熵。

当联合观察的符号序列长度趋于无穷时，有

$$H_\infty(X)=\lim_{L\to\infty}H_L(X)=\lim_{L\to\infty}\frac{1}{L}H(X_1,X_2,\cdots,X_L)$$

称为离散信源的极限熵。

信源符号之间的依赖关系越强，每个符号提供的平均信息量越小。常用信源的剩余度（冗余度）衡量信源的相关程度。信源剩余度定义为：

$$\gamma=\frac{H_0-H_\infty}{H_0}=1-\frac{H_\infty}{H_0}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

例如：英文26个字母加空格共27个符号，在完全等概且无记忆的情况下，其最大熵为：

$$H_{\max}(X)=\log_2 27=4.755\text{比特/字母}$$

然而，根据统计，这27个符号的出现概率如下页表1所示，由此可得实际熵（不考虑记忆）为：

$$H_1(X)=-0.1817\log_2 0.1817-0.1073\log_2 0.1073-\ldots$$
$$-0.00063\log_2 0.00063\approx 4\text{比特/字母}$$

如果考虑记忆，两个字母的组合。出现频度最高的20个：th, he, in, er, an, re, ed, on , es, st, en, at, to, nt, ha, nd, ou, ea, ng, as. 则可得：$H_2(X)\approx 3.3$比特/字母

根据有关研究：$H_3(X)\approx 3.1$比特/字母 … $H_\infty(X)\approx 1.4$比特/字母

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

表 1 英文 26 个字母和空格出现概率的一种统计结果

序号	字母	出现概率	序号	字母	出现概率
1	空格	0.1817	15	M	0.02105
2	E	0.1073	16	U	0.02010
3	T	0.0856	17	G	0.01633
4	A	0.0668	18	Y	0.01623
5	O	0.0654	19	P	0.01623
6	N	0.0581	20	W	0.01260
7	R	0.0559	21	B	0.01179
8	I	0.0510	22	V	0.00752
9	S	0.0499	23	K	0.00344
10	H	0.04305	24	X	0.00136
11	D	0.03100	25	J	0.00108
12	L	0.02775	26	Q	0.00099
13	F	0.02395	27	Z	0.00063
14	C	0.02260			

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

因此，该信源的冗余度为：

$$R=\frac{H_{\max}-H_\infty}{H_{\max}}\times 100\%=\frac{4.755-1.4}{4.755}\approx 71\%$$

可见，自然语言的冗余度是很大的，因为语言本身有很多固定的约束，它对于信息传输是“多余的”。因此从信息传输的角度才把它定义为“冗余”。

例如，由表1可知：$P(u)=0.0201$

又根据统计，当出现字母q时，其后出现字母u的概率近似为1，即：

$$P(u\mid q)\approx 1$$

如：freq ency, techniq e, req est, q ickly, q ite等。这意味着，当传输q时后面的u可以不传，接收端在收到的q后面直接加u即可。

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

中文的冗余度也很大。中文的最大熵是一个变量，每一个单字都具有明确的意义，再由字组词，字词之间的相关性千变万化。

尚未找到给出中文实际熵和统计方法的文献，但根据目前广泛使用的文本压缩软件的压缩率来看，中文的实际熵应该不会大于5比特/汉字，估计中文的冗余度大约也会在80%左右。

一种简单的近似，假设常用的汉字约10000个，则有：

$$H_0(X)=\log_2 10000\approx 13.288 \text{ 比特/汉字}$$

进一步统计汉字出现的概率，将其分为四类。第一类140个，出现概率占50%；第二类485个，出现概率占35%；第三类1775个，出现概率占14.7%；第四类7600个，出现概率占0.3%。为简单，假设每类汉字中各字出现等概，通过计算可得：

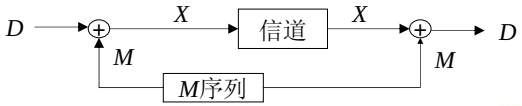
$$H_1(X)\approx 9.773 \text{ 比特/汉字}$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

无记忆的情况简单，容易分析，能否想办法把有记忆的变成无记忆的？

在实际应用中，可以通过扰码将有记忆信源变成无记忆的。然后用无记忆信源的分析方法处理记忆信源，从而使问题得到简化。



$$D\oplus M\oplus M=D\oplus(M\oplus M)=D\oplus 0=D$$

2.2 信源的种类及信息率

XI'AN JIAOTONG UNIVERSITY

扰乱后每数符的概率变为：

$$\begin{aligned}p(X=1)&=p(D=0,M=1)+p(D=1,M=0)\\&=p(D=0)\cdot p(M=1)+p(D=1)\cdot p(M=0)\\&=p(D=0)\cdot\frac{1}{2}+p(D=1)\cdot\frac{1}{2}\\&=\frac{1}{2}\left[p(D=0)+p(D=1)\right]=\frac{1}{2}\end{aligned}$$

$$p(X=0)=1-p(X=1)=\frac{1}{2}$$

2.3 本章小结

XI'AN JIAOTONG UNIVERSITY

- 信源（文字、语音、图像）
- 消息的数值表示
- 信源种类及信息率
- 离散无记忆信源

$$H(X)=\sum_i p_i\log\frac{1}{p_i}\qquad R=rH(X)$$

- 连续无记忆信源

$$H_{abs}(X)=H(X)+\infty\qquad H(X)=\int p(x)\log\frac{1}{p(x)}dx$$

- 离散记忆信源

$$H(X,Y)\qquad H(X\mid x_j)\qquad H(X_n\mid X_{n-1})$$

- 熵的链接准则