第四章 信息论基础

一 通信系统的数学模型

二 信息的度量

三 联合熵与条件熵

四 互信息与平均互信息

五 信息论在密码学中的应用

4.1 通信系统的数学模型

现实生活中的各种通信系统，如电报、电话、图像、计算机、雷达等系统，本质上都是如下图所示的通信系统模型。

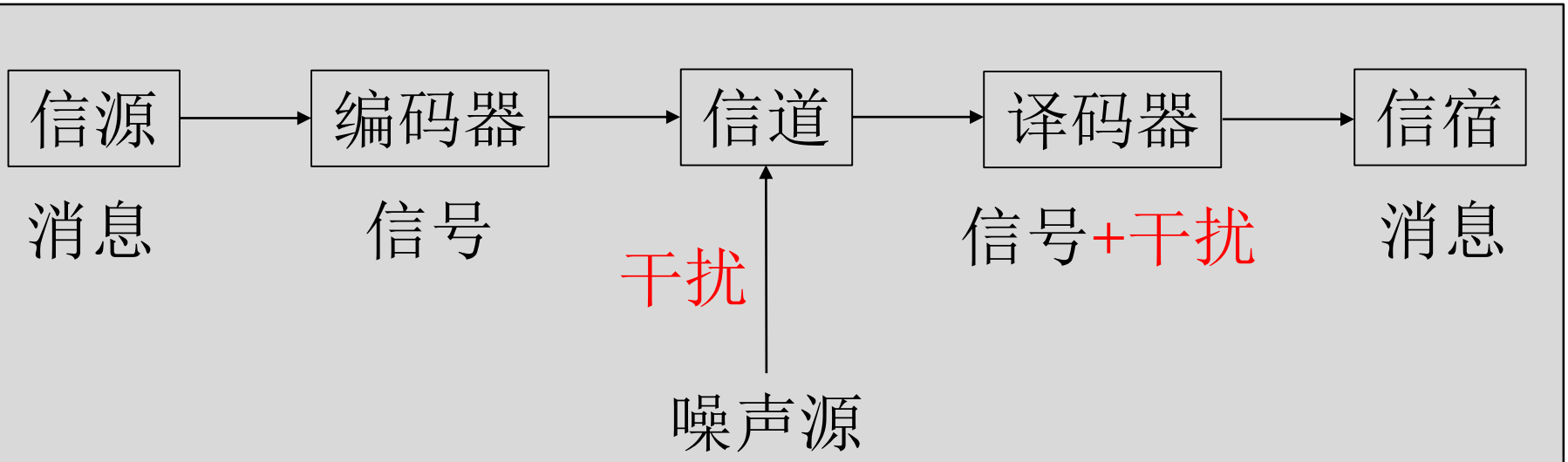


图 4.1　通信系统模型

4.1 通信系统的数学模型

- **信源**

产生消息和消息序列的源。信源按输出的符号取值，可分为“离散信源”和“连续信源”。

4.1 通信系统的数学模型

- **编码器**

将信源发出的消息转换成适合在信道上传送的信号设备。

编码器包括“信源编码器”、“信道编码器”、“调制器”。

- 信源编码器：提高信息传输有效性
- 信道编码器：解决信息传输的可靠性
- 调制器：提高传输效率

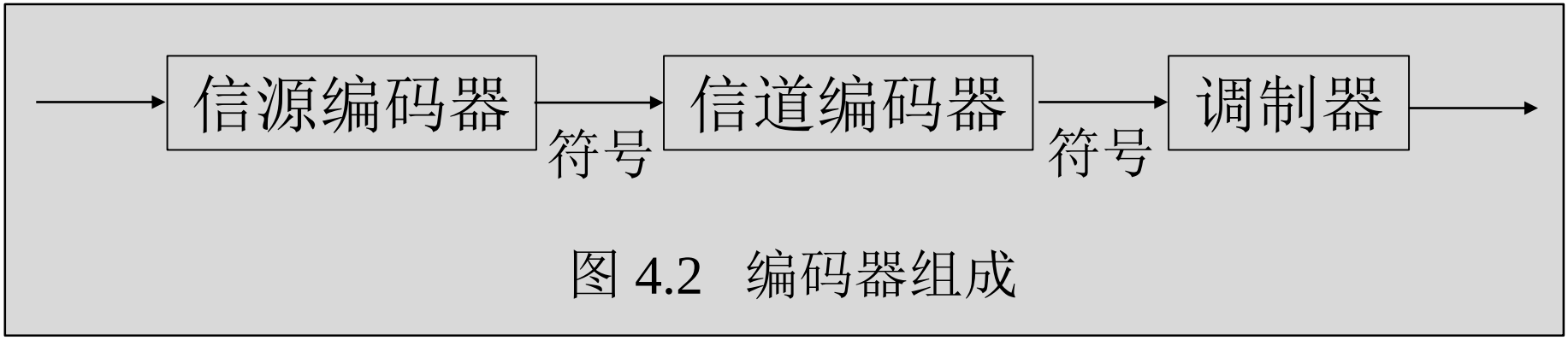


图 4.2　编码器组成

4.1 通信系统的数学模型

- **信道**

信息传输的媒质或通道。种类包括：电缆、光纤、无线电波等。信道干扰分为两大类：

- 由外界引入的随机干扰，如电磁干扰、设备内部噪声等。这类干扰信号叠加在信道信号上，称为“**加性干扰**”。
- 由物理条件的变化引起的参量随机变化，如温度、电离层位置等引起的信号频率、幅度、相位等变化。这类随机干扰与信道信号相乘，称为“**乘性干扰**”。

4.1 通信系统的数学模型

- **译码器**

把受干扰的信道输出尽可能恢复成原来的信源输出，并传给信宿。译码器包括“解调器”、“信道译码器”、“信源译码器”。

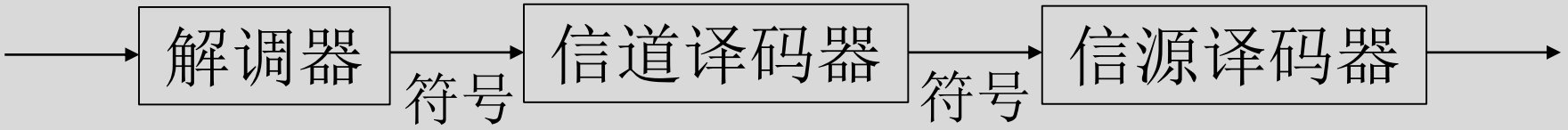


图 4.3　译码器组成

4.1 通信系统的数学模型

- **信宿**

  消息的接受者。从收到的消息试图恢复出原来的消息。

第四章 信息论基础

一 通信系统的数学模型

二 信息的度量

三 联合熵与条件熵

四 互信息与平均互信息

五 信息论在密码学中的应用

4.2 信息的度量

离散信源的信息度量：

- 消息或消息集合本身所含信息量多少的度量。

  使用“自信息”和“信息熵”描述。

- 消息之间或消息集合之间互相提供信息量多少的度量。

  使用“互信息”和“平均互信息”描述。

4.2 信息的度量

- 例如：抛硬币的过程可以看作一个信源，用随机变量 X 表示，而它输出的两种消息（正面、反面）看做两个基本事件 a_1, a_2，概率模型为

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0.5 & 0.5 \end{bmatrix}$$

我们有

$$\sum_{i=1}^{2} p(a_i) = 1$$

4.2 信息的度量

- 对于一般的离散信源 X，假设可能取到的符号有 q 个，其数学模型就是离散概率空间

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & \cdots & p(a_q) \end{bmatrix}$$

我们有

$$\sum_{i=1}^{q} p(a_i) = 1$$

4.2 信息的度量

自信息 $I(x)$ 应满足如下条件：

- $I(x)$ 是 $p(x)$ 的严格递减函数，即小概率事件的不确定度更大，事件发生以后包含的自信息值越大。
- 当 $p(x_1)=0$ 时，$I(x_1)\to\infty$，

  当 $p(x_1)=1$ 时，$I(x)=0$。
- 两个不相关独立事件的总信息量等于各自提供的信息量之和。

可以证明：满足上述公理化的函数形式是对数函数。

4.2 信息的度量

定义4.1 假设离散信源 X 的概率空间为

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & \cdots & p(a_q) \end{bmatrix}$$

事件 a_i 的自信息 $I(a_i)$ 定义为

$$I(a_i) = -\log p(a_i)$$

注意：

- $I(a_i) \geq 0$
- 对数底通常为 2，此时信息量的单位为 bit（比特）。

4.2 信息的度量

例4.1 英文字母 e 出现的概率为 0.123，z 出现的概率为 0.0008，分别计算它们的自信息。

解：根据自信息的定义可得

$$I(e) = -\log 0.123 = 3.02 \text{ bit}$$

$$I(z) = -\log 0.0008 = 13.61 \text{ bit}$$

4.2 信息的度量

定义4.2 随机变量 X 的每个可能取值的自信息 $I(a_i)$ 的统计平均值定义为随机变量 X 的平均自信息，

$$H(X) = E\big(I(X)\big) = -\sum_{i=1}^{q} p(a_i)\log p(a_i)$$

这里 q 为 X 的所有可能取值的个数。

注意：

- 平均自信息又被称为信息熵、信源熵，简称熵。

4.2 信息的度量

例4.3 随机变量 X_1, X_2 的概率空间为

$$\begin{bmatrix} X_1 \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0.01 & 0.99 \end{bmatrix}, \quad \begin{bmatrix} X_2 \\ p(x) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ 0.25 & 0.75 \end{bmatrix}$$

分别求 X_1, X_2 的熵。

解：根据平均自信息的定义可得

$$H(X_1) = -0.01 \times \log 0.01 - 0.99 \times \log 0.99 = 0.081 \text{ bit}$$

$$H(X_2) = -0.25 \times \log 0.25 - 0.75 \times \log 0.75 = 0.881 \text{ bit}$$

上述结果表明：信源 X_1 的不确定性比 X_2 要小。

4.2 信息的度量

- 信息熵 $H(X)$ 是随机变量 X 的概率分布函数，所以又称熵函数。
- 如果把概率分布 $p(a_1), \ldots, p(a_q)$ 记为 $p_1, \ldots, p_q$，则熵函数可以写为概率矢量 $\overline{p} = (p_1, \ldots, p_q)$ 的函数，记为 $H(\overline{p})$。
- 注意到 $\sum_{i=1}^{q} p_i = 1$，因此

$$H(X) = -\sum_{i=1}^{q} p_i \log(p_i) = H(\overline{p})$$

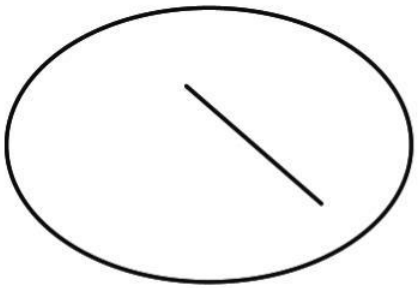
是一个 $q-1$ 元函数。

4.2 信息的度量

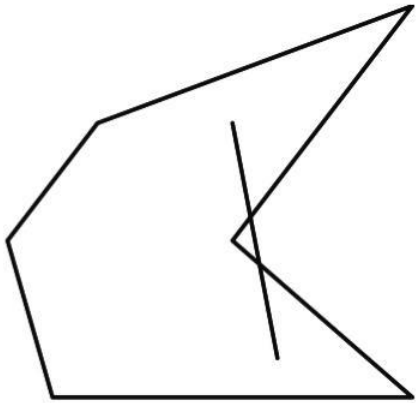
定义4.3 设 D 是一个平面区域，如果满足$\forall x_1 \in D, x_2 \in D$，以及 $\forall\, 0 \le \lambda \le 1$，均有

$$\lambda x_1 + (1-\lambda)x_2 \in D$$

则称区域 D 是凸域。



(a) 凸域



(b) 非凸域

4.2 信息的度量

定义4.3 设 D 是一个平面区域，如果满足$\forall x_1 \in D, x_2 \in D$，以及 $\forall\, 0 \le \lambda \le 1$，均有

$$\lambda x_1 + (1-\lambda)x_2 \in D$$

则称区域 D 是凸域。

注意：

- 凸域中任意两点间的线段也都在凸域内。
- 实数集合和复数集合是常见的凸域。
- 有理数集合和整数集合不是凸域。

4.2 信息的度量

定义4.4 如果凸域 D 上的函数 $f(x)$ 对任意的 $x_1, x_2 \in D, 0 \leq \lambda \leq 1$ 均满足如下性质

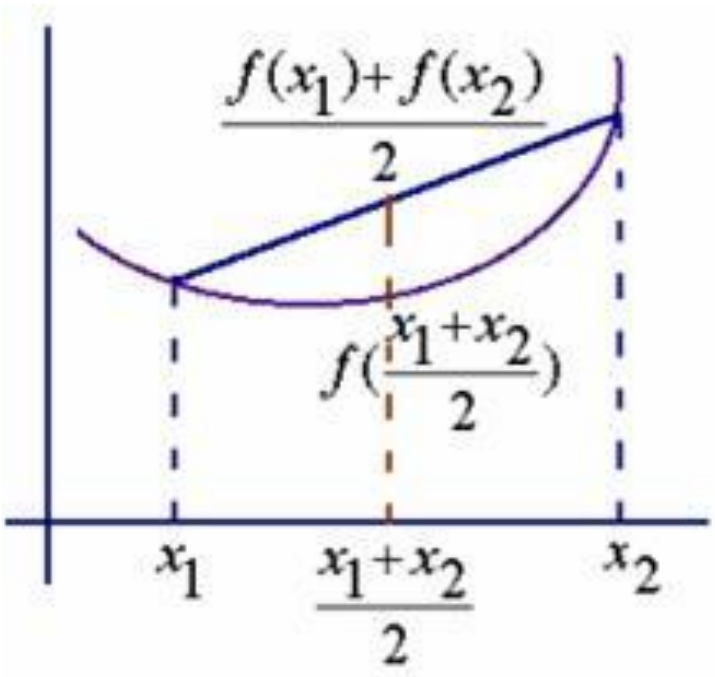
$$\lambda f(x_1) + (1-\lambda) f(x_2) \leq f(\lambda x_1 + (1-\lambda) x_2) \quad (1)$$

则称 $f(x)$ 是 D 上的上凸函数。如果上式当且仅当 $x_1 = x_2$ 时等号成立，则称 $f(x)$ 是 D 上的严格上凸函数。

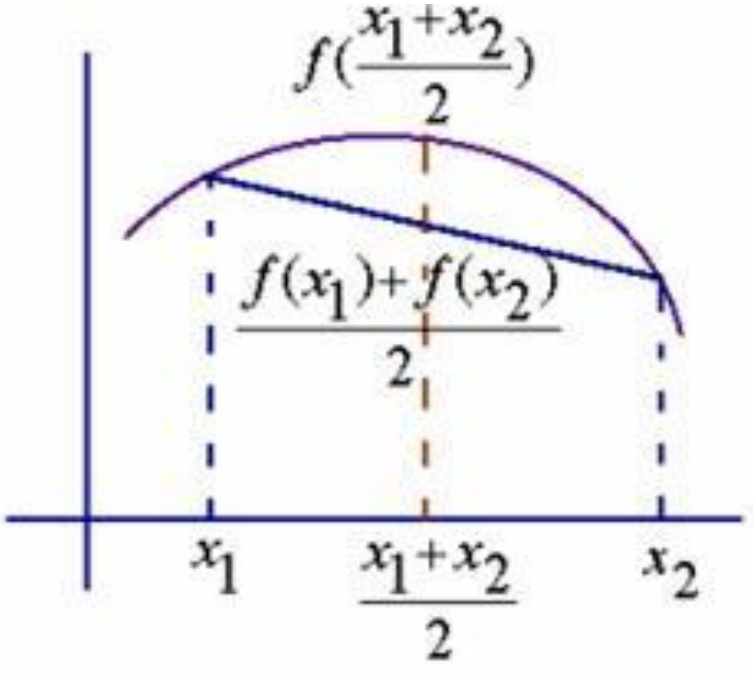
定义4.5 若定义4.4中不等式 (1) 取反方向，则称 $f(x)$ 是 D 上的下凹函数及严格下凹函数。

4.2 信息的度量

下凹函数



上凸函数



4.2 信息的度量

引理（Jensen不等式） 设 f 是区间 I 上的一个严格上凸函数，$\sum_{i=1}^{n} a_i = 1$，$a_i > 0$，$1 \le i \le n$，则

$$\sum_{i=1}^{n} a_i f(x_i) \le f\left(\sum_{i=1}^{n} a_i x_i\right) \qquad (1)$$

其中 $x_i \in I$，$1 \le i \le n$。(1) 式等号成立当且仅当 $x_1 = x_2 = \cdots = x_n$。

4.2 信息的度量

推论 已知对数函数 $f(x)=\log_2 x$ 在 $I=(0,+\infty)$ 上是严格上凸的，则熵函数

$$H(p_1,\cdots,p_n)=\sum_{i=1}^{n}p_i\log\frac{1}{p_i}\le\log\left(\sum_{i=1}^{n}p_i\cdot\frac{1}{p_i}\right)$$

$$=\log n$$

即

$$0\le H(\overline{p})\le\log n$$

上式等号成立当且仅当 $p_1=\cdots=p_n$

4.2 信息的度量

例子 考虑如下分布的随机变量

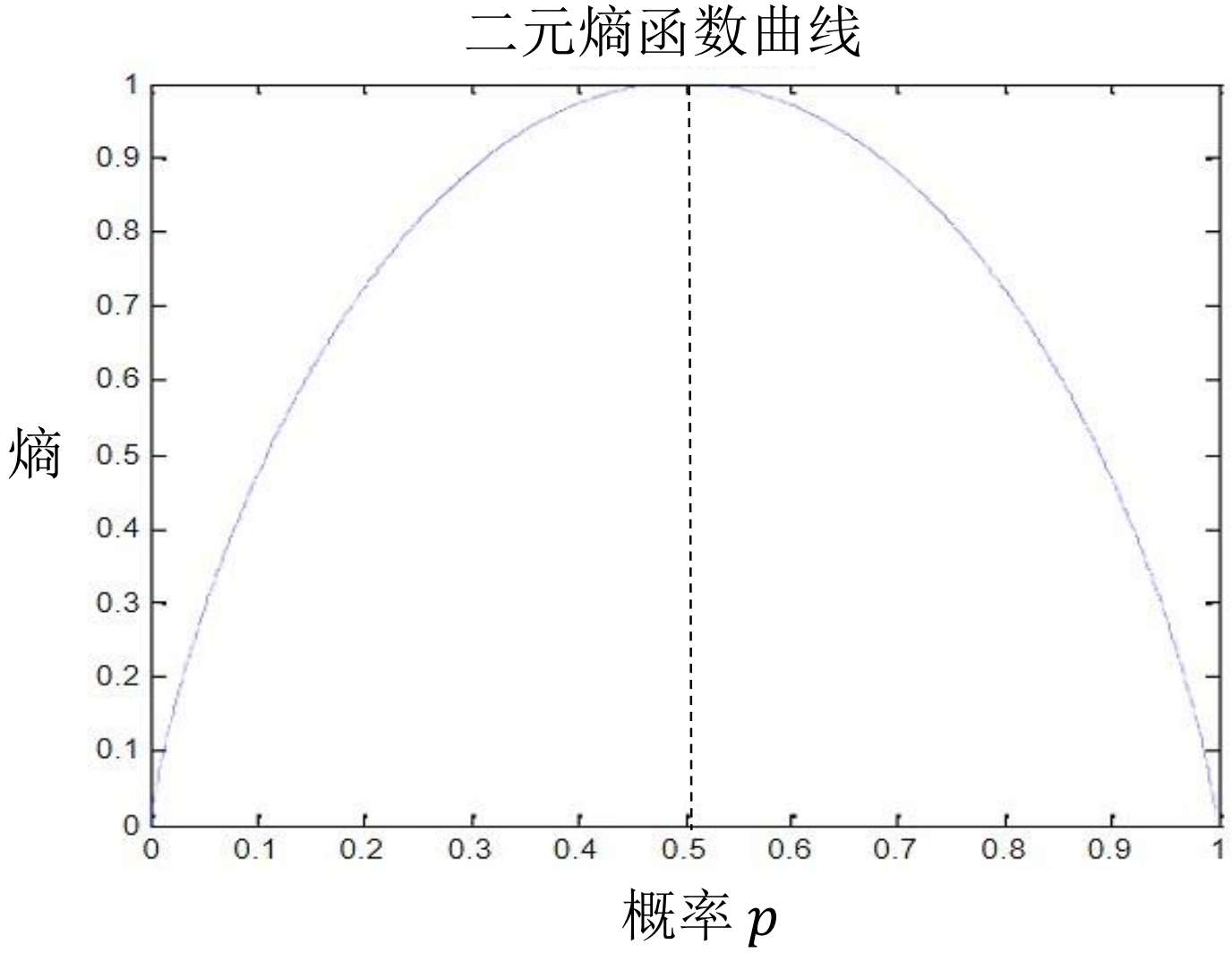
$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$$

它的熵简记为 $H(p)$，则

$$H(p) = -p\log p - (1-p)\log(1-p)$$

当等概分布时（$p = 0.5$），熵达到最大值 1 bit，这时随机变量具有最大的不确定性。当 $p = 0$ 或 1 时，熵等于 0，即没有不确定性。

4.2 信息的度量



4.2 信息的度量

- 熵函数 $H(\overline{p})$ 的性质：

1. 非负性：$H(\overline{p}) \geq 0$

证：$H(\overline{p}) = -\sum_{i=1}^{q} p_i \log(p_i)$，由于 $0 \leq p_i \leq 1$，因此 $\log(p_i) \leq 0$，即 $H(\overline{p}) \geq 0$。

2. 对称性：

$$H(p_1, p_2, \cdots, p_q) = H(p_2, p_1, \cdots, p_q) = H(p_q, p_1, \cdots, p_{q-1})$$

即概率矢量 $\overline{p}$ 各分量的次序可以任意变更，熵值不变。

4.2 信息的度量

- 熵函数 $H(\bar{p})$ 的性质：

3. 确定性：$H(1,0)=H(1,0,0)=H(1,0,\dots,0)=0$

证：$H(\bar{p})=-\sum_{i=1}^{q}p_i\log(p_i)=1\cdot\log 1+0+\cdots+0=0$

4. 扩展性：

$$\lim_{\epsilon\to 0}H_{q+1}(p_1,p_2,\dots,p_q-\epsilon,\epsilon)=H_q(p_1,\dots,p_q)$$

注意到 $\lim_{\epsilon\to 0}\epsilon\log\epsilon=0$。说明增加一个基本不会出现的小概率事件，信源的熵保持不变。

4.2 信息的度量

- 熵函数 $H(\bar{p})$ 的性质：

5. 连续性：

$$\lim_{\epsilon \to 0} H_q(p_1, p_2, \cdots, p_{q-1} - \epsilon, p_q + \epsilon) = H_q(p_1, \cdots, p_q)$$

6. 递推性：

$$H(p_1, p_2, \cdots, p_n) = H[(p_1 + p_2 + \cdots + p_k), p_{k+1}, \cdots, p_n] + (p_1 + p_2 + \cdots + p_k) H(p'_1, p'_2, \cdots, p'_k)$$

其中 $p'_l = p_l / (p_1 + p_2 + \cdots + p_k)$，$l = 1, 2, \cdots, k$ 。

说明信源中某个消息 x_i 被划分后，新信源的熵由于 x_i 被划分产生的不确定性而增加了。

4.2 信息的度量

递推性证明：

$$H(p_1, p_2, \dots, p_n) - H[(p_1 + p_2 + \cdots + p_k), p_{k+1}, \dots, p_n]$$

$$= \sum_{i=1}^{k} p_i \log \frac{1}{p_i} - (p_1 + p_2 + \cdots + p_k) \log \frac{1}{p_1 + p_2 + \cdots + p_k}$$

$$= (p_1 + p_2 + \cdots + p_k) \left(\sum_{l=1}^{k} \frac{p_l}{p_1 + p_2 + \cdots + p_k} \log \frac{p_1 + p_2 + \cdots + p_k}{p_l} \right)$$

$$= (p_1 + p_2 + \cdots + p_k) \sum_{l=1}^{k} p'_l \log \frac{1}{p'_l}$$

$$= (p_1 + p_2 + \cdots + p_k) H(p'_1, p'_2, \dots, p'_k)$$

4.2 信息的度量

- 熵函数 $H(\bar{p})$ 的性质：

7. 可加性：

$$H_{nm}(p_1q_1, p_1q_2, \cdots, p_1q_m, p_2q_1, \cdots, p_2q_m, p_nq_1, \cdots, p_nq_m)$$
$$= H_n(p_1, p_2, \cdots, p_n) + H_m(q_1, q_2, \cdots, q_m)$$

两个事件的联合熵等于各自熵之和。证：

$$H_{nm}(p_1q_1, p_1q_2, \cdots, p_1q_m, p_2q_1, \cdots, p_2q_m, p_nq_1, \cdots, p_nq_m)$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_i\, q_j \log\left(p_i q_j\right)$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_i\, q_j \log\left(p_i\right) - \sum_{i=1}^{n}\sum_{j=1}^{m} p_i\, q_j \log\left(q_j\right)$$
$$= -\sum_{j=1}^{m} q_j\left(\sum_{i=1}^{n} p_i \log p_i\right) - \sum_{i=1}^{n} p_i\left(\sum_{j=1}^{m} q_j \log q_j\right)$$
$$= H_n(p_1, p_2, \cdots, p_n) + H_m(q_1, q_2, \cdots, q_m)$$

第四章 信息论基础

一 通信系统的数学模型

二 信息的度量

三 联合熵与条件熵

四 互信息与平均互信息

五 信息论在密码学中的应用

4.3 联合熵与条件熵

- 联合事件的自信息称为联合自信息。

定义4.6 两个随机事件的联合自信息定义为这两个事件同时发生的联合概率的对数的负值，简称联合自信息。设事件 x_i, y_j 的联合概率为 $p(x_i y_j)$，则它们的联合自信息定义为

$$I(x_i y_j) = -\log p(x_i y_j)$$

4.3 联合熵与条件熵

例4.6 已知箱中有10个红球，5个白球。现从箱中随机地取出两个球。求

1. 事件“两个球中有红、白各一个”的不确定性；
2. 事件“两个球都是白球”所提供的信息量；
3. 事件“两个球都是白球”与“两个球都是红球”的发生，哪个事件更难猜测？

4.3 联合熵与条件熵

例4.6 解1：事件“两个球中有红、白各一个”的不确定性；

$$p(1,1)=\frac{C_{10}^{1}C_{5}^{1}}{C_{15}^{2}}=\frac{10}{21},$$

$$I(1,1)=-\log\frac{10}{21}=1.07\ \text{bit}$$

4.3 联合熵与条件熵

例4.6 解2：事件“两个球都是白球”所提供的信息量；

$$p(0,2)=\frac{C_5^2}{C_{15}^2}=\frac{2}{21},$$

$$I(0,2)=-\log\frac{2}{21}=3.39\ \text{bit}$$

4.3 联合熵与条件熵

例4.6 解3：事件“两个球都是白球”与“两个球都是红球”的发生，哪个事件更难猜测？

两个球都是红球的信息量：$p(2,0)=\frac{C_{10}^{2}}{C_{15}^{2}}=\frac{9}{21}$，

$$I(2,0)=-\log\frac{9}{21}=1.22\ \text{bit}$$

因此，两个球都是白球更难猜测。

4.3 联合熵与条件熵

- 两个事件的条件概率给出的信息称为条件自信息。

定义4.7 两个随机事件的条件自信息定义为在一个事件发生的条件下，另一个事件发生的概率对数的负值，简称条件自信息。设事件 $y = b_j$ 发生后，事件 $x = a_i$ 发生的概率为 $p(a_i \mid b_j)$，相应的条件自信息定义为

$$I(a_i \mid b_j) = -\log p(a_i \mid b_j)$$

4.3 联合熵与条件熵

- 条件自信息含义为：
  - 在事件 $y = b_j$ 给定的条件下，事件 $x = a_i$ 发生前的不确定性；
  - 在事件 $y = b_j$ 给定的条件下，事件 $x = a_i$ 发生后所得到的信息量。
- 利用定义容易证明

$$I(a_i b_j) = I(a_i) + I(b_j \mid a_i)$$

注意到，我们有 $p(a_i b_j) = p(a_i) \cdot p(b_j \mid a_i)$

4.3 联合熵与条件熵

例4.7 已知箱中有10个红球，5个白球。现从箱中先后拿出两球。求

1. 事件“第一个是红球条件下，第二个是白球”的不确定性；
2. 事件“第一个是红球条件下，第二个是红球”所提供的信息量；

4.3 联合熵与条件熵

例4.7 解1：事件“第一个是红球条件下，第二个是白球”的不确定性；

$$I(y = \text{white} \mid x = \text{red})$$

$$= -\log p(y = \text{white} \mid x = \text{red}) = -\log \frac{5}{14} = 1.49 \text{ bit}$$

例4.7 解2：事件“第一个是红球条件下，第二个是红球”所提供的信息量；

$$I(y = \text{red} \mid x = \text{red})$$

$$= -\log p(y = \text{red} \mid x = \text{red}) = -\log \frac{9}{14} = 0.64 \text{ bit}$$

4.3 联合熵与条件熵

- 一维随机变量的不确定性可以用信息熵来表示。
- 两个随机事件集合可用二维随机变量表示，二维随机变量的平均不确定性可以推广为联合熵。

4.3 联合熵与条件熵

定义4.8 设二维随机变量 (X,Y) 的概率空间为

$$p(x_iy_j)=p\{X=x_i,Y=y_j\},$$

$$x=x_1,\dots,x_n,y=y_1,\dots,y_m$$

其中 $0\le p(x_iy_j)\le 1$，$\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)=1$，则二维随机变量 (X,Y) 的**联合熵** $H(X,Y)$ 定义为联合自信息 $I(X,Y)$ 在 (X,Y) 的联合概率空间上的数学期望，它是 (X,Y) 平均不确定性的度量，即

$$H(X,Y)=E[I(XY)]=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log p(x_iy_j)$$

4.3 联合熵与条件熵

例4.9 计算下列事件的联合自信息：

1. 英文字母 e 出现的概率是 0.123，z 出现的概率是 0.0008，分别计算它们的自信息量。假定字母出现是相互独立事件，计算 ez 的联合自信息。
2. 4人抓阄，设 x_1 表示第一个人没抓到，y_1 表示第二个人没抓到，计算 x_1y_1 的联合自信息。

   X 表示第一个人抓阄情况，Y 表示第二个人的抓阄情况，计算 XY 的联合熵。

4.3 联合熵与条件熵

例4.9 解1：英文字 e 出现的概率是 0.123，z 出现的概率是0.0008，它们的自信息量：

$$I(e) = -\log 0.123 = 3.02 \text{ bit},$$

$$I(z) = -\log 0.0008 = 10.29 \text{ bit}$$

假定字母出现是相互独立事件， ez 的联合自信息：

$$I(ez) = -\log (0.123 \times 0.0008) = 13.31 \text{ bit}$$

两个独立事件的联合自信息等于两个事件各自自信息之和。

4.3 联合熵与条件熵

例4.9 解2：4人抓阄，x_1 表示第一个人没抓到，y_1 表示第二个人没抓到，x_1y_1 的联合自信息：

$$I(x_1y_1) = -\log p(x_1y_1) = -\log\left(\frac{3}{4}\times\frac{2}{3}\right) = 1 \text{ bit}$$

注意到，$I(x_1) = I(y_1) = -\log 3/4$=0.415 bit，故

$$I(x_1y_1) \neq I(x_1) + I(y_1)$$

两个非独立事件的联合自信息不等于两个事件各自自信息之和。

4.3 联合熵与条件熵

例4.9 解2： X 表示第一个人抓阄情况， Y 表示第二个人的抓阄情况，XY 的联合熵：

XY 共有3种可能（1表示没抓到，0表示抓到了），则

$$p(x_1y_1)=\frac{1}{2},\ p(x_0y_1)=\frac{1}{4},\ p(x_1y_0)=\frac{3}{4}\times\frac{1}{3}=\frac{1}{4}$$

在 XY 的概率空间上取平均，得到联合熵

$$H(XY)=E[I(XY)]=\frac{1}{2}\times 1+\frac{1}{4}\times 2+\frac{1}{4}\times 2=\frac{3}{2}\ \text{bit}$$

4.3 联合熵与条件熵

定义4.9 在给定 Y 时，X 的**条件熵** $H(X \mid Y)$ 被定义为

$$H(X \mid Y) = E[I(X \mid Y)]$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i y_j)\log p(x_i \mid y_j)$$

4.3 联合熵与条件熵

条件熵 $H(X \mid Y)$ 可以根据如下取两次统计平均得到：

1. 在给定 $Y = y_j$ 条件下，X 的不确定性可定义为 $I(X \mid Y)$ 在 $X \mid Y$ 条件概率空间上的统计平均：

$$H(X \mid y_j) = -\sum_{i=1}^{n} p(x_i \mid y_j) \log p(x_i \mid y_j)$$

2. 对不同的 y_j，将 $H(X \mid y_j)$ 在 X 的概率空间上取统计平均：

$$\begin{aligned} H(X \mid Y) &= -\sum_{j=1}^{m} p(y_j) H(X \mid y_j) \\ &= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(y_j) p(x_i \mid y_j) \log p(x_i \mid y_j) \\ &= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) \log p(x_i \mid y_j) \end{aligned}$$

4.3 联合熵与条件熵

定理4.1 熵、联合熵、条件熵满足如下等式：

1. $H(XY) = H(X) + H(Y \mid X)$；
2. $H(XY) = H(Y) + H(X \mid Y)$

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

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

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

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

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

4.3 联合熵与条件熵

定理4.1 熵、联合熵、条件熵满足如下等式：

1. $H(XY) = H(X) + H(Y \mid X)$；
2. $H(XY) = H(Y) + H(X \mid Y)$

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

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

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

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

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

X 已知下 Y 的熵

X 的熵

XY 的联合熵

4.3 联合熵与条件熵

定理4.2 熵、联合熵、条件熵满足如下不等式：

1. $\max\big(H(X),H(Y)\big)\le H(XY)\le H(X)+H(Y)$；
2. $0\le H(X\mid Y)\le H(X)$；
3. $0\le H(Y\mid X)\le H(Y)$；

证：因为 $p(x\mid y)\le 1$，所以 $H(X\mid Y)\ge 0$。

由 $H(XY)=H(Y)+H(X\mid Y)\ge H(X)$，则

（上式说明：联合熵不小于独立熵）

$$H(XY)\ge \max\big(H(X),H(Y)\big)$$

4.3 联合熵与条件熵

证（续）：

$$H(XY)-H(X)-H(Y)$$

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

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

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

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

$$\le\log\sum_x\sum_y p(x)p(y)$$

$$=\log 1=0$$

由Jensen不等式，

$\log x$ 是严格上凸函数

4.3 联合熵与条件熵

证（续）：

$$H(X \mid Y) - H(X)$$

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

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

$$= \sum_x \sum_y p(x,y) \log \frac{p(x)}{p(x|y)}$$

$$\leq \log \sum_x \sum_y \frac{p(x,y)p(x)}{p(x|y)}$$

$$= \log \sum_x \sum_y p(x)p(y)$$

$$= 0$$

由Jensen不等式，
$\log x$ 是严格上凸函数

4.3 联合熵与条件熵

下面讨论两种极端情况：

1. 当 X 与 Y 互相独立时，容易证明：

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

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

$$H(XY) = H(X) + H(Y)$$

2. 当 X 与 Y 之间有一一对应关系时，容易证明：

$$H(XY) = H(X) = H(Y)$$

$$H(X \mid Y) = H(Y \mid X) = 0$$

4.3 联合熵与条件熵

例4.10 随机变量 X, Y 的联合分布如下：

X \ Y	y_1	y_2	y_3
x_1	$7/24$	$1/24$	0
x_2	$1/24$	$1/4$	$1/24$
x_3	0	$1/24$	$7/24$

1. 如果知道 X 和 Y 的结果，求得到的平均信息量；
2. 如果知道 Y 的结果，求得到的平均信息量;
3. 在已知 Y 的情况下，知道 X 的结果，求得到的平均信息量。

4.3 联合熵与条件熵

例4.10 解1：

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

$$=2\times\frac{7}{24}\log\frac{24}{7}+4\times\frac{1}{24}\log 24+\frac{1}{4}\log 4=2.3\ \text{bit}$$

解2： X,Y 的概率分布如下

$$\begin{bmatrix}X\\p(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3\\ \frac{8}{24} & \frac{8}{24} & \frac{8}{24}\end{bmatrix},\quad \begin{bmatrix}Y\\p(Y)\end{bmatrix}=\begin{bmatrix}y_1 & y_2 & y_3\\ \frac{8}{24} & \frac{8}{24} & \frac{8}{24}\end{bmatrix}$$

所以 $H(Y)=3\times\frac{1}{3}\log 3=1.58\ \text{bit}$。

解3： $H(X\mid Y)=H(XY)-H(Y)=2.3-1.58=0.72\ \text{bit}$

4.3 联合熵与条件熵

例4.10 解3（续）：

事实上，X, Y 的条件概率分布如下

$p(x \mid y)$	y_1	y_2	y_3
x_1	$\frac{7}{8}$	$\frac{1}{8}$	0
x_2	$\frac{1}{8}$	$\frac{6}{8}$	$\frac{1}{8}$
x_3	0	$\frac{1}{8}$	$\frac{7}{8}$

$$H(X \mid Y) = \sum_x \sum_y p(x,y) \log \frac{1}{p(x|y)}$$

$$= 2 \times \frac{7}{24} \times \log \frac{8}{7} + 2 \times \frac{1}{24} \times \log \frac{8}{1} + 2 \times \frac{1}{24} \times \log \frac{8}{1} + \frac{1}{4} \times \log \frac{8}{6}$$

$$= 0.72 \text{ bit}$$

第四章 信息论基础

一 通信系统的数学模型

二 信息的度量

三 联合熵与条件熵

四 互信息与平均互信息

五 信息论在密码学中的应用

4.4 互信息与平均互信息

定义4.10 一个事件 y_j 给出关于另一个事件 x_i 的信息称为互信息，用 $I(x_i;y_j)$ 表示，即

$$I(x_i;y_j)=I(x_i)-I(x_i\mid y_j)=\log\frac{p(x_i|y_j)}{p(x_i)}$$

- 互信息 $I(x_i;y_j)$ 是已知事件 y_j 后所消除的关于事件 x_i 的不确定性。
- 在通信系统中，互信息的物理意义是，信道输出端接收到某消息 y_j 后获得的关于输入端某消息 x_i 的信息量。

4.4 互信息与平均互信息

例4.11 一个有80个彩球的箱子里分别有黄球40个，红球20个，蓝球10个，绿球10个。如果消息 y_1 表示摸出的球不是黄球，求收到 y_1 后，y_1 与摸到各种颜色的球的互信息量。

解 记消息 x_1（黄），x_2（红），x_3（蓝），x_4（绿），

$$p(x_1 \mid y_1) = 0,\ \ p(x_2 \mid y_1) = \frac{p(x_2 y_1)}{p(y_1)} = \frac{1}{2}$$

$$p(x_3 \mid y_1) = \frac{p(x_3 y_1)}{p(y_1)} = \frac{1}{4},\ \ p(x_4 \mid y_1) = \frac{p(x_4 y_1)}{p(y_1)} = \frac{1}{4}$$

y_1 与摸到各种颜色的球的互信息量：

$$I(x_1; y_1) = \log \frac{p(x_1|y_1)}{p(x_1)} = \infty,\ \ I(x_2; y_1) = \log \frac{p(x_2|y_1)}{p(x_2)} = 1 \text{ bit},$$

$$I(x_3; y_1) = \log \frac{p(x_3|y_1)}{p(x_3)} = 1 \text{ bit}, I(x_4; y_1) = \log \frac{p(x_4|y_1)}{p(x_2)} = 1 \text{ bit}$$

4.4 互信息与平均互信息

定理4.3 互信息具有以下几点性质：

1. 互易性：$I(x;y)=I(y;x)$；
2. 当事件 x,y 统计独立时，互信息为零，$I(x;y)=0$；
3. 互信息取值可能是正数或者负数；
4. 任何两事件之间的互信息不可能大于其中任意事件的自信息。

4.4 互信息与平均互信息

证1：互易性：$I(x;y)=I(y;x)$。

$$I(x;y)=\log\frac{p(x|y)}{p(x)}=\log\frac{p(xy)}{p(x)p(y)}=\log\frac{p(y|x)}{p(y)}=I(y;x)$$

证2： x,y 统计独立时，$I(x;y)=0$。

$$I(x;y)=\log\frac{p(x|y)}{p(x)}=\log 1=0$$

证3：互信息取值可能是正数或者负数。

$p(x\mid y)>p(x)$ 时 $I(x;y)>0$，$p(x\mid y)<p(x)$ 时 $I(x;y)<0$。

证4：任何两事件之间的互信息小于等于其中任意事件的自信息。

由 $I(x\mid y)\geq 0$， $I(x;y)=I(x)-I(x\mid y)\leq I(x)$。

4.4 互信息与平均互信息

定义4.11 随机变量 X 与 Y 之间的平均互信息被定义为

$$I(X;Y) = E[I(x;y)] = \sum_i \sum_j p(x_i y_j) I(x_i;y_j)$$

$$= \sum_i \sum_j p(x_i y_j) \log \frac{p(x_i|y_j)}{p(x_i)}$$

4.4 互信息与平均互信息

定理4.4 随机变量 X 与 Y 之间的平均互信息满足

$$I(X;Y) = H(X) - H(X \mid Y)$$

证：

$$I(X;Y)$$

$$= E[I(x;y)] = \sum_i \sum_j p(x_i y_j) I(x_i;y_j)$$

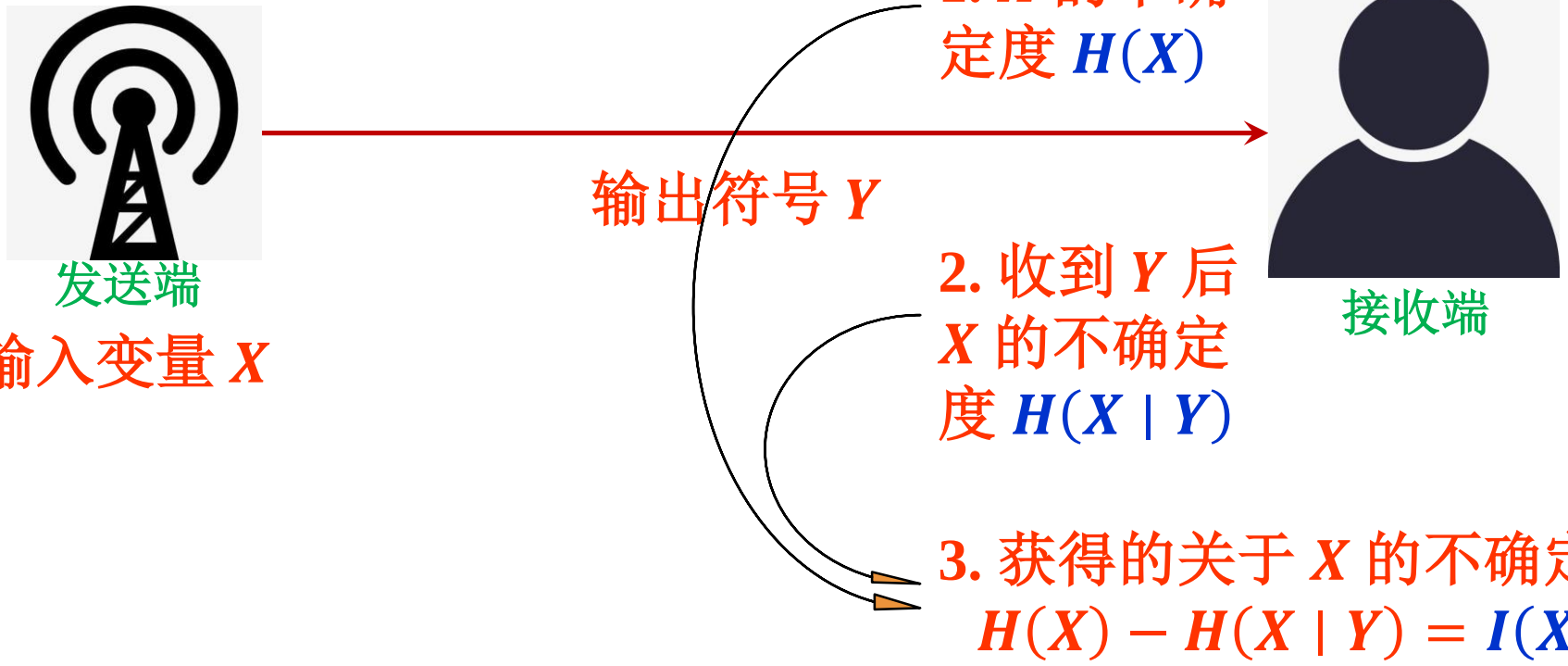
$$= \sum_i \sum_j p(x_i y_j) \log \frac{p(x_i|y_j)}{p(x_i)}$$

$$= \sum_i \sum_j p(x_i y_j) \log \frac{1}{p(x_i)} - \sum_i \sum_j p(x_i y_j) \log \frac{1}{p(x_i|y_j)}$$

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

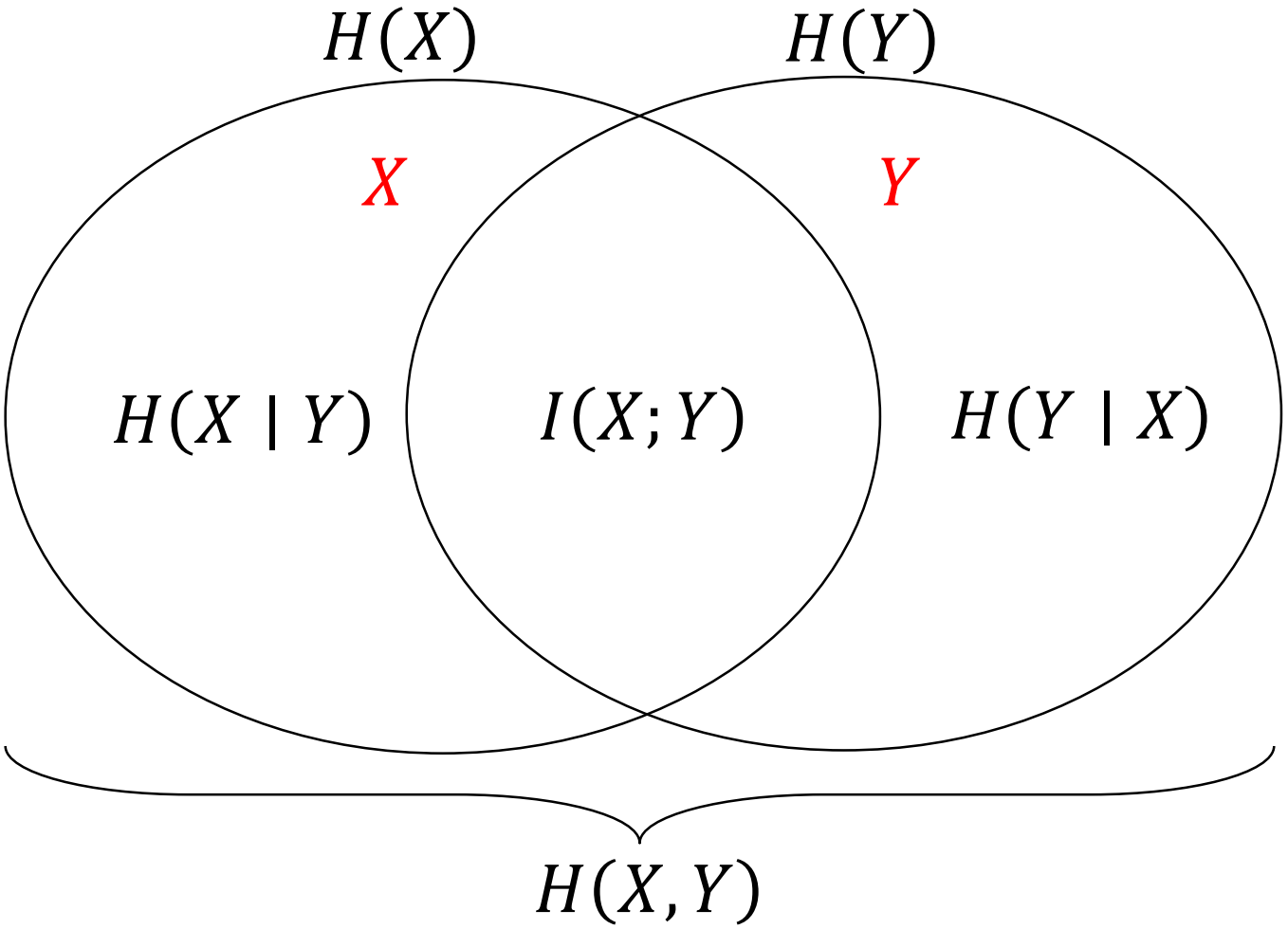
4.4 互信息与平均互信息

- 随机变量 Y 关于 X 的平均互信息 $I(X;Y)$ 是收到 Y 后关于 X 的不确定性减少的量，即从 Y 所获得的关于 X 的平均信息量。



4.4 互信息与平均互信息

- 平均互信息、熵、条件熵的关系图：



4.4 互信息与平均互信息

定理4.5 平均互信息具有如下性质：

1. 非负性：$I(X;Y) \geq 0$，等式成立当且仅当 X, Y 独立；
2. 对称性：$I(X;Y) = I(Y;X)$

证1：$-I(X;Y) = \sum_i \sum_j p(x_i y_j) \log \dfrac{p(x_i)p(y_j)}{p(x_i y_j)}$

$$\leq \log\left(\sum_i \sum_j p(x_i y_j) \frac{p(x_i)p(y_j)}{p(x_i y_j)}\right)$$

$$= \log \sum_i \sum_j p(x_i)p(y_j)$$

$$= 0$$

4.4 互信息与平均互信息

证2：对称性： $I(X;Y) = I(Y;X)$

$$
\begin{aligned}
I(Y;X) &= H(Y) - H(Y \mid X) \\
&= H(Y) - \big(H(X,Y) - H(X)\big) \\
&= H(X) + H(Y) - H(X,Y) \\
&= H(X) - \big(H(X,Y) - H(Y)\big) \\
&= H(X) - H(X \mid Y) \\
&= I(X;Y)
\end{aligned}
$$

4.4 互信息与平均互信息

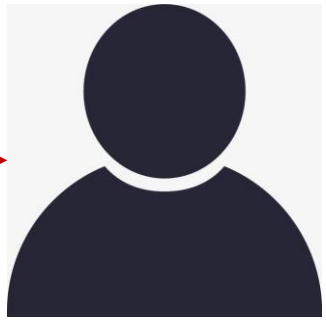
- 在通信系统中，平均意义上，信道每通过一条信息，总能传递一定的信息量，即接收端每收到一条消息，总能提取到信源 X 的信息量，总能使信源的不确定性有所降低。

发送端

输入变量 X

输出符号 Y

X ？ ？ ？



接收端

获得的关于 X 的不确定度

$I(X;Y) \geq 0$

4.4 互信息与平均互信息

例4.12 某地一段时间内出现晴、阴、雨的概率均为 1/3，

- 如果某天不是雨天，则抛 1 次硬币；
- 如果某天是雨天，则抛 2 次硬币；

试计算从抛硬币出现正面的次数可以得到多少关于天气的信息量。

解：根据抛硬币的正面次数 Y 来获得关于天气 X 的信息。

$X=1$ 表示雨天，$X=0$ 表示不是雨天，

$Y=0$ 表示硬币正面 0 次，$Y=1$ 表示硬币正面 1 次，$Y=2$ 表示硬币正面 2 次。

4.4 互信息与平均互信息

例4.12 解（续） X 的概率分布 $\begin{bmatrix} X \\ p(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2/3 & 1/3 \end{bmatrix}$,

Y 的条件概率分布 $P_{Y|X} = \begin{matrix} \\ X=0 \\ X=1 \end{matrix} \begin{matrix} \begin{matrix} Y=0 & Y=1 & Y=2 \end{matrix} \\ \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix} \end{matrix}$

求得 Y 的概率分布

$$P_Y = P_{Y|X}^T P_X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}^T \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{pmatrix} \frac{5}{12} & \frac{1}{2} & \frac{1}{12} \end{pmatrix}^T$$

Y 的熵：$H(Y) = \sum_y p(y) \log 1/p(y) = 1.325$ bit

4.4 互信息与平均互信息

例4.12 解（续）

由 $P_{Y|X}$ 求得已知 X 时 Y 的条件熵：

$$H(Y \mid X) = \sum_x \sum_y p(x,y) \log 1/p(y \mid x)$$

$$= \frac{2}{3}\left(\frac{1}{2}\log 2 + \frac{1}{2}\log 2\right) + \frac{1}{3}\left(\frac{1}{4}\log 4 + \frac{1}{2}\log 2 + \frac{1}{4}\log 4\right)$$

$$= 1.166 \text{ bit}$$

故

$$I(X;Y) = H(Y) - H(Y \mid X) = 1.325 - 1.166 = 0.159 \text{ bit}$$

即从抛硬币出现正面次数平均得到关于天气的信息量为 0.159 bit。

第四章 信息论基础

一 通信系统的数学模型

二 信息的度量

三 联合熵与条件熵

四 互信息与平均互信息

五 信息论的应用

4.5 离散信源

根据信源输出消息的不同随机性质，信源可分为如下两类：

- 离散信源

信源输出的消息数是有限的或可数的，每次只输出一个消息。

  - 如果信源符号集为有限集，则称为有限离散信源。
  - 如果信源符号集为无限集，则称为无限离散信源。

4.5 离散信源

- 连续信源

信源输出的消息数是无限的或不可数的，每次只输出一个消息。

  - 如果一个信源输出仅发生在离散时间间隔而取值是连续的，称为离散时间连续信源。
  - 如果一个信源输出是时间和取值都是连续的波形，称为连续时间信源，简称波形信源或模拟信源。

把波形信源在时间上离散化就可以得到离散时间连续信源。连续信源的熵可以通过在信源取值上的离散化（变成离散信源）实现。

4.5 离散信源

- 离散信源的概率空间表示：

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & \cdots & p(a_q) \end{bmatrix}, \qquad \sum_{i=1}^{q} p(a_i) = 1$$

- 连续信源的概率空间表示：

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix}, \qquad \int_a^b p(x)\mathrm{d}x = 1$$

4.5 离散信源

根据信源输出消息的平稳与否，信源可分为如下两类：

- 非平稳信源

  输出随机序列的统计特征比较复杂，分析困难。

- 平稳信源
  - 连续平稳信源
  - 离散无记忆平稳信源
  - 离散无记忆扩展信源

4.5 离散信源

- **连续平稳信源：**信源输出的随机序列 $X=(X_1,X_2,\dots,X_N)$ 中的 $X_i(i=1,\dots,N)$ 为取值连续的随机变量，任意两个不同时刻 X 的各维概率密度函数都相同。
- **离散无记忆平稳信源：**信源输出的随机序列 $X=(X_1,X_2,\dots,X_N)$ 中的随机变量 $X_i(i=1,\dots,N)$ 是同一概率空间上统计独立的。
- **离散无记忆扩展信源**

4.5 离散信源

- **离散无记忆扩展信源：** X 是一个离散无记忆平稳信源，X 的符号集为 $A=\{a_1,\dots,a_q\}$，信源每次发一个符号，符号间统计独立，X 的概率空间为

$$\begin{bmatrix} X \\ p(x) \end{bmatrix}=\begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & \cdots & p(a_q) \end{bmatrix},\ \sum_{i=1}^{q}p(a_i)=1$$

X 的 N 次扩展信源 X^N 是具有 q^N 个符号的离散信源，其 N 重概率空间为

$$\begin{bmatrix} X^N \\ p(x^N) \end{bmatrix}=\begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_2 & \cdots & \bar{\alpha}_{q^N} \\ p(\bar{\alpha}_1) & p(\bar{\alpha}_2) & \cdots & p(\bar{\alpha}_{q^N}) \end{bmatrix},\ \bar{\alpha}_i=(a_{i_1},a_{i_2},\dots,a_{i_N})$$

4.5 离散信源

- 离散无记忆扩展信源（续）：X^N 的概率空间为

$$\begin{bmatrix} X^N \\ p(x^N) \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_2 & \cdots & \bar{\alpha}_{q^N} \\ p(\bar{\alpha}_1) & p(\bar{\alpha}_2) & \cdots & p(\bar{\alpha}_{q^N}) \end{bmatrix}, \quad \bar{\alpha}_i = (a_{i_1}, a_{i_2}, \ldots, a_{i_N})$$

此时，$p(\bar{\alpha}_i) = p(a_{i_1})p(a_{i_2})\cdots p(a_{i_N})$，且 $\sum_{i=1}^{q^N} p(\bar{\alpha}_i) = 1$。

利用熵的定义，可得 N 次扩展信源 X^N 的熵为

$$H(X^N) = -\sum_{X^N} p(x^N)\log p(x^N) = -\sum_{i=1}^{q^N} p(\bar{\alpha}_i)\log p(\bar{\alpha}_i)$$

4.5 离散信源

例 4.13 设信源 X 的概率空间为 $\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}$，

试求 $H(X)$ 和 $H(X^2)$。

解： $H(X) = -\sum_{i=1}^{3} p(a_i)\log p(a_i) = 1.5 \text{ bit}$

X^2 的概率空间为 $\begin{bmatrix} X^2 \\ p(x^2) \end{bmatrix} =$

$$\begin{bmatrix} a_1a_1 & a_1a_2 & a_1a_3 & a_2a_1 & a_2a_2 & a_2a_3 & a_3a_1 & a_3a_2 & a_3a_3 \\ 1/4 & 1/8 & 1/8 & 1/8 & 1/16 & 1/16 & 1/8 & 1/16 & 1/16 \end{bmatrix}$$

$$H(X^2) = -\sum_{i=1}^{q^N} p(\overline{\alpha}_i)\log p(\overline{\alpha}_i)$$

$$= -\sum_{i_1=1}^{3}\sum_{i_2=1}^{3} p(a_{i_1})p(a_{i_2})\log\, p(a_{i_1})p(a_{i_2}) = 3 \text{ bit}$$

注意到，$H(X^2) = 2H(X)$。

4.5 离散信源

定理 4.6 离散无记忆信源 X 的 N 次扩展信源 X^N 的熵满足：

$$H(X^N) = N \times H(X)$$

- 对于离散平稳有记忆信源，我们有如下定义：

定义 4.12 长度为 N 的信源符号序列中平均每个信源符号所携带的信息量为：$H(X^N) = \frac{1}{N} H(X_1 X_2 \cdots X_N)$，称为平均符号熵。

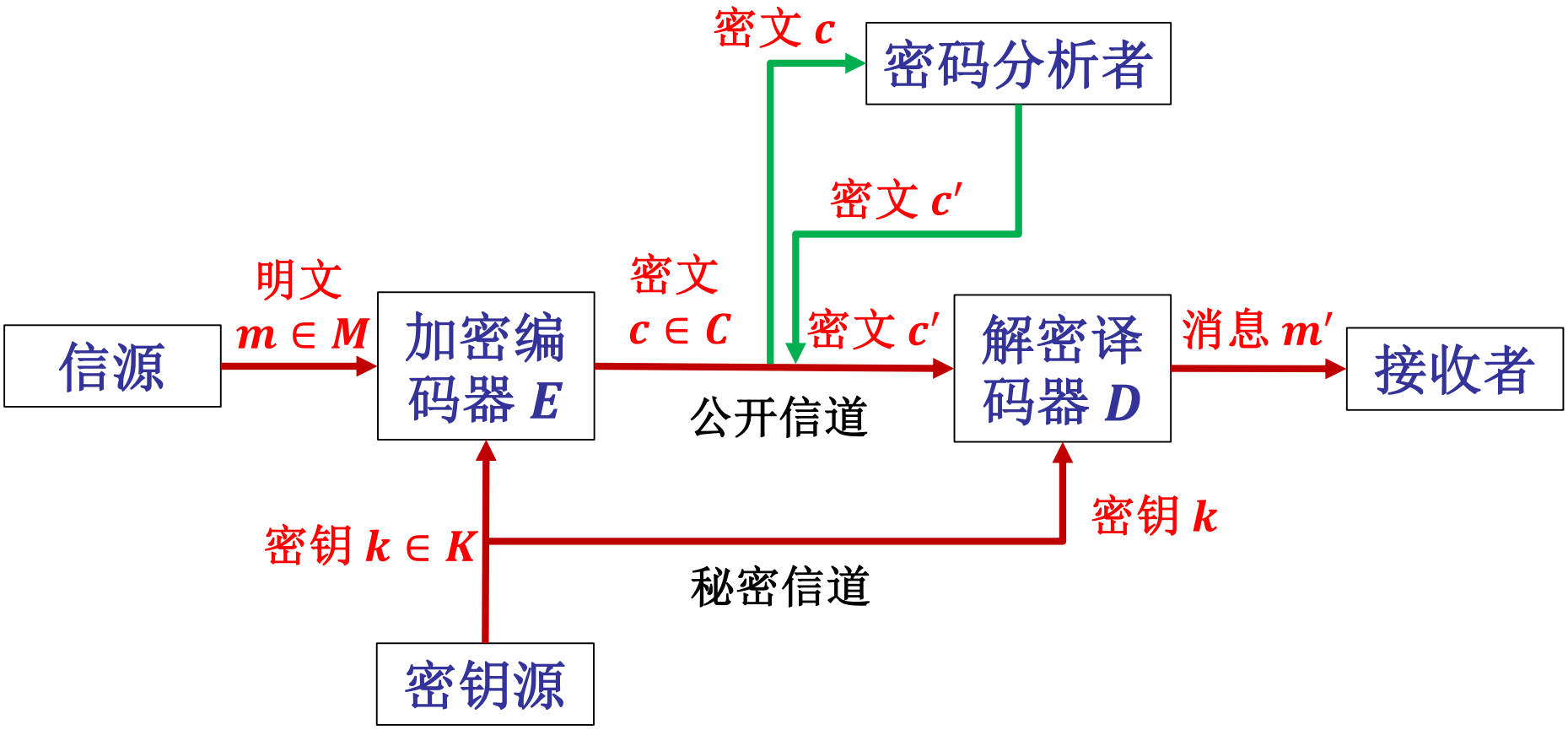
定义 4.13 信源 X 的极限符号熵定义为：

$$H_\infty(X) = \lim_{N\to\infty} H_N(X) = \lim_{N\to\infty} \frac{1}{N} H(X_1 X_2 \cdots X_N)$$

4.5 信息论在密码学中的应用

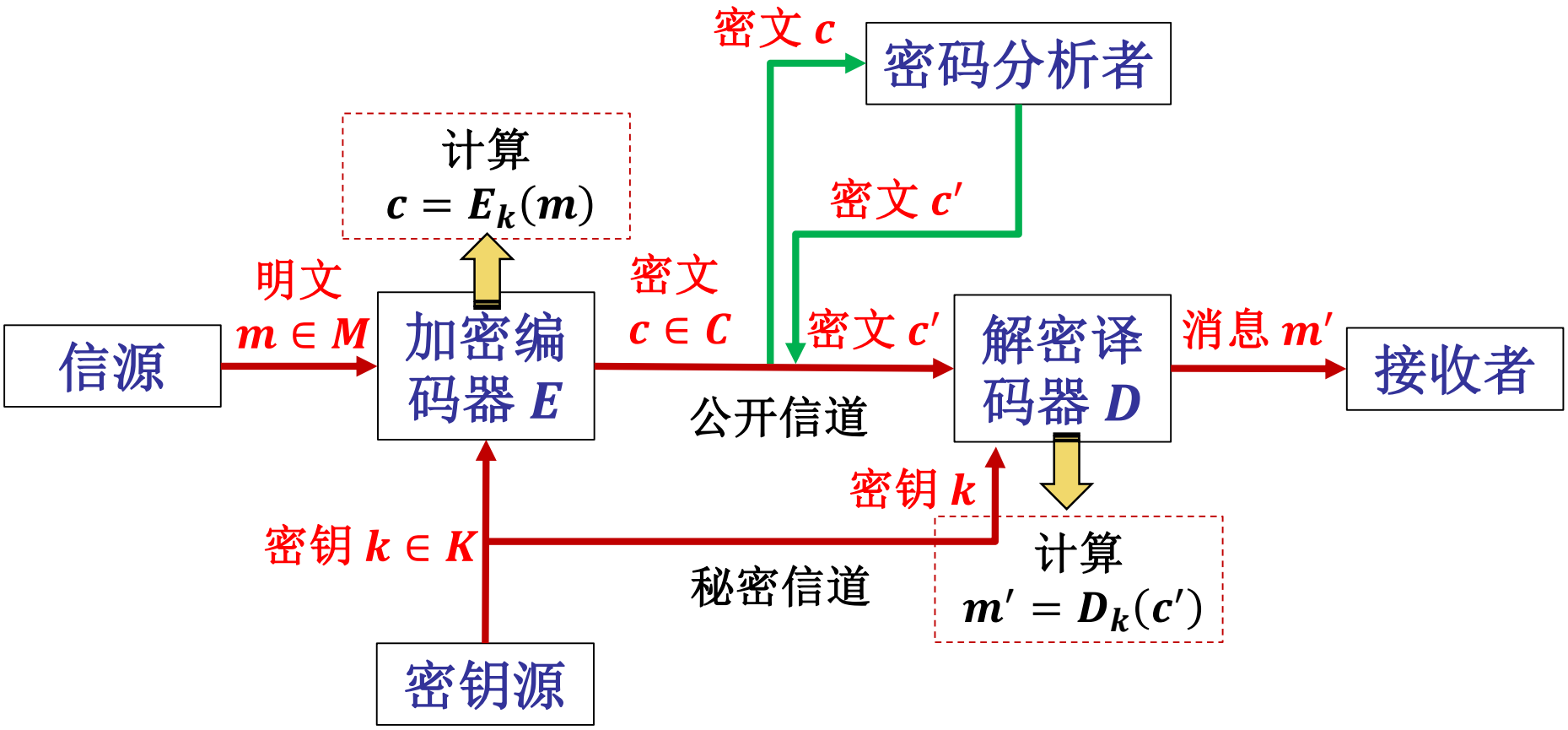
- 1948年，香农用概率论的观点对信息源、密钥源、接收和截获的密文进行数学描述和定量分析，提出了通用密码系统模型。
- M：信源的明文空间，C：密文空间，K：密钥空间。给定密钥 $k \in K$，加密编码 E_k 是 $M \to C$ 的一一映射，D_k 是 E_k 的逆变换，(M, C, K, E_k, D_k) 表示一个密码系统。
- 密码系统的两种安全性标准：
  - 理论安全性：破译者具有无限时间和无限计算资源下的抗破译能力。
  - 实用安全性：破译者仅有一定计算资源及其他实际限制下的抗破译能力。

4.5 信息论在密码学中的应用



通用密码系统模型

4.5 信息论在密码学中的应用



通用密码系统模型

4.5 信息论在密码学中的应用

令明文空间 $M=\{m_1,\dots,m_n\}$，密文空间 $C=\{c_1,\dots,c_n\}$，明文 $m_i\in M$ 出现的概率为 $p(m_i)$，密钥 $k\in K$ 出现的概率为 $p(k)$，密文 $c_j\in C$ 出现的概率为 $p(c_j)$。密钥是独立于明文随机选取的。收到密文 c_j 后明文 m_i 的条件概率为 $p(m_i\mid c_j)$。

定义 4.18 一个密码系统 (M,C,K,E_k,D_k) 称为完全保密系统，指对任意 $m_i\in M$, $c_j\in C$, $p(c_j)>0$ 有

$$p(m_i\mid c_j)=p(m_i)$$

即

$$H(M\mid C)=H(M)$$

也就是敌手截获一个密文不能得到任何明文信息。

4.5 信息论在密码学中的应用

定理 4.9 一个密码系统是完全保密的充分必要条件为对任意的 $m_i \in M, c_j \in C$ 有

$$p(c_j \mid m_i) = p(c_j)$$

也即

$$p(m_i, c_j) = p(m_i)p(c_j)$$

也就是明文和密文统计独立。

证：由贝叶斯公式 $p(c_j \mid m_i) = \frac{p(m_i|c_j)p(c_j)}{p(m_i)}$，所以

$$p(m_i \mid c_j) = p(m_i) \Leftrightarrow p(c_j \mid m_i) = p(c_j)$$

定理 4.10 一个密码系统是完全保密的充分必要条件为对任意的 $m_0, m_1 \in M, c \in C$ 有 $p(c \mid m_1) = p(c \mid m_0)$。

4.5 信息论在密码学中的应用

1917年，Vernam创造了一种完全保密的加密方法，称为“一次一密”，之后香农证明了“一次一密”的完全保密性。

“一次一密”加密方案：

1. 设整数 $l > 0$，明文空间 M，密钥空间 K 和密文空间 C 都等于 $\{0,1\}^l$（长度为 l 的二进制比特串集合）。
2. 密钥生成算法 Gen： 从 $\{0,1\}^l$ 中依均匀分布选择一个二进制比特串。
3. 加密算法 Enc：给定一个密钥 $k \in \{0,1\}^l$，一个明文 $m \in \{0,1\}^l$，输出 $c := k \oplus m$。
4. 解密算法 Dec：给定一个密钥 $k \in \{0,1\}^l$，一个密文 $c \in \{0,1\}^l$，输出 $m := k \oplus c$。

4.5 信息论在密码学中的应用

定理 4.11 “一次一密”是完全保密的加密体制。

定理 4.13（香农定理） 设一个密码系统的明文数目、密文数目和密钥数目相等，即 $|M| = |C| = |K| = n$，则该密码为完全保密的充分必要条件为：

1. 有且仅有一个密钥将每个给定的明文加密为每个给定的密文。
2. 所有的密钥选取是等概率的。

注意： 定理 4.13 给出了完全保密的完整特征，可以看出，密码体制的完全保密性与明文的概率分布无关。

4.5 信息论在密码学中的应用

定理 4.14 对密码系统 (M, C, K, E_k, D_k)，明文熵定义为

$$H(M) = -\sum_{m \in M} p(m) \log p(m)$$

密钥熵定义为

$$H(K) = -\sum_{k \in K} p(k) \log p(k)$$

令 X^t 表示 M 中所有长度为 t 的明文构成的集合，Y^n 表示 C 中所有长度为 n 的密文构成的集合，则在已知密文下对明文和密钥的疑义度为

$$H(X^t \mid Y^n) = -\sum_{m \in X^t, c \in Y^n} p(c, m) \log p(m \mid c)$$

$$H(K \mid Y^n) = -\sum_{k \in K, c \in Y^n} p(c, k) \log p(k \mid c)$$

4.5 信息论在密码学中的应用

- 疑义度 $H(X^t \mid Y^n)$ 和 $H(K \mid Y^n)$ 分别给出了截获长度为 n 的密文后，关于长度为 t 的明文和密钥的不确定度。
- 若 $N \geq n$，则

$$H(X^t \mid Y^n) \geq H\left(X^t \mid Y^N\right)$$

$$H(K \mid Y^n) \geq H\left(K \mid Y^N\right)$$

即随着截获密文的增加，获得关于明文或密钥的信息量会增加。

- 若 n 充分大，使得 $H(K \mid Y^n) = 0$，则此时可唯一确定密钥，这时的 n 在密码学中有重要意义。

4.5 信息论在密码学中的应用

定义 4.19 一个密码系统的**唯一解距离** N_0 定义为使 $H(K \mid Y^n) = 0$ 的最小正整数 n。

香农提出利用随机密码模型来给出 $H(K \mid Y^n) = 0$ 的近似估计。对 (M, C, K, E_k, D_k) 做如下假设。

1. 明文字符表和密文字符表均为 A，长度为 n。令 B^n 是有意义的明文集合，$\overline{B^n} = A^n - B^n$是无意义明文集合。当 $n \to \infty$时，$\overline{B^n}$ 中明文出现的概率可忽略不计。
2. 密钥为等概分布，即 $p(k) = |K|^{-1} = 2^{-H(K)}$ 。
3. 对 $k \in K$ 的加密变换 f_k 为 $A^n \to A^n$ 的一一映射，解密变换为 $\phi_k = f_k^{-1}$。
4. B^n 中每个明文等概出现。

4.5 信息论在密码学中的应用

在唯密文攻击下，我们假定密码分析者具有无穷的计算资源，并且知道明文是哪一种“自然语言”。密码分析者可能分析得到许多可能的密钥，但其中只有一个正确的密钥，其他的密钥都是不正确的，称为“伪密钥”。

例：假定密码分析者得到了一个密文字符串WNAJW，它是某个明文字符串由

$$c_i = m_i + k \bmod 26$$

得到的密文。容易算出，明文“river”（密钥 $k = 5$）和“arena”（密钥 $k = 22$）均对应密文“WNAJW”，但事实上只有一个密钥是正确的，另一个是伪密钥。

4.5 信息论在密码学中的应用

定理 4.15 若密码系统 (M, C, K, E_k, D_k) 将长度为 n 的明文 X^n 加密成长度为 n 的密文 Y^n，则

$$H(K \mid Y^n) = H(K) + H(X^n) - H(Y^n)$$

证：$H(X^n, K, Y^n) = H(K, Y^n) + H(X^n \mid K, Y^n)$

$$H(X^n, K, Y^n) = H(K, X^n) + H(Y^n \mid K, X^n)$$

所以，$H(K, Y^n) - H(K, X^n) = H(Y^n \mid K, X^n) - H(X^n \mid K, Y^n)$

由于 $H(Y^n \mid K, X^n) = H(X^n \mid K, Y^n) = 0$ （持密钥者可加解密）

从而，$H(K, Y^n) = H(K, X^n)$。又

$$H(K \mid Y^n) = H(K, Y^n) - H(Y^n) = H(K, X^n) - H(Y^n)$$

$$= H(K) + H(X^n) - H(Y^n)$$

最后一个等式由于明文 X^n 和密钥 K 相互独立。

4.5 信息论在密码学中的应用

- 记 $R_0 = \log |A|$，称为信源的绝对信息率，
- 记 $R_n = \log |B_n|/n$，称为信源的近似信息率，
- 记 $\lim\limits_{n\to\infty} R_n = H_\infty$，称为信源信息率。
- 记 $d_0 = R_0 - R_n$，称为信源的近似剩余度。

由假设条件知 $H(X^n) = nR_n$，所以

$$H(K \mid Y^n) = H(K) + nR_n - nR_0$$

（事实上，$H(Y^n) \le n\log |A| = nR_0$）

令 $H(K \mid Y^n) = 0$，得 $H(K) + nR_n - nR_0 = 0$，故

$$N_0 = n = \frac{H(K)}{R_0 - R_n}$$

当 $n \to \infty$ 时，可用 H_∞ 代替 R_n，故 $N_0 = \dfrac{H(K)}{R_0 - H_\infty}$。

4.5 信息论在密码学中的应用

- 我们已经证明了唯一解距离

$$N_0 = \frac{H(K)}{R_0 - H_\infty}$$

$R_0 - H_\infty$ 称为明文冗余度，上述公式表明明文冗余度是密码分析的基础，即

“明文冗余度越大，唯一解距离就越短”。

- 提高密码体制的安全性，应该尽量减少明文冗余度。减少明文冗余度的一种方法是在对

“加密之前先采用哈夫曼编码（Huffman）对明文进行压缩”

4.5 信息论在密码学中的应用

例 4.17 利用 DES 密码体制加密由“26 个英文字母加空格”组成的英文文本。明文分组长度是 64 bit，密钥长度 56 bit，认为 DES 是一种相当接近随机密码的系统。英文文本的信源信息率是 $H_\infty = 1.40$ bit。求唯一解距离 N_0。

解：已知 $H(K) = 56$ bit，$R_0 = \log 27 = 4.76$ bit，所以

$$N_0 \approx n = \frac{H(K)}{R_0 - H_\infty} = \frac{56}{4.76 - 1.40} = 16.96 = 17 \text{ bit}$$

4.5 信息论在密码学中的应用

注意：

- 唯一解距离 N_0 表明：当截获的密文长度大于 N_0 时，原则上可以唯一确定所使用的密钥，从而破译密码。
- 事实上，上述是在假定破译者拥有无限的计算能力，并充分利用信源的全部统计知识的前提下成立的。
- 在实际中，破译者收到各种限制，不能做到完全利用资源。
- 如果破译密码的代价超出破译信息的价值，或者破译成功所花的事件超出了所得的信息时效，则破译是无用的。
- 密码系统在时间、人力、物力等限制条件下安全性更有意义，称为实用安全性。

4.5 信息论在密码学中的应用

定义 如果对一个密码系统，

$$\lim_{n\to\infty} H(K \mid Y^n) \neq 0$$

则对该系统来说，截获更多的密文并不能消除关于密钥的不确定性，这种密码系统称为理想密码系统。