# 应用密码学（第三讲）— 信息论基础

林东岱

信息安全国家重点实验室

2022年9月

# 本节概要


1. 概率论基础
2. 完善保密性
3. 信息的度量（熵）
4. 伪密钥与唯一解距离
5. 乘积密码体制

# 第一部分

## 概率论基础

## 定义 1.1

样本空间$S$上的一个离散随机变量$\mathbf{X}$，用$P(\mathbf{X}=x)$表示随机变量取$x$时的概率，简记为$P(x)$， 对于任意的$x \in S$,则有$0 \leq P(\mathbf{X}=x) \leq 1$，而且: $\sum_{x \in S} P(x)=1$

## 定义 1.2

一个事件$E$是样本空间$S$的一个子集，事件发生的概率记为$P(E)$。特别的，当$E$是一个简单事件$x$时$P(E)=P(x)$。事件$E$发生的概率$P(E)$为：$P(E)=\sum_{x \in E} P(x)$.

## 定义 1.3

假设$\mathbf{X}$和$\mathbf{Y}$分别是定义在样本空间$S_1$和$S_2$上的随机变量。联合概率$P(x,y)$是$X$取$x$且 $Y$取$y$时的概率。条件概率$P(x|y)$表示当$\mathbf{Y}$取$y$时$\mathbf{X}$取$x$的概率。 如果对于任意的$x \in S_1$和$y \in S_2$, 都有$P(x,y)=P(x)P(y)$, 则称$\mathbf{X}$和$\mathbf{Y}$是统计独立的。

## 定理 1.1 (Bayes定理)

如果$p(y) > 0$，那么

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)}.$$

## 推论 1.1

$\mathbf{X}$和$\mathbf{Y}$是两个独立的随机变量，当且仅当对任何的$x, y$, 有 $p(x|y) = p(x)$ 成立.

## 定义 1.4

设$S$是一个样本空间，$\mathbf{X}$是$S$上的一个随机变量，且$\mathbf{X}$是一个从样本空间$S$到实数集$R$的函数； 对于每一个简单事件$x \in S$,$\mathbf{X}$分配一个实数$\mathbf{X}(x)$。$\mathbf{X}$的数学期望定义为:

$$E(\mathbf{X}) = \sum_{x \in S} \mathbf{X}(x)P(x)$$

# 本节概要


1 概率论基础

2 完善保密性

3 信息的度量（熵）

4 伪密钥与唯一解距离

5 乘积密码体制

# 第二部分

## 完善保密性

# 完善保密性 I

假设在密码系统$(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$中，明文$x$出现的先验概率为 $p_{\mathcal{P}}(x)$, 密钥$k$被使用的概率为$p_{\mathcal{K}}(k)$. 一般来说，密钥$k$是在知道明文之前选定的，因此我们假设$k$和$x$是相互独立的事件.

在知道明文和密钥概率分布的情况下，我们 很容易导出密文的概率分布如下：

**定理 2.1**

$$p_{\mathcal{P}}(x|y) = \frac{p_{\mathcal{P}}(x)\sum_{\{k:x=D_k(y)\}} p_{\mathcal{K}}(k)}{\sum_{\{k:y\in C(k)\}} p_{\mathcal{K}}(k)p_{\mathcal{P}}(D_k(y))},$$

其中$C(K)=\{e_k(x):x\in\mathcal{P}\}$.

**证明：** $\forall k\in\mathcal{K}$, 定义$C(k)=\{E_k(x)|x\in\mathcal{P}\}$, 那么$\forall y\in\mathcal{C}$, 我们有：

$$p_{\mathcal{C}}(Y=y)=\sum_{k:y\in C(k)} p_{\mathcal{K}}(K=k)p_{\mathcal{P}}(x=D_k(y))$$

# 完善保密性 II

同时，我们有

$$p_{\mathcal{C}}(Y=y|X=x)=\sum_{k:x=D_k(y)} p_{\mathcal{K}}(K=k)$$

所以

$$\begin{aligned} p_{\mathcal{P}}(x|y) &= \frac{p_{\mathcal{P}}(X=x)p_{\mathcal{C}}(Y=y|X=x)}{p_{\mathcal{C}}(Y=y)} \\ &= \frac{p_{\mathcal{P}}(x)\sum_{\{k:x=d_k(y)\}} p_{\mathcal{K}}(k)}{\sum_{\{k:y\in C(k)\}} p_{\mathcal{K}}(k)p_{\mathcal{P}}(d_K(y))}. \end{aligned}$$

□

# 计算实例 I

设$\mathcal{P}=\{a,b\}$, $p(a)=\frac{1}{4}, p(b)=\frac{3}{4}$; $\mathcal{K}=\{K_1,K_2,K_3\}$, $p(K_1)=\frac{1}{2}, p(K_2)=p(K_3)=\frac{1}{4}$; 设$\mathcal{C}=\{1,2,3,4\}$，并假设 $E_{K_1}(a)=1$, $E_{K_1}(b)=2$, $E_{K_2}(a)=2$, $E_{K_2}(b)=3$, $E_{K_3}(a)=3$, $E_{K_3}(b)=4$.

| | $a$ | $b$ |
|---|---|---|
| $K_1$ | 1 | 2 |
| $K_2$ | 2 | 3 |
| $K_3$ | 3 | 4 |

则

$$
\begin{aligned}
p_{\mathcal{C}}(1) &= \frac{1}{2}\times\frac{1}{4}=\frac{1}{8} \\
p_{\mathcal{C}}(2) &= \frac{1}{2}\times\frac{3}{4}+\frac{1}{4}\times\frac{1}{4}=\frac{7}{16} \\
p_{\mathcal{C}}(3) &= \frac{1}{4}\times\frac{3}{4}+\frac{1}{4}\times\frac{1}{4}=\frac{1}{4} \\
p_{\mathcal{C}}(4) &= \frac{1}{4}\times\frac{3}{4}=\frac{3}{16}
\end{aligned}
$$

# 计算实例 II

利用上面的定理，我们可以计算在收到一定密文下明文的条件概率分布如下：

$$
\begin{aligned}
p_{\mathcal{P}}(a|1) &= 1 & P_{\mathcal{P}}(b|1) &= 0 \\
p_{\mathcal{P}}(a|2) &= \tfrac{1}{7} & P_{\mathcal{P}}(b|2) &= \tfrac{6}{7} \\
p_{\mathcal{P}}(a|3) &= \tfrac{1}{4} & P_{\mathcal{P}}(b|3) &= \tfrac{3}{4} \\
p_{\mathcal{P}}(a|4) &= 0 & P_{\mathcal{P}}(b|4) &= 1
\end{aligned}
$$

只有密文$y=3$时，$p(a)=p(a/3), p(b)=p(b/3)$

**定义 2.1 (完善保密)**

一个密码系统$(\mathcal{P},\mathcal{C},\mathcal{K},E,D)$称为完善保密的，如果对任何的$x\in\mathcal{P}$和$y\in\mathcal{P}$，$p_{\mathcal{P}}(x|y)=p_{\mathcal{P}}(x)$成立．也就是说在接收到密文$y$ 的情况下，明文$x$的后验概率和其先验概率是相同的．

# 移位密码的完善保密性 I

## 定理 2.2

*假设移位密码系统中的26个密钥被等概率地使用．则对明文的任何概率分布，移位密码 都是完全保密的．*

**证明：** 因为$\mathcal{P}=\mathcal{C}=\mathcal{K}=\mathbf{Z}_{26}$，且对任何的 $0 \leq K \leq 25, x \in \mathbf{Z}_{26}$，$e_K(x)=x+K \bmod 26$，所以对 任何的$y \in \mathbf{Z}_{26}$，

$$\begin{aligned}
p_{\mathcal{P}}(y) &= \sum_{K \in \mathbf{Z}_{26}} p_{\mathcal{K}}(K) p_{\mathcal{P}}(d_K(y)) \\
&= \sum_{K \in \mathbf{Z}_{26}} \frac{1}{26} p_{\mathcal{P}}(y-K) \\
&= \frac{1}{26} \sum_{K \in \mathbf{Z}_{26}} p_{\mathcal{P}}(y-K) \\
&= \frac{1}{26} \sum_{y \in \mathbf{Z}_{26}} p_{\mathcal{P}}(y) \\
&= \frac{1}{26}.
\end{aligned}$$

# 移位密码的完善保密性 II

另外，由于对任何的明文$x$和密文$y$，满足$e_K(x)=y$的唯一密钥是$K=y-x \bmod 26$，所以

$$p_{\mathcal{C}}(y|x)=p_K(y-x \bmod 26)=\frac{1}{26}.$$

从而根据定理1.1我们有

$$\begin{aligned} p_{\mathcal{P}}(x|y) &= \frac{p_{\mathcal{P}}(x)p_{\mathcal{C}}(y|x)}{p_{\mathcal{C}}(y)} \\ &= \frac{p_{\mathcal{P}}(x)\frac{1}{26}}{\frac{1}{26}} \\ &= p_{\mathcal{P}}(x). \end{aligned}$$

定理得证。 □

上述定理说明了只要每次都用新的密钥加密明文字符，移位密码就是不可破的。

# 完善保密系统的性质 I

- $\forall x \in \mathcal{P}, y \in \mathcal{C}$, $p_{\mathcal{P}}(x|y) = p_{\mathcal{P}}(x) \Leftrightarrow p_{\mathcal{C}}(y|x) = p_{\mathcal{C}}(y)$.
- 不妨设$\forall y \in \mathcal{C}$, $p_{\mathcal{C}}(y) > 0$, 否则，我们可以从$\mathcal{C}$去掉那些从不被用的密文$y$.
  固定$x \in \mathcal{P}$. 则对每一 $y \in \mathcal{C}$, 我们有$p_{\mathcal{C}}(y|x) = p_{\mathcal{C}}(y) > 0$. 因此，对每一$y \in \mathcal{C}$, 一定存在一个$K \in \mathcal{K}$, 使得$e_K(x) = y$.
- $|\mathcal{K}| \geq |\mathcal{C}| \geq |\mathcal{P}|$, 也就是说， 密钥量一定要比密文量大。

**定理 2.3 (Shannon)**

假设$(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$是一密码系统，满足 $|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|$。则该密码系统具有完全保密性的充分必要条件是：密钥空间$\mathcal{K}$中的每一密钥都被等概率地使用，且对任何$x \in \mathcal{P}$和 $y \in \mathcal{C}$，存在唯一的密钥$K$使得$e_K(x) = y$.

# 完善保密系统的性质 II

**证明：** 假设给定的密码系统是完善保密的，那么对任意$x \in \mathcal{P}$和$y \in \mathcal{C}$，必存在一个密钥$K$使得$e_K(x) = y$. 所以我们有

$$|\mathcal{C}| = |\{e_K(x) : K \in \mathcal{K}\}| \leq |\mathcal{K}|.$$

据题设，$|\mathcal{C}| = |\mathcal{K}|$, 所以不存在两个密钥$K_1, K_2$, 使得$e_{K_1}(x) = e_{K_2}(x) = y$. 因此，对任何 $x \in \mathcal{P}$和 $y \in \mathcal{C}$，存在唯一的密钥$K$使得$e_K(x) = y$.

设$n = |\mathcal{K}|, \mathcal{P} = \{x_i : 1 \leq i \leq n\}$, $y \in \mathcal{C}$是一固定元素。则我们可以 命名密钥$K_1, K_2, \cdots, K_n$, 使得$e_{K_i}(x_i) = y, 1 \leq i \leq n$. 利用定理1.1, 我们有

$$p_{\mathcal{P}}(x_i|y) = \frac{p_{\mathcal{C}}(y|x_i)p_{\mathcal{P}}(x_i)}{p_{\mathcal{C}}(y)} = \frac{p_{\mathcal{K}}(K_i)p_{\mathcal{P}}(x_i)}{p_{\mathcal{C}}(y)}.$$

根据完全保密性条件我们有$p_{\mathcal{P}}(x_i|y) = p_{\mathcal{P}}(x_i)$, 所以$p_{\mathcal{K}}(K_i) = p_{\mathcal{C}}(y)$对所有的$1 \leq i \leq n$成立。 因此，每一密钥都被等概率地使用。 $\square$

# 维尔南(Gilbert Vernam)一次一密密码体制

## Vernam一次一密密码体制（1917）

取$n \geq 1$为一整数，$\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbf{Z}_2)^n$. 对任意 $K \in (\mathbf{Z}_2)^n$, $e_K(x)$定义为$K$和$x$向量的模2加。即，如果 $x = (x_1, \cdots, x_n)$, $K = (K_1, \cdots, K_n)$, 那么

$$e_K(x) = (x_1 + K_1, \cdots, x_n + K_n) \bmod 2.$$

解密算法和加密算法是相同的，即如果$y = (y_1, \cdots, y_n)$, 那么

$$d_K(y) = (y_1 + K_1, \cdots, y_n + K_n) \bmod 2.$$

## 推论 2.1

*Vernam一次一密密码体制具有完善保密性*

# 本节概要


1 概率论基础

2 完善保密性

3 信息的度量（熵）

4 伪密钥与唯一解距离

5 乘积密码体制

# 第三部分

## 信息的度量（熵）

# 不确定性和信息

从几个例子谈起...

- 明天太阳将从东边升起。

# 不确定性和信息

从几个例子谈起...

- 明天太阳将从东边升起。
- 明天北京会下雨。

# 不确定性和信息

从几个例子谈起...

- 明天太阳将从东边升起。
- 明天北京会下雨。
- 中国足球队获得了世界杯冠军。

# 不确定性和信息

从几个例子谈起...

- 明天太阳将从东边升起。
- 明天北京会下雨。
- 中国足球队获得了世界杯冠军。

信息度量的定义要满足的三个性质：

1. 信息量应是概率分布的连续函数；
2. 对有$n$个等概率结果的试验，信息量应是$n$的单调上升函数；
3. 一个试验分成相继的两个试验时，未分之前的信息量应是既分之后的加权和。

# 不确定性和信息

- 单符号离散信源：如果信源发出的消息是离散的、有限或无限可列的符号或数字， 且一个符号代表一条完整的消息，则称这种信源为单符号离散信源。
- 信源空间：若信源的输出是随机事件x，其出现概率为p(x)，则它们所构成的集合， 称为信源的概率空间或简称为信源空间。

**定义 3.1 (自信息)**

设有一个离散信源发出的符号消息的集合为$X=\{x_1,x_2,x_3,\cdots,x_n\}$,其概率分布为 $P=\{p(x_1),p(x_2),p(x_3),\cdots,p(x_n)\}$, 一个随机事件$x_i$的自信息定义为其出现概率对数的负值, 即$I(x_i)=-\log p(x_i)$.

$I(x)$的值取决于对数的底数，通常取为2或$e$. 当以2为底时，单位为**比特**(bit), 而当以$e$为底时，单位为**奈特**(nat).

## 例 3.1

一个等概率的二进制随机序列，求任一码元的自信息量。

解：因为二进制序列只有0和1，而且等概率$P(0)=P(1)=1/2$ 所以有$I(0)=I(1)=-\log_2(1/2)=\log_2 2=1$ **bit**

## 例 3.2

对于$n$位的2进制数，假设每一符号的出现完全随机且概率相等，求任一符号的自信息量。

解：因为对于一个$n$位的2进制数的每一位可以从$0,1$两个数字中任取一个，所以有$2^n$个等概率的可能组合。所以，

$$p(x_i)=\frac{1}{2^n}$$

$$I(x_i)=-\log_2 p(x_i)=-\log_2(\frac{1}{2^n})=n\ \textbf{bit}$$

## 定义 3.2 (联合自信息)

若有两个消息$x, y$同时出现，其发生的概率可用联合概率$p(x, y)$表示，这时的**联合自信息量**定义为$I(x, y) = -\log p(x, y)$.

当$x$和$y$相互独立时，有$p(x, y) = p(x)p(y)$,那么根据对数运算的性质，有$I(x, y) = I(x) + I(y)$。

## 定义 3.3 (条件自信息量)

我们用$p(x|y)$表示事件$y$发生的情况下事件$x$发生的条件概率，则在事件$y$发生的情况下事件$x$的**条件自信息量**定义为：$I(x|y) = -\log p(x|y)$.
$x$和$y$的**互信息量**定义为$I(x; y) = I(x) - I(x|y)$.

容易推出:

$$I(x|y) = -\log p(x|y) = -\log \frac{p(x,y)}{p(y)} \quad = -\log p(x, y) + \log p(y) = I(x, y) - I(y).$$

## 定义 3.4 (信息熵)

设$\mathbf{X}$是一按照概率分布$p(\mathbf{X})$在某一有限集取值的随机变量,并设$\mathbf{X}$所有可能的取值为$x_i, 1 \leq i \leq n$, 我们定义

$$
\begin{aligned}
H(\mathbf{X}) = E(I(\mathbf{X})) &= \sum_{i=1}^{n} p(\mathbf{X} = x_i) I(\mathbf{X} = x_i) \\
&= -\sum_{i=1}^{n} p(\mathbf{X} = x_i) \log p(\mathbf{X} = x_i)
\end{aligned}
$$

称为事件$\mathbf{X}$的熵。其中我们约定$\log 0 = 0$.

在上面的定义中，对数log的底可以任意选取。容易看出，不同的底对熵值的影响只是一个常数因子。

## 例 3.3

设信源符号集$X = \{x_1, x_2, x_3, x_4\}$每个符号发生的概率分别为$p(x_1) = 1/2, p(x_2) = 1/4, p(x_3) = 1/8, p(x_4) = 1/8$. 则信源$X$的熵为$H(X) = 1.78$。

# 延伸阅读

## 信源编码定理

设$X$为离散无记忆信源的字母组合，熵为$H(X)$，而且其输出符号$x_k$发生的概率为$p(x_k), k=1,2,\cdots,L$. 则一定可以构造满足前缀条件的码，其平均码长$\overline{R}$满足：

$$H(X) \leq \overline{R} < H(X)+1.$$

考虑一个以概率$p_1=0.5, p_2=0.3, p_3=0.1, p_4=0.1$产生四个符号的信源$X$. 该信源的熵为

$$H(X) = -\sum_{k=1}^{4} p_k \log_2 p_k = 1.685(\mathrm{bit})$$

假定我们使用前缀码$\{0,10,110,111\}$, 则平均码字长度$\overline{R}$为

$$\overline{R} = \sum_{k=1}^{4} n(x_k)P(x_k) = 1\times 0.5 + 2\times 0.3 + 3\times 0.1 + 3\times 0.1 = 1.7(\mathrm{bit})$$

## 定义 3.5

**联合熵**是联合符号集合$X,Y$上的每个元素对$x_i,y_j$的自信息量的概率加权统计平均值，定义为：

$$H(X,Y)=\sum_{i,j}p(x_i,y_j)I(x_i,y_j)=-\sum_{i,j}p(x_i,y_j)\log p(x_i,y_j)$$

## 定义 3.6

假设**X**和**Y**是两个随机变量，我们定义**条件熵**$H(X|Y)$为：

$$\begin{aligned}H(X|Y)&=\sum_j p(y_j)H(X|y_j)\\&=\sum_{i,j}p(y_j)p(x_i|y_j)I(x_i|y_j)\\&=\sum_{i,j}p(x_i,y_j)I(x_i|y_j)\end{aligned}$$

条件熵表示发生了事件**Y**后，**X**还保留的信息量。

# 熵的性质

## 定义 3.7 (凸函数)

我们称一个实值函数$f$在区间$I$上是凸的，如果 $f(\frac{x+y}{2}) \geq \frac{f(x)+f(y)}{2}$ 对任意的$x, y \in I$成立. $f$被称作严格凸的如果 $f(\frac{x+y}{2}) > \frac{f(x)+f(y)}{2}$ 对任意的$x, y \in I, x \neq y$成立.

## 定理 3.1 (Jessen不等式)

假设$f$是区间$I$上的一个严格凸的连续函数, $\sum_{i=1}^{n} a_i = 1$, $a_i > 0, 1 \leq i \leq n$.那么

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

其中$x_i \in I, 1 \leq i \leq n$. 进一步，等式成立当且仅当$x_1 = x_2 = \cdots = x_n$.

## 定理 3.2 (非负性)

$$H(X)=H(x_1,x_2,\cdots,x_n)\geq 0$$

其中等号只有在某$p(x_i)=1$,其他$p(x_j)=0$时成立。

## 定理 3.3

假设$\mathbf{X}$是一具有概率分布$p_1,p_2,\cdots,p_n$的随机变量，其中$p_i>0,1\leq i\leq n$. 那么我们有$H(\mathbf{X})\leq\log_2 n$, 等式成立当且仅当$p_i=\frac{1}{n}$, $1\leq i\leq n$.

**证明：** 应用Jensen不等式，我们有

$$\begin{aligned} H(\mathbf{X}) &= -\sum_{i=1}^{n} p_i\log_2 p_i=\sum_{i=1}^{n} p_i\log_2\frac{1}{p_i} \\ &\leq \log_2\sum_{i=1}^{n}(p_i\times\frac{1}{p_i})=\log_2 n. \end{aligned}$$

进一步，等式成立当且仅当$p_i=\frac{1}{n}$, $1\leq i\leq n$. □

## 定理 3.4

$H(\mathbf{X},\mathbf{Y}) \leq H(\mathbf{X}) + H(\mathbf{Y})$, 等式成立当且仅当$\mathbf{X}$和$\mathbf{Y}$ 是相互独立的事件.

**证明：** 假设**X**的取值为$x_i, 1 \leq i \leq n$, **Y**的取值为$y_j, 1 \leq j \leq m$, 则

$$p(x_i) = \sum_{j=1}^{m} p(x_i, y_j), 1 \leq i \leq n$$

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j), 1 \leq j \leq m$$

所以

$$\begin{aligned} H(X) + H(Y) &= -\left[\sum_{i=1}^{n} p(x_i)\log p(x_i) + \sum_{j=1}^{m} p(y_j)\log p(y_j)\right] \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i)p(y_j) \end{aligned}$$

由联合熵的定义和Jensen不等式可知：

$$
\begin{aligned}
&H(X,Y)-H(X)-H(Y)= \\
&=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log\frac{1}{p(x_i,y_j)}+\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log p(x_i)p(y_j) \\
&=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log\frac{p(x_i)p(y_j)}{p(x_i,y_j)} \\
&\le\log\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)=\log 1=0
\end{aligned}
\tag{1}
$$

上面等式成立当且仅当对任意的$1\le i\le n, 1\le j\le m$,

$$\frac{p(x_i)p(y_j)}{p(x_i,y_j)}=C$$

其中$C$为某一常数。因为

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

所以$C=1$, 即对任意的$1\le i\le n, 1\le j\le m$,

$$p(x_i,y_j)=p(x_i)p(y_j).$$

因此，当且仅当$X$与$Y$相互独立时，式(1)中等号成立。 □

**定理 3.5**

$H(\mathbf{X},\mathbf{Y})=H(\mathbf{Y})+H(\mathbf{X}|\mathbf{Y})=H(\mathbf{X})+H(\mathbf{Y}|\mathbf{X})$.

**证明：**

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

同理可证$H(X,Y)=H(X)+H(Y|X)$. □

## 推论 3.1

$H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$, 等式成立当且仅当$\mathbf{X}$和$\mathbf{Y}$ 是相互独立的事件.

## 例 3.4

在第9页的计算实例中，假设$P$是明文空间上的随机变量，$C$是密文空间上的随机变量， $K$是密钥空间上的随机变量。下面我们来计算$H(P), H(K), HC)$:

$$
\begin{aligned}
H(P) &= -p(a)\log_2 p(a) - p(b)\log_2 p(b) \\
&= -(1/4)\log_2(1/4) - (3/4)\log_2(3/4) = 0.81 \\
H(K) &= -p(k_1)\log_2 p(k_1) - p(k_2)\log_2 p(k_2) - p(k_3)\log_2 p(k_3) \\
&= -(1/2)\log_2(1/2) - (1/4)\log_2(1/4) - (1/4)\log_2(1/4) \\
&= 1.5 \\
H(C) &= -p(1)\log_2 p(1) - p(2)\log_2 p(2) - p(3)\log_2 p(3) - p(4)\log_2 p(4) \\
&= -\tfrac{1}{8}\log_2(\tfrac{1}{8}) - \tfrac{7}{16}\log_2(\tfrac{7}{16}) - \tfrac{1}{4}\log_2(\tfrac{1}{4}) - \tfrac{3}{16}\log_2(\tfrac{3}{16}) \\
&= 1.85
\end{aligned}
$$

# 本节概要

1 概率论基础

2 完善保密性

3 信息的度量（熵）

4 伪密钥与唯一解距离

5 乘积密码体制

# 第四部分

## 伪密钥与唯一解距离

## 定理 4.1 (密钥的爱昧度)

设$(\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$是一密码系统. 则

$$H(\mathbf{K}|\mathbf{C}) = H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C}).$$

**证明：** 首先我们有

$$H(\mathbf{P}, \mathbf{K}, \mathbf{C}) = H(\mathbf{P}|\mathbf{K}, \mathbf{C}) + H(\mathbf{K}, \mathbf{C})$$

$$H(\mathbf{P}, \mathbf{K}, \mathbf{C}) = H(\mathbf{C}|\mathbf{K}, \mathbf{P}) + H(\mathbf{K}, \mathbf{P})$$

由于密钥和明文唯一确定密文，密钥和密文唯一确定明文，所以

$$H(\mathbf{P}|\mathbf{K}, \mathbf{C}) = 0$$

$$H(\mathbf{C}|\mathbf{K}, \mathbf{P}) = 0$$

注意到$\mathbf{K}$和$\mathbf{P}$是相互独立的，因此我们有

$$\begin{aligned} H(\mathbf{K}|\mathbf{C}) &= H(\mathbf{K},\mathbf{C}) - H(\mathbf{C}) \\ &= H(\mathbf{K},\mathbf{P}) - H(\mathbf{C}) \\ &= H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C}). \end{aligned}$$

条件熵$H(K/C)$称为密钥暧昧度，它表示在已知密文的情况下能泄露多少密钥信息的一种度量。 □

在一个自然语言中，连续的字母之间并不是相互 独立的，正是连续的字母之间的这种相关性降低了自然语言的熵。

**定义 4.1 (自然语言的熵)**

设$L$是一自然语言．自然语言$L$的熵$H_L$定义为：

$$H_L = \lim_{n\to\infty} \frac{H(\mathbf{P}^n)}{n}$$

其中，$\mathbf{P}^n$为长为$n$个字符的明文的全体。

# 自然语言的冗余度

自然语言的熵刻画了自然语言中一个字符所包含的平均信息量. 以英语为例，$H_L$介于 1.0和1.5之间，也就是说，在英语中一个字符所包含的平均信息量只有1个到1.5比特. 但如果每个字符都是等概率使用的话，每个字符所包含的信息量应为$\log_2 |\mathcal{P}|$. 我们把

$$D = \log_2 |\mathcal{P}| - H_L$$

称为语言的冗余度。语言的冗余度刻画了语言中每个字符所包含的冗余信息量。以英语的语言熵为 1.3为例，其冗余度为3.4/字符. 也就是说每个英语字符包含了3.4比特的冗余信息.

假设$\mathbf{x}$是密码系统$(\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$中的一明文串，$\mathbf{y}$是用某一固定的密钥加密$\mathbf{x}$后得到的密文串。我们考虑唯密文攻击，并假设攻击者具有无限的 计算资源并且知道明文是某一自然语言，一般来说，攻击者可以找到多个密钥，每一个密钥都能 把密文$\mathbf{y}$解密成有意义的明文串，但其中只有一个是正确的. 我们把其他不正确的， 但能把密文解密成有意义的明文的那些密钥称为**伪密钥**. 我们希望利用语言的熵或多余度 推导出伪密钥个数的下界.

我们用$\mathbf{C}^n$表示长度为$n$的密文构成的随机变量. 设$\mathbf{y} \in \mathbf{C}^n$, 定义

$$K(\mathbf{y}) = \{K \in \mathcal{K} : \exists \mathbf{x} \in \mathcal{P}^n, p_{\mathcal{P}^n}(\mathbf{x}) > 0, e_K(\mathbf{x}) = \mathbf{y}\}.$$

也就是说，$K(\mathbf{y})$是所有可将某一长度为$n$的有意义的明文串加密成$\mathbf{y}$的密钥的 全体组成的集合, 即给定密文串$\mathbf{y}$的条件下，所有可能的密钥的集合. 如果$\mathbf{y}$是 接收到的密文串，那么伪密钥的个数是$|K(\mathbf{y})| - 1$, 因为其中只有一个是正确的密钥. 所以，伪密钥个数的均值为

$$\begin{aligned} \bar{s}_n &= \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})(|K(\mathbf{y})| - 1) \\ &= \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| - \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y}) \\ &= \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| - 1. \end{aligned}$$

根据定理4.1，我们有

$$H(\mathbf{K}|\mathbf{C}^n) = H(\mathbf{K}) + H(\mathbf{P}^n) - H(\mathbf{C}^n).$$

同时，在$n$非常大时$H(\mathbf{P}^n)$我们可以估算如下

$$H(\mathbf{P}^n) \approx nH_L = n(\log_2 |\mathcal{P}| - D).$$

显然$H(\mathbf{C}^n) \leq n\log_2 |\mathcal{C}|$, 所以在$|\mathcal{C}| = |\mathcal{P}|$时，我们有

$$\begin{aligned} H(\mathbf{K}|\mathbf{C}^n) &\approx H(\mathbf{K}) + n(\log_2 |\mathcal{P}| - D) - H(\mathbf{C}^n) \\ &\geq H(\mathbf{K}) + n(\log_2 |\mathcal{P}| - D) - n\log_2 |\mathcal{C}| \\ &= H(\mathbf{K}) - nD. \end{aligned}$$

另一方面，我们有

$$\begin{aligned} H(\mathbf{K}|\mathbf{C}^n) &= \textstyle\sum_{\mathbf{y}\in\mathcal{C}^n} p(\mathbf{y})H(\mathbf{K}|\mathbf{y}) \\ &\leq \textstyle\sum_{\mathbf{y}\in\mathcal{C}^n} p(\mathbf{y})\log_2 |K(\mathbf{y})| \\ &\leq \log_2 \textstyle\sum_{\mathbf{y}\in\mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| \\ &= \log_2(\bar{s}_n + 1), \end{aligned}$$

所以

$$\log_2(\bar{s}_n + 1) \geq H(\mathbf{K}) - nD.$$

$$\bar{s}_n \geq 2^{H(\mathbf{K})-nD} - 1.$$

这样我们得到

## 定理 4.2

设$(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$是一密码系统，$|\mathcal{C}| = |\mathcal{P}|$。那么 对于一个给定的充分长的密文串（设其长度为$n$)，伪密钥的期望个数满足

$$\bar{s}_n \geq 2^{H(\mathbf{K})-nD} - 1. \quad (2)$$

在不等式(2)中，令$\bar{s}_n = 0$, 我们可得有关密文长度的一个近似值 $n_0 = \frac{H(\mathbf{K})}{D}$, 当截获的密文量大于$n_0$时，原则上可以破译该密码， 当截获的密文量小于$n_0$时，就可能存在多种可能的密钥解，密码分析者无法从中确定 哪一个是正确的。关于具有上述性质的$n_0$, 我们有如下定义

## 定义 4.2

一个密码系统的**唯一解距离**定义为使得伪密钥的期望数等于0的密文平均长度，记为$n_0$, 即在给定足够的计算时间下，分析者能唯一地计算出密钥所需要的密文的平均量。

唯一解距离给出了用来估计强力攻击密码系统时，要解出具有唯一有意义的明文可能需 要的最短的密文数。当然，密码系统的唯一解距离只是一个理论值，并不能做出确切的 预测，只能给出概率性的预测，一般来说，破译密码系统所需要的密文量都远大于唯一 解距离给出的密文量。对一个密码系统来说，唯一解距离越大，密码系统就越好。

唯一解距离与语言的冗余度成反比，当冗余度接近于0时，即使一个简单的密码系统， 对于唯密文攻击也是不可破的。C. E. Shannon将一个唯一解距离无限大的密码系统 称为理想安全的密码系统。理想安全的密码系统不一定是绝对安全的，但绝对安全的 密码系统一定是理想安全的。若一个密码系统具有理想安全性，则再好的密码分析员 也不能确定恢复的明文就是真正的明文。

# 本节概要

1 概率论基础
2 完善保密性
3 信息的度量（熵）
4 伪密钥与唯一解距离
5 乘积密码体制

# 第五部分

## 乘积密码体制

假设$\mathcal{C}=\mathcal{P}$，这样的密码体制称为满的。

设$\mathbf{S_1}=(\mathcal{P},\mathcal{P},\mathcal{K}_1,E_1,D_1)$和 $\mathbf{S_2}=(\mathcal{P},\mathcal{P},\mathcal{K}_2,E_2,D_2)$是两个满的 密码体制。则$\mathbf{S_1}$和$\mathbf{S_2}$的乘积$\mathbf{S_1}\times\mathbf{S_2}$定义为密码系统

$$(\mathcal{P},\mathcal{P},\mathcal{K}_1\times\mathcal{K}_2,E,D).$$

对乘积系统中的每一密钥$K=(K_1,K_2)\in\mathcal{K}_1\times\mathcal{K}_2$, 加密 规则$e_K$定义为

$$e_K(x)=e_{(K_1,K_2)}(x)=e_{K_2}(e_{K_1}(x)),$$

解密规则$d_K$定义为

$$d_K(x)=d_{(K_1,K_2)}(x)=d_{K_1}(d_{K_2}(x)).$$

即先用$e_{K_1}$加密，再用$e_{K_2}$加密，而解密正好反过来。

显然，关于密钥概率分布我们有

$$p_{\mathcal{K}}(K_1,K_2)=p_{\mathcal{K}_1}(K_1)\times p_{\mathcal{K}_12}(K_2),$$

且

$$(\mathbf{S_1} \times \mathbf{S_2}) \times \mathbf{S_3} = \mathbf{S_1} \times (\mathbf{S_2} \times \mathbf{S_3}).$$

即密码系统的乘积运算是结合的。对两个密码系统$\mathbf{S_1}$和$\mathbf{S_2}$， 如果$\mathbf{S_1} \times \mathbf{S_2} = \mathbf{S_2} \times \mathbf{S_3}$，我们则称$\mathbf{S_1}$和$\mathbf{S_2}$ 是**可交换的**。有时我们也把$\mathbf{S} \times \mathbf{S}$记作$\mathbf{S}^2$. 类似地，我们也可以定义$\mathbf{S}^n$, 这样的密码我们称为**迭代密码**。一个密码体制称为幂等的，如果$\mathbf{S}^2 = \mathbf{S}$。

- 混乱(confusion): 用于掩盖明文和密文之间的关系。这可以挫败通过研究密文以获取冗余度和 统计模式的企图。做到这点最容易的方法是通过代替，简单的如将一个确定的明文 字符代替成一个密文字符，复杂的则可以将一个长的明文分组替代成一个不同的 密文分组，并且替代的机制随明文或密钥中的每一位发生变化。
- 扩散(difussion): 就是明文冗余度分散到密文中使之分散开来，使密码分析者寻求这些冗余度 更加困难。产生扩散最简单的方法就是通过换位，也称为置换。

# 应用密码学作业（二） I

1. 考虑一个密码体制$M = \{a, b, c\}, K = \{k_1, k_2, k_3\}$和$C = \{1, 2, 3, 4\}$。假设加密矩阵为:

| | $a$ | $b$ | $c$ |
|---|---|---|---|
| $k_1$ | 2 | 3 | 4 |
| $k_2$ | 3 | 4 | 1 |
| $k_3$ | 1 | 2 | 3 |

已知密钥概率分布为：$p(k_1) = 1/2, p(k_2) = p(k_3) = 1/4$，且明文概率分布为$p(a) = 1/3, p(b) = 8/15, p(c) = 2/15$， 计算$H(M)$, $H(K)$, $H(C)$, $H(M|C)$, $H(K|C)$.

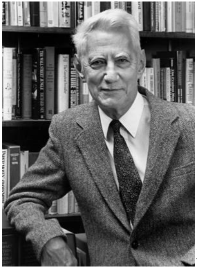
2. 考虑一个密码系统$(P, C, K, E, D)$。
   a) 说明为什么$H(P, K) = H(C, P, K) = H(P) + H(K)$.
   b) 假设这个系统具有完善保密性。证明$H(C, P) = H(C) + H(P)$和$H(C) = H(K) - H(K|C, P)$.

# 应用密码学作业（二） II

c) 假设这个系统具有完善保密性，并且对每一个明文密文对，最多只有一个相应的密钥能够加密。 证明$H(C) = H(K)$.

3. 假设$S_1$是移位密码（密钥等概率），$S_2$是密钥满足概率分布$p_k$（不必是等概率的）的移位密码。 证明$S_1 * S_2 = S_1$(这里用等号不一定准确，请思考什么叫相等或等价，给出你的定义并证明之)。

---

香农(Claude Elwood Shannon)(1916-2001)的主要贡献是创立了经典信息论。1948年他在《贝尔系统技术杂志》上发表"通信的数学理论"。 这篇分两期刊出、长达80余页的文章成了信息论的开端。