第 2 章　Shannon 理论

2.1　引言

1949 年，Claude Shannon 在 *Bell Systems Technical Journal* 上发表了题为“Communication Theory of Secrecy Systems”的论文。这篇论文对密码学的研究产生了巨大的影响。在本章中，我们讨论了若干 Shannon 的思想。首先让我们看看几个评价密码体制安全性的不同方法。现在定义几个最常用的准则。

计算安全性(computational security)

这种度量涉及的是破译一个密码体制所做的计算上的努力。如果使用最好的算法破译一个密码体制至少需要 N 次操作，这里的 N 是一个特定的非常大的数字，我们可以定义这个密码体制是计算安全的。问题是没有一个已知的实际的密码体制在这个定义下可以被证明是安全的。实际中，人们经常通过几种特定的攻击类型来研究计算上的安全性，例如穷尽密钥搜索攻击。当然对一种类型的攻击是安全的，并不表示对其他类型的攻击也是安全的。

可证明安全性(provable security)

另外一种方法是通过归约的方式为安全性提供证据。换句话讲，如果可用某一具体的方法破译一个密码体制，那么就有可能有效地解决某一被认为困难的经过深入研究的数学问题。例如，可以证明这样一类命题：如果给定的整数 n 是不可分解的，那么给定的密码体制是安全的。我们称这种类型的密码体制是可证明安全的。但是必须注意的是，这种方法只是说明了安全和某一个问题是相关的，并没有完全证明是安全的。这和证明一个问题是 NP 完全的有些类似：证明给定的问题至少和任何其他的 NP 完全问题的难度是一样的，但是并没有完全证明这个问题的计算难度。

无条件安全性(unconditional security)

这种度量考虑的是对攻击者 Oscar 的计算量没有限制时的密码体制的安全性。即使提供了无限的计算资源，也是无法被攻破的，我们定义这种密码体制是无条件安全的。

在讨论密码体制的安全性的时候，我们同时也规定了正在考虑的攻击类型。例如，在第 1 章中，我们说移位密码、代换密码和维吉尼亚密码对唯密文攻击都不是计算上安全的(如果给定了足够长的密文)。

我们将在 2.3 节研究一个对唯密文攻击是无条件安全的密码体制。这个理论从数学上证明了：如果给定的密文足够短，某些密码体制是安全的。例如，可以证明，如果只有单个的明文用给定的密钥加密，移位密码和代换密码都将是无条件安全的。类似地，如果密钥用于加密一个单位的明文(由 m 个字母组成)，使用密钥字长度为 m 的维吉尼亚密码是无条件安全的。

2.2　概率论基础

密码体制的无条件安全性当然不能用计算复杂性的观点来研究，因为我们允许计算时间是无限的。研究无条件安全性的合适框架是概率论。我们只需要概率论的基本事实；现在复习一下其主要内容。首先定义随机变量的概念。

定义 2.1　一个离散的随机变量，比方说 $\boldsymbol{X}$[①]，由有限集合 X 和定义在 X 上的概率分布组成。我们用 $\Pr[\mathbf{X}=x]$ 表示随机变量 $\mathbf{X}$ 取 x 时的概率。如果随机变量是固定的，我们有时缩写成 $\Pr[x]$。对任意的 $x\in X$，有 $0\leqslant \Pr[x]\leqslant 1$，并且

$$\sum_{x\in X}\Pr[x]=1$$

举一个例子，我们可以把抛硬币看成是定义在集合 {heads, tails} 上的随机变量。相关的概率分布可以是 $\Pr[\text{heads}]=\Pr[\text{tails}]=1/2$。

假设我们有定义在集合 X 上的随机变量 $\mathbf{X}$，$E\subseteq X$。$\mathbf{X}$ 在子集 E 上取值的概率可由下式计算：

$$\Pr[x\in E]=\sum_{x\in E}\Pr[x] \tag{2.1}$$

子集 E 通常称为事件。

例 2.1　假设随机抛一对骰子。我们可以用一个随机变量 $\mathbf{Z}$ 来刻画这个过程。这个随机变量定义在集合

$$Z=\{1,2,3,4,5,6\}\times\{1,2,3,4,5,6\}$$

上，并且对任意的 $(i,j)\in Z$，$\Pr[(i,j)]=1/36$。考虑两个骰子的和。每一个可能的和都定义了一个事件，这些事件的概率可以通过式(2.1)计算出来。例如，要计算和为 4 的概率，相应的事件是

$$S_4=\{(1,3),(2,2),(3,1)\}$$

因此 $\Pr[S_4]=3/36=1/12$。

所有和的概率都可以用类似的公式计算。如果用 S_j 表示和为 j 的事件，可以得到如下结果：$\Pr[S_2]=\Pr[S_{12}]=1/36$，$\Pr[S_3]=\Pr[S_{11}]=1/18$，$\Pr[S_4]=\Pr[S_{10}]=1/12$，$\Pr[S_5]=\Pr[S_9]=1/9$，$\Pr[S_6]=\Pr[S_8]=5/36$，$\Pr[S_7]=1/6$。

既然事件 $S_2,S_3,\cdots,S_{12}$ 是集合 S 的一个划分，我们可以把这对骰子和的值当成一个随机变量，概率分布是以上计算的结果。

下面考虑联合概率和条件概率的概念。

① 本书原英文版中用黑体表示“随机变量”，与集合意义不同。为与原英文版图书保持一致，此处未作修改——编者注。

定义 2.2 假设 **X** 和 **Y** 是分别定义在有限集合 X 和 Y 上的随机变量。联合概率 $\Pr[x, y]$ 是 X 取 x 并且 Y 取 y 的概率。条件概率 $\Pr[x \mid y]$ 表示 **Y** 取 y 时 **X** 取 x 的概率。如果对任意的 $x \in X$ 和 $y \in Y$，都有 $\Pr[x, y] = \Pr[x]\Pr[y]$，则称随机变量 **X** 和 **Y** 是统计独立的。

联合概率和条件概率可以通过下面的公式联系起来

$$\Pr[x, y] = \Pr[x \mid y]\Pr[y]$$

交换 x 和 y，我们有

$$\Pr[x, y] = \Pr[y \mid x]\Pr[x]$$

从这两个表达式，立即得到下面的结果，也就是 Bayes 定理。

定理 2.1(Bayes 定理) 如果 $\Pr[y] > 0$，那么

$$\Pr[x \mid y] = \frac{\Pr[x]\Pr[y \mid x]}{\Pr[y]}$$

推论 2.2 **X** 和 **Y** 是统计独立的随机变量，当且仅当对所有的 $x \in X$ 和 $y \in Y$，都有 $\Pr[x \mid y] = \Pr[x]$。

例 2.2 假设我们随机地抛一对骰子。如同例 2.1 那样，考虑两个骰子的和，得到一个定义在集合 $X = \{2, 3, \cdots, 12\}$ 上的随机变量 **X**。进一步，假设随机变量 **Y**，当骰子的值相等时，取 D；不等时，取 N。我们有 $\Pr[D] = 1/6$，$\Pr[N] = 5/6$。

可以直接算出这些随机变量的联合概率和条件概率。例如，$\Pr[D|4] = 1/3$，$\Pr[4|D] = 1/6$，所以

$$\Pr[D \mid 4]\Pr[4] = \Pr[D]\Pr[4 \mid D]$$

符合 Bayes 定理。

2.3 完善保密性

这一节，我们假设 $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ 是一个指定的密码体制，一个特定的密钥 $K \in \mathcal{K}$ 只用于一次加密。假设明文空间 $\mathcal{P}$ 存在一个概率分布。因此明文元素定义了一个随机变量，用 $\mathbf{x}$ 表示。$\Pr[\mathbf{x} = x]$ 表示明文 x 发生的先验概率。还假设 Alice 和 Bob 以某种固定的概率分布选取密钥 K（通常密钥是随机选取的，因此所有的密钥都是等概率的，但是这里不需要这样）。所以密钥也定义了一个随机变量，用 $\mathbf{K}$ 表示。$\Pr[\mathbf{K} = K]$ 表示密钥 K 发生的概率。回忆一下，密钥是在Alice知道明文之前选取的。因此，我们可以合理地假设密钥和明文是统计独立的随机变量。

$\mathcal{P}$ 和 $\mathcal{K}$ 的概率分布导出了 $\mathcal{C}$ 的概率分布。因此，同样可以把密文元素看成随机变量，用 $\mathbf{y}$ 表示。不难计算出密文 y 的概率 $\Pr[\mathbf{y} = y]$。对于密钥 $K \in \mathcal{K}$，定义

$$C(K) = \{e_K(x) : x \in \mathcal{P}\}$$

也就是说 $C(K)$ 代表密钥是 K 时所有可能的密文。对于任意的 $y \in \mathcal{C}$，我们有

$$\Pr[\mathbf{y}=y]=\sum_{\{K:y\in C(K)\}}\Pr[\mathbf{K}=K]\Pr[\mathbf{x}=d_K(y)]$$

同样可以观察到，对于任意的 $y\in\mathcal{C}$ 和 $x\in\mathcal{P}$，可按以下方式计算条件概率 $\Pr[\mathbf{y}=y\,|\,\mathbf{x}=x]$（也就是给定明文 x，密文 y 的概率）

$$\Pr[\mathbf{y}=y\,|\,\mathbf{x}=x]=\sum_{\{K:x=d_K(y)\}}\Pr[\mathbf{K}=K]$$

现在可以用 Bayes 定理计算条件概率 $\Pr[\mathbf{x}=x\,|\,\mathbf{y}=y]$（也就是给定密文 y，明文 x 的概率）。计算公式如下：

$$\Pr[\mathbf{x}=x\,|\,\mathbf{y}=y]=\frac{\Pr[\mathbf{x}=x]\times\sum_{\{K:x=d_K(y)\}}\Pr[\mathbf{K}=K]}{\sum_{\{K:y\in C(K)\}}\Pr[\mathbf{K}=K]\Pr[\mathbf{x}=d_K(y)]}$$

注意到只要知道了概率分布任何人都可以完成这些计算。

我们给出一个例子，看看具体如何计算这些概率。

例 2.3　假设 $\mathcal{P}=\{a,b\}$ 满足 $\Pr[a]=1/4$，$\Pr[b]=3/4$。设 $\mathcal{K}=\{K_1,K_2,K_3\}$ 满足 $\Pr[K_1]=1/2$，$\Pr[K_2]=\Pr[K_3]=1/4$。设 $\mathcal{C}=\{1,2,3,4\}$，加密函数定义为 $e_{K_1}(a)=1$，$e_{K_1}(b)=2$；$e_{K_2}(a)=2$，$e_{K_2}(b)=3$；$e_{K_3}(a)=3$，$e_{K_3}(b)=4$。这个密码体制可以用如下加密矩阵表示：

	a	b
K_1	1	2
K_2	2	3
K_3	3	4

$\mathcal{C}$ 的概率分布计算如下：

$$\Pr[1]=\frac{1}{8}$$
$$\Pr[2]=\frac{3}{8}+\frac{1}{16}=\frac{7}{16}$$
$$\Pr[3]=\frac{3}{16}+\frac{1}{16}=\frac{1}{4}$$
$$\Pr[4]=\frac{3}{16}$$

现在计算给定密文后，明文空间上的条件概率分布为：

$$\Pr[a\,|\,1]=1\qquad \Pr[b\,|\,1]=0$$
$$\Pr[a\,|\,2]=\frac{1}{7}\qquad \Pr[b\,|\,2]=\frac{6}{7}$$
$$\Pr[a\,|\,3]=\frac{1}{4}\qquad \Pr[b\,|\,3]=\frac{3}{4}$$
$$\Pr[a\,|\,4]=0\qquad \Pr[b\,|\,4]=1$$

我们将要定义完善保密性的概念。通俗地讲，完善保密性就是说 Oscar 不能通过观察密文获得明文的任何信息。用概率分布的术语，精确的定义如下。

定义 2.3 一个密码体制具有完善保密性，如果对于任意的 $x \in \mathcal{P}$ 和 $y \in \mathcal{C}$，都有 $\Pr[x \mid y] = \Pr[x]$。也就是说，给定密文 y，明文 x 的后验概率等于明文 x 的先验概率。

在例 2.3 中，对于密文 $y = 3$ 满足完善保密性的特性，但是对于其他的密文不满足。

现在证明移位密码的完善保密性。从直觉上看这是很显然的。因为对于任意给定的密文 $y \in \mathbb{Z}_{26}$，任何 $n!$ 都可能是对应的明文。下面的定理用概率分布给出了正式的陈述和证明。

定理 2.3 假设移位密码的26个密钥都是以相同的概率 1/26 使用的，则对于任意的明文概率分布，移位密码具有完善保密性。

证明 这里 $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$，对于 $0 \leqslant K \leqslant 25$，加密函数 e_K 定义为 $e_K(x) = (x + K) \bmod 26 (x \in \mathbb{Z}_{26})$。首先，计算 $\mathcal{C}$ 的概率分布。假设 $y \in \mathbb{Z}_{26}$，则

$$\begin{aligned}\Pr[\mathbf{y} = y] &= \sum_{K \in \mathbb{Z}_{26}} \Pr[\mathbf{K} = K] \Pr[\mathbf{x} = d_K(y)] \\ &= \sum_{K \in \mathbb{Z}_{26}} \frac{1}{26} \Pr[\mathbf{x} = y - K] \\ &= \frac{1}{26} \sum_{K \in \mathbb{Z}_{26}} \Pr[\mathbf{x} = y - K]\end{aligned}$$

现在固定 y，值 $(y–K) \bmod 26$ 构成 $\mathbb{Z}_{26}$ 的一个置换。因此有

$$\sum_{K \in \mathbb{Z}_{26}} \Pr[\mathbf{x} = y - K] = \sum_{x \in \mathbb{Z}_{26}} \Pr[\mathbf{x} = x] = 1$$

从而，对于任意的 $y \in \mathbb{Z}_{26}$，有

$$\Pr[y] = \frac{1}{26}$$

接下来我们有，对于任意的 x, y

$$\begin{aligned}\Pr[y \mid x] &= \Pr[\mathbf{K} = (y - x) \bmod 26] \\ &= \frac{1}{26}\end{aligned}$$

（这是因为对于任意的 x, y，满足 $e_K(x) = y$ 的 K 是唯一的 $K = (y - x) \bmod 26$）。现在应用 Bayes 定理，很容易计算出

$$\begin{aligned}\Pr[x \mid y] &= \frac{\Pr[x] \Pr[y \mid x]}{\Pr[y]} \\ &= \frac{\Pr[x] \frac{1}{26}}{\frac{1}{26}} \\ &= \Pr[x]\end{aligned}$$

所以这个密码体制是完善保密的。

因此，如果新的随机密钥用来加密每个明文字母，则移位密码是“不可攻破的”。

下面考察一般的完善保密性。如果对某一 x_0，有 $\Pr[x_0]=0$，则显然对所有的 $y\in\mathcal{C}$，都有 $\Pr[x_0\,|\,y]=\Pr[x_0]$。因此，我们只需考虑 $\Pr[x]>0$ 的那些明文 $x\in\mathcal{P}$。对这样的明文，我们注意到使用 Bayes 定理，条件“对于所有的 $y\in\mathcal{C}$，$\Pr[x\,|\,y]=\Pr[x]$”，等价于“对于所有的 $y\in\mathcal{C}$，$\Pr[y\,|\,x]=\Pr[y]$”。我们可以合理地假设对于所有的 $y\in\mathcal{C}$，$\Pr[y]>0$（如果 $\Pr[y]=0$，则密文 y 从不会使用，可以从 $\mathcal{C}$ 中去掉）。

对于任意固定的 $x\in\mathcal{P}$。对每个 $y\in\mathcal{C}$，我们有 $\Pr[y\,|\,x]=\Pr[y]>0$。因此，对于每个 $y\in\mathcal{C}$，一定至少存在一个密钥 K 满足 $e_K(x)=y$。这样就有 $|\mathcal{K}|\geqslant|\mathcal{C}|$。在任意一个密码体制中，因为加密函数是单射，一定有 $|\mathcal{C}|\geqslant|\mathcal{P}|$。当 $|\mathcal{K}|=|\mathcal{C}|=|\mathcal{P}|$ 时，我们有一个关于什么时候取得完善保密性的很好的性质。这个性质最早是由 Shannon 提出来的。

定理 2.4　假设密码体制 $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ 满足 $|\mathcal{K}|=|\mathcal{C}|=|\mathcal{P}|$。该密码体制是完善保密的，当且仅当每个密钥被使用的概率都是 $1/|\mathcal{K}|$，并且对于任意的 $x\in\mathcal{P}$ 和 $y\in\mathcal{C}$，存在唯一的密钥 K 使得 $e_K(x)=y$。

证明　假设给定的密码体制是完善保密的。由上面的观察可知，对于任意的 $x\in\mathcal{P}$ 和 $y\in\mathcal{C}$，一定至少存在一个密钥 K 满足 $e_K(x)=y$。因此有不等式：

$$\begin{aligned}|\mathcal{C}|&=|\{e_K(x):K\in\mathcal{K}\}|\\&\leqslant|\mathcal{K}|\end{aligned}$$

但是我们假设 $|\mathcal{C}|=|\mathcal{K}|$。因此一定有

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

也就是说，不存在两个不同的密钥 K_1 和 K_2 使得 $e_{K_1}(x)=e_{K_2}(x)=y$。因此对于 $x\in\mathcal{P}$ 和 $y\in\mathcal{C}$，刚好存在一个密钥 K 使得 $e_K(x)=y$。

记 $n=|\mathcal{K}|$。设 $\mathcal{P}=\{x_i:1\leqslant i\leqslant n\}$ 并且固定一个密文 $y\in\mathcal{C}$。设密钥为 $K_1,K_2,\cdots,K_n$，并且 $e_{K_i}(x_i)=y$，$1\leqslant i\leqslant n$。使用 Bayes 定理，我们有

$$\begin{aligned}\Pr[x_i\,|\,y]&=\frac{\Pr[y\,|\,x_i]\Pr[x_i]}{\Pr[y]}\\&=\frac{\Pr[\mathbf{K}=K_i]\Pr[x_i]}{\Pr[y]}\end{aligned}$$

考虑完善保密的条件 $\Pr[x_i\,|\,y]=\Pr[x_i]$。从这里，我们有 $\Pr[K_i]=\Pr[y]$，$1\leqslant i\leqslant n$。也就是说所有的密钥都是等概率使用的（概率为 $\Pr[y]$）。但是密钥的数目为 $|\mathcal{K}|$，于是我们得到对任意的 $K\in\mathcal{K}$，$\Pr[K]=1/|\mathcal{K}|$。

相反地，如果两个假设的条件是成立的。类似于定理 2.3 的证明，很容易推出，该密码体制是完善保密的。证明细节留给读者自己完成。

一个著名的具有完善保密性的密码体制是“一次一密”密码体制。这个体制最先由 Gilbert Vernam 于 1917 年用于报文消息的自动加密和解密。有意思的是“一次一密”很多年来被认

为是不可破的，但是一直都没有一个数学的证明，直到 30 年后 Shonnon 提出了完善保密性的概念之后才得到证明。“一次一密”密码体制的描述如下所述。

密码体制 2.1 一次一密

假设 $n \geqslant 1$ 是正整数， $\mathcal{P}=\mathcal{C}=\mathcal{K}=(\mathbb{Z}_2)^n$ 。对于 $K \in (\mathbb{Z}_2)^n$ ，定义 $e_K(x)$ 为 K 和 x 的模2的向量和(或者说是两个相关比特串的异或)。因此，如果 $x=(x_1, x_2, \cdots, x_n)$ 并且 $K=(K_1, K_2, \cdots, K_n)$ ，则

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

解密与加密是一样的。如果 $y=(y_1, \cdots, y_n)$ ，则

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

由定理 2.4，容易看出“一次一密”提供了完善保密性。这个密码体制的加密和解密都很容易，有一定的吸引力。Vernam 对此还申请了专利，希望会有广泛的商业用途。遗憾的是，“一次一密”存在一个较大的不利因素。$|\mathcal{K}| \geqslant |\mathcal{P}|$ 意味着秘密使用的密钥量必须至少和明文的数量一样多。例如，在“一次一密”密码体制中，我们要求用 n 比特的密钥加密 n 比特的明文。如果相同的密钥可以用于加密不同的消息，这就不是一个重要的问题，但是密码体制的无条件安全性是基于每个密钥仅用一次这样的一个事实。

例如，“一次一密”对已知明文攻击是很脆弱的，因为 K 可以通过异或比特串 x 和 $e_K(x)$ 得到。因此，新生成的密钥需要为将要发送的明文在安全的通道上传输。这就带来了一个严峻的密钥管理问题，因此限制了“一次一密”在商业上的应用。但是“一次一密”在要求无条件安全的军事和外交环境中有着很重要的应用。

在密码学的发展历史中，人们试图设计一个密钥可以加密相对长的明文的密码体制(也就是一个密钥可以用于加密许多消息)，并且仍然可以保持一定的计算安全性。这种类型的密码体制有数据加密标准(DES)和高级加密标准(AES)，我们将在下一章讨论。

2.4 熵

上一节，我们讨论了完善保密性的概念。我们将注意力限制在密钥只能用于一次加密的特殊情况下。现在看看用一个密钥加密多个消息会发生什么，并且还要看看给密码分析人员足够的时间，进行一次成功的唯密文攻击有多大的可能性。

研究这个问题的基本工具是熵(Entropy)。这个概念来自于 Shannon 于 1948 年创建的信息论。熵可以看做是对信息或不确定性的数学度量，是通过一个概率分布的函数来计算的。

假设离散的随机变量 $\mathbf{X}$ 根据特定的概率分布从有限集合 X 中取值。我们可以从按照这个概率分布的实验结果中得到什么信息？或者说，在实验发生之前，结果的不确定性是什么？这个性质称为 $\mathbf{X}$ 的熵，用 $H(\mathbf{X})$ 表示。

这些描述看起来还相当抽象，让我们看看更具体的例子。假设随机变量 $\mathbf{X}$ 代表掷硬币。如前面提到的那样，相关的概率分布是 $\Pr[\text{heads}]=\Pr[\text{tails}]=1/2$ 。因为我们可以将 heads 编码为 1，将 tails 编码成 0，所以我们说一次掷硬币的信息或者熵是 1 比特是合理的。用同样

的方式， n 次独立的掷硬币的熵是 n 比特，因为 n 次投掷可以用长为 n 比特的比特串进行编码。

再看一个稍微复杂一点的例子。假设有一个随机变量 $\mathbf{X}$，取三种可能的值 x_1， x_2， x_3，概率分别为 1/2, 1/4, 1/4。对这三种结果最有效的编码是将 x_1 表示成 0， x_2 表示成 10， x_3 表示成 11。那么编码的平均比特长度是

$$\frac{1}{2}\times 1+\frac{1}{4}\times 2+\frac{1}{4}\times 2=\frac{3}{2}$$

上面的例子说明以概率 2^{-n} 发生的事件可以编码成长度 n 的比特串。更一般地，我们可以想象以概率 p 发生的事件可以编码成长大约为 $-\mathrm{lb}\,p$ 的比特串。对于任意的概率分布为 p_1, $p_2,\cdots, p_n$ 的随机变量 $\mathbf{X}$，将 $-\mathrm{lb}\,p_i$ 的加权平均值作为对信息的度量。正式的定义如下所述。

定义 2.4　假设随机变量 $\mathbf{X}$ 在有限集合 X 上取值，则随机变量 $\mathbf{X}$ 的熵定义为

$$H(\mathbf{X})=-\sum_{x\in X}\Pr[x]\mathrm{lb}\Pr[x]$$

注：注意到 $y=0$ 时 $\mathrm{lb}\,y$ 没有定义。因此熵有时定义为所有非零概率的相关和。但是因为 $\lim\limits_{y\to 0} y\,\mathrm{lb}\,y=0$，实际上对于某些 x 可以令 $\Pr[x]=0$。

同样地，我们注意到选择2作为指数的底也是任意的，选取其他的底仅仅改变了熵的常数因子。

如果 $|X|=n$ 并且对于所有的 $x\in X$， $\Pr[x]=1/n$，那么 $H(\mathbf{X})=\mathrm{lb}\,n$。同样地，容易观察到，对于任意的随机变量 $\mathbf{X}$， $H(\mathbf{X})\geqslant 0$。 $H(\mathbf{X})=0$ 当且仅当对于某个 $x_0\in X$， $\Pr[x_0]=1$，对于所有的 $x\neq x_0$， $\Pr[x]=0$。

让我们看看密码体制不同部分的熵。可以把密钥看成是在 $\mathcal{K}$ 上取值的随机变量 $\mathbf{K}$，因此可以计算熵 $H(\mathbf{K})$。同样地，可以分别计算同明文和密文相关的熵 $H(\mathbf{P})$ 和 $H(\mathbf{C})$。

作为示例，我们计算例 2.3 中的密码体制的熵。

例 2.3(续)　计算如下:

$$\begin{aligned}H(\mathbf{P})&=-\frac{1}{4}\mathrm{lb}\frac{1}{4}-\frac{3}{4}\mathrm{lb}\frac{3}{4}\\&=-\frac{1}{4}(-2)-\frac{3}{4}(\mathrm{lb}\,3-2)\\&=2-\frac{3}{4}(\mathrm{lb}\,3)\\&\approx 0.81\end{aligned}$$

类似地， $H(\mathbf{K})$=1.5， $H(\mathbf{C})\approx 1.85$。

2.4.1　Huffman 编码

这一节，我们简要地讨论熵和Huffman 编码之间的联系。这一节的结果与熵在密码学中的应用没有什么关系，可以跳过去，不失连续性。但是这一节有助于理解熵的概念。

我们将通过根据特定的概率分布对随机事件进行编码来介绍熵。首先让这个想法更精确些。和以前一样，**X** 是一个在有限集合 X 上取值的随机变量， p 是相关的概率分布。

对 **X** 的编码是任何映射

$$f : X \to \{0,1\}^*$$

其中 $\{0,1\}^*$ 代表所有的 0 和 1 的有限串的集合。对一串有限的事件 $x_1 \cdots x_n$， $x_i \in X$，我们用一个明显的方式扩展编码 f，定义如下

$$f(x_1 \cdots x_n) = f(x_1) \| \cdots \| f(x_n)$$

其中 $\|$ 表示串的连接。这样，就可以将 f 看成映射

$$f : X^* \to \{0,1\}^*$$

现在假设串 $x_1 \cdots x_n$ 是由无记忆源生成的，串中的每一个 x_i 都按照 X 上特定的概率分布发生。也就是说，任意串 $x_1 \cdots x_n$ 的概率可以计算如下

$$\Pr[x_1 \cdots x_n] = \Pr[x_1] \times \cdots \times \Pr[x_n]$$

注：因为源是无记忆的，串 $x_1 \cdots x_n$ 没有必要由不同的值组成。举一个简单的例子，考虑 n 次公平地掷硬币。如果我们用 1 表示正面(heads)，用 0 表示反面(tails)，每一个长为 n 的二进制串对应于 n 次掷硬币的序列。

现在，假设我们要使用映射 f 对串进行编码，以一种明确的方式编码是很重要的。因此要求 f 是单射的。

例 2.4 假设 $X = \{a, b, c, d\}$，考虑下列三种编码:

$$f(a) = 1 \quad f(b) = 10 \quad f(c) = 100 \quad f(d) = 1000$$

$$g(a) = 0 \quad g(b) = 10 \quad g(c) = 110 \quad g(d) = 111$$

$$h(a) = 0 \quad h(b) = 01 \quad h(c) = 10 \quad h(d) = 11$$

可以看出， f 和 g 是单射的，但 h 不是。任何使用 f 的编码都可以从字母编码的尾部开始向后解码: 每遇到一个 1，就预示一个当前字母的开始。

使用 g 的编码可以从头开始依次进行解码。一看到子串是 a， b， c 或者 d 的编码，就马上解码，并且去掉这个子串。例如，给定串 10101110，我们将 10 解码成 b，然后 10 解码成 b，接下来 111 解码成 d，最后 0 解码成 a。所以解码得到的字符串是 $bbda$。

h 不是单射的，一个例子就可以说明:

$$h(ac) = h(ba) = 010$$

考虑到解码的简单性，我们喜欢编码 g 胜于 f。这是因为 g 的解码可以从开头到结尾依次完成，不需要记忆。这种允许简单的按次序的解码性质称为无前缀特性[我们说 g 的解码是无前缀的，如果不存在两个元素 $x, y \in X$ 和 $z \in \{0,1\}^*$，满足 $g(x) = g(y) \| z$]。

讨论到目前为止还没有涉及熵。不用奇怪，熵和解码的效率有关。我们将与以前一样度量 f 的解码效率：**X** 的元素的编码的加权平均长度(用 $\ell(f)$ 表示)。因此我们有如下定义：

$$\ell(f)=\sum_{x\in X}\Pr[x]\,|\,f(x)\,|$$

其中$|y|$表示串 y 的长度。

现在，我们的基本问题是找到一个单射编码 f，使得 $\ell(f)$ 最小。著名的 Huffman 算法达到了这个目标。另外，由 Huffman 算法得到的 f 解码是无前缀的，并且有

$$H(X)\leqslant\ell(f)\leqslant H(X)+1$$

因此，**X** 的熵值是对最佳单射编码的平均长度的精确估计。

我们不证明上述结果，但是会给出 Huffman 算法的简短的非正式的描述。Huffman 算法从集合 X 上的概率分布开始，每个元素的码字最初是空的。在每一个步骤中，两个概率最小的元素组合成一个元素，并以这两个概率的和作为新元素的概率。概率最小的元素赋值为“0”，另一个元素赋值为“1”。当只剩下一个元素时，每个 $x\in X$ 的编码可以通过从最后一个元素到最初的元素 x“回溯”记录下得到的 0、1 序列而构造出来。

举个例子很容易说明这个算法。

例 2.5　假设 $X=\{a,b,c,d,e\}$ 有如下概率分布：$\Pr[a]=0.05$，$\Pr[b]=0.10$，$\Pr[c]=0.12$，$\Pr[d]=0.13$，$\Pr[e]=0.60$。Huffman 算法按如下表格进行：

<table>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td></tr>
<tr><td>0.05</td><td>0.10</td><td>0.12</td><td>0.13</td><td>0.60</td></tr>
<tr><td>0</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">0.15</td><td>0.12</td><td>0.13</td><td>0.60</td></tr>
<tr><td colspan="2"></td><td>0</td><td>1</td><td></td></tr>
<tr><td colspan="2">0.15</td><td colspan="2">0.25</td><td>0.60</td></tr>
<tr><td colspan="2">0</td><td colspan="2">1</td><td></td></tr>
<tr><td colspan="4">0.40</td><td>0.60</td></tr>
<tr><td colspan="4">0</td><td>1</td></tr>
<tr><td colspan="5">1.0</td></tr>
</table>

可得到如下编码：

x	$f(x)$
a	000
b	001
c	010
d	011
e	1

因此，编码的平均长度是

$$\begin{aligned}\ell(f)&=0.05\times3+0.10\times3+0.12\times3+0.13\times3+0.60\times1\\&=1.8\end{aligned}$$

和熵比较：

$$H(\mathbf{X}) = 0.2161 + 0.3322 + 0.3671 + 0.3842 + 0.4422$$
$$= 1.7402$$

可以看出，编码的平均长度和熵十分接近。

2.5 熵的性质

这一节，我们证明一些和熵有关的基本结果。首先，叙述一个基本结果，称为 Jensen 不等式。这个不等式对我们很有用。Jensen 不等式涉及凸函数，下面给出定义。

定义 2.5 一个区间 I 上的实值函数 f 为凸函数，如果对任意的 $x, y \in I$ 满足

$$f\left(\frac{x+y}{2}\right) \geqslant \frac{f(x)+f(y)}{2}$$

称 f 是区间 I 上的严格凸函数，如果对任意的 $x, y \in I$， $x \neq y$ 满足

$$f\left(\frac{x+y}{2}\right) > \frac{f(x)+f(y)}{2}$$

下面是 Jensen 不等式，在此不给出证明。

定理 2.5（Jensen 不等式） 假设 f 是区间 I 上的连续的严格凸函数，且

$$\sum_{i=1}^{n} a_i = 1$$

其中 $a_i > 0$，$1 \leqslant i \leqslant n$。那么

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

其中 $x_i \in I$，$1 \leqslant i \leqslant n$。当且仅当 $x_1 = \cdots = x_n$ 等式成立。

我们现在继续给出几个熵的结果。下面的定理利用了函数 $\mathrm{lb}\, x$ 在区间 $(0,\infty)$ 上是严格凸的这一事实(事实上，对数函数的二阶导函数在 $(0,\infty)$ 上是负的，很容易就得到这个结论)。

定理 2.6 假设 $\mathbf{X}$ 是一个随机变量，概率分布为 $p_1, p_2, \cdots, p_n$，其中 $p_i > 0$，$1 \leqslant i \leqslant n$。那么 $H(\mathbf{X}) \leqslant \mathrm{lb}\, n$，当且仅当 $p_i = 1/n$，$1 \leqslant i \leqslant n$ 时等式成立。

证明 应用 Jensen 不等式，我们有如下结果：

$$H(\mathbf{X}) = -\sum_{i=1}^{n} p_i \mathrm{lb}\, p_i$$
$$= \sum_{i=1}^{n} p_i \mathrm{lb} \frac{1}{p_i}$$

$$\leqslant \mathrm{lb}\sum_{i=n}^{n}\left(p_i \times \frac{1}{p_i}\right)$$
$$= \mathrm{lb}\ n$$

等式成立当且仅当 $p_i = 1/n$，$1 \leqslant i \leqslant n$。

定理 2.7　$H(\mathbf{X},\mathbf{Y}) \leqslant H(\mathbf{X}) + H(\mathbf{Y})$，当且仅当 $\mathbf{X}$ 和 $\mathbf{Y}$ 统计独立时等号成立。

证明　假设 $\mathbf{X}$ 取值 x_i，$1 \leqslant i \leqslant m$，$\mathbf{Y}$ 取值 y_j，$1 \leqslant j \leqslant n$。记 $p_i = \Pr[\mathbf{X} = x_i]$，$1 \leqslant i \leqslant m$；$q_j = \Pr[\mathbf{Y} = y_j]$，$1 \leqslant j \leqslant n$。记 $r_{ij} = \Pr[\mathbf{X} = x_i, \mathbf{Y} = y_j]$，$1 \leqslant i \leqslant m$，$1 \leqslant j \leqslant n$（这是联合概率分布）。

注意到

$$p_i = \sum_{j=1}^{n} r_{ij}$$

（$1 \leqslant i \leqslant m$），且

$$q_j = \sum_{i=1}^{m} r_{ij}$$

（$1 \leqslant j \leqslant n$）。计算如下：

$$H(\mathbf{X}) + H(\mathbf{Y}) = -\left(\sum_{i=1}^{m} p_i \mathrm{lb}\ p_i + \sum_{j=1}^{n} q_j \mathrm{lb}\ q_j\right)$$
$$= -\left(\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} \mathrm{lb}\ p_i + \sum_{j=1}^{n}\sum_{i=1}^{m} r_{ij} \mathrm{lb}\ q_j\right)$$
$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} \mathrm{lb}\ p_i q_j$$

另一方面

$$H(\mathbf{X},\mathbf{Y}) = -\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} \mathrm{lb}\ r_{ij}$$

于是有：

$$H(\mathbf{X},\mathbf{Y}) - H(\mathbf{X}) - H(\mathbf{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} \mathrm{lb}\frac{1}{r_{ij}} + \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} \mathrm{lb}\ p_i q_j$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij} \mathrm{lb}\frac{p_i q_j}{r_{ij}}$$
$$\leqslant \mathrm{lb}\sum_{i=1}^{m}\sum_{j=1}^{n} p_i q_j$$
$$= \mathrm{lb}\ 1$$
$$= 0$$

（以上的计算使用了 Jensen 不等式，用到了 r_{ij} 是正实数并且和是 1 这一事实）。

我们同样可以得出什么时候使等式成立：此时存在常数 c 使得对于所有的 i, j，都有 $p_i q_j / r_{ij} = c$。利用事实

$$\sum_{j=1}^{n}\sum_{i=1}^{m} r_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} p_i q_j = 1$$

于是有 $c = 1$。因此，等式成立当且仅当 $r_{ij} = p_i q_j$，也就是

$$\Pr[\mathbf{X} = x_i, \mathbf{Y} = y_j] = \Pr[\mathbf{X} = x_i]\Pr[\mathbf{Y} = y_j]$$

$1 \leqslant i \leqslant m$，$1 \leqslant j \leqslant n$。这说明 $\mathbf{X}$ 和 $\mathbf{Y}$ 是统计独立的。

接下来定义条件熵。

定义 2.6 假设 $\mathbf{X}$ 和 $\mathbf{Y}$ 是两个随机变量。对于 $\mathbf{Y}$ 的任何固定值 y，得到一个 $\mathbf{X}$ 上的(条件)概率分布；记相应的随机变量为 $\mathbf{X}|y$。显然

$$H(\mathbf{X}|y) = -\sum_{x} \Pr[x|y] \text{lb } \Pr[x|y]$$

定义条件熵 $H(\mathbf{X}|\mathbf{Y})$ 为熵 $H(\mathbf{X}|y)$ 取遍所有的 y 的加权平均值。计算公式为

$$H(\mathbf{X}|\mathrm{Y}) = -\sum_{y}\sum_{x} \Pr[y]\Pr[x|y] \text{lb } \Pr[x|y]$$

条件熵度量了 $\mathbf{Y}$ 揭示的 $\mathbf{X}$ 的平均信息量。

接下来的两个结果很容易得到，我们将其证明留作练习。

定理 2.8 $H(\mathbf{X}, \mathbf{Y}) \leqslant H(\mathbf{Y}) + H(\mathbf{X}|\mathbf{Y})$。

推论 2.9 $H(\mathbf{X}|\mathbf{Y}) \leqslant H(\mathbf{X})$，等式成立当且仅当 $\mathbf{X}$ 和 $\mathbf{Y}$ 统计独立。

2.6 伪密钥和唯一解距离

这一节，应用证明过的有关密码体制的熵的结果。首先，给出了密码体制的各组成部分的熵的基本关系。条件熵 $H(\mathbf{K}|\mathbf{C})$ 称为密钥含糊度，度量了给定密文下密钥的不确定性。

定理 2.10 设 $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ 是一个密码体制，那么

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

证明 首先注意到 $H(\mathbf{K}, \mathbf{P}, \mathbf{C}) = H(\mathbf{C}|\mathbf{K}, \mathbf{P}) + H(\mathbf{K}, \mathbf{P})$。因为 $y = e_K(x)$，所以密钥和明文唯一决定密文。这说明 $H(\mathbf{C}|\mathbf{K}, \mathbf{P}) = 0$，因此 $H(\mathbf{K}, \mathbf{P}, \mathbf{C}) = H(\mathbf{K}, \mathbf{P})$。但是 $\mathbf{K}$ 和 $\mathbf{P}$ 是统计独立的，所以 $H(\mathbf{K}, \mathbf{P}) = H(\mathbf{K}) + H(\mathbf{P})$。因此

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

同样地，由于密钥和密文唯一决定明文(即 $x=d_K(y)$)，我们有 $H(\mathbf{P}\mid\mathbf{K},\mathbf{C})=0$，因此有 $H(\mathbf{K},\mathbf{P},\mathbf{C})=H(\mathbf{K},\mathbf{C})$。

所以

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

让我们回到例 2.3 解释这个结论。

例 2.3(续)　我们已经计算出 $H(\mathbf{P})\approx 0.81$，$H(\mathbf{K})\approx 1.5$，$H(\mathbf{C})\approx 1.85$。由定理 2.10 可知，$H(\mathbf{K}\mid\mathbf{C})\approx 1.5+0.81-1.85\approx 0.46$。可以通过条件熵的定义直接验证。首先需要计算概率 $\Pr[\mathbf{K}=K_i\mid y=j]$，$1\leqslant i\leqslant 3$，$1\leqslant j\leqslant 4$。这可以使用 Bayes 定理完成，计算结果如下:

$$\begin{array}{lll}\Pr[K_1\mid 1]=1 & \Pr[K_2\mid 1]=0 & \Pr[K_3\mid 1]=0\\ \Pr[K_1\mid 2]=\dfrac{6}{7} & \Pr[K_2\mid 2]=\dfrac{1}{7} & \Pr[K_3\mid 2]=0\\ \Pr[K_1\mid 3]=0 & \Pr[K_2\mid 3]=\dfrac{3}{4} & \Pr[K_3\mid 3]=\dfrac{1}{4}\\ \Pr[K_1\mid 4]=0 & \Pr[K_2\mid 4]=0 & \Pr[K_3\mid 4]=1\end{array}$$

所以

$$H(\mathbf{K}\mid\mathbf{C})=\frac{1}{8}\times 0+\frac{7}{16}\times 0.59+\frac{1}{4}\times 0.81+\frac{3}{16}\times 0=0.46$$

这与由定理 2.10 得到的结果相符合。

设 $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ 是正在使用的密码体制，明文串为

$$x_1x_2\cdots x_n$$

用一个密钥加密，得到密文串为

$$y_1y_2\cdots y_n$$

密码分析的基本目的是确定密钥。我们考虑唯密文攻击，并且假设分析者 Oscar 拥有无限的计算资源。同时假定 Oscar 知道明文是某一自然语言，如英语。一般来说，Oscar 能排除某些密钥，但是许多可能的密钥存在，只有一个密钥是正确的。那些可能的但不正确的密钥称为伪密钥。

例如，假设 Oscar 获得密文串 WNAJW，这个密文串是使用移位密码得到的。容易知道，只有两个有意义的明文串，即 river 和 arena，分别对应于可能的加密密钥 $F(=5)$ 和 $W(=22)$。这两个密钥，一个是正确的密钥，一个是伪密钥(实际上，对长为5的移位密码的密文找到两个有意义的解密是有些困难的；读者可以寻找其他一些例子)。

我们的目的是推导伪密钥的期望数的下界。首先，定义自然语言 L 的熵(每字母)的含义，记为 H_L。H_L 应该是对有意义的明文串中的每个字母平均信息的度量(注意，一个随机的字

母串具有的熵(每字母)是 $\mathrm{lb}\,26 \approx 4.70$)。作为 H_L 的“一阶”近似值，我们使用 $H(\mathbf{P})$。L 是英语的时候，我们使用表 1.1 给出的概率分布得到 $H(\mathbf{P}) \approx 4.19$。

当然，语言中相继的字母不是统计独立的，相继字母的相关性减少了熵。例如在英语中，字母“Q”总是跟着字母“U”。作为“二阶”近似值，我们计算所有两字母组的概率分布的熵，然后除以 2。一般地，定义 $\mathbf{P}^n$ 为所有 n 字母组的概率分布构成的随机变量。我们使用以下定义。

定义 2.7 假设 L 是自然语言，语言 L 的熵定义为

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

语言 L 的冗余度(redundancy)定义为

$$R_L = 1 - \frac{H_L}{\mathrm{lb}\,|\mathcal{P}|}$$

注： H_L 度量了语言 L 的每个字母的平均熵。一个随机语言具有熵 $\mathrm{lb}\,|\mathcal{P}|$。因此 R_L 度量了“多余字母”的比例，即冗余度。

在英语中，可以通过大量的两字母组的列表和它们的频率给出 $H(\mathbf{P}^2)$ 的估计。$H(\mathbf{P}^2)/2 \approx 3.90$ 就是这样得出的。可以继续对三字母组列表，等等，最后得到对 H_L 的估计。事实上，各种实验已经得出了一个经验性的结果：$1.0 \leqslant H_L \leqslant 1.5$。也就是说，英语的平均信息内容大概是每字母 1.5 比特。

使用 1.25 作为 H_L 的估计值时，冗余度大约为 0.75。这意味着英语有 75%的冗余度！(这不是说可以任意地从英语文本中去掉 3/4 的字母，希望还可以阅读。而是说，对一个充分大的 n，可以找到一个 n 字母组的 Huffman 编码，将英语原文压缩到原来长度的 1/4)。

给定 $\mathcal{K}$ 和 $\mathcal{P}^n$ 的概率分布，我们可以定义导出的 $\mathcal{C}^n$ 上的概率分布，其中 $\mathcal{C}^n$ 是密文的 n 字母组(已经讨论过 $n=1$ 时的情况)。定义 $\mathbf{P}^n$ 为代表明文 n 字母组的随机变量。类似地，定义 $\mathbf{C}^n$ 为代表密文 n 字母组的随机变量。

给定 $y \in \mathbf{C}^n$，定义

$$K(y) = \{K \in \mathcal{K} : \exists x \in \mathcal{P}^n \text{ 使得 } \Pr[x] > 0, e_K(x) = y\}$$

也就是说，$K(y)$ 是一个密钥的集合，在这些密钥下 y 是长为 n 的有意义的明文串的密文；或者说，$K(y)$ 是密文为 y 的可能密钥的集合。如果 y 是被观察的密文串，那么伪密钥的个数是 $|K(y)|-1$，因为只有一个可能的密钥是正确的密钥。伪密钥的平均数目(在所有可能的长为 n 的密文串上)记为 $\overline{s}_n$，其值计算如下：

$$\begin{aligned}
\overline{s}_n &= \sum_{y\in\mathcal{C}^n} \Pr[y](|K(y)|-1) \\
&= \sum_{y\in\mathcal{C}^n} \Pr[y]|K(y)| - \sum_{y\in\mathcal{C}^n} \Pr[y] \\
&= \sum_{y\in\mathcal{C}^n} \Pr[y]|K(y)| - 1
\end{aligned}$$

根据定理 2.10，我们有

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

我们也用到了估计

$$H(\mathbf{P}^n)\approx nH_L=n(1-R_L)\mathrm{lb}\,|\mathcal{P}|$$

其中 n 充分的大。当然

$$H(\mathbf{C}^n)\leqslant n\mathrm{lb}\,|\mathcal{C}|$$

如果 $|\mathcal{C}|=|\mathcal{P}|$，则有

$$H(\mathbf{K}\,|\,\mathbf{C}^n)\geqslant H(\mathbf{K})-nR_L\mathrm{lb}\,|\mathcal{P}| \tag{2.2}$$

接下来，将 $H(\mathbf{K}\,|\,\mathbf{C}^n)$ 和伪密钥的个数 $\overline{s}_n$ 关联起来。计算如下：

$$\begin{aligned}H(\mathbf{K}\,|\,\mathbf{C}^n)&=\sum_{y\in\mathcal{C}^n}\Pr[y]H(\mathbf{K}\,|\,y)\\&\leqslant\sum_{y\in\mathcal{C}^n}\Pr[y]\mathrm{lb}\,|K(y)|\\&\leqslant\mathrm{lb}\sum_{y\in\mathcal{C}^n}\Pr[y]\,|K(y)|\\&=\mathrm{lb}(\overline{s}_n+1)\end{aligned}$$

其中我们对函数 $f(x)=\mathrm{lb}\,x$ 用到了 Jensen 不等式（参见定理 2.5）。因此得到不等式

$$H(\mathbf{K}\,|\,\mathbf{C}^n)\leqslant\mathrm{lb}(\overline{s}_n+1) \tag{2.3}$$

将式（2.2）和式（2.3）结合起来，我们有

$$\mathrm{lb}(\overline{s}_n+1)\geqslant H(\mathbf{K})-nR_L\mathrm{lb}|\mathcal{P}|$$

考虑等概率选取密钥的情况（此时 $H(\mathbf{K})$ 取最大值），可得到如下结果。

定理 2.11　假设 $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ 是一个密码体制，$|\mathcal{C}|=|\mathcal{P}|$ 并且密钥是等概率选取的。设 R_L 表示明文的自然语言的冗余度，那么给定一个充分长（长为 n）的密文串，伪密钥的期望数满足

$$\overline{s}_n\geqslant\frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}}-1$$

当 n 增加时，$|\mathcal{K}|/|\mathcal{P}|^{nR_L}-1$ 以指数速度趋近零。同时注意到，如果 n 很小，$H(\mathbf{P}^n)/n$ 可能对 H_L 的估计不够好，此时上面的估计可能不精确。

还有一个概念需要定义。

定义 2.8　一个密码体制的唯一解距离定义为使得伪密钥的期望数等于零的 n 的值，记为 n_0，即在给定的足够的计算时间下分析者能唯一计算出密钥所需密文的平均量。

如果在定理 2.11 中令 $\overline{s}_n=0$，解出 n，我们可以得到唯一解距离的一个近似估计，即

$$n_0 \approx \frac{\text{lb}|\mathcal{K}|}{R_L \text{lb}|\mathcal{P}|}$$

作为一个例子，考虑代换密码。在这种密码体制中 $|\mathcal{P}|=26$，$|\mathcal{K}|=26!$。如果取 $R_L=0.75$，就得到唯一解距离的估计为

$$n_0 \approx 88.4/(0.75\times 4.7) \approx 25$$

这意味着，给定的密文串的长至少是 25 时，通常解密才是唯一的。

2.7 乘积密码体制

Shannon 在其1949年的论文中介绍的另一个新思想是通过“乘积”组合密码体制。这种思想在现代密码体制的设计中非常重要，比如 AES 的设计，我们将在下一章进行研究。

为简单起见，这一节中我们将注意力集中在 $\mathcal{C}=\mathcal{P}$ 的密码体制：这种类型的密码体制称为内嵌式密码体制(endomorphic cryptosystem)。设 $S_1=(\mathcal{P},\mathcal{P},\mathcal{K}_1,\mathcal{E}_1,\mathcal{D}_1)$，$S_2=(\mathcal{P},\mathcal{P},\mathcal{K}_2,\mathcal{E}_2,\mathcal{D}_2)$ 是两个具有相同明文空间(密文空间)的内嵌式密码体制。那么 S_1 和 S_2 的乘积密码体制 $S_1\times S_2$ 定义为

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

乘积密码体制的密钥形式是 $K=(K_1,K_2)$，其中 $K_1\in\mathcal{K}_1$，$K_2\in\mathcal{K}_2$。加密和解密的规则定义为：对于任意的 $K=(K_1,K_2)$，e_K 定义为

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

d_K 定义为

$$d_{(K_1,K_2)}(y)=d_{K_1}(d_{K_2}(y))$$

也就是说，我们首先用 e_{K_1} 加密 x，然后用 e_{K_2} 再加密一次得到的密文。解密是类似的，但是必须以相反的次序进行：

$$\begin{aligned} d_{(K_1,K_2)}(e_{(K_1,K_2)}(x)) &= d_{(K_1,K_2)}(e_{K_2}(e_{K_1}(x))) \\ &= d_{K_1}(d_{K_2}(e_{K_2}(e_{K_1}(x)))) \\ &= d_{K_1}(e_{K_1}(x)) \\ &= x \end{aligned}$$

密码体制有与密钥空间相关的概率分布，因此，需要定义密钥空间 $\mathcal{K}$ 的概率分布。很自然地定义为：

$$\Pr[(K_1,K_2)]=\Pr[K_1]\times\Pr[K_2]$$

换句话说，分别根据定义在 $\mathcal{K}_1$ 和 $\mathcal{K}_2$ 上的概率分布，独立地选取 K_1 和 K_2。

下面用一个简单的例子来说明乘积密码体制的定义。密码体制 2.2 是一个乘法密码(Multiplicative Cipher)。

密码体制 2.2　乘法密码

设 $\mathcal{P}=\mathcal{C}=\mathbb{Z}_{26}$，并且

$$\mathcal{K}=\{a\in\mathbb{Z}_{26}:\gcd(a,26)=1\}$$

对于 $a\in\mathcal{K}$，定义

$$e_a(x)=ax \bmod 26$$

和

$$d_a(y)=a^{-1}y \bmod 26$$

$(x,y\in\mathbb{Z}_{26})$。

假设 M 是乘法密码(密钥等概率选取)，S 是移位密码(密钥等概率选取)。很容易看出 $M\times S$ 也是仿射密码(同样，密钥等概率选取)。要说明 $S\times M$ 是密钥等概率选取的仿射密码要略微困难一些。

现在我们证明这个论断。移位密码的密钥是 $K\in\mathbb{Z}_{26}$，相应的加密规则是 $e_K(x)=(x+K) \bmod 26$。乘法密码的密钥是 $a\in\mathbb{Z}_{26}$，$\gcd(a,26)=1$；相应的加密规则是 $e_a(x)=ax \bmod 26$。因此，乘积密码 $M\times S$ 的密钥具有 (a,K) 的形式，并且

$$e_{(a,K)}(x)=(ax+K) \bmod 26$$

但是这就是仿射密码的定义。另外，仿射密码的密钥的概率是密钥 a 和 K 的概率乘积：$1/312=1/12\times 1/26$。因此 $M\times S$ 是仿射密码。

现在考虑 $S\times M$。这个密码的密钥具有形式 (K,a)，并且

$$e_{(K,a)}(x)=a(x+K) \bmod 26=(ax+aK) \bmod 26$$

这样乘积密码 $S\times M$ 的密钥 (K,a) 等同于仿射密码的密钥 (a,aK)。剩下要证明的是密钥在仿射密码中具有和在乘积密码 $S\times M$ 中相同的概率 $1/312$。注意到，$aK=K_1$ 当且仅当 $K=a^{-1}K_1$（$\gcd(a,26)=1$，故 a 对模 26 有乘积逆)。换句话说，仿射密码中的密钥 (a,K_1) 等同于乘积密码 $S\times M$ 中的密钥 $(a^{-1}K_1,a)$。这样我们就在两个密钥空间之间建立了一个双射。因为每个密钥都是等概率的，$S\times M$ 确实是仿射密码。

我们已经证明了 $M\times S=S\times M$，因此可以说密码体制 M 和 S 是可交换的。但是不是所有的密码体制都是可交换的，很容易找出反例。另外，乘积运算是可结合的：

$$(S_1\times S_2)\times S_3=S_1\times(S_2\times S_3)$$

如果将(内嵌式)密码体制和自己做乘积，我们得到密码体制 $S\times S$，记为 S^2。如果做 n 重乘积，得到的密码体制记为 S^n。

一个密码体制是幂等的，如果 $S^2=S$。我们在第 1 章研究的许多密码体制都是幂等的。例如移位密码、代换密码、仿射密码、希尔密码、维吉尼亚密码和置换密码都是幂等的。当然如果一个密码体制是幂等的，使用乘积体制 S^2 就毫无意义，因为这需要多余的密钥但没有提供更高的安全性。

如果密码体制不是幂等的，那么多次迭代有可能提高安全性。这个思想在 DES 中使用过了，其中包括了16轮的迭代。但是，这种方法显然要求以非幂等的密码体制开始。一种构造简单的非幂等的密码体制的方法是对两个不同的(简单的)密码体制做乘积。