# 维吉尼亚密码及其破译

Breaking Vigenère Cipher

刘卓

维吉尼亚密码（Breaking Vigenère Cipher）是使用一系列凯撒密码组成密码字母表的加密算法，即多字母替代凯撒密码。属于多表密码的一种简单形式。

## 1 排列

排列 (Permutation), 是将相异对象或符号根据确定的顺序重排。每个顺序都称作一个排列。即：从 n 个对象取 k 个对象的有序排列，其中 $k \leq n$。一共有

$$P(n,k) := n \cdot (n-1) \cdot (n-2) \cdots (n-k+2) \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

种方法.

## 2 组合

组合 (Combination)，一个集的元素的组合是一个子集。若两个子集的元素完全相同并顺序相异，它仍视为同一个组合。即：从 n 个对象取 k 个对象的无序排列，其中 $k \leq n$。读作 n 取 k。

$$C(n,k) = \frac{P(n,k)}{k!} = \begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$$

## 3 概率

对于一项实验，其中存在 n 种不同的可能性，然后以 k 种可能的方式出现结果的概率为 $\frac{k}{n}$

**例 1**

生日问题。一年有 365 天，在一个班级中，假设这个班有 n 个人，求这个班至少有两人在同一天生日的概率。

$$1 - \frac{P(365,n)}{365^n} = 1 - \frac{365 \cdot 364 \cdots (365-n+1)}{365^n}$$

| $n$ | 1 | 2 | 3 | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 0.0027 | 0.0082 | 0.117 | 0.411 | 0.706 | 0.891 | 0.97 |

即代表如果一个班有 50 个人，基本可以确定至少有两人在同一天生日。

**例 2**

假设甲板上有 45 张卡。在这些卡中，有 20 张标有 “X”，15 张标有 “Y”，以及 10 张 “Z”。随机选择一张牌，放回去，然后随机洗牌并随机选择另一张牌。

(1) 找出第一张是 X 第二张是 Z 的概率。

$$P(X\ and\ Z) = \mathbb{P}(x) \cdot \mathbb{P}(z) = \frac{20}{45} \cdot \frac{10}{45} = \frac{8}{81}$$

(2) 两张卡分别是 X 和 Z

$$\mathbb{P}\left(X\ and\ Z\right) = \mathbb{P}(X\ and\ Z) + \mathbb{P}(Z\ and\ X) = \frac{8}{81} + \frac{8}{81} = \frac{16}{81}$$

(3) 两张都是 Y

$$\mathbb{P}(Y\ and\ Y) = \mathbb{P}(Y) \cdot \mathbb{P}(Y) = \frac{15}{45} \cdot \frac{15}{15} = \frac{1}{9}$$

□

下表列出了 7834 个字母的英语写作样本中的字母的相对频率。

| $Letter$ | $Relative frequency(\%)$ | $Letter$ | $Relative frequency(\%)$ |
|---|---|---|---|
| $A$ | 8.399 | $N$ | 6.778 |
| $B$ | 1.442 | $O$ | 7.493 |
| $C$ | 2.527 | $P$ | 1.991 |
| $D$ | 4.800 | $Q$ | 0.077 |
| $E$ | 12.150 | $R$ | 6.063 |
| $F$ | 2.132 | $S$ | 6.319 |
| $G$ | 2.323 | $T$ | 8.999 |
| $H$ | 6.025 | $U$ | 2.783 |
| $I$ | 6.485 | $V$ | 0.996 |
| $J$ | 0.102 | $W$ | 2.464 |
| $K$ | 0.689 | $X$ | 0.204 |
| $L$ | 4.008 | $Y$ | 2.157 |
| $M$ | 2.566 | $Z$ | 0.025 |

随机取两个字母，其两个字母相同的概率为 $\mathbb{P}(2Letter) = \mathbb{P}(2A) + \mathbb{P}(2B) + \cdots + \mathbb{P}(2Z) = 0.08399^2 + 0.01442^2 + \cdots + 0.00025^2 = 6.5\%$

**例 3**

1. 如果**凯撒密码**转化字符的方法为 $A \to D$, 密文中随机选择一个字母 A 的概率是多少?

$$\mathbb{P}(\text{密文中的}A) = \mathbb{P}(\text{明文中的}X) = \mathbb{P}(X) = 0.204\%$$

2. 密文中随机选择一个字母 B 的概率是多少?

$$\mathbb{P}(\text{密文中的}B) = \mathbb{P}(\text{明文中的}Y) = \mathbb{P}(Y) = 2.157\%$$

□

**例 4** 假设使用维吉尼亚密码 (Vigenère Cipher)，密钥 (Key) 是 $DN$。

（1）密文中 A 的概率?

$$\mathbb{P}(\text{密文中的}A) = \mathbb{P}(\text{明文中奇数位是}X) + \mathbb{P}(\text{明文中偶数位是}N) = \frac{\mathbb{P}(X)}{2} + \frac{\mathbb{P}(Y)}{2} = 3.491\%$$

（2）密文中 B 的概率?

$$\mathbb{P}(\text{密文中的}B) = \mathbb{P}(\text{明文中奇数位是}Y) + \mathbb{P}(\text{明文中偶数位是}O) = \frac{\mathbb{P}(Y)}{2} + \frac{\mathbb{P}(O)}{2} = 4.825\% \qquad \square$$

以此列推，即使使用长密钥，在密文中看到任何字母的概率将收敛为 $\frac{1}{26}$

**条件概率**是指事件 A 已经发生了，发生事件 B 的概率。

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B\text{和}A}{\mathbb{P}(A)}$$

# 4 重合因子

重合因子 (Index of Coincidence) 是密文中两个随机选择的字母相同的概率，记为 $I$。

- 如果 $I \approx 0.065$，则表示密码很有可能是**单字母**代替
- 对于多字母代替，$I$ 的范围是 $\frac{1}{26} = 0.0385 \leq I \leq 0.065$

$$I = \frac{1}{n(n-1)} \sum_{i=0}^{25} n_i(n_i - 1), 0 \leq i \leq 25$$

其中 $n$ 为文本中所有字符总和。$n_i$ 为每个字母出现的个数。如 A 出现五次，记 $n_0 = 5$。

如果在拉丁字母表中使用维吉尼亚密码 (Vigenère Cipher)，其密钥长度为 $k$,，为了估计密钥长度，则可以使用弗里德曼检验 (Friedman Test):

$$\text{重合因子} = I \approx \frac{0.0385 \times n(k-1) + 0.065(n-k)}{k(n-1)}$$

$$\text{密钥长度} = k \approx \frac{0.0265n}{(0.065 - I) + n(I - 0.0385)}$$

弗里德曼检验仅仅只能估计密钥长度 $k$，但密文长度也**不能太短**。

**例 5**

已知密文使用维吉尼亚密码 (Vigenère Cipher) 加密，密文总长为 $n = 337$，每个字母出现频率如表格所示。估计密钥长度 $k$ 是多少。

解：

| 字母 | 数量 | 字母 | 数量 |
|---|---|---|---|
| $A$ | 13 | $N$ | 11 |
| $B$ | 18 | $O$ | 17 |
| $C$ | 12 | $P$ | 21 |
| $D$ | 15 | $Q$ | 9 |
| $E$ | 26 | $R$ | 16 |
| $F$ | 4 | $S$ | 7 |
| $G$ | 15 | $T$ | 8 |
| $H$ | 9 | $U$ | 7 |
| $I$ | 16 | $V$ | 8 |
| $J$ | 8 | $W$ | 14 |
| $K$ | 9 | $X$ | 8 |
| $L$ | 18 | $Y$ | 20 |
| $M$ | 22 | $Z$ | 6 |

$$I = \frac{1}{n(n-1)} = \sum_{i=0}^{25} n_i(n_i - 1) = \frac{1}{337 \times 336}[13 \times 12 + 18 \times 17 + \cdots + 6 \times 5 = 0.0428$$

$$k = \frac{0.0265 \times 337}{(0.065 - 0.0428) + 337 \times (0.0428 - 0.0385)} \approx 6.2 \approx 6 \qquad \square$$

# 5 卡西斯基检验

卡西斯基检验 (Kasiski Test) 是另一种维吉尼亚密码 (Vigenère Cipher) 中估算密钥长度的方法。它从密文中重复字母组之间的最大公约数 (gcd) 中获得可能的密钥长度。

**例 6**

估算密文 $IVEVYGARMLMYIVEKFDIVEFRL$ 密钥长度

解：

$$\underbrace{IVEVYGARMLMY}_{12}\underbrace{IVEKFD}_{6}\underbrace{IVEFRL}_{6}$$

$$k \approx gcd(12, 6) = 6$$

# 6 密码分析

弗里德曼检验 (Friedman Test) 和卡西斯基检验 (Kasiski Test) 只能估计密钥长度，而不能直接猜出密钥本身。而且有一定局限性，通常少于 400 个字符时，检验的准确率不高。