# Cryptography: Birthday Paradox

Yiheng Lin, Zhihao Jiang

## 1

### 1

Theorem 1.1. Let $S = \{1, 2, \cdots, N\}$. For n times, uniformly randomly draw one element from set S with replacement. Let $x_t$ be the element we draw at time t. Then $\forall p > 0$, there exists a constant $C_1$ such that when $n \geq C_1\sqrt{N}$, we have

$$Pr[\exists 1 \leq i, j \leq n, i \neq j \text{ such that } x_i = x_j] > p$$

.

*Proof.* Let X denote the event that $\exists i, j \leq t, i \neq j$ such that $x_i = x_j$, then we have

$$\begin{aligned} Pr[\bar{X}] &= \frac{N(N-1)\cdots(N-n+1)}{N^n} \\ &= \prod_{i=1}^{n-1}(1 - \frac{i}{N}) \\ &\leq \prod_{i=1}^{n-1} exp(-\frac{i}{N}) \\ &= exp(-\frac{n(n-1)}{2N}) \end{aligned} \tag{1}$$

Let $C_1 = \sqrt{-2ln(1-p)} + 1$. Then when $n \geq C_1\sqrt{N}$, we have

$$n(n-1) > (1 + \sqrt{-2ln(1-p) \cdot N})(\sqrt{-2ln(1-p) \cdot N}) > 2ln(2) \cdot N \tag{2}$$

Which is equivalent to $-\frac{n(n-1)}{2N} < ln(1-p)$. Thus use (1) we have

$$Pr[\bar{X}] \leq exp(-\frac{n(n-1)}{2N}) < 1 - p$$

So we have

$$Pr[X] > p$$

□

### 2

Lemma 1.1. For positive integer $n < N$, we have

$$\sum_{i=1}^{n-1} ln(1 - \frac{i}{N}) > -\frac{n^2}{N}$$

*Proof.* Notice that $\forall x \in [1-\frac{i+1}{N}, 1-\frac{i}{N}]$ $(0 \le i \le n)$, we have $ln(x) < ln(1-\frac{i}{N})$. Thus

$$\frac{1}{N} ln(1-\frac{i}{N}) \ge \int_{1-\frac{i+1}{N}}^{1-\frac{i}{N}} ln(x)dx$$

Thus we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{n-1} ln(1-\frac{i}{N}) &\ge \int_{1-\frac{n}{N}}^{1} ln(x)dx \\
&= (xln(x)-x)|_{1-\frac{n}{N}}^{1} \\
&= -\frac{n}{N} - (1-\frac{n}{N})ln(1-\frac{n}{N}) \\
&> -\frac{n}{N} - (1-\frac{n}{N})(-\frac{n}{N}) \\
&= -\frac{n^2}{N^2}
\end{aligned} \tag{3}$$

Thus

$$\sum_{i=1}^{n-1} ln(1-\frac{i}{N}) > -\frac{n^2}{N}$$

□

Theorem 1.2. Let $S = \{1, 2, \cdots, N\}$. For n times, uniformly randomly draw one element from set S with replacement. Let $x_t$ be the element we draw at time t. Then $\forall p > 0$, there exists a constant $C_2$ such that when $n \le C_2\sqrt{N}$, we have

$$Pr[\exists 1 \le i, j \le n, i \ne j \text{ such that } x_i = x_j] < p$$

.

*Proof.* Let X denote the event that $\exists i, j \le t, i \ne j$ such that $x_i = x_j$.

Use Lemma 1.1, we have

$$\begin{aligned}
Pr[\bar{X}] &= \frac{N(N-1)\cdots(N-n+1)}{N^n} \\
&= \prod_{i=1}^{n-1}(1-\frac{i}{N}) \\
&= exp(\sum_{i=1}^{n-1} ln(1-\frac{i}{N})) \\
&> exp(-\frac{n^2}{N})
\end{aligned} \tag{4}$$

Let $C_2 = \sqrt{-ln(1-p)}$. Then when $n \le C_2\sqrt{N}$, we have

$$exp(-\frac{n^2}{N}) \ge 1-p$$

So $Pr[\bar{X}] > 1-p$, thus

$$Pr[X] < p$$

□

# 2

Theorem 2.1. Let $S = \{1, 2, \cdots, N\}$. Let $D_1 : S \to R^+ \cup \{0\}$ be a discrete probability distribution over S. For n times, randomly draw one element from set S according to distribution $D_1$ with replacement . Let $x_t$ be the element we draw at time t. Let $D_0$ be the uniform distribution over S, which satisfies $\forall i \in S, D_0(i) = \frac{1}{N}$. Then we have

$$Pr_{D_1^n}[\exists 1 \leq i, j \leq n, i \neq j \text{ such that } x_i = x_j] \geq Pr_{D_0^n}[\exists 1 \leq i, j \leq n, i \neq j \text{ such that } x_i = x_j]$$

.

*Proof.* Let X denote the event that $\exists i, j \leq t, i \neq j$ such that $x_i = x_j$. Let $X_m$ denote the event that $\exists 1 \leq i, j \leq n, i \neq j$ such that $x_i = x_j = m$.

First, to change $D_1$ to $D_0$, we can apply the following algorithm:

1. $t := 1$
2. While $D_t \neq D_0$:
3. find $i, j \in S$ such that $D_t[i] < \frac{1}{N} < D_t[j]$
4. let $D_{t+1}[j] := D_t[i] + D_t[j] - \frac{1}{N}, D_{t+1}[i] := \frac{1}{N}, \forall k \neq i, j, D_{t+1}[k] := D_t[k]$
5. $t++$
6. End While

Since the number of $\frac{1}{N}$ in D increases at each iteration, this algorithm will terminate in N steps.

We only need to prove that

$$\forall t, Pr_{D_t^n}[X] \geq Pr_{D_{t+1}^n}[x]$$

Without losing generality, suppose when generate $D_{t+1}$ from $D_t$, we choose $i = 1, j = 2$.

Let $Y$ be the number of times that the element we draw is in $\{1, 2\}$.

$$\begin{aligned} Pr_{D^n}[X] &= Pr_{D^n}[\bigcup_{k=3}^{N} X_k] + (1 - Pr_{D^n}[\bigcup_{k=3}^{N} X_k])Pr_{D^n}(X_1 \cup X_2 | \bigcap_{k=3}^{N} \bar{X}_k) \\ &= Pr_{D^n}[\bigcup_{k=3}^{N} X_k] + (1 - Pr_{D^n}[\bigcup_{k=3}^{N} X_k]) \sum_{i=0}^{\infty} Pr_{D^n}(Y = i | \bigcap_{k=3}^{N} \bar{X}_k) Pr_{D^n}(X_1 \cup X_2 | Y = i) \end{aligned} \tag{5}$$

The last equation holds because $\forall i, Pr_{D^n}(X_1 \cup X_2 | Y = i) = Pr_{D^n}(X_1 \cup X_2 | Y = i, \bigcap_{k=3}^{N} \bar{X}_k)$.

Notice that

$$\begin{aligned} & Pr_{D_t^n}(X_1 \cup X_2 | Y = 2) - Pr_{D_{t+1}^n}(X_1 \cup X_2 | Y = 2) \\ &= \frac{1}{(D_t[1] + D_t[2])^2}(D_t[1]^2 + D_t[2]^2 - (\frac{1}{N})^2 - (D_t[1] + D_t[2] - \frac{1}{N})^2) \\ &= -\frac{2}{(D_t[1] + D_t[2])^2}(D_t[1] - \frac{1}{N})(D_t[2] - \frac{1}{N}) \\ &> 0 \end{aligned} \tag{6}$$

And for any distribution D over S we have

$$Pr_{D^n}(X_1 \cup X_2 | Y = i) = \begin{cases} 0 & i = 0, 1 \\ 1 & i \geq 3 \end{cases} \tag{7}$$

Thus

$$\forall i, Pr_{D_t^n}(X_1 \cup X_2 | Y = i) \geq Pr_{D_{t+1}^n}(X_1 \cup X_2 | Y = i)$$

Since we only adjust $D_t[0], D_t[1]$,

$$Pr_{D_t^n}[\bigcup_{k=3}^{N} X_k] = Pr_{D_{t+1}^n}[\bigcup_{k=3}^{N} X_k]$$

$$\forall i, Pr_{D_t^n}(Y = i| \bigcap_{k=3}^{N} \bar{X}_k) = Pr_{D_{t+1}^n}(Y = i| \bigcap_{k=3}^{N} \bar{X}_k)$$

So consider equation (6) and we get

$$Pr_{D_t^n}[X] \geq Pr_{D_{t+1}^n}[x]$$

$\square$

**Acknowledgement:**