The Birthday Paradox

(Slides for ESTR2102)

Yufei Tao

Department of Computer Science and Engineering
Chinese University of Hong Kong

The Birthday Paradox

Suppose that we choose n people uniformly at random. We **succeed** if all these people have distinct birthdays. Obviously, the larger n is, the less likely we would succeed.

We will show that, even with $n = 23$, the probability of succeeding already drops to below 50%.

The Birthday Paradox — In General

Suppose that we pick n integers from the domain $[1, u]$. What is the probability that all the integers are distinct?

Answer:

$$\frac{u-1}{u} \cdot \frac{u-2}{u} \cdot \ldots \cdot \frac{u-n+1}{u}$$

The Birthday Paradox — In General

Let us upper bound the probability:

$$\begin{aligned} & \left(1-\frac{1}{u}\right)\left(1-\frac{2}{u}\right)\dots\left(1-\frac{n-1}{u}\right) \\ < \quad & e^{-\frac{1}{u}} \cdot e^{-\frac{2}{u}} \cdot \ldots \cdot e^{-\frac{n-1}{u}} \\ = \quad & e^{-\frac{n(n-1)}{2u}}. \end{aligned}$$

For all $x \neq 0$, it holds that $1 + x < e^x$.

Back to the "birthday question" we asked in the first place. When $n = 23, u = 365$:

$$e^{-\frac{n(n-1)}{2u}} \quad < \quad 0.5.$$

Random IDs with No Collisions

In practice, we often need to assign random ids to n objects (e.g., for generating a random URL). But we do not want the ids to collide. One easy way to do so is the following algorithm:

1. **for** $i = 1$ to n
2. $\quad x = \text{RANDOM}(1, u)$
3. $\quad$ assign x as the id to the i-th object

We want to make sure that the probability of collision is at most $1/2$. How large should u be?

> **Note:** Our earlier analysis gives an upper bound of the distinct probability. But here we need a lower bound. Next, we will achieve the purpose with a different analysis.

Random IDs with No Collisions

For any distinct $i, j \in [1, n]$, define $X_{ij} =$

- 1, if objects i and j have the same id.
- 0, otherwise.

Clearly:

$$\begin{aligned} \boldsymbol{Pr}[X_{ij} = 1] &= 1/u \\ \boldsymbol{E}[X_{ij}] &= 1/u \end{aligned}$$

Random IDs with No Collisions

Define:

$$X = \sum_{\text{all distinct } i,j} X_{ij}$$

Note that $X = 0$ implies no collisions.

We know:

$$\begin{aligned} \boldsymbol{E}[X] &= \sum_{\text{all distinct } i,j} \boldsymbol{E}[X_{ij}] \\ &= \frac{n(n-1)}{2u} \end{aligned}$$

Random IDs with No Collisions

Let us set $u = n^2$ so that

$$\boldsymbol{E}[X] \;=\; \frac{n(n-1)}{2u} < 1/2.$$

Next we prove that $\boldsymbol{Pr}[X = 0] \geq 1/2$, namely, all ids are distinct with probability at least $1/2$.

Random IDs with No Collisions

Lemma: $\boldsymbol{Pr}[X \geq 1] \leq 1/2$.
Proof: If this is not true, then

$$\begin{aligned} \boldsymbol{E}[X] &= \sum_{i=1}^{\infty} i \cdot \boldsymbol{Pr}[X = i] \\ &\geq \sum_{i=1}^{\infty} \boldsymbol{Pr}[X = i] \\ &= \boldsymbol{Pr}[X \geq 1] \\ &> 1/2 \end{aligned}$$

giving a contradiction. $\square$

Random IDs with No Collisions

Let us set $u = n^3$ so that

$$\boldsymbol{E}[X] \;=\; \frac{n(n-1)}{2u} < 1/n.$$

Next we prove that $\boldsymbol{Pr}[X = 0] \geq 1 - 1/n$, namely, all ids are distinct with probability at least $1 - 1/n$ (the probability approaches 1 very quickly as n grows).

Random IDs with No Collisions

Lemma: $\boldsymbol{Pr}[X \geq 1] \leq 1/n$.
Proof: If this is not true, then

$$\begin{aligned} \boldsymbol{E}[X] &= \sum_{i=1}^{\infty} i \cdot \boldsymbol{Pr}[X = i] \\ &\geq \sum_{i=1}^{\infty} \boldsymbol{Pr}[X = i] \\ &= \boldsymbol{Pr}[X \geq 1] \\ &> 1/n \end{aligned}$$

giving a contradiction. $\square$