Coupon Collector’s Problem

Lecture Notes of Randomized Algorithms

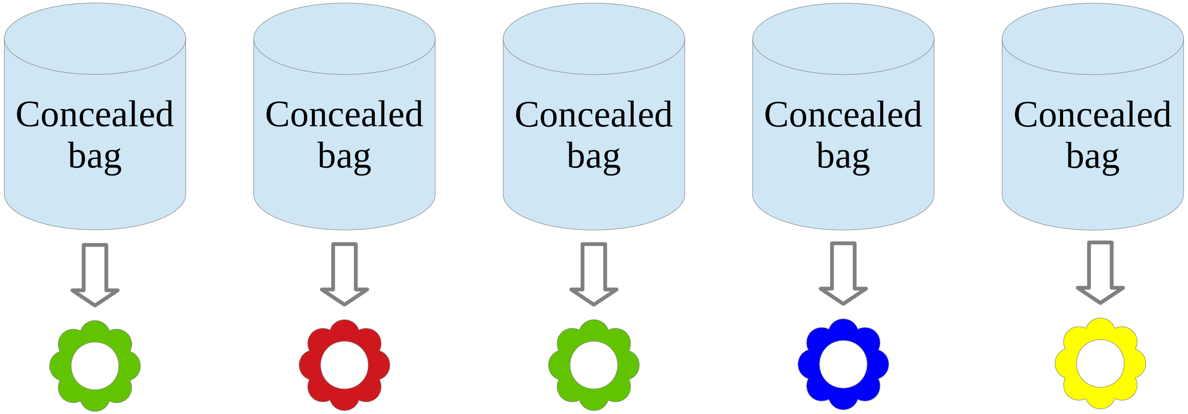
Joseph Chuang-Chieh Lin
Dept. CSIE, Tamkang University

Coupon Collector’s Problem

▲一名網友在臉書社團「爆廢公社公開版」表示，多年前為了收集連鎖便利商店 7-11 的一款贈品，花了不少金額，還拿到許多重複的款式，貼文引發 3 千多名網友共鳴。（圖／翻攝自爆廢公社公開版）

Coupon Collector's Problem

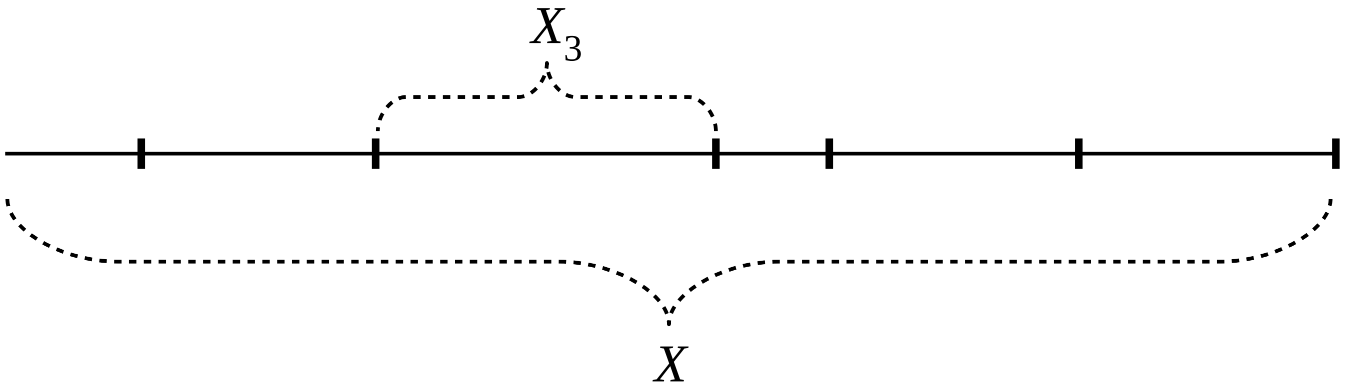
- Have you already got all of them (totally *n* types)?
- Have you ever thought about how much you should pay for them?



- Each bag is chosen independently and uniformly at random from the *n* possibilities.

Coupon Collector's Problem

- Let X be the number of bags bought until every type of coupon is obtained.
- Let X_i be the number of bags bought while you had already got exactly $i-1$ different coupons.



Coupon Collector’s Problem

- Let X be the number of bags bought until every type of coupon is obtained.
- Let X_i be the number of bags bought while you had already got exactly $i-1$ different coupons.
  - Geometric random variables?!

Geometric Distribution

- Imagine: *flip a coin* ***until it lands on a head***.
  - What's the distribution of the number of flips?

Geometric Distribution

- Imagine: *flip a coin* ***until it lands on a head***.
  - What's the distribution of the number of flips?

- Definition. A geometric random variable *X* with parameter *p* is

$$\Pr[X = n] = (1 - p)^{n-1} p.$$

for $n = 1, 2, \ldots$

Memoryless

- Let X be a geometric random variable X with parameter $p > 0$.
- For any n, $k > 0$, $\Pr[X = n + k \mid X > k] = \Pr[X = n]$.

The mean of a geometric r.v. $X(p)$

$$\begin{aligned}\mathbf{E}[X] &= \sum_{j=1}^{\infty} j\,\Pr[X=j] \\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{j}\Pr[X=j] \\ &= \sum_{i=1}^{\infty}\sum_{j=i}^{\infty}\Pr[X=j] \\ &= \sum_{i=1}^{\infty}\Pr[X\ge i].\end{aligned}$$

$$\mathbf{E}[X] = \sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

$$\Pr[X\ge i] = \sum_{k=i}^{\infty}(1-p)^{k-1}p = (1-p)^{i-1}$$

Coupon Collector's Problem

- Let X be the number of bags bought until every type of coupon is obtained.
- Let X_i be the number of bags bought while you had already got exactly $i-1$ different coupons.
  - Geometric random variables?!
  - What about

$$X = \sum_{i=1}^{n} X_i?$$

Coupon Collector's Problem

- When exactly $i-1$ coupons have been collected, the probability of obtaining a new one is

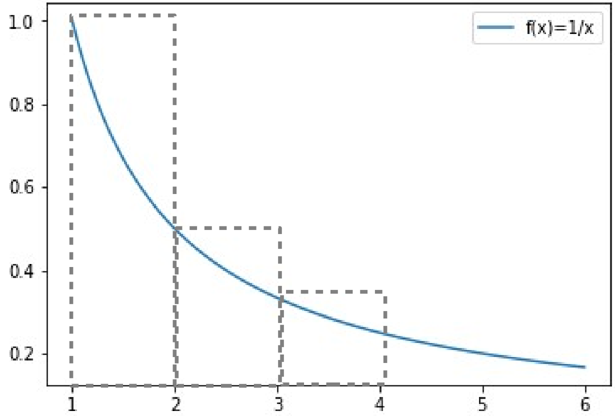
$$p_i = 1 - \frac{i-1}{n}$$

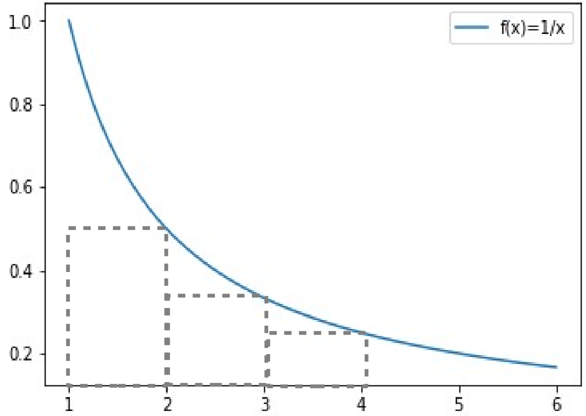
- X_i is a geometric random variable, so

$$\mathbf{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}.$$

Coupon Collector's Problem (contd.)

- $$\begin{aligned}\mathbf{E}[X] &= \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\ &= \sum_{i=1}^{n} \frac{n}{n-i+1} \\ &= n \cdot \boxed{\sum_{i=1}^{n} \frac{1}{i}}.\end{aligned}$$





$$\sum_{k=1}^{n} \frac{1}{k} \geq \int_{x=1}^{n} \frac{1}{x} dx = \ln n \qquad \sum_{k=2}^{n} \frac{1}{k} \leq \int_{x=1}^{n} \frac{1}{x} dx = \ln n$$

$$H(n) = \sum_{i=1}^{n} \frac{1}{i} = \ln n + \Theta(1).$$

What if somebody wants to sell you the whole set for NTD $ 30,000?

- You are about to buy $n \ln n + \Theta(n)$ bags for collecting all the coupons (stickers)!
  - 34 different coupons (each costs NTD $77) require you

    $\approx \text{NTD\$}\, 77 \times 34 \ln 34 \approx \text{NTD\$}\, 9,232.$

or $\approx 2 \times \text{NTD\$}\, 77 \times 34 \ln 34 \approx \text{NTD\$}\, 18,464.$

What if somebody wants to sell you the whole set for NTD $ 30,000?

- You are about to buy $n \ln n + \Theta(n)$ bags for collecting all the coupons (stickers)!
  - 34 different coupons (each costs NTD $77) require you

$$\approx \text{NTD\$}\, 77 \times 34 \ln 34 \approx \text{NTD\$}\, 9,232.$$

or $$\approx 2 \times \text{NTD\$}\, 77 \times 34 \ln 34 \approx \text{NTD\$}\, 18,464.$$

- How likely is this to happen?

Be a smart buyer!

- You are about to buy $n \ln n + \Theta(n)$ bags for collecting all the coupons (stickers)!
  - 34 different coupons (each costs NTD $77) require you

$$\approx \text{NTD\$}\, 77 \times 34 \ln 34 \approx \text{NTD\$}\, 9,232.$$

or $$\approx 2 \times \text{NTD\$}\, 77 \times 34 \ln 34 \approx \text{NTD\$}\, 18,464.$$

- How likely is this to happen?

$$\Pr[X \geq 2nH_n] \leq \frac{1}{2}.$$

Markov's Inequality

- Let X be a random variable that assumes only non-negative values. Then, for all $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

Andrei Andreyevich Markov (Wikipedia)
1856–1922