# Randomized Algorithms

## Coupon Collector’s Problem
## and
## Conditional Expectation

Joseph Chuang-Chieh Lin
Dept. CSIE, Tamkang University

# Review

- Expectation of discrete random variables
- Linearity of expectation.
- Bernoulli and Binomial random variable

# Expectation

- The expectation of a discrete random variable $X$, denoted by $\mathbf{E}[X]$, is

$$\mathbf{E}[X] = \sum_i i \cdot \Pr[X = i]$$

- Example: Let $X$ denote the sum of of dices:

$$\mathbf{E}[X] = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \cdots + \frac{1}{36} \cdot 12 = 7.$$

Adam Fejes@adamfjs

# Linearity of Expectation

- For any finite collection of discrete random variables $X_1$, $X_2$, ..., $X_n$ with finite expectations,

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

- For any constant c and discrete random variable $X$,

$$\mathbf{E}[cX] = c \cdot \mathbf{E}[X].$$

- Why is it useful?

# Example

- Consider the dice-throwing example again.
  - $X_1$ : the outcome of die 1
  - $X_2$ : the outcome of die 2

$$\mathbf{E}[X_1] = \mathbf{E}[X_2] = \frac{1}{6} \cdot \sum_{j=1}^{6} j = \frac{7}{2}$$

$$\mathbf{E}[X] = \mathbf{E}[X_1 + X_2] = 7.$$

Adam Fejes@adamfjs

# Bernoulli random variable

- Suppose we run an experiment that succeeds with probability $p$ and fails with probability $1-p$.

$$Y = \begin{cases} 1 & \text{if the experiment succeeds,} \\ 0 & \text{otherwise.} \end{cases}$$


- $Y$: Bernoulli random variable.
  - or *indicator random variable*.

$$\mathbf{E}[Y] = 1 \cdot \Pr[Y = 1] + 0 \cdot \Pr[Y = 0] = \Pr[Y = 1] = p.$$

# Binomial random variable

- A binomial random variable $X$ with parameters $n$ and $p$, denoted by $B(n, p)$, is defined as

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

for $j = 0, 1, 2, \ldots, n$.

- Exercise: Show that $\sum_{j=0}^{n} \Pr[X = j] = 1.$

# Binomial random variable

- A binomial random variable $X$ with parameters $n$ and $p$, denoted by $B(n, p)$, is defined as

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

for $j = 0, 1, 2, \ldots, n$.

$X$: the number of successful trials in the $n$ experiments.

- <u>Exercise</u>: Show that $\sum_{j=0}^{n} \Pr[X = j] = 1.$

# Binomial random variable (expectation)

- $$\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \\
&= np.
\end{aligned}$$

# Binomial random variable (expectation)

- $$\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \\
&= np.
\end{aligned}$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

# Let's make it simpler!

- Denote a set of $n$ Bernoulli random variables $X_1, X_2, \ldots, X_n$.
  - $X_i = 1$ if the $i$th trial is successful and 0 otherwise.
  - $X = X_1 + X_2 + \ldots + X_n$

# Let's make it simpler!

- Denote a set of $n$ Bernoulli random variables $X_1, X_2, \ldots, X_n$.
  - $X_i = 1$ if the $i$th trial is successful and 0 otherwise.
  - $X = X_1 + X_2 + \ldots + X_n$
  - Compute $\mathbf{E}[X]$ using linearity of expectation:

# Let's make it simpler!

- Denote a set of $n$ Bernoulli random variables $X_1, X_2, \ldots, X_n$.
  - $X_i = 1$ if the $i$th trial is successful and 0 otherwise.
  - $X = X_1 + X_2 + \ldots + X_n$
  - Compute $\mathbf{E}[X]$ using linearity of expectation:

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np.$$

# 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?

- <u>Definition</u>. A geometric random variable *X* with parameter *p* is

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

  for $n = 1, 2, \ldots$

- <u>Exercise</u>. Show that $\sum_{n\geq 1} \Pr[X = n] = 1.$

# Memoryless

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

# Memoryless

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

$$\begin{aligned}
\Pr[X = n + k \mid X > k] &= \frac{\Pr[(X = n + k) \cap (X > k)]}{\Pr[X > k]} \\
&= \frac{\Pr[X = n + k]}{\Pr[X > k]} \\
&= \frac{(1-p)^{n+k-1}p}{\sum_{i=k}^{\infty}(1-p)^i p} \\
&= \frac{(1-p)^{n+k-1}p}{(1-p)^k} \\
&= (1-p)^{n-1}p = \Pr[X = n].
\end{aligned}$$

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

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

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

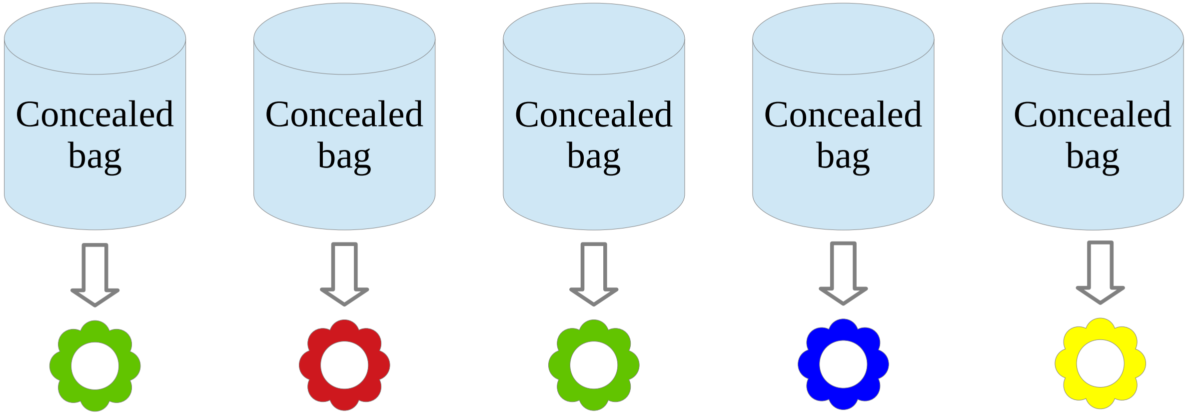
$$\Pr[X \ge i] = \sum_{k=i}^{\infty} (1-p)^{k-1} p = (1-p)^{i-1}$$

# Coupon Collector’s Problem

▲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

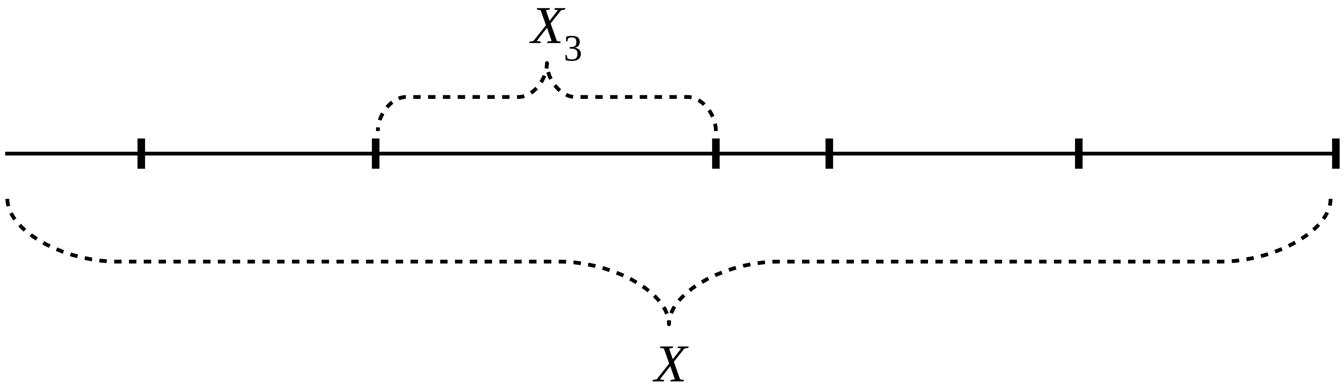
- Have you already got all of them (totally $n$ types)?
- Have you ever thought about how much you should pay for them?




- 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?!

# 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

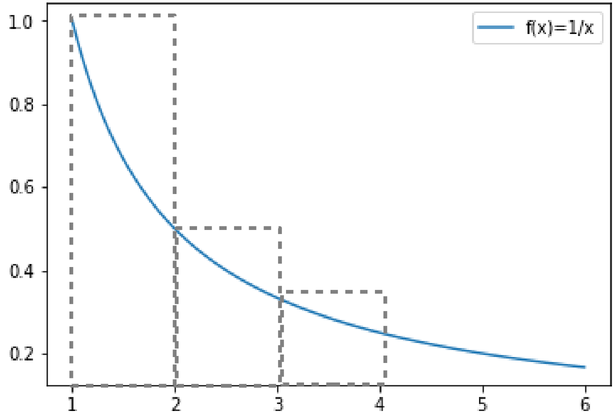
$$p_i = 1 - \frac{i-1}{n}$$

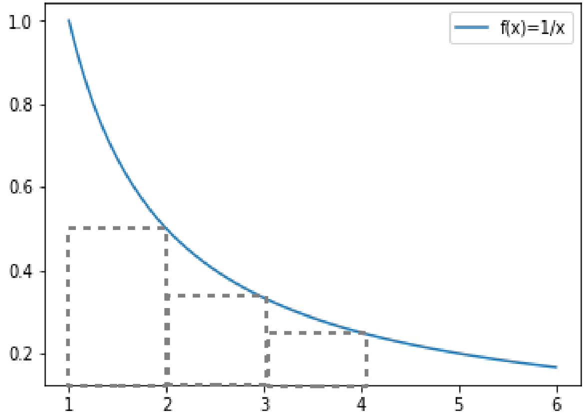
- $X_i$ is a geometric random variable, so

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

# Coupon Collector's Problem (contd.)

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




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

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

# Coupon Collector's Problem (contd.)

- So, you are about to buy $n \ln n + \Theta(n)$ bags for collecting all the coupons (stickers)!

# On conditional expectation

- There are terminologies which may confusing you.

# Conditional Expectation

- Definition.

$$\mathbf{E}[Y \mid Z = z] = \sum_{y} y \Pr[Y = y \mid Z = z].$$

- Example of two dices.
  - $X_1$: the number showing on the first die
  - $X_2$: the number showing on the second die
  - $X = X_1 + X_2$

$$\mathbf{E}[X \mid X_1 = 2] = \sum_{x=3}^{8} x \cdot \frac{1}{6} = \frac{11}{2}.$$

# Conditional Expectation (contd.)

- Lemma. For any random variables $X$ and $Y$,

$$\mathbf{E}[X] = \sum_y \Pr[Y = y] \mathbf{E}[X \mid Y = y].$$

- *Proof.*

$$\begin{aligned}
\sum_y \Pr[Y = y] \cdot \mathbf{E}[X \mid Y = y] &= \sum_y \Pr[Y = y] \cdot \sum_x x \Pr[X = x \mid Y = y] \\
&= \sum_x \sum_y x \Pr[X = x \mid Y = y] \cdot \Pr[Y = y] \\
&= \sum_x \sum_y x \Pr[X = x \cap Y = y] \\
&= \sum_x x \Pr[X = x] \\
&= \mathbf{E}[X].
\end{aligned}$$

# Conditional Expectation

- Lemma. For any finite collection of discrete random variables $X_1$, $X_2$, ..., $X_n$ with finite expectations and for any random variable $Y$,

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i \,\middle|\, Y = y\right] = \sum_{i=1}^{n} \mathbf{E}[X_i \mid Y = y].$$

# Conditional Expectation (contd.)

- A weird definition.
- $\mathbf{E}[Y \mid Z]$ : regarded as a **random variable** $f(Z)$.
  - It takes on the value $\mathbf{E}[Y \mid Z = z]$ when $Z = z$.

- In the previous example,

$$\mathbf{E}[X \mid X_1] = \sum_x x \cdot \Pr[X = x \mid X_1] = \sum_{X_1+1}^{X_1+6} x \cdot \frac{1}{6} = X_1 + \frac{7}{2}.$$

- So it makes sense that

$$\mathbf{E}[\mathbf{E}[X \mid X_1]] = \mathbf{E}\left[X_1 + \frac{7}{2}\right] = \frac{7}{2} + \frac{7}{2} = 7.$$

# Conditional Expectation (contd.)

- Theorem. $\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y \mid Z]]$.

- *Proof.*

$$\mathbf{E}[\mathbf{E}[Y \mid Z]] = \sum_z \mathbf{E}[Y \mid Z = z] \cdot \Pr[Z = z] = \mathbf{E}[Y].$$