Randomized Algorithms

Moments and Deviations

Joseph Chuang-Chieh Lin
Dept. CSIE, Tamkang University

25th March 2021

Review

- The Secretary Problem.

The Secretary Problem

- Consider the problem of hiring an office secretary.
  - We interview candidates, coming one by one, on a rolling basis.
  - Let's say the *i*th candidate has a value $v_i \in \mathbf{R}$ which stands for how much we like her.
  - At some time point, we would like to hire the best candidate we have seen so far.
  - Suppose we can fire the old one and hire a new better candidate.
  - Assume that we only want to interview at most *n* candidates.

The Secretary Problem (contd.)

- The whole hiring process will be just like:

```
Randomly shuffle the n candidates.
Set TheOne ← 0
for i ← 1 to n do:
    interview candidate i
    if v_i > v_TheOne then:
        TheOne ← i
        Hire candidate i
```

- Isn't it very simple?

The cost

- No pain, no gain. We cannot reap without sowing.

- Let c_I be the cost associated with interviewing a candidate.
- Let c_H be the cost associated with hiring a candidate.

- So, if totally we have ever hired m people ($m-1$ was fired though…), the total cost will be $O(c_I n + c_H m)$

The expected cost analysis

- Let X_i be an indicator random variable such that

$$\begin{cases} X_i = 1 & \text{if candidate } i \text{ is hired} \\ X_i = 0 & \text{otherwise} \end{cases}$$

- $X = \sum_{i=1}^{n} X_i$: the number of times we hire a new candidate.

- $\Pr[X_i] = ?$

The expected cost analysis

- Let X_i be an indicator random variable such that

$$\begin{cases} X_i = 1 & \text{if candidate } i \text{ is hired} \\ X_i = 0 & \text{otherwise} \end{cases}$$

- $X = \sum_{i=1}^{n} X_i$: the number of times we hire a new candidate.

- $\Pr[X_i] = \Pr[\text{candidate } i \text{ is better than previous } i-1 \text{ candidates}] = \dfrac{1}{i}.$

The expected cost analysis

- Let X_i be an indicator random variable such that

$$\begin{cases} X_i = 1 & \text{if candidate } i \text{ is hired} \\ X_i = 0 & \text{otherwise} \end{cases}$$

- $X = \sum_{i=1}^{n} X_i$: the number of times we hire a new candidate.

- $\Pr[X_i] = \Pr[\text{candidate } i \text{ is better than previous } i-1 \text{ candidates}] = \frac{1}{i}.$
  - For randomly chosen *i* numbers, the probability that the *i*th number is the biggest.

The expected cost analysis

- Therefore,

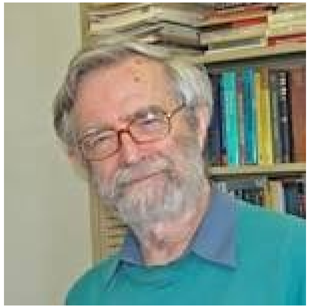
$$\begin{aligned}\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] &= \sum_{i=1}^{n} \Pr[X_i = 1] \\ &= \sum_{i=1}^{n} \frac{1}{i} \\ &= H_n \\ &= \ln n + \Theta(1).\end{aligned}$$

- The expected cost is

$$O(c_H \ln n + c_I n).$$

The classic version

- Reference:
  - Thomas S. Ferguson: Who solved the Secretary Problem? *Statistical Science*, Vol. 4 (1989), pp. 282−289.



2. STATEMENT OF THE PROBLEM

The reader's first reaction to the title might well be to ask, "Which secretary problem?". After all, as I have just implied, there are many variations on the problem. The secretary problem *in its simplest form* has the following features.

1. There is one secretarial position available.
2. The number n of applicants is known.
3. The applicants are interviewed sequentially in random order, each order being equally likely.
4. It is assumed that you can rank all the applicants from best to worst without ties. The decision to accept or reject an applicant must be based only on the relative ranks of those applicants interviewed so far.
5. An applicant once rejected cannot later be recalled.
6. You are very particular and will be satisfied with nothing but the very best. (That is, your payoff is 1 if you choose the best of the n applicants and 0 otherwise.)

A simple solution

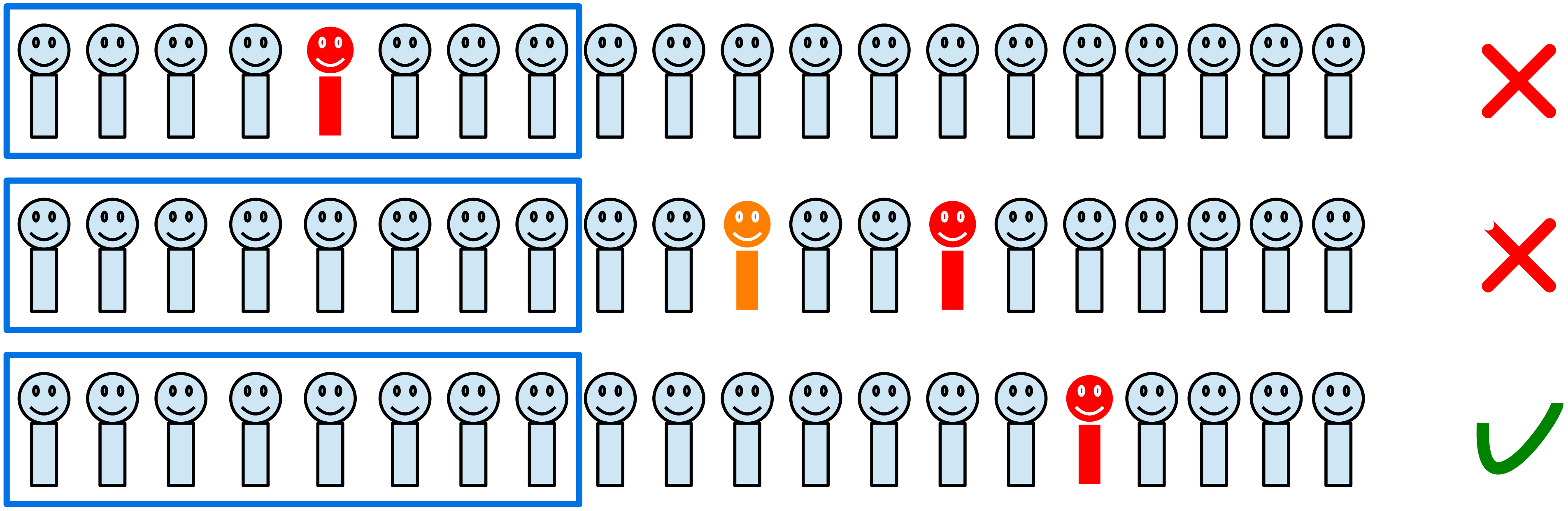
- Reject the first $r-1$ applicants.
- Choose the next applicant who is the best in the relative ranking of the observed applicants.

- The famous 37% rule.

$$\frac{r}{n} \approx 37\% \quad \text{as } n \text{ is large}$$

Refer to https://www.books.com.tw/products/F014054315

Illustration



- Let’s start the mathematics…
- Moments and Deviations.
- Basic tail bounds.

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

Markov's Inequality

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

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

- *Proof*:

Let $I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$. Since $X \geq 0$, $I \leq \dfrac{X}{a}$

$$\Pr[X \geq a] = \Pr[I = 1] = \mathbf{E}[I] \leq \mathbf{E}\left[\frac{X}{a}\right] = \frac{\mathbf{E}[X]}{a}.$$

Andrei Andreyevich Markov (Wikipedia)
1856–1922

Example: 75% heads in fair coin flips

- What's the probability of obtaining $> 3n/4$ heads in a sequence of n fair coin flips?

Example: 75% heads in fair coin flips

- What's the probability of obtaining > $3n/4$ heads in a sequence of n fair coin flips?

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2}$$

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- ✓ Applying Markov's inequality,

Example: 75% heads in fair coin flips

- What's the probability of obtaining > 3*n*/4 heads in a sequence of *n* fair coin flips?

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2}$$

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

  - ✓ Applying Markov's inequality,

$$\Pr[X \geq 3n/4] \leq \frac{\mathbf{E}[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}.$$

Scenarios of applying Markov's inequality

- Markov's inequality gives the best tail bound when:
  - All we know is the expectation of the random variable
  - The random variable is non-negative.

The kth moment

- <u>Definition</u>. The kth of a random variable X is $\mathbf{E}[X^k]$.
  - So, the expectation is the *first moment* of X.

The variance

- Definition. The variance of a random variable X is defined as

$$\mathbf{Var}[X] = \mathbf{E}[(X - \boldsymbol{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

The variance

- <u>Definition</u>. The variance of a random variable X is

$$\mathbf{Var}[X] = \mathbf{E}[(X - \boldsymbol{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

- <u>Definition</u>. The standard deviation of a random variable X is

$$\sigma[X] = (\mathbf{Var}[X])^{1/2}.$$

- <u>Definition</u>. The covariance of two random variables X and Y is

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \boldsymbol{E}[X])(Y - \boldsymbol{E}[Y])].$$

Variance of the sum of two random variables

- Theorem. For any two random variables X and Y,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\,\mathbf{Cov}(X, Y).$$

Variance of the sum of two random variables

- <u>Theorem</u>. For any two random variables X and Y,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\,\mathbf{Cov}(X, Y).$$

Not as nice as the linearity of expectation!

Variance of the sum of two random variables

- Theorem. For any two random variables X and Y,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\,\mathbf{Cov}(X, Y).$$

- *Proof*:

$$\begin{aligned}
\mathbf{Var}[X+Y] &= \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2] \\
&= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2] \\
&= \mathbf{E}[((X-\mathbf{E}[X])+(Y-\mathbf{E}[Y]))^2] \\
&= \mathbf{E}[(X-\mathbf{E}[X])^2+(Y-\mathbf{E}[Y])^2+2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])] \\
&= \mathbf{E}[(X-\mathbf{E}[X])^2]+\mathbf{E}[(Y-\mathbf{E}[Y])^2]+2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])] \\
&= \mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}(X,Y).
\end{aligned}$$

Expectation of product of two random variables

- Theorem. If X and Y are two **independent** random variables, then

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

Expectation of product of two random variables

- <u>Theorem</u>. If X and Y are two **independent** random variables, then

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

Proof:

$$\begin{aligned}
\mathbf{E}[X \cdot Y] &= \sum_i \sum_j (i \cdot j) \cdot \Pr[(X = i) \cap (Y = j)] \\
&= \sum_i \sum_j (i \cdot j) \cdot \Pr[X = i] \cdot \Pr[Y = j] \\
&= \left(\sum_i i \cdot \Pr[X = i] \right) \cdot \left(\sum_j \cdot \Pr[Y = j] \right) \\
&= \mathbf{E}[X] \cdot \mathbf{E}[Y].
\end{aligned}$$

Independence → Linearity of Variance

- Corollary. If X and Y are independent random variables, then

$$\mathbf{Cov}(X, Y) = 0$$

and

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Independence → Linearity of Variance

- Corollary. If X and Y are independent random variables, then

$$\mathbf{Cov}(X, Y) = 0$$

and

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Proof:

$$\begin{aligned}
\mathbf{Cov}[X, Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\
&= \mathbf{E}[X - \mathbf{E}[X]] \cdot \mathbf{E}[Y - \mathbf{E}[Y]] \\
&= 0.
\end{aligned}$$

Remark

- Theorem. For mutually independent random variables X_1, X_2, …, X_n

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

Example: Variance of a binomial random variable

- **Recall**: a binomial random variable X can be regarded as the sum of n independent Bernoulli trials (Y's), each with success probability p.

Example: Variance of a binomial random variable

- **Recall**: a binomial random variable X can be regarded as the sum of n independent Bernoulli trials (Y's), each with success probability p.

$$\mathbf{Var}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])^2] = p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2 = p(1-p).$$

Example: Variance of a binomial random variable

- **Recall**: a binomial random variable X can be regarded as the sum of n independent Bernoulli trials (Y's), each with success probability p.

$$\mathbf{Var}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])^2] = p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2 = p(1-p).$$

- By the theorem in p.19,

$$\mathbf{Var}[X] = n \cdot (p(1-p)) = np(1-p).$$

Chebyshev’s Inequality

- A stronger tail bound if you have the expectation and the variance.

- Theorem [Chebyshev’s Inequality]. For any $a > 0$,

$$\Pr[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{Var}[X]}{a^2}.$$

Chebyshev's Inequality

- <u>Theorem</u> [Chebyshev's Inequality]. For any $a > 0$,

$$\Pr[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{Var}[X]}{a^2}.$$

- *Proof*:

$$\Pr[|X - \mathbf{E}[X]| \geq a] = \Pr[(X - \mathbf{E}[X])^2 \geq a^2]$$

Apply Markov's inequality,

$$\Pr[(X - \mathbf{E}[X])^2 \geq a^2] \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} = \frac{\mathbf{Var}[X]}{a^2}.$$

Chebyshev's Inequality

- Corollary. For any $t > 1$,

$$\Pr[|X - \mathbf{E}[X])| \geq t \cdot \sigma[X]] \leq \frac{1}{t^2}.$$

$$\Pr[|X - \mathbf{E}[X])| \geq t \cdot \mathbf{E}[X]] \leq \frac{\mathbf{Var}[X]}{t^2(\mathbf{E}[X])^2}.$$

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually, $$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \frac{n}{4}.$$

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually, $$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \frac{n}{4}.$$

$$\Pr[X \geq 3n/4] \leq \Pr[|X - \mathbf{E}[X]| \geq n/4] \leq \frac{\mathbf{Var}[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

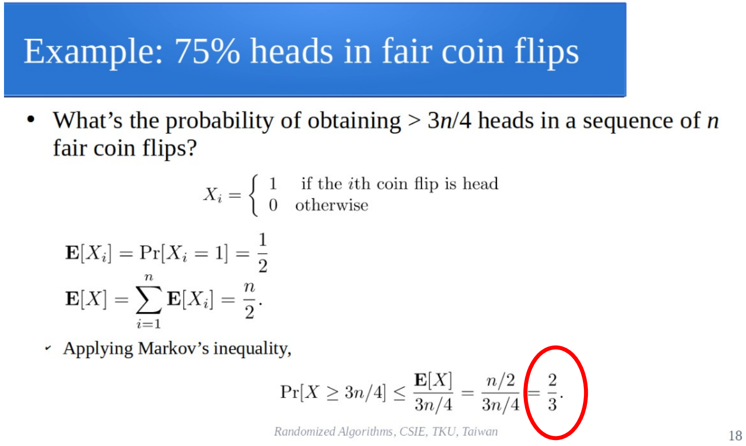
- Actually,

$$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \frac{n}{4}.$$

$$\Pr[X \geq 3n/4] \leq \Pr[|X - \mathbf{E}[X]| \geq n/4] \leq \frac{\mathbf{Var}[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$


Example: 75% heads in fair coin flips

- What's the probability of obtaining > $3n/4$ heads in a sequence of n fair coin flips?

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2}$$

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- Applying Markov's inequality,

$$\Pr[X \geq 3n/4] \leq \frac{\mathbf{E}[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}.$$

Randomized Algorithms, CSIE, TKU, Taiwan 18

Example: Coupon Collector's Problem

- Recall:
  - X: the time to collect n coupons.
  - $\mathbf{E}[X] \approx nH_n$.

- Using Markov's inequality,

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

Example: Coupon Collector's Problem

- Recall:
  - X: the time to collect n coupons (sum of geometric random variables).
  - $\mathbf{E}[X] \approx nH_n$.

- Using Markov's inequality,

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

- It seems to be weak. Can we get a better bound?

Example: Coupon Collector's Problem

- Consider a geometric random variable Y.
  - Y: the number of flips until the first heads (head: appear with prob. p).
- $\mathbf{E}[Y] = 1/p$. How about $\mathbf{E}[Y^2]$?
  - We try to calculate it using conditional expectation.
  - Could be easier (?)

Example: Coupon Collector's Problem

- Consider a geometric random variable Y.
  - Y: the number of flips until the first heads (head: appear with prob. p).
- $\mathbf{E}[Y] = 1/p$. How about $\mathbf{E}[Y^2]$?
  - We try to calculate it using conditional expectation.
  - Could be easier (?)

$$
\begin{aligned}
\mathbf{E}[Y^2] &= \Pr[X=0]\cdot\mathbf{E}[Y^2 \mid X=0] + \Pr[X=1]\cdot\mathbf{E}[Y^2 \mid X=1] \\
&= (1-p)\cdot\mathbf{E}[Y^2 \mid X=0] + p\cdot\mathbf{E}[Y^2 \mid X=1].
\end{aligned}
$$

Example: Coupon Collector's Problem

- Consider a geometric random variable Y.
  - Y: the number of flips until the first heads (head: appear with prob. p).
- $\mathbf{E}[Y] = 1/p$. How about $\mathbf{E}[Y^2]$?
  - We try to calculate it using conditional expectation.
  - Could be easier (?)

$$
\begin{aligned}
\mathbf{E}[Y^2] &= \Pr[X=0]\cdot\mathbf{E}[Y^2 \mid X=0] + \Pr[X=1]\cdot\mathbf{E}[Y^2 \mid X=1] \\
&= (1-p)\cdot\boxed{\mathbf{E}[Y^2 \mid X=0]} + p\cdot\boxed{\mathbf{E}[Y^2 \mid X=1]}.
\end{aligned}
$$

  - If $X = 1$, then $Y = 1$; if $X = 0$, then $Y > 1$. So,

$$\boxed{\mathbf{E}[Y^2 \mid X=1]} = 1 \qquad \boxed{\mathbf{E}[Y^2 \mid X=0]} = \mathbf{E}[(Z+1)^2]$$

***Z*: another geometric random variable with parameter *p*.**

Example: Coupon Collector's Problem

- Hence,

$$\begin{aligned}
\boxed{\mathbf{E}[Y^2]} &= \Pr[X=0]\cdot\mathbf{E}[Y^2 \mid X=0] + \Pr[X=1]\cdot\mathbf{E}[Y^2 \mid X=1] \\
&= (1-p)\cdot\mathbf{E}[Y^2 \mid X=0] + p\cdot\mathbf{E}[Y^2 \mid X=1] \\
&= (1-p)\cdot\mathbf{E}[(Z+1)^2] + p\cdot 1 \\
&= (1-p)\cdot\mathbf{E}[Z^2] + 2(1-p)\mathbf{E}[Z] + 1 \\
&= (1-p)\cdot\mathbf{E}[Y^2] + 2(1-p)\cdot\frac{1}{p} + 1 \\
&= \boxed{(1-p)\mathbf{E}[Y^2]} + \frac{2-p}{p}.
\end{aligned}$$

$$\Rightarrow \mathbf{E}[Y^2] = \frac{2-p}{p^2}.$$

$$\mathbf{Var}[Y] = \mathbf{E}[Y^2] - (\mathbf{E}[Y])^2 = \frac{1-p}{p^2} < \frac{1}{p^2}.$$

Example: Coupon Collector's Problem

- Let's go back to the coupons…

$$\mathbf{Var}[X] = \sum_{i=1}^{n} \mathbf{Var}[X_i] \le \sum_{i=1}^{n} \left(\frac{1}{p^2}\right) = \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right)^2$$
$$= n^2 \sum_{i=1}^{n} \left(\frac{1}{i}\right)^2$$
$$\le \frac{\pi^2 n^2}{6}.$$

- Let's welcome Chebyshev!

$$\Pr[|X - nH_n| \ge nH_n] \le \frac{\pi^2 n^2}{(nH_n)^2} = \frac{\pi^2}{6(H_n)^2} = O\left(\frac{1}{\ln^2 n}\right).$$

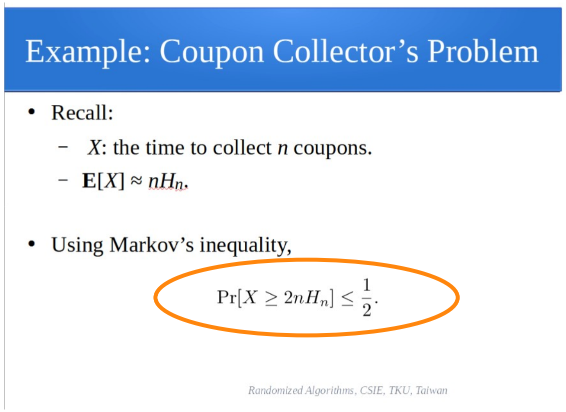
Example: Coupon Collector's Problem

- Let's go back to the coupons…

$$\mathbf{Var}[X] = \sum_{i=1}^{n} \mathbf{Var}[X_i] \le \sum_{i=1}^{n} \left(\frac{1}{p^2}\right) = \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right)^2$$

$$= n^2 \sum_{i=1}^{n} \left(\frac{1}{i}\right)^2$$

$$\le \frac{\pi^2 n^2}{6}.$$



Example: Coupon Collector's Problem

- Recall:
  - X: the time to collect n coupons.
  - $\mathbf{E}[X] \approx nH_n$.
- Using Markov's inequality,

$$\Pr[X \ge 2nH_n] \le \frac{1}{2}.$$

Randomized Algorithms, CSIE, TKU, Taiwan

- Let's welcome Chebyshev!

$$\Pr[|X - nH_n| \ge nH_n] \le \frac{\pi^2 n^2}{(nH_n)^2} = \frac{\pi^2}{6(H_n)^2} = O\left(\frac{1}{\ln^2 n}\right).$$

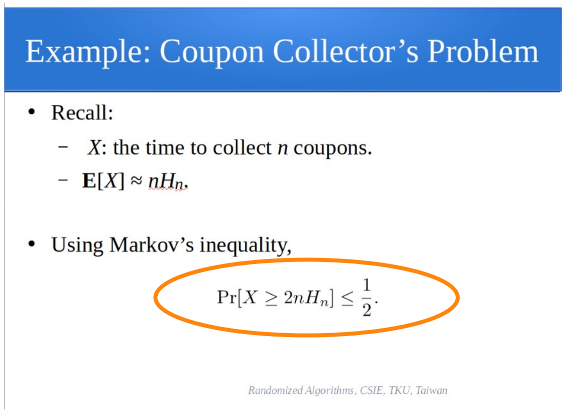
Example: Coupon Collector's Problem

- Let's go back to the coupons…

$$\textbf{Var}[X] = \sum_{i=1}^{n} \textbf{Var}[X_i] \le \sum_{i=1}^{n} \left(\frac{1}{p^2}\right) = \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right)^2$$

$$= n^2 \sum_{i=1}^{n} \left(\frac{1}{i}\right)^2$$

$$\le \frac{\pi^2 n^2}{6}.$$



- Let's welcome Chebyshev!

$$\Pr[|X - nH_n| \ge nH_n] \le \frac{\pi^2 n^2}{(nH_n)^2} = \frac{\pi^2}{6(H_n)^2} = O\left(\frac{1}{\ln^2 n}\right).$$

< 0.22 when n=100
< 0.15 when n=1,000

Basel problem

Euler's approach [edit]

Euler's original derivation of the value $\frac{\pi^2}{6}$ essentially extended observations about finite polynomials and assumed that these same properties hold true for infinite series.

Of course, Euler's original reasoning requires justification (100 years later, Karl Weierstrass proved that Euler's representation of the sine function as an infinite product is valid, by the Weierstrass factorization theorem), but even without justification, by simply obtaining the correct value, he was able to verify it numerically against partial sums of the series. The agreement he observed gave him sufficient confidence to announce his result to the mathematical community.

To follow Euler's argument, recall the Taylor series expansion of the sine function

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

Dividing through by x, we have

$$\frac{\sin x}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots$$

Using the Weierstrass factorization theorem, it can also be shown that the left-hand side is the product of linear factors given by its roots, just as we do for finite polynomials (which Euler assumed as a heuristic for expanding an infinite degree polynomial in terms of its roots, but in fact is not always true for general $P(x)$):[4]

$$\begin{aligned}\frac{\sin x}{x} &= \left(1-\frac{x}{\pi}\right)\left(1+\frac{x}{\pi}\right)\left(1-\frac{x}{2\pi}\right)\left(1+\frac{x}{2\pi}\right)\left(1-\frac{x}{3\pi}\right)\left(1+\frac{x}{3\pi}\right)\cdots \\ &= \left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\cdots\end{aligned}$$

If we formally multiply out this product and collect all the x^2 terms (we are allowed to do so because of Newton's identities), we see by induction that the x^2 coefficient of $\frac{\sin x}{x}$ is [5]

$$-\left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \cdots\right) = -\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

But from the original infinite series expansion of $\frac{\sin x}{x}$, the coefficient of x^2 is $-\frac{1}{3!} = -\frac{1}{6}$. These two coefficients must be equal; thus,

$$-\frac{1}{6} = -\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

Multiplying both sides of this equation by $-\pi^2$ gives the sum of the reciprocals of the positive square integers.

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}.$$

Refer to Wikipedia for more details and more approaches.
https://en.wikipedia.org/wiki/Basel_problem

Assignment 03

1. Let X be a number chosen uniformly at random from $[1, n]$. Find $\mathrm{Var}[X]$.
2. Suppose that we roll a standard fair die 100 times. Let X be the sum of the numbers that appear over the 100 rolls. Use Chebyshev's inequality to bound

$$\Pr[|X - 350| \geq 50].$$