Advanced Algorithms (II)

Shanghai Jiao Tong University

Chihao Zhang

March 9th, 2020

Random Variables

Random Variables

Recall that a probability space is a tuple $(\Omega, \mathscr{F}, \Pr)$

Random Variables

Recall that a probability space is a tuple $(\Omega, \mathcal{F}, \Pr)$

In this course, we mainly focus on countable Ω

Random Variables

Recall that a probability space is a tuple $(\Omega, \mathcal{F}, \Pr)$

In this course, we mainly focus on countable Ω

A random variable X is a function $X : \Omega \to \mathbb{R}$

Random Variables

Recall that a probability space is a tuple $(\Omega, \mathcal{F}, \Pr)$

In this course, we mainly focus on countable Ω

A random variable X is a function $X : \Omega \to \mathbb{R}$

The expectation $\mathbf{E}[X] = \sum_{a \in \Omega : \Pr[X=a] > 0} a \cdot \Pr[X = a]$

Linearity of Expectations

Linearity of Expectations

For any n random variables $X_1, \ldots, X_n$

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

Linearity of Expectations

For any n random variables $X_1, \ldots, X_n$

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

$$\begin{aligned}
\mathbf{E}[X_1 + X_2] &= \sum_{a,b} (a+b) \cdot \Pr[X_1 = a, X_2 = b] \\
&= \sum_{a,b} a \cdot \Pr[X_1 = a, X_2 = b] + \sum_{a,b} b \cdot \Pr[X_1 = a, X_2 = b] \\
&= \sum_{a} a \cdot \Pr[X_1 = a] + \sum_{b} b \cdot \Pr[X_2 = b] = \mathbf{E}[X_1] + \mathbf{E}[X_2]
\end{aligned}$$

Coupon Collector

Coupon Collector

There are n coupons to collect…

Coupon Collector

There are n coupons to collect…

Each time one coupon is drawn independently uniformly at random

Coupon Collector

There are n coupons to collect…

Each time one coupon is drawn independently uniformly at random

How many times one needs to draw to collect all coupons?

Let X_i be the number of draws between i-th distinct coupon to the $i + 1$-th distinct coupon

Let X_i be the number of draws between i-th distinct coupon to the $i+1$-th distinct coupon

$$X := \text{Number of draws} = \sum_{i=0}^{n-1} X_i$$

Let X_i be the number of draws between i-th distinct coupon to the $i + 1$-th distinct coupon

$$X := \text{Number of draws} = \sum_{i=0}^{n-1} X_i$$

For any i, X_i follows *geometric distribution* with probability $\dfrac{n-i}{n}$

Geometric Distribution

Geometric Distribution

Let X be a random variable following geometric distribution with probability p.

Geometric Distribution

Let X be a random variable following geometric distribution with probability p.

Namely, we toss a coin who comes to HEAD with probability p, X is the number of tosses to see the first HEAD.

Geometric Distribution

Let X be a random variable following geometric distribution with probability p.

Namely, we toss a coin who comes to HEAD with probability p, X is the number of tosses to see the first HEAD.

It is not hard to see that $\mathbf{E}[X] = \frac{1}{p}$

Back to Coupon Collector…

Back to Coupon Collector…

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=0}^{n-1} X_i\right] = \sum_{i=0}^{n-1} \mathbf{E}[X_i]$$

$$= \sum_{i=0}^{n-1} \frac{n}{n-i} = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{1}$$

$$= n \cdot H(n) \to n \log n + \gamma n$$

Back to Coupon Collector…

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=0}^{n-1} X_i\right] = \sum_{i=0}^{n-1} \mathbf{E}[X_i]$$

$$= \sum_{i=0}^{n-1} \frac{n}{n-i} = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{1}$$

$$= n \cdot H(n) \to n \log n + \gamma n$$

The constant $\gamma = 0.577...$ is called Euler constant

Linearity may fail when…

Linearity may fail when…

- $n = \infty$

Linearity may fail when…

- $n = \infty$

St. Petersburg paradox

Each stage of the game a fair coin is tossed and a gambler guesses the result. He wins the amount he bet if his guess is correct and lose the money if he is wrong. He bets $\$1$ *at the first stage. If he loses, he doubles the money and bets again. The game ends when the gambler wins.*

What is the expected money he wins?

What is the expected money he wins?

- In stage i, he wins X_i with $\mathbf{E}[X_i] = 0$,
- so $\sum_{i=1}^{\infty} \mathbf{E}[X_i] = 0$

What is the expected money he wins?

- In stage i, he wins X_i with $\mathbf{E}[X_i] = 0$,
- so $\sum_{i=1}^{\infty} \mathbf{E}[X_i] = 0$
- On the other hand, he eventually wins \$1,
- so $\mathbf{E}\left[\sum_{i=1}^{\infty} X_i\right] = 1 \neq \sum_{i=1}^{\infty} \mathbf{E}[X_i]!$

Linearity may fail when…

Linearity may fail when…

- $n = N$ is random

Linearity may fail when…

- $n = N$ is random

Suppose we draw a number N and toss N dices $X_1, \ldots, X_N$, what is $\mathbf{E}\left[\sum_{i=1}^{N} X_N\right]$?

Each X_i is uniform in $\{1,\ldots,6\}$, one might expect

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbf{E}[N] \cdot \mathbf{E}[X_1] = 3.5 \times 3.5 = 12.25$$

Each X_i is uniform in $\{1,\ldots,6\}$, one might expect

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbf{E}[N] \cdot \mathbf{E}[X_1] = 3.5 \times 3.5 = 12.25$$

If N itself is drawn by tossing a dice and let

$X_1 = X_2 = \ldots = X_N = N$

Each X_i is uniform in $\{1,\ldots,6\}$, one might expect

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbf{E}[N] \cdot \mathbf{E}[X_1] = 3.5 \times 3.5 = 12.25$$

If N itself is drawn by tossing a dice and let

$X_1 = X_2 = \ldots = X_N = N$

Then $\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbf{E}[N \cdot N] = 15.166..$

Wald’s Equation

Wald's Equation

If the variables satisfy

Wald's Equation

If the variables satisfy

- N and all X_i are independent and finite;
- All X_i are identically distributed

Wald's Equation

If the variables satisfy

- N and all X_i are independent and finite;
- All X_i are identically distributed

$$\sum_{i=1}^{N} \mathbf{E}\left[X_i\right] = \mathbf{E}[N] \cdot \mathbf{E}[X_1]$$

Wald's Equation

If the variables satisfy

- N and all X_i are independent and finite;
- All X_i are identically distributed

$$\sum_{i=1}^{N} \mathbf{E}\left[X_i\right] = \mathbf{E}[N] \cdot \mathbf{E}[X_1]$$

More generally if N is a *stopping time*

Application: Quick Select

Application: Quick Select

Find the k-th largest number in an unsorted array A

Application: Quick Select

Find the k-th largest number in an unsorted array A

Find(A, k)

Randomly choose a pivot $x \in A$

1. Partition $A - \{x\}$ into A_1, A_2 such that $\forall y \in A_1, y < x$, $\forall z \in A_2, z > x$
2. If $|A_1| = k - 1$, return x
3. If $|A_1| \geq k$, return **Find**(A_1, k)
4. return **Find**$(A_2, k - |A_1| - 1)$

The partition step takes $O(|A|)$ time

The partition step takes $O(|A|)$ time

What is the total time cost *in expectation?*

The partition step takes $O(|A|)$ time

What is the total time cost *in expectation?*

X_i - size of A at i-th round

$X_1 = n$ and $\mathbf{E}[X_{i+1} \mid X_i] \leq \frac{3}{4} X_i$

The time cost is $\sum_{i=1}^{\infty} X_i$

$$\mathbf{E}[X_{i+1} \mid X_i] \leq \frac{3}{4} X_i$$

$$\implies \mathbf{E}[X_{i+1}] = \mathbf{E}[\mathbf{E}[X_{i+1} \mid X_i]] \leq \frac{3}{4}\mathbf{E}[X_i] \leq \left(\frac{3}{4}\right)^i n$$

$$\mathbf{E}[X_{i+1} \mid X_i] \leq \frac{3}{4} X_i$$

$$\implies \mathbf{E}[X_{i+1}] = \mathbf{E}[\mathbf{E}[X_{i+1} \mid X_i]] \leq \frac{3}{4}\mathbf{E}[X_i] \leq \left(\frac{3}{4}\right)^i n$$

$$\begin{aligned}
\mathbf{E}\left[\sum_{i=1}^{\infty} X_i\right] &= \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i] \leq \sum_{i=1}^{n} \left(\frac{3}{4}\right)^{i-1} n \\
&= 4n\,.
\end{aligned}$$

KUW inequality

KUW inequality

While analyzing random algorithms, a common recursion is $T(n) = 1 + T(n - X_n)$ for random X_n

KUW inequality

While analyzing random algorithms, a common recursion is $T(n) = 1 + T(n - X_n)$ for random X_n

Theorem. (Karp-Upfal-Wigderson Inequality)

Assume for every n, $0 \leq X_n \leq n - a$ is an integer for some a such that $T(a) = 0$. If $\mathbf{E}[X_n] \geq \mu(n)$ for all $n > a$, where $\mu(n)$ is positive and increasing , then

$$\mathbf{E}[T(n)] \leq \int_a^n \frac{1}{\mu(t)} \, \mathrm{d}t$$

Application: Expectation of Geometric Variables

Application: Expectation of Geometric Variables

$T(1) = 1 + T(1 - X_1)$, where $\mathbf{E}[X_1] = p$

Application: Expectation of Geometric Variables

$T(1) = 1 + T(1 - X_1)$, where $\mathbf{E}[X_1] = p$

Choosing $\mu(n) = p$ gives $\mathbf{E}[T(1)] \leq \int_0^1 \frac{1}{p}\,\mathrm{d}t = \frac{1}{p}$.

Application: Rounds of Quick Select

Application: Rounds of Quick Select

In our **Find**(A, k) algorithm, we have

$$T(n) = 1 + \max\{T(m), T(n - m - 1)\},$$

Application: Rounds of Quick Select

In our **Find**(A, k) algorithm, we have

$$T(n) = 1 + \max\{T(m), T(n - m - 1)\},$$

where m is in $\{1,2,\ldots, n-1\}$ uniformly at random.

Application: Rounds of Quick Select

In our **Find**(A, k) algorithm, we have

$$T(n) = 1 + \max\{T(m), T(n-m-1)\},$$

where m is in $\{1,2,\ldots,n-1\}$ uniformly at random.

We can choose $\mu(n) = \dfrac{n}{4}$ (Why?)

Application: Rounds of Quick Select

In our **Find**(A, k) algorithm, we have

$$T(n) = 1 + \max\{T(m), T(n - m - 1)\},$$

where m is in $\{1,2,\ldots,n-1\}$ uniformly at random.

We can choose $\mu(n) = \dfrac{n}{4}$ (Why?)

KUW implies $\mathbf{E}[T(n)] \leq \displaystyle\int_1^n \frac{4}{t}\,\mathrm{d}t = 4\log n$

Application: Coupon Collector

Application: Coupon Collector

$T(m) = 1 + T(n - X_m)$ where $X_m \sim \text{Ber}(m/n)$

Application: Coupon Collector

$T(m) = 1 + T(n - X_m)$ where $X_m \sim \text{Ber}(m/n)$

So we can choose $\mu(m) = \dfrac{\lceil m \rceil}{n}$

Application: Coupon Collector

$T(m) = 1 + T(n - X_m)$ where $X_m \sim \mathrm{Ber}(m/n)$

So we can choose $\mu(m) = \dfrac{\lceil m \rceil}{n}$

KUW implies $\mathbf{E}[T(n)] \leq \displaystyle\int_0^n \frac{n}{\lceil t \rceil}\,\mathrm{d}t = n \cdot H_n$

Proof of KUW inequality