{ csci 3|6907 | Lecture 2 }

Hoeteck Wee · hoeteck@gwu.edu

Announcements

- Homework 1 is out, due next Wed

  feel free to discuss in groups

  homework must be written up **individually**

- online form on course webpage (complete by Fri)

PART I | two distributions & coupon-collecting

Random variables

- random experiment is a process that produces some uncertain outcome e.g. throwing a die or tossing coin
- use D to denote outcome of a die roll

  D could take values $1, 2, 3, 4, 5, 6$, all equally likely
- random variable is the outcome of a random experiment

  e.g. $\Pr[D = 5] = 1/6$ (we call "$D = 5$" an event)

Random variables

- throw a die twice and let D_0 denote the sum

  e.g. first roll 5, second roll is 2, then $D_0 = 7$.
- let D_1 and D_2 denote first and second die roll respectively
- FACT 1. random variables D_1 and D_2 are independent
- FACT 2. $D_0 = D_1 + D_2$

  (useful trick – decomposing a r.v. as sum of r.v.'s)

Expectation

Definition (expectation)

The expectation $\mathrm{E}[X]$ of a discrete random variable X is given by

$$\sum_i i \Pr[X = i].$$

- EXAMPLE 1: $\mathrm{E}[D] = \frac{1}{6}(1+2+3+4+5+6) = 3.5$
- EXAMPLE 2: toss a fair coin twice. Let Υ denote the number of heads.

  $\Pr[\Upsilon = 0] = 1/4.$ $\Pr[\Upsilon = 1] = 1/2.$ $\Pr[\Upsilon = 2] = 1/4.$

  $\mathrm{E}[\Upsilon] = 1.$
- EXAMPLE 3: toss a fair coin 100 times. Let Z denote # of heads.

Expectation

Fact (linearity of expectations)

Given any finite collection of r.v. $X_1, \ldots, X_n$, we have

$$\mathrm{E}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mathrm{E}[X_i]$$

- EXAMPLE 3: toss a fair coin 100 times. Let Z denote # of heads.
- define r.v. X_i

$$X_i = \begin{cases} 1 & \text{if } i\text{'th toss comes up heads} \\ 0 & \text{otherwise} \end{cases}$$

- what is $X_1 + \cdots + X_{100}$? what is $\mathrm{E}[X_1]$?

Variance

Definition (variance)

The variance $\mathrm{Var}[X]$ of a random variable X is given by

$$\mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2.$$

Fact

Given any finite collection of *independent* r.v. $X_1, \ldots, X_n$, we have

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$$

Bernoulli and Binomial random variables

- Consider an experiment with success probability p; call Υ the r.v.:

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

  Then, Υ is called a *Bernoulli* or *indicator* r.v.

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

Bernoulli and Binomial random variables

- Consider an experiment with success probability p; call Υ the r.v.:

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

  Then, Υ is called a *Bernoulli* or *indicator* r.v.

- *Binomial* r.v. $X - B(n, p)$ – denotes # successes in n independent trials.

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

- Can write X as sum of indicator r.v. $\Upsilon_1 + \Upsilon_2 + \cdots + \Upsilon_n$
- $\mathrm{E}[X] = np$ and $\mathrm{Var}[X] = np(1-p)$.

Geometric random variable

- Perform a sequence of independent trials until the first success.

  EXAMPLE 1. toss a fair coin until you get a head.

  EXAMPLE 2. roll a die until you get a six.

  How many tosses/rolls do you need?

- *Geometric* r.v. X with parameter p denotes # trials until first success.

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

- $\mathrm{E}[X] = \sum_{n=1}^{\infty} n(1 - p)^{n-1} p = 1/p^2 \cdot p = 1/p$

Coupon collector's problem

Coupon Collector's Problem

- Each box of cereals contains one of *n* different coupons.
- Coupon in each box is independently & uniformly random.

How many boxes must we buy to obtain $\geq$ one coupon of every type?

- e.g. collecting a coupon = passing a required course?
- need at least *n* boxes. expected number? (hint: decompose as sum)
- attempt #1: Let Υ_i denote # boxes until you get coupon of type i
- example: $1, 1, 2, 1, 3$

  $(\Upsilon_1 = 1, \Upsilon_2 = 3, \Upsilon_3 = 5)$

Coupon collector's problem

Coupon Collector's Problem

- Each box of cereals contains one of n different coupons.
- Coupon in each box is independently & uniformly random.

How many boxes must we buy to obtain $\geq$ one coupon of every type?

- Let X_i be # boxes to go from exactly $i-1$ different coupons to i.
- X_i is a geometric r.v. with parameter $p_i = \frac{n-(i-1)}{n}$.
- Total # of boxes $X = X_1 + X_2 + \cdots + X_n$.
- $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot \sum_{i=1}^{n} \frac{1}{i} = n(\ln n + \Theta(1))$

  (FACT. $1 + 1/2 + 1/3 + \cdots + 1/n = \ln n + \Theta(1)$)

PART 2 | quicksort

Sorting

SORTING Problem

Input: a list of n distinct values $x_1, \ldots, x_n$ from some domain

Goal: output a sorted list

- Total order on the domain.
- Complexity measure: # of comparisons
- $\log(n!) = \Omega(n \log n)$ comparisons are *necessary*

Quicksort

QUICKSORT Algorithm

Input: A list $S = \{x_1, \ldots, x_n\}$

1. if $|S| = 0$ or $|S| = 1$, return S and halt.
2. Choose a random element x of S as pivot.
3. Let $S_1 = \{y \in S \mid y < x\}$ and $S_2 = \{y \in S \mid y > x\}$
4. Return the list $\textsc{QuickSort}(S_1), x, \textsc{Quicksort}(S_2)$.

- Example: $\textsc{QuickSort}(4, 3, 2, 1)$
- Compute expectation of X, # of comparisons.
- Call $y_1, \ldots, y_n$ the sorted list.
- Fact: any pair $y_i \neq y_j$ is compared at most once (when/why?).

Analysis of QUICKSORT

1. Write $X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij}$ where X_{ij} indicates y_i and y_j are compared.
2. Compute $\mathrm{E}[X_{ij}]$:
   - Consider first element x in $S_{ij} = \{y_i, y_{i+1}, \ldots, y_j\}$ to be used as pivot.
   - If x equals y_i or y_j, then $X_{ij} = 1$.
   - If $x \neq y_i, y_j$, then $X_{ij} = 0$.
   - $\Pr[X_{ij} = 1] = \frac{2}{|S_{ij}|} = \frac{2}{j-i+1}$.
3. $$\begin{aligned} \mathrm{E}[X] &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{|S_{ij}|} = \sum_{i=1}^{n} \Big(\frac{2}{2} + \frac{2}{3} + \cdots + \frac{2}{n-i+1} \Big) \\ &\leq \sum_{i=1}^{n} \Big(\frac{2}{2} + \frac{2}{3} + \cdots + \frac{2}{n} \Big) \\ &= 2n \ln n + \Theta(n) \end{aligned}$$