

## Randomness and Computation

or, “Randomized Algorithms”

Heng Guo

(Based on slides by M. Cryan)

RC (2019/20) – Lecture 4 – slide 1

## Discrete random variables

Our main focus in RC is on *random variables*, especially *discrete random variables* (when  $X$  can take on a finite or countable number of values).

Not all random variables have **bounded** expectation. Expectation is *finite* if  $\sum_i |i| \Pr[X = i]$  *converges* as a series; otherwise **unbounded**. (note that  $EX$  cannot be unbounded unless it has infinite support).

RC (2019/20) – Lecture 4 – slide 2

## Discrete random variables

Our main focus in RC is on *random variables*, especially *discrete random variables* (when  $X$  can take on a finite or countable number of values).

Not all random variables have **bounded** expectation. Expectation is *finite* if  $\sum_i |i| \Pr[X = i]$  *converges* as a series; otherwise **unbounded**. (note that  $EX$  cannot be unbounded unless it has infinite support).

### Theorem (2.1, Linearity of Expectation)

For any finite collection of discrete random variables  $X_1, \dots, X_k$  with finite expectations,

$$E \left[ \sum_{j=1}^k X_j \right] = \sum_{j=1}^k E[X_j].$$

Theorem 2.1 holds regardless of whether the random variables are independent or not.

RC (2019/20) – Lecture 4 – slide 2

## Linearity of Expectation

Consider a uniformly at random permutation of  $n$  elements. What is the expected number of fixed points ( $x$  such that  $\sigma(x) = x$ )?

RC (2019/20) – Lecture 4 – slide 3

## Linearity of Expectation

Consider a uniformly at random permutation of  $n$  elements. What is the expected number of fixed points ( $x$  such that  $\sigma(x) = x$ )?

Let  $X_i$  be the indicator variable of the event  $\sigma(i) = i$ , and  $X = \sum_{i=1}^n X_i$ . Then

$$EX_i = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Thus,

$$EX = \sum_{i=1}^n EX_i = 1.$$

RC (2019/20) – Lecture 4 – slide 3

## Discrete random variables ...

### Lemma (2.2)

For any discrete random variable  $X$ , any constant  $c$ ,  $E[c \cdot X] = c \cdot E[X]$ .

### Definition (2.2)

A collection  $X_1, \dots, X_k$  of random variables are said to be **mutually independent** if and only if, for every subset  $I \subseteq \{1, \dots, k\}$ , and every tuple of values  $a_i, i \in I$ , we have

$$\Pr[\cap_{i \in I} (X_i = a_i)] = \prod_{i \in I} \Pr[X_i = a_i].$$

RC (2019/20) – Lecture 4 – slide 4

## Independence

“Mutually independent” is **stronger** than “pairwise independent” - a collection of random variables can be pairwise independent but *not* mutually independent.

### Example

Two fair coins, values 1 and 0.

$A$  value of the first flip;

$B$  value of the second flip;

$C$  absolute difference of two values.

Pairwise independence works out but

$$\Pr[(A = 1) \cap (B = 1) \cap (C = 1)] = ?$$

RC (2019/20) – Lecture 4 – slide 5

## Variance and second moment

### Definition

The  $k$ -th moment is defined as

$$E[X^k] := \sum_i i^k \Pr[X = i].$$

The **variance** is defined as

$$\text{Var}[X] := E[(X - E[X])^2] = \sum_i (i - E[X])^2 \Pr[X = i].$$

RC (2019/20) – Lecture 4 – slide 6

## Variance

### Lemma

For any discrete random variable  $X$ ,  $E[X^2] \geq E[X]^2$ .

RC (2019/20) – Lecture 4 – slide 7

## Variance

### Lemma

For any discrete random variable  $X$ ,  $E[X^2] \geq E[X]^2$ .

### Proof.

Consider the variance  $\text{Var}[X] = E[(X - E[X])^2]$ . Since it only takes non-negative values,  $\text{Var}[X] \geq 0$ . Moreover,

RC (2019/20) – Lecture 4 – slide 7

## Variance

### Lemma

For any discrete random variable  $X$ ,  $E[X^2] \geq E[X]^2$ .

### Proof.

Consider the variance  $\text{Var}[X] = E[(X - E[X])^2]$ . Since it only takes non-negative values,  $\text{Var}[X] \geq 0$ . Moreover,

$$\begin{aligned}\text{Var}[X] &= E[X^2 - 2E[X] \cdot X + E[X]^2] \\ &= E[X^2] - E[2E[X] \cdot X] + E[X]^2 && \text{(Thm 2.1)} \\ &= E[X^2] - 2E[X] \cdot E[X] + E[X]^2 && \text{(Lemma 2.2)} \\ &= E[X^2] - E[X]^2.\end{aligned}$$

By  $\text{Var}[X] \geq 0$ , we have  $E[X^2] \geq E[X]^2$ . □

RC (2019/20) – Lecture 4 – slide 7

## Jensen's Inequality

### Definition

A function  $f: \mathbb{R} \rightarrow \mathbb{R}$  is said to be *convex* if it is the case that for every  $x_1, x_2 \in \mathbb{R}$  and every  $\lambda \in [0, 1]$ ,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

### Lemma (2.3)

For any  $f$  which is twice differentiable,  $f$  is convex around  $x$  if and only if  $f''(x) \geq 0$ .

### Theorem (2.4, Jensen's Inequality)

If  $f$  is a convex function, then

$$E[f(X)] \geq f(E[X]).$$

RC (2019/20) – Lecture 4 – slide 8

## Jensen's Inequality

### Theorem (2.4, Jensen's Inequality)

If  $f$  is a convex function, then

$$E[f(X)] \geq f(E[X]).$$

### Proof.

Let  $\mu = E[X]$ . Assuming that  $f$  is twice differentiable on its domain, then Taylor's expansion implies  $\forall x$ , there is some  $c$  between  $\mu$  and  $x$  such that

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + f''(c) \frac{(x - \mu)^2}{2}.$$

By convexity of  $f$ , we know  $f''(\cdot) \geq 0$  throughout domain, so  $f(x) \geq f(\mu) + f'(\mu)(x - \mu)$  for all  $x$ . Take expectation and apply Thm 2.1, Lem 2.2,

$$\begin{aligned} E[f(X)] &\geq E[f(\mu)] + E[f'(\mu)(X - \mu)] \\ &= f(\mu) + f'(\mu) \cdot (E[X] - E[X]), \end{aligned}$$

so the  $f'$  term disappears and  $E[f(X)] \geq f(\mu) = f(E[X])$ .  $\square$

RC (2019/20) – Lecture 4 – slide 9

## Simple distributions

### Definition

The *Bernoulli distribution* (biased coin-flip) is the random variable  $Y$  such that  $Y = 1$  with probability  $p$  and  $Y = 0$  with probability  $1 - p$ .

Notice  $E[Y] = p$  when  $Y$  is Bernoulli.

### Definition (2.5)

The *binomial distribution* for  $n, p$ , written  $B(n, p)$ , is the random variable  $X$  which takes values in  $\{0, 1, \dots, n\}$  with the probabilities

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

We can prove  $E[X] = np$  for  $X$  being  $B(n, p)$  in *two ways*:

- ▶ Directly, using Definition 2.5 and simplifying/summing the series.
- ▶ Binomial distribution  $B(n, p)$  is the probabilities of getting  $j$  flips from  $n$  independent trials of a Bernoulli. Then use linearity of expectation.

RC (2019/20) – Lecture 4 – slide 10

## Conditional Expectation

### Definition (2.6)

For two random variables  $X, Y$ ,

$$E[X | Y = y] = \sum_x x \cdot \Pr[X = x | Y = y],$$

summation being taken over all  $x$  in the support of  $X$ .

### Definition (2.7)

We also use the expression  $E[X | Y]$ , where  $X, Y$  are random variables.

$E[X | Y]$  itself is a random variable and a function of  $Y$ . If  $Y = y$ , it takes value  $E[X | Y = y]$ . In other words, it takes value  $x$  with probability

$$\sum_{y: E[X|Y=y]=x} \Pr[Y = y].$$

RC (2019/20) – Lecture 4 – slide 11

## Conditional Expectation

### Observation

For any finite collection of discrete random variables  $X_1, \dots, X_n$  with finite expectations, and for any random variable  $X$ ,

$$E \left[ \left( \sum_{i=1}^n X_i \right) | Y = y \right] = \sum_{i=1}^n E[X_i | Y = y].$$

### Lemma (2.5)

For any random variables  $X$  and  $Y$ , such that  $E[X | Y = y]$  is always bounded

$$E[E[X | Y]] = E[X].$$

RC (2019/20) – Lecture 4 – slide 12

## Conditional Independence

Proof.

$$\begin{aligned} E[E[X | Y]] &= \sum_y \Pr[Y = y] E[X | Y = y] \\ &= \sum_y \Pr[Y = y] \sum_x x \Pr[X = x | Y = y] \\ &= \sum_y \sum_x x \Pr[Y = y] \Pr[X = x | Y = y] \\ &= \sum_y \sum_x x \Pr[X = x \cap Y = y] \\ &= \sum_x \sum_y x \Pr[X = x \cap Y = y] \\ &= \sum_x x \Pr[X = x] = E[X]. \quad \square \end{aligned}$$

RC (2019/20) – Lecture 4 – slide 13

## Geometric distributions

Imagine we flip a biased coin many times (success with prob.  $p$ ), and stop when we see the first success (heads, or alternatively 1). What is the distribution of the number of flips?

### Definition (2.8)

A *geometric* random variable  $X$  with parameter  $p$  is given by the following probability distribution on  $\mathbb{N}$ :

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

Should verify that  $\sum_{j=1}^{\infty} \Pr[X = j] = 1$ .

### Lemma (2.8)

For a geometric random variable  $X$  with parameter  $p$ , and for any  $j > 0, k \geq 0$ ,

$$\Pr[X = j + k | X > k] = \Pr[X = j].$$

RC (2019/20) – Lecture 4 – slide 14

## Geometric distributions

### Lemma (2.9)

For any discrete random variable  $X$  that only takes *non-negative integer* values, we have the following:

$$E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i].$$

Proof.

Consider the indicator variables  $Y_i = \mathbf{1}_{X \geq i}$ . Then  $X = \sum_{i=1}^{\infty} Y_i$ . Take expectation and use linearity.  $\square$

### Observation

If  $X$  is a geometric random variable  $X$  with parameter  $p$ , then for any  $i \geq 0$ ,  $\Pr[X \geq i] = (1 - p)^{i-1}$ .

Proof.

The event that  $X \geq i$  is exactly the event that the first  $(i - 1)$  trials all fail. (You can also directly calculate it.)  $\square$

RC (2019/20) – Lecture 4 – slide 15

## Geometric distributions

### Lemma

If  $X$  is a geometric random variable  $X$  with parameter  $p$ , then  $E[X] = p^{-1}$ .

Proof.

By Lemma 2.9, we have  $E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$ . For a geometric random variable, parameter  $p$ ,

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} (1 - p)^{i-1} = \sum_{i=0}^{\infty} (1 - p)^i \\ &= \frac{1}{1 - (1 - p)} = \frac{1}{p}. \quad \square \end{aligned}$$

RC (2019/20) – Lecture 4 – slide 16

## Alternatively

Let  $Y$  be the indicator variable of whether the first flip succeeds.

$$\begin{aligned} E[X] &= E[E[X | Y]] && \text{(Lemma 2.5)} \\ &= \Pr[Y = 1]E[X | Y = 1] + \Pr[Y = 0]E[X | Y = 0] \\ &= p + (1 - p) \sum_{i=1}^{\infty} i \Pr[X = i | X > 1] \\ &= p + (1 - p) \sum_{i=1}^{\infty} i \Pr[X = i - 1] && \text{(Lemma 2.8)} \\ &= p + (1 - p) \sum_{i=1}^{\infty} (i - 1) \Pr[X = i - 1] + (1 - p) \sum_{i=1}^{\infty} \Pr[X = i - 1] \\ &= p + (1 - p)E[X] + 1 - p = 1 + (1 - p)E[X]. \end{aligned}$$

We can solve that  $E[X] = \frac{1}{p}$ .

RC (2019/20) – Lecture 4 – slide 17

## Coupon Collector Problem

“Coupon collecting” is the activity of buying cereal-packets, each of which will have a coupon inside. There are  $n$  different types of “coupon” (eg cards with a photo of a footballer) and the goal is to collect one copy of each . . . then stop buying.

How many packets do we (expect to) need to buy?

Assumptions:

- ▶ Items are randomly and identically distributed in packets (one card per packet). So when buying a box the probability of any particular card being inside is  $1/n$ .

RC (2019/20) – Lecture 4 – slide 18

## Coupon Collector Analysis

How to analyse the process?

Could evaluate *expected number of purchases* to get card  $i$ . For any  $i$ , the “number of steps”  $Y_i$  is a **geometric random variable** with parameter  $1/n$  such that  $E[Y_i] = \frac{1}{(1/n)} = n$ .

**But** the total number the purchases in expectation is not the summation of all  $Y_i$  (WHY?). Worse, they are also not independent! Better to find another angle and not focused on any **particular** card . . .

- ▶ At any stage of the process (having found some cards already), analyse the “further purchases” to get a *card not seen before*.
- ▶ Let  $X_i$  be the number of packets bought (after having  $i - 1$  different cards) to get the  $i$ th new card.
- ▶ Let  $X$  be the number of packets bought to get all cards.
- ▶ Clearly  $X = \sum_{i=1}^n X_i$ .

RC (2019/20) – Lecture 4 – slide 19

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n - (i - 1)}{n}$ .

RC (2019/20) – Lecture 4 – slide 20

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n-(i-1)}{n}$ .

Linearity of  $E[\cdot]$  says  $E[X] = \sum_{i=1}^n E[X_i]$ .

RC (2019/20) – Lecture 4 – slide 20

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n-(i-1)}{n}$ .

Linearity of  $E[\cdot]$  says  $E[X] = \sum_{i=1}^n E[X_i]$ .

By Lemma on geometric random variables  $E[X_i] = \frac{n}{n-(i-1)}$  for every  $i$ .

RC (2019/20) – Lecture 4 – slide 20

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n-(i-1)}{n}$ .

Linearity of  $E[\cdot]$  says  $E[X] = \sum_{i=1}^n E[X_i]$ .

By Lemma on geometric random variables  $E[X_i] = \frac{n}{n-(i-1)}$  for every  $i$ .

Hence  $E[X] = \sum_{i=1}^n \frac{n}{n-(i-1)} = \sum_{i=n}^1 \frac{n}{i} = n(\sum_{i=1}^n \frac{1}{i})$ .

RC (2019/20) – Lecture 4 – slide 20

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n-(i-1)}{n}$ .

Linearity of  $E[\cdot]$  says  $E[X] = \sum_{i=1}^n E[X_i]$ .

By Lemma on geometric random variables  $E[X_i] = \frac{n}{n-(i-1)}$  for every  $i$ .

Hence  $E[X] = \sum_{i=1}^n \frac{n}{n-(i-1)} = \sum_{i=n}^1 \frac{n}{i} = n(\sum_{i=1}^n \frac{1}{i})$ .

$H(n) = \sum_{i=1}^n \frac{1}{i}$  is a crude “Riemann sum” to approximate  $\int_{x=1}^n \frac{1}{x}$ .

Can show  $\int_{x=1}^n \frac{1}{x} < \sum_{i=1}^n \frac{1}{i}$  and  $\sum_{i=2}^n \frac{1}{i} < \int_{x=1}^n \frac{1}{x}$  (Fig 2.1 in book).

RC (2019/20) – Lecture 4 – slide 20

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n-(i-1)}{n}$ .

Linearity of  $E[\cdot]$  says  $E[X] = \sum_{i=1}^n E[X_i]$ .

By Lemma on geometric random variables  $E[X_i] = \frac{n}{n-(i-1)}$  for every  $i$ .

Hence  $E[X] = \sum_{i=1}^n \frac{n}{n-(i-1)} = \sum_{i=n}^1 \frac{n}{i} = n(\sum_{i=1}^n \frac{1}{i})$ .

$H(n) = \sum_{i=1}^n \frac{1}{i}$  is a crude “Riemann sum” to approximate  $\int_{x=1}^n \frac{1}{x}$ .

Can show  $\int_{x=1}^n \frac{1}{x} < \sum_{i=1}^n \frac{1}{i}$  and  $\sum_{i=2}^n \frac{1}{i} < \int_{x=1}^n \frac{1}{x}$  (Fig 2.1 in book).

Hence  $\ln(n) < \sum_{i=1}^n \frac{1}{i} \leq \ln(n) + 1$ .

RC (2019/20) – Lecture 4 – slide 20

## Coupon Collector Analysis

$X_i$  can also be modelled as a *geometric random variable*; if we own  $i - 1$  different cards, and buy one more packet, the (conditional) probability  $p_i$  that we get a *new* card is  $p_i = \frac{n-(i-1)}{n}$ .

Linearity of  $E[\cdot]$  says  $E[X] = \sum_{i=1}^n E[X_i]$ .

By Lemma on geometric random variables  $E[X_i] = \frac{n}{n-(i-1)}$  for every  $i$ .

Hence  $E[X] = \sum_{i=1}^n \frac{n}{n-(i-1)} = \sum_{i=n}^1 \frac{n}{i} = n(\sum_{i=1}^n \frac{1}{i})$ .

$H(n) = \sum_{i=1}^n \frac{1}{i}$  is a crude “Riemann sum” to approximate  $\int_{x=1}^n \frac{1}{x}$ .

Can show  $\int_{x=1}^n \frac{1}{x} < \sum_{i=1}^n \frac{1}{i}$  and  $\sum_{i=2}^n \frac{1}{i} < \int_{x=1}^n \frac{1}{x}$  (Fig 2.1 in book).

Hence  $\ln(n) < \sum_{i=1}^n \frac{1}{i} \leq \ln(n) + 1$ .

So the expected time  $E[X]$  to collect all cards is  $\sim n \ln(n)$ .

RC (2019/20) – Lecture 4 – slide 20

## Is “expected” the same as “typical”?

All we know (for Coupon collecting) is the “average” (weighted over random choices) number of cards.

We don’t know how likely one “run” of the process is to come close to that value.

Concentration inequalities help us bound the *deviation from the mean*:

- ▶ **Markov’s Inequality**;
- ▶ **Chebyshev’s Inequality**;
- ▶ **Chernoff Bound / Hoeffding inequality**.

RC (2019/20) – Lecture 4 – slide 21