Randomness and Computation

or, “Randomized Algorithms”

Heng Guo

(Based on slides by M. Cryan)

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 $\mathrm{E}X$ cannot be unbounded unless it has infinite support).

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 $\mathrm{E}X$ cannot be unbounded unless it has infinite support).

Theorem (2.1, Linearity of Expectation)

For any finite collection of discrete random variables $X_1, \ldots, X_k$ 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.

Linearity of Expectation

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

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

$$\mathrm{E}X_i = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Thus,

$$\mathrm{E}X = \sum_{i=1}^{n} \mathrm{E}X_i = 1.$$

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

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

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

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

Variance and second moment

Definition

The *k*-th moment is defined as

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

The variance is defined as

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

Variance

Lemma

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

Variance

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

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

Variance

Lemma

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

Proof.

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

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

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

Jensen's Inequality

Definition

A function $f : \mathbb{R} \to \mathbb{R}$ 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

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

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

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

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

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 $\mathrm{E}[Y] = p$ 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 $\{0, 1, \ldots, n\}$ with the probabilities

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

We can prove $\mathrm{E}[X] = np$ 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.

Conditional Expectation

Definition (2.6)

For two random variables X, Y,

$$\mathrm{E}[X \mid Y = y] \;=\; \sum_{x} x \cdot \Pr[X = x \mid Y = y],$$

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

Definition (2.7)

We also use the expression $\mathrm{E}[X \mid Y]$, where X, Y are random variables.

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

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

Conditional Expectation

Observation

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

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

Lemma (2.5)

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

$$E[E[X \mid Y]] = E[X].$$

Conditional Independence

Proof.

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

□

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 $j > 0$, $k \geq 0$,

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

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.) $\square$

Geometric distributions

Lemma

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

Proof.

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

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

$\square$

Alternatively

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

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

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

Coupon Collector Problem

"Coupon collecting" is the activity of buying cereal-packets, each of which will have a coupon inside. There are be 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 $1/n$.

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 $\mathrm{E}[Y_i] = \frac{1}{(1/n)} = n$.

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 ith new card.
- Let X be the number of packets bought to get all cards.
- Clearly $X = \sum_{i=1}^{n} X_i$.

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 $p_i = \frac{n-(i-1)}{n}$.

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

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

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

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

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

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

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

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

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

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ 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).*

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

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

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

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

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

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

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

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ 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 $\mathrm{E}[X]$ to collect all cards is $\sim n \ln(n)$.

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.