

MATH 407 Lecture 1

Chapter 1 : combinatorics

Basic counting principles

- perform two experiments. If experiment 1 has m possible outcomes, and experiment 2 has n possible outcomes, then there are mn possible outcomes from the two experiments
- more generally, if have r experiments
first has n_1 outcomes
second has n_2 outcomes
etc,

Then the total number of possible outcomes is $n_1 n_2 \dots n_r$.

Example

- (a) How many different 7-place license plates are possible if the first two places ~~are~~ are for letters and the

other five are for numbers?

Sol.

$$26 \times 26 \times 10 \times 10 \times 10 \times 10 \times 10 = 67600000$$

(b) Repeat (a) under the assumption that no letter or number can be repeated in a single license plate.

Sol.

$$26 \times 25 \times 10 \times 9 \times 8 \times 7 \times 6 = 19656000$$

Example

How many outcome sequences are possible when a die is rolled 4 times, and we say for instance that the outcome is 3, 4, 3, 1 if first roll is a 3, the second a 4, the third a 3, and the fourth a 1?

Sol. $6 \times 6 \times 6 \times 6 = 1296$

Permutations

How many different ordered arrangements of 1, 2, 3 are possible?

Sol. the possibilities are

123, 132, 213, 231, 312, 321.

so have 6 possibilities, which we call the permutations of 1,2,3.

In general, the number of permutations of n distinct objects is

$$n(n-1) \cdots 3(2)(1) = n!$$

we define $0! = 1$.

Example

John, Jim, Jay, Jack form a band with 4 instruments.

(a) If each boy can play all 4 instruments, how many arrangements are possible?

(b) What if John and Jim can play all 4 instruments, but Jay and Jack can only play the piano and drums?

4

Sol.

$$(a) 4! = 24$$

$$(b) 2! \times 2 \times 1 = 4$$

$\uparrow$ $\uparrow$ $\nwarrow$
 For Jay For For
 and Jack John Jim

Example

(a) In how many ways can 3 boys and 3 girls sit in a row?

$$\text{Sol. } 6! = 720$$

(b) In how many ways can 3 boys and 3 girls sit in a row, if the boys and girls are each to sit together?

Sol.

BBB GGG or GGG BBB

$\uparrow$ $\uparrow$
 $3! \times 3!$ $3! \times 3!$

$$\text{So } 36 + 36 = 72$$

5

(c) In how many ways if only boys must sit together?

Sol.

$$\begin{array}{ll}
 BBB GGG & 3! \times 3! = 36 \\
 G BBB GG & = 36 \\
 GG BBB G & = 36 \\
 GGG BBB & = 36
 \end{array}$$

$$\text{So } 36 + 36 + 36 + 36 = 144$$

(d) In how many ways if no 2 people of same gender are allowed to sit together?

Sol.

$$\begin{array}{ll}
 G B G B G B & 3 \times 3 \times 2 \times 2 \times 1 \times 1 = 36 \\
 B G B G B G & = 36
 \end{array}$$

$$\text{So total is } 36 + 36 = 72.$$

Example How many length 9 orderings of two 1's, four 2's, and three 3's are there?
(for example 1 3 2 2 1 2 3 3 2)

Sol.

Let $S_9 =$ permutations of $1, 2, \dots, 9$

Define a map $f: S_9 \longrightarrow$ orderings of
 two 1's
 four 2's
 three 3's.

Put 1's in positions of 1, 2
 2's in positions of 3, 4, 5, 6
 3's in positions of 7, 8, 9.

Thus

1 7 3 4 2 5 8 9 6

$\longrightarrow$ 1 3 2 2 1 2 3 3 2

The map f is onto, and each ordering of two 1's, four 2's, three 3's is hit by

$$2! \times 4! \times 3!$$

many permutations.

So the total number of orderings is

$$\frac{9!}{2! 4! 3!} = 1260.$$

The reasoning for this example yields

Principle : There are $\frac{n!}{n_1! n_2! \dots n_r!}$

different permutations of n objects, of which n_1 are alike, n_2 are alike, $\dots$ n_r are alike.

Example

How many different letter arrangements can be made from the letters, where we don't distinguish between capital and lower case?

(a) Fluke

$$\text{Sol. } \frac{5!}{1!1!1!1!1!} = 120$$

(b) Propose

$$\text{Sol. } \frac{7!}{2!2!1!1!1!1!} = 1260$$

(c) Mississippi

$$\text{Sol. } \frac{11!}{1!2!4!4!} = 34650$$

(d) Arrange

$$\text{Sol. } \frac{7!}{2!2!1!1!1!} = 1260$$

Example

A child has 12 blocks: 6 black
4 red
1 white
1 blue

If child puts the blocks in a line, how many arrangements are possible?

$$\text{Sol. } \frac{12!}{6!4!1!1!} = 27720$$

Combinations

How many ways can you choose 3 items from the 6 items a, b, c, d, e, f?

There are 6 ways of choosing 1st item
5 ways of choosing 2nd item
4 ways of choosing 3rd item

So have $6 \times 5 \times 4$ ways if order is relevant

If order isn't relevant, need to divide

9

by $3 \times 2 \times 1$ since $abc, acb, bac, bca, cab, cba$ give same subset of $\{a, b, c, d, e, f\}$.

In general, number of ways of choosing r items from n (where order is irrelevant) is

$$\frac{n(n-1) \cdots (n-r+1)}{r!} = \frac{n!}{(n-r)! r!}$$

Notation For $0 \leq r \leq n$, define

$$\binom{n}{r} \text{ as } \frac{n!}{(n-r)! r!}$$

(call this "n choose r".)

Example

Five separate awards (best scholarship, best leadership, etc.) are to be presented to selected students from a class of 30. How many different outcomes are possible if

(a) a student can receive any number of awards?

Sol. ↓ choices for leadership 10

$$30 \times 30 \times 30 \times 30 \times 30 = 24300000$$

↑
choices
for scholarship

(b) each student can receive at most one award?

Sol. $30 \times 29 \times 28 \times 27 \times 26 = 17100720.$

Alternately, pick 5 students to get the awards then assign the 5 awards to the five students.

Get $\binom{30}{5} 5! = 17100720.$

Example

Consider a group of 20 people. If everyone shakes hands with everyone else, how many handshakes take place?

Sol. $\binom{20}{2} = \frac{20(19)}{2} = 180.$

Example How many 5 card poker hands are there?

Sol. $\binom{52}{5} = \frac{(52)(51) \cdots (48)}{(5)(4) \cdots (1)}$

$$= 2593960$$

Example

A student has to select 2 books from a collection of 6 math, 7 science, and 4 economics books. How many choices are possible if

- Both books are on the same subject?
- The books are to be on different subjects?

Sol.

$$(a) \quad \binom{6}{2} + \binom{7}{2} + \binom{4}{2} = 15 + 21 + 6 = 42$$

$$(b) \quad \binom{6}{1}\binom{7}{1} + \binom{6}{1}\binom{4}{1} + \binom{7}{1}\binom{4}{1} = 94$$

$\uparrow$
 math and
 science

$\uparrow$
 math
 and econ

$\uparrow$
 science
 and econ

Example

Suppose have n balls, k red, $n-k$ white that all reds are considered indistinguishable

and all whites are considered indistinguishable.
How many linear orderings are there in which no 2 reds are adjacent?

Sol. Have $n-k$ whites lined up



So there are $n-k+1$ places where you can put the k reds. So the total number of orderings is $\binom{n-k+1}{k}$.

Useful identity

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \quad \text{if } 1 \leq r \leq n$$

Proof 1

$$\begin{aligned} \binom{n-1}{r-1} + \binom{n-1}{r} &= \frac{(n-1) - (n-r+1)}{(r-1) - (1)} + \frac{(n-1) - (n-r)}{r(r-1) - (1)} \\ &= \frac{(n-1) - (n-r+1)r}{r!} + \frac{(n-1) - (n-r+1)(n-r)}{r!} \end{aligned}$$

$$= \frac{n(n-1)\cdots(n-r+1)}{r!}$$

$$= \binom{n}{r}$$

Proof 2

Label n objects $1, 2, \dots, n$.

How many subsets of these objects have size r ?

on one hand, get $\binom{n}{r}$

on other hand,

$$\text{number of subsets of size } r \text{ containing } 1 = \binom{n-1}{r-1}$$

↑
choose $r-1$ elements
from $\{2, \dots, n\}$

$$\text{number of subsets of size } r \text{ not containing } 1 = \binom{n-1}{r}$$

↑
choose r elements
from $\{2, \dots, n\}$

The result follows.

Binomial theorem

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

Combinatorial proof

$$(x+y)(x+y) \cdots (x+y)$$

What is the coefficient of $x^k y^{n-k}$?

Have n terms $(x+y)$

choose k for which use x (so the other $n-k$ terms use y).

Thus coefficient of $x^k y^{n-k}$ is equal to $\binom{n}{k}$

next time: give an inductive proof

MATH 407 Lecture 2

From last time, binomial theorem says

$$\text{that } (x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

Inductive proof (ignore if you haven't seen induction)

Base case $n=1$

$$(x+y) = \binom{1}{0} x + \binom{1}{1} y$$

Induction step

Assume the result for $n-1$. Then

$$\begin{aligned} (x+y)^n &= (x+y)(x+y)^{n-1} \\ &= (x+y) \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-1-k} \end{aligned}$$

$$\begin{aligned} &= \sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1} y^{n-1-k} \\ &\quad + \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k} \end{aligned}$$

So if $0 \leq k < n$, the coefficient of

$x^k y^{n-k}$ in $(x+y)^n$ is

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}$$

$\uparrow$
 by a result from
 last time

Finally, note that

$$\text{Coefficient of } x^n \text{ in } (x+y)^n = 1$$

$$\text{Coefficient of } y^n \text{ in } (x+y)^n = 1 \quad \text{End of proof.}$$

Example

Expand $(3x^2+y)^5$

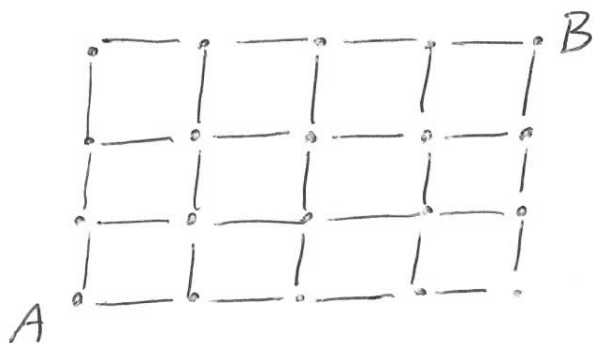
Sol.

$$(3x^2+y)^5 = \sum_{k=0}^5 \binom{5}{k} (3x^2)^k y^{5-k}$$

$$\begin{aligned}
 &= y^5 + 3\binom{5}{1}x^2y^4 + 9\binom{5}{2}x^4y^3 \\
 &\quad + 27\binom{5}{3}x^6y^2 + 81\binom{5}{4}x^8y \\
 &\quad + 243\binom{5}{5}x^{10}
 \end{aligned}$$

$$= y^5 + 15x^2y^4 + 90x^4y^3 + 270x^6y^2 + 405x^8y + 243x^{10} \quad 17$$

Example

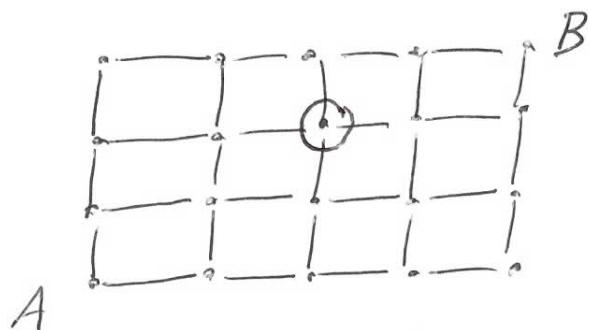


How many paths are there from A to B that at each step go up or to the right?

Solution

The path has 7 steps, and need to choose 3 which go up. So total is $\binom{7}{3} = 35$

Example



How many of the paths in previous example go through the circled point?

Sol.

$$\begin{aligned} \text{Answer} &= \left(\begin{array}{c} \text{number of paths} \\ \text{from A to O} \end{array} \right) \times \left(\begin{array}{c} \text{number of paths} \\ \text{from O to B} \end{array} \right) \\ &= \binom{4}{2} \binom{3}{1} = 18 \end{aligned}$$

multinomial coefficients

Have a set of n distinct items.

In how many ways can you divide it into groups $1, 2, \dots, r$ so that the group sizes are $n_1, n_2, \dots, n_r$, where $\sum n_i = n$.

Sol. 1

$$\begin{array}{l} \text{Number of choices} \\ \text{for first group} \end{array} = \binom{n}{n_1}$$

For each choice of first group, have $\binom{n-n_1}{n_2}$ choices for the second group.

For each choice of first two groups, have $\binom{n-n_1-n_2}{n_3}$ choices for third group.

etc.

Thus, the total number of choices is

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots$$

$$= \frac{n!}{(n-n_1)! n_1!} \frac{(n-n_1)!}{(n-n_1-n_2)! n_2!} \dots \frac{(n-n_1-\dots-n_{r-1})!}{0! n_r!}$$

$$= \frac{n!}{n_1! n_2! \dots n_r!}$$

Sol. 2

items	1	2	3	...	n
group item is in	2	3	3	...	2

So answer = number of words on ~~n~~ symbols with n_1 1's, n_2 2's, ..., n_r r's.

by earlier result $\rightarrow = \frac{n!}{n_1! n_2! \dots n_r!}$

Example

If 12 people are to be divided into 3 committees of sizes 3, 4, 5, how many divisions are possible?

we use notation

20

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$$

Sol.

$$\binom{12}{3, 4, 5} = \frac{12!}{3! 4! 5!} = 27,720$$

Example

(a) If 8 new teachers are divided among 4 schools, how many divisions are possible?

(b) What if each school must receive 2 teachers?

Sol.

4 choices for teacher 2

(a) $4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 = 65536$

4 choices for teacher 1

(b) $\binom{8}{2, 2, 2, 2} = \frac{8!}{2^4} = 2520$

Multinomial theorem

$$(x_1 + \dots + x_r)^n = \sum_{\substack{(n_1, \dots, n_r) \\ n_1 + \dots + n_r = n}} \binom{n}{n_1, n_2, \dots, n_r} x_1^{n_1} x_2^{n_2} \dots x_r^{n_r}$$

Proof Use combinatorial argument
as in case $r=2$.

Proposition There are $\binom{n-1}{r-1}$ distinct positive
integer valued vectors $(x_1, \dots, x_r)$ satisfying
 $x_1 + x_2 + \dots + x_r = n$, all $x_i > 0$.

Proof

$$\bullet \mid \bullet \bullet \mid \bullet \bullet \mid \bullet \bullet \Leftrightarrow 1+2+2+2=7$$

Have n dots and $r-1$ bars.

There are $n-1$ possible positions for the $r-1$
bars.

Prop. There are $\binom{n+r-1}{r-1}$ distinct non-negative
integer valued vectors $(x_1, \dots, x_r)$ satisfying

$$x_1 + x_2 + \dots + x_r = n, \text{ all } x_i \geq 0.$$

Proof Let $y_i = x_i + 1$ for all i .

Then solutions $(x_1, \dots, x_r)$ correspond to positive
integer solutions $(y_1, \dots, y_r)$ of

$$y_1 + \dots + y_r = n+r$$

By previous proposition, the number of such solutions is $\binom{n+r-1}{r-1}$

Example

- (a) If 8 identical blackboards are to be divided among 4 schools, how many divisions are possible?
- (b) How many if each school must receive at least 1 blackboard?

Sol.

- (a) seek number of solutions to

$$x_1 + x_2 + x_3 + x_4 = 8$$

with the x_i non-negative integers.

$$= \binom{8+3}{3} = \binom{11}{3} = 165$$

- (b) seek number of solutions to

$$x_1 + x_2 + x_3 + x_4 = 8$$

with the x_i positive integers

$$= \binom{8-1}{4-1} = \binom{7}{3} = 35$$

Example

An elevator starts at basement with 8 people (not including the operator) and discharges them all by the time it reaches the top floor, number 6.

(a) In how many ways could the operator have perceived people leaving the elevator if all people look alike to him?

Sol.

seek number of solutions to

$$x_1 + \dots + x_6 = 8, \text{ all } x_i \geq 0 \text{ are integers}$$

$$= \binom{8+5}{5} = \binom{13}{5} = 1287$$

(b) What if the 8 people consisted of 5 men and 3 women and the operator could tell a man from a woman?

Sol.

Number of solutions to $x_1 + \dots + x_6 = 5$, $x_i \geq 0$ is $\binom{10}{5}$

Number of solutions
to $x_1 + \dots + x_6 = 3$, $x_i \geq 0$ is $\binom{8}{5}$

So answer is $\binom{10}{5} \binom{8}{5} = (252/56)$
 $= 14, 112.$

Chapter 2 : axioms of probability

Sample space S = set of possible outcomes
of an experiment

Examples

1) Gender of a newborn child = $\{g, b\}$

2) Roll a die 3 times

$$S = \{(i, j, k) : 1 \leq i, j, k \leq 6\}$$

$$|S| = 6 \times 6 \times 6 = 216$$

3) Toss a coin twice

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

4) Lifetime of a light bulb (in hours)

$$S = \{x : 0 \leq x < \infty\}$$

5) Ranking of 3 athletes in long jump
competition

$S =$ all permutations of $\{1, 2, 3\}$

$$= \{123, 132, 213, 231, 312, 321\}$$

↑
means 2 comes in first
1 comes in second
3 comes in third

Definition

An event E is a subspace of sample space.

Examples

- 1) $E = \{b\}$ is event that child is a boy
- 2) $E = \{(1, 1, 5), (2, 2, 5), (3, 3, 5), (4, 4, 5), (5, 5, 5), (6, 6, 5)\}$
is the event that the first two rolls are the same, and third roll is a 5.
- 3) $E = \{(T, H), (T, T)\}$
is event that first toss is tails
- 4) $E = \{1 \leq x \leq 2.5\}$ is event bulb lasts between 1 and 2.5 hours
- 5) $E = \{213, 231\}$ is event that athlete 2 wins

Def. $E \cup F =$ outcomes either in E or F
or in both E and F

So $E \cup F$ occurs if either E or F does. 26

Example

Let $E = (T, H), (T, T)$ = event first coin is tails

Let $F = (H, T), (T, T)$ = event second coin is tails

Then $E \cup F = (T, H), (T, T), (H, T)$
= event at least one of tosses is tails

Def. $E \cap F$ = outcomes in both E and F

So if $E = (T, H), (T, T)$, $F = (H, T), (T, T)$
then $E \cap F = (T, T)$ = event both tosses are tails

Example When you roll 3 dice, if

$E = (1, 1, 2), (1, 1, 3), (1, 2, 4)$

$F = (2, 1, 1), (2, 2, 2)$

Then $E \cap F = \emptyset$

Def. If $E \cap F = \emptyset$, say that E, F are mutually exclusive

In other words, E, F can't both happen.

27

Similarly,

$\bigcup_{n=1}^{\infty} E_n$ is union of $E_1, E_2, \dots$

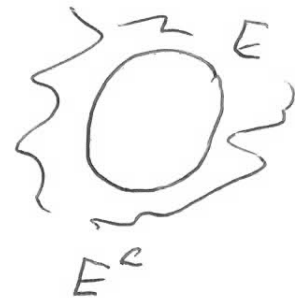
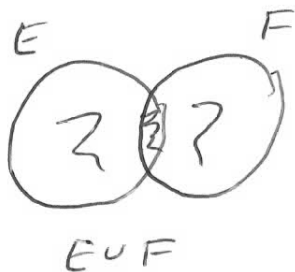
$\bigcap_{n=1}^{\infty} E_n$ is intersection of $E_1, E_2, \dots$

Given E , let $E^c =$ complement of E
= all outcomes in sample space that aren't in E

Example If toss two coins and $E = (T, T)$
then $E^c = (H, H), (H, T), (T, H)$.

If all outcomes in E are also in F , we say E is contained in F , and write $E \subset F$.

If $E \subset F$ and $F \subset E$, we say E, F are equal and write $E = F$



Rules

$$E \cup F = F \cup E$$

$$E \cap F = F \cap E$$

$$(E \cup F) \cup G = E \cup (F \cup G)$$

$$(E \cap F) \cap G = E \cap (F \cap G)$$

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$$

$$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$$

To prove these rules, must show that event on left of a rule is contained in event on right, and that event on right is contained in event on left.

Example

$$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$$

Proof

Step 1 Show $(E \cap F) \cup G \subseteq (E \cup G) \cap (F \cup G)$

If $x \in (E \cap F) \cup G$ then $x \in E \cap F$ or $x \in G$

If $x \in E \cap F$, then $x \in E \cup G$ and $x \in F \cup G$,

so $x \in (E \cup G) \cap (F \cup G)$

If $x \in G$, then $x \in (E \cup G) \cap (F \cup G)$

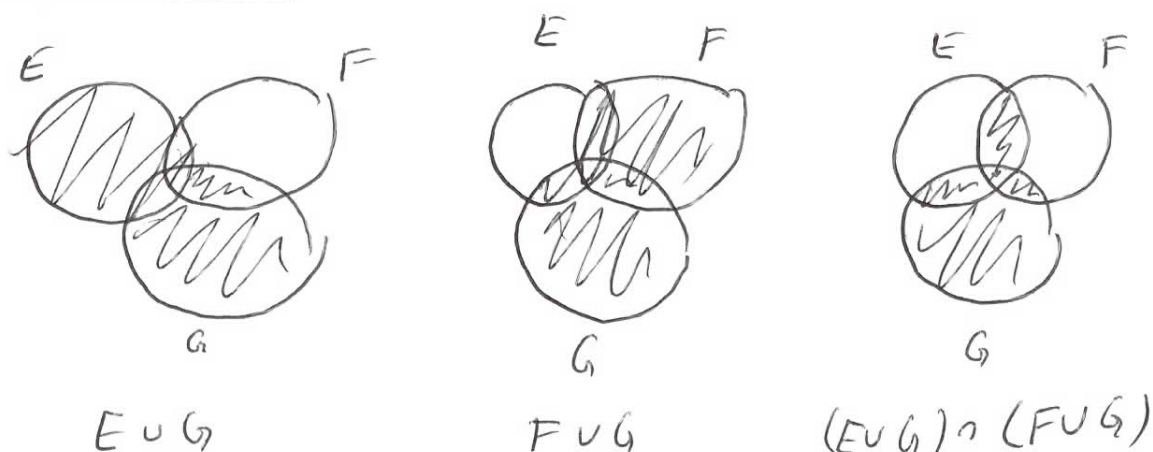
Step 2 Show $(E \cup G) \cap (F \cup G) \subseteq (E \cap F) \cup G$

If $x \in (E \cup G) \cap (F \cup G)$, then $x \in E \cup G$
and $x \in F \cup G$.

If $x \in G$, then $x \in (E \cap F) \cup G$

If $x \notin G$, then $x \in E$ and $x \in F$, so
 $x \in (E \cap F) \cup G$.

Visual proof



↑

From diagram, this is
equal to $(E \cap F) \cup G$

De Morgan's laws (prove next time)

$$1) \left(\bigcup_{i=1}^n E_i \right)^c = \bigcap_{i=1}^n E_i^c$$

$$2) \left(\bigcap_{i=1}^n E_i \right)^c = \bigcup_{i=1}^n E_i^c$$

MATH 407 Lecture 3De Morgan's Laws

$$1) \left(\bigcup_{i=1}^n E_i \right)^c = \bigcap_{i=1}^n E_i^c$$

$$2) \left(\bigcap_{i=1}^n E_i \right)^c = \bigcup_{i=1}^n E_i^c$$

Exercise: prove these for $n=2$ using Venn diagrams

Proof (part 1)

Suppose $x \in \left(\bigcup_{i=1}^n E_i \right)^c$. Thus x is not in any of the E_i 's. So $x \in E_i^c$ for all i . So $x \in \bigcap_{i=1}^n E_i^c$.

Thus $\left(\bigcup_{i=1}^n E_i \right)^c \subset \bigcap_{i=1}^n E_i^c$.

Suppose $x \in \bigcap_{i=1}^n E_i^c$. Then $x \in E_i^c$ for all i , so x is not in any E_i . Thus x is not in $\bigcup_{i=1}^n E_i$. So $x \in \left(\bigcup_{i=1}^n E_i \right)^c$. Thus $\bigcap_{i=1}^n E_i^c \subset \left(\bigcup_{i=1}^n E_i \right)^c$.

This proves 1).

Proof of part 2

Replace E_i by E_i^c in part 1.

Thus
$$\left(\bigcup_{i=1}^n E_i^c \right)^c = \bigcap_{i=1}^n E_i$$

Taking complements gives

$$\bigcup_{i=1}^n E_i^c = \left(\bigcap_{i=1}^n E_i \right)^c$$

Example

A box contains 3 marbles: 1 red, 1 green, 1 blue. Consider an experiment that takes one marble from the box, then replaces it and draws a second marble from the box.

- Describe the sample space
- Repeat when the second marble is drawn without replacing the first marble

Sol.

a) RR RG RB GR GG GB BR BG BB

b) RG RB GR GB BR BG

Example Two dice are rolled, let

E = event sum of rolls is odd

F = event at least one of dice lands on 1

G = event sum of dice is 5.

calculate $E \cap F$, $E \cap F \cap G$, $F \cap G$, $E \cup F$, $E \cap F^c$

We do the first 3, leave last 2 as exercise.

Sol. $E \cap F = (1,2), (1,4), (1,6), (2,1), (4,1), (6,1)$

$$E \cap F \cap G = (1,4), (4,1)$$

$$F \cap G = (1,4), (4,1)$$

Axioms of probability

1) $0 \leq P(E) \leq 1$ for all events E

2) $P(S) = 1$

3) If $E_1, E_2, \dots$ are mutually exclusive, then

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

there have consequences, such as

Prop. $P(E^c) = 1 - P(E)$

This is intuitively obvious.

Formal proof

$E \cup E^c = S$, and E, E^c are mutually exclusive.

Thus

$$P(E \cup E^c) = P(S) = 1$$

and $P(E \cup E^c) = P(E) + P(E^c)$

So $P(E) + P(E^c) = 1$

Example

If toss a coin 3 times, let

E = event all heads.

$$\begin{aligned} \text{Then } P(\text{at least one tail}) &= P(E^c) \\ &= 1 - P(E) \\ &= 1 - 1/8 = 7/8 \end{aligned}$$

Theorem

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Proof

$E \cup F = E \cup (E^c \cap F)$. Since $E, E^c \cap F$ are mutually exclusive,

$$P(E \cup F) = P(E) + P(E^c \cap F)$$

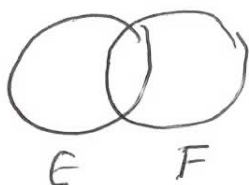
Also, $F = (E \cap F) \cup (E^c \cap F)$. Since $E \cap F$ and $E^c \cap F$ are mutually exclusive,

$$P(F) = P(E \cap F) + P(E^c \cap F)$$

Combining the two boxed equations proves the proposition.

End of proof

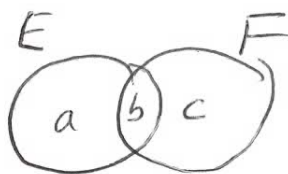
Intuitively,



$$\text{Area of } E \cup F = \text{Area of } E + \text{area of } F - \text{area of } E \cap F$$

↑
to correct for fact that $E \cap F$ is double counted.

More rigorously,



Since a, b, c are mutually exclusive,

$$P(E \cup F) = P(a) + P(b) + P(c)$$

$$P(E) = P(a) + P(b)$$

$$P(F) = P(b) + P(c)$$

$$\begin{aligned} \text{So } P(E \cup F) &= P(E) + P(F) - P(b) \\ &= P(E) + P(F) - P(E \cap F) \end{aligned}$$

Example

Suppose that A, B are mutually exclusive events for which $P(A) = .3$, $P(B) = .5$.

What is the probability that

- (a) either A or B occurs?
- (b) A occurs but B does not?
- (c) Both A and B occur?

Sol.

$$(a) .3 + .5 = .8$$

$$(b) P(A) - P(A \cap B) = .3 - 0 = .3$$

$$(c) 0$$

Example

36

A store accepts AMEX or VISA.

24%	of customers carry	AMEX
61%	- - -	VISA
11%	- - -	both

What percent of customers carry AMEX or VISA?

Sol. Let E = event that a random customer carries AMEX

F = event that a random customer carries VISA

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(E \cap F) \\ &= .24 + .61 - .11 = .74 \end{aligned}$$

So answer is 74%.

Example

Sixty percent of students at a school wear neither a ring nor a necklace.

Twenty percent wear a ring.

Thirty percent wear a necklace.

If a student is chosen randomly, what is chance student wears

(a) a ring or a necklace?

(b) a ring and a necklace?

Sol.

Let E = event a random student wears a ring

F = event a random student wears a necklace

$$(a) P(E \cup F) = 1 - .60 = .4$$

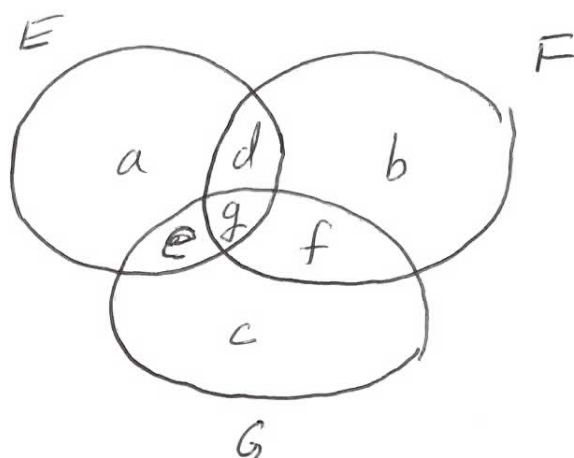
$$(b) P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$\uparrow$	$\uparrow$	$\uparrow$
.4	.2	.3

$$\text{So } P(E \cap F) = .5 - .4 = .1$$

Inclusion-exclusion

$$\begin{aligned}
 P(E \cup F \cup G) &= P(E) + P(F) + P(G) \\
 &\quad - P(E \cap F) - P(E \cap G) - P(F \cap G) \\
 &\quad + P(E \cap F \cap G)
 \end{aligned}$$

Proof

If $x \notin E \cup F \cup G$, it's counted 0 times on each side.

a is counted once on right (in term $P(E)$)

b is counted once on right

c is counted once on right

d is counted once on right, since

$$\begin{array}{ccccccc}
 & 1 & + & 1 & - & 1 & = & 1 \\
 \nearrow & & \nearrow & & \nearrow & & & \\
 \text{from} & & \text{from} & & \text{from} & & & \\
 E & & F & & E \cap F & & &
 \end{array}$$

Similarly,

e, f are each counted once on right
 g is counted once on right, since

$$\begin{array}{ccccccccccc}
 1 & + & 1 & + & 1 & - & 1 & - & 1 & - & 1 & + & 1 & = & 1 \\
 \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\
 \text{from } E & & \text{from } F & & \text{from } G & & \text{from } E \cap F & & \text{from } E \cap G & & \text{from } F \cap G & & \text{from } E \cap F \cap G
 \end{array}$$

In general,

Theorem For any events $E_1, \dots, E_n$

$$P(E_1 \cup \dots \cup E_n) = \sum_i P(E_i)$$

$$- \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2})$$

$$\vdots$$

$$+ (-1)^{n+1} \sum_{i_1 < i_2 < \dots < i_n} P(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_n})$$

$$\vdots$$

$$+ (-1)^{n+1} P(E_1 \cap \dots \cap E_n)$$

Proof

Consider $x \in S = \text{sample space}$.

If $x \notin E_1 \cup \dots \cup E_n$, then it contributes 0 to both sides.

Suppose $x \in E_1 \cup \dots \cup E_n$, and that x is in exactly $m > 0$ of the E_i 's.

on left side, x is counted once.

on right side, x is counted

$$\binom{m}{1} - \binom{m}{2} + \dots + (-1)^{m+1} \binom{m}{m} \text{ many times}$$

So we need

$$1 = \binom{m}{1} - \binom{m}{2} + \dots + (-1)^{m+1} \binom{m}{m}$$

$$\text{Since } 1 = \binom{m}{0}, \text{ we need } \sum_{i=0}^m \binom{m}{i} (-1)^i = 0.$$

By binomial theorem,

$$\begin{aligned} 0 &= (-1+1)^m = \sum_{i=0}^m \binom{m}{i} (-1)^i (1)^{m-i} \\ &= \sum_{i=0}^m \binom{m}{i} (-1)^i \end{aligned}$$

End of proof.

Remark

$$P(E_1 \cup \dots \cup E_n) \leq \sum_{i=1}^n P(E_i)$$

$$P(E_1 \cup \dots \cup E_n) \geq \sum_{i=1}^n P(E_i) - \sum_{i < j} P(E_i \cap E_j)$$

$$P(E_1 \cup \dots \cup E_n) \leq \sum_{i=1}^n P(E_i) - \sum_{i < j} P(E_i \cap E_j) \\ + \sum_{i < j < k} P(E_i \cap E_j \cap E_k)$$

etc.

Example

A school offers 3 classes: Spanish, French, German

$\begin{matrix} P & P & P \\ E & F & G \end{matrix}$

The school has 100 students:

28 in Spanish
 26 in French
 16 in German
 12 in Spanish and French
 4 in Spanish and German
 6 in French and German
 2 in all 3.

(a) If student is chosen randomly, what's

the probability that she is in no classes?

Sol.

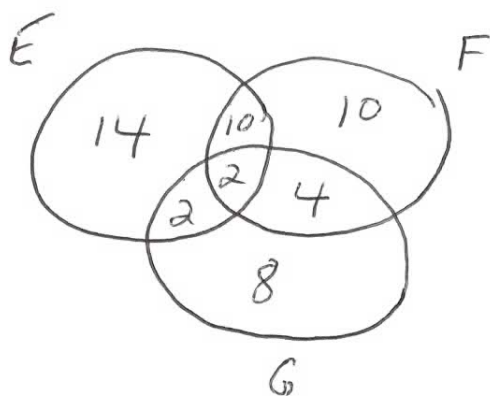
By inclusion-exclusion,

$$|E \cup F \cup G| = 28 + 26 + 16 - 12 - 4 - 6 + 2 = 50$$

So $100 - 50 = 50$ students are in no classes,
so chance is $50/100 = .5$.

(b) What is the probability that a student is in exactly 1 class.

Sol. Working from the center outward, we get the diagram



$$\text{Now } \frac{14 + 10 + 8}{100} = .32$$

Example

A town with 100,000 people has 3 newspapers: I, II, III. The proportions of people who read these papers are

$$I = 10\%$$

$$I \text{ and } II = 8\%$$

$$II = 30\%$$

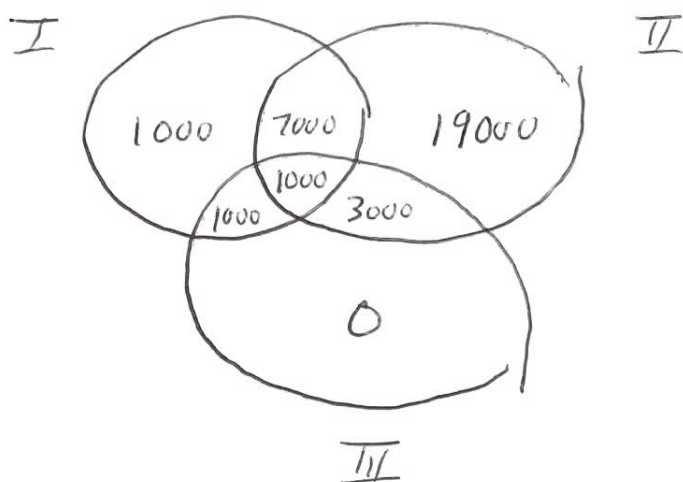
$$I \text{ and } III = 2\%$$

$$I, II, \text{ and } III = 1\%$$

$$III = 5\%$$

$$II \text{ and } III = 4\%$$

Thus, starting with the triple intersection, can fill in the Venn diagram



This allows us to answer many questions.

(a) How many people read only one news paper?

Sol. $1000 + 19000 = 20000.$

(b) How many people read at least two newspapers?

Sol. $1000 + 1000 + 7000 + 3000 = 12000$

(c) If I, III are morning papers and II is an evening paper, how many ~~paper~~ people read at least one morning paper plus an evening paper?

Sol. $7000 + 3000 + 1000 = 11000.$

(d) How many people do not read any newspaper?

Sol.

$$\begin{aligned} 100000 - (1000 + 7000 + 1000 + 1000 \\ + 19000 + 3000) \\ = 68000 \end{aligned}$$

(e) How many people read only one morning paper and one evening paper?

Sol. $7000 + 3000 = 10000.$

MATH 407 Lecture 4Sample spaces with equally likely outcomes

In many problems, the outcomes in a sample space are equally likely. Then for any event E ,

$$P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}$$

Example Roll two dice. What is probability that sum is 8?

Sol.

$$S = \{(x, y) : 1 \leq x, y \leq 6\}$$

So $|S| = 36$ and all outcomes are equally likely.

$$E = (2, 6), (3, 5), (4, 4), (5, 3), (6, 2)$$

$$\text{So } P(E) = |E|/|S| = 5/36$$

Note: why is $(3, 5) \neq (5, 3)$?

- Can imagine rolling dice sequentially
Then clearly $(3, 5) \neq (5, 3)$

46

Or can color one die white and other black. Then clearly $(3,5) \neq (5,3)$.

Example (birthday problem)

If n people are in a room, what's the probability that no 2 have the same birthday? How large need n be so that this probability is less than $1/2$? Let's assume that each year has 365 days.

Sol. Assuming each birthday is equally likely, the chance that no 2 have the same birthday is:

$$\frac{365}{365} \cdot \frac{364}{365} \cdots \frac{(365-n+1)}{365}$$

When $n \geq 23$, this is $< 1/2$.

Remark There are many variations on the birthday problem. For example,

- what if all birthdays aren't equally likely?
- what if want the chance that no 3 people have the same birthday?

Example (matching problem)

Suppose have n people at a party. Each throws a hat into center of room. The hats are mixed up, then each person selects a random hat? What is probability that no one gets own hat back?

Sol.

Let E_i = event person i gets own hat back

By inclusion-exclusion,

$$\begin{aligned}
 P\left(\bigcup_{i=1}^n E_i\right) &= \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) \\
 &\quad \dots \\
 &\quad + (-1)^{n+1} P(E_1 \cap E_2 \cap \dots \cap E_n)
 \end{aligned}$$

Observe that for any r , and $i_1, \dots, i_r$

$$\begin{aligned} P(E_{i_1} \cap \dots \cap E_{i_r}) &= P(\text{people } i_1, \dots, i_r \text{ get own hat back}) \\ &= (n-r)!/n! \end{aligned}$$

We have $\binom{n}{r}$ terms of this form.

Thus,

$$\begin{aligned} P\left(\bigcup_{i=1}^n E_i\right) &= \binom{n}{1} \frac{(n-1)!}{n!} - \binom{n}{2} \frac{(n-2)!}{n!} \\ &\quad + \binom{n}{3} \frac{(n-3)!}{n!} - \dots \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + \frac{(-1)^{n+1}}{n!} \end{aligned}$$

So $P(\text{no one gets hat back})$

$$\begin{aligned} &= 1 - P\left(\bigcup_{i=1}^n E_i\right) \\ &= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!} \end{aligned}$$

Since $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$, this probability

tends to $1/e = .3677$ as $n \rightarrow \infty$

49

Example

Assume all $\binom{52}{5}$ poker hands are equally likely. What is the probability of being dealt:

(a) Flush (all 5 cards are of same suit)

Sol.

$$\frac{\binom{4}{1} \binom{13}{5}}{\binom{52}{5}} = .0020$$

pick suit then pick 5 ranks

(b) One pair (cards have ranks a, a, b, c, d where a, b, c, d are distinct)

Sol.

$$\frac{\binom{13}{1} \binom{4}{2} \binom{12}{3} \binom{4}{1}^3}{\binom{52}{5}} = .4226$$

rank of a suits for the pair other 3 ranks suits for the other 3 ranks

(c) Two pairs(ranks are a, a, b, b, c with a, b, c distinct)Sol.

$$\frac{\overbrace{\binom{13}{2} \binom{4}{2}^2}^{\substack{\text{ranks for the pairs} \\ \text{suits for the pairs}}} \underbrace{44}_{\substack{\text{remaining cards} \\ \text{with ranks} \\ \text{different from } a, b}}}{\binom{52}{5}} = .0475$$

(d) Three of a kind(ranks are a, a, a, b, c with a, b, c distinct)Sol.

$$\frac{\overbrace{\binom{13}{1} \binom{4}{3} \binom{12}{2} 4^2}^{\substack{\text{rank of } a \\ \text{suits for } a \\ \text{choose ranks for } b, c}}}{\binom{52}{5}} \underbrace{\quad}_{\substack{\text{choose suits for } b, c}} = .0211$$

(e) Four of a kind(ranks are a, a, a, a, b with $a \neq b$)Sol.

$$\frac{\overbrace{\binom{13}{1} 48}^{\substack{\text{choose } a \\ \text{any card with rank } \neq a}}}{\binom{52}{5}} = .00024$$

Example

2 cards are randomly selected from an ordinary deck. What is the chance they form a blackjack (one card is an ace, other is a 10, J, Q, or K)?

Sol. ↙ suit for ace
 ↘ number of options
 for 10, J, Q, or K

$$\frac{\binom{4}{1} 16}{\binom{52}{2}} = .048$$

Example Roll 2 dice. What is the probability that the second die lands on a higher number than the first?

Sol.

Sample space $S = \{(x, y) : 1 \leq x, y \leq 6\}$

So $|S| = 36$ and all outcomes in S are equally likely.

$$E = (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 3), (2, 4), (2, 5), (2, 6), (3, 4), (3, 5), (3, 6), (4, 5), (4, 6), (5, 6)$$

$$\text{So chance} = |E|/|S| = 15/36 = 5/12 \quad 52$$

Example

A pair of dice is rolled until a sum of either 5 or 7 appears. Find the probability that a 5 occurs first.

Sol. For $n \geq 1$,

let $E_n = 5$ on n th roll, no 5 or 7 on first $n-1$ rolls

$$\text{Then } P(E_n) = \left(\frac{26}{36}\right)^{n-1} \frac{4}{36}$$

Since the E_n 's are mutually exclusive,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} E_n\right) &= \frac{4}{36} \sum_{n=1}^{\infty} \left(\frac{26}{36}\right)^{n-1} \\ &= \frac{4}{36} \left[1 + \frac{26}{36} + \left(\frac{26}{36}\right)^2 + \dots\right] \\ &= \frac{4}{36} \frac{1}{1 - 26/36} \\ &= \frac{4}{10} = .4 \end{aligned}$$

Note: This is $< 1/2$, as expected.

Example

An urn contains 5 red, 6 blue, 8 green balls.
Choose 3 balls at random. What is
chance that

- (a) balls are of same color
 - (b) balls are of different colors
 - (c) } repeat (a), (b) when the balls
 - (d) } are sampled with replacement
- (so when a ball is selected, its
color is noted and then it's
replaced in urn before next
selection)

Sol.

(a) RRR, BBB, GGG

$$\swarrow \qquad \downarrow \qquad \searrow$$

$$\left(\frac{5}{19}\right)\left(\frac{4}{18}\right)\left(\frac{3}{17}\right) + \left(\frac{6}{19}\right)\left(\frac{5}{18}\right)\left(\frac{4}{17}\right) + \left(\frac{8}{19}\right)\left(\frac{7}{18}\right)\left(\frac{6}{17}\right)$$

$$= \frac{516}{5814} = .0888$$

(b) RBG, RGB, BRG, BGR, GRB, GBR⁵⁴

Each of these six outcomes occurs with probability $(\frac{5}{19})(\frac{6}{18})(\frac{8}{17})$

$$\text{So get } 6(\frac{5}{19})(\frac{6}{18})(\frac{8}{17}) = .2477$$

(c) RRR, BBB, GGG

$$\begin{array}{ccc} \downarrow & \downarrow & \downarrow \\ (\frac{5}{19})^3 & + & (\frac{6}{19})^3 + (\frac{8}{19})^3 = .124 \end{array}$$

(d) RBG, RGB, BRG, BGR, GRB, GBR

Each of these six outcomes occurs with probability $(\frac{5}{19})(\frac{6}{19})(\frac{8}{19})$,

$$\text{so get } 6(\frac{5}{19})(\frac{6}{19})(\frac{8}{19}) = .2099$$

Example

Compute the probability that a bridge hand (that is a random 13 card hand) is void

in at least one suit (that is it has no cards in that suit).

55

Note that the answer is not

$$\binom{4}{1} \binom{39}{13} / \binom{52}{13}.$$

Why not?

Sol.

Let E_1 = void in clubs

E_2 = void in diamonds

E_3 = void in hearts

E_4 = void in spades

If E_1, E_2, E_3, E_4 were mutually exclusive, we'd have

$$\begin{aligned} P(\text{void in at least one suit}) &= P(E_1 \cup E_2 \cup E_3 \cup E_4) \\ &= 4P(E_1) \\ &= 4 \binom{39}{13} / \binom{52}{13} \end{aligned}$$

But E_1, E_2, E_3, E_4 aren't mutually exclusive

By inclusion-exclusion principle, we have

$$P(E_1 \cup E_2 \cup E_3 \cup E_4)$$

$$= P(E_1) + P(E_2) + P(E_3) + P(E_4)$$

$$- P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_1 \cap E_4)$$

$$- P(E_2 \cap E_3) - P(E_2 \cap E_4) - P(E_3 \cap E_4)$$

$$+ P(E_1 \cap E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_4)$$

$$+ P(E_2 \cap E_3 \cap E_4) + P(E_1 \cap E_3 \cap E_4)$$

$$- P(E_1 \cap E_2 \cap E_3 \cap E_4)$$

$$\begin{aligned} & \xrightarrow{\text{by symmetry}} = 4 P(E_1) - 6 P(E_1 \cap E_2) + 4 P(E_1 \cap E_2 \cap E_3) \\ & \quad - P(E_1 \cap E_2 \cap E_3 \cap E_4) \end{aligned}$$

$$= 4 \frac{\binom{39}{13}}{\binom{52}{13}} - 6 \frac{\binom{26}{13}}{\binom{52}{13}} + 4 \frac{\binom{13}{13}}{\binom{52}{13}} - 0$$

$$= .051$$

Chapter 3 : conditional probability
and independence

Example

If you bet two or more heads will appear in 3 tosses of a fair coin, you are more likely to win the bet given that the first toss is heads than given that the first toss is tails.

To be precise, assume the 8 patterns of heads and tails in 3 tosses

$\{hhh, hht, hth, htt, thh, tht, tth, ttt\}$

are equally likely. Then the overall or unconditional probability of the event

$A = 2 \text{ or more heads in 3 tosses}$

$= \{hhh, hht, hth, thh\}$

is $P(A) = 4/8 = 1/2$.

But given that first toss lands heads (call this event H), the event A occurs if and only if there is at least one head in the

next two tosses, an event of chance $3/4$. 58

So we say that "conditional probability of A given H " is $3/4$, and we write

$$P(A|H) = 3/4.$$

To see this another way, note that $H = \{hhh, hht, hth, htt\}$ can occur in 4 ways, and that 3 of these outcomes make A occur.

Counting formula for $P(A|B)$

For a finite set S of equally likely outcomes, and events A and B represented by subsets of S ,

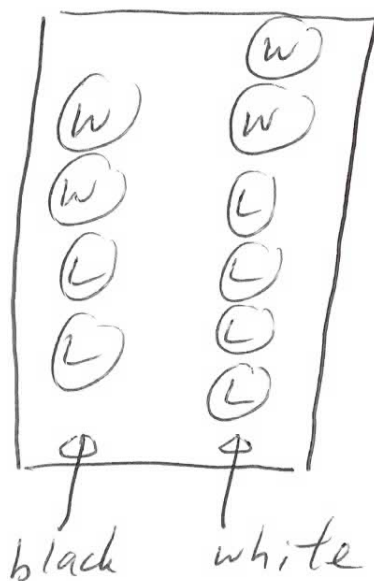
$$P(A|B) = |A \cap B| / |B|$$

Example

A box contains 10 capsules, similar except 4 are black and 6 are white. You can't see inside capsule without breaking it open.

Pick a random capsule from the box, then break it open to read the result.

If it says win, you win a prize;
otherwise, you win nothing. The number of
winning and losing tickets of each color
are given in diagram



W = win
L = lose

Suppose a capsule is randomly drawn, but
not yet broken open to read the result.
The capsule is black. What is the probability
that you win a prize?

Sol.

$$P(\text{win} | \text{black}) = \frac{\text{number of win and black capsules}}{\text{number of black capsules}} \\ = 2/4 = .5$$

MATH 407 Lecture 5General definition of conditional probability

Don't need to assume that all outcomes are equally likely.

$$\text{If } P(F) > 0, \text{ define } P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Example

Roll two fair dice Find the conditional probability that at least one die lands on a 6 given that they land on different numbers.

Sol.

Let E = at least one lands on 6

F = dice land on different numbers

$$P(E|F) = P(E \cap F) / P(F)$$

$$\text{Now } P(E \cap F) = 10/36 \quad \text{since } E \cap F \Leftrightarrow \begin{matrix} (6,1), (6,2), (6,3), (6,4), \\ (6,5), (1,6), (2,6) \\ (3,6), (4,6), (5,6) \end{matrix}$$

$$\text{Also, } P(F) = 1 - \text{chance dice land on same number} = 1 - 1/6 = 5/6$$

Thus $P(E|F) = \frac{10/36}{5/6} = 1/3$

Example

Roll 2 dice. What is the conditional probability that first die is 6 given that sum of dice is i ? Compute this for all values of i from 2 to 12

Sol.

Let E = first die is 6

F = sum of dice is i

If $i = 2, 3, 4, 5$, or 6 , then $P(E \cap F) = 0$,
so $P(E|F) = 0$.

$i = 7$ $P(E \cap F) = 1/36$, $P(F) = 6/36$

so $P(E|F) = 1/6$

$i = 8$ $P(E \cap F) = 1/36$, $P(F) = 5/36$

so $P(E|F) = 1/5$

Similarly, if $i = 9, 10, 11, 12$,

$P(E|F) = 1/4, 1/3, 1/2, 1$ respectively

Multiplication rule

Since $P(E|F) = \frac{P(E \cap F)}{P(F)}$, it follows that

$$P(E \cap F) = P(F) P(E|F).$$

More generally,

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) P(E_2|E_1) P(E_3|E_1 \cap E_2) \\ \dots P(E_n|E_1 \cap \dots \cap E_{n-1})$$

Proof

The right hand side is

$$\frac{P(E_1)}{P(E_1)} \frac{P(E_1 \cap E_2)}{P(E_1)} \frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1 \cap E_2)} \dots \frac{P(E_1 \cap \dots \cap E_n)}{P(E_1 \cap \dots \cap E_{n-1})} \\ = P(E_1 \cap \dots \cap E_n)$$

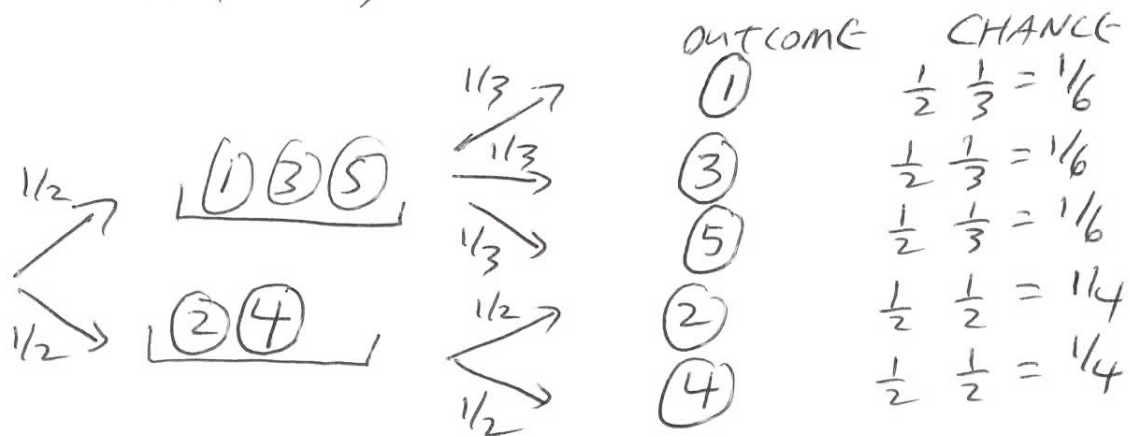
Example

Suppose there are 2 boxes, labeled odd and even. Odd box contains balls numbered 1, 3, 5. Even box contains balls numbered 2, 4.

One of the boxes is picked at random by tossing a fair coin. Then a ball is picked at

random from the box. What is the probability that the ball drawn is a 3? 63

Sol. Use a tree diagram. Each branch represents possible ways things might turn out. Probabilities and conditional probabilities are indicated along branch.



Remarks

1) A naive approach to above problem would be to assume all outcomes are equally likely, but this would imply

$$P(\text{odd box}) = 3/5, \quad P(\text{even box}) = 2/5$$

which is inconsistent with box being chosen by a fair coin toss.

2) The problem could also be solved without conditional probabilities by a symmetry argument, assuming that

$$P(1) = P(3) = P(5), \quad P(2) = P(4)$$

$$P(1) + P(3) + P(5) = P(2) + P(4) = 1/2$$

These equations yield same answer as above.

Example (Monty Hall problem)

Suppose you're on Monty Hall's "Let's make a deal". You're given the choice of 3 doors: behind one is a car, behind the others goats.

You pick a door, say 1. Monty opens another door, say 3, which has a goat. Monty says to you: do you want to switch to door 2? Should you switch?

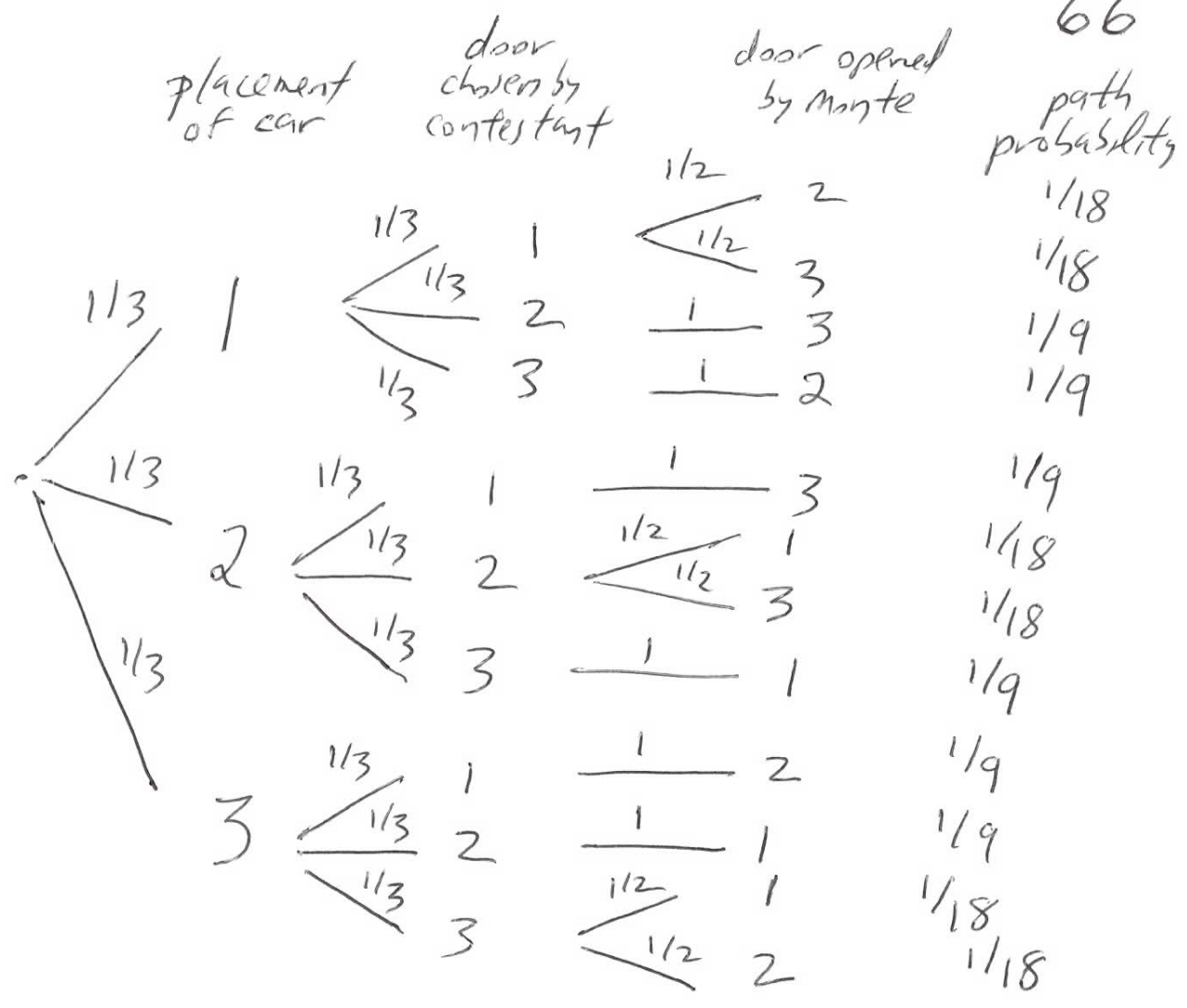
Assume

- car was put behind a random door ($1/3$ chance of each door)
- Monty knows where the ~~car~~ car is and always opens a door with a goat behind it
- If Monty has a choice of doors (that is the contestant picks door with car behind it), then Monty chooses each door with probability $1/2$.

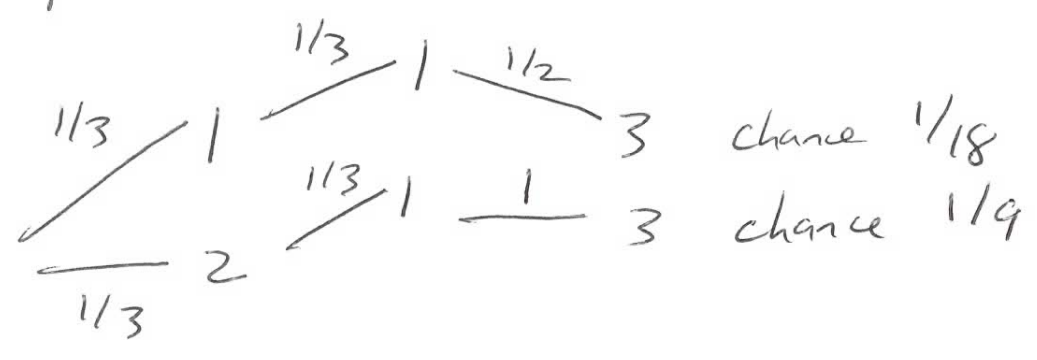
Note

- 1) Turns out you should switch, in which case the winning probability is $2/3$.
- 2) Incorrect argument: since initial guess is correct with probability $1/3$ chance of winning when stay is $1/3$.

This answers the wrong question. Want conditional probability of winning if switch, given that you chose door 1 and Monty opens door 3. More on this coming.



Now given that contestant chose door 1 and Monty opened door 3, only two paths are possible



67

So conditional probability of first path is $1/18 / (1/18 + 1/9) = 1/3$ and conditional probability of second path is $1/9 / (1/18 + 1/9) = 2/3$.

So you should switch, in which case you win with probability $2/3$.

Equivalently

$$\begin{aligned}
 & P(\text{car behind door one} \mid \text{you chose door 1 and Monty opened door 3}) \\
 &= \frac{P(\text{car behind door 1 and you chose door 1 and Monty opened door 3})}{P(\text{you chose door 1 and Monty opened door 3})} \\
 &= \frac{1/18}{1/18 + 1/9} = 1/3
 \end{aligned}$$

Note Suppose when Monty has a choice, he

chooses the door with the larger number with probability $3/4$, not $1/2$. Then if the contestant switches, wins with probability $4/7$, not $2/3$. This shows that earlier "Bogus argument" is wrong, since $4/7 \neq 2/3$.

Example

Consider a family with 2 children. Given that at least one of children is a boy, what's the chance both are boys?

Sol.

method 1

Let E = both are boys

F = at least one is a boy

$$P(E \cap F) = 1/4, \quad P(F) = 3/4, \quad \text{so } P(E|F) = \frac{1/4}{3/4} = 1/3$$

method 2 Use tree diagram

first child	second child	unconditional probability
$\frac{1}{2}$ b	$\frac{1}{2}$ b	$\frac{1}{4}$
	$\frac{1}{2}$ g	$\frac{1}{4}$
$\frac{1}{2}$ g	$\frac{1}{2}$ b	$\frac{1}{4}$
	$\frac{1}{2}$ g	$\frac{1}{4}$

first child	second child	uncond. prob.	conditional prob. given at least 1 boy
$\frac{1}{2}$ b	$\frac{1}{2}$ b	$\frac{1}{4}$	$\frac{1}{3}$ boy
	$\frac{1}{2}$ g	$\frac{1}{4}$	$\frac{1}{3}$
$\frac{1}{2}$ g	$\frac{1}{2}$ b	$\frac{1}{4}$	$\frac{1}{3}$

↑
since

$$\frac{1/4}{(1/4 + 1/4 + 1/4)} = 1/3$$

Example

Mr-Smith is the father of two. You meet him walking along street with a young boy

who is his son. What is the chance 70
that Mr. Smith's other child is also a
boy?

Assume: if Mr. Smith has a boy and a girl,
he is equally likely to choose either one
to accompany him on walk.

Sol.

Mr. Smith's children	walking with Mr. Smith	unconditional probability
$\frac{1/4}{1/4}$ bb	$\frac{1}{1/2}$ b	$1/4$
$\frac{1/4}{1/4}$ bg	$\frac{1/2}{1/2}$ b	$1/8$
$\frac{1/4}{1/4}$ gb	$\frac{1/2}{1/2}$ g	$1/8$
$\frac{1/4}{1/4}$ gg	$\frac{1}{1/2}$ b	$1/8$
	$\frac{1}{1/2}$ g	$1/8$
		$1/4$

Mr. Smith's children	walking with Mr. Smith	uncond. prob.	cond. prob.
$\frac{1/4}{1/4}$ bb	$\frac{1}{1/2}$ b	$1/4$	$1/2$
$\frac{1/4}{1/4}$ bg	$\frac{1/2}{1/2}$ b	$1/8$	$1/4$
$\frac{1/4}{1/4}$ gb	$\frac{1/2}{1/2}$ b	$1/8$	$1/4$

So answer is $\frac{1/4}{(1/4 + 1/8 + 1/8)} = 1/2$

Averaging conditional probabilities

Let A, B be events with $P(B), P(B^c) > 0$.

Then

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Proof

Right hand side

$$= \frac{P(A \cap B)}{P(B)} P(B) + \frac{P(A \cap B^c)}{P(B^c)} P(B^c)$$

$$= P(A \cap B) + P(A \cap B^c)$$

$$= P(A), \text{ since } A \text{ is disjoint union of } A \cap B \text{ and } A \cap B^c.$$

More generally, if state space is disjoint union of $B_1, \dots, B_n$, then

$$P(A) = P(A|B_1)P(B_1) + \dots + P(A|B_n)P(B_n)$$

Proof Same argument

Example Suppose two cards are dealt from a well-shuffled deck. What is the probability

that the second card is black?

72

Sol.

A common response to this question is that you can't say: it depends on whether the first card is black or not.

If first card is black, the chance the second card is black is $25/51$.

If first card is red, the chance the second card is black is $26/51$.

These are the conditional probabilities of black on second card given black and red respectively on first card.

But the question doesn't refer to the first card at all.

The overall probability of black on the second card is the average of these conditional probabilities.

$$P(\text{2nd black}) = P(\text{2nd black} \mid \text{1st black}) \times P(\text{1st black}) \\ + P(\text{2nd black} \mid \text{1st red}) \times P(\text{1st red})$$

$$= \left(\frac{25}{51}\right) \frac{1}{2} + \left(\frac{26}{51}\right) \frac{1}{2} = \frac{1}{2}$$

Remark: Can also compute this by symmetry.

Since there are an equal number of black and red cards in the deck,

$$P(\text{2nd black}) = P(\text{2 red})$$

Since $P(\text{2nd black}) + P(\text{2nd red}) = 1$, the result follows.

Bayes' formula

Let S = sample space

Have a set of events $H_1, H_2, \dots, H_m$ that are pairwise disjoint and such that $S = H_1 \cup \dots \cup H_m$.

Call these events hypotheses.

Also have an event E that gives some information about which hypothesis is correct. Call E the evidence.

Before we receive the evidence, have a 74
set of prior probabilities $P(H_1), P(H_2), \dots, P(H_m)$
for the hypotheses. If we know the correct
hypothesis, we know the probability of the
evidence. That is, we know $P(E|H_i)$ for all i .

want to find the probabilities for the
hypotheses given the evidence. That is, we want
to find the conditional probabilities $P(H_i|E)$
These are called "posterior probabilities".

$$\text{Now } P(H_i|E) = P(H_i \cap E) / P(E)$$

$$P(H_i \cap E) = P(H_i) P(E|H_i)$$

$$\begin{aligned} P(E) &= P(H_1 \cap E) + \dots + P(H_m \cap E) \\ &= P(H_1)P(E|H_1) + \dots + P(H_m)P(E|H_m) \end{aligned}$$

thus

$$P(H_i|E) = \frac{P(H_i) P(E|H_i)}{\sum_{k=1}^m P(H_k) P(E|H_k)}$$

Bayes' formula

MATH 407 Lecture 6Bayes' formula continued

From last time, have hypotheses $H_1, \dots, H_m$ and evidence E . Then

$$P(H_i | E) = \frac{P(H_i) P(E | H_i)}{\sum_{k=1}^m P(H_k) P(E | H_k)}$$

Example

A doctor is trying to decide if a patient has one of 3 diseases d_1, d_2, d_3 (assume patient doesn't have more than 1 disease). Two tests are to be carried out, each of which results in a positive (+) or a negative (-) outcome.

There are 4 possible test patterns:

$++$, $+-$, $-+$, $--$

National records indicate that for 10,000 people having one of these diseases, the

76

distribution of diseases and test results
is in the following table

Disease	Number having disease	++	+-	-+	--
d_1	3215	2110	301	704	100
d_2	2125	396	132	1187	410
d_3	4660	510	3568	73	509
Total	10,000				

From this table, we can estimate the prior probabilities for each of the diseases, and given a particular disease, the probability of a particular test outcome.

For example, the prior probability of disease d_1 may be estimated as $3215/10000 = .3215$.

The probability of the test result $+-$ given disease d_1 may be estimated to be $301/3215 = .094$.

Using Bayes' formula, we can compute the posterior probabilities

	d_1	d_2	d_3
++	.700	.132	.168
+-	.076	.033	.891
-+	.357	.605	.038
--	.066	.600	.334

Table of posterior probabilities

For example,

$$P(d_1 | ++)= \frac{P(d_1)P(++|d_1)}{P(d_1)P(++|d_1) + P(d_2)P(++|d_2) + P(d_3)P(++|d_3)}$$

$$= \frac{\frac{3215}{10000} \cdot \frac{2110}{3215}}{\frac{3215}{10000} \cdot \frac{2110}{3215} + \frac{2125}{10000} \cdot \frac{396}{2125} + \frac{4660}{10000} \cdot \frac{510}{4660}}$$

$$= \frac{\frac{3215}{10000} \cdot \frac{2110}{3215}}{\frac{3215}{10000} \cdot \frac{2110}{3215} + \frac{2125}{10000} \cdot \frac{396}{2125} + \frac{4660}{10000} \cdot \frac{510}{4660}}$$

$$= .700$$

$$P(d_2 | -+) = \frac{P(d_2)P(-+|d_2)}{P(d_1)P(-+|d_1) + P(d_2)P(-+|d_2) + P(d_3)P(-+|d_3)}$$

$$= \frac{\frac{2125}{10000} \cdot \frac{1187}{2125}}{\frac{3215}{10000} \cdot \frac{704}{3215} + \frac{2125}{10000} \cdot \frac{1187}{2125} + \frac{4660}{10000} \cdot \frac{73}{4660}}$$

$$= \frac{\frac{2125}{10000} \cdot \frac{1187}{2125}}{\frac{3215}{10000} \cdot \frac{704}{3215} + \frac{2125}{10000} \cdot \frac{1187}{2125} + \frac{4660}{10000} \cdot \frac{73}{4660}}$$

$$= .605$$

Example

A doctor gives a patient a test for a particular cancer. Before the test, the only evidence the doctor has to go on is that 1 woman in 1000 has the cancer. Experience has shown that in 99% of cases where ~~the~~ cancer is present, the test is positive, and in 95% of cases where cancer isn't present, the test is negative.

If test turns out positive, what probability should the doctor assign to event cancer is present?

Sol.

↓ prior

$$P(\text{cancer}) = .001,$$

↓ prior

$$P(\text{no cancer}) = .999$$

$$P(+ | \text{cancer}) = .99, \quad P(- | \text{cancer}) = .01$$

$$P(+ | \text{no cancer}) = .05, \quad P(- | \text{no cancer}) = .95$$

By Baye's formula,

$P(\text{cancer} | +)$ is equal to

$$\begin{aligned}
 & \frac{p(\text{cancer}) p(+|\text{cancer})}{p(\text{cancer}) p(+|\text{cancer}) + p(\text{no cancer}) p(+|\text{no cancer})} \\
 &= \frac{(.001)(.99)}{(.001)(.99) + (.999)(.05)} \\
 &= .019
 \end{aligned}$$

So the probability of cancer given a positive test has only increased from .001 to .019. While this is nearly a 20-fold increase, the probability the patient has cancer is still small.

When a group of 2nd year medical students was asked this question, over $\frac{1}{2}$ of the students incorrectly guessed the probability to be greater than .5.

Example

In a criminal investigation, the inspector in charge is 60% convinced of guilt of a certain suspect. There is a new piece of evidence: the criminal is bald. Suppose 20% of population is bald.

How certain of guilt should the inspector be if it turns out the suspect is bald?

Sol.

Let G = suspect is guilty
 C = suspect is bald

Then

$$P(G|C) = \frac{P(C|G)P(G)}{P(C|G)P(G) + P(C|G^c)P(G^c)}$$

$$= \frac{(1)(.6)}{(1)(.6) + (.2)(.4)}$$

We supposed that
 $P(C|G^c) = .2$,

$= .882$

that is if suspect is innocent chance of suspect being bald is the proportion of population that is bald.

Independence

Suppose that knowing that event

81

F occurs doesn't change probability that event E occurs, that is,

$$P(E|F) = P(E).$$

Since $P(E|F) = P(E \cap F) / P(F)$, we conclude that $P(E \cap F) = P(E)P(F)$.

We say that E, F are independent if

$$P(E \cap F) = P(E)P(F).$$

Example Toss a coin 3 times

Let E = 1st ~~the~~ toss is heads

F = 2nd toss is tails and 3rd toss is heads

$$\text{Then } P(E \cap F) = \frac{1}{2} \frac{1}{4} = P(E)P(F)$$

Intuitively, first toss doesn't affect 2nd, 3rd tosses.

Example Pick a random card from a 52

card deck. Let E = card is a 3, F = card is a heart

82

Then $P(E) = 1/3$, $P(F) = 1/4$

$$P(E \cap F) = P(\text{card is 3 of hearts}) = 1/52$$

Thus $P(E \cap F) = P(E)P(F)$, so E and F are independent.

Intuitively, knowing the suit of a card tells you nothing about its rank.

Example Roll two dice.

Let E = get 1 on first die

F = even number on second die

$$P(E) = 1/6, \quad P(F) = 1/2$$

$P(E \cap F) = 3/36 = 1/12 = P(E)P(F)$, so E, F are independent.

Intuitively, outcome of first die has no effect on outcome of second die.

Example

In a random permutation of 4 letters a, b, c, d , the events "a precedes b" and

83

"c precedes d" are independent. This is intuitively clear and easily verified.

Example

Consider families with 3 children. Assume each of 8 possibilities

bbb, bbg, ..., ggg has probability $1/8$

Let H = family has children of both genders

A = family has at most 1 girl

Then

$$P(H) = 6/8, \quad P(A) = 4/8$$

$$P(A \cap H) = P(\text{one girl, 2 boys}) = 3/8$$

So $P(A \cap H) = P(A)P(H)$, and events A, H are independent.

Note = independence holds for families with

3 children, but not in families with 2 or 4 children. So it's not always obvious when have independence.

Prop. If A, H are independent, then so are A, H^c .

Proof

Since A is disjoint union of $A \cap H$ and $A \cap H^c$,

$$P(A \cap H^c) = P(A) - P(A \cap H)$$

$$\begin{aligned} & \xrightarrow{\text{since } A, H \text{ are independent}} = P(A) - P(A)P(H) \\ & = P(A)(1 - P(H)) \\ & = P(A)P(H^c) \end{aligned}$$

Next, suppose 3 events A, B, C are pairwise independent:

$$P(A \cap B) = P(A)P(B)$$

$$P(A \cap C) = P(A)P(C)$$

$$P(B \cap C) = P(B)P(C)$$

This does not imply that $A \cap B$ is independent of C .

Example Roll two dice.

Let $A = 1\text{st die odd}$

$B = 2\text{nd die odd}$

$C = \text{sum of dice odd}$

Can check that A, B, C are pairwise independent.

But they can't all happen, so

$$P(A \cap B \cap C) \neq P(A)P(B)P(C)$$

Remark: Say 3 events A, B, C are mutually independent if

$$P(A \cap B) = P(A)P(B)$$

$$P(A \cap C) = P(A)P(C)$$

$$P(B \cap C) = P(B)P(C)$$

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

In general, call $A_1, \dots, A_n$ mutually independent if for all $1 \leq i < j < k < \dots \leq n$

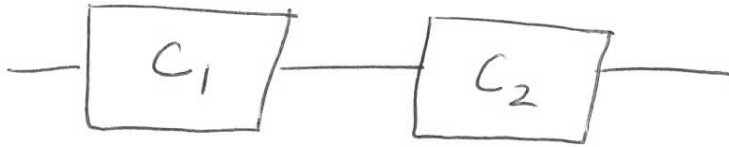
$$P(A_i \cap A_j) = P(A_i)P(A_j) \leftarrow \binom{n}{2} \text{ equations}$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \leftarrow \binom{n}{3} \text{ equations}$$

$\vdots$

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2)\dots P(A_n)$$

Example : reliability of components
in series



Let W_i = component C_i works without
Failure for one day $i=1, 2$

System works if and only if both components
work

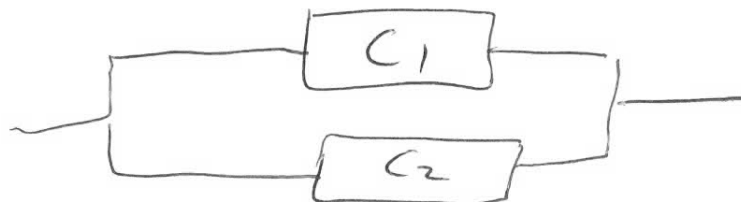
Suppose $P(W_1) = .9$, $P(W_2) = .8$

If W_1, W_2 are independent,

$$\begin{aligned} P(\text{system works}) &= P(W_1 \cap W_2) \\ &= P(W_1)P(W_2) \\ &= (.9)(.8) = .72 \end{aligned}$$

Example : reliability of components
in parallel

Let W_1, W_2 be as in last example



The system works if and only if at least 1 component works. Thus,

$$\begin{aligned}
 P(\text{system works}) &= P(W_1 \cup W_2) \\
 &= P(W_1) + P(W_2) - P(W_1 \cap W_2) \\
 \text{by independence} \rightarrow &= P(W_1) + P(W_2) - P(W_1)P(W_2) \\
 &= .9 + .8 - (.9)(.8) \\
 &= .98
 \end{aligned}$$

Example A contractor is planning a construction project to be completed in 3 stages. The contractor figures that

- (i) $P(\text{first stage on time} \mid \text{completed}) = .7$
- (ii) $P(\text{second stage completed} \mid \text{first stage on time}) = .8$
- (iii) $P(\text{third stage completed} \mid \text{first two stages on time}) = .9$

What is the chance all 3 stages will be completed on time? 88

Sol.

Let A_i = chance stage i is completed on time ($i=1, 2, 3$)

By multiplication rule

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \\ &= (.7)(.8)(.9) = .504 \end{aligned}$$

Remark The data determine the probability of some other events, such as event that first and second stages are completed on time, but the third is not. This is:

$$P(A_1 \cap A_2 \cap A_3^c) = (.7)(.8)(.1) = .056$$

But data do not determine the probability that second stage is late. To calculate this, it would be necessary to know $P(\text{2nd stage late} | \text{1st stage late})$. Then $P(\text{2nd stage late})$ could be obtained by

rule of average conditional probabilities. ⁸⁹

Example : geometric distribution

A die has proportion p of its faces painted white, and proportion q of its faces painted black, where $q = 1 - p$. The die is rolled until first time a white face shows up.

(a) What is chance this takes 3 or fewer rolls?

Sol.

Let W_i = i th roll is white

B_i = i th roll is black

Then by independence of the rolls,

$P(\text{white in 3 or fewer rolls})$

$$= P(W_1) + P(B_1 \cap W_2) + P(B_1 \cap B_2 \cap W_3)$$

$$= p + qp + q^2p$$

$$= (1 + q + q^2)p$$

(b) What is chance it takes 4 or more rolls to get a white face?

MATH 407 Lecture 7

90

Example (continued)

A die has proportion p of its faces painted white, and proportion q painted black, where $q = 1 - p$. Die is rolled until first time a white face shows up.

(b) What is chance it takes 4 or more rolls to get a white face?

Sol.

$$P(\text{4 or more rolls to get white}) = 1 - P(\text{white in 3 or fewer rolls})$$

$$\begin{array}{l} \text{by} \\ \text{part a} \end{array} \quad \rightarrow = 1 - (1 + q + q^2)p$$

Remark If set $p = 1 - q$ in this formula

and simplify, you get q^3 . To see this directly, note that it takes 4 or more rolls to get a white if and only if the first

3 rolls are black. This gives the simplest 91
solution to the questions above:

$$P(4 \text{ or more rolls to get white}) = q^3$$

$$P(3 \text{ or fewer rolls to get a white}) = 1 - q^3$$

The distribution for number of rolls required to get a white face is:

number of rolls	1	2	3	...	k	...
-----------------	---	---	---	-----	---	-----

probability	p	qp	q ² p	...	q ^{k-1} p	...
-------------	---	----	------------------	-----	--------------------	-----

This is the geometric distribution with parameter p, and we'll study it later.

Prop. Suppose $P(F) > 0$. If $E_i, i \geq 1$ are mutually exclusive, then

$$P\left(\bigcup_{i \geq 1} E_i \mid F\right) = \sum_{i \geq 1} P(E_i \mid F)$$

Proof

$$P\left(\bigcup_{i \geq 1} E_i \mid F\right) = \frac{P\left(\left(\bigcup_{i \geq 1} E_i\right) \cap F\right)}{P(F)}$$

$$= P(U(E_i \cap F)) / P(F)$$

since $\rightarrow$ the $E_i \cap F$ are mutually exclusive

$$= \sum_i P(E_i \cap F) / P(F)$$

$$= \sum_i P(E_i | F)$$

Chapter 4 : Random variables

Bernoulli(p) trial : on each trial have success probability p and failure probability $q = 1 - p$.

Example

<u>Nature of trial</u>	<u>Meaning of success</u>	<u>Meaning of failure</u>	<u>p, q</u>
toss a fair coin	head	tail	$1/2, 1/2$
rolling a die	6	not 6	$1/6, 5/6$
rolling a pair of dice	double 6	not double six	$1/36, 35/36$
birth of child	girl	boy	$.487, .513$

Binomial distribution

For n independent trials, with probability p of success and $q=1-p$ of failure on each trial, the chance of exactly k successes is given by the binomial probability formula

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}$$

Proof The number of patterns $SFFSSSF$ with k S's and $n-k$ F's is $\binom{n}{k}$. Letting S mean success and F mean failure, each such pattern occurs with probability $p^k q^{n-k}$. Since different patterns are mutually exclusive, the result follows.

Special case = fair coin tossing

$$\text{Let } p=q=1/2, \text{ so } p^k q^{n-k} = 1/2^n$$

$$\text{Then } P(k \text{ heads in } n \text{ fair coin tosses}) = \binom{n}{k} / 2^n, 0 \leq k \leq n$$

Example

Find $P(4 \text{ or more heads in } 6 \text{ tosses})$
of a fair coin

Sol. We seek $P(4) + P(5) + P(6)$

where $P(k) = P(k \text{ heads in } 6 \text{ tosses}) = \binom{6}{k} / 2^6$

So answer is $(15 + 6 + 1) / 2^6 = 11/32$

Example

What is probability that among 5 families, each with 6 children, at least 3 of families have 4 or more girls.

Sol. Assume each child in each family is equally likely to be a boy or a girl, independent of all other children.

By previous example, chance any particular family has 4 or more girls is $11/32$. Call this a success in the current problem. Then the sought probability is the chance of at least

3 successes in $n=5$ trials, with 95
success probability $p=11/32$ on each trial.

So the required probability is

$$\begin{aligned} & \binom{5}{3} \left(\frac{11}{32}\right)^3 \left(\frac{21}{32}\right)^2 + \binom{5}{4} \left(\frac{11}{32}\right)^4 \left(\frac{21}{32}\right) + \binom{5}{5} \left(\frac{11}{32}\right)^5 \\ & \quad \uparrow \qquad \qquad \uparrow \qquad \qquad \uparrow \\ & \quad P(3 \text{ successes}) \quad P(4 \text{ successes}) \quad P(5 \text{ successes}) \\ & = .226 \end{aligned}$$

Remarks

- 1) most likely number of success = "mode" of binomial distribution

Fact

For $0 < p < 1$, the most likely number of successes in n independent trials with probability of success p on each trial is

$m =$ greatest integer less than
or equal to $np + p$

If $np + p$ is an integer, as in case $p=1/2$, n odd

then there are two most likely numbers, m and $m-1$. Otherwise, there is a unique most likely number.

In either case, the probabilities in the binomial (n, p) distribution are strictly increasing before they reach the maximum, and strictly decreasing after the maximum.

2) expected number of successes = "mean" of binomial distribution

Fact The mean μ of a binomial (n, p) distribution is $\mu = np$.

Later we will carefully define the mean, and prove this fact.

Normal curve with mean μ , standard deviation σ

is given by $y = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$



Note The area under the normal curve = 1

Standard normal has $\mu = 0, \sigma = 1$

The standard normal has "cumulative distribution function"

$\Phi(z)$ = area to left of z under the
 $\uparrow$
 definition standard normal curve

$$= \int_{-\infty}^z \phi(y) dy, \text{ where } \phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

For normal distribution with mean μ , standard deviation σ , the probability of being between a and b is

$$\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Indeed, if X is normal with mean μ and standard deviation σ , then

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

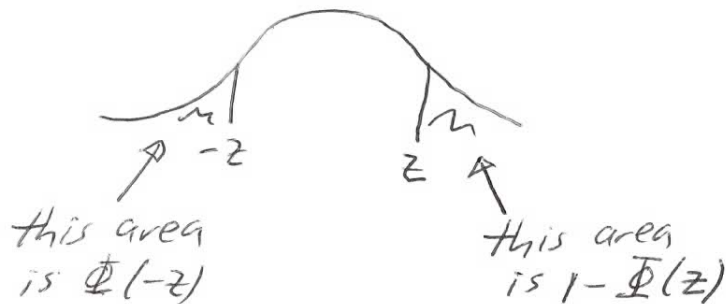
where $z = \frac{X-M}{\sigma}$ is a standard normal. ⁹⁸

There is no simple formula for $\Phi(z)$, but there are tables of values of $\Phi(z)$.

Note

- By symmetry of the normal curve,

$$\Phi(-z) = 1 - \Phi(z) \text{ for } z \geq 0$$



So in particular, $\Phi(0) = 1/2$, and only need a table of values of $\Phi(z)$ for $z > 0$.

- The probability of the interval (a, b) is denoted $\Phi(a, b)$ and is equal to $\Phi(b) - \Phi(a)$.

In particular, for $z \geq 0$,

$$\begin{aligned} \Phi(-z, z) &= \Phi(z) - \Phi(-z) \\ &= \Phi(z) - (1 - \Phi(z)) \\ &= 2\Phi(z) - 1 \end{aligned}$$

worth remembering

99

$\Phi(-1, 1) \approx .68 =$ probability of being within one standard deviation of mean

$\Phi(-2, 2) \approx .95 =$ probability of being within two standard deviations of mean

$\Phi(-3, 3) \approx .997 =$ probability of being within three standard deviations of mean

From these you can easily find $\Phi(a, b)$ for some other intervals. For example,

$$\Phi(0, 1) = \frac{1}{2} \Phi(-1, 1) \approx \frac{1}{2} (.68) = .34$$

$$\Phi(2, \infty) = \frac{1}{2} (1 - \Phi(-2, 2)) \approx \frac{1}{2} (1 - .95) = .025$$

Normal approximation to binomial distribution

For n independent trials with success probability p ,

$$P(a \text{ to } b \text{ successes}) \approx \Phi\left(\frac{b + 1/2 - np}{\sigma}\right) - \Phi\left(\frac{a - 1/2 - np}{\sigma}\right)$$

where $\mu = np$ is mean

100

$\sigma = \sqrt{npq}$ is standard deviation
of binomial (n, p) distribution.

Remark : use of $b+1/2, a-1/2$ is called the
continuity correction. It's essential to use the
correction to obtain good approximations for
small values of $\sqrt{npq}$. For large $\sqrt{npq}$ it
makes little difference unless a, b are very close.

Example Find, approximately, chance of exactly
50 heads in 100 tosses of a fair coin.

Sol.

Let $n=100, p=1/2$, so $\mu = 50, \sigma = 5$

The normal approximation with $a=b=50$
gives

$$\begin{aligned} P(\text{exactly } 50 \text{ heads}) &\approx \Phi\left(\frac{50+1/2-50}{5}\right) - \Phi\left(\frac{50-1/2-50}{5}\right) \\ &= \Phi(0.1) - \Phi(-0.1) \end{aligned}$$

$$= .0796$$

101

Note: exact value is .0795892

Similarly,

$$P(45 \text{ to } 55 \text{ heads}) \approx \Phi\left(\frac{55.5-50}{5}\right) - \Phi\left(\frac{44.5-50}{5}\right) \\ = .7286 \quad (\text{exact value is } .728747)$$

Similarly,

$$P(40 \text{ to } 60 \text{ heads}) \approx .9642 \quad (\text{exact value is } .9648)$$

$$P(35 \text{ to } 65 \text{ heads}) \approx .9980 \quad (\text{exact value is } .99821)$$

Poisson approximation to binomial distribution

If n is large and p is small, the distribution of the number of successes in n independent trials is largely determined by the value of the mean $\mu = np$, according to the Poisson approximation

$$P(k \text{ successes}) \approx \frac{1}{e^\mu} \frac{\mu^k}{k!}$$

Remarks

102

1) It can be shown that the accuracy of the approximation depends largely on the value of p , and hardly at all on n . Roughly speaking, absolute errors in using this approximation will be of the same order of magnitude as p .

2) Sketch of proof of Poisson approximation to binomial

Let $\mu = np$, n large, p small.

Suppose μ is fixed.

Then

$$P(\text{binomial}(n, p) = 0) = (1-p)^n$$

$$= (1 - \mu/n)^n$$

$$\xrightarrow{\text{as } n \rightarrow \infty} e^{-\mu} \text{ as } n \rightarrow \infty$$

$$\text{Since } \log(1 - \mu/n)^n = n \log(1 - \mu/n)$$

$$= n(-\mu/n - (\mu/n)^2/2 - (\mu/n)^3/3 - \dots)$$

$$\rightarrow -\mu \text{ as } n \rightarrow \infty$$

observe that

$$\frac{P(\text{binomial}(n, p) = k)}{P(\text{binomial}(n, p) = k-1)} = \frac{\mu - (k-1)p}{k(1-p)} \\ \approx \mu/k \text{ for small } p$$

Thus $P(\text{binomial}(n, p) = 1) \approx \mu e^{-\mu}$

$$P(\text{binomial}(n, p) = 2) \approx \mu^2 e^{-\mu} / 2$$

and continuing in this manner,

$$P(\text{binomial}(n, p) = k) \approx \mu^k e^{-\mu} / k!$$

Example

Suppose that over long run a manufacturing process produces 1% defective items. What is the chance of getting 2 or more defective items in a sample of 200 items produced by the process?

Sol.

Consider a binomial distribution with parameters $n=200$, $p=1/100$. Approximate it by a Poisson

distribution with $\mu = np = 2$

104

$$P(\text{2 or more defectives}) = 1 - P(0 \text{ defective}) - P(1 \text{ defective})$$

$$\approx 1 - e^{-2} 2^0/0! - e^{-2} 2^1/1!$$

$$= 1 - 3e^{-2} \approx .594$$

Def. The Poisson distribution with parameter μ is the distribution of probabilities

$P_\mu(k)$ on $\{0, 1, 2, \dots\}$ defined by

$$P_\mu(k) = e^{-\mu} \mu^k / k! \quad k = 0, 1, 2, \dots$$

Note

$$\sum_{k \geq 0} P_\mu(k) = e^{-\mu} (1 + \mu + \mu^2/2! + \mu^3/3! + \dots)$$
$$= e^{-\mu} e^{\mu} = 1$$

Poisson approximation is a widely studied topic. Here, at USC, some researchers in this topic are Arratia, Goldstein, Waterman, me.

MATH 407 Lecture 8

105

Sampling with and without replacement

Suppose a population of size N has G good and B bad elements, where $N = G + B$. For a sample of size $n = g + b$, with $0 \leq g \leq n$, the chance of getting g good elements and b bad elements is

- (for sampling with replacement)

$$P(g \text{ good}, b \text{ bad}) = \binom{n}{g} \frac{G^g B^b}{N^n}$$

- (for sampling without replacement)

$$P(g \text{ good}, b \text{ bad}) = \binom{G}{g} \binom{B}{b} / \binom{N}{n}$$

Proof

- (with replacement case)

This follows from binomial distribution with success chance $= G/N$, failure chance $= B/N$

$$\binom{n}{g} \left(\frac{G}{N}\right)^g \left(\frac{B}{N}\right)^b = \binom{n}{g} G^g B^b / N^n$$

- (without replacement case)

Have $\binom{N}{n}$ many n element subsets of the population, each equally likely. Of these, the number consisting of g good elements and b bad elements is $\binom{G}{g}\binom{B}{b}$.

Remark

The hypergeometric distribution is name of distribution of number of good elements in a sample of size n without replacement from a population of G good and $N-G$ bad elements. The distribution has 3 parameters: n, N, G . The probability this distribution assigns to g is $\binom{G}{g}\binom{B}{b}/\binom{N}{n}$ where $B = N - G$.

- It's true, but not obvious from the formula that

$$\sum_{g \geq 0} \binom{G}{g}\binom{B}{b}/\binom{N}{n} = 1.$$

- If N, G, B are large compared to n, g, b , then there is little difference in sampling with or without replacement. So in this case can approximate the hypergeometric distribution by binomial distribution.

The number of heads in 4 tosses of a coin can be 0, 1, 2, 3, 4. A random variable will denote something like number of heads, which might be one of several possible values, with a distribution of probabilities over the set of values.

Typically capital letters X, Y, Z , etc. are used to denote random variables.

eg.

X = number obtained by rolling a die

Y = number of heads in 4 coin tosses

Z = suit of a card from a well shuffled deck

The range of a random variable is the set of possible values that X might produce:

	<u>description</u>	<u>range</u>
X	number on a die	$\{1, 2, 3, 4, 5, 6\}$
Y	number of heads in 4 coin tosses	$\{0, 1, 2, 3, 4\}$
Z	suit of a card	$\{C, D, H, S\}$

Example

Let X = number of heads in two tosses of a fair coin

The distribution of X is

x	0	1	2
$P(X=x)$	$1/4$	$1/2$	$1/4$

Joint distributions

Say (X, Y) has value (x, y) if $X=x$ and $Y=y$.

The range of (X, Y) is the set of all ordered pairs (x, y) with $x \in \text{Range}(X)$, $y \in \text{Range}(Y)$ and

$P(X=x, Y=y) > 0$. The distribution of (X, Y) is called the joint distribution of X and Y . 109

Example

Let X, Y be the first and second draws made at random without replacement from a box containing three tickets 1, 2, 3. Let $P(x, y) = P(X=x, Y=y)$.

X values

		1	2	3
Y values →	3	$1/6$	$1/6$	0
	2	$1/6$	0	$1/6$
	1	0	$1/6$	$1/6$

Marginal probabilities

$$P(X=x) = \sum_y P(x, y), \quad P(Y=y) = \sum_x P(x, y)$$

In example

	1	2	3
distribution of X	$1/3$	$1/3$	$1/3$
	1	2	3
distribution of Y	$1/3$	$1/3$	$1/3$

110

Remark In this example, X and Y have the same distribution, but aren't equal. Indeed, X, Y are called equal if $P(X=Y)=1$. In this example, $P(X=Y)=0$.

Probabilities of events determined by X and Y

The probability that X, Y satisfy some condition is the sum of $P(x, y)$ over (x, y) satisfying that condition. For example,

$$P(X < Y) = \sum_{\substack{(x, y) \\ x < y}} P(x, y) = \sum_x \sum_{y > x} P(x, y)$$

$$P(X = Y) = \sum_{\substack{(x, y) \\ x = y}} P(x, y) = \sum_x P(x, x)$$

The distribution of any function of (X, Y) such as $X+Y$, $X-Y$, XY , $\min(X, Y)$, $\max(X, Y)$ can be obtained from the joint distribution of X and Y . For example,

$$P(X+Y=z) = \sum_{\substack{(x, y) \\ x+y=z}} P(x, y)$$

111

Example Calculate distribution of $X+Y$ for two random draws X, Y from box containing $\{1, 2, 3\}$

(a) without replacement

(b) with replacement

Sol.

(a) From earlier table, the possible values of $S = X+Y$ are 3, 4, 5.

By inspection of the table, each possible value s of S corresponds to exactly two possible pairs, each with probability $1/6$.

Hence the distribution of S is given by

S	3	4	5
$P(S=s)$	$1/3$	$1/3$	$1/3$

(b) If draws are made with replacement, the joint probabilities are $P(x, y) = 1/9$ for $1 \leq x, y \leq 3$.

112

Now there is one possible pair adding to 2

two	- - - - -	3
three	- - - - -	4
two	- - - - -	5
one	- - - - -	6

Thus the distribution of S is given by

s	2	3	4	5	6
$P(S=s)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

Remark: Example shows that to find distribution of a function of two random variables X, Y , such as their sum, must think in terms of the joint distribution of X and Y . Knowing that X, Y are both uniform on $\{1, 2, 3\}$ isn't enough to determine the distribution of $X+Y$.

Example let X be minimum, Y be maximum of 3 digits picked at random without replacement from $\{0, 1, \dots, 9\}$.

(a) Find joint distribution of X and Y

Sol.

Since sampling is done without replacement, the 3 digits must be distinct, so the only possible values of x, y for minimum and maximum are

$$0 \leq x \leq 7, x+2 \leq y, y \leq 9$$

To find chance of such a pair, forget the order in which the digits come, and think about what subset of 3 digits is chosen from $\{0, 1, \dots, 9\}$. Every subset has the same chance of being chosen:

$$1/\binom{10}{3} = 1/120$$

To illustrate, for a minimum of 4 and a maximum of 7, there are just two possible subsets $\{4, 5, 7\}$ and $\{4, 6, 7\}$. In this case there are $7-4-1=2$ ways to pick the intermediate number.

$$\text{So } P(X=4, Y=7) = \frac{7-4-1}{120}$$

In general, for a minimum of x and maximum of y , there are $y-x-1$ possible subsets, one for each possible value of the third number between x and y . Thus for possible pairs (x, y) as above,

$$P(X=x, Y=y) = (y-x-1)/120.$$

(b) Find distribution of $Z = Y - X$, the maximum minus the minimum.

Sol.

The possible values of Z are $2, 3, \dots, 9$.

Any one of these possible values z say, must come from a (min, max) pair (x, y) with $y-x=z$. Every such pair has the same probability

$$\frac{y-x-1}{120} = \frac{z-1}{120}$$

For $z=9$, there is one pair $(0, 9)$

$z=8$, there are two pairs $(0, 8), (1, 9)$

etc.

In general, there are $10-z$ possible pairs (x,y) with $y-x=z$.

Thus

$$P(Z=z) = (10-z)(z-1)/120 \quad z=2, 3, \dots, 9$$

As a check, the sum of these probabilities from $z=2$ to 9 is

$$\frac{(8 \times 1) + (7 \times 2) + (6 \times 3) + (5 \times 4) + (4 \times 5) + (3 \times 6) + (2 \times 7) + (1 \times 8)}{120} = 1$$

Multinomial distribution

Let N_i denote the number of results in category i in a sequence of n independent trials with probability p_i for a result in i th category on each trial, $1 \leq i \leq m$ with $p_1 + \dots + p_m = 1$.

Then for every m -tuple of non-negative integers $(n_1, \dots, n_m)$ with sum n ,

$$P(N_1 = n_1, \dots, N_m = n_m) = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}$$

116

Remark : The derivation is similar to that of binomial distribution, which was the special case $m=2$.

Example

Suppose a fair die is rolled 10 times. What's the chance of one 4, two 5's and three 6's ?

Sol.

Use multinomial distribution with $n=10$ trials.
 $m=4$ categories : "four", "five", "six", "other"

with probabilities $\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6}$

So answer is $\frac{10!}{1!2!3!4!} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^3 \left(\frac{3}{6}\right)^4$

Expectation

Recall that the mean (or average) of a list of numbers $x_1, \dots, x_n$ is $(x_1 + \dots + x_n)/n$

more generally

117

- The mean μ of a probability distribution P over a set of numerical values x is average of values x weighted by their probabilities (assuming set is finite)

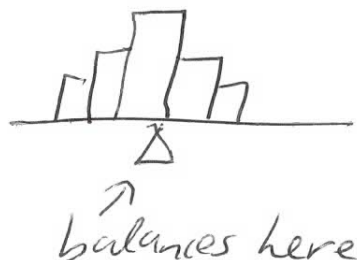
$$\mu = \sum_x x P(x)$$

- The expectation (also called the expected value or mean) of a random variable X is the mean of the distribution of X , denoted $E(X)$. That is,

$$E(X) = \sum_x x P(X=x)$$

the average of all possible values of X , weighted by their probabilities.

Note: mean of a distribution is center of gravity, point where the histogram of the distribution balances



Fact (to be proved later) : the binomial (n, p) distribution has mean $\mu = np$.

Example Suppose n tickets numbered $x_1, \dots, x_n$ are put in a box and a ticket is drawn at random. Then letting $X =$ value of ticket drawn,

$$E(X) = \frac{x_1}{n} + \dots + \frac{x_n}{n} = \text{average of numbers in the box}$$

Example Suppose X takes two possible values, say a and b , with probabilities $P(a)$ and $P(b)$. Then $E(X) = aP(a) + bP(b)$. This average is a number between a and b . The larger $P(a)$, the closer $E(X)$ is to a , and the larger $P(b)$, the closer $E(X)$ is to b .

Example

Suppose $X = I_A$ is indicator of event A

Thus $I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{else} \end{cases}$

$$\begin{aligned}\text{Then } E(I_A) &= 0 P(I_A=0) + 1 P(I_A=1) \\ &= P(A)\end{aligned}$$

Example (rolling a die)

Let X be number produced by rolling a fair die. Then

$$\begin{aligned}E(X) &= 1 P(X=1) + 2 P(X=2) + \dots + 6 P(X=6) \\ &= (1)(\frac{1}{6}) + (2)(\frac{1}{6}) + \dots + (6)(\frac{1}{6}) \\ &= 3.5\end{aligned}$$

Remark Should not expect a single roll to be 3.5. But if roll a die a large number of times, should expect average to be close to 3.5. Why?

$$\text{sum of rolls} = 1 \left(\text{number of } 1\text{'s} \right) + \dots + 6 \left(\text{number of } 6\text{'s} \right)$$

$$\text{So average of rolls} = 1 \left(\text{proportion of } 1\text{'s} \right) + \dots + 6 \left(\text{proportion of } 6\text{'s} \right)$$

close to $1/6$ by "law of large numbers"
if have a large number of rolls

So average should be close to $E(X) = 3.5$

MATH 407 Lecture 9

120

Remark So from example at end of last class, if the probability values of a random variable X are approximate long-run frequencies, then $E(X)$ is approximately the long-run average value of X .

Remark The mean is one way to locate a central point in the distribution of X . There are other ways, for example the mode and the median

A mode is a most likely value of X
(there may be more than one)

A median is a number m so that $P(X \leq m)$ and $P(X \geq m)$ are at least $1/2$ (there may be more than one median).

For example, if X is roll of a fair die, then 1, 2, 3, 4, 5, 6 are all modes, and every number between 3 and 4 is a median of X .

121

The mean, median, and mode may be quite different. But if the distribution is symmetric about some point m , and has a single mode, the 3 quantities all equal m .

The mean is the most important in theory. This is due to the close connection between means and long-run averages, and fact (proved later) that the mean of a sum of random variables is sum of means. There is no such simple rule for modes or medians

Def. Random variables X, Y are independent if $P(X=x, Y=y) = P(X=x)P(Y=y)$ for all x, y .

If X, Y are independent random variables, then every event determined by X is independent of every event determined by Y :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Indeed,

$$P(X \in A, Y \in B) = \sum_{\substack{x \in A \\ y \in B}} P(X=x, Y=y)$$

$$= \sum_{\substack{x \in A \\ y \in B}} P(X=x) P(Y=y)$$

122

$$= P(X \in A) P(Y \in B)$$

Example Roll two dice. Let

X = outcome of die 1

Y = outcome of die 2

Then X, Y are independent

Remark

For possible value x of X , as y ranges over the range of Y , the probabilities $P(Y=y | X=x)$ define a probability distribution over the range of Y . This distribution, which may depend on $x \in X$, is called the conditional distribution of Y given $X=x$.

X, Y are independent $\iff$ the conditional distribution of Y given $X=x$ does not depend on x .

Addition rule for expectation

For any two random variables X, Y

$$E(X+Y) = E(X) + E(Y)$$

even if X, Y aren't independent.

Thus for a sequence $X_1, \dots, X_n$ of random variables, however dependent,

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$$

Example Let T_n be the sum of rolls of n dice. Find $E(T_n)$

Sol. Let $X_1, \dots, X_n$ be numbers obtained from n die rolls. Then $T_n = X_1 + \dots + X_n$.

$$\text{Thus } E(T_n) = E(X_1) + \dots + E(X_n)$$

$$= 3.5 + \dots + 3.5$$

by
earlier
example

$$= 3.5(n)$$

Remark The distribution of T_n becomes increasingly difficult to calculate as n increases

124

but the formula for $E(T_n)$ is simple.

As a check, $E(T_2)$ can be found from the distribution of T_2 .

$$\begin{aligned} E(T_2) &= 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} \\ &\quad + 7 \times \frac{6}{36} + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} \\ &= 7 \end{aligned}$$

method of indicators

Recall that if I_A is the indicator of event A , then $E(I_A) = P(A)$

Example Suppose a system has n components and that at a particular time, the j th component is working with probability p_j , $1 \leq j \leq n$. Let X = number of components working at that time. Find $E(X)$

Sol.

$$X = I_1 + \dots + I_n \text{ where } I_j = \begin{cases} 1 & \text{if } j\text{th component is working} \\ 0 & \text{else} \end{cases}$$

$$\begin{aligned}\text{Thus } E(X) &= E(I_1) + \dots + E(I_n) \\ &= p_1 + p_2 + \dots + p_n\end{aligned}$$

Remark: We didn't need to assume independence between components.

Example

Suppose X is number of successes in n independent trials with success probability p on each trial, so that X has binomial (n, p) distribution.

Show that $E(X) = np$

Sol.

$$X = I_1 + \dots + I_n \quad \text{where} \quad I_j = \begin{cases} 1 & \text{if success on trial } j \\ 0 & \text{else} \end{cases}$$

$$\begin{aligned}\text{Then } E(X) &= E(I_1) + \dots + E(I_n) \\ &= p + \dots + p = np\end{aligned}$$

Remark: This is not so obvious from definition of $E(X)$:

$$E(X) = \sum_x x P(X=x) = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x}$$

Example

Let X = number of aces in a 5-card poker hand. Find $E(X)$.

Sol. Label the cards 1, 2, 3, 4, 5.

Then $X = I_1 + \dots + I_5$ where $I_j = \begin{cases} 1 & \text{if } j\text{th card is ace} \\ 0 & \text{else} \end{cases}$

Thus $E(X) = E(I_1) + \dots + E(I_5)$

Now $E(I_j) = P(j\text{th card is ace}) = \frac{4}{52}$

Thus $E(X) = 5 \frac{4}{52} = 5/13$

Remark : Computing $E(X)$ directly from the definition gives

$$\begin{aligned} E(X) &= \sum_{x=0}^4 x P(X=x) \\ &= \sum_{x=0}^4 x \frac{\binom{4}{x} \binom{48}{5-x}}{\binom{52}{5}} \end{aligned}$$

This does give the same answer, but other method is quicker.

Tail sum formula for expectation

127

Theorem For X with possible values $\{0, 1, \dots, n\}$,

$$E(X) = \sum_{j=1}^n P(X \geq j)$$

Proof 1

Observe that $X = X_1 + \dots + X_n$ where

$$X_j = \begin{cases} 1 & \text{if } X \geq j \\ 0 & \text{else} \end{cases}$$

Indeed, if $X = i$, then $X \geq 1, \dots, i$ and $X < i+1, \dots, n$.

$$\begin{aligned} \text{Thus } E(X) &= \sum_{j=1}^n E(X_j) \\ &= \sum_{j=1}^n P(X \geq j) \end{aligned}$$

Proof 2

Define $p_j = P(X=j)$. Then

$E(X) = 1p_1 + 2p_2 + 3p_3 + \dots + np_n$ is the following sum:

$$\begin{aligned} & p_1 \\ & + p_1 + p_2 \\ & + p_1 + p_2 + p_3 \\ & \vdots \\ & + p_1 + p_2 + p_3 + \dots + p_n \end{aligned}$$

The sum of first column is $P(X \geq 1)$. 128

The sum of second column is $P(X \geq 2)$

etc.

So the sum of all columns is $\sum_{j=1}^n P(X \geq j)$.

Example

Roll 4 dice. Let M = minimum of the four numbers rolled

Find $E(M)$

Sol. For any j , $1 \leq j \leq 6$, the event $M \geq j$ means that each die roll X_i is at least j .

Thus

$$\begin{aligned} P(M \geq j) &= P(X_1 \geq j, X_2 \geq j, X_3 \geq j, X_4 \geq j) \\ &= \left(\frac{6-j+1}{6} \right)^4 \end{aligned}$$

by independence of the X_i 's and the fact that there are $6-j+1$ possible values for each X between j and 6. The tail sum formula gives

$$E(M) = P(M \geq 1) + P(M \geq 2) + \dots + P(M \geq 6)$$

129

$$= \left(\frac{6}{6}\right)^4 + \left(\frac{5}{6}\right)^4 + \left(\frac{4}{6}\right)^4 + \left(\frac{3}{6}\right)^4 + \left(\frac{2}{6}\right)^4 + \left(\frac{1}{6}\right)^4$$

$$= 1.755$$

Remark: The point of using the tail sum formula in this example is that the tail probabilities $P(M \geq j)$ are simpler than the individual probabilities

$$P(M = m) = P(M \geq m) - P(M \geq m+1)$$

If you substitute this in the definition

$$E(M) = \sum_m m P(M = m)$$

and simplify, you will find the coefficient of $P(M \geq j)$ is 1 for each j from 1 to n . That is the substance of the tail sum formula.

Example Let S be the largest 3 numbers among 4 dice. Find $E(S)$.

Sol. Note that $S = T - M$, where T is sum

of all 4 numbers and M is minimum number.

$$\text{Thus } E(S) = E(T) - E(M)$$

$$\text{From earlier examples, } E(T) = 4(3.5) = 14$$

$$E(M) = 1.755$$

$$\text{Thus } E(S) = 14 - 1.755 = 12.245$$

Remark : It is much harder to find $E(S)$ via the distribution of S .

Markov's inequality

Theorem If $X \geq 0$, then $P(X \geq a) \leq \frac{E(X)}{a}$

for all $a > 0$.

Proof (Discrete case)

$$\begin{aligned} E(X) &= \sum_{j \geq 0} j P(X=j) \\ &\geq \sum_{j \geq a} j P(X=j) \\ &\geq a \sum_{j \geq a} P(X=j) \end{aligned}$$

$$= a P(X \geq a)$$

131

Thus $P(X \geq a) \leq E(X)/a$.

Expectation of a function of X

Typically, $E[g(X)] \neq g[E(X)]$. Instead,

$$E[g(X)] = \sum_x g(x) P(X=x)$$

This formula is valid for any numerical function g defined on set of possible values of X .

In particular, for $g(x) = x^k$ with $k=1, 2, \dots$ the number

$$E(X^k) = \sum_x x^k P(X=x)$$

is called the k th moment of X .

Proof of

$$\begin{aligned} \sum_x g(x) P(X=x) &= \sum_y \sum_{x: g(x)=y} g(x) P(X=x) \\ &= \sum_y y \sum_{x: g(x)=y} P(X=x) \end{aligned}$$

$$= \sum_y y P(g(X)=y)$$

$$= E[g(X)]$$

Remark : If c is constant, $E(cX) = cE(X)$,
as can see by applying the formula with $g(x) = cx$.

Example Let X be uniform on $\{-1, 0, 1\}$

$$\text{Then } E(X) = \frac{1}{3}(-1) + \frac{1}{3}(0) + \frac{1}{3}(1) = 0$$

$$\text{So } (E(X))^2 = 0.$$

By the formula for $E[g(X)]$ with $g(x) = x^2$,

$$E(X^2) = \frac{1}{3}(-1)^2 + \frac{1}{3}(0)^2 + \frac{1}{3}(1)^2 = 2/3$$

$$\text{So } E(X^2) \neq (E(X))^2$$

Expectations of functions of two or more
random variables

$$E[g(X, Y)] = \sum_{x, y} g(x, y) P(X=x, Y=y)$$

Proof of addition rule ($E(X+Y) = E(X) + E(Y)$)

$$E(X) = \sum_{x,y} x P(x,y)$$

133

$$E(Y) = \sum_{x,y} y P(x,y)$$

$$E(X+Y) = \sum_{x,y} (x+y) P(x,y), \text{ so result follows}$$

Multiplication rule for expectation

If X, Y are independent, then $E(XY) = E(X)E(Y)$

Proof

$$E(XY) = \sum_{x,y} xy P(X=x, Y=y)$$

by independence $\rightarrow$

$$= \sum_{x,y} xy P(X=x) P(Y=y)$$

$$= \left[\sum_x x P(X=x) \right] \left[\sum_y y P(Y=y) \right]$$

$$= E(X) E(Y)$$

Standard deviation and normal approximation

Def. $\text{Var}(X)$ = mean squared deviation
 of X from its
 expected value $\mu = E(X)$

$\nearrow$
 "variance of X "

$$= E[(X-\mu)^2]$$

The standard deviation of X , denoted $SD(X)$ is the square root of the variance of X :

$$SD(X) = \sqrt{\text{Var}(X)}$$

Remark

Intuitively, $SD(X)$ is a measure of how spread out the distribution of X is around its mean μ . Variance is harder to interpret than SD , but has simpler algebraic properties.

Computational formula for variance

$$\begin{aligned}\text{Var}(X) &= E(X^2) - (E(X))^2 \\ &= \sum_x x^2 P(X=x) - \left(\sum_x x P(X=x) \right)^2\end{aligned}$$

Remark From its definition, $\text{Var}(X) \geq 0$, and $\text{Var}(X) = 0$ if and only if $P(X=\mu) = 1$. So computational formula shows that $E(X^2) \geq (E(X))^2$ with equality if and only if X is a constant.

MATH 407 Lecture 10

Proof that $\text{Var}(X) = E(X^2) - (E(X))^2$

$$\text{Var}(X) = E[(X - \mu)^2] \quad \text{where } \mu = E(X)$$

$$\begin{aligned} \text{addition} \rightarrow &= E[X^2] - 2\mu E(X) + E[\mu^2] \\ \text{rule} & \\ \text{for expectations} & \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - (E(X))^2 \end{aligned}$$

Remarks

- 1) If distribution of X is approximately normal, then chance of being within one standard deviation of the mean is roughly .68.
- 2) If distribution of X is not roughly normal, there is no simple way to visualize $\text{SD}(X)$. But no matter what the distribution of X , one should expect X to be around

136

$E(X)$, plus or minus a few times
 $SD(X)$. This is made more precise later
by Chebyshev's inequality.

Like the mean $E(X)$, the standard deviation
 $SD(X)$ can be interpreted in terms of a sum
 $S_n = X_1 + \dots + X_n$ of a large number of
independent random variables X_i with the
same distribution as X . What happens is that
for large n , the distribution of S_n is
approximately normal with parameters determined
by $E(X)$, $SD(X)$, n . This is made precise later
by the central limit theorem.

Example: random sampling

Suppose n tickets numbered $X_1, \dots, X_n$ are
put in a box, and a ticket is drawn at
random. Let X = number on the ticket.

$$\text{Then } E(X) = \frac{X_1 + \dots + X_n}{n} = \bar{X}$$

↗ average of
numbers in
the box

$$\text{and } \text{Var}(X) = \frac{1}{n} \sum_i (X_i - \bar{X})^2$$

↗
by
definition

$$= \frac{1}{n} \sum_i X_i^2 - \bar{X}^2$$

↗
from
computational
formula for variance

Example : indicators

Suppose X is indicator of an event with probability p . Find $SD(X)$.

Sol. Since $0^2=0$ and $1^2=1$, we have that $X^2=X$.

$$\text{Thus } E(X^2) = E(X) = 1(p) + 0(1-p) = p$$

$$\begin{aligned} \text{Thus } \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

$$\text{So } SD(X) = \sqrt{\text{Var}(X)} = \sqrt{p(1-p)}$$

Remark: Since X has binomial $(1, p)$ distribution, this agrees with the formula that $\sqrt{np(1-p)}$ is the SD of the binomial (n, p) distribution.

Example: Let X be the number on a fair die. Find $SD(X)$.

Sol.
$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{1^2 + 2^2 + \dots + 6^2}{6} - (3.5)^2 \\ &= 35/12 \end{aligned}$$

Thus $SD(X) = \sqrt{35/12} \approx 1.71$

Scaling and shifting

For constants a and b , $SD(aX + b) = |a| SD(X)$

Shifting by a constant doesn't change the spread of the distribution, but multiplying by a or $-a$ spreads out the distribution by a factor of $|a|$.

139

You can check this from the definition of SD, using properties of expectation.

Compare this with the corresponding formula for expectation:

$$E(aX + b) = a E(X) + b.$$

Example (Celsius to Fahrenheit)

Let X represent a temperature in Celsius and let Y be the same temperature in Fahrenheit.

$$\text{So } Y = \frac{9}{5} X + 32.$$

How are $E(Y)$, $SD(Y)$ related to $E(X)$, $SD(X)$?

Sol. $E(Y) = \frac{9}{5} E(X) + 32$

$$SD(Y) = \frac{9}{5} SD(X)$$

Note that SD is only affected by the scale factor $\frac{9}{5}$, and not by the shift of 32.

Example (successes and failures)

Let X = number of successes in
 n trials of some kind

Y = number of failures in
same sequence of trials

Assuming every trial results in either success
or failure, how are $E(Y)$, $SD(Y)$ related to
 $E(X)$, $SD(X)$?

Sol. $Y = n - X$

Then $E(Y) = n - E(X)$

$$SD(Y) = |-1| SD(X) = SD(X).$$

Standardization

When making normal approximation, it's
convenient to transform a random variable X
into a standardized variable X^* , which gives
the number of SD's by which X differs
from its expected value.

more precisely,

141

If a random variable X has
 $E(X) = \mu$, $SD(X) = \sigma > 0$, then the random
variable

$$X^* = (X - \mu) / \sigma,$$

called X in standard units, has $E(X^*) = 0$,
 $SD(X^*) = 1$.

Remark

Any event determined by value of X can
be written in terms of X^* . Usually this is
done by manipulating inequalities. For
example for any real number b ,

$$P(X \leq b) = P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = P(X^* \leq \frac{b - \mu}{\sigma}).$$

If X is approximately normal, then X^* is
approximately a standard normal. Then the
above probability can be approximated by

$\Phi[(b - \mu) / \sigma]$, where Φ is the standard normal
cdf. For a binomial random variable X ,

142

this is the normal approximation discussed earlier, except that we are ignoring the correction from b to $b + 1/2$, called the "continuity correction" which is appropriate only if the range of possible values of X is a sequence of consecutive integers.

Example

A person is picked at random from a population of individuals with heights distributed approximately to a normal curve. If the population mean height is 5 feet 10 inches, and the SD of height is 2 inches, approximate the chance that the person is over 6 feet tall.

Sol. Let X = height of individual

Then $E(X) = 70$ inches, $SD(X) = 2$ inches

Converting to standard units gives

143

$$P(X > 6 \text{ feet}) = P(X > 72 \text{ inches})$$

$$= P\left(\frac{X-70}{2} > \frac{72-70}{2}\right)$$

$$= P\left(\frac{X-70}{2} > 1\right)$$

$$= P(X^* > 1)$$

$$= 1 - \Phi(1) = .16$$

Example In a recent study, the average height of men age 18 and over was 69 inches and SD was 3 inches. Use normal approximation to estimate percentage of these men with heights between 63 and 72 inches.

Sol. $P(63 < X < 72)$

$$= P\left(\frac{63-69}{3} < \frac{X-69}{3} < \frac{72-69}{3}\right)$$

$$= P(-2 < X^* < 1)$$

$$= \Phi(1) - \Phi(-2) \approx .82$$

Tail probabilities

144

Chebyshev's inequality

For any random variable X , and $k > 0$,

$$P(|X - E(X)| \geq k \text{SD}(X)) \leq 1/k^2$$

In words: the chance that a random variable differs from its expected value by k standard deviations or more is at most $1/k^2$.

Proof

$$\text{Let } \mu = E(X), \sigma = \text{SD}(X)$$

$$\text{Then } |X - \mu| \geq k\sigma \iff (X - \mu)^2 \geq k^2\sigma^2$$

$$\text{Let } Y = (X - \mu)^2, a = k^2\sigma^2$$

Then

$$P(|X - \mu| \geq k\sigma) = P(Y \geq a)$$

by Markov's
inequality using
 $Y \geq 0$

$$\begin{aligned} &\leq E(Y) / a \\ &= \sigma^2 / k^2 \sigma^2 \\ &= 1/k^2 \end{aligned}$$

Comparison of Chebyshev bound with normal probabilities

145

	<u>cheb.</u>	<u>standard normal value</u>
$P(X-\mu \geq 6)$	1	.3173
$P(X-\mu \geq 26)$	$1/4 = .25$	.0465
$P(X-\mu \geq 36)$	$1/9 \approx .11$	.00270
$P(X-\mu \geq 46)$	$1/16 \approx .06$	.000063

Thus Chebyshev bound is very crude for a distribution which is approximately normal. Its importance is that it holds whatever the shape of the distribution.

Example

The average of a list of a million numbers is 10, and the average of the squares of the numbers is 101. Find an upper bound on how

146

many of the entries on the list are 14 or more.

Sol. Let X = random number from the list

$$\text{Then } \mu = E(X) = 10, E(X^2) = 101$$

$$\text{Thus } \sigma = SD(X) = \sqrt{101 - (10)^2} = 1$$

$$\text{So } P(X \geq 14) = P(X - \mu \geq 4)$$

$$\leq P(|X - \mu| \geq 4)$$

$$\text{Chebyshev's inequality} \rightarrow \leq 1/4^2 = 1/16$$

Thus the number of entries 14 or more is at most

$$10^6 P(X \geq 14) \leq 10^6 / 16 = 62,500.$$

Remark If the distribution of the list were known to be symmetric about 10, the probabilities $P(X \geq 14)$ and $P(X \leq 6)$ would be equal. Since

it is the sum of these two probabilities 147
which is at most $1/16$, the bound in this
case could be reduced by a factor of 2
to $31,250$.

Sums and averages of independent random variables

Addition rule for variances

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \quad \text{if } X, Y \text{ independent}$$

More generally,

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

if $X_1, \dots, X_n$ are independent (meaning that

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n)$$

for all $x_1, \dots, x_n$).

Remark : Variances do not always add for
dependent random variables. For example if

$$X=Y, \text{ then } \text{Var}(X+Y) = \text{Var}(2X) = 4 \text{Var}(X)$$

$$\text{but } \text{Var}(X) + \text{Var}(Y) = 2 \text{Var}(X)$$

Proof (of addition rule for variances)

Let $S = X + Y$. Then $E(S) = E(X) + E(Y)$

Thus $S - E(S) = [X - E(X)] + [Y - E(Y)]$

Hence

$$\begin{aligned} (S - E(S))^2 &= (X - E(X))^2 + (Y - E(Y))^2 \\ &\quad + 2(X - E(X))(Y - E(Y)) \end{aligned}$$

Taking expectations gives

$$\begin{aligned} \text{Var}(S) &= \text{Var}(X) + \text{Var}(Y) \\ &\quad + 2E((X - E(X))(Y - E(Y))) \end{aligned}$$

If X, Y are independent, then so are $X - E(X)$ and $Y - E(Y)$. Thus

$$\begin{aligned} E((X - E(X))(Y - E(Y))) &= E(X - E(X)) E(Y - E(Y)) \\ &= (0)(0) = 0 \quad \text{as needed.} \end{aligned}$$

To get addition rule for n variables, apply

the addition rule for 2 variables 149
repeatedly.

Sum of independent random variables with
same distribution

Let $X_1, \dots, X_n$ be independent with same
distribution as X . Then for $S_n = X_1 + \dots + X_n$,

$$E(S_n) = n E(X) \quad \text{addition rule for expectations}$$

$$SD(S_n) = \sqrt{\text{Var}(S_n)}$$

$$= \sqrt{n \text{Var}(X)} = \sqrt{n} SD(X)$$

addition rule for variances

Square-root law

Let S_n be the sum, $\bar{X}_n = S_n/n$ be the
average, of n independent random variables
 $X_1, \dots, X_n$ each with same distribution as X .

Then

$$E(S_n) = n E(X), \quad SD(S_n) = \sqrt{n} SD(X)$$

$$E(\bar{X}_n) = \frac{1}{n} n E(X) = E(X), \quad SD(\bar{X}_n) = \frac{SD(X)}{\sqrt{n}}$$

Example derive formula $\sqrt{np(1-p)}$ for SD of binomial (n, p) distribution

Sol. This is distribution of the sum

$$S_n = X_1 + \dots + X_n$$

of n indicators of independent events, each with probability p of occurring.

So $\sqrt{np(1-p)}$ follows from the square root law for $SD(S_n)$ and the formula $\sqrt{p(1-p)}$ for the SD of an indicator, found earlier.

Law of averages

Let $X_1, X_2, \dots$ be a sequence of independent random variables each with the same distribution as X . Let $\mu = E(X)$ be the common expected value of the X_i 's, and let

$$\bar{X}_n = (X_1 + \dots + X_n) / n$$

be the average of $X_1, \dots, X_n$.

Then, for every $\epsilon > 0$, no matter how small,

$$P(|\bar{X}_n - \mu| < \epsilon) \longrightarrow 1 \text{ as } n \rightarrow \infty$$

In words: as the number of variables increases, with probability approaching 1, the average will be arbitrarily close to the expected value.

Proof

By square root law, $E(\bar{X}_n) = \mu$, $SD(\bar{X}_n) = \sigma/\sqrt{n}$ where $\sigma = SD(X)$. Applying Chebyshev's inequality to $\bar{X}_n$ gives

$$\begin{aligned} P(|\bar{X}_n - \mu| \geq \epsilon) &= P(|\bar{X}_n - \mu| \geq \frac{\epsilon}{SD(\bar{X}_n)} SD(\bar{X}_n)) \\ &\leq \left(SD(\bar{X}_n) / \epsilon \right)^2 \\ &= \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

152

For fixed ϵ , the right hand side tends to 0 as $n \rightarrow \infty$, and so does the left hand side since probabilities are non-negative. Taking complements yields the result.

Normal approximation (central limit theory)

Let $S_n = X_1 + \dots + X_n$ be the sum of n independent random variables, each with the same distribution over some finite set of values. For n large, the distribution of S_n is approximately normal, with mean $E(S_n) = n\mu$ and standard deviation $SD(S_n) = \sigma\sqrt{n}$, where $\mu = E(X_i)$, $\sigma = SD(X_i) \neq 0$. That is to say, for all $a \leq b$,

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

where Φ is the standard normal cumulative distribution function, namely $\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-x^2/2} dx$

No matter what the distribution of the X_i , for every $a \leq b$, the error in using this normal approximation tends to 0 as $n \rightarrow \infty$. The same result holds for X_i with an infinite range of possible values, provided that the standard deviation σ is non-0 and finite.

Remarks

- 1) The random variable $(S_n - n\mu)/\sigma\sqrt{n}$ appearing in the central limit theorem is S_n in standard units. It has mean 0 and variance 1.
- 2) If the possible values of X_i form a sequence of consecutive integers, the continuity correction, discussed earlier, should be used to get a better approximation.
- 3) Normal approximation works just as well for averages as for sums, since the factor of n has no effect on the standardized variable

4) For any distribution of X_i with just two possible values, the above normal approximation follows from the normal approximation to the binomial distribution, used earlier, by using scaling properties of the mean and standard deviation to reduce to the case when the possible values are 0,1. But a full proof of the central limit theorem is beyond scope of book.

Example

(Random walk)

physicists use random walk to model the process of diffusion, or random motion of particles. The position S_n of a particle at time n can be thought of as a sum of displacements $X_1, \dots, X_n$. Assuming the displacements are independent and identically distributed, we can use the above ideas

Problem Suppose at each step a particle on sites labelled by the integers is equally likely to move 1 step to right, 1 step to left, or to stay where it is



Approximate the probability that after 10,000 steps, the particle ends up more than 100 sites to the right of its starting point.

Sol.

Let X represent a single step.

Then $E(X) = 0$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - 0^2 = \frac{(-1)^2}{3} + \frac{0^2}{3} + \frac{(1)^2}{3} \\ &= 2/3 \end{aligned}$$

$$\text{SD}(X) = \sqrt{2/3} \approx .8165$$

The problem is to find $P(S_{10,000} > 100)$

$$\text{Here } E(S_{10,000}) = 10,000 E(X) = 0$$

$$\begin{aligned} SD(S_{10,000}) &= \sqrt{10,000} \cdot SD(X) \\ &= (100)(0.8165) \approx 81.65 \end{aligned}$$

Normal approximation gives

$$\begin{aligned} P(S_{10,000} > 100) &= P\left(\frac{S_{10,000}}{81.65} > \frac{100}{81.65}\right) \\ &\approx 1 - \Phi\left(\frac{100}{81.65}\right) \\ &\approx .11 \end{aligned}$$

Discrete distributions

Until now, we assumed random variables only have a finite number of possible values. Probabilities and expected values were calculated as finite sums. Here we consider ~~discrete~~ discrete distributions on the set $\{0, 1, 2, \dots\}$ of non-negative integers: a sequence of probabilities $p_0, p_1, p_2, p_3, \dots$ such that $p_i \geq 0$ for all i , and $\sum_i p_i = 1$

Example

Consider the number of times T need to roll a fair die to get a 6

$$P(T=i) = q^{i-1} p \text{ where } p=1/6, q=1-p=5/6$$

This is the "geometric" distribution on $\{1, 2, 3, \dots\}$ with parameter $p=1/6$.

$$\text{Note: } P(T=\infty) = 0$$

$$\begin{aligned} \text{Indeed, } 0 \leq P(T=\infty) &\leq P(\text{first } n \text{ rolls aren't 6}) \\ &= (5/6)^n \\ &\rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

Other important examples to be discussed are the negative binomial and Poisson distributions. Later we'll study continuous distributions such as uniform distribution on $(0,1)$ and the normal distribution.

Example

Suppose you and I take turns rolling a die, to see who can first roll a 6. Suppose I roll first, then you roll, then I roll, and so on, until one of us has rolled a 6. What is the chance you roll first 6?

Sol.

In terms of $T = \text{number of rolls to get first 6}$, we seek $P(T \text{ is even})$. Clearly

$$P(T \text{ even}) = P(T=2) + P(T=4) + P(T=6) + \dots$$

$$= q \cdot p + q^3 p + q^5 p + \dots$$

$$\text{where } p = 1/6, q = 5/6$$

$$= q \cdot p (1 + q^2 + q^4 + \dots)$$

$$= q \cdot p / (1 - q^2) = \frac{5/6 \cdot 1/6}{(1 - 25/36)} = 5/11$$

What is the chance that I roll first 6?

Sol.

This is $P(T \text{ is odd})$. A similar calculation could be done again, but there is no need.

Since we argued that T is certain to be finite, and T must be either even or odd,

$$P(T \text{ odd}) = 1 - 5/11 = 6/11$$

Example

Suppose A, B play over and over, a game which results in a win for A , a win for B , or a draw (meaning no winner) with probabilities $P(A), P(B), P(D)$. Suppose they keep playing until the first game that does not result in a draw, and call the winner of that game the overall winner.

1) Show that

$$P(A \text{ wins overall}) = P(A) / (P(A) + P(B))$$

$$P(B \text{ wins overall}) = P(B) / (P(A) + P(B))$$

Sol. $P(A \text{ wins at game } n)$

$$= P(\text{first } n-1 \text{ games are draws and } A \text{ wins})$$

game n

$$= P(D)^{n-1} P(A)$$

$$\text{Thus } P(A \text{ wins}) = \sum_{n \geq 1} P(D)^{n-1} P(A)$$

$$= P(A) / (1 - P(D))$$

$$= P(A) / (P(A) + P(B))$$

$$\text{Similarly, } P(B \text{ wins}) = P(B) / (P(A) + P(B))$$

2) Let G = number of games playedand X = name of winner

Show that G has geometric distribution, and
that G, X are independent.

Sol. G is geometric with parameter $p = 1 - P(D)$
(the waiting time until first non-draw).

$P(G=n \text{ and } A \text{ is winner})$

$= P(n-1 \text{ draws, then } A \text{ wins})$

$= P(D)^{n-1} P(A)$

$= P(D)^{n-1} (1 - P(D)) P(A) / (1 - P(D))$

$= P(G=n) P(A \text{ wins})$

Thus $P(G=n | A \text{ wins}) = P(G=n)$

Similarly, $P(G=n | B \text{ wins}) = P(G=n)$

Thus G, X are independent.

Expectation

Def. The expectation of a discrete random variable X is

$$E(X) = \sum_x x P(X=x)$$

provided that the series is absolutely convergent,

that is provided that $\sum_x |x| P(X=x) < \infty$.

Remarks

- 1) X can have both positive and negative values
- 2) Absolute convergence guarantees that get the same value of $E(X)$, regardless of order in which you sum
- 3) If $g(X)$ is a function of a discrete random variable X , then

$$E[g(X)] = \sum_x g(x) P(X=x)$$

- 4) Chebyshev's inequality, law of averages, and normal approximation all hold for discrete random variables X with $E(X^2) < \infty$. In fact the law of averages holds for identically distributed independent random variables $X_1, X_2, \dots$, provided that $E(X_1)$ is defined. Proof is beyond scope of the course.

Example (moments of geometric distribution)

Let T = waiting time until first success
in a sequence of independent trials
each of which succeeds with
probability p or fails with probability
 $q = 1 - p$.

So T is geometric on $\{1, 2, \dots\}$ with parameter p .

Problem Find $E(T)$

Sol.

$$E(T) = \sum_{n=1}^{\infty} n P(T=n)$$

$$= \sum_{n=1}^{\infty} n q^{n-1} p = p \sum_{n=1}^{\infty} n q^{n-1}$$

$$\text{Now } \sum_n n q^{n-1} = 1 + 2q + 3q^2 + \dots$$

$$q \sum_n n q^{n-1} = q + 2q^2 + \dots$$

$$\Rightarrow (1-q) \sum_n n q^{n-1} = 1 + q + q^2 + \dots = 1/(1-q)$$

$$\Rightarrow \sum_n n q^{n-1} = 1/(1-q)^2$$

Thus $E(T) = p/(1-q)^2 = 1/p$

Problem Find $SD(T)$

Sol. $SD(T) = \sqrt{E(T^2) - (E(T))^2}$ where $E(T) = 1/p$
from above

Now $E(T^2) = \sum_{n=1}^{\infty} n^2 P(T=n) = p \sum_{n=1}^{\infty} n^2 q^{n-1}$

$$\sum_{n=1}^{\infty} n^2 q^{n-1} = 1 + 4q + 9q^2 + \dots + n^2 q^{n-1} + \dots$$

$$q \sum_{n=1}^{\infty} n^2 q^{n-1} = q + 4q^2 + \dots + (n-1)^2 q^{n-1} + \dots$$

$$\Rightarrow (1-q) \sum_{n=1}^{\infty} n^2 q^{n-1} = 1 + 3q + 5q^2 + \dots + (2n-1)q^{n-1} + \dots$$

$$= 2 \sum_{n=1}^{\infty} n q^{n-1} - (1 + q + q^2 + \dots)$$

by
evaluation
of $E(T)$

$$= 2/(1-q)^2 - 1/(1-q) = \frac{1+q}{(1-q)^2}$$

$$\text{Thus } \sum_{n=1}^{\infty} n^2 q^{n-1} = (1+q)/(1-q)^3$$

$$\Rightarrow E(T^2) = p(1+q)/(1-q)^3 = (1+q)/p^2$$

$$\text{Thus } SD(T) = \sqrt{\frac{(1+q)}{p^2} - \frac{1}{p^2}} = \sqrt{q}/p$$

The end.

MATH 407, lecture 12

Example : waiting time until r th success
(negative binomial distribution)

Let T_r = number of trials until r th success
when trials are independent with
success probability p

For instance, letting 1 = success, 0 = failure

000101001 ...

has $T_1 = 4, T_2 = 6, T_3 = 9$.

1) What is the distribution of T_r ?

Sol.

The possible values of T_r are $r, r+1, r+2, \dots$

For t in this range,

$$\begin{aligned}
 P(T_r = t) &= P\left(\begin{array}{l} r-1 \text{ successes in first } t-1 \text{ trials} \\ \text{and then success on trial } t \end{array}\right) \\
 &= \binom{t-1}{r-1} p^{r-1} (1-p)^{(t-1)-(r-1)} p
 \end{aligned}$$

$$= \binom{t-1}{r-1} p^r (1-p)^{t-r}$$

166

2) Find $E(T_r)$ and $SD(T_r)$.

Sol. Direct calculation from the formula for the distribution is tedious. For a quick solution, note that

$$T_r = w_1 + w_2 + \dots + w_r$$

where w_i = waiting time after $i-1$ st success to i th success.

It's intuitively clear, and not hard to check, that $w_1, w_2, w_3, \dots$ are independent, each with geometric (p) distribution.

So by results of last example, the addition rule for expectation, and the square root law,

$$E(T_r) = r/p, \quad SD(T_r) = \sqrt{rq}/p$$

Remarks

1) As $r \rightarrow \infty$, the distribution of T_r becomes asymptotically normal, another example of

the central limit theorem. But due to 167
skewness of the geometric distribution of the
terms being added, the approach to normality
is rather slow. Particularly for p near .5,
better approximations are obtained using relation

$$P(T_r \geq n) = P(S_n \leq r)$$

where S_n is number of successes in first n trials,
and the normal approximation to the binomial(n, p)
distribution of S_n .

2) The distribution of $T_r - r$, the number
of failures before r th success, is called
negative binomial with parameters r and p .
This is just the distribution of T_r shifted
from $\{r, r+1, r+2, \dots\}$ to $\{0, 1, 2, \dots\}$.

$$P(T_r - r = n) = P(T_r = n+r) = \binom{n+r-1}{r-1} p^r (1-p)^n$$

Example : Coupon collector's problem

Each box of a particular brand of cereal

contains one out of a set of n distinct plastic animals. Suppose that the animal in each box is equally likely to be any one of the set of n , independently of what animals are in other boxes.

Problem : What is the expected number of cereal boxes a collector must buy in order to obtain the complete set of animals?

Sol.

Collector gets one of the n animals in the first box.

Each subsequent box contains an animal that is different from this first one with probability $(n-1)/n$ and same with probability $1/n$. Using the independence assumption, the additional number of boxes required to get 2 different animals is geometric with parameter $p = (n-1)/n$ and mean $1/p = n/(n-1)$. So the number of boxes required

to get 2 different animals has
mean $1 + n/(n-1)$.

Once 2 different animals are obtained, each box contains a new animal with probability $\frac{n-2}{n}$ and one of the old ones with probability $2/n$.

So the additional time to get 3 different animals once 2 have been obtained is a geometric random variable with parameter $p = \frac{n-2}{n}$ and mean $1/p = n/(n-2)$. So the number of boxes required to get 3 different animals has mean

$$1 + n/(n-1) + n/(n-2).$$

Continuing in this way, the mean μ_n of the overall waiting time for the set of all n animals is the sum of n terms

$$\begin{aligned}\mu_n &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \dots + \frac{n}{2} + \frac{n}{1} \\ &= n(1 + 1/2 + \dots + 1/(n-1) + 1/n)\end{aligned}$$

Remarks

1) For example, with $n=6$ animals, the expected number of boxes required is

$$\mu_6 = 6 \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \right) \\ \approx 14.7$$

2) For large n , approximate values of μ_n can be obtained using Euler's approximation for the harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \log(n) + \gamma + \frac{1}{2n}$$

$\gamma = .57721 \dots$
is Euler's
constant

$$\text{So } \mu_n \approx n \log(n) + \gamma n + \frac{1}{2}$$

This approximation is good even for small n , as you can check on a calculator.

Poisson distribution

The Poisson distribution is an approximation to the number N of occurrences of events of

some kind, when the events all have 17/
small probabilities, and are independent or
nearly so. For example, N might be one of
the following counting random variables:

N_{wins} = number of wins in n games of roulette
for a gambler who bets on a single
number each game.

N_{drops} = number of raindrops that fall on a
particular square inch of roof during
a one-second interval of time

$N_{\text{particles}}$ = number of radioactive particles
emitted by a piece of radioactive
material during an interval of time

If there are n independent events with
probability p , the exact distribution of the
number N that occur is binomial (n, p) . As
shown earlier, if p is small this distribution is

172
closely approximated by the Poisson
distribution with parameter $\mu = E(N) = np$.

$$P(N=k) \approx \frac{1}{e^\mu} \frac{\mu^k}{k!}, \quad k=0, 1, 2, \dots$$

This justifies use of the Poisson distribution in each case above. For instance in the raindrops example, imagine a square inch divided into 100 hundredths of a square inch, each of which may or may not be hit by a raindrop. Suppose each hundredth of a square inch has the small small chance of being hit by a raindrop in the given second, independently of what happens elsewhere on the roof, and ignore the extremely small chance that some hundredth of square inch is hit more than once. Then N_{drops} is the number of successes in 100 independent trials, with small probability on each trial.

Can think of $N_{\text{particles}}$ in a similar way by dividing time into small units

Features of the Poisson(μ) distribution come from corresponding features of the binomial(n, p) distribution, by passage to the limit $n \rightarrow \infty, p \rightarrow 0$ with $\mu = np$ fixed.

It was shown earlier that in the limit the probabilities of individual values converge:

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{1}{e^\mu} \frac{\mu^k}{k!} \quad \text{as } \begin{matrix} n \rightarrow \infty \\ p \rightarrow 0 \\ np = \mu \end{matrix}$$

Since the binomial distribution has mean $np = \mu$, it is natural that the Poisson(μ) limit should also have mean μ . And the SD of the binomial(n, p) distribution is $\sqrt{np(1-p)}$, which tends to $\sqrt{\mu}$ as $n \rightarrow \infty, p \rightarrow 0$ with $np = \mu$.

Theorem If N has Poisson(μ) distribution,
then $E(N) = \mu, SD(N) = \sqrt{\mu}$

Proof $E(N) = \sum_{k=0}^{\infty} k P(N=k)$

$$= \sum_{k \geq 1} k e^{-\mu} \mu^k / k!$$

$$= e^{-\mu} \mu \sum_{k \geq 1} \frac{\mu^{k-1}}{(k-1)!}$$

$$\begin{aligned} & \xrightarrow{\text{set } j=k-1} e^{-\mu} \mu \sum_{j \geq 0} \frac{\mu^j}{j!} \\ & = e^{-\mu} \mu e^{\mu} = \mu \end{aligned}$$

Derivation of $SD(N)$

A direct attempt to find $E(N^2)$ would be to try to repeat last calculation with $k^2 P(N=k)$ instead of $k P(N=k)$. This gives terms of a constant times $\mu^k k^2 / k!$, which are not easy to sum. But $\mu^k k(k-1) / k!$ can easily be summed, and this solves the problem.

$$E(N(N-1)) = \sum_{k \geq 0} k(k-1) e^{-\mu} \mu^k / k!$$

$$= e^{-\mu} \sum_{k \geq 2} k(k-1) \mu^k / k!$$

$$= e^{-\mu} \mu^2 \sum_{k \geq 2} \mu^{k-2} / (k-2)!$$

$$\begin{aligned} & \nearrow = e^{-\mu} \mu^2 \sum_{j \geq 0} \mu^j / j! \\ \text{set } j &= k-2 \\ &= e^{-\mu} \mu^2 e^{\mu} = \mu^2 \end{aligned}$$

Thus,

$$\begin{aligned} E(N^2) &= E(N(N-1) + N) \\ &= E(N(N-1)) + E(N) \\ &= \mu^2 + \mu \end{aligned}$$

$$\begin{aligned} \text{So } \text{Var}(N) &= E(N^2) - (E(N))^2 \\ &= \mu^2 + \mu - \mu^2 = \mu \end{aligned}$$

$$\text{So } \text{SD}(N) = \sqrt{\mu}$$

Remarks : Let N_{μ} be Poisson with parameter μ .

1) Normal approximation

For μ large enough that the standard deviation $\sqrt{\mu}$ of the Poisson distribution is small in comparison to its mean μ , the

176

distribution starts to become normal in shape. The distribution of the standardized Poisson variable $(N_\mu - \mu) / \sqrt{\mu}$ approaches a standard normal as $\mu \rightarrow \infty$. This can be proved by direct formulas, and is yet another case of central limit theorem, since as we shall soon see, sums of independent Poisson variables are Poisson.

2) Law of large numbers

~~State~~ Since $E(N_\mu/\mu) = \mu/\mu = 1$ and

$$SD(N_\mu/\mu) = \sqrt{\mu}/\mu = 1/\sqrt{\mu} \rightarrow 0 \text{ as } \mu \rightarrow \infty$$

we get that $N_\mu/\mu \approx 1$ for large μ in the probabilistic sense that N_μ/μ will most likely be very close to 1. This is a special case of more general results for sums of independent random variables, due to the next result.

Theorem

If $N_1, \dots, N_j$ are independent Poisson random variables with parameters $\mu_1, \dots, \mu_j$ then $N_1 + \dots + N_j$ is a Poisson variable with parameter $\mu_1 + \dots + \mu_j$.

Proof 1

If N_1, N_2 are independent with binomial (n_1, p) and binomial (n_2, p) distributions, then $N_1 + N_2$ has binomial $(n_1 + n_2, p)$ distribution.

Let $n_1, n_2 \rightarrow \infty$ and $p \rightarrow 0$ with $n_1 p \rightarrow \mu_1$ and $n_2 p \rightarrow \mu_2$. Then $(n_1 + n_2)p \rightarrow \mu_1 + \mu_2$. So N_1, N_2 approach independent Poisson random variables with parameters μ_1, μ_2 , while $N_1 + N_2$ approaches a Poisson $(\mu_1 + \mu_2)$ distribution.

Repeated application of the result for two terms gives the result for any number of terms.

Proof 2

To simplify notation, let $\alpha = \mu_1$, $\beta = \mu_2$

Then

$$P(N_1 + N_2 = k) = \sum_j P(N_1 = j, N_2 = k-j)$$

$$= \sum_{j=0}^k P(N_1 = j) P(N_2 = k-j)$$

$$= \sum_{j=0}^k e^{-\alpha} \frac{\alpha^j}{j!} e^{-\beta} \frac{\beta^{k-j}}{(k-j)!}$$

$$= e^{-(\alpha+\beta)} \frac{(\alpha+\beta)^k}{k!} \left[\sum_{j=0}^k \frac{k!}{j!(k-j)!} \left(\frac{\alpha}{\alpha+\beta}\right)^j \left(\frac{\beta}{\alpha+\beta}\right)^{k-j} \right]$$

$$= e^{-(\alpha+\beta)} (\alpha+\beta)^k / k!$$

Since sum in box is the sum of terms in a binomial $(k, \frac{\alpha}{\alpha+\beta})$ distribution, hence 1

Example Suppose a gambler bets:

- 10 times on events of probability $1/10$
- 20 times on events of probability $1/20$
- 30 times on events of probability $1/30$
- 40 times on events of probability $1/40$

Assuming the events are independent, 179
what is the approximate distribution of the
number of times the gambler wins?

Sol.

Let N_1 = number of times win on first
10 events of probability $1/10$

N_2 = number of times win on 20 events
of probability $1/20$

N_3 = number of times win on 30 events
of probability $1/30$

N_4 = number of times win on 40 events
of probability $1/40$.

the exact distribution of the number of wins
is $N = N_1 + N_2 + N_3 + N_4$. The random variables
 N_i are independent, and each N_i is
binomial $(10i, 1/10i)$, hence approximately Poisson(1).

So the distribution of N is approximately

Poisson $(1+1+1+1) = \text{Poisson}(4)$

the end.

MATH 407, Lecture 13

180

Remark As previous example suggests, the Poisson approximation to binomial distribution extends to the case of independent trials with possibly different probabilities of success. It can be shown that if N is the number of events which occur among n independent events with probabilities $p_1, \dots, p_n$ then if all p_i are small, the distribution of N is approximately $\text{Poisson}(\mu)$, where

$$\mu = E(N) = p_1 + p_2 + \dots + p_n$$

Example Suppose the number of typographical errors on a single page of our textbook is $\text{Poisson}(1/2)$ calculate the probability of at least one error on this page.

Sol. Let $X = \text{number of errors on this page}$

Then $P(X \geq 1) = 1 - P(X = 0)$
 $= 1 - e^{-1/2} \approx .393$

Chapter 5

study continuous random variables

- normal, uniform on interval,
 exponential, etc.

Continuous distributions arise both

(i) as limits from discrete models

(eg. normal approximation to binomial
 or exponential approximation to geometric)

(ii) directly from physical phenomena modeled
 by continuous variables

(eg. normal distribution as model for height,
 or exponential distribution as model for
 lifetime of an atom)

Probability density

X has density $f(x)$ if for all $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



mean, variance, and standard deviation

These are defined as before in terms of expectations

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

If the second integral is finite, then so is the first, and then $E(X)$, $E(X^2)$ can be used to calculate $\text{Var}(X)$, $\text{SD}(X)$ in the usual way:

$$\text{Var}(X) = E(X^2) - (E(X))^2, \quad \text{SD}(X) = \sqrt{\text{Var}(X)}$$

Basic properties of expectation, variance, and standard deviation are the same as in discrete case. For example, Chebyshev's inequality holds just as well for X with a density as for a discrete random variable X . Proofs of such things parallel the discrete case, using properties of integrals instead of properties of sums.

Note : $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx,$

provided that the integral converges absolutely,
(that is $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$)

Independence

Two random variables X, Y are called independent if for all $A, B,$

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

Basic properties of independent random variables are same in continuous case as in discrete case.

For example, if X, Y are independent and $E(X), E(Y)$ are defined and finite, then $E(XY) = E(X)E(Y),$

The addition rule for variance of sum of independent random variables follows from this.

Uniform distribution

Say X is uniform on $[a, b]$ if X has density function $f(x)$ which is constant on $[a, b]$ and 0 elsewhere:

$$f(x) = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & \text{else} \end{cases}$$

Note that the constant value c on (a, b) is $1/(b-a)$, since $\int_a^b f(x) dx = 1$ 184

• If $a < x < y < b$, then

$$P(x < X < y) = \int_x^y \frac{1}{b-a} dx = \frac{y-x}{b-a}$$

For example, if X has uniform $[0, 2]$ distribution, then probability that $X = 1.23$ correct to 2 decimal places is

$$P(1.225 < X < 1.235) = \frac{1.235 - 1.225}{2} = .005$$

Note let $X = \text{uniform on } [a, b]$
 $U = \text{uniform on } [0, 1]$

$$\text{Then } X = a + (b-a)U$$

Indeed, if $a < x < y < b$,

$$\begin{aligned} P(x < a + (b-a)U < y) \\ &= P\left(\frac{x-a}{b-a} < U < \frac{y-a}{b-a}\right) \\ &= \frac{y-x}{b-a} \end{aligned}$$

which is same as probability that 185
a uniform on $[a, b]$ is between x and y .

This kind of linear change of variable
is a basic technique for reducing problems
to simplest case to avoid unnecessary calculation.

Example

$$\begin{aligned} E(X) &= E[a + (b-a)U] \\ &= a + (b-a)E(U) \quad \text{since} \\ &= a + (b-a)(1/2) \quad E(U) = \int_0^1 u f(u) du \\ &= (a+b)/2 \quad = \int_0^1 u du = 1/2 \end{aligned}$$

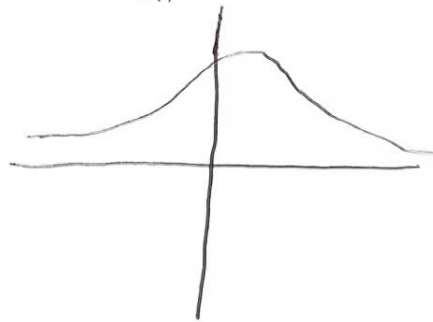
$$\begin{aligned} \text{Var}(X) &= \text{Var}[a + (b-a)U] \\ &= (b-a)^2 \text{Var}(U) \\ &= (b-a)^2 [E(U^2) - (E(U))^2] \\ &= (b-a)^2 [1/3 - (1/2)^2] \quad E(U^2) = \int_0^1 u^2 f(u) du \\ &= (b-a)^2/12 \quad = \int_0^1 u^2 du = 1/3 \end{aligned}$$

Normal distribution

A random variable Z has standard normal

distribution if its probability density 186
is the standard normal density

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (-\infty < z < \infty)$$



standard normal integrals:

$$\int_{-\infty}^{\infty} \phi(z) dz = 1$$

$$\int_{-\infty}^{\infty} z \phi(z) dz = 0$$

$$\int_{-\infty}^{\infty} z^2 \phi(z) dz = 1$$

Later we'll prove the first and third of these integrals. The second integral = 0 by symmetry of ϕ around 0. The second and third integrals show that standard normal has mean 0 and variance 1.

There is no simple formula for $\Phi(a,b) = P(a < Z < b)$
$$= \int_a^b \phi(z) dz.$$

Instead, use a table of standard normal cumulative distribution function:

$$\Phi(b) = \Phi(-\infty, b) = P(Z \leq b) = \int_{-\infty}^b \phi(z) dz$$

$$\Phi(a, b) = \Phi(b) - \Phi(a)$$

Note For $x > 0$, $\Phi(-x) = 1 - \Phi(x)$.

So table only gives $\Phi(x)$ for $x \geq 0$

Normal (μ, σ^2) distribution

Let Z be a standard normal, and μ, σ constants with $\sigma > 0$. Then $X = \mu + \sigma Z$ has mean μ , $SD = \sigma$, $Var = \sigma^2$. The distribution of X is called normal distribution with mean μ and variance σ^2 , abbreviated normal (μ, σ^2) .

So X has normal (μ, σ^2) distribution if and only if $Z = (X - \mu)/\sigma$ is a standard normal.

To find $P(c < X < d)$, change to standard units and use standard normal table

$$\begin{aligned} P(c < X < d) &= P\left(\frac{c - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{d - \mu}{\sigma}\right) \\ &= P(a < Z < b) = \Phi(b) - \Phi(a) \end{aligned}$$

$$\text{where } a = \frac{c - \mu}{\sigma}, \quad Z = \frac{X - \mu}{\sigma}, \quad b = \frac{d - \mu}{\sigma}$$

Formula for normal (μ, σ^2) density

188

For $\sigma > 0$, the formula is

$$\frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \quad -\infty < x < \infty$$

Remarks

- 1) This is the transformation of the standard normal density $\phi(z)$ corresponding to the linear change of variable from z to $x = \mu + \sigma z$. Later we'll see how changes of variables affect densities.
- 2) We rarely use this formula. It's usually simpler to transform to standard units as in next example.

Example : repeated measurements

Suppose a long series of repeated measurements of the weight of a standard kilogram yield results that are normally distributed with a mean of 1 kilogram and a SD of 20 micograms.

↑
one millionth of a gram

189

(a) About what proportion of measurements are accurate to within 10 micrograms?

Sol.

Let X = measurement in micrograms

μ = mean in micrograms

$$P(\mu - 10 \leq X \leq \mu + 10) = P(-10 \leq X - \mu \leq 10)$$

$$= P\left(\frac{-10}{20} \leq \frac{X - \mu}{20} \leq \frac{10}{20}\right)$$

$$\begin{aligned} &= P(-.5 \leq Z \leq .5) = \Phi(.5) - \Phi(-.5) \\ &= \Phi(.5) - (1 - \Phi(.5)) \\ &= .3829 \end{aligned}$$

standard normal table $\rightarrow$

(b) In the long series of measurements, some errors are positive and some are negative. What is the approximate average absolute size of these errors?

Sol.

Here X = observed weight - 1 kilogram (in micrograms)

has a normal $(0, 20^2)$ distribution

190

We want $E(|X|)$. In terms of a standard normal, Z , $X = 20Z$

$$\text{Thus } E(|X|) = 20E(|Z|) = 20 \int_{-\infty}^{\infty} |z| \phi(z) dz$$

$$\begin{array}{c} \text{by} \\ \text{symmetry} \end{array} \rightarrow = 40 \int_0^{\infty} z \phi(z) dz$$

$$= 40 \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= -\frac{40}{\sqrt{2\pi}} e^{-z^2/2} \Big|_0^{\infty}$$

$$= \frac{40}{\sqrt{2\pi}} = 15.96 \text{ micrograms}$$

Example : radial distance

Suppose a bacterial colony appears at a point uniformly distributed at random on a circular plate of radius 1. Let R be the distance of the point from the center of the plate

(a) Find the probability density of R

Proof (sketch)

Basic assumption is that probability of colony

appearing in any particular region of $|9|$
the plate is proportional to the area of
the region. Thus, for $0 < r < 1$



$$P(R \in (r, r+dr)) = \frac{\text{area of annulus from } r \text{ to } r+dr}{\text{total area}}$$

$$= \frac{\pi(r+dr)^2 - \pi(r^2)}{\pi}$$

$$\approx 2rdr$$

ignoring term with $(dr)^2$

So R has density $f(r) = \begin{cases} 2r & 0 < r < 1 \\ 0 & \text{else} \end{cases}$

(b) Find $P(a \leq R \leq b)$ for $0 < a < b < 1$.

Sol.

$$P(a \leq R \leq b) = \int_a^b 2r dr = r^2 \Big|_a^b = b^2 - a^2$$

(c) Find mean and variance of R

Sol. $E(R) = \int_{-\infty}^{\infty} r f(r) dr = \int_0^1 2r^2 dr = \frac{2}{3}$

$$E(R^2) = \int_{-\infty}^{\infty} r^2 f(r) dr = \int_0^1 2r^3 dr = \frac{1}{2}$$

$$\text{Thus } \text{Var}(R) = E(R^2) - (E(R))^2 \\ = 1/2 - 4/9 = 1/18$$

192

(d) Suppose 100 bacterial colonies are distributed independently and uniformly at random on a circular plate of radius 1. What is the probability that the mean distance of the colonies from center of plate is at least .7?

Sol.

Goal is to find $P(A_{100} \geq .7)$ where

$$A_{100} = (R_1 + R_2 + \dots + R_{100}) / 100$$

where the R_i are independent random variables with the same distribution as that of R in part (a).

Basic formulas for means and SD in previous chapter still apply, and

$$E(A_{100}) = E(R) = .667$$

$$SD(A_{100}) = \frac{SD(R)}{\sqrt{100}} = \sqrt{\frac{1}{8} \cdot \frac{1}{100}} \approx .0236$$

$$\text{Thus } P(A_{100} \geq .7) = P\left(\frac{A_{100} - .667}{.0236} \geq \frac{.7 - .667}{.0236}\right)$$

$$\begin{aligned}
 & \approx P\left(Z \geq \frac{.7 - .667}{.0236}\right) \quad \text{where } Z \text{ is a} \\
 & \quad \text{standard} \\
 & \quad \text{normal} \\
 & = 1 - \Phi\left(\frac{.7 - .667}{.0236}\right) \\
 & = 1 - \Phi(1.40) = .087.
 \end{aligned}$$

Example : A probability distribution with infinite mean

Suppose X has probability density

$$f(x) = \begin{cases} 1/(1+x)^2 & x > 0 \\ 0 & \text{else} \end{cases}$$

(a) Find $P(X > 3)$

Sol.

$$P(X > 3) = \int_3^{\infty} \frac{1}{(1+x)^2} dx = \left. -\frac{1}{1+x} \right|_3^{\infty} = 1/4$$

(b) Let X_1, X_2, X_3, X_4 be independent random variables with same distribution as X . Find the chance that exactly two of these variables are greater than 3.

194

Sol. The events $X_i > 3$, $i=1,2,3,4$ are independent, each with probability $1/4$. So the number of events which occur is binomial(4, $1/4$). Call this random variable N .

$$\text{Then } P(N=2) = \binom{4}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^2 = 27/128$$

(c) Find $E(X)$

$$\begin{aligned} \text{Sol. } E(X) &= \int_0^{\infty} \frac{x}{(1+x)^2} dx = \int_0^{\infty} \frac{1}{1+x} - \frac{1}{(1+x)^2} dx \\ &= \int_0^{\infty} \frac{1}{1+x} dx - 1 \\ &= \log(1+x) \Big|_0^{\infty} - 1 = \infty \end{aligned}$$

Remark The long run interpretation is that the average $(X_1 + \dots + X_n)/n$ of independent random variables chosen according to this distribution will, with overwhelming probability, tend to increase beyond all finite bounds as $n \rightarrow \infty$.

The end.

MATH 407 Lecture 14

195

Exponential distribution

One of the things most commonly described by a distribution with a density is a random time of some kind. Some examples are:

- (1) lifetime of an individual picked at random from some population
- (2) time until decay of a radioactive atom
- (3) length of time a patient survives after an operation of some kind
- (4) time it takes a computer to process a job of some kind

Such random times will be regarded as random variables with range the interval $[0, \infty)$.

The simplest model for a random time with no upper bound on its range is the exponential distribution.

Def. A random time T has exponential distribution

with rate λ , denoted exponential(λ), where λ is a positive parameter, if T has probability density $f(t) = \lambda e^{-\lambda t}$, $t \geq 0$.

Equivalently, for $0 \leq a < b < \infty$,

$$\begin{aligned} P(a < T \leq b) &= \int_a^b \lambda e^{-\lambda t} dt \\ &= -e^{-\lambda t} \Big|_a^b = e^{-\lambda a} - e^{-\lambda b} \end{aligned}$$

To see that $f(t)$ is a probability density on $[0, \infty)$, let $a=0$, $b \rightarrow \infty$ to find total probability of 1 on $[0, \infty)$.

Set $a=t$ and let $b \rightarrow \infty$ to get the next formula for the survival function. Calculation of mean and standard deviation is left as an exercise.

EXPONENTIAL SURVIVAL FUNCTION

A random time T is exponential(λ) if and only if T has survival function

$$P(T > t) = e^{-\lambda t}, \quad t \geq 0$$

$$E(T) = SD(T) = 1/\lambda$$

note that rate λ is inverse of mean, 197
so an exponential random time with large
rate is likely to be small, and one with
small rate is likely to be large.

memoryless property of the exponential distribution

A positive random variable T has exponential
distribution for some $\lambda > 0$ if and only if T
has the 'memoryless property'

$$P(T > t+s \mid T > t) = P(T > s) \quad s, t \geq 0.$$

In words: given survival to time t , the
chance of surviving a further time s is
the same as surviving time s in first place.

Indeed, suppose T is exponential(λ). Then

$$\begin{aligned} P(T > t+s \mid T > t) &= \frac{P(T > t+s \text{ and } T > t)}{P(T > t)} \\ &= P(T > t+s) / P(T > t) \\ &= e^{-\lambda(t+s)} / e^{-\lambda t} \end{aligned}$$

$$= e^{-\lambda s} = P(T > s) \quad 198$$

Conversely, if T has memoryless property, the survival function $G(t) = P(T > t)$ must be a solution of functional equation

$$G(t+s) = G(t)G(s) \quad t, s > 0.$$

with $G(t)$ decreasing and bounded between 0 and 1. It can be shown (see the book) that every such function $G(t)$ is of the form $e^{-\lambda t}$ for some λ .

Example Suppose the average lifetime of a particular kind of transistor is 100 working hours, and that the lifetime is approximately exponential. Estimate the probability that the transistor will work for at least 50 hours.

Sol.

Since the mean of exponential distribution is $1/\lambda$, get $\lambda = .01$. We calculate

$$P(T > 50) = e^{-\lambda(50)} = e^{-1/2} = .606.$$

Example

199

Suppose length of phone call in minutes is exponential (λ) with $\lambda = 1/10$. If someone arrives immediately ahead of you at a public phone booth, find the probability you will have to wait

(a) more than 10 minutes

(b) between 10 and 20 minutes

Sol.

$$(a) P(T > 10) = e^{-\lambda(10)} = e^{-1} \approx .368$$

$$(b) P(10 < T < 20) = P(T > 10) - P(T > 20) \\ = e^{-1} - e^{-2} \approx .233$$

Other continuous distributions (see book for details)

Gamma distribution

For parameters $\lambda, \alpha > 0$, the density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} / \Gamma(\alpha) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$\text{where } \Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} dy$$

when $\alpha = 1$, get exponential distribution. Can prove (see book) that for integers n , $\Gamma(n) = (n-1)!$

200

Beta distribution (with parameters a, b)

Has density

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$

where $B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$

• When $a=b=1$, get uniform distribution on $[0,1]$

Change of variables

Let X be a random variable with density $f_X(x)$ on range (a,b) . Let $Y=g(X)$ where g is either strictly increasing or strictly decreasing on (a,b) . For example, X might have exponential distribution on $(0, \infty)$ and Y might be X^2 , $\sqrt{X}$, or $1/X$. The range of Y is then an interval

with endpoints $g(a)$ and $g(b)$.

201

Theorem With these assumptions, the density of Y on this interval is

$$f_Y(y) = f_X(x) \left| \frac{dy}{dx} \right| \quad \text{where } y = g(x).$$

The equation $y = g(x)$ must be solved for x in terms of y , and this value of x substituted into $f_X(x)$ and dy/dx . This will leave an expression for $f_Y(y)$ entirely in terms of y .

Example

Let X be exponential with rate 1. That is, $f_X(x) = e^{-x}$. Find density of $Y = \sqrt{X}$

Sol.

1) Find range of Y . Here $0 < x < \infty$, $y = \sqrt{x}$, so $0 < y < \infty$.

2) Check function is one-to-one by solving for x in terms of y . Here $x = y^2$.

3) calculate dy/dx . Here

$$\frac{dy}{dx} = \frac{d}{dx}(\sqrt{x}) = \frac{1}{2\sqrt{x}}$$

4) Plug density of X and result of step

$$3 \text{ into } f_Y(y) = f_X(x) / |dy/dx|$$

$$\text{Let } f_Y(y) = e^{-x} / \frac{1}{2\sqrt{x}}$$

5) Use step 2 to eliminate x from right hand side. Let

$$f_Y(y) = e^{-y^2} / \frac{1}{2\sqrt{y^2}} = 2ye^{-y^2}, y > 0.$$

Example

Let $Y = aX + b$, where X is a standard normal and $a > 0$

Sol.

Step 1 range of $Y = (-\infty, \infty)$

Step 2 $X = \frac{Y-b}{a}$

Step 3 $\frac{dy}{dx} = \frac{d}{dx}(ax+b) = a$

$$\begin{aligned}\text{Step 4 } f_Y(y) &= f_X(x) / |dy/dx| \\ &= \frac{1}{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\end{aligned}$$

$$\begin{aligned}\text{Step 5 } \text{Eliminate } x \text{ from right hand side} \\ &= \frac{1}{a} \frac{1}{\sqrt{2\pi}} e^{-(y-b)^2/2a^2}\end{aligned}$$

Example

Let X have uniform $[0, 1]$ distribution.
Find distribution of $Y = -\frac{1}{\lambda} \log(X)$,
where $\lambda > 0$.

Sol.

$$\text{Step 1 } \text{image of } Y = (0, \infty)$$

$$\text{Step 2 } X = e^{-\lambda Y}$$

$$\text{Step 3 } \frac{dy}{dx} = -\frac{1}{\lambda x}$$

$$\text{Step 4 } f_Y(y) = f_X(x) / |dy/dx| = 1 / |-1/\lambda x| = \lambda x$$

$$\text{Step 5 } f_Y(y) = \lambda e^{-\lambda y}$$

Thus Y is exponential (λ).

204

Remark This way of obtaining an exponential as a function of a uniform $(0,1)$ random variable is standard method of simulating exponential variables by computer.

Cumulative distribution functions of continuous random variables

Def. Let X be a continuous, real-valued random variable. Then the cdf (cumulative distribution function) of X is defined by equation

$$F_X(x) = P(X \leq x)$$

Theorem Let X be a continuous real-valued random variable with density function $f(x)$. Then the function defined by

$$F(x) = \int_{-\infty}^x f(t) dt$$

is the cumulative distribution function of X . Furthermore,

$$\frac{d}{dx} F(x) = f(x).$$

Proof

By definition,

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]) = \int_{-\infty}^x f(t) dt$$

By the fundamental theorem of calculus,

$$\frac{d}{dx} F(x) = \frac{d}{dx} \int_{-\infty}^x f(t) dt = f(x)$$

Note In some cases the density function of relevant random variable is easy to write down. But often the cdf is easier to obtain than the density function. (Of course, once you have the cdf, can obtain the density function by differentiation).

Example

A real number is chosen uniformly at random in $[0, 1]$, and then this number is squared.

Let X be the result. What is the cdf of X ?

What is the density of X ?

206

Sol.

Let U represent the chosen real number.

So $X = U^2$.

If $0 \leq X \leq 1$, then we have

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(U^2 \leq x) \\ &= P(U \leq \sqrt{x}) \\ &= \sqrt{x} \end{aligned}$$

Since $0 \leq X \leq 1$, the cdf of X is given by

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \sqrt{x} & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

Taking derivatives shows that the density function of X is

$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ 1/(2\sqrt{x}) & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$

Note that $F_X(x)$ is continuous, but $f_X(x)$ is not.

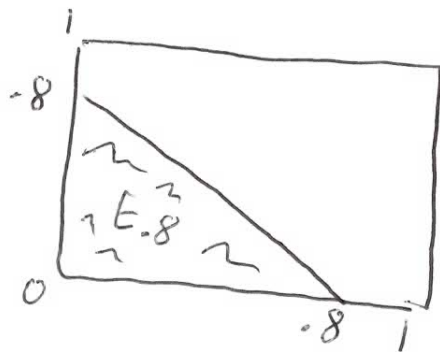
Example

Let X, Y be uniform $[0, 1]$ and independent. Let $Z = X + Y$. Derive the cdf and density of Z

Sol.

Let $E_z = \text{event that } Z \leq z$.

If $0 < z < 1$, this looks like the shaded set in the figure



If $1 < z \leq 2$, this looks like the square with a triangle removed from upper right corner.

$$F_Z(z) = P(Z \leq z) \\ = \text{Area of } E_z = \begin{cases} 0 & z < 0 \\ z^2/2 & 0 \leq z \leq 1 \\ 1 - \frac{1}{2}(2-z)^2 & 1 \leq z \leq 2 \\ 1 & z > 2 \end{cases}$$

Differentiating gives that

$$f_Z(z) = \begin{cases} 0 & z < 0 \\ z & 0 \leq z \leq 1 \\ 2-z & 1 \leq z \leq 2 \\ 0 & z > 2 \end{cases}$$

Example

Let sample space $S = \{(x, y) = x^2 + y^2 \leq 1\}$

Throw a dart at target and assume for

$$E \subseteq S, \quad P(E) = \frac{\text{area of } E}{\pi \times \text{area of target } S}$$

Let $Z = \sqrt{x^2 + y^2}$ = distance from the center of target

What is the cdf of Z ? What is density of Z ?

Sol.

Let E be event $Z \leq z$. Then

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = \frac{\text{area of } E}{\text{area of target}} \\ &= \begin{cases} 0 & z \leq 0 \\ z^2 & 0 \leq z \leq 1 \\ 1 & z > 1 \end{cases} \end{aligned}$$

Differentiating gives that

$$f_Z(z) = \begin{cases} 0 & z \leq 0 \\ 2z & 0 \leq z \leq 1 \\ 0 & z > 1 \end{cases}$$

Chapter 6 - jointly distributed random variables

Earlier we discussed joint distributions for discrete random variables.

Uniform distributions

Earlier discussed uniform distribution on an interval. The idea extends to higher dimensions.

For example, a random point (X, Y) in plane has uniform distribution on D , where D is a region of the plane with finite area, if

- (i) (X, Y) is certain to lie in D
- (ii) chance that (X, Y) falls in a subregion C of D is proportional to area of C :

$$P((X, Y) \in C) = \frac{\text{area of } C}{\text{area of } D} \quad \text{for } C \subset D.$$

Independent uniform variables

If X, Y are independent random variables, each uniformly distributed on an interval, then (X, Y) is uniformly distributed on a rectangle.

MATH 407, Lecture 15

Theorem If X, Y are independent random variables, each uniformly distributed on an interval, then (X, Y) is uniformly distributed on a rectangle.

Proof Suppose X, Y are independent and uniform on $[0, a], [0, b]$. For any rectangle $A \times B$,

$$P((X, Y) \in A \times B) = P(X \in A, Y \in B)$$

$$\begin{aligned} &\xrightarrow[\text{of } X, Y]{\text{by independence}} = P(X \in A) P(Y \in B) \\ &= \frac{\text{length}(A)}{a} \frac{\text{length}(B)}{b} \\ &= \frac{\text{area of } A \times B}{ab} \end{aligned}$$

So $P((X, Y) \in C)$ is the relative area of C in $[0, a] \times [0, b]$ for every rectangle C .

The same must be true for finite unions of rectangles, by addition rule for probability and area, hence for any set C whose area can

be defined by approximating with unions of rectangles. The result follows

211

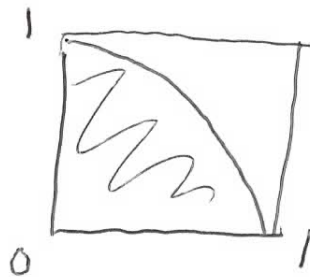
Example

Let X, Y be independent uniform $[0, 1]$ random variables.

(a) Find $P(X^2 + Y^2 \leq 1)$

Sol. (X, Y) is uniform on the unit square.

Event $X^2 + Y^2 \leq 1$ corresponds to region inside both circle and square



$$\text{So } P(X^2 + Y^2 \leq 1) = \frac{\text{area of } \boxed{\text{shaded region}}}{\text{area of } \boxed{\text{square}}} = \frac{\pi/4}{1} = \pi/4$$

(b) Find the conditional probability

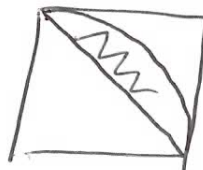
$$P(X^2 + Y^2 \leq 1 \mid X + Y \geq 1)$$

Sol.

The probability we seek is

$$P(X^2 + Y^2 \leq 1, X + Y \geq 1) / P(X + Y \geq 1)$$

Now $P(X^2 + Y^2 \leq 1, X + Y \geq 1) = \text{area of}$



$= \text{area of}$  $-\text{area of}$ 

$$= \pi/4 - 1/2$$

and $P(X + Y \geq 1) = \text{area of}$  $= 1/2$

So the answer is $\frac{\pi/4 - 1/2}{1/2} = \pi/2 - 1$

(c) Find $P(Y \leq X^2)$

Sol. This is area of 

$$= \int_0^1 x^2 dx = \frac{1}{3} x^3 \Big|_0^1 = \frac{1}{3}$$

Densities

Let $f(x, y) = \text{joint probability density function}$

$$\bullet P((X,Y) \in B) = \iint_B f(x,y) dx dy$$

$$\bullet f(x,y) \geq 0, \iint_{-\infty}^{\infty} f(x,y) dx dy = 1$$

• marginals

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

• independence means

$$f(x,y) = f_X(x) f_Y(y) \text{ for all } x,y$$

• expectation of a function g of (X,Y)

$$E[g(X,Y)] = \iint g(x,y) f(x,y) dx dy$$

Example

Let (X,Y) be uniformly distributed on
 $\{(x,y) : 0 < x < y < 1\}$

(a) Find joint density of (X,Y)

Sol.

$$f(x,y) = \begin{cases} c & 0 < x < y < 1 \\ 0 & \text{else} \end{cases}$$

Since area of triangle $0 < x < y < 1$ is $\frac{1}{2}$, $c = 2$.

(b) Find marginal densities $f_X(x)$, $f_Y(y)$. 214

Sol.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y) dy \\ &= \int_{y=x}^{y=1} 2 dy \quad \text{since } f(x,y)=2 \text{ for } 0 < x < y < 1 \\ &\quad \text{and } f(x,y)=0 \text{ else} \\ &= 2(1-x) \quad 0 < x < 1, 0 \text{ else} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y) dx \\ &= \int_{x=0}^{x=y} 2 dx \quad \text{since } f(x,y)=2 \text{ for } 0 < x < y < 1 \\ &\quad \text{and } f(x,y)=0 \text{ else} \\ &= 2y \quad 0 < y < 1, 0 \text{ else} \end{aligned}$$

(c) Are X, Y independent?

Sol. No, since $f(x,y) \neq f_X(x)f_Y(y)$

(d) Find $E(X)$, $E(Y)$

Sol.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 2x(1-x) dx = 1/3$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 2y^2 dy = 2/3$$

(e) Find $E(XY)$

Sol.

$$\begin{aligned}
 E(XY) &= \iint_{\mathbb{R}^2} xy f(x,y) dx dy \\
 &= 2 \int_{y=0}^1 \int_{x=0}^y xy dx dy \\
 &= 2 \int_{y=0}^1 y^3/2 dy = 1/4
 \end{aligned}$$

Remark : can show that joint distribution of

X, Y considered here is that of $X = \min(U, V)$,

$Y = \max(U, V)$, where ~~XXXX~~ U, V are independent

~~the~~ uniform $[0, 1]$ random variables.

Example

Let X, Y be independent and exponentially distributed random variables with parameters λ, μ respectively. calculate $P(X < Y)$.

Sol. By independence, the joint density of X, Y is

$$f(x,y) = (\lambda e^{-\lambda x})(\mu e^{-\mu y})$$

$$= \lambda \mu e^{-\lambda x - \mu y}$$

And $P(X < Y)$ is found by integrating the joint density over the set $\{(x,y) : 0 \leq x < y\}$

$$P(X < Y) = \iint_{0 \leq x < y} \lambda \mu e^{-\lambda x - \mu y} dx dy$$

$$= \int_{x=0}^{\infty} dx \int_{y=x}^{\infty} \lambda \mu e^{-\lambda x - \mu y} dy$$

$$= \int_{x=0}^{\infty} \lambda e^{-\lambda x - \mu x} dx = \lambda / (\lambda + \mu)$$

Note - if λ is small, X is likely to be large
so $P(X < Y)$ is likely to be small

Remark Done in the other order, the integral is $\int_{y=0}^{\infty} dy \int_{x=0}^y \lambda \mu e^{-\lambda x - \mu y} dx$

which simplifies to the same answer.

Independent normal variables

217

Suppose X, Y are independent, each with standard normal density function $\phi(z) = ce^{-z^2/2}$ where the formula $c = 1/\sqrt{2\pi}$ taken for granted until now, will be verified below.

$$\begin{aligned}\text{Then } f(x, y) &= \phi(x)\phi(y) \\ &= c^2 e^{-1/2(x^2+y^2)} \\ &= c^2 e^{-r^2/2}\end{aligned}$$

$$\text{Let } R = \sqrt{x^2 + y^2}$$



$$P(R \in (r, r+dr)) \approx 2\pi r dr c^2 e^{-r^2/2}$$

So R has density function $f_R(r) = 2\pi r c^2 e^{-r^2/2}$

Thus

$$\begin{aligned}1 &= \int_0^\infty f_R(r) dr = \int_0^\infty 2\pi r c^2 e^{-r^2/2} dr \\ &= -2\pi c^2 e^{-r^2/2} \Big|_0^\infty = 2\pi c^2\end{aligned}$$

$$\text{Thus } c = 1/\sqrt{2\pi}$$

218

Remark The distribution of R appearing here,
with density function

$$f_R(r) = r e^{-r^2/2}, \quad r > 0$$

is called the Rayleigh distribution.

Calculating variance of standard normal distribution

Since $E(X) = 0$ by symmetry, the variance is

$$\sigma^2 = E(X^2) = \frac{1}{\sqrt{2\pi}} \int x^2 e^{-x^2/2} dx$$

Can evaluate this using integration by parts
to reduce to the integral of the standard normal
density (exercise).

Another approach

Let X, Y be independent standard normals.

$$\text{Let } R^2 = X^2 + Y^2.$$

$$\text{Then } E(R^2) = E(X^2) + E(Y^2) = 2E(X^2).$$

$$\text{So } \sigma^2 = E(X^2) = \frac{1}{2} E(R^2)$$

But $S = R^2$ has density given by
change of variable formula

$$\begin{aligned} f_S(s) &= f_R(r) / |ds/dr| \\ &= r e^{-r^2/2} / 2r \\ &= \frac{1}{2} e^{-s/2} \quad s > 0. \end{aligned}$$

This is the exponential density with parameter $\lambda = 1/2$,

so that $E(R^2) = E(S) = 1/\lambda = 2$

Thus $\sigma^2 = \frac{1}{2} E(R^2) = 1$.

Sums of independent normal variables

If X, Y are independent with normal (λ, σ^2)
and normal (μ, τ^2) distributions, then $X+Y$
has normal $(\lambda+\mu, \sigma^2+\tau^2)$ distribution.

- will prove this later
- the result extends to more terms

Example let $X_1 = \text{normal}(0, 1)$, $X_2 = \text{normal}(0, 4)$ ^{variance}
 $X_3 = \text{normal}(0, 9)$ be independent.

(a) Find $P(X_1 + X_2 + X_3 < 4)$

Sol.

Let $S = X_1 + X_2 + X_3$. Then S is normal with mean 0 and variance $1 + 4 + 9 = 14$.

$$\text{So } P(S < 4) = P\left(\frac{S}{\sqrt{14}} < \frac{4}{\sqrt{14}}\right)$$

$$= P\left(Z < \frac{4}{\sqrt{14}}\right) = \Phi\left(\frac{4}{\sqrt{14}}\right) \approx .857$$

standard normal $\nearrow$

(b) Find $P(4X_1 - 10 < X_2 + X_3)$

Sol.

$$= P(4X_1 - X_2 - X_3 < 10)$$

$$= P(L < 10)$$

Here $L = 4X_1 - X_2 - X_3$, so L has mean 0

and variance $4^2(1^2) + (-1)^2 2^2 + (-1)^2 3^2$

$$= 29$$

$$\text{So } P(L < 10) = P\left(Z < \frac{10}{\sqrt{29}}\right) = \Phi\left(\frac{10}{\sqrt{29}}\right)$$

$\uparrow$
standard
normal

$$\approx .968$$

Distribution of sums

221

Some examples (see book for details)

Distribution of terms

Distribution of sum

n indep. Bernoulli(p)

binomial(n, p)

indep. Poisson(μ_i)

Poisson($\sum \mu_i$)

indep. normal(μ_i, σ_i^2)

normal($\sum \mu_i, \sum \sigma_i^2$)

r indep. geometric p

negative binomial(r, p)

r indep. exponential(λ)

gamma(r, λ)

In discrete case,

$$P(X+Y=z) = \sum_x P(X=x, Y=z-x)$$

In continuous case,

- If (X, Y) has density function $f(x, y)$ in the plane, then $X+Y$ has density

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x) dx$$

- If X, Y are independent with density functions $f_X(x), f_Y(y)$, then $X+Y$ has density function $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$

Proof (see book)

Example Sum of two independent uniforms on $[0, 1]$

Sol.

Let X, Y be independent uniforms on $[0, 1]$

$$\text{Then } f_X(x) = f_Y(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$$

$$\text{Let } Z = X + Y.$$

$$\text{Then } f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$

$$= \int_0^1 f_X(z-y) dy$$

$$\begin{aligned} &\text{since} \\ f_Y(y) &= 1 \text{ if } 0 \leq y \leq 1 \\ &= 0 \text{ else} \end{aligned}$$

Case 1 $0 \leq z \leq 1$

$$\text{then } 0 \leq y \leq z, \text{ so } f_Z(z) = \int_0^z dy = z$$

Case 2 $1 < z \leq 2$

$$\text{Then } z-1 \leq y \leq 1, \text{ so } f_Z(z) = \int_{z-1}^1 dy = 2-z$$

$$\begin{aligned} &\text{since} \\ 0 &\leq z-y \leq 1 \end{aligned}$$

If $z < 0$ or $z > 2$, then $f_z(z) = 0$. Hence,

223

$$f_z(z) = \begin{cases} z & 0 \leq z \leq 1 \\ 2-z & 1 < z \leq 2 \\ 0 & \text{else} \end{cases}$$

Example (sum of two independent exponentials)

Let X, Y be independent exponentials with parameter λ . What is density of $X+Y$?

Sol.

$$f_X(x) = f_Y(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

So if $z > 0$,

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \\ &= \int_0^{\infty} f_X(z-y) \lambda e^{-\lambda y} dy \\ &= \int_0^z \lambda e^{-\lambda(z-y)} \lambda e^{-\lambda y} dy \\ &= \int_0^z \lambda^2 e^{-\lambda z} dy = \lambda^2 z e^{-\lambda z} \end{aligned}$$

If $z < 0$, $f_z(z) = 0$. Hence

$$f_z(z) = \begin{cases} \lambda^2 z e^{-\lambda z} & z \geq 0 \\ 0 & \text{else} \end{cases}$$

Example {sum of two independent
standard normals}

224

Let X, Y be standard normals, so

$$f_X(x) = f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Let $Z = X + Y$. What is density of Z ?

Sol.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(z-y)^2/2} e^{-y^2/2} dy$$

$$= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{\infty} e^{-(y-z/2)^2} dy$$

$$= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{\infty} e^{-y^2} dy$$

$$= \frac{1}{2\pi} e^{-z^2/4} \sqrt{\pi} \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} dy \right]$$

$$= \frac{1}{\sqrt{4\pi}} e^{-z^2/4}$$

= 1 since it's integral
of a normal density,
with $\mu=0$, $\sigma=1/\sqrt{2}$

So Z is normal with mean 0, variance 2.

Example

The joint density of X and Y is

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x, y < \infty \\ 0 & \text{else} \end{cases}$$

Find density of X/Y .

Sol.

$$F_{X/Y}(a) = P(X/Y \leq a)$$

$$= \iint_{\substack{X/Y \leq a \\ 0 < X, Y < \infty}} e^{-(x+y)} dx dy$$

$$= \int_0^\infty \int_0^{ay} e^{-(x+y)} dx dy$$

$$= \int_0^\infty (1 - e^{-ay}) e^{-y} dy$$

$$= \left[-e^{-y} + e^{-(a+1)y} / (a+1) \right]_0^\infty$$

$$= 1 - \frac{1}{a+1}$$

Differentiating gives that

$$f_{X|Y}(a) = \frac{1}{(a+1)^2} \quad 0 \leq a < \infty$$

226

can check that $\int_0^{\infty} \frac{1}{(a+1)^2} da = 1$

Conditional distributions: discrete case

Example

Suppose a fair die is rolled. Then as many fair coins are tossed as there are spots showing on the die.

Find the distribution of the number of heads showing among the coins.

Sol.

Let Y = number of heads showing among coins

X = number showing on the die

Then $P(Y=y | X=x) = \binom{x}{y} / 2^x$ ← probability of y heads when toss x coins

So conditional distribution of Y given $X=x$ is binomial $(x, 1/2)$.

Now $P(Y=y) = P(y \text{ heads})$

$$\begin{aligned}
 &= \sum_{x=1}^6 P(\text{die rolls } x, \text{ and } y \text{ heads}) \\
 &= \sum_{x=1}^6 P(y \text{ heads} \mid \text{die rolls } x) P(\text{die rolls } x) \\
 &= \sum_{x=1}^6 P(Y=y \mid X=x) P(X=x) \\
 &= \frac{1}{6} \sum_{x=1}^6 \binom{x}{y} \frac{1}{2^x} \quad 0 \leq y \leq 6
 \end{aligned}$$

For example,

$$P(Y=0) = \frac{1}{6} \left(\frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^6} \right) = 63/384$$

$$P(Y=4) = \frac{1}{6} \left(\binom{4}{4} \frac{1}{2^4} + \binom{5}{4} \frac{1}{2^5} + \binom{6}{4} \frac{1}{2^6} \right) = 29/384$$

Table

y	0	1	2	3	4	5	6
$P(Y=y)$	$\frac{63}{384}$	$\frac{120}{384}$	$\frac{99}{384}$	$\frac{64}{384}$	$\frac{29}{384}$	$\frac{8}{384}$	$\frac{1}{384}$

For each possible value of X , as y ranges over all possible values of Y , the probabilities

$P(Y=y | X=x)$ form a probability distribution depending on x , called the conditional distribution of Y given that $X=x$.

Remark The given value of X can be thought of as a parameter in the distribution of Y given that $X=x$. In previous example, the distribution of Y given $X=x$ is binomial with parameters $n=x$, $p=1/2$.

Rule of average conditional probabilities (implicitly used in last example)

$$P(Y=y) = \sum_x P(Y=y | X=x) P(X=x)$$

Proof

$$\begin{aligned} \sum_x P(Y=y | X=x) P(X=x) \\ &= \sum_x \frac{P(Y=y, X=x)}{P(X=x)} P(X=x) \\ &= \sum_x P(Y=y, X=x) = P(Y=y) \end{aligned}$$

Example

As in previous example, let Y be number of heads on X fair coin tosses, where X is uniformly distributed on $\{1, 2, \dots, 6\}$. Find conditional distribution of X given $Y=y$, for $y=2$.

Sol.

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = P(Y=y|X=x)P(X=x)$$

get by summing numerator over x

So if $y=2$,

$$P(X=x, Y=2) = \begin{cases} 0 & x=1 \\ 1/24 & x=2 \\ 1/16 & x=3 \\ 1/16 & x=4 \\ 5/96 & x=5 \\ 5/128 & x=6 \end{cases}$$

Thus $P(Y=2) = 0 + 1/24 + 1/16 + 1/16 + 5/96 + 5/128$

$$= 99/384$$

Table

x	1	2	3	4	5	6
$P(X=x Y=2)$	0	$\frac{16}{99}$	$\frac{24}{99}$	$\frac{24}{99}$	$\frac{20}{99}$	$\frac{15}{99}$

Continuous case = conditional density
of Y given $X=x$

230

For random variables X, Y with joint density $f(x, y)$, for each x such that $f_X(x) > 0$, the conditional density of Y given $X=x$ is the probability density function

$$f_Y(y | X=x) = f(x, y) / f_X(x)$$

Example

Suppose a point (X, Y) is chosen uniformly at random from triangle $\{(x, y) : x, y \geq 0, x+y \leq 2\}$.

Find $P(Y > 1 | X=x)$

Sol.

The uniform distribution on triangle has joint density

$$f(x, y) = \begin{cases} 1/2 & x, y \geq 0, x+y \leq 2 \\ 0 & \text{else} \end{cases}$$

So for $0 \leq x \leq 2$,

$$f_X(x) = \int_0^{\infty} f(x, y) dy = \int_0^{2-x} \frac{1}{2} dy = \frac{1}{2} (2-x)$$

$$\text{and } f_Y(y|X=x) = \frac{f(x,y)}{f_X(x)}$$

231

$$= \begin{cases} \frac{1}{2-x} & 0 \leq y \leq 2-x \\ 0 & \text{else} \end{cases}$$

That is, given $X=x$, for $0 \leq x \leq 2$, Y has uniform $[0, 2-x]$ distribution, as one intuitively expects.

$$\begin{aligned} \text{So } P(Y > 1 | X=x) &= \int_1^{\infty} f_Y(y|X=x) dy \\ &= \int_1^{2-x} \frac{dy}{2-x} = \frac{1-x}{2-x} \quad 0 \leq x \leq 1. \end{aligned}$$

$$\text{and } P(Y > 1 | X=x) = 0 \text{ otherwise}$$

(since if $x > 1$ and $f_Y(y|X=x) \neq 0$
then $y < 1$)

Multiplication rule for densities

$$f(x,y) = f_X(x) f_Y(y|X=x)$$

Proof This is a rearrangement of formula

$$f_Y(y|X=x) = f(x,y) / f_X(x)$$

Example Suppose X has gamma(2, λ) distribution and that given x , Y has uniform $[0, x]$ distribution.

(a) Find joint density of X and Y .

Sol. By definition of gamma distribution,

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

From uniform $[0, x]$ distribution of Y given $X=x$,

$$f_Y(y|X=x) = \begin{cases} 1/x & 0 \leq y \leq x \\ 0 & \text{else} \end{cases}$$

So by multiplication rule for densities,

$$f(x, y) = f_X(x) f_Y(y|X=x) = \begin{cases} \lambda^2 e^{-\lambda x} & 0 < y < x \\ 0 & \text{else} \end{cases}$$

(b) Find marginal density of Y

Sol.

$$\begin{aligned} f_Y(y) &= \int_0^{\infty} f(x, y) dx = \int_y^{\infty} \lambda^2 e^{-\lambda x} dx \\ &= \lambda e^{-\lambda y} \quad (y > 0). \end{aligned}$$

$$f_Y(y) = 0 \quad \text{for } y \leq 0.$$

233

Thus Y has exponential(λ) distribution.

Useful formula

$$P(A) = \int P(A | X=x) f_X(x) dx$$

Example

Pick π uniformly from $(0,1)$. Given that $\pi=p$, let S_n be number of successes in n independent trials with probability of success p on each trial. Find distribution of S_n .

Sol. Density of π is $f_\pi(p) = \begin{cases} 1 & 0 < p < 1 \\ 0 & \text{else} \end{cases}$

Thus

$$\begin{aligned} P(S_n=k) &= \int P(S_n=k | \pi=p) f_\pi(p) dp \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp \end{aligned}$$

$$\begin{aligned} &= 1/(n+1) \\ \text{omit} \rightarrow & \text{proof} \end{aligned}$$

So S_n is uniform on $\{0, 1, \dots, n\}$

Theorem Let X_1, X_2 have joint probability density f_{X_1, X_2} . Let $Y_1 = g_1(X_1, X_2)$
 $Y_2 = g_2(X_1, X_2)$

$$\text{Define } J(X_1, X_2) = \begin{vmatrix} \frac{\partial g_1}{\partial X_1} & \frac{\partial g_1}{\partial X_2} \\ \frac{\partial g_2}{\partial X_1} & \frac{\partial g_2}{\partial X_2} \end{vmatrix}$$

This is a 2×2 determinant of matrix $\begin{pmatrix} \frac{\partial g_1}{\partial X_1} & \frac{\partial g_1}{\partial X_2} \\ \frac{\partial g_2}{\partial X_1} & \frac{\partial g_2}{\partial X_2} \end{pmatrix}$

Then under mild technical conditions (see book for details)

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{|J(X_1, X_2)|}$$

must express this in terms of Y_1, Y_2

Example

Let X_1, X_2 have joint density function f_{X_1, X_2} .

Let $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$

Find joint density of Y_1, Y_2 .

Sol.

$$\text{Let } y_1 = x_1 + x_2, \quad y_2 = x_1 - x_2.$$

$$\text{Then } x_1 = (y_1 + y_2)/2, \quad x_2 = (y_1 - y_2)/2$$

$$\text{Let } g_1 = x_1 + x_2, \quad g_2 = x_1 - x_2$$

$$\text{Then } |J(x_1, x_2)| = |(1)(-1) - (1)(1)| = |-2| = 2$$

$$\text{So } f_{y_1, y_2} = \frac{f_{x_1, x_2} \left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2} \right)}{2}$$

see book for more examples

Chapter 7 Properties of expectation

Prop.

If X, Y have joint probability mass function

$$p(x, y), \text{ then } E[g(X, Y)] = \sum_y \sum_x g(x, y) p(x, y)$$

If X, Y have joint probability density function

$$f(x, y), \text{ then } E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

Example

An accident occurs at a point X uniformly distributed on a road of length L . At time of accident, an ambulance is at location Y that is also uniformly distributed on the road. Assuming X, Y are independent, find expected distance between ambulance and point of accident.

Sol.

need to compute $E[|X-Y|]$

Now joint density of X, Y is $f(x, y) = \begin{cases} 1/L^2 & 0 \leq x, y \leq L \\ 0 & \text{else} \end{cases}$

Thus

$$E[|X-Y|] = \frac{1}{L^2} \int_0^L \int_0^L |x-y| dy dx$$

$$\text{Now } \int_0^L |x-y| dy = \int_0^x (x-y) dy \quad \swarrow \text{case: } y < x \quad + \quad \int_x^L (y-x) dy \quad \swarrow \text{case: } y > x$$

after some simplification $\rightarrow = L^2/2 + x^2 - xL$

$$\text{Thus } E[|X-Y|] = \frac{1}{L^2} \int_0^L \left(\frac{L^2}{2} + x^2 - xL \right) dx$$

$$= L/3$$

Note $E(X+Y) = E(X) + E(Y)$, even if X, Y are not independent. We proved this earlier in the discrete case. In continuous case,

$$E(X+Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f(x,y) dx dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy$$

$$= E(X) + E(Y)$$

Example (mean of binomial random variable)

Let X be binomial (n, p) . Thus

$X = X_1 + \dots + X_n$ where X_i 's are independent
with $P(X_i=1)=p$, $P(X_i=0)=1-p$.

Clearly $E(X_i) = 1(p) + 0(1-p) = p$ for all i

Thus $E(X) = E(X_1) + \dots + E(X_n) = np$

Example

If n balls are randomly selected from an urn containing N balls of which m are white, find the expected number of white balls selected.

Method 1

Let X = number of white balls selected

$$= X_1 + \dots + X_m$$

where $X_i = \begin{cases} 1 & \text{if } i\text{th white ball is selected} \\ 0 & \text{else} \end{cases}$

Then $E(X_i) = P(X_i=1) = P(\text{ith white ball is selected})$

$$= \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

Thus $E(X) = E(X_1) + \dots + E(X_m) = \frac{mn}{N}$

Method 2

$$X = Y_1 + \dots + Y_n$$

$$\text{where } Y_i = \begin{cases} 1 & \text{if } i\text{th ball selected is white} \\ 0 & \text{else} \end{cases}$$

Since the i th ball selected is equally likely to be any of the N balls,

$$E(Y_i) = P(Y_i = 1) = \frac{m}{N} \text{ for all } i.$$

$$\text{Thus } E(X) = E(Y_1) + \dots + E(Y_n) = \frac{nm}{N}$$

Example What is the expected number of fixed points of a random permutation of $\{1, \dots, n\}$.

Sol. Let X = number of fixed points

$$\text{So } X = X_1 + \dots + X_n \text{ where } X_i = \begin{cases} 1 & \text{if } i \text{ is fixed} \\ 0 & \text{else} \end{cases}$$

$$\text{Now } E(X_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

number of permutations fixing i

number of permutations

$$\begin{aligned} \text{So } E(X) &= E(X_1) + \dots + E(X_n) \\ &= \frac{1}{n} + \dots + \frac{1}{n} = 1. \end{aligned}$$

The end.

MATH 407, Lecture 17

240

Example : coupon collecting problem

Suppose there are n types of coupons, and each time one obtains a coupon, it's equally likely to be any one of the n types. Find the expected number of coupons one needs to obtain a complete set of at least one of each type.

Sol. Let X = number of coupons collected until get complete set

Let X_i , $i=0, 1, \dots, n-1$ be the number of additional coupons that need to be obtained after i distinct types have been collected to obtain another distinct type. Note that

$$X = X_0 + X_1 + \dots + X_{n-1}$$

$$P(X_i = k) = \frac{n-i}{n} \left(\frac{i}{n}\right)^{k-1}, \quad k \geq 1$$

So X_i is geometric with parameter $\frac{n-i}{n}$. Thus $E(X_i) = \frac{n}{n-i}$, so that

$$E(X) = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \dots + \frac{n}{1}$$

$$= n \left(1 + \dots + \frac{1}{n-1} + \frac{1}{n} \right)$$

Example Consider a random permutation of $\{1, \dots, n\}$. What is variance of number of fixed points?

Sol.

As earlier, $X = X_1 + \dots + X_n$ where $X_i = \begin{cases} 1 & \text{if } i \text{ is fixed} \\ 0 & \text{else} \end{cases}$

We saw that $E(X) = 1$.

$$E(X^2) = E(X_1^2 + \dots + X_n^2) + 2 \sum_{i < j} E(X_i X_j)$$

since $X_i = 0 \text{ or } 1$ $\rightarrow$

$$= E(X_1 + \dots + X_n) + 2 \sum_{i < j} E(X_i X_j)$$

$$= n \left(\frac{1}{n} \right) + 2 \binom{n}{2} E(X_1 X_2)$$

$$= 1 + 2 \binom{n}{2} P(1, 2 \text{ are both fixed})$$

$$= 1 + 2 \binom{n}{2} \frac{(n-2)!}{n!} = 1 + 1 = 2$$

$$\begin{aligned}\text{Thus } \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= 2 - 1 = 1\end{aligned}$$

Remark: As $n \rightarrow \infty$, distribution of fixed points tends to a Poisson (1) random variable.

Covariance and Correlation

Earlier we showed that if X, Y are independent then $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. This followed from more general formula

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2E[(X-\mu_X)(Y-\mu_Y)]$$

where $\mu_X = E(X)$, $\mu_Y = E(Y)$.

Def. The covariance of X, Y , denoted $\text{Cov}(X, Y)$ is the number

$$\text{Cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

where $\mu_X = E(X)$, $\mu_Y = E(Y)$

Alternative formula

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Variance of a sum (proved earlier)

243

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

Proof of alternative formula

$$(X - \mu_X)(Y - \mu_Y) = XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y$$

Take expectations to get

$$\begin{aligned} E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y \\ = E[XY] - 2E[X]E[Y] + E[X]E[Y] \\ = E[XY] - E[X]E[Y] \end{aligned}$$

Remarks

- 1) $\text{Cov}(X, X) = \text{Var}(X)$, so the above formulas for covariance are extensions of old formulas for variance
- 2) If X, Y are independent, $\text{Cov}(X, Y) = 0$
- 3) $\text{Cov}(X, Y) = 0$ does not imply X, Y are independent.

Indicators

244

Let $X = I_A$ be indicator of event A

$Y = I_B$ be indicator of event B

Then $XY = I_A I_B = I_{A \cap B}$. Thus,

$$E(I_A) = P(A), E(I_B) = P(B), E(I_A I_B) = P(A \cap B)$$

$$\text{So } \text{Cov}(I_A, I_B) = P(A \cap B) - P(A)P(B)$$

This covariance is

positive $\Leftrightarrow P(A \cap B) > P(A)P(B)$, call A, B positively dependent

zero $\Leftrightarrow P(A \cap B) = P(A)P(B)$, when A, B are independent

negative $\Leftrightarrow P(A \cap B) < P(A)P(B)$, call A, B negatively dependent

In case of positive dependence, learning that B occurred increases the chance of A :

$$P(A|B) > P(A) \text{ and vice-versa } P(B|A) > P(B)$$

For negative dependence learning that B occurred decreases chance of A : 245

$$P(A|B) < P(A) \text{ and vice-versa } P(B|A) < P(B)$$

Example : draws with and without replacement

A box has b black balls, w white balls, where $b, w > 0$. Make two draws from the box. Let B_i, W_i be events of getting a black or white ball on i th draw, $i=1,2$. Can check the following, which is intuitively obvious.

<u>Pair of events</u>	<u>Sampling with replacement</u>	<u>Sampling without replacement</u>
B_1, B_2	indep.	- dependent
B_1, W_2	indep.	+ dependent
W_1, W_2	indep.	- dependent
W_1, B_2	indep.	+ dependent

Sign of covariance

The sign of $\text{Cov}(X, Y)$ is positive if above average

values of X tend to be associated with above average values of Y , and below average values of X with below average values for Y . (For example X =height, Y =weight of random individual). The random variable $(X-\mu_X)(Y-\mu_Y)$ is then most likely positive, with a positive expectation.

The sign of $\text{Cov}(X, Y)$ is negative if above average values of X tend to be associated with below average values of Y and vice-versa. (For example age of a car and its fuel economy). Then $(X-\mu_X)(Y-\mu_Y)$ is most likely negative, with a negative expectation.

While the sign of covariance can be interpreted as above, its magnitude is hard to interpret. It's easier to interpret the correlation of X and Y , denoted $\text{Corr}(X, Y)$ and defined as:

Def.

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X) \text{SD}(Y)}$$

Assume that neither X nor Y is constant, so $\text{SD}(X) \text{SD}(Y) > 0$. The sign of $\text{Cov}(X, Y)$ is then the same as sign of $\text{Corr}(X, Y)$.

Conditions for X, Y to be uncorrelated

The following 3 conditions are equivalent:

- 1) $\text{Corr}(X, Y) = 0$
- 2) $\text{Cov}(X, Y) = 0$
- 3) $E(XY) = E(X)E(Y)$

in which case call X, Y uncorrelated.

Independent variables are uncorrelated, but uncorrelated variables are not necessarily independent.

let X^*, Y^* denote X, Y rescaled to standard units. So

$$X^* = \frac{X - \mu_X}{\text{SD}(X)}, \quad Y^* = \frac{Y - \mu_Y}{\text{SD}(Y)}$$

Then $E(X^*) = E(Y^*) = 0$, and

$$SD(X^*) = SD(Y^*) = 1$$

Claim $\text{Corr}(X, Y) = \text{Cov}(X^*, Y^*) = E(X^* Y^*)$,
so correlation is a kind of "standardized covariance".

Proof

$$\begin{aligned} \text{Cov}(X^*, Y^*) &= E(X^* Y^*) - E(X^*) E(Y^*) \\ &= E(X^* Y^*) \\ &= \frac{1}{SD(X)} \frac{1}{SD(Y)} E[(X - \mu_X)(Y - \mu_Y)] \\ &= \frac{\text{Cov}(X, Y)}{SD(X) SD(Y)} \\ &= \text{Corr}(X, Y). \end{aligned}$$

Theorem

$-1 \leq \text{Corr}(X, Y) \leq 1$, no matter what the
joint distribution of
 X and Y

Proof

Since $E(X^{*2}) = E(Y^{*2}) = 1$,

$$0 \leq E[(X^* - Y^*)^2] = 1 + 1 - 2E(X^*Y^*)$$

$$0 \leq E[(X^* + Y^*)^2] = 1 + 1 + 2E(X^*Y^*)$$

Thus $-1 \leq E(X^*Y^*) \leq 1$, so result follows

since $\text{Corr}(X, Y) = E(X^*Y^*)$ by previous result.

Correlations of ± 1

The proof that correlations are between ± 1 show that $\text{Corr}(X, Y) = +1$ if and only if $E[(X^* - Y^*)^2] = 0$, that is if and only if $X^* = Y^*$ with probability 1. This means there are constants a, b with $a > 0$ such that $Y = aX + b$ with probability 1.

So a correlation of $+1$ determines a linear relationship between X, Y with positive slope, and a correlation of -1 indicates a linear relationship between X, Y with negative slope.

Example (Black and white)

let N_R = number of reds that appear
 N_B = number of blacks that appear

in n spins of a roulette wheel that
has proportion r of its numbers red,
proportion b black, and the rest green.

250

(So $r+b < 1$. In many casinos, $r=b=18/38$).

Problem Find $\text{Corr}(N_R, N_B)$.

Sol. Note if $r+b=1$ (no greens on wheel), then
 $N_B = n - N_R$, so $\text{Corr}(N_R, N_B) = -1$. For $r+b$
close to 1, this relation is still approximately
correct, so expect a correlation close to -1 .

Since N_R is binomial (n, r) ,

$$E(N_R) = nr, \quad SD(N_R) = \sqrt{nr(1-r)}$$

Similarly

$$E(N_B) = nb, \quad SD(N_B) = \sqrt{nb(1-b)}$$

$$\text{Since } \text{Cov}(N_R, N_B) = E(N_R N_B) - E(N_R) E(N_B),$$

$$\text{to calculate } \text{Corr}(N_R, N_B) = \frac{\text{Cov}(N_R, N_B)}{SD(N_R) SD(N_B)}$$

the only missing ingredient is $E(N_R N_B)$.

To compute $E(N_R N_B)$ directly from the joint distribution of N_R, N_B is very difficult.

Instead, we compute $\text{Cov}(N_R, N_B)$ using

$$\text{Var}(N_R + N_B) = \text{Var}(N_R) + \text{Var}(N_B) + 2 \text{Cov}(N_R, N_B)$$

The point is that $N_R + N_B$ is the number of spins which are red or black, which is a binomial $(n, r+b)$ random variable with variance $n(r+b)(1-r-b)$. Rearranging $\square$ and substituting variances gives

$$\begin{aligned} \text{Cov}(N_R, N_B) &= \frac{1}{2} n [(r+b)(1-r-b) - r(1-r) - b(1-b)] \\ &= -nrb \end{aligned}$$

$$\begin{aligned} \text{Hence } \text{Corr}(N_R, N_B) &= \frac{-nrb}{\sqrt{nr(1-r)} \sqrt{nb(1-b)}} \\ &= -\sqrt{\frac{rb}{(1-r)(1-b)}} \end{aligned}$$

Remark: if $r=b=18/38$, then $\text{Corr}(N_R, N_B) = -.9$

Conditional expectation: discrete case

def. The conditional expectation of a random

variable Y given an event A , denoted by 252

$E(Y|A)$, is the expectation of Y under the conditional probability distribution given A :

$$E(Y|A) = \sum_y y P(Y=y|A)$$

Example

Let Y be number of heads in 4 tosses of a fair coin. What is the conditional expectation of Y given 2 or fewer heads?

Sol.

Let A = event 2 or fewer heads

Then $P(Y=y|A) = P(Y=y, A) / P(A)$

$$\text{Now } P(A) = \frac{\binom{4}{0} + \binom{4}{1} + \binom{4}{2}}{2^4} = 11/16$$

$$P(Y=y|A) = \begin{cases} \binom{4}{y}/2^4 & \text{if } y=0,1,2 \\ 0 & \text{else} \end{cases}$$

thus,

$$E(Y|A) = \sum_y y P(Y=y|A)$$

$$= \sum_{y=0}^2 \frac{y \binom{4}{y} 12^y}{11/16}$$

253

$$= (1 \cdot 4 + 2 \cdot 6) / 11 = 16/11$$

Long-run interpretation:

If you repeatedly toss 4 coins, the long run average number of heads, averaging only over trials that produce 0, 1, or 2 heads, will be $16/11$.

Properties of conditional expectation

$$1) E(X+Y|A) = E(X|A) + E(Y|A)$$

2) If $A_1, \dots, A_n$ is a partition of the sample space into disjoint sets,

$$E(Y) = \sum_{i=1}^n E(Y|A_i) P(A_i)$$

• In the special case $Y = I_B$, the indicator of event B , we get the rule

$$P(B) = \sum_{i=1}^n P(B|A_i) P(A_i)$$

• For partition generated by values of X , get

$$E(Y) = \sum_x E(Y|X=x) P(X=x).$$

Definition of $E(Y|X)$

The conditional expectation of Y given X , denoted $E(Y|X)$ is the function of X whose value is $E(Y|X=x)$ if $X=x$.

Note that $E(Y|X)$ is a random variable, since it's a function of the random variable X .

Can prove $E(Y) = E(E(Y|X))$

Example Let Y = number of heads in X tosses of a fair coin, where X is generated by a fair die roll.

1) What is $E(Y|X=x)$?

Sol. The conditional distribution of Y given $X=x$ is binomial($x, 1/2$). So the conditional expectation is $x/2$.

To be continued...

From last time, let Y = number of heads
in X tosses of a
fair coin, where X is
generated by a fair
die roll

1) What is $E(Y|X=x)$?

~ we solved this last time, showing that

$$E(Y|X=x) = x/2$$

2) Find $E(Y)$

we know that $E(Y|X) = X/2$, $E(X) = 3.5$

Thus,

$$E(Y) = E[E(Y|X)]$$

$$= E(X/2) = 1/2 E(X) = \frac{3.5}{2} = 1.75$$

Remark: can also calculate $E(Y)$ from the
distribution of Y , calculated in an earlier
example

3) Find $E(X|Y=2)$

In earlier example, we calculated 256
table for $P(X=x|Y=2)$

x	1	2	3	4	5	6
$P(X=x Y=2)$	0	$\frac{16}{99}$	$\frac{24}{99}$	$\frac{24}{99}$	$\frac{20}{99}$	$\frac{15}{99}$

$$\begin{aligned}
 \text{Thus, } E(X|Y=2) &= 2\left(\frac{16}{99}\right) + 3\left(\frac{24}{99}\right) + 4\left(\frac{24}{99}\right) \\
 &\quad + 5\left(\frac{20}{99}\right) + 6\left(\frac{15}{99}\right) \\
 &= 390/99 \approx 3.94
 \end{aligned}$$

Example

Suppose number of children in a family is a random variable X with mean μ , and given $X=n$, each of the n children in the family is a girl with probability p and a boy with probability $1-p$. What is the expected number of girls in the family?

Sol. Intuitively, should be $p\mu$.

To see this is correct, let G = number of girls in family

Then $E(G|X=n) = np$, since given $X=n$, G is binomial (n, p) .

257

$$\begin{aligned}\text{Thus, } E(G) &= \sum_n E(G|X=n) P(X=n) \\ &= \sum_n np P(X=n) \\ &= p \sum_n n P(X=n) = p \mu.\end{aligned}$$

Example

Let S_n = number of successes in n independent trials with probability of success p on each trial

Calculate $E(S_m | S_n = k)$ for $m \leq n$

Sol.

Since $S_m = X_1 + \dots + X_m$ where X_j = indicator of success on trial j

$$E(S_m | S_n = k) = \sum_{j=1}^m E(X_j | S_n = k)$$

$$\begin{aligned}\text{Now } E(X_j | S_n = k) &= 1 \cdot P(X_j = 1 | S_n = k) + 0 \cdot P(X_j = 0 | S_n = k) \\ &= P(j\text{th trial is success} | S_n = k) \\ &= \frac{P(j\text{th trial is success, } S_n = k)}{P(S_n = k)}\end{aligned}$$

$$= \frac{P(\text{jth trial is success, } k-1 \text{ of other } n-1 \text{ trials are successes})}{P(S_n = k)} \quad 258$$

$$\stackrel{\uparrow}{=} \frac{P \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}}$$

using independence and binomial distribution

$$= k/n$$

$$\text{Thus } E(S_m | S_n = k) = mk/n$$

Remark This formula can be derived in other ways.

For example, $E(X_j | S_n)$ must be the same for all j by symmetry, and equal to $E(X_1 | S_n)$.

$$\text{Since } S_m = E(S_n | S_n) = \sum_{j=1}^n E(X_j | S_n) = n E(X_1 | S_n)$$

it follows that $E(X_1 | S_n) = S_n/n$.

$$\begin{aligned} \text{Hence } E(S_m | S_n) &= \sum_{j=1}^m E(X_j | S_n) = m E(X_1 | S_n) \\ &= \frac{m}{n} S_n \end{aligned}$$

Example Suppose X, Y are independent.

Find $E(X+Y | X=x)$

259

Sol.

$$E(X+Y | X=x) = E(X | X=x) + E(Y | X=x)$$

$$= x + E(Y)$$

⚡
since X, Y are independent

Conditional expectation : continuous case (very brief treatment)

Suppose X, Y have joint probability density function $f(x, y)$. Then if $f_Y(y) > 0$, the conditional density of X given that $Y=y$ is defined by

$$f_{X|Y}(x|y) = f(x, y) / f_Y(y)$$

Define conditional expectation of X given that $Y=y$

by $E[X | Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$

provided that $f_Y(y) > 0$.

Example

Suppose joint density of X and Y is given by

$$f(x, y) = e^{-x/y} e^{-y} / y \quad 0 < x, y < \infty.$$

Compute $E[X|Y=y]$

260

Sol. Compute the conditional density

$$\begin{aligned} f_{X|Y}(x|y) &= f(x,y)/f_Y(y) \\ &= f(x,y) / \int_{-\infty}^{\infty} f(x,y) dx \\ &= \frac{(1/y) e^{-x/y} e^{-y}}{\int_0^{\infty} \frac{1}{y} e^{-x/y} e^{-y} dx} \\ &= \frac{1/y e^{-x/y}}{\int_0^{\infty} 1/y e^{-x/y} dx} \\ &= \frac{1}{y} e^{-x/y} \end{aligned}$$

$$\text{Thus } E[X|Y=y] = \int_0^{\infty} \frac{x}{y} e^{-x/y} dx = y$$

Remark More generally, $E[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx$

Conditional variance

$$\text{Var}(X|Y) = E[(X - E[X|Y])^2 | Y]$$

Useful formula:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

Example

Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables, and let N be a non-negative integer valued random variable that is independent of the X_i 's.

Claim $\text{Var} \left(\sum_{i=1}^N X_i \right) = E(N) \text{Var}(X) + (E(X))^2 \text{Var}(N)$

where X has the distribution of the X_i 's.

Proof

Let $X = X_1 + \dots + X_N$, $Y = N$.

Then by the "useful formula",

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^N X_i \right) &= E \left[\text{Var} \left(\sum_{i=1}^N X_i \mid N \right) \right] + \text{Var} \left(E \left(\sum_{i=1}^N X_i \mid N \right) \right) \\ &= E [N \text{Var}(X)] + \text{Var} (N E(X)) \\ &= \text{Var}(X) E(N) + (E(X))^2 \text{Var}(N) \end{aligned}$$

Moment generating functions

Define $M(t) = E[e^{tX}]$

$$= \begin{cases} \sum_x e^{tx} p(x) & \text{if } X \text{ is discrete with mass} \\ & \text{function } p(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x) & \text{if } X \text{ is continuous with} \\ & \text{density } f(x) \end{cases}$$

Claim $M'(0) = E(X)$

Proof (discrete case)

$$\begin{aligned} M'(t) &= d/dt E[e^{tX}] \\ &= E\left[\frac{d}{dt} e^{tX}\right] \\ &= E[Xe^{tX}] \end{aligned}$$

Thus $M'(0) = E[X]$.

Claim $M''(0) = E(X^2)$

$$\begin{aligned} M''(t) &= \frac{d}{dt} E[Xe^{tX}] \\ &= E\left[X \frac{d}{dt} e^{tX}\right] \\ &= E[X^2 e^{tX}]. \end{aligned}$$

Thus $M''(0) = E(X^2)$

Similarly n th derivative of $M(t)$ evaluated at $t=0 = E[X^n]$

Examples

263

Binomial (n, p)

Let X be binomial (n, p) . Then

$$m(t) = E[e^{tX}]$$

$$= \sum_{k=0}^n e^{tk} P(X=k)$$

$$= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k}$$

$$\stackrel{\substack{\text{binomial} \\ \text{theorem}}}{=} (pe^t + 1 - p)^n$$

$$\text{Thus, } m'(t) \stackrel{\substack{\text{chain rule}}}{=} n(pe^t + 1 - p)^{n-1} \frac{d}{dt} (pe^t + 1 - p)$$

$$= n(pe^t + 1 - p)^{n-1} pe^t$$

$$\text{Hence } E(X) = m'(0) = np$$

By product rule and chain rule,

$$m''(t) = n(n-1)(pe^t + 1 - p)^{n-2} (pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

$$\text{Hence } E(X^2) = m''(0) = n(n-1)p^2 + np$$

$$\text{So } \text{Var}(X) = E(X^2) - (E(X))^2$$

264

$$= n(n-1)p^2 + np - n^2p^2$$

$$= np(1-p) \quad \text{as we proved earlier in the course}$$

Example: Poisson(λ)

Let X be Poisson(λ). Then

$$M(t) = E[e^{tX}]$$

$$= \sum_{n \geq 0} e^{tn} p(X=n)$$

$$= \sum_{n \geq 0} e^{tn} e^{-\lambda} \lambda^n / n!$$

$$= e^{-\lambda} \sum_{n \geq 0} (\lambda e^t)^n / n!$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$$M'(t) = \lambda e^t e^{\lambda(e^t - 1)} \quad \text{by chain rule}$$

$$M''(t) = (\lambda e^t)^2 e^{\lambda(e^t - 1)} + \lambda e^t e^{\lambda(e^t - 1)} \quad \text{by chain rule and product rule}$$

$$\text{Hence } E(X) = M'(0) = \lambda$$

$$E(X^2) = M''(0) = \lambda^2 + \lambda$$

$$\text{So } \text{Var}(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

as proved earlier

Exponential (X)

Let X be exponential with parameter λ .

Then

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-(\lambda-t)x} dx = \frac{\lambda}{\lambda-t} \quad \text{for } t < \lambda \end{aligned}$$

$$M'(t) = \lambda/(\lambda-t)^2$$

$$M''(t) = 2\lambda/(\lambda-t)^3$$

$$\text{Thus } E(X) = M'(0) = 1/\lambda$$

$$E(X^2) = M''(0) = 2/\lambda^2$$

$$\text{So } \text{Var}(X) = E(X^2) - (E(X))^2 = 1/\lambda^2$$

Standard normal

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2 - 2tx)/2} dx \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2 + t^2/2} dx$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

change
of
variables

$$= e^{t^2/2} \quad (\text{since integral of normal density } \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \text{ is equal to } 1)$$

Now $M'(t) = te^{t^2/2}$ by product rule

$M''(t) = e^{t^2/2} + t^2 e^{t^2/2}$ by product, chain rules

Thus $M'(0) = 0$, $M''(0) = 1$,

so $E(X) = 0$, $\text{Var}(X) = 1$.

Remarks

i) If X, Y are independent, with moment generating functions $M_X(t)$, $M_Y(t)$, then

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

Proof

$$M_{X+Y}(t) = E[e^{t(X+Y)}]$$

$$= E[e^{tX} e^{tY}]$$

by independence $\rightarrow = E[e^{tX}] E[e^{tY}] = M_X(t) M_Y(t)$

2) If $M_X(t)$ exists and is finite in some region about $t=0$, then the distribution of X is uniquely determined.

Example

Calculate distribution of $X+Y$, where X, Y are independent Poissons with parameters λ_1, λ_2 .

Sol.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \\ &= e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

Now this is the generating function for a Poisson

with parameter $\lambda_1 + \lambda_2$, so $X+Y$ is
Poisson ($\lambda_1 + \lambda_2$).

268

3) Can use moment generating functions to prove
central limit theorem.

Chapter 8 : Limit theorems

Recall inequalities proved earlier:

- Markov's inequality

If X is a random variable that takes
only non-negative values, then for any $a > 0$,
 $P(X \geq a) \leq E[X]/a$.

- Chebyshev's inequality

If X is a random variable with finite
mean μ and variance σ^2 , then for any $k > 0$,

$$P(|X - \mu| \geq k) \leq \sigma^2/k^2$$

Weak law of large numbers

Theorem Let $X_1, X_2, \dots, X_n$ be independent 269
with same distribution. Suppose $\mu = E(X_j)$ and
 $\sigma^2 = \text{Var}(X_j)$ are both finite. Let $S_n = X_1 + \dots + X_n$
then for any $\epsilon > 0$, $P(|\frac{S_n}{n} - \mu| < \epsilon) \rightarrow 1$ as $n \rightarrow \infty$

proof

Since $X_1, \dots, X_n$ are independent with the same
distribution, $\text{Var}(S_n) = n\sigma^2$, so $\text{Var}(S_n/n) = \sigma^2/n$

Also, $E(S_n/n) = \mu$.

By Chebyshev's inequality, for any $\epsilon > 0$,

$$\begin{aligned} P(|\frac{S_n}{n} - \mu| \geq \epsilon) &\leq \frac{\text{Var}(S_n/n)}{\epsilon^2} \\ &= \sigma^2/n\epsilon^2 \end{aligned}$$

For ϵ fixed, this $\rightarrow 0$ as $n \rightarrow \infty$.

Remark The strong law of large numbers says
that with probability 1,

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mu \text{ as } n \rightarrow \infty$$

The strong law implies the weak law.

Example (uniform case)

Suppose choose independently at random n numbers from uniform distribution on $[0, 1]$. Then if X_i denotes i th choice, we have

$$\mu = E[X_i] = \int_0^1 x dx = 1/2$$

$$\sigma^2 = \text{Var}(X_i) = \int_0^1 x^2 dx - \mu^2 = 1/3 - 1/4 = 1/12$$

$$\text{Hence } E\left(\frac{S_n}{n}\right) = 1/2, \quad \text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{12n}$$

So for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{12n\epsilon^2}$$

Example (normal case)

Suppose choose n real numbers independently at random from a standard normal distribution. Call them $X_1, \dots, X_n$.

$$\text{Then } \mu = E(X_i) = 0, \quad \sigma^2 = \text{Var}(X_i) = 1$$

271

Hence $E\left(\frac{S_n}{n}\right) = 0$, $\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n}$.

Chebyshev's inequality gives that for any

$$\epsilon > 0, \quad P\left(\left|\frac{S_n}{n} - 0\right| > \epsilon\right) \leq \frac{1}{n\epsilon^2}$$

Table of Chebyshev estimates

<u>n</u>	<u>$P(S_n/n \geq .1)$</u>	<u>Chebyshev bound</u>
100	.317	1.000
200	.157	.500
300	.083	.3333
400	.045	.2500
500	.025	.1667

Central limit theorem

Bernoulli trials

Let S_n = number of successes in n independent trials with probability p of success on each trial. Let $a < b$ be two fixed real numbers. Define

$$a^* = \frac{a - np}{\sqrt{np(1-p)}}, \quad b^* = \frac{b - np}{\sqrt{np(1-p)}}$$

Then

$$\lim_{n \rightarrow \infty} P(a \leq S_n \leq b) = \int_{a^*}^{b^*} \phi(x) dx$$

$\phi(x)$ = normal density function

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Note A slightly more accurate approximation, if i, j are integers, is given by

$$P(i \leq S_n \leq j) \approx \int_{i^*}^{j^*} \phi(x) dx$$

$$\text{where } i^* = \frac{i - 1/2 - np}{\sqrt{np(1-p)}}, \quad j^* = \frac{j + 1/2 - np}{\sqrt{np(1-p)}}$$

Example A coin is tossed 100 times. Estimate probability that the number of heads is between 40 and 60 inclusive.

Sol. Let S_n = number of heads, $n = 100$

$$E(S_n) = 100(1/2) = 50, \quad SD(S_n) = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$$

273

Since $n = 100$ is large,

$$P(40 \leq S_n \leq 60) \approx P\left(\frac{39.5 - 50}{5} \leq \frac{S_n - 50}{5} \leq \frac{60.5 - 50}{5}\right)$$

$$= P(-2.1 \leq \frac{S_n - 50}{5} \leq 2.1)$$

$$\approx \int_{-2.1}^{2.1} \phi(x) dx$$

$$\approx .9642$$

The actual value is .96480 --

Example

Dartmouth college would like to have 1050 freshmen. The college can not accommodate more than 1060. Assume each applicant accepts with probability .6, and that the acceptances can be modeled by independent Bernoulli trials. If college accepts 1700, what is the probability that too many students accept?

Sol. If college accepts 1700 students, the expected number of students who accept is

$.6(1700) = 1020$, and the standard ²⁷⁴
deviation for number who accept is

$\sqrt{1700(.6)(.4)} \approx 20$. So we estimate

$$P(S_{1700} > 1060) = P(S_{1700} \geq 1061)$$

$$\approx P\left(\frac{S_{1700} - 1020}{20} \geq \frac{1060.5 - 1020}{20}\right)$$

$$= P\left(\frac{S_{1700} - 1020}{20} \geq 2.025\right)$$

$$\approx \int_{2.025}^{\infty} \phi(x) dx \approx .0216$$

More general central limit theorem

Let $S_n = X_1 + X_2 + \dots + X_n$ be sum of n
independent discrete random variables, with
common distribution having expected value μ
and variance σ^2 . Then for $a < b$,

$$\lim_{n \rightarrow \infty} P\left(a < \frac{S_n - n\mu}{\sqrt{n}\sigma} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

275
Example A die is rolled 420 times.

What is the probability that the sum of the rolls lies between 1400 and 1550 inclusive?

Sol. The sum is a random variable

$$S_{420} = X_1 + X_2 + \dots + X_{420}$$

where the X_i 's are independent with distribution

1	2	3	4	5	6
$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

One calculates that $\mu = E(X_i) = 7/2$
 $\sigma^2 = \text{Var}(X_i) = 35/12$

$$\text{Thus } E(S_{420}) = 420 (7/2) = 1470$$

$$\sigma^2(S_{420}) = 420 (35/12) = 1225,$$

$$\text{so } \sigma(S_{420}) = 35.$$

Thus

$$P(1400 \leq S_{420} \leq 1550)$$

$$\text{continuity correction} \rightarrow \approx P\left(\frac{1395.5 - 1470}{35} \leq \frac{S_{420} - 1470}{35} \leq \frac{1550.5 - 1470}{35}\right)$$

$$= P(-2.01 \leq \frac{5420 - 1470}{35} \leq 2.30) \quad 276$$

$$\approx \int_{-2.01}^{2.30} \phi(x) = .9670$$

Example

A student's GPA is average of grades in 30 courses. Grades are based on 100 possible points and are recorded as integers. Assume in each course, instructor makes an error of grading of k with probability $|p/k|$ where $k = \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$. The probability of no error is then $1 - \frac{137}{30}p$. (The parameter p represents the inaccuracy of the instructor's grading).

There are two average grades for the student: the average of the correct grades and average of the recorded grades.

We wish to estimate the probability that

277

these two average grades differ by less than .05 for a given student. Now assume that $p = 1/20$. Also assume that the total error is the sum S_{30} of 30 independent random variables each with distribution

	-5	-4	-3	-2	-1	0	1	2	3	4	5
X	$\frac{1}{100}$	$\frac{1}{80}$	$\frac{1}{60}$	$\frac{1}{40}$	$\frac{1}{20}$	$\frac{463}{600}$	$\frac{1}{20}$	$\frac{1}{40}$	$\frac{1}{60}$	$\frac{1}{80}$	$\frac{1}{100}$

Can easily calculate $E(X) = 0$, $\sigma^2(X) = 1.5$

Then we have

$$P(-.05 \leq \frac{S_{30}}{30} \leq .05)$$

$$= P(-1.5 \leq S_{30} \leq 1.5)$$

$$= P\left(\frac{-1.5}{\sqrt{30(1.5)}} \leq \frac{S_{30}}{\sqrt{30(1.5)}} \leq \frac{1.5}{\sqrt{30(1.5)}}\right)$$

$$= P(-.224 \leq \frac{S_{30}}{\sqrt{30(1.5)}} \leq .224)$$

$$\approx \int_{-.224}^{.224} \phi(x) dx \approx .1772$$

So there is only a 17.7% chance 278
that a student's GPA is accurate to
within .05 (out of 100).

More general central limit theorem

Let $X_1, \dots, X_n$ be a sequence of independent
discrete random variables, and let

$$S_n = X_1 + \dots + X_n$$

Define the mean and variance of S_n to
be μ_n and σ_n^2 respectively, and assume that
 $\sigma_n \rightarrow \infty$. If there exists a constant A
such that $|X_n| \leq A$ for all n , then for $a \leq b$,

$$\lim_{n \rightarrow \infty} P(a < \frac{S_n - \mu_n}{\sigma_n} < b)$$

$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

CLT for continuous random variables

Let $S_n = X_1 + X_2 + \dots + X_n$ be sum of
 n independent continuous random variables
with common density function f having expected

value μ and variance σ^2 . Then for all $a < b$,

$$\lim_{n \rightarrow \infty} P\left(a < \frac{S_n - n\mu}{\sqrt{n}\sigma} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Other inequalities

One-sided Chebyshev inequality

Theorem If X is a random variable with mean 0 and finite variance σ^2 , then for any $a > 0$,

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Proof

$$P(X \geq a) = P(X+b \geq a+b) \quad \text{any } b$$

$$\leq P((X+b)^2 \geq (a+b)^2)$$

$$\begin{aligned} \text{Markov's inequality} \rightarrow & \leq E[(X+b)^2] / (a+b)^2 \\ & = (\sigma^2 + b^2) / (a+b)^2 \end{aligned}$$

Calculus gives that the b minimizing this

expression is $b = 6^2 a$, and the result follows.

280

Example

The number of items produced in a factory during one week is a random variable with mean 100 and variance 400. Upper bound the probability that this week's production will be at least 120.

Sol.

$$P(X \geq 120) = P(X - 100 \geq 20)$$

one sided
Chebyshev, noting
that $\text{Var}(X - 100)$
 $= \text{Var}(X)$

$$\leq \frac{400}{400 + (20)^2} = 1/2$$

Remark: Applying Markov's inequality gives

$$P(X \geq 120) \leq \frac{E(X)}{120} = 5/6,$$

a weaker bound.

Remark Suppose X has mean μ , variance σ^2 .

281

Since both $X - \mu$, $\mu - X$ have mean 0 and variance σ^2 , it follows from 1-sided Chebyshev inequality that for $a > 0$,

$$P(X - \mu \geq a) \leq \sigma^2 / (\sigma^2 + a^2)$$

$$P(\mu - X \geq a) \leq \sigma^2 / (\sigma^2 + a^2)$$

Thus $P(X \geq \mu + a)$ and $P(X \leq \mu - a)$ are both at most $\sigma^2 / (\sigma^2 + a^2)$

Chebyshev bound

Theorem Given a random variable X , let $M(t) = E[e^{tX}]$.

$$1) P(X \geq a) \leq e^{-ta} M(t) \text{ for all } t > 0$$

$$2) P(X \leq a) \leq e^{-ta} M(t) \text{ for all } t < 0$$

Proof (of part 1)

Markov's inequality

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq E[e^{tX}] e^{-ta}$$

since
 $t > 0$

$$= e^{-ta} M(t)$$

Example

282

Let Z be a standard normal.

Then $M(t) = e^{t^2/2}$ by an earlier calculation

Thus $P(Z \geq a) \leq e^{-ta} e^{t^2/2}$ for all $t > 0$

By calculus, the value of t minimizing

$t^2/2 - ta$ is $t = a$. Then

$$P(Z \geq a) \leq e^{-a^2/2}$$

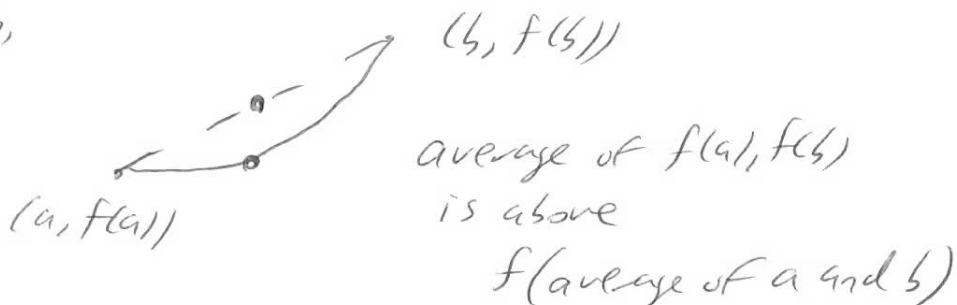
Jensen's inequality

Def. A twice differentiable real valued function $f(x)$ is called convex if $f''(x) \geq 0$.

(For example, $f(x) = x^2$ is convex.)

Prop. If $f(x)$ is convex, $E[f(X)] \geq f(E[X])$

Intuitively,



Proof

Taylor series expansion of f about $\mu = E(X)$ gives

$$f(x) = f(\mu) + f'(\mu)(x-\mu) + f''(\xi)(x-\mu)^2/2$$

where ξ is between x, μ .

Since $f''(\xi) \geq 0$,

$$f(x) \geq f(\mu) + f'(\mu)(x-\mu) \text{ all } x.$$

thus

$$f(X) \geq f(\mu) + f'(\mu)(X-\mu)$$

Taking expectations gives $E[f(X)] \geq f(\mu)$,
as claimed.

This completes all material that will
be on the final (Chapters 1-8).

Next topic : cycles of random permutations

284

$$\text{Let } \pi = \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 2 & 3 & 6 & 5 & 1 & 8 & 7 \end{array}$$

↗
2-line form of π



write this as $(146)(2)(3)(5)(78)$

Let $N_i(\pi)$ = number of cycles of π of length i

So in example,

$$N_1(\pi) = 3, N_2(\pi) = 1, N_3(\pi) = 1$$

Let S_n = all permutations on n symbols

$$S_n \quad |S_n| = n!$$

Problem: Study cycles of a random element of S_n . For example, we know that $E(N_1) = 1$, and will see that N_1 is asymptotically Poisson(1) as $n \rightarrow \infty$.

MATH 407, Lecture 20

285

Recall, S_n = permutations on n symbols

Prop.

The number of $\pi \in S_n$ with $\begin{cases} n_1 & 1\text{-cycles} \\ n_2 & 2\text{-cycles} \\ & \text{etc.} \end{cases}$

$$= \begin{cases} 0 & \text{if } \sum n_i \neq n \\ \frac{n!}{\prod n_i! (n_i)!} & \text{if } \sum n_i = n \end{cases}$$

Proof

Clear if $\sum n_i \neq n$.

So assume $\sum n_i = n$. Define a map

$\phi: S_n \longrightarrow$ permutations with $\begin{cases} n_1 & 1\text{-cycles} \\ n_2 & 2\text{-cycles} \\ & \text{etc.} \end{cases}$

We define ϕ "by example".

Suppose $n=9$, $n_1=2$, $n_2=2$, $n_3=1$

4 1 9 8 3 5 6 7 2 $\rightarrow$ (4)(1)(98)(35)(672)

first 1-cycles, next 2-cycles, then 3-cycles, etc.

Now each π with $\begin{cases} n_1 \text{ 1-cycles} \\ n_2 \text{ 2-cycles} \\ \text{etc.} \end{cases}$

286

is hit exactly $\prod_i n_i (n_i)!$ many times.

Thus the number of $\pi \in S_n$ with $\begin{cases} n_1 \text{ 1-cycles} \\ n_2 \text{ 2-cycles} \\ \text{etc.} \end{cases}$

is $\frac{n!}{\prod_i n_i (n_i)!}$ as needed.

Theorem (cycle index)

$$1 + \sum_{n \geq 1} \frac{u^n}{n!} \sum_{\pi \in S_n} \prod_i x_i^{n_i(\pi)} = e^{\sum_{m \geq 1} x_m u^m / m}$$

Proof

Take coefficient of $u^n \prod_i x_i^{n_i}$ on both sides.

Case 1 $\sum n_i \neq n$.

Then get 0 on both sides

Case 2 $\sum n_i = n$

$$\text{left hand side} = \frac{1}{n!} \left(\begin{array}{l} \text{number of } \pi \in S_n \\ \text{with } n_1 \text{ 1-cycles} \\ n_2 \text{ 2-cycles} \\ \text{etc.} \end{array} \right)$$

by $\rightarrow$ Theorem $= \frac{1}{n!} \frac{n!}{\prod i^{n_i} (n_i)!} = \frac{1}{\prod i^{n_i} (n_i)!}$

Right hand side $\rightarrow$ $= \frac{1}{\prod i^{n_i} (n_i)!}$
by Taylor expansion of e^z

Example (fixed points)

- 1) Compute the number of ΠES_n with no fixed points. (We did this earlier using inclusion-exclusion)

Sol. In the cycle index, set $x_1 = 0$ and all other $x_i = 1$.

We conclude that

$$1 + \sum_{n \geq 1} \frac{u^n}{n!} \left(\begin{array}{l} \text{number of } \Pi ES_n \\ \text{with no fixed} \\ \text{points} \end{array} \right) = e^{\sum_{m \geq 2} u^m / m}$$

$$= \frac{1}{e^u} e^{\sum_{m \geq 1} u^m / m}$$

$$= \frac{1}{e^u} e^{-\log(1-u)}$$

since
 $\log(1-u)$

$$= -u - u^2/2 - u^3/3 - \dots$$

$$= \frac{1}{(1-u)e^u}$$

So the number of $\pi \in S_n$ with no fixed points

$$= n! \text{ coeff. } u^n \text{ in } \frac{1}{(1-u)e^u}$$

$$= n! \text{ coeff. } u^n \text{ in } (1+u+u^2+\dots)/e^u$$

$$= n! \sum_k \text{coeff. } u^k \text{ in } e^{-u}$$

$$= n! (1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!})$$

So the proportion of $\pi \in S_n$ with no fixed points is $1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}$

This converges to $1/e$ as $n \rightarrow \infty$

2) let's compute $E[\text{number of fixed points}]$

of a random $\pi \in S_n$. we did this 289
earlier in the course by other methods

Sol.

In the cycle index, set $x_i = x$ and all
other $x_i = 1$. we conclude that

$$\begin{aligned}
 1 + \sum_{n \geq 1} \frac{u^n}{n!} \sum_{\pi \in S_n} x^{N_1(\pi)} &= e^{xu} \prod_{i \geq 2} e^{u/i} \\
 &= \frac{e^{xu}}{e^u} \prod_{i \geq 1} e^{u/i} \\
 &= \frac{e^{xu}}{(1-u)e^u} \quad \begin{array}{l} \text{Since} \\ \log(1-u) \\ = -u - u^2/2 - u^3/3 \dots \end{array}
 \end{aligned}$$

Now differentiate with respect to
 x and set $x=1$.

$$\text{Left hand side} = \sum_{n \geq 1} \sum_{\pi \in S_n} \frac{u^n}{n!} N_1(\pi)$$

So coefficient of u^n on left hand
side = $E[N_1]$

$$\begin{array}{l} \text{Right hand} \\ \text{side} \end{array} = \frac{u}{1-u} = u + u^2 + u^3 + \dots$$

290

So for all $n \geq 1$, coefficient of u^n on right hand side = 1

We conclude that if $n \geq 1$, then

$$E[N_1(\pi)] = 1$$

π
random in S_n

3) Claim

As $n \rightarrow \infty$, the distribution of number of fixed points of a random $\pi \in S_n$ converges to a Poisson(1) limit.

Proof Suppose Z is Poisson(1). Then for all positive integers k ,

$$E(Z(Z-1)\dots(Z-k+1)) = 0$$

as you can prove using moment generating functions.

From above, we know that

$$1 + \sum_{n \geq 1} \frac{u^n}{n!} \sum_{\pi \in S_n} x^{N_1(\pi)} = \frac{e^{xu}}{e^u(1-u)}$$

Differentiate k times with respect to x and then set $x=1$. We get

$$\begin{aligned} \sum_{n \geq 1} u^n E[N_1(\pi)(N_1(\pi)-1) \cdots (N_1(\pi)-k+1)] \\ = u^k / (1-u). \end{aligned}$$

So if $n \geq k$, and π is random in S_n ,

$$E[N_1(N_1-1) \cdots (N_1-k+1)] = 1,$$

as can be seen by taking coefficients of u^n .

The "method of moments" implies that

$$N_1 \rightarrow \text{Poisson}(1) \text{ as } n \rightarrow \infty$$

Remark More generally, for i fixed,

$$N_i \rightarrow \text{Poisson}(1/i) \text{ as } n \rightarrow \infty$$

Example (Number of cycles)

Set all $x_i = x$ in the cycle index.

Let $c(\pi)$ = number of cycles of π

We obtain

$$\begin{aligned}
 1 + \sum_{n \geq 1} \frac{u^n}{n!} \sum_{\pi \in S_n} x^{c(\pi)} &= e^{\sum_{m \geq 1} x u^m / m} \\
 &= \left(e^{\sum_{m \geq 1} u^m / m} \right)^x \\
 &= \left(e^{-\log(1-u)} \right)^x \\
 &= \left(\frac{1}{1-u} \right)^x
 \end{aligned}$$

Take coefficient of u^n on both sides.

On left, get $\frac{1}{n!} \sum_{\pi \in S_n} x^{c(\pi)}$

On right, get $\frac{x(x+1) \cdots (x+n-1)}{n!}$

Since if $f(u) = \sum_n c_n u^n$, then

$$c_n = \frac{1}{n!} f^{(n)}(0)$$

Thus, $\frac{1}{n!} \sum_{\pi \in S_n} x^{c(\pi)} = \frac{x(x+1) \cdots (x+n-1)}{n!}$

Can use this to prove that $C(\pi)$ is asymptotically normal with mean $1 + \frac{1}{2} + \dots + \frac{1}{n}$ and known variance.

Cycles and records

We define a bijection $\phi: S_n \rightarrow S_n$.

Given π in cycle form, start each cycle with its largest value, and order the cycles in increasing order of the largest value in each cycle. Finally, remove the parentheses.

Example

$$\phi = (4)(1)(98)(35)(672)$$

$$\rightarrow (1)(4)(53)(726)(98)$$

$$\rightarrow 145372698$$

Note

1) ϕ is a bijection, since can invert

ϕ by inserting parentheses where you have left-to-right records.

$$2) \quad C(\pi) = \begin{array}{c} \uparrow \\ \text{number of} \\ \text{cycles of } \pi \end{array} \text{ number of left-to-right records of } \phi(\pi).$$

The number of records of a random permutation is of interest in sports.

We conclude that

$$E \left[\begin{array}{c} \text{number of records} \\ \text{of } \pi \in S_n \end{array} \right] = 1 + \frac{1}{2} + \dots + \frac{1}{n}$$

$$\sim \log(n).$$

Next topic = Markov chain theory

First, we review some basic linear algebra

Def. An $m \times n$ matrix is a rectangular array of numbers arranged in m horizontal

rows and n vertical columns.

For example $A = \begin{bmatrix} 3/2 & 7 & 1/3 \\ 0 & -2 & 6 \end{bmatrix}$ is a 2×3

matrix, and

$B = \begin{bmatrix} 4 & 1 & 8 \\ 2 & 3 & -1 \\ 0 & -6 & 2 \end{bmatrix}$ is a 3×3

matrix. Given a matrix A , let a_{ij} denote the entry in i th row and j th column of A .

For example, $a_{12} = 7$ and $a_{23} = 6$

def. If A, B are both $m \times n$ matrices, then define $A+B$ to be matrix obtained by adding the corresponding elements of A and B .

For example,

$$\begin{bmatrix} 2 & 1 \\ 4 & 5 \\ 3 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 6 \\ 0 & 2 \\ 4 & -3 \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 4 & 7 \\ 7 & -3 \end{bmatrix}$$

Next goal: define matrix multiplication

Given an n row vector and an n column

vector, define

$$[a_1 \ a_2 \ \dots \ a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \dots + a_n b_n$$

This is also denoted $a \cdot b$, and is called the dot product of a and b . For example,

$$\begin{aligned} [2 \ 4 \ 1 \ -3] \begin{bmatrix} 1 \\ -2 \\ 3 \\ 1 \end{bmatrix} &= 2(1) + 4(-2) + 1(3) + (-3)(1) \\ &= -6 \end{aligned}$$

Def. We define the product of an $m \times n$ matrix A and an $n \times p$ matrix B to be the $m \times p$ matrix whose i, j entry is the dot product of row i of A with column j of B .

We denote this product matrix by AB .

For example, if $A = \begin{bmatrix} 1 & 4 & 3 \\ 0 & 2 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 10 & 1 \\ 0 & 3 \\ 2 & 6 \end{bmatrix}$

then

$$AB = \begin{bmatrix} 1(10) + 4(0) + 3(2) & 1(1) + 4(3) + 3(6) \\ 0(10) + 2(0) + 6(2) & 0(1) + 2(3) + 6(6) \end{bmatrix}$$

297

$$= \begin{bmatrix} 16 & 31 \\ 12 & 42 \end{bmatrix}$$

Note If A is an $m \times n$ matrix, and B is an $n \times p$ matrix, then the $m \times p$ matrix $C = AB$ satisfies

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

Def. Let I_n be the $n \times n$ matrix with 1's on the diagonal and 0's elsewhere.

If A, B ~~are~~ $n \times n$ matrices such that $AB = I_n$ and $BA = I_n$, we say that B is the inverse of A and write $B = A^{-1}$.

For example, $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and if

$$A = \begin{bmatrix} \frac{1}{2} & 0 & \frac{2}{3} \\ 1 & 0 & -1 \end{bmatrix}, \text{ then } A^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ -\frac{5}{3} & 1 & -\frac{4}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{bmatrix}$$

Markov chains

Have a set of states $S = \{s_1, \dots, s_r\}$

298

The process starts in one of these states and moves successively from one state to another. Each move is called a step. If chain is currently in state s_i , then it moves to state s_j at next step with probability p_{ij} , and this probability does not depend on which states the chain was in before the current state. The probabilities p_{ij} are called the transition probabilities of the Markov chain. The process can remain in the state it is in, and this occurs with probability p_{ii} . An initial probability distribution, defined on S , specifies the starting state. Usually, this is done by specifying a particular state as the starting state.

Example

Pittsburgh is blessed by many things, 299
 but not good weather. One never has 2
 nice days in a row. If have a nice day,
 we are equally likely to have snow or rain
 the next day. If have snow or rain, then
 have $1/2$ chance of same the next day. If
 there is a change from snow or rain, only
 $1/2$ the time is this a change to a nice day.

Summarizing,

$$P = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{matrix}$$

Powers of transition matrix P

We will see that the ij entry of P^n
 gives the probability that the Markov chain
 started in i th state will be in j th state
 after n steps.

MATH 407 Lecture 2/

300

Recall Pittsburgh weather Markov chain (R =rain, N =nice, S =snow)
with transition matrix

$$P = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{matrix}$$

Let $p_{ij}^{(2)}$ = probability that if chain is in state i today, that it will be in j th state in 2 days.

For example, suppose it's raining today. Then the chance it's snowy 2 days from now can be computed by adding the probabilities of 3 events:

- 1) Rain tomorrow and snow in 2 days
(probability $p_{11}p_{13}$)

2) nice tomorrow and snow in
2 days (probability $p_{12}p_{23}$)

3) snow tomorrow and snow in 2
days (probability $p_{13}p_{33}$)

Thus,
$$p_{13}^{(2)} = p_{11}p_{13} + p_{12}p_{23} + p_{13}p_{33}$$

$$= \text{the } 1,3 \text{ entry of } P^2.$$

more generally,

Theorem Let P be the transition matrix of a Markov chain. Then the ij entry of P^n gives the probability that the Markov chain started in i th state will be in j th state after n steps.

Proof Use induction.

In the Pittsburgh example, the first 302
6 powers of P are:

$$P^1 = \begin{pmatrix} .500 & .250 & .250 \\ .500 & 0 & .500 \\ .250 & .250 & .500 \end{pmatrix}$$

$$P^2 = \begin{pmatrix} .438 & .188 & .375 \\ .375 & .250 & .375 \\ .375 & .188 & .438 \end{pmatrix}$$

$$P^3 = \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix}$$

$$P^4 = \begin{pmatrix} .402 & .199 & .398 \\ .398 & .203 & .398 \\ .398 & .199 & .402 \end{pmatrix}$$

$$P^5 = \begin{pmatrix} .400 & .200 & .399 \\ .400 & .199 & .400 \\ .399 & .200 & .400 \end{pmatrix}$$

$$P^6 = \begin{pmatrix} .400 & .200 & .400 \\ .400 & .200 & .400 \\ .400 & .200 & .400 \end{pmatrix}$$

So after 6 days, our weather predictions are, to 3 decimal places, independent of today's weather.

A probability ~~with~~ vector with r components is a row vector whose entries are non-negative and sum to 1. If u is a probability vector which

303
represents the starting state of a Markov chain, its i th component is the chance that the chain starts in the i th state.

Theorem let P be transition matrix of a Markov chain, and let u be the probability vector which represents the starting distribution. Then the probability that the chain is in the j th state after n steps is the j th entry of the vector uP^n .

Proof (Exercise)

Note: If Markov chain starts in i th state, we simply choose u to be the vector with i th entry $= 1$ and all other entries $= 0$.

Example In Pittsburgh example, let the initial probability vector u be $(1/3, 1/3, 1/3)$. Then we can calculate the distribution of

304

states after 3 days as:

$$uP^3 = \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right) \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix}$$

$$= (.401 \quad .198 \quad .401)$$

Example (Ehrenfest model)

The following model has been used to study the diffusion of gases. Have 2 urns which between them contain 4 balls. At each step, one of the 4 balls is chosen at random and moved from the urn it is in to the other urn. We choose as states the number of balls in the first urn. The transition matrix is then

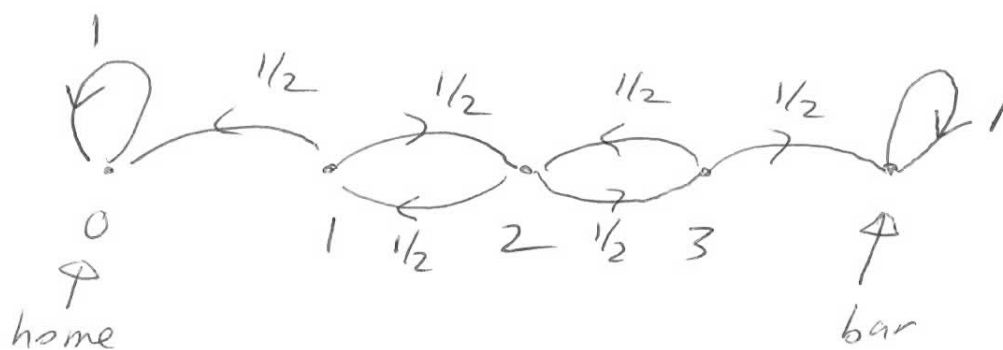
$$\begin{matrix} & 0 & 1 & 2 & 3 & 4 \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

Absorbing Markov chains

305

def. A state s_i of a Markov chain is called absorbing if it is impossible to leave it (i.e. $p_{ii} = 1$). A Markov chain is called absorbing if it has at least one absorbing state, and if from every state it is possible to get to an absorbing state (not necessarily in 1 step). In an absorbing Markov chain, a state which is not absorbing is called transient.

Example : Drunkard's walk



A man walks on a 4-block stretch of

Main street. If he is at corner 1, 2, or 3, he walks to the left or right with equal probability. He continues until he reaches corner 4 (a bar) or corner 0 (his home). If he reaches either home or the bar, he stays there.

Clearly, this Markov chain is absorbing. States 0, 4 are absorbing. The transition matrix is

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Some questions about absorbing chains

- What is the probability that the process will end in a given absorbing state?
- On average, how long will it take for the process to be absorbed?

c) On average, how many times will ³⁰⁷
the process be in each transient state?

The answers depend both on the starting state and on the transition matrix P .

Canonical Form

Consider an absorbing Markov chain.
Reorder the states so that the transient states come first. If there are r absorbing states and t transient states, the transition matrix will have canonical form

$$P = \begin{matrix} & \begin{matrix} \text{Tr} & \text{Abs} \end{matrix} \\ \begin{matrix} \text{Tr} \\ \text{Abs} \end{matrix} & \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \end{matrix}$$

Here, $I = r \times r$ identity matrix

$0 = r \times t$ zero matrix

$R = \text{non-zero } t \times r$ matrix

$Q = t \times t$ matrix

Thus,

$$P^n = \begin{pmatrix} Q^n & * \\ 0 & I \end{pmatrix}$$

some trsr
matrix

308

Theorem In an absorbing Markov chain,
the probability of being absorbed is 1
(that is, $Q^n \rightarrow 0$ as $n \rightarrow \infty$).

Pf. Intuitively obvious

Theorem For an absorbing Markov chain, the
matrix $I - Q$ has an inverse N , and
 $N = I + Q + Q^2 + \dots$. The ij entry n_{ij} of N
is the expected number of times the chain is
in state s_j , given that it starts in state
 s_i . The initial state is counted if $i=j$.

Proof

From linear algebra, if $(I - Q)x = 0$ implies
that $x = 0$, then $I - Q$ is invertible. So suppose
that $(I - Q)x = 0$. Then $x = Qx$, so $x = Q^n x$.

309

Since $Q^n \rightarrow 0$, we have $Q^n x \rightarrow 0$, so $x = 0$.

Thus $(I - Q)^{-1} = N$ exists.

Next note that

$$(I - Q)(I + Q + \dots + Q^n) = I - Q^{n+1}$$

multiplying both sides by N gives

$$I + Q + \dots + Q^n = N(I - Q^{n+1})$$

Letting $n \rightarrow \infty$ gives $N = I + Q + Q^2 + \dots$

Let s_i, s_j be two transient states, and assume for rest of proof that i, j are fixed.

Let $X^{(k)} = \begin{cases} 1 & \text{if chain is in state } s_j \\ & \text{after } k \text{ steps} \\ 0 & \text{else} \end{cases}$

Thus $P(X^{(k)} = 1) = q_{ij}^{(k)}$ $\swarrow$ ij entry of Q^k

$$P(X^{(k)} = 0) = 1 - q_{ij}^{(k)}$$

These hold for $k=0$ since $Q^0 = I$.

Since $X^{(k)}$ is always 0 or 1, $E(X^{(k)}) = q_{ij}^{(k)}$

The expected number of times the chain is in state s_j in the first n steps, given that it started in state s_i is clearly

$$E(X^{(0)} + X^{(1)} + \dots + X^{(n)}) \\ = q_{ij}^{(0)} + q_{ij}^{(1)} + \dots + q_{ij}^{(n)}$$

Letting $n \rightarrow \infty$, we have

$$E(X^{(0)} + X^{(1)} + \dots) = q_{ij}^{(0)} + q_{ij}^{(1)} + \dots = n_{ij}$$

Def. For an absorbing Markov chain with transition matrix P , the matrix $N = (I - Q)^{-1}$ is called the fundamental matrix of P . The entry n_{ij} of N is the expected number of times the process is in transient state s_j , if it is started in transient state s_i .

Example

In Drunkard's walk, P has canonical form

$$P = \begin{array}{c} 1 \quad 2 \quad 3 \quad 0 \quad 4 \\ \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

311

Thus, $Q = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}$ and $I - Q = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}$

Computing $(I - Q)^{-1}$, we find

$$N = (I - Q)^{-1} = \begin{array}{c} 1 \quad 2 \quad 3 \\ \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array}$$

From middle row of N , we see that if chain is started in state 2, the expected numbers of times in states 1, 2, 3 before being absorbed are 1, 2, 1 respectively.

Time to absorption

Consider the question: given that the chain starts in state s_i , what is the expected number of steps before the chain is absorbed?

Theorem

312

Let t_i = expected number of steps before chain is absorbed, given that it starts in transient state s_i .

Let t = column vector whose i th entry is t_i .

Then $t = Nc$, where c is column vector all of whose entries are 1.

Proof If add all the entries in i th row of N , we get the expected number of times in any of the transient states for a given starting state s_i . This is the expected time until absorption. Thus t_i is the sum of entries in i th row of N , proving the result.

In the drunkard's walk example, we found

$$N = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{matrix}$$

$$\text{Thus } t = Nc = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ = \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix}$$

thus starting in states 1, 2, 3, the expected times to absorption are 3, 4, 3 respectively.

Absorption probabilities

Theorem

Let b_{ij} = probability an absorbing chain will be absorbed in state s_j if it starts in transient state s_i .

Let B = matrix with ij entry b_{ij}

Then $B \equiv \underset{\substack{\uparrow \\ \text{fundamental} \\ \text{matrix}}}{N} R \leftarrow \begin{matrix} \text{matrix from} \\ \text{canonical form of } P \end{matrix}$

Proof Let s_k be the final transient state the chain is in before transitioning to the absorbing state s_j .

Then

314

$$B_{ij} = \sum_n \sum_k q_{ik}^{(n)} r_{kj}$$

$$= \sum_k \sum_n q_{ik}^{(n)} r_{kj}$$

$$= \sum_k n_{ik} r_{kj}$$

$$= ij \text{ entry of } NR$$

Example For Drunkard's walk, $R = \begin{matrix} & 0 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix} \end{matrix}$

$$\text{So } B = NR = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$= \begin{matrix} & 0 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{matrix}$$

So the first row tells us that starting from state 1, there is probability $3/4$ of absorption in state 0, and probability $1/4$ of absorption in state 4. The second row tells us (as is obvious from symmetry), that starting from state 2, one is equally likely to be absorbed at state 0 and at state 4.

Markov chains, continued

Def. A Markov chain is irreducible if can go from every state to every state (not necessarily in 1 step).

Def. A Markov chain is regular if some power of its transition matrix has all entries positive.

Clearly every regular matrix is irreducible. But not every irreducible matrix is regular, as following example shows

Example Let the transition matrix of a Markov chain on $\{1, 2\}$ be

$$P = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The chain is clearly irreducible, but if n is odd, can't go from 1 to 1 in steps. And

if n is even, can't go from 1 to 2
in n steps. So chain isn't regular. 316

Example Recall Ehrenfest urn model. This
has transition matrix

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

Starting at 0, we'll be at state 0, 2, or 4
after an even number of steps, and at state
1 or 3 after an odd number of steps. So the
chain is irreducible, but not regular.

Regular Markov chains

Any transition matrix with no zeros
determines a regular Markov chain. But a
regular transition matrix could have zeros.

For example the Pittsburgh weather matrix has

$P_{N,N} = 0$, but P^2 has no zeros, so the Markov

chain is regular.

317

Theorem

Let P be the transition matrix of a regular Markov chain. Then as $n \rightarrow \infty$, the powers P^n approach a limiting matrix W with all rows the same vector w . The vector w is a strictly positive probability vector (i.e. the components of w are all positive and sum to 1).

Idea of proof

want to show that P^n tends to a matrix with all rows equal. This is the same as showing that P^n converges to a matrix with constant columns. Now the j th column of P^n is $P^n y$ where y is a column vector with 1 in j th entry and 0 in other entries. So it suffices to prove that for any column vector y , $P^n y$ approaches a constant vector as $n \rightarrow \infty$.

Since each row of P is a probability vector,

P_y replaces y by averages of its components. For example,

$$\begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{4}(3) \\ \frac{1}{3}(1) + \frac{1}{3}(2) + \frac{1}{3}(3) \\ \frac{1}{2}(1) + \frac{1}{2}(2) + 0(3) \end{pmatrix} \\ = \begin{pmatrix} 7/4 \\ 2 \\ 3/2 \end{pmatrix}$$

The result of averaging is to make the components of P_y more similar than those of y . In particular, the maximum component decreases (from 3 to 2) and the minimum component increases (from 1 to $3/2$). As we do more and more of this averaging to get $P^n y$, the difference between the maximum and minimum component will tend to 0 as $n \rightarrow \infty$. Thus $P^n y$ tends to a constant vector, as claimed.

Note The ij entry of P^n is the chance of being in state j after n steps started from state i . Denoting the common row of W by w , the theorem says that the probability of being in j th state will be approximately w_j , the j th entry of w , and is independent of starting state

Example In the Pittsburgh weather example, recall that to 3 decimal places,

$$P^6 = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{pmatrix} .4 & .2 & .4 \\ .4 & .2 & .4 \\ .4 & .2 & .4 \end{pmatrix} \end{matrix}$$

Thus to this degree of accuracy,

$$P(\text{rain 6 days after a rainy day})$$

$$= P(\text{rain 6 days after a nice day})$$

$$= P(\text{rain 6 days after a snowy day}) = .4$$

The previous theorem gives that for 320
large n , the rows of P^n approach a common
vector. It's interesting that this happens
so soon in our example.

Theorem Let P be a regular transition matrix.
Let $W = \lim_{n \rightarrow \infty} P^n$, and let w be the common
row of W . Let c be the column vector all
of whose components are 1. Then

- a) $wP = w$, and any row vector v such that
 $vP = v$ is a constant multiple of w
- b) $Pc = c$, and any column vector x such that
 $Px = x$ is a multiple of c .

Proof

a) Since $P^n \rightarrow W$, $P^{n+1} = P^n P \rightarrow WP$. But

$P^{n+1} \rightarrow W$. Thus $W = WP$, so $w = wP$. We
don't prove claim about v being a multiple of w .

b) can check that $Pc=c$. Don't prove ³²¹
the claim about x being a multiple of c .

Example (of calculating w for a regular chain)

Recall Pittsburgh weather example. By theorem, can find the limiting vector w from equations,

$$w_1 + w_2 + w_3 = 1 \quad \text{and} \quad (w_1 \ w_2 \ w_3) \begin{pmatrix} \cancel{1/2} & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} = (w_1 \ w_2 \ w_3)$$

we get 4 equations in 3 unknowns:

$$w_1 + w_2 + w_3 = 1$$

$$1/2 w_1 + 1/2 w_2 + 1/4 w_3 = w_1$$

$$1/4 w_1 + 1/4 w_3 = w_2$$

$$1/4 w_1 + 1/2 w_2 + 1/2 w_3 = w_3$$

The theorem guarantees a unique solution.

Solving, we get $w = (.4, .2, .4)$, in agreement

with that predicted from P^6 .

322

Remark Can also set $w_1 = 1$, then solve all but 1 of the linear equations from $wP = w$. Can obtain w from this solution by dividing each component by their sum.

In example, set $w_1 = 1$.

Solve first and second equations from $wP = w$.

$$1/2 + 1/2 w_2 + 1/4 w_3 = 1$$

$$1/4 + 1/4 w_3 = w_2$$

Solving gives $(1, 1/2, 1)$. Dividing this vector by sum of its components gives $w = (.4, .2, .4)$

New topic : Riffle shuffling

These results made it to front page of the New York Times

Let n = size of deck

Sequential description of a 2-shuffle

Choose k according to the binomial distribution $\binom{n}{k}/2^n$. Cut the top k cards from the deck, yielding packets of size k and $n-k$. When you have A_1 cards in packet 1 and A_2 cards in packet 2, drop the next card from packet 1 with probability A_1/A_1+A_2 , and from packet 2 with probability A_2/A_1+A_2 . Continue until all cards are dropped.

Geometric description

Place n points uniformly and independently in $[0,1]$. Points are labeled in order $x_1 < x_2 < \dots < x_n$. The map $x \rightarrow 2x \bmod 1$ rearranges these points, giving a random TTE S_n .

More generally, define an a -shuffle, where a is a positive integer.

324

sequential description of an a -shuffle

Choose integers $k_1, k_2, \dots, k_a$ according to the multinomial distribution $\binom{n}{k_1, \dots, k_a} / a^n$.

Given k_1 , cut off the top k_1 cards, the next k_2 cards, and so on, producing a packets (some of them possibly empty). When you have A_1 cards in packet 1, $\dots$, A_a cards in packet a , the next card is dropped from packet i with probability $A_i / (A_1 + \dots + A_a)$. Continue until all cards are dropped.

Geometric description

Use $x \rightarrow ax \pmod{1}$ instead of $x \rightarrow 2x \pmod{1}$.

Theorem The sequential and geometric descriptions are equivalent.

Def. Say that a permutation π has a descent

at position i , if $\pi(i) > \pi(i+1)$, $1 \leq i \leq n-1$. 325

For example,

5 1 9 8 6 7 3 4 2

has 5 descents. Let $d(\pi) =$ number of descents of π

Theorem

The chance that an a -shuffle of a deck of n cards yields a permutation π is

$$\binom{n+a-d(\pi^{-1})-1}{n} / a^n$$

Theorem Performing r 2-shuffles is the same as performing a single 2^r shuffle.

Proof (sketch)

Clear from geometric description of riffle shuffling.

want to study how many 2-shuffles are

needed to mix a deck of n cards. 326

To quantify this, use notion of total variation distance.

Let P, Q be probability distributions on a finite set X . Define the total variation distance between P and Q by

$$\|P - Q\|_{TV} = \frac{1}{2} \sum_{x \in X} |P(x) - Q(x)|$$

Remarks

1) $0 \leq \|P - Q\|_{TV} \leq 1.$

2) $\|P - Q\|_{TV} = \max_{A \subseteq X} |P(A) - Q(A)|$

where $P(A) = \sum_{a \in A} P(a)$

$$Q(A) = \sum_{a \in A} Q(a)$$

So if $\|P - Q\|_{TV}$ is small, the probability of event A under P is close to probability of event A under Q , for any A .

So by our earlier results, the total variation distance after r 2-shuffles to the uniform distribution is equal to

$$\begin{aligned}
 & \frac{1}{2} \sum_{\pi} \left| \frac{\binom{n+2^r-d(\pi)-1}{n}}{2^{rn}} - \frac{1}{n!} \right| \\
 &= \frac{1}{2} \sum_{\pi} \left| \frac{\binom{n+2^r-d(\pi)-1}{n}}{2^{rn}} - \frac{1}{n!} \right| \\
 &= \frac{1}{2} \sum_{i=0}^{n-1} A(n,i) \left| \frac{\binom{n+2^r-i-1}{n}}{2^{rn}} - \frac{1}{n!} \right|
 \end{aligned}$$

where $A(n,i)$ = number of permutations on n symbols with i descents

since we have the recursion

$$A(n,k) = (n-k+1)A(n-1,k-1) + kA(n-1,k)$$

we can quickly calculate the $A(n,k)$, so the total variation distance. For example, when $n=52$, we get the table

	328									
	1	2	3	4	5	6	7	8	9	10
total variation to uniform after r 2-shuffles of 52 cards	1.000	1.000	1.000	1.000	.924	.614	.334	.167	.085	.043
	hardly random at all!				starts to fall by a factor of 2 after each shuffle					

Using asymptotics of the numbers $A(n, k)$, one can prove the following result.

Theorem For $r = \log_2(n^{3/2}c)$, $0 < c < \infty$ fixed, as $n \rightarrow \infty$, the total variation distance after r 2-shuffles to uniform converges to

$$1 - 2\Phi(-1/4\sqrt{c})$$

$$\text{with } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

For c large, this is close to 0.

Remarks

a) There is a nice theory of cycle

structure after an a -shuffle,
 corresponding to the theory for uniform
 permutations (which we studied earlier)
 as $a \rightarrow \infty$. For example,

329

1) The expected number of fixed points after
 an a -shuffle is $1 + \frac{1}{a} + \frac{1}{a^2} + \dots + \frac{1}{a^{n-1}}$.

This is n if $a=1$, and 1 if $a=\infty$

2) After an a -shuffle, the chance that it
 has n_1 fixed points, n_2 2-cycles, etc. is

$$\frac{1}{a^n} \prod_{i=1}^n \binom{f_i(a) + n_i - 1}{n_i} \quad \text{with}$$

$$f_i(a) = \frac{1}{i} \sum_{d|i} \mu(d) a^{i/d} \quad \text{where } \mu \text{ is the}$$

Mobius function of number theory

$$\mu(d) = \begin{cases} 1 & \text{if } d=1 \\ (-1)^j & \text{if } d \text{ is product of } j \text{ distinct primes} \\ 0 & \text{else} \end{cases}$$

MATH 407 lecture 23

Last time we gave a formula for the chance of given cycle type after an a -shuffle.

If $a = q$ is a prime power, the chance of n_1 one-cycles, n_2 two-cycles, etc. is equal to the chance that a random monic degree n polynomial with coefficients in a finite field F_q factors into n_1 factors of degree 1, n_2 factors of degree 2, etc.

These results show that functions of π depending only on small cycles (such as fixed points), have the right distribution (approximately) after $c(n)$ steps, where $c(n) \rightarrow \infty$ arbitrarily slowly.

Functions which depend on large cycles
(such as length of longest cycle) have the
right distribution after 1 shuffle (for
 n large).

b) The transition matrix of a 2-shuffle
of n cards has nice eigenvalues, namely

$$1, 1/2, 1/4, \dots, 1/2^{n-1}$$

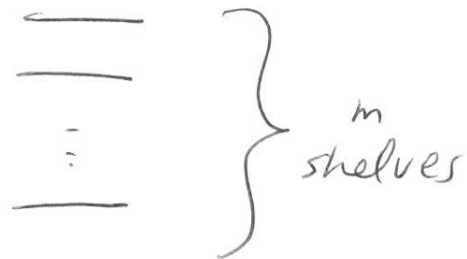
where $1/2^i$ has multiplicity = number of
permutations on
 n symbols with
 $n-i$ many cycles

It's remarkable that this
matrix has real eigenvalues, since it's not
symmetric.

Analysis of casino shelf-shuffling machines

332

1) cards are dealt sequentially from bottom of deck onto random shelves



2) each card is placed randomly above or below the previous cards on the shelf with probability $1/2$.

3) at end the piles are assembled into 1 pile, with cards on top shelf on top, followed by cards on 2nd shelf, etc.

Question: For 52 cards, is 10 shelves enough to produce a well-shuffled deck?

Answer: No

Note If there is only 1 shelf, the resulting permutation is unimodal.

Here unimodal means that π has 0 valleys. We call i a valley of π if $1 < i < n$ and $\pi(i-1) > \pi(i) < \pi(i+1)$.

333
1
2
5
6
8
7
4
3

← a unimodal permutation

So $\pi = \underline{5} 1 3 6 7 \underline{2} 4$ has two valleys.

Theorem Let $v(\pi)$ = number of valleys of π .

The chance that a shelf shuffler with m shelves and n cards produces π is

$$P_m(v) = \frac{4^{v(\pi)+1}}{2(2m)^n} \sum_{a=0}^{m-1} \binom{n+m-a-1}{n} \binom{n-1-2v(\pi)}{a-v(\pi)}$$

Example If have $m=1$ shelf, and $P_m(v) \neq 0$

must have $a = v(\pi) = 0$. Then

$$P(\pi) = \frac{4}{2 \cdot 2^n} \binom{n-1}{0} = \frac{1}{2^{n-1}},$$

and we get the uniform distribution on the set of 2^{n-1} unimodal permutations.

Remark: unimodal permutations arise
in social choice theory, dynamical
systems, and magic tricks

Corollary of theorem The total variation distance
of a permutation with n cards produced by
a shelf-shuffler with m shelves to the
uniform distribution is

$$\|P_m - U\|_{TV} = \frac{1}{2} \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} v(n, k) |P_m(k) - \frac{1}{n!}|$$

where $v(n, k)$ = number of permutations on n
symbols with k valleys

we have the recursion $v(1, 0) = 1$

$$v(n, k) = (2k+2) v(n-1, k) + (n-2k) v(n-1, k-1)$$

Let's use this to compute $\|P_m - U\|_{TV}$ for $n=52$ cards

m	10	15	20	25	30	35	50	100	150	200	250
$\ P_m - U\ _{TV}$	1	.943	.72	.544	.391	.299	.159	.041	.018	.010	.007

Thus 10 shelves are very insufficient
to mix 52 cards.

Distribution of cycle type

mobius function
defined earlier

Theorem Let $f_{i,m} = \frac{1}{2i} \sum_{\substack{d|i \\ d \text{ odd}}} \mu(d) (2m)^{i/d}$

Let $P_m(\pi)$ be the probability that a shelf
shuffler with m shelves produces π . Then

$$1 + \sum_{n \geq 1} u^n \sum_{\pi \in S_n} P_m(\pi) \prod_{i \geq 1} x_i^{N_i(\pi)} \quad \text{is equal to}$$

number of i -cycles of π

$$\prod_{j \geq 1} \left(\frac{1 + x_j (u/2m)^j}{1 - x_j (u/2m)^j} \right)^{f_{j,m}}$$

Iterated shuffling

Theorem Using a shelf shuffler with m_1
shelves and then a shelf shuffler with m_2
shelves is same as using a shelf shuffler with $2m_1 m_2$

shelves.

336

Corollary Using a 10 shelf shuffler twice is same as using a 200 shelf shuffler once.

The total variation distance to uniform for a 200 shelf shuffler with 52 cards is, by our earlier table, .010. So using a 10 shelf shuffle twice does adequately mix 52 cards.

Card guessing with feedback

A deck of cards is dealt face up, one card at a time, onto a table. Before each card is shown, a guess is made as to the value of the card. Let X_i ($1 \leq i \leq n$) be 1 if i th guess is correct, 0 else. Then

$$T_n = X_1 + \dots + X_n$$

is the total number of correct guesses.

337

If cards are perfectly mixed, the best strategy is to guess a card you haven't seen yet. then $P(X_1=1) = 1/n$, $P(X_2=1) = 1/(n-1)$, etc. and the X_i 's are independent.

Prop. For a perfectly mixed deck,

$$1) E(T_n) = \frac{1}{n} + \frac{1}{n-1} + \dots + 1$$
$$\sim \log(n) + \gamma + \text{term going to } 0 \text{ as } n \rightarrow \infty$$

where $\gamma = .577$ (Euler's constant)

$$2) \text{Var}(T_n) = \frac{1}{n} (1 - \frac{1}{n}) + \frac{1}{n-1} (1 - \frac{1}{n-1}) + \dots$$
$$\sim \log(n) + \gamma - \frac{\pi^2}{6} + \text{term going to } 0 \text{ as } n \rightarrow \infty$$

3) Normalized by its mean and variance,

T_n is approximately normal

Thus with $n=52$, T_n has mean approximately

338

4.5, standard deviation approximately

$\sqrt{2.9}$, and T_n is between 2.7 and 6.3 roughly 70% of the time.

For a shelf shuffler with m shelves, there is a conjectured optimal guessing strategy.

Let Z_n be the number of correct guesses with this strategy. Then

m	1	2	4	10	20	64
$E(Z_n)$	3.9	2.7	17.6	9.3	6.2	4.7
$\text{Var}(Z_n)$	3.2	5.6	6.0	4.7	3.8	3.1

thus for the actual shelf-shuffler (with $m=10$ shelves), you get about 9.3 correct guesses versus 4.5 for a well-shuffled deck.

Markov chains and linear algebra

Have a Markov chain with a finite state space X . Let $K(x, y)$ = chance go from x to y in one step.

(earlier we denoted this $P(x, y)$)

def. let π be a probability distribution on X .

we say that π is stationary for the Markov chain K if

$$\pi(y) = \sum_x \pi(x) K(x, y) \text{ for all } y.$$

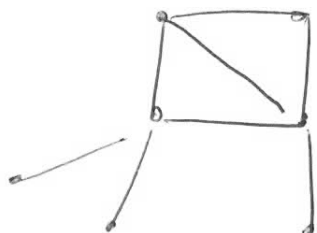
(we will see that under mild conditions, running the chain for a long time gets you close to the stationary distribution π on X).

Def. Let π be a probability distribution on X

we say that K is reversible with respect to π if

$$\pi(x) K(x, y) = \pi(y) K(y, x) \text{ for all } x, y$$

Example Random walk on a graph.



From a given vertex, move to a random neighbor.

So $X =$ vertices of the graph. For a vertex v ,

let $\deg(v) =$ "degree of v "

$=$ number of neighbors of v

Claim Let $\pi(v) = \deg(v) / 2|E|$, where $E =$ edges of graph

$$a) \sum_v \pi(v) = 1$$

b) K is reversible with respect to π .

Proof

a) count the number of ordered pairs (x, y)
where x, y are adjacent vertices.

on one hand,

the number of pairs is $\sum_x \deg(x)$ 341

on other hand, the number of pairs is $2|E|$.

The result follows.

b) Case 1 x, y aren't adjacent.

Then $\pi(x)K(x, y) = 0$ and $\pi(y)K(y, x) = 0$

$$\text{So } \pi(x)K(x, y) = \pi(y)K(y, x)$$

Case 2 x, y are adjacent

$$\text{Then } \pi(x)K(x, y) = \frac{\deg(x)}{2|E|} \frac{1}{\deg(x)} = \frac{1}{2|E|}$$

$$\pi(y)K(y, x) = \frac{\deg(y)}{2|E|} \frac{1}{\deg(y)} = \frac{1}{2|E|}$$

So $\pi(x)K(x, y) = \pi(y)K(y, x)$, completing the proof.

Prop. If K is reversible with respect to π ,
then π is stationary for K .

Proof

342

$$\sum_x \pi(x) K(x, y) \stackrel{\text{by reversibility}}{=} \sum_x \pi(y) K(y, x)$$

$$= \pi(y) \sum_x K(y, x) = \pi(y)$$

Recall, the total variation distance between probability distributions P, Q on a finite set X is

$$\|P - Q\|_{TV} = \frac{1}{2} \sum_x |P(x) - Q(x)| = \max_{A \subseteq X} |P(A) - Q(A)|$$

exercise

Take $P = K_x^r$, $Q = \pi$

φ
probability distribution
after r steps started
from x .

stationary
distribution
(if it's unique)

GOAL Study decay of $\|K_x^r - \pi\|_{TV}$ as

a function of r .

343

Example of a theorem in the subject

Consider "random transposition" shuffle on n cards. At each step, pick 2 cards at random (possibly the same card), then swap them

Theorem Let $r = \frac{1}{2} n \log(n) + cn$

1) For $c > 0$, $\|K^r - U\|_{TV} \leq Ae^{-2c}$ ↙ uniform distribution on S_n

For a universal constant A

2) For $c < 0$, $\|K^r - U\|_{TV} \geq \frac{1}{e} - e^{-e^{-2c}}$ + term going to 0 as $n \rightarrow \infty$

Note

1) Starting state is irrelevant in this example

2) Theorem says " $\frac{1}{2} n \log(n) + cn$ steps are necessary and sufficient for randomness"

Cauchy - Schwarz inequality says that 344

for real vectors u, v ,

$$u \cdot v \leq |u| |v|$$

$\uparrow$ $\nwarrow$
 $\sum u_i v_i$ $\sqrt{\sum u_i^2 v_i^2}$

True since $|u \cdot v| = |u| |v| \cos(\theta)$ where θ is angle between u, v in plane spanned by u and v .

thus,

$$\begin{aligned}
 2 \|K_X^v - \pi\|_{TV} &= \sum_y |K_X^v(y) - \pi(y)| \\
 &= \sum_y \left| \frac{K_X^v(y)}{\sqrt{\pi(y)}} - \sqrt{\pi(y)} \right| \sqrt{\pi(y)} \\
 &\leq \sqrt{\sum_y \left[\frac{K_X^v(y)}{\sqrt{\pi(y)}} - \sqrt{\pi(y)} \right]^2} \sum_y [\sqrt{\pi(y)}]^2 \\
 &= \sqrt{\sum_y \left(\frac{K_X^v(y)}{\pi(y)} - 1 \right)^2 \pi(y)}.
 \end{aligned}$$

$\nrightarrow$

we will bound this using linear algebra

From last time,

$$\|K_X^r - \pi\|_{TV}^2 \leq \boxed{\frac{1}{4} \sum_y \left(\frac{K_X^r(y)}{\pi(y)} - 1 \right)^2 \pi(y)}$$

✦

will bound using linear algebra (eigenvalues and eigenvectors)

For a Markov chain K on a finite state space X , define

$$V = \{f: X \rightarrow \mathbb{R}\} = \text{real valued functions on } X$$

For $f, g \in V$, define "inner product" with respect to π by $\langle f, g \rangle = \sum_x f(x)g(x)\pi(x)$
a probability distribution on X

Given $f \in V$, define $Kf \in V$ by

$$Kf(x) = \sum_y K(x, y) f(y)$$

Lemma Suppose K is reversible with respect to π . Then $\langle Kf, g \rangle = \langle f, Kg \rangle$ for all f, g .

Proof

$$\langle Kf, g \rangle = \sum_x Kf(x) g(x) \pi(x) = \sum_x g(x) \pi(x) \sum_y K(x, y) f(y)$$

$$\langle f, Kg \rangle = \sum_y f(y) Kg(y) \pi(y) = \sum_y f(y) \pi(y) \sum_x K(y, x) g(x)$$

By reversibility, $\pi(x) K(x, y) = \pi(y) K(y, x)$,

so we are done.

Def. We say that a function $\phi \in V$ is an eigenvector of K with eigenvalue β if $K\phi = \beta\phi$.

Def. A set of functions $f_1, f_2, \dots, f_{|X|}$ in V is called a basis of V , if every

347
element of V can be uniquely
written as $c_1 f_1 + \dots + c_{|X|} f_{|X|}$ with
 $c_1, \dots, c_{|X|}$ real numbers.

Cor. Suppose K is reversible with respect
to Π . Then (by linear algebra and previous
lemma), V has an orthonormal basis of
eigenvectors of K , with real eigenvalues.
That is to say, there is a basis $\phi_0, \dots, \phi_{|X|-1}$
of V with $K\phi_i = \beta_i \phi_i$ where β_i are real
and
$$\langle \phi_i, \phi_j \rangle = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$$

Remark Corollary follows from fact that a
symmetric real matrix has an orthonormal
basis of real eigenvectors with real eigenvalues.
(A matrix M is symmetric if $m_{ij} = m_{ji}$ all i, j .)

So $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 5 & 4 \\ 2 & 4 & 7 \end{bmatrix}$ is symmetric,

Theorem Let K be reversible with respect to π . Let ϕ_i, β_i be as in previous corollary. So $K\phi_i = \beta_i\phi_i$ with β_i real. Order the eigenvalues

$$\beta_0 \geq \beta_1 \geq \dots \geq \beta_{|X|-1}$$

and define $\beta_* = \max(|\beta_1|, |\beta_{|X|-1}|)$

Suppose $\pi(x) \neq 0$ for all x . Then

$$1) |\beta_i| \leq 1 \text{ for all } i, \beta_0 = 1$$

$$2) \sum_{i=0}^{|X|-1} \phi_i^2(x) = \frac{1}{\pi(x)} \text{ all } x.$$

$$3) K_x^v(y) = \sum_{i=0}^{|X|-1} \beta_i^v \phi_i(x) \phi_i(y) \pi(y)$$

$$4) \sum_y \left(\frac{K_x^v(y)}{\pi(y)} - 1 \right)^2 \pi(y) = \sum_{i=1}^{|X|-1} \phi_i(x)^2 \beta_i^{2v}$$

$$\leq \frac{1 - \pi(x)}{\pi(x)} \beta_*^{2v} \leq \frac{\beta_*^{2v}}{\pi(x)}$$

Proof

1) Let ϕ be an eigenvector of K with eigenvalue

β . Choose x^* so that $|\phi(x^*)| \geq |\phi(y)|$ for all y . Then

$$\beta \phi(x^*) = K\phi(x^*) = \sum_y K(x^*, y) \phi(y)$$

$$\Rightarrow |\beta| |\phi(x^*)| \leq |\phi(x^*)| \sum_y |K(x^*, y)|$$

$$= |\phi(x^*)|$$

$$\Rightarrow |\beta| \leq 1$$

To see that $\beta_0 = 1$, note that $\phi = 1$ for all x is an eigenvector of K with eigenvalue 1.

2) Define $d_x(y) = \begin{cases} 1/\pi(x) & \text{if } y=x \\ 0 & \text{else} \end{cases}$

Define d_x in basis of orthonormal eigenvectors.

$$d_x(y) = \sum_{i=0}^{|x|-1} \phi_i(y) \langle d_x, \phi_i \rangle$$

just as $(3, 1, 2)$ in $\mathbb{R}^3$ = $\langle (3, 1, 2), (1, 0, 0) \rangle (1, 0, 0)$
 $+ \langle (3, 1, 2), (0, 1, 0) \rangle (0, 1, 0)$
 $+ \langle (3, 1, 2), (0, 0, 1) \rangle (0, 0, 1)$

using the inner product in $\mathbb{R}^3$

$$\langle (a, b, c), (d, e, f) \rangle = ad + be + cf.$$

$$\text{Thus } d_x(y) = \sum_{i=0}^{|X|-1} \phi_i(y) \phi_i(x)$$

$$\text{Setting } y=x \text{ gives } \frac{1}{\pi(x)} = \sum_{i=0}^{|X|-1} \phi_i^2(x)$$

3) Can prove by induction that

$$K_X^r(y) = K^r(\delta_y)(x)$$

$$\delta_y = \begin{cases} 1 & \text{at } y \\ 0 & \text{else} \end{cases}$$

$$\text{Claim } \delta_y = \pi(y) \sum_{i=0}^{|X|-1} \phi_i(y) \phi_i$$

$$\text{Indeed, } \delta_y = \sum_i \phi_i \frac{\langle \delta_y, \phi_i \rangle}{\pi(y)}$$

$$= \sum_x \delta_y(x) \phi_i(x) \pi(x)$$

$$= \phi_i(y) \pi(y)$$

Now the claim implies

$$K^r(\delta_y) = \pi(y) \sum_{i=0}^{|X|-1} \beta_i^r \phi_i(y) \phi_i$$

35/

Thus

$$K^r(\sigma_y)(x) = \pi(y) \sum_{i=0}^{|x|-1} \beta_i^r \phi_i(y) \phi_i(x)$$

$$\Rightarrow K_x^r(y) = K^r(\sigma_y)(x) = \sum_{i=0}^{|x|-1} \beta_i^r \phi_i(x) \phi_i(y) \pi(y)$$

$$4) \sum_y \left[\frac{K_x^r(y)}{\pi(y)} - 1 \right]^2 \pi(y)$$

by parts 1,3

$$\nearrow = \sum_y \left[\sum_{i=1}^{|x|-1} \beta_i^r \phi_i(x) \phi_i(y) \right]^2 \pi(y)$$

$$= \sum_y \sum_{i=1}^{|x|-1} \beta_i^{2r} \phi_i^2(x) \phi_i^2(y) \pi(y)$$

since
 $\langle \phi_i, \phi_j \rangle$
 $= \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$

$$= \sum_{i=1}^{|x|-1} \beta_i^{2r} \phi_i^2(x)$$

$$\leq \beta_*^{2r} \sum_{i=1}^{|x|-1} \phi_i^2(x)$$

by parts 1,2

$$\nearrow = \beta_*^{2r} \left[\frac{1}{\pi(x)} - 1 \right] \leq \beta_*^{2r} / \pi(x)$$

End of proof

REMEMBER $4 \|K_X^r - \pi\|_{TV}^2 \leq \beta_*^{2r} / \pi(x).$

Example (where can calculate everything)

Let $X = \mathbb{Z}_2^n = n$ -tuples of 0's, 1's
 = vertices of n -dimensional hypercube

$$K(x, y) = \begin{cases} 1/(n+1) & \text{if } x+y = 0, e_1, \dots, e_n \\ 0 & \text{else} \end{cases}$$

where addition
is mod 2 and

$$e_i = (0, \dots, 0, 1, 0, \dots, 0)$$

$\uparrow$

Equivalently, from x , stay
 at x with probability $1/(n+1)$
 and move to a random neighbor
 with probability $1/(n+1)$.

1 at position i
 and 0 else

Can easily check that K is reversible
 with respect to π , where $\pi(x) = 1/2^n$ for all x .

Theorem

353

A basis of orthonormal eigenvectors of K is given by

$$\phi_x(y) = (-1)^{x \cdot y} \quad \text{where } x \cdot y = \sum_{i=1}^n x_i y_i$$

and $x \in \mathbb{R}_2^n$.

$$K \phi_x = \left(1 - \frac{2|x|}{n+1}\right) \phi_x \quad \text{where } |x| = \text{number of 1's in } x$$

Proof

$$K \phi_x(z) = \sum_y K(z, y) \phi_x(y)$$

$$= \sum_y K(z, y) (-1)^{x \cdot y}$$

$$= \frac{1}{n+1} \left((-1)^{x \cdot z} + (-1)^{x \cdot (z + e_1)} + \dots + (-1)^{x \cdot (z + e_n)} \right)$$

$$= \frac{1}{n+1} \phi_x(z) \left(1 + (-1)^{x_1} + (-1)^{x_2} + \dots + (-1)^{x_n} \right)$$

$$= \left(1 - \frac{2|x|}{n+1}\right) \phi_x(z)$$

We omit the proof of orthonormality

end of proof

Thus the eigenvalues of K are

354

$(1 - \frac{2j}{n+1})$ with multiplicity $\binom{n}{j}$, $0 \leq j \leq n$.

$$\text{Thus } 4 \|K_X^r - \pi\|_{TV}^2 \leq \sum_{i=1}^{|X|-1} \phi_i(x)^2 \beta_i^{2r}$$

Let's start at $x = (0, 0, \dots, 0)$. Then $\phi_i(x) = 1$ all i .

Thus

$$4 \|K_X^r - \pi\|_{TV}^2 \leq \sum_{i=1}^n \binom{n}{i} \left(1 - \frac{2i}{n+1}\right)^{2r}$$

using $\leq \sum_{i=1}^{\infty} \frac{n^i}{i!} e^{-2i/(n+1) 2r}$

$$\binom{n}{i} \leq n^i / i!$$

and $1-x \leq e^{-x}$
for $x > 0$

$$\text{Taking } r = \frac{1}{4} (n+1) (\log(n) + 2c)$$

$$\leq \sum_{i=1}^{\infty} \frac{1}{i!} e^{i \log(n) - i (\log(n) + 2c)}$$

$$= \sum_{i=1}^{\infty} e^{-2ic} / i!$$

$$= e e^{-2c} - 1$$

thus, if $r = \frac{1}{4} (n+1) (\log(n) + 2c)$,

$$\|K_X^r - \pi\|_{TV} \leq \frac{1}{2} \sqrt{e^{e^{-2c}} - 1} \quad 355$$

$$\approx \frac{1}{2} e^{-c} \text{ for } c \text{ large}$$

Remark This bound used all the eigenvalues and eigenvectors. If we just use bound

$$\|K_X^r - \pi\|_{TV}^2 \leq \frac{1}{4\pi(X)} \beta_*^{2r}$$

with $\beta_* = (1 - 2/(n+1))$, $\pi(X) = 1/2^n$, then

$$\text{we get } \frac{2^n}{4} \left(1 - \frac{2}{n+1}\right)^{2r} = \frac{1}{4} \left[e^{n \log 2 + 2r \log \left(1 - \frac{2}{n+1}\right)} \right]$$

Since $\log(1-y) = -y - y^2/2 - y^3/3 \dots$,
this expression is small for r of order n^2 .

Note

1) Not all Markov chains are reversible.

For example, top to random shuffle isn't reversible. But by a miracle, it does have real eigenvalues.

2) Call a Markov chain on the set $\{0, \dots, k\}$ a birth-death chain if

$$K(i, j) > 0 \text{ when } |i - j| = 1$$

$$\text{and } K(i, j) = 0 \text{ when } |i - j| \geq 2.$$

Exercise: Find a distribution π with respect to which the birth-death chain is reversible.

Theorem Let K be a reversible Markov chain, with eigenvalues

$$1 = \beta_0 \geq \beta_1 \geq \dots \geq \beta_{|X|-1} \geq -1.$$

a) If K is irreducible, then $\beta_1 < 1$.

b) Call K aperiodic if all states i have period 1, where period of state i is $\gcd \{n \geq 1 \text{ such that } K^n(i, i) > 0\}$.

So random walk on  is not aperiodic.

If K is aperiodic, then $\beta_{|X|-1} > -1$.

c) If K is irreducible and aperiodic, then it has a unique stationary distribution

π , and $\|K_X^r - \pi\|_{TV} \rightarrow 0$ as $r \rightarrow \infty$

Proof

a) Suppose K is irreducible. Suppose that $K\phi = \phi$.

We need to conclude that ϕ is constant.

There exists x^* such that $|\phi(x^*)| \geq |\phi(y)|$ for all y .

$$\text{Now } \phi(x^*) = K\phi(x^*) = \sum_y K(x^*, y) \phi(y)$$

It follows that $\phi(y) = \phi(x^*)$ for all y such

that $K(x^*, y) > 0$. Since the chain is irreducible, we conclude that ϕ is constant.

b) If -1 is an eigenvalue, we have ϕ with

$K\phi = -\phi$. As in part a), choose x^* so that

$|\phi(x^*)| \geq |\phi(y)|$ for all y . So it can get

to y in 1 step from x^* , then $\phi(y) = -\phi(x^*)$. More generally, if you can get to a state y from x in m steps, then $\phi(y) = (-1)^m \phi(x)$. It follows that the state x^* has even period, so if -1 is an eigenvalue, the chain is periodic.

c) If K is irreducible and aperiodic, then by parts a), b), $\beta_1 < 1$ and $\beta_{|X|-1} > -1$. So

$$\beta_* = \max(|\beta_1|, |\beta_{|X|-1}|) < 1.$$

$$\text{Since } 4 \|K_x^r - \pi\|_{TV}^2 \leq \beta_*^{2r} / \pi(x)$$

it follows that $\lim_{r \rightarrow \infty} \|K_x^r - \pi\|_{TV} = 0$.

Remark

part c) of the theorem is true even if K is not reversible.

Metropolis
algorithm

— an application of Markov chains

359

The metropolis algorithm is widely regarded as the most important algorithm of the 20th century.

Have a finite set X and a probability distribution π on X . The Metropolis algorithm is a procedure for drawing samples from X , approximately distributed according to π .

The algorithm requires an irreducible, aperiodic Markov chain K on X . K need not be reversible, but we must have

$$K(x, y) > 0 \iff K(y, x) > 0.$$

The chain K is modified by using coin tossing to create a new Markov chain M which has stationary distribution π . So if you run M long enough, you get approximate samples from π .

-To be continued

Metropolis algorithm continued

Recall that have a distribution π on a finite state space X , and we want to draw approximate samples from π .

Have an irreducible, aperiodic Markov chain on X , which need not be reversible but which satisfies

$$K(x, y) > 0 \iff K(y, x) > 0.$$

We use K to create a new Markov chain M on X which has stationary distribution π .

We define M in words as follows.

If chain is currently at x , choose y with probability $K(x, y)$. Define acceptance ratio

$$\text{by } A(x, y) = \frac{\pi(y) K(y, x)}{\pi(x) K(x, y)}$$

361

If $A(x, y) \geq 1$, the chain moves to y .

If $A(x, y) < 1$, flip a coin with heads probability $A(x, y)$. If coin comes up heads, move to y . Otherwise, stay at x .

Formally,

$$M(x, y) = \begin{cases} K(x, y) & \text{if } A(x, y) \geq 1 \text{ and } y \neq x \\ K(x, y) A(x, y) & \text{if } A(x, y) < 1 \\ K(x, y) + \sum_{\substack{z \text{ such} \\ \text{that } A(x, z) < 1}} K(x, z) (1 - A(x, z)) & \text{if } y = x \end{cases}$$

Lemma The Markov chain M is irreducible, aperiodic, and reversible with respect to π (that is $\pi(x)M(x, y) = \pi(y)M(y, x)$ for all x, y).

Thus by our earlier theorems,

$$\lim_{r \rightarrow \infty} \|M_x^r - \pi\|_{TV} = 0.$$

Proof Exercise

Remark In applications X is often a huge set, and the distribution π is given as

$$\pi(x) = \frac{1}{Z} e^{-H(x)}$$

Often $H(x)$ is easy to calculate, but the normalizing constant $Z = \sum_x e^{-H(x)}$ is impossible to compute. Note that Z cancels out of the acceptance ratio $A(x, y)$, so the Markov chain is easy to run.

Example the symmetric group S_n = all permutations on n symbols

For statistical purposes, it's useful to study the distribution on permutations which chooses a permutation σ with probability

$$\pi_\theta(\sigma) = \theta^{d(\sigma)} / Z.$$

363

Here $0 < \theta \leq 1$ is fixed, $d(\sigma) = \sum_i |\sigma(i) - i|$

and Z is an intractable normalizing constant ($Z = \sum_{\sigma} \theta^{d(\sigma)}$). For large n

(say $n=52$), Z is hard to compute, so we sample from π using the Metropolis algorithm with the chain K being the "random transposition" chain on S_n .

Thus, if chain is currently at σ , the Metropolis chain proceeds by choosing i, j at random in $\{1, \dots, n\}$ and transposing (formally, $\sigma' = (ij)\sigma$).

If $d(\sigma') \leq d(\sigma)$, the chain moves to σ' .

If $d(\sigma') > d(\sigma)$, a coin is flipped with heads probability $\theta^{d(\sigma') - d(\sigma)}$. If heads, the chain moves to σ' . Otherwise it stays at σ .

364

Note : The convergence rate of this Markov chain has not been rigorously analyzed.

Example : Ising model from statistical physics

Let Λ be a finite, connected subset of the lattice $\mathbb{Z}^2$. Let X = labelings of vertices of Λ by $+1, -1$.

Ising model choose a labeling x with probability

$$\pi(x) = \frac{1}{Z} \exp \left[\beta \left(\sum_{(i,j)} x_i x_j + h \sum_i x_i \right) \right]$$

$\nearrow$
intractable
normalizing
constant

— first sum is over
neighboring pairs with
 $i, j \in \Lambda$

— $\beta > 0$ is "inverse temperature"

— h is "external field strength"

Physicists want to draw samples from π

and see what they "look like". Often 365
 Λ is a square grid of size 64×64 or
 128×128 .

To get approximate samples from π ,
use the Metropolis algorithm. As the
chain K , use the following:

"Pick in Λ a vertex i uniformly at
random, and change x_i to $-x_i$ ".

Clearly K is irreducible. K is periodic,
but the Metropolis chain M has some holding
probability, so is irreducible and aperiodic.

So $\lim_{r \rightarrow \infty} \|M_x^r - \pi\|_{TV} = 0$ for any starting
state x .

Simulated annealing — an application of
Markov chain theory

366

Have a set $S = \{s_1, \dots, s_k\}$ and a function $f: S \rightarrow \mathbb{R}$. Goal is to find an s_i which minimizes f . How do we do this when k is large?

Example Traveling salesman problem

Suppose there are m cities and a symmetric positive $m \times m$ matrix D with entries giving the distances between the cities. Imagine a salesman living in one of the cities, needing to visit the other $m-1$ cities and return home. In which order should he visit the cities in order to minimize the total distance traveled? This is equivalent to finding the permutation

$\gamma = (\gamma_1, \dots, \gamma_m)$ of the set $\{1, \dots, m\}$ which

minimizes
$$f(\gamma) = \sum_{i=1}^{m-1} D_{\gamma_i, \gamma_{i+1}} + D_{\gamma_m, \gamma_1}$$

Idea of simulated annealing

Run a Markov chain with state space S , whose unique stationary distribution places most of its probability on states $s \in S$ with a small value of $f(s)$. If run the chain long enough, we're likely to end up in such a state s .

Now switch to running a Markov chain which has unique stationary distribution concentrating even more of its probability on states s that minimize $f(s)$, so that after a while we are even more likely to be in an f -minimizing state s .

Then switch to a Markov chain with even stronger preferences for states minimizing f , and so on. It seems reasonable that if the above scheme is constructed with some care, then the chance of being in an f -minimizing state at time n tends to 1 as $n \rightarrow \infty$.

Def. The Boltzman distribution $\pi_{f,T}$ on the finite set S with energy function $f: S \rightarrow \mathbb{R}$ and temperature $T > 0$ is defined by

$$\pi_{f,T}(s) = \frac{1}{Z_{f,T}} \exp(-f(s)/T)$$

Here $Z_{f,T} = \sum_S \exp(-f(s)/T)$ is the normalizing constant.

Theorem Let S be a finite set, and $f: S \rightarrow \mathbb{R}$ be arbitrary. For $T > 0$, let $d(T)$ be the probability that a random element Y chosen according to $\pi_{f,T}$ satisfies $f(Y) = \min_S f(s)$.

Then $\lim_{T \rightarrow 0} d(T) = 1$.

Proof Consider the case where S has a unique f -minimizer (leave case of several

f -minimizers as an exercise). Let
 $k = |S|$.

369

Let s be the unique f -minimizer. Let
 $a = f(s)$, $b = \min_{s' \in S-s} f(s')$. Note that $a < b$,

$$\text{so } \lim_{T \rightarrow 0} \exp\left(\frac{a-b}{T}\right) = 0.$$

$$\text{We get } \pi_{f,T}(s) = \frac{1}{Z_{f,T}} \exp(-a/T)$$

$$= \frac{\exp(-a/T)}{\sum_{s' \in S} \exp(-f(s')/T)}$$

$$= \frac{\exp(-a/T)}{\exp(-a/T) + \sum_{s' \in S-s} \exp(-f(s')/T)}$$

$$\geq \frac{\exp(-a/T)}{\exp(-a/T) + (k-1) \exp(-b/T)}$$

$$= \frac{1}{1 + (k-1) \exp\left(\frac{a-b}{T}\right)}$$

This tends to 1 as $T \rightarrow 0$, proving the theorem.

Remark Need a Markov chain with limit distribution $\pi_{f,T}$ for general choice of T .

Often use Metropolis algorithm. Choose a

cooling schedule $T_1 > T_2 > T_3 > \dots$ with

$T_i \rightarrow 0$ as $i \rightarrow \infty$, and positive integers

$N_1, N_2, \dots$. Run N_1 steps at temperature T_1 , then N_2 steps at temperature T_2 , etc.

Note: If cool too quickly, you risk getting stuck in a local minimum rather than a global minimum.

Example : simulated annealing for the traveling salesman problem

Need a Metropolis chain with stationary distribution $\pi_{f,T}$.

Base chain K is defined as follows. 371

Declare the permutations γ, γ' to be neighbors if exist $i < j$ such that γ' is obtained by reversing the segment $(\gamma_i, \dots, \gamma_j)$ meaning that

$$\gamma' = (\gamma'_1, \dots, \gamma'_m)$$

$$= (\gamma_1, \dots, \gamma_{i-1}, \gamma_j, \gamma_{j-1}, \dots, \gamma_{i+1}, \gamma_i, \gamma_{j+1}, \dots, \gamma_m)$$

The chain K moves from γ to a randomly chosen neighbor. The Metropolis chain M has transition probabilities

$$M_{\gamma, \gamma'} = \begin{cases} \frac{2}{m(m-1)} \min \left\{ \exp \left(\frac{f(\gamma) - f(\gamma')}{T} \right), 1 \right\} & \text{if } \gamma, \gamma' \text{ are neighbors} \\ 0 & \text{if } \gamma \neq \gamma' \text{ aren't neighbors} \\ 1 - \sum_{\substack{\gamma'' \\ \text{neighbor} \\ \text{of } \gamma}} \frac{2}{m(m-1)} \min \left\{ \exp \left(\frac{f(\gamma) - f(\gamma'')}{T} \right), 1 \right\} & \text{if } \gamma = \gamma' \end{cases}$$

Note : The normalizing constant $Z_{f, T}$ cancels

out of $M_{3,3}$. So the Metropolis chain can be easily run.

372

Strong uniform times

Consider mixing a deck of n cards by repeatedly removing the top card and inserting it at random.

The following argument will be used to show that $n \log(n)$ shuffles suffice to mix up the cards. Consider the bottom card of the deck. This card stays at the bottom until the first time a card is inserted below it. This is a geometric waiting time with mean n . As the shuffles continue, eventually a second card is inserted below the original bottom card (this takes about $n/2$ further shuffles). The two cards below the original bottom card are equally likely to be in relative order low-high or

high-low.

373

Similarly, the first time a 3rd card is inserted below the original bottom card, each of the 6 possible orders of the 3 bottom cards are equally likely. Now consider the first time T that the original bottom card comes to top and is reinserted. By an inductive argument, when the bottom card comes to the top, all $(n-1)!$ arrangements of the lower cards are equally likely. So at time T , all $n!$ orders of the cards are equally likely.

When the original bottom card is at position k from the bottom, the waiting time for a new card to be inserted below it is geometric with mean n/k . Thus the waiting time T has

mean $n + n/2 + n/3 + \dots + n/n \approx n \log(n)$.

Def Let $G = S_n =$ permutations of $\{1, \dots, n\}$.

Let Q be a probability on G , and X_k be the associated random walk. We say that T is a strong uniform time if

$$P(X_k = s | T = k) = 1/|G|$$

Example In the top to random example, the time T that the original bottom card takes to come to the top and be reinserted is a strong uniform time.

Theorem Let Q be a probability on G , and let T be a strong uniform time for Q . Then for all k ,

$$\|Q^{*k} - u\|_{TV} \leq P(T > k)$$

$\nwarrow$ $\nearrow$
 distribution after k steps uniform distribution

MATH 407 Lecture 26

375

From last time, say T is a strong uniform time for the random walk on S_n (or more generally a group G) corresponding to the probability measure Q if

$$P(X_k = s \mid T = k) = 1/|G|$$

↑
state after
 k steps

Theorem Let Q be a probability on G , and let T be a strong uniform time for Q . Then for all k ,

$$\|Q^{*k} - U\|_{TV} \leq P(T > k)$$

↑ ↑
distribution uniform
after k steps distribution

Proof For any $A \subset G$,

$$\mathbb{Q}^{*k}(A) = P(X_k \in A)$$

376

$$= \sum_{j \leq k} P(X_k \in A, T=j) + P(X_k \in A, T > k)$$

$$= \sum_{j \leq k} P(X_k \in A | T=j) P(T=j) \\ + P(X_k \in A | T > k) P(T > k)$$

$$= \sum_{j \leq k} U(A) P(T=j) \\ + P(X_k \in A | T > k) P(T > k)$$

$$= U(A) + [P(X_k \in A | T > k) - U(A)] P(T > k)$$

Thus $|\mathbb{Q}^{*k}(A) - U(A)| \leq P(T > k),$

since $|P(X_k \in A | T > k) - U(A)| \leq 1.$

As an application, we have the following result.

Theorem

377

For the top to random shuffle on
 n cards, let $k = n \log(n) + cn$. Then for
 $c \geq 0$ and $n \geq 2$, $\|p^{*k} - u\|_{TV} \leq e^{-c}$

Proof

Let $T =$ first time that the original bottom
 card comes to top and is reinserted.

From last time, T is a strong uniform time.

Also from last time,

$$T = T_1 + (T_2 - T_1) + \dots + (T_{n-1} - T_{n-2}) + (T - T_{n-1})$$

where T_i is the time the first card is placed
 under the bottom card and $T_{i+1} - T_i$ has a
 geometric distribution

$$P(T_{i+1} - T_i = j) = \frac{i+1}{n} \left(1 - \frac{i+1}{n}\right)^{j-1}, j \geq 1.$$

Further, these differences are independent.

The time T has the same distribution
 as the waiting time in the coupon collectors

problem. To define this, consider a random sample with replacement from an urn with n balls. Let V be the number of balls required until each ball has been drawn at least once. Let $m = n \log(n) + cn$. For each ball b , let A_b be the event "ball b is not drawn in the first m draws". Then

$$P(V > m) = P\left(\bigcup_b A_b\right) \leq \sum_b P(A_b)$$

$$= n(1 - 1/n)^m$$

$$\leq n e^{-m/n}$$

$$= e^{-c}$$

since
 $1 - x \leq e^{-x}$
 for $x \geq 0$

So we only need to see that T, V have the same distribution. Now V can be written

$$V = (V - V_{n-1}) + (V_{n-1} - V_{n-2}) \cdots + (V_2 - V_1) + V_1$$

where V_i is the number of draws required until i distinct balls have been drawn at least once. After i distinct balls have been drawn, the chance that a draw produces a new ball is $\frac{n-i}{n}$, so $V_{i+1} - V_i$ is geometric

$$P(V_{i+1} - V_i = j) = \frac{n-i}{n} \left(1 - \frac{n-i}{n}\right)^{j-1}, j \geq 1.$$

so T, V have the same distribution, and the result follows.

Remark For the top t random shuffle,

let $k = n \log(n) + cn$. Can prove the lower bound

$$\|P^{*k} - U\|_{TV} \rightarrow 1 \text{ as } n \rightarrow \infty$$

for $c = c(n) \rightarrow -\infty$.

Practice final exam (10 questions)

Problem 1

380

The joint density function of X and Y is

$$f(x,y) = \begin{cases} x+y & 0 < x < 1, 0 < y < 1 \\ 0 & \text{else} \end{cases}$$

a) Are X, Y independent?

Sol. No, can't factor $x+y$ as a product of a function of x and a function of y .

b) Find the density function of X .

Sol.

For $0 < x < 1$,

$$\begin{aligned} \int_0^1 (x+y) dy &= xy + y^2/2 \Big|_{y=0}^{y=1} \\ &= x + 1/2 \end{aligned}$$

c) Find $P(X+Y < 1)$.

Sol.

$$P(X+Y < 1) = \int_0^1 \int_0^{1-x} (x+y) dy dx$$

$$\begin{aligned}
 &= \int_0^1 \int_0^{1-x} x \, dy \, dx + \int_0^1 \int_0^{1-x} y \, dy \, dx \quad 381 \\
 &= \int_0^1 (x - x^2) \, dx + \int_0^1 (1-x)^2 / 2 \, dx \\
 &= 1/3
 \end{aligned}$$

Problem 2

From a group of 3 first year students, 4 sophomores, 4 juniors, and 3 seniors, a committee of size 4 is randomly selected. Find the probability that the committee will consist of:

a) 1 from each class

Sol. $3 \cdot 4 \cdot 4 \cdot 3 / \binom{14}{4} = .1439$

b) 2 sophomores and 2 juniors

Sol. $\binom{4}{2} \cdot \binom{4}{2} / \binom{14}{4} = .0360$

c) only sophomores or juniors

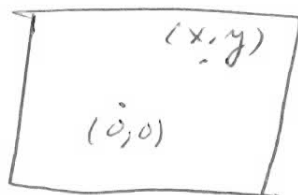
382

Sol. $\binom{8}{4} / \binom{14}{4} = .0699$

Problem 3

The county hospital is located at the center of a square whose sides are 3 miles wide. If an accident occurs within this square, then the hospital sends out an ambulance. The road network is rectangular, so the travel distance from the hospital, whose coordinates are $(0,0)$, to the point (x,y) is $|x|+|y|$. If an accident occurs at a point that is uniformly distributed in the square, find the expected travel distance of the ambulance.

Sol.



383

$$\frac{1}{9} \int_{-3/2}^{3/2} \int_{-3/2}^{3/2} |x| + |y| \, dx \, dy$$

$$= \frac{2}{9} \int_{-3/2}^{3/2} \int_{-3/2}^{3/2} |x| \, dx \, dy$$

$$= \frac{4}{9} \int_{-3/2}^{3/2} \int_0^{3/2} x \, dx \, dy$$

$$= \frac{4}{9} \frac{(3/2)^2}{2} (3) = 3/2.$$

Problem 4

Three cards are chosen without replacement from a deck of 52 cards. Let X denote the number of aces chosen. Compute the expected value of X given that the ace of spades is chosen.

Sol.

$$P(X=1 \mid \text{ace of spades chosen})$$

$$= P(X=1, \text{ace of spades chosen}) / P(\text{ace of spades chosen})$$

384

$$= \frac{\binom{48}{2} / \binom{52}{3}}{3/52}$$

$3/52 \leftarrow$ since $P(\text{ace of spades})$
chosen

$$= P(\text{1st card is A spades}) \\ + P(\text{2nd card is A spades}) \\ + P(\text{3rd card is A spades}) \\ = 1/52 + 1/52 + 1/52$$

$$= .8847$$

Similarly,

$$P(X=2 | \text{ace of spades chosen})$$

$$= \frac{\binom{3}{1} \binom{48}{1} / \binom{52}{3}}{3/52} = .1129$$

$$P(X=3 | \text{ace of spades chosen})$$

$$= \frac{\binom{2}{2} / \binom{52}{3}}{3/52} = .0023$$

Thus $E(X | \text{ace of spades chosen})$

$$= \sum_{i=0}^3 i P(X=i | \text{ace of spades chosen})$$

$$= .8847 + 2(-.1129) + 3(-.0023)$$

$$= 1.1176$$

(note: exact value is $19/17$).

Problem 5

If X is binomial random variable with expected value 6 and variance 2.4, find $P(X=5)$.

Sol.

Suppose X is binomial (n, p)

$$E(X) = np = 6$$

$$\text{Var}(X) = np(1-p) = 2.4$$

$$\text{so } 1-p = .4, \text{ so } p = .6, n = 10$$

Thus
$$P(X=5) = \binom{10}{5} (0.6)^5 (0.4)^5$$

$$\approx .2007$$

Problem 6

Let $X_1, \dots, X_{20}$ be independent Poisson random variables with mean 1.

a) Use Markov's inequality to upper bound

$$P\left(\sum_{i=1}^{20} X_i \geq 15\right)$$

b) Use central limit theorem to approximate

$$P\left(\sum_{i=1}^{20} X_i \geq 15\right)$$

Hint: A Poisson random variable with parameter λ has variance λ .

Sol.

a) $\sum_i X_i$ has mean 20.

$$\text{So } P\left(\sum X_i \geq 15\right) \leq 20/15 = 4/3$$

387

b) $\sum X_i$ has mean 20 and variance 20.

$$P\left(\frac{\sum X_i - 20}{\sqrt{20}} \geq \frac{14.5 - 20}{\sqrt{20}}\right)$$

$$\approx P(Z \geq -1.23) = P(Z \leq 1.23)$$

$\begin{matrix} \text{standard} \\ \text{normal} \end{matrix}$

$$= .8907$$

Problem 7

Twelve percent of the population is left-handed. Approximate the probability that there are at least 20 left-handers in a school of 200 students. State your assumptions.

Sol.

$$\text{mean} = 200 (.12) = 24$$

$$SD = \sqrt{200 (.12)(.88)} = 4.59$$

So using normal approximation.

388

$$P(X \geq 19.5) \approx P\left(\frac{X-24}{4.59} \geq \frac{19.5-24}{4.59}\right)$$

$$\approx P(Z \geq -0.98)$$

$$= P(Z \leq 0.98) = 0.8365$$

Problem 8

If $E(X)=75$, $E(Y)=75$, $\text{Var}(X)=17$, $\text{Var}(Y)=12$
and $\text{Cov}(X, Y) = -3$, upper bound

a) $P(|X-Y| > 15)$

b) $P(X > Y + 15)$

c) $P(Y > X + 15)$

Hint: Chebyshev's inequality and its one-sided version may be helpful.

Sol.

$$E(X-Y) = E(X) - E(Y) = 75 - 75 = 0$$

$$\begin{aligned} \text{Var}(X-Y) &= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \\ &= 28 \end{aligned}$$

$$\text{Since } \text{Cov}(X, -Y) = -\text{Cov}(X, Y)$$

a)

$$P(|X-Y-0| \geq 15)$$

$$\leq 28/(15)^2 = .124$$

by Chebyshev's inequality

b) By one-sided Chebyshev inequality

$$P(X-Y \geq 15) \leq 28/(28+(15)^2)$$

$$= 28/253 = .111$$

c) By one-sided Chebyshev inequality,

$$P(Y-X \geq 15) \leq 28/(28+(15)^2)$$

$$= 28/253 = .111$$

Problem 9

A standard Cauchy random variable has density function $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.

Show that if X is a standard Cauchy, then so is $1/X$.

MATH 407 lecture 27

Problem 9 A standard Cauchy random variable has density function $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.

Show that if X is a standard Cauchy, then so is $1/X$.

Sol. Let $Y = 1/X$.

$$f_Y(y) = f_X(x) / \left| \frac{dy}{dx} \right|$$

$$y = 1/x \Rightarrow \frac{dy}{dx} = -\frac{1}{x^2} \Rightarrow \left| \frac{dy}{dx} \right| = \frac{1}{x^2}$$

$$\text{So } f_Y(y) = \frac{1}{\pi(1+x^2)} / \frac{1}{x^2}$$

$$= \frac{1}{\pi(1+1/y^2)}$$

$$= \frac{1}{\pi(1+y^2)}$$

Problem 10

391

Suppose X, Y, Z are independent random variables that are each equally likely to be either 1 or 2. Find the probability mass function of

a) XYZ

b) $XY + XZ + YZ$

Sol.

a) Let $p_j = P(XYZ = j)$

$$X=1, Y=1, Z=1 \rightarrow XYZ=1 \quad 1/8$$

$$X=1, Y=1, Z=2 \rightarrow XYZ=2 \quad 1/8$$

$$X=1, Y=2, Z=1 \rightarrow XYZ=2 \quad 1/8$$

$$X=1, Y=2, Z=2 \rightarrow XYZ=4 \quad 1/8$$

$$X=2, Y=1, Z=1 \rightarrow XYZ=2 \quad 1/8$$

$$X=2, Y=1, Z=2 \rightarrow XYZ=4 \quad 1/8$$

$$X=2, Y=2, Z=1 \rightarrow XYZ=4 \quad 1/8$$

$$X=2, Y=2, Z=2 \rightarrow XYZ=8 \quad 1/8$$

Thus

$$p_1 = 1/8$$

$$p_2 = 3/8$$

$$p_3 = 3/8$$

$$p_4 = 1/8$$

b) Arguing as in a), we calculate

$$p_3 = 1/8$$

$$p_5 = 3/8$$

$$p_8 = 3/8$$

$$p_{10} = 1/8$$

Practice Final 2 (6 problems)

Problem 1

Suppose a laboratory blood test gives one of two results (positive or negative). It is found that 95 percent of people with a certain disease

produce a positive result. But 2 percent
of the people without the disease also produce
a positive result (called a false positive).

Suppose that one percent of the population
actually has the disease. What is the probability
that a person chosen at random from the population
will have the disease, given that the person's
blood test is positive.

Sol. Test results are +, -

Let D = event have disease.

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} \\ \text{Bayes' formula} \quad \nearrow & \\ &= \frac{(.95)(.01)}{(.95)(.01) + (.02)(.99)} \\ &= 95/293 \approx .32 \end{aligned}$$

Problem 2

Five cards are dealt from a standard deck of 52 cards.

a) What is $P(\text{5th card is a queen})$?

b) What is $P(\text{3rd and 5th cards are red})$?

Sol.

a) Fifth card is equally likely to be anything, so chance is $4/52 = 1/13$.

b) By symmetry,

$$P(\text{3rd and 5th cards are red})$$

$$= P(\text{1st and 2nd cards are red})$$

$$= 26/52 \times 25/51$$

Problem 3

Let $X_1, X_2, \dots, X_n$ be independent exponential random variables, and suppose X_i is exponential with parameter λ_i .

- a) Find the distribution of $X_{\min}$, the minimum of $X_1, \dots, X_n$.
- b) Compute the expected value of $X_{\min}$.

Sol.

Since X_i is exponential(λ_i),

$$P(X_i > x) = e^{-\lambda_i x}$$

$$a) X_{\min} > x \Leftrightarrow X_i > x \text{ for all } i$$

$$\begin{aligned} \text{So } P(X_{\min} > x) &= (e^{-\lambda_1 x}) (e^{-\lambda_2 x}) \dots (e^{-\lambda_n x}) \\ &= e^{-(\lambda_1 + \dots + \lambda_n)x} \end{aligned}$$

So $X_{\min}$ is exponential ($\lambda_1 + \dots + \lambda_n$)

$$b) E(X_{\min}) = \frac{1}{\lambda_1 + \dots + \lambda_n}$$

Problem 4

If X and Y are independent poisson random variables with respective parameters λ_1 and λ_2 ,

396

calculate the conditional expected value of X given that $X+Y=n$. (Hint: find the conditional distribution of X given that $X+Y=n$).

Sol.

First, find the conditional distribution of X given that $X+Y=n$.

$$P(X=k | X+Y=n) = P(X=k, X+Y=n) / P(X+Y=n)$$

$$= P(X=k, Y=n-k) / P(X+Y=n)$$

$$= P(X=k) P(Y=n-k) / P(X+Y=n)$$

by
independence

Since $X+Y$ is Poisson $(\lambda_1 + \lambda_2)$, we get

$$\frac{e^{-\lambda_1} \lambda_1^k}{k!} \frac{e^{-\lambda_2} \lambda_2^{n-k}}{(n-k)!} / \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^n}{n!}$$

$$= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k}$$

So the conditional distribution of 397
 X given $X+Y=n$ is binomial $(n, \lambda_1/(\lambda_1+\lambda_2))$

$$\text{So } E(X | X+Y=n) = \frac{n \lambda_1}{(\lambda_1 + \lambda_2)}$$

Problem 5

Flip a fair coin 100 times. Use the central limit theorem to estimate the probability that the number of heads is between 48 and 53 inclusive

Sol. We want to approximate

$$P(48 \leq X \leq 53), \text{ where } X = \text{number of heads}$$

$$E(X) = 100(1/2) = 50$$

$$SD(X) = \sqrt{100(1/2)(1/2)} = 5$$

Since 48, 53 are close to 50, use the continuity correction. So

$$P(47.5 \leq X \leq 53.5)$$

398

$$= P\left(\frac{47.5 - 50}{5} \leq \frac{X - 50}{5} \leq \frac{53.5 - 50}{5}\right)$$

$$\approx P(-.5 \leq Z \leq .7)$$

↖ standard normal

$$= \Phi(.7) - \Phi(-.5) = .4495$$

Problem 6

Let X be uniform on $[0, 1]$. Find the distribution of $-\frac{1}{\lambda} \log(1-X)$, where $\lambda > 0$.

Sol.

$$y = -\frac{1}{\lambda} \log(1-x)$$

$$\frac{dy}{dx} = \frac{-1}{\lambda(1-x)} (-1) = \frac{1}{\lambda(1-x)}$$

$$\text{So } f_Y(y) = f_X(x) / \left| \frac{dy}{dx} \right| = \lambda(1-x)$$

$$\text{Now } e^{-\lambda y} = (1-x)$$

$$\text{So } f_Y(y) = \lambda e^{-\lambda y}, \text{ so } Y \text{ is exponential } (\lambda)$$

Practice Final 3 (6 problems)

399

Problem 1

Let X_2 and X_3 be indicators of independent events with respective probabilities $1/2$ and $1/3$.

a) Display the joint distribution table of $X_2 + X_3$ and $X_2 - X_3$

b) calculate $E(X_2 - X_3)$

c) calculate $SD(X_2 - X_3)$

Sol.

a)

		$X_2 + X_3$		
		0	1	2
$X_2 - X_3$	-1	0	$1/6$	0
	0	$1/3$	0	$1/6$
	1	0	$1/3$	0

b)

$$P(X_2 - X_3 = -1) = 1/6$$

$$P(X_2 - X_3 = 0) = 1/2$$

$$P(X_2 - X_3 = 1) = 1/3$$

$$\begin{aligned} \text{So } E[X_2 - X_3] &= -1(1/6) + 0(1/2) + 1(1/3) \\ &= 1/6 \end{aligned}$$

$$c) \text{Var}(X_2 - X_3) = E[(X_2 - X_3)^2] - [E(X_2 - X_3)]^2$$

$$= 1/6 + 1/3 - 1/36 = 17/36$$

$$\text{So } SD(X_2 - X_3) = \sqrt{17/36}$$

Problem 2

Five cards are dealt from a standard deck of 52 cards. Write down expressions for

- a) $P(\text{3rd card is a queen})$
- b) $P(\text{3 card is queen given that last 2 cards are not queens})$
- c) $P(\text{all cards of same suit})$
- d) $P(\text{2 or more queens})$

Sol.

a) $1/13$, since 3rd card is equally likely to be anything

b) $P(\text{3rd is queen} \mid \text{cards 4, 5 aren't queens})$

401

by $\nearrow$ symmetry $= P(\text{1st is queen} \mid \text{2nd, 3rd aren't queens})$

$$= \frac{P(\text{1st queen and 2nd, 3rd aren't queens})}{P(\text{2nd, 3rd aren't queens})}$$

$$= \frac{(\frac{1}{13})(\frac{48}{51})(\frac{47}{50})}{(\frac{48}{52})(\frac{47}{51})} = \frac{4}{50}$$

c) $4 \binom{13}{5} / \binom{52}{5}$

$\uparrow$ choose suit
 $\uparrow$ choose 5 cards in the suit

d) $1 - P(0 \text{ or } 1 \text{ queen})$

$$= 1 - \left[\frac{\binom{4}{0} \binom{48}{5} + \binom{4}{1} \binom{48}{4}}{\binom{52}{5}} \right]$$

Problem 3

A random variable N is uniformly distributed on the set $\{1, 2, \dots, 10\}$. Let X be the indicator

of the event $N \leq 5$, and let Y be the indicator of event N is even.

- Find $E(X)$ and $E(Y)$
- Are X, Y independent? Explain
- Find $\text{Cov}(X, Y)$
- Find $E[(X+Y)^2]$

Sol.

a) $E(X) = E(Y) = 1/2$

b) No

$$P(X=1, Y=1) = P(N=2 \text{ or } 4) = 2/10$$

But $P(X=1)P(Y=1) = (1/2)(1/2)$

c) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

Now $E(XY) = P(X=1, Y=1) = 2/10$

So $\text{Cov}(X, Y) = 2/10 - 1/2(1/2) = -1/20$

d) $E[(X+Y)^2] = E(X^2) + 2E(XY) + E(Y^2)$

Since X, Y only take values 0, 1 $\rightarrow = \cancel{E(X)} + 2E(XY) + E(Y)$

$$= 1/2 + 4/10 + 1/2 = 1.4$$

Problem 4

Suppose that T is a random variable such that $P(T > t) = e^{-t}$, $t \geq 0$. Derive a formula for the density function f_X of the random variable $X = 1/T$.

Sol. $P(T > t) = e^{-t}$.

We have $X = 1/T$, so

$$\begin{aligned} f_X(x) &= f_T(t) \left| \frac{dx}{dt} \right| \\ &= e^{-t} / \left| -1/t^2 \right| \\ &= t^2 e^{-t} = \frac{1}{x^2} e^{-1/x} \quad x > 0 \end{aligned}$$

$$f_X(x) = 0 \quad \text{if } x \leq 0$$

Problem 5

Let X, Y, Z be 3 independent standard normals

a) calculate $P(|X| \leq 1, |Y| \leq 2, |Z| \leq 3)$

Sol.

$$\text{Get } [\Phi(1) - \Phi(-1)] [\Phi(2) - \Phi(-2)] [\Phi(3) - \Phi(-3)]$$

$$\approx (.68)(.95)(.9974)$$

$$\approx .65$$

b) calculate $E[(X+Y+Z)^2]$

Sol.

$$E[(X+Y+Z)^2] = E(X^2) + E(Y^2) + E(Z^2) \\ + 2E(XY) + 2E(XZ) + 2E(YZ)$$

by
independence

$$\rightarrow = 1 + 1 + 1$$

$$+ 2E(X)E(Y) + 2E(X)E(Z) + 2E(Y)E(Z)$$

$$= 3 + 0 + 0 + 0 = 3$$

c) calculate $P(X+Y \leq 2Z)$

Sol.

$$P(X+Y \leq 2Z) = P(X+Y-2Z \leq 0)$$

Now $X+Y-2Z$ is normal with mean 0, variance

$$(1) + (1) + (-2)^2 = 6$$

So answer is $1/2$.

MATH 407 Lecture 28Problem 6 from 3rd practice final

The joint density of X and Y is given by

$$f(x,y) = \frac{1}{2} y e^{-xy} \quad \begin{array}{l} 0 < x < \infty \\ 0 < y < 2 \end{array}$$

and $f(x,y) = 0$ else

Calculate $E[e^{X/2} | Y=1]$

Sol.

The conditional density of X , given $Y=1$, is

$$f(x,1)/f_Y(1) = \frac{1/2 e^{-x}}{\int_0^{\infty} 1/2 e^{-x} dx} = e^{-x}$$

$$\begin{aligned} \text{So } E[e^{X/2} | Y=1] &= \int_0^{\infty} e^{x/2} \frac{f(x,1)}{f_Y(1)} dx \\ &= \int_0^{\infty} e^{x/2} e^{-x} dx \\ &= 2 \end{aligned}$$

Back to Markov chains

Coupling

Example

Take top card off deck and put it in a random position. Then take bottom card off and put it in a random position. Continue, alternately placing top in at random, bottom in at random.

There is no longer an obvious strong stationary time. Let's work with the inverse shuffle that removes a random card and alternately places it on top or bottom (the inverse shuffle and original shuffle have the same convergence rates).

To describe a "coupling", consider a 2nd deck of cards. The first deck starts in order $123 \dots n$. The second deck starts in a random order. Think of a random card, say 6 of hearts.

Remove it from deck 1 and place it on top. Remove it from deck 2 and place it on top. Note that from each deck's marginal vantage point, a card was removed at random and placed on top.

Next, think of a random card, say Ace of spades, and remove it and place it at bottom of both decks. Continue in this way, each time thinking of a random card, removing the card from the 2 decks, and placing the card alternately on top and bottom.

As process continues, decks 1 and 2 match up. If same card is chosen again, the procedure keeps the ~~same~~ same number of matches. Let T be first time all cards have been touched. Clearly the 2 decks are in same order at time T .

But deck 2 started at random, and so remains random. Thus deck 1 is random at time T .

The bound on coupon collector's problem yields

Theorem For this shuffle, if $k = n \log(n) + cn$, and $c > 0$,

$$\| p^{*k} - u \|_{TV} \leq e^{-c}$$

$\nwarrow$ $\nearrow$
 distribution after k steps uniform distribution

This follows from two lemmas

Lemma 1

Let X be a finite set. Let P_1, P_2 be probability measures on X . Let Q be a probability on $X \times X$ with margins P_1 and P_2 . Let Δ be the diagonal

$$\Delta = \{(x, x) : x \in X\}.$$

Then $\|P_1 - P_2\|_{TV} \leq Q(\Delta^c)$ $\nwarrow$ complement of Δ

Proof For all $A \subseteq X$,

$$|P_1(A) - P_2(A)| = |Q(A \times X) - Q(X \times A)|$$

$$= |Q(A \times X \cap \Delta) + Q(A \times X \cap \Delta^c) \\ - Q(X \times A \cap \Delta) - Q(X \times A \cap \Delta^c)|$$

The first and third numbers in the absolute value sign are equal. The second and fourth terms give a difference between two numbers, both non-negative and smaller than $Q(\Delta^c)$.

Lemma 2

$$\|p^{*h} - \pi\|_{TV} \leq P(T > k) \quad \begin{array}{l} \swarrow \text{stationary distribution} \\ \text{first deck} \end{array}$$

Proof Let Q be distribution of (X^1, X^2) . ↙ ↘ second deck

Thus Q has marginals p^{*h} and π . So by previous lemma,

$$\|p^{*h} - \pi\|_{TV} \leq Q(\Delta^c) = P(T > k)$$

Note : Unlike our linear algebra techniques, coupling does not require reversibility.

410

Carries Markov chainExample Look at carries when you add

3 40-digit numbers (chosen randomly)

Carries $\rightarrow$

$$\begin{array}{rcccccccc}
 & & & & 1 & 1 & 1 & 2 & 1 & 0 & 2 & 1 & 1 & 2 \\
 83621 & - & - & - & 73497 & 38426 \\
 46700 & - & - & - & 17255 & 19649 \\
 31026 & - & - & - & 43856 & 37376 \\
 \hline
 & & & & 34608 & 95451
 \end{array}$$

The carries are only 0, 1, or 2. In the example

$$0's : 10/40 = 25\%$$

$$1's : 23/40 = 57.5\%$$

$$2's : 7/40 = 17.5\%$$

As 40 grows,

$$0's \rightarrow A(3,0)/6 = 1/6$$

$$1's \rightarrow A(3,1)/6 = 2/3$$

$$2's \rightarrow A(3,2)/6 = 1/6$$

Here $A(n, j) =$ number of $\pi \in S_n$
with j descents

(a descent at position i , $1 \leq i \leq n-1$
means that $\pi(i) > \pi(i+1)$).

$$\text{So } A(3, 0) = 1 \quad 123$$

$$A(3, 1) = 4 \quad 132, 231, 213, 312$$

$$A(3, 2) = 1 \quad 321$$

Work of Holte (1997)

If one adds n integers base b , by
choosing digits uniformly at random in
 $\{0, 1, \dots, b-1\}$, the sequence of carries
 $K_0 = 0, K_1, K_2, \dots$ is a Markov chain with
values in $\{0, 1, \dots, n-1\}$

The Markov property holds since to
compute carry to next column, one only
needs to know the carry and numbers in
current column: past doesn't matter.

412

$$\text{let } P(i, j) = P(K' = j \mid K = i)$$

denote entry of transition matrix between successive carriers K and K' .

Holte showed:

1) For $0 \leq i, j \leq n-1$,

$$P(i, j) = \frac{1}{b^n} \sum_{r=0}^{j-L(i)/b} (-1)^r \binom{n+1}{r} \binom{n-1-i+(j+1-r)b}{n}$$

we will prove this later.

Example

$$n=2 \quad (P(i, j)) = \frac{1}{2b} \begin{pmatrix} b+1 & b-1 \\ b-1 & b+1 \end{pmatrix}$$

$$n=3 \quad (P(i, j)) = \frac{1}{6b^2} \begin{pmatrix} b^2+3b+2 & 4b^2-4 & b^2-3b+2 \\ b^2-1 & 4b^2+2 & b^2-1 \\ b^2-3b+2 & 4b^2-4 & b^2+3b+2 \end{pmatrix}$$

He also showed that when $b=2$, for

$$\text{any } n, \quad P(i, j) = \frac{1}{2^n} \binom{n+1}{2j-i+1} \quad 0 \leq i, j \leq n-1$$

Denote these matrices by P_b . Holte
also showed:

- 2) Matrix P_b has stationary vector
(left eigenvector with eigenvalue 1)
independent of base b :

$$\pi(j) = A(n, j) / n!$$

- 3) Matrix P_b has eigenvalues $1, 1/b, 1/b^2, \dots, 1/b^{n-1}$
with explicitly computable left and right
eigenvectors independent of b .

- 4) $P_a P_b = P_{ab}$ for all real a, b .

Recall : b -riffle shuffle on S_n . Let

$Q_b(\sigma) = P(\sigma \text{ after a } b\text{-shuffle})$. Recall

that

- 1) Let $*$ be convolution of measures

Then $Q_a * Q_b = Q_{ab}$ (in words,
an a -shuffle
followed by a b -shuffle
is an ab shuffle)

2) Eigenvalues of Markov chain

induced by Q_b are $1, 1/b, \dots, 1/b^{n-1}$

(Here the multiplicity of $1/b^i$ is the number of $\pi \in S_n$ with $n-i$ cycles, instead of multiplicity 1).

Properties 1), 2) and the appearance of descents in both carries and shuffling convinced us that there must be an intimate connection between carries and shuffling.

Theorem The number of descents in successive b -shuffles of n cards forms a Markov chain on $\{0, 1, \dots, n-1\}$ with transition matrix the same as Holte's carries matrix.

Note Have a bijective proof of theorem and also a generating function proof.

Cor. stationary distribution of carries

chain is $\pi(j) = A(n, j)/n!$

Cor. Let a Markov chain on symmetric group begin at identity and proceed by successive b -shuffles. Then $d(\pi)$, the number of descents, forms a Markov chain.

Note: usually a function of a Markov chain isn't Markov.

Let's give another proof that $\pi(j) = A(n, j)/n!$ is the stationary distribution for the carries chain.

Lemma Let $X_1, X_2, \dots, X_n$ be independent uniform $(0, 1)$ random variables. Then for all integers j ,

$$\frac{A(n, j)}{n!} = P\left(j \leq \sum_{i=1}^n X_i < j+1\right)$$

(will prove this later)

Theorem $\pi(j) = A(n, j)/n!$ is the stationary distribution of carries chain.

Proof By Holte, r steps of the base b carries chain is equivalent to one step of the base b^r carries chain. Letting $Y_1, \dots, Y_n$ be independent discrete uniforms on $\{0, 1, \dots, b^r - 1\}$, it follows that

$$p^r(0, j) = P(jb^r \leq \sum_{i=1}^n Y_i < (j+1)b^r)$$

Letting $U_1, \dots, U_n$ be independent continuous uniforms on $[0, b^r]$, it follows that

$$p^r(0, j) = P(jb^r \leq \sum_{i=1}^n \lfloor U_i \rfloor < (j+1)b^r)$$

$$= P(jb^r \leq \sum_{i=1}^n u_i - \sum_{i=1}^n (u_i - \lfloor u_i \rfloor) < (j+1)b^r)$$

$$= P(j \leq \sum_{i=1}^n X_i - E < j+1)$$

Here the $X_i = u_i / b^r$ are independent uniforms on $[0, 1]$ and $E = 1/b^r \sum_{i=1}^n (u_i - \lfloor u_i \rfloor)$.

Although E isn't independent of the X_i 's, note that when n is fixed and $r \rightarrow \infty$, E converges in probability to 0. Indeed,

$|E| \leq n/b^r$ with probability 1. Thus Slutsky's theorem implies that

$$\lim_{r \rightarrow \infty} P^r(0, j) = P(j \leq \sum_{i=1}^n X_i < j+1),$$

so the theorem follows from the previous lemma.

Convergence rate of carries chain

separation distance

$$\text{sep}(r) = \max_j \left[1 - \frac{P^r(0,j)}{\pi(j)} \right]$$

Lemma Call a matrix TP_2 if all of its 2×2 minors are non-negative. Then Holte's matrix $(P(i,j))$ is TP_2 .

Theorem For any $b \geq 2$ and $n \geq 2$, Holte's matrix $(P(i,j))$ satisfies:

- 1) For all $r \geq 0$, the separation $\text{sep}(r)$ (started at 0) is attained at $j = n-1$
- 2) For $r \leq \lfloor 2 \log_b(n) + \log_b(4) \rfloor$,

$$\text{sep}(r) \rightarrow 1 - e^{-1/2c} \quad \text{if } c > 0 \text{ is fixed and } n \rightarrow \infty$$

Proof

1) uses the fact that $(P(i,j))$ is TP_2 .

2) By connection between carries and shuffling,

$p^r(0, n-1)$ = chance of being at unique permutation with $n-1$ descents after r many b -shuffles.

Thus by our discussion of riffle shuffles,

$$p^r(0, n-1) = b^{-rn} \binom{b^r}{n}$$

$$\text{So } \text{sep}(r) = 1 - p^r(0, n-1) / \pi(n-1)$$

$$= 1 - \prod_{i=0}^{n-1} (1 - i/b^r)$$

$$= 1 - \exp\left(\sum_{i=0}^{n-1} \log(1 - i/b^r)\right)$$

Letting $b^r = cn^2$ with $c > 0$ fixed, this becomes (using $\log(1-x) = -x - x^2/2 - x^3/3 \dots$)

$$1 - \exp\left(-\sum_{i=0}^{n-1} \frac{i}{cn^2} + O(i^2/n^4)\right)$$

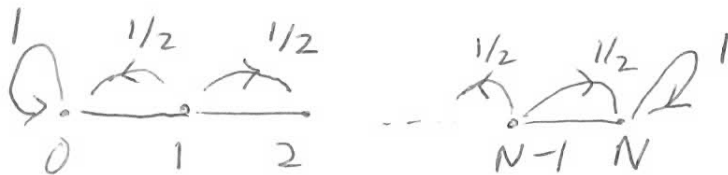
$$\rightarrow 1 - e^{-1/2c} \text{ as } n \rightarrow \infty$$

"big O notation"

Math 407 Lecture 29

Random walks and electrical networks

Consider random walk on $\{0, 1, 2, \dots, N\}$



1st interpretation: Have a drunk man walking along a stretch of Park Avenue. Corner N is his home and corner 0 is a bar. If the walker is at position i , $1 \leq i \leq N-1$, half the time the walker moves to the left, half the time to the right. If the walker reaches 0 or N , the walker stays there.

2nd interpretation: Peter and Paul have a total of N dollars. At each turn, Peter wins a dollar from Paul with

probability $1/2$ and loses a dollar 421
to Paul with probability $1/2$. The game
continues until Peter's fortune is 0
(he is ruined) or his fortune is N
(he wins all of Paul's money).

Let $p(x)$ = Probability that starting at x ,
reach N before 0.

Then

- 1) $p(0) = 0$
- 2) $p(N) = 1$
- 3) $p(x) = \frac{1}{2}p(x-1) + \frac{1}{2}p(x+1), 1 \leq x \leq N-1$

Indeed, we have a basic fact :

Let E be any event

Let F, G be events so that one and
only one of F, G will occur.

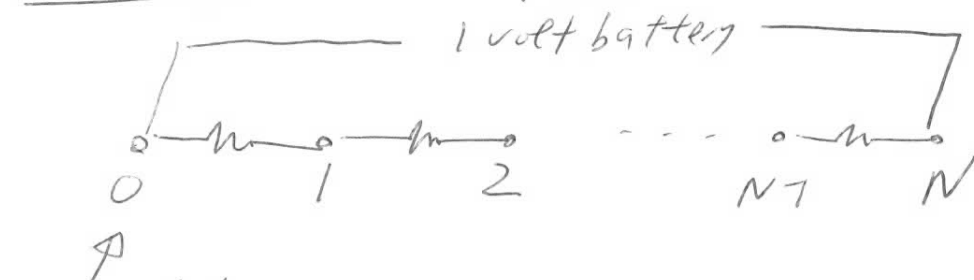
Then $P(E) = P(F)P(E|F) + P(G)P(E|G)$

Now, let E = starting from x , hit
 N before 0

F = first step to left

G = first step to right

Electrical network problem



grounded
so that
voltage = 0

Connect equal resistors
in series and put a unit
voltage across the ends.

Voltages will be established at the points $x = 0, 1, \dots, N$. We have grounded the point $x = 0$ so that $v(0) = 0$. We ask for the voltages $v(x)$. We know $v(0) = 0$ and $v(N) = 1$, so $v(x)$ satisfies 1), 2) above. We claim that $v(x)$ also satisfies 3).

By "Kirchoff's law", for $x = 1, \dots, N-1$ the current flowing into x must be equal to

423

the current flowing out. By 'Ohm's law', if x, y are connected by a resistance of magnitude R , the current flowing from x to y is

$$i_{xy} = \frac{v(x) - v(y)}{R}$$

Thus for $x = 1, 2, \dots, N-1$

$$\frac{v(x-1) - v(x)}{R} + \frac{v(x+1) - v(x)}{R} = 0$$

Rearranging, we get

$$v(x) = \frac{v(x+1) + v(x-1)}{2} \quad \text{for } x = 1, 2, \dots, N-1$$

thus $v(x)$ satisfies 3) above.

Are $p(x), v(x)$ equal? we shall see that the answer is yes.

Harmonic functions in one-dimension:
the uniqueness principle

424

Let $S = \{0, 1, \dots, N\}$

Let $D = \{1, 2, \dots, N-1\}$ be the interior points of S

Let $B = \{0, N\}$ be the boundary points of S

call a function f on S harmonic if, at all points x of D , it satisfies

$$f(x) = \frac{f(x-1) + f(x+1)}{2}$$

As we have seen, $p(x)$ and $v(x)$ are harmonic functions on S having the same values on the boundary: $p(0) = v(0) = 1$, $p(N) = v(N) = 1$. We will see that can't have different harmonic functions with the same boundary values.

Lemma A harmonic function defined on S

425
takes its maximum value M and minimum value m on the boundary.

Proof

Let M be the largest value of f . Then if $f(x) = M$ for x in D , the same must be true for $f(x-1)$ and $f(x+1)$, since $f(x)$ is the average of these two values. If $x-1$ is still an interior point, the same argument implies that $f(x-2) = M$. Continuing in this way, we eventually conclude that $f(0) = M$. The same argument works for the minimum value m .

Theorem

If $f(x)$, $g(x)$ are harmonic on S such that $f(x) = g(x)$ on B , then $f(x) = g(x)$ for all x .

Proof

Let $h(x) = f(x) - g(x)$. Then if x is

any interior point,

$$\frac{h(x-1) + h(x+1)}{2}$$

$$= \frac{f(x-1) + f(x+1)}{2} - \frac{g(x-1) + g(x+1)}{2}$$

$$= f(x) - g(x) = h(x).$$

So h is harmonic. But $h(x) = 0$ on B , so by the lemma, the maximum and minimum values of h are 0. Thus $h(x) = 0$ for all x , and $f(x) = g(x)$ for all x . End of proof

So $p(x) = v(x)$, but what does $v(x)$ equal?

By uniqueness principle, we can guess!

A natural guess is that $f(x)$ is linear,

say $f(x) = x/N$. Then $f(0) = 0$ and $f(N) = 1$,

$$\text{and } \frac{f(x-1) + f(x+1)}{2} = \frac{x-1 + x+1}{2N} = \frac{x}{N} = f(x).$$

Thus $p(x) = v(x) = x/N$.

Another application of the uniqueness principle: we prove that walker eventually reaches 0 or N . Let $h(x)$ = probability that starting from x , the walker never reaches 0 or N . Clearly $h(0) = h(N) = 0$.

Also $h(x) = \frac{1}{2} h(x+1) + \frac{1}{2} h(x-1)$ if $0 < x < N$, as can be seen by conditioning on the first step of the walk. So h is harmonic, and by the lemma, $h(x) = 0$ for all x .

Solution as a fair game (martingale)

Think of random walk on $\{0, 1, \dots, N\}$ as a fair game. Peter and Paul have between them N dollars. At each time Peter wins a dollar from Paul with probability $1/2$ or loses a dollar with probability $1/2$. They play until Peter is

ruined (he reaches 0), or he wins 428 all of Paul's money (he reaches N).

then

$p(x)$ = probability that Peter wins if he starts with x dollars.

Now, suppose that Peter has k dollars.

His expected fortune after the next step is

$$\frac{1}{2}(k-1) + \frac{1}{2}(k+1) = k.$$

So the game is fair. A chance process that can be interpreted as a player's fortune in a fair game is called a martingale.

Let $p(x)$ = probability that when Peter starts with x dollars, he wins.

Then Peter's expected final fortune is

$$(1-p(x))0 + p(x)N = p(x)N$$

If we could be sure a fair game remains fair to the end, then Peter's expected final fortune is equal to his starting fortune x .

$$x = p(x)N.$$

$$\text{So } p(x) = x/N$$

Note: this game ends at a random time.

Does fairness of a game ending at a random time get preserved?

Not necessarily For example,

Assume Peter is allowed to go into debt and can play as long as he wants. He starts with 0 dollars and decides to play until his fortune is 1 dollar, and then quit. We will see later that a random walk on the

430

integers, started at 0, will hit 1 if we wait long enough. So with this system of stopping, Peter will always end up a dollar ahead.

BUT we do have the following result.

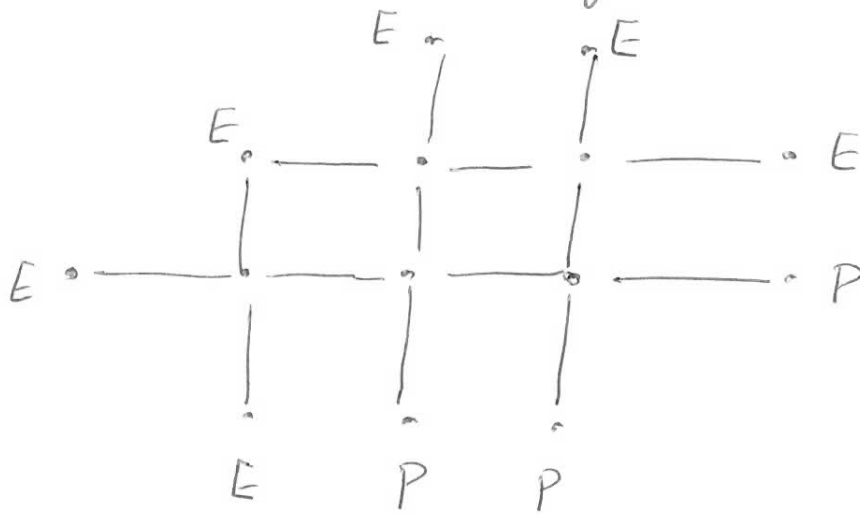
Theorem $\Leftarrow$ "Martingale stopping theorem"

A fair game that is stopped at a random time will remain fair to the end of the game if it is assumed there is a finite amount of money in the world and a player must stop if he wins all this money or goes into debt by this amount.

Note: This theorem justifies the above argument to obtain $p(x) = x/N$.

Random walks in 2 dimensions

An example of random walk on 431
a 2-dimensional array.

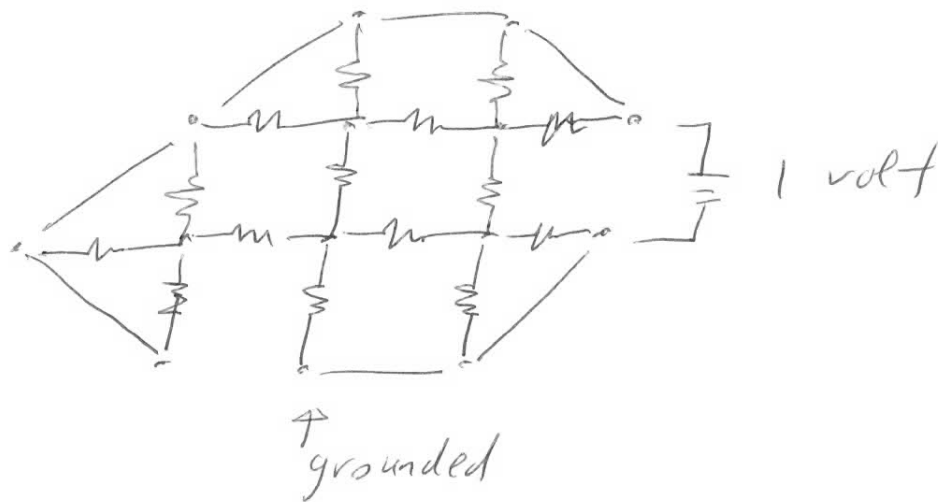


$E = \text{escape}$, $P = \text{police}$

Find $p(x)$, the probability, that started from an interior point x , that reach escape before reaching a policeman. The walker moves from $x = (a, b)$ to each of the four neighboring points $(a+1, b)$, $(a-1, b)$, $(a, b+1)$, $(a, b-1)$ with probability $1/4$. If walker reaches a boundary point, he remains there.

432

The corresponding voltage problem is shown below. The boundary points P are grounded and points E are connected and fixed at one volt by a one-volt battery. We ask for the voltage $v(x)$ at interior points:



Harmonic Functions in 2-dimensions

Have a set of points S , and let D, B be interior and boundary points, defined in the obvious way. A function f defined on S is harmonic if for all points $(a, b) \in D$,

it has the averaging property

433

$$f(a,b) = \frac{f(a+1,b) + f(a-1,b) + f(a,b+1) + f(a,b-1)}{4}$$

We want that $p(x) = v(x)$, as in the 1-dimensional case. That $p(x)$ is harmonic follows again by considering all four possible first steps. That $v(x)$ is harmonic follows again from Kirchhoff's Law, since the current coming into $x = (a,b)$ is

$$\begin{aligned} & \frac{v(a+1,b) - v(a,b)}{R} + \frac{v(a-1,b) - v(a,b)}{R} \\ & + \frac{v(a,b+1) - v(a,b)}{R} + \frac{v(a,b-1) - v(a,b)}{R} \\ & = 0. \end{aligned}$$

Rearranging, we get

$$v(a, b) = \frac{v(a+1, b) + v(a-1, b) + v(a, b+1) + v(a, b-1)}{4}$$

Thus $p(x)$, $v(x)$ are harmonic functions with the same boundary values. We deduce that $p(x) = v(x)$ by extending the maximum principle to two dimensions. If M is the maximum value of f and if $f(P) = M$ for P an interior point, then since $f(P)$ is the average of the values of f at its neighbors, these values must all equal M also. By working our way due South, say, repeating this argument at each step, we eventually reach a boundary point Q for which we can conclude $f(Q) = M$.

To be continued...

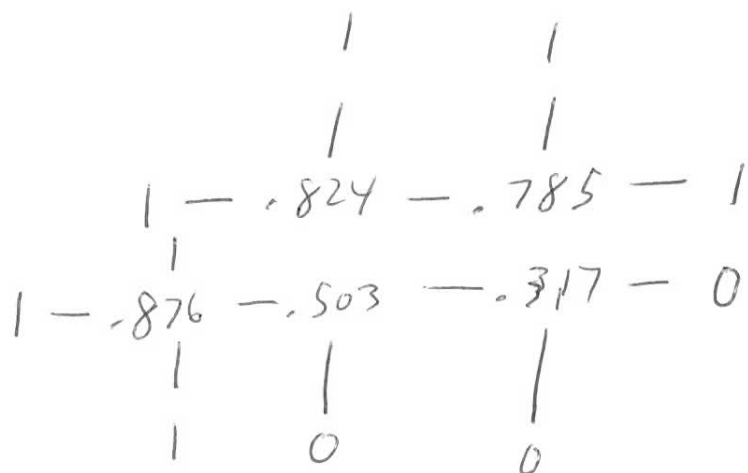
Monte Carlo solution

Starting at x , carry out a random walk until reaching a boundary point. The player then gets paid an amount $f(y)$ if y is the first boundary point reached.

Thus to find $f(x)$, (the expected final winning started at x), we can start many random walks at x and find the average final winnings for these walks. By the law of large numbers, the estimate that we obtain this way will approach the true expected final winning $f(x)$.

Here are some estimates obtained this way by starting 10,000 random walks from each of the interior points

436



This method is quite inefficient. To quantify this, let's look at case with boundary values 1 or 0 as in our example. For each x , let $p(x)$ = probability that hit a 1 started from x .

We estimate this by $\bar{p}$, the fraction of walkers that end in a 1.

Let S_n = number of walkers that hit 1 in n independent trials

By central limit theorem, letting $q = 1 - p$, with $p = p(x)$,

$$P\left(-k < \frac{S_n - np}{\sqrt{npq}} < k\right) \approx A(k)$$

where $A(k)$ is the area under the normal curve between $-k$ and k . For $k=2$, this area is approximately .95.

What does this say about $\bar{p} = S_n/n$? Doing a little rearranging, we see that

$$P\left(-2 < \frac{\bar{p} - p}{\sqrt{pq/n}} < 2\right) \approx .95$$

$$\text{or } P(-2\sqrt{pq/n} < \bar{p} - p < 2\sqrt{pq/n}) \approx .95$$

Since $\sqrt{pq} \leq 1/2$,

$$P(-1/\sqrt{n} < \bar{p} - p < 1/\sqrt{n}) \approx .95$$

So if we choose $1/\sqrt{n} = .01$, or $n = 10,000$ there is a 95 percent chance that our estimate $\bar{p} = S_n/n$ will not be off by more than .01. This is a large number for modest accuracy.

method of relaxations

Recall that we want a function with given boundary values for which the value at any interior point is the average of values at its neighbors.

Begin with any function having the specified boundary values (for instance having value 0 at all interior points). Pick an interior point, and adjust its value to be the average of the values of its neighbors. Run through all the interior points, repeating this process. The function that results won't be harmonic, but will be more nearly harmonic than the function we started with. Keep repeating this averaging process, running through all the interior points again and again.

439

So starting with all interior points

0, we get, after one iteration

$$\begin{array}{ccccccc}
 & & & & 1 & & 1 \\
 & & & & 1 & & 1 \\
 & & 1 & - & .547 & - & .648 & - & 1 \\
 & 1 & - & .75 & - & .188 & - & .047 & - & 0 \\
 & & 1 & & 1 & & 1 \\
 & & 1 & & 0 & & 0
 \end{array}$$

Note that we go from left to right moving up each column replacing each value by the average of its four neighboring values. The computations for the first iteration are:

$$.75 = (1 + 1 + 1 + 0) / 4$$

$$.1875 = (.75 + 0 + 0 + 0) / 4$$

$$.5469 = (.1875 + 1 + 1 + 0) / 4$$

$$.0469 = (.1875 + 0 + 0 + 0) / 4$$

$$.64844 = (.0469 + .5469 + 1 + 1) / 4$$

440

we have written the results to
 3 decimal places. Continue until we
 obtain the same results to 3 decimal
 places. This occurs at iterations 8 and 9.
 We get

$$\begin{array}{ccccccc}
 & & 1 & & 1 & & \\
 & & | & & | & & \\
 1 & - & .823 & - & .787 & - & 1 \\
 & & | & & | & & \\
 1 & - & .826 & - & .506 & - & .323 & - & 0 \\
 & & | & & | & & | & & \\
 & & 1 & & 1 & & 1 & & \\
 & & 1 & & 0 & & 0 & &
 \end{array}$$

Solution by solving linear equations

with same example,

$$\begin{array}{ccccccc}
 & & 1 & & 1 & & \\
 & & | & & | & & \\
 1 & - & .823 & - & .787 & - & 1 \\
 & & | & & | & & \\
 1 & - & .826 & - & .506 & - & .323 & - & 0 \\
 & & | & & | & & | & & \\
 & & 1 & & 1 & & 1 & & \\
 & & 1 & & 0 & & 0 & &
 \end{array}$$

We get the equations

441

$$x_a = \frac{x_b + x_d + 2}{4}$$

$$x_b = \frac{x_a + x_e + 2}{4}$$

$$x_c = \frac{x_d + 3}{4}$$

$$x_d = \frac{x_a + x_c + x_e}{4}$$

$$x_e = \frac{x_b + x_d}{4}$$

can write this as

$$\begin{pmatrix} 1 & -1/4 & 0 & -1/4 & 0 \\ -1/4 & 1 & 0 & 0 & -1/4 \\ 0 & 0 & 1 & -1/4 & 0 \\ -1/4 & 0 & -1/4 & 1 & -1/4 \\ 0 & -1/4 & 0 & -1/4 & 1 \end{pmatrix} \begin{pmatrix} x_a \\ x_b \\ x_c \\ x_d \\ x_e \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \\ 3/4 \\ 0 \\ 0 \end{pmatrix}$$

write this as $Ax = u$. Since there is a unique solution, ~~we can~~ $x = A^{-1}u$. Carrying this out, we calculate

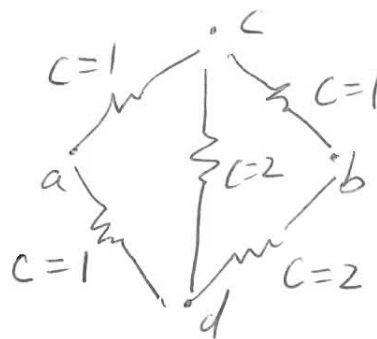
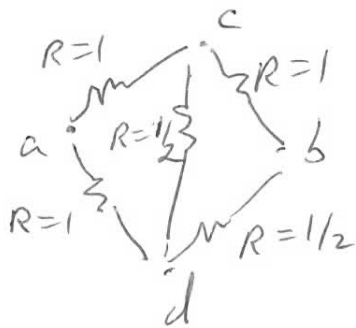
$$x = \begin{pmatrix} .823 \\ -.787 \\ -.876 \\ -.506 \\ -.323 \end{pmatrix}$$

So our Monte Carlo approximation was very good, and our relaxed calculation was perfect.

Random walks on more general networks

Let G be a connected graph, and assign to each edge xy a resistance R_{xy} . Define the conductance of edge xy by $C_{xy} = 1/R_{xy}$.

For example,



Define an associated random walk by

$$P_{xy} = \frac{C_{xy}}{C_x} \quad \text{where} \quad C_x = \sum_y C_{xy}$$

For our example,

443

$$C_a = 2, C_b = 3, C_c = 4, C_d = 5$$

and the transition matrix P is

$$P = \begin{matrix} & \begin{matrix} a & b & c & d \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \end{matrix} & \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 2/3 \\ 1/4 & 1/4 & 0 & 1/2 \\ 1/5 & 2/5 & 2/5 & 0 \end{pmatrix} \end{matrix}$$

Claim: This Markov chain is reversible with respect to the distribution

$$w_x = C_x / C$$

$$\text{where } C = \sum_x C_x.$$

In our example, $C = 14$, so

$$w = \left(\frac{2}{14}, \frac{3}{14}, \frac{4}{14}, \frac{5}{14} \right)$$

In general,

$$\begin{aligned} w_x P_{xy} &= \frac{C_x}{C} \frac{C_{xy}}{C_x} = \frac{C_{xy}}{C} = \frac{C_{yx}}{C} \\ &= \frac{C_y}{C} \frac{C_{yx}}{C_y} = w_y P_{yx} \end{aligned}$$

444

Note: When all conductances on a network are equal, we get simple random walk on the graph G : at each step, move from a vertex to a uniformly chosen random neighbor.

Probabilistic interpretation of voltages

Choose 2 vertices a and b , and put a 1-volt battery across these points establishing a voltage $V_a = 1$, $V_b = 0$. We are interested in finding the voltages V_x and the currents i_{xy} in the circuit, and in giving a probabilistic interpretation to these quantities.

We start with a probabilistic interpretation of voltage. It comes as no surprise that we will interpret the voltage as a hitting probability, observing that both functions are harmonic and have the same boundary values.

By Ohm's law

$$i_{xy} = \frac{v_x - v_y}{R_{xy}} = (v_x - v_y) C_{xy}$$

Note that $i_{xy} = -i_{yx}$

By Kirchhoff's law, for $x \neq a, b$

$$\sum_y i_{xy} = 0.$$

This will be true if $\sum_y (v_x - v_y) C_{xy} = 0$

$$\text{or } v_x \sum_y C_{xy} = \sum_y C_{xy} v_y$$

$$\text{So } v_x = \frac{1}{C_x} \sum_y C_{xy} v_y = \sum_y p_{xy} v_y$$

So v_x is harmonic, for $x \neq a, b$.

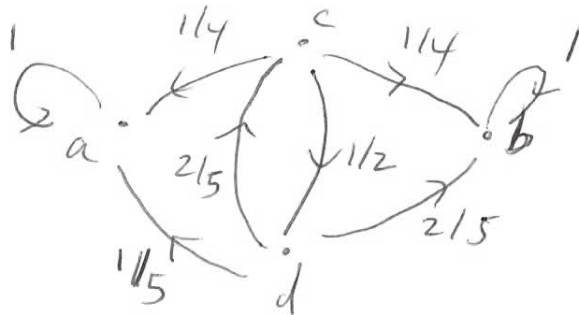
Let h_x be the probability, starting at x , that state a is reached before state b .
Then h_x is also harmonic for $x \neq a, b$ and

$$v_a = h_a = 1, v_b = h_b = 0$$

Thus $v = h$.

Summarizing: when a unit voltage 446
 is applied between points a and b , making
 $V_a = 1$ and $V_b = 0$, the voltage V_x at any
 point x is the probability that a walker
 starting from x will hit a before b .

In our example (modify P to make
 a, b be absorbing)



$$V_a = 1, V_b = 0, V_c = 1/4 + 1/2 V_d, V_d = \frac{1}{5} + \frac{2}{5} V_c$$

solving, we get $V_c = 7/16, V_d = 3/8$

Knowing voltages, can obtain all the currents.

$$\text{For example, } i_{cd} = (7/16 - 3/8) 2 = 1/8$$

Probabilistic interpretation of current

The walker begins at a and walks 447 until he reaches b . Note that if he reaches a before reaching b , he keeps going.

Let

$u_x =$ expected number of visits to state x before reaching b .

Then $u_b = 0$ and for $x \neq a, b$

$$u_x = \sum_y u_y P_{yx} \quad \left(\begin{array}{l} \text{exercise:} \\ \text{to prove this, use fact} \\ \text{that for } x \neq a, b, \text{ every} \\ \text{entrance to } x \text{ must} \\ \text{come from some } y \end{array} \right)$$

We have seen that $C_x P_{xy} = C_y P_{yx}$

$$\text{Thus } u_x = \sum_y u_y P_{xy} C_x / C_y$$

$$\text{or } \frac{u_x}{C_x} = \sum_y P_{xy} \frac{u_y}{C_y}$$

Thus $v_x = u_x / C_x$ is harmonic for $x \neq a, b$.

We also have $v_b = 0$ and $v_a = u_a / C_a$.

448

Thus v_x is the voltage at x when
put a battery from a to b that establishes
a voltage u_a/c_a at a , and a voltage
0 at b .

We are interested in the current that flows
from x to y (with these voltages). That is,

$$i_{xy} = (v_x - v_y) c_{xy} \\ = \left(\frac{u_x}{c_x} - \frac{u_y}{c_y} \right) c_{xy}$$

$$= u_x \frac{c_{xy}}{c_x} - u_y \frac{c_{yx}}{c_y} = u_x p_{xy} - u_y p_{yx}$$

Now $u_x p_{xy}$ is the expected number of times
the walker goes from x to y , and $u_y p_{yx}$
is the expected number of times the walker
goes from y to x . So the current i_{xy} is
the expected value for the net number of
times the walker passes along edge from x to y .

As we have noted, the currents i_{xy} here are not those of the original electric problem, where we used a 1-volt battery.

But they are proportional to the original currents. To find the constant of proportionality, note following property of the new currents i_{xy} :

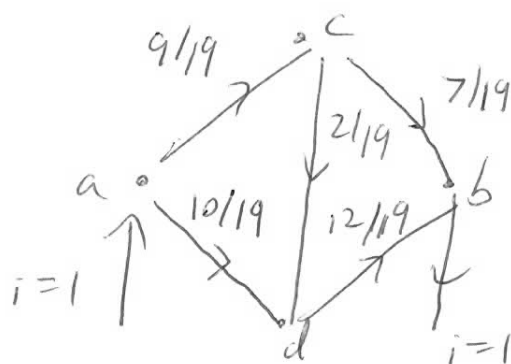
the total current flowing into the network at a (and out at b) is 1. In symbols,

$$\sum_y i_{ay} = 1.$$

Indeed, from the probabilistic interpretation, this sum is the expected value of the difference between the number of times the walker leaves a and enters a . This is necessarily 1.

To be continued

From last time, this unit current flow from a to b can be obtained from the currents in the original circuit corresponding to a 1-volt battery by dividing through by $\sum i_{ay}$, the total current flowing into a . In our example,



This shows that the constant of proportionality we were seeking is the reciprocal of the amount of current that flows through the circuit when a 1-volt battery is applied between a and b . This quantity is called the effective resistance between a and b .

To summarize, we have the following:

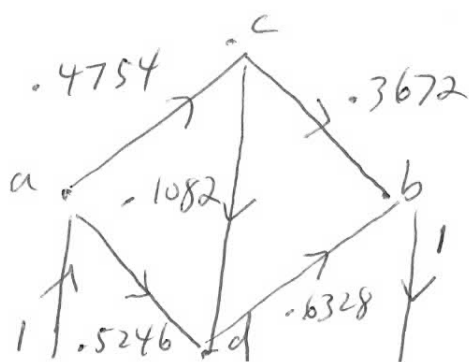
451

Interpretation of current

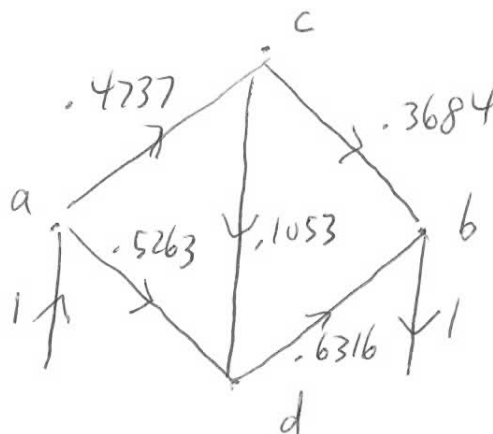
When a unit current flows into a and out of b , the current i_{xy} flowing through edge connecting x to y is equal to the expected net number of times that a walker, starting at a and walking until he reaches b , will move along the edge ~~from~~ x to y . These currents are proportional to the currents that arise when a unit voltage is applied between a and b , the constant of proportionality being the effective resistance of the network.

Note we have seen that we can estimate voltages by simulation. we can now do the same for currents. we have to estimate the expected value for the net number of crossings of the

edge xy . To do this, start a large 452 number of walks at a and, for each one, record the net number of crossings of edges and average there as an estimate for the expected value. Doing this for 10,000 walks yielded the following results.



currents
estimated by
simulation



calculated
currents

Effective resistance and escape probability

When we impose a voltage v between points a and b , a voltage $v_a = v$ is established at a and $v_b = 0$, and a current $i_a = \sum_x i_{ax}$

will flow into the circuit from the outside source. The amount of current that flows depends on the overall resistance of the circuit. We define the effective resistance R_{EFF} between a and b by $R_{EFF} = V_a / i_a$. The quantity $G_{EFF} = 1/R_{EFF} = i_a / V_a$ is called the effective conductance. If the voltage between a and b is multiplied by a constant, then the currents are multiplied by the same constant. So R_{EFF} depends only on the ratio of V_a to i_a .

In our example, when a unit voltage was imposed, the total current flowing into the circuit is $i_a = 9/16 + 10/16 = 19/16$.

~~Therefore~~ So $R_{EFF} = \frac{V_a}{i_a} = \frac{1}{(19/16)} = 16/19$

454

we can interpret the effective conductance probabilistically as an escape probability. When $v_a = 1$, the effective conductance is the total current i_a flowing into a . This current is

$$\begin{aligned} i_a &= \sum_y (v_a - v_y) C_{ay} \\ &= \sum_y (v_a - v_y) \frac{C_{ay}}{C_a} C_a \\ &= C_a \left(1 - \sum_y P_{ay} v_y \right) \\ &= C_a P_{\text{escape}} \end{aligned}$$

where p_{escape} = probability that starting at a , the walker reaches b before returning to a

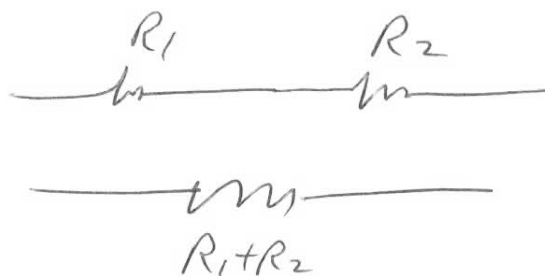
thus $C_{\text{eff}} = C_a p_{\text{escape}}$, and $p_{\text{escape}} = \frac{C_{\text{eff}}}{C_a}$

In our example, $C_a = 2$ and we found that $i_a = 19/16$. So $p_{\text{escape}} = 19/32$.

455

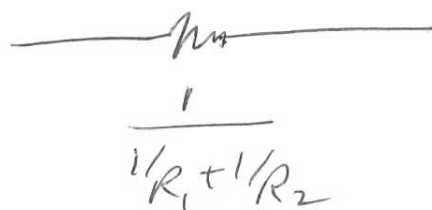
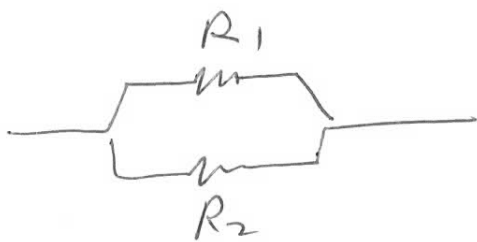
In calculating effective resistances,
we use two important facts about electrical
networks.

First, if two resistors are placed in series,
they may be replaced by a single resistor whose
resistance is the sum of the two resistances.



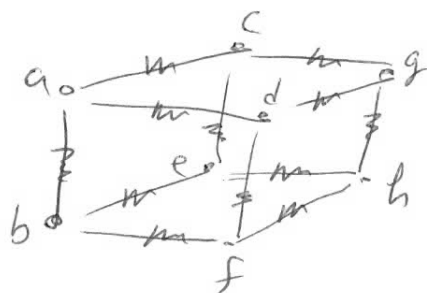
Second, two resistors in parallel may be
replaced by a single resistor whose resistance

R satisfies
$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

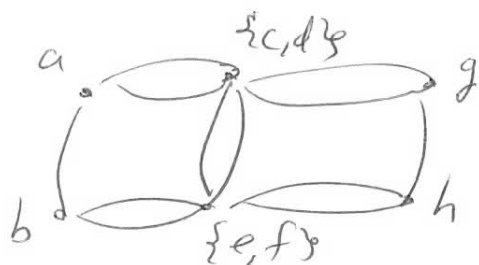


In terms of conductances, two resistors in parallel can be replaced by a single resistor whose conductance is the sum of the two conductances.

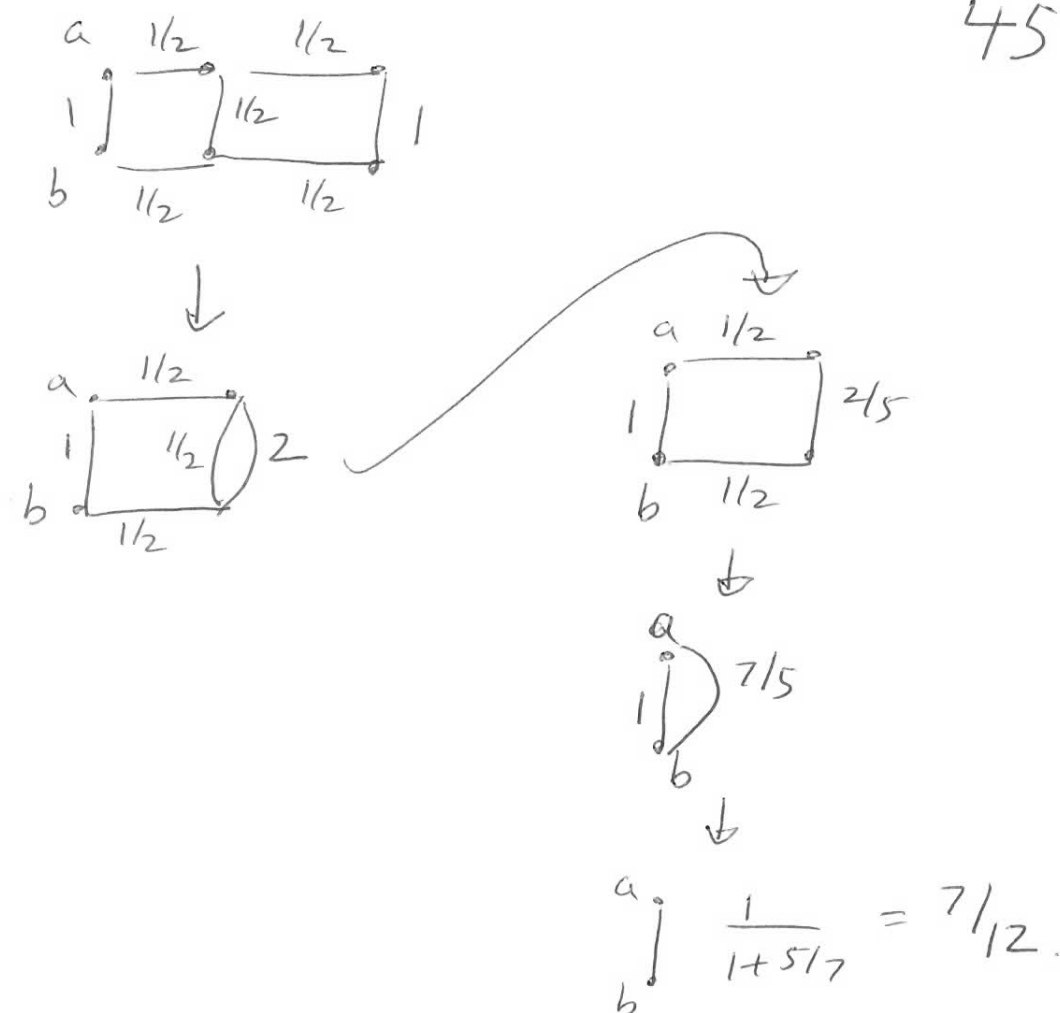
Let's illustrate the use of these ideas to compute the effective resistance between two adjacent points of a unit cube of unit resistors.



Put a unit battery between a and b .
By symmetry, voltages at c, d will be the same as will the voltages at e, f .



457



The effective resistance is $7/12$. The current flowing into a from the battery will be $i_a = 1/R_{\text{eff}} = 12/7$. The probability that a walk starting at a will reach b before returning to a is

$$p_{\text{escape}} = \frac{i_a}{C_a} = \frac{12/7}{3} = 4/7$$

currents minimize energy dissipation

When a current i_{xy} flows through a resistor, the energy dissipated is $i_{xy}^2 R_{xy}$ (this is the product of the current i_{xy} and the voltage difference $v_x - v_y = i_{xy} R_{xy}$).

The total energy dissipated is

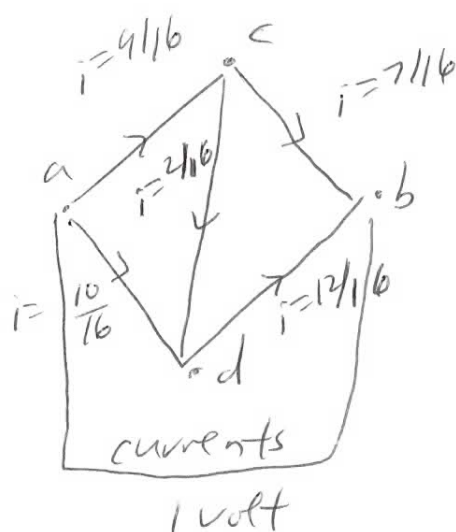
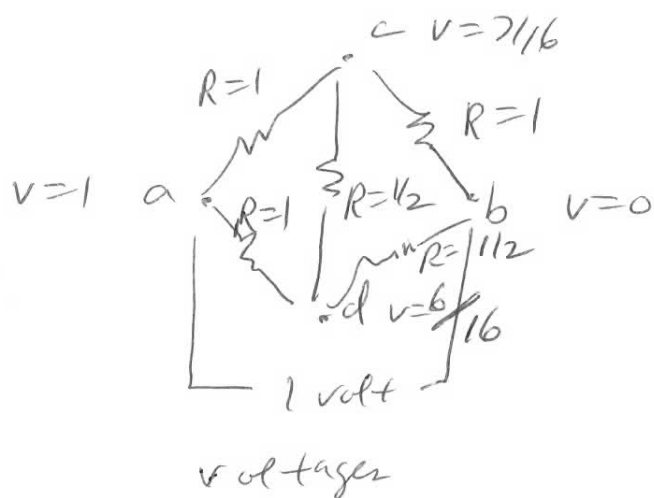
$$E = \frac{1}{2} \sum_{x,y} i_{xy}^2 R_{xy}$$

$$= \frac{1}{2} \sum_{x,y} i_{xy} (v_x - v_y)$$

⚡

We have the factor $1/2$ since each edge is counted twice in the sum.

In our example,



$$\begin{aligned} \text{Thus } E &= 1\left(\frac{9}{16}\right)^2 + 1\left(\frac{10}{16}\right)^2 + 1\left(\frac{7}{16}\right)^2 \\ &\quad + \frac{1}{2}\left(\frac{2}{16}\right)^2 + \frac{1}{2}\left(\frac{12}{16}\right)^2 \\ &= 19/16. \end{aligned} \quad 459$$

If a source (battery) establishes voltages v_a and v_b at a and b , then the energy supplied is $(v_a - v_b) i_a$, where $i_a = \sum_x i_{ax}$.

By conservation of energy, we would expect this to be equal to the energy dissipated.

In our example, $v_a - v_b = 1$ and $i_a = 19/16$,

so this is the case. We shall show this is true in a more general setting.

Define a flow j from a to b to be an assignment of numbers j_{xy} to pairs x, y such that

$$(a) \quad j_{xy} = -j_{yx}$$

$$(b) \quad \sum_y j_{xy} = 0 \quad \text{for } x \neq a, b$$

(c) $j_{xy} = 0$ if x, y aren't adjacent.

We denote by $j_x = \sum_y j_{xy}$ the flow into x from outside. By (b), $j_x = 0$ for $x \neq a, b$. Of course $j_b = -j_a$. To verify this, note that

$$\begin{aligned} j_a + j_b &= \sum_x j_x = \sum_x \sum_y j_{xy} \\ &= \frac{1}{2} \sum_{x,y} (j_{xy} + j_{yx}) \\ &= 0, \text{ since } j_{xy} = -j_{yx}. \end{aligned}$$

We now prove:

CONSERVATION OF ENERGY

Let w be any function on vertices of a graph. Let j be a flow from a to b . Then

$$(w_a - w_b) j_a = \frac{1}{2} \sum_{x,y} (w_x - w_y) j_{xy}$$

Proof

461

$$\sum_{x,y} (w_x - w_y) j_{xy}$$

$$= \sum_x (w_x \sum_y j_{xy}) - \sum_y (w_y \sum_x j_{xy})$$

$$= w_a \sum_y j_{ay} + w_b \sum_y j_{by}$$

$$- w_a \sum_x j_{xa} - w_b \sum_x j_{xb}$$

$$= w_a j_a + w_b j_b - (w_a)(-j_a) - (w_b)(-j_b)$$

$$= 2(w_a - w_b) j_a$$

✓

If we now impose a voltage v_a between a and b with $v_b = 0$, we obtain voltages v_x and currents i_{xy} . The currents give a flow from a to b , so by the previous result,

$$v_a i_a = \frac{1}{2} \sum_{x,y} (v_x - v_y) i_{xy} = \frac{1}{2} \sum_{x,y} i_{xy}^2 R_{xy}.$$



Recall that $R_{\text{eff}} = v_a / i_a$.

Thus we can write

$$i_a^2 R_{\text{eff}} = \frac{1}{2} \sum_{x,y} i_{xy}^2 R_{xy}$$

If we adjust V_a so that $i_a = 1$,
 call the resulting flow the unit current flow
 from a to b . This is a particular example
 of a unit flow from a to b , which we
 define to be any flow i_{xy} from a to b
 for which $i_a = -i_b = 1$. The above formula
 shows that the energy dissipated by the
 unit current flow is just R_{EFF} . By
Thompson's principle, this value is smaller
 than the energy dissipated by any other unit
 flow from a to b .

From earlier example, true currents
 have energy dissipation

$$E = R_{EFF} = 16/19 = -8421053 \dots$$

The approximate currents also form a unit
 flow, and for these, the energy dissipation
 (from simulation)

is

463

$$\begin{aligned}\bar{E} &= 1(.4754)^2 + 1(.5246)^2 + 1(.3672)^2 \\ &\quad + \frac{1}{2}(.1082)^2 + \frac{1}{2}(.6328)^2 \\ &= .8421177\end{aligned}$$

So $\bar{E} \geq E$, though just barely.

We omit the ~~proof~~

New topic : Generating functions

- useful in probability and combinatorics
- we already discussed moment generating functions and cycle indices of the symmetric group
- we will use generating functions to prove the central limit theorem

Suppose want to study a sequence of numbers $a_0, a_1, a_2, \dots$. If you're lucky, there's a

simple formula for a_n , such as $a_n = n^2 + 3$. 464

But often there is not a simple formula.

But maybe there is a simple formula for the sum of a power series, whose coefficients are the sequence that we're looking for.

For example, suppose we want to study the Fibonacci numbers $F_0, F_1, F_2, \dots$ and what we know about them is that they satisfy the recurrence

$$F_{n+1} = F_n + F_{n-1} \quad (n \geq 1, F_0 = F_1 = 1)$$

So the sequence begins

1, 1, 2, 3, 5, 8, 13, 21, ...

(There are exact formulas for F_n which we'll derive later). But using generating functions, one might say that F_n is the coefficient of x^n

in Taylor expansion of $\frac{1}{1-x-x^2}$ about the origin.

Here are some things you'll often be able to do with generating functions of your sequence:

- (a) Find an exact formula for the numbers of your sequence.
- (b) Find a recurrence formula for your sequence
- (c) Find averages and other statistical properties of your sequence
- (d) Find asymptotic formulas for your sequence
- (e) prove unimodality, log-concavity, etc.

(a sequence is unimodal if its entries rise to a maximum and then decrease. for example the sequence $\binom{n}{k}$, $0 \leq k \leq n$ is unimodal).

(f) Prove identities, such as

$$\sum_{j=0}^n \binom{n}{j}^2 = \binom{2n}{n} \quad n=0,1,2,\dots$$

Idea is to show that the sequence on left and sequence on right have same generating function.

Example : an easy two term recurrence

A sequence of numbers $a_0, a_1, \dots$ satisfies the conditions $a_{n+1} = 2a_n + 1$, $n \geq 0$, $a_0 = 0$.

Find the sequence

Sol.

The first terms of the sequence are

0, 1, 3, 7, 15, 31, ...

These look like 1 less than the powers of 2.

So we conjecture that $a_n = 2^n - 1$ and prove this easily by induction using the recurrence.

How can one approach this problem using

generating functions?

467

Let's try to solve for the generating function $A(x) = \sum_{n \geq 0} a_n x^n$. Once we know this, we'll be able to find the explicit formula for the a_n 's by expanding $A(x)$ in a series.

To find $A(x)$, multiply both sides of the recursion $\boxed{a_{n+1} = 2a_n + 1}$ by x^n and sum over the values of n for which the recursion is valid, namely over $n \geq 0$.

Do this to left side of $\boxed{\phantom{a_{n+1} = 2a_n + 1}}$. You get

$$\sum_{n \geq 0} a_{n+1} x^n = a_1 + a_2 x + a_3 x^2 + \dots$$

easy $\rightarrow$
$$= \frac{(a_0 + a_1 x + a_2 x^2 + a_3 x^3) - a_0}{x}$$

since $a_0 = 0 \rightarrow = A(x)/x.$

468

Do this to right side of $\square$. You get

$$\begin{aligned}\sum_{n \geq 0} (2a_{n+1})x^n &= 2A(x) + \sum_{n \geq 0} x^n \\ &= 2A(x) + \frac{1}{1-x}\end{aligned}$$

Thus

$$\frac{A(x)}{x} = 2A(x) + \frac{1}{1-x}$$

Solving for $A(x)$, we get

$$A(x) = \frac{x}{(1-x)(1-2x)}$$

To find an explicit formula for the a_n 's, we expand $A(x)$ in a series. This isn't hard in our example, since the partial fraction expansion is

$$\begin{aligned}\frac{x}{(1-x)(1-2x)} &= x \left[\frac{2}{1-2x} - \frac{1}{1-x} \right] \\ &= (2x + 2^2x^2 + 2^3x^3 + \dots) \\ &\quad - (x + x^2 + x^3 + \dots) \\ &= (2-1)x + (2^2-1)x^2 + (2^3-1)x^3 + \dots\end{aligned}$$

So clearly the coefficient of x^n is 469
equal to $2^n - 1$, for each $n \geq 0$.

Example : a slightly harder two term
recurrence

A sequence $a_0, a_1, \dots$ satisfies the
conditions $\boxed{a_{n+1} = 2a_n + n}$ $n \geq 0, a_0 = 1$

Sol.

The first members of the sequence are

1, 2, 5, 12, 27, 58, 121, ...

and it's not obvious what an explicit formula
is for this sequence.

$$\text{let } A(x) = \sum_{n \geq 0} a_n x^n$$

Multiply left hand side of $\boxed{\phantom{a_{n+1} = 2a_n + n}}$ by x^n
and sum over all n for which the recurrence
is valid.

$$\text{we get } a_1 + a_2 x + a_3 x^2 + a_4 x^3 + \dots$$

470

$$= (A(x) - a_0) / x$$

$$= (A(x) - 1) / x$$

Now do the same for right hand side of

$$\boxed{}. \text{ You get } 2A(x) + \sum_{n \geq 0} nx^n.$$

So we need to identify the series

$$\sum_{n \geq 0} nx^n = x + 2x^2 + 3x^3 + \dots$$

You can either look this up or work it out.

To work it out,

$$\sum_{n \geq 0} nx^n = \sum_{n \geq 0} x \frac{d}{dx} (x^n)$$

$$= x \frac{d}{dx} \sum_{n \geq 0} x^n$$

$$= x \frac{d}{dx} \frac{1}{1-x} = \frac{x}{(1-x)^2}$$

$$\text{Thus } \frac{(A(x)-1)}{x} = 2A(x) + \frac{x}{(1-x)^2}$$

$$\text{Hence } A(x) = (1 - 2x + 2x^2) / ((1-x)^2(1-2x))$$

So we found the generating function ⁴⁷¹ of our sequence. Let's use it to find an explicit formula for a_n .

Let's expand $(1-2x+2x^2)/(1-x)^2(1-2x)$ in partial fractions:

$$(*) \quad \frac{1-2x+2x^2}{(1-x)^2(1-2x)} = \frac{A}{(1-x)^2} + \frac{B}{(1-x)} + \frac{C}{(1-2x)}$$

One way to find A, B, C is as follows.

Multiply both sides of $*$ by $(1-x)^2$ and set $x=1$.

You instantly get $A = -1$. Next multiply

$*$ by $1-2x$ and set $x=1/2$. You instantly get

$C=2$. The hard one to find is B , so let's

do that by cheating. Setting $x=0$ in $*$, and

knowing the values of A and C , we

find that $B=0$.

Thus

$$A(x) = \frac{-1}{(1-x)^2} + \frac{2}{1-2x}$$

Coefficient of x^n in $\frac{2}{1-2x}$ is 2^{n+1}

Coefficient of x^n in $\frac{-1}{(1-x)^2}$ is $-(n+1)$

So $a_n = 2^{n+1} - n - 1, n \geq 0.$

Notation : Let $f(x)$ be a power series in x .

Let $[x^n] f(x)$ denote the coefficient of x^n in the series $f(x)$. For example if s is a positive integer, $[x^n] (1+x)^s = \binom{s}{n}$

Example : A three term recurrence

Consider the Fibonacci recurrence

$$\boxed{F_{n+1} = F_n + F_{n-1}} \quad n \geq 1, F_0 = F_1 = 1$$

$$\text{let } F(x) = \sum_{n \geq 0} F_n x^n.$$

multiply left hand side of $\square$ by x^n
and sum over all $n \geq 1$. we get

$$F_2 x + F_3 x^2 + F_4 x^3 + \dots = \frac{F(x) - 1 - x}{x}$$

On right side, we get

$$\begin{aligned} (F_1 x + F_2 x^2 + F_3 x^3 + \dots) + (F_0 x + F_1 x^2 + F_2 x^3 + \dots) \\ = (F(x) - 1) + (x F(x)) \end{aligned}$$

$$\text{Thus } \frac{F(x) - 1 - x}{x} = F(x) - 1 + x F(x),$$

$$\text{so } F(x) = \frac{1}{1 - x - x^2}$$

To use partial fractions, write

$$1 - x - x^2 = (1 - x r_+)(1 - x r_-)$$

$$\text{where } r_+ = (1 + \sqrt{5})/2, \quad r_- = (1 - \sqrt{5})/2$$

$$\text{Thus, } \frac{1}{1 - x - x^2} = \frac{1}{(1 - x r_+)(1 - x r_-)}$$

$$= \frac{1}{(r_+ - r_-)} \left(\frac{r_+}{1 - x r_+} - \frac{r_-}{1 - x r_-} \right)^{474}$$

$$= \frac{1}{\sqrt{5}} \left(\sum_{j \geq 0} r_+^{j+1} x^j - \sum_{j \geq 0} r_-^{j+1} x^j \right)$$

$$\text{So } F_n = \frac{1}{\sqrt{5}} (r_+^{n+1} - r_-^{n+1})$$

Note

1) Since $r_+ > 1$ and $|r_-| < 1$, we get that

$$\text{for large } n, \quad F_n \sim \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{n+1}$$

is asymptotic
to

2) In fact, since F_n is an integer and

$$\frac{r_-^{n+1}}{\sqrt{5}} < 1/2, \text{ it follows that } F_n \text{ is the}$$

nearest integer to $\frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{n+1}$

Example (with two independent variables)

475

Let $f(n, m) = \binom{n}{m}$.

It's easy to verify that

$$\boxed{f(n, m) = f(n-1, m) + f(n-1, m-1)}, \quad (n, m) \neq (0, 0) \\ f(0, 0) = 1.$$

There are at least 3 choices of generating functions that come to mind:

$$F(x, y) = \sum_{n, m} f(n, m) x^n y^m$$

$$G_n(y) = \sum_m f(n, m) y^m \quad (\text{integer } n)$$

$$H_m(x) = \sum_n f(n, m) x^n \quad (\text{integer } m)$$

Let's work with each of these.

working with $F(x, y)$

Multiply $\boxed{}$ by $x^n y^m$ and sum over all $(n, m) \neq (0, 0)$. We get

$$F(x, y) - 1 = \sum_{n, m} f(n-1, m) x^n y^m \\ + \sum_{n, m} f(n-1, m-1) x^n y^m$$

$$\text{So } F(x, y) - 1$$

$$= x \sum_{r, m} f(r, m) x^r y^m + xy \sum_{r, s} f(r, s) x^r y^s$$

$$= x F(x, y) + xy F(x, y)$$

$$\text{So } F(x, y) = \frac{1}{1-x-xy} \text{ and thus}$$

$$f(n, m) = [x^n y^m] \frac{1}{1-x-xy}$$

working with $G_n(y)$

Fix $n \neq 0$, multiply $\square$ by y^m and sum over all m . The result would be

$$G_n(y) = G_{n-1}(y) + y G_{n-1}(y) \quad (n \neq 0)$$

From which each $G_n(y)$ is $(1+y)$ times the preceding $G_{n-1}(y)$. Since $G_0(y) = 1$, it follows that $G_n(y) = (1+y)^n$, $n \geq 0$, so

$$f(n, m) = [y^m] (1+y)^n \quad n \geq 0$$

Working with $H_m(x)$

Fix $m \neq 0$, multiply $\square$ by x^n and sum over all n . We get

$$H_m(x) = x H_m(x) + x H_{m-1}(x) \quad m \neq 0$$

$$\text{So } H_m(x) = \frac{x}{1-x} H_{m-1}(x) \text{ for } m \geq 1$$

But what is H_0 ? It is

$$H_0(x) = \sum_n f(n,0) x^n = \sum_{n \geq 0} x^n = \frac{1}{1-x}$$

$$\text{Thus } H_m(x) = \frac{x^m}{(1-x)^{m+1}} \text{ for all } m \geq 0,$$

$$\begin{aligned} \text{So } f(n,m) &= [x^n] \frac{x^m}{(1-x)^{m+1}} \\ &= [x^{n-m}] \left(\frac{1}{1-x} \right)^{m+1} \end{aligned}$$

Let's see how to get formulas for $f(n,m)$ from these generating functions.

We need to find series expansions of a given function. By calculus, if $f(x)$ is given and we want to write $f(x) = \sum_{n \geq 0} a_n x^n$, use Taylor's formula

$$a_n = \frac{1}{n!} \left(\frac{d}{dx} \right)^n f(x) \Big|_{x=0}.$$

The problem with this is that successive derivatives might become harder and harder to handle, for instance for the Fibonacci generating function $1/(1-x-x^2)$.

But in the present example, Taylor's formula works fine: using formula for $H_m(x)$, we get

$$f(n, m) = [x^{n-m}] \left(\frac{1}{1-x} \right)^{m+1}$$

$$\text{Now } \left(\frac{d}{dx} \right)^n \frac{1}{(1-x)^{m+1}} = \frac{(m+1)(m+2) \cdots (m+n)}{(1-x)^{m+n+1}}$$

$$\text{So } [x^n] \left(\frac{1}{(1-x)^{m+1}} \right) = \frac{(m+n)!}{m! n!} \quad m, n \geq 0$$

Thus $f(n, m) = [x^{n-m}] \left(\frac{1}{1-x} \right)^{m+1}$

$$= \frac{n!}{(n-m)! m!} = \binom{n}{m}$$

Similarly, could use Taylor's formula on $G_n(y)$.

Using $F(x, y)$, get

$$\begin{aligned} \frac{1}{1-x-xy} &= \sum_{r \geq 0} (x+xy)^r \\ &= \sum_{r \geq 0} x^r (1+y)^r \end{aligned}$$

$$\begin{aligned} \text{So } f(n, m) &= [x^n y^m] \frac{1}{1-x-xy} \\ &= [x^n y^m] \sum_{r \geq 0} x^r (1+y)^r \\ &= [y^m] (1+y)^n. \end{aligned}$$

Next time : study set partitions

MATH 407 Lecture 33

480

A set partition of a finite set S is a collection of nonempty, pairwise disjoint sets whose union is S . For example, there are 5 set partitions of $\{1, 2, 3\}$, namely:

$$\{1, 2, 3\}, \{1\}\{2, 3\}, \{2\}\{1, 3\}, \{3\}\{1, 2\}, \{1\}\{2\}\{3\}$$

Note that the order of the elements in each block don't matter, and neither does the order of the blocks.

There are 7 set partitions of $\{1, 2, 3, 4\}$ into 2 blocks, namely

$$\{1, 2\}\{3, 4\}, \{1, 3\}\{2, 4\}, \{1, 4\}\{2, 3\}, \{1, 2, 3\}\{4\}, \{1, 2, 4\}\{3\}, \\ \{1, 3, 4\}\{2\}, \{1\}\{2, 3, 4\}$$

Let $\{S_k^n\}$ denote the number of set partitions of $\{1, \dots, n\}$ into k blocks.

Let's find a recurrence for the numbers $\{S_k^n\}$.

481

Split the set partitions of $\{1, \dots, n\}$ into two types: those in which n lives in a block by itself, and those in which n lives in a block with other numbers.

Clearly the number of set partitions in which n is its own block is $\begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix}$.

Also, the number of set partitions into k blocks where n isn't in its own block is $k \begin{Bmatrix} n-1 \\ k \end{Bmatrix}$, since we have $\begin{Bmatrix} n-1 \\ k \end{Bmatrix}$ set partitions of $\{1, \dots, n-1\}$ into k blocks, and have k ways of adding ~~the~~ n to one of these blocks.

Thus

$$\begin{Bmatrix} n \\ k \end{Bmatrix} = \begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix} + k \begin{Bmatrix} n-1 \\ k \end{Bmatrix}, \quad (n, k) = \text{????}$$

to determine the range of n and k , extend the definition of $\begin{Bmatrix} n \\ k \end{Bmatrix}$ to all pairs of integers. we put $\begin{Bmatrix} n \\ k \end{Bmatrix} = 0$ if $k > n$

or if $n < 0$ or if $k < 0$. Further,

$\{n \atop 0\} = 0$ if $n \neq 0$ and we take $\{0 \atop 0\} = 1$.

With these conventions, the above recurrence is valid for all (n, k) other than $(0, 0)$, and we have

$$\boxed{\{n \atop k\} = \{n-1 \atop k-1\} + k \{n-1 \atop k\}}, \quad (n, k) \neq (0, 0)$$

$\{0 \atop 0\} = 1.$

To study this by generating functions, one could use several approaches:

$$A_n(y) = \sum_k \{n \atop k\} y^k$$

$$B_k(x) = \sum_n \{n \atop k\} x^n$$

$$C(x, y) = \sum_{n, k} \{n \atop k\} x^n y^k$$

Let's think about which choice is likely to succeed. To find $A_n(y)$ will involve multiplying

by y^k and summing in k . So we

would have to deal with terms like $\sum_k k \{k\}^n y^k$. This is possible to think about since it's related to the derivative of $A_n(y)$, but it is a complication. As we will see, this complication goes away if we deal with $B_k(x)$.

So multiplying $\square$ by x^n and summing over n , we get

$$B_k(x) = x B_{k-1}(x) + kx B_k(x) \quad k \geq 1, \\ B_0(x) = 1.$$

this leads to

$$B_k(x) = \frac{x}{1-kx} B_{k-1}(x) \quad k \geq 1, B_0(x) = 1$$

and finally to the evaluation

$$B_k(x) = \sum_n \{k\}^n x^k = \frac{x^k}{(1-x)(1-2x)\cdots(1-kx)} \quad k \geq 0$$

let's study this using partial fractions,
not Taylor's formula.

The partial fraction expansion has the form

484

$$\frac{1}{(1-x)(1-2x)\cdots(1-kx)} = \sum_{j=1}^k \frac{\alpha_j}{1-jx}$$

To find the α_j 's, for $1 \leq r \leq k$, multiply both sides by $(1-rx)$ and let $x=1/r$. The result is that

$$\begin{aligned} \alpha_r &= \frac{1}{(1-1/r)(1-2/r)\cdots(1-\frac{r-1}{r})(1-\frac{r+1}{r})\cdots(1-k/r)} \\ &= (-1)^{k-r} r^{k-1} / (r-1)!(k-r)! \quad 1 \leq r \leq k \end{aligned}$$

thus

$$\begin{aligned} \left\{ \begin{matrix} n \\ k \end{matrix} \right\} &= [X^n] \frac{x^k}{(1-x)(1-2x)\cdots(1-kx)} \\ &= [X^{n-k}] \frac{1}{(1-x)(1-2x)\cdots(1-kx)} \\ &= [X^{n-k}] \sum_{r=1}^k \frac{\alpha_r}{1-rx} \quad (k \geq 1) \\ &= \sum_{r=1}^k \alpha_r [X^{n-k}] \frac{1}{1-rx} \\ &= \sum_{r=1}^k \alpha_r r^{n-k} \end{aligned}$$

$$= \sum_{r=1}^k (-1)^{k-r} \frac{r^{k-1}}{(r-1)!(k-r)!} r^{n-k}$$

485

$$= \sum_{r=1}^k (-1)^{k-r} \frac{r^{n-1}}{(r-1)!(k-r)!} \quad (n, k > 0)$$

You can check this explicit formula to see that $\{4\}_2 = 7$. Also note that the formula gives $\{n\}_2 = 2^{n-1} - 1$ ($n > 0$). Can you see this directly?

Suppose we want to study the Stirling numbers using the functions $A_n(y)$. This is harder to carry out, but yields different insights.

Fix $n > 0$, multiply $\boxed{}$ by y^k and sum over k . The result is

$$\begin{aligned} A_n(y) &= \sum_k \{n-1\}_{k-1} y^k + \sum_k k \{n-1\}_k y^k \\ &= y A_{n-1}(y) + (y \frac{d}{dy}) A_{n-1}(y) \\ &= \{y(1 + D_y)\} A_{n-1}(y), \quad n > 0 \\ &\quad A_0(y) = 1 \end{aligned}$$

starting with $A_0 = 1$, we obtain successively $y, y + y^2, y + 3y^2 + y^3, \dots$

but as far as an explicit formula is concerned, we find only that

$$A_n(y) = \{y(1 + D_y)\}^n 1 \quad n \geq 0.$$

Can use this to prove that the sequence $\{y_k^n\}_{k=0}^n$ is unimodal.

Let's now apply our formula

$$\boxed{\{y_k^n\} = \sum_{r=1}^k (-1)^{k-r} \frac{r^{n-1}}{(r-1)!(k-r)!}} \quad n, k > 0$$

Let $b(n) = n$ th "Bell number"

= number of set partitions
of $\{1, \dots, n\}$

$$= \sum_{k=1}^n \{y_k^n\}$$

The sequence of Bell numbers begins

1, 1, 2, 5, 15, 52, ...

Thus for any $M \geq n$, $\boxed{}$ gives

$$b(n) = \sum_{k=1}^M \sum_{r=1}^k (-1)^{k-r} \frac{r^{n-1}}{(r-1)!(k-r)!}$$

$$= \sum_{r=1}^M \frac{r^{n-1}}{(r-1)!} \sum_{k=r}^M \frac{(-1)^{k-r}}{(k-r)!}$$

$$= \sum_{r=1}^M \frac{r^{n-1}}{(r-1)!} \sum_{s=0}^{M-r} \frac{(-1)^s}{s!}$$

Letting $M \rightarrow \infty$, we get

$$\boxed{b(n) = \frac{1}{e} \sum_{r=1}^{\infty} \frac{r^{n-1}}{(r-1)!}}, \quad n \geq 1$$

This formula for the Bell numbers, while lovely, isn't so useful for computation.

However, we can use it to derive a nice formula for the exponential generating

function

$$B(x) = \sum_{n \geq 0} \frac{b(n)}{n!} x^n$$

Multiply $\boxed{}$ by $x^n/n!$ and sum over $n \geq 1$

This gives

$$\begin{aligned}
 B(x)-1 &= \frac{1}{e} \sum_{n \geq 1} \frac{x^n}{n!} \sum_{r \geq 1} \frac{r^{n-1}}{(r-1)!} \\
 &= \frac{1}{e} \sum_{r \geq 1} \frac{1}{r!} \sum_{n \geq 1} \frac{(rx)^n}{n!} \\
 &= \frac{1}{e} \sum_{r \geq 1} \frac{1}{r!} (e^{rx} - 1) \\
 &= \frac{1}{e} (e^{e^x} - e) \\
 &= e^{e^x-1} - 1
 \end{aligned}$$

We have thus proved:

Theorem The exponential generating function of the Bell numbers is e^{e^x-1} , that is, the coefficient of $x^n/n!$ in the power-series expansion of e^{e^x-1} is the number of set partitions of an n -element set.

✓ Thus, while the Bell numbers are complicated, their exponential generating function is simple.

489

Let's use this theorem to find a recursion for the Bell numbers. First, the theorem tells us that

$$\sum_{n \geq 0} \frac{b(n)}{n!} x^n = e^{e^x - 1}$$

Carry out the " $x \frac{d}{dx} \log$ " operation

- 1) Take the logarithm of both sides of the equation
- 2) Differentiate both sides and multiply through by x
- 3) Clear the equation of fractions
- 4) For each n , find the coefficients of x^n on both sides of the equation, and equate them.

Why is this useful? The best reason is that it works. The point of taking logarithms is to simplify the function $e^{e^x - 1}$, whose

power series coefficients are quite mysterious before taking the logarithm, but quite obvious after doing so. The price we pay is that the left side becomes a log of a sum. The next step, the differentiation, changes the log of the sum into the ratio of two sums, which is nicer. The reason for multiplying by x at the end is that the differentiation dropped the power of x by 1, and it's handy to restore it.

In our example, step 1 gives

$$\log\left(\sum_{n \geq 0} \frac{b(n)}{n!} x^n\right) = e^x - 1$$

Step 2 gives

$$\frac{\sum_n n b(n) x^n / n!}{\sum_n b(n) x^n / n!} = x e^x$$

To clear fractions, multiply both sides by the

denominator on the left, obtaining

$$\sum_n \frac{n b(n) x^n}{n!} = (x e^x) \sum_n \frac{b(n) x^n}{n!}$$

Comparing coefficients of x^n on both sides of this equation, we conclude, after some work, that

$$b(n) = \sum_k \binom{n-1}{k} b(k) \quad n \geq 1, \quad b(0) = 1.$$

Snake oil method for doing combinatorial sums

- Identify the free variable, say n , that the sum depends on. Give a name to the sum that you're working on; call it $f(n)$.
- Let $F(x) = \sum_n x^n f(n)$
- Multiply the sum by x^n and sum over n .
You get a double sum over n and over whatever variable was first used

as a dummy summation variable. 492

d) Interchange the order of summation and try to perform the inner summation in simple closed form.

e) try to identify the coefficients of the generating function of the answer, since these are the coefficients you want to find

Example

Consider the sum $\sum_{k \geq 0} \binom{k}{n-k}$ ($n=0, 1, 2, \dots$)

The free variable is n , so let's call the sum $f(n)$. Write it as

$$f(n) = \sum_{k \geq 0} \binom{k}{n-k}$$

Now multiply both sides by x^n and sum over n . So we get

$$\begin{aligned}
F(x) &= \sum_n x^n \sum_{k \geq 0} \binom{k}{n-k} \\
&= \sum_{k \geq 0} \sum_n \binom{k}{n-k} x^n \\
&= \sum_{k \geq 0} x^k \sum_n \binom{k}{n-k} x^{n-k} \\
&= \sum_{k \geq 0} x^k \sum_r \binom{k}{r} x^r \\
&= \sum_{k \geq 0} x^k (1+x)^k \\
&= \sum_{k \geq 0} (x+x^2)^k = \frac{1}{1-x-x^2}
\end{aligned}$$

We recognize this as the generating function for Fibonacci numbers, so

$$\sum_{k \geq 0} \binom{k}{n-k} = F_n \quad n=0, 1, 2, \dots$$

Example

Consider the sum $\sum_k \binom{n+k}{m+2k} \binom{2k}{k} \frac{(-1)^k}{k+1}$.

Then multiplying by x^n and summing over $n \geq 0$, we get

$$F(x) = \sum_{n \geq 0} x^n \sum_k \binom{n+k}{m+2k} \binom{2k}{k} \frac{(-1)^k}{k+1}$$

$$= \sum_k \binom{2k}{k} \frac{(-1)^k}{k+1} x^{-k} \sum_{n \geq 0} \binom{n+k}{m+2k} x^{n+k}$$

$$= \sum_k \binom{2k}{k} \frac{(-1)^k}{k+1} x^{-k} \sum_{r \geq k} \binom{r}{m+2k} x^r$$

(since $\sum_{r \geq 0} \binom{r}{k} x^r = \frac{x^k}{(1-x)^{k+1}}$ we get)

$$= \sum_k \binom{2k}{k} \frac{(-1)^k}{k+1} x^{-k} \frac{x^{m+2k}}{(1-x)^{m+2k+1}}$$

$$= \frac{x^m}{(1-x)^{m+1}} \sum_k \binom{2k}{k} \frac{1}{k+1} \left(\frac{-x}{(1-x)^2} \right)^k$$

exercise $\frac{-x^{m-1}}{2(1-x)^{m-1}} \left[1 - \sqrt{1 + \frac{4x}{(1-x)^2}} \right]$

$$= \frac{-x^{m-1}}{2(1-x)^{m-1}} \left[1 - \frac{1+x}{1-x} \right] = x^m / (1-x)^m$$

So sum = $[x^m] \frac{x^m}{(1-x)^m} = \binom{n-1}{m-1}$.

Example

Evaluate the sums

$$f_n = \sum_k \binom{n+k}{2k} 2^{n-k}, \quad n \geq 0.$$

Sol. Let $F = \sum_{n \geq 0} f_n x^n$.

$$\text{So } F = \sum_k 2^{-k} \sum_{n \geq 0} \binom{n+k}{2k} 2^n x^n$$

$$= \sum_k 2^{-k} (2k)^{-k} \sum_{n \geq 0} \binom{n+k}{2k} (2x)^{n+k}$$

$$= \sum_{k \geq 0} 2^{-k} (2x)^{-k} \frac{(2x)^{2k}}{(1-2x)^{2k+1}}$$

↗

$$\left(\text{since } \sum_{r \geq 0} \binom{r}{k} x^r = \frac{x^k}{(1-x)^{k+1}} \right)$$

$$= \frac{1}{1-2x} \sum_{k \geq 0} \left[\frac{x}{(1-2x)^2} \right]^k$$

$$= \frac{1}{1-2x} \frac{1}{1 - \frac{x}{(1-2x)^2}}$$

496

$$= \frac{1-2x}{(1-4x)(1-x)}$$

$$= \frac{2}{3(1-4x)} + \frac{1}{3(1-x)}$$

Taking the coefficient of x^n on both sides gives

$$\sum_k \binom{n+k}{2k} 2^{n-k} = \frac{2^{2n+1} + 1}{3} \quad (n \geq 0).$$

Example

Let's try to prove that

$$\boxed{\sum_k \binom{m}{k} \binom{n+k}{m} = \sum_k \binom{m}{k} \binom{n}{k} 2^k} \quad m, n \geq 0$$

without being able to evaluate either of these sums.

Multiply the left of $\boxed{\phantom{\sum_k \binom{m}{k} \binom{n+k}{m}}}$ by x^n , sum over $n \geq 0$, and interchange the summations to arrive at

$$\sum_k \binom{m}{k} x^{-k} \sum_{n \geq 0} \binom{n+k}{m} x^{n+k}$$

$$= \sum_k \binom{m}{k} x^{-k} \frac{x^m}{(1-x)^{m+1}}$$

$$= \frac{x^m}{(1-x)^{m+1}} (1 + 1/x)^m$$

$$= (1+x)^m / (1-x)^{m+1}$$

Doing the same to the right of $\square$,
we obtain

$$\sum_k \binom{m}{k} 2^k \sum_{n \geq 0} \binom{n}{k} x^n$$

$$= \frac{1}{(1-x)} \sum_k \binom{m}{k} \left(\frac{2x}{(1-x)} \right)^k$$

$$= \frac{1}{(1-x)} \left(1 + \frac{2x}{1-x} \right)^m$$

$$= (1+x)^m / (1-x)^{m+1}$$

Hence the two sums are equal, even if we don't know what they are.

Example Proving the central limit theorem

using generating functions.

498

Lemma

Assume that the moment generating functions of the random variables $Y_1, Y_2, Y_3, \dots$

satisfy

$$\lim_{n \rightarrow \infty} M_{Y_n}(t) = e^{t^2/2}$$

↑
(the moment generating function of a standard normal).

for all t in an interval $(-t, t)$ for some $\epsilon > 0$. Then for any real number a ,

$$\lim_{n \rightarrow \infty} P(Y_n \leq a) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

Central limit theorem

Suppose we have independent, identically distributed random variables $X_1, X_2, X_3, \dots$ with finite mean $E[X_1] = \mu$ and finite

variance $\text{Var}(X_i) = \sigma^2$. Let

$S_n = X_1 + \dots + X_n$. Then for any fixed

$-\infty \leq a < b \leq \infty$, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \\ = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \end{aligned}$$

Sketch of proof

Assume that $M_{X_i}(t)$ is finite and take

$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. We prove that

$$\lim_{n \rightarrow \infty} M_{Y_n}(t) = e^{t^2/2},$$

and the result follows from the lemma.

From the definition of the moment generating function and the fact that

$X_1, \dots, X_n$ are independent,

$$M_{Y_n}(t) = E[e^{tY_n}] = E\left[e^{t \frac{S_n - n\mu}{\sigma\sqrt{n}}}\right]$$

$$\begin{aligned}
 &= E \left[\exp \left(\frac{\pm}{\sigma \sqrt{n}} \sum_{k=1}^n (X_k - \mu) \right) \right] \\
 &= E \left[\prod_{k=1}^n e^{\frac{\pm}{\sigma \sqrt{n}} (X_k - \mu)} \right] \\
 &= \left[\prod_{k=1}^n E \left[e^{\frac{\pm}{\sigma \sqrt{n}} (X_k - \mu)} \right] \right]
 \end{aligned}$$

Inside each expectation we insert the second order Taylor approximation

$$e^x \approx 1 + x + x^2/2$$

which is valid near 0 and then use additivity of expectation :

$$\begin{aligned}
 &E \left[e^{\frac{\pm}{\sigma \sqrt{n}} (X_k - \mu)} \right] \\
 &\approx E \left[1 + \frac{\pm}{\sigma \sqrt{n}} (X_k - \mu) + \frac{\pm^2}{2\sigma^2 n} (X_k - \mu)^2 \right] \\
 &= 1 + \frac{\pm}{\sigma \sqrt{n}} E(X_k - \mu) + \frac{\pm^2}{2\sigma^2 n} E[(X_k - \mu)^2] \\
 &= 1 + \frac{\pm}{\sigma \sqrt{n}} 0 + \frac{\pm^2}{2\sigma^2 n} \sigma^2 = 1 + \frac{\pm^2}{2n}
 \end{aligned}$$

501
plus this back into $\square$ to
obtain

$$M_{Y_n}(t) \approx \left(1 + \frac{t^2}{2n}\right)^n \rightarrow e^{t^2/2} \text{ as } n \rightarrow \infty.$$

Making this proof rigorous requires
bounds for the $\approx$ steps. Note that as $n \rightarrow \infty$
the Taylor approximation gains accuracy
because the quantity $\frac{t}{6\sqrt{n}} (X_k - \mu)$ is
converging to 0.

Back to Markov chains

Consider Markov chains on a finite or
countable number of states.

Classification of states

State j is said to be accessible from
state i if $P_{ij}^n > 0$ for some $n \geq 0$
 $\nearrow$
= chance of going from i to j in n steps

Two states that are accessible from each other are said to communicate.

Two states that communicate are said to be in the same class. It's easy to see that any two classes of states are either identical or disjoint. In other words, the concept of communication divides the state space up into a number of separate classes. The Markov chain is called irreducible if there is only one class, that is if all states communicate with each other.

For any state i , let f_i be probability that starting at state i , the process will eventually reenter state i . State i is said to be recurrent if $f_i = 1$ and transient if $f_i < 1$.

Suppose the process starts in state i and is recurrent. Hence with probability 1, the process will eventually reenter state i . By definition of a Markov chain, the process will start over again when the chain reenters state i , and therefore state i will be visited yet again. Continuing, we conclude that if state i is recurrent, it will reenter state i infinitely often.

On the other hand, suppose state i is transient. Hence each time the process enters state i , there is a positive probability, namely $1 - f_i$, that it never returns to state i .

Thus, starting in state i , the probability that the chain will be in state i for exactly n time periods is $f_i^{n-1}(1 - f_i)$. Thus

504

if state i is transient then
starting in state i , the number of time
periods the process will be in state i
is geometric with finite mean $1/(1-f_i)$.

From the preceding two paragraphs, it follows that state i is recurrent if and only if, starting in state i , the expected number of time periods the process is in state i is infinite. But letting

$$I_n = \begin{cases} 1 & \text{if } X_n = i \\ 0 & \text{if } X_n \neq i \end{cases}$$

Then $\sum_{n \geq 0} I_n$ is the number of time periods the chain is in state i . Also,

$$E \left[\sum_{n \geq 0} I_n \mid X_0 = i \right] = \sum_{n \geq 0} E [I_n \mid X_0 = i]$$

$$= \sum_{n \geq 0} P(X_n = i \mid X_0 = i)$$

$$= \sum_{n \geq 0} P_{ii}^n.$$

So we have proved

Theorem State i is

recurrent if $\sum_{n \geq 1} P_{ii}^n = \infty$

transient if $\sum_{n \geq 1} P_{ii}^n < \infty$.

Cor- (recurrence is a class property)

If state i is recurrent, and state i communicates with state j , then state j is recurrent.

Proof Since state i communicates with state j , there exist integers k and m such that $P_{ij}^k > 0$ and $P_{ji}^m > 0$. Now for any integer n ,

$$P_{jj}^{m+n+k} \geq P_{ji}^m P_{ii}^n P_{ij}^k.$$

506

Indeed, P_{jj}^{m+n+k} is the chance of going from j to j in $m+n+k$ steps, while the right side is the chance of going from j to j in $m+n+k$ steps, via a path that goes from j to i in m steps, then from i to i in n steps, then from i to j in k steps.

Summing over n , we obtain

$$\sum_{n \geq 1} P_{jj}^{m+n+k} \geq P_{ji}^m P_{ij}^k \sum_{n \geq 1} P_{ii}^n = \infty,$$

since $P_{ji}^m P_{ij}^k > 0$, and $\sum_{n \geq 1} P_{ii}^n = \infty$ since state i is recurrent. Thus state j is recurrent by the previous theorem.

Example

Consider random walk on the integers

such that $P_{i,i+1} = p$, $P_{i,i-1} = 1-p$,
and $0 < p < 1$.

Since all states clearly communicate,
they are all either transient or recurrent.
So let us consider state 0 and attempt to
determine if $\sum_{n \geq 1} P_{00}^n$ is finite or not.

Since it's impossible to be ^{at 0} ~~even~~ after an
odd number of steps, we have $P_{00}^{2n-1} = 0$ $n=1, 2, \dots$

On the other hand, to be at 0 after $2n$
steps, we must have exactly n steps to the
right and n steps to the left. So

$$P_{00}^{2n} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!} (p(1-p))^n$$

"Stirling's formula" says that

$$n! \sim n^{n+1/2} e^{-n} \sqrt{2\pi}$$

where we say that $a_n \sim b_n$ when

508

$\lim_{n \rightarrow \infty} a_n/b_n = 1$. We obtain

$$p_{00}^{2n} \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

It's easy to verify that if $a_n \sim b_n$, then $\sum a_n < \infty$ if and only if $\sum b_n < \infty$. Hence

$\sum_{n \geq 1} p_{00}^n$ will converge if and only if

$$\sum_{n \geq 1} \frac{(4p(1-p))^n}{\sqrt{\pi n}} \text{ converges.}$$

But $4p(1-p) \leq 1$ with equality if and only if $p = 1/2$. So the chain is recurrent when $p = 1/2$ and transient when $p \neq 1/2$.

When $p = 1/2$, this is called symmetric random walk. Let's look at symmetric random walk on the integer lattice. So

$$\frac{1}{4} = P((i,j), (i+1,j)) = P((i,j), (i-1,j)) = P((i,j), (i,j+1)) = P((i,j), (i,j-1))$$

Since this chain is irreducible, all 509 states are recurrent if $(0,0)$ is recurrent.

After $2n$ steps, the chain will be back at its original location if for some i , $0 \leq i \leq n$, the $2n$ steps consist of i steps to the left, i to the right, $n-i$ up, and $n-i$ down.

So

$$\begin{aligned} P_{(0,0),(0,0)}^{2n} &= \sum_{i=0}^n \frac{(2n)!}{i! i! (n-i)! (n-i)!} \left(\frac{1}{4}\right)^{2n} \\ &= \sum_{i=0}^n \frac{(2n)!}{n! n!} \frac{n!}{(n-i)! i!} \frac{n!}{(n-i)! i!} \left(\frac{1}{4}\right)^{2n} \\ &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \sum_{i=0}^n \binom{n}{i} \binom{n}{i} \\ &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \binom{2n}{n} \end{aligned}$$

using identity 2
 $\binom{2n}{n} = \sum_{i=0}^n \binom{n}{i}^2$

By Stirling's formula, $\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$

Thus $P_{(0,0),(0,0)}^{2n} \sim \frac{1}{\pi n}$, so $\sum_{n \geq 0} P_{(0,0),(0,0)}^n = \infty$

and all states are recurrent. This fails in dimensions 3 and higher.

MATH 407 Lecture 35New topic = combinatorics of permutationsDescents

Def. Let $p = p_1 p_2 \dots p_n$ be a permutation, and let $i < n$ be a positive integer. We say that i is a descent of p if $p_i > p_{i+1}$.

Similarly, we say that i is an ascent of p if $p_i < p_{i+1}$.

Example

Let $p = 3412576$. Then 2 and 6 are descents of p , while 1, 3, 4, 5 are ascents of p .

Notation $D(p)$ = set of all descents of p
 $d(p)$ = number of descents of p

Lemma Let $S = \{s_1, \dots, s_k\} \subseteq \{1, 2, \dots, n-1\}$, and let $\alpha(S)$ be the number of permutations on n symbols with descent set contained

in S . Then

$$\alpha(S) = \binom{n}{s_1} \binom{n-s_1}{s_2-s_1} \binom{n-s_2}{s_3-s_2} \cdots \binom{n-s_k}{n-s_k}$$

Proof

The crucial idea of proof is the following. We arrange our n elements into $k+1$ segments so that the first i segments together have s_i entries for each i . Then within each segment, we put our entries in increasing order. Then the only places where the resulting permutation has a chance to have a descent is where 2 segments meet, that is at $s_1, s_2, \dots, s_k$. Thus the descent set of the resulting permutation is contained in S .

In how many ways can one arrange our entries in these segments? The first segment has length s_1 , so can be chosen in $\binom{n}{s_1}$ ways. The next segment has length $s_2 - s_1$ and is

disjoint from the first one, so can be chosen in $\binom{n-s_1}{s_2-s_1}$ ways. Continue in this way to complete the proof.

Theorem Let $S \subseteq \{1, \dots, n-1\}$. Then the number of permutations on n symbols with descent set

$$S \text{ is } \beta(S) = \sum_{T \subseteq S} (-1)^{|S-T|} \alpha(T)$$

Proof This is immediate from the principle of inclusion and exclusion.

Eulerian numbers

Let $A(n, k)$ be the number of permutations on n symbols with k -descents. (This differs slightly from our earlier notation). So

$$A(3, 1) = 1 \quad 123$$

$$A(3, 2) = 4 \quad 132, 213, 231, 312$$

$$A(3, 3) = 1 \quad 321$$

Note that if p has $k-1$ descents, then p is the union of k increasing subsequences of consecutive entries. These are called the ascending runs of p .

Example

The three ascending runs of $p = 2415367$ are 24, 15, 367.

So the permutations with k ascending runs are the same as permutations with $k-1$ descents, providing one reason for the notation $A(n, k)$.

Call the $A(n, k)$ the Eulerian numbers.

Let's study these, starting with a simple recursive relation.

Theorem For all positive integers k and n satisfying $k \leq n$, we have

$$A(n, k+1) = (k+1)A(n-1, k+1) + (n-k)A(n-1, k)$$

Proof

There are two ways we can get an n -permutation p with k descents from an $(n-1)$ -permutation p' by inserting the entry n into p' . Either p' has k descents and the insertion of n does not form a new descent or p' has $k-1$ descents, and the insertion of n does form a new descent.

In the first case, we have to put the entry n at the end of p' , or we have to insert n between two entries that form one of the k descents of p' . This means we have $k+1$ choices for the position of n . As we have $A(n-1, k+1)$ choices for p' , the first term on the right hand side is explained.

In the second case, we have to put the entry n at the front of p' or between two entries that form one of the $n-k-1$ ascents of p' . This means that we have $n-k$ choices

for the position of n . As we have $A(n-1, k)$ choices for p' , the second part of the right hand side is explained, completing the proof.



Note that $A(n, k+1) = A(n, n-k)$, since if $p = p_1 \cdots p_n$ has k descents, then its reversal $p' = p_n \cdots p_1$ has $n-k-1$ descents

Theorem

Set $A(0,0) = 1$ and $A(n,0) = 0$ for $n > 0$. Then for all non-negative integers n , and for all real numbers x , we have

$$x^n = \sum_{k=1}^n A(n, k) \binom{x+n-k}{n}$$

Example Let $n = 3$. Then $A(3,1) = 1$, $A(3,2) = 4$, $A(3,3) = 1$, so

$$x^3 = \binom{x+2}{3} + 4 \binom{x+1}{3} + \binom{x}{3}$$

Proof (omit)

Cor- For all positive integers n ,

$$x^n = \sum_{k=0}^n A(n, k) \binom{x+k-1}{n}$$

Proof Replace x by $-x$ in the previous theorem. We get

$$x^n (-1)^n = \sum_{k=0}^n A(n, k) \binom{-x+n-k}{n}$$

Now note that

$$\begin{aligned} \binom{-x+n-k}{n} &= \frac{(-x+n-k)(-x+n-k-1) \cdots (-x+1-k)}{n!} \\ &= (-1)^n \binom{x+k-1}{n} \end{aligned}$$

Comparing these two identities yields the desired result.

~

Question : Is there an explicit formula for the numbers $A(n, k)$? Yes, we have the following

Theorem For all non-negative integers

n and k satisfying $k \leq n$, we have 517

$$A(n, k) = \sum_{i=0}^k (-1)^i \binom{n+1}{i} (k-i)^n.$$

Proof (hard to omit)

Stirling numbers and Eulerian numbers

Let $S(n, k)$ = number of set partitions of $\{1, \dots, n\}$ into k blocks.

We studied $S(n, k)$ earlier in our discussion of generating functions. Call $S(n, k)$ Stirling numbers of the second kind.

Example: $S(4, 3) = 6$ corresponding to
 $\{1, 2\}\{3\}\{4\}$, $\{1, 3\}\{2\}\{4\}$, $\{1, 4\}\{2\}\{3\}$,
 $\{2, 3\}\{1\}\{4\}$, $\{2, 4\}\{1\}\{3\}$, $\{3, 4\}\{1\}\{2\}$

Lemma For all positive integers n and r ,

$$S(n, r) = \frac{1}{r!} \sum_{i=0}^r (-1)^i \binom{r}{i} (r-i)^n$$

Proof An ordered partition of $\{1, 2, \dots, n\}$

518

into r blocks is a set partition of $\{1, \dots, n\}$ into r blocks in which the set of blocks is totally ordered.

So on one hand, the number of ordered partitions of $\{1, \dots, n\}$ into r blocks is

$$r! S(n, r).$$

On other hand, an ordered partition of $\{1, \dots, n\}$ into r blocks is the same as the number of onto maps from $\{1, \dots, n\}$ to $\{1, \dots, r\}$.

To enumerate all such onto maps, let

$A_i =$ functions from $\{1, \dots, n\}$ to $\{1, \dots, r\}$

whose image does not contain i . Then

$f: \{1, \dots, n\} \rightarrow \{1, \dots, r\}$ is onto if and only if

it is not contained in $A_1 \cup \dots \cup A_r$, and the

theorem now follows by a standard application of the principle of inclusion and exclusion

Theorem For all positive integers n and r , we have

$$S(n, r) = \frac{1}{r!} \sum_{k=0}^r A(n, k) \binom{n-k}{r-k}$$

Cor. For all positive integers n and k ,

$$A(n, k) = \sum_{r=1}^k S(n, r) r! \binom{n-r}{k-r} (-1)^{k-r}$$

(we omit the proofs)

Generating functions and Eulerian numbers

define the "Eulerian polynomial"

$$A_n(x) = \sum_{k=1}^n A(n, k) x^k.$$

Theorem For all positive integers n , the n th Eulerian polynomial has the alternative description

$$A_n(x) = (1-x)^{n+1} \sum_{i \geq 0} i^n x^i$$

Note: this is actually how Euler defined the Eulerian polynomial

Example

For $n=1$, we have

$$\begin{aligned} A_1(x) &= (1-x)^2 \sum_{i \geq 0} i x^i \\ &= (1-x)^2 \frac{x}{(1-x)^2} = x \end{aligned}$$

For $n=2$, we have

$$\begin{aligned} A_2(x) &= (1-x)^3 \sum_{i \geq 0} i^2 x^i \\ &= (1-x)^3 \left(\frac{2x^2}{(1-x)^3} + \frac{x}{(1-x)^2} \right) = x + x^2 \end{aligned}$$

Proof By an earlier result (the explicit formula for Eulerian numbers),

$$\sum_{k=1}^n A(n, k) x^k = \sum_{k=1}^n \sum_{0 \leq i \leq k} (-1)^i \binom{n+1}{i} (k-i)^n x^k$$

$$= \sum_{k=1}^n \left(\sum_{0 \leq i \leq k} (-1)^{k-i} \binom{n+1}{k-i} i^n x^k \right) \quad 521$$

Changing the order of summation, and noting that the sum in parentheses, being equal to a multiple of $A(n, k)$ vanishes for $k > n$, we get

$$\begin{aligned} \sum_{i \geq 0} i^n x^i \sum_{k \geq i} \binom{n+1}{k-i} (-x)^{k-i} \\ = (1-x)^{n+1} \sum_{i \geq 0} i^n x^i. \end{aligned}$$

It is often useful to collect the Eulerian numbers $A(n, k)$ for all n and all k into a single generating function. The next result does this.

Theorem

$$\text{Let } r(t, u) = \sum_{n \geq 0} \sum_{k \geq 0} A(n, k) t^k u^n / n!$$

$$\text{Then } r(t, u) = \frac{1-t}{1-te^{u(1-t)}}$$

Proof

By the previous result, we see that

$$\begin{aligned}
 r(t, u) &= \sum_{n \geq 0} \left((1-t)^{n+1} \sum_{i \geq 0} t^i u^i \right) \frac{u^n}{n!} \\
 &= (1-t) \sum_{i \geq 0} t^i \sum_{n \geq 0} \frac{(i u (1-t))^n}{n!} \\
 &= (1-t) \sum_{i \geq 0} t^i e^{i u (1-t)} \\
 &= (1-t) \sum_{i \geq 0} (t e^{u(1-t)})^i \\
 &= (1-t) / (1 - t e^{u(1-t)})
 \end{aligned}$$

Inversions

Let $p = p_1 p_2 \dots p_n$ be a permutation. We say that (p_i, p_j) is an inversion of p if $i < j$ but $p_i > p_j$.

Example

$p = 31524$ has four inversions: $(3, 1), (3, 2), (5, 2), (5, 4)$

Remark : inversions are useful in statistics

Let $i(p)$ = number of inversions of p .

Theorem

For all positive integers $n \geq 2$,

$$\sum_{p \in S_n} x^{i(p)} = (1+x)(1+x+x^2) \cdots (1+x+x^2+\cdots+x^{n-1})$$

This theorem can be proved by induction. It implies that $i(p)$ has mean $\binom{n}{2}/2$ and variance $n(n-1)(2n+5)/24$, and that normalized by its mean and variance, $i(p)$ is asymptotically normal as $n \rightarrow \infty$.

Why are metrics such as $i(p)$ statistically useful? Define $f(p) = \text{"spearman's footrule"}$

$$= \sum_{i=1}^n |p_i - i|$$

524

Then if p is a uniformly chosen random permutation,

$$E(f) = \frac{1}{3} (n^2 - 1)$$

$$\text{Var}(f) = \frac{1}{45} (n+1)(2n^2+7)$$

and normalized by its mean and standard deviation, f is asymptotically normal.

Look at draft lottery data

<u>Month</u>	J	F	M	A	M	J	J	A	S	O	N	D
rank of month in <u>lottery</u>	5	4	1	3	2	6	8	9	10	7	11	12

So people born in early months seem to be drafted more often. How can you test this?

$$f(\text{above permutation}) = 18$$

If permutation were random, it would have mean = 47.7, SD = 9.23. Now 18 is more than

3 standard deviations away from the mean, so by Chebyshev's inequality unlikely to arise by chance.

MATH 407 lecture 36

Another metric on S_n : Ulam's distance U

define $U(p) = n - \text{length of longest increasing subsequence of } p.$

For example if $p = 321459678$
then 145678 is an increasing subsequence of maximum length (6).

To motivate definition of U , consider n books on a shelf in order p . We want to sort the books by deletion-insertion operations: taking a book out and inserting it in another place. Thus 3 moves are required to sort p above.

Lemma The smallest number of moves

required to sort p is

n - length of longest increasing subsequence of p .

Proof

If $p_{i_1} < p_{i_2} < \dots < p_{i_k}$ is a longest increasing subsequence, then clearly inserting and deleting other letters doesn't change the ordering of this subsequence. It follows that $n-k$ moves suffice. Since each move can increase the longest increasing subsequence by at most 1, $n-k$ moves are required.

Remark Ulam's metric is related to solitaire and to airline boarding. Ask me for details some time.

Back to random walks

Let $X_1, X_2, X_3, \dots$ be independent
 with each $X_i = \begin{cases} 1 & \text{probability } 1/2 \\ -1 & \text{probability } 1/2 \end{cases}$

Let $S_n = X_1 + \dots + X_n$

Can think of this as random walk on the integers, started at 0. We say that there is a return to the origin at time n if $S_n = 0$.

Theorem

The probability of a return to the origin at time $2m$ is given by

$$u_{2m} = \binom{2m}{m} / 2^{2m}$$

the probability of a return to the origin at odd time is 0.

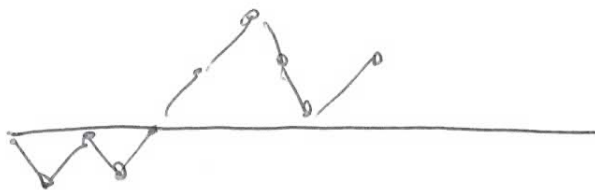
Proof the second assertion is clear.

For the first assertion, must choose m of the $X_1, \dots, X_{2m}$ to be equal to $+1$ (so the rest are equal to -1).

528

this can be done in $\binom{2m}{m}$ ways,
and each way occurs with probability
 $\frac{1}{2^{2m}}$. ~~note~~ note: we define $u_0 = 1$.

We say that the random walk has a first return to the origin at time $2m$ if $m > 0$ and $S_{2k} \neq 0$ for all $k < m$. Let f_{2m} be the probability of this event. (We also define $f_0 = 0$). One can think of the expression $f_{2m} 2^{2m}$ as the number of paths of length $2m$ between the points $(0, 0)$ and $(2m, 0)$ that do not touch the horizontal axis except at the endpoints



Theorem For $n \geq 1$, the probabilities $\{a_{2k}\}$

and $\{f_{2k}\}$ are related by the equation

$$u_{2n} = f_0 u_{2n} + f_2 u_{2n-2} + \dots + f_{2n} u_0.$$

Proof

There are $u_{2n} 2^{2n}$ paths of length $2n$ which have endpoints $(0,0)$ and $(2n,0)$.

The collection of such paths can be partitioned into n sets, depending on the time of the first return to the origin. A path in this collection which has a first return to the origin at time $2k$ consists of an initial segment from $(0,0)$ to $(2k,0)$ in which no interior points are on the horizontal axis, and a terminal segment from $(2k,0)$ to $(2n,0)$ with no further restrictions on this segment. Thus, the number of paths in this collection which have a first return to the

origin at time $2k$ is given by 530

$$f_{2k} 2^{2k} u_{2n-2k} 2^{2n-2k} = f_{2k} u_{2n-2k} 2^{2n}$$

If we sum over k , we obtain the equation

$$u_{2n} 2^{2n} = f_0 u_{2n} 2^{2n} + f_2 u_{2n-2} 2^{2n} \\ + \dots + f_{2n} u_0 2^{2n}$$

Dividing both sides of this equation by 2^{2n} completes the proof.

Theorem For $m \geq 1$, the probability of a first return to the origin at time $2m$ is given by

$$f_{2m} = \frac{u_{2m}}{2^{2m-1}} = \frac{\binom{2m}{m}}{(2m-1) 2^{2m}}$$

Proof

we begin by defining the generating functions

$$U(x) = \sum_{m \geq 0} u_{2m} x^m$$

$$F(x) = \sum_{m \geq 0} f_{2m} x^m$$

The previous result gives

$$U(x) = 1 + U(x)F(x)$$

(The presence of the 1 on the right-hand side is due to the fact that u_0 is defined to be 1, but previous result only holds for $n \geq 1$).

Solving for $F(x)$, one obtains

$$F(x) = \frac{U(x) - 1}{U(x)}$$

So let's find a closed-form expression for $U(x)$. We know that

$$U(x) = \sum_{m \geq 0} \binom{2m}{m} 2^{-2m} x^m$$

It's known that $\frac{1}{\sqrt{1-4x}} = \sum_{m \geq 0} \binom{2m}{m} x^m$.

Thus, $U(x) = \frac{1}{\sqrt{1-x}}$

Thus $F(x) = \frac{u(x)-1}{u(x)}$

532

$$= \frac{(1-x)^{-1/2} - 1}{(1-x)^{-1/2}}$$

$$= 1 - (1-x)^{1/2}$$

Although it's possible to compute f_{2m} using Taylor's theorem, it's easier to note that

$F'(x) = u(x)/2$, so that the coefficients f_{2m} can be found by integrating the series for

$u(x)$. We obtain, for $m \geq 1$

$$f_{2m} = u_{2m-2} / 2m$$

$$= \frac{\binom{2m-2}{m-1}}{m 2^{2m-1}}$$

$$= \frac{\binom{2m}{m}}{(2m-1) 2^{2m}}$$

$$= \frac{u_{2m}}{(2m-1)}$$

Probability of eventual return

we discussed this earlier. Let's revisit.

Example Eventual return in 1-dimension

Intuitively,

$$P(\text{eventually return to } 0) = \lim_{n \rightarrow \infty} P(\text{return to } 0 \text{ by time } n)$$

$$= \lim_{n \rightarrow \infty} \sum_{i=1}^{n/2} f_{2i}$$

$$= \sum_{i \geq 1} f_{2i}$$

Now with $F(x) = 1 - (1-x)^{1/2}$ as above,

$\sum_{i \geq 1} f_{2i} = F(1) = 1$. So with probability 1, the particle returns to the origin.

Example

(The expected number of returns to the origin ~~is~~ up to time $2m$.)

534

Let g_{2m} be the number of returns to the origin among all random walks of length $2m$. Then the expected number of returns to the origin up to time $2m$ is $g_{2m} / 2^{2m}$.

$$\text{Define } G(x) = \sum_{k \geq 0} g_{2k} x^k$$

Letting $F(x), U(x)$ be as above, one can prove that $G(x) = F(4x)G(x) + \frac{1}{1-4x} - U(4x)$

Solving this recursion and simplifying, we get

$$G(x) = \frac{1}{(1-4x)^{3/2}} - \frac{1}{(1-4x)}$$

$$\text{Thus } g_{2m} = \frac{1}{2} \binom{2m+2}{m+1} - 2^{2m}$$

So the expected number of returns to 0

by time $2m$ is

535

$$\frac{g_{2m}}{2^{2m}} \sim \sqrt{\frac{2}{\pi}} \sqrt{2m} \quad (\text{by Stirling's formula})$$

Remark: Letting M_{2m} be the maximum value of the random walk up to time $2m$, one can show that $M_{2m} \sim \sqrt{\frac{2}{\pi}} \sqrt{2m}$.

Probability of long leads: the first arcsine law

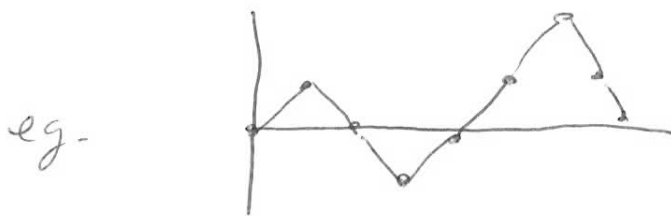
Theorem

Let $p_{2k, 2n}$ be the probability that in the time interval from 0 to $2n$, the particle spends $2k$ time units on the positive side and $2n-2k$ time units on the negative side.

Then $p_{2k, 2n} = u_{2k} u_{2n-2k}$.

Note We say that the particle spends the time

from $k-1$ to k on the positive side 536
if the k th side of its path lies above
the x -axis.



spends 6 time
units on positive
side and 2 time
units on negative
side

Proof (omit)

We feel intuitively that the fraction k/n of
the total time spent on the positive side is
most likely to be close to $1/2$. However the
opposite is true: the possible values close
to $1/2$ are least probable and the values $k/n=0$
and $k/n=1$ have the greatest probability. This
is easily checked using the explicit formula
for $p_{2k, 2n}$ in the theorem.

Table below illustrates the paradox. In betting terminology, it reveals the startling fact that in $2n=20$ tossings of a fair coin, with probability .3524, the less fortunate player is never in the lead. In most cases (with probability .5379) the accumulated gain of the less fortunate player will be positive just once or never. By contrast, an equal division 10:10 of the leads has a probability of only .0606

TABLE (distribution of leads in 20 tosses of a fair coin)

	$k=0$	$k=2$	$k=4$	$k=6$	$k=8$	$k=10$
	$k=20$	$k=18$	$k=16$	$k=14$	$k=12$	
$P_{k,20}$	.1762	.0927	.0736	.0655	.0617	.0606

The formula for $p_{2k, 2n}$ in the previous result, though exact, is not very revealing, so we replace it by a simpler approximation. Stirling's formula shows that $u_{2n}(\sqrt{\pi n}) \rightarrow 1$ as $n \rightarrow \infty$. It follows that

$$p_{2k, 2n} \sim \frac{1}{\pi k^{1/2} (n-k)^{1/2}}$$

where the ratio of the two sides tends rapidly to 1 as $k \rightarrow \infty$ and $n-k \rightarrow \infty$. The probability that the fraction k/n of time spent on the positive side lies between $\frac{1}{2}$ and α ($\frac{1}{2} < \alpha < 1$) is given by

$$\sum_{\frac{1}{2}n < k < \alpha n} p_{2k, 2n} \sim \frac{1}{\pi n} \sum_{\frac{1}{2}n < k < \alpha n} \left| \frac{k}{n} \left(1 - \frac{k}{n} \right) \right|^{-1/2}.$$

On the right side we recognize the Riemann

sum approximating the integral

$$\pi^{-1} \int_{1/2}^{\alpha} \frac{dx}{[x(1-x)]^{1/2}} = 2\pi^{-1} \arcsin(\alpha^{1/2}) - 1/2.$$

For reasons of symmetry, the probability that $k/n \leq 1/2$ tends to $1/2$ as $n \rightarrow \infty$. Adding this, we get

Theorem (first arcsine law)

For fixed α ($0 < \alpha < 1$) and $n \rightarrow \infty$, the chance that the fraction k/n of time spent on the positive side is $< \alpha$ tends to

$$\pi^{-1} \int_0^{\alpha} \frac{dx}{[x(1-x)]^{1/2}} = \frac{2}{\pi} \arcsin(\sqrt{\alpha})$$

Remark This is an excellent approximation even for values of n as small as $n = 20$. With probability .20 the particle stays for about 97.6 percent of the time on the same side of the origin.

Math 407 Lecture 37

540

Review : Approximations of the binomial distribution

Normal approximation of the binomial distribution

Suppose that S_n has the binomial (n, p) distribution with n large and p not too close to 0 or 1. Then

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \text{ is close to } \Phi(b) - \Phi(a)$$

where $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.

As a rule of thumb, the approximation is good if $np(1-p) > 10$.

Example A fair coin is flipped 10,000 times.

Estimate the probability that the number of

heads is between 4850 and 5100
inclusive.

541

Sol. Let S = number of heads. Then S is

binomial $(10,000, 1/2)$, with $E(S) = \frac{10,000}{2} = 5000$

and $\text{Var}(S) = 10,000 \cdot \frac{1}{2} \cdot \frac{1}{2} = 2500$. The probability

in question is $P(4850 \leq S \leq 5100)$.

Subtracting the mean and dividing by the standard deviation, we get

$$P(4850 \leq S \leq 5100)$$

$$= P\left(\frac{4850 - 5000}{\sqrt{2500}} \leq \frac{S - 5000}{\sqrt{2500}} \leq \frac{5100 - 5000}{\sqrt{2500}}\right)$$

$$= P\left(-3 \leq \frac{S - 5000}{\sqrt{2500}} \leq 2\right)$$

$$\approx \Phi(2) - \Phi(-3) = .9759$$

The true answer is .9765

542

Example

Suppose a game piece moves along a board according to the following rule. If a die roll gives 1 or 2, then take two steps ahead. Otherwise, take 3 steps. Approximate the chance that after 120 rolls, the piece has moved more than 315 steps.

Sol.

Let X_n = number of steps taken up to and including the n th die roll. X_n is not a binomial random variable, but can be easily related to one.

Let S_n = number of 1's and 2's among the first n rolls.

$$\text{Then } X_n = 2(S_n) + 3(n - S_n) = 3n - S_n.$$

So we need to estimate

$$P(X_{120} > 315) = P(360 - S_{120} > 315)$$

543

$$= P(S_{120} < 45)$$

Now apply normal approximation to

$S_{120} \sim \text{Bin}(120, 1/3)$. The mean and variance are

$$E(S_{120}) = 40, \quad \text{Var}(S_{120}) = \cancel{40} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{80}{3}$$

Thus

$$P(S_{120} < 45) = P\left(\frac{S_{120} - 40}{\sqrt{80/3}} < \frac{5}{\sqrt{80/3}}\right)$$

$$\approx \Phi\left(\frac{5}{\sqrt{80/3}}\right) \approx .834$$

Continuity correction

If S_n is binomial (n, p) then it can only take integer values. Thus if k_1, k_2 are integers,

$$P(k_1 \leq S_n \leq k_2) = P(k_1 - 1/2 \leq S_n \leq k_2 + 1/2)$$

Since the interval $[k_1, k_2]$ contains the same integers ~~as~~ as $[k_1 - 1/2, k_2 + 1/2]$, It turns out that if we apply normal approximation

to the modified interval $[k_1 - 1/2, k_2 + 1/2]$,⁵⁴⁴
we get a slightly better approximation of
the exact binomial probability. In other
words,

$$\Phi\left(\frac{k_2 + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

gives a better estimate for $P(k_1 \leq S_n \leq k_2)$ than
the formula without the $1/2$ terms.

Switching from $[k_1, k_2]$ to $[k_1 - 1/2, k_2 + 1/2]$
is called the continuity correction. It is
especially important if k_1, k_2 are close to each
other or if $np(1-p)$ is not very large.

Example

Roll a fair die 720 times. Estimate the
probability of getting exactly 113 sixes

Sol.

545

Denote the number of sixes by S .

This is a binomial $(720, 1/6)$ random variable with mean 120 and variance $720 \cdot \frac{1}{6} \cdot \frac{5}{6} = 100$.

To estimate $P(S=113)$, we write it as

$$P(S=113) = P(112.5 \leq S \leq 113.5)$$

Normal approximation now gives

$$P(S=113) \approx \Phi\left(\frac{113.5-120}{10}\right) - \Phi\left(\frac{112.5-120}{10}\right)$$

$$= \Phi(-.65) - \Phi(-.75)$$

$$\approx (1-.7422) - (1-.7734)$$

$$=.0312.$$

The exact probability is

$$P(S=113) = \binom{720}{113} \frac{5^{607}}{6^{720}} \approx .0318$$

Note that without using continuity correction, our approximation would be

546

$$P(S=113) = P\left(\frac{113-120}{10} \leq \frac{S-120}{10} \leq \frac{113-120}{10}\right)$$

$$= P(-.7 \leq \frac{S-120}{10} \leq .7)$$

$$= \Phi(-.7) - \Phi(-.7) = 0.$$

In this example $np(1-p) = 100$, so we are well within the range of earlier rule of thumb.

Partial proof of central limit theorem for binomial

For ease of notation, let $q = 1-p$.

Then

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = P(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq})$$

$$= \sum_{np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}} \frac{n!}{(n-k)!k!} p^k q^{n-k}$$

where the sum is over integers k between $np + a\sqrt{npq}$ and $np + b\sqrt{npq}$.

If c_n, d_n are sequences, we write 547

$a_n \sim b_n$ to mean that $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$

Recall Stirling's formula:

$$\text{as } n \rightarrow \infty, n! \sim n^n e^{-n} \sqrt{2\pi n}$$

Thus

$$\frac{n!}{(n-k)! k!} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

From this follows the approximate equality

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right)$$

$$\approx \sum_{np+a\sqrt{npq} \leq k \leq np+b\sqrt{npq}} \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right)$$

The sum on the right can be cast in the familiar form of a Riemann sum over the interval $[a, b]$ by defining partition points

$$x_k = \frac{k - np}{\sqrt{npq}}$$

the summation limits

548

$$np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}$$

specify exactly $a \leq x_k \leq b$, and the length of the subintervals is $\Delta x = \frac{1}{\sqrt{npq}}$. The

approximation above becomes

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right)$$

$$\approx \sum_{k: a \leq x_k \leq b} \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} \Delta x$$

$$\xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

The last step is just the Riemann sum approximation of the integral. To make the proof completely rigorous requires a more quantitative version of Stirling's formula.

Applications of normal approximation

Confidence intervals

Suppose we have a biased coin and we do not know the true probability p that it lands on heads. How can we estimate p ?

The law of large numbers suggests a natural approach: flip the coin n times, count the number S_n of heads, and take the observed frequency $\hat{p} = S_n/n$ as an estimate for p .

In practice we can't flip a coin forever. Can we estimate the quality of approximation for finite n ?

$$\begin{aligned}
 P(|\hat{p} - p| < \varepsilon) &= P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) \\
 &= P(-n\varepsilon < S_n - np < n\varepsilon) \\
 &= P\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} < \frac{S_n - np}{\sqrt{np(1-p)}} < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)
 \end{aligned}$$

$$\text{normal approximation } p \approx \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \quad 550$$

$$= 2 \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1.$$

We don't know the value of p , so how can we evaluate the last expression? We can get a lower bound that works for all p . The maximum of $p(1-p)$ is $1/4$ which is achieved at $p = 1/2$. Consequently, $\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq 2\varepsilon\sqrt{n}$.

Since the function Φ is increasing,

$$2 \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 2 \Phi(2\varepsilon\sqrt{n}) - 1.$$

Summarizing,

$$P(|\hat{p} - p| < \varepsilon) \geq 2 \Phi(2\varepsilon\sqrt{n}) - 1.$$

Example

How many times should we flip a coin with unknown success probability p so that the estimate $\hat{p} = S_n/n$ is within .05 of the true p with probability at least .99?

Sol.

Using $\square$, to ensure a lower bound of at least .99, we need

$$P(|\hat{p} - p| < \varepsilon) \geq 2\Phi(2\varepsilon\sqrt{n}) - 1 \geq .99$$

The last inequality is satisfied if

$$\Phi(2\varepsilon\sqrt{n}) \geq .995$$

From the table for Φ , this is equivalent to $2\varepsilon\sqrt{n} \geq 2.58$, which is the same as

$$n \geq \frac{(2.58)^2}{4\varepsilon^2} = \frac{(2.58)^2}{4(.05)^2} \approx 665.64$$

where we used $\varepsilon = .05$. The conclusion

is that if we flip the coin 666 552
times, the estimate $\hat{p}$ is within 0.05 of
the real p , with probability at least .99



Another task is to find the confidence interval
around $\hat{p}$ that captures the true p with a
given (high) probability. The 100% or 90
confidence interval for the unknown success
probability p is given by $(\hat{p} - \epsilon, \hat{p} + \epsilon)$ where
 ϵ is chosen to satisfy $P(|\hat{p} - p| < \epsilon) \geq r$. In
other words, the random interval $(\hat{p} - \epsilon, \hat{p} + \epsilon)$
contains the true p with probability at least r .

Example We repeat a trial 1000 times and
observe 450 successes. Find the 95%
confidence interval for the true success

probability p .

553

Sol.

This time n is given and we look for ε such that $P(|\hat{p} - p| < \varepsilon) \geq .95$. So we need to solve (by $\Rightarrow$), the inequality $2\Phi(2\varepsilon\sqrt{n}) - 1 \geq .95$ for ε . First simplify and then turn to the Φ table:

$$\Phi(2\varepsilon\sqrt{n}) \geq .975 \Leftrightarrow 2\varepsilon\sqrt{n} \geq 1.96$$

$$\Leftrightarrow \varepsilon \geq \frac{1.96}{2\sqrt{1000}} \approx .031$$

Thus if $n=1000$, then with probability at least .95, the random quantity $\hat{p}$ satisfies $|\hat{p} - p| < .031$.

Polling Suppose that the fraction of the population who like broccoli is p . We want

to estimate p . It's too slow and 554 expensive to ask everyone whether they like broccoli, so instead we take a random sample of n individuals, ask each of them whether they like broccoli, and estimate p with the ratio $\hat{p}$ who said yes.

When we take a poll we are sampling without replacement. So the number of people who say yes is not binomially distributed, and instead follows the hypergeometric distribution discussed earlier. So can't use our above approach to estimating p .

BUT, if the sample size n is small compared to the population size, sampling with and without replacement is very close, so we can pretend the sample was with replacement and use our techniques for the binomial.

Math 407 Lecture 38

555

Recall that X is said to be Poisson(λ)

if for all $k \in \{0, 1, 2, \dots\}$,

$$P(X=k) = e^{-\lambda} \lambda^k / k!$$

Theorem Let $\lambda > 0$ and consider positive integers n for which $\lambda/n < 1$. Let S_n be Binomial($n, \lambda/n$). Then

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \lambda^k / k!$$

Proof We rearrange the binomial probability and then simply observe that the different pieces have limits as $n \rightarrow \infty$.

$$P(S_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1) \cdots (n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k}$$

$$= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left[1 - \left(1 - \frac{\lambda}{n}\right)\left(1 - \frac{2\lambda}{n}\right) \cdots \left(1 - \frac{(k-1)\lambda}{n}\right)\right] \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k}$$

$$\xrightarrow[\substack{\text{as} \\ n \rightarrow \infty}]{\frac{\lambda^k}{k!}} e^{-\lambda} \cdot 1 \cdot 1 = e^{-\lambda} \frac{\lambda^k}{k!}$$



The previous result says that the binomial $(n, \lambda/n)$ distribution gets closer and closer to $\text{Poisson}(\lambda)$ as $n \rightarrow \infty$. But what if we want a statement for fixed n ?

Theorem

Let X be binomial (n, p) and Y be $\text{Poisson}(np)$. Then for any subset A of $\{0, 1, 2, \dots\}$, we have $|P(X \in A) - P(Y \in A)| \leq np^2$.

The theorem suggests that Poisson approximation is good when you count rare occurrences in a large number of trials that can be assumed approximately independent.

Example

Suppose a factory experiences on average 3 accidents each month. What is the probability that during a particular month there are exactly two accidents?

Sol. We need to make an assumption about the distribution of X , the number of accidents during a month. If the factory has many workers, each having a small chance of an accident independent of other works, can model X as Poisson. Since $E(X)=3$, we model X by a $\text{Poisson}(3)$ random variable. So

$$P(X=2) = e^{-3} 3^2 / 2! \approx .224.$$

Example

Let X be the number of fixed points of a random permutation on n symbols. We saw earlier in the course that X is approximately $\text{Poisson}(1)$.

Why do we expect this? well,

$$X = X_1 + \dots + X_n \text{ where } X_i = \begin{cases} 1 & \text{if } \pi(i) = i \\ 0 & \text{if } \pi(i) \neq i \end{cases}$$

so the X_i 's are rare events (each occurring with probability $1/n$), and are approximately independent of each other.

Comparing the normal and Poisson approximations to the binomial

By our earlier discussion, normal approximation should be safe when $np(1-p) > 10$. In fact we have

Theorem Let S_n be binomial (n, p) . Then for all x ,

$$\left| P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) - \Phi(x) \right| \leq \frac{3}{\sqrt{np(1-p)}}.$$

And if np^2 is small, we know that Poisson approximation works well.

Let's illustrate this in two examples.

Example

Let X be binomial $(10, \frac{1}{10})$. Compare the Poisson and normal approximations to $P(X \leq 1)$.

Sol.

The exact value is

$$\begin{aligned} P(X \leq 1) &= P(X=0) + P(X=1) \\ &= \left(\frac{9}{10}\right)^{10} + 10 \frac{1}{10} \left(\frac{9}{10}\right)^9 \approx .7361 \end{aligned}$$

Now $E(X) = np = 10 \cdot \frac{1}{10} = 1$, so the correct Poisson approximation to use is Y is Poisson(1).

We have $np^2 = 10 \left(\frac{1}{10}\right)^2 = \frac{1}{10}$, so by an earlier result, the Poisson approximation will be within $1/10$ of the true probability. In fact it's much better than that:

$$\begin{aligned} P(Y \leq 1) &= P(Y=0) + P(Y=1) \\ &= e^{-1} + e^{-1} \approx .7358 \text{ which is} \\ &\text{within .001 of the exact value.} \end{aligned}$$

560

Next we see how normal approximation performs. Since the number of trials is small, we use the continuity correction to enhance accuracy. With $E(X)=1$ and $\text{Var}(X)=10 \cdot \frac{1}{10} \cdot \frac{9}{10} = \frac{9}{10}$, normal approximation gives

$$\begin{aligned} P(X \leq 1) &= P(X \leq 3/2) \\ &= P\left(\frac{X-1}{\sqrt{9/10}} \leq \frac{3/2-1}{\sqrt{9/10}}\right) \\ &\approx P\left(\frac{X-1}{\sqrt{9/10}} \leq .53\right) \approx \Phi(.53) \\ &\approx .7019 \end{aligned}$$

The answer is clearly not as accurate as Poisson approximation. Since $np(1-p) = 9/10$ is far below 10, the suboptimal performance of the normal approximation isn't a surprise. Note that without the continuity correction, the normal approximation gives .5 (do the calculation as practice), which is a poor estimate.

The next example reverses the 561
roles: Poisson approximation is poor but
normal approximation is good.

Example

Estimate the probability that 40 flips of
a fair coin give exactly 20 heads.

Sol.

Let S = number of heads. Then S is
binomial $(40, 1/2)$ with $E(S) = 20$ and

$$\text{Var}(S) = 40 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = 10.$$

A computer calculation gives the exact
probability

$$P(S=20) = \binom{40}{20} / 2^{40} \approx .1254$$

Normal approximation with the continuity
correction gives

$$\begin{aligned} P(S=20) &= P(19.5 \leq S \leq 20.5) \\ &= P\left(\frac{19.5-20}{\sqrt{10}} \leq \frac{S-20}{\sqrt{10}} \leq \frac{20.5-20}{\sqrt{10}} \right) \end{aligned}$$

562

$$\approx P\left(-.16 \leq \frac{S-20}{\sqrt{10}} \leq .16\right)$$

$$\approx \Phi(.16) - \Phi(-.16)$$

$$= 2\Phi(.16) - 1 \approx .1272.$$

This is quite accurate.

For the Poisson approximation, we compare S to a Poisson(20) random variable Y . A computer calculation gives

$$P(Y=20) = \frac{e^{-20} 20^{20}}{20!} \approx .089$$

This is a poor approximation. The reason is simply that we are not in the regime of rare events, since successes with probability $1/2$ are not rare. We have $np^2 = 40 \cdot 1/4 = 10$, so our earlier upper bound is larger than 1, so useless.

I enjoyed teaching this class; thank you!

Some references for future study

Probability

- 1) An introduction to probability theory and its applications, Volumes 1 and 2, by Feller
- 2) Probability and measure by Billingsley

Markov chains

- 1) Markov chains and mixing times, by Levin, Peres, and Wilmer.
- 2) Finite Markov chains and algorithmic applications by Haggstrom

Combinatorics

- 1) Enumerative combinatorics, Volumes 1 and 2, by Stanley