Math 340: Advanced introduction to probability

Today: basic probabilistic objects

- Ω = sample space / outcome space → elements $\omega \in \Omega$ are outcomes
  = some mathematical set
  = "all the ways your experiment could unfold"
- Events: subsets of Ω
  $A \subseteq \Omega$
  ↳ collection of outcomes
- $\mathbb{P}$, a probability measure.
  Assigns number $\in [0,1] \subset \mathbb{R}$ to
  each event: $A \subseteq \Omega \to \mathbb{P}(A) \in [0,1]$
  ↳ "Probability that A occurs"

Ex: toss a coin 3 times. What is the probability of ≥ 2 heads?

~~Idea 0: $\Omega = \{0,1,2,3\}$~~
~~(loses a lot of information, not the best approach)~~

Idea 1: $\Omega = \{\text{all sequences of tosses}\}$
$= \{H,T\}^3$
$= \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

richer interpretation!

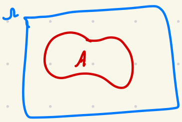
Event "≥ 2 heads": $A = \{HHH, HHT, HTH, THH\}$
C = "tails occurs on toss #2": $C = \{HTH, HTT, TTH, TTT\}$ both events can occur

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{4}{8} = \frac{1}{2}$$
$$= \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \frac{1}{8} \times 4 = \frac{1}{2}$$
} assumption that each outcome has equal probability of occurring

Any probability measure $\mathbb{P}$ must satisfy some axioms:

1) $0 \leq \mathbb{P}(A) \leq 1 \quad \forall A \subseteq \Omega$
2) $\mathbb{P}(\Omega) = 1$
3) (Countable additivity) For any countable collection of events $A_1, \dots, A_n$ which are disjoint $A_i \cap A_j = \emptyset$
Then $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$

$\rightsquigarrow$ $\mathbb{P}$ behaves like "area" or "volume"



$\mathbb{P}(A)$ is "size" of A relative to "size" of Ω
(so Ω has "size" scaled to 1)

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B) \quad \text{(in general)}$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

for any countable collection,

$$\mathbb{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathbb{P}(A_i)$$

"union bound"

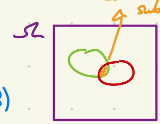
equality when A_i disjoint

Lots follow from above: $\quad A^c := \{\omega \in \Omega \mid \omega \notin A\}$

4) for any $A \subseteq \Omega$, $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$

5) If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

6) for any events, A, B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$



Examples: Tickets in a box $0, 1, \ldots, 9$

Draw a ticket at random. $\mathbb{P}(\text{drawing } 7, 8, \text{ or } 9) = 3/10$

$\Omega = \{0, \ldots, 9\}$
$A = \{7, 8, 9\}$ $\quad \mathbb{P}(A) = 3/10$

Assumption: all outcomes equally likely

Suppose Ω is a discrete set (either finite or countable)
Then for any event A,

$$A = \bigcup_{\omega \in A} \{\omega\}$$

countable union of disjoint events

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$$

Ex: $\mathbb{P}(\{7, 8, 9\}) = \mathbb{P}(\{7\}) + \mathbb{P}(\{8\}) + \mathbb{P}(\{9\})$

Special case: Suppose Ω is a finite set

& all $\omega \in \Omega$ are equally likely i.e. $\omega, \omega' \in \Omega \Rightarrow \mathbb{P}(\{\omega\}) = \mathbb{P}(\{\omega'\})$

Then, for any event A, $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$

Pf: $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = |A| \cdot c$, $\mathbb{P}(\Omega) = 1 = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = |\Omega| \cdot c \Rightarrow c = \frac{1}{|\Omega|} \Rightarrow \mathbb{P}(A) = \frac{|A|}{|\Omega|}$ $\boxtimes$

Ex: Tickets $0, \dots, 9$ in box

1) Draw 3 tickets with replacement
2) Draw 3 tickets without replacement

In case 1: $\mathbb{P}(\text{drawing } 789)$

$\Omega = \{0, \dots, 9\}^3 \to$ equally likely

so $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{|\Omega|} = \frac{1}{1000}$

In Case 2: $\mathbb{P}(\text{drawing } 789 \text{ in any order})$

$\Omega = \{0, \dots, 9\}^3 \setminus \{(d_1, d_2, d_3) \mid d_1 = d_2 \text{ or } d_1 = d_3 \text{ or } d_2 = d_3\}$

or Ω = all subsets of $\{0, \dots, 9\}$ having size 3

$|\Omega| = \binom{10}{3}$

$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{\binom{10}{3}} = \frac{3!\,7!}{10!}$

fact: # of ways to choose k things from a set of n distinct things

is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ "n choose k"

Ex: (Coin tossing)

Toss a fair coin N times ("independent" tosses)

$\mathbb{P}(\underbrace{\text{exactly } k \text{ heads occur}}_{A_k})$

1 = heads
0 = tails

$\Omega = \{0,1\}^N$ = binary sequences of length N

equally likely outcomes

$\mathbb{P}(A_k) = \frac{|A_k|}{|\Omega|} = \frac{|A_k|}{2^N} = \frac{\binom{N}{k}}{2^N}$ → # of ways to choose k slots for 1's among N slots

$|A_k| \to$ choose k spots in length N

$\mathbb{P}(A_k) = 2^{-N}\binom{N}{k}$

Next, $\mathbb{P}(\underbrace{\text{tossing} \geq l \text{ heads}})$

$\mathbb{P}\left(\bigcup_{k=l}^{N} A_k\right)$

$= \mathbb{P}(A_l \cup A_{l+1} \cup \dots \cup A_N)$

$= \sum_{k=l}^{N} \mathbb{P}(A_k)$

$= \sum_{k=l}^{N} 2^{-N}\binom{N}{k}$

fact: if $l = \alpha N$ (fraction of heads), fix $\alpha \in (\frac{1}{2}, 1]$

$\mathbb{P}(\text{tossing} \geq \alpha N \text{ heads}) = \sum_{k \geq \alpha N}^{N} 2^{-N}\binom{N}{k}$ → goes to zero exponentially quickly

Ex: fix a sequence of bits $s = (0,1,1,0,0,0,1,0,1,\dots)$ of length N

m people try to guess it.

each guess is by tossing coins. $\alpha \in (0,1)$

$\mathbb{P}($ someone guesses at least αN bits correctly$)$

$= \mathbb{P}\left(\bigcup_{j=1}^{m} C_j\right)$

C_j: person j gets $\geq \alpha N$ correct

union bound

$\leq \sum_{j=1}^{m} \mathbb{P}(C_j)$

$= m\,\mathbb{P}($a particular person guesses $\geq \alpha N$ correct$)$

$= m \sum_{k=\alpha N}^{N} 2^{-N} \binom{N}{k}$

9/13

Recall: if Ω is finite and all outcomes are equally likely, then $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$

Conditional probability

Let $A, B \subseteq \Omega$ be events with $\mathbb{P}(B) \neq 0$

def $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

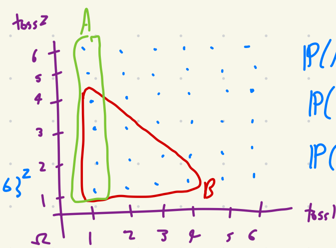
"Probability of A given B"

Rearranging this gives: $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$

Example: toss two fair six-sided dice $\Omega = \{1, \dots, 6\}^2$

A = event that first toss is 1

B = sum of rolls is ≤ 5



$\mathbb{P}(A \cap B) = 4/36$

$\mathbb{P}(B) = \frac{10}{36}$

$\mathbb{P}(A|B) = \frac{4}{10}$

$\mathbb{P}(A|B)$ "Redefine your outcome space"

Notice: $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$

$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \neq \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$

Bayes' formula: $(\mathbb{P}(A) \neq 0 \neq \mathbb{P}(B))$

$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \rightarrow \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$

$= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \frac{\mathbb{P}(A)}{\mathbb{P}(B)}$

$= \mathbb{P}(B|A) \frac{\mathbb{P}(A)}{\mathbb{P}(B)}$

Ex: (Canonical Bayes' example)

Boxes & marbles example:

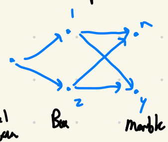
There are two boxes. In box 1, there are 5 red marbles, 10 yellow marbles
In box 2, there are 5 red marbles, 1 yellow marble

Roll a six-sided die. If roll 1, 2, pick a marble from box 1
Else, pick from box 2

$\mathbb{P}(Y) = ?$
event that yellow is drawn

$\Omega := \{(1,r),(1,y),(2,r),(2,y)\}$
$Y = \{(1,y),(2,y)\}$

or, Ω is all paths (left-to-right) in diagram



event that Box 1 is drawn

$\mathbb{P}(B_1) = \frac{1}{3}$ $\qquad$ $\mathbb{P}(Y|B_1) = \frac{10}{15}$

$\mathbb{P}(B_2) = \frac{2}{3}$ $\qquad$ $\mathbb{P}(Y|B_2) = \frac{1}{6}$

What is $\mathbb{P}(Y)$?

Partition rule:

Suppose events $B_1, \dots, B_n \subseteq \Omega$ partition Ω (i.e. $B_i \cap B_j = \emptyset$ if $i \neq j$ and $\bigcup_i B_i = \Omega$)

Then for any event A, A is partitioned by $\{(A \cap B_k)\}_{k=1}^{n}$

By Additivity,

$$\mathbb{P}(A) = \sum_{k=1}^{n} \overbrace{\mathbb{P}(A \cap B_k)}^{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}$$

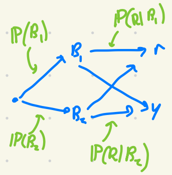
So,
$$\boxed{\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(A|B_k)\mathbb{P}(B_k)}$$
Partition rule! very important

So $\mathbb{P}(Y) = \mathbb{P}(Y|B_1)\mathbb{P}(B_1) + \mathbb{P}(Y|B_2)\mathbb{P}(B_2)$

$= \frac{10}{15} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{2}{3}$

$= \boxed{\frac{1}{3}}$

Ω:



$\mathbb{P}(R) = \mathbb{P}(R|B_1)\mathbb{P}(B_1) + \mathbb{P}(R|B_2)B_2$

= "sum of weight of each path to r"

Consider: $\mathbb{P}(B_k|Y)$ = "Probability of Box k given that yellow marble was drawn"

$$\mathbb{P}(B_1|Y) = \mathbb{P}(Y|B_1)\frac{\mathbb{P}(B_1)}{\mathbb{P}(Y)} = \frac{\mathbb{P}(Y|B_1)\mathbb{P}(B_1)}{\sum_{k=1}^{2}\mathbb{P}(Y|B_k)\mathbb{P}(B_k)}$$

by partition rule

$$= \frac{10}{15}\cdot\frac{1/3}{1/3}$$

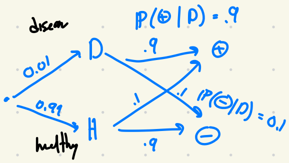
$$= 2/3$$

Diagnosis example:

Suppose 1% of population has disease/trait

A test for disease is 90% accurate

Suppose you test positive. What is $\mathbb{P}$(you have disease)?



$$\mathbb{P}(D|\oplus) = \mathbb{P}(\oplus|D)\frac{\mathbb{P}(D)}{\mathbb{P}(\oplus)}$$

$$= \frac{\mathbb{P}(\oplus|D)\mathbb{P}(D)}{\mathbb{P}(\oplus|D)\mathbb{P}(D) + \mathbb{P}(\oplus|H)\mathbb{P}(H)}$$

$$= \frac{0.9\cdot 0.01}{0.9\cdot 0.01 + 0.1\cdot 0.99}$$

$$= \frac{1}{12} \approx 0.083$$

Suppose Ω is discrete, but outcomes <u>not</u> equally likely.

How do we define $\mathbb{P}$?

Suppose $p: \Omega \to [0,1]$ and $\sum_{\omega\in\Omega} p(\omega) = 1$

Such a function is called a probability mass function

Then <u>def</u> $\mathbb{P}(A) := \sum_{\omega\in A} p(\omega)$ for any $A \subseteq \Omega$.

This <u>defines</u> a probability measure $\mathbb{P}$.

So to define $\mathbb{P}$, it suffices to find the mass function p.

9/5

<u>Recall</u>:

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

<u>Independence</u>: Event A is independent of B

if $\mathbb{P}(A|B) = \mathbb{P}(A)$

$\Leftrightarrow \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$

$\Leftrightarrow \mathbb{P}(A\cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ (symmetric in A, B)

"Events A, B are independent

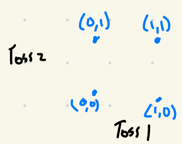
if $\mathbb{P}(A\cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$"

fact: if A, B are independent then so are A^c and B

A^c and B^c

$\mathbb{P}(A^c\cap B) = \mathbb{P}(A^c)\,\mathbb{P}(B)$

<u>Ex.</u> Toss two (fair) coins $\Omega = \{0, \underset{\text{"heads"}}{1}\}^2$



Consider a few simple events: H_1 = first toss is heads

T_2 = second toss is tails

B = both land tails

Intuitively, H_1 & T_2 independent

$\mathbb{P}(H_1 \cap T_2) = \mathbb{P}(\{(1,0)\}) = \frac{1}{4}$

$\mathbb{P}(H_1) = \frac{1}{2}$ $\rightarrow$ $\mathbb{P}(H_1)\mathbb{P}(T_2) = \frac{1}{4}$

$\mathbb{P}(T_2) = \frac{1}{2}$

Claim: H_1, B <u>not</u> independent $\mathbb{P}(H_1 \cap B) = 0$

but $\mathbb{P}(H_1) \neq 0$

$\mathbb{P}(B) \neq 0$

so these events are dependent

Independence of multiple events: Suppose $A_1, A_2, \dots, A_n$ are some events. They are <u>independent</u> if for any set of indices $J \subseteq \{1, 2, \dots, n\}$

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

So if $J = \{1, \dots, n\}$ then $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$

Does pairwise independence imply independence as a collection?

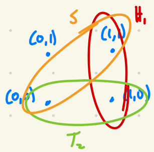
Suppose (A,B), (B,C), (C,A) independent. Must A, B, C be independent as a collection of three events?

No. Ex: toss two fair coins $\Omega = \{0,1\}^2$

Consider events $\overset{\{(1,0),(1,1)\}}{H_1}$, $\overset{\{(0,0),(1,0)\}}{T_2}$, S

S = two tosses are same $= \{(0,0), (1,1)\}$

Claim: pairwise independent, but not as triple.



$$\mathbb{P}(H_1 \cap T_2 \cap S) = \mathbb{P}(\emptyset) = 0$$

But $\mathbb{P}(H_1 \cap T_2) = \frac{1}{4} = \underset{1/2}{\mathbb{P}(H_1)}\, \underset{1/2}{\mathbb{P}(T_2)}$

$\mathbb{P}(T_2 \cap S) = \frac{1}{4} = \underset{1/2}{\mathbb{P}(T_2)}\, \underset{1/2}{\mathbb{P}(S)}$

$\mathbb{P}(H_1 \cap S) = \frac{1}{4} = \underset{1/2}{\mathbb{P}(H_1)}\, \underset{1/2}{\mathbb{P}(S)}$

<u>Fact</u>: suppose $A_1, \dots, A_n$ are independent. Then $B_1, \dots, B_n$ are also independent, when either $B_i = A_i$ or A_i^c

Also, for any $k \in \{1, \dots, n\}$
if C is "built from $A_1, \dots, A_k$"
and D is "built from $A_{k+1}, \dots, A_n$"
then C and D are independent.

<u>Ex</u>: toss n fair coins H_k = event that heads occurs on jth toss

$$H_j = \{\omega \mid \omega = (\_\_\_\_ \underset{j\text{th position}}{1} \_\_\_\_)\}$$

$H_1, \ldots, H_n$ are independent (assuming equally likely outcomes)

$N=100$

C = event that ≤ 10 heads among tosses $1, \ldots, 99$

D = event that last 20 tosses are tails

$$D = \bigcap_{j=81}^{100} H_j^c$$

Independent "coin" toss (i.e. independent repeated trials)

<u>Goal:</u> build probability model for "tossing a <u>biased</u> coin N times s.t. tosses are independent and $\mathbb{P}(H_i) = q$" $\longleftrightarrow$ "N tosses of a q-coin"

$\mathbb{P}(H_i) = q$

$\mathbb{P}(T_i) = 1-q$

<u>Ex</u>. Box w/ 7 blue marbles, 3 red, 10 white

If thinking about # of blue

H_k = "heads on kth draw"

= "drawing blue on kth draw"

T_k = "tails on kth draw"

= "drawing red/white on kth draw"

Let $\Omega = \{0,1\}^N$ binary sequences of length N.

Need to define $\mathbb{P}$

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} p(\omega)$$

for fair coin (all outcomes are equally likely)

$$p(\omega) = \frac{1}{|\Omega|} = 2^{-N}$$

for q-coin consider a particular sequence.

Suppose $N=5$ Consider

$$\omega = (0,1,1,0,1)$$

$$\{\omega\} = T_1 \cap H_2 \cap H_3 \cap T_4 \cap H_5$$

if independence is to hold,

$$\mathbb{P}(T_1 \cap H_2 \cap H_3 \cap T_4 \cap H_5) = \mathbb{P}(T_1)\,\mathbb{P}(H_2)\,\mathbb{P}(H_3)\,\mathbb{P}(T_4)\,\mathbb{P}(H_5)$$
$$= q^3(1-q)^2$$

In general, if ω has k heads and $N-k$ tails, define

$$\mathbb{P}(\{\omega\}) = p(\omega) = q^k(1-q)^{N-k} \in (0,1)$$

Need $\sum_{\omega\in\Omega} p(\omega) = 1$

Let A_k be the event that exactly k heads occur ($N-k$ tails)

$$\mathbb{P}(A_k) = \sum_{\omega\in A_k} p(\omega)$$
$$= \sum_{\omega\in A_k} q^k(1-q)^{N-k}$$
$$= |A_k|\, q^k(1-q)^{N-k}$$
$$= \binom{N}{k} q^k(1-q)^{N-k}$$

$$\mathbb{P}(A_k) = \binom{N}{k} q^k(1-q)^{N-k}$$

then $\Omega = \bigcup_{k=0}^{N} A_k$ so

$$\mathbb{P}(\Omega) = \sum_{k=0}^{N} \mathbb{P}(A_k)$$
$$= \sum_{k=0}^{N} \binom{N}{k} q^k(1-q)^{N-k}$$
$$= (q+1-q)^N$$
$$= 1 \quad \text{as desired} \ \square$$

Binomial thm:
$$(x+y)^N = \sum_{k=0}^{N}\binom{N}{k} x^k y^{N-k}$$

9/10

Bean Problem (#6 on HW1)

G_k = event that person k draws gold bean

$\{G_k\}_{k=1}^{n}$ are independent

$$\mathbb{P}\left(\bigcup_{k=1}^{n} G_k\right) \le \sum_{k=1}^{n} \mathbb{P}(G_k) = \frac{n}{1000}$$

$$\mathbb{P}\left(\bigcup_{k=1}^{n} G_k\right) = 1 - \mathbb{P}\left(\left(\bigcup_{k=1}^{n} G_k\right)^c\right)$$

de Morgan's law

$$= 1 - \mathbb{P}\Big(\bigcap_{k=1}^{n} \underbrace{G_k^c}_{\text{independent}}\Big)$$

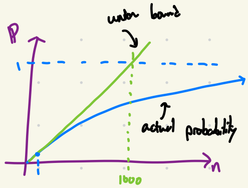
$$= 1 - \prod_{k=1}^{n} \mathbb{P}(G_k^c)$$

$$= 1 - \prod_{k=1}^{n} \left(1 - \mathbb{P}(G_k)\right)$$

$$= 1 - \prod_{k=1}^{n} \left(1 - \tfrac{1}{1000}\right)$$

$$= 1 - \left(1 - \tfrac{1}{1000}\right)^n$$

$$= 1 - e^{n \underbrace{\ln\left(1 - \frac{1}{1000}\right)}_{<0}}$$



Birthday Problem

a) n people: what is prob that there is a birthday match? ← bigger, like a lot...

b) what is prob that someone has my birthday?

n people draw labeled beans $1, \dots, 365$ with replacement

a) two people draw same bean

b) at least one person draws specific bean

b) $\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{k=1}^{n} G_k\right) = 1 - \left(1 - \frac{1}{365}\right)^n \leq \frac{n}{365}$

a) $\Omega = \{1, 2, \dots, 365\}^n$

D_k = event that first k draws are distinct

Want: $\mathbb{P}(D_n)$

$\mathbb{P}(D_1) = 1$

$\mathbb{P}(D_2) = \frac{365-1}{365}$

$$\mathbb{P}(D_{k+1} \cap D_k) = \mathbb{P}(D_{k+1}) = \underbrace{\mathbb{P}(D_{k+1} \mid D_k)}_{\frac{365-k}{365}} \mathbb{P}(D_k)$$

Because $D_{k+1} \subset D_k$

$$\mathbb{P}(D_{k+1}) = \left(\frac{365-k}{365}\right) \mathbb{P}(D_k)$$

$$\mathbb{P}(D_n) = \prod_{k=1}^{n-1} \left(\frac{365-k}{365}\right)$$

$$= \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right)$$

$\mathbb{P}(D_{366}) = 0 \qquad \mathbb{P}(D_{22}) \approx 0.5$

Generalization: m distinct marbles (HW problem)

Last time: coin tossing, p-coin

$\mathbb{P}(\text{"heads"}) = p$

$\Omega = \{0,1\}^n$

A_k = event that exactly k heads (1's) are tossed

$\mathbb{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$

# of such ω

$\mathbb{P}(\{\omega\}) = p^k(1-p)^{n-k}$ if ω has exactly k 1's

particular sequence

(note: if $p = \frac{1}{2}$, $\mathbb{P}(A_k) = 2^{-n}\binom{n}{k}$ which we've shown)

Large Deviation Estimate

- fair coin ($p = \frac{1}{2}$)
- make n indep. tosses.

Goal: make sense of "probably about half tosses will be heads"

Theorem:
Fix $\varepsilon > 0$ small.

$$\mathbb{P}\Bigg(\underbrace{\bigcup_{k \geq n(\frac{1}{2}+\varepsilon)} A_k}_{\text{event that at least } n(\frac{1}{2}+\varepsilon) \text{ heads are tossed}}\Bigg) \leq e^{-\varepsilon^2 n}$$

Corollary: $$\mathbb{P}\Bigg(\underbrace{\bigcup_{n(\frac{1}{2}-\varepsilon) \leq k \leq n(\frac{1}{2}+\varepsilon)}}_{\text{fraction of heads between } \frac{1}{2} \pm \varepsilon} A_k\Bigg) \geq 1 - 2e^{-\varepsilon^2 n}$$

Pf.

$$= 1 - \mathbb{P}\left(\left(\bigcup_{n(\frac{1}{2}-\varepsilon) \leq k \leq n(\frac{1}{2}+\varepsilon)} A_k\right)^c\right)$$

$$= 1 - \mathbb{P}\left(\left(\bigcup_{k \geq n(\frac{1}{2}+\varepsilon)} A_k\right) \cup \left(\bigcup_{k \leq n(\frac{1}{2}-\varepsilon)} A_k\right)\right)$$

$$= 1 - \left(\mathbb{P}\left(\bigcup_{k \geq n(\frac{1}{2}+\varepsilon)} A_k\right) + \mathbb{P}\left(\bigcup_{k \leq n(\frac{1}{2}-\varepsilon)} A_k\right)\right)$$

same b/c symmetric w.r.t. heads/tails

$$\geq 1 - 2e^{-\varepsilon^2 n}$$

Pf. (Thm) See canvas I ain't writing allat. $\boxed{\Sigma}$

Law of Small numbers

→ large n (many tosses)

→ small p (unlikely success)

"np is not too big"

for $p \in [0,1]$, $n \geq 1$, $k = 0, 1, \dots, n$

define $f(n,p,k) = \binom{n}{k} p^k (1-p)^{n-k}$

Thm: for any $\lambda > 0$,

$$\lim_{n \to \infty} f(n, \tfrac{\lambda}{n}, k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{for } k = 0, 1, \dots$$

choose $\lambda = np$

$k \mapsto e^{-\lambda} \frac{\lambda^k}{k!}$ defines the Poisson (λ) distribution.

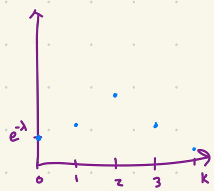
it is a probability mass function on non-neg. integers.

$$e^{-\lambda} \frac{\lambda^k}{k!} > 0 \text{ for all } k \in \mathbb{N} \cup \{0\}$$

$$\sum_{k=0}^{\infty} = e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = e^{-\lambda} e^{\lambda} = e^0 = 1$$

$$g(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$g(0) = e^{-\lambda} \quad g(1) = e^{-\lambda} \lambda \quad g(2)\ e^{-\lambda} \frac{\lambda^2}{2}$$



Pf When $n \gg 1$ large, p small,

$$\mathbb{P}(A_k) = f(n,p,k) \approx e^{-\lambda} \frac{\lambda^k}{k!} \text{ when } \lambda = np.$$

$$f(n, \tfrac{\lambda}{n}, k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$\frac{1}{n^k} \frac{n!}{(n-k)!} = \frac{1}{n^k} \underbrace{(n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1))}_{k \text{ terms}}$$

$$= \prod_{j=0}^{k-1} \left(\frac{n-j}{n} \right) = \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \xrightarrow{n \to \infty} 1$$

k fixed, so as $n \to \infty$

$1 - \frac{\lambda}{n} \to 1$

$(\quad)^{-k} \to 1$

$$\ln\left(1 - \frac{\lambda}{n}\right)^n = n \ln\left(1 - \frac{\lambda}{n}\right) = \frac{\frac{d}{dn} \ln\left(1 - \frac{\lambda}{n}\right)}{\frac{d}{dn} \frac{1}{n}} \to \frac{\frac{1}{1 - \frac{\lambda}{n}} \frac{\lambda}{n^2}}{-1/n^2} = \frac{-\lambda}{1 - \frac{\lambda}{n}} \xrightarrow{n \to \infty} -\lambda \text{ so } \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

All together, limit goes to $e^{-\lambda} \frac{\lambda^k}{k!}$

n tosses of a p-coin

$\mathbb{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$

exactly k heads

$$\sum_{k\geq 0} \binom{n}{k} p^k (1-p)^{n-k} = (p+(1-p))^n = 1$$

$$\sum_{k\geq 0} \mathbb{P}(A_k) = \mathbb{P}\Big(\bigcup_{k\geq 0} A_k\Big) = \mathbb{P}(\Omega) = 1$$

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ how big? need approx. for $n!$

Stirling: $n! \underset{n\to\infty}{\sim} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

meaning: $\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1$

equivalently: $\ln(n!) - \ln\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right) = 0$

so for n large, $\ln(n!) \approx \ln\left(\sqrt{2\pi e}\left(\frac{n}{e}\right)^n\right)$

$= n\ln\left(\frac{n}{e}\right) + \ln\left(\sqrt{2\pi n}\right)$

Simpler version: for any $n \geq 2$

$$\left(\frac{n}{e}\right)^n e \leq n! \leq n\left(\frac{n}{e}\right)^n e^2$$

equivalently:

$$(n\ln(n) - n) + 1 \leq \ln(n!) \leq (n\ln n - n) + 2 + \ln(n)$$

gap: $\ln(n) + 1$

when $n \gg 1$, $\ln(n!)$ is well-approximated by $n\ln(n) - n$

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

$\ln(n!) = \sum_{k=1}^{n} \ln(k)$

$$\int_1^n \ln(x)\,dx \leq \sum_{k=1}^{n} \ln(k) \leq \int_2^{n+1} \ln(x)\,dx$$

$n\ln n - n + 1 \leq \qquad \leq n\ln n - n + 2 + \ln n$

9/1

Chapter 2: random variables n shit

Random Variables

$X = X(\omega)$ is a random variable

a function on Ω

$X : \Omega \to \mathbb{R}$ could be something else; i.e., vector, pt. on manifold

Given $\omega \in \Omega$ $X(\omega)$ is a number

We write $\{X = a\} = \{\omega \in \Omega \mid X(\omega) = a\}$

$$\mathbb{P}(X=a) = \mathbb{P}(\{X=a\}) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = a\})$$

<u>Ex</u>: toss fair coin n times

$\Omega = \{0,1\}^n$

$\omega \in \Omega$ is a sequence

Let X = # heads tossed

Y = # tails

Z = max # consecutive heads

$n = 8$

$\omega = 00111101$ $\quad$ $\omega' = 00000000$

$X(\omega) = 5$ $\quad$ $X(\omega') = 0$

$Y(\omega) = 3$ $\quad$ $Y(\omega') = 8$

$Z(\omega) = 4$ $\quad$ $Z(\omega') = 0$

If X, Y random variables, then

$Z := X+Y$ also random variable!

$Z(\omega) = X(\omega) + Y(\omega)$

Space of random variables forms a ring?

Range of X: values that X can take:

$\mathcal{R}(X)$ = all # X can take

$= \{a \in \mathbb{R} \mid \mathbb{P}(X=a) \neq 0\}$

only in discrete case

Notice: these events

$\{X = a\}$ $\quad a \in \mathcal{R}(X)$

are disjoint events.

We say X is <u>discrete</u> if $\mathcal{R}(X)$ is countable.

In this case, $\{\{X=a\}\}_{a \in \mathcal{R}(X)}$ partitions Ω

So, $\sum_{a \in R(X)} \mathbb{P}(X=a)$

$$= \sum_{a \in R(X)} \mathbb{P}(\{\omega \in \Omega \mid X(\omega)=a\})$$

$$= \mathbb{P}\Big(\bigcup_{a \in R(X)} \{X=a\}\Big) = \mathbb{P}(\Omega) = 1$$

The <u>distribution</u> of a discrete random variable X is the assignment of probabilities to events

$\{X=a\}$ $a \in R(X)$

$\mathbb{P}(X=a)$ $a \in R(X)$

<u>fact</u>: two variables X & Y may have same distribution, even though they are different random variables.

Toss a (fair) coin 4 times.

$\Omega = \{0,1\}^4$

Let $X = X(\omega) = \#$ heads $\quad R(X) = \{0,1,2,3,4\}$

$Y = Y(\omega) = \#$ tails $\quad R(Y) = \{0,1,2,3,4\}$

$\mathbb{P}(X=0) = 2^{-4}$	$\mathbb{P}(Y=0) = 2^{-4}$
$\mathbb{P}(X=1) = \binom{4}{1}2^{-4}$	$\mathbb{P}(Y=1) = \binom{4}{1}2^{-4}$
$\mathbb{P}(X=2) = \binom{4}{2}2^{-4}$	$\mathbb{P}(Y=2) = \binom{4}{2}2^{-4}$
$\mathbb{P}(X=3) = \binom{4}{3}2^{-4}$	$\mathbb{P}(Y=3) = \binom{4}{3}2^{-4}$

<u>Special Named Distributions:</u>

- Bernoulli(p) distribution, $p \in [0,1]$

  $\mathbb{P}(X=0) = (1-p)$

  $\mathbb{P}(X=1) = p \qquad R(X) = \{0,1\}$

  e.g. toss a p-coin n times

  let $X = \begin{cases} 0 & \text{tails} \\ 1 & \text{heads} \end{cases}$ on toss 7

  $Y = \begin{cases} 0 \\ 1 \end{cases}$ " " " 8

  diff variables, same distribution

- Binomial(n,p) distribution

  $R(X) = \{0,1,\dots,n\} \ni k$

  $\mathbb{P}(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$

  e.g. # heads in n tosses of a p-coin

Sps $X \sim \text{Binomial}(n,p)$

Consider $Y = n - X$ $\quad R(Y) = \{0, \dots, n\}$

$Y \sim \text{Binomial}(n, 1-p)$

- Indicator function Let $A \subset \Omega$ be any event

define a r.v.: $X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{else} \end{cases}$

"indicator of A"

"characteristic function " " "

$$\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(\{\omega \mid \mathbb{1}_A(\omega) = 1\})$$
$$= \mathbb{P}(A)$$
$$\mathbb{P}(\mathbb{1}_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

So $\mathbb{1}_A \sim \text{Bernoulli}(\mathbb{P}(A))$

Sps. A_1, A_2, A_n independent events.

$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_n)$

Let $Z(\omega) = \sum_{k=1}^{n} \mathbb{1}_{A_k}(\omega)$

$= \mathbb{1}_{A_1} + \dots + \mathbb{1}_{A_n}$

$Z \sim \text{Binomial}(n, \mathbb{P}(A_1))$

- $X \sim \text{Poisson}(\lambda)$

if $R(X) = \mathbb{N} \cup \{0\}$ $\quad \lambda > 0$

and

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$k = 0, 1, \dots$

- $X \sim \text{Geometric}(p)$ $\quad p \in (0,1)$

$\mathbb{P}(X = k) = (1-p)^{k-1} p$ $\quad k = 1, 2, \dots$

$= \underbrace{(1-p)(1-p)\cdots(1-p)}_{k-1} p$

$R(X) = \{1, 2, \dots\} = \mathbb{N}$

Check:

$$\sum_{k \in R(X)} \mathbb{P}(X = k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p$$
$$= \sum_{l=0}^{\infty} (1-p)^l p$$
$$= p \frac{1}{1-(1-p)} = 1$$

e.g. toss a p-coin repeatedly

until heads occur.

Let X = # of tosses until (including) first heads

$$\mathbb{P}(X = k) = \overbrace{(1-p)^{k-1}}^{\text{first } k-1 \text{ tosses tails}} \underset{k^{\text{th}} \text{ toss heads}}{p}$$

CDF

Def: the <u>cumulative distribution function</u> of an r.v. X is

$$F(a) = \mathbb{P}(X \leq a)$$

$$F: \mathbb{R} \to [0,1]$$

9/19

Midterm in 2 weeks Mecster 1.1 – 2.3

Random variables: functions on outcome space Ω

$$X: \Omega \to \mathbb{R}$$
$$\mathcal{R}(X) = \text{range of } X$$

Distribution of X (discrete case)

- $\mathbb{P}(X = a) \quad a \in \mathcal{R}(X)$

doesn't tell us everything about X

e.g. Ω = people in Durham

$X(\omega)$ = age

moreso general info

$$\mathbb{P}(X = 25) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = 25\})$$

CDF: cumulative distribution function

for any r.v. X,

$$F_X(r) = F(r) = \mathbb{P}(X \leq r) \quad r \in \mathbb{R}$$

notation

$$F: \mathbb{R} \to [0,1]$$

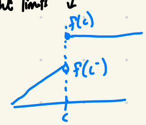
<u>Properties of CDFs</u>

1) $r \mapsto F(r)$ non-decreasing
2) $\lim_{r \to \infty} F(r) = 1$
3) $\lim_{r \to -\infty} F(r) = 0$
4) for $a < b$, $F(b) - F(a) = \mathbb{P}(a < X \leq b)$

5) $r \mapsto F(r)$ is $\overbrace{\text{right-continuous, has left-side limits}}^{\text{closed circle on right side}}$

$F(c^-) = \lim_{r \nearrow c^-} F(r)$

$F(c) = \lim_{r \searrow c^+} F(r)$



6) $\mathbb{P}(X=c) = \overbrace{F(c) - F(c^-)}^{\text{size of jump at } c}$

By (6) if F is continuous at $r=c$, then

$F(c^+) = F(c) = F(c^-)$ then

$\mathbb{P}(X=c) = 0$

<u>Def</u>: A r.v. X is <u>continuously distributed</u> if $F_X(r)$ is continuous on $\mathbb{R}$

<u>Ex</u> Toss <u>fair</u> ($p=\frac{1}{2}$) coin 3 times

$X = \#$ heads tossed

$\Omega = \{0,1\}^3$

$\mathbb{P}(X=0) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$

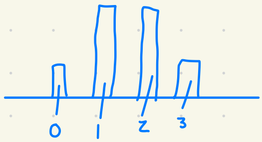
$\mathbb{P}(X=1) = 3\left(\frac{1}{2}\right)^3 = \frac{3}{8}$

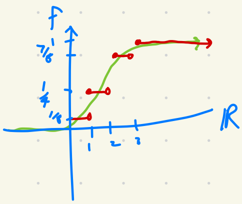
$\mathbb{P}(X=2) = 3\left(\frac{1}{2}\right)^3 = \frac{3}{8}$

$\mathbb{P}(X=3) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$

$F_X(r) = \mathbb{P}(X \le r)$

Visualization (bar graph)





<u>Ex</u>: toss p-coin indefinitely until heads occurs

$X = \#$ of tosses until (including) first head occurs

$R(X) = \{1, 2, 3, \dots\}$

$\mathbb{P}(X > k) = \mathbb{P}(k \text{ tails in a row})$

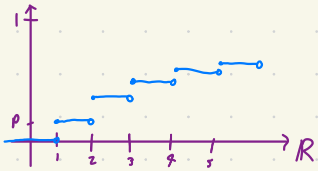
$= (1-p)^k$

$F_X(k) = \mathbb{P}(X \le k) = 1-(1-p)^k$

for $k = 1, 2, \dots,$

$F(k) = \mathbb{P}(X \leq k) = 1 - (1-p)^k$

$F(n)$ constant on $\mathbb{R} \setminus \mathbb{N}$



$$\mathbb{P}(X = k) = F(k) - F(k^-)$$
$$= \mathbb{P}(X \leq k) - \mathbb{P}(X \leq k-1)$$
$$= 1 - (1-p)^k - (1 - (1-p)^{k-1})$$
$$= (1-p)^{k-1} p \longleftarrow \text{Geometric}(p)$$

<u>Independence of R.V.s</u>

Let $X_1, X_2, \dots X_n$ be real-valued random variables

We say they are <u>independent</u> if

for any intervals $I_1, I_2, \dots, I_n \subseteq \mathbb{R}$

$\star$ $\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n)$

$$= \mathbb{P}(X_1 \in I_1)\mathbb{P}(X_2 \in I_2) \cdots \mathbb{P}(X_n \in I_n)$$
$$= \prod_{j=1}^{n} \mathbb{P}(X_j \in I_j)$$

Thm: TFAE:

1) $X_1, \dots, X_n$ are independent

2) $\star$

3) for any $a_1, a_2, \dots, a_n \in \mathbb{R}$

$$\mathbb{P}(X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n)$$
$$= \prod_{j=1}^{n} \mathbb{P}(X_j \leq a_j)$$

and if X_i is discrete,

4) for any $v_1, v_2, \dots, v_n \in R(X)$

$$\mathbb{P}(X_1 = v_1, X_2 = v_2, \dots, X_n = v_n)$$
$$= \prod_{j=1}^{n} \mathbb{P}(X_j = v_j)$$

Example Toss a p-coin N times $p \in (0,1)$

$X = \#$ heads

$Y = \#$ tails

$X \sim \text{Binomial}(N,p)$

$Y \sim \text{Binomial}(N,1-p)$

If X,Y independent, then

$$\mathbb{P}(X=k, Y=j) = \mathbb{P}(X=k)\mathbb{P}(Y=j) \quad \forall k,j$$

Note: $X_{(\omega)} + Y_{(\omega)} = N$

$$\mathbb{P}(X=N, Y=N) = 0$$
$$\ne$$
$$\mathbb{P}(X=N)\,\mathbb{P}(Y=N) = p^N(1-p)^N$$

Let N be random.

$N \sim \text{Poisson}(\lambda)$ → not true for, say, geometric distribution

$X = \#$ heads $\quad Y = \#$ tails

Given $N=\ell$, toss ℓ p-coins independently

Fact: X,Y independent.

$$\mathbb{P}(X=k, Y=j) = \mathbb{P}(X=k)\mathbb{P}(Y=j)$$

$$\mathbb{P}(X=k) = \sum_{\ell=0}^{\infty} \mathbb{P}(X=k \mid N=\ell)\mathbb{P}(N=\ell)$$

$$= \sum_{\ell=k}^{\infty} \binom{\ell}{k} p^k (1-p)^{\ell-k} \frac{e^{-\lambda}\lambda^{\ell}}{\ell!}$$

summing over ℓ

$$\mathbb{P}(X=k, Y=j) = \sum_{\ell=0}^{\infty} \mathbb{P}(X=k, Y=j \mid \underbrace{N=\ell}_{N=k+j})\,\mathbb{P}(\underbrace{N=\ell}_{N=k+j})$$

$$\binom{k+j}{k} p^k (1-p)^j \frac{e^{-\lambda}\lambda^{k+j}}{(k+j)!}$$

9/24

Expectation

Weighted avg. of values a r.v. can take

$$\mathbb{E}(X) = \sum_{a \in R(X)} a\,\mathbb{P}(X=a)$$ well defined as long as $\sum_{a \in R(X)} |a|\,\mathbb{P}(X=a)$ converges → absolute convergence

Properties $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$

if X,Y independent $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

9/26

Expectation (discrete r.v.'s)

$$\mathbb{E}(X) = \sum_{a \in R(X)} a\,\mathbb{P}(X=a)$$

depends only on distribution

Key property: linearity $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$

Ex: $X \sim \text{Binomial}(n,p)$

$\mathbb{E}(X) = np$

Ex: Suppose $X \sim \text{Poisson}(\lambda)$

$R(X) = \mathbb{Z}_{\geq 0}$

$\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-1)!} \\
&= \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^{j+1}}{j!} \\
&= \lambda e^{-\lambda}\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

Tail Sum formula

Sps $\underbrace{X \text{ nonnegative, } X \text{ integer valued}}_{R(X) \subseteq \mathbb{Z}_{\geq 0}}$

then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

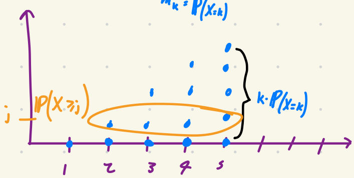
for this special scenario:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\underbrace{\mathbb{P}(X=k)} \overset{\text{or}}{=} \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(X \geq k)}$$

not true that equal

Proof by picture:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \underbrace{\mathbb{P}(X=k)}_{m_k = \mathbb{P}(X=k)} \qquad \sum_{k=1}^{\infty} m_k = 1$$



$$\sum_{k=j}^{\infty} \mathbb{P}(X=k) = \mathbb{P}(X \geq j)$$

Pf (for real this time):

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\,\mathbb{P}(X=k)$$
$$= \sum_{k=1}^{\infty}\left(\sum_{j=1}^{k} \mathbb{P}(X=k)\right)$$
$$= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \mathbb{1}_{j \leq k}\, \mathbb{P}(X=k)$$
$$= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \mathbb{1}_{j \leq k}\, \mathbb{P}(X=k)$$
$$= \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} \mathbb{P}(X=k)$$
$$= \sum_{j=1}^{\infty} \mathbb{P}(X \geq j)$$

Ex: $X \sim \text{Geometric}(p)$

$$R(X) = \{1, 2, 3, \cdots\}$$
$$\mathbb{P}(X=k) = (1-p)^{k-1} p$$

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

Tedious. Let's use tail sum

$$\mathbb{P}(X \geq k) = (1-p)^{k-1}$$

($k-1$ tails in a row, don't care what happens after)

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$
$$= \sum_{k=1}^{\infty} (1-p)^{k-1}$$
$$= \sum_{j=0}^{\infty} (1-p)^{j}$$
$$= \frac{1}{1-(1-p)} = \frac{1}{p}$$

Ex: 1000 marbles

1 gold 999 blue

prize: $Y = \begin{cases} \$1 & \text{gold} \\ \$0 & \text{blue} \end{cases}$

fair price? $\mathbb{E}(Y) = \sum_{k=0}^{1} k\,\mathbb{P}(Y=k)$

$= 0\,\mathbb{P}(Y=0) + 1\,\mathbb{P}(Y=1)$

$= 1 \cdot \frac{1}{1000} = \frac{1}{1000}$

fair price

Different prize

X = # draws until gold

get \$1 each time you draw, draw until you get gold

$X \sim \text{Geometric}\left(\frac{1}{1000}\right)$

$\mathbb{E}(X) = \frac{1}{p} = \$1000$

Ex: toss a fair coin $\left(p = \frac{1}{2}\right)$

Each time tails is tossed, double pot

when you toss heads, take home prize

$X = \$\text{prize}$ $\quad Y$ = # tosses until heads

$= 2^{Y-1}$ $\quad Y \sim \text{Geometric}\left(\frac{1}{2}\right)$

$\mathbb{E}(X) = \mathbb{E}\left(2^{Y-1}\right)$

$= \sum_{k=1}^{\infty} 2^{k-1}\,\mathbb{P}(Y=k)$

$= \sum_{k=1}^{\infty} 2^{k-1}(1-p)^{k-1}\,p$ → what about p-coin? When is this finite? Converges only if $p > \frac{1}{2}$

$= \sum_{k=1}^{\infty} 2^{k-1}\left(\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right)$

$= \sum_{k=1}^{\infty} \frac{1}{2}$ diverges to ∞

Weak Law of Large Numbers

Suppose $X_1, X_2, \cdots$ i.i.d. (independent & identically distributed)

Suppose $\mathbb{E}(X) = \mu$. Consider:

$$\frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}(X_1 + \cdots + X_n)$$

Statement: for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) - \mu\right| > \varepsilon\right) = 0$$

equivalently:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) - \underset{\mathbb{E}(X)}{\mu}\right| \le \varepsilon\right) = 1$$

Special case:

Y = fraction of heads in n tosses of a fair coin

$\mathbb{P}(Y_n > \varepsilon + \frac{1}{2}) \le e^{-\varepsilon^2 n}$

if $X_k = \mathbb{1}_{H_k}(\omega)$

heads on k^{th} toss

$\sum_{k=1}^{n} X_k$ = # heads in n tosses

$\frac{1}{n}\sum X_k$ = fraction of heads in n tosses $= Y_n$

$$\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2} = \frac{1}{2}$$

$$\mathbb{P}\left(\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) - \overset{1/2}{\mu} > \varepsilon\right) \le e^{-\varepsilon^2 n}$$

fraction of heads

satisfy WLLN $\lim_{n\to\infty} e^{-\varepsilon^2 n} = 0$

more precise: we have a specific bound.

10/01

Midterm Practice

First 6 problems were previous midterm

1.) 6-sided die rolled 1 time

D = value rolled on die,

fair coin tossed D times

Given all tosses same, $\mathbb{P}(D=6)$?

$$\mathbb{P}(D=6 \mid \text{all toss same}) = \frac{\mathbb{P}(D=6 \cap \text{all toss same})}{\mathbb{P}(\text{same toss})} = \frac{2\cdot\frac{1}{2^6}}{\sum_{i=1}^{6}\mathbb{P}(\text{same toss}\mid D=i)\,\mathbb{P}(D=i)} = \frac{1/2^5}{\sum \frac{2}{2^i}\cdot\frac{1}{6}} = \boxed{\frac{1/2^5}{1+\frac{1}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^4}+\frac{1}{2^5}}}$$

2.] $p \in (0,1)$ toss p-coin 3 times

A = first toss heads

B = third toss tails

$X = 5\mathbb{1}_A - 3\mathbb{1}_B$

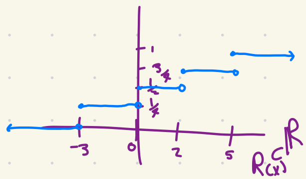
$\mathcal{R}(X) = \{0, -3, 2, 5\}$

$\mathbb{P}(X=0) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = p(1-p)$

$\mathbb{P}(X=-3) = (1-p)^2$

$\mathbb{P}(X=2) = p(1-p)$

$\mathbb{P}(X=5) = p^2$



3.] 100 beans

90 red, 10 white

100 people draw a bean, no replacement

$\mathbb{P}(\underbrace{\text{20th bean = first white bean}}_{A})$

Ω = 90 r's 10 w's (red, white)

$\{\underbrace{0, 0, \dots, 0}_{19}, 1, \overbrace{\dots}^{\text{can be anything}}\}$ ↳ arrange 9 white beans,

$= \binom{80}{9}\binom{100}{10}^{-1}$

ways, $|\Omega|$

F = event that person 1 draws red bean

L = last person draws red bean

$\mathbb{P}(L|F) = \frac{\mathbb{P}(L \cap F)}{\mathbb{P}(F)} = \frac{\binom{98}{10}\binom{100}{10}^{-1}}{90/100}$

$\binom{98}{10}\binom{100}{10}^{-1}$ not equal!

4.]

10/08

Today
- Chebychev's Inequality
- Variance
- Weak LLN (finite variance)

Thm: (Chebychev's Inequality)

Sps X r.v., X nonnegative ≥ 0

Then for any $v > 0$,

$$\mathbb{P}(X \geq v) \leq \frac{1}{v}\mathbb{E}(X)$$

mean

Note: only useful when <1, i.e. $v > \mathbb{E}(X)$

e.g. let $v = k\,\mathbb{E}(X)$

$$\mathbb{P}(X \geq k\,\mathbb{E}(X)) \leq \frac{1}{k}$$

Ex: tickets in a box with nonnegative $\$$.

$\mathbb{P}(X \geq \$1{,}000{,}000)$?

If $\mathbb{E}(X) = \$100$, then → more likely to draw $\$1M$ $\Rightarrow \mathbb{E}(X)$ would be greater

$$\mathbb{P}(X \geq 10^6) \leq \frac{10^2}{10^6} = \frac{1}{10^4}$$

Corollary: Sps Y is any r.v. with $\mathbb{E}(Y) = \mu$ $\quad |\mu| < \infty$

Let $t > 0$

$$\mathbb{P}(|Y-\mu| \geq t)$$

$$= \mathbb{P}(\underbrace{|Y-\mu|^2}_{X} \geq \underbrace{t^2}_{v}) \leq \frac{1}{t^2}\underbrace{\mathbb{E}(|Y-\mu|^2)}_{\substack{\text{Variance of } Y \\ VAR(Y)}}$$

Thm: If Y has finite mean $\mathbb{E}(Y) = \mu$,

then $\mathbb{P}(|Y-\mu| \geq t) \leq \frac{1}{t^2} VAR(Y)$.

Pf: (Chebychev)

Let $X \geq 0$. for $v > 0$ define

$$A_v = \{X \geq v\}$$
$$A_v^c = \{X < v\}$$

Note: $\mathbb{1}_{A_v} + \mathbb{1}_{A_v^c} = 1$

So $X(\omega) = X(\omega)\cdot 1 = X(\omega)\left(\mathbb{1}_{A_v}(\omega) + \mathbb{1}_{A_v^c}(\omega)\right)$

$$= X\mathbb{1}_{A_v} + X\mathbb{1}_{A_v^c}$$

$$\mathbb{E}(X) \overset{\text{linearity}}{=} \mathbb{E}(X\mathbb{1}_{A_v}) + \mathbb{E}(X\mathbb{1}_{A_v^c})$$

Therefore

$$\mathbb{E}(X) \geq \mathbb{E}(X\mathbb{1}_{A_v})$$

$$X(\omega)\mathbb{1}_{A_v}(\omega) = \begin{cases} 0 & \omega \notin A_v \\ \underbrace{X(\omega)}_{X(\omega)\geq v} & \omega \in A_v \end{cases}$$

So claim:

$$X(\omega)\mathbb{1}_{A_v}(\omega) \geq v\mathbb{1}_{A_v}(\omega)$$

$$\begin{aligned}\mathbb{E}(X) \geq \mathbb{E}(X\mathbb{1}_{A_v}) \geq \mathbb{E}(v\mathbb{1}_{A_v}) &= v\,\mathbb{E}(\mathbb{1}_{A_v}) \\ &= v\,\mathbb{P}(A_v) \\ &= v\,\mathbb{P}(X \geq v)\end{aligned}$$

$$\mathbb{P}(X \geq v) \leq \frac{1}{v}\mathbb{E}(X)$$

If $\mathbb{E}(Y) = \mu$

$$\mathbb{P}(|Y-\mu| \geq t) \leq \frac{1}{t^2}VAR(Y) := \frac{1}{t^2}\mathbb{E}(|Y-\mu|^2)$$

<u>Def</u>: Suppose Y is r.v. with $\mathbb{E}(Y) = \mu$ finite

$$\begin{aligned} VAR(Y) &= \mathbb{E}(|Y-\mu|^2) \geq 0 \\ &= \mathbb{E}\left((Y-\mathbb{E}(Y))^2\right)\end{aligned}$$

<u>Properties:</u>

0) If $\mathbb{E}(Y^2) < \infty$ then $VAR(Y) < \infty$ and $VAR(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$
1) Scaling: for any $\alpha \in \mathbb{R}$, $VAR(\alpha Y) = \alpha^2 VAR(Y)$
2) Shift: for any $c \in \mathbb{R}$ $VAR(Y+c) = VAR(Y)$
3) Sums: sps X, Y r.v.s with finite variance.

   Then $VAR(X+Y) < \infty$ and

   $$VAR(X+Y) = VAR(X) + VAR(Y) + 2COV(X,Y)$$

   Where $COV(X,Y) := \mathbb{E}((X-\mu_X)(Y-\mu_Y))$

4) If X, Y independent, then $COV(X,Y)=0$ so "uncorrelated", doesn't imply independence

$$VAR(X+Y) = VAR(X)+VAR(Y)$$ $\longrightarrow$ "pythagorean thm for Random variables"

Properties:

$$\mathbb{E}((Y-\overbrace{\mathbb{E}(Y)}^{\mu})^2)$$
$$= \mathbb{E}(Y^2 - 2Y\mu + \mu^2)$$
$$= \mathbb{E}(Y^2) - 2\mu\mathbb{E}(Y) + \mu^2$$
$$= \mathbb{E}(Y^2) - \mu^2$$
$$= \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$$

Scaling:
$$Var(\alpha Y) = \mathbb{E}((\alpha Y)^2) - (\mathbb{E}(\alpha Y))^2$$
$$= \alpha^2(\mathbb{E}(Y^2) - (\mathbb{E}(Y))^2)$$
$$= \alpha^2 VAR(Y)$$

Shifts:
$$Var(\overbrace{Y+c}^{Z}) = \mathbb{E}(|Z-\mu_Z|^2)$$
$$= \mathbb{E}(|Y+c-\mu_Y-c|^2)$$
$$= VAR(Y)$$

$$\mu_Z = \mathbb{E}(Z) = \mathbb{E}(Y+c) = \mathbb{E}(Y)+c = \mu_Y + c$$

Sums:
$$VAR(X+Y)$$
$$= \mathbb{E}(|X+Y-\mathbb{E}(X+Y)|^2)$$
$$= \mathbb{E}(|(X-\mu_X)+(Y-\mu_Y)|^2)$$
$$= \mathbb{E}((X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y))$$
$$= \mathbb{E}((X-\mu_X)^2) + \mathbb{E}((Y-\mu_Y)^2) + 2\mathbb{E}((X-\mu_X)(Y-\mu_Y))$$
$$= VAR(X) + VAR(Y) + 2COV(X,Y)$$

If X, Y indep, then

$$\mathbb{E}((X-\mu_X)(Y-\mu_Y)) = \overbrace{\mathbb{E}(X-\mu_X)}^{\mathbb{E}(X)-\mu_X=0}\mathbb{E}(Y-\mu_Y) = 0$$

Standard deviation: $SD(X) = \sqrt{VAR(X)}$

"$VAR(X) = \sigma^2$" ← standard deviation squared

Corollary: If $X_1, X_2, \dots, X_n$ independent,

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

So if $X_1, \dots, X_n$ iid with $VAR(X_i) = \sigma^2$

Then $VAR(X_1, \dots, X_n) = n\sigma^2$

Suppose $X_1, \dots, X_n$ independent with same mean, variance

$\mathbb{E}(X) = \mu \quad Var(X) = \sigma^2$

Then $\mathbb{E}(X_1 + \dots + X_n) = n\mu$

$VAR(X_1 + \dots + X_n) = n\sigma^2$

Markov:

$$\mathbb{P}\left(\left|(X_1 + \dots + X_n) - n\mu\right| \geq t\right) \leq \frac{1}{t^2} Var(X_1 + \dots + X_n)$$
$$= \frac{n\sigma^2}{t^2}$$

Let $t = \varepsilon n$

$$\mathbb{P}\left(\left|(X_1 + \dots + X_n) - n\mu\right| \geq \varepsilon n\right) \leq \frac{n\sigma^2}{(\varepsilon n)^2} = \frac{\sigma^2}{\varepsilon^2 n}$$

Weak LLN:

Suppose $X_1, X_2, \dots$ is a sequence of independent variables with same distribution and $\mathbb{E}(X_i) = \mu$

(Sps $VAR(X_i) = \sigma^2 < \infty$)

Then for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(\left|\left(\frac{X_1 + \dots + X_n}{n}\right) - \mu\right| \geq \varepsilon\right) = 0$$

Pf:

$$\mathbb{P}\left(\left|\left(\frac{X_1 + \dots + X_n}{n}\right) - \mu\right| \geq \varepsilon\right) = \mathbb{P}\left(\left|(X_1 + \dots + X_n) - n\mu\right| \geq n\varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

which goes to zero as $n \to \infty$.

Multiple r.v.'s + Joint distribution

Sps X, Y pair of discrete r.v.s

Ω = outcome space

$X: \Omega \to \mathbb{R} \quad Y: \Omega \to \mathbb{R}$

Their joint distribution is the assignment

$$(a, b) \longmapsto \mathbb{P}(X = a, Y = b)$$

for all $a \in \mathbb{R}(X) \quad b \in \mathbb{R}(Y)$

Meester notation: $p_{XY}(a, b) = \mathbb{P}(X = a, Y = b)$

defines a mass function on $\mathbb{R}^2$

$$\sum_{a \in \mathbb{R}(X)} \sum_{b \in \mathbb{R}(Y)} p_{XY}(a, b) = 1$$

Note: $\mathbb{P}(X=a) = \sum_{b \in \mathcal{R}(Y)} \mathbb{P}(X=a, Y=b)$

marginal distribution of X

" " " Y

↳ $\mathbb{P}(Y=b) = \sum_{a \in \mathcal{R}(X)} \mathbb{P}(X=a, Y=b)$

$= \sum_{a \in \mathcal{R}(X)} \mathbb{P}(Y=b \mid X=a)\mathbb{P}(X=a)$

Given $a \in \mathcal{R}(X)$,

$b \mapsto \mathbb{P}(Y=b \mid X=a)$

defines the <u>conditional distribution</u> of Y given $X=a$

Similarly, conditional distribution of X given $Y=b$ is

$$a \mapsto \mathbb{P}(X=a \mid Y=b) = \frac{\mathbb{P}(X=a, Y=b)}{\mathbb{P}(Y=b)}$$

Recall: X, Y independent $\iff$

$$\mathbb{P}(X=a, Y=b) = \mathbb{P}(X=a)\mathbb{P}(Y=b) \quad \forall a \in \mathcal{R}(X), b \in \mathcal{R}(Y)$$

If X, Y independent, then

$$\mathbb{P}(X=a \mid Y=b) = \mathbb{P}(X=a)$$

Conditional = marginal

* Marginal distribution of a pair (X, Y) do <u>not</u> determine their joint distribution

To reconstruct joint from marginals, we need more info (e.g. independence)

<u>Ex</u>: toss fair coin 4 times

$\Omega = \{0,1\}^4$

X_1 = # heads in tosses 1 & 2

X_2 = " " " " 2 & 3

X_3 = " " " " 3 & 4

$\omega \in \Omega \rightarrow \omega = (\omega_1, \omega_2, \omega_3, \omega_4) \quad \omega_i \in \{0,1\}$

$X_1(\omega) = \omega_1 + \omega_2$

$X_2(\omega) = \omega_2 + \omega_3$

$X_3(\omega) = \omega_3 + \omega_4$

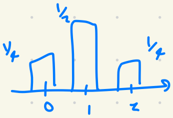
Distribution of X_1:

$X_1 \sim \text{Binomial}(2, \frac{1}{2})$

$X_2 \sim \text{Binomial}(2, \frac{1}{2})$

$X_3 \sim \text{Binomial}(2, \frac{1}{2})$

$\mathbb{P}(X_i = k)$



Joint distribution of pair (X_1, X_2)

$(a, b) \mapsto \mathbb{P}(X_1 = a, X_2 = b)$

2	0	1/8	1/4
1	1/4	2/8	1/8
0	1/8	1/4	0
	0	1	2

Joint distribution of (X_1, X_3)

X_3			
2	1/16	2/16	1/16
1	2/16	4/16	2/16
0	1/16	2/16	1/16
	0	1	2
		X_1	

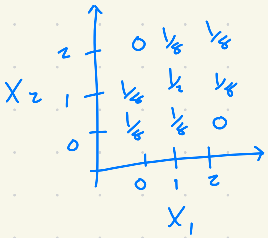
$$\mathbb{P}(X_1 = 0, X_3 = 2) \neq 0 = \mathbb{P}(X_1 = 0, X_2 = 2)$$

$$\mathbb{P}(X_1 = k) = \sum_{j \in R(X_2)} \mathbb{P}(X_1 = k, X_2 = j)$$

X_3			
2	1/16	2/16	1/16
1	2/16	4/16	2/16
0	1/16	2/16	1/16
	0	1	2
		X_1	

project mass down

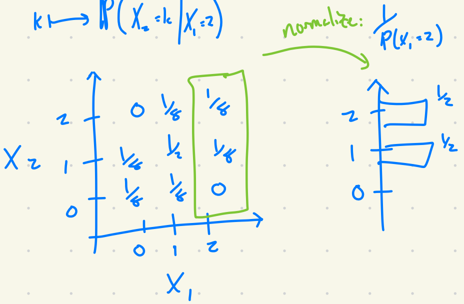
Conditional distributions



Conditional distribution of X_2 given $X_1 = 2$

$$\{X_1 = 2\} = \{(\cdot, \cdot, \cdot)\}$$

$$k \mapsto \mathbb{P}(X_2 = k \mid X_1 = 2)$$



Def Conditional expectation of X given $Y = b$ is

$$\mathbb{E}(X \mid Y = b)$$
$$= \sum_{a \in \mathbb{R}(X)} a\,\mathbb{P}(X = a \mid Y = b)$$

10/17

Continuously Distributed r.v.'s

A random variable $X : \Omega \to \mathbb{R}$ is continuously distributed if $r \mapsto \mathbb{P}(X \le r)$ is continuous for $r \in \mathbb{R}$

CDF

i.e. CDF $F_X(r)$ is a continuous function

i.e. $\lim_{a \to b} F(a) = F(b)$

A continuously distributed r.v. X has a density $f(y)$ if $\mathbb{P}(a < X \le b) = \int_a^b f(y)\,dy$ for all $a < b$, $a, b \in \mathbb{R}$

$$F(b)-F(a) = \mathbb{P}(a<X\leq b) = \int_a^b f(y)\,dy$$

fact if X has CDF F & density f, then

$$F(b)-F(a) = \int_a^b f(y)\,dy$$

$\forall a<b$

$$\Rightarrow F'(y) = f(y)$$

function f is a probability density on $\mathbb{R}$ iff:

1) $f(y) \geq 0 \quad \forall y$

2) $\int_{\mathbb{R}} f(y)\,dy = 1$

Note: $\sup_{y\in\mathbb{R}} f(y)$ may be >1

$$\mathbb{P}(a<X\leq b) = \int_a^b f(y)\,dy$$

$= \mathbb{P}(a \leq X \leq b)$

only for continuously distributed r.v.s

$= \mathbb{P}(a \leq X < b)$

Let $1 \gg \varepsilon > 0$, then $\mathbb{P}(X\in[a,a+\varepsilon]) = \int_a^{a+\varepsilon} f(r)\,dr$

$\rightsquigarrow f(a) \approx \dfrac{\mathbb{P}(X\in[a,a+\varepsilon])}{\varepsilon} \qquad \approx \varepsilon \cdot f(a)$

$$f(a) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_a^{a+\varepsilon} f(r)\,dr$$

Special Cases/Examples

EX1

$X \sim \text{Uniform}([0,1])$

X has density $f(y) = \begin{cases} 1 & y\in[0,1] \\ 0 & \text{else} \end{cases}$

CDF:

$$\begin{aligned} F(r) &= \mathbb{P}(X\leq r) \\ &= \mathbb{P}(X\in(-\infty, r]) \\ &= \int_{-\infty}^{r} f(y)\,dy \\ &= \begin{cases} 0 & r<0 \\ r & r\in[0,1] \\ 1 & r>1 \end{cases} \end{aligned}$$

Ex 2 $\ell < r$

$X \sim \text{Uniform}([\ell, r])$

if it has density

$$f(n) = \begin{cases} 0 & y \notin [\ell, r] \\ \frac{1}{r-\ell} & y \in [\ell, r] \end{cases}$$

let $\ell < a < b < r$

$$\mathbb{P}(X \in (a,b)) = \int_a^b f(y)\,dy = \frac{b-a}{r-\ell}$$

General fact: if X is continuously distributed, then for any $v \in \mathbb{R}$, $\mathbb{P}(X=v) = 0$

for any ε, $\mathbb{P}(X=v) \le \mathbb{P}(X \in [v, v+\varepsilon])$

$$= \int_v^{v+\varepsilon} f(y)\,dy \quad \text{(if density)}$$

$$= F(v+\varepsilon) - F(v) \quad \text{(if not)}$$

$$\mathbb{P}(X=v) \le \lim_{\varepsilon \to 0} \left(F(v+\varepsilon) - F(v)\right) = 0$$

Prop: if $G : \mathbb{R} \to [0,1]$ is a function satisfying:

1) G is nondecreasing
2) $\lim_{n \to \infty} G(n) = 1$
3) $\lim_{n \to -\infty} G(n) = 0$

Then G is a valid CDF

Additivity?

Recall if X discrete, then

$$\sum_{v \in \mathcal{R}(X)} \mathbb{P}(X=v) = 1$$

countable

But in continuous case,

$$\sum_{v \in \mathbb{R}} \mathbb{P}(X=v) = 0 \ ?$$

uncountable

$$1 = \mathbb{P}(X \in \mathbb{R})$$

$$= \mathbb{P}\left(\bigcup_{v \in \mathbb{R}} \{X=v\}\right)$$

$$= \sum_{v \in \mathbb{R}} \mathbb{P}(X=v) = 0 \ ?$$

Suppose $\{A_\alpha\}_\alpha$ are disjoint events.

$$\mathbb{P}\left(\bigcup_\alpha A_\alpha\right) = \sum_\alpha \mathbb{P}(A_\alpha)$$

iff collection is countable

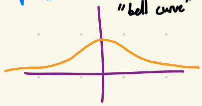
Ex 3 $X \sim Normal(\underset{\text{mean } 0}{0}, \underset{\text{variance } 1}{1})$

"Standard normal"

if $\mathbb{P}(a < X < b) = \int_a^b \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$

"bell curve"

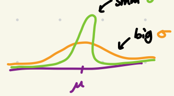
Density: $\frac{e^{-y^2/2}}{\sqrt{2\pi}}$



Ex 4. $X \sim Normal(\mu, \sigma^2)$

mean, variance

if it has density $\frac{e^{-\frac{(y-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$

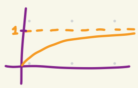


Ex 5 Let $\lambda > 0$. $X \sim Exponential(\lambda)$ → analog of geometric distribution

if $\mathbb{P}(X \geq t) = \begin{cases} e^{-\lambda t} & t \geq 0 \\ 1 & t < 0 \end{cases}$

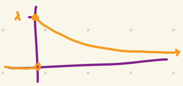
CDF:

$$\mathbb{P}(X < t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$



density:

$$f(t) = F'(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$



$$\int_{\mathbb{R}} f(r)\,dr = \int_0^\infty \lambda e^{-\lambda r}\,dr = 1$$

Suppose $G: \mathbb{R} \to [0,1]$ is any CDF

Define $G^{-1}(t) = \min\{s \in \mathbb{R} \mid G(s) \geq t\}$

"right continuous inverse"

(if G invertible, $G(G^{-1}(t)) = t$)

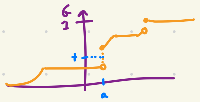
Prop: Let $X \sim Uniform([0,1])$

Then let $Y = G^{-1}(X)$

Then $\mathbb{P}(Y \le t) = G(t)$

i.e., CDF of Y is G.

Picture:



$G^{-1}(t) = a$

$Y = G^{-1}(X)$

Pf: $\mathbb{P}(Y \le a) = \mathbb{P}(G^{-1}(X) \le a)$

Fact: by def of G^{-1},

$G^{-1}(t) \le a$ iff $G(a) \ge t$

if $G^{-1}(t) \le a$

$\Rightarrow \min\{s \mid G(s) \ge t\} \le a$

$\Rightarrow \exists s \le a$ s.t. $G(s) \ge t$

$\Rightarrow G(a) \ge t$ since G nondecreasing

Conversely, if $G(a) \ge t$

then $a \in \{s \mid G(s) \ge t\}$

So $\min\{s \mid G(s) \ge t\} \le a$

$\Rightarrow G^{-1}(t) \le a$

$= \mathbb{P}(G(a) \ge X)$

$= \mathbb{P}(X \in [0, G(a)])$

$= \dfrac{G(a) - 0}{1}$

$= G(a)$

So for any a, $\mathbb{P}(Y \le a) = G(a)$

i.e., Y has CDF G.

Ex: consider $G(s) = \begin{cases} 1 - e^{-\lambda s} & s \ge 0 \\ 0 & s < 0 \end{cases}$

CDF for Exponential(λ)

compute: $G^{-1}(t) = -\frac{1}{\lambda}\ln(1-t)$

So if $X \sim$ Uniform$([0,1])$,

then $Y = -\frac{1}{\lambda}\ln(1-X) \sim$ Exponential(λ)

10/22 Continuously distributed r.v.s

X is continuously distributed if

$t \mapsto F(t) = \mathbb{P}(X \le t) \quad t \in \mathbb{R}$

is continuous in t.

$\Rightarrow \mathbb{P}(X = r) = 0$ for any $r \in \mathbb{R}$

so we ask

$\mathbb{P}(a < X \le b) = F(b) - F(a)$

$= \int_a^b f(r)\,dr$

density

Sps X has a density $f(r)$. Expectation of X is

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} r f(r)\, dr$$

Recall:

$$\int_{-\infty}^{\infty} f(r)\, dr = 1$$

compare:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} r f(r)\, dr$$

to discrete case

$$\sum_{r \in \mathbb{R}(Y)} r\, \mathbb{P}(Y = r)$$

discrete set

Density case: $\mathbb{E}(X)$ is well-defined

if $\int_{-\infty}^{\infty} |r| f(r)\, dr < \infty$

converges absolutely

If $g: \mathbb{R} \to \mathbb{R}$ and X has density f, then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(r) f(r)\, dr$$

recall discrete:

$$\mathbb{E}(g(Y)) = \sum_{r \in \mathbb{R}(Y)} g(r)\, \mathbb{P}(Y = r)$$

Moments:

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} r^k f(r)\, dr$$

$$\mathrm{VAR}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Properties:

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$$

Examples

1) Let $X \sim \text{Exponential}(\lambda)$

density: $f(r) = \begin{cases} 0 & r < 0 \\ \lambda e^{-\lambda r} & r \geq 0 \end{cases}$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} r f(r)\, dr$$

$$= \int_{0}^{\infty} r \lambda e^{-\lambda r}\, dr \qquad u = -\lambda r$$

$$= -\int_{0}^{\infty} r \frac{d}{dr}\left(e^{-\lambda r}\right) dr$$

$$= \int_0^\infty -\frac{d}{dr}\left(re^{-\lambda r}\right) dr + \int_0^\infty \left(\frac{d}{dr} r\right) e^{-\lambda r} dr$$

$$= -re^{-\lambda r}\Big|_0^\infty + \int_0^\infty e^{-\lambda r} dr$$

$$= 0 - 0 + \frac{1}{\lambda}$$

So if $X \sim \text{Exponential}(\lambda)$, then $\mathbb{E}(X) = \frac{1}{\lambda}$

$$Var(X) = \mathbb{E}(X^2) - \underbrace{\mathbb{E}(X)^2}_{1/\lambda^2}$$

$$\mathbb{E}(X^2) = \int_{-\infty}^\infty r^2 f(r)\, dr$$

$$= \int_0^\infty r^2 \lambda e^{-\lambda r} dr$$

$$= -\int_0^\infty r^2 \frac{d}{dr}\left(e^{-\lambda r}\right) dr$$

$$= -\int_0^\infty \frac{d}{dr}\left(r^2 e^{-\lambda r}\right) dr + \int_0^\infty 2r e^{-\lambda r} dr$$

$$= r^2 e^{-\lambda r}\Big|_0^\infty + \frac{2}{\lambda}\int_0^\infty \lambda r e^{-\lambda r} dr$$

$$= 0 - 0 + \frac{2}{\lambda^2}$$

$$VAR(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$= 2/\lambda^2 - 1/\lambda^2 = \frac{1}{\lambda^2}$$

"Memoryless property"

$X \sim Exp(\lambda)$

$\mathbb{P}(X > t+s \mid X > t)$

$= \mathbb{P}(X > s) \quad \forall t,s > 0$

$$\mathbb{P}(X > t+s \mid X > t) = \mathbb{P}(\{X > t+s\} \cap \{X > t\}) \Big/ \mathbb{P}(X > t)$$

$$= \mathbb{P}(X > t+s) \Big/ \mathbb{P}(X > t)$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}}$$

$$= e^{-\lambda s} = \mathbb{P}(X > s)$$

Ex: Let $X \sim \text{Normal}(\mu, \sigma^2)$

density: $f(x) = \dfrac{e^{-\frac{(x-\mu)^2}{2\sigma}}}{\sqrt{2\pi\sigma}}$

Fact: Let $Y = \dfrac{X-\mu}{\sigma}$

then $Y \sim \text{Normal}(0,1)$

Thus $X = \sigma Y + \mu$

$$\Rightarrow \mathbb{E}(X) = \mathbb{E}(\sigma Y + \mu) = \sigma \cancel{\mathbb{E}(Y)}^{0} + \mu = \mu$$

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \underbrace{\frac{e^{-y^2/2}}{\sqrt{2\pi}}}\, dy = 0$$ odd symmetry about 0.

$$\mathrm{VAR}(X) = \mathrm{VAR}(\sigma Y + \mu) = \sigma^2 \mathrm{VAR}(Y)$$

$$\mathrm{VAR}(Y) = \mathbb{E}(Y^2) - \cancel{\mathbb{E}(Y)^2}^{0} = \int_{-\infty}^{\infty} y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy = 1 \quad \text{(trust!)}$$

So if $X \sim N(\mu, \sigma^2)$

then $\mathbb{E}(X) = \mu$

$\mathrm{VAR}(X) = \sigma^2$.

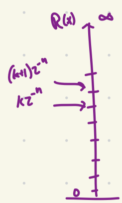
Suppose X is a random variable

$\mathbb{P}(X \geq 0) = 1$ fix $n \geq 1$

define two approximating r.v.s, Y_n, Z_n

$Y_n(\omega) = k2^{-n}$ if $k2^{-n} \leq X(\omega) \leq (k+1)2^{-n}$

$Z_n(\omega) = (k+1)2^{-n}$ if " " " " "



So by definition:

$Y_n(\omega) \leq X \leq Z_n(\omega)$

for any ω

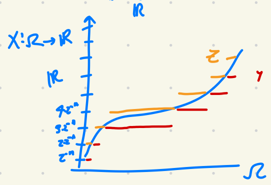
discrete

might define

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(Y_n) \overset{\text{claim}}{=} \lim_{n\to\infty} \mathbb{E}(Z_n)$$

Claim: limit exists + agrees w/ formula

$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\, dx$ when X has density f



Notice: Range of Y, Z is $\{0, 1\cdot 2^{-n}, 2\cdot 2^{-n}, \dots\}$

so $\mathbb{E}(Y_n) = \sum_{k=0}^{\infty} k 2^{-n} \mathbb{P}(k 2^{-n} \le X < (k+1) 2^{-n})$

$\mathbb{E}(Z_n) = \sum_{k=0}^{\infty} k 2^{-n} \mathbb{P}(k 2^{-n} \le X < (k+1) 2^{-n})$

$$\mathbb{E}(Z_n) - \mathbb{E}(Y_n) = \sum_{k=0}^{\infty} ((k+1)2^{-n} - k2^{-n}) \mathbb{P}(\dots)$$
$$= \sum_{k=0}^{\infty} 2^{-n} \mathbb{P}(\dots)$$
$$= 2^{-n} \sum_{k=0}^{\infty} \mathbb{P}(k 2^{-n} \le X < (k+1) 2^{-n})$$
$$= 2^{-n}$$

As $n \to \infty$

$\lim_{n\to\infty} \mathbb{E}(Z_n) - \mathbb{E}(Y_n)$

$= \lim_{n\to\infty} 2^{-n} = 0$

$\mathbb{E}(Y_n)$ increasing

$\mathbb{E}(Z_n)$ decreasing

$\Rightarrow$ limit $\mathbb{E}(X)$ exists

$\lim_{n\to\infty} \mathbb{E}(Y_n) = \lim_{n\to\infty} \mathbb{E}(Z_n) = \mathbb{E}(X)$

Same strategy: derive tail sum

$$\mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X > t)\,dt$$

(above becomes riemann sums)

For general r.v., write

$X = X_+ - X_-$ $\quad X_-, X_+ \geq 0$

X_+: positive part, X_-: negative part

$$\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-)$$

10/24

Tail Sum formula

Let $X \geq 0$ be any r.v. then

$$\mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X > t)\,dt$$

fix $n = 1, 2, \dots$

Define $Y_n = k2^{-n}$ if $k2^{-n} \le X \le (k+1)2^{-n}$

$Z_n = (k+1)2^{-n}$ " " " "

For $k = 0, 1, 2, \dots$

$\Rightarrow Y_n(\omega) \le X(\omega) \le Z_n(\omega)$

then $\mathbb{E}(X) = \lim\limits_{n\to\infty} \mathbb{E}(Y_n) = \lim\limits_{n\to\infty} \mathbb{E}(Z_n)$

range of Y_n, Z_n is

$\{k2^{-n} \mid k=0,1,2,\dots\}$

range of 2^nY_n and 2^nZ_n

is $\{k \mid k=0,1,2,\dots\}$ tail sum

So $\mathbb{E}(2^nY_n) = \sum_{j=0}^{\infty} \mathbb{P}(2^nY_n > j)$

$\Rightarrow 2^n\mathbb{E}(Y_n) = \sum_{j=0}^{\infty} \mathbb{P}(2^nY_n > j)\,2^{-n}$

Since $Y_n \le X \le Z_n$

$\mathbb{P}(Y_n > a) \le \mathbb{P}(X > a) \le \mathbb{P}(Z_n > a)$

So $\mathbb{E}(Y_n) = \sum_{j=0}^{\infty} \mathbb{P}(Y_n > j2^{-n})\,2^{-n}$

$\le \sum_{j=0}^{\infty} \mathbb{P}(X > j2^{-n})\,2^{-n}$

$\le \mathbb{E}(Z_n)$

both converge to $\mathbb{E}(X)$ as $n\to\infty$

So $\mathbb{E}(X) = \lim\limits_{n\to\infty} \sum_{j=0}^{\infty} \mathbb{P}(X > j2^{-n})\,2^{-n}$

$= \int_0^\infty \mathbb{P}(X > t)\,dt$

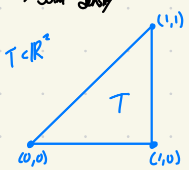
If X has density $f(x)$, then

$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\,dx$ assuming

the integral converges absolutely.

Two or more continuously distributed r.v.'s

$\to$ Joint density



$T = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ 0 \le y \le x\}$

two ways of picking a random point:

1) Let (X_1, Y_1) be uniformly distributed on T

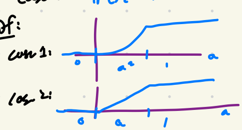
2) Let $X_2 \sim \text{Uniform}([0,1])$ then given $X_2 = r$,
let $Y_2 \sim \text{Uniform}([0,r])$

Let $a \in (0,1)$. $\mathbb{P}(X \le a)$?

Case 1: $\dfrac{\frac{a^2}{2}}{\frac{1^2}{2}} = a^2$

Case 2: $\mathbb{P}(X_2 \le a) = a$

CDF:



Densities:

case 1: $f(x) = \begin{cases} 0 & x \notin (0,1) \\ 2x & x \in (0,1) \end{cases}$

case 2: $f(x) = \begin{cases} 0 & x \notin (0,1) \\ 1 & x \in (0,1) \end{cases}$

Def A pair of real-valued r.v.'s (X,Y) has a joint density $f(x,y)$ if $\mathbb{P}((X,Y) \in B) = \iint_B f(x,y)\,dx\,dy$ for any open $B \subset \mathbb{R}^2$

in particular, $\Rightarrow \iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$

Back to triangle example:

case 1: $(X,Y) \sim \text{Uniform}(T)$

$\mathbb{P}((X,Y) \in B) = \frac{|B|}{|T|}$

if $B \subseteq T$

(X,Y) have joint density

$f(x,y) = \begin{cases} 0 & (x,y) \notin T \\ c & (x,y) \in T \end{cases}$

$\iint_{\mathbb{R}} f\,dx\,dy = \iint_T c\,dx\,dy = c|T| \Rightarrow c = \frac{1}{|T|} = 2$

Case 2:

recall discrete joint distributions

$$\mathbb{P}(X=k, Y=j) = \underbrace{\mathbb{P}(X=k \mid Y=j)}_{\text{Conditional distribution}}\mathbb{P}(Y=j)$$

Continuous case:

If (X,Y) have a joint density $f(x,y)$ then the conditional density for X given $Y=y$ is

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

$\leftarrow$ joint density

marginal density of Y $\nearrow$

Similarly,

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}$$

$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$

$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy$

are marginal densities of Y, X, respectively

Notice: $f(x,y) = f_{X|Y}(x|y) f_Y(y)$

Similarly: $f(x,y) = f_{Y|X}(y|x) f_X(x)$

Return to triangle (case 2):

Let $X \sim \text{Uniform}([0,1])$

marginal density is $f_X(x) = \begin{cases} 0 & x \notin (0,1) \\ 1 & x \in (0,1) \end{cases}$

Given $X=x$, $Y \sim \text{Uniform}((0,x))$

$$f_{Y|X}(y|x) = \begin{cases} 0 & y \notin (0,x) \\ 1/x & y \in (0,x) \end{cases}$$

conditional density

$$\Rightarrow f(x,y) = f_{Y|X}(y|x) f_X(x)$$

$$= \begin{cases} 0 & (x,y) \notin T \\ 1/x & (x,y) \in T \end{cases}$$

10/24 - got notes from gc

10/31

CLT

Standing assumption:

$X_1, X_2, X_3, \dots$

are iid (independent, identically distributed)

Let $\mu = E(X_i)$ $\sigma^2 = VAR(X_i)$

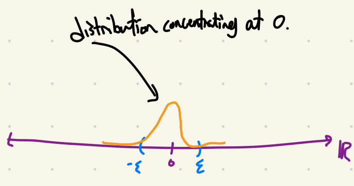
Law of Large #'s:

$$E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$$

for $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\left(\frac{1}{n}\sum X_i\right) - \mu\right| < \varepsilon\right) = 1$$

$$= \mathbb{P}\Big(-\varepsilon < \underbrace{\frac{X_1 + \dots + X_n - \mu n}{n}}_{A_n = \text{shifted average}} < \varepsilon\Big) \to 1$$



Consider:

$$\underset{\text{growing with } n}{\frac{\sqrt{n}}{\sigma}} \cdot \underset{\text{typically small}}{\frac{(X_1 + \dots + X_n - n\mu)}{n}} = \underbrace{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}}_{S_n}$$

$$VAR(S_n) = VAR\left(\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right) = \frac{1}{\sigma^2 n} VAR(X_1 + \dots + X_n)$$

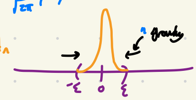
$$= \frac{n\sigma^2}{\sigma^2 n} = 1$$

$$\lim_{n\to\infty} \mathbb{P}\Big(a \le \underbrace{\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}}_{S_n} \le b\Big)$$

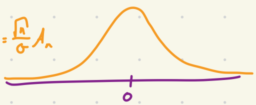
$$= \Phi(b) - \Phi(a) \qquad \Phi(b) = \int_{-\infty}^{b} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$$

$$= \int_a^b \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$$

LLN: A_n



CLT: $S_n = \frac{\sqrt{n}}{\sigma} A_n$



Proof: Note: redefine $X_k' = \frac{X_k - \mu}{\sigma}$

So may assume $\mu = 0$, $\sigma = 1$

Doesn't change A_n.

Lets prove case $a = -\infty$:

$$\underbrace{\mathbb{P}\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \le b\Big)}_{F_n(b) \to \Phi(b)} \xrightarrow{n\to\infty} \Phi(b)$$

CDF

Thm: Assume $\mu = 0$, $\sigma = 1$, $\mathbb{E}(|X_i|^3) < \infty$

Then

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \le b\Big) = \Phi(b)$$

Let $f_b(x) = \begin{cases} 1 & x \le b \\ 0 & x > b \end{cases}$

$f_b(x) = \mathbb{1}_{(-\infty, b]}(x)$

$$\mathbb{P}\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \le b\Big) = \mathbb{E}\Big(f_b\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big)$$

$$= \mathbb{E}\Big(\mathbb{1}_{(-\infty, b]}\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big)$$

CLT: for any b: $\displaystyle\lim_{n\to\infty} \mathbb{E}\Big(f_b\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big) = \Phi(b)$

Fact: let $Y_1, Y_2, Y_3, \ldots$ be indep

$Y_i \sim N(0,1)$

Fact: $\frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \sim N(0,1)$ standard gaussian

$\mathbb{P}\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}} \le b\right) = \Phi(b)$

$= \mathbb{E}\left(f_b\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)\right)$

So CLT equivalent to: might as well replace the X's with gaussians Y

$$\lim_{n\to\infty} \mathbb{E}\left(f_b\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right) - \mathbb{E}\left(f_b\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)\right) = 0$$

Special case of Lindenberg-Lévy CLT

Maybe this holds for "any" f?

Note: f is discontinuous

Let $f^l(x) \le f(x) \le f^u(x)$ $\qquad$ f^l, f^u thrice-diff'able

f^l, f^u agree with f outside of ε-radius

Let $S_n = \frac{X_1+\cdots+X_n}{\sqrt{n}}$

$G_n = \frac{Y_1+\cdots+Y_n}{\sqrt{n}}$

$$\mathbb{E}(f^l(S_n)) \le \mathbb{E}(f_b(S_n)) \le \mathbb{E}(f^u(S_n))$$

Claim: $\mathbb{E}(f^l(S_n)) - \mathbb{E}(f^l(G_n)) \xrightarrow{n\to\infty} 0$

$\mathbb{E}(f^u(S_n)) - \mathbb{E}(f^u(G_n)) \xrightarrow{n\to\infty} 0$

Since f^l, f^u arbitrarily close to f_b, this implies

$\mathbb{E}(f_b(S_n)) - \mathbb{E}(f_b(G_n)) \to 0$

So CLT reduces to thinking about

$\mathbb{E}(\hat{f}(S_n)) - \mathbb{E}(\hat{f}(G_n))$ ← $G_n \sim N(0,1)$

for some function (either $\hat{f} = f^l$ or $\hat{f} = f^u$)

$\hat{f}$ continuous.

Idea: consider effect of swapping X_k and Y_k ← $\sim N(0,1)$

Define $Q_{n,k} = \frac{X_1+\cdots+X_k+\overbrace{Y_{k+1}+\cdots+Y_n}^{\text{swap for Gaussians}}}{\sqrt{n}}$

$Q_{n,0} = G_n \qquad Q_{n,n} = S_n$

So can write: $\mathbb{E}(\hat{f}(S_n)) - \mathbb{E}(\hat{f}(G_n))$

$$= \mathbb{E}(\hat{f}(S_n) - \hat{f}(G_n))$$
$$= \mathbb{E}(\hat{f}(Q_{n,n}) - \hat{f}(Q_{n,0}))$$
$$= \mathbb{E}\left(\sum_{k=0}^{n-1} \hat{f}(Q_{n,k+1}) - \hat{f}(Q_{n,k})\right)$$
$$= \sum_{k=0}^{n-1} \mathbb{E}(\hat{f}(Q_{n,k+1}) - \hat{f}(Q_{n,k}))$$

$$Q_{n,k} = \frac{X_1 + \dots + X_k + Y_{k+1} + \dots + Y_n}{\sqrt{n}}$$
$$Q_{n,k+1} = \frac{X_1 + \dots + X_{k+1} + X_{k+1} + Y_{k+2} + \dots + Y_n}{\sqrt{n}}$$

<u>Def</u>: $U_k = \dfrac{X_1 + \dots + X_k + \overset{\text{leaves out term of contention}}{Y_{k+2}} + \dots + Y_n}{\sqrt{n}}$

So
$$Q_{n,k+1} = U_k + \frac{X_{k+1}}{\sqrt{n}}$$
$$Q_{n,k} = U_k + \frac{Y_{k+1}}{\sqrt{n}}$$

So $\hat{f}(Q_{n,k+1}) - \hat{f}(Q_{n,k})$
$$= \hat{f}\left(U_k + \frac{X_{k+1}}{\sqrt{n}}\right) - \hat{f}\left(U_k + \frac{Y_{k+1}}{\sqrt{n}}\right)$$

now estimate this difference $\longrightarrow$ calculus

<u>Taylor's thm</u>:

Suppose $\hat{f}$ is 3 times diff'ble. Then
$$\hat{f}(x+h_1) - \hat{f}(x+h_2)$$
$$= (h_1 - h_2)\hat{f}'(x) + (h_1^2 - h_2^2)\frac{\hat{f}''(x)}{2} + R$$
where $|R| \le C\left(\min(h_1^2, |h_1|^3) + \min(h_2^2, |h_2|^3)\right)$

$$\hat{f}\left(U_k + \frac{X_{k+1}}{\sqrt{n}}\right) - \hat{f}\left(U_k + \frac{Y_{k+1}}{\sqrt{n}}\right)$$

$$= \left(\frac{X_{k+1} - Y_{k+1}}{\sqrt{n}}\right)\hat{f}'(U_k) + \left[\left(\frac{X_{k+1}}{\sqrt{n}}\right)^2 - \left(\frac{Y_{k+1}}{\sqrt{n}}\right)^2\right]\hat{f}''(U_k) + R$$

$$\mathbb{E}(\hat{f}(Q_{n,k+1}) - \hat{f}(Q_{n,k}))$$
$$= \mathbb{E}\left(\left(\frac{X_{k+1} - Y_{k+1}}{\sqrt{n}}\right)\hat{f}(U_k)\right) + \mathbb{E}\left(\left(\left(\frac{X_{k+1}}{\sqrt{n}}\right)^2 - \left(\frac{Y_{k+1}}{\sqrt{n}}\right)^2\right)\hat{f}''(U_k)\right) + \mathbb{E}(R)$$
$$= \mathbb{E}\left(\frac{\overset{\text{same mean + variance}}{X_{k+1} - Y_{k+1}}}{\sqrt{n}}\right)\mathbb{E}(\hat{f}(U_k))$$
$$= \mathbb{E}(R)!$$

$$|\mathbb{E}(R)| \le C\,\mathbb{E}\min\left(\left(\frac{X_{k+1}}{\sqrt{n}}\right)^2, \left(\frac{|X_{k+1}|}{\sqrt{n}}\right)^3\right) + C\,\mathbb{E}\min\left(\left(\frac{Y_{k+1}}{\sqrt{n}}\right)^2, \left(\frac{|Y_{k+1}|}{\sqrt{n}}\right)^3\right)$$

So $\left|\mathbb{E}\left(\hat{f}(S_n)-\hat{f}(G_n)\right)\right| \le \sum_{k=0}^{n-1} \left(\mathbb{E}(R_n)\right) \le \sum_{k=0}^{n-1} \frac{C}{n^{3/2}} \le \frac{C}{\sqrt{n}} \to 0$

11/05

Recall CLT:

Thm: Let $X_1, X_2, X_3, \dots$ be iid. with $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = \mathrm{VAR}(X_i) < \infty$.

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{X_1+\dots+X_n - n\mu}{\sqrt{n\sigma^2}} \le b\right) = \Phi(b)-\Phi(a) = \int_a^b \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$$

note: $\sqrt{n\sigma^2}$ is the standard deviation of sum $Z_n = (X_1+\dots+X_n)$

$$a \le \frac{Z_n - n\mu}{\sqrt{n\sigma^2}} \le b$$

$\iff n\mu + a\sqrt{n\sigma^2} \le Z_n \le n\mu + b\sqrt{n\sigma^2}$ all equivalent

$\iff a\sqrt{n\sigma^2} \le Z - n\mu \le b\sqrt{n\sigma^2}$

In particular, if $a = -k$, $b = k$ for some $k \in \{1,2,3,\dots\}$

$$-k\sqrt{n\sigma^2} \le Z_n - \mu n \le k\sqrt{n\sigma^2}$$

= event that "Z_n is within k standard deviations of its mean"

"when k is large, 3, say"

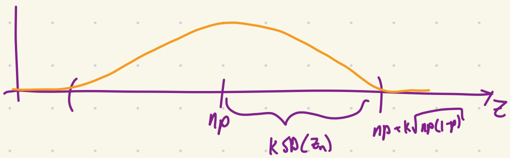
Consider scaling of distribution of Z_n.

Imagine $Z_n = \sum_{k=1}^{n} \mathbb{1}_{H_k}$

H_k = event that kth toss of a p-coin is heads

$\mathbb{E}(Z_n) = np$ $\quad \mathrm{VAR}(Z_n) = np(1-p)$

$\mathrm{SD}(Z_n) = \sqrt{np(1-p)}$



We know $Z_n \sim \mathrm{Binomial}(n,p)$

$$\mathbb{P}(Z_n = k) = \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{\text{complicated w.r.t. } k}$$

CLT:

$$\mathbb{P}\left(np - k\sqrt{np(1-p)} \le Z_n \le np + k\sqrt{np(1-p)}\right)$$

grows order $\sqrt{n}$

LLN:

$$\mathbb{P}(\mu n - n\varepsilon \le Z_n \le \mu n + n\varepsilon) \xrightarrow{n\to\infty} 1$$

grows order n

Poisson Approximation:

Recall for $Z_n \sim \text{Binomial}(n,p) = \text{Binomial}(n, \frac{\lambda}{n})$

with $p = \frac{\lambda}{n}$ $\quad \mathbb{E}(Z_n) = pn = \lambda$

(λ fixed)

big difference b/w this and CLT: distribution parameter changes with n

$$\lim_{n\to\infty} \mathbb{P}(Z_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Recall Gaussian thingy

$$\Phi(b) = \int_{-\infty}^{b} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$$

1.5 SD's above mean

$$\Phi(1.5) = 0.8531$$

$$\Phi(3) = 0.9987$$

Consider fair coin: $n=100$ tosses

$p = \frac{1}{2}$

$\text{VAR}(Z_n) = np(1-p) = \frac{n}{4}$

$\text{SD}(Z_n) = \frac{\sqrt{n}}{2}$

for $n=100$

$\text{SD}(Z_{100}) = \frac{\sqrt{100}}{2} = 5$

$\mathbb{E}(Z_{100}) = \frac{100}{2} = 50$

What is probability $Z_{100} \in (45, 60)$

$= \mathbb{P}(-1 \cdot \text{SD}(Z_{100}) \le Z_{100} \le 2 \cdot \text{SD}(Z_{100}))$

$\approx \Phi(2) - \Phi(1)$

error is order $\frac{1}{\sqrt{n}}$

$= \int_{-1}^{2} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$

Ex toss n fair coins, assume n even.

$\mathbb{E}(Z_n) = \frac{n}{2}$.

What is $\mathbb{P}(Z_n = \frac{n}{2})$

$k = \frac{n}{2}$

$= \binom{n}{n/2}\left(\frac{1}{2}\right)^{n/2}\left(1-\frac{1}{2}\right)^{n-n/2}$

$= \binom{n}{n/2}\left(\frac{1}{2}\right)^{n}$

$Z_n \in \{0, 1, 2, \dots, n\}$

$= \mathbb{P}(Z_n \in (\frac{n}{2} - \frac{1}{2}, \frac{n}{2} + \frac{1}{2}))$

mean

By CLT: $= \mathbb{P}(-\frac{1}{2} < Z_n - \frac{n}{2} < \frac{1}{2})$

$VAR(Z_n) = \frac{n}{4}$ $SD(Z_n) = \frac{\sqrt{n}}{2}$

$k\frac{\sqrt{n}}{2} = \frac{1}{2} \Rightarrow k = \frac{1}{\sqrt{n}}$

So

$$= \mathbb{P}\left(-\frac{1}{\sqrt{n}} SD(Z_n) \le Z_n - \mathbb{E}(Z_n) \le \frac{1}{\sqrt{n}} SD(Z_n)\right)$$

$$\approx \int_{-\frac{1}{\sqrt{n}}}^{\frac{1}{\sqrt{n}}} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$

CLT normal approximation

$$\approx \frac{2}{\sqrt{n}} \frac{1}{\sqrt{2\pi}} = \sqrt{\frac{2}{n\pi}}$$

Imagine a box with gold/silver marbles

We don't know fraction of gold.

Study the box by repeated independent sampling w/ replacement

make n observations. sps fraction of gold observed in draws

is $\alpha \in [0,1]$

So # gold drawn is $X = \alpha n$

What is best guess at population fraction of gold in box?

Idea: use Bayes' Thm

Imagine fraction of gold in box is random

Y = fraction of gold in box

Given $Y = p$ the conditional distribution of X is

$X \sim \text{Binomial}(n, p)$

Discrete case:

Assume M marbles in box, so

$Y \in \{0, \frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}, 1\}$

Assume $Y \sim \text{Unif}\{0, \frac{1}{M}, \dots\}$

Ask: given $X = \alpha n$ (observation)

$\mathbb{P}(Y = p \mid X = \alpha n) = ?$

$$\mathbb{P}(Y = p \mid X = \alpha n) = \frac{\mathbb{P}(X = \alpha n \mid Y = p)\,\mathbb{P}(Y = p)}{\mathbb{P}(X = \alpha n)}$$

$$= \frac{\binom{n}{\alpha n} p^{\alpha n} (1-p)^{n - \alpha n}\, \mathbb{P}(Y = p)}{\mathbb{P}(X = \alpha n)} = \text{"Posterior distribution of } Y \text{ given } X = \alpha n\text{"}$$

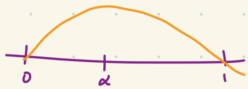
Assume $Y \sim \text{Unif}$

$$\mathbb{P}(Y=p) = c = \frac{1}{m+1}$$

As function of p, posterior has the form

$$p \mapsto \mathbb{P}(Y=p \mid X = \alpha n)$$

$$g(p) = C_{\alpha,n}\, p^{\alpha n}(1-p)^{n-\alpha n} = \text{polynomial in } p$$



max occurs at α

$$C_{\alpha,n}\, p^{\alpha n}(1-p)^{n-\alpha n} = C_{\alpha,n}\left(p^{\alpha}(1-p)^{1-\alpha}\right)^{n}$$

Idea: treat p as r.v., sample population as n increases, will center around true value of p

---

11/07 Midterm 2 moved to tuesday, 11/19

Strong Law of Large Numbers

Standing assumption: $X_1, X_2, X_3, \dots$ be a sequence of i.i.d. r.v. with $\mathbb{E}(X_i) = \mu < \infty$

Weak LLN: for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\underbrace{\frac{X_1+\cdots+X_n}{n}}_{\text{measured average}} \underbrace{-\mu\right| \geq \varepsilon}_{\text{close to mean}}\right) = 0$$

Strong LLN

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right) = 1$$

* Random variables are functions on Ω

$$X_i(\omega): \Omega \to \mathbb{R}$$

define: $Y_n(\omega) = \dfrac{X_1(\omega)+\cdots+X_n(\omega)}{n}$

$\{Y_n(\omega)\}_{n\geq 1}$ is a sequence of functions on Ω.

Both versions say something about convergence of

$$Y_n \longrightarrow \underset{\text{(constant function)}}{\mu}$$

Difference in notion of convergence.

Summarize:

WLLN $\leftrightarrow$ $Y_n \xrightarrow{n\to\infty} \mu$ $\underbrace{\text{"in probability"}}^{\text{this is just convergence in measure}}$ $\rightarrow$ ε - bad set $B_{n,\varepsilon} = \{\omega \in \Omega \mid |Y_n(\omega) - \mu| \geq \varepsilon\}$

has $\lim_{n\to\infty} |B_{n,\varepsilon}| = 0$ for any ε

$\lim_{n\to\infty} \mathbb{P}(B_{n,\varepsilon}) = 0$

SLLN $\leftrightarrow$ $Y_n \xrightarrow{n\to\infty} \mu$ for "a.e. Ω"

Claim: SLLN equivalent to:

for any $\varepsilon > 0$,

$$\mathbb{P}\left(\bigcap_{k\geq 1} \bigcup_{n\geq k} B_{n,\varepsilon}\right) = 0$$

$$\Leftrightarrow \lim_{k\to\infty} \mathbb{P}\left(\bigcup_{n\geq k} B_{n,\varepsilon}\right) = 0$$

WLLN is just $\lim_{n\to\infty}(B_{n,\varepsilon}) = 0$

so clearly follows from this.

Consider the event $\bigcap_{k\geq 1} \bigcup_{n\geq k} B_{n,\varepsilon}$

what does it mean for

$Y_n(\omega) \not\to \mu$

for some ε,

$|Y_n(\omega) - \mu| \geq \varepsilon$ for infinitely many of the n's

$\Rightarrow \omega \in \bigcup_{n\geq k} B_{n,\varepsilon}$ for all k

$$\Rightarrow \underbrace{\omega \in \bigcap_{k\geq 1} \bigcup_{n\geq k} B_{n,\varepsilon}}_{\text{set that occur infinitely often}}$$

i.e. $B_{n,\varepsilon}$ occurs infinitely often

Pf (SLLN):

for $B_{n,\varepsilon} = \{\omega \in \Omega \mid |Y_n(\omega) - \mu| \geq \varepsilon\}$

for any $\varepsilon > 0$,

WLLN: $\lim_{n\to\infty} \mathbb{P}(B_{n,\varepsilon}) = 0$

SLLN: $\mathbb{P}\left(\bigcap_{k\geq 1} \bigcup_{n\geq k} B_{n,\varepsilon}\right) = 0$

Borel-Cantelli Lemma:

Let $\{H_n\}_{n=0}^{\infty}$ be any sequence of events.

Suppose $\sum_{n=1}^{\infty} \mathbb{P}(H_n) < \infty$.

Then $\mathbb{P}\left(\underbrace{\bigcap_{k\geq 1} \bigcup_{n\geq k} H_n}_{\text{"}H_n\text{ infinitely often"}}\right) = 0$

Pf: for any $m \geq 1$,

$$\mathbb{P}\left(\bigcap_{k\geq 1} \bigcup_{n\geq k} H_n\right) \leq \mathbb{P}\left(\bigcup_{n\geq m} H_n\right)$$
$$\leq \underbrace{\sum_{n\geq m} \mathbb{P}(H_n)}$$

if $m \to \infty$, this goes to 0

doesn't depend on m, so we let $m \to \infty$

So remains to show $\sum_{n=1}^{\infty} \mathbb{P}(B_{n,\varepsilon}) < \infty$, we are done (by Borel-Cantelli).

11/12

1) finished proof of SLLN

2) Poisson Arrival Process

$W(t)$ random process $\to$ # of arrivals up to time t

$t \mapsto N(t, w)$ random function of time

(waiting time

Def: Let $W_1, W_2, \dots$ indep with $W_i \sim Exp(\lambda)$ for fixed $\lambda > 0$

Then let $A_n = \sum_{k=1}^{n} W_k$ = time of nth arrival

W_k = time b/w $(k-1)^{th}$ and k^{th} arrival.

Then $N(t) = \max\{n \geq 0 \mid A_n \leq t\}$

↳ # of arrivals that have occurred until time t.

$=$ # arrivals in $[0, t]$

$W(t) = \max\{n \mid \underbrace{\sum_{k=1}^{n} W_k} \leq t\}$

$A_n \to$ time of arrival is within t, max of arrival #

Def 2: $N(t)$ satisfies

1) $N(0) = 0$

2) for any $0 \leq s < t$

$N(t) - N(s) \sim Poisson(\lambda(t-s))$

# of arrivals in $(s, t]$

3) for any $0 \leq t_0 < t_1 < t_2 < \dots < t_n$

then the r.v.s $\{N(t_{k+1}) - N(t_k)\}_{k=0}^{n-1}$ are independent

Thm: these are equivalent def's

11/14

Joint Distribution (Density)

- Conditional Distribution (Density)
- Marginal Distribution (Density)

Discrete Case

X, Y

Joint distribution

$\mathbb{P}(X=k, Y=j)$

Marginal distribution

$\mathbb{P}(X=k) = \sum_{j \in R(Y)} \mathbb{P}(X=k, Y=j)$

Conditional distribution: given $Y=j$

$k \mapsto \mathbb{P}(X=k \mid Y=j)$

$= \dfrac{\mathbb{P}(X=k, Y=j)}{\mathbb{P}(Y=j)}$

$= \dfrac{\mathbb{P}(X=k, Y=j)}{\sum_{r \in R(X)} \mathbb{P}(X=r, Y=j)}$

Density Case

X, Y have joint density $f(x,y)$

$\mathbb{P}((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$

Marginal density

$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy$

Conditional density

$f_{X|Y}(x|y) = \dfrac{f(x,y)}{\int_{\mathbb{R}} f(r,y)\,dr} = \dfrac{f(x,y)}{f_Y(y)}$

Ex: Sps X, Y have joint density

$$f(x,y) = \begin{cases} ce^{-\lambda y} & 0<x<y<\infty \\ 0 & \text{else} \end{cases}$$

$$c = \frac{1}{\iint_{0<x<y<\infty} e^{-\lambda y}\,dx\,dy}$$

Marginal distribution of Y? of X?

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dy = \int_0^y ce^{-\lambda y}\,dx = \begin{cases} yce^{-\lambda y} & y > 0 \\ 0 & y < 0 \end{cases}$$

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \int_x^\infty ce^{-\lambda y}\,dy = \begin{cases} \frac{c}{\lambda}e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

Sps $X \sim Exp(\lambda)$ X, Z indepdt.
$Z \sim Exp(\lambda)$

Consider $Y = X + Z$ then (X, Y) has joint density as above.

2) Example w/ joint density

Sps X, Y are arrival times of two students at the library

$X \sim Uniform[0,1]$ X, Y indep.
$Y \sim Uniform[0,1]$

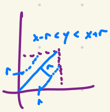
Let W be the time that one student waits for
the distribution of $W = |X - Y|$

independent $\longrightarrow$ joint density product of marginals

$$f(x,y) = \begin{cases} 1 & (x,y) \in [0,1]^2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{P}(W < r)$$
$$= \mathbb{P}(|X - Y| < r)$$
$$= \iint f(x,y)\,dx\,dy$$

$\{(x,y) \mid |x-y| < r\} = H$



$= \text{area of this } \frac{|H|}{|[0,1]^2|} = |H|$

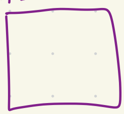
$= 1 - (1-r)^2$
$= 1 - (1 - 2r + r^2)$
$= 2r - r^2$
$= r(2-r)$

$$\mathbb{P}(W<r) = \begin{cases} 0 & r \le 0 \\ 1-(1-r)^2 & r \in (0,1) \\ 1 & r \ge 1 \end{cases}$$

density $\quad f(w) = \frac{d}{dr}\mathbb{P}(W<r) = \begin{cases} 2-2r & r \in (0,1) \\ 0 & \text{otherwise} \end{cases}$

<u>Borel-Cantelli:</u>

$[0,1]^2$



Pieces of tape (2-dim)

areas $|A_k|$ = area covered by kth piece

A_k = pts covered by kth piece

$\{A_k\}_{k=1}^{\infty}$ let W = set of pts covered by infinitely many pieces

Sps we know $\lim_{k\to\infty} |A_k| = 0$

Q: find condition on $\nearrow$ s.t. $|W|$ must be 0

<u>Note:</u> $\quad x \in W$ iff for any $k > 0$

$$x \in \bigcup_{n \ge k} A_n$$

$$\iff x \in \bigcap_{k>0} \bigcup_{n\ge k} A_n$$

<u>Borel-Cantelli:</u> If $\sum_{k=1}^{\infty} |A_w| < \infty$

then $|W| = 0$

$$|W| \le \left|\bigcup_{n\ge k} A_n\right|$$
$$\le \sum_{n \ge k} |A_k|$$

Sps $X_1, X_2, X_3, \dots$ i.i.d.

$X_i \sim Exp(\lambda)$

$R_n = \max(X_1, X_2, \dots, X_n)$

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{R_n}{n} = 0\right) = ?$$

fix ε. $\quad \mathbb{P}\left(\frac{R_n}{n} > \varepsilon\right)$

$= \mathbb{P}(R_n > n\varepsilon)$

$= 1 - \mathbb{P}(R_n \le n\varepsilon)$

$= 1 - \mathbb{P}(\max(X_1, \dots, X_n) \le n\varepsilon) = 1 - \mathbb{P}(X_i \le n\varepsilon)^n \xrightarrow{n\to\infty} 0$

$= 1 - (1 - e^{-\lambda n \varepsilon})^n$

$\mathbb{P}\left(\frac{R_n}{n} > \varepsilon \text{ for infinitely many } n \right)$

if $\sum_{n=1}^{\infty} a_n < \infty$ then by Borel-Cantelli, $= 0$

$\Rightarrow \mathbb{P}\left(\lim_{n\to\infty} \frac{R_n}{n} = 0 \right) = 1$

Sea turtle lays eggs. # of eggs is $N \sim \text{Poisson}(\lambda)$

$\mathbb{E}(N) = \lambda$

Each egg hatches with probability $p \in (0,1)$ independent of each other

X = # of eggs that hatch

Conditional distribution of X:

given $N=k$, is $\text{Binomial}(k,p)$

$\mathbb{E}(X \mid N=k) = kp$

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{E}(X \mid N=k)\mathbb{P}(N=k)$$
$$= \sum_{k=1}^{\infty} kp e^{-\lambda} \frac{\lambda^k}{k!}$$
$$= p \sum_{k=1}^{\infty} k \mathbb{P}(N=k)$$
$$= p\mathbb{E}(N) = p\lambda$$

$$\mathbb{E}(X^2) = \sum_{k=0}^{\infty} \mathbb{E}(X^2 \mid N=k)\mathbb{P}(N=k)$$
$$= \sum_{k=0}^{\infty} \left[kp(1-p) + (kp)^2 \right] \mathbb{P}(N=k)$$
$$= p(1-p)\,\mathbb{E}(N) + p^2\,\mathbb{E}(N^2)$$

11/21

<u>Markov Chain</u> — discrete

$$\{X_n(\omega)\}_{n=0}^{\infty}$$

$n = 0, 1, \ldots$ "time"

X_n = state of system at time n

$X_n = X_n(\omega)$ r.v., $X_n \in S$ "state space"

S is a countable set, often we assume finite

$X_n : \Omega \to S$

$n \mapsto X_n(\omega)$ defines a random path through S

Ex: $S = \{$"cloudy", "sunny"$\}$

Ex: $S = \{$letters in alphabet$\}$

X_n = letter in position n

$\Omega = \{\omega \mid \omega \in S^K\}$

Markov property

About how X at time $_{n+1}$ relates to its history.

(*) for any $n \geq 1$, $x, y \in S$ and history sequence $a_0, a_1, a_2, \dots a_{n-1} \in S$

$$\mathbb{P}(X_{n+1} = y \mid \underbrace{X_n = x, X_{n-1} = a_{n-1}, \dots, X_0 = a_0}_{\text{"entire history"}})$$
$$= \mathbb{P}(X_{n+1} = y \mid X_n = x)$$

Also for any n, $x, y \in S$, any sets $A_0, A_1, \dots, A_{n-1} \subset S$,

$$\mathbb{P}(X_{n+1} = y \mid X_n = x, X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0)$$
$$= \underbrace{\mathbb{P}(X_{n+1} = y \mid X_n = x)}_{P_n(x,y)}$$

"transition probability"

probability of a transition from $x \mapsto y$ given $X_n = x$

Time homogeneous:

$\mathbb{P}(X_{n+1} = y \mid X_n = x) = P(x,y)$

not dependent on n

$P(x,y) = P_{xy}$ ← matrix! $\qquad x, y \in S$

Any transition matrix must satisfy the following:

$0 \le P(x,y) \le 1$

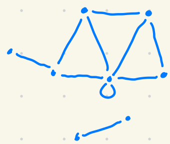
for any $x \in S$

$$\sum_{y \in S} P(x,y) = 1$$

<u>Ex</u>: Simple random walk on a graph

Let $G = (V, E)$ be an undirected graph, $|V| < \infty$,

no more than one edge per vertex pair.



Random walk:

$S = V$

Edges define allowable transitions

Given $X_n = x$ choose $X_{n+1} = y$

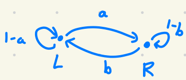
uniformly at random from

neighboring states $(x,y) \in E$

$$P(x,y) = \begin{cases} 0 & \text{if } (x,y) \notin E \\ \frac{1}{\deg(x)} & \text{if } (x,y) \in E \end{cases}$$

$\deg(x)$ = # of neighbors of x

<u>Ex</u>: 2-state model



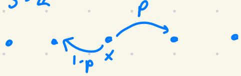
P = 2x2 matrix

$$\begin{array}{c|cc} & L & R \\ \hline L & 1-a & a \\ R & b & 1-b \end{array}$$

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

<u>Ex</u>: (biased) random walk on $\mathbb{Z}$.

$S = \mathbb{Z}$



$$P(x,y) = \begin{cases} p & y = x+1 \\ 1-p & y = x-1 \\ 0 & \text{else} \end{cases}$$

<u>Ex</u> Moran Model

N individuals in population

two types: red and blue

X_n = # of blue cells at "time" n

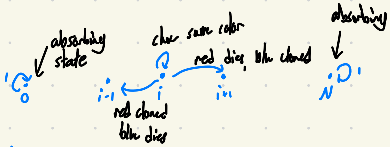
$N - X_n$ = # of red

At each time: an individual is chosen (uniformly at random) to be cloned

<u>and</u> an individual is chosen (uniformly at random) to be eliminated

independently

$S = \{0, 1, 2, \dots, N\}$



$$P(i,j) = \begin{cases} \left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right) & j = i+1 \\ \left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right) & j = i-1 \\ 1 - 2\left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right) & j = i \\ 0 & \text{else} \end{cases}$$

<u>Computations w/ P</u>

Suppose $X_0 = x$

$P(x,y) = \mathbb{P}(X_1 = y \mid X_0 = x)$

Suppose $X_0 \sim \mu$

μ = probability distribution on S

$0 \le \underset{\mathbb{P}(X_0 = x)}{\underset{\shortparallel}{\mu(x)}}, \quad \sum_{x \in S} \mu(x) = 1$

$$\mathbb{P}(X_1 = y) = \sum_{x \in S} \underbrace{\mathbb{P}(X_1 = y \mid X_0 = x)}_{P(x,y)} \underbrace{\mathbb{P}(X_0 = x)}_{\mu(x)}$$

$$= \sum_{x \in S} \mu(x) P(x,y)$$

$$= (\mu P)_y$$

row vector

$$\mathbb{P}(X_n = y) = \sum_{x \in S} \underbrace{\mathbb{P}(X_n = y \mid X_{n-1} = x)}_{P(x,y)} \mathbb{P}(X_{n-1} = x)$$

if μ_n is distribution of X_n,

$$\mu_n(y) = \sum_{x\in S} \mu_{n-1}(x) P(x,y)$$

So $\mu_n = \mu_{n-1} P$

$= \mu_0 P^{(n)}$ ← nth power of matrix P

$$\mathbb{P}(X_n = y) = \mu_n(y) = \sum_{x\in S} \mu_0(x) P^{(n)}(x,y)$$

Let $0 < m < n$

$\mathbb{P}(X_n = y, X_m = z \mid X_0 = x)$ ← start at x.

$P^{(m)}(x,z) P^{(n-m)}(z,y)$

Note:

$$\sum_{z\in S} \mathbb{P}(X_n = y, X_m = z \mid X_0 = x)$$
$$= \mathbb{P}(X_n = y \mid X_0 = x)$$

$$\sum_{z\in S} P^{(m)}(x,z) P^{(n-m)}(z,y)$$
$$= \left(P^{(m)} P^{(n-m)}\right)_{(x,y)}$$
$$= P^{(n)}_{(x,y)}$$

Let $a_1, a_2, \dots, a_n \in S$

$$\mathbb{P}(X_n = a_n, X_{n-1} = a_{n-1}, \dots, X_1 = a_1 \mid X_0 = x)$$
$$= P(x, a_1) P(a_1, a_2) \cdots P(a_{n-1}, a_n)$$

Stationary distribution

Def a probability distribution π on S is a stationary distribution for the chain if

$\pi P = \pi$

meaning $\sum_{x\in S} \pi(x) P(x,y) = \pi(y)$ for all $y \in S$

π is a left eigenvector w/ eigenvalue 1

If $X_0 \sim \pi$ then $\mathbb{P}(X_n = y \mid X_0 \sim \pi) = \pi(y)$

$$\mathbb{P}(X_n = y \mid X_0 = \pi) = \sum_{x\in S} \mathbb{P}(X_n = y \mid X_0 = x)\pi(x)$$

$= \sum_{x \in S} \pi(x) p^{(n)}(x,y)$

$= \pi(y)$

11/26

12/3

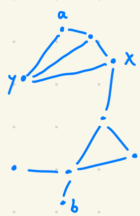
Calculations w/ Markov Chains

Exit problem: $|S| < \infty$, chain is irreducible

Let $a, b \in S$ be given.

Define $\tau_a = \min\{n \geq 0 \mid X_n = a\}$

= first time chain reaches a



Q: What is $\mathbb{P}(\tau_b < \tau_a \mid X_0 = x)$

$= \mathbb{P}(X$ reaches b before $a)$ — should depend on where you start

$=: h(x)$ — initial state

Claim: h satisfies linear system of equations

first: simple cases: $h(b) = 1 \quad h(a) = 0$

Think about conditioning on first step:

$$h(x) = \sum_{y \in S} \underbrace{\mathbb{P}(\tau_b < \tau_a \mid X_1 = y, \overbrace{X_0 = x}^{\text{don't need}})}_{h(y)} \underbrace{\mathbb{P}(X_1 = y \mid X_0 = x)}_{P(x,y)}$$

Thus for any $x \in S \setminus \{a,b\}$,

$$h(x) = \sum_{y \in S} P(x,y) h(y)$$

$|S|$ eqns in $|S|$ unknowns $h(x)$, $x \in S$

$$\begin{cases} h: S \to \mathbb{R} \text{ solves linear system for all } x \in S \setminus \{a,b\} \\ h(a) = 0 \\ h(b) = 1 \end{cases}$$

Exit Prize Problem

Let $D \subset S$ distinguished states

Let $g(y)$ be the prize for reaching $y \in D$ first

$g: D \to \mathbb{R}$

Let $\tau_D = \min\{n \geq 0 \mid X_n \in D\}$ = first time X arrives at distinguished state

Expected Prize: $\mathbb{E}\big(\overbrace{g(\underbrace{X_{\tau_D}}_{\text{state of } X \text{ when it first reaches } D})}^{\text{prize for first distinguished state reached}} \,\big|\, X_0 = x\big)$

Note: if $D = \{a,b\}$ and $g(y) = \begin{cases} 1 & \text{if } y=b \\ 0 & \text{if } y=a \end{cases}$

Then $\mathbb{E}(g(X_{\tau_D}) | X_0 = x)$

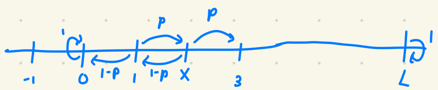
$= \mathbb{E}(\mathbb{1}_{\{\tau_b < \tau_a\}} | X_0 = x) = \mathbb{P}(\tau_b < \tau_a | X_0 = x)$

(general)

$h(x)$ satisfies linear system:

$$\begin{cases} h(x) = \sum_{y \in S} P(x,y) h(y) & \text{for } x \in S \setminus D \\ h(x) = g(x) & \text{for } x \in D \end{cases}$$

Ex: random walk on $\mathbb{Z}$. fix $p \in (0,1)$



Start at x $\quad 0 < x < L$

$h(x) = \mathbb{P}(\tau_L < \tau_0 | X_0 = x)$?

h satisfies linear system:

$h(0) = 0$

$h(L) = 1$

(*) $h(x) = \sum_{y \in S} P(x,y) h(y)$

for $x = 1, 2, \dots, L-1$

$h(0) = 0$

$h(L) = 1$

(*): $h(x) = p\, h(x-1) + (1-p) h(x-1)$

Case 1: unbiased walk $p = \frac{1}{2}$

$h(x) = \frac{1}{2} h(x-1) + \frac{1}{2} h(x-2)$

$= \frac{1}{2}\big(h(x-1) + h(x-2)\big)$

must be a line

$h(0) = 0 \quad h(L) = 1 \Rightarrow h(x) = \frac{x}{L} \quad x = 0, \dots, L$

If $\tau_L < \tau_0$ then $\tau_0 \geq L$

$\Rightarrow \mathbb{P}(\tau_0 \geq L | X_0 = 1) \geq \frac{1}{L}$

$\mathbb{E}(\tau_0 | X_0 = 1) = \sum_{k=1}^{\infty} \mathbb{P}(\tau_0 \geq k | X_0 = 1)$

$$\geq \sum_{k=1}^{\infty} \frac{1}{k}$$

$= +\infty$ oops

Case 2: biased walk

Assume $p \in (\frac{1}{2}, 1)$
(biased to the right)

System:

$h(0) = 0$

messes up here

$h(L) = 1$

$h(x) = p\, h(x+1) + (1-p)\, h(x-1)$

$h(x) = c$ $\quad = c(p + 1 - p)$

$= c$ ✓

Try: $h(x) = r^x$ for some r

r satisfies

$h(x) = r^x = p r^{x+1} + (1-p) r^{x-1}$

$\Rightarrow 1 = pr + (1-p) r^{-1}$

solvable quadratic eqn!

So $h(x) = r^x$ works so long as

r satisfies $r = pr^2 + (1-p)$

$r = 1$ is a root.

$pr^2 - r + (1-p) = 0$

$(r-1)(pr - (1-p)) = 0$

$r = \frac{1-p}{p}$ other root

General solution:

$h(x) = c_1 + c_2 \left(\frac{1-p}{p}\right)^x$ $\qquad h(0) = 0$

$h(L) = 1$

$c_1 = -c_2$

$h(L) = c - c\left(\frac{1-p}{p}\right)^L = 1$

$c\left(1 - \left(\frac{1-p}{p}\right)^L\right) = 1$

$c = \frac{1}{1 - \left(\frac{1-p}{p}\right)^L}$

So $h(x)$ has the form

$$h(x) = \frac{1 - \left(\frac{1-p}{p}\right)^x}{1 - \left(\frac{1-p}{p}\right)^L}$$

$p > \frac{1}{2} \Rightarrow 1-p < p$

$\Rightarrow \frac{1-p}{p} < 1$

$$h(1) = \mathbb{P}(\tau_L < \tau_0 \mid X_0 = 1) = \frac{1-\left(\frac{1-p}{p}\right)}{1-\left(\frac{1-p}{p}\right)^L} \xrightarrow[L\to\infty]{} 1-\left(\frac{1-p}{p}\right) \in (0,1)$$

$$\mathbb{P}(\tau_0 \geq L \mid X_0 = 1) \geq \mathbb{P}(\tau_L < \tau_0) \longrightarrow 1-\left(\frac{1-p}{p}\right)$$

$R_L = \{\tau_0 \geq L\}$

$\bigcap_{L\geq 1} R_L = \{\tau_0 = \infty\}$

$$\mathbb{P}(\tau_0 = \infty \mid X_0 = 1) = \lim_{L\to\infty} \mathbb{P}(\tau_0 \geq L) = 1-\left(\frac{1-p}{p}\right) = \frac{2p-1}{p}$$

for biased walk $p \neq \frac{1}{2}$, the origin (and every state $x \in S$) is transient $\longrightarrow$ $\exists$ positive probability that you'll never reach state

for <u>unbiased</u> walk, every state is recurrent, but expected time of return

$$\mathbb{E}(\tau_0 \mid X_0 = x) = \infty \quad x \neq 0$$

<u>Exit time problem</u>

Let $D \subset S$ distinguished states.

Assume $|S| < \infty$, irreducible chain

$$\tau_D = \min\{n \geq 0 \mid X_n \in D\}$$

let $h(x) = \mathbb{E}(\tau_D \mid X_0 = x)$

$h(x)$ satisfies linear system:

for $x \in D$: $\tau_D = 0$ $\quad h(x) = 0$

If $x \notin D$, must take at least one step. Condition on this: $\tau_D = 1 +$ remaining time τ_D'

$$\mathbb{E}(\tau_D \mid X_0 = x) = \sum_{y\in S} \mathbb{E}(\tau_D \mid X_1 = y, X_0 = x)\, \mathbb{P}(X_1 = y \mid X_0 = x)$$

$$= \sum_{y\in S} \mathbb{E}(1 + \tau_D' \mid X_1 = y, X_0 = x)\, P(x,y)$$

$$= \sum_{y\in S} \left(1 + \mathbb{E}(\tau_D' \mid X_1 = y, X_0 = x)\right) P(x,y)$$

$$= \sum_{y \in S} P(x,y) + \sum_{y \in S} P(x,y)\, h(y)$$

$$= \underset{\text{first step}}{1} + \underbrace{\sum_{y \in S} P(x,y)\, h(y)}_{\text{expected time remaining after first step}}$$

irreducible $\Rightarrow$ unique Solution
$$\begin{cases} h(x) = 1 + \sum_{y \in S} P(x,y)\, h(y) & x \in S \setminus D \\ h(x) = 0 & x \in D \end{cases}$$

12/5 Last day! <u>Markov Chain Monte-Carlo</u>

Let $|S| < \infty$. Suppose $f : S \longrightarrow [0,\infty)$ is some non-neg function on S.

Let $\pi(x) = \frac{f(x)}{c}$ $x \in S$ $c = \sum_{y \in S} f(y)$.

probability density

<u>Problem</u>: how to compute statistics of π without knowing c?

Typically, given x, $f(x)$ can be evaluated "easily"

$|S| \gg 1$ may be so large that we can't compute c.

ex) Compute $E_\pi(g(x)) = \sum_{x \in S} g(x)\pi(x)$

Idea: build a Markov Chain X_n on S so that π is the stationary distribution.

Then $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k) = E_\pi(g(X))$ (assuming chain is irred, aperiodic)

<u>Examples</u>

1) Bayesian Statistical models:

X = parameter in model

$f(x) = f(x,z)$ = likelihood of parameter x given data z

2) Statistical Physics

Ising Model (1920's)

X = configurations of spin of many atoms

$f(x) = e^{-\beta H(x)}$ "energy"

3) "pick a random sentence"

S = all 100 letter sentences

$= A^{100}$ $A = \{a, b, c, \ldots, z\}$

$= \{x = (l_1, \ldots, l_{100}) \mid l_i \in A\}$

$|S| = |A|^{100} = 26^{100}$

Suppose $f: S \to [0, \infty)$

$f(x) = e^{\beta(\#\text{ special words from a list})}$

$W = \{\text{"probability"}, \text{"Duke"}, \text{"finals"}\}$

<u>The algorithms</u>

1) Metropolis-Hastings

2) Gibbs sampling

1) Metropolis-Hastings (1950s)

Input: $f: S \to [0, \infty)$

$q(x,y)$ a transition probability on S

<u>Algorithm</u>: Given $X_n = x$, generate X_{n+1} as follows

1) propose a new state y according to $q(x,y)$

$x = (l_1, \dots, l_N)$

Pick index $k \in \{1, \dots, N\}$

Draw new letter $l_k' \in A$

replace $l_k \to l_k'$

So q is pretty simple (and sparse)

2) Accept or reject:

compute $\alpha(x,y) = \min\left(1, \dfrac{f(x)\, q(y,x)}{f(y)\, q(x,y)}\right)$

toss an α-coin:

heads $\to$ accept, set $X_{n+1} = y$

tails $\to$ reject, set $X_{n+1} = x$

<u>Claim</u>: $\pi(x) = \dfrac{f(x)}{c}$ is stationary for this chain.

transition probability implicitly defined by algo

pf: Need to show $\sum_x \pi(x) P(x,y) = \pi(y)$

suffices to show $\pi(x) P(x,y) = \pi(y) P(y,x)$ for all x, y — detailed balance condition

Because if holds, then

$$\sum_x \pi(x) P(x,y) = \sum_x \pi(y) P(y,x) = \pi(y) \underbrace{\sum_x P(y,x)}_{=1} = \pi(y)$$

Lets prove that detailed balance condition holds.

If $x = y$, obvious, so consider $x \neq y$. May assume $f(y) q(y,x) < f(x) q(x,y)$

$$P(x,y) = q(x,y) \overbrace{\min\left(1, \frac{\pi(y)\, q(y,x)}{\pi(x)\, q(x,y)}\right)}^{\alpha(x,y)}$$

probability that you propose

$$= q(x,y) \frac{\pi(y)\, q(y,x)}{\pi(x)\, q(x,y)}$$

So

$$\pi(x) P(x,y) = \cancel{\pi(x)\, q(x,y)} \frac{\pi(y)\, q(y,x)}{\cancel{\pi(x)\, q(x,y)}} = \pi(y)\, q(y,x)$$

$\parallel$ ✓ yay!

$$\pi(y) P(y,x) = \pi(y)\, q(y,x) \min\left(1, \frac{\pi(x)\, q(x,y)}{\pi(y)\, q(y,x)}\right) = \pi(y)\, q(y,x) = \pi(x) P(x,y)$$ ✓ $\blacksquare$

Gibbs Sampler:

Specific to S of the form $S = A^N = \{x = (l_1, \dots, l_N) \mid l_i \in A\}$

e.g. $A = \{+1\ -1\}$ $\quad l_i =$ spin state of atom i

$f(x) = e^{-\beta H(x)}$

Algorithm: given $X_n = x = (l_1, \dots, l_n)$ generate X_{n+1} as follows:

1) Pick index $k \in \{1, \dots, N\}$ uniformly at random

2) Resample $l_k \to l_k'$ s.t.

$$\mathbb{P}(l_k' = a) = \frac{\cancel{f}\,\pi(\underbrace{l_1, \dots, l_{k-1}}_{\text{frozen}}, a, \underbrace{l_{k+1}, \dots, l_N}_{\text{frozen}})}{\sum_{b \in A} \cancel{f}\,\pi(l_1, \dots, l_{k-1}, b, l_{i+1}, \dots, l_{k+1})}$$

Set $X_{n+1} = (l_1, \dots, l_{k-1}, l_k', l_{k+1}, \dots, l_N)$