11/17

LECTURE 19
Limit theorems – I
of 2

Back to basics
- when have lots of RVs

- **Readings:** Sections 5.1-5.3; start Section 5.4

"weak law of large #"

next time: central limit theorm

- $X_1, \ldots, X_n$ i.i.d.

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

What happens as $n \to \infty$?

gets close to correct

- Why bother?
- A tool: Chebyshev's inequality
- Convergence "in probability"
- Convergence of a RV ~~M_n~~ (weak law of large numbers)

- if want to know avg height of penguines
  - catch n penguines
  - then we get average over our sample

$E[X_i]$ "true mean" ↳ "sample mean"

as get 1-2 → fairly easy
as get 15 RVs - hard to do
lots of joint PDFs

as go $\lim_{\to\infty}$ RV then easy again

markov inequality - separate page

Chebyshev's inequality

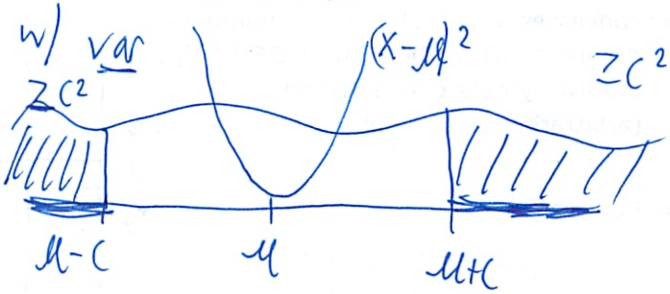
- Random variable X (with finite mean μ and variance σ^2)

Like markov except w/ $X-\mu$

Working w/ var

$$\begin{aligned}\sigma^2 &= \int (x-\mu)^2 f_X(x)\,dx \\ &\geq \int_{-\infty}^{-c+\mu} (x-\mu)^2 f_X(x)\,dx + \int_{\mu+c}^{\infty} (x-\mu)^2 f_X(x)\,dx \\ &\geq c^2 \cdot \mathbf{P}(|X-\mu| \geq c)\end{aligned}$$

$\mu+c$ ← correction



Only over region $|x-\mu| \geq c$

$-x+\mu = c$

$x = \mu - c$

$$\mathbf{P}(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

$$P((X-\mu)^2 \geq a) \leq \frac{Var(x)}{a}$$

$c = k\sigma$

$$\mathbf{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Probability that you fall at least 1 st dev < 1

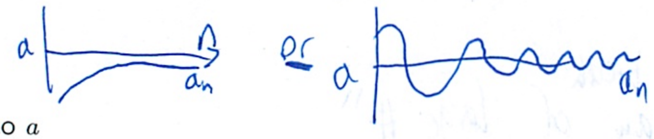
2st dev < ¼ from mean

obtaining bounds on distribution
1. Fix c to be certain #
2. P of being at least c away from μ
3. Look at var - when var small, small prob of being away from the mean

Deterministic limits

Refresher on limits

- Sequence a_n
  Number a

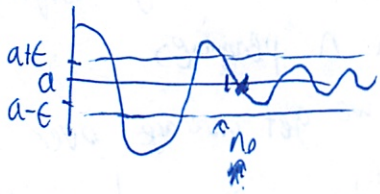


- a_n converges to a

$$\lim_{n\to\infty} a_n = a$$

"a_n eventually gets and stays (arbitrarily) close to a"

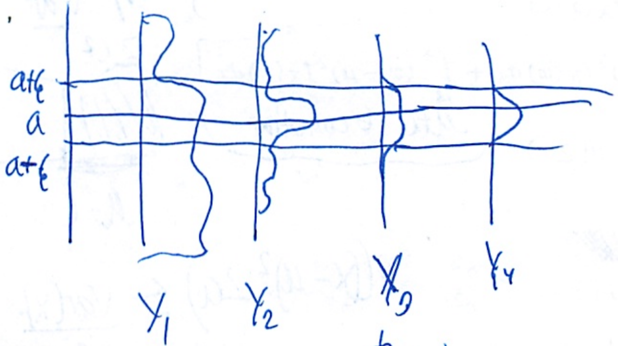
- For every $\epsilon > 0$,
  there exists n_0,
  such that for every $n \geq n_0$,
  we have $|a_n - a| \leq \epsilon$.



Convergence "in probability"

what does it mean for RVs to converge

- Sequence of random variables Y_n
- converges in probability to a number a:
  "(almost all) of the PMF/PDF of Y_n, eventually gets concentrated (arbitrarily) close to a"



lots of prob of being outside ε band

the prob distribution outside band falls to 0

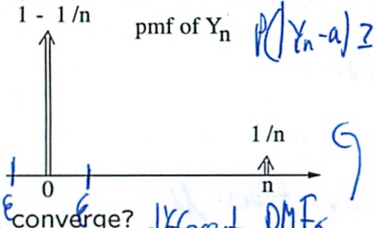
- For every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}(|Y_n - a| \geq \epsilon) = 0$$

the wrote very bad

$\forall \delta > 0 \ \forall_\epsilon > 0 \ \exists n_0$ s.t. $\forall n \geq n_0$ [illegible]

$P(|Y_n - a| \geq \epsilon) \leq \delta$



Does Y_n converge? different PMFs

$P(|Y_n - 0| \geq \epsilon) = \frac{1}{n} \xrightarrow{\text{as } n\to\infty} 0$

gets more and more concentrated at 0 as n goes to ∞

$E[Y_n] = 1$
$E[Y_n^2] = n$

- just tells us about probs - not about E[ ]

Convergence of the sample mean

(Weak law of large numbers)

- $X_1, X_2, \ldots$ i.i.d.
  finite mean μ and variance σ^2

Penguine heights

$$RV \rightarrow M_n = \frac{X_1 + \cdots + X_n}{n}$$

- $\mathrm{E}[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \overset{\text{true mean}}{\mu}$

\# won't have bias

On avg → # will = true mean = μ

- $\mathrm{Var}(M_n) = \dfrac{Var(X_1) + \cdots + var(X_n)}{n^2} = \dfrac{n\sigma^2}{n^2} = \sigma^2/n$

$$\mathrm{P}(|M_n - \mu| \geq \epsilon) \ \leq \ \frac{\mathrm{Var}(M_n)}{\epsilon^2} \ = \ \frac{\sigma^2}{n\epsilon^2}$$

← prob that sample mean makes a mistake bigger than ϵ

and $\forall \epsilon > 0 \quad \lim_{n \to \infty} = 0$ prob dist gets concentrated on ϵ band

- M_n converges in probability to μ

$$M_n \xrightarrow{i.p.} \mu = E[X]$$

i.p. = in probability

The pollster's problem

- f: fraction of population that "..." = true fraction
- ith (randomly selected) person polled:

- can't find w/o asking everyone

$$X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$$

- $M_n = (X_1 + \cdots + X_n)/n$
  fraction of "yes" in our sample

$E[X_i] = f$

- Goal: 95% confidence of $\leq$1% error

$$\mathrm{P}(\underbrace{|M_n - f| \geq .01}_{\text{accuracy}}) \leq \underbrace{.05}_{\text{confidence}} \quad \pm 1\%$$

- sample ⇒ want a representive answer

How confident are we that our results are accurate

- Use Chebyshev's inequality:

$$\mathrm{P}(|M_n - f| \geq .01) \ \leq \ \frac{\sigma^2_{M_n}}{(0.01)^2}$$
$$= \frac{\sigma^2_x}{n(0.01)^2} \leq \frac{1}{4n(0.01)^2}$$

↙ chose n so this is ≤ .05

- If $n = 50{,}000$,
  then $\mathrm{P}(|M_n - f| \geq .01) \leq .05$
  (conservative)

← but always polls w/ less → do accuracy 3%

just an inequality

- can be much smaller

$\sigma_x^2 = f(1-f)$ Bernoulli RV

- but don't know f - so don't know var

- so look at worst possible situation

$\sigma_x^2 \leq \frac{1}{4}$

Different scalings of M_n

- $X_1, \ldots, X_n$ i.i.d.
  finite variance σ^2

- Look at three variants of their sum:

- $S_n = X_1 + \cdots + X_n$ variance $n\sigma^2$

  as add more → var goes up

  $E[S_n] = n\mu$

- $M_n = \dfrac{S_n}{n}$ variance σ^2/n
  converges "in probability" to $\mathrm{E}[X]$ (WLLN)

- $\dfrac{S_n}{\sqrt{n}}$ ← Var will be constant constant variance σ^2

  $$\mathrm{Var}\left(\frac{S_n}{\sqrt{n}}\right) = \frac{\mathrm{Var}(S_n)}{n} = \sigma^2$$

  - Asymptotic shape?

  bell shape → normal distribution

The central limit theorem

- "Standardized" $S_n = X_1 + \cdots + X_n$:

  when scale → get a RV approx normal

  $$\overset{\uparrow \text{new rv}}{Z_n} = \frac{S_n - \mathrm{E}[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mathrm{E}[X]}{\sqrt{n}\,\sigma}$$

  how many st. dev is S_n away from mean

  - zero mean
  - unit variance

- Let Z be a standard normal r.v.
  (zero mean, unit variance)

- **Theorem:** For every c:

  $$\mathrm{P}(Z_n \le c) \to \mathrm{P}(Z \le c)$$

  converges st normal

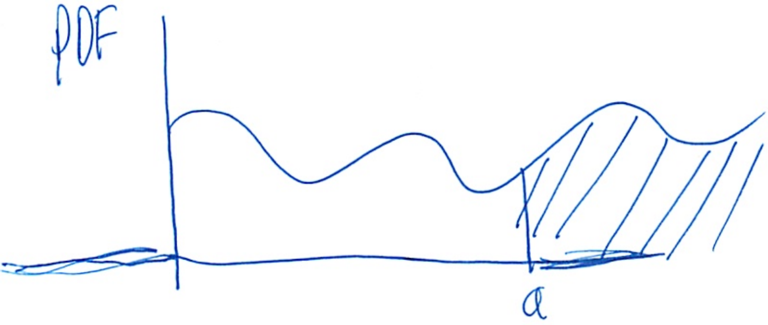
  no matter the distribution of X
  - as long as they are ind

- $\mathrm{P}(Z \le c)$ is the standard normal CDF, $\Phi(c)$, available from the normal tables

↑ preview - more next time

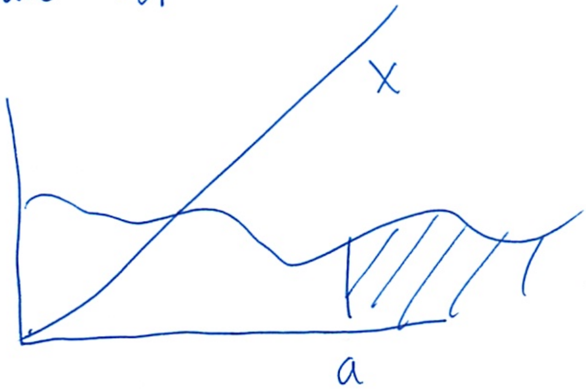
Markov inequality

$X \geq 0$



$$E[X] = \int_0^\infty x f_X(x)\, dx$$

Take PDF



$$E[X] \geq \int_a^\infty x f_X(x)\, dx$$

$$\geq a \int_a^\infty f_X(x)\, dx$$

$$= a P(X \geq a)$$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2010)

Recitation 20: November 18, 2010

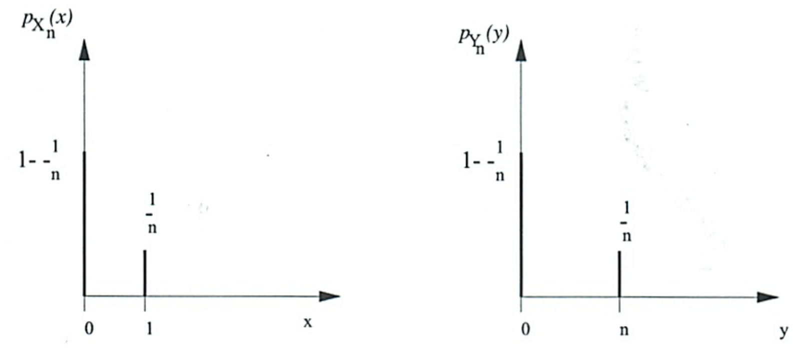
1. In your summer internship, you are working for the world's largest producer of lightbulbs. Your manager asks you to estimate the quality of production, that is, to estimate the probability p that a bulb produced by the factory is defectless. You are told to assume that all lightbulbs have the same probability of having a defect, and that defects in different lightbulbs are independent.

   (a) Suppose that you test n randomly picked bulbs, what is a good estimate Z_n for p, such that Z_n converges to p in probability?

   (b) If you test 50 light bulbs, what is the probability that your estimate is in the range $p \pm 0.1$?

   (c) The management asks that your estimate falls in the range $p \pm 0.1$ with probability 0.95. How many light bulbs do you need to test to meet this specification?

2.



   Let X_n and Y_n have the distributions shown above.

   (a) Find the expected value and variance of X_n and Y_n.

   (b) What does the Chebyshev inequality tell us about the convergence of X_n? Y_n?

   (c) Is Y_n convergent in probability? If so, to what value?

   (d) If a sequence of random variables converges in probability to a, does the corresponding sequence of expected values converge to a? Prove or give a counter example.

   A sequence of random variables is said to converge to a number c in the **mean square**, if

$$\lim_{n\to\infty} \mathbf{E}\left[(X_n - c)^2\right] = 0.$$

   (e) Use Markov's inequality to show that convergence in the mean square implies convergence in probability.

   (f) Give an example that shows that convergence in probability does not imply convergence in the mean square.

3. Random variable X is uniformly distributed between -1.0 and 1.0. Let $X_1, X_2, \ldots,$ be independent identically distributed random variables with the same distribution as X. Determine which, if any, of the following sequences (all with $i = 1, 2, \ldots$) are convergent in probability. Give reasons for your answers. Include the limits if they exist.

(a) X_i

(b) $Y_i = \dfrac{X_i}{i}$

(c) $Z_i = (X_i)^i$

Recitation 20

11/18

- reaching threshold from prob → stats
  - ^had first lecture
- Stats
  - deal w/ data
  - when trying to estimate E[ ] from sampling
  - consider as estimate of true mean

$X_1, X_2, \ldots$ sequence iid RVs

mean μ var σ^2

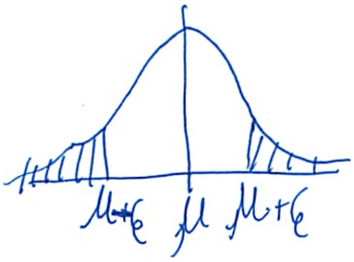
$$M_n = \underbrace{\frac{X_1 + X_2 + \ldots + X_n}{n}}_{\text{Sample mean}}$$

$E[M_n] = \mu$

$var(M_n) = \frac{\sigma^2}{n}$

Important Aim: Estimate confidence interval $P(|M_n - \mu| > \epsilon)$

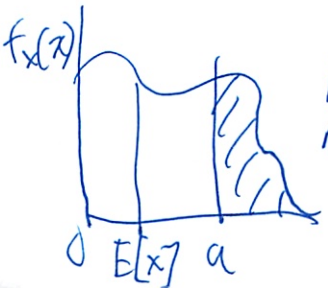
Markov Inequality: Suppose X RV that takes ≥ 0



If $a > 0$ then

$$\underset{x}{P}(X \geq a) \leq \frac{E[X]}{a}$$

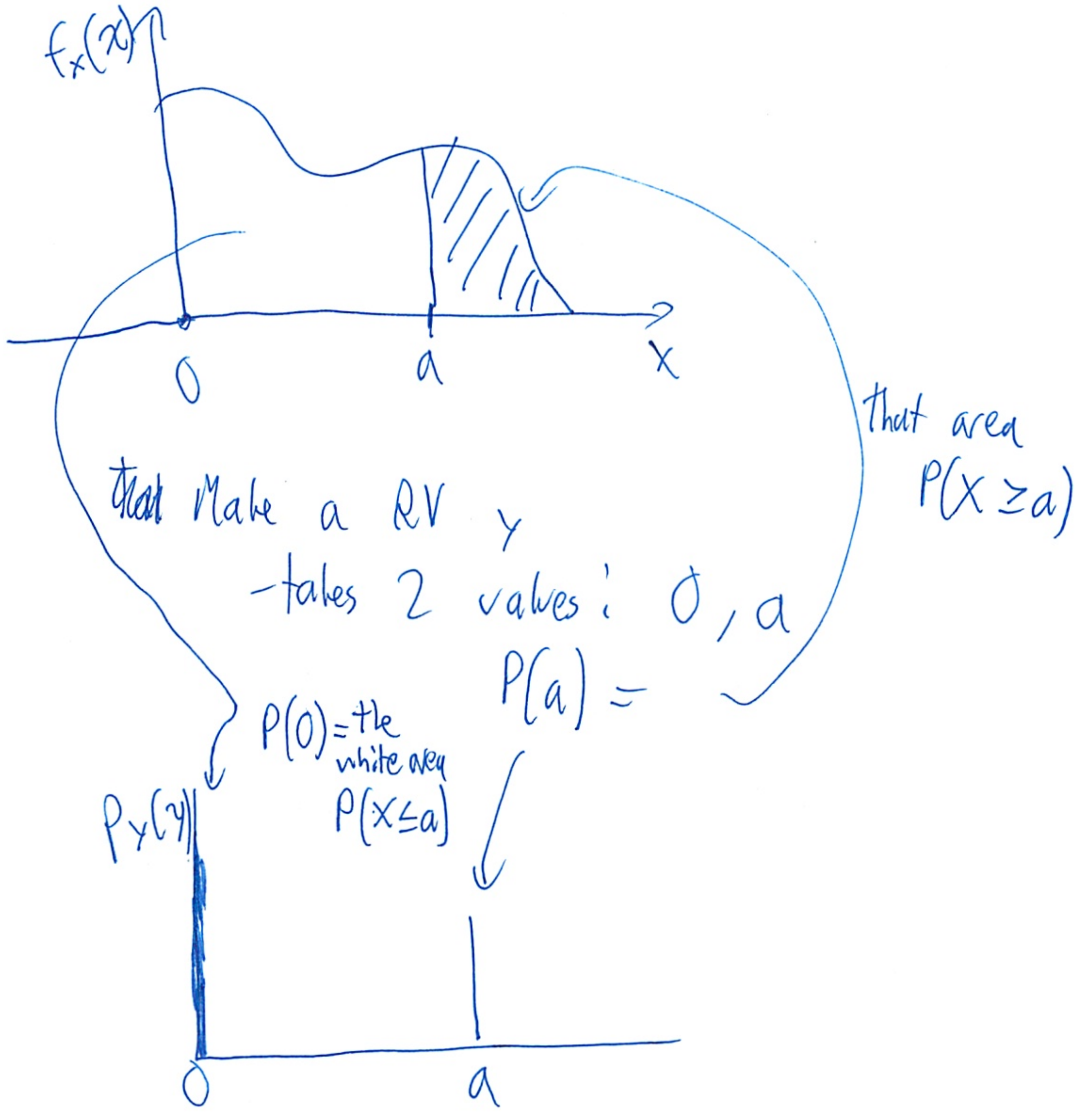
$|W_n - a|^2$



bounded by mean-ratio of a

②

Graphical Proof



that area $P(X \geq a)$

that Make a RV Y

- takes 2 values: 0, a

$P(a) =$

$P(0) =$ the white area $P(X \leq a)$

$$\underbrace{E[Y]}_{\text{will be } < E[x]} = \underbrace{0 \cdot P(X < a)}_{\text{dissappears}} + a \cdot P(X \geq a)$$

So get Markoff inequality

$$P(X \geq a) \leq \frac{E[X]}{a}$$

q.e.d ← (what does that mean again?)

③

Chebichev Inequality (less exact)

Take RV Y, $a > 0$

$$P(|Y - E[Y]| \geq a) \leq \frac{var(Y)}{a^2}$$

Proof

Take $X = (Y - E[Y])^2$

Apply Markov.

$$P(|Y - E[Y]| \geq a) = P(\underbrace{|Y - E[Y]|^2}_{X} \geq a^2) \leq$$

$$\frac{E[\overbrace{(Y - E[Y])^2}^{var(Y)}]}{\alpha^2}$$

Apply Chebichev to sample mean)

$$M_N = \frac{X_1 + \dots + X_n}{n} \qquad E[M_n] = \mu \qquad var(M_n) = \frac{\sigma^2}{n}$$

$$P(|M_n - \mu| \geq \underset{\substack{\uparrow \\ \text{Confidence} \\ \text{level}}}{c}) \leq \boxed{\frac{\sigma^2}{\underset{\substack{\uparrow \\ \text{\# of} \\ \text{Samples}}}{(n)} \cdot c^2}} \leftarrow \text{band on confidence probability}$$

④

1, Lightbulbs Probability lightbulbs are defective

- never know true probability of defectiveness

a) Lightbulbs good w/ prob p.

Want to estimate p

$$X_i = i\text{th sample} = \begin{cases} 1 & \text{if } i\text{th bulb good} \\ 0 & \text{if " " bad} \end{cases}$$

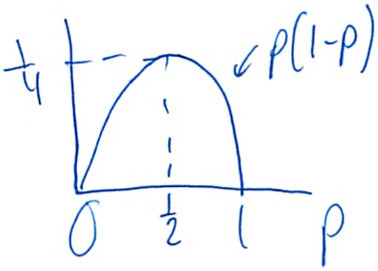
Bernoulli

$E[X_i] = p$ $var(X_i) = p(1-p)$

$M_n = \frac{X_1 + \dots + X_n}{n}$, estimate p w/ M_n

- But we don't know σ^2
  - need all parts to find bound
- But is upper bound to $p(1-p)$

$p(1-p) \leq \frac{1}{4}$



So will use $\frac{1}{4}$, since no better upper bound

Suppose $n = 50$

$$P(|M_n - p| \geq .1) \overset{?}{\leq} \frac{p(1-p)}{\underset{50}{n} \cdot (.1)} \leftarrow \text{var}$$

$$\leq \frac{\frac{1}{4}}{50 \cdot (.1)^2}$$

$$\leq \frac{1}{2}$$

No more than 50% chance error will be less than .1 if test 50 bulbs

But it is just a bound - true probability may be less

Typically use central limit theorm

c) Find n so that $|M_n - p| \leq .1$ w/ prob no more than .05

$$P(|M_n - p| \leq .1) \leq \frac{1/4}{n \cdot (.1)^2} \leq .05$$

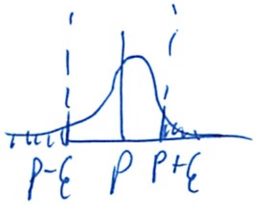
$$n \geq \frac{\frac{1}{4}}{.05 \cdot (.1)^2}$$

$= 500$ (need to test 500 bulbs

$.1 =$ error
$.05 =$ confidence } right?

From Chebichev inequality

$$P(|M_n - p| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$



as n increases, more and more prob will be inside the ε bounds

- "Convergence in probability"

$$P(|M_n - p| \geq \varepsilon) \to 0 \quad \text{as } n \to 0$$

for every $\varepsilon > 0$

Definition A seq. of RVs $Y_1, Y_2, \ldots$ Converges to a prob if ~~for~~ every $\varepsilon > 0$

$$P(|Y_n - a| \geq \varepsilon) \to 0$$

⑦ Weak Law of large #

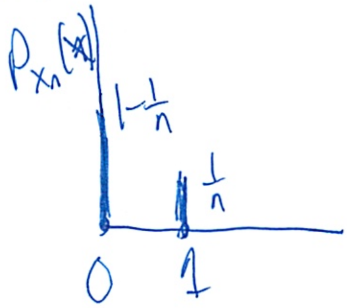
- Converges of Fnctions
  - for every sequence is a whole ensamble of #
- "Convergence" can have many different meanings
- this is weak convergence
  - not actual / strong convergence

(kinda confused)

- will do a function later that makes this more complex

$$\boxed{\text{WLLN} \rightarrow \{M_n\} \rightarrow p \text{ in probability}}$$

2.



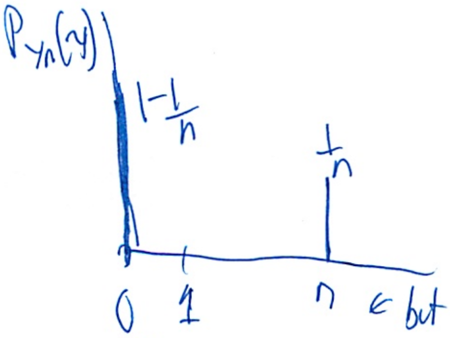
↑ as n ↑ more prob shifted to 0

$$E[X_n] = \frac{1}{n}$$

$$\text{var}(X_n) = \left(0-\frac{1}{n}\right)^2 \cdot \left(1-\frac{1}{n}\right) + \left(1-\frac{1}{n}\right)^2 \cdot \frac{1}{n}$$

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

var → 0 as $n \to \infty$



← but this is moving out →

- has counterbalance effect

↑ shifting more to 0

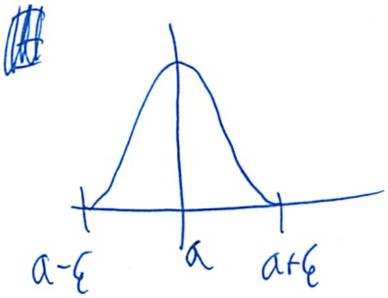
$$E[Y_n] = 1$$

$$\text{var}(Y_n) \geq \frac{1}{n}(n-1)^2$$

$\frac{1}{n}(n-1)^2 \to \infty$ as $n \to \infty$

⑧

Do sequences converge in probability?



Is it all within ε?

So $X_n \overset{\text{converges}}{\rightarrow} 0$ in probability

$Y_n \rightarrow 0$ " "

what is the likely limit in probability?

but $E[Y_n] \neq 0$ (so one does not imply the other)

- the means
- and converging in probability

<u>Define</u> $W_1, W_2, \ldots$ converge to a in <u>mean square</u> if $E[(W_n - a)^2] \rightarrow 0$ as $n \rightarrow \infty$

<u>Claim</u> Convergence in mean square

$\hookrightarrow$ Convergence in probability

Apply Markov $\neq$lity w/ $X = (W_n - a)^2 \geq 0$

$$P(|W_n - a|^2 \geq \epsilon^2) \leq \frac{E[|W_n - a|^2]}{\epsilon^2} \to 0$$

Converges $\to 0$ as $n \to 0$

But reverse is not true

Convergence in mean square $\not\Leftarrow$ Convergence in prob

$E[(Y_n - a)^2]$

$Y_n \to 0$ in prob

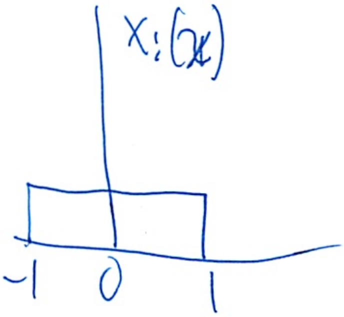
but $E[(Y_n)^2] \to \infty$

So <u>not</u> in mean square

3rd form of convergence

- Convergence almost everywhere
- will mention next week a bit
- leads to strong law of large #

3. Investigating convergace in probabilities



a) $\{X_i\} \xrightarrow{?} 0$ in prob

- <u>no</u> - Prob mass must be squishd to #

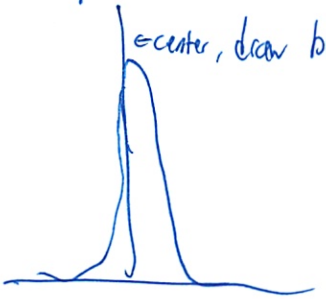
b) $\{\frac{X_i}{i}\} \xrightarrow{?} 0$ in prob

- Yes

smushes into 0

c) $\{(X_i)^i\} \xrightarrow{?} 0$ in prob

- absolutly



← center, draw better

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2010)

Tutorial 10
November 18/19, 2010

1. Define X as the height in meters of a randomly selected Canadian, where the selection probability is equal for each Canadian, and denote $\mathbf{E}[X]$ by h. Bo is interested in estimating h. Because he is sure that no Canadian is taller than 3 meters, Bo decides to use 1.5 meters as a conservative (large) value for the standard deviation of X. To estimate h, Bo averages the heights of n Canadians that he selects at random; he denotes this quantity by H.

   (a) In terms of h and Bo's 1.5 meter bound for the standard deviation of X, determine the expected value and standard deviation for H.

   (b) Help Bo by calculating a minimum value of n (with $n > 0$) such that the standard deviation of Bo's estimator, H, will be less than 0.01 meters.

   (c) Bo would like to be 99% sure that his estimate is within 5 centimeters of the true average height of Canadians. Using the Chebyshev inequality, calculate the minimum value of n that will make Bo happy.

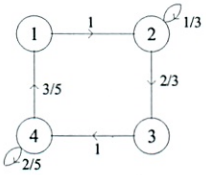
   (d) If we agree that no Canadians are taller than three meters, why is it correct to use 1.5 meters as an upper bound on the standard deviation for X, the height of any Canadian selected at random?

2. On any given week while taking 6.041, a student can be either up-to-date on learning the material, or she may have fallen behind. If she is up-to-date in a given week, the probability that she will be up-to-date (or behind) in the next week is 0.8 (or 0.2, respectively). If she is behind in the given week, the probability that she will be up-to-date (or behind) in the next week is 0.6 (or 0.4, respectively). We assume that these probabilities do not depend on whether she was up-to-date or behind in previous weeks, so we can model the situation as a 2-state Markov chain where State 1 is the case when the student is up-to-date and State 2 is the case when the student is behind.

   (a) Calculate the mean first passage time to State 1, starting from State 2.

   (b) Calculate the mean recurrence time to State 1.

3. Consider the following Markov chain:



The steady-state probabilities for this process are:

$$\pi_1 = \frac{6}{31} \qquad \pi_2 = \frac{9}{31} \qquad \pi_3 = \frac{6}{31} \qquad \pi_4 = \frac{10}{31}$$

Assume the process is in state 1 just before the first transition.

   (a) Determine the expected value and variance of K, the number of transitions up to and including the next transition on which the process returns to state 1.

(b) What is the probability that the state of the system resulting from transition 1000 is neither the same as the state resulting from transition 999 nor the same as the state resulting from transition 1001?

Tutorial 10

11/19

Markov Ineq

if $x \geq 0$

$$P(X \geq a) \leq \frac{E[X]}{a} \qquad \forall a > 0$$

Chepyshev Ineq

- small variance ⇒ not very spread out

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c2} \qquad \forall c > 0$$

if X's mean $= \mu$

Var $= \sigma^2$

Weak Law of Large

"$M_n \xrightarrow[\text{prob}]{} M$"

$$M_n = \frac{X_1 + \dots + X_n}{n} \qquad \text{sample mean}$$

$$P(|M_n - M| \geq \epsilon) \to 0$$

$n \to \infty$

$\forall \epsilon > 0$

Convergence in Prob

$Y_1, Y_2, \dots, Y_n$ seq of RVs is a real #

if $\forall \epsilon > 0$, $\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0$ ⇒ $Y_n \xrightarrow{\text{prop}} a$

②

1. Height of Canenadians

$$E[x] = h$$

$$\sigma_X = 1.5m$$

$$H = \frac{X_1 + \dots + X_n}{n}$$

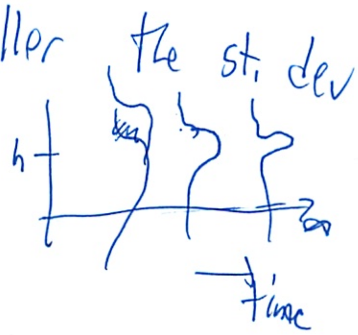
Good estimate is sample mean
-but can't take ∞ samples

a) $$E[h] = E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{nE[x]}{n} = h$$

$$Var(h) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}(n \, var(x) = \frac{var(x)}{n}$$

$$\sigma_H = \frac{1.5}{\sqrt{n}} \qquad = \frac{(1.5)^2}{n}$$

the more samples of n, the smaller the st. dev
→ Convergence in probability



b) want
$\sigma_H \leq .01$ So $\frac{1.5}{\sqrt{n}} \leq .01$

So n must be ≥ 22500

③

c) Wants 99% confidence of error = 5cm

↓ real quantity

-only % when the qu is about % like support Obama? poll

n=?

Use Chebethev

$$P(|H-h| \geq .05) \leq \frac{Var(h)}{c^2}$$

$$\leq \frac{(1.5)^2}{n(.05)^2} \leq .01$$

or

$$\frac{(1.5)^2}{n(.05)^2} \geq .99$$

$$\frac{(1.5)^2}{(.05)^2} \leq n .01$$

$$\frac{(1.5)^2}{(.01)(.05)^2} \leq n$$

$$n \geq 90000$$

You will always have some sort of error, ~~because~~ if you don't sample everyone

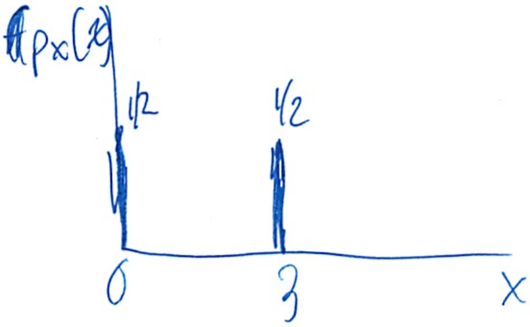
④

~~Sol~~

d)

Can't have an ∞ variance on something ~~[illegible]~~ that is bounded

So make a PDF?



$$var(x) = \frac{1}{2}(3-1.5)^2 + \frac{1}{2}\left(0-\frac{1}{2}\right)^2$$

$$= (1.5)2$$

So make $\sigma^2 = 1.5$

$$var(x) = \frac{1}{2}\left(b-\frac{b}{2}\right)^2 + \frac{1}{2}\left(0-\frac{b}{2}\right)^2$$

$$= \frac{b^2}{4}$$

So max $var(X) = \frac{(b-a)^2}{4}$

- if don't have var, can estimate, or use upper bound on variable

ie if $f_X(x)$

+ assuming ~~ha~~ symetrical

← told a + b are the bounds ✓??

Chebichev $P\left(\left|X - \frac{b+a}{2}\right| \geq c\right) \leq \frac{var(x)}{c^2}$

can still get inequality by splitting upper bound

$$\leq \frac{(b-a)^2}{4ac}$$

? statement only for this symetrical figure

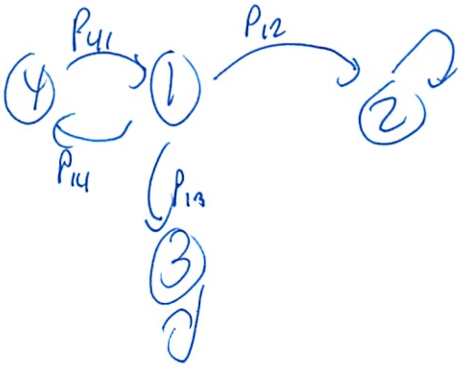
* We only know this for very specific plots *

- Chebichev - not required to know on exam
  - On grad problem
  - Could use 1 sided for not symetrical dists.

⑥

Markov

Asorbtion possibilities



2 and 3 are asorbing states

Starting from state i, what is prob of getting asorbed in s

- fix s
- a_i

$a_i = 0 \ \forall i$ ~~other~~ recurrent states

$a_s = 1$ of course stay

$a_i = \sum_j p_{ij} a_j$

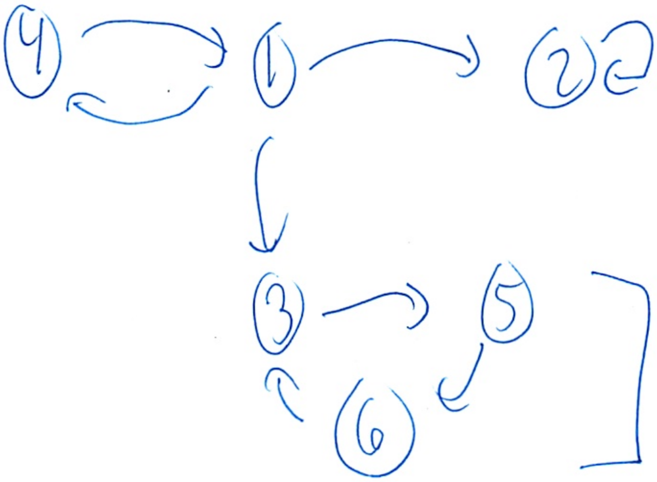
fix s=3

$a_3 = 1$

$a_2 = 0$

$a_1 = p_{13} \cdot 1 \cdot p_{14} \cdot a_4$

⑦ with
recurrent claseses



just pretend recurrent class is one state, like before

E[ ] to asorbtion

$\mu_i = E[\# \text{ of transitions until asorbtion} \mid \text{start } X_0 = i]$

$\mu_1 =$

$\mu_2 = 0$

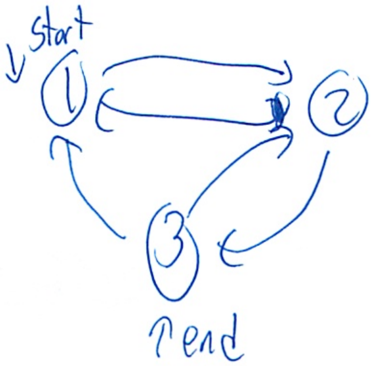
$\mu_3 = 0$

$\mu_4 =$

$\mu_i = 0$ ∀ recurrent states i

$\mu_i = 1 + \underbrace{\sum_j p_{ij}\mu_j}_{\text{total expectation}}$ transient

---



Wait expected time

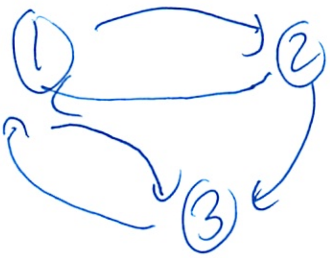
⑧ So basically

Mean 1st passage time

$$t_i = 1 + \sum_{j=1}^{m} P_{ij} t_j \quad \forall i \neq s$$

Mean recurrence time

if I start in ① how long till revisit?



essentially →

$$t_s^* = E[\# \text{ of transitions to return to } s \text{ starting from } s]$$

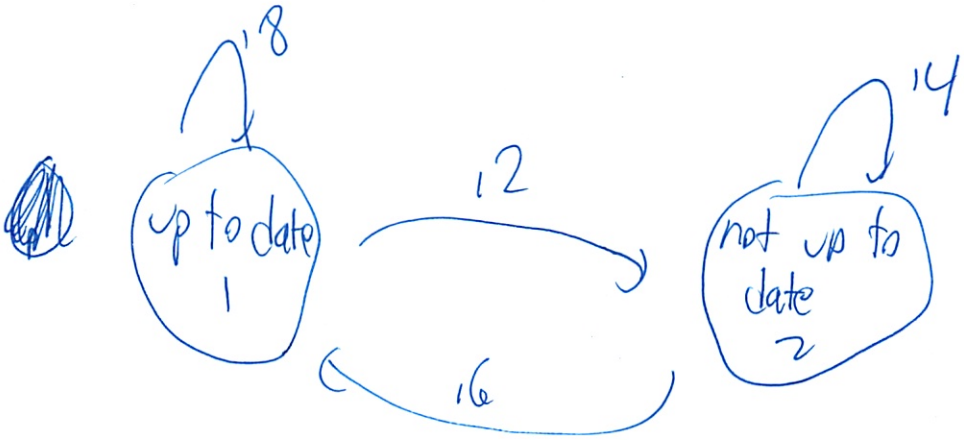
$$= 1 + \sum_j P_{sj} t_j$$

~~Mean 1st passage time~~

~~$t_i = 1 + \sum^m$~~

⑨

#2



a) Mean 1st passage time $2 \to 1$

w/o using formula (simpler)

keep looping in 2 then break away

$X = \text{geometric } (p = .6)$

So $E[X] = \frac{1}{.6}$

b) $t_1^* = 1 + \sum P_{1j} t_j$

$= 1 + P_{11} \underset{?}{t_1} + P_{12} t_2$

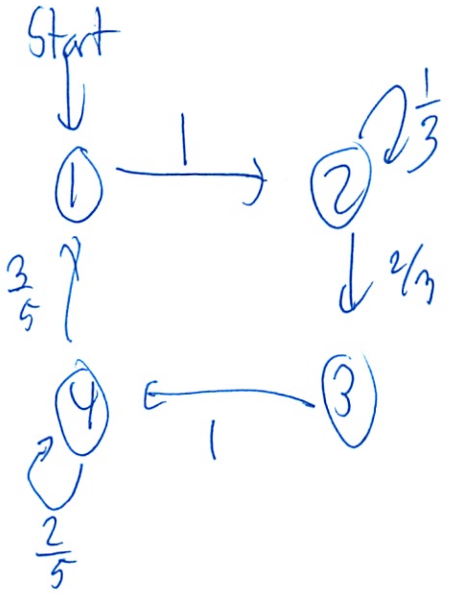
mean 1st passage time

$= 1 + 0 + .2\left(\frac{1}{.6}\right)$

$= \frac{4}{3}$

⑩

3.



a) k = time to go from 1 until 1

mean recurrence time

- i can do w/o equations

$k = \underline{1} + X_1 + \underline{1} + X_2$ ← # of steps

geometric(2/3) geometric(4/5)

$E[k] = 2 + E[X_1] + E[X_2]$

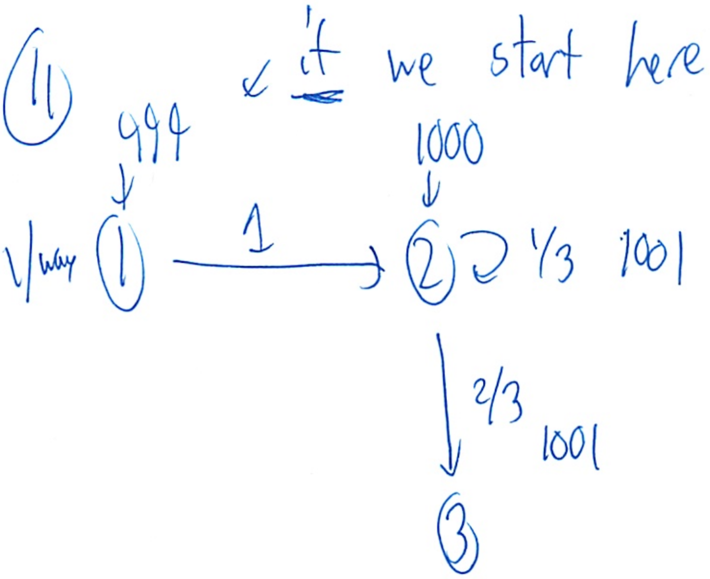
$2 + \frac{3}{2} + \frac{5}{3}$

$var(k) = var(X_1) + var(X_2)$

b) $P(Y_{999} \neq Y_{1000} \neq Y_{1001})$

- is an exclusive or problem
- but problem set up ⇒ does not matter

⑪



total prob theorm

$= \sum P(A | X_{999} = j)$

$\simeq = \pi_j$

$= \pi_1 \cdot \frac{2}{3} + \pi_2 \cdot \frac{2}{3} + \pi_3 \cdot \frac{3}{5} + \pi_4 \cdot \frac{3}{5}$

$\approx .6323$

Chap 5 Limit Theorms

11/20

Reading

-assymptotic behavior of seq of RVs

$S_n = X_2 + \dots + X_n$ mean $= \mu$

↑sum of first n RVs, independent var $= \sigma^2$

Concerned w/ properties of S_n as $n \to \infty$

$Var(S_n) = Var(X_1) + \dots + var(X_n) = n\sigma^2$

↑thus distribution can't have a meaningful limit

$$M_n = \frac{X_1 + \dots + X_n}{n} = \frac{S_n}{n}$$

$E[M_n] = \mu$ $var(M_n) = \frac{\sigma^2}{n}$

↑so var ↓ as $n \to \infty$

-the bulk of the distribution of M_n must be very close to μ

-law of large # → the (sample mean) $M_n \xrightarrow{\text{converges}} \mu$ (true mean)

↑RV ↑#

-aka $E[X] = \mu$ - avg of large # of samples drawn from the distribution of X

② Central Limit Theorm

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \qquad E[Z_n] = 0$$

$$\mathrm{var}(Z_n) = 1$$

quantity is intermediate b/w S_n and M_n

- subtract $n\mu$ from S_n = zero mean RV $S_n - n\mu$
- divide by $\sigma\sqrt{n}$ to form RV

(I don't think we have covered in class yet)

- CLT concerned w/ asymptotic shape of distribution of Z_n and asserts that it becomes the st. normal dist

Limit theorms useful for several reasons:

1) Provide interpretation of expectations + probabilities in terms of a long sequence of identical, ind. experiments
2) Allow for an approx. analysis of properties of RVs such as S_n
3) Play a big role in inference + stats

③ 5.1 Markov + Chebyshev Inequalities

- important inequalities
- use mean + variance to draw conclusions on P(certain events)
- useful when exact value / bounds for mean/var are easily computable, but dist X unavalible or hard to calculate

---

Markov inequality

*asserts than if a non negitive RVs has a small mean, then prob that it takes a large value must be small

~~If a RV X can only take~~

$$\boxed{P(X \geq a) \leq \frac{E[X]}{a} \quad \text{for } a > 0}$$

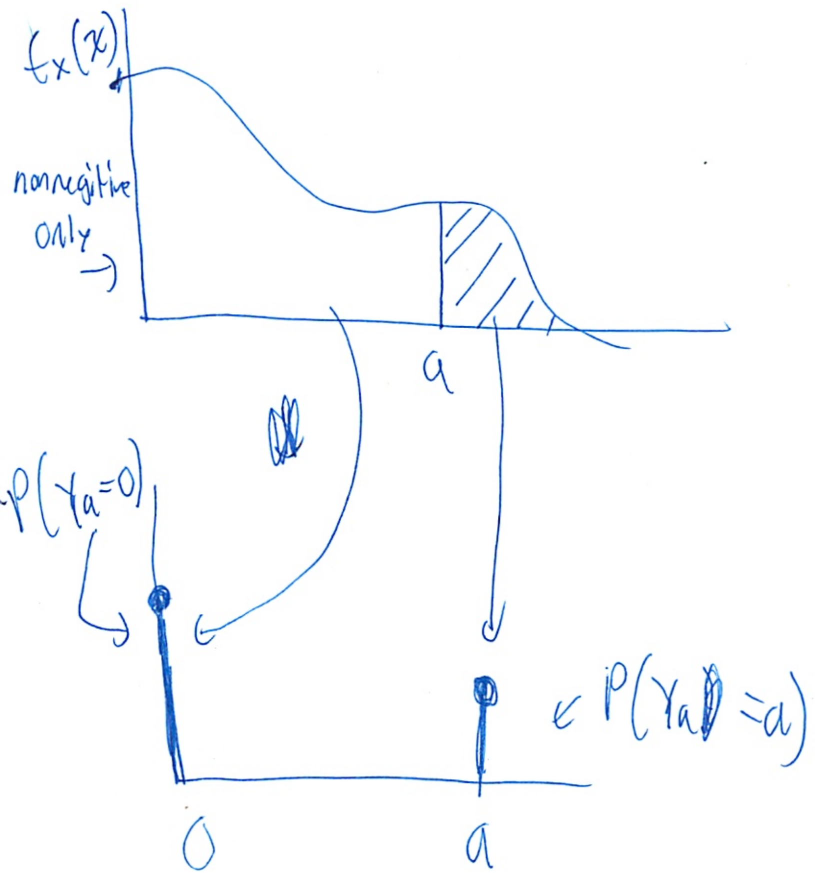
So if we take a RV Y_a,

$$Y_a = \begin{cases} 0 & \text{if } X < a \\ a & \text{if } X \geq a \end{cases}$$

$Y_a \leq X$ always holds

So $E[Y_a] \leq E[X]$

④



Since mass is shifted to the left in both cases, expectation can only decrease $E[X] \ge \underbrace{E[Y_a] = aP(Y_a=a) = aP(X \ge a)}_{=}$

so $aP(X \ge a) \le E[X]$

bounds can be quite lose

⑤

Chebyshev Inequality

- if a RV has a small var, then the prob that it takes a value far from its mean is also small.
- Note: no non-negitive requirement
- (I think I remember from class Markov was moreexact than Chebyshev - Markov built on Chebyshev) — Confirm - later it says other way around

- If X is a RV w/ mean μ and var σ^2

$$\boxed{P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2} \quad \text{for } c > 0}$$

To justify consider nonnegitive RV $(X-\mu)^2$ apply Markov inequality w/ $a = c^2$

$$P((X-\mu)^2 \geq c^2) \leq \frac{E[(X-\mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

or

$$\underset{\text{function?}}{g(x)} = \begin{cases} 0 & \text{if } |x-\mu| < c \\ c^2 & \text{if } |x-\mu| \geq c \end{cases}$$

and now note $(x-\mu)^2 \geq g(x)$ for all x

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_x(x)\,dx \geq \int_{-\infty}^{\infty} g(x) f_x(x)\,dx$$

$$= c^2 P(|X-\mu| \geq c)$$

alt form $c = \underset{\substack{\uparrow\\\text{positive}}}{k}\sigma$

$$P(|X-\mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Thus the P(RV takes a value more than k st.dev away from mean)

is $\underset{\substack{\uparrow\\\text{at most}}}{\leq} \frac{1}{k^2}$

Chebyshev is more powerful than Markov

- bounds more accurate
- b/c also uses info on var of X
- still only rough estimate

examples Upper bounds to Chebyshev Ineq.

X is inside bounds $[a, b]$

Claim $\sigma^2 \leq (b-a)^2/4$

if σ^2 is unknown, use max value $(b-a)^2/4$

So $P(|X-\mu| \geq c) \leq \frac{(b-a)^2}{4c^2}$ for all $c > 0$

Note for constant γ

$$E[(X-\gamma)^2] = E[X^2] - 2E[X]\gamma + \gamma^2$$

So quadratic minimized when $\gamma = E[X]$

So $\sigma^2 = E[(X - E[X])^2] \le E[(X-\gamma)^2]$ for all γ

by letting $\gamma = \frac{a+b}{2}$

$$\sigma^2 \le E\left[\left(X - \frac{a+b}{2}\right)^2\right] = E[(X-a)(X-b)] + \frac{(b-a)^2}{4} \le \frac{(b-a)^2}{4}$$

↑ $(x-a)(x-b) \le 0$ for all X in the range $[a,b]$

bound $\sigma^2 \le \frac{(b-a)^2}{4}$ is conservative -

- but can't be improved w/o extra info

5.2 Weak Law of Large

- sample mean of large iid RVs close to true mean w/ high probability
- $M_n = \frac{X_1 + \cdots + X_n}{n}$

$$E[M_n] = \frac{E[X_1] + \cdots + E[X_n]}{n} = \frac{n\mu}{n} = \mu$$

Since independent

$$var(M_n) = \frac{var(X_1 + \cdots + X_n)}{n} = \frac{var(X_1) + \cdots + var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

⑧ (I am confused on all of the parts + how they fit together)

Then use Chebishev inequality

$$P(|M_n-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \text{for any } \varepsilon > 0$$

- So for any fixed $\varepsilon > 0$ the right hand side $\to 0$ as $n \to \infty$
- get weak law of large #
- true even if X_i have ∞ var, but omitting argument
- only assumption $E[X_i]$ = well defined

Let $X_1, X_2, \ldots$ be iid mean $= \mu$ for $\varepsilon > 0$

$$P(|M_n-\mu| \geq \varepsilon) = P\left(\left|\frac{X_1+\ldots+X_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$

- So for large n, the bulk of the distribution of M_n is concentrated near μ
- So when consider pos interval $[\mu-\varepsilon, \mu+\varepsilon]$ around μ then high prob that M_n will fall in this interval as $n \to \infty$
- $P() \to 1$
- when ε is small, will have to wait longer (bigger n) till it converges

(did a lot of examples in class - have 2, find third)

-error, confidence, probability

5.3 Convergence in Probability

- Weak LL#: "M_n Converges to μ"
- but $M_1, M_2, \ldots$ is seq of RV ~~tha~~ need to clarify meaning of conversion

Convergence of Deterministic Sequence

Let $a_1, a_2, \ldots$ be seq of real #

a be another real #

a_n converges to a

$\Big\} \lim_{n \to \infty} a_n = a$

if for every $\epsilon > 0$ there exists some n_0 such that

$$|a_n - a| \le \epsilon \quad \text{for all } n \ge n_0$$

Convergence in Probability

Let $Y_1, Y_2, \ldots$ be seq of RV (not nessarlly ind.)

a = ~~b~~ real #

Y_n converges to a in prob if for every $\epsilon > 0$ have

$$\lim_{n \to \infty} P(|Y_n - a| \ge \epsilon) = 0$$

⑩

- WLL# simply states that sample mean conveges in prob to the true mean μ
- Chebsher inequality implies that if all Y_n have same mean μ and $\mathrm{var}(Y_n)$ converges to 0, then Y_n converges to μ in prob.
- If RVs $Y_1, Y_2, \dots$ have a PMF or a PDF and converge in prob to a, then by def, "almost all" of the PMF/PDF of Y_n is concentrated within ϵ of a for large values of n.
- For every $\epsilon > 0$, and for any $\sigma > 0$, there exists some n_0 such that $P(|Y_n - a| \geq \epsilon) \leq \sigma$ for all $n \geq n_0$
- ϵ = accuracy level
- σ = Confidence level
- given a certain accuracy + confidence, Y_n will = to a within these levels of accuracy + confidence, provided that n is large enoug

Michael Plasmeier

6.5/10

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2010)

Problem Set 9
Due November 22, 2010

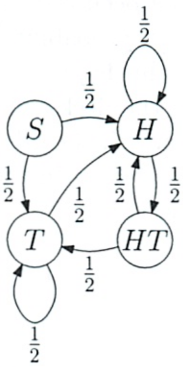
1. Random variable X is uniformly distributed between -1.0 and 1.0. Let $X_1, X_2, \ldots$, be independent identically distributed random variables with the same distribution as X. Determine which, if any, of the following sequences (all with $i = 1, 2, \ldots$) are convergent in probability. Fully justify your answers. Include the limits if they exist.
   (a) $U_i = \dfrac{X_1 + X_2 + \ldots + X_i}{i}$
   (b) $W_i = \max(X_1, \ldots, X_i)$
   (c) $V_i = X_1 \cdot X_2 \cdot \ldots \cdot X_i$

2. Demonstrate that the Chebyshev inequality is tight, that is, for every μ, $\sigma > 0$, and $c \geq \sigma$, construct a random variable X with mean μ and standard deviation σ such that

$$\mathbf{P}(|X - \mu| \geq c) = \frac{\sigma^2}{c^2}$$

   Hint: You should be able to do this with a discrete random variable that takes on only 3 distinct values with nonzero probability.

3. Assume that a fair coin is tossed repeatedly, with the tosses being independent. We want to determine the expected number of tosses necessary to first observe a head directly followed by a tail. To do so, we define a Markov chain with states S, H, T, HT, where S is a starting state, H indicates a head on the current toss, T indicates a tail on the current toss (without heads on the previous toss), and HT indicates heads followed by tails over the last two tosses. This Markov chain is illustrated below:



   We can find the expected number of tosses necessary to first observe a heads directly followed by tails by solving a mean first passage time problem for this Markov chain.

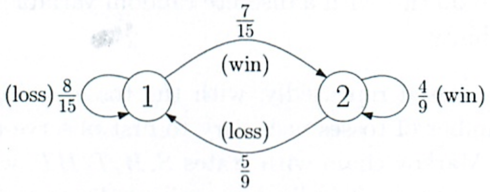
   (a) What is the expected number of tosses necessary to first observe a head directly followed by tails?
   (b) Assuming we have just observed a head followed by a tail, what is the expected number of additional tosses until we again observe a head followed directly by a tail?

Next, we want to answer the same questions for the event tails directly followed by tails. Set up a different Markov chain from which we could calculate the expected number of tosses necessary to first observe tails directly followed by tails.

(c) What is the expected number of tosses necessary to first observe a tail directly followed by a tail?

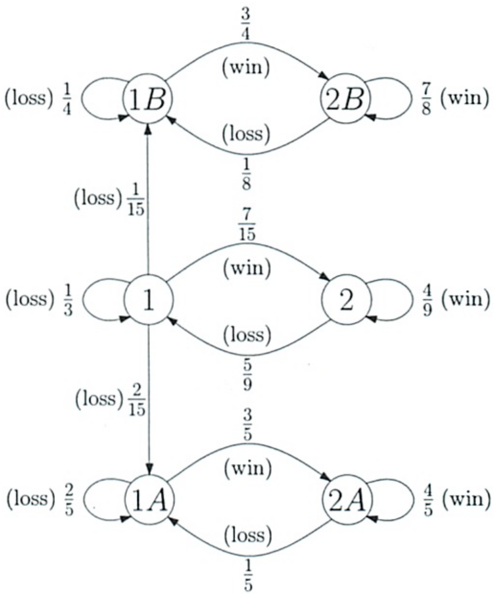
(d) Assuming we have just observed a tail followed by a tail, what is the expected number of additional tosses until we again observe a tail followed directly by a tail? Note that the number of additional tosses could be as little as one, if tails were to come up again.

4. Jack is a gambler who pays for his MIT tuition by spending weekends in Las Vegas. Lately he's been playing 21 at a table that returns cards to the deck and reshuffles them all before each hand. As he has a fixed policy in how he plays, his probability of winning a particular hand remains constant, and is independent of all other hands. There is a wrinkle, however; the dealer switches between two decks (deck #2 is more unfair to Jack than deck #1), depending on whether or not Jack wins. Jack's wins and losses can be modeled via the transitions of the following Markov chain, whose states correspond to the particular deck being used.



(a) What is Jack's long term probability of winning?

Given that Jack loses and the dealer is not occupied with switching decks, with probability $\frac{2}{8}$ the dealer looks away for one second and with probability $\frac{1}{8}$ the dealer looks away for two seconds, independently of everything else. When this happens, Jack secretly inserts additional cards into both of the dealer's decks, transforming the decks into types $1A$ & $2A$ (when he has 1 second) or $1B$ & $2B$ (when he has 2 seconds). Jack slips cards into the decks at most once. The process can be described by the modified Markov chain in the picture. Assume in all future problems that play begins with the dealer using deck #1.



(b) What is the probability of Jack eventually playing with decks $1A$ and $2A$? p asorbtion

(c) What is Jack's long-term probability of winning?

(d) What is the expected time (as in number of hands) until Jack slips additional cards into the deck? expected time to asorbtion

(e) What is the distribution of the number of times that the dealer switches from deck 2 to deck 1?

(f) What is the distribution of the number of wins that Jack has before he slips extra cards into the deck? *Hint*: Note that after some conditioning, we have a geometric number of geometric random variables, all of which are independent.

(g) What is the average net losses (number of losses minus the number of wins, sometimes negative) prior to Jack slipping additional cards into the deck?

(h) Given that after a very long period of time Jack is playing a hand with deck $1A$, what is the approximate probability that his previous hand was played with deck $2A$?

G1[†]. Show the following one-sided version of Chebyshev's inequality:

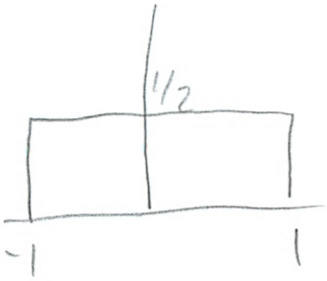
$$\mathbf{P}\left(X - \mu \geq a\right) \leq \frac{\sigma^2}{(\sigma^2 + a^2)}$$

where μ and σ^2 are the mean and variance of X respectively, and $a > 0$. Hint: Start by finding a bound on $\mathbf{P}\left(X - \mu + c \geq a + c\right)$ with $c \geq 0$. Then find the c that 'tightens' your bound.

[†] Required for 6.431; optional challenge problem for 6.041

Michael Plasmeier   6.041 P-Set 9   11/22

1. RV X is uniformly distributed b/w [-1, 1]



Let $X_1, X_2, \ldots$ be iid RVs w/ same distribution as X. Determine which are convergent in probability

a) $U_i = \frac{X_1 + X_2 + \ldots + X_i}{i}$ weak law large #

What does it mean to converge in probability?

- PDF gets concentrated to 1 point
- if $\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0$ as $Y_n \xrightarrow{Prob} a$

+0.5

We did this exact qu in Tutorial 10.

$$E[U_i] = E\left[\frac{X_1 + X_2 + \ldots + X_i}{i}\right] = i \frac{E[U]}{i} = U$$

$$var(U_i) = var\left(\frac{X1 + X2 + \ldots + X_i}{i}\right) = \frac{1}{i^2} \cdot i \, var(X)$$

$$= \frac{Var(X)}{i}$$

What to? (-0.5)

②

$$\sigma_U = \sqrt{var(U)}$$

$$\sigma_U = \frac{\sqrt{var(x)}}{\sqrt{i}}$$

So how does σ change as i ↑

It gets smaller exponentially, so <u>will</u> converge in probability

i hope I interpreted it right

③

b) $W_i = \max(X_1, \ldots, X_i)$

$var(W) = Var(\max(x_1, x_2 \ldots x_i))$

= Variance of largest RV

but all iid, So just

= Var(x)

$\sigma_W = \sqrt{var(x)}$

So what happens as $\lambda \to \infty$?

~~Nothing!~~

~~Does not converge in probability~~

- Something like this in example 5.6 but w/ min

This converges to var(x) which is $\frac{(1--1)^2}{12} = \frac{1}{3}$

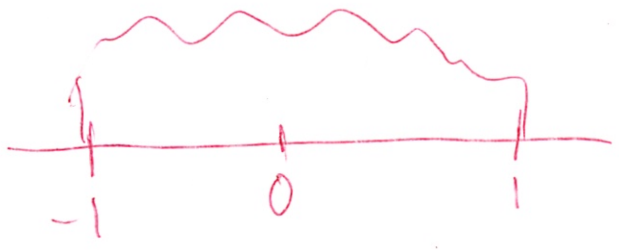
So Converges in prob to $\frac{1}{3}$ (not the mean)

(I don't get the notation they use to prove it)

Converges to 1

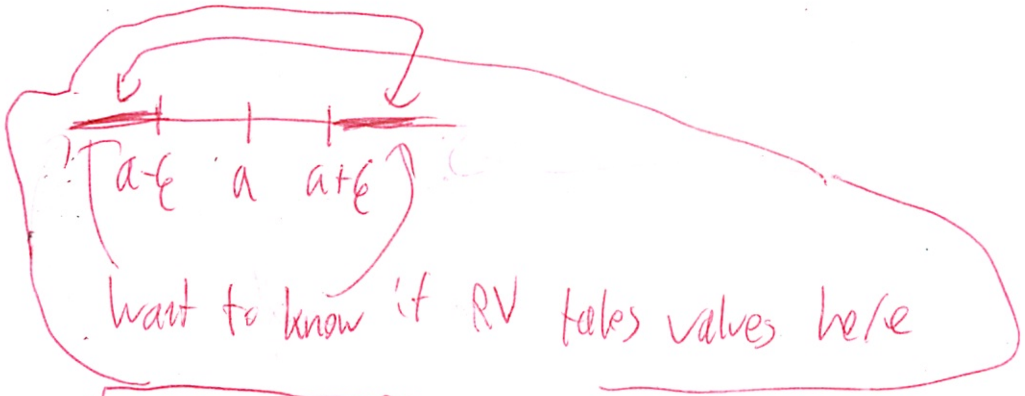
③b

$P(|\max\{X_1, \ldots, X_n\} - 1| \geq \epsilon)$



$P(|\overset{W_2}{\max}\{X_1, X_2\} - 1 \geq \epsilon)$

- don't know which is larger



want to know if RV takes values here

$P(\overbrace{W_2 - 1 \geq \epsilon} \cup \overbrace{W_2 - 1 \leq -\epsilon})$

Means RV is ε more than W_1

- but W_1 is between [-1, 1]

~~$P(W_2 - 1 \geq \epsilon)$~~ $+ P(W_2 - 1 \leq -\epsilon)$

W2 can never be > 1

$P(W_2 - 1 \leq -\epsilon]$

(3c)

$P(W_2 \le 1-\epsilon)$

$a = 1$

So prob

$P(\max\{X_1, X_2\} \le (1-\epsilon))$

$P(X_1 \le (1-\epsilon) \cap (X_2 \le 1-\epsilon))$

can bound - bound what?

$\le$ something

+1

④

c) $V_i = X_1 \cdot X_2 \cdot \ldots X_i$

$Var(V_i) = Var(X_1 \cdot X_2 \cdot \ldots \cdot X_i)$

? so what is the rule again?

$Var(x+y) = Var(x) + Var(y)$

but this is multiply

guessing

~~$= Var(x_1) \cdot Var(x_2) \cdot \ldots \cdot Var(x_n)$~~

~~$= Var(x)^i$~~

~~As i increases, Varience increases exponentially,~~

~~So no it diverges in probability~~

Or Recitation 20 #3 b which is same thing

most certainly converges

? so all 3 converges -confused- need to see pic

Is it about var? -yeah var ↓ as i↑

? Depends on $Var(x^i)$ or $var(x)^i$

?oh duh it should be this

So will absolutly converge in prob to mean

OH! They do all converge

4b

Chebsley $P(|X_n - \mu_{X_n}| \geq c) \leq \frac{\sigma_{X_n}{}^2}{c^2}$

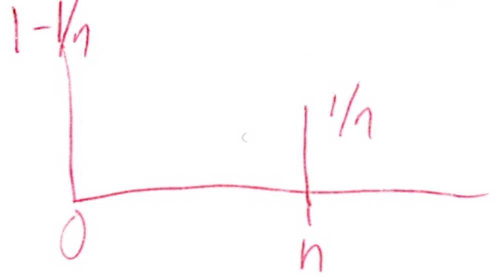
as $X_n \to \infty$

then σ goes to 0

So then . . .

$X_n \to$ converges to mean

~~So~~ Converge in prob does not ressarily converge to mean



mean = 1

Converges in prob to 0

To check convergance in prob

$$P(|X_n - \underset{\uparrow}{a}| \geq \epsilon) \to 0 \quad \text{as } n \to \infty \quad \forall \epsilon > 0$$

what it converges to

- will be in terms of X_n, ϵ
- take limit to see if converge

(4c)

Chebyshev helped in 1a on p-set w/ weak LLN

But not always will it work

- no inequality will work

Use st. conversion case like we did in OH for

1b and started w/ 1c

↳ next pg

(4d)

OH: Similar to 1b

$P(|X_1 \circ X_2 \circ \ldots \circ X_i - a| \geq \epsilon)$

- think it converges to 0
- always fractional # or 1
- gets smaller + smaller

So how write this math?

But just ≤ 1, not helpful

But also less than $\min \{X_1, X_2, \ldots X_i\}$

Chebyshev [illegible] helpful

+1

2.5/3

⑮

2. Demonstrate that a Chebyshev inequality is tight.

So for every μ, $\sigma > 0$ $c \geq \sigma$

Construct a RV X w/ mean μ and st. dev σ such that

$$P(|X-\mu| \geq c) = \frac{\sigma^2}{c^2}$$

Hint: You should be able to do this w/ a discrete RV that takes on only 3 distinct values

? Guess + check

?? unsure on this

mean = 1

$\text{Var} = \frac{1}{3}(0)^2 + \frac{1}{3}(1)^2 + \frac{1}{3}(2)^2 \quad -(1)^2 = \frac{2}{3}$

$\sigma = \sqrt{2/3} = .816$

$c =$? what - just find that is larger

$$P(|X-1| \geq c) = \frac{2/3}{c^2}$$

$$\begin{cases} X=2 \to 1 \geq c \\ \ \ =1 \to 0 \geq c \\ \ \ =0 \to 1 \geq c \end{cases} = \frac{2/3}{c^2}$$

the p that is $> C$ = $\begin{cases} 2/3 \text{ true } \epsilon \\ 1/3 \text{ false,} \end{cases}$ for $0 < c < 1$

$\{ 1$ true for $c < 0$

$$\frac{2}{3} = \frac{\frac{2}{3}}{c^2} \quad 0 < c < 1$$

⑥

C=1 for that to be true

Lets check

$$P(|X-1| \geq 1) = \frac{2/3}{1} = \frac{2}{3}$$

$$\begin{cases} 2 \to 1 \to \text{true} & P = 1/3 \\ 1 \to 0 \to \text{false} & P = 1/3 \\ 0 \to 1 \to \text{true} & P = 1/3 \end{cases}$$

but now $P() = \frac{2}{3} = \frac{2}{3}$

✓

So does that work?

But how does demonstrate tightness?

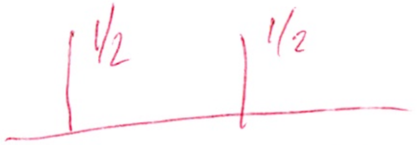
What does it have to do w/ Chebishev?

didn't do upper bands

didn't do error/confidence

These inequality qs hard - should have asked easier ones

(6b) OH

3. 

Discrete w/ 2 spikes

Maximize var if $\frac{1}{2}$ on each spike

Chebshev

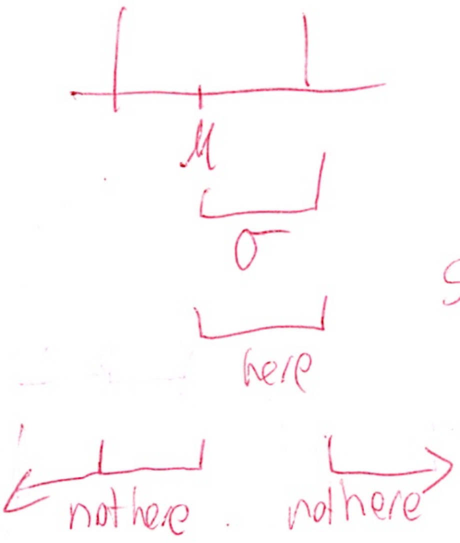
$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

We know σ is 

Prob $x-\mu$ = any c inside that σ interval

But c can be $> \sigma$, then "this" $=0$

Satisfied inequality, but have not met w/ equality



So true if c

(6C)

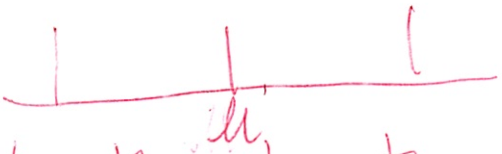
There is nothing wrong w/ picking C

With any σ, c, There is some RV where this holds in equality

So RV wer are finding <u>should</u> depend on C

Hint: 3 spikes

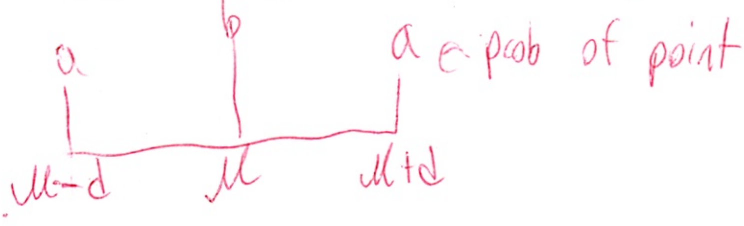
Maximize var by putting 2 as far apart as possible, one in the middle.
^ the mean



Make the ends the same height, so mean stays in middle.

Want to max $(\text{distance from mean})^2 = \sigma^2 = \text{var}$

What height should middle/ends be



$\sigma^2 = 2ad^2$

(6d)

c should represent some distance from mean.

Where plug it in?

$d = kc$

?constant $k > 1$

$\sigma^2 = 2a(kc)^2 \overset{?}{=} var()$

$b = 1 - 2a$ (so adds to 1)

Now need to compute prob more than c out from mean (left hand side)

Solve the eq for k

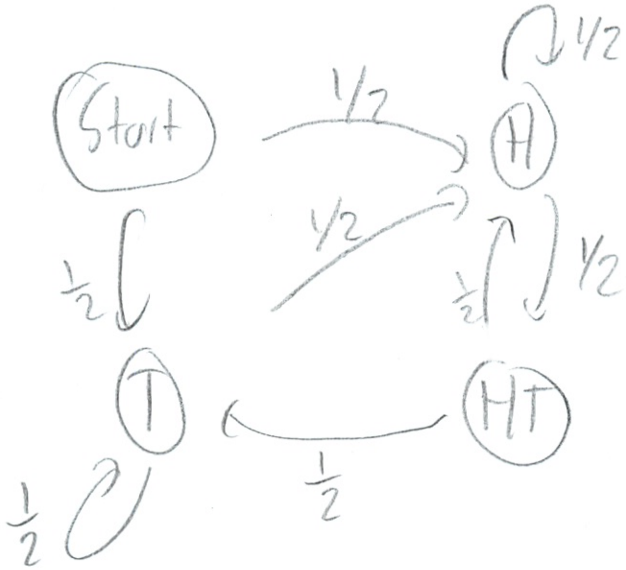
then everything should come out in terms of c, μ, σ

Solve left in terms of c

$P(|x - \mu| \geq c)$

??

3. Assume a fair coin is tossed repeatedly w/tosses naturally being indep. We want to determine the E[# of tosses HT]. Use markov chain.



?: why no connection H->T - oh duh that is our pattern

? Could we have use Markoff chains to solve problems from earlier in the semester?

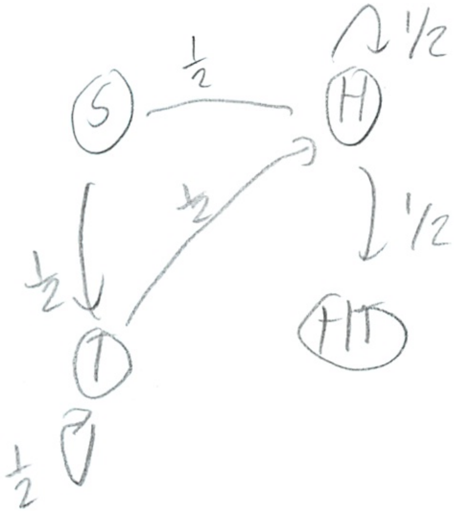
We can find the answer by using mean 1st passage time for this Markov chain.

a) What is E[ ]?

- So this is expected time to asorbtion

$$\underset{\substack{\uparrow \\ \text{start state}}}{\mu_i} = 1 + \sum_{j=1}^{m} P_{ij} \underset{\substack{\uparrow \\ \text{all of the leaving arrows}}}{\mu_j} \text{ for all non asorbing } i$$

but that i is the starting state — we only have one redraw making aserbing state recurrent



? is it simpler to do formula, or just looking at it

$\mu_{HT} = 0$ ← if they start there

but told start at S, so

↙ all of the arrows leaving

$$\mu_S = 1 + \frac{1}{2}\mu_T + \frac{1}{2}\mu_H$$

↑ ↑ now need to find these

$$\mu_T = 1 + \frac{1}{2}\mu_T + \frac{1}{2}\mu_H$$

$$\mu_H = 1 + \frac{1}{2}\mu_H + \frac{1}{2}\mu_{HT}$$

↑ now solve

↑0 we know remember

⑨

For this one will try to solve by hand

$\frac{1}{2}\mu_T = 1 + \frac{1}{2}\mu_H$ $\frac{1}{2}\mu_H = 1 + 0$

$\mu_T = 2 + \mu_H$ $\mu_H = 2$

$$\mu_s = 1 + \frac{1}{2}(2+2) + \frac{1}{2}\cdot 2$$

$$= 4$$

Confirm w/ Wollfram alpha

$\mu_s = 4$ $\mu_T = 4$ $\mu_H = 2$ $\mu_{HT} = 0$

so ✓ out

b) Assuming we just saw HT what is the expected # of additional tosses until HT

-(does the fact that HT is 2 tosses change anything? don't think - every transition is a coin flip)

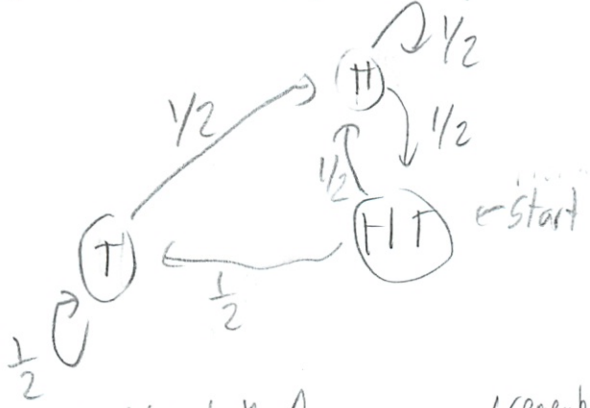
So this is mean recurrence time

$$t_s^* = E[\min\{n \geq 1 \text{ such that } X_n = s \mid X_0 = s\}]$$

$$t_s^* = 1 + \sum_j \underset{\uparrow \text{leaving } s}{p_{sj}} t_j$$

⑩

So this one is similar, except with a different picture



↓ dont forget the 1 ↓ remember leaving

~~$t_{HT} = 1 + \frac{1}{2} t_H + \frac{1}{2} t_T$~~ ← new

~~$t_T = 1 + \frac{1}{2} t_T + \frac{1}{2} t_H$~~

~~$t_H = 1 + \frac{1}{2} t_H + \frac{1}{2} t_{HT}$~~

} same as before

Solve (wolfram alpha) 3 eq, 3 unknowns

- no solutions

- Oh if see example 7.13 - Use the old values

↖ confused different notation used in recitation 19

$t_{HT} = 0 \quad t_T = 4 \quad t_H = 2$

$t_{HT}^{*} = 1 + \frac{1}{2} t_H + \frac{1}{2} t_T$

↑ new ↑ use previously

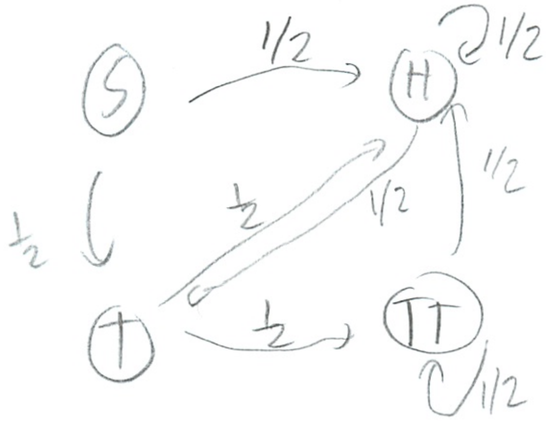
$t_{HT}^{*} = 1 + \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 4$

$= 1 + 1 + 2$

$= 4$

⑪

Now same question for TT. New markov chain



- kinda like 6.01 state machines

c) E[# of tosses to TT asorbtion | start at S] =

$$\mu_S = 1 + \frac{1}{2}\mu_H + \frac{1}{2}\mu_T$$

$$\mu_H = 1 + \frac{1}{2}\mu_H + \frac{1}{2}\mu_T$$

$$\mu_T = 1 + \frac{1}{2}\mu_H + \frac{1}{2}\mu_{TT}$$

$\mu_{TT} = 0$ treating as recurrent state

Wolfram alpha solve

$\mu_S = 6$ $\mu_h = 6$ $\mu_t = 4$ $\mu_{TT} = 0$

note recitation 11 notation issue

$\mu = t$

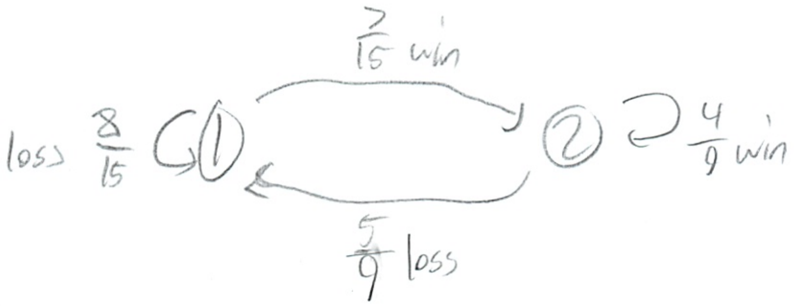
d) Assuming we just saw TT, what is expected #
of additional tosses untill TT again?

- can be as little as 1, if T again
  - which I did account for on diagram

$$t^*_{TT} = 1 + \frac{1}{2} t_{TT} + \frac{1}{2} t_{H}$$
$$= 1 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 4$$
$$= 1 + 2$$
$$= 3$$

⑬

4. Jack is a gambler who pays for his MIT tuition by spending weekends in Vegas playing Blackjack. Table reshuffles card after every hand. His P(winning) remained constant + was ind. from other hands, but dealer shifts decks depending if he wins



a) What is Jack's long term prob of winning?

- ? so this is transition probability
- No, we can define state 1 as loss
  2 as win
- So want long term prob/freq of state 2 = π_2
- So balance eq

$$\pi_1 = \frac{8}{15}\pi_1 + \frac{5}{9}\pi_2$$

? all of the ones towards this

$$\pi_2 = \frac{7}{15}\pi_1 + \frac{4}{9}\pi_2$$

$$\pi_1 + \pi_2 = 1$$

(14).

Solve w/ wolfram alpha

$\pi_1 = \frac{25}{46}$ $\pi_2 = \frac{21}{46}$

↑ long term prob of winning

+1

Given that Jack loses and dealer is looking away w/ prob $\frac{2}{8}$ 1 sec, $\frac{1}{8}$ 2 sec ind.

Jack inserts extra cards into decks 1A + 2A (when 1sec) or 1B + 2B (when has 2 sec). Slips cards in once

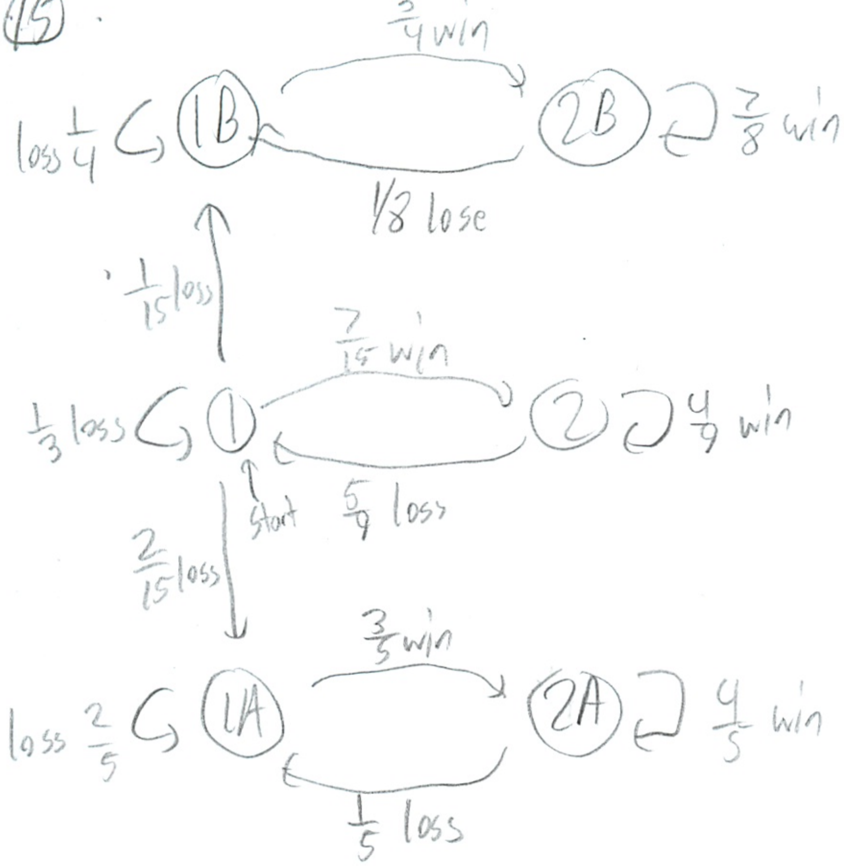
Dealer always starts w/ deck 1.

- ? so does dealer still switch decks?
- if he has inserted 1 card now will he ever insert the other?
- if he has messed up deck 1, will he mess up deck 2

- ah I see they provide us with a nice picture next pg

⑮.

I really like Markov Chains



b) What is the prob of Jack eventually playing w/ 1A, 2A?

- So this is P(asorbed by A recurrent class)

- So $a_1 + a_2$ for $s = 1A, 2A$

$a_{1B} = a_{2B} = 0$

$a_{1A} = a_{1B} = 1$

) not starting here but need to write in case others mention it

$$a_1 = \frac{1}{3}a_1 + \frac{7}{15}a_2 + \frac{1}{15}a_{1B} + \frac{2}{15}a_{1A}$$

↑ leaving state ↑ 0 ↑ 1

$$a_2 = \frac{4}{9}a_2 + \frac{5}{4}a_1$$

Solve w/ wolfram alpha

(16)

Gives two answers - but we only care for positive answers

$a_1 = \frac{2}{3}$

$a_2 = \frac{2}{3}$

} makes sense that is same

+0.5

c) What is Jack's long term prob of winning

- so this is $\pi_2 + \pi_{2A} + \pi_{2B}$

$\uparrow$ 0 since transient

$\pi_{2A} =$ prob of being asorbed in a $\cdot$ long term prob of 2A inside 1A, 2A class

$\uparrow$ last ans

$\pi_{1A} = \frac{2}{5}\pi_{1A} + \frac{1}{5}\pi_{2A}$

$\uparrow$ coming to it

$\pi_{1A} + \pi_{2A} = 1$

$\pi_{2A} = \frac{3}{5}\pi_{1A} + \frac{4}{5}\pi_{2A}$

Solve w/ wolfram alpha

$\pi_{1A} = \frac{1}{4}$ $\pi_{2A} = \frac{3}{4}$

$\pi_{2A} = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$

⑰

π_{2B} = prob of being absorbed in B · prob of 2B inside 1B, 2B class

↑ is this $1-\frac{2}{3}$?

verify by doing from scratch

↓

$$a_{1B} = a_{2B} = 1$$

$$a_{1A} = a_{2A} = 1$$

$$a_1 = \underset{\uparrow \text{leaving}}{\frac{1}{3} a_1} + \frac{7}{15} a_2 + \frac{1}{15} \underset{\uparrow 1}{a_{1B}} + \frac{2}{15} \underset{\uparrow 0}{a_{1A}}$$

$$a_2 = \frac{4}{9} a_2 + \frac{5}{9} a_1$$

$$a_1 = \frac{1}{3}$$

$$a_2 = \frac{1}{3}$$

✓ verifies

now other part: prob of 2B inside 1B, 2B class

$$\pi_{1B} = \frac{1}{4} \overset{\text{arriving}\downarrow}{\pi_{1B}} + \frac{1}{8} \pi_{2B} \qquad \pi_{1B} + \pi_{2B} = 1$$

$$\pi_{2B} = \frac{3}{4} \pi_{1B} + \frac{7}{8} \pi_{2B}$$

$$\pi_{1B} = \frac{1}{7} \qquad \pi_{2B} = \frac{6}{7}$$

So π_{2B} complete system $= \frac{1}{3} \cdot \frac{6}{7} = \frac{2}{7}$

So add $0 + \frac{1}{2} + \frac{2}{7} = \frac{11}{14}$ +1

(18).

d) E[# hands until slips extra cards into deck]

- So E[time till either absorbtion]
- will use notation t, not μ

$t_{1B} = t_{2B} = t_{1A} = t_{2A} = 0$ recurrent states

$$t_1 = 1 + \underset{\uparrow \text{ time leaving}}{\frac{1}{3}t_1} + \frac{7}{15}t_2 + \frac{1}{15}\underset{\uparrow 0}{t_{1B}} + \frac{2}{15}\underset{0}{t_{1A}}$$

$$t_2 = 1 + \frac{4}{9}t_2 + \frac{5}{9}t_1$$

Solve

$$t_1 = \frac{46}{5} = 9.2 \quad \underset{\uparrow \text{ told starts here}}{t_2 = 11}$$ ← depends where you start

+0.5

e) What is distribution of # of time dealer switches from 2 to 1 ↑?? does this just mean probability distribution?

- Short term

$p_{21} = \frac{5}{9}$

$r_{ij}(n)$ ← related to this right?, want dist for each n

Or is this mean recurrence time
↑ not mean though

- what is this?
- look in book for order of material
- likely new material
- like slide 3
- or some sort of geometric exponential
- ? like in Tutorial 10?

⑲

Geometric w/ param $(5/9)$

↳ like what we did in tutorial

is # of trials till 1st sucess

w/ sucess $= \frac{5}{9} = p_{21}$

failure $\frac{4}{9} = p_{22}$

$E[X] = \frac{1}{p}$, $\quad var(x) = \frac{1-p}{p^2}$ $\quad p = \frac{5}{9}$

But this is starting at state 2

Really we have to include periodic, asorbing

Is this $r_{21}(n)$?

What if do manually?

$n=1 \rightarrow \frac{7}{15} \cdot 0$

$n=2 \rightarrow \frac{7}{15} \cdot \frac{5}{9}$

$n=3 \rightarrow \frac{7}{15} \cdot \frac{4}{9} \cdot \frac{5}{9}$

$n=4 \Rightarrow \frac{7}{15} \cdot \frac{5}{9} \cdot \frac{7}{15} \cdot \frac{5}{9} + \frac{7}{15} \left(\frac{4}{9}\right)^2 \cdot \frac{5}{9}$

$n=5 \Rightarrow \frac{7}{15} \cdot \frac{5}{9} \cdot \frac{7}{15} \cdot \frac{4}{9} \cdot \frac{5}{9} + \frac{7}{15} \left(\frac{4}{9}\right)^3 \cdot \frac{5}{9}$

asorbing included w/ 7/15

So $n =$ even $\rightarrow \left(\frac{7}{15} \cdot \frac{5}{9}\right)^{n/2} + \frac{7}{15}\left(\frac{4}{9}\right)^{n-2} \cdot \frac{5}{9}$

$n =$ odd $\rightarrow \xcancel{\left(\frac{7}{15} \cdot \frac{5}{9}\right)^{\frac{n-3}{2}}} + \left(\frac{7}{15} \cdot \frac{4}{9} \cdot \frac{5}{9}\right)^{\frac{n}{3}} + \frac{7}{15}\left(\frac{4}{9}\right)^{n-2} \cdot \frac{5}{9}$

Still think I am missing stuff

-ask in OH

OH: Consier getting rid...

$= P(\text{go to } 2 \mid \text{given leave } 1) = \frac{7/15}{10/15} = \frac{7}{10} = .7$

↑ Must go back then ↑ remove self transition

Until asorbtion

- prob, not E[]

Want dist of times 1 to 2

-know prob

Do something over + over till something happens

- ? geometric

- $\left(\frac{7}{10}\right)^k \underbrace{+\frac{3}{10}}$ k=0,1,2,...

but must add this so almost geometric

Q: How can you ignore self transition

- self transition does not add anything - like nothing happened

Not # trials - a lot more complex -> 2 variables (what I tried)

(21).

f) What is dist of # of wins Jack has before he slips extra cards into deck?

Hint: After some conditioning have a geometric # of geometric RVs - all of which are ind.

- similar to previous
- do conditioning
- win = transition into 2
  - don't think
- given what → is that a geometric

How many times enter 2 in a cycle

$\# $ times until sucess, $p(\text{sucess}) = \frac{5}{4}$
- then out of 2

+1

$\#$ times enter 2 is geometric

↓ given that I am here

how many times stay(reenter) 2
geometric$\left(\frac{5}{9}\right)$

- $\#$ times transition $1 \to 2$ so ans from e

$$\left(\frac{7}{10}\right)^{k} + \frac{3}{10}$$

What would need to condition on

\# times transition 1 → 2

← so can make it actually geometric not shifted geometric

So conditioned on ↗ would be geometric

Given I am not asorbed in 1st time

End = geometric sum of geometric RVs

Lets say after W = # wins before asorbtion

$P_W(w) = P(\text{absorbed in 1st state change } A) \cdot P_{W|A}(w)$ ← 0

$+ P(\text{Not asorbed 1st time } A^c)\, P_{W|A^c}(w)$

↗ but this is PMF of Sum of geometric # of geometric RV - Bernoulli processes

but PMF $\begin{cases} \text{prob} = 1 \text{ when } w = 0 \\ 0 \text{ when } w \neq 0 \end{cases}$

interarrival times = geometric

(23)

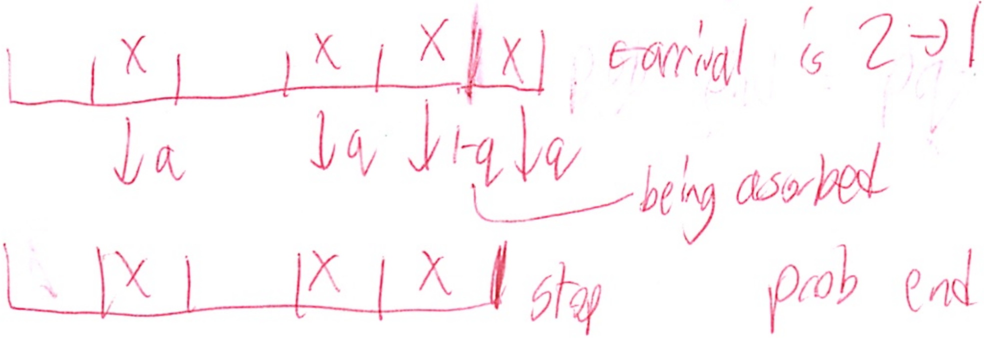
Sum intercarrival times

= Sum of geometric # of geometric RV

But splliting

w/ prob q put in process

prob $1-q$ stop adding



prob end = pq

So easy to add back together

$$P_w(w) = \frac{3}{15} \cdot \begin{cases} 1 & \text{when } w=0 \\ 0 & \text{when } w \neq 0 \end{cases} + \frac{12}{15}\left(q \cdot \left[\left(\frac{7}{10}\right)^k + \frac{3}{10}\right]\right)$$

(−1)

㉔

g) What is the average net loss (losses − wins, can be −) prior to Jack slipping in extra cards ?

- average = mean = expected value
- remember expected time to asorbtion was 9.2 (ans from d)
- So what is this exactly - something w/ π ?
  - but its not long term
- ? something connected to e and f ?

\# wins = E[ans to last qu]

\# losses = E[1 − previous − asobtion]

$$E[win] = \frac{3}{15} + \frac{12}{15}\left(q \cdot \left[\frac{1-7/10}{\left(\frac{7}{10}\right)^2} + \frac{3}{10}\right]\right)$$

$$E[loss] = 1 - \frac{12}{15}\left(q \cdot \left[\frac{1-7/10}{\left(\frac{7}{10}\right)^2} + \frac{3}{10}\right]\right)$$

$$E[net\ loss] = 1 + \frac{3}{15}$$

(−1)

don't care anymore

㉕.

h) Given that after a very long period of time Jack is playing a hand w/ deck 1A, what is approx prob that previous hand was played w/ 2A

?, so is this just $P_{2A\,1A} = \frac{1}{5}$

- seems too easy...
- no it must be prob that came from that

$$\frac{1/5}{\frac{1}{5}+\frac{2}{5}} = \frac{1}{3}$$

(-1)

4/7

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2010)

Problem Set 9 Solutions

1. (a) Yes, to 0. Applying the weak law of large numbers, we have

$$\mathbf{P}(|U_i - \mu| > \epsilon) \to 0 \text{ as } i \to \infty, \text{ for all } \epsilon > 0$$

Here $\mu = 0$ since $X_i \sim U(-1.0, 1.0)$.

(b) Yes, to 1. Since $W_i \leq 1$, we have for $\epsilon > 0$,

$$\begin{aligned}\lim_{i\to\infty} \mathbf{P}(|W_i - 1| > \epsilon) &= \lim_{i\to\infty} \mathbf{P}(\max\{X_1, \cdots, X_i\} < 1-\epsilon) \\ &= \lim_{i\to\infty} \mathbf{P}(X_1 < 1-\epsilon) \cdots \mathbf{P}(X_i < 1-\epsilon)\} \\ &= \lim_{i\to\infty} (1 - \frac{\epsilon}{2})^i \\ &= 0.\end{aligned}$$

(c) Yes, to 0.

$|V_n| \leq \min\{|X_1|, |X_2|, \cdots, |X_n|\}$
but $\min\{|X_1|, |X_2|, \cdots, |X_n|\}$ converges to 0 in probability. So, since $|V_n| \geq 0$, $|V_n|$ converges to 0 in probability. To see why $\min\{|X_1|, |X_2|, \cdots, |X_n|\}$ converges to 0 in probability, note that:

$$\begin{aligned}\lim_{i\to\infty} \mathbf{P}(|\min\{|X_1|, \cdots, |X_i|\} - 0| > \epsilon) &= \lim_{i\to\infty} \mathbf{P}(\min\{|X_1|, \cdots, |X_i|\} > \epsilon) \\ &= \lim_{i\to\infty} \mathbf{P}(|X_1| > \epsilon) \cdot \mathbf{P}(|X_2| > \epsilon) \cdots \mathbf{P}(|X_i| > \epsilon) \\ &= \lim_{i\to\infty} (1+\epsilon)^i \text{ since } |X_i| \text{ is uniform between 0 and 1} \\ &= 0.\end{aligned}$$

2. Consider a random variable X with PMF

$$p_X(x) = \begin{cases} p, & \text{if } x = \mu - c; \\ p, & \text{if } x = \mu + c; \\ 1-2p, & \text{if } x = \mu. \end{cases}$$

The mean of X is μ, and the variance of X is $2pc^2$. To make the variance equal σ^2, set $p = \frac{\sigma^2}{2c^2}$. For this random variable, we have

$$\mathbf{P}(|X - \mu| \geq c) = 2p = \frac{\sigma^2}{c^2},$$

and therefore the Chebyshev inequality is tight.

3. (a) Let t_i be the expected time until the state HT is reached, starting in state i, i.e., the mean first passage time to reach state HT starting in state i. Note that t_S is the expected number of tosses until first observing heads directly followed by tails. We have,

$$\begin{aligned} t_S &= 1 + \frac{1}{2}t_H + \frac{1}{2}t_T \\ t_T &= 1 + \frac{1}{2}t_H + \frac{1}{2}t_T \\ t_H &= 1 + \frac{1}{2}t_H \end{aligned}$$

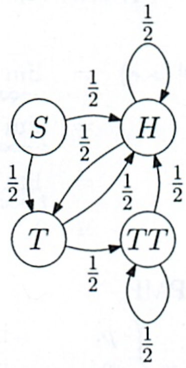
and by solving these equations, we find that the expected number of tosses until first observing heads directly followed by tails is

$$t_S = 4\,.$$

(b) To find the expected number of additional tosses necessary to again observe heads followed by tails, we recognize that this is the mean recurrence time t^*_{HT} of state HT. This can be determined as

$$\begin{aligned} t^*_{HT} &= 1 + p_{HT,H} t_H + p_{HT,T} t_T \\ &= 1 + \frac{1}{2}\cdot 2 + \frac{1}{2}\cdot 4 \\ &= 4\,. \end{aligned}$$

(c) Let's consider a Markov chain with states S, H, T, TT, where S is a starting state, H indicates heads on the current toss, T indicates tails on the current toss (without tails on the previous toss), and TT indicates tails over the last two tosses. The transition probabilities for this Markov chain are illustrated below in the state transition diagram:



Let t_i be the expected time until the state TT is reached, starting in state i, i.e., the mean first passage time to reach state TT starting in state i. Note that t_S is the expected number of tosses until first observing tails directly followed by tails. We have,

$$\begin{aligned} t_S &= 1 + \frac{1}{2} t_H + \frac{1}{2} t_T \\ t_T &= 1 + \frac{1}{2} t_H \\ t_H &= 1 + \frac{1}{2} t_H + \frac{1}{2} t_T \end{aligned}$$

and by solving these equations, we find that the expected number of tosses until first observing two consecutive tails is

$$t_S = 6\,.$$

(d) To find the expected number of additional tosses necessary to again observe heads followed by tails, we recognize that this is the mean recurrence time t^*_{TT} of state TT. This can be determined as

$$\begin{aligned} t_{TT}^{*} &= 1 + p_{TT,H} t_H + p_{TT,TT} t_T T \\ &= 1 + \frac{1}{2} \cdot 6 + \frac{1}{2} \cdot 0 \\ &= 4 \, . \end{aligned}$$

It may be surprising that the average number of tosses until the first two consecutive tails is greater than the average number of tosses until heads is directly followed by tails, considering that the mean recurrence time between pairs of tosses with heads directly followed by tails equals the mean recurrence time between pairs of tosses that are both tails (or equivalently, the long-term frequency of pairs of tosses with heads followed by tails equals the long-term frequency of pairs of tosses with two consecutive tails[1]). This is a start-up artifact. Note that the distribution of the first passage time to reach state HT (or TT) starting in state S is the same as the conditional distribution of the recurrence time of state HT (or TT), given that it is greater than 1. Although in both cases the *expected values* of the recurrence times are equal (this is what parts (b) and (d) tell us), the conditional expected values of the recurrence time given that it is greater than 1 is not the same in both cases (possible, because the unconditional distributions are not equal).

4. (a) The long-term frequency of winning can be found as sum of the long-term frequency of transitions from 1 to 2 and 2 to 2. These can be found from the steady-state probabilities π_1 and π_2, which are known to exist as the chain is aperiodic and recurrent. The local balance and normalization equations are as follows:

$$\begin{aligned} \frac{7}{15}\pi_1 &= \frac{5}{9}\pi_2 \, , \\ \pi_1 + \pi_2 &= 1 \, . \end{aligned}$$

Solving these we obtain,

$$\pi_1 = \frac{25}{46} \approx 0.54, \ \pi_2 = \frac{21}{46} \approx 0.46 \, .$$

The probability of winning, which is the long-term frequency of the transitions from 1 to 2 and 2 to 2, can now be found as

$$\mathbf{P}(\text{winning}) = \pi_1 p_{12} + \pi_2 p_{22} = \frac{25}{46}\frac{7}{15} + \frac{21}{46}\frac{4}{9} = \frac{21}{46} \approx 0.46 \, .$$

Note that from the balance equation for state 2,

$$\pi_2 = \pi_1 p_{12} + \pi_2 p_{22} \, ,$$

the long-term probability of winning always equals π_2.

(b) This question is one of determining the probability of absorption into the recurrent class $\{1A, 2A\}$. This probability of absorption can be found by recognizing that it will be the ratio of probabilities

$$\frac{p_{1,1A}}{p_{1,1A} + p_{1,1B}} = \frac{\frac{2}{15}}{\frac{2}{15} + \frac{1}{15}} = \frac{2}{3} \, .$$

---

[1] See problem 7.34 on page 399 of the text for a detailed explanation of this correspondence between mean recurrence times and steady-state probabilities.

More methodically, if we define a_i as the probability of being absorbed into the class $\{1A, 2A\}$, starting in state i, we can solve for the a_i by solving the system of equations

$$\begin{aligned} a_1 &= p_{1,1A} + p_{11}a_1 + p_{12}a_2 \\ &= \frac{2}{15} + \frac{1}{3}a_1 + \frac{7}{15}a_2 \\ a_2 &= p_{21}a_1 + p_{22}a_2 \\ &= \frac{5}{9}a_1 + \frac{4}{9}a_2 \, , \end{aligned}$$

from which we determine that $a_1 = \frac{p_{1,1A}}{p_{1,1A}+p_{1,1B}} = \frac{2}{3}$.

(c) Let A, B be the events that Jack eventually plays with decks $1A$ & $2A$, $1B$ & $2B$, respectively, when starting in state 1. From part (b), we know that $\mathbf{P}(A) = a_1 = \frac{2}{3}$ and $\mathbf{P}(B) = 1 - a_1 = \frac{1}{3}$. The probability of winning can be determined as

$$\mathbf{P}(\text{winning}) = \mathbf{P}(\text{winning}|A)\mathbf{P}(A) + \mathbf{P}(\text{winning}|B)\mathbf{P}(B) \, .$$

By considering the corresponding the appropriate recurrent class and solving a problem similar to part (a), $\mathbf{P}(\text{winning}|A)$ and $\mathbf{P}(\text{winning}|B)$ can be determined; in these cases, the steady-state probabilities of each recurrent class are defined under the assumption of being absorbed into that particular recurrent class. Let's begin with $\mathbf{P}(\text{winning}|A)$. The local balance and normalization equations for the recurrent class $\{1A, 2A\}$ are

$$\begin{aligned} \frac{3}{5}\pi_{1A} &= \frac{1}{5}\pi_{2A} \, , \\ \pi_{1A} + \pi_{2A} &= 1 \, . \end{aligned}$$

Solving these we obtain,

$$\pi_{1A} = \frac{1}{4}, \; \pi_{2A} = \frac{3}{4} \, ,$$

and hence conclude that

$$\mathbf{P}(\text{winning}|A) = p_{1A,2A}\pi_{1A} + p_{2A,2A}\pi_{2A} = \pi_{2A} = \frac{3}{4} \, .$$

Similarly, the local balance and normalization equations for the recurrent class $\{1B, 2B\}$ are

$$\begin{aligned} \frac{3}{4}\pi_{1B} &= \frac{1}{8}\pi_{2B} \, , \\ \pi_{1B} + \pi_{2B} &= 1 \, . \end{aligned}$$

Solving these we obtain,

$$\pi_{1B} = \frac{1}{7}, \; \pi_{2B} = \frac{6}{7} \, ,$$

and hence conclude that

$$\mathbf{P}(\text{winning}|B) = p_{1B,2B}\pi_{1B} + p_{2B,2B}\pi_{2B} = \pi_{2B} = \frac{6}{7} \, .$$

Putting these pieces together, we have that

$$\begin{aligned}\mathbf{P}(\text{winning}) &= \mathbf{P}(\text{winning}|A)\mathbf{P}(A) + \mathbf{P}(\text{winning}|B)\mathbf{P}(B) \\ &= \frac{3}{4}\cdot\frac{2}{3} + \frac{6}{7}\cdot\frac{1}{3} \\ &= \frac{11}{14} \approx 0.79\ ,\end{aligned}$$

meaning that Jack substantially increases the odds to his favor by slipping additional cards into the decks.

(d) The expected time until Jack slips cards into the deck is the same as the expected time until the Markov chain enters a recurrent state. Let μ_i be the expected amount of time until a recurrent state is reached from state i. We have the equations

$$\mu_1 = 1 + p_{11}\mu_1 + p_{12}\mu_2 = 1 + \frac{1}{3}\mu_1 + \frac{7}{15}\mu_2$$
$$\mu_2 = 1 + p_{21}\mu_1 + p_{22}\mu_2 = 1 + \frac{5}{9}\mu_1 + \frac{4}{9}\mu_2\ ,$$

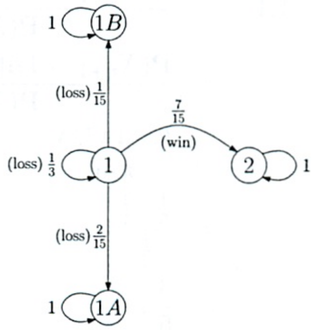
which when solved, yields the expected time until Jack slips cards into the deck,

$$\mu_1 = 9.2\ .$$

(e) Let S be the number of times that the dealer switches from deck #2 to deck #1, which equals the number of times that he/she switches from deck #1 to deck #2. Let p be the probability that $S = 0$, which is the sum of the probability of all ways for the first change of state to be from state 1 to state 1A or state 1B,

$$p = \left(\frac{2}{15}+\frac{1}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{15}+\frac{1}{15}\right) + \left(\frac{1}{3}\right)^2\left(\frac{2}{15}+\frac{1}{15}\right) + \ldots = \frac{1}{1-1/3}\cdot\frac{3}{15} = \frac{3}{10}\ .$$

Alternatively, p is the probability of absorption of the following modified chain into an absorbing state ($1A$ or $1B$), when started in state 1:



As $\mathbf{P}(S > 0) = 1 - p$, and similarly, $\mathbf{P}(S > k+1|S > k) = 1 - p$, it should be clear that S will be a shifted geometric, and thus

$$p_S(k) = \left(\frac{7}{10}\right)^k \frac{3}{10} \qquad k = 0, 1, 2, \ldots\ .$$

(f) Note that S from part (e) is the total number of cycles from 1 to 2 and back to 1. During the ith cycle, the number of wins, W_i, is a geometric random variable with parameter $q = \frac{5}{9}$. Thus the total number of wins by Jack before he slips extra cards into the deck is

$$W = W_1 + W_2 + \ldots + W_S \ ,$$

which is a random number of random variables, all of which are independent. Conditioned on $S > 0$, W is a geometric (with parameter p) number of geometric (with parameter q) random variables, all conditionally independent, and thus from the theory of splitting Bernoulli processes,

$$p_{W|S>0}(k) = (1-pq)^{k-1}pq \quad k = 1, 2, \ldots \ ,$$

where $pq = \frac{3}{10} \cdot \frac{5}{9} = \frac{1}{6}$. When $S = 0$, it follows that $W = 0$, and thus by total probability,

$$p_W(k) = \begin{cases} \frac{3}{10} & k = 0 \\ (\frac{7}{10})(\frac{5}{6})^{k-1}\frac{1}{6} & k = 1, 2, \ldots \ . \end{cases} \ .$$

(g) Let W be the total number of wins before slipping cards into the deck (as in part (f)), and similarly let L be the total number of losses before absorption. We know from part (d) that $\mathbf{E}[W + L] = \mu_1 = 9.2$. From part (f) we can find $\mathbf{E}[W]$ by total expectation,

$$\mathbf{E}[W] = E[W|S=0]\mathbf{P}(S=0) + E[W|S>0]\mathbf{P}(S>0) = \frac{7/10}{1/6} = \frac{42}{10} = 4.2 \ ,$$

because when conditioned on $S > 0$, the number of wins, W, is a geometric random variable with parameter $pq = \frac{1}{6}$. From linearity of expectation, we find

$$\mathbf{E}[L - W] = \mathbf{E}[W + L] - 2\mathbf{E}[W] = 9.2 - 2 \cdot 4.2 = 0.8 \ .$$

(h) Using A to again denote the probability of being absorbed into the recurrent class $\{1A, 2A\}$, starting in state 1,

$$\begin{aligned} \mathbf{P}(X_n = 2A | X_{n+1} = 1A) &= \frac{\mathbf{P}(X_{n+1} = 1A | X_n = 2A)\mathbf{P}(X_n = 2A)}{\mathbf{P}(X_{n+1} = 1A)} \\ &= \frac{\mathbf{P}(X_{n+1} = 1A | X_n = 2A)\mathbf{P}(X_n = 2A | A)\mathbf{P}(A)}{\mathbf{P}(X_{n+1} = 1A | A)\mathbf{P}(A)} \\ &\approx \frac{p_{2A,1A}\pi_{2A}}{\pi_{1A}} \\ &= \frac{\frac{1}{5} \cdot \frac{3}{4}}{\frac{1}{4}} \\ &= \frac{3}{5} \ . \end{aligned}$$

Note that the right hand side above equals $p_{1A,2A}$, as clear from the local balance equation $\pi_{1A}p_{1A,2A} = \pi_{2A}p_{2A,1A}$.

G1[†]. With $a > 0$ and $c \geq 0$,

$$\begin{aligned}\mathbf{P}(X-\mu \geq a) &= \mathbf{P}(X-\mu+c \geq a+c)\\ &\leq \mathbf{P}\left((X-\mu+c)^2 \geq (a+c)^2\right)\\ &\leq \frac{\mathbf{E}\left((X-\mu+c)^2\right)}{(a+c)^2}\\ &= \frac{(\sigma^2+c^2)}{(a+c)^2}\end{aligned}$$

where the first inequality follows from the fact that $a + c > 0$, and the second inequality follows from the Markov inequality.
To tighten the bound, we treat $(\sigma^2 + c^2)/(a + c)^2$ as a function of c, and find c such that the derivative is 0. The minimum occurs at $c = \sigma^2/a$. Therefore,

$$\mathbf{P}(X-\mu \geq a) \leq \frac{(\sigma^2+\frac{(\sigma^4)}{a^2})}{(a+\frac{\sigma^2}{a})^2} = \frac{\sigma^2}{(\sigma^2+a^2)}$$

[†]Required for 6.431; optional challenge problem for 6.041

1/22

LECTURE 20
THE CENTRAL LIMIT THEOREM

- Readings: Section 5.4
- $X_1, \ldots, X_n$ i.i.d., finite mean + variance σ^2
- "Standardized" $S_n = X_1 + \cdots + X_n$: Sum is rougly a normal RV

$$Z_n = \frac{S_n - \mathrm{E}[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mathrm{E}[X]}{\sqrt{n}\sigma}$$

  - $\mathrm{E}[Z_n] = 0, \qquad \mathrm{var}(Z_n) = 1$

but as you add more + more it spreads out more + more

- Flatens out
- does not converge to anything meaningful

So do stuff to it

So will standrdize to normal

- Let Z be a standard normal r.v. (zero mean, unit variance)
- **Theorem:** For every c:

$$\mathrm{P}(Z_n \le c) \to \mathrm{P}(Z \le c)$$

- $\mathrm{P}(Z \le c)$ is the standard normal CDF, $\Phi(c)$, available from the normal tables

Settles to limit

- Can calulate prob for Z_n by looking at st. normal table for Z

Usefulness

- universal; only means, finite variances matter
- accurate computational shortcut
- justification of normal models

any dist

↑ why we can use normal - lots of ind. pieces of noise

What exactly does it say?

- CDF of Z_n converges to normal CDF
  - not a statement about convergence of PDFs or PMFs

not about sum but normalized sum and about its CDF

↑ often happens, not going to talk about.

Normal approximation

- Treat Z_n as if normal
  - also treat S_n as if normal

↑ Since Z_n is linear function of S_n

Can we use it when n is "moderate"?

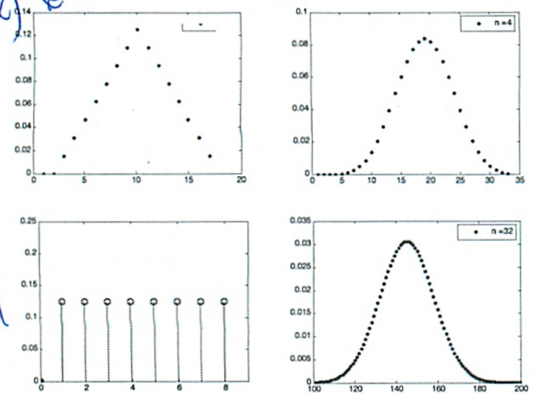
tells us limit when $n \to \infty$

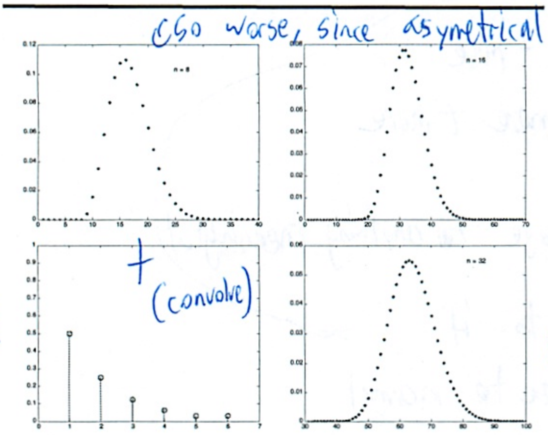
but can you use it when small

- Yes, but no nice theorems to this effect
- Symmetry helps a lot

↑ Depends on dist of Xs

takes effect for much smaller values of n





Conversion so good - Computer random # Generator → adds a relatively small (~16) # of uniform RVs to get a normal dist

if n is in 1000s - have no doubt central limit theorm applies

The pollster's problem using the CLT Poll

- f: fraction of population that "..."
- ith (randomly selected) person polled: Pick n people

$$X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$$

- $M_n = (X_1 + \cdots + X_n)/n$
- Suppose we want:

$$P(|M_n - f| \geq \underset{\text{error}}{.01}) \leq \underset{\text{confidence}}{.05}$$

Prob of .01 error should be ≤ .05

How large a n do we need to satisfy specs?

- Event of interest: $|M_n - f| \geq .01$

$$\left|\frac{X_1 + \cdots + X_n - \overset{\text{same}}{nf}}{n}\right| \geq .01$$

↓ standrize so var( )=1, ~~so~~ "normalize" ↑ unit var

$$|Z_n| = \left|\frac{X_1 + \cdots + X_n - nf}{\sqrt{n}\sigma}\right| \geq \frac{.01\sqrt{n}}{\sigma}$$

$$\begin{aligned} P(|M_n - f| \geq .01) &\approx P(|Z| \geq .01\sqrt{n}/\sigma) \\ &\leq P(|Z| \geq .02\sqrt{n}) \end{aligned}$$

Prob that st. normal is larger than this # (event)

$.01\sqrt{n}/\sigma$

$.02\sqrt{n}$ ↑ bigger than

prob bigger also

other paper →

↑ Use central limit theorm to approx for all n but don't know σ - pollster will be conservative → worst case

$\sigma^2 = f(1-f) \leq \frac{1}{4}$ so $\sigma \leq \frac{1}{2}$ ← maximum randomness

Apply to binomial

- Fix p, where $0 < p < 1$
- X_i: Bernoulli(p)
- $S_n = X_1 + \cdots + X_n$: Binomial(n, p) ← binomial = sum of Bernoullis
  - mean np, variance $np(1-p)$
- CDF of $\dfrac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{} $ standard normal
  approx

Example

- $n = 36$, $p = 0.5$; find $\mathrm{P}(S_n \le 21)$

$P(S_n \le 21)$

↑ take event trying to analize + rewrite to look like st. normal

$E[S_n] = 18$

$var(S_n) = np(1-p) = 9$

$\sigma_{S_n} = 3$

do to both sides

$$P\Big(\underbrace{\frac{S_n-18}{3}}_{Z_n} \le \frac{21-18}{3}\Big)$$

approx to st normal

$\approx P(Z \le 1)$

Look at st. normal table

$= .8413$

- Exact answer:

$$\sum_{k=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = 0.8785$$

↑ by comparison - can calculate exact w/ binomial prob w/ computer

What did we do?

The 1/2 correction for binomial approximation

- $\mathrm{P}(S_n \le 21) = \mathrm{P}(S_n < 22)$, because S_n is integer ← but its the same thing
  - but use same normal approx - gives diff values
- Compromise: consider $\mathrm{P}(S_n \le 21.5)$

↑ so split the difference

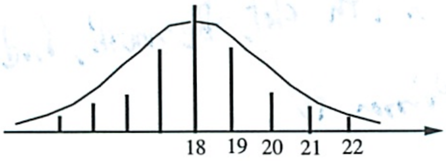
$P\left(\frac{S_n-18}{3} \le \frac{21.5-18}{3}\right)$

$= P(Z \le 1 + \frac{.5}{3})$

$= P(Z \le 1.17)$

$\approx .8790$

↑ closer to real ans



↑ sum of all bars ≤ 21

so pretend its normal

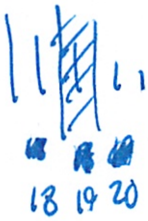
the ± [illegible] on .5 is a good trick for more accuracy

De Moivre–Laplace CLT (for binomial)

- When the 1/2 correction is used, CLT can also approximate the binomial p.m.f. (not just the binomial CDF)

is part of a theorm
but just for binomial for PDF
CDF always works

$$P(S_n = 19) = P(18.5 \le S_n \le 19.5)$$



rewrite (, so standrized

$$18.5 \le S_n \le 19.5 \quad \Longleftrightarrow$$

$$\frac{18.5-18}{3} \le \frac{S_n-18}{3} \le \frac{19.5-18}{3} \quad \Longleftrightarrow$$

$$0.17 \le Z_n \le 0.5$$

$$\begin{aligned} P(S_n = 19) &\approx P(0.17 \le Z \le 0.5) \\ &= P(Z \le 0.5) - P(Z \le 0.17) \\ &= 0.6915 - 0.5675 \\ &= 0.124 \end{aligned}$$

- Exact answer:

$$\binom{36}{19}\left(\frac{1}{2}\right)^{36} = 0.1251$$

Poisson vs. normal approximations of the binomial

lots of little Bernoulli trials in tiny timeslots

- Poisson arrivals during unit interval equals: sum of n (independent) Poisson arrivals during n intervals of length $1/n$
  - Let $n \to \infty$, apply CLT (??)
  - Poisson=normal (????)

X_i = # arrivals in i th slot, ~Bernoulli, ind.
$X = \sum_i X_i$ ~Poisson RV

- Binomial(n, p)
  - p fixed, $n \to \infty$: normal
  - np fixed, $n \to \infty$, $p \to 0$: Poisson
- $p = 1/100$, $n = 100$: Poisson
- $p = 1/10$, $n = 500$: normal

but add items, $p = \frac{1}{n}$ is also changing
so central limit theorm does not apply

In real life

Which to estimate to

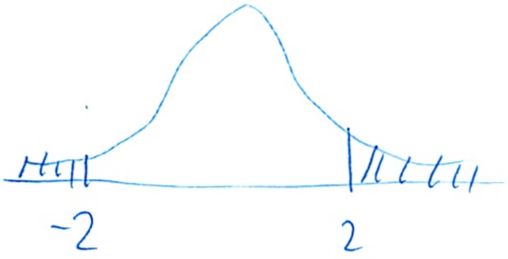
What if $n = 10{,}000$

$P(|z| \geq .02 \cdot 100)$

$= P(|z| \geq 2)$

need to look at normal tables

but they give values for the cumulative



$= 2P(z \geq 2)$

$= 2(1 - P(z \leq 2))$

now can go to table

$= 2(1 - .9767)$

$= 0.045$

? boss wanted 5% - we are way under

so can decrease n?

→ next pg

②

Find n such that $P(|z| \geq 0.02\sqrt{n}) = .05$

sum of shaded area must = .05

So each side should be .025

so CDF here = .975

So look on normal table for .975

We see that this = 1.96 on table

$$P(|z| \geq 1.96) = .05$$

$$\boxed{.02\sqrt{n} = 1.96}$$

$$\Rightarrow n = 9604$$

(what did he do here! – study!)

– this was what I was confused on

Newspapers have accuracy ~2-3%, so sample size ~2000

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2010)

Recitation 21
November 23, 2010

1. Let $X_1, \ldots, X_{10}$ be independent random variables, uniformly distributed over the unit interval [0,1].
    (a) Estimate $\mathbf{P}(X_1 + \cdots + X_{10} \geq 7)$ using the Markov inequality.
    (b) Repeat part (a) using the Chebyshev inequality.
    (c) Repeat part (a) using the central limit theorem.

2. **Problem 10 in the textbook (page 290)**
A factory produces X_n gadgets on day n, where the X_n are independent and identically distributed random variables, with mean 5 and variance 9.
    (a) Find an approximation to the probability that the total number of gadgets produced in 100 days is less than 440.
    (b) Find (approximately) the largest value of n such that
    $$\mathbf{P}\left(X_1 + \cdots + X_n \geq 200 + 5n\right) \leq 0.05.$$
    (c) Let N be the first day on which the total number of gadgets produced exceeds 1000. Calculate an approximation to the probability that $N \geq 220$.

3. Let $X_1, X_2, \ldots,$ be independent Poisson random variables with mean and variance equal to 1. For any $n > 0$, let $S_n = \sum_{i=1}^{n} X_i$.
    (a) Show that S_n is Poisson with mean and variance equal to n. Hint: Relate $X_1, X_2, \ldots, X_n$ to a Poisson process with rate 1.
    (b) Show how the central limit theorem suggests the approximation
    $$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$
    for large values of the positive integer n.

Recitation 21

11/23

Central Limit Theorm

$X_1, X_2, \ldots\ldots$ IID RVs

w/ $E[X_i] = \mu$ $\quad var(X_i) = \sigma^2$

$S_n = X_1 + \ldots + X_n$

$E[S_n] = n\mu$ $\quad var(S_n) = n\sigma^2$

Standardize S_n i.e. transform to

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

so $E[Z_n] = 0$ $\quad var(Z_n) = 1$ st dormal

CLT: $CDF(Z_n) \xrightarrow[\lim_{n\to\infty}]{} CDF(st. normal)$

today most important day in carse

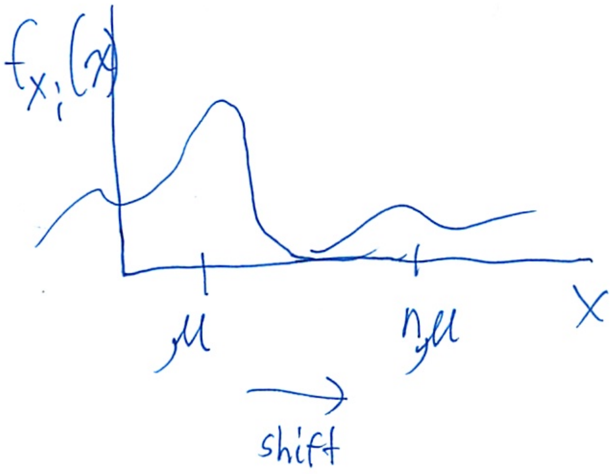
Most important theorm in class

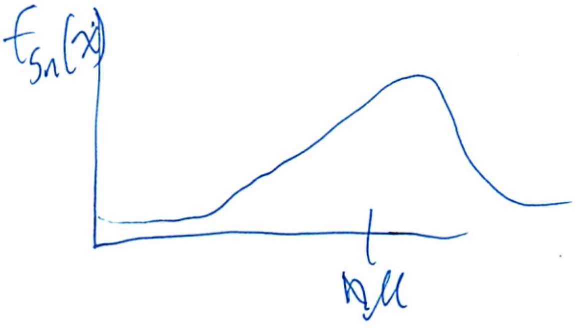
---

estimating prob that would be hard to estimate oter way

* don't have to be sums

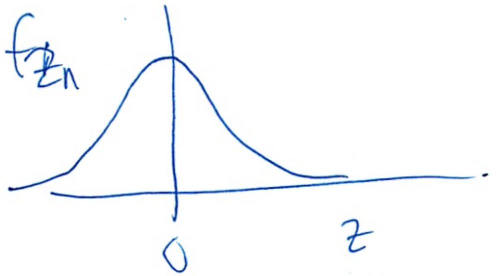
- as $n \uparrow$, becomes more spread out
- but divide by n -> then squishes
- Converges to st. normal as $n \uparrow$

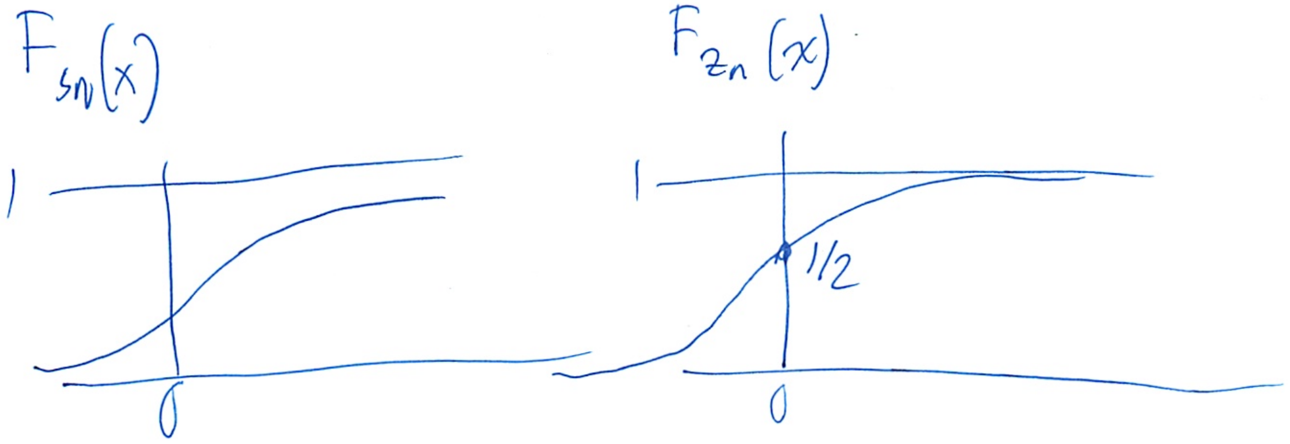




Varience will be bigger

will approch st. normal





Then look up table for correct value

this is for contineas

if discrete → PDF would have spikes
CDF has steps
will resemble the contineas curve

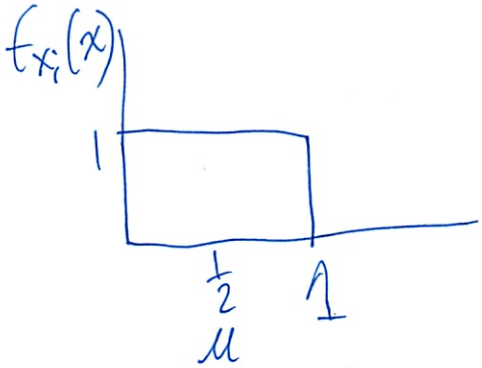
$\forall -\infty < z < \infty$

How to use

- estimate prob of sums
  - markov + chebshev give upper bands
  - CLT gives estimate, which is much better

④

1. $X_1, \dots X_{10}$ iid



$X = X_1 + \dots + X_{10}$

$E[X_1 + \dots + X_{10}] = .5 \cdot 10 = 5$

$Var(X_1 + \dots + X_{10}) = \frac{10}{12} = \frac{5}{6}$

a) Estimate $p(X_1 + \dots X_{10} > 7)$ by Markov inequality

$$P(X \geq a) \leq \frac{E[X]}{a}$$

$$P(X \geq 7) \leq \frac{E[X]}{7}$$

$$= \frac{5}{7}$$

$$= .07142$$

⑤

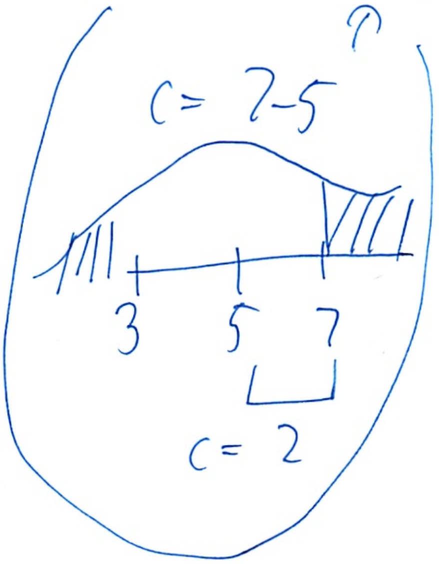
b) Estimate w/ Chebschev $P(|X-\mu| \ge c) \le \frac{\sigma^2}{c^2}$

$\downarrow$

$P(X \ge 7)$

- prob that I exceed mean on eiter side by certain value
- **but** this is one sided

$$2 \cdot P(X \ge 7) = P(|X-5| \ge 2) \le \frac{\frac{10}{12}}{4}$$



* $P(X \ge 7) \le \frac{10/12}{2 \cdot 4} = .295$ ← is closer ~~more accurate~~

~~So~~ smaller

also uses var

not. ness. more accurate

* is upper bound estimate

(I think I get it much better now!)

⑥

c) Estimate $P(X \geq 7)$ Using CLT

We want Complementry prob - b/c we need other way)

$$P(X \geq 7) = 1 - P(X \leq 7)$$

$$= 1 - P\left(\frac{X - E[X]}{\sqrt{Var(X)}} \leq \underbrace{\frac{7 - E[X]}{\sqrt{Var(X)}}}_{2,91}\right)$$

$$= 1 - \Phi(2,91)$$

go to table

; but why not $\frac{S_n - nE[X]}{\sqrt{n}\,\sigma}$

$$\approx 1 - ,9$$

$$\approx ,1$$

← much smaller

actual estimate (not upper bound)

pretty close to actual value

Could do actual by convolving

- with a computer
- too hard by hand

---

no way of knowing when n is big enough

- judge ; ~~Is it close to normal?~~ ↓

Does the dist look like a normal dist?

Depends if close to normal

Binomial 30

100 of just about anything

Next week: how to estimate var by taking samples

-can use to make a confidence interval

Now moving away from sums (S_n)

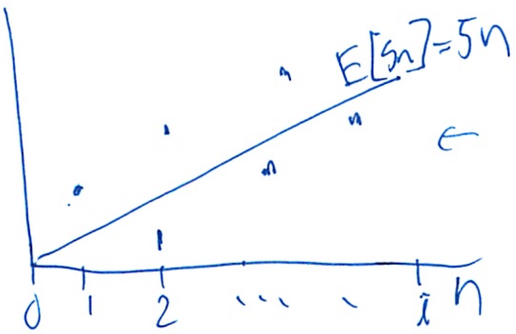
2. Factory produces gadgets

X_i = # of gadgets produced in day i

$E[X_i] = 5 \quad var(X_i) = 9$

$S_n = X_1 + \dots + X_n$

$E[S_n] = 5n \qquad var(S_n) = 9n$



← actual # will be close to line

⑧

Lets look at day 100

- expected # = 400

But there is probability - won't make exactly 100

We want to know

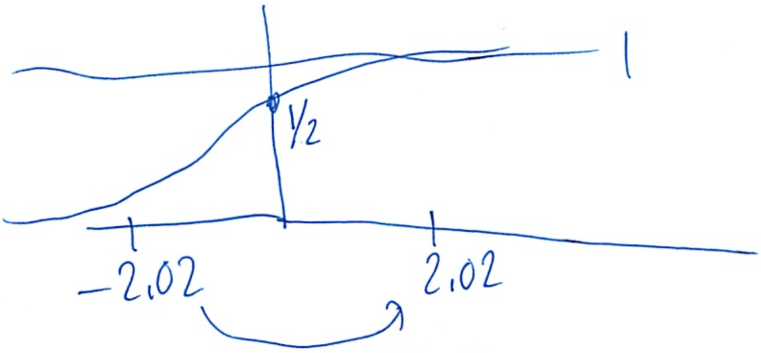
$$P(S_{100} \leq 440) =$$

go down to $\frac{1}{2}$ to be more accurate

$$P(S_{100} \leq 439.5) =$$

$$= P\left(\frac{S_{100} - 5 \cdot 100}{3\sqrt{100}} \leq \underbrace{\frac{439.5 - 500}{30}}_{-2.02}\right)$$

$$\approx \cancel{P}\,\Phi(-2.02)$$



go to 2.02
and $1 - \Phi(2.02)$

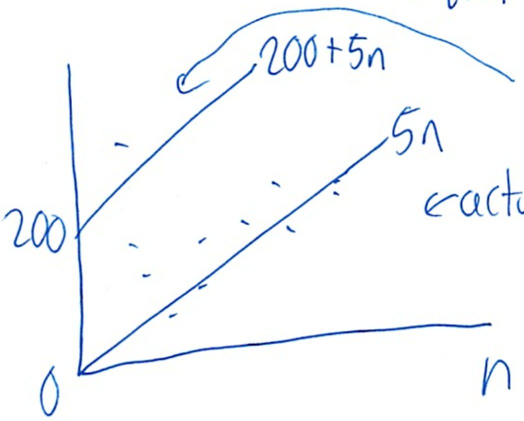
$= 1 - \phi(2.02)$

$= .0217$

---

b) Now a different RV

Find n s.t. $P(S_n \geq 200 + 5n) \leq .05)$



Prob that I will be way off will be 5%

actual will be around

-find # days off avg

Use central limit approximation

$$P\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}\right) \leq .05$$

$$\approx 1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) \leq .05$$

or $\Phi\left(\frac{200}{3\sqrt{n}}\right) \geq .95$.

Solve for n

look at table for value $= \phi(.95)$

So

$$\frac{200}{3\sqrt{n}} \geq 1.65$$

$$n \leq 1632$$

will take 1632 days to go over 200+5n line
w/ prob .95

.95 confidence level

How many days will be safe to that extent.

C. similar - ship

---

3. ~~Similar Stirling~~ Stirling approximation

Uses factorial

$$n! \approx \underbrace{n^{n+\frac{1}{2}} e^{-n} \sqrt{2n}}_{\text{eaiser to calculate}}$$

What does it mean for them to be close

relative error/ratio is small (goes to 1)
as $n \to \infty$

to prove: calculus + expansion series

How in all world made up?

- same person as central limit theory

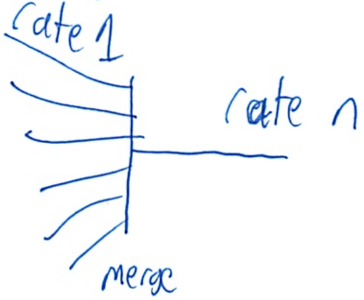
But can't prove w/ central limit theory

Person who invented it was a consultant to gamblers

Tracing line of Thought:

- Consider Poisson RVs w/ ~~param~~ mean (1)
- If you add these ind. poisson, get poisson

$$\underbrace{X_1 + \dots + X_n}_{\substack{\text{poisson} \\ \text{mean } n \\ \text{var } n}}$$



$$- P(S_n = n) = P\left(n - \tfrac{1}{2} < S_n < n + \tfrac{1}{2}\right)$$

standarize

$$= P\left(-\frac{1}{2\sqrt{n}} < \underbrace{\frac{S_n - n}{2\sqrt{n}}}_{\approx \text{st. normal}} < \frac{1}{2\sqrt{n}}\right)$$

$\approx$ Guassian

$$\Phi(\ ) - \Phi(\ )$$

$$\approx \underbrace{\int \frac{1}{\sqrt{2\pi}} \int_{-\frac{1}{2\sqrt{n}}}^{\frac{1}{2\sqrt{n}}} e^{-z^2/2} \, dz}_{\text{central limit approximation}}$$

$$\overset{\text{large } n}{\approx} \frac{1}{\sqrt{2}\,n} \cdot \frac{1}{\sqrt{n}} \underbrace{e\Big|^{-z^2/2}_{z=0}}_{1}$$

$$= \frac{1}{\sqrt{2\pi n}}$$

$$\approx \underbrace{e^{-n} \frac{n^n}{n!}}_{\text{poisson pmf}}$$

$$\downarrow$$

$$\frac{1}{\sqrt{2\pi n}} \approx \frac{e^{-n} n^n}{n!}$$ stirling approx

just suggestive

not a proof (need more analysis to prove accurately

this is just central limit theorm in disguise