6.041 Read Notes

12/13

- analytical framework

Ω = sample space

prob of exact pt = 0 ← don't waste time rewriting stuff I know

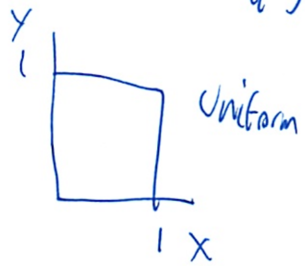
$\cap$ = and

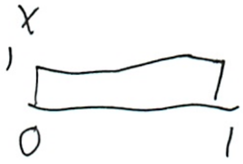
$\cup$ = union/or

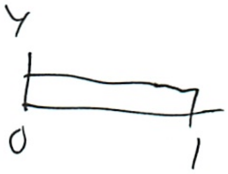
A^c

$\emptyset$ = disjoint

$P(X+Y \le 1/2) =$

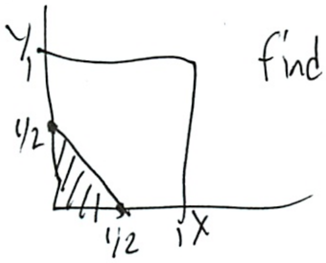






i have no clue - even now on day 1

Calculate area where true



find area of that

$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \boxed{\frac{1}{8}}$

See I do know it

just need to think about it

Sum disjoint pieces

(will just do lectures for now)

1. Specify sample space
2. Define prob law
3. Identify event of interest
4. Calculate

Conditional $P(A|B)$ = prob of A given B has occured

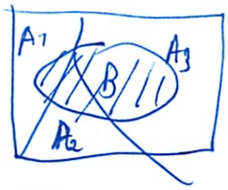
$P(A \cap B) = P(A|B)\,P(B)$

- ~~Bin~~ - are the further out branches of a tree

Partitions

Total prob theorm

- just got this recently



$P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$

Bayes Rule

- went over in 6.01 - get much more now

Conditioning affects independence

- does it provide any info?

Counting

Permutations - order matters?

(this has all been compressed into geometric, etc)

- abstracted away

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

Binomial probabilities

$\binom{n}{k} p^k (1-p)^{n-k}$

Random Variable

↳ assignment of a value to every possible outcome

- function that maps sample space to the real #s

PMF

↳ how much weight

(we got into topics fast)

Binomial PMF

Expected value

- center of gravity on PMF
- average outcome w/ ∞ trials

$E[\alpha] = \alpha$ α, β constants

$E[\alpha X] = \alpha E[X]$

$E[\alpha X + \beta] = \alpha E[X] + \beta$

Variance

↳ aka 2nd moment

how far X is from the average

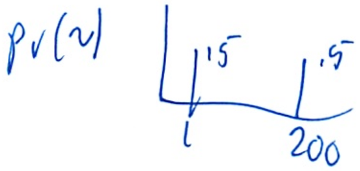
- make sure can still calc manually

St Dev $\sigma_x = \sqrt{Var(x)}$

④

Example

Go. 200 miles at constant speed V



$E[V] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 200 = 100.5$

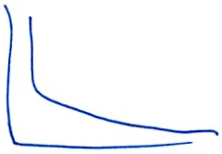
$var(V) = \frac{1}{2}(1-100.5)^2 + \frac{1}{2}(200-100.5)^2 = 10,000$

st dev $= \sqrt{10,000} = 100$

$E[T/V]$ ← always need to study

Conditional expectations
- new universe
- normalize so still adds to 1

Geometric



Memoryless

not writing PMF, E[ ], var( ) down on cheat sheet

Total Expectation Theorm

"just add the expectations on each"

$P(A_1) E[X|A_1] + \dots + P(A_n) E[X|A_n]$

Joint PMF

- that chart thing
- make sure am able to do
- smushing + all

⑤

begining of year was def. much eaiser

- or at least later material built upon so I got lots of practice

$E[X+Y+Z] = E[X]+E[Y]+E[Z]$

If X, Y ind

$E[XY] = E[X]E[Y] = \sum_x \sum_y xy\, P_{X,Y}(x,y)$

$= \sum_y \sum_x p_X(x)\, p_Y(y)$

$Var(aX) = a^2 Var(X)$

$Var(X+a) = Var(X)$

a lot of things on formula sheet I just make sense now

exam 1

Wow that was fast

Contineous RVs PDF

$P(a \le X \le b) = \int_a^b f_X(x)\,dx$ never really solidly learned

work w/ little slices

$P(x \le X \le x+\delta) = \int_x^{x+\delta} f_X(s)\,ds \approx f_X(x)\cdot\delta$

disjoint - so just add

Uniform RVs

$f_X(x) = \frac{1}{b-a}$ $E[X] = \int_a^b x \cdot \frac{1}{b-a}\,dx$

⑥

CDFs

$$\underset{\text{capital}}{F_x}(x) = P(X \le x) = \int_{-\infty}^{x} f_x(t)\,dt$$

I think my math reasoning got a lot better w/ this class

Can mix discrete + continuous

Normal

$$(\underset{\text{mean}}{0}, \overset{\text{var} = \sigma^2}{1}) \leftarrow \text{standard}$$

Look at table for CDF

Joint PDF

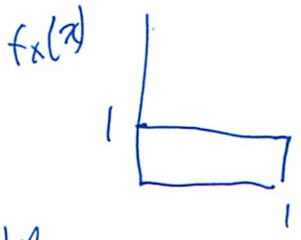
- just sum both ways

Conditioning - same as discrete but w/ an f

Stick Breaking example

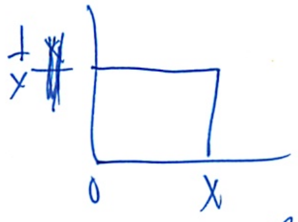
- prob density may be darker in some areas (aka is 3D)
- make sure to think about
- integrate over interval

Start w/ stick $l = 1$, break uniform $(0, 1]$



[illegible]

Break again uniform [0, x]

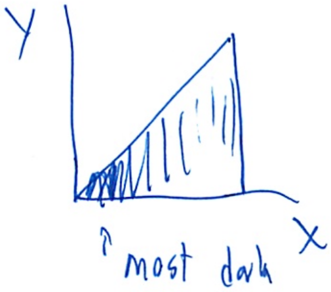


So uniform ~~[illegible]~~ $f_{X,Y}(x,y) =$

$f_X(x)\,f_{Y|X}(y|x) = \frac{1}{l}\cdot\frac{1}{x}$ $\quad x\in[0,l]$

↑ ahh! remember $\quad y\in[0,x]$

↑ include



$0 \le y \le x \le l$

Then they can ask questions

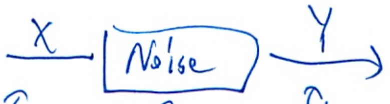
↓ since $y=x$ $\quad$ ↙ the value itself w/ E[ ]

$E[Y|X=x] =$

$= \int_0^x y\cdot\frac{1}{l}\cdot\frac{1}{x}\,dy$

$= \frac{x}{2}$ ← uniform along the slice

Inference - what did in 6.01



↑ know initial dist want to know what actual value is

↑ know

↑ know actual value

?, Need a 1-1 mapping for ?

(I don't really get what they are trying to do here)

Actually if just think about concepts better than memoizing how to do each problem

Oh Derived dist was also in here - that confused me

~~Can Convolve~~

To find joint density → convolution

- marginal · conditional → conditional

Cross multiply + integrate

In practice is difficult

Sum of ind normal RVs → end result is normal

- end of course featured this

<u>Cov</u>

$$\mathrm{Cov}(X,Y) = E[(X-E[X])\cdot(Y-E[Y])]$$

$$= E[XY]-E[X]E[Y]$$

0 if ind

Correlation coefficent

$$\frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$ - dimensionless version of cov

⑨

Conditional expectation

- just expectation but only inside the condition

Law of iterated expectations

- still don't get this on

$$E[E[X|Y]] = \sum_{Y} E[X \mid Y = y]\, p_Y y = E[X]$$

Stick example

- ~~All~~ view Y as random too

find ~~A~~ E[Y]

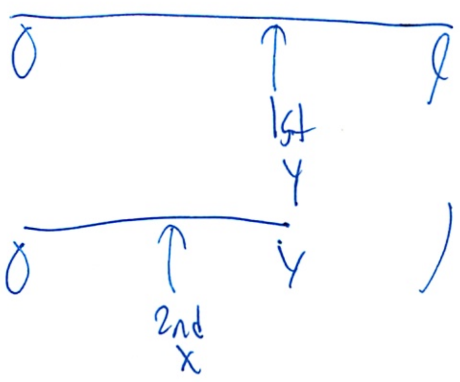
then calc expected value of x

$$E[X] = E[E[X|Y]] = E\left[\frac{Y}{2}\right] = \frac{E[Y]}{2} = \frac{\overset{\downarrow E[Y]}{l/2}}{2} = \frac{l}{4}$$

$\uparrow$

convert ↑ $E[X|Y] = \frac{Y}{2}$ RV

do first → $E[X \mid Y = y] = \frac{y}{2}$ #



) they flipped 1st + 2nd break variables — confusing

Could find PDF of X

- but not that easy to do
- So instead

$$E[X] = E[\underbrace{E[X|Y]}_{\frac{Y}{2}}]$$

$$= \frac{E[Y]}{2}$$

$$= \frac{\ell/2}{2} =$$

$$= \frac{\ell}{4}$$

Another example

$$E[g(X)\cdot h(Y)|X] = g(X)\cdot E[h(Y)|X]$$

Have for any possible value $X=x$

$$E[g(X)\cdot h(Y)|X]$$

$$= \int_{-\infty}^{\infty} g(x)\, h(y)\, f_{Y|X}(y|x)\, dy$$

$$= g(x)\, \underbrace{E[h(Y)|X=x]}_{\text{Constant comes out}}$$

Still don't get

Try to find other examples

~~Var(X|Y)~~ Law of Total Var

$$var(X) = E[var(X|Y)] + var(E[X|Y])$$

- how do we use this again?

$$var(X|Y=y) = E[(X - E[X|Y=y])^2 | Y=y]$$

Proof

$$Var(X) = E[X^2] - (E[X])^2$$

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

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

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

Since $Var(X) = E[X^2] - (E[X])^2$

---

example test scores

Y = section # (Y = 1 10 students
Y = 2 20 students } sections
X = quiz score

given [$E[X|Y=1] = 90$ $E[X|Y=2] = 60$

always confused on what is given

$$E[X|Y] = \begin{cases} 90 & \text{w/ prob } 1/3 \\ 60 & 2/3 \end{cases}$$ } due to students

? assume given

⑬

$$E[E[X|Y]] = \frac{1}{3} \cdot 90 + \frac{\cancel{1}2}{3} \cdot 60$$

$$= 70$$

$= E[X]$ - test score of any section

ok makes for more sense now!

$$Var(E[X|Y]) = \frac{1}{3}(90-70)^2 + \frac{2}{3}(60-70)^2$$

$$= \frac{600}{3} = 200$$

ok see this one as well

Variability b/w Sections

var of each piece

I guess it depends on what data you have

given

$$\left[\begin{array}{l} Var(X|Y=1) = 10 \\ Var(X|Y=2) = 20 \end{array}\right.$$

| for each section

$$Var(X|Y) = \begin{cases} 10 & \text{w/ prob } 1/3 \\ 20 & \text{w/ prob } \underbrace{2/3}_{\text{extra info}} \end{cases}$$

$$E[Var(X|Y)] = \frac{1}{3} \cdot 10 + \frac{2}{3} \cdot 20 = \frac{50}{3}$$

avg variability within section

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

now just plug in

$$\frac{50}{3} + 200$$

Sum of a R # of RVs

$Y = X_1 + \dots + X_n$

$E[Y | N = n] = n E[X]$

but ensure of

$E[Y] = E[n] E[X]$

Var = law of total var before

Quiz 2

- Seemed to come very fast
- did a lot of review for it
- almost as much as the work from that chap
- well p-sets were missing

Bernoulli process

need to do major cheat sheet now

- really this was not on quiz 2?

$P(\text{success}) = P(X_i = 1) = p$

$P(\text{failure}) = P(X_i = 0) = 1-p$

Series of independent Bernoulli RVs - 0 or 1

$E[X_t] = p$

$Var(X_t) = p(1-p)$

15

Random Processes

- Bernoulli discrete
- Poisson contineous) memoryless
- Markov w/ memory

\# sucesses in n time slots = binomial

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

$E[S] = np$

$var(S) = np(1-p)$

Doing rest on Formula sheet so don't write 2x

Yeah all these things make more sense now
- The qu is - can I do the problems?

Don't really need to know proofs
- Don't confuse yourself w/
- I think I don't seperate out in my mind what is a proof

Merging + splitting

Poisson

- interval duration τ
- arrival rate λ

Leture notes not very symetical - challenging

Splitting + Merging

(doing all this on cheat sheet)

Poisson Fishing example

$\lambda = .6$ hr

Fish for 2 hrs

- if catch, stop now

- if nothing, continue to 1st catch

a) P(fish for more than 2hrs)

= P no arrivals in 2hrs

binomial $$p(0,2) = \frac{(.6\cdot 2)^0 e^{-.6\cdot 2}}{0!}$$

Then evaluate
don't need to on exam

$e^{-\lambda t} = e^{-2\cdot .6}$

-oh just simplified of what of I had

or $P(T_1 > 2) = \int_2^\infty f_{T_1}(t)\,dt$

↑ CDF of exponential?

b) P(fish for more than 2 hrs and less than 5)

$P(0,2) \cap P(1,5)$ Can be more than 1??

$P(0,2)\,(1-P(0,3))$

- but if 1, you will stop
- then its not 5 hrs

↑ need to do the complement

or $P(2 < T_1 < 5) = \int_2^5 f_{T_1}(t)\,dt$

(17)

c) P(Catch at least 2 Fish) =

- must be in the 2 hrs

$P(\geq 2, 2)$

↑ can't do

$1 - P(0,2) - P(1,2)$ ✓

or $P(Y_2 \leq 2) = \int_0^2 f_{Y_2}(y)\,dy$

↑erlang

↑CDF or erlang

d) $E[\#\text{ of fish}]$

$= E[\#\text{ fish 2hrs}] + E[\text{fish after}] \cdot P_{no}$ fish 2hrs ✓ total prob theorm

$E[N_T] = \lambda t$

$= .6 \cdot 2 + P(0,2) \cdot 1$ ✓

↑always 1 fish

e) E[future fishing time | fished for 4 hrs] =

- meant 0 fish caught up to then
- but memoryless

└→ exponential till then $E[T_1] = \frac{1}{\lambda} = \frac{1}{.6}$ ✓

E) E[total fishing time] =

$$2 \text{ hrs} \cdot (1-p(0,2)) + \underset{\uparrow \frac{1}{\lambda}}{E[T_1]} \cdot p(0,2)$$

$$\underbrace{2\cancel{(1-e^{-2\cdot .6})}}_{\text{they don't have this}} + \frac{1}{.6} \cdot \underbrace{e^{-2\cdot .6}}_{}$$

[illegible]

{always there

↑ extra time

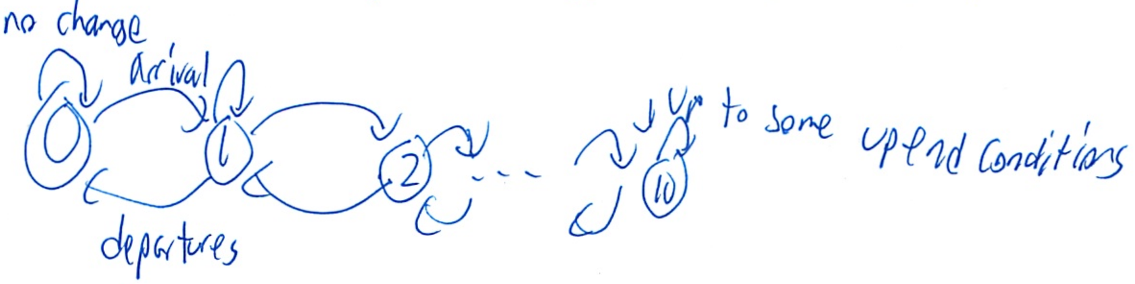
nice - doing pretty good

Random Incidence

- the tricky bus thing "paradox"
- erlang order 2

Markov

- need to write on cheatsheet what all of the letters mean
- State ~~[illegible]~~ - define
  - ↳ in one example # customers at time n



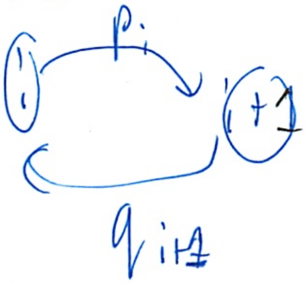
↳ can define prob of each transition

like here and time served is ~~[illegible]~~ geometric(a)

Only current state matters for next transition probabilities

See cheatsheet

Still kinda confused on birth-death



↳ just a simplified / narrower case of long run trans prob?

Just write formulas down I guess

Limit Theorems

- So this is Chebshev
  Markov
  CLT
- than what?
- This is the 1 section in the book I am most unsure of
  11/17 - so all in last month
  ~ post thanksgiving

Work from textbook here - clear

Provide an interpretation of expectations

Allow approx estimate of S_n (height of penguins in antartica)
w/o an explicit PMF/PDF

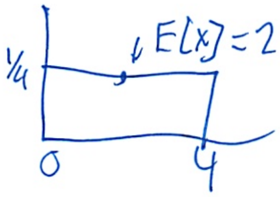
Yeah helpful w/ large data sets

⑳

Markov ineq - if a nonneg RV has small mean, than prob that it takes a small value must also be small

See cheat sheet

example X



So Markov

$P(X \ge 2) \boxed{\le} \frac{2}{2} = 1$ ← So say this prob must be than less than 1 (duh in this case)

actual $P(X \ge 2) = .5$

(kinda getting it now)

$P(X \ge 3) \le \frac{2}{3} = .67$ $\quad P(X \ge 3) = .25$

$P(X \ge \underset{\overset{\uparrow}{a}}{4}) \le \frac{2}{4} = .5$ $\quad P(X \ge 4) = 0$

See I need to see an easy example + think about that is what this inequality means!

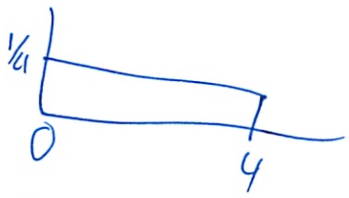
Chebshev

cheat sheet

↓

㉑

Same example



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

$$\mu = \frac{b-a}{2} = \frac{4-0}{2} = 2$$

$$\text{var} = \sigma^2 = \frac{(b-a)^2}{12} = \frac{(4-0)^2}{12} = \frac{4}{3}$$

Now what is c?

Want to test prob |X-2| ≥ 1

$$P(|X-2| \geq 1) \leq \frac{4}{3}$$

- uninformate

I don't get what they are asking

Howabout that thing we tested w/ Markov?

Upper bound

- if don't know σ^2, or range $= [a,b]$
- put in bound

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

~~qudratic minimized when~~

Put in γ for constant

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

I dont get this at all

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

$$\sigma^2 = E[(X - E[X])^2] \le E[(X-\gamma)^2] \text{ for all } \gamma$$

By letting $\gamma = \frac{(a+b)}{2}$ get

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

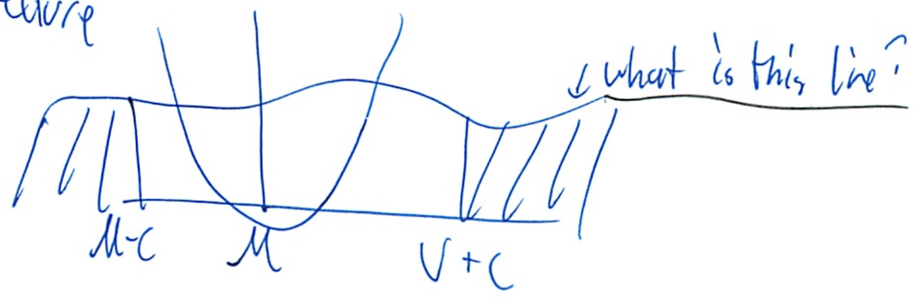
↑ from $(X-a)(X-b) \le 0$ for all X in range $[a,b]$

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

but can't improved

Any better examples in notes?

Lecture



↓ what is this line?

㉓

Steps

1. Fix c to be a certain #
   ↳ How? What?
2. P of being at least c away from μ
3. Look at var when var small - small prob of being away from mean

WP: No more than $\frac{1}{k^2}$ of values can be more than k st. dev from mean
$\uparrow c$

↳ oh is in notes

Lecture Example: Pollsters Problem

f = true fraction

ith person polled

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

$M_n = \frac{(x_1 + \dots + x_n)}{n}$ ← experimental mean

Goal 95% confidence of ≤ 1% error
↳ never fully got difference

$$P(|M_n - f| \geq .01) \leq .05$$

㉔ Chebsher

$$P(|M_n - \mathcal{E}| \geq .01) \leq \frac{\sigma_{M_n}^2}{(0.01)^2}$$

$$\underbrace{= \frac{\sigma_x^2}{n(.01)^2} \leq \frac{1}{4n(0.01)^2}}_{\text{why are we doing this?}}$$

Choose n so this is $\leq .05$

?key

Diff Scaling of M_n

- look at var of sum

$\frac{S_n}{\sqrt{n}}$ → will be constant var σ^2

(don't really get - what are the conditions?)

CLT

- Standrize

C = Confidence

$$\frac{\sigma^2}{n \; C^2}$$

↑ # samples

- yeah have this in pollster problem

㉕

Oh right have to estimate var sometimes

– had a nice table somewhere at end

Example lightbulbs

$$X_i = \text{ith sample} = \begin{cases} 1 & \text{if bulb good} \quad w/p = p \\ 0 & \text{bad} \quad w.p = 1-p \end{cases}$$

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

But don't know var $= \sigma^2$

So estimate ...

Suppose $n = 50$

$$\underset{\text{? confidence}}{P(|M_n - p| \geq .1)} \leq \frac{p(1-p)^{\leftarrow \text{var}}}{n \cdot c^2}$$

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

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

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

? make sure can put into words

✱ just a upper bound – better to use CLT

I never fully got that M_n meant sample mean

CLT

-standardize

-when n is "large"

Ex Pollster again

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

-event of interest $|M_n - f| \geq .01$

$$\left.\begin{array}{l} \text{standize} \\ \geq \dfrac{.01\sqrt{n}}{\sigma} \end{array}\right\} \text{details}$$

$$P(|M_n - f| \geq .01) \approx P\left(|Z| \geq \frac{.01\sqrt{n}}{\sigma}\right)$$

again may need to estimate σ

Binomial

CDF of $\dfrac{S_n - np}{\sqrt{np(1-p)}}$

ex find $P(S_n \leq 21)$ $n = 36$ $p = .5$

㉗

exact ans $\sum_{k=6}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = .87785$

but $E[n] = S18$

$\text{var}(Sn) = np(1-p) = 9$

$\sigma_{Sn} = 3$

$P\left(\frac{Sn - 18}{3} \leq \frac{21 - 18}{3}\right)$

$P(Z \leq 1)$

$\Phi(1)$ normal table

Study this!

But usually $\frac{1}{2}$ correction

so do 21.5 instead

$\Phi\left(\frac{21.5 - 18}{3}\right)$

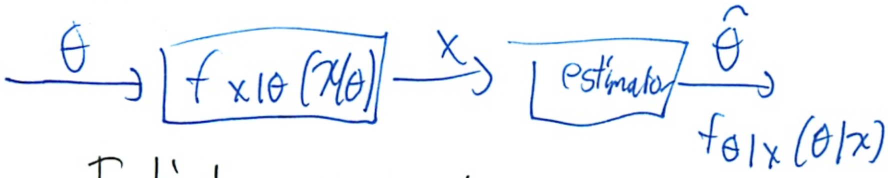
~~Skipping~~ De Moivre Laplace

is interval in middle

prob in width b/w 18.5, 19.5

$P\left(Z \leq \frac{19.5 - E[\,]}{\sigma}\right) - P\left(Z \leq \frac{18.5 - E[\,]}{\sigma}\right)$

Now estimators (11/29)



I kinda get this chap

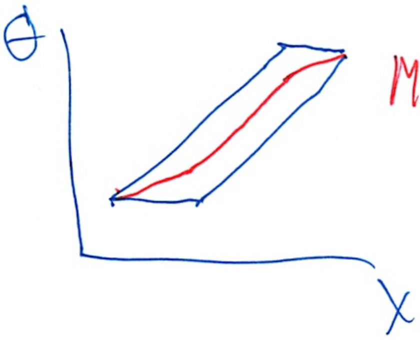
MAP - easy peasy
↳ just watch will get wrong

LSE/CMS

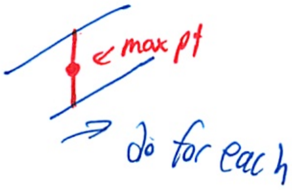
- - .

Pay attention to their example

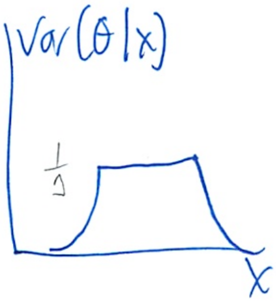
MAP



MAP - where higest pt is at each slice



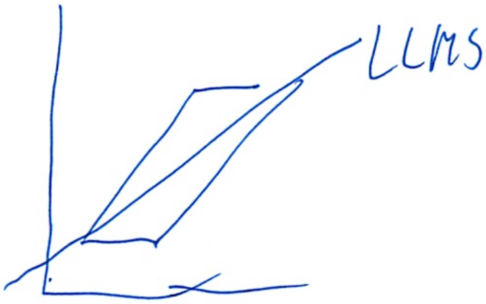
LSC



? what would this be?

? Same - but never wrote it!

LLMS

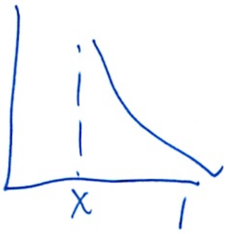


Romeo + Juliet are also examples

- should look over

More complex math

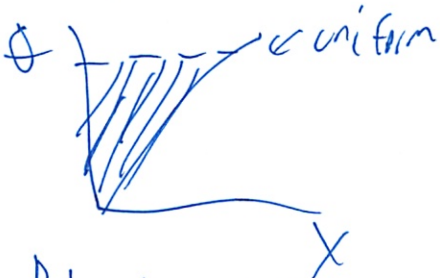
$f_{\theta|x}(\theta|x)$



← ?what estimator?

Then get ops in

Sketch joint dist



But then later pts have diff densities

LLMS just do math

- make sure am able to do
- review later

㉚

Classical stats

here you have to try all the θ or something

Problem Types

- Hyp Testing

  $H_0 = \frac{1}{2}$ vs $H_1 = \frac{3}{4}$

- Composite Hyp

  $H_0 = \frac{1}{2}$ vs $H_1 \neq \frac{1}{2}$

- Estimation

  design estimator $\hat{\theta}$ to keep estimation error $\hat{\theta} - \theta$ small

$\hat{\theta}_{ML}$ = Pick θ that makes data most likly

But usually log

Confidence Intervals

(oh last chap responsible for!
- the last 2 lectures don't count

- estimator $\hat{\theta}_n$ may not be infomative enough
- $1 - \alpha$ CI

$$P(\hat{\theta}_n^- \leq \theta \leq \hat{\theta}_n^+) \geq 1 - \alpha \quad \forall \theta$$

Work backwards $\Phi(\ ) = 1 - \frac{.05}{2}$ CI

㉛

I think I am going too fast through this
- well never did any problems on this

Where was the review of possible estimators
? tutorial or recitation
↳ recitation 23

estimate V_n =

1, Bernoulli only

$\hat{\theta}(1-\hat{\theta})$

2. Sample var
- good for normal or large # samples

$$V_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\theta}_n)^2$$

3. Or use upper bound
- will over estimate, conservative
- Bernoulli $= \frac{1}{4}$

Practice 12/13

6.041 Fall 2009 Final Exam
Tuesday, December 15, 1:30 - 4:30 PM.

DO NOT TURN THIS PAGE OVER UNTIL YOU ARE TOLD TO DO SO

Name: ____________________

Recitation Instructor: ____________________

TA: ____________________

Question	Score	Out of
1		
2 (a)		5
2 (b)		5
2 (c)		5
2 (d)		5
3 (a)		5
3 (b)		5
3 (c)		5
3 (d)		5
3 (e)		5

Question	Score	Out of
4 (a)		5
4 (b)		5
4 (c)		5
4 (d)		5
4 (e)		5
4 (f)		5
5 (a)		5
5 (b)		5
5 (c)		5
5 (d)		5
5 (e)		5
Your Grade		100

- This exam has 5 problems, worth a total of 100 points.
- **When giving a formula for a PDF, make sure to specify the range over which the formula holds.**
- Write your solutions in this quiz booklet, only solutions in this quiz booklet will be graded. Be neat! You will not get credit if we can't read it.
- You are allowed three two-sided, handwritten, formula sheets plus a calculator.
- You may give an answer in the form of an arithmetic expression (sums, products, ratios, factorials) of numbers that could be evaluated using a calculator. Expressions like $\binom{8}{3}$ or $\sum_{k=0}^{5}(1/2)^k$ are also fine.
- The last page of this final contains a standard normal table.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2009)

Problem 1: (incorrect answers: -1 point) Write your name, your assigned recitation instructor's name, and assigned TA's name on the cover of the quiz booklet. The Instructor/TA pairing is listed below.

Recitation Instructor	TA	Recitation Time
Jeffrey Shapiro	Jimmy Li	10 & 11 AM
Danielle Hinton	Uzoma Orji	1 & 2 PM
William Richoux	Ulric Ferner	2 & 3 PM
John Wyatt (6.431)	Aliaa Atwi	11 & 12 PM

Problem 2. (20 points)
A pair of jointly continuous random variables, X and Y, have a joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} c, & \text{in the shaded region of Fig. 1} \\ 0, & \text{elsewhere.} \end{cases}$$

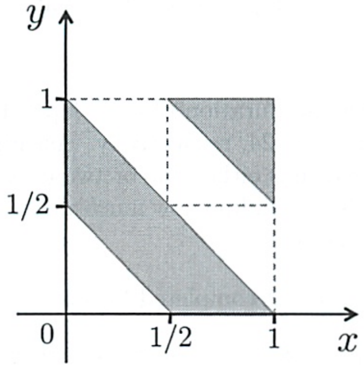


Figure 1: The shaded region is the domain in which $f_{X,Y}(x,y) = c$.

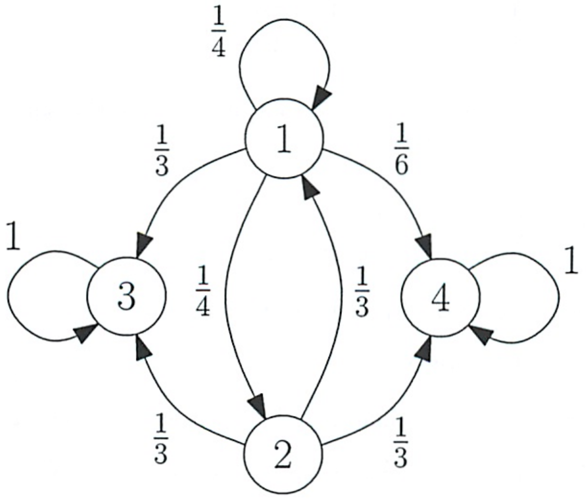
(a) (5 points) Find c.

(b) (5 points) Find the marginal PDFs of X and Y, i.e., $f_X(x)$ and $f_Y(y)$.

(c) (5 points) Find $\mathbf{E}[X \mid Y = 1/4]$ and $\text{Var}[X \mid Y = 1/4]$, that is, the conditional mean and conditional variance of X given that $Y = 1/4$.

(d) (5 points) Find the conditional PDF for X given that $Y = 3/4$, i.e., $f_{X|Y}(x \mid 3/4)$.

Problem 3. (25 points)
Consider a Markov chain X_n whose one-step transition probabilities are shown in the figure.



(a) (5 points) What are the recurrent states?

(b) (5 points) Find $\mathbf{P}(X_2 = 4 \mid X_0 = 2)$.

(c) (5 points) Suppose that you are given the values of $r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i)$. Give a formula for $r_{11}(n+1)$ in terms of the $r_{ij}(n)$.

(d) (5 points) Find the steady-state probabilities $\pi_j = \lim_{n\to\infty} \mathbf{P}(X_n = j \mid X_0 = i)$, or explain why they do not exist.

(e) (5 points) What is the probability of eventually visiting state 4, given that the initial state is $X_0 = 1$?

Problem 4. (30 points)
Al, Bonnie, and Clyde run laps around a track, with the duration of each lap (in hours) being exponentially distributed with parameters $\lambda_A = 21$, $\lambda_B = 23$, and $\lambda_C = 24$, respectively. Assume that all lap durations are independent. At the completion of each lap, a runner drinks either one or two cups of water, with probabilities 1/3 and 2/3, respectively, independent of everything else, including how much water was consumed after previous laps. (The time spent drinking is negligible, assumed zero.)

(a) (5 points) Write down the PMF of the total number of completed laps over the first hour.

(b) (5 points) What is the expected number of cups of water to be consumed by the three runners, in total, over the first hour.

(c) (5 points) Al has amazing endurance and completed 72 laps. Find a good approximation for the probability that he drank at least 130 cups. (You do not have to use 1/2-corrections.)

(d) (5 points) What is the probability that Al finishes his first lap before any of the others?

(e) (5 points) Suppose that the runners have been running for a very long time when you arrive at the track. What is the distribution of the duration of Al's current lap? (This includes the duration of that lap both before and after the time of your arrival.)

(f) (5 points) Suppose that the runners have been running for 1/4 hours. What is the distribution of the time Al spends on his second lap, given that he is on his second lap?

Problem 5. (25 points)
A pulse of light has energy X that is a second-order Erlang random variable with parameter λ, i.e., its PDF is

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & \text{for } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

This pulse illuminates an ideal photon-counting detector whose output N is a Poisson-distributed random variable with mean x when $X = x$, i.e., its conditional PMF is

$$p_{N|X}(n \mid x) = \begin{cases} \dfrac{x^n e^{-x}}{n!}, & \text{for } n = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

(a) (5 points) Find $\mathbf{E}[N]$ and $\mathrm{Var}[N]$, the unconditional mean and variance of N

(b) (5 points) Find $p_N(n)$, the unconditional PMF of N.

(c) (5 points) Find $\hat{X}_{\text{lin}}(N)$, the linear least-squares estimator of X based on an observation of N.

(d) (5 points) Find $\hat{X}_{\text{MAP}}(N)$, the MAP estimator of X based on an observation of N.

(e) (5 points) Instead of the prior distribution in Eq. (1), we are now told that

$$\mathbf{P}(X=2)=3^3/35, \qquad \mathbf{P}(X=3)=2^3/35.$$

Given the observation $N=3$, and in order to minimize the probability of error, which one of the two hypotheses $X=2$ and $X=3$ should be chosen?

Useful integral and facts:

$$\int_0^\infty y^k e^{-\alpha y}\,dy=\frac{k!}{\alpha^{k+1}}, \quad \text{for } \alpha>0 \text{ and } k=0,1,2,\ldots \text{ (recall that 0!=1)}$$

The second-order Erlang random variable satisfies:

$$\mathbf{E}[X]=2/\lambda, \qquad \mathrm{Var}(X)=2/\lambda^2.$$

is useful

why do they do this preview

- don't miss on exam

Each question is repeated in the following pages. Please write your answer on the appropriate page.

On own - after review session + reviewing notes

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2009)

Problem 2. (20 points)
A pair of jointly continuous random variables, X and Y, have a joint probability density function given by

no calculators

$$f_{X,Y}(x,y) = \begin{cases} c, & \text{in the shaded region of Fig. 1} \\ 0, & \text{elsewhere.} \end{cases}$$

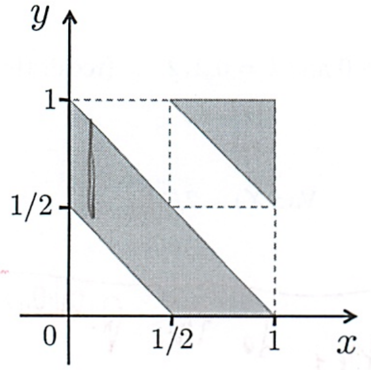


Figure 2: The shaded region is the domain in which $f_{X,Y}(x,y) = c$.

(a) (5 points) Find c.

based on area

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

$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

$\frac{1}{2} - \frac{1}{8} + \frac{1}{8}$

$\frac{1}{2}$

so $c = 2$

(b) (5 points) Find the marginal PDFs of X and Y, i.e., $f_X(x)$ and $f_Y(y)$.

flatten

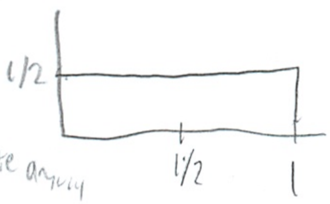
but which way

X ↓↓↓

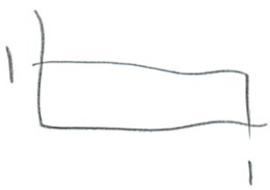
Y ←←←

$f_X(x)$

have a



to renormalize anyway



✓

Same

b/c extra triangle

$f_X(x)$ normalize

write at

$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

$f_Y(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

(c) (5 points) Find $\mathbf{E}[X \mid Y = 1/4]$ and $\mathrm{Var}[X \mid Y = 1/4]$, that is, the conditional mean and conditional variance of X given that $Y = 1/4$.

So at $y = 1/4$ what is $E[x]$

Uniform
but what are the values

$(\frac{1}{4}, \frac{1}{4})$ $(\frac{3}{4}, \frac{1}{4})$

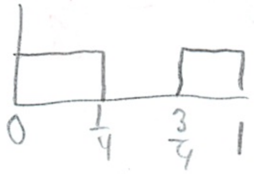
$\frac{1}{2}$ ✓

Var = uniform $\frac{(b-a)^2}{12}$

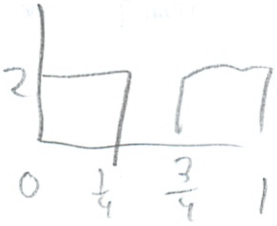
$\frac{\left(\frac{3}{4} - \frac{1}{4}\right)^2}{12}$

$\frac{\left(\frac{1}{2}\right)^2}{12} = \frac{\frac{1}{4}}{12} = \frac{1}{4 \cdot 12} = \frac{1}{48}$ ✓

(d) (5 points) Find the conditional PDF for X given that $Y = 3/4$, i.e., $f_{X|Y}(x \mid 3/4)$.



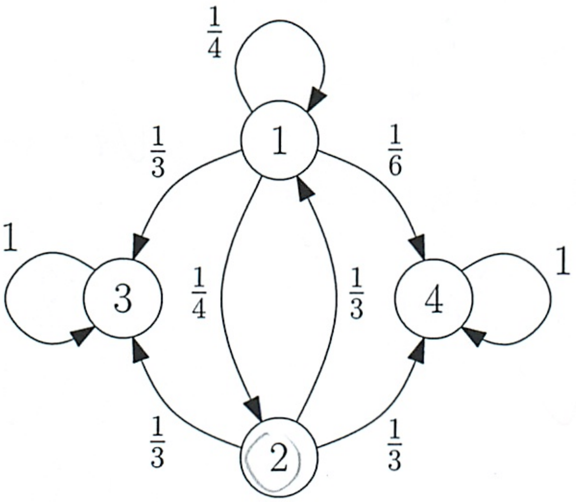
normalize



Write it out $f_{x|y}(x|\frac{3}{4}) = \begin{cases} 2 & \text{when } 0 \le x \le \frac{1}{4} \text{ and } \frac{3}{4} \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ ✓

That went much faster than in review session

Problem 3. (25 points)
Consider a Markov chain X_n whose one-step transition probabilities are shown in the figure.



(a) (5 points) What are the recurrent states?

3, 4

✓

(b) (5 points) Find $\mathbf{P}(X_2 = 4 \mid X_0 = 2)$.

$P_{24}(2) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{6} = \frac{7}{18}$ ✓

(no other possibilities

must not simplify
don't rush algebra

(c) (5 points) Suppose that you are given the values of $r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i)$. Give a formula for $r_{11}(n+1)$ in terms of the $r_{ij}(n)$.

abstract qu

$$r_{11}(n+1) = \sum_{k=1}^{m} r_{1k}(n)\, p_{k1}$$

~~[illegible]~~ fill in ~~ex~~ ← ✓except

$$= \sum_{j=1}^{4} p_{ij}\, r_{j1}(n)$$

Since 3,4 asorbing "simplifies" to

$$= \frac{1}{4} r_{11}(n) + \frac{1}{4} r_{21}(n)$$

← would you have had to include that

(d) (5 points) Find the steady-state probabilities $\pi_j = \lim_{n\to\infty} \mathbf{P}(X_n = j \mid X_0 = i)$, or explain why they do not exist.

Well multiple recurrent classes, so generally not easy to find depend on start state

~~- instead it is the asorbtion prob~~ - don't think have to do - do for practice

~~- add all the lines to it~~

don't need - wrong thing

~~$a_i = \sum_j p_{ij} a_j$ for all i~~ ← what is this again - add going to it

Book all the starts

~~need to define to which one~~

~~$a_2 = \frac{1}{3} a_3 + \frac{1}{3} a_1 + \frac{1}{3} a_4$~~ ↑0 ↑1

~~$a_1 = \frac{1}{4} a_3 + \frac{1}{4} a_3 + \frac{1}{8} a_4$~~

☆ Prob that are asorbed in given state, given starting state

(e) (5 points) What is the probability of eventually visiting state 4, given that the initial state is $X_0 = 1$?

This is p asorbed in 4, given start at 1

aka $a_1 = ?$

$a_4 = 1$

$a_3 = 0$

$$a_1 = \frac{1}{3}\underset{\uparrow 0}{a_3} + \frac{1}{4}a_2 + \frac{1}{4}a_1 + \frac{1}{6}\underset{\uparrow 1}{a_4}$$

↑must include self

$$a_2 = \frac{1}{3}\underset{\uparrow 0}{a_3} + \frac{1}{3}a_1 + \frac{1}{3}a_4$$

$$a_1 = \frac{1}{4}a_2 + \frac{1}{4}a_1 + \frac{1}{6}$$

$$a_2 = \frac{1}{3}a_1 + \frac{1}{3}$$

– solve for a_1

$$a_1 = \frac{1}{4}\left(\frac{1}{3}a_1 + \frac{1}{3}\right) + \frac{1}{4}a_1 + \frac{1}{6}$$

$$a_1 = \frac{1}{12}a_1 + \frac{1}{12} + \frac{1}{4}a_1 + \frac{1}{6}$$

$$a_1 - \frac{1}{12}a_1 - \frac{1}{4}a_1 = \frac{1}{6} + \frac{1}{12}$$

$$\frac{12}{12} - \frac{1}{12} - \frac{3}{12}$$

$$\frac{8}{12}a_1 = \frac{3}{12}$$

$$\cdot\frac{12}{8} \qquad \cdot\frac{12}{8}$$

$$a_1 = \frac{3}{12}\cdot\frac{12}{8} = \boxed{\frac{3}{8}}$$ ✓

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Department of Electrical Engineering & Computer Science

6.041/6.431: Probabilistic Systems Analysis

(Fall 2009)

Problem 4. (30 points) poisson — here it gets tricky

Al, Bonnie, and Clyde run laps around a track, with the duration of each lap (in hours) being exponentially distributed with parameters $\lambda_A = 21$, $\lambda_B = 23$, and $\lambda_C = 24$, respectively. Assume that all lap durations are independent. At the completion of each lap, a runner drinks either one or two cups of water, with probabilities 1/3 and 2/3, respectively, independent of everything else, including how much water was consumed after previous laps. (The time spent drinking is negligible, assumed zero.)

(remember from review)

(a) (5 points) Write down the PMF of the total number of completed laps over the first hour.

So merge the 3 poisson processes

$\sim$ poisson$(\lambda_A + \lambda_B + \lambda_C)$

$\sim$ poisson$(21+23+24)$

$\sim$ poisson(68)

arrivals in n time slots = binomial

~~X~~ want $E[T] = \lambda T$

$= 68 \cdot 1$

$= \boxed{68}$

No they do want the PMF

$$p(k,1) = \begin{cases} \dfrac{(68\cdot 1)^k e^{-68\cdot 1}}{k!} & k=0,1,\ldots \\ 0 & \text{otherwise} \end{cases}$$

and they have k as l

(b) (5 points) What is the expected number of cups of water to be consumed by the three runners, in total, over the first hour. E[

$\begin{cases} 1 & w/p=\frac{1}{3} \\ 2 & w\ p=\frac{2}{3} \end{cases}$ at each arrival

E[# arrivals] · E[drinks per arrival]

$\lambda T \cdot \frac{1}{3}\cdot 1 + \frac{2}{3}\cdot 2$

$68 \cdot \frac{5}{3}$

$113\frac{1}{3}$ ✓

$\frac{340}{3}$

They did it more formally than I did

$E[C] = E[E[C|L]] = E[LC_i]$

$= E[L]\cdot E[C_i]$

↑ I skipped to this

I remember these iterated expectation qu

(c) (5 points) Al has amazing endurance and completed 72 laps. Find a good approximation for the probability that he drank at least 130 cups. (You do not have to use 1/2-corrections.)

Ok - Can I do it ↑ hint to use CLT

$$\Phi\left(\frac{130-72E[x]}{\sqrt{n}\,\sigma_x}\right)$$

$E[x]=\frac{5}{3}$

$Var(x)=\frac{1}{3}\left(1-\frac{5}{3}\right)^2+\frac{2}{3}\left(2-\frac{5}{3}\right)^2$

$\sigma = var(x)^2$

$$\Phi\left(\frac{130-72\cdot\frac{5}{3}}{\sqrt{72}\cdot\frac{4}{81}}\right)$$

$Var(x)=\frac{4}{27}+\frac{2}{27}=\frac{6}{27}=\frac{2}{9}$

...

⊗ more complex

- Seperate paper

$\sigma=\left(\frac{2}{9}\right)^2=\frac{4}{81}$

(d) (5 points) What is the probability that Al finishes his first lap before any of the others?

-remember

P(arrival in A before B, C)

So p first arrival is A

$$\frac{\lambda_A}{\lambda_A+\lambda_B+\lambda_C}=\frac{21}{62}$$ ✓

c) X = # of laps out of 72 where Al drank 2 cups

In order for him to drink at least 130, must have

$$1(72 - x) + 2 \cdot x > 130$$

That implies

$$x \geq 58$$

← and why do we have to do this? ¿ to get it binomial?

X_i = iid RV of # of cups

↳ binomial $n = 72$ $p = 2/3$

So actual is $P(x \geq 58) = \sum_{k=58}^{72} \binom{72}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{72-k}$

But thats hard so CLT

↳ normal RV mean $= np = 72 \cdot \frac{2}{3} = 48$

var $np(1-p) = 16$

← oh special for Binomial - wrote on cheat sheet. but did not use now

$$P(x \geq 58) = 1 - P(x < 58)$$

$$= 1 - \Phi\left(\frac{58-48}{\sqrt{16}}\right)$$

$$= 1 - \Phi(2.5)$$

$$\approx .0062$$

Getting things wrong undermines my confidence

↙ steady state

(e) (5 points) Suppose that the runners have been running for a very long time when you arrive at the track. What is the distribution of the duration of Al's current lap? (This includes the duration of that lap both before and after the time of your arrival.)

-remem

? derived dist

erlang order 2

$$f_{Y_2}(y) = \frac{68^2 y^1 e^{-68 y}}{1!}$$

$\otimes$

✓ just al -not all runners

Close thought error

$$= \begin{cases} 68^2 y e^{-68 y} & y \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

→ must say 0 otherwise!

(f) (5 points) Suppose that the runners have been running for 1/4 hours. What is the distribution of the time Al spends on his second lap, given that he is on his second lap?

I remember this was the very tricky problem

Skip for now

~~just skip - will never get challenge anyways~~

Seperate paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2009)

Problem 5. (25 points) did not do in review session

A pulse of light has energy X that is a second-order Erlang random variable with parameter λ, i.e., its PDF is

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & \text{for } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (1)$$

This pulse illuminates an ideal photon-counting detector whose output N is a Poisson-distributed random variable with mean x when $X = x$, i.e., its conditional PMF is

$$p_{N|X}(n \mid x) = \begin{cases} \dfrac{x^n e^{-x}}{n!}, & \text{for } n = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

? Inference
- always be able to identify problem

$$X \rightarrow \boxed{p_{N|X}(n|x)} \xrightarrow{0} \boxed{\text{estimator}} \rightarrow$$

??

Useful integral and facts:

$$\int_0^\infty y^k e^{-\alpha y}\, dy = \frac{k!}{\alpha^{k+1}}, \quad \text{for } \alpha > 0 \text{ and } k = 0, 1, 2, \ldots \text{ (recall that 0!=1)}$$

← how does that help

need obs?

The second-order Erlang random variable satisfies:

$$\mathbf{E}[X] = 2/\lambda, \qquad \text{Var}(X) = 2/\lambda^2.$$

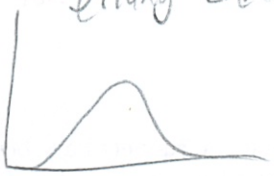
- ? CDF
? convolution - why can't I get started

(a) (5 points) Find $\mathbf{E}[N]$ and $\text{Var}[N]$, the unconditional mean and variance of N

erlang order 2

$$f_{X|N}(x|n) = \frac{f_X(x) \cdot p_{N|X}(n|x)}{f_N(n)}$$ ← just want

$$= \sum f_X(x)\, p_{N|X}(n|x)$$



Each moment $\lambda = \dfrac{X^n e^{-X}}{n!}$ ↙ does depend on X

Law of iterated exp.

$$E[N] = E[E[N|X]]$$ ← duh

$$= E[X]$$

$$= \frac{2}{\lambda}$$

never get this

even after spending time on it today

Law of iterated var

$$Var(N) = E[var(N|X)] + var(E[N|X])$$

$$= E[X] + var(X)$$

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

where. $var(N|X) = E[N|X] = X$

f) First work w/ CDF

$$P(X \leq x) = P(\text{The 1st to arrive occured less than } x \text{ hrs from time } \tfrac{1}{4})$$

$$= \frac{P(\text{1 arrival in interval } [\frac{1}{4}-x, \frac{1}{4}] \text{ and no } \overset{\text{arrivals}}{} [0, \frac{1}{4}-x)}{P(\text{1 arrival } [0, \frac{1}{4}])}$$

$$= \frac{P(\text{1 arrival } [\frac{1}{4}-x, \frac{1}{4}]) \cdot P(\text{no arrivals } [0, \frac{1}{4}-x))}{P(\text{1 arrival } [0, \frac{1}{4}])}$$

$$= \frac{P(1, x)\, P(0, \frac{1}{4}-x)}{P(1, 1/4)}$$

$$= \frac{e^{-21x}(21x)\, e^{-21(1/4-x)}}{e^{-21/4}(21/4)}$$

$$= \begin{cases} 0 & x < 0 \\ 4x & x \in [0, 1/4) \\ 1 & x > 1/4 \end{cases}$$

Find X is uniform over $[0, \frac{1}{4})$ w/ PDF

$$f_X(x) = \begin{cases} 4 & x \in [0, 1/4] \\ 0 & \text{otherwise} \end{cases}$$

②

Total time ~~spends~~ Al spends on 2nd lap → $T = X + Y$

- disjoint
- so <u>convolve</u>

$$f_T(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t-x)\, dx$$

$$= \int_0^{\min(1/4,t)} 4 \cdot 21 e^{-21(t-x)}\, dx$$

$$= \begin{cases} 4e^{-21t}\left(e^{21\min(1/4,t)} - 1\right) & t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Would never have gotton close!

(b) (5 points) Find $p_N(n)$, the unconditional PMF of N. See I was thinking last one was based on this

G - like last one

$$p_N(N) = \int_x \underbrace{f_x(x)}_{\text{Continuous}} p_{N|X}(n|x)\, dx \quad \checkmark \text{hard}$$

↑ ? but here discrete

$$= \int_{x=0}^{\infty} \frac{\lambda^2}{n!} x^{n+1} e^{-(1+\lambda)x}\, dx$$

← ~~did not do~~ ~~whatso~~ ever no clue

$$= \frac{\lambda^2}{n!} \cdot \frac{(n+1)!}{(1+\lambda)^{n+2}}$$

$$= \begin{cases} \frac{\lambda^2 (n+1)}{(1+\lambda)^{n+2}} & n = 0, 1, 2 \ldots \\ 0 & \text{otherwise} \end{cases}$$

(c) (5 points) Find $\hat{X}_{\text{lin}}(N)$, the linear least-squares estimator of X based on an observation of N.

$$\hat{X}_{LLMS}(N) = E[X] + \frac{\text{Cov}(N,X)}{\text{var}(N)} (N - E[N]) \quad \checkmark$$

↑ relies on previous ans

Did not try to fill in

Only unknown $\text{cov}(X,N) = E[XN] - E[X]E[N]$

$$= E[XN] - (E[X])^2$$

$$E[XN] = E[E[XN|X]] \quad \text{conditional ex.}$$

$$= E[X E[N|X]]$$

$$= E[X^2] = \text{var}(X) + (E[X])^2$$

$$= \frac{6}{\lambda^2}$$

$$\text{Cov}(X,N) = \frac{6}{\lambda^2} - \frac{4}{\lambda^2} = \frac{2}{\lambda^2}$$

$$\hat{X}_{Lin}(X) = \boxed{\frac{2}{\lambda} + \frac{2/\lambda^2}{\frac{2}{\lambda} + \frac{2}{\lambda^2}} \left(N - \frac{2}{\lambda}\right)}$$

not enough practice w/

(d) (5 points) Find $\hat{X}_{\text{MAP}}(N)$, the MAP estimator of X based on an observation of N.

would not have remembered

$$\hat{X}_{MAP}(N) = \arg\max_x f_{X|N}(x|N)$$

need N
need previous ans

$$= \arg\max_x \frac{f_X(x)\, p_{N|X}(n|x)}{p_N(n)}$$

$$= \arg\max_x f_X(x)\, p_{N|X}(n|x)$$

$$= \arg\max_x \frac{\lambda^2}{n!} x^{n+1} e^{-(1+\lambda)x}$$

$$= \arg\max_x x^{n+1} e^{-(1+\lambda)x}$$

→ differentiate, set = 0

$$\hat{X}_{MAP} = \frac{1+N}{1+\lambda}$$

Conclude is max by ...

) again need prior ans

(e) (5 points) Instead of the prior distribution in Eq. (1), we are now told that

$$\mathbf{P}(X = 2) = 3^3/35, \qquad \mathbf{P}(X = 3) = 2^3/35.$$

Given the observation $N = 3$, and in order to minimize the probability of error, which one of the two hypotheses $X = 2$ and $X = 3$ should be chosen?

Classical stats - choose hyp w/ largest posterior prob

$\hat{\theta}_{ML} = \arg\max_x (x; N)$ pick X that makes data most likely

Pick X = 2 - data value most likely ??

✓ but a lot of work

I think I got lucky

$$P(X=2 \mid N=3) > P(X=3 \mid N=3)$$

$$\frac{P(X=2)P(N=3|X=2)}{P(N=3)} > \frac{P(X=3)P(N=3|X=3)}{P(N=3)}$$

again real values →

$$P(X=2)P(N=3|X=2) > P(X=3)P(N=3|X=3)$$

$$\frac{3^3}{35} \cdot \frac{2^3 e^{-2}}{3!} > \frac{2^3}{35} \cdot \frac{3^3 e^{-3}}{3!}$$

$$e^{-2} > e^{-3}$$

I think may have gotten B on this (perhaps C) - based on dist B/C for year

Final Fall 09 Solutions

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2009)

Final Solutions:
December 15, 2009

Problem 2. (20 points)

(a) (5 points)
We're given that the joint PDF is constant in the shaded region, and since the PDF must integrate to 1, we know that the constant must equal 1 over the area of the region. Thus,

$$c = \frac{1}{1/2} = 2.$$

(b) (5 points)
The marginal PDFs of X and Y are found by integrating the joint PDF over all possible y's and x's, respectively. To find the marginal PDF of X, we take a particular value x and integrate over all possible y values in that vertical "slice" at $X = x$. Since the joint PDF is constant, this integral simplifies to just multiplying the joint PDF by the width of the "slice". Because the width of the slice is always $1/2$ for any $x \in [0, 1]$, we have that the marginal PDF of X is uniform over that interval:

$$f_X(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Since the joint PDF is symmetric, the marginal PDF of Y is also uniform:

$$f_Y(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) (5 points)
To find the conditional expectation and variance, first we need to determine what the conditional distribution is given $Y = 1/4$. At $Y = 1/4$, we take a horizontal slice of a uniform joint PDF, which gives us a uniform distribution over the interval $x \in [1/4, 3/4]$. Thus, we have

$$\mathbf{E}[X \mid Y = 1/4] = \frac{1}{2},$$

$$\text{var}(X \mid Y = 1/4) = \frac{(1/2)^2}{12} = \frac{1}{48}.$$

(d) (5 points)
At $Y = 3/4$, we have a horizontal slice of the joint PDF, which is nonzero when $x \in [0, 1/4] \cup [3/4, 1]$. Since the joint PDF is uniform, the slice will also be uniform, but only in the range of x where the joint PDF is nonzero (i.e. where (x, y) lies in the shaded region). Thus, the conditional PDF of X is

$$f_{X|Y}(x \mid 3/4) = \begin{cases} 2, & x \in [0, 1/4] \cup [3/4, 1], \\ 0, & \text{otherwise.} \end{cases}$$

Problem 3. (25 points)

(a) (5 points)
The recurrent states are {3,4}.

(b) (5 points)
The 2-step transition probability from State 2 to State 4 can be found by enumerating all the possible sequences. They are $\{2 \to 1 \to 4\}$ and $\{2 \to 4 \to 4\}$. Thus,

$$\mathbf{P}(X_2 = 4 \mid X_0 = 2) = \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot 1 = \frac{7}{18}.$$

(c) (5 points)
Generally,

$$r_{11}(n+1) = \sum_{j=1}^{4} p_{1j} r_{j1}(n).$$

Since states 3 and 4 are absorbing states, this expression simplifies to

$$r_{11}(n+1) = \frac{1}{4} r_{11}(n) + \frac{1}{4} r_{21}(n).$$

Alternatively,

$$\begin{aligned} r_{11}(n+1) &= \sum_{k=1}^{4} r_{1k}(n) p_{k1} \\ &= r_{11}(n) \cdot \frac{1}{4} + r_{12}(n) \cdot \frac{1}{3}. \end{aligned}$$

(d) (5 points)
The steady-state probabilities do not exist since there is more than one recurrent class. The long-term state probabilities would depend on the initial state.

(e) (5 points)
To find the probability of being absorbed by state 4, we set up the absorption probabilities. Note that $a_4 = 1$ and $a_3 = 0$.

$$\begin{aligned} a_1 &= \frac{1}{4} a_1 + \frac{1}{4} a_2 + \frac{1}{3} a_3 + \frac{1}{6} a_4 \\ &= \frac{1}{4} a_1 + \frac{1}{4} a_2 + \frac{1}{6} \end{aligned}$$

$$\begin{aligned} a_2 &= \frac{1}{3} a_1 + \frac{1}{3} a_3 + \frac{1}{3} a_4 \\ &= \frac{1}{3} a_1 + \frac{1}{3} \end{aligned}$$

Solving these equations yields $a_1 = \frac{3}{8}$.

Problem 4. (30 points)

(a) (5 points)
Given the problem statement, we can treat Al, Bonnie, and Clyde's running as 3 independent Poisson processes, where the arrivals correspond to lap completions and the arrival rates indicate the number of laps completed per hour. Since the three processes are independent, we can merge them to create a new process that captures the lap completions of all three runners. This merged process will have arrival rate $\lambda_M = \lambda_A + \lambda_B + \lambda_C = 68$. The total number of completed laps, L, over the first hour is then described by a Poisson PMF with $\lambda_M = 68$ and $\tau = 1$:

$$p_L(\ell) = \begin{cases} \frac{68^\ell e^{-68}}{\ell!}, & \ell = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

(b) (5 points)
Let L be the total number of completed laps over the first hour, and let C_i be the number of cups of water consumed at the end of the ith lap. Then, the total number of cups of water consumed is

$$C = \sum_{i=1}^{L} C_i,$$

which is a sum of a random number of i.i.d. random variables. Thus, we can use the law of iterated expectations to find

$$\mathbf{E}[C] = \mathbf{E}[\mathbf{E}[C \mid L]] = \mathbf{E}[LC_i] = \mathbf{E}[L]\mathbf{E}[C_i] = (\lambda_M \tau) \cdot \left(1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3}\right) = 68 \cdot \frac{5}{3} = \frac{340}{3}.$$

(c) (5 points)
Let X be the number of laps (out of 72) after which Al drank 2 cups of water. Then, in order for him to drink at least 130 cups, we must have

$$1 \cdot (72 - X) + 2 \cdot X \geq 130,$$

which implies that we need

$$X \geq 58.$$

Now, let X_i be i.i.d. Bernoulli random variables that equal 1 if Al drank 2 cups of water following his ith lap and 0 if he drank 1 cup. Then

$$X = X_1 + X_2 + \cdots + X_{72}.$$

X is evidently a binomial random variable with $n = 72$ and $p = 2/3$, and the probability we are looking for is

$$\mathbf{P}(X \geq 58) = \sum_{k=58}^{72} \binom{72}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{72-k}.$$

This expression is difficult to calculate, but since we're dealing with the sum of a relatively large number of i.i.d. random variables, we can invoke the Central Limit Theorem to approximate this probability using a normal distribution. In particular, we can approximate X as a normal random variable with mean $np = 72 \cdot 2/3 = 48$ and variance $np(1-p) = 16$ and approximate the desired probability as

$$\mathbf{P}(X \geq 58) = 1 - \mathbf{P}(X < 58) \approx 1 - \Phi\left(\frac{58-48}{\sqrt{16}}\right) = 1 - \Phi(2.5) \approx 0.0062.$$

(d) (5 points)
The event that Al is the first to finish a lap is the same as the event that the first arrival in the merged process came from Al's process. This probability is

$$\frac{\lambda_A}{\lambda_A + \lambda_B + \lambda_C} = \frac{21}{68}.$$

(e) (5 points)
This is an instance of the random incidence paradox, so the duration of Al's current lap consists of the sum of the duration from the time of your arrival until Al's next lap completion and the duration from the time of your arrival back to the time of Al's previous lap completion. This is the sum of 2 independent exponential random variables with parameter $\lambda_A = 21$ (i.e. a second- order Erlang random variable):

$$f_T(t) = \begin{cases} 21^2 t e^{-21t}, & t \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

(f) (5 points)
As in the previous part, the duration of Al's second lap consists of the time remaining from $t = 1/4$ until he completes his second lap and the time elapsed since he began his second lap until $t = 1/4$. Let X be the time elapsed and Y be the time remaining. We can still model the time remaining Y as an exponential random variable. However, we can no longer do the same for the time elapsed X because we know X can be no larger than $1/4$, whereas the exponential random variable can be arbitrarily large.

To find the PDF of X, let's first consider its CDF.

$$\begin{aligned}
\mathbf{P}(X \leq x) &= \mathbf{P}(\text{The 1 arrival occurred less than } x \text{ hours ago from time } 1/4) \\
&= \frac{\mathbf{P}(1 \text{ arrival in the interval } [1/4 - x, 1/4] \text{ and no arrivals in the interval } [0, 1/4 - x])}{\mathbf{P}(1 \text{ arrival in the interval } [0, 1/4])} \\
&= \frac{\mathbf{P}(1 \text{ arrival in the interval } [1/4 - x, 1/4])\mathbf{P}(\text{no arrivals in the interval } [0, 1/4 - x])}{\mathbf{P}(1 \text{ arrival in the interval } [0, 1/4])} \\
&= \frac{\mathbf{P}(1, x)\mathbf{P}(0, 1/4 - x)}{\mathbf{P}(1, 1/4)} \\
&= \frac{e^{-21x}(21x)e^{-21(1/4-x)}}{e^{-21/4}(21/4)} \\
&= \begin{cases} 0, & x < 0, \\ 4x, & x \in [0, 1/4], \\ 1, & x > 1/4. \end{cases}
\end{aligned}$$

Thus, we find that the X is uniform over the interval $[0, 1/4]$, with PDF

$$f_X(x) = \begin{cases} 4, & x \in [0, 1/4], \\ 0, & \text{otherwise.} \end{cases}$$

The total time that Al spends on his second lap is $T = X + Y$. Since X and Y correspond to disjoint time intervals in the Poisson process, they are independent, and therefore we can use convolution to find the PDF of T:

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_X(x) f_Y(t-x)\,dx \\ &= \int_0^{\min(1/4,t)} 4 \cdot 21 e^{-21(t-x)}\,dx \\ &= \begin{cases} 4e^{-21t}\left(e^{21\min(1/4,t)} - 1\right), & t \geq 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Problem 5. (25 points)

(a) (5 points)
Using the law of iterated expectations and the law of total variance,

$$\begin{aligned} \mathbf{E}[N] &= \mathbf{E}[\mathbf{E}[N \mid X]] \\ &= \mathbf{E}[X] \\ &= \frac{2}{\lambda}, \end{aligned}$$

$$\begin{aligned} \text{var}(N) &= \mathbf{E}[\text{var}(N \mid X)] + \text{var}(\mathbf{E}[N \mid X]) \\ &= \mathbf{E}[X] + \text{var}(X) \\ &= \frac{2}{\lambda} + \frac{2}{\lambda^2}, \end{aligned}$$

where $\text{var}(N \mid X) = \mathbf{E}[N \mid X] = X$.

(b) (5 points)

$$\begin{aligned} p_N(n) &= \int_x f_X(x) p_{N|X}(n \mid x) dx \\ &= \int_{x=0}^{\infty} \frac{\lambda^2}{n!} x^{n+1} e^{-(1+\lambda)x} dx \\ &= \frac{\lambda^2}{n!} \cdot \frac{(n+1)!}{(1+\lambda)^{n+2}} \\ &= \begin{cases} \frac{\lambda^2(n+1)}{(1+\lambda)^{n+2}} & n = 0, 1, 2 \ldots \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(c) (5 points)
The equation for $\hat{X}_{\text{lin}}(N)$, the linear least-squares estimator of X based on an observation of N, is

$$\hat{X}_{\text{lin}}(N) = \mathbf{E}[X] + \frac{\text{cov}(X, N)}{\text{var}(N)}(N - \mathbf{E}(N)).$$

The only unknown quantity is $\mathrm{cov}(X,N) = \mathbf{E}[XN] - \mathbf{E}[X]\mathbf{E}[N] = \mathbf{E}[XN] - (\mathbf{E}[X])^2$. Using the law of iterated expectations again,

$$\begin{aligned}\mathbf{E}[XN] &= \mathbf{E}[\mathbf{E}[XN \mid X]] \\ &= \mathbf{E}[X\mathbf{E}[N \mid X]] \\ &= \mathbf{E}[X^2] = \mathrm{var}(X) + (\mathbf{E}[X])^2 \\ &= \frac{6}{\lambda^2}.\end{aligned}$$

Thus, $\mathrm{cov}(X,N) = 6/\lambda^2 - 4/\lambda^2 = 2/\lambda^2$. Combining this result with those from (a),

$$\begin{aligned}\hat{X}_{\mathrm{lin}}(N) &= \frac{2}{\lambda} + \frac{\frac{2}{\lambda^2}}{\frac{2}{\lambda} + \frac{2}{\lambda^2}}\left(N - \frac{2}{\lambda}\right) \\ &= \frac{2+N}{1+\lambda}.\end{aligned}$$

(d) (5 points)
The expression for $\hat{X}_{\mathrm{MAP}}(N)$, the MAP estimator of X based on an observation of N is

$$\begin{aligned}\hat{X}_{\mathrm{MAP}}(N) &= \arg\max_x f_{X|N}(x \mid n) \\ &= \arg\max_x \frac{f_X(x)p_{N|X}(n \mid x)}{p_N(n)} \\ &= \arg\max_x f_X(x)p_{N|X}(n \mid x) \\ &= \arg\max_x \frac{\lambda^2}{n!}x^{n+1}e^{-(1+\lambda)x} \\ &= \arg\max_x x^{n+1}e^{-(1+\lambda)x},\end{aligned}$$

where the third equality holds since $p_N(n)$ has no dependency on x and the last equality holds by removing all quantities that have no dependency on x. The max can be found by differentiation and the result is:

$$\hat{X}_{\mathrm{MAP}}(N) = \frac{1+N}{1+\lambda}.$$

This is the only local extremum in the range $x \in [0, \infty)$. Moreover, $f_{X|N}(x \mid n)$ equals 0 at $x = 0$ and goes to 0 as $x \to \infty$ and $f_{X|N}(x \mid n) > 0$ otherwise. We can therefore conclude that $\hat{X}_{\mathrm{MAP}}(N)$ is indeed a maximum.

(e) (5 points)
To minimize the probability of error, we choose the hypothesis that has the larger posterior

probability. We will choose the hypothesis that $X = 2$ if

$$\begin{aligned}
\mathbf{P}(X = 2 \mid N = 3) &> \mathbf{P}(X = 3 \mid N = 3) \\
\frac{\mathbf{P}(X = 2)\mathbf{P}(N = 3 \mid X = 2)}{\mathbf{P}(N = 3)} &> \frac{\mathbf{P}(X = 3)\mathbf{P}(N = 3 \mid X = 3)}{\mathbf{P}(N = 3)} \\
\mathbf{P}(X = 2)\mathbf{P}(N = 3 \mid X = 2) &> \mathbf{P}(X = 3)\mathbf{P}(N = 3 \mid X = 3) \\
\frac{3^3}{35} \cdot \frac{2^3 e^{-2}}{3!} &> \frac{2^3}{35} \cdot \frac{3^3 e^{-3}}{3!} \\
e^{-2} &> e^{-3}.
\end{aligned}$$

The inequality holds so we choose the hypothesis that $X = 2$ to minimize the probability of error.

6.041 Day Before Final

12/14

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

$P(A \cap B) = P(A) + P(B)$ if disjoint

Derived Distribution

PDF of RV X w/ known PDF $Y = g(X)$

Method Get CDF

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x \mid x(g) x \leq y} f_X(x)\,dx$$

differentiate

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

(so just apply function?
(?? only on the range?)

Special case if $Y = g(X) = aX + b$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right)$$

So this is dist after the function

②

<u>See example</u>

$X = \text{Uniform}[0,1]$

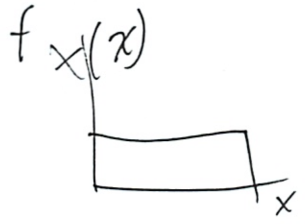
$Y = \sqrt{X}$

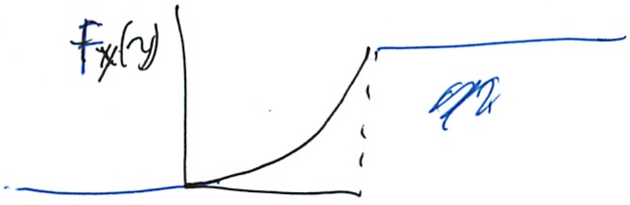
$F_Y(y) = P(Y \leq y) = P(\sqrt{X} \leq y) = P(X \leq y^2)$

$= y^2$

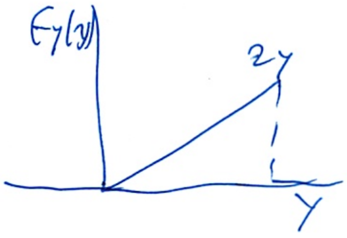
how to get here?

$\int_0^1 y = \frac{y^2}{2}$





$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d(y^2)}{dy} = 2y \quad 0 \leq y \leq 1$



③

CDF

$$P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt$$

So uniform 0 → 1

$$\int_{-\infty}^{x} \frac{1}{1}\,dt$$

x ✓

differenitate

1

same again

uniform 5→10

$f_X(x) = \frac{1}{10-5} = \frac{1}{5}$ from $5 \le x \le 10$

0 otherwise

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{5}\,dt$$

$$= \frac{x}{5}\Big|_{-\infty}^{x}$$

$\frac{x}{5} \neq \frac{x-a}{b-a} = \frac{x-5}{10-5} = \frac{x-5}{5}$ ⊗ does not match

do better bounds

$$\int_{5}^{10} \frac{1}{5}\,dt$$

④

$$\frac{x}{5}\Big|_{5}^{10}$$

$$\frac{10}{5}-\frac{5}{5} = 2-1$$
$$= 1$$

true at that pt
- but w/ variable

$$\frac{x}{5}\Big|_{5}^{x}$$ ← make sure to put in a <u>bottom band</u>

$$\frac{x}{5}-\frac{5}{5} = \frac{x-5}{5}$$ ✓ there we go

Now back to the other one

$\cancel{\int_0^{y^2}}$ Well y^2 is the <u>variable</u>

$$\int_a^{y^2} \frac{1}{b-a} = \int_0^{y^2} \frac{1}{1-0}\,dx$$

$$= x\Big|_0^{y^2}$$

$$y^2-0$$

$$y^2$$ ✓ there we go

then don't forget to differentiaty

$$\frac{d\,y^2}{dy} = \boxed{2y}$$

Convolution

$W = X + Y$ w/ x, y independent

- when need to convolve
  - ↳ Sum of 2 RV, independent

    (did a bit in 6.01)
- Flip + slide
- always a disaster
- get how to do it, but mistake city

Law of Iterated Expection

$E[X \mid Y = y]$ is a #

↳ I think you use this if you have conditional probs - like those examples yesterday

$E[X \mid Y] \neq f(Y)$ is a RV

↳ First compute $E[X \mid Y = y]$ as a function of y

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

↳ I remember this kinda from yesterday

6.

~~Lots~~ Lots of qv from chap 4 - special RV topics

↳ think b/c not much builds on

<u>Cov</u>

$\mathrm{Cov}(x,y) = 0$ uncorrelated

$\mathrm{Cov}(x,x) = \mathrm{var}(x)$

$\mathrm{cov}(x, aY+b) = a\,\mathrm{cov}(x,Y)$

$\mathrm{cov}(x, Y+Z) = \mathrm{cov}(x,Y) + \mathrm{cov}(x,Z)$

$\mathrm{Cov}(x,y) = E[XY] - E[X]\,E[Y]$

Now what is this $E[X \cdot Y]$ ← what is it even called?

wp: Marginal PDF

$$\int_X \int_Y x \cdot y \; p(x,y)\, dy\, dx$$

Oh here it is in book

$$E[XY] = \sum_Y y\, p_Y(y)\, E[X \mid Y=y]$$

total expectation theorem - but here are <u>ind</u>

What when not ind

Look at quiz solutions

$$E[XN] = E[E[XN|X]$$

$$= E[X\,E[N|X]$$

$$= E[X^2]$$

look in recent class/recitation notes from LLMs

$E[X(X+W_3)]$

ind here

"called 'cross second moment"

$= E[X^2] + E[X]\cancel{E}[W_3]$

web: normal 2nd moment $\int_a^b x^{\textcircled{2}} f(x)dx$

mean = first moment

but no normal second moment results

Var is <u>second moment</u>

---

here is a more general case: the Romeo + Juliet

$$E[X\theta] = E[E[X\theta|\theta]]$$

condition on θ - easier

first find w/ little θ, then replace w/ big θ

↓

⑧

$$E[x\Theta \mid \Theta = \theta] = E[\Theta x \mid \Theta = \theta]$$

$=$

$$= \theta \, E[x \mid \theta = \theta]$$

$$= \theta \cdot \frac{\theta}{2}$$

↑ in this case

$$E[x\Theta] = E[\tfrac{1}{2}\Theta^2]$$

$$= \frac{1}{2} \cdot \frac{1}{3}$$

↑ again from previous

$$= \frac{1}{6}$$

, so law of iterated expectation is best answer,

I will be Ok at this

Ok that was my punch list

Quiz 2 Mine Problems

X = continuous uniform $[0,4]$ } ind
$Y = \exp(2)$

1. ~~Mean~~ $E[X-3Y]$
 - derived dist
 - or convolution
 - gets messy, are some tricks can do 1st (remember)

$E[X] - 3E[Y]$ say

$\frac{4-0}{2} - 3\,\frac{1}{2}$

$2 - \frac{3}{2}$

$\boxed{\frac{1}{2}}$ ✓

2. $\mathrm{Var}(X - 3Y)$

$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\,\mathrm{cov}(X,Y)$

$\mathrm{var}(aX) = a^2\,\mathrm{var}(X)$

but what about the minus
 - roll into the $a = (-3)$ so squared it goes away

$\mathrm{var}(x) = \frac{(b-a)^2}{12} = \frac{(4-0)^2}{12} = \frac{16}{12} = \frac{4}{3}$

⑩

$var(y) = \frac{1}{\lambda^2} = \frac{1}{2^2} = \frac{1}{4}$

$cov(X,Y) = 0$ ind

$var(x+y) = var(x) + a^2 var(y)$

$= \frac{4}{3} + (-3)^2 \frac{1}{4}$

$= \frac{4}{3} + 9 \frac{1}{4}$

$= \frac{4}{3} + \frac{9}{4}$

$= \frac{43}{12}$ ✓ yeah I learned something did w/o looking

2. Find $P(Y \geq X)$

Now how to do this?

?, joint

total probability theorm

$P(Y \geq X) = \int_{x} f_X(x) P(Y \geq X | X = x) dx$

$= \int_0^4 \frac{1}{4} (1 - F_Y(x)) dx$

(ii)

$$= \int_0^4 \frac{1}{4} e^{-2x}\,dx$$

$$= \frac{1}{8}\int_0^4 2e^{-2x}dx$$

$$= \frac{1}{8}(1-e^{-8})$$

¿¿ what is this

Got 0 points on it

Some CDF > than

- like ta derived dist

¿shall I just do other practice test - material focuses on that

4. Find PDF $z = x + y$

- Convolution

3. Find conditional joint given event

$$f_{x,y|A}(x,y) = \frac{f_{x,y}(x,y)}{P(A)} = \frac{f_x(x)f_y(y)}{P(A)}$$

$$= \begin{cases} \frac{4e^{-2y}}{1-e^{-8}} & 0 \le x \le 4 \quad y \ge x \\ 0 & \text{otherwise} \end{cases}$$

will skip around

⑫

6. Find $E[Z \mid Y=y]$ and $E[Z \mid Y]$

Conditional dis is uniform b/w y y+4

So y+2 ← I saw on previous qu

and then as RV Y+2

7. Joint PDF $f_{Z,Y}$

$f_{Z|Y}(z|y)\, f_Y(y)$ ✓ or $f_{Y|Z}(y|z)\, f_Z(z)$

↑have ↑given — exp(2) = $\lambda e^{-\lambda x}$

$2e^{-2x}$ $\cancel{x} \geq 0$

~~4~~

~~$\begin{cases} \frac{1}{4} & y \leq z \leq y+4 \\ 0 & \text{else} \end{cases}$~~

~~↑how combine~~

f $\begin{cases} 2e^{-2x} & 0 \leq \cancel{x} \leq 4 \\ \frac{1}{4} & y \leq z \leq y+4 \end{cases}$

~~↑but z in other one~~
Can do I think

$$= \begin{cases} \frac{1}{2} e^{-2y} & y \geq 0,\ y \leq z \leq y+4 \\ 0 & \text{otherwise} \end{cases}$$

↑ 0 elsewhere

↑ only where they multiply together

$\frac{1}{4} \cdot 2 = \frac{1}{2}$

My other thought

⑬

8. - seperate problem?

$$W = QV$$

$$L \begin{cases} Y & w/ \ p \ 1/2 \\ Y+2 & w/ \ p \ 1/2 \end{cases}$$

$P(\text{heads} \mid W=3)$

$$X = \text{Bernolli} \begin{cases} 1 & w/ \ p \ 1/2 \\ 0 & w/ \ p \ 1/2 \end{cases}$$

$$f_{w|x=1}(w) = f_y(w)$$

$$f_{w|x=0}(w) = f_y(w-2)$$

Bayes

$$P(x=1 \mid W=3) = \frac{P(X=1) f_{w|x=1}(3)}{P(X=1) f_{w|x=1}(3) + P(X=0) f_{w|x=0}}$$

$$= \frac{P(x=1) f_y(3)}{P(x=1) f_y(3) + P(x=0) f_y(1)}$$

↑ back it out

$$= \frac{e^{-6}}{e^{-6} + e^{-2}}$$

⑭ 2. $X, X_1, X_2, \ldots$ normal RV mean 0
var 9

N = pos integer RV

$E[N] = 2 \quad E[N^2] = 5$

$$S = \sum_{i=1}^{N} X_i$$

- random sum of RVs

1. If σ is small pos $P(1 \leq |X| \leq 1+\sigma) \approx a\sigma$
for some constant a

- would have never gotten this

don't forget either side

$$P(1 \leq |x| \leq 1+\sigma) = 2P(1 \leq x \leq 1+\sigma)$$

$$\approx 2f_x(1)\sigma$$

↑ actually makes sense

$$a = 2f_x(1)$$

convert to #

$$= 2 \cdot \frac{1}{\sqrt{9 \cdot 2\pi}} e^{-1/2 \cdot \frac{(1-0)^2}{9}}$$

$$= \frac{2}{3\sqrt{2\pi}} e^{-\frac{1}{18}}$$

⑮

3. Are N, S uncorrelated

they do math to prove

$$Cov(S,N) = E[SN] - E[S]E[N]$$

$$= E[E[SN|N]] - E[E[S|N]]E[N]$$

$$= E[E[\sum_{i=1}^{N} X_i N|N]] - E[E[\sum_{i=1}^{N} X_i|N]]E[N]$$

$$= E[X_i]E[N^2] - E[X_i]E[N]$$

$$= 0$$

4. are Ind

- here I screwed up w/ wording
- I don't get ~~the~~ their eplination

12/14 Practice

6.041/6.431 Spring 2009 Final Exam
Thursday, May 21, 1:30 - 4:30 PM.

Name: ____________________

Recitation Instructor: ____________________

Question	Part	Score	Out of
0			2
1	all		18
2	all		24
3	a		4
	b		4
	c		4
4	a		6
	b		6
	c		6
5	a		6
	b		6
6	a		4
	b		4
	c		4
	d		5
	e		5
7	a		6
	b		6
Total			120

- Write your solutions in this quiz packet, only solutions in the quiz packet will be graded.
- You are allowed three two-sided 8.5 by 11 formula sheet plus a calculator.
- You have 180 minutes to complete the quiz.
- Be neat! You will not get credit if we can't read it.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 0 (2 pts.)
Write your name, and your assigned recitation instructor's name, on the cover of the quiz booklet. The Instructors are listed below.

Recitation Instructor	Recitation Time
Devavrat Shah	10 & 11 AM
Shivani Agarwal	11AM & 12PM
Asu Ozdaglar	12 & 1 PM
Pablo Parrilo (6.431)	10 & 11AM

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 1: True or False (2pts. each, 18 pts. total)

No partial credit will be given for individual questions in this part of the quiz. d not uniform

a. Let $\{X_n\}$ be a sequence of i.i.d random variables taking values in the interval $[0, 0.5]$. Consider the following statements:

will not converge - i 125 - but asks if - definition

(A) If $\mathbf{E}[X_n^2]$ converges to 0 as $n \to \infty$ then X_n converges to 0 in probability.

(B) If all X_n have $\mathbf{E}[X_n] = 0.2$ and $\text{var}(X_n)$ converges to 0 as $n \to \infty$ then X_n converges to 0.2 in probability.

going off intuition

(C) The sequence of random variables Z_n, defined by $Z_n = X_1 \cdot X_2 \cdots X_n$, converges to 0 in probability as $n \to \infty$.

T multiply

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

for all $\varepsilon > 0$ since $Z_n \le (1/2)^n$
$P(|Z_n - 0| \ge \varepsilon) = 0$ for $n > \frac{-\log \varepsilon}{\log 2}$

will never get these

A: F T	**B:** T ✓	**C:** F T

b. Let X_i $(i = 1, 2, \dots)$ be i.i.d. random variables with mean 0 and variance 2; Y_i $(i = 1, 2, \dots)$ be i.i.d. random variables with mean 2. Assume that all variables X_i, Y_j are independent. Consider the following statements:

something w/ 2

(A) $\frac{X_1 + \cdots + X_n}{n}$ converges to 0 in probability as $n \to \infty$.

(B) $\frac{X_1^2 + \cdots + X_n^2}{n}$ converges to 2 in probability as $n \to \infty$. still add to 0

(C) $\frac{X_1 Y_1 + \cdots + X_n Y_n}{n}$ converges to 0 in probability as $n \to \infty$. ~~mean~~

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: T	**B:** F F	**C:** F F

c. We have i.i.d. random variables $X_1 \dots X_n$ with an unknown distribution, and with $\mu = \mathbf{E}[X_i]$. We define $M_n = (X_1 + \ldots + X_n)/n$. Consider the following statements:

sample mean not mode

(A) M_n is a maximum-likelihood estimator for μ, irrespective of the distribution of the X_i's.

(B) M_n is a consistent estimator for μ, irrespective of the distribution of the X_i's.

(C) M_n is an asymptotically unbiased estimator for μ, irrespective of the distribution of the X_i's.

? can't tell

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: F ✓	**B:** T ✓	**C:** T F

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 2: Multiple Choice (4 pts. each, 24 pts. total)

Clearly circle the appropriate choice. **No partial credit** will be given for individual questions in this part of the quiz.

a. Earthquakes in Sumatra occur according to a Poisson process of rate $\lambda = 2$/year. Conditioned on the event that exactly two earthquakes take place in a year, what is the probability that both earthquakes occur in the first three months of the year? (for simplicity, assume all months have 30 days, and each year has 12 months, i.e., 360 days).

✓

- (i) 1/12
- [✓ (ii) 1/16]
- (iii) 64/225
- (iv) $4e^{-4}$
- (v) There is not enough information to determine the required probability.
- (vi) None of the above.

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

$P($2 arrivals in 60 days | 2 arrivals 360 days$)$

$= \frac{P(\text{2 arrivals in 60})}{P(\text{2 arrivals in 360})}$

$\frac{P(2;60)}{P(2,360)} = \frac{\frac{(2\cdot 60)^2 e^{-2\cdot 60}}{2}}{\frac{(2\cdot 360)^2 e^{-2\cdot 360}}{2}}$

Binomial

b. Consider a continuous-time Markov chain with three states $i \in \{1, 2, 3\}$, with dwelling time in each visit to state i being an exponential random variable with parameter $\nu_i = i$, and transition probabilities p_{ij} defined by the graph

↳ dwell time mean

$\frac{(120)^2 e^{-120}}{2} \cdot \frac{2}{720^2 e^{-720}}$

? seems to complex w/ou calc

$\frac{e^{600}}{36} \approx 1.04$

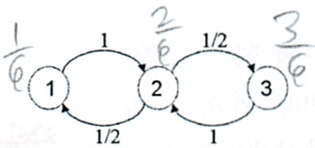
? why is it > 1 ?

? or smaller #s ?

$\frac{(2\cdot\frac{1}{4})^2 e^{2\cdot\frac{1}{4}}}{2^2 e^2}$

$\frac{\frac{1}{4}\cancel{e^2}}{4\cancel{e^2}} = \frac{1}{16}$

Why did that work differently?



What is the long-term expected fraction of time spent in state 2?

- (i) 1/2
- (ii) 1/4
- (iii) 2/5
- [✓ (iv) 3/7]
- [✓ (v) None of the above.]

← if all dwell times same $\pi_2 = ?$

$\pi_2 = 1\cdot\pi_1 + 1\cdot\pi_3$

$\pi_1 = .5\pi_2$

$\pi_3 = .5\pi_2$

$\pi_2 = .5\pi_2 + .5\pi_2$

$\pi_2 = \pi_2$

So much for that

↓ guessing ~~the~~ periodic-ish?

$\frac{1}{2}\cdot\frac{2}{6} = \frac{1}{6} = \frac{4}{24}$

$\frac{1}{4}\cdot\frac{1}{6} = \frac{1}{24}$

$\frac{1}{4}\cdot\frac{3}{6} = \frac{3}{24}$

normalize $\frac{4}{8} = \boxed{\frac{1}{2}}$

2b. $q_{ij} = V_i p_{ij}$

? no q on my cheat sheet

$q_{12} = q_{21} = q_{23} = 1$

$q_{32} = 3$

$\Big\}$ where did they get this?

birth-death Markov

$\pi_1 = \pi_2$

$\pi_2 = 3\pi_3$

normalize $\frac{3}{7}$

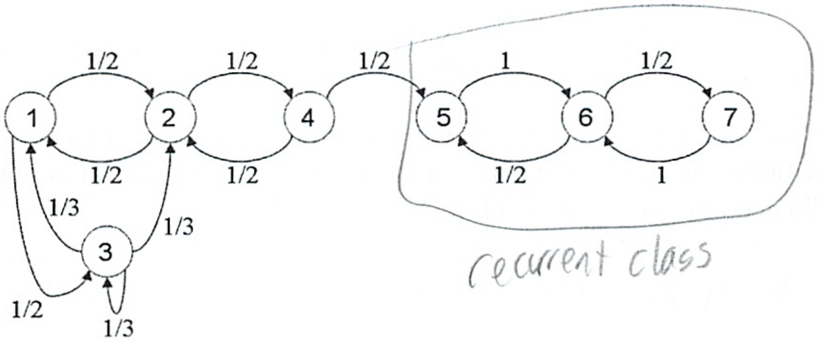
What is this q?

$V_{ij}(n)$ = expected # of visits to j within first n transitions starting from i

$q_{jk}(n)$ = expected # of transitions from j to k within n transitions do it before checking

– more realistic
does not get ya dan

c. Consider the following Markov chain:



does not matter where start
↳ unless part is recurrent

Starting in state 3, what is the steady-state probability of being in state 1?

π_1

(i) 1/3
(ii) 1/4
(iii) 1
[✓ (iv) 0] ✓
(v) None of the above.

d. Random variables X and Y are such that the pair (X, Y) is uniformly distributed over the trapezoid A with corners $(0,0)$, $(1,2)$, $(3,2)$, and $(4,0)$ shown in Fig. 1:

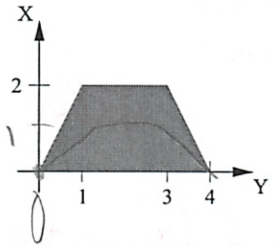


Figure 1: $f_{X,Y}(x,y)$ is constant over the shaded area, zero otherwise.

i.e.

$$f_{X,Y}(x,y) = \begin{cases} c, & (x,y) \in A \\ 0, & \text{else}. \end{cases}$$

We observe Y and use it to estimate X. Let $\hat{X}$ be the least mean squared error estimator of X given Y. What is the value of $\text{var}(\hat{X} - X|Y = 1)$? Cool - visual

but am so bad at

(i) 1/6 ✓ $\text{Var}(1 - X)$
(ii) 3/2 constant in var does not matter
[✓ (iii) 1/3] $\text{Var}(X)$ — uniform 0,2
(iv) The information is not sufficient to compute this value.
(v) None of the above. $\frac{(2-0)^2}{12}$ $\frac{1}{3}$

I no stupid math errors

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

---

e. $X_1 \dots X_n$ are i.i.d. normal random variables with mean value μ and variance v. Both μ and v are unknown. We define $M_n = (X_1 + \dots + X_n)/n$ and

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$$

(~~A~~ sample var)

except for

We also define $\Phi(x)$ to be the CDF for the standard normal distribution, and $\Psi_{n-1}(x)$ to be the CDF for the t-distribution with ~~$n-1$ degrees of freedom~~. Which of the following choices gives an exact 99% confidence interval for μ for all $n > 1$?

not responsible for

(i) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Phi(\delta) = 0.99$.

[✓ (ii)] $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Phi(\delta) = 0.995$.

(iii) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Psi_{n-1}(\delta) = 0.99$.

[✓ (iv)] $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Psi_{n-1}(\delta) = 0.995$.

(v) None of the above.

Out of scope

f. We have i.i.d. random variables X_1, X_2 which have an exponential distribution with unknown parameter θ. Under hypothesis H_0, $\theta = 1$. Under hypothesis H_1, $\theta = 2$. Under a likelihood-ratio test, the rejection region takes which of the following forms?

not very good at

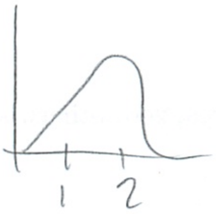
~~(i)~~ $R = \{(x_1, x_2) : x_1 + x_2 > \xi\}$ for some value ξ.

[✓ (ii)] $R = \{(x_1, x_2) : x_1 + x_2 < \xi\}$ for some value ξ.

(iii) $R = \{(x_1, x_2) : e^{x_1} + e^{x_2} > \xi\}$ for some value ξ.

(iv) $R = \{(x_1, x_2) : e^{x_1} + e^{x_2} < \xi\}$ for some value ξ.

(v) None of the above.



define $R = \{x = (x_1, x_2) \mid L(x) > \xi\}$

$$L(x) = \frac{f_X(x; H_1)}{f_X(x; H_0)} = \frac{\theta_1 e^{-\theta_1 x_1} e^{-\theta_1 x_2}}{\theta_2 e^{-\theta_0 x_1} e^{-\theta_0 x_2}}$$

$$= \frac{\theta_1^2}{\theta_3^2} e^{(\theta_0 - \theta_1)(x_1 + x_2)}$$

$$= 4e^{-(x_1 + x_2)}$$

Wald not def

So $R = \{(x_1, x_2) \mid x_1 + x_2 < -\log(\frac{\xi}{4})\}$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 3 (12 pts. total)

Aliens of two races (blue and green) are arriving on Earth independently according to Poisson process distributions with parameters λ_b and λ_a respectively. The Alien Arrival Registration Service Authority (AARSA) will begin registering alien arrivals soon.

Let T_1 denote the time AARSA will function until it registers its first alien. Let G be the event that the first alien to be registered is a green one. Let T_2 be the time AARSA will function until at least one alien of both races is registered.

(a) (4 points.) Express $\mu_1 = \mathbf{E}[T_1]$ in terms of λ_g and λ_b. **Show your work.**

merged poisson process

$$\frac{1}{\lambda_g+\lambda_B} \quad ✓$$

(b) (4 points.) Express $p = \mathbf{P}(G)$ in terms of λ_g and λ_b. **Show your work.**

$$\frac{\lambda_g}{\lambda_g+\lambda_B} \quad ✓$$

1

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

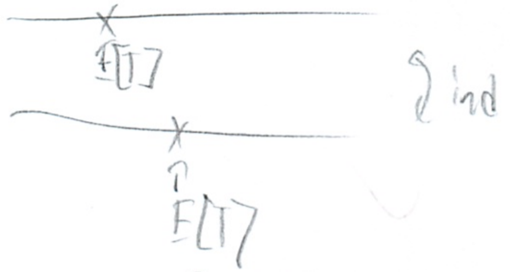
(c) (4 points.) Express $\mu_2 = \mathbf{E}[T_2]$ in terms of λ_g and λ_b. **Show your work.**

a bit more complex - since one of both

~~just~~ ~~first one is not~~ first

- look at each individual process

first one is just first
then second can be
same or diff



~~Must have~~ other one must have occured first

Or do look at merged

$P \frac{\lambda_g}{\lambda_g + \lambda_B}$

? some sum of both

$E[T] = \frac{1}{\lambda}$

$\max\{E[T_{1G}] = \frac{1}{\lambda_G}, E[T_{1B}] = \frac{1}{\lambda_B}\}$

? I think that's cheating, but clever

~~that was cheating~~

Oh they say this $\max(T_1^g, T_1^b)$ w/ PDF

Seems, so obvious now

$\mu_1 = E[T_1] = \frac{1}{\lambda_G + \lambda_B}$

Simply condition on 1st alien being Green or blue + memoryless property of Poisson

was kinda thinking this this but could not get it down on paper

$E[T_2] = E[T_1] + P(G) E[\text{Time until 1st } B | G] + P(G^c) E[\text{Time until 1st } G | B]$

$= E[T_1] + P(G) E[T_2^B) + (1 - P(G)) E[T_2^G]$ just add times

$\frac{1}{\lambda_B + \lambda_G} + \frac{\lambda_G}{\lambda_B + \lambda_G}\left(\frac{1}{\lambda_B}\right) + \frac{\lambda_B}{\lambda_B + \lambda_G}\left(\frac{1}{\lambda_G}\right)$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 4 (18 pts. total)

Researcher Jill is interested in studying employment in technology firms in Dilicon Valley. She denotes by X_i the number of employees in technology firm i and assumes that X_i are independent and identically distributed with mean p. To estimate p, Jill randomly interviews n technology firms and observes the number of employees in these firms.

(a) (6 points.) Jill uses

Estimation problem

Some sort of Chebsha Markov problem

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

true mean

some value

Only basic understanding

as an estimator for p. Find the limit of $\mathbf{P}(M_n \leq x)$ as $n \to \infty$ for $x < p$. Find the limit of $\mathbf{P}(M_n \leq x)$ as $n \to \infty$ for $x > p$. **Show your work.** $= a$

$$\lim_{n\to\infty} P(M_n \leq x) = 1 - P(M_n \geq x) = \frac{E[x)}{x}$$

$$\lim_{n\to\infty} P(|M_n - \mu) \leq x) = 1 - P(|M_n - \mu| \geq x) = \frac{\sigma^2}{x^2} \quad \text{for } x > p$$

Since M_i is iid M_n converges in prob - -

$$\lim_{n\to\infty} P(M_n \leq x) = \begin{cases} 0 & x < p \\ 1 & x > p \end{cases}$$

Oh just write this

this is likely Markov given others

(b) (6 points.) Find the smallest n, the number of technology firms Jill must sample, for which the Chebyshev inequality yields a guarantee

$$\mathbf{P}(|M_n - p| \geq 0.5) \leq 0.05.$$

↓error ↓confidence

$$\left(\frac{5}{100}\right)^2 = \frac{25}{10000} = ✓$$

Assume that $\text{var}(X_i) = v$ for some constant v. State your solution as a function of v. **Show your work.**

$$P(|M_n - p| \geq .5) \leq \frac{\sigma_{M_n}^2}{(\text{error})^2}$$

$$\leq \frac{\sigma_X^2}{n(\text{error}^2)}$$

$$\sigma_X^2 = \text{var}(X_i) = V$$

pick n so true

guessing - on Confidence

Fixed notes

$$.05 \leq \frac{V}{n\frac{25}{10{,}000}}$$

$$.05 = \frac{10{,}000V}{25n}$$

$$12.5n = 10000V$$

$$n = \frac{10000V}{12.5} = ~~800V~~$$

80V

~~Must be something w/ a decimal pt~~

(c) (6 points.) Assume now that the researcher samples $n = 5000$ firms. Find an approximate value for the probability

↓?, test or mean?

$$\mathbf{P}(|M_{5000} - p| \geq 0.5)$$

15 is really small???

using the Central Limit Theorem. Assume again that $\mathrm{var}(X_i) = v$ for some constant v. Give your answer in terms of v, and the standard normal CDF Φ. **Show your work.**

$$Z_n = \frac{S_n - E[S_n]}{\sigma_{S_n}} = \frac{S_n - nE[X]}{\sqrt{n}\,\sigma_X} \qquad \sigma_X = \sqrt{\mathrm{var}(X_i)} = \sqrt{v}$$

$$1 - \Phi\left(\frac{15 - 5000p}{\sqrt{5000}\sqrt{v}}\right)$$

$$P(|M_n - p| \geq .5) = 1 - P(|M_n - p| \leq .5)$$

✓

?? how discard stuff and n is given

But where is this from?

$$= P\left(\left|\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{nv}_{\,\varepsilon\, \mathrm{var}\, v}}\right| \geq \frac{.5\sqrt{n}}{\sqrt{v}}\right)$$

$$= 2 - 2\Phi\left(\frac{.5\sqrt{n}}{\sqrt{v}}\right)$$

because abs value

$$\frac{15 - np}{\sqrt{n}\sqrt{v}}$$

$$P(M_n - p \geq \epsilon) \leq 1 - \Phi(z)$$

$$= 1 - \Phi(2\epsilon\sqrt{n})$$

? where is this

where does $\sqrt{v}$ go?

I like this test so far (w/o ans) - seems fair + easish

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 5 (12 pts. total)

The RandomView window factory produces window panes. After manufacturing, 1000 panes were loaded onto a truck. The weight W_i of the i-th pane (in pounds) on the truck is modeled as a random variable, with the assumption that the W_i's are independent and identically distributed.

(a) (6 points.) Assume that the measured weight of the load on the truck was 2340 pounds, and that var $(W_i) \leq 4$. Find an approximate 95 percent confidence interval for $\mu = \mathbf{E}[W_i]$, using the Central Limit Theorem (you may use the standard normal table which was handed out with this quiz). **Show your work.**

bound

just copy

$$P\left(\mu - \frac{z\sqrt{4}}{\sqrt{1000}} < \theta < \mu + \frac{z\sqrt{4}}{\sqrt{1000}}\right) \approx 1-\alpha$$

$$z = 1 - \frac{\alpha}{2}$$

$\alpha = 5$

actually find sample mean estimator

$$\hat{\theta}_n = \frac{W_1 + \dots + W_n}{\quad}$$

$$\hat{\theta}_{1000} = \frac{2340}{1000} = 2.34$$

$\Phi(\underline{\quad}) = .975$

↑ 1.96 from table

? know how they did - if needed

$$P\left(\frac{|\hat{\theta}_{1000} - \mu|}{\sqrt{\mathrm{var}(W_i)/1000}} \leq 1.96\right) \approx .95$$

? plug in 4

→ get $\left[\hat{\theta}_{1000} - 1.96\sqrt{\frac{4}{1000}}, \hat{\theta}_{1000} + 1.96\sqrt{\frac{4}{1000}}\right]$

basically find this

(b) (6 points.) Now assume instead that the random variables W_i are i.i.d, with an exponential distribution with parameter $\theta > 0$, i.e., a distribution with PDF

$$f_W(w;\theta) = \theta e^{-\theta w}$$

convention

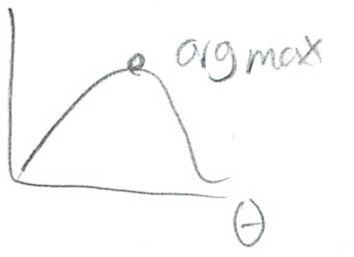
What is the maximum likelihood estimate of θ, given that the truckload has weight 2340 pounds? **Show your work.**

$$\hat{\theta}_{ML} = \underset{\theta}{\mathrm{argmax}}\ \theta e^{-\theta w}$$

? 2340

? which θ is most likely - try each one

$\theta e^{-\theta 2340}$



differentiate $1 - e^{\theta 2340}$

set = to 0 $\quad 0 = 1 - e^{2340\theta}$

$1 = e^{2340\theta}$

log log

$0 = 2340\theta$

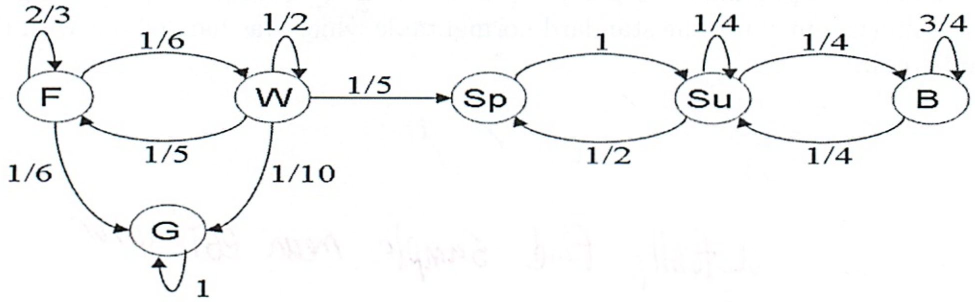
$\theta = 0$

oh - ML

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 6 (21 pts. total)

In Alice's Wonderland, there are six different seasons: Fall (F), Winter (W), Spring (Sp), Summer (Su), Bitter Cold (B), and Golden Sunshine (G). The seasons do not follow any particular order, instead, at the beginning of each day the Head Wizard assigns the season for the day, according to the following Markov chain model:



Thus, for example, if it is Fall one day then there is 1/6 probability that it will be Winter the next day (note that it is possible to have the same season again the next day). if $\circlearrowleft$

(a) (4 points.) For each state in the above chain, identify whether it is recurrent or transient. **Show your work.**

What is there to show

Sp, Su, B = recurrent class

G = recurrent class

F, W = transient ✓

(b) (4 points.) If it is Fall on Monday, what is the probability that it will be Summer on Thursday of the same week? **Show your work.**

0

M T W R
0 1 2 3

$r_{FS}(3) = ?$

$\frac{1}{6} \cdot \frac{1}{5} \cdot 1 = \frac{1}{30}$ ✓

Only 1 possible path

56.

(a) $f_w(w; \theta) = \prod_{i=1}^{n} f_{w_i}(w_i; \theta)$ ← oh that is a common formula?

$$= \prod_{i=1}^{n} \theta e^{-\theta w_i}$$

Then take log

$$\log f_w(w; \theta) = n \log \theta - \theta \sum_{i=1}^{n} w_i$$

Take deriv w/ respect to θ

$$\frac{n}{\theta} - \sum_{i=1}^{n} w_i$$

Set $= 0$, find max $\log f_w(w; \theta)$ over $\theta \geq 0$

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} w_i}$$

And finally plug in # don't forget

$$= \frac{1000}{2340} = .4274$$

hmh — reverse of before

— guess is accident

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

(c) (4 points.) If it is Spring today, will the chain converge to steady-state probabilities? If so, compute the steady-state probability for each state. If not, explain why these probabilities do not exist. **Show your work.**

Shold be all π

Yes - inside a reccurent class

$a_{sp} = \frac{1}{2} a_{su} + \cancel{\frac{1}{3} a_w}$

$a_{su} = a_{sp} + \frac{1}{4} a_{su} + \frac{1}{4} a_B$

$a_B = \frac{1}{4} a_{su} + \frac{3}{4} a_B$

Solve

$a_{su} = \frac{1}{2} a_{su} + \frac{1}{4} a_{su} + \frac{1}{4} a_b$

$\frac{1}{4} a_{su} = \frac{1}{4} a_b$

$\cdot 4 \quad \cdot 4$

$a_{su} = a_b$

$a_b = \frac{1}{4} a_b + \frac{3}{4} a_b$

$a_{sp} = \frac{1}{2} a_{su} = \frac{1}{2} a_b$

must add to 1 $a_{sp} + a_{su} + a_b = 1$

$a_{sp} = \frac{1}{5}$

$a_{su} = \frac{2}{5}$

$a_b = \frac{2}{5}$

✓

and show others $\pi_F = \pi_W = \pi_G = 0$

(d) (5 points.) If it is Fall today, what is the probability that Bitter Cold will never arrive in the future? **Show your work.**

P(asorbed into G) = a_F for G

$a_F = \frac{1}{6} \underset{1}{a_G} + \frac{2}{3} a_F + \frac{1}{6} a_w$ $\quad a_{sp} = a_{su} = a_B = 0$

$a_G = 1$

$a_w = \frac{1}{10} \underset{1}{a_g} + \frac{1}{2} a_w + \frac{1}{5} a_F$

$a_F = \frac{1}{6} + \frac{2}{3} a_F + \frac{1}{6} a_w$

$a_w = \frac{1}{10} + \frac{1}{2} a_w + \frac{1}{5} a_F$

$\frac{1}{3} a_F = \frac{1}{6} + \frac{1}{6} a_w$

$\frac{1}{2} a_w = \frac{1}{10} + \frac{1}{5} a_F$

$\cdot 2 \quad = \frac{1}{5} + \frac{2}{5} a_F$

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

$\frac{1}{3} a_F = \frac{1}{6} + \frac{1}{30} + \frac{2}{30} a_F$

$\frac{8}{30} a_F = \frac{6}{30}$

$\cdot \frac{30}{8} \quad \cdot \frac{30}{8}$

$a_F = \frac{6}{30} \cdot \frac{30}{8} = \frac{6}{8} = \frac{3}{4}$ ✓

woot - so happy Figured these out!

(e) (5 points.) If it is Fall today, what is the expected number of days till either Summer or Golden Sunshine arrives for the first time? **Show your work.**

$\mu_F = E[\text{asorbtion}]$

↳ either asorbtion

↳ with 1 +1 for Sp → Su

μ_F ✓ = ~~min(n≥0 such that $X_n = S$ | $X_0 = F$)~~

Wrong – that is mean 1st passage time

Misread cheatsheet

← where is this from?

$t_S = 0$ has nothing to do w/ – got confused

$t_i = 1 + \sum_j p_{ij} t_i$

$\mu_F = 1 + \frac{2}{3}\mu_F + \frac{1}{6}\mu_W + \frac{1}{6}0$

$\mu_W = 1 + \frac{1}{5}\mu_F + \frac{1}{2}\mu_W + \frac{1}{5}0$

didnot finish

$t_F = 1 + \frac{2}{3}t_F + \frac{1}{6}t_G + \frac{1}{6}t_W$

start

bingo 0

Solve

$\mu_F = 5.25$

↳ $\frac{1}{3}t_f = 1 + \frac{1}{6}t_w$

$t_f = 3 + \frac{1}{2}t_w$

Should have gotten

~~$t_G = 0$~~

$t_G = 0$

$t_W = 1 + \frac{1}{6}t_f + \frac{1}{2}t_w + \frac{1}{10}t_G$

0

Should really have gotten

$t_{sp} = 1 + \frac{1}{5}t_w$

↳ $\frac{1}{2}t_w = 1 + \frac{1}{6}t_f$

$t_w = 2 + \frac{1}{3}t_f$

? how do or

$t_{su} = 1 + t_{sp}$

$t_{su} = 2 + \frac{1}{5}t_w$

$= 2 + \frac{1}{5}\left(2 + \frac{1}{3}t_f\right)$

$= 2 + \frac{1}{5}\left(2 + \frac{1}{3}\left(3 + \frac{1}{2}t_w\right)\right)$

$= 2 + \frac{1}{5}\left(2 + 1 + \frac{1}{6}t_w\right)$

$= 2 + \frac{3}{5} + \frac{1}{30}t_w$

Hmm – check

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 7 (12 pts. total)

G G

A newscast covering the final baseball game between Sed Rox and Y Nakee becomes noisy at the crucial moment when the viewers are informed whether Y Nakee won the game.

Let a be the parameter describing the actual outcome: $a = 1$ if Y Nakee won, $a = -1$ otherwise. There were n viewers listening to the telecast. Let Y_i be the information received by viewer i ($1 \leq i \leq n$). Under the noisy telecast, $Y_i = a$ with probability p, and $Y_i = -a$ with probability $1 - p$. Assume that the random variables Y_i are independent of each other.

binomial

$E[\,] = p$

The viewers as a group come up with a joint estimator

Signal processing/inference

each viewer

↓ depending on mean

normalish

$$Z_n = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} Y_i \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

diff info Y_i - yeah

1 a win

(a) (6 points.)

Find $\lim_{n\to\infty} \mathbf{P}(Z_n = a)$ assuming that $p > 0.5$ and $a = 1$. **Show your work.**

CLT

$1 - \Phi\left(\frac{0 - np}{\sqrt{np(1-p)}}\right)$

$P(Y_i \geq 0)$

G Makov better

$P(Y_i \geq 0) \leq \frac{E[X] = p}{0.001}$ divide by 0

basically 0

so $p(\,)$ is $\textcircled{0}$

(b) (6 points.) Find $\lim_{n\to\infty} \mathbf{P}(Z_n = a)$, assuming that $p = 0.5$ and $a = 1$. **Show your work.**

exactly

$1 - \Phi\left(\frac{0 - np}{\sqrt{np(1-p)}}\right)$

7. a.

$$\lim_{n\to\infty} P(Z_n = 1) = \lim_{n\to\infty} P\left(\sum_{i=1}^{n} Y_i \geq 0\right) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} Y_i}{n} \geq 0\right)$$

I kinda over assumed all this

Since the Y_i are iid w/ mean $E[Y_i] = 2p-1$

$\text{var}(Y_i) = 1-(2p-1)^2$

One has by Chebshev inequality for all $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} Y_i}{n} - (2p-1)\right| \geq \varepsilon\right) = 0$$

How do you know when to use which?

↳ since asks for > ??

Take $\varepsilon = p - \frac{1}{2}$ and above is

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} Y_i}{n} \leq \frac{(2p-1)}{2}\right) = 0$$

So

$$\lim_{n\to\infty} P(Z_n = 1)$$

Would never get this

b)

Now $\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}} \geq 0\right)$

Since Y_i are iid w/ $E[Y_i]=0$

$Var(Y_i)=1$

Can approx $\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}}$ as Standard normal RV when $n \to \infty$

So $\lim_{n\to\infty} P(Z_n=1) = \boxed{\frac{1}{2}}$

?? are they using the table ??

Well it just falls about mean adds to $\frac{1}{2}$

Final Spring 09 Solutions

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2009)

Problem 1: True or False (2pts. each, 18 pts. total)

No partial credit will be given for individual questions in this part of the quiz.

a. Let $\{X_n\}$ be a sequence of i.i.d random variables taking values in the interval $[0, 0.5]$. Consider the following statements:

(A) If $\mathbf{E}[X_n^2]$ converges to 0 as $n \to \infty$ then X_n converges to 0 in probability.

(B) If all X_n have $\mathbf{E}[X_n] = 0.2$ and $\text{var}(X_n)$ converges to 0 as $n \to \infty$ then X_n converges to 0.2 in probability.

(C) The sequence of random variables Z_n, defined by $Z_n = X_1 \cdot X_2 \cdots X_n$, converges to 0 in probability as $n \to \infty$.

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: True	**B**: True	**C**: True

Solution:

(A) True. The fact that $\lim_{n\to\infty} \mathbf{E}[X_n^2] = 0$ implies $\lim_{n\to\infty} \mathbf{E}[X_n] = 0$ and $\lim_{n\to\infty} \text{var}(X_n) = 0$. Hence, one has

$$\begin{aligned}\mathbf{P}(|X_n - 0| \geq \epsilon) &\leq \mathbf{P}(|X_n - \mathbf{E}[X_n]| \geq \epsilon/2) + \mathbf{P}(|\mathbf{E}[X_n] - 0| \geq \epsilon/2) \\ &\leq \frac{\text{var}(X_n)}{(\epsilon/2)^2} + \mathbf{P}(|\mathbf{E}[X_n] - 0| \geq \epsilon/2) \to 0,\end{aligned}$$

where we have applied Chebyshev inequality.

(B) True. Applying Chebyshev inequality gives

$$\mathbf{P}(|X_n - \mathbf{E}[X_n]| \geq \epsilon) \leq \frac{\text{var}(X_n)}{\epsilon^2} \to 0.$$

Hence X_n converges to $\mathbf{E}[X_n] = 0.2$ in probability.

(C) True. For all $\epsilon > 0$, since $Z_n \leq (1/2)^n \Rightarrow \mathbf{P}(|Z_n - 0| \geq \epsilon) = 0$ for $n > -\log\epsilon/\log 2$.

b. Let X_i $(i = 1, 2, \ldots)$ be i.i.d. random variables with mean 0 and variance 2; Y_i $(i = 1, 2, \ldots)$ be i.i.d. random variables with mean 2. Assume that all variables X_i, Y_j are independent. Consider the following statements:

(A) $\frac{X_1+\cdots+X_n}{n}$ converges to 0 in probability as $n \to \infty$.

(B) $\frac{X_1^2+\cdots+X_n^2}{n}$ converges to 2 in probability as $n \to \infty$.

(C) $\frac{X_1Y_1+\cdots+X_nY_n}{n}$ converges to 0 in probability as $n \to \infty$.

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: True	B: True	C: True

Solution:

(A) True. Note that $\mathbf{E}[\frac{X_1+\cdots+X_n}{n}] = 0$ and $\mathrm{var}(\frac{X_1+\cdots+X_n}{n}) = \frac{n \cdot 2}{n^2} = \frac{2}{n}$. One can see $\frac{X_1+\cdots+X_n}{n}$ converges to 0 in probability.

(B) True. Let $Z_i = X_i^2$ and $\mathbf{E}[Z_i] = 2$. Note Z_i are i.i.d. since X_i are i.i.d., and hence one has that $\frac{Z_1+\cdots+Z_n}{n}$ converges to $\mathbf{E}[Z_i] = 2$ in probability by the WLLN.

(C) True. Let $W_i = X_i Y_i$ and $\mathbf{E}[W_i] = \mathbf{E}[X_i]\mathbf{E}[Y_i] = 0$. Note W_i are i.i.d. since X_i and Y_i are respectively i.i.d., and hence one has that $\frac{W_1+\cdots+W_n}{n}$ converges to $\mathbf{E}[W_i] = 0$ in probability by the WLLN.

c. We have i.i.d. random variables $X_1 \ldots X_n$ with an unknown distribution, and with $\mu = \mathbf{E}[X_i]$. We define $M_n = (X_1 + \ldots + X_n)/n$. Consider the following statements:

(A) M_n is a maximum-likelihood estimator for μ, irrespective of the distribution of the X_i's.

(B) M_n is a consistent estimator for μ, irrespective of the distribution of the X_i's.

(C) M_n is an asymptotically unbiased estimator for μ, irrespective of the distribution of the X_i's.

Which of these statements are **always** true? Write **True** or **False** in each of the boxes below.

A: False	B: True	C: True

Solution:

(A) False. Consider X_i follow a uniform distribution $U[\mu - \frac{1}{2}, \mu + \frac{1}{2}]$. The ML estimator for μ is any value between $\max(X_1, \cdots, X_n) - \frac{1}{2}$ and $\min(X_1, \cdots, X_n) + \frac{1}{2}$, instead of M_n.

(B) True. By the WLLN, M_n converges to μ in probability and hence it is a consistent estimator.

(C) True. Since $\mathbf{E}[M_n] = \mathbf{E}[X_i] = \mu$, M_n is unbiased estimator for μ and hence asymptotically unbiased.

Problem 2: Multiple Choice (4 pts. each, 24 pts. total)

Clearly circle the appropriate choice. **No partial credit** will be given for individual questions in this part of the quiz.

a. Earthquakes in Sumatra occur according to a Poisson process of rate $\lambda = 2/\text{year}$. Conditioned on the event that exactly two earthquakes take place in a year, what is the probability that both earthquakes occur in the first three months of the year? (for simplicity, assume all months have 30 days, and each year has 12 months, i.e., 360 days).

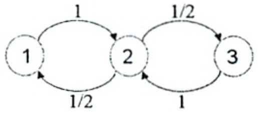
   (i) 1/12
   (ii) $\boxed{1/16}$
   (iii) 64/225
   (iv) $4e^{-4}$
   (v) There is not enough information to determine the required probability.
   (vi) None of the above.

   Solution: Consider the interval of a year be $[0, 1]$.

$$\begin{aligned}
\mathbf{P}\left(2 \text{ in } [0, \frac{1}{4}) \mid 2 \text{ in } [0, 1]\right) &= \frac{\mathbf{P}\left(2 \text{ in } [0, \frac{1}{4}), 0 \text{ in } [\frac{1}{4}, 1]\right)}{\mathbf{P}(2 \text{ in } [0, 1]))} \\
&= \frac{\frac{(\lambda \cdot 1/4)^2}{2!} e^{-\lambda \cdot 1/4} \cdot \frac{(\lambda \cdot 3/4)^0}{0!} e^{-\lambda \cdot 3/4}}{\frac{\lambda^2}{2!} e^{-\lambda}} \\
&= \frac{1}{16}
\end{aligned}$$

   (alternative explanation) Given that exactly two earthquakes happened in 12 months, each earthquake is equally likely to happen in any month of the 12, the probability that it happens in the first 3 months is $3/12 = 1/4$. The probability that both happen in the first 3 months is $(1/4)^2$.

b. Consider a continuous-time Markov chain with three states $i \in \{1, 2, 3\}$, with dwelling time in each visit to state i being an exponential random variable with parameter $\nu_i = i$, and transition probabilities p_{ij} defined by the graph



   What is the long-term expected fraction of time spent in state 2?

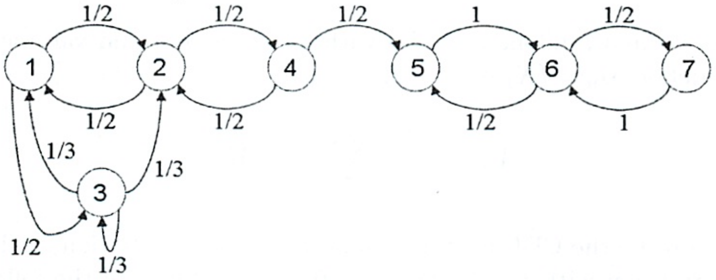
   (i) 1/2
   (ii) 1/4
   (iii) 2/5

(iv) $\boxed{3/7}$

(v) None of the above.

Solution: First, we calculate the $q_{ij} = \nu_i p_{ij}$, i.e., $q_{12} = q_{21} = q_{23} = 1$ and $q_{32} = 3$. The balance and normalization equations of this birth-death markov chain can be expressed as, $\pi_1 = \pi_2, \pi_2 = 3\pi_3$ and $\pi_1 + \pi_2 + \pi_3 = 1$, yielding $\pi_2 = 3/7$.

c. Consider the following Markov chain:



Starting in state 3, what is the steady-state probability of being in state 1?

(i) 1/3

(ii) 1/4

(iii) 1

(iv) $\boxed{0}$

(v) None of the above.

Solution: State 1 is transient.

d. Random variables X and Y are such that the pair (X, Y) is uniformly distributed over the trapezoid A with corners $(0,0)$, $(1,2)$, $(3,2)$, and $(4,0)$ shown in Fig. 1:

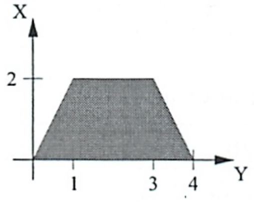


Figure 1: $f_{X,Y}(x,y)$ is constant over the shaded area, zero otherwise.

i.e.

$$f_{X,Y}(x,y) = \begin{cases} c\,, & (x,y) \in A \\ 0\,, & \text{else}\,. \end{cases}$$

We observe Y and use it to estimate X. Let $\hat{X}$ be the least mean squared error estimator of X given Y. What is the value of $\text{var}(\hat{X} - X | Y = 1)$?

(i) 1/6

(ii) 3/2

(iii) $\boxed{1/3}$

(iv) The information is not sufficient to compute this value.

(v) None of the above.

Solution: $f_{X|Y=1}(x)$ is uniform on $[0,2]$ therefore $\hat{X} = \mathbf{E}[X|Y=1] = 1$ and $\text{var}(\hat{X} - X|Y = 1) = \text{var}(X|Y=1) = (2-0)^2/12 = 1/3$.

e. $X_1 \ldots X_n$ are i.i.d. normal random variables with mean value μ and variance v. Both μ and v are unknown. We define $M_n = (X_1 + \ldots + X_n)/n$ and

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$$

We also define $\Phi(x)$ to be the CDF for the standard normal distribution, and $\Psi_{n-1}(x)$ to be the CDF for the t-distribution with $n-1$ degrees of freedom. Which of the following choices gives an exact 99% confidence interval for μ for all $n > 1$?

(i) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Phi(\delta) = 0.99$.

(ii) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Phi(\delta) = 0.995$.

(iii) $[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}]$ where δ is chosen to give $\Psi_{n-1}(\delta) = 0.99$.

(iv) $\boxed{[M_n - \delta\sqrt{\frac{V_n}{n}}, M_n + \delta\sqrt{\frac{V_n}{n}}] \text{ where } \delta \text{ is chosen to give } \Psi_{n-1}(\delta) = 0.995.}$

(v) None of the above.

Solution: See Lecture 23, slides 10-12.

f. We have i.i.d. random variables X_1, X_2 which have an exponential distribution with unknown parameter θ. Under hypothesis H_0, $\theta = 1$. Under hypothesis H_1, $\theta = 2$. Under a likelihood-ratio test, the rejection region takes which of the following forms?

(i) $R = \{(x_1, x_2) : x_1 + x_2 > \xi\}$ for some value ξ.

(ii) $\boxed{R = \{(x_1, x_2) : x_1 + x_2 < \xi\} \text{ for some value } \xi.}$

(iii) $R = \{(x_1, x_2) : e^{x_1} + e^{x_2} > \xi\}$ for some value ξ.

(iv) $R = \{(x_1, x_2) : e^{x_1} + e^{x_2} < \xi\}$ for some value ξ.

(v) None of the above.

Solution: We defined $R = \{x = (x_1, x_2) | L(x) > c\}$ where

$$L(x) = \frac{f_X(x; H_1)}{f_X(x; H_0)} = \frac{\theta_1 e^{-\theta_1 x_1}\theta_1 e^{-\theta_1 x_2}}{\theta_0 e^{-\theta_0 x_1}\theta_0 e^{-\theta_0 x_2}} = \frac{\theta_1^2}{\theta_0^2} e^{(\theta_0 - \theta_1)(x_1 + x_2)} = 4e^{-(x_1+x_2)}.$$

So $R = \{(x_1, x_2) | x_1 + x_2 < -\log{(c/4)}\}$

Problem 3 (12 pts. total)

Aliens of two races (blue and green) are arriving on Earth independently according to Poisson process distributions with parameters λ_b and λ_g respectively. The Alien Arrival Registration Service Authority (AARSA) will begin registering alien arrivals soon.

Let T_1 denote the time AARSA will function until it registers its first alien. Let G be the event that the first alien to be registered is a green one. Let T_2 be the time AARSA will function until at least one alien of both races is registered.

(a) (4 points.) Express $\mu_1 = \mathbf{E}[T_1]$ in terms of λ_g and λ_b. **Show your work.**

Answer: $\mu_1 = \mathbf{E}[T_1] = \frac{1}{\lambda_g+\lambda_b}$

Solution: We consider the process of arrivals of both types of Aliens. This is a merged Poisson process with arrival rate $\lambda_g+\lambda_b$. T_1 is the time until the first arrival, and therefore is exponentially distributed with parameter $\lambda_g+\lambda_b$. Therefore $\mu_1 = \mathbf{E}[T_1] = \frac{1}{\lambda_g+\lambda_b}$.

One can also go about this using derived distributions, since $T_1 = \min(T_1^g, T_1^b)$ where T_1^g and T_1^b are the first arrival times of green and blue Aliens respectively (i.e., T_1^g and T_1^b are exponentially distributed with parameters λ_g and λ_b, respectively.)

(b) (4 points.) Express $p = \mathbf{P}(G)$ in terms of λ_g and λ_b. **Show your work.**

Answer: $\mathbf{P}(G) = \frac{\lambda_g}{\lambda_g+\lambda_b}$

Solution: We consider the same merged Poisson process as before, with arrival rate $\lambda_g+\lambda_b$. Any particular arrival of the merged process has probability $\frac{\lambda_g}{\lambda_g+\lambda_b}$ of corresponding to a green Alien and probability $\frac{\lambda_b}{\lambda_g+\lambda_b}$ of corresponding to a blue Alien. The question asks for $\mathbf{P}(G) = \frac{\lambda_g}{\lambda_g+\lambda_b}$.

(c) (4 points.) Express $\mu_2 = \mathbf{E}[T_2]$ in terms of λ_g and λ_b.
Show your work.

Answer: $\frac{1}{\lambda_g + \lambda_b} + \frac{\lambda_g}{\lambda_g + \lambda_b}\left(\frac{1}{\lambda_b}\right) + \frac{\lambda_b}{\lambda_g + \lambda_b}\left(\frac{1}{\lambda_g}\right)$

Solution: The time T_2 until at least one green and one red Aliens have arrived can be expressed as $T_2 = \max(T_1^g, T_1^b)$, where T_1^g and T_1^b are the first arrival times of green and blue Aliens respectively (i.e., T_1^g and T_1^b are exponentially distributed with parameters λ_g and λ_b, respectively.)

The expected time till the 1st Alien arrives was calculated in (a), $\mu_1 = \mathbf{E}[T_1] = \frac{1}{\lambda_g+\lambda_b}$. To compute the remaining time we simply condition on the 1st Alien being green(e.g. event G) or blue(event G^c), and use the memoryless property of Poïsson, i.e.,

$$\begin{aligned}\mathbf{E}[T_2] &= \mathbf{E}[T_1] + \mathbf{P}(G)\mathbf{E}[\text{Time until first Blue arrives}|G] + \mathbf{P}(G^c)\mathbf{E}[\text{Time until first Green arrives}|G^c] \\ &= \mathbf{E}[T_1] + \mathbf{P}(G)\mathbf{E}[T_2^b] + (1 - \mathbf{P}(G))\mathbf{E}[T_2^g] \\ &= \frac{1}{\lambda_g + \lambda_b} + \frac{\lambda_g}{\lambda_g + \lambda_b}\left(\frac{1}{\lambda_b}\right) + \frac{\lambda_b}{\lambda_g + \lambda_b}\left(\frac{1}{\lambda_g}\right)\end{aligned}$$

Problem 4 (18 pts. total)

Researcher Jill is interested in studying employment in technology firms in Dilicon Valley. She denotes by X_i the number of employees in technology firm i and assumes that X_i are independent and identically distributed with mean p. To estimate p, Jill randomly interviews n technology firms and observes the number of employees in these firms.

(a) (6 points.) Jill uses

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

as an estimator for p. Find the limit of $\mathbf{P}(M_n \leq x)$ as $n \to \infty$ for $x < p$. Find the limit of $\mathbf{P}(M_n \leq x)$ as $n \to \infty$ for $x > p$. **Show your work.**

Solution: Since X_i is i.i.d., M_n converges to p in probability, i.e., $\lim_{n\to\infty} \mathbf{P}(|M_n - p| > \epsilon) = 0$, implying $\lim_{n\to\infty} \mathbf{P}(M_n < p - \epsilon) = 0$ and $\lim_{n\to\infty} \mathbf{P}(M_n > p + \epsilon) = 0$, for all $\epsilon > 0$. Hence

$$\boxed{\lim_{n\to\infty} \mathbf{P}(M_n \leq x) = \begin{cases} 0, & x < p, \\ 1, & x > p. \end{cases}}$$

(b) (6 points.) Find the smallest n, the number of technology firms Jill must sample, for which the Chebyshev inequality yields a guarantee

$$\mathbf{P}(|M_n - p| \geq 0.5) \leq 0.05.$$

Assume that $\text{var}(X_i) = v$ for some constant v. State your solution as a function of v. **Show your work.**

Solution: Since M_n converges to p in probability and $\text{var}(M_n) = \frac{n}{n^2} \cdot \text{var}(X_i) = v/n$, Chebyshev inequality gives

$$P(|M_n - p| \geq 0.5) \leq \frac{\text{var}(M_n)}{0.5^2} = \frac{v}{n \cdot 0.5^2} = 0.05$$

$\Rightarrow \boxed{n = 80v.}$

(c) (6 points.) Assume now that the researcher samples $n = 5000$ firms. Find an approximate value for the probability

$$\mathbf{P}(|M_{5000} - p| \geq 0.5)$$

using the Central Limit Theorem. Assume again that $\text{var}(X_i) = v$ for some constant v. Give your answer in terms of v, and the standard normal CDF Φ. **Show your work.**

Solution: By CLT, we can approximate by a standard normal distribution

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{nv}}$$

when n is large, and hence,

$$\mathbf{P}\left(|M_{5000} - p| \geq 0.5\right) = P\left(\left|\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{nv}}\right| \geq \frac{0.5\sqrt{n}}{\sqrt{v}}\right) = \boxed{2 - 2\Phi\left(\frac{0.5\sqrt{n}}{\sqrt{v}}\right)},$$

where $n = 5000$.

Problem 5 (12 pts. total)

The RandomView window factory produces window panes. After manufacturing, 1000 panes were loaded onto a truck. The weight W_i of the i-th pane (in pounds) on the truck is modeled as a random variable, with the assumption that the W_i's are independent and identically distributed.

(a) (6 points.) Assume that the measured weight of the load on the truck was 2340 pounds, and that $\text{var}(W_i) \leq 4$. Find an approximate 95 percent confidence interval for $\mu = \mathbf{E}[W_i]$, using the Central Limit Theorem (you may use the standard normal table which was handed out with this quiz). **Show your work.**

$\boxed{\textbf{Answer: } [2.216,\ 2.464]}$

Solution: The sample mean estimator $\hat{\Theta}_n = \frac{W_1+\cdots+W_n}{n}$ in this case is

$$\hat{\Theta}_{1000} = \frac{2340}{1000} = 2.34$$

Using the CDF $\Phi(z)$ of the standard normal available in the normal tables, we have $\Phi(1.96) = 0.975$, so we obtain

$$\mathbf{P}\Big(\frac{|\hat{\Theta}_{1000} - \mu|}{\sqrt{\text{var}(W_i)/1000}} \leq 1.96\Big) \approx 0.95.$$

Because the variance is less than 4, we have

$$\mathbf{P}(|\hat{\Theta}_{1000} - \mu| \leq 1.96\sqrt{\text{var}(W_i)/1000}) \leq \mathbf{P}(|\hat{\Theta}_{1000} - \mu| \leq 1.96\sqrt{4/1000}),$$

and letting the right-hand side of the above equation ≈ 0.95 gives a 95% confidence, i.e.,

$$[\hat{\Theta}_{1000} - 1.96\sqrt{\frac{4}{1000}}, \hat{\Theta}_{1000} + 1.96\sqrt{\frac{4}{1000}}] = [\hat{\Theta}_{1000} - 0.124, \hat{\Theta}_{1000} + 0.124]$$

(b) (6 points.) Now assume instead that the random variables W_i are i.i.d, with an exponential distribution with parameter $\theta > 0$, i.e., a distribution with PDF

$$f_W(w;\theta) = \theta e^{-\theta w}$$

What is the maximum likelihood estimate of θ, given that the truckload has weight 2340 pounds? **Show your work.**

Answer: $\hat{\Theta}^{mle}_{1000} = \frac{1000}{2340} = 0.4274$

Solution: The likelihood function is

$$f_W(w;\theta) = \prod_{i=1}^{n} f_{W_i}(w_i;\theta) = \prod_{i=1}^{n} \theta e^{-\theta w_i},$$

And the log-likelihood function is

$$\log f_W(w;\theta) = n \log\theta - \theta \sum_{i=1}^{n} w_i,$$

The derivative with respect to θ is $\frac{n}{\theta} - \sum_{i=1}^{n} w_i,$, and by setting it to zero, we see that the maximum of $\log f_W(w;\theta)$ over $\theta \geq 0$ is attained at $\hat{\theta}_n = \frac{n}{\sum_{i=1}^{n} w_i}$. The resulting estimator is

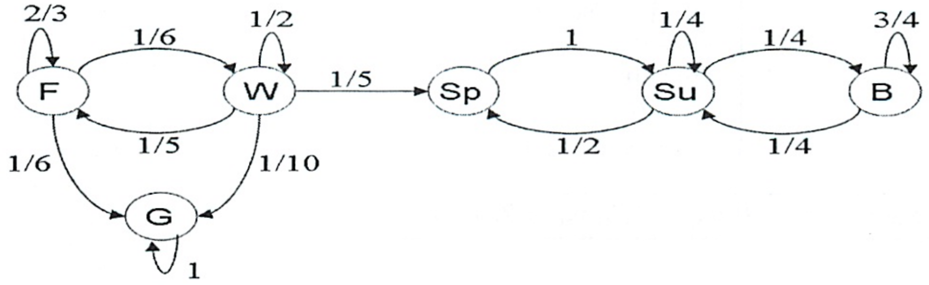
$$\hat{\Theta}^{mle}_n = \frac{n}{\sum_{i=1}^{n} W_i}.$$

In our case,

$$\hat{\Theta}^{mle}_{1000} = \frac{1000}{2340} = 0.4274$$

Problem 6 (21 pts. total)

In Alice's Wonderland, there are six different seasons: Fall (F), Winter (W), Spring (Sp), Summer (Su), Bitter Cold (B), and Golden Sunshine (G). The seasons do not follow any particular order, instead, at the beginning of each day the Head Wizard assigns the season for the day, according to the following Markov chain model:



Thus, for example, if it is Fall one day then there is 1/6 probability that it will be Winter the next day (note that it is possible to have the same season again the next day).

(a) (4 points.) For each state in the above chain, identify whether it is recurrent or transient. **Show your work.**

Solution: F and W are transient states; Sp, Su, B, and G are recurrent states.

(b) (4 points.) If it is Fall on Monday, what is the probability that it will be Summer on Thursday of the same week? **Show your work.**

Solution: There is only one path from F to Su in three days.

$$\begin{aligned}\mathbf{P}(S_4 = \text{Su}|S_1 = \text{F}) &= \mathbf{P}(S_2 = \text{W}|S_1 = \text{F}) \cdot \mathbf{P}(S_3 = \text{Sp}|S_2 = \text{W}) \cdot \mathbf{P}(S_4 = \text{Su}|S_3 = \text{Sp}) \\ &= \frac{1}{6} \cdot \frac{1}{5} \cdot 1 = \boxed{\frac{1}{30}}\end{aligned}$$

(c) (4 points.) If it is Spring today, will the chain converge to steady-state probabilities? If so, compute the steady-state probability for each state. If not, explain why these probabilities do not exist. **Show your work.**

Solution: The Markov chain will stay in the recurrent class $\{\mathrm{Sp}, \mathrm{Su}, \mathrm{B}\}$, and

$$\begin{cases} \pi_{\mathrm{Sp}} \cdot 1 = \pi_{\mathrm{Su}} \cdot \frac{1}{2} \\ \pi_{\mathrm{B}} \cdot \frac{1}{4} = \pi_{\mathrm{Su}} \cdot \frac{1}{4} \\ \pi_{\mathrm{F}} = 0 \\ \pi_{\mathrm{W}} = 0 \\ \pi_{\mathrm{G}} = 0 \\ \pi_{\mathrm{F}} + \pi_{\mathrm{W}} + \pi_{\mathrm{G}} + \pi_{\mathrm{Sp}} + \pi_{\mathrm{Su}} + \pi_{\mathrm{B}} = 1 \end{cases}$$

$$\Rightarrow \boxed{\pi_{\mathrm{F}} = 0, \pi_{\mathrm{W}} = 0, \pi_{\mathrm{G}} = 0, \pi_{\mathrm{Sp}} = 1/5, \pi_{\mathrm{Su}} = 2/5, \pi_{\mathrm{B}} = 2/5.}$$

(d) (5 points.) If it is Fall today, what is the probability that Bitter Cold will never arrive in the future? **Show your work.**

Solution: Let a_{F} and a_{W} be the probabilities that Bitter Cold will never arrive starting from Fall and Winter, respectively. This is equivalent to the Markov chain ends up in G.

$$\begin{cases} a_{\mathrm{F}} = \frac{2}{3} \cdot a_{\mathrm{F}} + \frac{1}{6} \cdot a_{\mathrm{W}} + \frac{1}{6} \cdot 1 \\ a_{\mathrm{W}} = \frac{1}{5} \cdot a_{\mathrm{F}} + \frac{1}{2} \cdot a_{\mathrm{W}} + \frac{1}{10} \cdot 1 \end{cases}$$

$$\Rightarrow \boxed{a_{\mathrm{F}} = 3/4.}$$

(e) (5 points.) If it is Fall today, what is the expected number of days till either Summer or Golden Sunshine arrives for the first time? **Show your work.**

Solution: Let μ_{F} and μ_{W} be expected number of days till either Summer or Golden Sunshine arrives for the first time, respectively.

$$\begin{cases} \mu_{\mathrm{F}} = 1 + \frac{2}{3} \cdot \mu_{\mathrm{F}} + \frac{1}{6} \cdot \mu_{\mathrm{W}} + \frac{1}{6} \cdot 0 \\ \mu_{\mathrm{W}} = 1 + \frac{1}{5} \cdot \mu_{\mathrm{F}} + \frac{1}{2} \cdot \mu_{\mathrm{W}} + \frac{1}{5} \cdot 1 \end{cases}$$

$\Rightarrow \boxed{\mu_{\mathrm{F}} = 5.25.}$

Problem 7 (12 pts. total)

A newscast covering the final baseball game between Sed Rox and Y Nakee becomes noisy at the crucial moment when the viewers are informed whether Y Nakee won the game.

Let a be the parameter describing the actual outcome: $a = 1$ if Y Nakee won, $a = -1$ otherwise. There were n viewers listening to the telecast. Let Y_i be the information received by viewer i ($1 \leq i \leq n$). Under the noisy telecast, $Y_i = a$ with probability p, and $Y_i = -a$ with probability $1 - p$. Assume that the random variables Y_i are independent of each other.

The viewers as a group come up with a joint estimator

$$Z_n = \begin{cases} 1 & \text{if } \sum_{i=1}^n Y_i \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

(a) (6 points.) Find $\lim_{n\to\infty} \mathbf{P}(Z_n = a)$ assuming that $p > 0.5$ and $a = 1$. **Show your work.**

Solution: Note that

$$\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = \lim_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^n Y_i \geq 0\right) = \lim_{n\to\infty} \mathbf{P}\left(\frac{\sum_{i=1}^n Y_i}{n} \geq 0\right).$$

Since Y_i are i.i.d. with mean $\mathbf{E}[Y_i] = 2p - 1$ and finite variance $\text{var}(Y_i) = 1 - (2p-1)^2$, one has, by Chebyshev inequality, for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{\sum_{i=1}^n Y_i}{n} - (2p-1)\right| \geq \epsilon\right) = 0.$$

Take $\epsilon = p - \frac{1}{2}$, and the above equation implies $\lim_{n\to\infty} \mathbf{P}\left(\frac{\sum_{i=1}^n Y_i}{n} \leq (2p-1)/2\right) = 0$. Therefore, $\boxed{\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = 1.}$

(b) (6 points.) Find $\lim_{n\to\infty} \mathbf{P}(Z_n = a)$, assuming that $p = 0.5$ and $a = 1$. **Show your work.**

Solution: Note that

$$\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = \lim_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^n Y_i \geq 0\right) = \lim_{n\to\infty} \mathbf{P}\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \geq 0\right).$$

Since Y_i are i.i.d. with $\mathbf{E}[Y_i] = 0$ and $\text{var}(Y_i) = 1$, we can approximate $\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}$ as a standard normal random variable when n goes to infinity. Thus, $\boxed{\lim_{n\to\infty} \mathbf{P}(Z_n = 1) = 1/2.}$

α,β constants

$E[\alpha]=\alpha$ $\sigma=\sqrt{var}$

$E[\alpha X]=\alpha E[X]$ $\sigma^2=var$

$E[\alpha X+\beta]=\alpha E[X]+\beta$

$Var(\alpha X)=a^2 var(x)$

$Var(X+\beta)=var(x)$

$E[XY]=\sum_x\sum_y x\,y\,p_{x,y}(x,y)$ ⎫ ind.

$=\sum_y\sum_x p_x(x)\,p_y(y)$ ⎭

$var(x+y) = E[(x+y)^2] = E[x^2]+E[y^2]+2\,E[XY]$

$= var(x)+var(y)+C$

Sum of R# RV

$var(Y)=E[var(Y|N)]+var(E[Y|N])$

$E[Y|N]=N\,E[X]$

$Var(E[Y|N])=(E[X])^2\,var(N)$

$Var(Y|N=n)=n\,var(x)$ ← just sum of var

$Var(Y|N)=N\,var(x)$

$E[var(Y|N)]=E[N]\,var(x)$

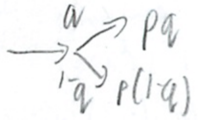
$var(Y)=E[var(Y|N)]+var(E[Y|N])$

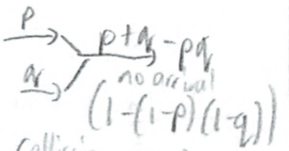
$=E[N]\,var(X)+(E[X])^2 var(N)$

Bernoulli Process

$P(\text{sucess}): P(X_i=1)=p$ $E[X_t]=p$

$p(\text{failure})=P(X_i=0)=1-p$ $Var(X_t)=p(1-p)$





- memoryless

collissions are 1

Poisson Process

λ = arrival rate

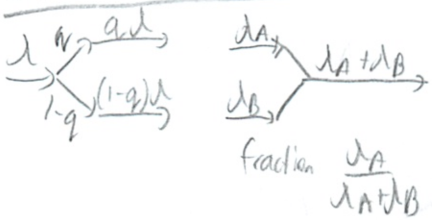
$P(k,\tau)$ = prob of k arrivals in interval lenght τ

- each section independent

$$P(k,\delta)=\begin{cases}1-\lambda\delta & \text{if } k=0\\ \lambda\delta & \text{if } k=1\\ 0 & \text{if } k>1\end{cases}$$

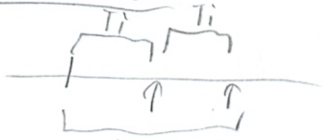
δ ← very small

6.041 Cheat Sheet 3



fraction $\frac{\lambda_A}{\lambda_A+\lambda_B}$

Random Incidence



erlang order 2

Markov - on back

Limit Theorms

X = RV w/ mean μ var σ^2

$S_n=X_1+\dots X_n$

$var(S_n)=n\sigma^2$ - dist spreads out as n↑

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

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

↳ weak law lg # / cov in prob

$Z_n=\frac{S_n-n\mu}{\sigma\sqrt{n}}$ normalize so $E[Z_n]=0$ $var(Z_n)=1$

Markov Ineq If a nonneg RV has small mean, than P(takes large value) must be small

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

Smushes the PDF

E[ ] must be lower

bounds very lose

prob will be less than $\frac{E[X]}{a}$

- RV must be non-neg

Chebsher If a RV has a small var, than P(value far (from mean)) is small

- RV can be negitive

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

1. Fix c to be a certain # c = confidence
2. P of being at least c away from μ

Pollster Problem $P(|M_n-f|\geq \text{error})\leq \text{confidence}$

$P(|M_n-f|\geq \text{error})\leq\frac{\sigma_{M_n}^2}{(\text{error})^2}$

confidence $=\frac{\sigma_x^2}{n(\text{error})^2}\leq\frac{1}{4n(\text{error})^2}$ ← pick n so true

---

Weak Law Large

12/13

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

↳ basically will be inside bounds

"M_n converges to μ"

CLT

when n is large

$Z_n=\frac{S_n-E[S_n]}{\sigma_{S_n}}=\frac{S_n-nE[X]}{\sqrt{n}\,\sigma_X}$

$E[Z_n]=0$ $var(Z_n)=1$

$P(Z\leq c)$ same as normal CDF

Binomial CDF of $\frac{S_n-np}{\sqrt{np(1-p)}}$ → st. normal

$P(S_n\leq 21)=\Phi\left(\frac{21-E[S_n]}{\sigma_{S_n}}\right)$ ← plug in

21 ← value trying to test

look at table

But usually $\frac{1}{2}$ correction!

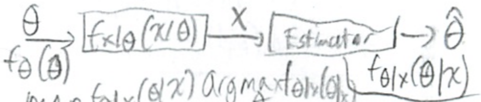
De Moivre Laplace

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

$=\Phi\left(\frac{19.5-E[\ ]}{\sigma}\right)-\Phi\left(\frac{18.5-E[\ ]}{\sigma}\right)$

prob in that gap

Baysean Estimators



MAP $f_{\Theta|X}(\theta|x)$ $\arg\max_\theta f_{\Theta|X}(\theta|x)$

minimizes prob error

LSE/LMS $=E[(\Theta-E[\Theta|X])^2|X=x]$

$=Var(\Theta|X=x)$

= estimation error

estimator $\hat{\Theta}=E[\Theta|X]$

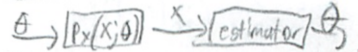
estimation error $\tilde{\Theta}=\hat{\Theta}-\Theta$

$E[\tilde{\Theta}]=0$ $cov(\tilde{\Theta},\hat{\Theta})=0$

$Var(\Theta)=var(\tilde{\Theta})+var(\hat{\Theta})$

LLMS $\hat{\Theta}_{LLMS}=E[\Theta]+\frac{cov(X,\Theta)}{var(X)}(X-E[X])$

Classical Estimation



$\hat{\Theta}_{ML}=\arg\max_\theta p_X(x;\theta)$ pick θ that makes data most likely

---

CI $P\left(\hat{\Theta}_n-\frac{z\sigma}{\sqrt{n}}\leq\Theta\leq\hat{\Theta}_n+\frac{z\sigma}{\sqrt{n}}\right)\approx 1-\alpha$

$\Phi(z)=1-\frac{\alpha}{2}$

arrival rate

# of arrivals in n time slots

discrete(S

contineus(N_t

Interarrival times/Time b/w arrivals

disct
- T_1 = # trials till 1st arrivals
- Memoryless
- T_2 = add. time to next sucess
  - Same dist
  - ind. of T_1

Time of kth arrival

Y_k = # of trials till 1st arrival $= T_1 + T_2 + \dots + T_k$

Estimator of Var
1. Bernoulli $\hat{\theta}(1-\hat{\theta})$
2. Sample var $\hat{V}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\theta}_n)^2$
3. Upper band (bad) Bernoulli $= \frac{1}{4}$

Discrete / Bernoulli

p = per time ; arrivals / unit time

Binomial

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

$E[S] = np$

$var(S) = np(1-p)$

Geometric

$P(T_1 = t) = (1-p)^{t-1} p$

$E[T_1] = \frac{1}{p}$

$var(T_1) = \frac{1-p}{p^2}$

Pascal

$P(Y_k = t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}$

$E[Y_k] = k \cdot \frac{1}{p}$

$var(Y_k) = k \cdot \frac{1-p}{p^2}$

(q) $q_{jk}(n) = E[\,]$ of transitions from j to k within n transitions

(v) $v_{ij}(n) = E[\,]$ visits to j w/in n transitions - i start

Contineous / Poisson

λ = contineous arrivals/unit time

Binomial $= N_T$

↳ since finely discretize $[0, t]$

Take $\delta \to 0$ or $n \to \infty$

$P(k, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$ $k = 0, 1, \dots$

$E[N_T] = \lambda t$ $var(N_t) = \lambda t$

Exponential

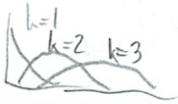
$f_{T_1}(y) = \lambda e^{-\lambda y}$ $y \geq 0$

$E[T_1] = \frac{1}{\lambda}$

Erlang

$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$ $y \geq 0$

$E[Y_k] = \frac{k}{\lambda}$ $var(Y_k) = k\,var(T_k) = \frac{k}{\lambda^2}$



Markov

define states

define transition possibilities

↳ can be geometric, etc

X_n = state at time n

X_0 either given or random

(p) transition prob = $p_{ij} = P(X_{n+1} = j \mid X_n = i)$

(r) State occupance prob = $r_{ij}(n) = P(X_n = j \mid X_0 = i)$

given initial state, after n transitions

$= \sum_{k=1}^{m} r_{ik}(n-1) p_{kj}$

w/ random initial state

$P(X_n = j) = \sum_{i=1}^{m} P(X_0 = i) r_{ij}(n)$

Consider all the posibilities it could have been

recurrent - where ever you can go, you can always come back

transient - otherwise $P(X_n = i) \to 0$

periodic - all transitions from 1 grop to another

(π) Steady state prob $\pi_j = \sum_k \pi_k p_{kj}$ for all j

$\lim_{n\to\infty} r_{ij}(n) = \pi_j$

$\sum_j \pi_j = 1$ ↑ going towards

this is the add up trans prob to + solve

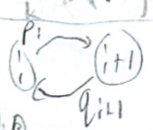
= long run freq of being in j

freq transitions $k \to j = \pi_k p_{kj}$

freq transitions $\to j = \sum_k \pi_k p_{kj}$

need aperiodic since recurrent class

Birth-Death



$\pi_i p_i = \pi_{i+1} q_{i+1}$

$\pi_{i+1} = \pi_{i+1} = \pi_i \frac{p}{q} = \pi_i \rho$

$\pi_i = \pi_0 \rho^i$ $i = 0, 1, \dots, m$ / Assume $p < q$ $m \approx \infty$

$\pi_0 = 1 - \rho$

$E[X_n] = \frac{\rho}{1-\rho}$ in steady state

(a) asorbtion prob $a_i = \sum_j p_{ij} a_j$ for all other i

-prob settles in a

i = start

↑ all tos

(μ) E[asorbtion] $= \mu_i = 1 + \sum_j p_{ij} \mu_j$

i = start $\mu_s = 0$ ↑ recurrent

solve

↑ leaving

(t) mean first passage time $= E[\min\{n \geq 0$ such that $X_n = s\} \mid X_0 = i]$

s = recurrent $i \to s$

i = start

$t_s = 0$ $t_i = 1 + \sum_j p_{ij} t_j$ ↓ away from

(t*) mean recurance time of s $= E[\min(n \geq 1$ such that $X_n = s \mid X_0 = i]$

$t_s^* = 1 + \sum_j p_{sj} t_j$ ↓ away from

Desirable Properties

Unbiased $E[\hat{\theta}_n] = \theta$

Consistent $\hat{\theta}_n \to \theta$ in prob

Small MSE $E[(\hat{\theta} - \theta)^2]$

$\log e^x = x$

$\log 1 = 0$

$\log_b(m^n) = n \log_b(m)$

6.041 Quiz 1

10/11

$\cap$ = and $\subset$ subset of

$\cup$ = or ← don't forget remove overlap

$A \subset C \rightarrow P(A) \leq P(B)$

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

$P(A \cup B) \leq P(A) + P(B)$

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

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

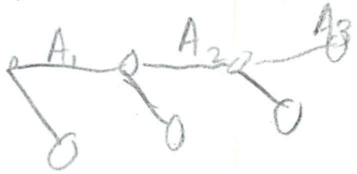
↳ $P(A|B) \geq 0$

$P(\Omega|B) = 1$

$P(A_1 \cup A_2 \cup A_3 \ldots |B) = P(A_1|B) + P(A_2|B) + \ldots$

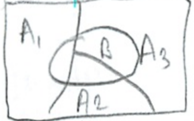
Multiplication Rule

$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$



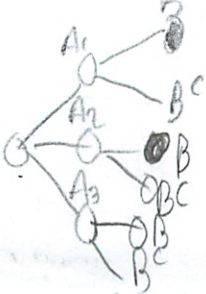
Total Probability Theorm

-add up each section



$P(B) = \sum_{i=1}^{n} P(B \cap A_i) =$

$\sum_{i=1} P(B/A_i)P(A_i)$



Independent

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

$P(A|B) = P(A)$

Conditionaly independent seperate

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

Counting Techniques

n items

realighn the universe

basic $n_1 n_2 n_3 .. n_m$

pormentations $n(n-1) \ldots (n-k+1) = \frac{n!}{(n-k)!}$ ↓$_nP_k$

└order important k = seq lenght =

Combonations $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

partitions

r-disjoint subsets

n_k elements in nth subset

$$\binom{n}{n_1, n_2, \ldots n_r} = \binom{n}{n_1}\binom{n-n_1}{n_2} \ldots \binom{n-n_1-\ldots n_r-1}{n_r}$$

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

$$= \frac{n!}{n_1! \circ \ldots \circ n_r!}$$ multinomial formual

Bayes Rule

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

↓ liklihood ← prior prob

↑ posterior prob

← normalization constant

Discrete Random Variables

$\{X = x\} = \{\omega \in \Omega | X(\omega) = x\} \subseteq \Omega$

$p_X(x) \geq 0$

$P(X \in S) = \sum_{x \in S} p_X(x)$

$Y = g(X)$ still works w/ everything

$E[X] = \sum_x x p_X(x)$

$E[aX+b] = aE[X]+b$ linearity of expectatio

$E[g(X)] = \sum_x g(x)p_X(x)$

$Var(X) = E[(X-E[X])^2] = E[X^2] - E[X]^2$

$var(aX+b) = a^2 var(X)$

joint

$p_{X,Y}(x,y) = P(\{X=x\} \cap \{Y=y\})$

$p_X(x) = \sum_y p_{X,Y}(x,y)$

$p_Y(y) = \sum_x p_{X,Y}(x,y)$) marginal

Total Expectation Theorm

$E[X] = \sum_{i=1}^{n} P(A_i)E[X|A_i]$

Independent

$p_{X|A} = p_X(x)$

$\forall x$

↑for all/given any

or

$p_{X,A}(x) = p_X(x)P(A)$

$\forall x$

$E[XY] = E[X]E[Y]$ inverse not true

$Var(X+Y) = var(X) + var(Y)$ - constants don't matter

knowing info abat 1 does not help you at all w/ other

$p_{X,Y}(x,y) = p_X(x)p_Y(y)$

$\binom{5}{1} = 5$

$\binom{5}{5} = 1$

$\frac{|P(A)|}{|P(\Omega)|}$ Discrete Uniform Law

Discrete Distributions

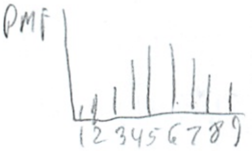
	X	PMF $p_x(k)$	$E[X]$	$var(x)$
Bernoulli	$\begin{cases}1 & \text{sucess}\\0 & \text{failure}\end{cases}$	$\begin{cases}p & k=1\\1-p & k=0\end{cases}$	p	$p(1-p)$
Binomial	# of sucesses in n Bernoulli trials	$\binom{n}{k} p^k (1-p)^{n-k}$ $k=0,1,\dots n$ ↑ # sucesses	np	$np(1-p)$
Geometric	# of trials till 1st sucess	$(1-p)^{k-1} p$ $k=1,2,3$ ↑ # of trials !!	$\frac{1}{p}$	$\frac{1-p}{p^2}$ expected trials till sucess
Uniform	An integer in $[a,b]$	$\begin{cases}\frac{1}{b-a+1} & k=a,a+1,\dots,b\\0 & \text{oterise}\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)(b-a+2)}{12}$
Poisson		$e^{-\lambda}\frac{\lambda^k}{k!}$ $k=0,1,2,$		

Bernoulli

- 2 outcomes
- like a coin flip
- used to construct more complex random variables

Binomial

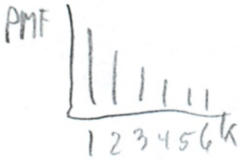
- coin tossed n coins
- X= # sucesses in n trials



$n=9$
$p=1/2$

Geometric

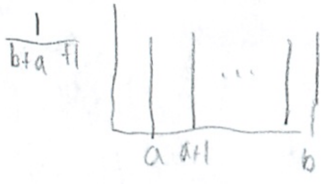
- #(x) tosses needed till 1st sucess



Uniform

- rolling a fair 6-side die

$$p_x(k) = \begin{cases} 1/6 & \text{if } k=1,2,3,4,5,6 \\ 0 & \text{otherwise}\end{cases}$$



Poisson

- like a binomial RV w/ very small p + large n
- good, quick approx for binomial

$$\lambda = \overset{\text{\# trials}}{\underset{\text{very large}}{n}}\, \overset{\text{epro'n}}{p} \leftarrow \text{very small}$$

Classical Part 2

$$\log f_w(w;\theta)^c = \prod_{i=1}^{n} f_{w_i}(w_i;\theta)$$

take deriv → set = to 0 w/respect to θ

6.041 Quiz 2

Rules

$P(a \le X \le b) = \int_a^b f_X(x)\,dx$

$P(X \in A) = \int_A f_X(x)\,dx$

Nonregitivity $f_X(x) \ge 0 \ \forall x$

Normalization $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

$P(X = x) = 0 \ \forall x$ since contineous

$f_X(x)$ can be ≥ 1

$P(x \le X \le x+\delta) \approx f_X(x)\delta$

$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$

$var(X) = E[(X - E[X])^2]$

$= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\,dx$

$= E[X^2] - (E[X])^2 \ge 0$

$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$

$E[aX+b] = aE[X] + b$

$Var(aX+b) = a^2 Var(X)$

CDF $F_X(x) = P(X \le x)$

monotonically increasing $0 \to 1$

Discrete $\sum_{k \le x} p_X(k)$ piecewise constant

$p_X(k) = F_X(k) - F_X(k-1)$

Contineous $\int_{-\infty}^{x} f_X(t)\,dt$

$f_X(x) = \frac{dF_X(x)}{dx}$

Joint PDF $f_{X,Y}(x,y)$

$P(A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$

?? $\to f_X(x) f_{Y|X}(y|x)$

Marginal: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$

Joint CDF: $F_{X,Y}(x,y) = P(X \le x, Y \le y)$

$\forall x, y$

Independence $f_{X,Y}(x,y) = f_X(x) f_Y(y)$

$E[XY] = E[X]E[Y]$

$g(X), h(Y)$ independent

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Conditioning A is an event $P(A) > 0$

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X \in A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

$P(X \in B | X \in A) = \int_B f_{X|A}(x)\,dx$

$E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x)\,dx$

$E[g(X)|A] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)\,dx$

partitioning $A_1, \ldots, A_n$ disjoint

$f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)$ total prob theorm

$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$ total expectation theorm

$E[g(X)] = \sum_{i=1}^{n} P(A_i) E[g(X)|A_i]$

Conditioning on RV X, Y contineous

$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ total prob theorm

$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y)\,dy$

$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$

$E[g(X)|Y=y] = \int_{-\infty}^{\infty} g(X) f_{X|Y}(x|y)\,dx$

$E[g(X,Y)|Y=y] = \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\,dx$

Total Expectation Theorm

$E[X] = \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y)\,dy$

$E[g(X)] = \int_{-\infty}^{\infty} E[g(X)|Y=y] f_Y(y)\,dy$

$E[g(X,Y)] = \int_{-\infty}^{\infty} E[g(X,Y|Y=y] f_Y(y)\,dy$

generic var

$var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$

$E[X+Y] = E[X] + E[Y]$ linearity of expectation

Contineous Bayes' Rule 10/31

X, Y contineous RV, N discrete, A event

$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$

contineous-contineous

$= \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|t) f_X(t)\,dt}$

$p_{X|Y}(x|y) = \frac{p_X(x) p_{Y|X}(y|x)}{p_Y(y)}$

discrete discrete

$\sum_x p_X(x) p_{Y|X}(y|x)$

$p_{X|Y}(x|y) = \frac{p_X(x) f_{Y|X}(y|x)}{f_Y(y)}$

discrete X contineous Y

$= \sum_x p_X(x) f_{Y|X}(y|x)$

$f_{X|Y}(x|y) = \frac{f_X(x) p_{Y|X}(y|x)}{p_Y(y)}$

contineous X discrete Y

$= \int_x f_X(x) p_{Y|X}(y|x)\,dx$

$P(A|Y=y) = \frac{P(A) f_{Y|A}(y)}{f_Y(y)}$

event

$= \frac{P(A) f_{Y|A}(y)}{f_{Y|A}(y)P(A) + f_{Y|A^c}(y)P(A^c)}$

$P(N=n|Y=y) = \frac{p_N(n) f_{Y|N}(y|n)}{f_Y(y)}$

$= \frac{p_N(n) f_{Y|N}(y|n)}{\sum_i p_N(i) f_{Y|N}(y|i)}$

$f_{X|A}(x) = \frac{f_X(x)}{P(A)}$ $X \in A$

event

$f_{Y|X \cap A}(y|x)$ $Y \in A$

$= \frac{f_{Y|X}(y|x)}{P(A|X=x)}$

$f_{X|A}(x) = \frac{P(A|X=x) f_X(x)}{\int P(A|X=x) f_X(x)\,dx}$

	PDF $f_X(x)$	CDF $F_X(x)$	$E[X]$	$var(X)$
Uniform Uniform over interval $[a,b]$	$\frac{1}{b-a}$ if $a \le x \le b$ 0 otherwise	0 if $\le a$ $\frac{x-a}{b-a}$ if $a \le x \le b$ 1 if $x > b$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$
Exponential(λ) exponential RV w/ param λ Memoryless given $X > t$, $X-t$ is an exponential RV w/ param λ	$\lambda e^{-\lambda x}$ if $x \ge 0$ 0 otherwise	$1-e^{-\lambda x}$ if $x \ge 0$ 0 otherwise	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal/Gaussian(μ, σ^2) $X \sim (\mu, \sigma^2)$ $Y = aX+b$ $Y \sim N(a\mu+b, a^2\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Φ See table $P(X \le 0) = 1-P(X>0) \leadsto 1-\Phi(X)$ $\phi\left(\frac{X-\mu}{\sigma}\right) = \phi\left(\frac{X-E[X]}{\sqrt{var(X)}}\right)$	μ	σ^2

Derived Distributions $Y = g(X)$

have PDF of X, want PDF of Y

1. $F_Y(y) = P(Y \le y) = P(g(X) \le y) =$

Get CDF $= \int_{x \mid g(x) \le y} f_X(x)\,dx$

2. Differenitate $f_Y(y) = \frac{dF_Y(y)}{dy}$

Special Case $Y = g(X) = aX + b$

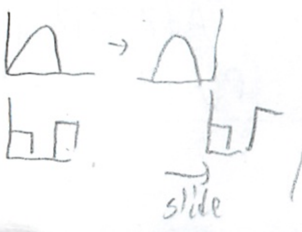
$$f_Y(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right)$$

Convolution $W = X + Y$ X,Y independent

discrete $p_W(w) = \sum_x p_X(x)\, p_Y(w-x)$

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\,dx$$

Flip + slide + Cross multiply + evaluate



each piecewise

if X, Y normal W will be normal

Law of Iterated Expectations

$E[X|Y] = f(Y) = RV$ (expetation w/ respect to X)

$E[X|Y=y] = \#$

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

Law of Total Variance

$var(X|Y) = RV$ (var w/ respect to x)

First $var(X|Y=y)$

$= E[(X - E[X|Y=y])^2 \mid Y=y]$

$var(X) = E[var(X|Y)] + var(E[X|Y])$

Sum of a Random # of RVs

N = discrete RV,

X_i = iid, independent of N

$Y = X_1 + X_2 + \ldots + X_n$

$E[Y] = E[X]E[N]$

$var(Y) = E[N]var(X) + (E[X])^2 var(N)$

Cov and Correlation

$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$

$= E[XY] - E[X]E[Y]$

$Cov(X,Y) = 0 \rightarrow$ uncorrelated

$= 1$

$= -1$

if X,Y ind $\rightarrow$ uncorrelated

$\not\leftarrow$

$var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$

Correlation coefficent (dimensionless)

$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \in [-1,1]$

$\rho = 0 \rightarrow$ uncorrelated

$|\rho| = 1 \rightarrow$ linearly related

$= X - E[X]$

$= c[Y - E[Y]]$

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

~~Make Cheat sheet~~

~~$E\{XY\}$~~

~~derived dist~~

~~Practice~~

~~- in Matlab~~

~~Cov~~

6.041 Debreif

12/15

left 15 min early - ~~ma~~ almost no one else did

much eaiser than I feared

but if easy - everyone does better - curve

- still likly same place on curve

- B- for year, hope hope get B!