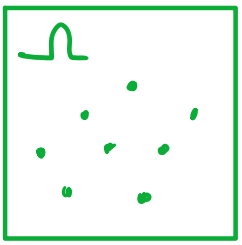
# Lecture 1: Probability Models and Axioms

Monday, September 12, 2016 7:04 AM

## Probabilistic Model

1) Describe possible outcomes → sample space




Mutually exclusive - (only one)

Collectively exhaustive - (all of them)

### Sample Space

- Rolling 3-sided die ← discrete and finite

  Sequential description

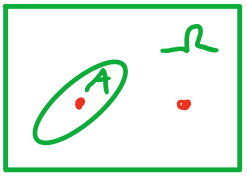
  (useful for experiments w/ stages)

- $(x,y)$ st $0 \le x, y \le 1$ ← continuous

2) Specify probability law

↳ Assign probability to outcomes

**Event** - A subset of the sample space



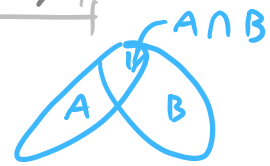

### Axioms:

- Nonnegativity $P(A) \ge 0$ ⇒ $P(\emptyset) = 0$
- Normalization $P(\Omega) = 1$ ⇒
- (Finite) additive — disjoint events

- Normalization $P(\Omega) = 1$ $\Rightarrow$
- (Finite) additive

disjoint events

IF $\underbrace{A \cap B = \emptyset}_{A \cap B}$ $\Rightarrow$ $P(A \cup B) = P(A) + P(B)$



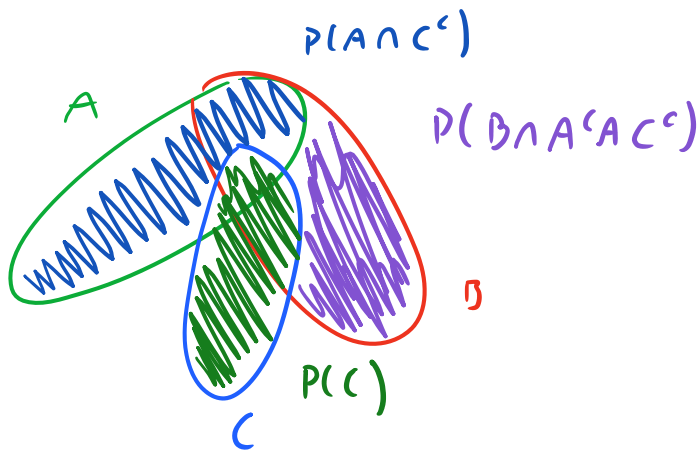

$\Rightarrow$

- IF $A \subset B$ $\Rightarrow$ $P(A) \leq P(B)$

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

  $\Rightarrow P(A \cup B) \leq P(A) + P(B)$ ← union bound

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




- Countable Additivity

  ↳ discrete but infinite sample space

IF $A_1, A_2 \ldots$ is an infinite sequence of disjoint events

$\Rightarrow P(A_1 \cup A_2 \cup \ldots A_3 \cup \ldots)$

$= P(A_1) + P(A_2) + \ldots$

# Mathematical Background

Thursday, September 15, 2016 8:44 PM

## Sets

- A collection of distinct elements

Notation: $x \in S$ "x in S" $x \notin S$ "x not in S"

$\{x \in \mathbb{R} : \cos(x) > 1/2\}$

"real #'s that satisfy"

$\Omega$ : universal set

$S^c$: $x \in S^c$ if $x \in \Omega$ and $x \notin S$

$(S^c)^c = S$

$\emptyset$: empty set $\Rightarrow \Omega^c = \emptyset$

$S \subset T$: $x \in S \Rightarrow x \in T$

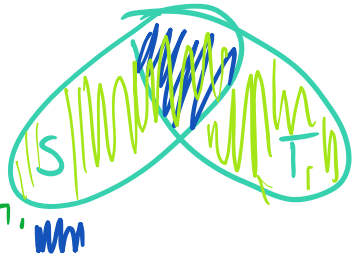
"S is subset of T"

$S \cup T$: $\Leftrightarrow x \in S$ or $x \in T$

"union of S and T"

$S \cap T \Leftrightarrow x \in S$ and $x \in T$

"intersection of S and T"



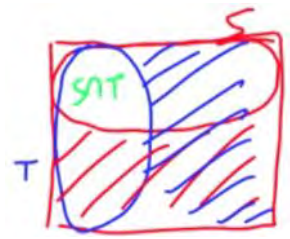
# Properties:

$$S \cup T = T \cup S, \qquad S \cup (T \cup U) = (S \cup T) \cup U,$$
$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U), \qquad S \cup (T \cap U) = (S \cup T) \cap (S \cup U),$$
$$(S^c)^c = S, \qquad S \cap S^c = \emptyset,$$
$$S \cup \Omega = \Omega, \qquad S \cap \Omega = S.$$

# De Morgan's Law

Go between unions $\cup$ $\Rightarrow$ and intersections $\cap$

$$(S \cap T)^c = S^c \cup T^c$$




Proof:

$$x \in (S \cap T)^c \Leftrightarrow x \in S \cap T \Leftrightarrow \begin{Bmatrix} x \notin S \\ \text{or} \\ x \notin T \end{Bmatrix} \Leftrightarrow \begin{Bmatrix} x \in S^c \\ \text{or} \\ x \in T^c \end{Bmatrix}$$

$$\Leftrightarrow x \in S^c \cup T^c$$

$\Rightarrow$

$$S^c \cap T^c = (S \cup T)^c$$

General Form $\Rightarrow$

$$\left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$$
$$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c$$

# Sequences

$a_1, a_2, a_3 \ldots\ldots$ $\qquad i \in \mathbb{N}$

$a_i \in S$

$a_1, a_2, a_3 \ldots\ldots$ $\quad i \in \mathbb{N}$

$a_i \in S$

Sequence $a_i$

Sequence: Function $\mathbb{N} \to S$

$f(i) = a_i$

Notation

$$a_i \xrightarrow[i\to\infty]{} a = \lim_{i\to\infty} a_i = a$$

converging sequence

Converging:

If $a_i \leq a_{i+1} \Rightarrow$ converges to $\infty$ or $\Rightarrow$ converges to $a$

If $|a_i - a| \leq b_i$ for all $i$ and $b_i \to 0 \Rightarrow a_i \to a$

# Series

## Infinite Series

$$\sum_{i=1}^{\infty} a_i = \lim_{n\to\infty} \sum_{i=1}^{n} a_i \qquad \text{(provided limit exists)}$$

If $a_i > 0 \Rightarrow$ limit exists

Otherwise:

If $\sum_{i=1}^{\infty} |a_i| < \infty \Rightarrow$ limit exists and independent order of summation

## Geometric Series

$$S = \sum_{i=0}^{\infty} \alpha^i = 1 + \alpha + \alpha^2 + \dots = \frac{1}{1-\alpha} \qquad |\alpha| < 1$$

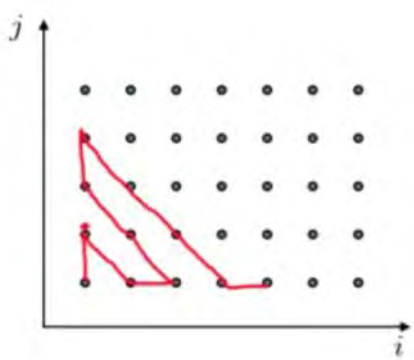
$$(1-\alpha)(1+\alpha+\alpha^2+\dots\alpha^n) = 1-\alpha^{n+1}$$

$n \to \infty$

$$(1-\alpha)S = 1 \Rightarrow S = 1/1-\alpha$$

## Order of summation

$$\sum_{i\geq 1, j\geq 1} a_{ij}$$




Many ways to add terms together

If $\sum |a_{ij}| < \infty$

order will not matter

## Countable and Uncountable Sets

Countable: can number every element

Ex: positive integers, integers, pairs of int... etc

uncountable:

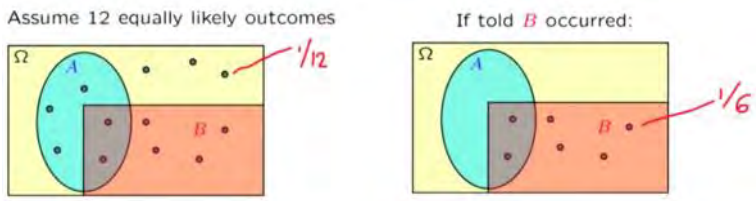
Ex: the interval $[0,1]$, the reals, the plane

# Lecture 2 : Conditioning and Bayes Rule

Monday, September 19, 2016 1:24 PM

## Conditional Probabilities

- Take into account other information and revise model




$P(A \mid B) =$ "probability of $A$, given that $B$ occurred"

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

New total probability

IF $P(B) > 0$ $\Rightarrow$ $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$

## Conditional Probabilities obay the same axioms

IF $A \cap C = \emptyset \Rightarrow P(A \cup C \mid B) = P(A \mid B) + P(C \mid B)$

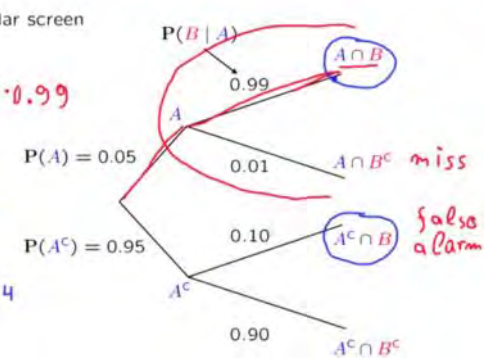
## Example:

**Models based on conditional probabilities** $\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad P(B \mid A) = \frac{P(A \cap B)}{P(A)}}$

Event $A$: Airplane is flying above

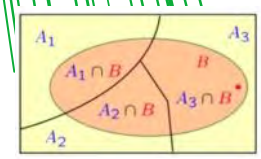
Event $B$: Something registers on radar screen

- $P(A \cap B) = P(A) \cdot P(B \mid A) = 0.05 \cdot 0.99$
- $P(B) = 0.05 \cdot 0.99 + 0.95 \cdot 0.1 = 0.1445$
- $P(A \mid B) = \dfrac{0.05 \cdot 0.99}{0.1445} = 0.34$



**Multiplication Rule** - Calculate Probabilities of leaves in tree by multiplying probability of branches

**Total Probability Theorem** - P(event) is weighted average of Prob of each possible scenario




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

Generalized

$$P(B) = \sum_i P(A_i) P(B \mid A_i)$$

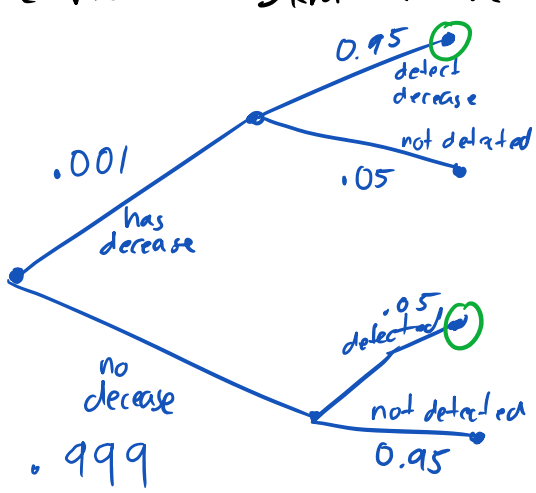
**Bayes Rule** • Way to incorporate new evidence

$$P(A_i \mid B) = \frac{P(A_i) P(B \mid A_i)}{\sum_j P(A_j) P(B \mid A_j)} \qquad P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

---

A test for a certain rare disease is assumed to be correct 95% of the time: if a person has the disease, the test result is positive with probability 0.95, and if the person does not have the disease, the test result is negative with probability 0.95. A person drawn at random from a certain population has probability 0.001 of having the disease.

Sample Problem Scratchwork FUS




Test Positive

$(.001)(.95)$
$+ (.999)(.05)$ ✓

$$P(D \mid P) = \frac{P(D \text{ and } P)}{P(P)}$$

$$= \frac{(.001)(.95)}{(.001)(.95) + (.999)(.05)} \checkmark$$

# Lecture 3: Independence

Monday, September 19, 2016 3:31 PM

$P(B|A) = P(B)$

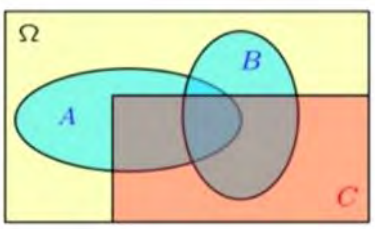
*occurance of A provides no new info about B*

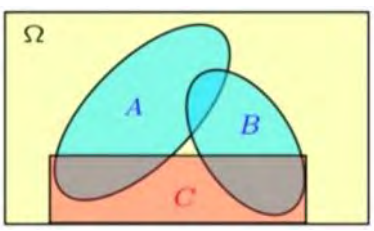
$P(A \cap B) = P(A)P(B|A) = P(A)P(B)$

$$\boxed{\text{Independence: } P(A \cap B) = P(A) \cdot P(B) \quad \text{disjoint} \neq \text{independen!}}$$

IF A and B are independent $\Rightarrow$ A and $B^c$ are independent

## Conditional Independence




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

not independent

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

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

Independence $\not\Rightarrow$ conditional independence

## Pairwise Independence

Pairwise Independence $\not\Rightarrow$ Independence

# Lecture 4: Counting

Tuesday, September 20, 2016 3:55 PM

## Basic Counting Principle

$r$ stages

$n_i$ choices as stage $i$

#choices: $n_1 \cdot n_2 \cdot n_3 \cdots n_r$

Permuations: # of ways of ordering $n$ elements

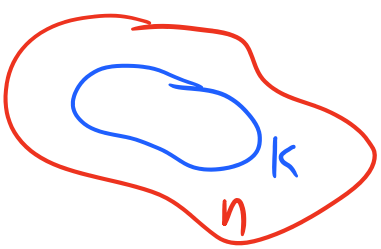
$n!$

#subsets: $\{1 \ldots n\} = 2^n$

$\underbrace{2 \cdot 2 \cdot 2 \cdot 2 \cdot \cdots}_{n}$

add it or not

## Combination




$\binom{n}{k}$ # of $k$ element subsets of given $n$ element set

n choose k

$\frac{n!}{(n-k)!\,k!}$

$0! = 1$

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

$$\sum_{k=0}^{n} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots \binom{n}{n} = \text{all subsets} = 2^n$$

## Partitions

$n$ distinct items

Give $m_i$ items to person $i$

$n$ distinct items  Give $n_i$ items to person $i$
$r$ people

ordering $n$ items: $n!$

$$\Rightarrow \#\text{partitions} = \frac{n!}{n_1!\, n_2! \ldots n_r!}$$

IF $r=2$ $n_1=k$ $n_2=n-k$
$\Rightarrow$ Partition of set into 2 sets

Example:
Deal 52 deck card to 4 Players
Find p(each player gets ace)

1) $\#\text{outcomes} = \dfrac{52!}{\underbrace{13!\,13!\,13!\,13!}_{\text{13 cards for 4 players}}}$ (1)

2) #outcomes where each person gets an ace.

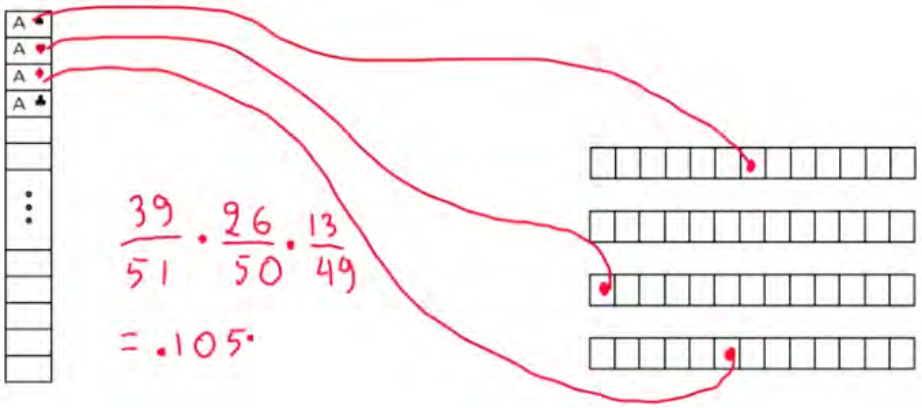
- Distribute aces: $4 \cdot 3 \cdot 2 \cdot 1$
- Distribute remaining 48 cards

$$\frac{48!}{12!\,12!\,12!\,12!} \quad (2)$$

$$\text{Answer} = \frac{\#\text{outcomes in event}}{\#\text{outcomes}} \quad \begin{matrix}(2)\\(1)\end{matrix}$$

— OR —

Stack the deck, aces on top




$$\frac{39}{51} \cdot \frac{26}{50} \cdot \frac{13}{49}$$

$$= .105$$

# Lecture 5: Probability Mass Functions and Expectations

Saturday, October 1, 2016 12:24 AM

⇒ Random Variables

## Random Variable

↳ numerical quantity that can take random values

Ex: weight of student chosen at random
W
↖ don't know until it is picked

height " " "
H

RV - Associate a number to every possible outcome
ie weight

$\Omega \to \mathbb{R}$

Can take discrete or continious values
↳ ...  ↳ "infinite precision" ~

Notation: $X$  $x$
RV ↗  ↖ numerical values

- Can have several RVs on same sample space
- Can make new RVs using others ← ie weight and height

## Probability Mass function (PMF)

↳ Probability distribution of discrete RV

IF we fix $x$, $X = x$ is an event
RV

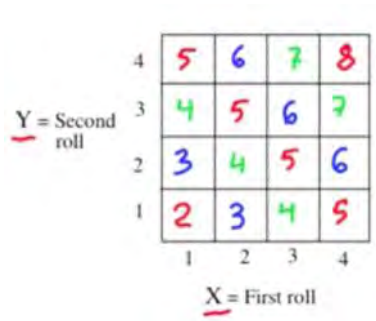
$x = 5 \Rightarrow X = 5$  $\{w : X(w) = 5\}$

Notation: $\underbrace{P_X(x)}_{\text{Probability mass function}} = P(X = x) = P(\{w \in \Omega \text{ s.t } X(w) = x\})$

Properties:

$P_X(x) \geq 0$  $\sum_x P_X(x) = 1$

# Example: PMF Calculation

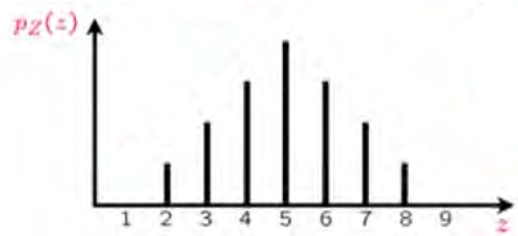



$Z = X + Y$

Find $p_Z(z)$ for all $z$

$p_Z(2) = P(Z=2) = 1/16$

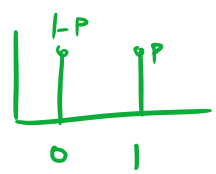
$p_Z(3) = P(Z=3) = 3/16$




# Bernoulli RV

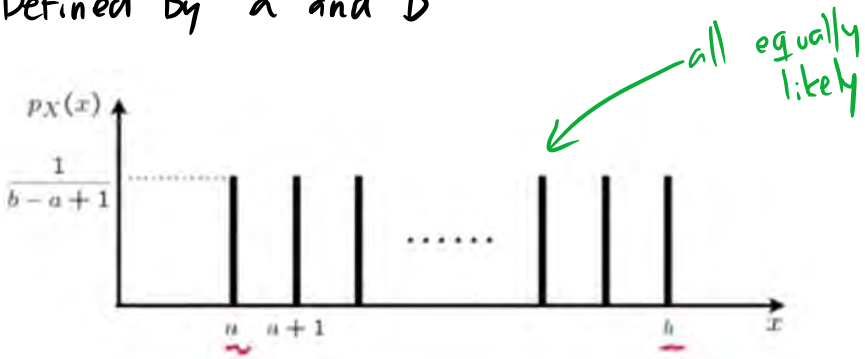
↳ Indicator RV

$$X = \begin{cases} 1, & \text{wp } p \\ 0, & \text{wp } 1-p \end{cases}$$




# Discrete Uniform RV

Defined by $a$ and $b$




# Binomial RV

Defined by $n$ and $p$

ie # tosses

ie. P(heads)

ie # tosses  ie. P(heads)

Model of: # of successes in a given # of independent trials.

## Geometric RV

Defined by $p$.

"Infinitely many independent tosses"

Model of: waiting for something to happen

EX:

$X$: # tosses until first Heads

$$P_X(k) = P(X=k)$$

$$= P(\underbrace{T \ldots T}_{k-1} H) = (1-p)^{k-1} p$$

# Expectation

Expected value of a random variable

$$\underbrace{E[X]}_{\text{average in large \# of independent reps}} = \sum_x x p_X(x)$$

$$E[\text{uniform RV}] = 0 \cdot \frac{1}{n+1} + 1 \cdot \frac{1}{n+1} + \ldots n \frac{1}{n+1},$$

$$= 1/n+1 \,(0+1+2+3 \ldots + n) = \frac{1}{n+1} \cdot \frac{n(n+1)}{2} = n/2$$

center of gravity of pmf

## Properties:

If $X > 0 \Rightarrow E[X] > 0$
If $a \le X \le b \Rightarrow a \le E[X] \le b$
If $c$ constant $\Rightarrow E[c] = c$

## Expected Value Rule

Find $E[Y]$ given $X$ and $Y = g(X)$

- Can average over $x$ or $y$

expected value rule

$$E[Y] = E[g(X)] = \sum_x g(x) \underbrace{p_X(x)}_{\text{only need original PMF}}$$

$$E[X^2] \neq (E[X])^2$$

in general

## Linearity of expectation

$$E[aX + b] = aE[X] + b$$

# Lecture 6: Variance, Conditioning, multiple RVs

Saturday, October 1, 2016 12:04 PM

## Variance

↳ Spread of PMF

mean $\mu = E[X]$

$$\text{var}(X) = E[\underbrace{(X-\mu)}_{\text{Distance from mean}}{}^2]$$

$$\text{var}(X) = E[g(X)] = \sum_x (x-\mu)^2 p_X(x)$$

$g(x) = (x-\mu)^2$

Standard Deviation: $\sigma_X = \sqrt{\text{var}(X)}$

in the right units since var is squared

## Properties

$$\text{var}(aX+b) = a^2 \text{var}(X)$$

⇒ Adding a constant b does not change variance

↳ since spread is still the same

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

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

## Variance of Bernoulli RV

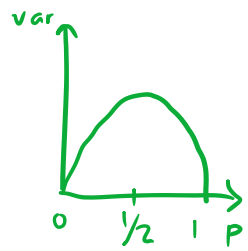
## Variance of Bernoulli RV

$E[X] = p$ From Lec 5

$$\mathrm{var}(X) = \sum_x (x - E[X])^2 p_X(x)$$

$$= \underbrace{(1-p)^2 p}_{x=1} + \underbrace{(0-p)^2(1-p)}_{x=0}$$

$$= p(1-p)$$




## Var of uniform RV

$$\mathrm{var}(X) = E[X^2] - (E[X])^2$$

midpoint

$$= \frac{1}{n+1}\underbrace{(0^2 + 1^2 + 2^2 + \dots n^2)}_{1/6\, n(n+1)(2n+1)} - \left(\frac{n}{2}\right)^2$$

$$= 1/12\, n(n+2)$$

# Conditional PMF and expectation

NORMAL $\Longrightarrow$ CONDITIONAL MODELS

$$p_X(x) = \mathbf{P}(X = x) \qquad p_{X|A}(x) = \mathbf{P}(X = x \mid A)$$

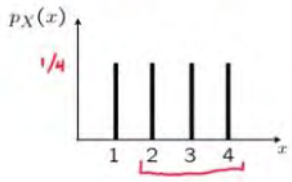
$$\sum_x p_X(x) = 1 \qquad \sum_x p_{X|A}(x) = 1$$

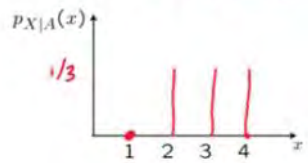
$$\mathbf{E}[X] = \sum_x x p_X(x) \qquad \mathbf{E}[X \mid A] = \sum_x x p_{X|A}(x)$$

$$\mathbf{E}[g(X)] = \sum_x g(x)\, p_X(x) \qquad \mathbf{E}[g(X) \mid A] = \sum_x g(x)\, p_{X|A}(x)$$

**Example of conditioning**

- Let $A = \{X \geq 2\}$




$\mathbf{E}[X] = 2.5$

$\text{var}(X) = \frac{1}{12}(b-a)(b-a+2)$

$= \frac{1}{12} 3 \cdot 5 = \frac{5}{4}.$

$\mathbf{E}[X \mid A] = 3$

$\text{var}(X \mid A) = \frac{1}{3}(4-3)^2 + \frac{1}{3}(3-3)^2 + \frac{1}{3}(2-3)^2 = \frac{2}{3}$

# Total Expectation Theorem

$$E[X] = P(A_1)E[X \mid A_1] + \ldots + P(A_n)E[X \mid A_n]$$

# Geometric RV

$$p_X(x) = (1-p)^{k-1} p$$

"Tosses until first heads"

**Memorylessness:**
Number of **remaining** coin tosses, conditioned on Tails in the first toss, is **Geometric**, with parameter $p$

$$P(\underbrace{X-1=3}_{\text{Need 3 more tosses}} \mid \underbrace{X>1}_{\text{already saw one tail}})$$

since independent

$$= P(T_2 T_3 H_4 \mid T_1) = P(T_2 T_3 H_4)$$

$$= P_X(3)$$

$\Rightarrow$ conditional PMF = unconditional PMF

# Multiple RV + Joint PMF

Joint PMF: $p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

| y \ x | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 4 | 1/20 | 2/20 | 2/20 | |
| 3 | 2/20 | 4/20 | 1/20 | 2/20 |
| 2 | | 1/20 | 3/20 | 1/20 |
| 1 | | 1/20 | | |

joint PMF

$$P_{X,Y}(1,3) = 2/20$$

$$P_X(4) = 1/20 + 2/20$$

Fix $x=4$ and add all

## More than 2 RVs

$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x,y,z)$$

$$p_{X,Y}(x,y) = \sum_z p_{X,Y,Z}(x,y,z)$$

## Functions of multiple RVs

$$Z = g(X,Y)$$

PMF: $p_Z(z) = P(Z=z)$

$$= P(g(X,Y)=z) = \sum_{(x,y):\, g(x,y)=z} p_{X,Y}(x,y)$$

**Expected value rule:** $\mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y)\, p_{X,Y}(x,y)$

## Linearity of Expectations

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

Ex: $E[2X + 3Y - Z]$

$$= E[2X] + E[3Y] - E[Z]$$

$$= 2E[X] + 3E[Y] - E[Z]$$

# Lecture 7: Conditioning on RV, Independence of RVs

Sunday, October 2, 2016 12:24 PM

## Conditional PMs

Recall:

$$p_{X|A}(x|A) = P(X=x|A)$$

- conditioned on a RV

$$p_{X|Y}(x|y) = P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$p_{X|Y}(x|y) = \frac{p_{X|Y}(x,y)}{p_Y(y)}$$

Fix y and treat as a function x

MULTIPLICATION RULE

$$\Rightarrow p_{X,Y}(x,y) = p_Y(y)\, p_{X|Y}(x|y)$$

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

$$\Rightarrow p_{X,Y,Z}(x,y,z) = p_X(x)\, p_{Y|X}(y \mid x)\, p_{Z|X,Y}(z \mid x,y)$$

## Conditional Expectations

$$\mathbf{E}[g(X)] = \sum_x g(x)\, p_X(x) \qquad \mathbf{E}[g(X) \mid A] = \sum_x g(x)\, p_{X|A}(x)$$

$$\mathbf{E}[g(X) \mid Y=y] = \sum_x g(x)\, p_{X|Y}(x \mid y)$$

## Total Probability

$$p_X(x) = \sum_y p_Y(y)\, p_{X|Y}(x|y)$$

## Total Expectation

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

# Independence of RVs

$$P(X=x \text{ and } A) = P(X=x)\cdot P(A) \quad \text{for all } x$$

$$p_{X|A}(x) = p_X(x) \quad \text{for all } x$$

$$P(A \mid X=x) = P(A) \quad \text{for all } x$$

## Multiple RVs

$$P(X=x \text{ and } Y=y) = P(X=x)\,P(Y=y)$$

$$p_{X,Y}(x,y) = p_X(x)\,p_Y(y) \quad \text{for all } x,y$$

$$p_{X|Y}(x|y) = p_X(x)$$

> $X, Y, Z$ are **independent** if:
> $p_{X,Y,Z}(x,y,z) = p_X(x)\,p_Y(y)\,p_Z(z)$, for all $x, y, z$

"no common sources of uncertaint y"

## Example:

| y \ x | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 4 | 1/20 | 2/20 | 2/20 | |
| 3 | 2/20 | 4/20 | 1/20 | 2/20 |
| 2 | | 1/20 | 3/20 | 1/20 |
| 1 | | 1/20 | | |

Independent?

$p_X(1) = 3/20$
$p_{X|Y}(1|1) = 0$ $\neq$ $\Rightarrow$ not independent

Knowing about y changed beliefs about x

Condition on $X \le 2$ $Y \ge 3$

marginal PMF

| Y \ X | | | |
|---|---|---|---|
| | 1/9 | 2/9 | 1/3 |
| | 2/9 | 4/9 | 2/3 |
| | 1/3 | 2/3 | |

$\frac{2}{9} = \frac{1}{3}\cdot\frac{2}{3}$

$4/9 = 2/3 \cdot 2/3$

$\Rightarrow$ independent

$\frac{2}{9} = \frac{1}{3} \cdot \frac{1}{3}$ — $\boxed{\text{[illegible]}}$ $\frac{1}{3}\ \frac{2}{3}$ $\times$ $4/9 = 2/3 \cdot 2/3$

# Independence and Expectations

- In general: $\mathbf{E}[g(X,Y)] \neq g\big(\mathbf{E}[X], \mathbf{E}[Y]\big)$
- Exceptions: $\mathbf{E}[aX+b] = a\mathbf{E}[X] + b$ $\quad$ always true $\mathbf{E}[X+Y+Z] = \mathbf{E}[X] + \mathbf{E}[Y] + \mathbf{E}[Z]$

If $X$, $Y$ are **independent:** $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$

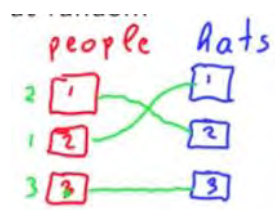
$g(X)$ and $h(Y)$ are also independent: $\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)] \cdot \mathbf{E}[h(Y)]$

# Independence and Variance

- Always true: $\text{var}(aX) = a^2\text{var}(X)$ $\quad$ $\text{var}(X+a) = \text{var}(X)$
- In general: $\text{var}(X+Y) \neq \text{var}(X) + \text{var}(Y)$

If $X$, $Y$ are independent: $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

# Example: The hat problem




n people throw their hat into box then pick up a random hat.

X: # of people that get their hat back

$E[X] = ??$ $\quad = E[X_1] + E[X_2]$

- calculating PMF $p_X(k)$ is too hard.

Let $X_i = \begin{cases} 1 & \text{if } i \text{ selects their hat} \\ 0 & \text{else} \end{cases}$

$$X = X_1 + X_2 + \dots + X_n$$

$$E[X_i] = E[X_1] = P(X_1 = 1) = 1/n$$

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_n]$$
$$= n\frac{1}{n} = 1$$

$\mathrm{var}(X) = ??$

$X_i$'s are <u>not</u> independent !

Recall: $\mathrm{var}(X) = E[X^2] - (E[X])^2$

$$X^2 = \underbrace{\sum_i X_i^2}_{n} + \underbrace{\sum_{i,j:i\neq j} X_i X_j}_{n^2 - n = n(n-1)}$$

since $X_1 = \begin{cases} 0 \\ 1 \end{cases}$

$$E[X_i^2] = E[X_1^2] = E[X_1] = 1/n$$

every person is equal

$$E[X_i X_j] = E[X_1 X_2] = P(X_1 X_2 = 1)$$
$$= P(X_1 = 1, X_2 = 1)$$

$$= \underbrace{P(X_1 = 1)}_{1/n} \underbrace{P(X_2 = 1 \mid X_1 = 1)}_{1/(n-1)}$$

$$\Rightarrow E[X^2] = \underbrace{n}_{\#\text{terms}} \cdot \underbrace{\frac{1}{n}}_{p} + \underbrace{n(n-1)}_{\#\text{ terms}} \frac{1}{n}\frac{1}{(n-1)} = 2$$

$$\Rightarrow \mathrm{var}(X) = 2 - 1 = 1$$

# Unit 4 Sample Problems

Tuesday, October 4, 2016 8:15 PM

$$p_X(x) = \begin{cases} \frac{x^2}{a}, & \text{for } x \in \{-3,-2,-1,1,2,3\}, \\ 0, & \text{otherwise,} \end{cases}$$

where $a > 0$ is a real parameter.

1. Find $a$.

2. What is the PMF of the random variable $Z = X^2$?

① $\sum_x P_X(x) = P_X(-3) + P_X(-2) + P_X(-1) + \ldots + P_X(3)$

$= 2\left(\frac{9}{a}\right) + 2\left(\frac{4}{a}\right) + 2\left(\frac{1}{a}\right)$

$= \frac{28}{a} = 1$ ← probability axiom

$\boxed{\Rightarrow a = 28}$

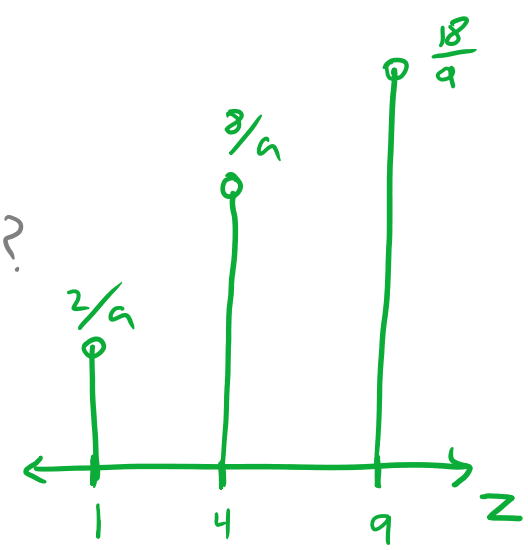
② $Z = X^2$

$P_Z(k) = P(Z = k)$

What could $Z$ take on?

$Z \in \{9, 4, 1\}$

$3^2, -3^2 \ldots$




Note that $Z = 1$ when $X = 1$ or $X = -1$

$$= \frac{1}{a} + \frac{1}{a} = 2/a$$

$$P(Z=4) = P(X=2) \cup P(X=-2)$$
$$= \frac{4}{a} + \frac{4}{a} = \frac{8}{a}$$

$$\boxed{P_Z(k) = \begin{cases} \frac{2}{28} & k=1 \\ 8/28 & k=4 \\ 18/28 & k=9 \\ 0 & \text{otherwise} \end{cases}}$$

Sanity check:

$$\frac{2}{28} + \frac{8}{28} + \frac{18}{28} = 1 \checkmark$$

**Sampling people on buses.** Four buses carrying 148 job-seeking MIT students arrive at a job convention. The buses carry 40, 33, 25, and 50 students, respectively. One of the students is randomly selected. Let $X$ denote the number of students that were on the bus carrying this randomly selected student. One of the 4 bus drivers is also randomly selected. Let $Y$ denote the number of students on his bus.

1. Which of $\mathbf{E}[X]$ or $\mathbf{E}[Y]$ do you think is larger? Give your reasoning in words.

2. Compute $\mathbf{E}[X]$ and $\mathbf{E}[Y]$.

1 Think of 2 busses one with $\boxed{1}$ $\boxed{1000}$

$$\Rightarrow X > Y$$

2

$$P_X(x) = \begin{cases} 40/148 & x=40 \\ 33/148 & x=33 \\ 25/148 & x=25 \\ 50/148 & x=50 \end{cases} \qquad P_Y(y) = \begin{cases} 1/4 & x=40 \\ 1/4 & x=33 \\ 1/4 & x=25 \end{cases}$$

$$\begin{cases} 25/148 & x=25 \\ 50/148 & x=50 \\ 0 & \text{otherwise} \end{cases} \qquad \begin{cases} 1/4 & x=33 \\ 1/4 & x=25 \\ 1/4 & x=50 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \sum_x x P_X(x)$$

$$= 40 \cdot \frac{40}{148} + 33 \cdot \frac{33}{148} + 25 \frac{25}{148} + 50 \cdot \frac{50}{147} = 39$$

$$E[Y] = 40 \cdot \frac{1}{4} \cdot 33 \cdot \frac{1}{4} + 25/4 + 50/4 = 27$$

1. Let $X$ be a random variable that takes nonnegative integer values. Show that

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

*Hint*: Express the right-hand side of the above formula as a double summation then interchange the order of the summations.

2. Use the formula in the previous part to find the expectation of a random variable $Y$ whose PMF is defined as follows:

$$p_Y(y) = \frac{1}{b-a+1}, \qquad y = a, a+1, \ldots, b$$

where $a$ and $b$ are nonnegative integers with $b > a$. Note that for $y = a, a+1, \ldots, b$, $p_Y(y)$ does not depend explicitly on $y$ since it is a uniform PMF.

1

$$E[X] = \sum_{k=-\infty}^{\infty} k P_X(k) = \sum_{k=1}^{\infty} k P_X(k)$$

want to show:

$k=1$

want to show:

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

$$= \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} P_X(i)$$

$$= \sum_{i=1}^{\infty} \sum_{k=1}^{i} P_X(i)$$

$$= \sum_{i=1}^{\infty} i P_X(i) \quad \blacksquare$$

(2) $P_Y(y) = 1/(b-a+1) \quad y = a, \ldots b \quad b \geq a \geq 0$ ints

$$P(Y \geq k) = \begin{cases} 1 & k \leq a \\ 0 & k > b \\ \frac{b-k+1}{b-a+1} & a+1 \leq k \leq b+1 \end{cases}$$

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

$$= \sum_{k=1}^{a} 1 + \sum_{k=a+1}^{b} \frac{b-k+1}{b-a}$$

$$= \frac{2+a}{2}$$

# Lecture 8: Continious RV PDF

Thursday, October 6, 2016 9:02 PM

↳ Probability density function

## Probability Density Function

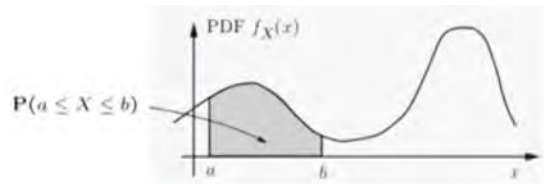
DISCRETE

$$\mathbf{P}(a \le X \le b) = \sum_{x:\, a \le x \le b} p_X(x)$$

$$p_X(x) \ge 0 \qquad \sum_x p_X(x) = 1$$

$\Rightarrow$

CONTINIOUS




$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\,dx$$

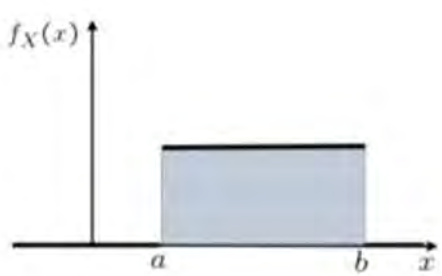
$$f_X(x) \ge 0 \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

CONTINIOUS = can be described by PDF

Probability/unit length

- Single Points have 0 probability

## Uniform RV




$\frac{1}{b-a}$

since total area must equal 1

## Expectation

$$E[X] = \int^{\infty} x f_X(x)\,dx$$

Assume

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

Same as center of gravity

Assume $\int_{-\infty}^{\infty} |g(x) f_X(x)|\,dx < \infty$

## Expected Value Rule

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$$

## Linearity

same as discrete

$$E[aX+b] = aE[X]+b$$

## Variance

$E[X]$

$$\mathrm{var}(X) = E[(X-\mu)^2]$$

$$\mathrm{var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$$

$$\mathrm{var}(X) = E[X^2] - (E[X])^2$$

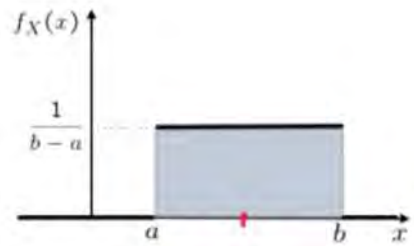
## Linearity

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

## EXAMPLE

$f_X(x)$ ↑

$$E[X] = \int_a^b \frac{1}{\ldots}\,dx \qquad \frac{a+b}{\ldots}$$




$$E[X] = \int_a^b x \frac{1}{b-a}\,dx = \frac{a+b}{2}$$

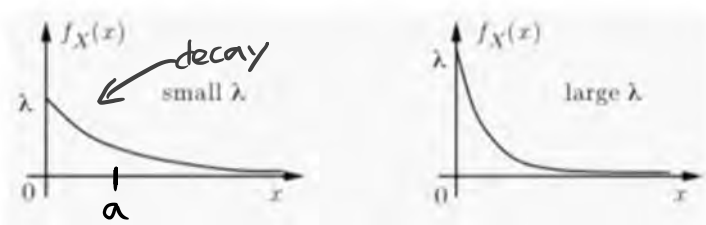
$$E[X^2] = \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{1}{b-a}\left(\frac{b^3}{3} - \frac{a^3}{3}\right)$$

$$\mathrm{var}(X) = E[X^2] - (E[X])^2$$

$$= (b-a)^2/12$$

# Exponential RV

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$




$$P(X \geq a) = \int_a^\infty \lambda e^{-\lambda x}\,dx$$

$$= \lambda\left(-\frac{1}{\lambda}\right)e^{-\lambda x}\Big|_a^\infty$$

$$= e^{-\lambda a}$$

$$E[X] = \int_0^\infty x \cdot \lambda e^{\lambda x}\,dx = 1/\lambda$$

integration by parts

$$E[X^2] = \int_0^\infty x^2 \lambda e^{\lambda x}\,dx = 2/\lambda^2$$

$$\mathrm{var}(X) = E[X^2] - (E[X])^2 = 1/\lambda^2$$

Models: Time to wait until something happens

# Cumulative distribution Function

# Cumulative distribution function (CDF)

$$F_X(x) = P(X \leq x)$$

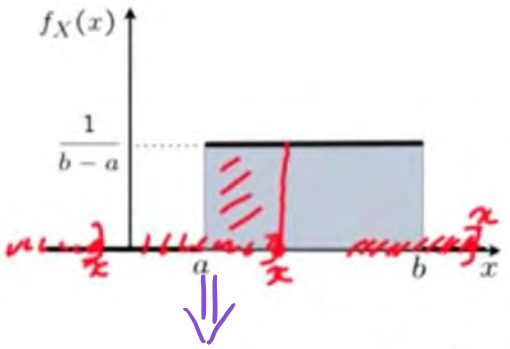
## Continious RV

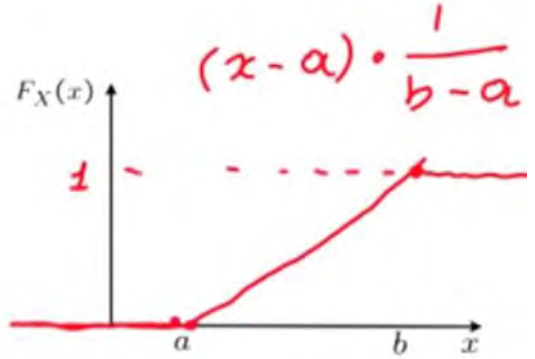
$$= \int_{-\infty}^{x} f_X(t)\,dt \quad \text{if continious RV}$$

$P(X \leq 4) = P(X \leq 3) + P(3 < X \leq 4)$

★ Can go between PDF $\Longleftrightarrow$ CDF

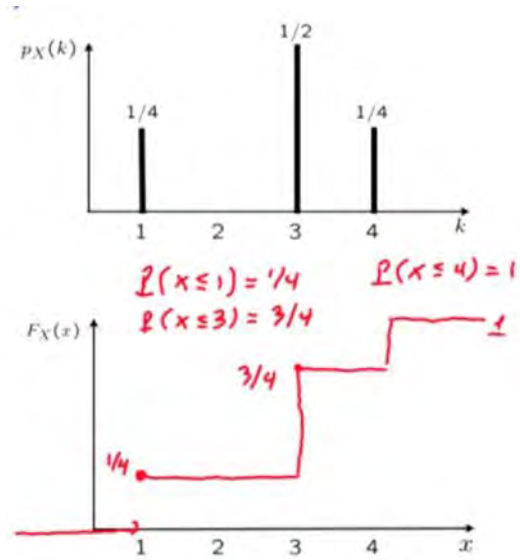
$$\frac{dF_X}{dx}(x) = f_X(x)$$




## Discrete RV

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(x)$$




## CDF Properties

- Non decreasing
  if $y \ge x$ $F_X(y) \ge F_X(x)$

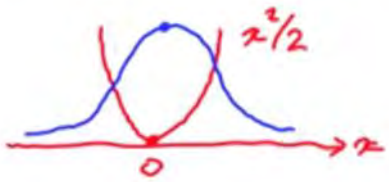
$F_X(x)$ tends to $1$ as $x \to \infty$
$F_X(x)$ tends to $0$ as $x \to -\infty$

★

## Normal RV Gaussian

Model of: Noise $N$

Standard normal $N(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$




calculus:

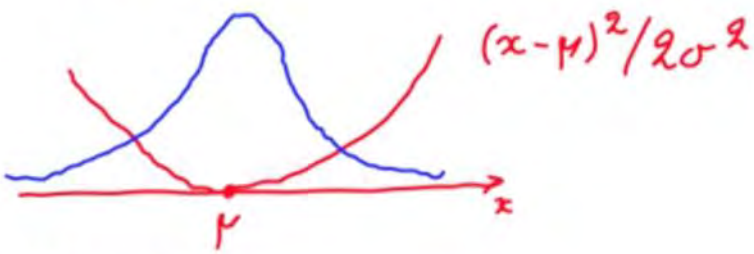
$$\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$$

$\mathbf{E}[X] = 0$ (by symmetry)

$\text{var}(X) = 1$ integrate by parts

General normal $N(\mu, \sigma^2)$: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$\sigma > 0$




$\mathbf{E}[X] = \mu$

$\text{var}(X) = \sigma^2$

$\sigma$ small

## Linear Function

EX: Let $Y = aX + b$ $\quad X \sim N(\mu, \sigma^2)$

$E[Y] = a\mu + b$

$\text{Var}(Y) = a^2\sigma^2$

FACT: Linearity is preserved

$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$

Note that if $a = 0 \Rightarrow Y = b$ ← discrete

$\Rightarrow N(b, 0)$

Note that if $a = 0 \Rightarrow Y = b$ ← discrete
$\Rightarrow N(b, 0)$

According to our conventions, a normal random variable $X \sim N(\mu, \sigma^2)$ is a continuous random variable

- ○ always.
- ◉ if and only if $\sigma \neq 0$. ✔
- ○ if and only if $\mu \neq 0$ and $\sigma \neq 0$.

Answer:
When $\sigma \neq 0$, the distribution of $X$ is described by a PDF, and so $X$ is a continuous random variable. But when $\sigma = 0$, then $X$ has all of its probability assigned to a single point, and therefore it is not a continuous random variable. (For continuous random variables, any single point must have zero probability.)

Standard Normal Table

| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |
| 3.3 | .9995 | .9995 | .9995 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9997 |
| 3.4 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9998 |

$$\Phi(y) = F_Y(y) = P(Y \le y)$$

$$\Phi(0) = P(Y \le 0) = \underline{0.5}$$

$$\Phi(1.16) = \underline{.8770}$$

$$\Phi(-2) = 1 - \Phi(2) = 1 - \underline{.9772}$$

## Standardizing an RV

Let $X$ have mean $\mu$ and var $\sigma^2 > 0$

Let $Y = \dfrac{X-\mu}{\sigma}$ ← # std dvs from mean

$E[Y] = 0$

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

$$\frac{\quad}{\sigma}$$

$$E[Y] = 0$$

$$\mathrm{var}(Y) = \frac{1}{\sigma^2}\,\mathrm{Var}(X) = 1$$

$$X = \mu + \sigma Y$$

If $X$ is normal $\Rightarrow$ $Y \sim N(0,1)$

can use table!

Express event of interest in terms of standard normal

$$X \sim N(6,4) \qquad \sigma = 2$$

st. normal

$$\frac{2-6}{2} \le \frac{X-6}{2} \le \frac{8-6}{2}$$

$$P(2 \le X \le 8) = P(-2 \le Y \le 1)$$

$$= P(Y \le 1) - P(Y \le -2)$$

$$= P(Y \le 1) - (1 - P(Y \le 2))$$

# Lecture 9: Conditioning on event, Multiple RVs

Friday, October 14, 2016 10:52 PM

## Conditional PDF

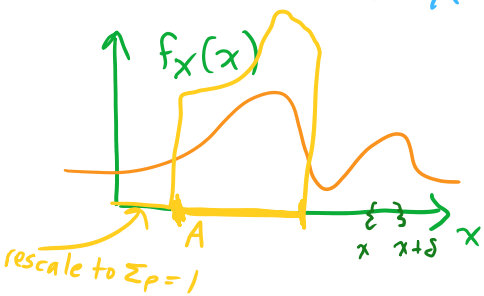
$$p_X(x) = \mathbf{P}(X = x) \qquad f_X(x) \cdot \delta \approx \mathbf{P}(x \le X \le x + \delta)$$

$$p_{X|A}(x) = \mathbf{P}(X = x \mid A) \qquad f_{X|A}(x) \cdot \delta \approx \mathbf{P}(x \le X \le x + \delta \mid A)$$

$$\mathbf{P}(X \in B) = \sum_{x \in B} p_X(x) \qquad \mathbf{P}(X \in B) = \int_B f_X(x)\, dx$$

$$\mathbf{P}(X \in B \mid A) = \sum_{x \in B} p_{X|A}(x) \qquad \mathbf{P}(X \in B \mid A) = \int_B f_{X|A}(x)\, dx$$

## Conditional PDF of $X \mid X \in A$




$$P(x \le X \le x+\delta \mid X \in A) \approx f_{X|X\in A}(x) \cdot \delta$$

$$\frac{P(x \le X \le x+\delta,\ X \in A)}{P(A)} = \frac{P(x \le X \le x+\delta)}{P(A)} = \frac{f_X(x)\,\cancel{\delta}}{P(A)} = f_{X|X\in A}(x) \cdot \cancel{\delta}$$

$$f_{X|X\in A}(x) = \begin{cases} 0, & \text{if } x \notin A \\ \dfrac{f_X(x)}{P(A)} & \text{if } x \in A \end{cases}$$

## Conditional Expectation

$$\mathbf{E}[X] = \sum_x x p_X(x) \qquad \mathbf{E}[X] = \int x f_X(x)\, dx$$

$$\mathbf{E}[X \mid A] = \sum_x x p_{X|A}(x) \qquad \mathbf{E}[X \mid A] = \int x f_{X|A}(x)\, dx$$

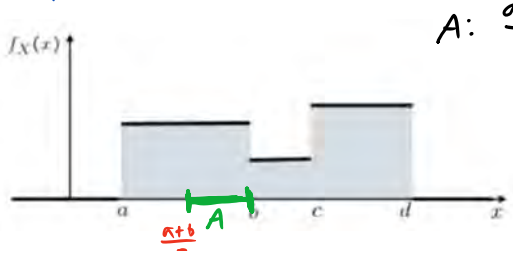
**Expected value rule:**

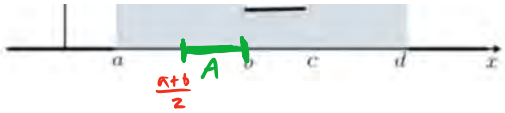
$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x) \qquad \mathbf{E}[g(X)] = \int g(x) f_X(x)\, dx$$

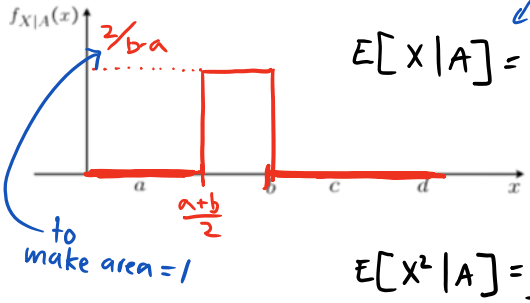
$$\mathbf{E}[g(X) \mid A] = \sum_x g(x) p_{X|A}(x) \qquad \mathbf{E}[g(X) \mid A] = \int g(x) f_{X|A}(x)\, dx$$

## Example

$A: \frac{a+b}{2} \le x \le b$






midpoint

$$E[X|A] = \frac{1}{2}\cdot\frac{a+b}{2} + \frac{1}{2}b = \frac{1}{4}a + \frac{3}{4}b$$

$$E[X^2|A] = \int_{\frac{a+b}{2}}^{b} \frac{2}{b-a}\cdot x^2\,dx =$$

Example (lightbulb)

Do you prefer a used or a new "exponential" light bulb?

Bulb lifetime $T$: exponential($\lambda$)

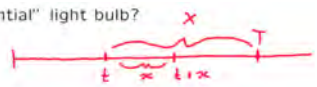
$\mathbf{P}(T > x) = e^{-\lambda x}$, for $x \geq 0$

– we are told that $T > t$ ← time has already passed

– r.v. $X$: remaining lifetime $= T - t$

bulb lives for $x$ for units

$\mathbf{P}(X > x | T > t) = e^{-\lambda x}$, for $x \geq 0$




$$= \frac{P(T-t>x,\ T>t)}{P(T>t)} = \frac{P(T>t+x,\ T>t)}{P(T>t)} = \frac{P(T>t+x)}{P(T>t)}$$

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

$\Rightarrow$ P(used lightbulb alive in next $x$ time) = P(new lightbulb alive in next $x$ time)

$\Rightarrow$ new and used bulbs probabilistically the same

$\Rightarrow$ Memorylessness

(can think of it as: every timestep bulb has probability $p$ of failing)

# Total Probability and Expectation Theorems

$$F_X(x) = P(X \leq x) = P(A_1)P(X \leq x | A_1) + \cdots + P(A_n)P(X \leq x | A_n)$$
$$= P(A_1)F_{X|A_1}(x) + \cdots$$

deriv of both sides

$$\int_{-\infty}^{\infty} x f_X(x)\,dx = P(A_1)\int_{-\infty}^{\infty} x f_{X|A_1}(x)\,dx + \cdots$$

$$\boxed{f_X(x) = \mathbf{P}(A_1)f_{X|A_1}(x) + \cdots + \mathbf{P}(A_n)f_{X|A_n}(x)}$$

$$\boxed{\mathbf{E}[X] = \mathbf{P}(A_1)\mathbf{E}[X \mid A_1] + \cdots + \mathbf{P}(A_n)\mathbf{E}[X \mid A_n]}$$

EXAMPLE $A_1$

## EXAMPLE

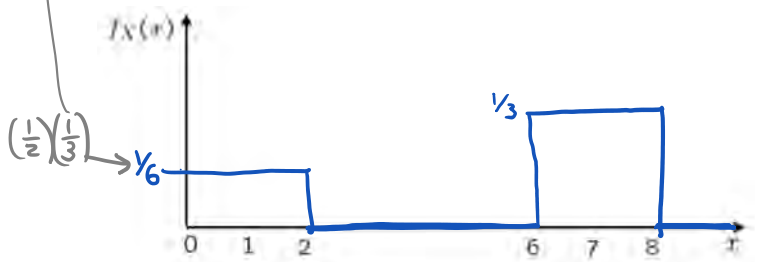
Bill goes to the supermarket shortly, with probability 1/3, at a time uniformly distributed between 0 and 2 hours from now; or with probability 2/3, later in the day at a time uniformly distributed between 6 and 8 hours from now

$A_1$

$P(A_1) = 1/3 \quad f_{X|A_1} \sim \text{uniform}[0,2]$ ← height should be $\frac{1}{2}\boxed{\phantom{x}} = 1$

$P(A_2) = 2/3 \quad f_{X|A_2} \sim \text{Uniform}[6,8]$ $h = 1$

Recall: $f_X(x) = P(A_1) f_{X|A_1}(x) + \ldots + P(A_n) f_{X|A_n}(x)$




Expectation = 1 Expectation = 7

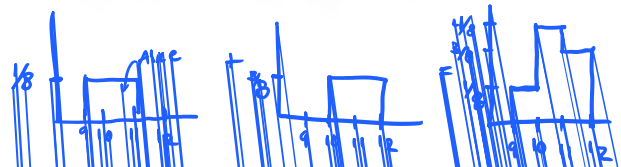
$\Rightarrow E[X] = \frac{1}{3}\cdot 1 + \frac{2}{3}\cdot 7$

## EXAMPLE

On any given day, mail gets delivered by either Alice or Bob. If Alice delivers it, which happens with probability $1/4$, she does so at a time that is uniformly distributed between $9$ and $11$. If Bob delivers it, which happens with probability $3/4$, he does so at a time that is uniformly distributed between $10$ and $12$. The PDF of the time $X$ that mail gets delivered satisfies

a) $f_X(9.5) =$ 1/8 ✔

b) $f_X(10.5) =$ 4/8 ✔




# Mixed Random Variables

↳ discrete and continious

$$X = \begin{cases} \text{uniform } [0,2] & \text{with } p = 1/2 \quad A_1 \\ 1 & \text{with } p = 1/2 \quad A_2 \end{cases}$$

X not discrete since it can take $[0,2] \Rightarrow$ no PMF

X not continious since $P(X=1) = 1/2 \Rightarrow$ no PDF

$\Rightarrow$ use CDF

↳ always well defined

$\Rightarrow$ X is mixed

More generally

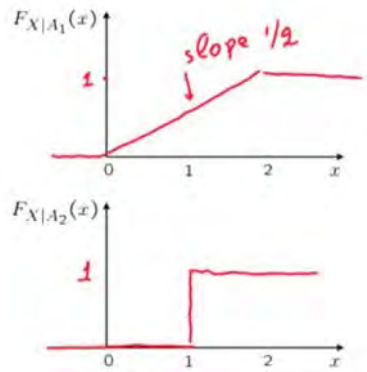
$Y$ discrete $\quad Y = \begin{cases} Y, & \text{with probability } p \\ Z, & \text{with probability } 1-p \end{cases}$

$Z$ continuous

$F_X(x) = p \cdot P(Y \le x) + (1-p) P(Z \le x)$

$= p F_Y(x) + (1-p) F_Z(x)$

$E[X] = pE[M] + qE[Z]$




$$F_X(x) = \mathbf{P}(A_1)F_{X|A_1}(x) + \mathbf{P}(A_2)F_{X|A_2}(x)$$

EXERCISE: A MIXED RANDOM VARIABLE (1 point possible)

A lightbulb is installed. With probability $1/3$, it burns out immediately ($Y$) when it is first installed. With probability $2/3$, it burns out after an amount of time that is uniformly distributed on $[0, 3]$. The expected value of the time until the lightbulb burns out is ($Z$)

$(2/3)1.5$ ✓

$E[X] = \frac{1}{3}E[Y] + \frac{2}{3}E[Z]$

$= (\frac{1}{3})(0) + \frac{2}{3}(1.5) = (2/3)(1.5)$

# Multiple continious RV

## Joint PDF

$p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y) \geq 0$ $\qquad$ $f_{X,Y}(x,y) \geq 0$

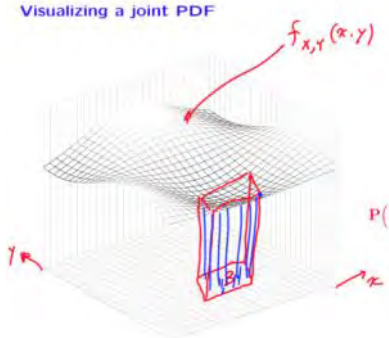
$\mathbf{P}\big((X,Y) \in B\big) = \sum\sum_{(x,y)\in B} p_{X,Y}(x,y)$ $\qquad$ $\mathbf{P}\big((X,Y) \in B\big) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$

$\sum_x\sum_y p_{X,Y}(x,y) = 1$ $\qquad$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$

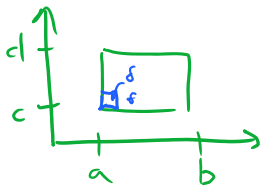
**Definition:** Two random variables are **jointly continuous** if they can be described by a joint PDF

Probability spread over two dimensions

Visualizing a joint PDF




$$\mathbf{P}\big((X,Y) \in B\big) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$




$P(a \leq X \leq c,\ c \leq Y \leq d)$

$= \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy$

$$\mathbf{P}(a \le X \le a+\delta,\ c \le Y \le c+\delta) \approx f_{X,Y}(a,c)\cdot\delta^2$$

$f_{X,Y}(x,y)$: probability per unit area

EXERCISES

EXERCISE: JOINTLY CONTINUOUS R.V.'S (2/2 points)

The random variables $X$ and $Y$ are continuous. Is this enough information to determine the value of $\mathbf{P}(X^2 = e^{2Y})$?

No ✔ Answer: No

The random variables $X$ and $Y$ are jointly continuous. Is this enough information to determine the value of $\mathbf{P}(X^2 = e^{2Y})$?

Yes ✔ Answer: Yes

Answer:

a) There is no information on the relation between the two random variables. If, for example, $X = \sqrt{e^{2Y}}$, the probability is 1, whereas if $X = \sqrt{e^{2Y}} + 1$, then the probability is zero.

b) The set of points on the $x$-$y$ plane that correspond to the event $X^2 = e^{2Y}$ is a one-dimensional curve, which has zero area, and therefore zero probability.

a) The probability of the event that $0 \le Y \le X \le 1$ is of the form $\int_a^b \left(\int_c^d f_{X,Y}(x,y)\,dx\right) dy$.

Find the values of $a, b, c, d$. Each one of your answers should be one of the following: $0$, $x$, $y$, or $1$.

a) For any given $y \in [0,1]$, $x$ ranges from $y$ to 1, yielding $\int_0^1 \int_y^1 f_{X,Y}(x,y)\,dx\,dy$.

b) The probability of the event that $0 \le Y \le X \le 1$ is also of the form $\int_a^b \left(\int_c^d f_{X,Y}(x,y)\,dy\right) dx$. Note the different order of integration as compared to part (a).

Find the values of $a, b, c, d$. Each one of your answers should be one of the following: $0$, $x$, $y$, or $1$.

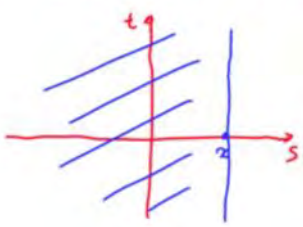
b) For any given $x \in [0,1]$, $y$ ranges from 0 to $x$, yielding $\int_0^1 \int_0^x f_{X,Y}(x,y)\,dy\,dx$.

## From the joint to the marginals

$$p_X(x) = \sum_y p_{X,Y}(x,y) \qquad f_X(x) = \int f_{X,Y}(x,y)\,dy$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y) \qquad f_Y(y) = \int f_{X,Y}(x,y)\,dx$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} f_{X,Y}(s,t)\,dt\right] ds$$



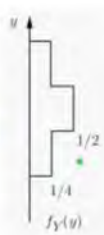

$$f_X(x) = \frac{dF_X}{dx}(x) = [\ \ ]$$

## Uniform joint PDF on a set $S$

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of } S}, & \text{if } (x,y) \in S, \\ 0, & \text{otherwise.} \end{cases}$$

$$= \frac{\text{area}(A \cap S)}{\text{area}(S)}$$

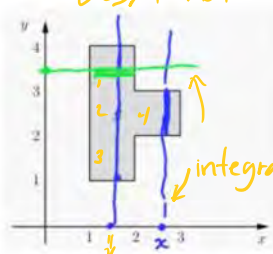
$P(A)$




PDF

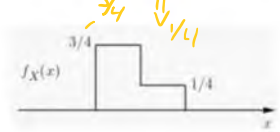
$f_{X,Y} = \frac{1}{4}$ height

JOINT PDF




3/4 1/4

PDF




Joint CDF

PDF $\xrightarrow{\text{integrate}}$ CDF

$$\text{PDF} \underset{\text{differentiate}}{\overset{\text{integrate}}{\rightleftarrows}} \text{CDF}$$

# Lecture 10: Continious RV: · conditioning
· independence
· Bayes Rule

Saturday, October 15, 2016 11:18 AM

## Conditional PDF given an RV

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)\, dx$$

$$f_{X,Y}(x,y) = f_Y(y) \cdot f_{X|Y}(x \mid y)$$ ← Multiplication Rule

← densities, not probabilities

**Total probability and expectation theorems**

| | | |
|---|---|---|
| $p_X(x) = \sum_y p_Y(y) p_{X\|Y}(x \mid y)$ | $f_X(x) = \int_{-\infty}^{\infty} \overbrace{f_Y(y) f_{X\|Y}(x \mid y)}^{f_{X,Y}(x,y)}\, dy$ | Thm. |
| $\mathbf{E}[X \mid Y = y] = \sum_x x p_{X\|Y}(x \mid y)$ | $\mathbf{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X\|Y}(x \mid y)\, dx$ | Def. |
| $\mathbf{E}[X] = \sum_y p_Y(y) \mathbf{E}[X \mid Y = y]$ | $\mathbf{E}[X] = \int_{-\infty}^{\infty} f_Y(y) \mathbf{E}[X \mid Y = y]\, dy$ | |

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

$$= \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y)\, dy}\, dx$$

$$= \int_{-\infty}^{\infty} x f_X(x)\, dx = E[X]$$

- Expected value rule...

$$E[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx$$

## Independence (continious RV)

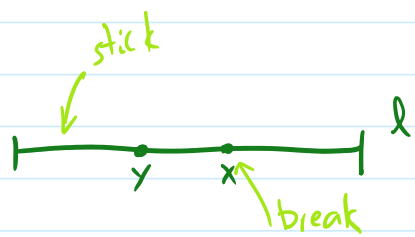
$$\boxed{f_{X,Y}(x,y) = f_X(x)\, f_Y(y), \quad \text{for all } x \text{ and } y}$$

If $X$, $Y$ are **independent**: $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$

$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

$g(X)$ and $h(Y)$ are also independent: $\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)] \cdot \mathbf{E}[h(Y)]$

## EXAMPLE: Stick breaking




- Break at random point x
- then break at y

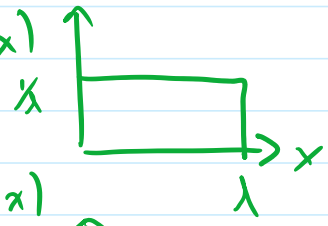
break at random point x

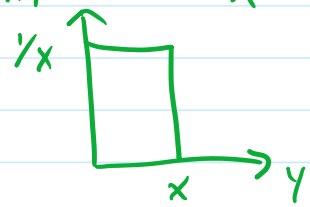
- then break at y

← stick of length $x$

can reconstruct joint using multiplication rule

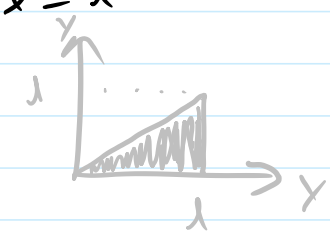
marginals → $f_X(x)$, $f_{Y|X}(y|x)$




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

$$= 1/\lambda x \quad \text{for } 0 \le x \le \lambda$$




Possible x's given y

$$f_Y(y) = \int f_{X,Y}(x,y)\,dx = \int_y^{\lambda} \frac{1}{\lambda x}\,dx = \frac{1}{\lambda}\log(\lambda/y)$$

$$E[Y] = \int y \frac{1}{\lambda} \log\left(\frac{\lambda}{y}\right) dy$$

← hard integral!

Using Total Expectation Theorem

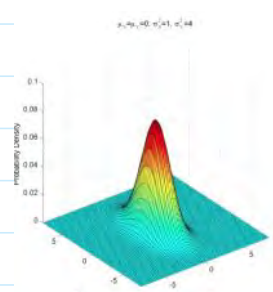
$$E[Y] = \int_0^{\lambda} \frac{1}{\lambda} E[Y|X=x]\,dx = \int_0^{\lambda} \underbrace{\frac{1}{\lambda}}_{f_X(x)} \frac{x}{2}\,dx$$

$$= \frac{1}{2}E[X] = \frac{1}{2}\cdot\frac{\lambda}{2} = \frac{\lambda}{4}$$

makes sense, break in half → half again

$\frac{1}{2} \to \frac{1}{2} = 1/4$

## Independent Normals



JOINT of two independent normals

- Shape of bell
- centered determined by means
- stretched in x and y by var(x) var(y)

· streched in x and y by var(x) var(y)

## Bayes Rule

$$p_{X,Y}(x|y) = \frac{\overbrace{p_X(x)\, p_{Y|X}(y|x)}^{\text{JOINT}}}{p_Y(y)}$$

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

## Mixed Bayes Rule

K- discrete  Y: continious

$$\boxed{p_{K|Y}(k\,|\,y) = \frac{p_K(k)\, f_{Y|K}(y\,|\,k)}{f_Y(y)}}$$

$$\boxed{f_{Y|K}(y\,|\,k) = \frac{f_Y(y)\, p_{K|Y}(k\,|\,y)}{p_K(k)}}$$

$$f_Y(y) = \sum_{k'} p_K(k')\, f_{Y|K}(y\,|\,k')$$

$$p_K(k) = \int f_Y(y')\, p_{K|Y}(k\,|\,y')\, dy'$$

# Lecture 11: Derived distributions

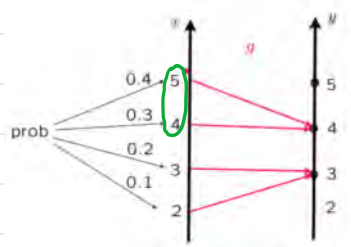
Wednesday, October 19, 2016 11:04 AM

## Derived Distributions

- Given distribution of $X$ find dist of $Y = g(X)$
- Given joint $X, Y$ find dist of $Z = g(X, Y)$

### Discrete case

$$Y = g(X)$$




$$P_Y(4) = P(Y=4)$$
$$= P(X=4) + P(X=5)$$
$$= P_X(4) + P_X(5)$$
$$= .3 + .4$$

$$\Rightarrow P_Y(y) = \sum_{x : g(x) = y} P_X(x)$$

### Linear functions

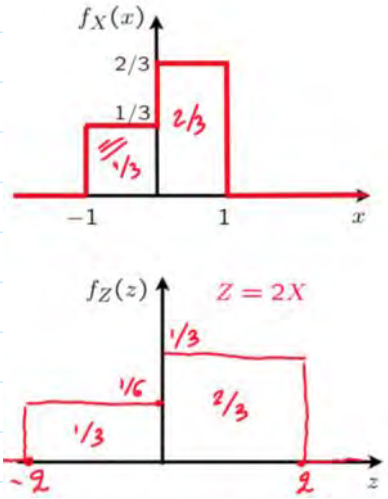
shift

$$\boxed{Y = aX + b: \quad p_Y(y) = p_X\left(\frac{y-b}{a}\right)}$$

stretch

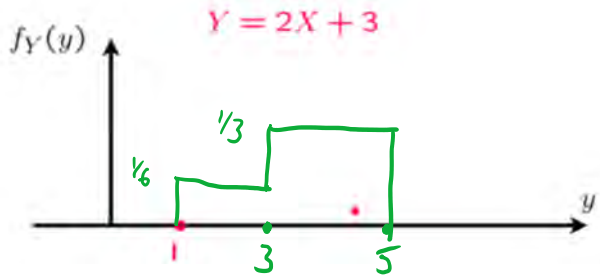
### Continious

#### Linear Function




stretched and scaled

to keep area = 1




same shape, just shifted

$Y = aX + b$

if $a > 0$ $\quad F_Y(y) = P(Y \le y) = P(aX + b \le y)$

$$= P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

(diff both sides

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

$a < 0$

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

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

vertical scaling

## Normal PDFs

$$\boxed{\text{If } X \sim \mathcal{N}(\mu, \sigma^2), \text{ then } aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)}$$

## General Function

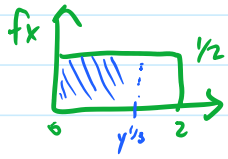
1. Find CDF of $Y$ $\quad F_Y(y) = P(Y \le y) = P(g(X) \le y)$
2. Differentiate $\quad f_Y(y) = \frac{dF_Y}{dy}(y)$

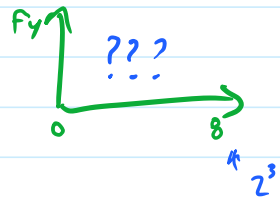
EXAMPLES

$Y = X^3$ $\quad X$ uniform on $[0, 2]$

EXAMPLES

$y = X^3$ $\quad$ X uniform on $[0, 2]$




① CDF of Y

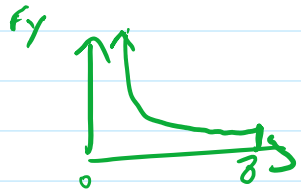
$$F_Y(y) = P(Y \le y) = P(X^3 \le y) = P(X \le y^{1/3})$$

For $0 \le y \le 8$

$$= (1/2)(y^{1/3})$$

② Differentiate

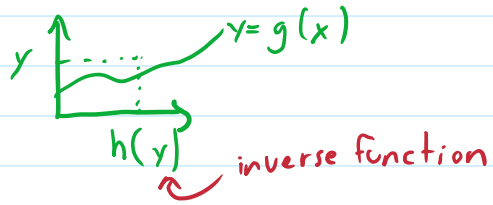
$$\frac{dF_Y}{dy}(y) = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) y^{-2/3} = 1/6 \cdot 1/y^{2/3}$$




## Monotomic function

PDF of $y = g(X)$ when $g$ is monotonic

Assume $g$ strictly increasing $\quad x < x' \Rightarrow g(x) < g(x')$




$$F_Y(y) = P(Y \le y) = P(X \le h(y)) = F_X(h(y))$$

$$f_Y(y) = f_X(h(y)) \frac{dh}{dy}(y)$$

Assume decreasing

$$F_Y(y) = P(Y \le y) = P(X \ge h(y))$$

$$f_Y(y) = -f_X(h(y)) \frac{dh}{dy}(y)$$

$$\Rightarrow \boxed{f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|}$$

inverse function

EXAMPLE

inverse function

EXAMPLE

$Y = X^2$ $\quad$ $X$ uniform on $[0,1]$

$y=x^2$ is monotonic in this range

$$y = x^2 \iff x = \sqrt{y} = h(y)$$

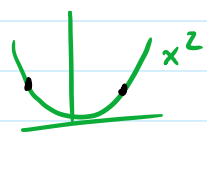
$$f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$$

$$f_Y(y) = \frac{1}{2\sqrt{y}} \quad \text{for } 0 < y < 1$$

EXAMPLE

NON MONOTOMIC

$Y = X^2$




Discrete case

$$P_Y(y) = P_X(\sqrt{y}) + P_X(-\sqrt{y})$$

Contineous Case

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(|X| \le \sqrt{y}) = P(-\sqrt{y} \le X \le \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

derivative

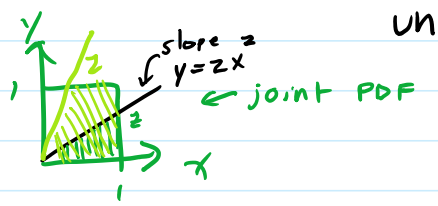
chain rule

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})\,{-1/2\sqrt{y}}$$

# Function of Multiple RVs $\quad Z = g(X, Y)$

EXAMPLE

$Z = Y/X$ $\quad$ $X, Y$ independent uniform on $[0,1]$




① Find CDF of $Z$

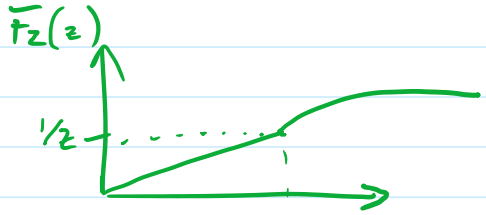
$$F_Z(z) = P\left(\frac{Y}{X} \le z\right) = 0 \quad \text{if } z < 0$$

$$F_Z(z) = \quad 0 < z < 1$$

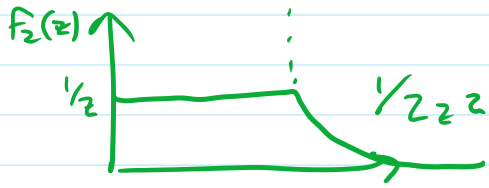
$$F_Z(z) = P\left(\frac{Y}{X} \le z\right) = 0 \quad \text{if } z < 0$$

$$= \underbrace{(1/2)z}_{\text{Area}} \quad 0 \le z \le 1$$

$$= 1 \cdot (1/2)(1/z) \quad z > 1$$




② Differentiate



# Lecture 12: Covariance and Correlation

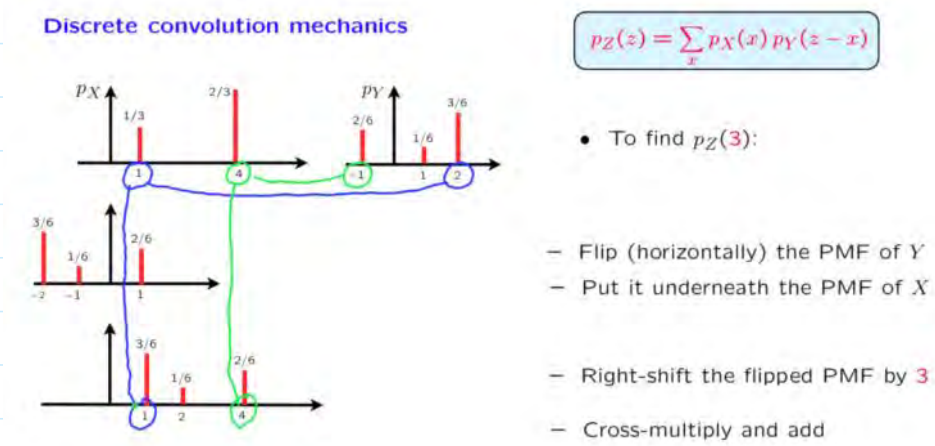
Saturday, October 22, 2016 6:47 PM

Sums of independent RVs

$Z = X + Y$ $\quad$ $X, Y$ independent

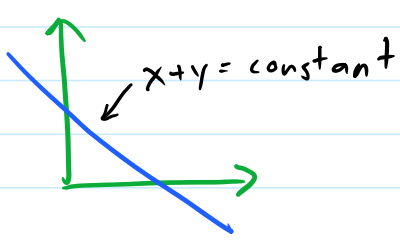
$$p_Z(3) = \ldots + P(X=0, Y=3) + P(X=1, Y=2) + \ldots$$
$$+ p_X(0)\,p_Y(3) + p_X(1)\,p_Y(2) + \ldots$$

$$p_Z(z) = \sum_x P(X=x, Y=z-x)$$
$$= \underbrace{\sum p_X(x)\,p_Y(z-x)}_{\text{convolution!}}$$




## Continuous case

$Z = X + Y$ $\quad$ $X, Y$ independent




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

PROOF

Recall $f_{X|Y}(x|y) = f_X(x)$

$f_{Z|X}(z|x) = f_Y(z-x)$

Assume

$f_{X,Z}(x,z) = f_X(x)\,f_Y(z-x)$

$f_Z(z) = \int f_X(x)\,f_Y(z-x)\,dx$

★ Sum of many independent normal RVs is normal

## Covariance

$$\text{cov}(X,Y) = E[(X - E[X]) \cdot (Y - E[Y])]$$

↳ Tells wether two RVs tend to move together

X,Y independent: $\text{cov}(X,Y) = 0$

$\text{cov}(X,Y) \neq 0 \not\Rightarrow$ independance

### Properties

$$\text{cov}(X,X) = E[(X - E[X])^2] = \text{var}(X) = E[X^2] - (E[X])^2$$

$$\Rightarrow \text{cov}(X,Y) = E[XY] - E[X]E[Y]$$

$$\text{cov}(aX + b, Y) = a \cdot \text{cov}(X,Y)$$

Assume 0 means

$$\text{cov}(X, Y+Z) = \text{cov}(X,Y) + \text{cov}(Y,Z)$$

expanding and substituting

$$\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2) + 2\,\text{cov}(X_1, X_2)$$

$$\boxed{\text{var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \text{var}(X_i) + \sum_{\{(i,j):\, i \neq j\}} \text{cov}(X_i, X_j)}$$

## The correlation coefficient

↳ Dimension less

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

$$-1 \leq \rho \leq 1$$

## Properties

Independent $\Rightarrow \rho = 0$ "uncorrelated"

$\rho(X,Y) = 0 \not\Rightarrow$ Independent

$\rho(X,X) = \mathrm{var}(X)/\sigma_X^2 = 1$ "extreme dependance"

$|\rho| = 1 \iff (X - E[X]) = c(Y - E[Y])$ "linearly dependent"

$\mathrm{cov}(aX + bY) = a \cdot \mathrm{cov}(X,Y) \Rightarrow \rho(aX+bY) = \mathrm{sign}(a) \cdot \rho(X,Y)$

$\hookrightarrow$ ie changing units does not change $\rho$

# Lecture 18: Inequalities, convergance and weak law of large numbers

Monday, October 31, 2016 3:55 PM

## The Markov Inequality

- Learn about probability of "extreme event"

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

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

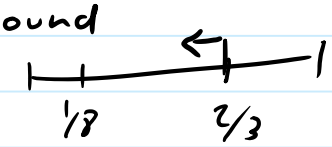
$$\Rightarrow aP(X \geq a) = E[Y] \leq E[X]$$

### Examples

$X$ uniform $[-4, 4]$

$$P(X \geq 3) \leq P(|X| \geq 3) \leq \frac{E[|X|]}{3} = 2/3$$

Actual: $1/8$ Bad bound




EXERCISE: MARKOV INEQUALITY (1 point possible)

Let $Z$ be a nonnegative random variable that satisfies $\mathbf{E}[Z^4] = 4$. Apply the Markov inequality to the random variable $Z^4$ to find the tightest possible (given the available information) upper bound on $\mathbf{P}(Z \geq 2)$.

$\mathbf{P}(Z \geq 2) \leq$ 4/2 ✖ Answer: 0.25

Answer:
We have

$$\mathbf{P}(Z \geq 2) = \mathbf{P}(Z^4 \geq 16) \leq \frac{\mathbf{E}[Z^4]}{16} = \frac{4}{16} = \frac{1}{4}.$$

## Chebyshev inequality

Positive number

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

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

$$P(|X-\mu| \geq k\sigma) \leq 1/k^2$$

↳ being k stdev from mean

# The Weak Law of Large numbers

- Start with distribution with mean $\mu$ and variance $\sigma^2$
- draw independent and identically distributed RVs $X_1, X_2 \ldots$ (iid)

an RV → not $\mu = E[X_i]$

Sample Mean: $M_n = \frac{X_1 + X_2 + \ldots X_n}{n}$ ← observed

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

← Theoretical not observed

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

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

Draw more RVs

**WLLN:** For $\epsilon > 0$, $\mathbf{P}(|M_n - \mu| \geq \epsilon) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$, as $n \to \infty$

- Sample mean unlikely to be far off from true mean

# Polling

p: Fraction of population that will vote yes

randomly selected, independantly

ith person polled $X_i = \begin{cases} 1 & \text{if yes} \\ 0 & \text{else} \end{cases}$

person polled $X_i = \{ 0$ else

$\Rightarrow E[X_i] = p$

$M_n = (X_1 + \dots + X_n)/n$ fraction of "yes"

confidence

- $\mathbf{P}(|M_{10,000} - p| \geq \underline{0.01}) \leq \frac{\sigma^2}{n\varepsilon^2} = \frac{p(1-p)}{10^4 \cdot 10^{-4}} \leq \frac{1}{4}$ ← want $\leq 5\%$

accuracy

$\frac{1/4}{n\, 10^{-4}} \leq \frac{5}{10^2} \iff n \geq \frac{10^6}{20} = 50{,}000$ ← will suffice

# Convergance of RVs

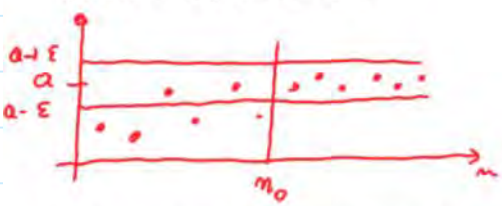
**Definition:** A sequence $Y_n$ **converges in probability** to a number $a$ if:

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

- Ordinary convergence
  - Sequence $a_n$; number $a$

    $a_n \to a$

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

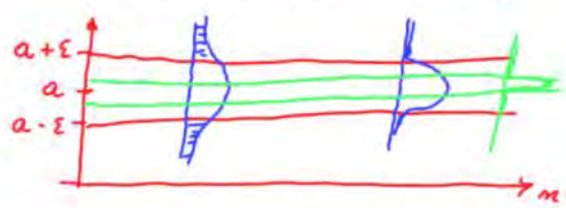



- For every $\epsilon > 0$, there exists $n_0$, such that for every $n \geq n_0$, we have $|a_n - a| \leq \epsilon$

- Convergence in probability
  - Sequence $Y_n$; number $a$

    $Y_n \to a$

- for any $\epsilon > 0$, $\mathbf{P}(|Y_n - a| \geq \epsilon) \to 0$




"(almost all) of the PMF/PDF of $Y_n$ eventually gets concentrated (arbitrarily) close to $a$"

## Properties:

Suppose $X_n \to a$ $Y_n \to b$

If $g$ continuous: $g(X_n) \to g(a)$ ie $X_n^2 \to a^2$

$X_n + Y_n \to a + b$

- convergance of RVs $\not\Rightarrow$ convergance of expectation

$E[X_n]$ need not converge to $a$

## Example

$\epsilon > 0 \quad P(|Y_n - 0| \geq \epsilon) = 1/n \xrightarrow[n\to\infty]{} 0$

$\Rightarrow Y_n \xrightarrow[n\to\infty]{i.p.} 0$

$E[Y_n] = n^2(1/n)$

$\to \infty$ as $n \to \infty$

# L13: Iterated Expectation
# Total Variance

Friday, November 4, 2016 7:43 PM

Recall:

$$E[X \mid Y=y] = \sum_x x\, p_{X,Y}(x \mid y) = \underbrace{g(y)}_{\text{Number}}$$

A function of $y$

$\underbrace{g(Y)}_{\text{Random Variable}}$: Takes value $E[X \mid Y=y]$ if $Y=y$

## Law of Iterated Expectations

Since $g(Y)$ is an r.v. it has all the properties of an r.v. and we can find its expectation/variance

$$g(y) = E[X \mid Y=y]$$

$$g(Y) \triangleq E[X \mid Y]$$

defined to be

$$E[E[X \mid Y]] = E[g(Y)]$$

$$= \sum_y g(y)\, p_Y(y)$$

Expected Value Rule

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

Total Expectation

$$= E[X]$$

* Expectation of conditional expectations = unconditional expectation

Exercise

(a) The law of iterated expectations tells us that $\mathbf{E}[\mathbf{E}[X \mid Y]] = \mathbf{E}[X]$. Suppose that we want apply this law in a conditional universe, given another random variable $Z$, in order to evaluate $\mathbf{E}[X \mid Z]$. Then:

$\mathbf{E}[\mathbf{E}[X \mid Y, Z] \mid Z] = \mathbf{E}[X \mid Z]$

False ✖ Answer: True

$\mathbf{E}[\mathbf{E}[X \mid Y] \mid Z] = \mathbf{E}[X \mid Z]$

True ✖ Answer: False

$\mathbf{E}[\mathbf{E}[X \mid Y, Z]] = \mathbf{E}[X \mid Z]$

False ✔ Answer: False

(a) The first statement is correct: it is just the law of iterated expectations where all the expectations now involve the additional conditioning on $Z$.

The second statement is incorrect because the inner conditional expectation should be evaluated in a conditional universe where $Z$ is given. For a concrete counterexample, suppose that $X$ and $Y$ are independent and zero mean, and that $X = Z$. Because of independence, $\mathbf{E}[X \mid Y] = \mathbf{E}[X] = 0$, and the left-jand side evaluates to zero. On the other hand, the right-hand side is equal to $Z$.

For the third statement, note that the left-hand side is a number (the ordinary expectation of the random variable $\mathbf{E}[X \mid Y, Z]$ ), whereas the right-hand side is a random variable (a function of $Z$). Hence the statement is incorrect.

## Stick breaking revisited

- Stick of length $\ell$, break at uniformly chosen point $Y$
  , break what is left " point $X$

$E[X, Y=y] = y/2$

$E[X \mid Y] = Y/2$

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

# Conditional Variance

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

$\text{var}(X \mid Y=y) = E[(X - E[X \mid Y=y])^2 \mid Y=y]$

$\text{var}(X \mid Y)$ : rv that takes value $\text{var}(X \mid Y=y)$ when $Y=y$

Example: $X$ uniform on $[0, Y]$

$\text{var}(X \mid Y=y) = y^2/12$

$\text{var}(X \mid Y) = Y^2/12$

**Law of total variance:** $\text{var}(X) = \mathbf{E}[\text{var}(X \mid Y)] + \text{var}(\mathbf{E}[X \mid Y])$

## Derivation of the law of total variance

$$\boxed{\text{var}(X) = \mathbf{E}\big[\text{var}(X \mid Y)\big] + \text{var}\big(\mathbf{E}[X \mid Y]\big)}$$

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

$\text{var}(X \mid Y = y) = E[X^2 \mid Y=y] - (E[X \mid Y=y])^2$ for all $y$

$\text{var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$

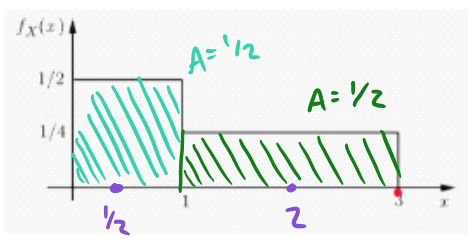
$\mathbf{E}\big[\text{var}(X \mid Y)\big] = E[X^2] - E\big[(E[X \mid Y])^2\big]$

$+$

$\text{var}\big(\mathbf{E}[X \mid Y]\big) = E\big[(E[X \mid Y])^2\big] - \big(E[E[X \mid Y]]\big)^2$

$(E[X])^2$

## Example




$E[X \mid Y] = \begin{cases} E[X \mid Y=1] = 1/2 & \text{w.p. } 1/2 \\ E[X \mid Y=2] = 2 & \text{w.p. } 1/2 \end{cases}$ midpoints

$E[E[X \mid Y]] = (1/2)(1/2) + (1/2)(2) = 5/4 = E[X]$

$\text{var}(X \mid Y) = \begin{cases} \text{var}(X \mid Y=1) = 1/12 & \text{w.p. } 1/2 \\ \text{var}(X \mid Y=2) = 4/12 & \text{w.p. } 1/2 \end{cases}$ uniform over length = 1, length = 2

$E[\text{var}(X \mid Y)] = (1/2)(1/12) + (1/2)(4/12) = 5/24$

$\text{var}(E[X \mid Y]) = \frac{1}{2}\left(\frac{1}{2} - \frac{5}{4}\right)^2 + \frac{1}{2}\left(2 - 5/4\right)^2 = 9/16$

mean

$\text{var}(X) = E[\text{var}(X \mid Y)] + \text{var}(E[X \mid Y])$

$5/24 + 9/16 = 37/48$

## Sum of a random number of independent rvs.

$N$: # of stores to visit

$X_i$: \$\$ spent on store $i$

- $X$: independent and identically distributed
- independent of $N$

Total \$\$ spent $= Y = X_1 + \dots X_N$ ← an RV

Find $E[Y]$

$$E[Y|N=n] = E[X_1 + \dots X_N \mid N=n] = E[X_1 + \dots + X_n]$$
$$= nE[X]$$

Total Expectation Thm

$$E[Y] = \sum p_N(n) E[Y|N=n] = \underbrace{\sum_n p_N(n) n}_{E[N]} E[X] = E[N]E[X]$$

Law of iterated expectations

$$E[Y] = E[E[Y|N]] = E[N \underbrace{E[X]}_{\text{constant}}] = E[N]E[X]$$

Find $var(Y)$

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

$$var(Y|N=n) = var(X_1 + \dots X_n) = n\, var(X)$$
$$var(Y|N) = N\, var(X)$$
$$E[var(Y|N)] = E[N\, var(X)] = E[N]\, var(X)$$

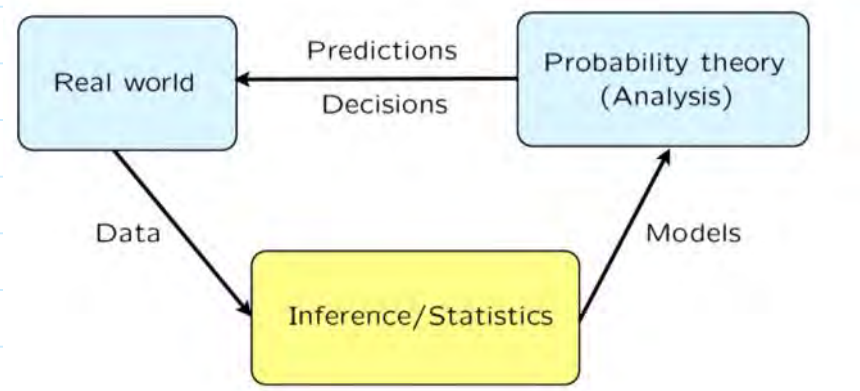
$$var(Y) = E[N]\, var(X) + (E[X])^2 var(N)$$

Two sources of randomness

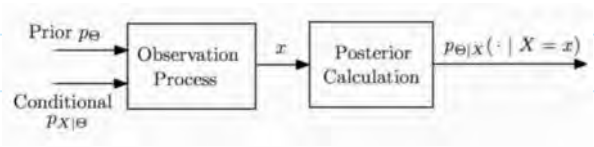
# L14: Introduction to Bayesian inference

Saturday, November 5, 2016 1:05 PM

## Inference




## Bayesian Inference Framework




1. Unknown $\theta$
   - treated as an RV
   - prior distribution $p_\theta$
2. Observation $X$
   - model $p_{X|\theta}$
3. Bayes Rule
   - find $p_{\theta|X}(\cdot \mid X = x)$ ← posterior

### Coming up with estimates

- Maximum a posteriori probability (MAP):

  $p_{\Theta|X}(\theta^* \mid x) = \max_\theta p_{\Theta|X}(\theta \mid x)$,

  $f_{\Theta|X}(\theta^* \mid x) = \max_\theta f_{\Theta|X}(\theta \mid x)$

- Conditional expectation: $\mathbf{E}[\Theta \mid X = x]$ (LMS: Least Mean Squares)

## Discrete $\Theta$, discrete $X$

alternative hypothesis

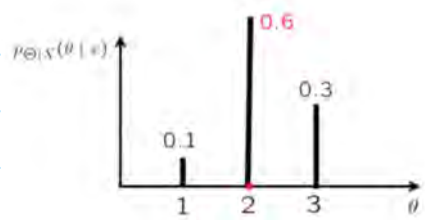
Parameter

observation

$$p_{\Theta|X}(\theta \mid x) = \frac{p_\Theta(\theta)\, p_{X|\Theta}(x \mid \theta)}{p_X(x)}$$

0.6

Parameter  observation




$$p_{\Theta|X}(\theta \mid x) = \frac{p_\Theta(\theta)\, p_{X|\Theta}(x \mid \theta)}{p_X(x)}$$

$$p_X(x) = \sum_{\theta'} p_\Theta(\theta')\, p_{X|\Theta}(x \mid \theta')$$

MAP Rule: $\hat{\Theta} = 2$

LMS Rule: $\hat{\Theta} = E[\Theta \mid X = x] \approx 2.2$

## Conditional Probability of error

$P(\hat{\Theta} \neq \Theta \mid X = x) = .4$  smallest If Map Rule

## Overall Probability of error

$$P(\hat{\Theta} \neq \Theta) = \sum_x P(\Theta \neq \Theta \mid X = x)\, P_X(x)$$

$$= \sum_\Theta P(\hat{\Theta} \neq \Theta \mid \Theta = \theta)\, P_\Theta(\theta)$$

## Discrete $\Theta$, continuous $X$

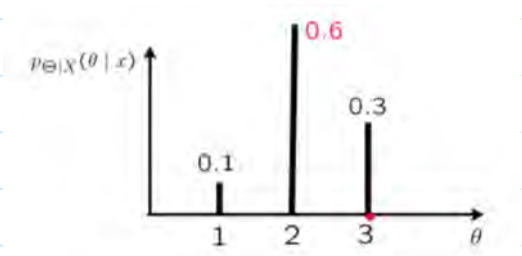
send signal $\Theta \in \{1, 2, 3\}$

$X = \Theta + W$  noise

observation

$W \sim N(0, \sigma^2)$ indep of $\Theta$

$$f_{X|\Theta}(x \mid \theta) = f_W(x - \theta)$$




MAP: $\hat{\Theta} = 2$

Overall P of error same as above.

Continuous X does not

Continuous X does not matter any more

## continuous $\Theta$, continuous $X$

- Happens in linear normal models

$$X = \Theta + W$$

Independent Normals

$$E = \int_0^\infty \theta 2 e^{-2\theta} d\theta$$

$$-\theta e^{-2\theta} \Big|_0^\infty = 0 - -\theta = \theta$$

## Example

discrete observations

# heads

### Unknown Bias of coin and Beta distribution

- coin with bias $\theta$
- fix $n$. $K = \#$ of heads

prior $f_\Theta(\cdot)$ uniform on $[0,1]$

ie complete ignorance

$$f_{\Theta|K}(\theta|k) = \frac{f_\Theta(\theta)\, p_{K|\Theta}(k|\theta)}{p_K(k)}$$

$$= \frac{1 \cdot \binom{n}{k} \theta^k (1-\theta)^{n-k}}{p_K(k)}$$

binomial Probabilities

constant

$$= \underbrace{\frac{1}{d(n,k)} \theta^k (1-\theta)^{n-k}}_{\text{Beta distribution}} \quad \text{if } 0 \le \theta \le 1 \text{ else } 0$$

add +1 to exponents. (because historical

Beta distribution

w/ parameters $(k+1,\ n-k+1)$ add +1 to exponents. (because historical reasons)

IF Prior is Beta: $f_\Theta(\theta) = \frac{1}{c}\theta^\alpha(1-\theta)^\beta$

$$f_{\Theta|K}(\theta|k) = \frac{\frac{1}{c}\theta^\alpha(1-\theta)^\beta \cdot \binom{n}{k}\theta^k(1-\theta)^{n-k}}{p_K(k)}$$

$$= d\,\theta^{\alpha+k}(1-\theta)^{\beta+n-k}$$ if $0 \le \theta < 1$ else $0$

$\hat{\theta}_{MAP}$ = value of $\theta$ that maximizes Posterior $\leftarrow f_{\Theta|K}(\theta|k) = \frac{1}{d(n,k)}\theta^k(1-\theta)^{n-k}$

$= \max\left[k\log\theta + (n-k)\log(1-\theta)\right]$ (log both sides)

$k/\theta - \frac{n-k}{1-\theta} = 0 \leftarrow$ max (derivative)

$\Rightarrow \hat{\theta}_{MAP} = k/n$

estimator

$\hat{\Theta} = K/n$

RV

LMS

$$E[\Theta|K=k] = \int_0^1 \theta f_{\Theta|K}(\theta|k)\,d\theta$$

$$= \frac{1}{d(n,k)}\int_0^1 \theta^{k+1}(1-\theta)^{n-k}\,d\theta$$

Magic Formula: $\boxed{\int_0^1 \theta^\alpha(1-\theta)^\beta\,d\theta = \frac{\alpha!\,\beta!}{(\alpha+\beta+1)!}}$ $\alpha \ge 0$, $\beta \ge 0$

$$= \frac{1}{\frac{k!\,\cancel{(n-k)!}}{(n+1)!}} \cdot \frac{(k+1)!\,\cancel{(n-k)!}}{(n+2)!} = \frac{k+1}{n+2} \approx k/n$$

when n is large so it's like

$(n+1)!$ $n+2$

when n is large so its like MAP estimator

# Summary

unknown

- Problem data: $\underbrace{p_\Theta(\cdot)}_{\text{Prior}}$, $\underbrace{p_{X|\Theta}(\cdot|\cdot)}_{\text{Observation}}$

  - Given value $x$ of $X$ find $\underbrace{p_{\Theta|X}(\cdot|x)}_{\text{Posterior}}$

    - Use Bayes Rule (4 choices)

## Estimator

- calculate certain value as function of observed data

$\hat{\Theta} = g(X)$ estimator

$\theta = g(x)$ estimate

MAP: $\hat{\theta}_{MAP} = g_{MAP}(x)$ - maximizes $p_{\Theta|X}(\theta|x)$

LMS: $\hat{\theta}_{LMS} = g_{LMS}(x) = E[\Theta | X = x]$

## Performance

$P(\hat{\Theta} \neq \Theta | X = x)$

$P(\hat{\Theta} \neq \Theta)$

## Summary

- Problem data: $p_\Theta(\cdot)$, $p_{X|\Theta}(\ \mid \cdot)$
- Given the value $x$ of $X$: find, e.g., $p_{\Theta|X}(\cdot \mid x)$
  - using appropriate version of the Bayes rule (4 choices)
- Estimator $\widehat{\Theta} = g(X)$ Estimate $\widehat{\theta} = g(x)$
  - MAP: $\widehat{\theta}_{\text{MAP}} = g_{\text{MAP}}(x)$ maximizes $p_{\Theta|X}(\theta \mid x)$
  - LMS: $\widehat{\theta}_{\text{LMS}} = g_{\text{LMS}}(x) = \mathbf{E}[\Theta \mid X = x]$
- Performance evaluation of an estimator $\widehat{\Theta}$

$\mathbf{P}(\widehat{\Theta} \neq \Theta \mid X = x)$ $\mathbf{E}\big[(\widehat{\Theta} - \Theta)^2 \mid X = x\big]$

$\mathbf{P}(\widehat{\Theta} \neq \Theta)$ $\mathbf{E}\big[(\widehat{\Theta} - \Theta)^2\big]$

# L15: Linear models with normal noise

Saturday, November 12, 2016 4:36 PM

## Recognizing Normal PDFs

$X \sim N(\mu, \sigma^2)$ $\qquad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$c \cdot e^{-8(x-3)^2}$ : $\mu = 3$ $\quad \frac{1}{2\sigma^2} = 8 \Rightarrow \sigma^2 = 16$ $\quad c = \frac{1}{4\sqrt{2\pi}}$

since Integral needs to be 1

$f_X(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)}$ : $\alpha > 0$

exponent: $\alpha x^2 + \beta x + \gamma = \alpha\left(x^2 + \frac{\beta}{\alpha}x + \frac{\gamma}{\alpha}\right)$

completing the square

$= \alpha\left(\left(x + \frac{\beta}{2\alpha}\right)^2 - \frac{\beta^2}{4\alpha^2} + \frac{\gamma}{\alpha}\right)$

$\Rightarrow f_X(x) = e^{-\alpha\left(x + \frac{\beta}{2\alpha}\right)^2} \underbrace{e^{-\alpha\left(-\beta^2/4\alpha^2 - \frac{\gamma}{\alpha}\right)}}_{\text{constant}}$

$\mu = -\beta/2\alpha$ $\quad \frac{1}{2\sigma^2} = \alpha \Rightarrow \sigma^2 = 1/2\alpha$

OR If we know it is normal

mean = peak of pdf $\Rightarrow$ maximize pdf $\Rightarrow$ minimize exponent

$\Downarrow$

$\frac{d}{dx} = 0$

exponent: $\alpha x^2 + \beta x + \gamma$

$2\alpha x + \beta = 0$

$\Rightarrow \mu = \beta/2\alpha$ ✓

$-4x^2 - 24x + 30$

$-8x - 24 = 0$

$-8x = 24$

$x = -24/8$

## Estimating a normal RV w/ normal noise

$X = \Theta + W$

$\Theta$: $N(0,1)$ $\quad W$: $N(0,1)$

$X$: observation

$\Theta$: unknown that we want to estimate

$W$: Noise

Recall Bayes Rule

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)\, d\theta$$

$$f_{\Theta|X}(\theta|x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x|\theta)}{f_X(x)}$$

$$f_{\Theta|X}(\theta|x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x|\theta)}{f_X(x)}$$

mean of $W+\theta$

var of $W$

$$f_{X|\Theta}(x|\theta) = \theta + W \quad : N(\theta, 1)$$

density of $\Theta \Rightarrow$ has some constant $c$

$$f_{\Theta|X}(\theta|x) = \frac{1}{f_X(x)}\, c e^{-1/2\,\theta^2} \cdot c e^{-1/2(x-\theta)^2}$$

some function of $x$

$$= \underbrace{c(x)}\, e^{-\text{quadratic}(\theta)}$$

Fix $x \Rightarrow$ constant, Normal $\Rightarrow$ posterior is normal

$\Rightarrow E = \text{peak}$

$$\Rightarrow \hat{\Theta}_{MAP} = \hat{\Theta}_{LMS} = E[\Theta | X = x]$$

Peak:

exponent: $\frac{1}{2}\theta^2 + \frac{1}{2}(x-\theta)^2$

chain Rule

$\cancel{2(1/2)}\theta + \cancel{2(1/2)}(x-\theta)(-1)$

$\theta + (\theta - x)$

$$\theta + \theta - x = 0$$

$$\theta = x/2 = E[\Theta | X = x]$$

estimate

estimator

$$\hat{\Theta}_{MAP} = E[\Theta | X] = X/2$$

- Even with general means and variances:
  - posterior is normal
  - LMS and MAP estimators coincide
  - these estimators are "linear," of the form $\widehat{\Theta} = aX + b$

## Multiple Observations

$X_1 = \Theta + W_1$

$\vdots$

$X_n = \Theta + W_n$

Assume:

$\Theta \sim N(x_0, \sigma_0^2) \quad W_i \sim N(0, \sigma_i^2)$

$\Theta, W_1, \ldots W_n$ independent

individual case

$$f_{X_i|\Theta}(x_i|\theta) = c_i e^{-(x_i-\theta)^2/2\sigma_i^2}$$

given $\Theta = \theta \Rightarrow X_i = \theta + W_i \sim N(\theta, \sigma_i^2)$

Just like above

joint PDF of observations

$$f_{X|\Theta}(x|\theta) = f_{X_1, X_2 \ldots X_n|\Theta}(x_1, \ldots x_n | \theta)$$

$$f_{X|\Theta}(x|\theta) = f_{X_1,X_2\ldots X_n|\Theta}(x_1,\ldots x_n|\theta)$$

given $\Theta=\theta$: $W_i$ still independent $\Rightarrow$ $X_i$ independent

$$f_{X|\Theta}(x|\theta) = \prod_{i=1}^{n} f_{X_i|\Theta}(x_i|\theta)$$

can just multiply individual PDFs since

$$\Rightarrow f_{\Theta|X}(\theta|x) = \frac{1}{f_X(x)} \cdot c_0 e^{-(\theta-x_0)^2/2\sigma_0^2} \cdot \prod_{i=0}^{n} c_i e^{-(x_i-\theta)^2/2\sigma_i^2}$$

$\Rightarrow$ NORMAL

NORMAL SINCE

$$f_{\Theta|X}(\theta|x) = c \cdot \exp\{-\text{quad}(\theta)\} \qquad \text{quad}(\theta) = \frac{(\theta-x_0)^2}{2\sigma_0^2} + \frac{(\theta-x_1)^2}{2\sigma_1^2} + \cdots + \frac{(\theta-x_n)^2}{2\sigma_n^2}$$

Mean = peak = where quadratic is smallest

$$\frac{d}{d\theta}\text{quad}(\theta) = 0 \;:\; \sum_{i=0}^{n} \frac{(\theta-x_i)}{\sigma_i^2} = 0 \Rightarrow \theta \sum_{i=0}^{n} \frac{1}{\sigma_i^2} = \sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}$$

$$\boxed{\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X = x] = \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}}$$

- Key conclusions:
  - posterior is normal
  - LMS and MAP estimates coincide
  - these estimates are "linear," of the form $\hat{\theta} = a_0 + a_1 x_1 + \cdots + a_n x_n$
- Interpretations:
  - estimate $\hat{\theta}$: weighted average of $x_0$ (prior mean) and $x_i$ (observations)
  - weights determined by variances

# Mean Squared Error (MSE)

Performance Metric:

mean of posterior distribution

$$E[(\Theta-\hat{\Theta})^2 | X=x] = E[(\Theta-\hat{\theta})^2 | X=x]$$

$$= \text{var}(\Theta | X=x)$$

coefficient next to the $x^2$ term in $f_X(x)$

(From last section)

$$\alpha = \frac{1}{2\sigma_0^2} + \cdots \frac{1}{2\sigma_n^2} \Rightarrow \text{var}(\Theta|X=x) = 1 \Big/ \sum_{i=0}^{n} 1/\sigma_i^2$$

$$\alpha = \frac{1}{2\sigma_0^2} + \cdots \frac{1}{2\sigma_n^2} \Rightarrow \mathrm{var}(\theta \mid X = x) = \frac{1}{\sum_{i=0}^{n} 1/\sigma_i^2}$$

overall error (without observations)

$$E[(\theta - \hat{\theta})^2] = \int E[(\theta - \hat{\theta})^2 \mid X = x] f_X(x)\,dx$$

$$= \frac{1}{\sum_0^n 1/\sigma_i^2} \underbrace{\int f_X(x)\,dx}_{1} = \frac{1}{\sum_0^n 1/\sigma_i^2}$$

same as conditional!

one low noise measurement helps a lot

some $\sigma_i^2$ small $\rightarrow$ MSE small

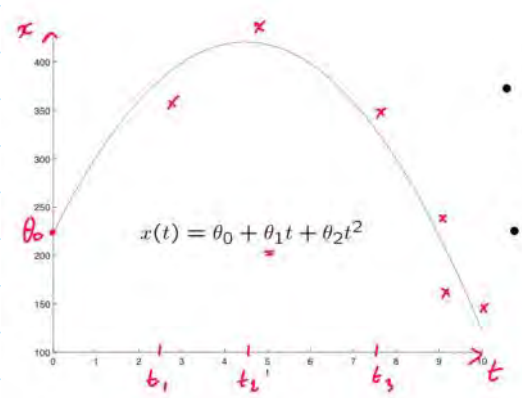
all $\sigma_i^2$ large $\rightarrow$ MSE large

Example: $\sigma_0^2 = \sigma_1^2 = \ldots = \sigma_n^2 = \sigma^2$ (same variance)

$$E[(\theta - \hat{\theta})^2 \mid X = x] = \frac{1}{(n+1)\frac{1}{\sigma^2}} = \sigma^2/n+1$$

The more observation ($n$) we improve performance (var of posterior goes down)

## Multiple Parameters: Trajectory estimation




- Random variables $\Theta_0, \Theta_1, \Theta_2$ independent; priors $f_{\Theta_j}$
- Measurements at times $t_1, \ldots, t_n$
  $X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$
  noise model: $f_{W_i}$
  independent $W_i$; independent from $\Theta_j$

**A model with normality assumptions**

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)\, d\theta$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i \qquad i = 1, \ldots, n$$

- assume $\Theta_j \sim N(0, \sigma_j^2)$, $W_i \sim N(0, \sigma^2)$; independent
- Given $\Theta = \theta = (\theta_0, \theta_1, \theta_2)$, $X_i$ is: $N(\theta_0 + \theta_1 t_i + \theta_2 t_i^2, \sigma^2)$

$$f_{X_i|\Theta}(x_i \mid \theta) = c \cdot \exp\left\{ -(x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 / 2\sigma^2 \right\}$$

- posterior: $f_{\Theta|X}(\theta \mid x) = \frac{1}{f_X(x)} \prod_{j=0}^{2} f_{\Theta_j}(\theta_j) \prod_{i=1}^{n} f_{X_i|\Theta}(x_i \mid \theta)$

$$c(x) \exp\left\{ -\frac{1}{2}\left(\frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \right\}$$

- MAP estimate: maximize over $(\theta_0, \theta_1, \theta_2)$;
  (minimize quadratic function)

$\frac{\partial}{\partial \theta_j}(\text{quad}(\theta)) = 0$ 3 equations, 3 unknowns
linear

# Linear Normal Models

$\Theta_j$ and $X_i$ are linear function of independent normals

Posterior: $f_{\Theta|X}(\theta|x) = c(x) \exp\{-\text{quadratic}(\theta_1 \ldots \theta_n)\}$

MAP: maximize over $\theta_1 \ldots \theta_n$
↳ minimize quadratic

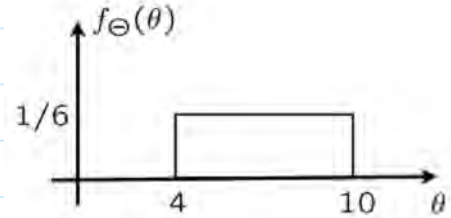
$\hat{\theta}_{MAP,j}$ is linear in $X = (X_1 \ldots X_n)$

- Facts:
  - $\widehat{\Theta}_{\text{MAP},j} = \mathbf{E}[\Theta_j \mid X]$
  - marginal posterior PDF of $\Theta_j$: $f_{\Theta_j|X}(\theta_j \mid x)$, is normal
  - MAP estimate based on the joint posterior PDF:
    same as MAP estimate based on the marginal posterior PDF
  - $\mathbf{E}\left[(\widehat{\Theta}_{i,\text{MAP}} - \Theta_i)^2 \mid X = x\right]$: same for all $x$

# L16: Least Mean Squares (LMS) Estimator

Sunday, November 13, 2016 9:06 PM

## LMS estimation without observations




Unknown: $\Theta$ Prior: $p_\Theta(\theta)$

MAP Rule: Any $\hat{\theta} \in [4, 10]$ just as good

Conditional expectation: $\hat{\theta} = 7$

Criteria: Mean Squared Error (MSE)

$$E[(\Theta - \hat{\theta})^2]$$

Find $\hat{\theta}$ that minimizes MSE

$$E[(\Theta-\hat{\theta})^2] = E[\Theta]^2 - 2E[\Theta]\hat{\theta} + \hat{\theta}^2$$

$$\frac{d}{d\hat{\theta}} = 0: \; -2E[\Theta] + 2\hat{\theta} = 0 \Rightarrow \hat{\theta} = E[\Theta]$$

OR

$$E[(\Theta-\hat{\theta})^2] = \mathrm{var}(\Theta - \hat{\theta}) + \underbrace{(E[\Theta-\hat{\theta}])^2}_{\text{minimized if } \hat{\theta} = E[\Theta]}$$

constant

$$= \mathrm{var}(\Theta)$$

so now....

$$E[(\Theta-\hat{\theta})^2] = E[(\Theta - E[\Theta])^2] = \mathrm{var}(\Theta)$$

## LMS Estimation Based on X

unknown: $\Theta$ Prior: $p_\Theta(\theta)$ Observation: $X$ observe: $X = x$

want to find $\hat{\theta}$ Model: $p_{X|\Theta}(x|\theta)$

minimize conditional mean squared error, $\mathbf{E}\left[(\Theta - \hat{\theta})^2 \mid X = x\right]$: $\hat{\theta} = \mathbf{E}[\Theta \mid X = x]$

Estimator: $E[\Theta|X]$

Since $E[\Theta]$ minimizes $E[(\Theta-\hat{\theta})^2]$

$$\Rightarrow E[(\Theta - E[\Theta])^2] \leq E[(\Theta - c)^2] \text{ for all } c$$

Similarly since $E[\Theta|X=x]$ minimizes $E[(\Theta-\hat{\theta})^2|X=x]$

$$\Rightarrow E[(\Theta - E[\Theta|X=x])^2] < E[(\Theta - a(x))^2 | X = x]$$

Similarly since $E[\Theta|X=x]$ minimizes $E[(\Theta - g(x))^2 | X=x]$

$$\Rightarrow E[(\Theta - E[\Theta|X=x])^2] \le E[(\Theta - g(x))^2 | X=x]$$

for all $x$

$$\Rightarrow E[(\Theta - E[\Theta|X])^2 | X] \le E[(\Theta - g(X))^2 | X]$$

rewrite

$$E[(\Theta - E[\Theta|X])^2] \le E[(\Theta - g(X))^2]$$

take expectations of both sides

$\hat{\Theta}_{LMS} = \mathbf{E}[\Theta \mid X]$ minimizes $\mathbf{E}\left[(\Theta - g(X))^2\right]$, over all estimators $\hat{\Theta} = g(X)$

## LMS Performance Evaluation

LMS Estimate:

$$\hat{\theta} = E[\Theta | X=x]$$

LMS Estimator:

$$\hat{\Theta} = E[\Theta | X]$$

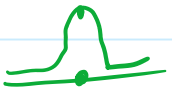
Expected Performance:

* $$MSE = E[(\Theta - E[\Theta|X=x])^2] = var(\Theta | X=x)$$

without a measurement...

* $$MSE = E[(\Theta - E[\Theta|X])^2] = E[var(\Theta|X)]$$

- LMS relevant to estimation (not hypothesis testing)



- Same as MAP if posterior has one peak symmetric around mean
  - when posterior is normal ("linear normal" models)

### Example:

EXERCISE: LMS ESTIMATION ERROR (3/3 points)

As in the previous exercise, let $\Theta$ be the bias of a coin, i.e., the probability of Heads at each toss. We assume that $\Theta$ is uniformly distributed on $[0, 1]$. Let $K$ be the number of Heads in $9$ independent tosses. We have seen that the LMS estimate of $K$ is $\mathbf{E}[K \mid \Theta = \theta] = n\theta$.

a) Find the conditional mean squared error $\mathbf{E}\left[(K - \mathbf{E}[K \mid \Theta = \theta])^2 \mid \Theta = \theta\right]$ if $\theta = 1/3$.

9*1/3*(2/3) ✔ Answer: 2

$= np(1-p) = var(K|\Theta=\theta)$ ← since binomial

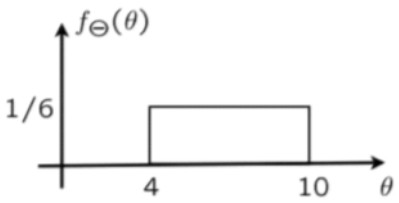
b) Find the overall mean squared error of this estimation procedure.

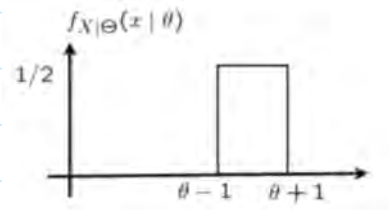
1.5 ✔ Answer: 1.5

$= E[var(K|\Theta)]$ ← all possible vars

$= \int_0^1 f_\Theta(\theta)\, var(K|\Theta=\theta)\, d\theta = 1.5$

# Example: LMS Estimate




$X = \Theta + U \quad U \sim \text{unif}[-1, 1]$

① Find posterior

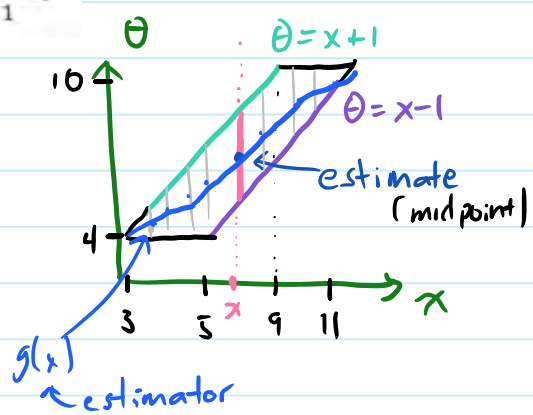
1.5 Find joint

We know: $\theta - 1 \le x \le \theta + 1$

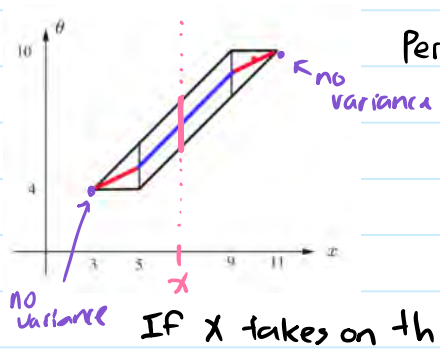
$\theta \le x + 1 \quad \theta \ge x - 1$

uniform

Joint = 12



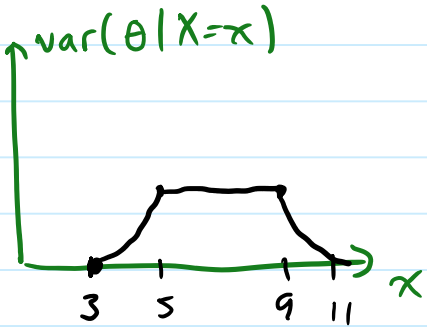

## Performance Evaluation




$$\text{Performance} = E\left[\left(\theta - E[\theta \mid X = x]\right)^2 \mid X = x\right]$$
$$= \text{var}(\theta \mid X = x)$$

If $X$ takes on this value, variance is uniform over | interval

Recall: $U[a, b]$ var $= \frac{(b-a)^2}{12} = 1/3$




can get since we have joint

$$E[\text{var}(\theta \mid X)] = \int f_X(x)\, \text{var}(\theta \mid X = x)\, dx$$

# LMS Estimation with multiple observations or unknowns

## LMS Estimation with multiple observations or unknowns

Observations: $X = (X_1, X_2 \dots X_n)$ Model: $P_{X|\Theta}(x|\theta)$

LMS Estimate: $E[\Theta \mid x_1 \dots x_n, X_n = x_n]$

If $\Theta$ is a vector apply to each component separately
$\Theta = (\theta_1 \dots \theta_n)$

$$\hat{\theta}_j = E[\theta_j \mid x_1 \dots x_n, X_n = x_n]$$

### Properties of the estimation error in LMS estimation

- Estimator: $\widehat{\Theta} = \mathbf{E}[\Theta \mid X]$
- Error: $\underline{\widetilde{\Theta} = \widehat{\Theta} - \Theta}$

$E[\hat{\Theta}] = E[\Theta]$

$E[\tilde{\Theta}] = 0$

$\boxed{\mathbf{E}[\widetilde{\Theta} \mid X = x] = 0}$

$E[\hat{\Theta} - \Theta \mid X = x] = \hat{\Theta} - E[\Theta \mid X = x] = 0$

$\boxed{\text{cov}(\widetilde{\Theta}, \widehat{\Theta}) = 0}$

$E[\tilde{\Theta}\hat{\Theta}] - \cancel{E[\tilde{\Theta}]}E[\hat{\Theta}]$ 0

$E[\tilde{\Theta}\hat{\Theta} \mid X = x] = \hat{\Theta} E[\tilde{\Theta} \mid X = x] = 0$

$\boxed{\text{var}(\Theta) = \text{var}(\widehat{\Theta}) + \text{var}(\widetilde{\Theta})}$

$\Theta = \hat{\Theta} - \tilde{\Theta}$

# L17: LLMS: Linear Least Mean Squares

Monday, November 14, 2016 9:10 AM

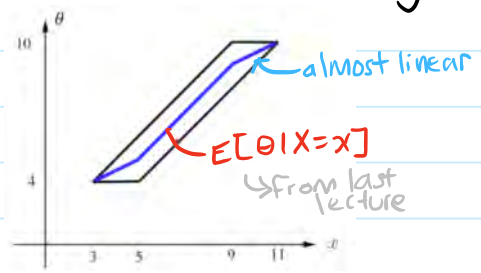
Sometimes conditional expectations are hard to compute
$\Rightarrow$ use a simpler linear estimator

## LLMS

Unknown: $\Theta$ Observation: $X$

want to minimize $E[(\hat{\Theta}-\Theta)^2]$

Estimator $\hat{\Theta} = g(X) \rightarrow \hat{\Theta}_{LMS} = E[\Theta|X]$




Only consider estimators that depend linearly on $X$

$$\hat{\Theta} = aX + b$$

$\Rightarrow$ Minimize $E[(\Theta - aX - b)^2]$

- only need to find $a$ and $b$

IF $E[\Theta|X]$ is linear in $X \Rightarrow \hat{\Theta}_{LMS} = \hat{\Theta}_{LLMS}$

## LLMS Solution

Find $a, b$ to minimize $E[(\Theta - aX - b)^2]$

- If we know $a \Rightarrow b = E[\Theta] - aE[X]$

$\Rightarrow$ minimize $E[(\Theta - aX - E[\Theta - aX])^2] = var(\Theta - aX)$ dependent

$= var(\Theta) + a^2 var(X) - 2a\,cov(\Theta, X)$

$\frac{d}{da} = 0: \; 2a\,var(X) - 2\,cov(\Theta, X) = 0$

$\Rightarrow a = cov(\Theta, X) / var(X)$

$$\boxed{\hat{\Theta}_L = E[\Theta] + \frac{Cov(\Theta, X)}{var(X)}(X - E[X]) = E[\Theta] + \rho\frac{\sigma_\Theta}{\sigma_X}(X - E[X])}$$

- only means, variances and covariances matter

- only means, variances and covariances matter
- $\rho > 0$: $X > E[X] \Rightarrow \hat{\Theta}_L > E[\Theta]$ $\quad X < E[X] \Rightarrow \Theta_L < E[X]$
- $\rho < 0$: $X > E[X] \Rightarrow \hat{\Theta}_L < E[\Theta]$ $\quad X < E[X] \Rightarrow \Theta_L > E[X]$
- $\rho = 0$: $\hat{\Theta}_L = E[\Theta]$

$$\boxed{\mathbf{E}\left[(\widehat{\Theta}_L - \Theta)^2\right] = \underbrace{(1-\rho^2)\,\mathrm{var}(\Theta)}_{\text{uncertainty gets reduced by } (1-\rho^2)}}$$

EXERCISE: LLMS WITHOUT A CONSTANT TERM (2 points possible)

Suppose that instead of estimators of the form $aX + e$, we consider estimators of the form $\widehat{\Theta} = aX$ and ask for the value of $a$ that minimizes the mean squared error. Mimic the derivation you have just seen and find the optimal value of $a$. Your answer should be an algebraic expression involving some of the constants $b$, $c$, $d$, where $b = \mathbf{E}[\Theta^2]$, $c = \mathbf{E}[\Theta X]$, $d = \mathbf{E}[X^2]$.

b + d^2 ✖ Answer: c/d

$b + d^2$

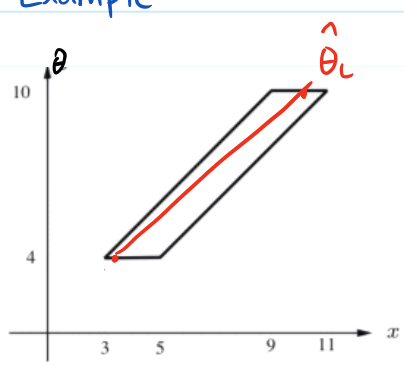
Answer:
The mean squared error is

$$\mathbf{E}\left[(\Theta - aX)^2\right] = \mathbf{E}[\Theta^2] - 2a\mathbf{E}[\Theta X] + a^2\mathbf{E}[X^2].$$

By setting to zero the derivative with respect to $a$, we find that

$$a = \frac{\mathbf{E}[\Theta X]}{\mathbf{E}[X^2]} = \frac{c}{d}.$$

Example:




$\Theta$: uniform $[4, 10]$

$X = \Theta + U$ $\quad$ uniform$[-1, 1]$

Independent

$E[\Theta] = \text{midpoint}(4, 10) = 7$

$E[U] = 0 \qquad E[X] = 7$

$\mathrm{var}(\Theta) = \frac{6^2}{12} = 3 \qquad \mathrm{var}(U) = 2^2/12 = 1/3$

$\mathrm{var}(X) = \mathrm{var}(\Theta) + \mathrm{var}(U) = 10/3$

$\mathrm{cov}(\Theta, X) = \mathrm{cov}(\Theta, \Theta + U)$

$= \mathrm{cov}(\Theta, \Theta) + \cancel{\mathrm{cov}(\Theta, U)}^{\,0} = 3$

Positive since $X, \Theta$ positively correlated

$= \mathrm{cov}(\Theta, \Theta) + \cancel{\mathrm{cov}(\Theta, U)} = 3$

Positive since $X, \Theta$ positively correlated

$\Rightarrow \hat{\Theta}_L = 7 + 9/10\,(x - 7)$

EXERCISE: LLMS DRILL (2 points possible)

Suppose that $\Theta$ and $W$ are independent, both with variance 1, and that $X = \Theta + W$. Furthermore, $\mathbf{E}[\Theta] = 1$ and $\mathbf{E}[W] = 2$. The LLMS estimator $\widehat{\Theta} = aX + b$ has

$a =$ 1 ✖ Answer: 0.5

$b =$ 1 ✖ Answer: -0.5

*Hint:* Remember the formula $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$.

Answer:
We have $\mathbf{E}[X] = \mathbf{E}[\Theta] + \mathbf{E}[W] = 3$ and $\mathrm{var}(X) = \mathrm{var}(\Theta) + \mathrm{var}(W) = 2$. Also,

$$\mathrm{cov}(X, \Theta) = \mathrm{cov}(\Theta, \Theta) + \mathrm{cov}(\Theta, W) = \mathrm{var}(\Theta) + 0 = 1.$$

Therefore, the LLMS estimator is

$$\widehat{\Theta} = 1 + \frac{1}{2}(X - 3) = \frac{1}{2}X - \frac{1}{2}.$$

## Coin Bias Example

coin with unknown bias $\Theta$ prior $f_\Theta(\cdot)$

Flip $n$ times. $X = \#$ heads

Assume prior $f_\Theta(\theta)$ uniform $[0,1]$

$$\hat{\Theta}_{LMS} = \frac{X+1}{n+2}$$ From before

$= \hat{\Theta}_{LLMS}$ since its linear

Eventhough we know the answer, lets find $\hat{\Theta}_{LLMS}$

① $E[\Theta] = 1/2$ $\quad \mathrm{var}(\Theta) = 1/12$ $\quad P_{X|\Theta} : \mathrm{Bin}(n, \Theta)$

$E[X|\Theta] = n\Theta$ $\quad \mathrm{var}(X|\Theta) = n\,\Theta(1-\Theta)$

② $E[X] = E\big[E[X|\Theta]\big] = E[n\Theta] = n/2$

③ $\text{var}(X) = E[X^2] - (E[X])^2$

$E[X^2] = E[E[X^2 | \theta]]$ Iterated Expectations

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

$= n\theta(1-\theta) + n^2\theta^2$

$E[X^2] = E[n\theta(1-\theta) + n^2\theta^2]$

$= E[n\theta + (n^2 - n)\theta^2]$

$E[\theta^2] = \text{var}(\theta) + (E[\theta])^2 = 1/3$

$E[X^2] = E[n\theta] + E[(n^2-n)/\theta^2]$

$= n/2 + \frac{n^2-n}{3} = \frac{n}{6} + \frac{n^2}{3}$

$\text{var}(X) = \frac{n}{6} + \frac{n^2}{3} - \frac{n^2}{4} = n/6 + n^2/12 = \frac{n(n+2)}{12}$

④ $\text{cov}(\theta, X) = E[\theta X] - E[\theta]E[X]$

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

$E[\theta X | \theta] = \theta E[X | \theta] = n\theta^2$

$E[\theta X] = E[n\theta^2] = n/3$

$\text{cov}(\theta, X) = n/3 - n/4 = n/12$

$\hat{\theta}_{LMS} = E[\theta] + \frac{\text{cov}(\theta, X)}{\text{var}(X)}(X - E[X]) = \frac{X+1}{n+1}$ ✓

## LLMS with multiple observations

Unknown $\theta$. observations $X = (X_1, X_2, \ldots X_n)$

Find $a_1 \ldots a_n, b$ to minimize $E[(a_1X_1 + \ldots + a_nX_n + b - \theta)^2]$

If $E[\theta | X]$ linear in $X \Rightarrow \hat{\theta}_{LMS} = \hat{\theta}_{LLMS}$

- Solve linear system in b and the $a_i$ (with computer)
- Only means, variances and covariances matter
- If multiple unknown $\theta_j$, apply to each one separately

Example:

$X_1 = \theta + W_1$ $\quad \theta \sim x_0, \sigma_0^2$ $\quad W_i \sim 0, \sigma_i^2$

$X_n = \theta + W_n$ $\quad \theta, W_1 \ldots W_n$ uncorrelated

$X_1 = \Theta + W_1$ $\quad \Theta \sim x_0, \sigma_0^2 \quad W_i \sim 0, \sigma_i^2$

$X_n = \Theta + W_n$ $\quad \Theta, W_1, \ldots, W_n$ uncorrelated

Trick:

- Suppose $\Theta, W_1, \ldots, W_n$ are independent normal

Recall:

$$\hat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X = x] = \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} \qquad \widehat{\Theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X] = \frac{\frac{x_0}{\sigma_0^2} + \sum_{i=1}^{n} \frac{X_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} = \hat{\Theta}_{\text{LLMS}}$$

- Suppose general (not normal) distributions, but same means, variances, as in normal example
  - all covariances also the same
  - solution must be the same

EXERCISE: LLMS WITH MULTIPLE OBSERVATIONS (1/3 points)

Suppose that $\Theta$, $X_1$, and $X_2$ have zero means. Furthermore,

$$\text{var}(X_1) = \text{var}(X_2) = \text{var}(\Theta) = 4,$$

and

$$\text{cov}(\Theta, X_1) = \text{cov}(\Theta, X_2) = \text{cov}(X_1, X_2) = 1.$$

The LLMS estimator of $\Theta$ based on $X_1$ and $X_2$ is of the form $\widehat{\Theta} = a_1 X_1 + a_2 X_2 + b$. Find the coefficients $a_1$, $a_2$, and $b$. *Hint:* To find $b$, recall the argument we used for the case of a single observation.

$a_1 =$ (2/4) ✖ **Answer:** 0.2

$a_2 =$ (2/4) ✖ **Answer:** 0.2

$b =$ 0 ✔ **Answer:** 0

Answer:
By the same argument as in the case of a single observation, we will have $b = \mathbf{E}[\Theta - a_1 X_1 - a_2 X_2] = 0$. Using the variance and covariance information we are given, the expression we want to minimize is

$$\mathbf{E}\left[(a_1 X_1 + a_2 X_2 - \Theta)^2\right] = 4a_1^2 + 4a_2^2 + 4 + 2a_1 a_2 - 2a_1 - 2a_2.$$

Because of symmetry, we see that the optimal solution will satisfy $a_1 = a_2 = a$, so the expression is of the form $8a^2 + 4 + 2a^2 - 4a$. By setting the derivative to zero, we find that $20a = 4$, or $a = 1/5$.

## The representation of the data matters in LLMS

- Estimation based on $X$ versus $X^3$
  - LMS: $\mathbf{E}[\Theta \mid X]$ is the same as $\mathbf{E}[\Theta \mid X^3]$
  - LLMS is different: estimator $\widehat{\Theta} = aX + b$ versus $\widehat{\Theta} = aX^3 + b$

    $\text{cov}(\Theta, X^3)$ $\text{var}(X^3)$

  - can also consider $\widehat{\Theta} = a_1 X + a_2 X^2 + a_3 X^3 + b$
  - can also consider $\widehat{\Theta} = a_1 X + a_2 e^X + a_3 \log X + b$

# L19: Central Limit Theorem (CLT)

Monday, November 21, 2016 11:02 AM

← iid RV

$S_n = X_1 + X_2 + \ldots X_n$ variance: $n\sigma^2$

$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ $E[Z_n] = 0$ $var(Z_n) = 1$

## Central Limit theorem:

$$\lim_{n\to\infty} P(Z_n \le z) = P(Z \le z) = \Phi(z)$$

CDF converges to normal CDF

- Only means and variances matter
- Fairly accurate and quick
- Results for convergance of PDF or PMs (with more assumptions)

EXERCISE: CLT (2/2 points)

Let $X_n$ be i.i.d. random variables with mean zero and variance $\sigma^2$. Let $S_n = X_1 + \cdots + X_n$. Let $\Phi$ stand for the standard normal CDF. According to the central limit theorem, and as $n \to \infty$, $\mathbf{P}(S_n \le 2\sigma\sqrt{n})$ converges to $\Phi(a)$, where:

$a =$ 2 ✔ Answer: 2

Furthermore,

$\mathbf{P}(S_n \le 0)$ converges to: .5 ✔ Answer: 0.5

(Here, enter the numerical value of the probability.)

Answer:
We have

$$\lim_{n\to\infty} \mathbf{P}(S_n \le 2\sigma\sqrt{n}) = \lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - 0}{\sigma\sqrt{n}} \le 2\right) = \Phi(2).$$

Similarly,

$$\lim_{n\to\infty} \mathbf{P}(S_n \le 0) = \lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - 0}{\sigma\sqrt{n}} \le 0\right) = \Phi(0) = \frac{1}{2}.$$

- So we can treat $Z_n$ as normal

$\Rightarrow S_n$ is linear combination of $Z_n$

$$S_n = \sqrt{n}\,\sigma Z_n + n\mu$$

$\rightarrow$ can treat $S_n$ as normal: $N(n\mu, n\sigma^2)$

## Example

$P(S_n \le a) \le b$ · given 2 parameters, find third

- Package weight $X_i$ iid exponential $\lambda = 1/2$
- Load container with $n = 100$ packages

Given $a, n$
Find $b$

$P(S_n \ge 210)$ max capacity of container

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} \qquad \mu = \sigma = 2$$

$$\Rightarrow P\left(\frac{S_n - 200}{20} \ge \frac{210 - 200}{20}\right)$$

$$= P(Z_n \ge .5) \approx P(Z \ge .5)$$

$$= 1 - P(Z < 0.5) = 1 - \Phi(.5) = 1 - .6195$$

$$= .3085$$

- Now choose the capacity $a$ so that $P(S_n \ge a) \approx .05$
- Find $a$

Given $b, n$
Find $a$

$$.05 = P\left(\frac{S_n - 200}{20} \ge \frac{a - 200}{20}\right)$$

$$= 1 - \Phi\left(\frac{a - 200}{20}\right)$$

$$\Phi\left(\frac{a-200}{20}\right) = .95$$

$$\frac{a - 200}{20} \approx 1.65 \Rightarrow a = 232.9$$

- How large can $n$ be so that

Given $a, b$
Find $n$

$$P(S_n \ge 210) \approx .05$$

$$P\left(\frac{S_n - 2n}{2\sqrt{n}} \ge \frac{210 - 2n}{2\sqrt{n}}\right)$$

$$= 1 - \Phi\left(\frac{210 - 2n}{2\sqrt{n}}\right) \approx .05$$

$$= 1 - \Phi\left(\frac{210-2n}{2\sqrt{n}}\right) \approx .05$$

$$\Rightarrow \Phi\left(\frac{210-2n}{2\sqrt{n}}\right) = .95$$

$$\Rightarrow \frac{210-2n}{2\sqrt{n}} = 1.645$$

$$\Rightarrow n = 89$$

- Load container until weigh exceeds 210

N: # packages loaded

$P(N > 100)$

not the sum of independent RVs

express event in terms of $X_i$

$$= P\left(\sum_{i=0}^{100} X_i \leq 210\right)$$

$$\approx \Phi\left(\frac{210-200}{20}\right) = \Phi(0.5) = 0.6915$$

## Normal approximation to the binomial

$X_i$: independent, Bernoulli(p)

$S_n = X_1 + \ldots X_n$ Binomial(n,p)

CDF of $\frac{S_n - np}{\sqrt{np(1-p)}} \sim$ standard normal

- $n = 36$ $p = .5$ Find $P(S_n \leq 21)$

$$\sum_{i=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = .8785$$

OR use CLT

$np = 18$ $\sqrt{np(1-p)} = 3$

$$P\left(\frac{S_n - 18}{3} \leq \frac{21-18}{3}\right) \approx \Phi(1) = .8413$$

Get Better approx

$.84 = P(S_n \leq 21) = P(S_n \leq 22) = .9082$

underestimate over estimate

$P(S_n \leq 21.5) = .8790$ ✓

← underestimate  overestimate →

$P(S_n \le 21.5) = .8790$ ✓

← remember $S_n$ is discrete

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

$$= P\left(\frac{18.5-18}{3} \le Z_n \le \frac{19.5-18}{3}\right) = P(.17 \le Z_n \le .5)$$

$$\approx \Phi(.5) - \Phi(.17) = 0.124$$

Exact answer: $\binom{30}{19}\left(\frac{1}{2}\right)^{30} = .1251$

# Polling

- $p$: fraction of population that will vote "yes" in a referendum
- $i$th (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$  $E[X_i] = p = \mu$, $\sigma = \sqrt{p(1-p)}$
- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$

$\mathbf{P}(|M_n - p| \ge .01)$

$$\left|\frac{S_n - np}{n}\right| \ge .01 \qquad \left|\overbrace{\frac{S_n - np}{\sqrt{n}\,\sigma}}^{Z_n}\right| \ge \frac{.01\sqrt{n}}{\sigma}$$

$$= P\left(|Z_n| \ge \frac{.01\sqrt{n}}{\sigma}\right)$$

$$\approx P\left(|Z| \ge \frac{.01\sqrt{n}}{\sigma}\right)$$

standard normal

$\sigma = \sqrt{p(1-p)} \le 1/2$

$$\le P(|Z| \ge .2\sqrt{n})$$

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

Try $n = 10{,}000$

$$\text{prob} \le 2(1 - \Phi(2)) = 0.046$$

# L21: The Bernoulli Process

Monday, November 21, 2016 3:32 PM

## The Bernoulli Process

- Sequence of independent Bernoulli trials $X_i$

  $P(X_i = 1) = p \qquad P(X_i = 0) = (1-p)$

Key assumptions

- Each trial independent
- Time homogeneity

Model of:

- Sequence of lottery wins/losses
- Arrivals (at each timeslots) to server

## The Stochastic Process

- Sequence of Random Variables

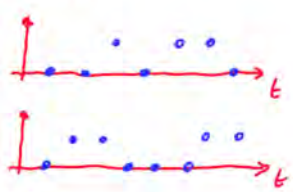
- First view: sequence of random variables $X_1, X_2, \ldots$ (infinite)

  Interested in: $\mathbf{E}[X_i] = p \qquad \text{var}(X_i) = p(1-p) \qquad p_{X_i}(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$

  $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$ for all $n$

- Second view – sample space:

  $\Omega =$ set of infinite sequences of 0's and 1's




- Example (for Bernoulli process):

  $\mathbf{P}(X_i = 1 \text{ for all } i) =$ $(p < 1)$

  $\leq P(X_1 = 1, \ldots, X_n = 1) = p^n$, for all $n$

## Bernoulli Review

$S = X_1 + X_2 + \ldots X_n \qquad E[S] = np$

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

$T_1 = \min\{i : X_1 = 1\}$

$P(T_1 = k) = P(\underbrace{0\,0\,0\,0\,0}_{k-1}\,1) = \underbrace{(1-p)^{k-1}}\,p$ Geometric Distribution

## The Fresh start properties

since each trial

# The Fresh start properties

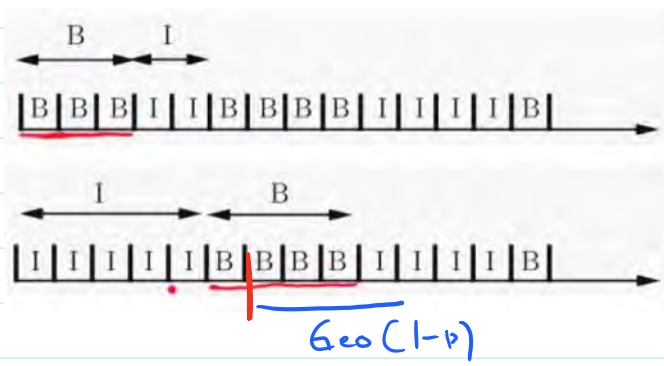
- Fresh start after time $N$ ← since each trial is independent
  - $N$ can be = time of 3rd success
    - = time of 3 successes in a row } Causally determined
    - .... etc
  - The Process $X_{N+1}, X_{N+2}$ is:
    - A bernoulli process
    - independent of $N, X_1 \ldots X_N$

$N$ = the time just before first occurance !!! } Not causally determined
- No Fresh start
- Will not observe a Bernoulli Process

## EXAMPLE

Bernoulli Process

- At each slot, a server is busy or idle
- First busy period
  - starts w/ first busy slot
  - ends just before the first idle slot

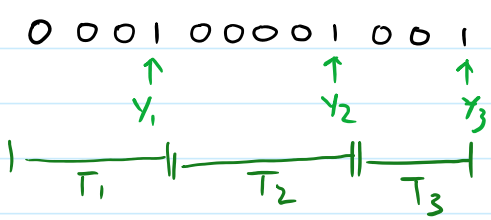



idle = success

=

⇒ so first busy period is geometric(1-p)

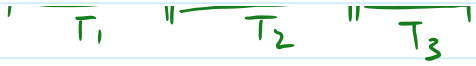
## Time of $k^{th}$ success/arrival




- $Y_k$ = time of $k^{th}$ arrival
- $T_k$ = $k^{th}$ inter-arrival time

$$T_k = Y_k - Y_{k-1} \quad (k \geq 2)$$

- Process starts fresh after $T_1$



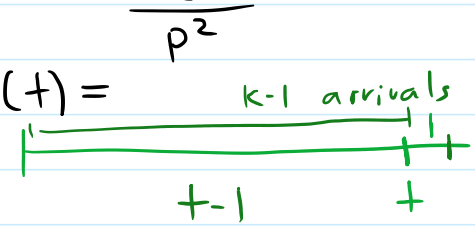

- Process starts fresh after $T_1$
  $\Rightarrow T_2$ indep $T_1$
  Geometric $(p)$

- $Y_k = T_1 + T_2 + \dots + T_k$

$$E[Y_k] = \frac{k}{p}$$

$$\mathrm{var}(Y_k) = \mathrm{var}(T_1) + \mathrm{var}(T_2) + \dots$$
$$= \frac{k(1-p)}{p^2}$$

$P_{Y_k}(t) =$

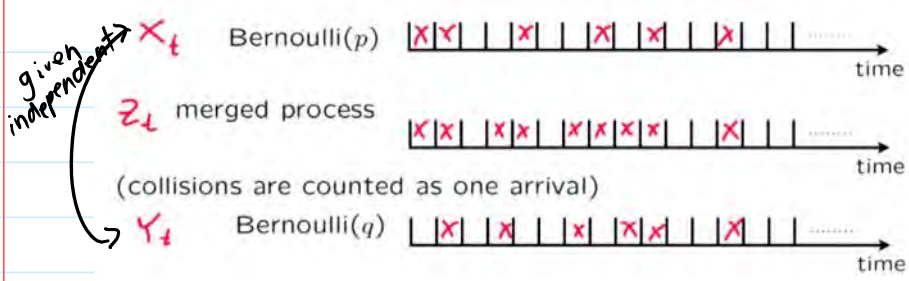



$= P(k-1 \text{ arrivals in } t-1 \text{ slots}) \cdot P(\text{arrival at time } t)$

$\leftarrow$ independent $\rightarrow$ $= p$

$$= \binom{t-1}{k-1} p^{k-1} (1-p)^{t-k} p = \begin{cases} \binom{t-1}{k-1} p^k (1-p)^{t-k} & t = k, k+1 \dots \\ 0 & \text{else} \end{cases}$$

## Merging of independent Bernoulli Processes




$Z_t = g(X_t, Y_t)$
$Z_{t+1} = g(X_{t+1}, Y_{t+1})$ $\Rightarrow Z_1 \dots Z_t$ independent

| Y \ X | 0 | 1 |
|---|---|---|
| 1 | $(1-p)q$ | $pq$ |
| 0 | $(1-p)(1-q)$ | $p(1-q)$ |

$= 1$ in merged process $= 1-(1-p)(1-q)$

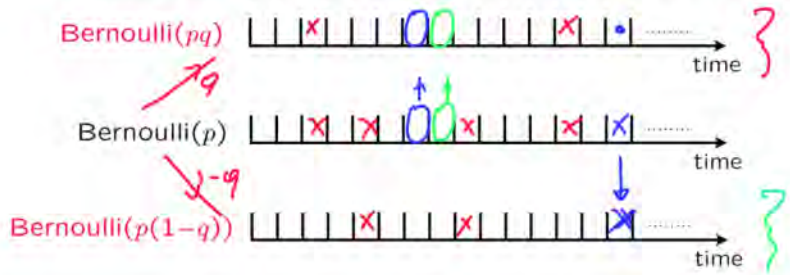
$\Rightarrow$ merged process also Bernoulli$(p+q-pq)$

$$P(\text{arrival in First process} \mid \text{arrival}) = \frac{p}{p+q-pq}$$

## Splitting a Bernoulli Process

# Splitting a Bernoulli Process

- Split successes into two streams, using independent flips of a coin with bias $q$
  - assume that coin flips are independent from the original Bernoulli process




- Are the two resulting streams independent? No

EXERCISE: SPLITTING (1 point possible)

For each exam, Ariadne studies with probability $1/2$ and does not study with probability $1/2$, independently of any other exams. On any exam for which she has not studied, she still has a $0.20$ probability of passing, independently of whatever happens on other exams. What is the expected number of total exams taken until she has had 3 exams for which she did not study but which she still passed?

✖ Answer: 30

Answer:
The sequence of exams for which she does not study and passes can be modeled as follows. We look at the exams for which she has not studied (a Bernoulli process with parameter $1/2$) and "split" it according to whether she passes or not. This creates a new Bernoulli process for the exams for which she does not study and passes, with parameter $(1/2) \cdot 0.20 = 0.10$. The expected time until 3 successes in this process is $3/0.10 = 30$.

# Poisson Approximation to Binomial

- Interesting regime: large $n$, small $p$, moderate $\lambda = np$ $\quad n \to \infty \quad p \to 0 \quad p = \frac{\lambda}{n}$

- Number of arrivals $S$ in $n$ slots: $p_S(k) = \frac{n!}{(n-k)!\,k!} \cdot p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$

For fixed $k = 0, 1, \ldots,$
$p_S(k) \to \frac{\lambda^k}{k!} e^{-\lambda},$

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

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

$$\xrightarrow[n \to \infty]{} 1 \cdot 1 \cdots 1 \cdot \frac{\lambda^k}{k!} e^{-\lambda} \cdot 1$$

- Fact: $\lim_{n \to \infty} (1 - \lambda/n)^n = e^{-\lambda}$

# L22: The Poisson Process

Monday, November 28, 2016 8:50 PM

## Poisson Process

continuous version of Bernoulli Process

### Assumptions

- # arrivals in disjoint time intervals are independent
- $P(k,T)$ = Prob $k$ arrivals in any interval of duration $T$
  - ↳ location does not matter

- **Small interval probabilities:**
  For VERY small $\delta$:

$$P(k,\delta) \approx \begin{cases} 1-\lambda\delta & \text{if } k=0 \\ \lambda\delta & \text{if } k=1 \\ 0 & \text{if } k>1 \end{cases} \qquad P(k,\delta) = \begin{cases} 1-\lambda\delta+O(\delta^2) & \text{if } k=0 \\ \lambda\delta+O(\delta^2) & \text{if } k=1 \\ 0+O(\delta^2) & \text{if } k>1 \end{cases} \qquad \frac{O(\delta^2)}{\delta} \xrightarrow[\delta\to 0]{} 0$$

$\lambda$ = arrival rate

**Applications**: Financial markets, placement of calls, requests, etc ...

## Poisson PMF

$P(k \text{ arrivals in Poisson}) \approx P(k \text{ slots have arrival})$

since $P(\text{some slot contains} \geq 2 \text{ arrivals}) = \frac{T}{\delta} O(\delta)^2 \xrightarrow[\delta\to 0]{} 0$

$N_T$: # arrivals in $[0,T]$

= # slots that have an arrival

$\Rightarrow$ Binomial $p = \lambda\delta + O(\delta^2)$

$np \approx \lambda T$

From last lecture

**Bernoulli**

$$p_S(k) = \frac{n!}{(n-k)!\,k!} \cdot p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$$

$\lambda = np \quad n \to \infty \quad p \to 0$

For fixed $k = 0, 1, \ldots,$

$$p_S(k) \to \frac{\lambda^k}{k!} e^{-\lambda},$$

$$\Rightarrow P(k,T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$$

**Mean:** $E[N_T] = \lambda T$ (arrival rate, length of time interval)

↳ since $N_T \approx \text{Binomial}(n, p)$ $\quad n = T/\delta \quad p = \lambda\delta + O(\delta^2)$

$np = \lambda T$ as $\delta \to 0$

**Variance:** $\text{var}(N_T) = \lambda T$

## Example

Get email according to Poisson Proccess $\lambda = 5$ messages/hour

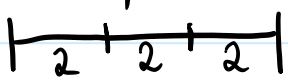
- Mean and variance of mail in one day:

$$E[N_T] = \lambda T = var(N_T) = 5 \cdot 24$$

- P( one new message in next hour)

$$P(k,T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \qquad P(1,1) = 5e^{-5}$$

- P( exactly 2 messages in each of next three hours)




$$(P(2,1))^3 = \left(\frac{5^2 e^{-5}}{2}\right)^3$$

EXERCISE: POISSON PRACTICE (1/1 point)

Consider a Poisson arrival process with rate $\lambda$ per hour. To simplify notation, we let $a = P(0,1)$, $b = P(1,1)$, and $c = P(2,1)$, where $P(k,1)$ is the probability of exactly $k$ arrivals over an hour-long time interval.

What is the probability that we will have "at most one arrival between 10:00 and 11:00 and exactly two arrivals between 10:00 and 12:00"? Your answer should be an algebraic function of $a$, $b$, and $c$ in standard notation .

(a*c) + (b*b) ✔ **Answer:** a*c+b^2

$(a \cdot c) + (b \cdot b)$

Answer:
The event of interest can happen in two ways:
(i) Zero arrivals during the first hour and two arrivals over the second hour; this has probability $ac$.
(ii) One arrival during each one of the two hours; this has probability $b^2$.
Thus, the answer is $ac + b^2$. (Note that for both scenarios, we have used independence to find the associated probabilities.)

## Time until $k^{th}$ arrival

$k=1$ case:

$T_1$. Find CDF: $P(T_1 \le t) = 1 - P(T_1 > t)$

$$= 1 - \underbrace{P(0,t)}_{\text{PMF}} = 1 - e^{-\lambda t}$$

Derivative:

$$= \underbrace{\lambda e^{-\lambda t}}_{\text{Exponential}(\lambda)} \quad \text{for } t > 0$$

Memorylessness

$$P(k,\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, \qquad k = 0, 1, \ldots$$

## Time $Y_k$ of $k_{TH}$ arrival

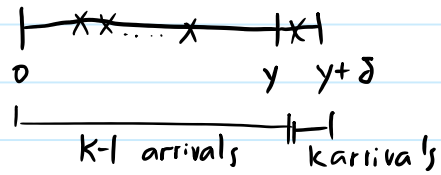
- Find CDF $P(Y_k \le y) = \sum_{n=k}^{\infty} P(n,y)$

Time $Y_k$ of $k$th arrival

- Find CDF $P(Y_k \le y) = \sum_{n=k}^{\infty} P(n, y)$ ..... derivative .....

OR

Recall $f_{Y_k}(y)\delta \approx P(y \le Y_k \le y+\delta)$




$$P(k-1, y)\,P(1,\delta) \approx P(k-1, y)\,\lambda\delta$$

$$+$$

$$P(k-2, y)\,O(\delta^2) \to 0$$

↳ 2 arrivals on $[y, y+\delta]$

$+ \ldots 3 \ldots 4 \ldots \to 0$

$$\Rightarrow \cancel{\delta} f_{Y_k}(y) = P(k-1, y)\,\lambda \cancel{\delta}$$

$$= \boxed{\textbf{Erlang distribution:} \quad f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}, \quad y \ge 0}$$

$E[Y_k] = k/\lambda$

$var(Y_k) = k/\lambda^2$

## Memorylessness and Freshstart

- If we start watching at time $t$
  - $\Rightarrow$ Poisson Process starts fresh at time $t$
  - time until next arrival : $Exp(\lambda)$ indep of past
- If we start watching at time $T_1$ ← first arrival
  - $\Rightarrow$ we see a Poisson Process independent of history
  - time until next arrival : $T_2 = Y_2 - Y_1$ : $Exp(\lambda)$ independent of $T_1$
- same for $T_k$
  - $Y_k = T_1 + T_2 + \ldots + T_k$ sum of iid exponentials

    $E[Y_k] = k/\lambda$ $\quad var(Y_k) = k/\lambda^2$

## Summary

# Summary

| | POISSON | BERNOULLI |
|---|---|---|
| Times of Arrival | Continuous | Discrete |
| Arrival Rate | $\lambda$/unit time | $p$/per trial |
| PMF of # of Arrivals | Poisson | Binomial |
| Interarrival Time Distr. | Exponential | Geometric |
| Time to $k$-th arrival | Erlang | Pascal |

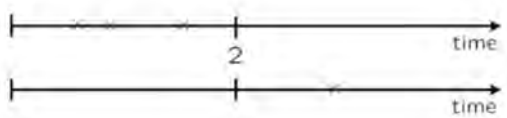
$$P(k,\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$$

$$\mathbf{E}[N_\tau] = \lambda\tau$$

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$$

# Example: Poisson Fishing

- Fish are caught as a Poisson process, $\lambda = 0.6$/hour
  - fish for two hours;
  - if you caught at least one fish, stop
  - else continue until first fish is caught




- P( fish for more than 2 hours ) : $P(0,2)$

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

- P( Fish more than 2 and less than 5 ) : $P(0,2)\,(1 - P(0,3))$ (independent)

  $\Rightarrow$ no fish in first 2 hours and (at least) one b/w [2, 5] (length = 3)

  OR

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

- P( catch at least 2 fish ):

  $$\sum_{k=2}^{\infty} P(k,2) = 1 - P(0,2) - P(1,2)$$

  OR

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

- E[ Future Fishing time | already fished for 3 hours ]

  : Exponential $(\lambda)$ $= 1/\lambda$

- E[ total Fishing time ] $= E[F] = 2 + E[F-2]$

- $E[\text{total Fishing time}] = E[F] = 2 + E[F-2]$

$$= 2 + P(F=2)\cdot 0 + P(F>2)\, E[F-2 \mid F>2]$$

$$= 2 + P(0,2)\cdot 1/\lambda$$

- $E[\#\text{Fish}] = \underbrace{\lambda_T}_{\text{first two hours}} + \underbrace{P(0,2)}_{\text{no fish at first 2 hours}} \cdot 1$ ← will catch 1

$.6 \quad 2$

# L23: More Poisson

Thursday, December 1, 2016 6:23 PM

## Sum of Poisson RV
↳ Independent

> The sum of independent Poisson random variables, with means/parameters $\mu$ and $\nu$, is Poisson with mean/parameter $\mu + \nu$

Recall

Poisson $(\lambda T)$

$$P(k, T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$$

EXERCISE: PEOPLE IN THE PARK (2 points possible)

Busy people arrive at the park according to a Poisson process with rate $\lambda_1 = 3$/hour and stay in the park for exactly $1/6$ of an hour. Relaxed people arrive at the park according to a Poisson process with rate $\lambda_2 = 2$/hou and stay in the park for exactly half an hour. The arrivals of busy and relaxed people are independent processes. An observer visits the park at a specific time and sees $B$ busy and $R$ relaxed people at the park at that moment.

For both parts below, use standard notation . If your answer involves the exponential function, use notation such as e^(3).

a) Find that probability that $B = 0$. *Hint:* Think about what must have happened in the immediate past. Recall also the formula for the Poisson PMF with parameter $\lambda$:

$$\frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } k = 0, 1, 2, \ldots.$$

$\mathbf{P}(B = 0) =$ 3^(1/6)*e^3*1/(1/6) ✖ Answer: e^(-0.5)

$3^{\frac{1}{6}} \cdot e^3 \cdot \frac{1}{\frac{1}{6}}$

You entered:

$$6\sqrt[6]{3}e^3$$

b) Find the probability that $B + R = 1$.

$\mathbf{P}(B + R = 1) =$ 5*e^-5 ✖ Answer: 1.5*e^(-1.5)

$5 \cdot e^{-5}$

You entered:

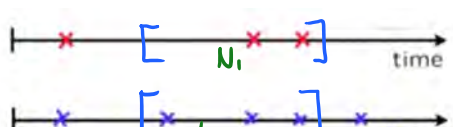
$$\frac{5}{e^5}$$

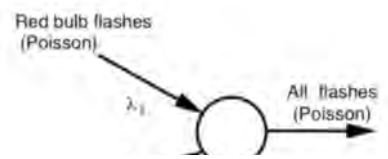
Answer:
a) The busy people that the observer sees are exactly those busy people who arrived during the last (1/6)th of an hour. It is therefore a Poisson random variable with parameter $3 \cdot (1/6) = 1/2$. The desired probability is $e^{-1/2}$.

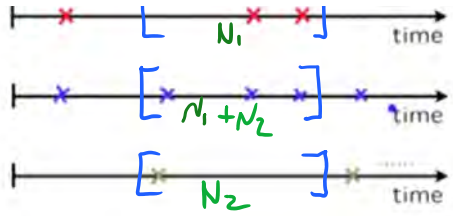
b) By the same argument, $R$ is an independent Poisson random variable with parameter $2 \cdot (1/2) = 1$. Thus, $B + R$ is a Poisson random variable with parameter $1.5$. Using the formula for the Poisson PMF,

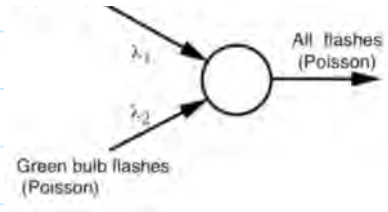
$$\mathbf{P}(B + R = 1) = 1.5e^{-1.5}.$$

## Merging of independent Poisson processes






RED

| | | $1-\lambda_1\delta$ | $\lambda_1\delta$ | $O(\delta^2)$ |
|---|---|---|---|---|
| | | 0 | 1 | $\geq 2$ |
| $1-\lambda_2\delta$ | 0 | $1-(\lambda_1+\lambda_2)\delta$ | $\lambda_1\delta$ | |
| GREEN $\lambda_2\delta$ | 1 | $\lambda_2\delta$ | $O(\delta^2)$ | |
| $O(\delta^2)$ | $\geq 2$ | | | |

Merged process: Poisson$(\lambda_1+\lambda_2)$

$P(\text{Red} \mid \text{arrival at time } t) = \lambda_1/(\lambda_1+\lambda_2)$

$P(k^{th} \text{ arrival is red}) = \lambda_1/(\lambda_1+\lambda_2)$

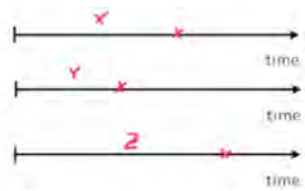
$P(4 \text{ out of } 10 \text{ arrivals are Red})$

↳ 10 Bernoulli trials

$$\Rightarrow \binom{10}{4}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^4\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^6$$

## Example: Light Bulbs

- Three lightbulbs
  - independent lifetimes $X$, $Y$, $Z$; exponential($\lambda$)
- Find expected time until first burnout




too hard

$$E[\min\{x,y,z\}] = \int_0^\infty\int_0^\infty\int_0^\infty \min\{x,y,z\}\,\lambda e^{-\lambda x}\lambda e^{-\lambda y}\lambda e^{-\lambda z}\,dx\,dy\,dz$$

$$P(\min\{x,y,z\} \geq t) = P(x \geq t)\,P(y \geq t)\,P(z \geq t)$$
$$= e^{-\lambda t}e^{-\lambda t}e^{-\lambda t} = e^{-3\lambda t}$$

$$\Rightarrow \min\{x,y,z\} \sim \text{Exp}(3\lambda) \Rightarrow E[\min\{x,y,z\}] = \boxed{1/3\lambda}$$

OR

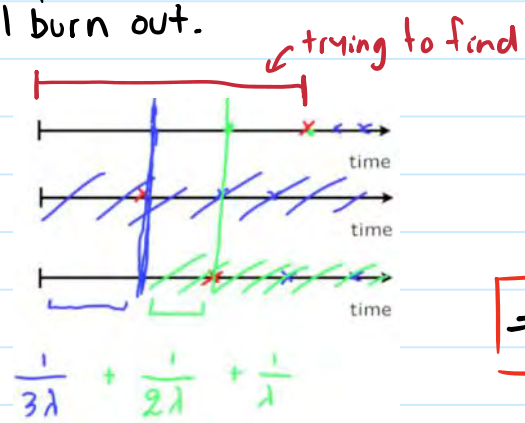
Merge all thee poisson processes: Poisson$(3\lambda)$
and find first arrival
$= \text{Exp}(3\lambda)$

$$\Rightarrow E[\min\{x,y,z\}] = 1/3\lambda \checkmark$$

Find expected time until all burn out.

trying to find

Find expected time until all burn out.




- Break up into sections

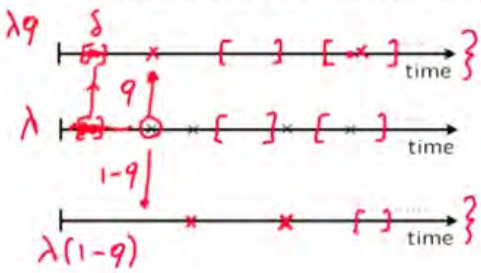
$$\boxed{= \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda}}$$

↳ Merge remaining 2 processes, find first arrival

↳ Merge 3 processes, find first arrival

# Splitting a Poisson Process

- Split arrivals into two streams, using independent coin flips of a coin with bias $q$
  - assume that coin flips are independent from the original Poisson process




$\geq 2$ $\quad O(\delta^2)$

$= 1$ $\quad \lambda \delta q$

Resulting streams are Poisson, rates $\lambda q$, $\lambda(1-q)$

- Are the two resulting streams independent?

  Surprisingly, yes!

# Random Incidence - Sampling Matters!

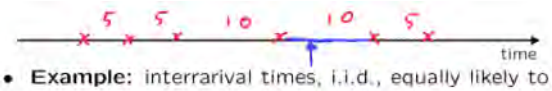
EXERCISE: RANDOM INCIDENCE (1/1 point)

Consider an arrival process whose interarrival times are independent exponential random variables with mean 2 (and consequently variance equal to 4), and consider the interarrival interval $S$ seen by an observer who arrives at a fixed time $t^*$, as in the preceding video. What is the variance of $S$?

8 ✔ Answer: 8

Answer:
As discussed in the preceding video, such an interval is the sum of two independent exponential random variables. Its variance is the sum of the variances of these two exponentials: $4 + 4 = 8$.

Example:




- **Example:** interrarival times, i.i.d., equally likely to be 5 or 10 minutes

  expected value of $k$th interarrival time: $\frac{1}{2}\cdot 5 + \frac{1}{2}\cdot 10 = 7.5$

- you show up at a "random time"

  $\mathbf{P}$(arrive during a 5-minute interarrival interval) $= \frac{1}{3}$

  expected length of interarrival interval during which you arrive $= \frac{1}{3}\cdot 5 + \frac{2}{3}\cdot 10$

  $\approx 8.3 > 7.5$

since "random time" is biased towards longer interarrival times

# Sampling Matters

**Different sampling methods can give different results**

- Average family size?

  $\frac{3}{4}\cdot 1 + \frac{1}{4}\cdot 6$

  - look at a "random" family (uniformly chosen)
  - look at a "random" person's (uniformly chosen) family $\frac{3}{9}\cdot 1 + \frac{6}{9}\cdot 6$

- Average bus occupancy?

  $\boxed{0}$ $\boxed{50}$

  - look at a "random" bus (uniformly chosen)
  - look at a "random" passenger's bus

- Average class size?

# L24: Finite-State Markov Chains

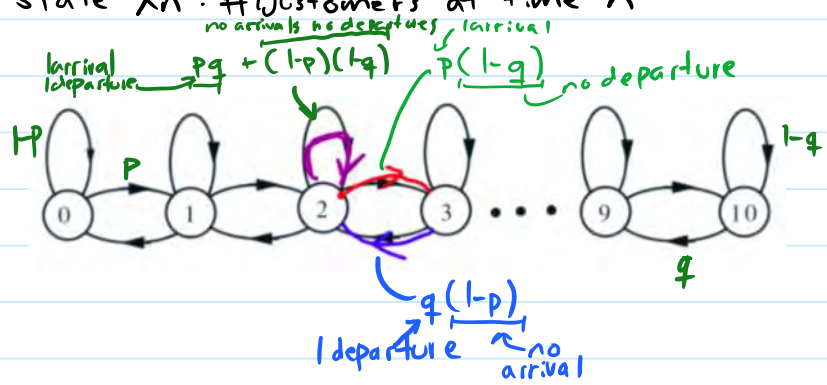
Friday, December 2, 2016 12:07 PM

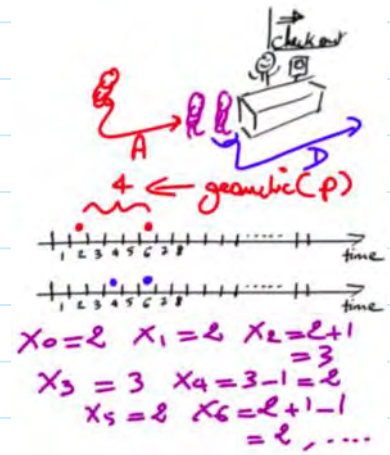
## Markov Processes

- Allow future to depend on the past
- States

Example: Checkout counter example

- customers arrive and leave queue at supermarket
- customer arrivals: Bernoulli(p)
- customer service times: geometric(q)
- state $X_n$: #Customers at time n




* Transition Probabilities of a node must sum to 1

## Markov Chains

Markov Property - Given current state, the past does not matter

## n-step transition probabilities

- state probabilities, given initial state i

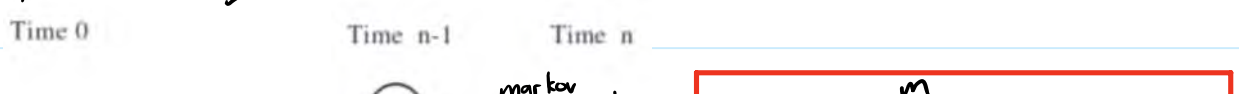
$$r_{ij}(n) = P(X_n = j \mid X_0 = i)$$

$$= P(X_{n+s} = j \mid X_s = i)$$

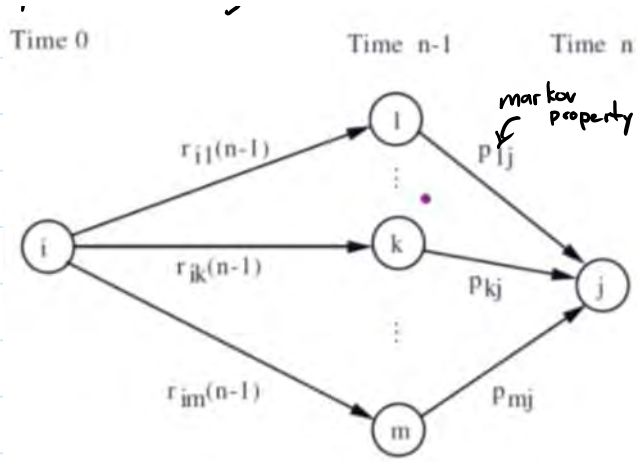
$i \xleftrightarrow{\text{n steps}} j$

Note: $r_{ij}(0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$ $\quad r_{ij}(1) = P_{ij}$

ways to reach j from i




markov

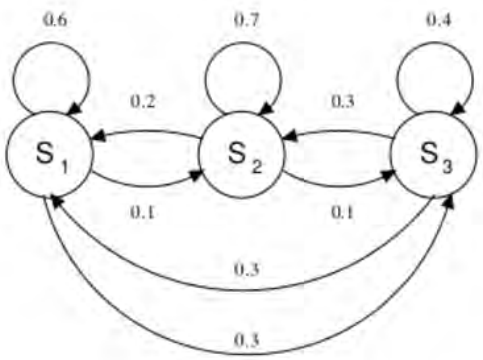



$$\boxed{r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)\, p_{kj}}$$

$$= \sum_{k=1}^{m} p_{ik}\, r_{kj}(n-1)$$

EXERCISE: N-STEP CALCULATION (2/2 points)

Consider the following transition probability graph:




Calculate the three-step transition probability $r_{11}(3)$ by using the recursion formula

$$r_{ij}(n) = \sum_{k=1}^{3} r_{ik}(n-1)p_{kj}.$$

$r_{11}(3) =$ (.47*.6)+(.22*.2)+(.31*.

$$r_{11}(3) = \sum_{k=1}^{3} r_{1k}(2)\, p_{k1}$$

$$= r_{11}(2)\, p_{11} + r_{12}(2)\, p_{21} + r_{13}(2)\, p_{31}$$

$$r_{11}(2) = \sum_{k=1}^{3} r_{1k}(1)\, p_{k1}$$

$$= r_{11}(1)\, p_{11} + r_{12}(1)\, p_{21} + r_{13}(1)\, p_{31}$$

$$r_{12}(2) = \sum_{k=1}^{3} r_{1k}(1)\, p_{k2}$$

$$= r_{11}(1)\, p_{12} + r_{12}(1)\, p_{22} + r_{13}(1)\, p_{32}$$

$$r_{13}(2) = \sum_{k=1}^{3} r_{1k}(1)\, p_{k3}$$

$$= r_{11}(1)\, p_{13} + r_{12}(1)\, p_{23} + r_{13}(1)\, p_{33}$$

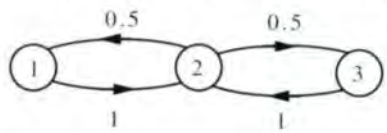
$$r_{11}(3) = (.47*.6)+(.22*.2)+(.31*.3)$$

Convergance

Probability of starting at i, taking n steps, and ending at j

$$r_{ij}(n) \xrightarrow[n\to\infty]{} \pi_j$$

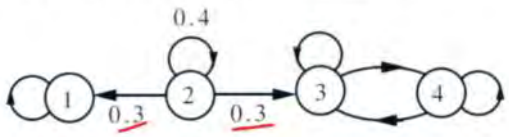
- does $r_{ij}(n)$ converge to something?




$n$ odd: $r_{22}(n) = 0$

$n$ even: $r_{22}(n) = 1$

- does the limit depend on initial state?



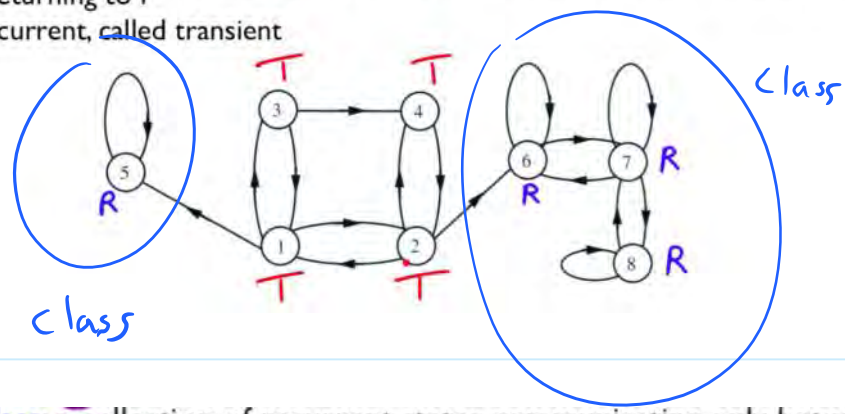

$r_{11}(n) = 1$

$r_{31}(n) = 0$

$r_{21}(n) = 1/2$

# Recurent and transient state

- state i is recurrent if "starting from i, and from wherever you can go, there is a way of returning to i"
- if not recurrent, called transient



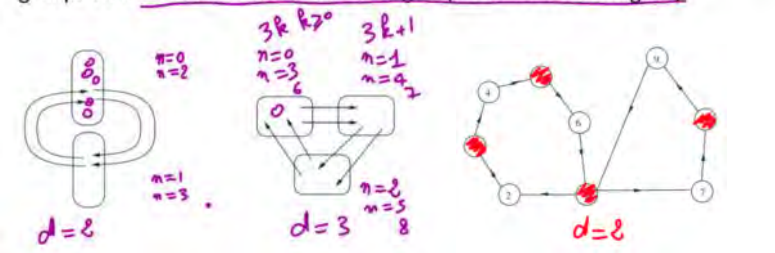

recurrent class: a collection of recurrent states communicating only between each other

# L25: Steady-state behavior of Markov Chains

Saturday, December 3, 2016 11:27 AM

## Periodic state in a recurrent class

The states in a recurrent class are periodic if they can be grouped into d > 1 groups so that all transitions from one group lead to the next group




- self transition ⟹ not periodic

## Steady state Probabilities

- does $r_{ij}(n)$ converge to $\pi_j$?

THM: If

- Recurrent states are all in single class and
- Single reccurent class is not periodic

$$\Rightarrow r_{ij}(n) \to \pi_j = \sum_k \pi_k P_{kj} \quad \text{for } j \in [0, m]$$

(initial state does not matter)

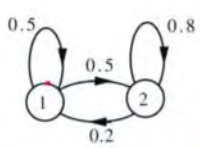
- m equations
- m unknowns

$$\sum_{j=1}^{m} \pi_j = 1$$

### Example

$\pi_1$ = steady state probability of being in state 1.

example




- single recurrent class
- not periodic

$$\begin{cases} \pi_j = \sum_{k=1}^{m} \pi_k P_{kj}, & j = 1, \dots m \\ m = 2 \end{cases}$$

$$\begin{cases} \pi_1 = \pi_1 \cdot 0.5 + \pi_2 \cdot 0.2 \\ \pi_2 = \pi_1 \times 0.5 + \pi_2 \times 0.8 \end{cases} \Rightarrow \begin{cases} \pi_1 \times 0.5 = \pi_2 \times 0.2 \\ \pi_2 \times 0.2 = \pi_1 \times 0.5 \end{cases}$$

same

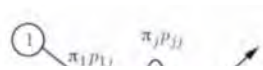
$$\begin{cases} \pi_1 \times \frac{1}{2} = \pi_2 \times \frac{1}{5} \\ \pi_1 + \pi_2 = 1 \end{cases} \Rightarrow \begin{cases} \pi_1 = \frac{2}{5}\pi_2 \\ \pi_2\left(\frac{2}{5} + 1\right) = 1 \end{cases} \Rightarrow \begin{cases} \pi_1 = \frac{2}{5} \times \frac{5}{7} = \frac{2}{7} \\ \pi_2 = \frac{5}{7} \end{cases}$$

$$\frac{2}{5}\pi_2 + \pi_2 = 1 \Rightarrow \pi_2 = 5/7 \Rightarrow \pi_1 = \left(\frac{2}{5}\right)\left(\frac{5}{7}\right)$$

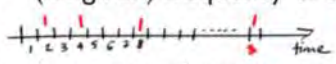
## Frequency Interpretation

### Balance equation:

- (long run) frequency of being in $j$: $\pi_j$

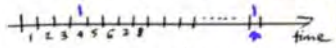


- (long run) frequency of being in $j$: $\pi_j$



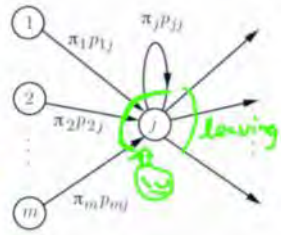

$\frac{Y_j(n)}{n} \approx \pi_j$ (n large)

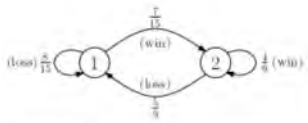
- frequency of transitions $1 \to j$: $\pi_1 p_{1j}$




$\frac{Z_{1j}(n)}{n} \approx \pi_1 \cdot p_{1j}$ (n large)

- frequency of transitions into $j$: $\sum_k \pi_k p_{kj}$




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

.456 ✔ Answer: 0.45652

Answer:
The long-term frequency of winning can be found as the sum of the long-term frequency of transitions from state 1 to state 2 and from state 2 back to itself. These can be found from the steady-state probabilities $\pi_1$ and $\pi_2$, which are known to exist as the chain is aperiodic and recurrent. The local balance and normalization equations are

$$\frac{7}{15}\pi_1 = \frac{5}{9}\pi_2$$
$$\pi_1 + \pi_2 = 1,$$

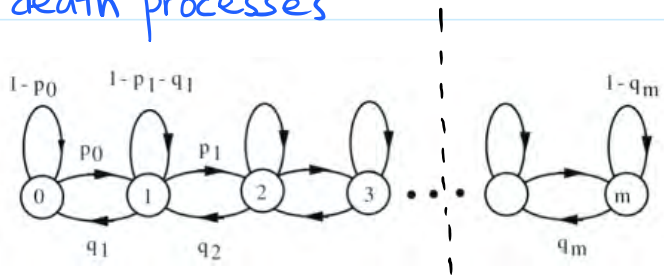
which lead to the solution $\pi_1 = 25/46$ and $\pi_2 = 21/46$.

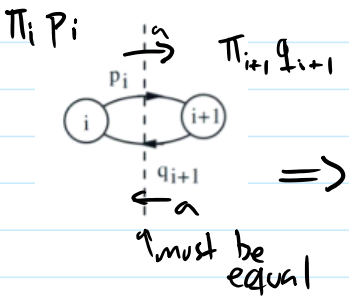
The probability of winning can now be found as

$$\pi_1 p_{12} + \pi_2 p_{22} = \frac{25}{46} \cdot \frac{7}{15} + \frac{21}{46} \cdot \frac{4}{9} = \frac{21}{46} = \pi_2,$$

and so for this particular game of blackjack, the long-term probability of winning also happens to equal the steady-state probability of being in state 2.

# Birth-death processes




$\Rightarrow$ $\boxed{\pi_i p_i = \pi_{i+1} q_{i+1}}$

- Need $\pi_0$

$$\pi_0 + \pi_0 \times \frac{p_0}{q_1} + \pi_0 \times \frac{p_0 \cdot p_1}{q_1 q_2} + \cdots$$

special case: $p_i = p$ and $q_i = q$ for all $i$

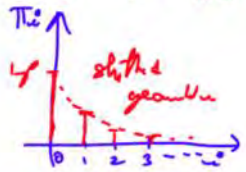
$\rho = p/q \qquad \pi_{i+1} = \pi_i \frac{p}{q} = \pi_i \rho \qquad \pi_1 = \pi_0 \rho,\ \pi_2 = \pi_1 \rho = \pi_0 \rho^2, \ldots$

$\pi_i = \pi_0 \rho^i \quad i = 0, 1, \ldots, m \qquad \sum_j \pi_j = 1 \Rightarrow \pi_0 [1 + \rho + \rho^2 + \cdots + \rho^m] = 1$

- assume $p = q$ $\Rightarrow \pi_i = \pi_0 \quad i = 0, \ldots, m \qquad \pi_0 [1 + m] = 1,\ \pi_0 = \frac{1}{1+m}$
- assume $p < q$ and $m \approx \infty$ $\Rightarrow \sum_{i=1}^{\infty} \rho^i = \frac{1}{1-\rho}$

$\boxed{\pi_0 = 1 - \rho}$, $\boxed{\mathbf{E}[X_n] = \frac{\rho}{1-\rho}}$ (in steady-state) $\quad \frac{\rho}{1-\rho}$

$\pi_i = \pi_0 \rho^i = (1-\rho)\rho^i, \ i = 0, \ldots$

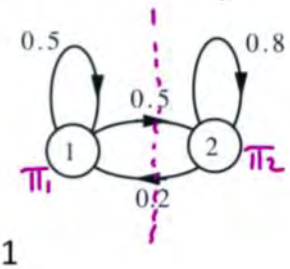


# L26: Absorption Probabilities and expected time to absorption

Wednesday, December 14, 2016 5:30 PM

$$\begin{cases} \pi_1 \times 0.5 = \pi_2 \times 0.2 \\ \pi_1 + \pi_2 = 1 \end{cases}$$

$$\boxed{\pi_1 = 2/7,\ \pi_2 = 5/7}$$




assume process starts in state 1

$P(X_1 = 1 \text{ and } X_{100} = 1 \mid X_0 = 1) =$

$\mathbb{P}(X_1 = 1 \mid X_0 = 1) \times \mathbb{P}(X_{100} = 1 \mid X_1 = 1, \cancel{X_0 = 1})$

$p_{11} \times r_{11}(99) \approx p_{11} \times \pi_1 = 0.5 \times \frac{2}{7}$

$P(X_{100} = 1 \text{ and } X_{101} = 2 \mid X_0 = 1) =$

$\mathbb{P}(X_{100} = 1 \mid X_0 = 1) \times \mathbb{P}(X_{101} = 2 \mid X_{100} = 1, \cancel{X_0 = 1})$

$r_{11}(100) \times p_{12} \approx \pi_1 \times p_{12} = \frac{2}{7} \times 0.5$

$P(X_{100} = 1 \text{ and } X_{200} = 1 \mid X_0 = 1) =$

$\mathbb{P}(X_{100} = 1 \mid X_0 = 1) \times \mathbb{P}(X_{200} = 1 \mid X_{100} = 1, \cancel{X_0 = 1})$

$= r_{11}(100) \times r_{11}(100) \approx \pi_1 \times \pi_1 = \pi_1^2 = (2/7)^2$

Is $n = 99, 100$ big enough?

Mixin Timescale: How long it takes for initial state to be forgotten

## Design of a phone system

How many lines to have in a phone system:

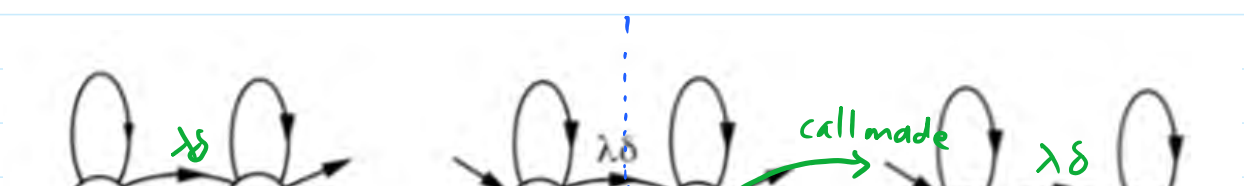
too many $\rightarrow$ waste of money

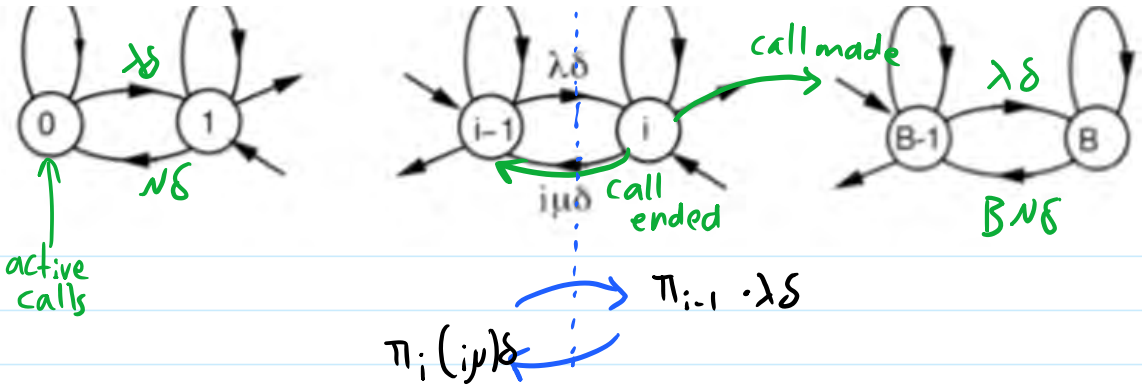
too little $\rightarrow$ bad performance

- calls originate as $\sim \text{Poi}(\lambda)$
- each duration $\sim \text{Exp}(\mu)$ Given/estimated

$P(1 \text{ new call}) \approx \lambda \delta$

$P(1 \text{ call ends} \mid i \text{ active calls}) \approx i\mu\delta$






$$\pi_{i-1} \cdot \lambda\delta$$

$$\pi_i (i\mu)\delta$$

$$\pi_i (i\mu)\cancel{\delta} = \pi_{i-1} \cdot \lambda\cancel{\delta}$$

$$\Rightarrow \pi_i = \pi_0 \frac{\lambda^i}{\mu^i i!}$$

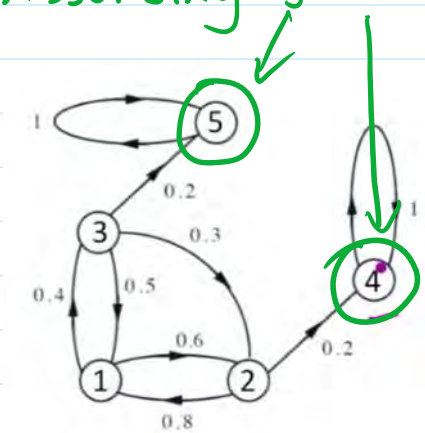
$$\pi_0 = 1 \Big/ \sum_{i=0}^{B} \frac{\lambda^i}{\mu^i i!}$$

$$P(\text{arriving customer finds busy system}) = \pi_B$$

$$\Rightarrow \text{If } \pi_B \le 1\% \Rightarrow B \ge 106$$

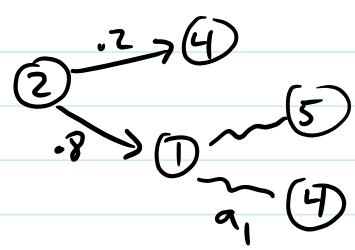
## Absorbing Probabilities

**Absorbing state:** Recurrent state $k$ with $P_{kk} = 1$




Let $a_i$ be probability of settling in 4 given you started in $i$

$a_4 = 1$
$a_5 = 0$



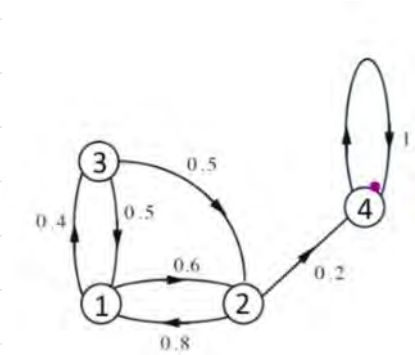

$$a_2 = .7\, a_1 + .8\, a_1$$

Can solve this system
$$\begin{cases} a_2 = .2a_4 + .8a_1 \\ a_1 = .6a_2 + .4a_3 \\ a_3 = .3a_2 + .5a_1 + \cancel{.2a_5} \end{cases}$$

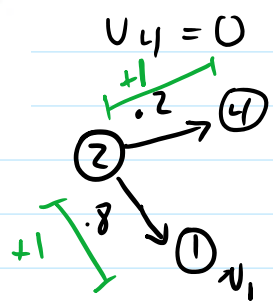
$$a_i = \sum_{j=1}^{m} P_{ij} a_j$$

* Can think of recurring class as one big absorbing state.

## Expected time to absorption




Let $v_i$ = expected # transitions given initial state is $i$




Can solve system
$$\begin{cases} v_2 = 1 + (.2)v_4 + (.8)v_1 \\ v_1 = 1 + .6v_2 + .4v_3 \\ v_3 = 1 + .5v_1 + .5v_2 \end{cases}$$

## Expected Time to absorption

- Combine all absorbing states
- Combine branches
- Solve as above

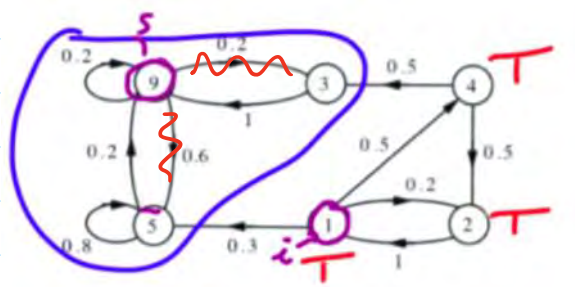
## Mean first Passage

$v_i = E[\min\{n \geq 0 \text{ s.t. } X_n = s\} \mid X_0 = i]$

## Mean First Passage

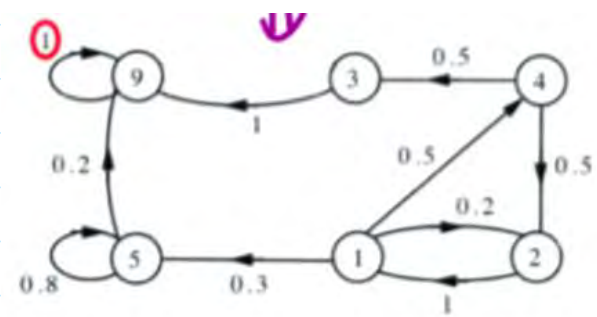
$$t_i = E[\min\{n \geq 0 \text{ s.t. } X_n = s\} \mid X_0 = i]$$

↳ Expected # steps to get from $i \to s$ for the first time




- Since we don't care what happens after (9) we can remove arcs from (9)

$\Downarrow$




- 9 is now an absorbing state

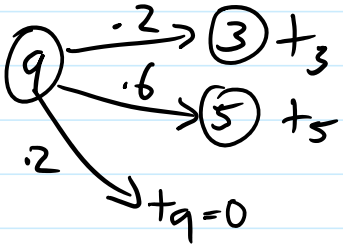
$\Rightarrow$ can solve as above

$$t_s = 0$$
$$t_i = 1 + \sum_{i \neq s} p_{ij} t_j$$

## Mean Recurrance Time

- IF you leave (9), how long will it take to get back to (9)

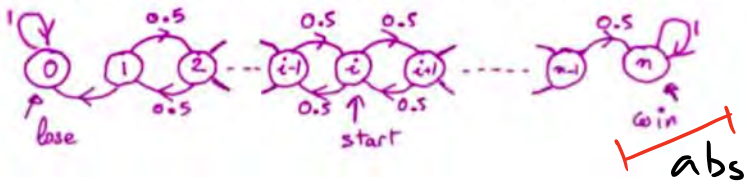
$$t_s^* = E[\min\{n \geq 1 \text{ s.t } X_n = s\} \mid X_0 = s]$$




$$t_9^* = 1 + .2t_3 + .6t_5$$

## EXAMPLE: Gambler's Run

- a gambler starts with $i$ dollars; each time, she bets \$1 in a fair game, until she either has 0 or $n$ dollars.
- what is the probability $a_i$ that she ends up with having $n$ dollars?




$a_0 = 0$

$a_n = 1$

$$a_i = .5 a_{i+1} + .5 a_{i-1}$$

$$\Rightarrow a_i = i/n$$

Expected wealth at end?

$$= 0 \cdot (1 - a_i) + n \cdot a_i = n(i/n) = i$$

How long will gambler stay in game

$\Rightarrow$ expected time to absorption

$N_i$ = expected # plays starting from $i$

$N_0 = 0$

$N_n = 0$ $\qquad N_i = 1 + .5 N_{i+1} + .5 N_{i-1}$

$$N_i = 1 + \sum_j p_{ij} N_j \qquad \Rightarrow N_i = i(n-i)$$

- in case of unfavorable odds? $p \neq 0.5$

$r = \frac{1-p}{p}$ $\qquad a_i = \frac{1-r^i}{1-r^n}$ $\qquad \mu_i = \left(\frac{r+1}{r-1}\right)\left(i - n \cdot \frac{1-r^i}{1-r^n}\right)$

# Final Review

Thursday, December 15, 2016 12:04 PM

## Conditional Expectation

$$E[X \mid Y=y] = g(y)$$

$$g(Y) \rightarrow E[E[X \mid Y]] = E[X]$$

$$Y = \sum_{i=0}^{n} X_i \rightarrow E[Y] = \sum_{i=0}^{n} E[X_i]$$

does not work if random sum

$$Z = \sum_{i=0}^{N} X_i \rightarrow E[Z] = E[E[Z \mid N]] = E[N E[X]]$$

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

$$\text{var}(Z) = E[\text{var}(Z \mid N)] + \text{var}(E[Z \mid N])$$

$$= E[N \,\text{var}(X)] + \text{var}(E[X]N)$$

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

## Inference

$\Theta$ Posterior: $P_{\Theta \mid X}$

$X$ $f_{\Theta \mid X}$

$\hat{\Theta}(X)$ : Estimator

$\theta(x)$ : Estimate given specific observation

### MAP estimator

$$\hat{\theta}(x) = \underset{\theta}{\text{argmax}}\ P_{\Theta \mid X}(\theta \mid x)$$

or $f_{\Theta \mid X}(\theta \mid x)$

$$p(\hat{\Theta} = \theta \mid x)$$

$$P(\hat{\Theta} = \theta) = \sum P(\hat{\Theta} = \theta \mid \Theta = \theta)\, P_{\Theta}(\theta)$$

### LMS Estimator

$$\hat{\Theta}(X) = \min_{\theta} E[(\Theta - \hat{\theta}(x))^2 \mid X]$$

$$\hat{\Theta}(X) = E[\Theta \mid X]$$

$$E[\text{Error}] = E[\text{var}(\Theta|X)]$$

$$\text{Error} = E[(\Theta - \hat{\Theta}(x))^2]$$

$$f_X(x) = Ke^{-(\alpha x^2 + \beta x + \gamma)}$$

$$X \sim N\left(\frac{-\beta}{2\alpha}, \frac{1}{2\alpha}\right)$$

$$\Theta \sim N(y_0, \sigma^2) \quad \text{indep} \quad W_i \sim N(0, \sigma_i^2)$$

$$Y_i = \Theta + W_i$$

① Calculate Posterior

$$f_{\Theta|Y_1^n}(\theta, y_1 \ldots) = \frac{f_{Y_1^n|\Theta}(y_1 \ldots) f_\Theta(\theta)}{f_{Y_1 \ldots}(y_1 \ldots)}$$

Proportional to

$$\propto \prod_{i=1}^{n} e^{-(y_i-\theta)^2/2\sigma_i^2} \cdot e^{-\frac{(\theta - y_0)^2}{2\sigma_0^2}}$$

Conditioned on $\theta$:
Y's are independent: Gaussian
Mean: 0
var: W

$$= \prod_{i=0}^{n} e^{\frac{-(\theta - y_i)^2}{2\sigma_i^2}} = h(\theta)$$

② Find max

$$\frac{\delta \log(h\theta)}{\delta\theta} = \frac{\delta}{\delta\theta}\left(\sum_{i=0}^{n} \frac{(y_i - \theta)^2}{2\sigma_i^2}\right)$$

$$= \sum_{i=0}^{n} \frac{(y_i - \theta)}{\sigma_i^2} = 0$$

$$\Rightarrow \sum_{i=0}^{n} \frac{y_i}{\theta_i} = \theta \sum_{i=0}^{n} \frac{1}{\sigma_i^2} \Rightarrow \boxed{\hat{\theta}_{MAP} = \frac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}}$$

LLMS Estimator

$$\hat{\theta}(x) = E[\theta] = \frac{cov(\theta, x)}{var(x)} \left(x - E[x]\right)$$

## Convergance

### WLLN

Sequence of iid $X_1, \ldots X_n$

$E[X_i] = \mu \quad var(X_i) = \sigma_0^2$

$$M_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

close to mean by WLLN

Average of $n$ $X_i$'s

$$\lim_{n \to \infty} P(|M_n - \mu| \geq \epsilon) = 0$$

### Central Limit Theorem

seq iid $X_i$

$E[X_i] = \mu \quad var(X) = \mu_0^2$

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

$$\lim_{n \to \infty} (Z_n \leq z) = P(\overset{N(0,1)}{Z} \leq z)$$

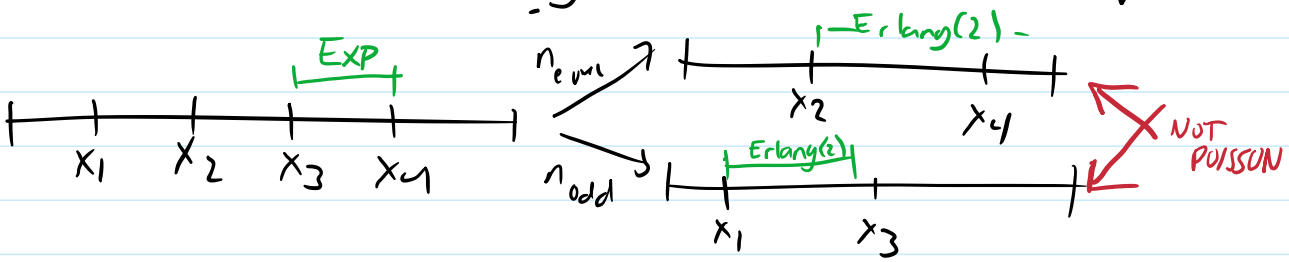
## Stoichastic Processes

$$\sum_{i=0}^{k} Geom(p) = Pascal(K, p)$$

| | Poissom | Bernoulli |
|---|---|---|
| Time of arrival | cont | discrete |
| arrival rate | $\lambda$ | $p$ |
| ↳ PMF | Poiss | Bin |
| Interarrival | EXP | Geom |

| | | |
|---|---|---|
| ↳ PMF | Poiss | Bin |
| Interarrival | EXP | Geom |
| Time until $k^{th}$ arrival | Erlang | Pascal |

Splitting a Poisson process in a dependent manner resulting processes are not Poisson




# Markov Chains

9/19/16

# 6.041 Problem Set #1

$\frac{50}{50}$ nice!

Fernando Trujo

## Exercise 1

a) $P(H_1)P(H_2)P(H_3) + P(T_1)P(T_2)P(T_3) = 1/4$

b) $P(H_1)P(H_2)P(T_3) + P(H_1)P(T_2)P(H_3) + P(T_1)P(H_2)P(H_3)$
$P(T_1)P(T_2)P(H_3) + P(T_1)P(H_2)P(T_3) + P(H_1)P(T_2)P(T_3) = 3/4$

c) $P(H_1)P(H_2)P(T_3) + P(T_1)P(H_2)P(H_3) + P(T_1)P(T_2)P(H_3) + P(H_1)P(T_2)P(T_3) = 1/2$

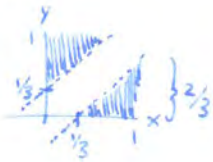
d) $P(H_1)P(H_2)P(T_3) + P(H_1)P(T_2)P(H_3) + P(T_1)P(H_2)P(H_3) = 3/8$

e) $P(H_1)P(H_2)P(T_3) + P(H_1)P(T_2)P(H_3) + P(T_1)P(H_2)P(H_3) + P(H_1)P(H_2)P(H_3) = 1/2$

f) $P(H_1)P(T_2)P(H_3) + P(T_1)P(H_2)P(T_3) = 1/4$

## Exercise 2

1) $P(A)$




$y - x > 1/3$ OR $y > x + 1/3$
$x - y > 1/3$ $\quad y < x - 1/3$

$P(A) = (2/3)^2$

2) $P(B)$

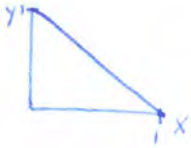
$x > 1/4$
$y > 1/4$

$P(B) = 1 - (1/4)^2 = 15/16$

3) $P(A \cap B)$

$= P(A) = (2/3)^2$
since $A$ is subset of $B$

4) $P(C)$




$x + y = 1$
$y = -x + 1$

Area of line $= P(C) = 0$

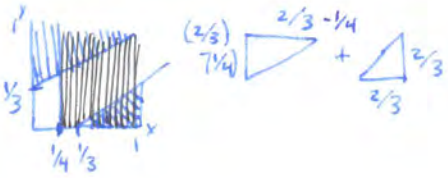
5) $P(D)$

Let Alice be $x$.

$x > 1/4$

$P(D) = 3/4$

6) $P(A \cap D)$




$= \left(\frac{2}{3} - \frac{1}{4}\right)^2 \frac{1}{2} + \left(\frac{2}{3}\right)^2 \frac{1}{2} = 89/288$

Exercise 3

1) $P(A \cap B \cap C^c) = P(A \cap B)\, P(C^c \mid A \cap B)$

Let $D = A \cap B$

$P(D \cap C^c) = P(D)\, P(C^c \mid D)$ (Associative Property)

$\dfrac{P(D \cap C^c)}{P(D)} = P(C^c \mid D)$ (Baye's Rule)

$\square$ $\boxed{\text{TRUE}}$

2) $P(A \cap B \cap C^c) = P(A)\, P(C^c \mid A)\, P(B \mid A \cap C^c)$

$P(A \cap B \cap C^c) = P(A \cap B)\, P(C^c \mid A \cap B)$ (From 1)

$= P(A \cap C^c)\, P(B \mid A \cap C^c)$ (Rename events)

$= P(A)\, P(C^c \mid A)\, P(B \mid A \cap C^c)$ (Baye's Rule)

$\boxed{\text{TRUE}}$ $\square$

3) $P(A \cap B \cap C^c) = P(A)\, P(C^c \cap A \mid A)\, P(B \mid A \cap C^c)$

$P(A \cap B \cap C^c) = P(A)\, \underbrace{P(C^c \mid A)}_{P(C^c \mid A) = P(C^c \cap A \mid A)}\, P(B \mid A \cap C^c)$ (From 2)

since we already know A occurred intersecting $C^c$ with A does not change anything

$\Rightarrow P(A \cap B \cap C^c) = P(A)\, P(C^c \cap A \mid A)\, P(B \mid A \cap C^c)$ $\square$ $\boxed{\text{TRUE}}$

4) $P(A \cap B \mid C) = P(A \mid C)\, P(B \mid A \cap C)$

$= \dfrac{P(C \mid A)\, P(A)}{P(C)} \cdot \dfrac{P(A \cap B \cap C)}{P(A \cap C)}$ (Bayes and conditional probability)

$P(A \cap B \mid C) = \cancel{\dfrac{P(C \mid A \cap B)\, P(A \cap B)}{P(C)}}$ $= \dfrac{P(A \cap B \cap C)}{P(A \cap C)} \cdot \dfrac{P(A \cap C)}{P(C)}$

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

$= \dfrac{P(B \mid A \cap C)\, \cancel{P(A \cap C)}}{\cancel{P(A \cap C)}} \cdot \dfrac{P(C \mid A)\, P(A)}{P(C)}$ (Bayes)

$= P(B \mid A \cap C)\, P(A \mid C)$ (Bayes)

$= P(A \mid C)\, P(B \mid A \cap C)$ $\boxed{\text{TRUE}}$

Exercise 4

a) Pick old

$P(\text{defective 1st pick}) = 75/500$

$P(\text{defective 2nd pick} \mid \text{defective first pick}) = 74/499$

$P(\text{both defective}) = \left(\frac{75}{500}\right)\left(\frac{74}{499}\right)$

Pick new

$P(\text{defective 1st pick}) = 75/1500$

$P(\text{defective 2nd pick} \mid \text{defective 1st pick}) = 74/1499$

$P(\text{both defective}) = \left(\frac{75}{1500}\right)\left(\frac{74}{1499}\right)$

$\Rightarrow P(\text{both defective}) = \left(\frac{1}{2}\right)\left(\frac{75}{500}\right)\left(\frac{74}{499}\right) + \frac{1}{2}\left(\frac{75}{1500}\right)\left(\frac{74}{1499}\right)$

b) $P(\underset{A}{\text{old widget}} \mid \underset{B}{\text{both defective}}) = \dfrac{P(A \overset{\text{and}}{\cap} B)}{P(B)}$

$P(A \text{ and } B)$

$= \dfrac{P(B \mid A)\,P(A)}{P(B)}$ (Bayes Rule)

$= \dfrac{\left(\frac{75}{500}\right)\left(\frac{74}{499}\right)\frac{1}{2}}{P(\text{both defective})}$

$\boxed{\approx .90}$

# 6.041 Problem Set #2

$\frac{49}{50}$

Fernando Trujan
Sep 24 2016

## Exercise 1

1.

a) A and B are NOT independent since knowing B changes the probability of A from $1/25$ to $1/5$.

b) A and C are NOT independent at all since knowing C changes the probability of A to 0. (since the events are disjoint.

2. a) $P(D) = 3/25$
$P(E) = 5/25$
$P(D \cap E) = 1/25$ (each die is 4)
$P(D \cap E) \neq P(D)P(E)$
$\Rightarrow$ NOT independent

b) Event F: At least one 3 rolled

$$P(F|D) = \frac{P(F \cap D)}{P(D)} = \frac{2/25}{3/25} = 2/3$$

## Exercise 2

There are 3 cases

Case 1: After n tosses Alice has more heads than Bob $\leftarrow$ Bob can't win
Case 2: After n tosses Alice is tied with Bob $\leftarrow$ Bob may win (with last toss)
Case 3: After n tosses Bob has more heads than Alice $\leftarrow$ Bob already won

By symmetry $P(C1) = P(C3)$

Also $P(C2) = 1 - P(C1) - P(C3) = 1 - 2P(C3)$

So $P(\overset{Bw}{\text{Bob}} \text{ wins}) = P(C1)\underbrace{P(Bw \mid C1)}_{0} + P(C2)\underbrace{P(Bw \mid C2)}_{1/2} + P(C3)\underbrace{P(Bw \mid C3)}_{1}$ (Total Probability

$= \left(\frac{1}{2}\right)P(C2) + P(C3)$

$= \left(\frac{1}{2}\right)(1 - 2P(C3)) + P(C3)$

$= \frac{1}{2} - P(C3) + P(C3)$

$= 1/2$

$\square$

Exercise 3

1. $4/52$

2. $\left(\frac{4}{52}\right)\left(\frac{3}{51}\right) + \left(\frac{48}{52}\right)\left(\frac{4}{51}\right) = \frac{204}{2652} = .0769 = \frac{4}{52}$

First card is an Ace; second card is an ace; First card is not an ace; second card is ace

3. $\frac{4}{51}$ ← # aces / ← # cards left

4.

a) $\left(\frac{4}{52}\right)\left(\frac{3}{51}\right)\left(\frac{2}{50}\right)\left(\frac{48}{49}\right)\left(\frac{47}{48}\right)\left(\frac{46}{47}\right)\left(\frac{45}{46}\right)$ } Probability of 1 specific sequence w/ exactly 3 aces

First card is ace; second card is ace; third card is Ace; Rest of cards not aces

$\Rightarrow$ P(specific sequence) $\binom{7}{3}$ = P(7 cards include exactly 3 aces)

# of sequences that there are

$$\frac{112078080}{674274182400} \cdot 35 = 9/1547 = \frac{\binom{4}{3}\binom{48}{4}}{\binom{52}{7}}$$

b) $\left(\frac{4}{52}\right)\left(\frac{3}{51}\right)\left(\frac{48}{50}\right)\left(\frac{47}{49}\right)\left(\frac{46}{48}\right)\left(\frac{45}{47}\right)\left(\frac{44}{46}\right)$ } Probability of 1 specific sequence w/ exactly 2 kings

First two cards are kings; Rest of cards not kings

$\Rightarrow$ P(specific sequence) $\binom{7}{2}$ # sequences that there are

$$= 594/7735 = \frac{\binom{4}{2}\binom{48}{5}}{\binom{52}{7}}$$

c) Event A = 7 cards include exactly 3 aces $\Rightarrow P(A) = 9/1547$

Event B = 7 cards include exactly 2 kings $\Rightarrow P(B) = 594/7735$

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

$P(A \cap B)$

$\left(\frac{4}{52}\right)\left(\frac{3}{51}\right)\left(\frac{2}{50}\right)\left(\frac{4}{49}\right)\left(\frac{3}{48}\right)\left(\frac{44}{47}\right)\left(\frac{43}{46}\right)$ } P of 1 specific sequence

First 3 cards ace; Next two cards king

$\Rightarrow P(A \cap B) = P(\text{specific sequence})\binom{7}{5} = \frac{1419}{83615350} = \frac{\binom{4}{2}\binom{4}{3}\binom{44}{2}}{\binom{52}{7}}$

off by factor of 10

$\Rightarrow P(A \cup B) = 9/1547 + 594/7735 - 1419/83615350 = 6906171/83615350 \approx 0.08259$

2/3

Exercise 4

1. $\binom{n}{1} \overbrace{\sum_{i=0}^{n-1} \binom{n-1}{i}}^{\text{Pick rest of team}}$

Pick leader

$n \quad 2^{n-1}$ since each member can be in or out

$$\boxed{= n\,2^{n-1}}$$

2. $\sum_{k=1}^{n} k \binom{n}{k}$

can be seen as forming a team of $k$ (the $\binom{n}{k}$ part) and selecting a leader (the $(k)$ part)

since this is the same as 1 we can equate the two

$$\Rightarrow \sum_{k=1}^{n} k \binom{n}{k} = n\,2^{n-1}$$

3. $\binom{n}{1}\binom{n-1}{1} \overbrace{\sum_{i=0}^{n-2} \binom{n-2}{i}}^{\text{Rest of team either in or out} \Rightarrow 2^{n-2}}$

Pick leader

pick vice leader

$$\Rightarrow n \cdot (n-1) \cdot 2^{n-2}$$

Fernando Trujando
Collab: Erika Lu

# 6.041 Problem Set #3

$\frac{48}{50}$

10/1/16

a) $E[X^2] = \sum_x x P_X(x^2)$

$g(X) = x^2$

$\Rightarrow E[X^2] = E[g(X)] = \sum_x g(x) P_X(x) = \sum_x x^2 P_X(x) \neq \sum_x x P_X(x^2)$

$\Rightarrow$ False

b) $E[X^2] = \sum_x x^2 P_X(x)$

$\Rightarrow$ True (see explanation for a)

c) $E[X^2] = \sum_x x^2 P_{X,Y}(x)$ $P_{X,Y}$ is a function of two variables, so $P_{X,Y}(x)$ is nonsense, so it is false

-2

We don't care about Y in this case so the PMF essentially collapses from $\rightarrow$ so by a) $\Rightarrow$ True

d) $E[X^2] = \sum_x x^2 P_{X,Y}(x,y)$

False since the right hand side is a function of y (and left is not)

e) $E[X^2] = \sum_x \sum_y x^2 P_{X,Y}(x,y)$

True - this is essentially the same as as b just adding up the values a bit differently.

f) $E[X^2] = \sum_z z P_{X^2}(z)$

$z = x^2$

$g(X) = z$

$E[X^2] = E[g(X)] = \sum_x g(X) P_X(x) = \sum_x x^2 P_X(x) = E[X^2]$ by b) $\Rightarrow$ true

Exercise 2

$X = X_2 - X_1$

1) PMF:

$X = X_2 - X_1$

| $X_2$ \ First Roll $X_1$ | 1 | 2 | 3 |
|---|---|---|---|
| 3 | 2 | 1 | 0 |
| 2 | 1 | 0 | -1 |
| 1 | 0 | -1 | -2 |

1/9 each

$$P_X(x) = \begin{cases} -2, & 1/9 \\ -1, & 2/9 \\ 0, & 3/9 \\ 1, & 2/9 \\ 2, & 1/9 \end{cases}$$

Expectation:

$$E[X] = \sum_x x P_X(x)$$

$$= \cancel{(-2)(1/9)} + \cancel{(2)(1/9)} + \cancel{(-1)(2/9)} + \cancel{1(2/9)} + 0(3/9) = \boxed{0}$$

Variance:

$$Var(X) = E[(X-\mu)^2] = E[X^2] = \sum_x x^2 P_X(x)$$

$$(4)(1/9) + (1)(2/9) + \cancel{(0)(3/9)} + (1)(2/9) + (4)(1/9)$$

$$4/9 + 2/9 + 2/9 + 4/9 = \boxed{12/9}$$

2) $y = x^2$

$$P_Y(y) = \begin{cases} 4, & 2/9 \\ 1, & 4/9 \\ 0, & 3/9 \end{cases}$$

square $P_X(x)$ and combine

Exercise 3

1. $P(B \mid A) = P(B) = p$

since the events are independent

2. $P(\underbrace{\text{3 heads in first 4 tosses}}\ \text{and}\ \underbrace{\text{exactly 2 H in last 3}})$

$$= \binom{4}{3} p^3(1-p) \cdot \binom{3}{2} p^2(1-p) \Rightarrow 12\, p^5 (1-p)^2$$

Where heads could end up

since independent

Where heads could end up

3.

$P(\underbrace{\text{2nd H at 4th toss}}_{\text{event } A} \mid \underbrace{\text{4 heads in first 7 tosses}}_{\text{event } B})$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) = \binom{7}{4} p^4 (1-p)^3$$

$P(A \cap B) =$

$$\underbrace{_\ _\ _}_{1H}\ \underset{p}{\underline{H}}\ \underbrace{_\ _\ _}_{2H}$$

$$\binom{3}{1} p(1-p)^2 \qquad \binom{3}{2} p^2(1-p)$$

$$= \overset{3}{\binom{3}{1}} p(1-p)^2\, p \overset{3}{\binom{3}{2}} p^2(1-p)$$

$$= 9p^4(1-p)^3$$

$$P(A \mid B) = \frac{9\cancel{p^4}\cancel{(1-p)^3}}{\binom{7}{4}\cancel{p^4}\cancel{(1-p)^3}} = \frac{9}{\binom{7}{4}} = \boxed{9/35}$$

4) A = exactly 5 heads in first 8 tosses
B = exactly 3 heads in last 5 tosses

$P(A \cap B) =$

$$\underbrace{_\ _\ _\ _\ _}_{5H}\ \underbrace{_\ _\ _}_{3H}$$

Must have at least 1 head

$\Rightarrow$ (next page)

3 cases:

c1 $\quad$ 2H $\quad$ H H H | T T $\quad$ 2H $\quad$ 1H

c2 $\quad$ 3H $\quad$ H H | T

c3 $\quad$ H H | H H $\quad$ 4H $\quad$ 1H

$$\binom{5}{2} p^2(1-p)^3 \cdot p^3 \cdot (1-p)^2$$
$$+$$
$$\binom{5}{3} p^3(1-p)^2 \cdot \binom{3}{2} p^2(1-p)$$
$$+$$
$$\binom{5}{4} p^4(1-p) \binom{3}{1} p(1-p)^2\, p^2$$

4) (cont)

$$P(A \cap B) = P(C_1) + P(C_2) + P(C_3)$$

$$= 10p^5(1-p)^5 + 30p^5(1-p)^3 2p(1-p) + 5p^7(1-p)^3 3$$

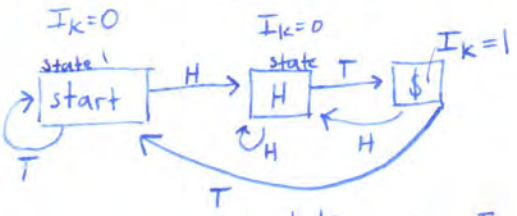
$$\boxed{= 10p^5(1-p)^5 + 60p^6(1-p)^4 + 15p^7(1-p)^3}$$

Exercise 4

1) a) $\sum_k I_k = R$

b) $E[R] = E\left[\sum_k I_k\right] \overset{\text{by linearity of expectation}}{=} \sum_k E[I_k]$

(state diagram)




cant get to $ from start

$$E[I_k] = P(\text{state H at } k-1)\, E[I_k \mid \text{state H at } k-1] + \cancel{P(\text{start stage at } k-1)\, E[I_k \mid \text{start state at}_{k-1}]} \to 0$$

$$+ \cancel{P(\text{stage \$ at } k-1)\, E[I_k \mid \text{state \$ at } k-1]} \to 0$$

can't win twice in a row

Note that

$P(\text{being at state H}) = P(H \mid \text{Any state}) = P(H)$

since tossing H from any state leads to state H

$$E[I_k] = (p)\, \underbrace{E[I_k \mid \text{state H at } k-1]}_{(1-p) \leftarrow \text{rolling a tails} = 1 \text{ else } 0}$$

$$E[I_k] = p(1-p)$$

$$E[R] = \sum_k E[I_k] = \sum_{k=0}^{n} E[I_k] = np(1-p) \qquad = \boxed{np(1-p)}$$

2) a) $E[I_k^2]$ Note that $I_k^2 = I_k$ since $I_k = \begin{cases} 0 \text{ or} \\ 1 \end{cases}$ $\Rightarrow E[I_k^2] = E[I_k] = \boxed{p(1-p)}$

b) $E[I_k I_{k+1}] = 0$ since you cant get a point twice in a row (see state diagram)

c) $E[I_k I_{k+\ell}]$ for $\{(k,\ell) : k \geq 1, \ell \geq 2 \text{ and } k+\ell \leq n\}$

"any 2 at least 2 apart from each other" $\Rightarrow$ independent

$$= E[I_k]\, E[I_{k+\ell}] \overset{\text{independent}}{=} \underbrace{(E[I_k])}_{\text{ok}}^2 = \boxed{(p^2(1-p))^2}$$

Fernando Trujano

# 6.041A Problem Set 4

$\frac{50}{50}$ nice!

10/8/16

Collab: Erika Lu

Exercise 1.

1) $P_{X,Y,Z}(x,y,z) = \underbrace{P_Y(y)\,P_{Z|Y}(z|y)}\,P_{X|Y,Z}(x|y,z)$

$P_{Z,Y}(z,y)\,P_{X|Y,Z}(x|y,z)$ (General Multiplication Rule)

$= P_{X,Y,Z}(x,y,z)$ (again)

True

2) $P_{X,Y,Z}(x,y|z) = P_X(x)\,P_{Y|Z}(y|z)$

False Would have to be conditionally independent.

Counterexample if $P_Y(y) > 0$ and $P_X(x) = 0$

3) $P_{X,Y,Z}(x,y|z) = P_{X|Z}(x|z)\,P_{Y|X,Z}(y|x,z)$

$\underbrace{P_Z(z)P_{X,Y,Z}(x,y|z)} = \underbrace{P_{X|Z}(x|z)}\,P_{Y|X,Z}(y|x,z)\,\underbrace{P(z)}$

$P_{X,Y,Z}(x,y,z) = P_{XZ}(x,z)\,P_{Y|X,Z}(y|x,z)$

$= P_{XYZ}(x,y,z)$

$\Rightarrow$ True

4) $\sum_x P_{X,Y|Z}(x,y|z) = 1$ $\Rightarrow$ False

LHS is a function of y

5) $\sum_x \sum_y P_{X,Y|Z}(x,y|z) = 1$

$\Rightarrow$ True

Since you know that z happened and are counting all possibilities of x and y

6) $P_{X,Y|Z}(x,y|z) = P_{X,Y,Z}(x,y,z) / P_Z(z)$

$\Rightarrow$ True by definition of conditional probability

7) $P_{X|Y,Z}(x|y,z) = P_{X,Y,Z}(x,y,z) / P_{Y,Z}(y,z)$

$\Rightarrow$ True using a supervariable $Q = Y,Z$ then same as 6)

## Exercise 2

1) a) False X and Y could have different probabilities of success so not all trials would have the same P(success) thereby breaking a property of a binomial random variable.

b) False. Just because n and p are the same does not meat A and B are independent. and each trial must be independent from each other to be a binomial RV.

c) True Since X (m trials) and Y (n trials) have the same Probability of success ∴ all trials will have the same P(success) as each other so this property of a binomial RV is preserved.
m+n

2) False: counterexample. consider a uniform distribution centered at 0 E[X] will be 0 but X(ω) might not = 0.

3) True. Since $X^2$ must be positive the only way to have $E[X^2]=0$ is to have all values of $X = 0$

## Exercise 3

1) Let X: # trials needed to open the door.

$E[X] = \sum_x x\,P(x)$

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

pick wrong key first time

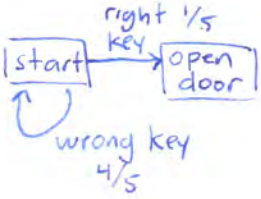
$= \frac{1}{5}(1+2+3+4+5) = 3$

$E[X^2] = \frac{1}{5}(1^2+2^2+3^2+4^2+5^2) = 11$

$var(X) = E[X^2] - (E[X])^2 = 11 - 9 = 2$

2) $E[X] = 1 + \frac{4}{5}E[X]$ $\Rightarrow \boxed{E[X]=5}$

* Geometric Random Variable $E[p] = 1/p$ (success) $var[p] = \frac{1-p}{p^2}$




$var(X) = \frac{1-1/5}{(1/5)^2} = \frac{4/5}{1/25} = \frac{20/25}{1/25} = \boxed{20}$

Exercise 4

$$f_Z(z) = \begin{cases} 5(1-z)^4 & 0 < z < 1 \\ 0 & \text{otherwise} \end{cases}$$

1) $E[Z] = \int_{-\infty}^{\infty} z f_Z(z)\,dz = \int_0^1 z f_Z(z)\,dz = \int_0^1 z5(1-z)^4\,dz$

$$= 5\left(\frac{z^6}{6} - \frac{4z^5}{5} + \frac{5z^4}{2} - \frac{4z^3}{3} + \frac{z^2}{2}\right)\Big|_0^1 = \boxed{1/6}$$

$E[Z^2] = \int_{-\infty}^{\infty} z^2 f_Z(z)\,dz = \int_0^1 z^2 5(1-z)^4\,dz = 1/21$

$Var(Z) = E[Z^2] - (E[Z])^2 = \frac{1}{21} - \frac{1}{36} = \boxed{5/252}$

2) $0.99 = F(Z) = \int_0^{x\,\text{total \$\$}} 5(1-z)^4\,dz = -(1-z)^5\Big|_0^x = (x-1)^5 + 1 = .99$

$\Rightarrow x = 1 - 1/10^{2/5} \approx .60189$

$\boxed{\approx \$60{,}190}$

Exercise 5

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

1. $F_X(x) = \int f_X(x)\,dx = \int_0^x \lambda e^{-\lambda t}\,dt = 1 - e^{-x\lambda}$ $\boxed{F_X(x) = 1 - e^{-x\lambda}}$

$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = 1/\lambda$ (From lecture) $\boxed{E[X] = 1/\lambda}$

$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx = 2/\lambda^2$

$Var(X) = E[X^2] - E[X]^2 = \left(\frac{2}{\lambda^2}\right) - \left(\frac{1}{\lambda^2}\right) = 1/\lambda^2$ $\boxed{Var(X) = 1/\lambda^2}$ $\boxed{P(X \geq E[X]) = 1/e}$

$P(X \geq E[X]) = P(X \geq 1/\lambda) = \underbrace{1 - P(X < 1/\lambda)}_{\text{hint}} = 1 - F_X(1/\lambda) = 1 - (1 - e^{-\lambda 1/\lambda}) = 1/e$

2. $P(X \geq t + z \mid X > t)$

$$\frac{P(X \geq t+z)}{P(X > t)} = \frac{1 - F_X(t+z)}{1 - F_X(t)} = \frac{1 - (1 - e^{-(t+z)\lambda})}{1 - (1 - e^{-\lambda t})} = e^{-\lambda z}$$

not dependent on $t$!

This means it does not matter what $t$ is / where you start

$P(X \geq t + z \mid X > t) = e^{\lambda z}$ for all $t$

Fernando Trujano — 6.041 Problem Set 5 — $\frac{50}{50}$ nice! — 10/15/16

Collab: Erika Lu

## Exercise 1

a) $X: N(0,1)$ $\quad P(X > 0.75) = 1 - \Phi(0.75) = .2266$

$Y: N(2,9)$ $\quad P(X \le -1.25) = \Phi(-1.25) = 1 - \Phi(1.25) = .1056$

b) $Z = (Y-3)/4 = \frac{Y}{4} - \frac{3}{4}$

$E[aX+b] = aE[X]+b$ $\quad$ Mean $Z = \frac{1}{4}E[Y] - 3/4 = 2/4 - 3/4 = -1/4$

$var(aX+b) = a^2 var(X)$ $\quad var(Z) = \left(\frac{1}{4}\right)^2 var(Y) = (1/16)\,9 = 9/16$

c) $P(-1 \le Y \le 2) = P(\underset{-1-2/3}{-1} \le X \le \underset{\frac{2-2}{3}}{0}) = P(X \le 0) - P(X \le -1) = \Phi(0) - \Phi(-1)$

$X = \frac{Y-\mu}{\sigma} = \frac{Y-2}{3}$ $\quad = \Phi(0) - (1 - \Phi(1))$

$= .5 - (1 - .8413) = .3413$

## Exercise 2

a) $\boxed{\text{No}}$ $F_X(x)$ takes on a value greater than 1.

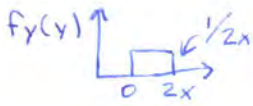
b) $\boxed{\text{Yes}}$ Values never decrease or go above 1

c) $\boxed{\text{No}}$ In a discontinuity, the rightmost value should be taken

d) $\boxed{\text{No}}$ Values decrease. not allowed by CDF since probabilities can't be negative.

## Exercise 3

a) $f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x)$




$f_{Y|X}(y|x) = \begin{cases} 1/2x & \text{for } 0 \le y \le 2x \\ 0 & \text{else} \end{cases}$

$f_X(x) = \begin{cases} x/800 & \text{for } 0 \le x \le 40 \\ 0 & \text{else} \end{cases}$

$\Rightarrow f_{X,Y}(x,y) = \begin{cases} 1/1600 & \text{for } 0 \le x \le 40 \text{ and } 0 < y < 2x \\ 0 & \text{else} \end{cases}$

b) $P(Z > 0) = 1/2$

$Z = Y - X$ $\quad$ since regardless of $X$ (money taken) there is a positive profit w/ probability $1/2$

c) $f_Z(z) = \int f_{Z|X}(z|x) f_X(x)\,dx$ $\quad$ (uniform $[-x, x]$)

$f_{Z|X} = \begin{cases} 1/2x & \text{for } -x \le z \le x \\ 0 & \text{else} \end{cases}$

$= \int_{|z|}^{40} (1/2x)(x/800)\,dx = \int_{|z|}^{40} 1/1600\,dx$ $\quad$ (if profit is $z$ $x$ is bounded.)

$f_Z(z) = \begin{cases} \frac{40-|z|}{1600} & \text{for } |z| < 40 \\ 0 & \text{else} \end{cases}$ $\quad$ since can lose/win at most \$40

d) $E[Z] = 0$ $\quad = \int z \frac{40-|z|}{1600}\,dz = 0$

↳ since regardless of amount of money brought, you will be winning and losing the same amount.

1/2

Exercise 4

a) X and Y are <u>not</u> independent

For example, if we know $X=0$ then $Y$ has to be $0$. $\Rightarrow$ not independent

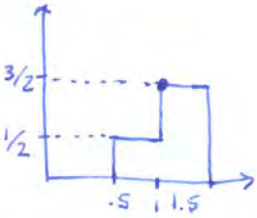
b) $f_X(x) = \begin{cases} x/2 & \text{for } 0 \le x \le 1 \\ -\frac{3x}{2}+3 & \text{for } 1 < x \le 2 \\ 0 & \text{else} \end{cases}$

c)


$f_{Y|X}(y|.5) = \begin{cases} 2 & 0 \le y \le .5 \\ 0 & \text{else} \end{cases}$ ← normalized

$\frac{1}{2}h = 1$
$\Rightarrow h = 2$

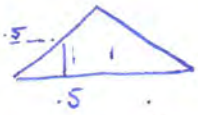
d)


$A = (.5)(3/2) + (.5)(.5) = 1$ ✓

$f_{X|Y}(x|.5) = \begin{cases} 1/2 & \text{for } .5 \le x \le 1 \\ 3/2 & \text{for } 1 < x < 3/2 \\ 0 & \text{else} \end{cases}$

e) $\mathbb{E}[XY \mid X < .5] = \int_0^{.5}\int_0^{x} xy\, f_{X,Y|A}(x,y|A)\, dy\, dx$




$= \int_0^{.5}\int_0^{x} xy\, 8\, dy\, dx$

$(.5)(.5)(.5) = 1/8 \rightarrow$ need to scale by 8

$\int_0^{.5} x\, 4y^2 \Big|_0^x dx = \int_0^{.5} 4x^3 dx = (.5)^4 = 1/16$

Exercise 5

a) $X: 0 \to c$ To get max variance

$= X_m$

$\mathrm{var}(X_m) = E[X_m^2] - (E[X_m])^2$ ($E[X_m] = c/2$)

$= E[(x-\mu)^2] = E[(X_m - \frac{c}{2})^2]$ ($\mu = c/2$)

$= \frac{1}{2}(0-\frac{c}{2})^2 + \frac{1}{2}(c-\frac{c}{2})^2 = c^2/4$

$\Rightarrow \mathrm{var}(X) \le c^2/4$ □

b) X is just another random variable with mean $\mu$ and variance $\sigma^2$

$\Rightarrow$ by linearity $Y = aX+b$ will have mean $= a\mu + b$ and variance $= a^2\sigma^2$

by linearity of Expectation $E[Y] = aE[X]+b$

↑ stnd div scales by $a$ $\Rightarrow$ var scales by $a^2$ since stdiv $= \sqrt{\sigma^2}$

c)

$\int_0^\infty P(X>t)\,dt = \int_0^\infty \int_t^\infty f_X(t)\,dx\,dt = \int_0^\infty \int_0^x f_X(x)\,dt\,dx = \int_0^\infty x f_X(x)\,dx = \mathbb{E}[X]$

Fernando Trujano

# 6.041A Problem Set 6

10/23/16

Collab: Erika (

47

## Exercise 1

$f_X(x) = e^{-x}$ for $x \geq 0$
$\quad 0$ else

$f_{Y|X}(y|x) = xe^{-yx}$ for $x > 0$ for $y > 0$ $\quad E[Y] = 1/x$
$\quad 0$ else

1. $f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x) = e^{-x} x e^{-yx} = xe^{-x-yx}$ for $x, y \geq 0$
$\quad 0$ else

2. $f_Y(y) = \int_0^\infty f_{X,Y}(x,y)\,dy = \int_0^\infty xe^{-x-yx}\,dy = \frac{1}{(1+y)^2}$ for $y > 0$ $\quad 0$ else ($\lambda = 1+y$)

3. $f_{X|Y=2}(x|2) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{1/(1+2)^2} = (xe^{-x-yx})\,9 = 9(xe^{-3x})$ for $y > 0$ $\quad 0$ else ($y=2$)

4. $E[X|Y=2] = \int_{-\infty}^{\infty} x f_{X|Y=2}(x|y=2)\,dx = \int_{-\infty}^{\infty} x\,9xe^{-3x} = \int_0^\infty 9x^2 e^{-3x} = 9(2/3^3) = 2/3$ ($\lambda = 3$)

5. $f_{Y|X=2,Y\geq 3}(y|x=2, y \geq 3) =$

$f_{Y|X=2}(y|x=2) = \frac{f_{X,Y}(x,y)}{f_X(x=2)} = \frac{xe^{-x-yx}}{e^{-2}} = \frac{2e^{-2-2y}}{e^{-2}} = 2e^{-2y}$

$f_{Y|X=2,Y\geq 3}(y|x=2, y\geq 3) = \frac{f_{Y|X=2}(y|x=2)}{\int_3^\infty f_{Y|X=2}(y|x=2)\,dy} = \frac{2e^{-2y}}{\int_3^\infty 2e^{-2y}\,dy} = \frac{2e^{-2y}}{-e^{-2y}\big|_3^\infty} = \frac{2e^{-2y}}{-e^{-2\infty} - -e^{-6}} = 2e^{-2y+6}$ ($\lambda = 2$)

6. $f_Z(z) =$ $\quad z = e^{2x}$ $\quad x = \frac{\ln z}{2}$

$F_Z(z) = P(Z \leq z) = P(e^{2X} \leq z) = P(x \geq \frac{\ln z}{2}) = F_X(\frac{\ln z}{2})$

$f_Z(z) = f_X(\frac{\ln z}{2}) \Big| \frac{d}{dz}(\frac{\ln z}{2}) = (e^{-\ln z/2})(1/2z) = \frac{1}{\sqrt{z}\,2z} = \frac{1}{2}z^{-1.5}$

## Exercise 2

1. Event A: $5 < V < 10$

a) $E[V|A] = \int_{-\infty}^{\infty} v f_{V|A}(v|A)\,dv = \int_5^{10} v\frac{1}{5}\,dv = 7.5$ miles

$f_{V|A} = \begin{cases} 1/5 & 5 < v < 10 \\ 0 & \text{otherwise} \end{cases}$

b) $E[T|A]$ $\quad T = 200/V \Rightarrow E[T|A] = \int_{-\infty}^{\infty} \frac{200}{v} f_{V|A}(v)\,dv = \int_5^{10} \frac{200}{v}(\frac{1}{5})\,dv = 27.725$ hours

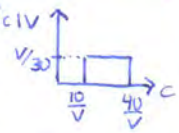
c) $\sigma_{T|A}$:

$var(T) = E[T^2|A] - (E[T|A])^2 = \sigma^2_{T|A} = 800 - (27.725)^2 = 31.32$

$E[T^2|A] = \int_5^{10} (\frac{200}{v})^2 (\frac{1}{5})\,dv = 800$ $\quad \Rightarrow \sigma_{T|A} = 5.596$

2. $f_{C|V}: [10/v, 40/v]$ uniform $\quad V: [5, 25]$ uniform

a) $f_{C|V}$

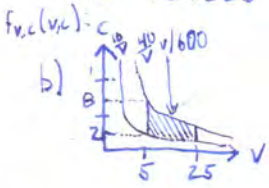



$(\frac{40}{v} - \frac{10}{v})h = 1$
$h = v/30$

$f_{C|V} = \begin{cases} v/30 & \text{for } 5 \leq v \leq 25 \\ 0 & \text{else} \end{cases}$ $\quad \frac{10}{v} \leq c \leq \frac{40}{v}$

for $5 \leq v \leq 25$

b)

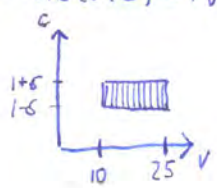



$f_V(v)$

$f_{V,C}(v,c) = \begin{cases} \frac{1}{20}\frac{v}{30} & \text{for } \frac{10}{v} \leq c \leq \frac{40}{v},\ 5 \leq v \leq 25 \\ 0 & \text{else} \end{cases}$

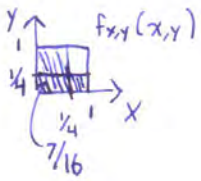
c) Event B: consumed 1 kilocalorie

$f_{V|B}(v|B) = f_{V|C=1}(v|c=1) = \frac{f_{V,C}(v,c)}{f_C(1)}$



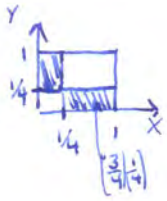
Exercise 3

1. $P(0 \le \overset{\text{RV message received}}{Z} \le 1) = \iint f_{X,Y}(x,y)\,dx\,dy$



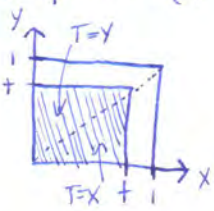

$= \boxed{7/16}$

2.




$A = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)2 = \boxed{3/8}$

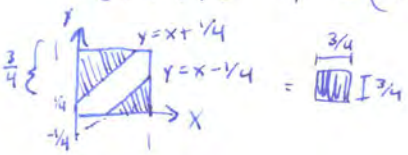
3. $F_T(t) = P(T \le t) = P(\max(X,Y) \le t) = t^2$




$f_T(t) = \frac{d}{dt} t^2 = 2t$

4. aka: "given than one channel gets it find P that other chanel doesn't get it for 15 mins"

$P(|X-Y| \ge 1/4) = (3/4)^2$

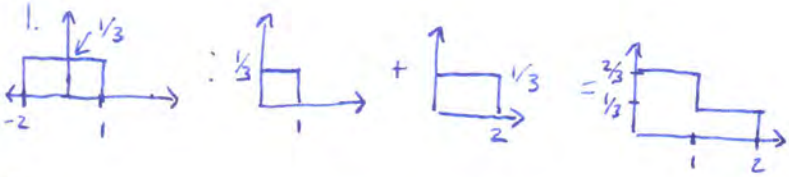



5. $P(\max(X,Y) \le 3/4)$

$F_T(3/4) = (3/4)^2 = P(|X-Y| \ge 1/4) \Rightarrow$ It is actually the same so $\boxed{\text{No}}$

Exercise 4

1.



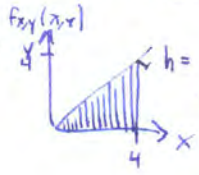

$f_Y(y) = \begin{cases} 2/3 & 0 \le y \le 1 \\ 1/3 & 1 < y \le 2 \\ 0 & \text{else} \end{cases}$

2. $Y = X$ when $x \ge 0 \Rightarrow f_X(x) = f_Y(y) = \begin{cases} 2e^{-2x} & \text{if } x \ge 0 \\ 0 & \text{else} \end{cases}$

3. $f_Y(y) = \begin{cases} f_X(y) + f_X(-y) & \text{if } y \ge 0 \\ 0 & \text{else} \end{cases}$

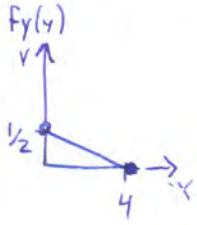
Exercise 5

1. No. Information about $X$ can reveal information about $Y$ - for example if I am given that $x=3$ I know $y \neq 5$. ie $y$ has to be $[0,4]$

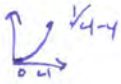
2. 


$\frac{1}{2}4^2h=1$
$8h=1$
$h=1/8$

$f_Y(y)=\int f_{X,Y}(x,y)\,dx=\int_y^4 1/8\,dx=\frac{1}{2}-\frac{y}{8}$ for $0\le y\le 4$

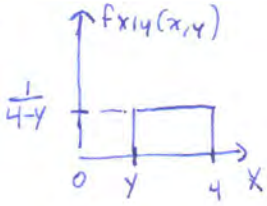



$f_{X,Y}(x,y)=f_Y(y)\,f_{X|Y}(x|y)$

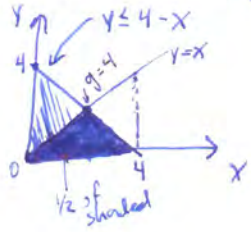



$f_{X|Y}(x|y)=\dfrac{f_{X,Y}(x,y)}{f_Y(y)}=\dfrac{1/8}{\frac{1}{2}-\frac{y}{8}}=\dfrac{1}{4-y}$ for $0\le y\le 4$

$0$ otherwise




3. $P(A)=P(G\le 4)=P(X+Y\le 4)=P(Y\le 4-X)=F_Y(4-X)$




$P(A)=1/2$

$f_{G|A}(g|A)$

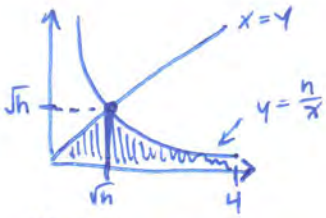
$F_{G|A}(g|A)=\dfrac{g\cdot\frac{g}{2}\cdot\frac{1}{2}}{4\cdot 2\cdot\frac{1}{2}}=g^2/16$

$f_{G|A}(g|A)=g/8$

4. $H=XY$

$E[H^2]=E[(XY)^2]=E[X^2Y^2]=\int h^2 f_H(h)\,dh$

$F_H(h)=F_H(XY)=P(XY\le xy)=P(Y\le \frac{xy}{x})=P(Y\le \frac{h}{x})$




$y=x=\frac{h}{x}\Rightarrow x=\sqrt{h}$
$y=\sqrt{h}$

$A=$

$A=\frac{h}{2}+\int_{\sqrt{h}}^4 \frac{h}{x}\,dx$

$A=\frac{h}{2}+h(\ln 4-\ln\sqrt{h})=F_H(h)=\frac{h}{2}+h\ln 4-\frac{h\ln h}{2}$

$f_H(h)=\frac{1}{2}+\ln 4-\frac{1}{2}(\ln h+1)=\ln 4-\frac{1}{2}\ln h=\ln(4/\sqrt{h})$

$E[H^2]=\int h^2 f_H(h)\,dh=\int_0^{16} h^2\ln(4/\sqrt{h})\approx 227.56$

Fernando Trujano
Collab: Erika Lu

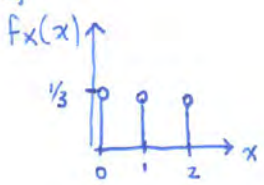
# 6.041B Problem set #7

11/5/16

39

## Exercise 1

a)



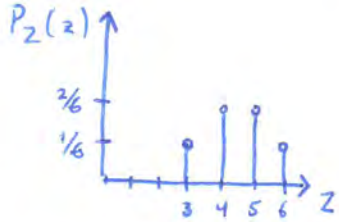
$Z = X + Y$ $\qquad$ x: 0, 1, 2,

$$P_Z(z) = \sum_x P_X(x)\, P_Y(z-x)$$

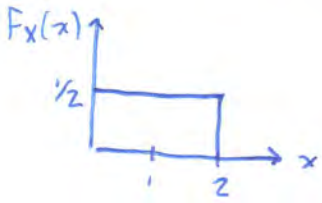
$$= P_X(0)P_Y(z) + P_X(1)P_Y(z-1) + P_X(2)P_Y(z-2)$$

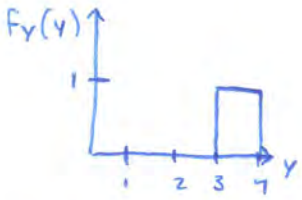
$$= \tfrac{1}{3}P_Y(z) + \tfrac{1}{3}P_Y(z-1) + \tfrac{1}{3}P_Y(z-2)$$

$$P_Z(z) = \begin{cases} (1/3)(1/2) & \text{if } x+y=3 \quad 1/6 \\ (1/3)(1/2)+(1/3)(1/2) & \text{if } x+y=4 \quad 2/6 \\ (1/3)(1/2)+(1/3)(1/2) & \text{if } x+y=5 \quad 2/6 \\ (1/3)(1/2) & \text{if } x+y=6 \quad 1/6 \\ 0 & \text{else} \end{cases}$$




b)




$$\cancel{f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx}$$

$$\cancel{= \int_0^2 (1/2)(1)\,dx}$$

$$\cancel{= \tfrac{1}{2}x\Big|_0^2 = 1 \quad \text{if } 0 \le x \le 2,\ 3 \le z-x \le 4}$$

$$f_{X,Z}(x,z) = f_X(x) f_Y(z-x)$$

$$= (1/2)(1) \quad \text{if } 0 \le x \le 2,\ \text{if } 3 \le z-x \le 4$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)\,dx \qquad 3-z \le -x \le 4-z$$

$$z-3 \ge x \ge z-4$$

$$\cancel{\int_{z-4}^{z-3} 1/2\,dx = \tfrac{1}{2}x\Big|_{z-4}^{z-3} = \tfrac{1}{2}(z-3) - \tfrac{1}{2}(z-4)}$$

$$f_Z(z) = \int_0^2 f_X(x) f_Y(z-x)\,dx$$

$$= \int_0^2 \tfrac{1}{2} f_Y(z-x)\,dx \qquad 3 \le z-x \le 4$$

$$z-4 \le x \le z-3$$

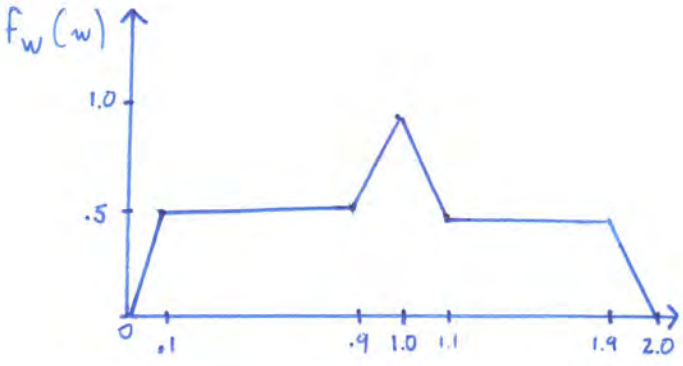
If ~~$0 < z-3$~~

$$f_Z(z) = \int_0^{z-3} \tfrac{1}{2}\,dx \qquad 3 \le z \le 5 \qquad = \tfrac{1}{2}x\Big|_0^{z-3} = \tfrac{1}{2}(z-3)$$

$$= \int_{z-4}^{2} \tfrac{1}{2}\,dx \qquad 4 \le z \le 6 \qquad = \tfrac{1}{2}x\Big|_{z-4}^{2} = 1 - \tfrac{1}{2}(z-4)$$

$$f_Z(z) = \begin{cases} 1/2(z-3) & \text{if } 3 \le z \le 5 \\ 1 - \tfrac{1}{2}(z-4) & \text{if } 4 \le z \le 6 \\ 0 & \text{else} \end{cases}$$

42

c) 


Excersise 2

a) (iii) $X_1$ and $X_2$ are negatively corrolated.

As $X_1$ goes up $X_2$ is likely to go down. If $X_1 = n$ then $X_2 = 0$

$X_1 = x$ then $X_2$ can be at most $n-x$

b) $\text{cov}(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2]$

$I_{1_i} = \begin{cases} 1 & \text{if roll } i = 1 \\ 0 & \text{else } 0 \end{cases}$

$I_{2_i} = \begin{cases} 1 & \text{if roll } i = 2 \\ 0 & \text{else} \end{cases}$

$X_1 = \sum_n I_{1_i}$ — Binomial w/ $p(\text{success}) = 1/n$

$X_2 = \sum_n I_{2_i}$

By linearity

$E[X_1] = n\frac{1}{k} = E[X_2]$

$E[X_1 X_2] = E\left[\sum_n I_{1_i} \sum I_{2_k}\right]$

$= \sum_{i=1}^{n}\sum_{j=1}^{n} E[I_{1_i} I_{2_j}] = \cancel{\sum_{i=1}^{n} E[I_{1_i} I_{2_i}]} + \sum_{(i,j):i\neq j} E[I_{1_i} I_{2_j}] = \sum E[I_{1_i}][I_{2_j}]$

By linearity

each term $\frac{1}{k^2}$

0 can't both be 1 on the same roll

$n^2 - n$ terms

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

$\text{cov}(X_1, X_2) = \frac{n^2-n}{k^2} - \left(\frac{n}{k}\right)^2 = \boxed{\frac{-n}{k^2}}$

c) $E[X_1] = np_1 \quad E[X_2] = np_2$

$E[X_1 X_2] \overset{\text{from } p}{=} \underbrace{\sum_{(i,j):i\neq j} E[I_{1_i}]E[I_{2_j}]}_{n^2-n \text{ terms}} = (n^2-n)p_1 p_2$

$\text{cov}(X_1, X_2) = \boxed{(n^2-n)p_1 p_2 - n^2 p_1 p_2} = \boxed{-n p_1 p_2}$

Exercise 3

$E[X]=E[Y]=E[Z]=0 \qquad E[X^3]=E[Y^3]=E[Z^3]=0$

$E[X^2]=E[Y^2]=E[Z^2]=1 \qquad E[X^4]=E[Y^4]=E[Z^4]=3$

$W=a+bX+cX^2 \qquad V=dX$

$$\rho(X-Y, X+Y) = \frac{\operatorname{cov}(X-Y, X+Y)}{\sigma_{X-Y}\,\sigma_{X+Y}} = \frac{\overset{FOIL}{E[(X-Y)(X+Y)]} - E[X-Y]\,E[X+Y]}{\sigma_{X-Y}\,\sigma_{X+Y}}$$

$$= \frac{E[X^2+\cancel{XY}-\cancel{XY}-Y^2] - E[X-Y]\,E[X+Y]}{\sigma_{X+Y}\,\sigma_{X+Y}} = \frac{E[\cancel{X^2}]-E[\cancel{Y^2}] - E[X-Y]\,E[X+Y]}{\sigma_{X-Y}\,\sigma_{X+Y}}$$

$$= (E[X]-E[Y])(E[\cancel{X}]+E[\cancel{Y}]) = \boxed{0}$$

$$\rho(X+Y, Y+Z) = \frac{\operatorname{cov}(X+Y, Y+Z)}{g} = \frac{E[(X+Y)(Y+Z)] - \cancel{E[X+Y]\,E[Y+Z]}^{\,0}}{g}$$

$g = \sigma_{X+Y}\sigma_{Y+Z}$

$\operatorname{cov}(X,Z) = E[XZ] - E[X]E[Z] = 0$ & symmetry

since uncorrelated

$$= \frac{E[X^2+XZ+Y^2+YZ]}{g} = \frac{E[X^2]+\cancel{E[XZ]}^{\,0}+E[Y^2]+\cancel{E[YZ]}}{g}$$

$$= 2/g = \frac{2}{\sigma_{X+Y}\sigma_{Y+Z}} = 2/2 = \boxed{1}$$

$\sigma_{X+Y} = \sqrt{\operatorname{var}(X+Y)} = \sqrt{E[(X+Y)^2] - E[(X+Y)]}$

$= \sqrt{E[X^2+2XY+Y^2] - E[\cancel{X}]+E[\cancel{Y}]}$

$= \sqrt{E[X^2] + 2E[\cancel{XY}] + E[Y^2]}$

$= \sqrt{2} = \sigma_{Y+Z}$ By symmetry

$$\rho(X, Y+Z) = \frac{\operatorname{cov}(X, Y+Z)}{\sigma_X\,\sigma_{Y+Z}} = \frac{E[X(Y+Z)] - \cancel{E[X]}^{\,0}E[Y+Z]}{\sqrt{2}\,\sigma_X} = \frac{E[\cancel{XY}]^{\,0} + E[\cancel{XZ}]^{\,0}}{} = \boxed{0}$$

$$\rho(W,V) = \rho(a+bX+cX^2, dX) \overset{\text{since scaling does not change correlation}}{=} \rho(X+X^2, X) = \frac{\operatorname{cov}(X+X^2, X)}{\sigma_{X+X^2}\,\sigma_X} = \frac{E[X^2+X^3] - \cancel{E[X+X^2]}^{\,0}}{\sqrt{1}\sqrt{\ }}$$

$\sigma_X = \sqrt{\operatorname{var}(X)} = \sqrt{E[X^2]-(E[X])^2} = \sqrt{1}$

$\sigma_{X+X^2} = \sqrt{\operatorname{var}(X+X^2)} = E[(X+X^2)^2] - E[X+X^2]^2$

$E[X^2+2X^3+X^4] - 1$

$0+2(0)+3 = \sqrt{3}$

$$= \frac{E[X^2]+E[X^3]}{\sqrt{1}\sqrt{3}} = \frac{1}{\sqrt{1}\sqrt{3}} = \boxed{1/\sqrt{3}}$$

Exercise 4: $E[X_i \mid i=k]$

$K$ - geometric RV with $p = .5$

$E[K] = 1/p = 1/.5 = 2$

$E[X_K] = 3/2 \leftarrow$ midpoint $[0,3]$

$$E[X] = E[K]E[X_K]$$
$$= 2 \cdot 3/2 = \boxed{3}$$

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

uniform

$$var(X) = E[K]\,var(X_K) + E[X_K]^2\,var(K)$$

From lecture 13 (near the end)

$$= (2)(3/12) + (3/2)^2 (1-p)/p^2$$
$$= (2)(3/12) + (3/2)^2 \cdot .5/(.5)^2$$
$$= \frac{6}{12} + \left(\frac{3}{2}\right)^2 (2)$$
$$= \frac{1}{2} + \frac{9}{4}\,2$$
$$= \frac{2}{4} + \frac{18}{4} = \frac{20}{4} = \boxed{5}$$

— 4

## Basic Probability

Sample space - set of all possible outcomes

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (AND, OR)

$(S \cup T)^c = S^c \cap T^c$ ← De Morgans

### Conditional Probability

$P(A|B) = P(A \cap B) / P(B)$ (multiply by B gives tree)

Product Rule: $P(A \cap B) = P(B)P(B|A)$

Total Probab: $P(B) = \sum_i P(A_i)P(B|A_i)$

Bayes Rule: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

### Independance

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

### Conditional Independance:

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

Independence $\not\Leftrightarrow$ conditional independance

### Pairwise Independence:

$P(A_i \cap A_j) = P(A_i)P(A_j)$ for each $i \neq j$

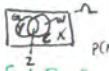
Pairwise independence $\not\Rightarrow$ Independance

↳ All events independent from each other

To maximize probability, take the derivative $\frac{d}{dp}$, set = 0 and find p.

### Sample Problems

Given: $P(G) = .6$, $P(C) = .7$, $P(G \cap C) = .4$ Find $P(G^c \cap C^c)$



$P(G) = P(Y \cup Z) = P(Y) + P(Z) \Rightarrow P(Y) = .2$ (P(Z) = .4)

$P(C) = P(W \cup Z) = P(W) + P(Z) \Rightarrow P(W) = .3$

$P(X) = P(G^c \cap C^c) = 1 - P(W) - P(Z) - P(Y) = .1$

### Set Proofs

Given: A, B, C independent

$A^c \cap B$ indep?

$A^c \cap B = B - A \cap B = B - AB$ (since $A^c \cap B \subseteq B$)

$= B(1-A) = BA^c \Rightarrow$ YES

Proove: $P(A \cap B \cap C^c) = P(A \cap B)P(C^c | A \cap B)$

Let $D = A \cap B$ $P(D \cap C^c) = P(D)P(C^c | D)$ □

## Counting

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

Partitions: $\#\text{parts} = \frac{n!}{n_1! n_2! \cdots n_r!}$ (n = distinct items, $n_i$ = size, r disjoint sets)

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

### Basic Counting Principle

For m-stage process with $n_i$ choices at stage i: $\#\text{choices} = n_1 n_2 \cdots n_m$

# ways to arrange n items were x of them are identical $= n!/x!$ (n = TOTAL, x = since they were repeated more than once)

### Sample Problems

P(3 aces and 2 kings) $= \binom{4}{2}\binom{4}{3}\binom{44}{2} / \binom{52}{7}$ (pick kings, pick aces, pick rest of cards, Total)

Draw 7 cards from deck Randomly:

P(exactly 3 aces): $\binom{4}{3}\binom{48}{4} / \binom{52}{7}$ (pick aces, pick non-aces, pick which cards, Total cards)

P(exactly 2 kings): $\binom{4}{2}\binom{48}{5} / \binom{52}{7}$

Probability of Failure: p(failure) = p/per hour

a) P(k failures in n hours) $= \binom{n}{k} p^k (1-p)^{n-k}$ (k failures, n-k success)

b) P(at least one failure in n hrs)

Let M = # Failures in n hrs

$P(M > 0) = 1 - P(\text{no mistake}) = 1 - (1-p)^n$

c) H = hrs until first failure. E[H] = ??

$E[H] = \sum_{i=0}^{\infty} P(H > i) = \sum_{i=0}^{\infty} (1-p)^i = (1-(1-p)^n)\frac{1}{p}$

### Password Counting

Passwords must be 6-8 chrs and start w/ letter. (upper/lower/digits)

Ω = all valid passwords

↳ Break into disjoint sets to make it easier

$\Sigma = \{a, b \ldots z, A, B, \ldots Z\}$ // letters

$\Phi = \{a, b \ldots z, A, B, \ldots Z, 0, 1 \ldots 9\}$ letters and digits

$\Omega = (|\Sigma \cdot \Phi^5|) \cup (|\Sigma \cdot \Phi^6|) \cup (|\Sigma \cdot \Phi^7|)$

$= |\Sigma|(|\Phi^5| + |\Phi^6| + |\Phi^7|)$

### Dice Rolling

Rolling 6 fair dice, two choices left

# ways all # distinct $= 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6!$

# ways exactly 5 distinct:

$= 6 \cdot 5 \cdot 6!/2$ (pick missing, pick repeated #, arrange 4 distinct and two identical)

### Birthday Problem

Find probability that n people have distinct birthday (pick for 1st person, pick for second person)

$|\Omega| = 365^n$ $P = 365 \cdot 364 \cdot 363 \cdots 365-n+1 / 365^n$

### Blue and Red Balls:

10 empty bins. 4 red balls placed into distinct bins, 4 blue placed in distinct bins independent

P(B, getting red ball) $= \frac{1}{10} + \frac{9}{10} \cdot \frac{1}{9} + \frac{8}{10} \cdot \frac{1}{8} + \ldots$ (First ball, second, first red makes it nicer)

$= 4/10$ OR total $= \binom{10}{4}$ # ways red $= \binom{9}{3}$ (distribute 3 remaining on 4 bins)

total/# ways $= 4/10 = P$

Expected # bins that get exactly 1 ball - N

$E[N] = \sum_{i=0}^{10} E[I_i] = 10E[I_1] = 10P(I_1 = 1)$ (# bins, since indep.)

$P(I_1 = 1) = p(1-p) + p(1-p)$ (one red ball, no blue, OR one blue, one red) $\Rightarrow E[N] = 20p \cdot (1-p)$

Pascal, many Bernoulli trials until r failures

$P_X(k) = \binom{k+r-1}{k} p^k (1-p)^r$ for $k \geq 0$ (# successes, p(success) = p)

$E[k|r] = rp/1-p$ $E[k+r|r] = r/1-p$ (# trials)

$\text{var}(k) = \frac{rp}{(1-p)^2}$ $P_X(k) = \binom{k+r-1}{r} p^k (1-p)^r$ (# fails, # successes)

## Discrete Random Variables

Random Variable: Associate # to every possible outcome. $X \leftarrow RV$ x ← numerical value

### Probability Mass Function (PMF)

↳ Probability distribution of discrete RV

$P_X(x) = P(X = x)$ $\sum_x P_X(x) = 1$

$P_{X,Y}(x,y) = P(X = x \text{ and } Y = y)$ ← Joint PMF

### Expectation ← Average

$E[X] = \sum_x x p_X(x)$ If $a \leq X \leq b \Rightarrow a \leq E[X] \leq b$

$E[g(X)] = \sum_x g(x) p_X(x)$ ← Expected value rule

$E[X^n] = \sum_x x^n p_X(x)$

$E[aX + b] = aE[X] + b$ ← Linearity of expectation

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

↳ Total Expectation Theorem

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

$E[X|A] = \sum_x x p_{X|A}(x)$

If X, Y independent $\Rightarrow E[XY] = E[X]E[Y]$

### Variance ← Spread of PMF

$\text{Var}(X) = E[(X - E[X])^2]$

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

$\text{var}(aX + b) = a^2 \text{var}(X)$ $\text{var}(XY) = \text{var}(X)\text{var}(Y)$

If X, Y independent $\Rightarrow \text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

standard deviation $\sigma_X = \sqrt{\text{var}(X)}$

### Multiple RV's

Joint PMF $p_{X,Y}(x,y) = P(X = x \text{ and } Y = y)$

Marginal PMF: $p_X(x) = \sum_y p_{X,Y}(x,y)$

$p_Y(y) = \sum_x p_{X,Y}(x,y)$

$E[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$

$p_{X|Y}(x|y) = p_{X,Y}(x,y) / p_Y(y)$

### Common Distributions

Bernoulli: (Indicator RV)

$X = \begin{cases} 1 & \text{(success)} \\ 0 & \text{(failure)} \end{cases}$

$p_X(k) = \begin{cases} p & k = 0 \\ 1-p & k = 1 \end{cases}$

$E[X] = p$

$\text{var}(X) = p(1-p)$

Uniform:

$X = [a, b]$

$p_X(k) = \begin{cases} \frac{1}{b-a} & k \in [a,b] \\ 0 & \text{else} \end{cases}$ (height)

$E[X] = (a+b)/2$

$\text{var}(X) = (b-a)^2/12$

Geometric

X = # trials till first success.

$P_X(k) = (1-p)^{k-1} p$ (first k-1 trials fails, last succeeds)

$E[X] = 1/p$

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

Binomial: X = # successes in n Bernoulli trials w/ p

$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = [0, n]$

$E[X] = np$

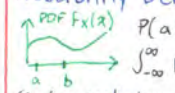
$\text{var}(X) = np(1-p)$

CDF: $F_X(x) = P(X \leq x) = \sum_k p_X(k)$ $P_X(k) = F_X(k) - F_X(k-1)$

## Continious Random Variables

Discrete: Continious.

### Probability Density Function (PDF)



$P(a \leq X \leq b) = \int_a^b f_X(x)dx$

$\int_{-\infty}^{\infty} f_X(x)dx = 1$

Single points have 0 probability. Need Area.

### Expectation

$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$

$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$

$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx$

$E[aX+b] = aE[X] + b$

Rest of properties same as discrete (same for variance)

### Cumulative Distribution Functions (CDF)

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$ (Integrate ⇒ CDF)

$f_X(x) = \frac{dF_X}{dx}(x)$ (differentiate ⇒ PDF)

As $x \to \infty$ $F_X(x) \to 1$. As $x \to -\infty$ $F_X(x) \to 0$

### Common Random Variables

Uniform:

$f_X(x) = \begin{cases} 1/(b-a) & \text{if } a \leq x \leq b \\ 0 & \text{else} \end{cases}$ $E[X] = \frac{b-a}{2}$

$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{else} \end{cases}$ $\text{var}(X) = \frac{(b-a)^2}{12}$

Exponential: Param λ

$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$ $E[X^2] = 2/\lambda^2$, $E[X] = 1/\lambda$

$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$ $\text{var}(X) = 1/\lambda^2$

Memoryless

Models time to wait until something happens

Normal/Gaussian

Standard Normal $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (mean, var) $N(0,1)$

$N(\mu, \sigma^2)$ $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ $E[X] = 0$ $\text{var}(X) = 1$

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$

$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$

Standard Normal Table

Standard Normal CDF $= \Phi(y)$: $F_Y(y) = P(Y \leq y)$

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

$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0,1)$ (Standardizing RV, can use SN Table)

JOINT of two independent normals shaped like 3D bell center = means stretch x,y ~ var(x) var(y)

### Conditioning

$f_{X|A}(x)\delta \approx P(x \leq X \leq \delta | A)$

$E[X|A] = \int x f_{X|A}(x)dx$

$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X \in A)} & \text{if } x \in A \\ 0 & \text{else} \end{cases}$

$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y)dy$

$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dy$

$P(A|Y=y) = P(A)F_{Y|A}(y)/F_Y(y)$

Bayes Rule:

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

$P(N = n | Y = y) = \frac{p_N(n) f_{Y|N}(y|n)}{f_Y(y)}$

Markov's Thm

$P(R \geq x) \leq \frac{E[R]}{x}$

↑ P that x exceeds threshold

Chebyshev's Thm

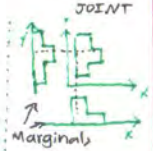
$P(|R - E[R]| \geq x) \leq \frac{\text{var}(R)}{x^2}$

Joint PDF

$P((X,Y) \in B) = \iint_{(x,y) \in B} f_{X,Y}(x,y)dxdy$

$f_X(x) = \int f_{X,Y}(x,y)dy$

$f_Y(y) = \int f_{X,Y}(x,y)dx$ ← marginal PDF



Marginals

## Derived Distributions and covariance

Discrete Case

$Y = aX + b \Rightarrow P_Y(y) = P_X(\frac{y-b}{a})$ (shift, stretch)

Continious Case

$Y = aX + b \Rightarrow f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$

General Case

$F_Y(y) = P(Y \leq y)$

① Find CDF of Y

② Differentiate $F_Y(y) = \frac{dF_Y}{dy}(y)$

Example: Given X uniform [0,2] and $Y = X^3$ Find $f_Y$

① $F_Y(y) = P(Y \leq y)$

$= P(X^3 \leq y) = P(X \leq y^{1/3})$

$= (1/2)y^{1/3}$

② $f_Y(y) = \frac{dF_Y}{dy}(y) = (1/6)(1/y^{2/3})$

Monotonic Case (inverse function)

$f_Y(y) = f_X(h(y)) \left|\frac{dh}{dy}(y)\right|$

Convolution: $W = X + Y$ X, Y indep

Discrete: $P_W(w) = \sum_x P_X(x)p_Y(w-x)$

Continuous: $f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w-x)dx$

Graphical Method

① Flip PMF of Y

② Shift PMF of Y by w

③ Cross Multiply and add.

* Sum of independent normal RV's is normal

### Covariance ← how much RVs change together

$\text{cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$

$\text{cov}(X,Y) = E[XY] - E[X]E[Y]$

If X, Y indep. $\text{cov}(X,Y) = 0$

but $\text{cov}(X,Y) = 0 \not\Rightarrow$ independance

If $\text{cov}(X,Y) = 0 \Rightarrow \text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

In general:

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

$\text{cov}(X,X) = \text{var}(X)$

$\text{cov}(aX + b, Y) = a \cdot \text{cov}(X,Y)$

$\text{cov}(X, Y+Z) = \text{cov}(X,Y) + \text{cov}(X,Z)$

### Correlation Coefficient

$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$ $-1 \leq \rho \leq 1$

$\rho = 0 \Leftrightarrow$ X, Y are uncorrelated

Independent $XY \Rightarrow \rho = 0$

$\rho = 0 \not\Rightarrow$ Indep X, Y

$|\rho| = 1 \Leftrightarrow$

$X - E[X] = c(Y - E[Y])$

↳ linearly dependent

$\rho(X,X) = \text{var}(X)/\sigma_X^2 = 1$ → extreme dependance

$\text{cov}(aX + bY) = a \cdot \text{cov}(X,Y) \Rightarrow$

$\rho(aX + bY) = \text{sign}(a)\rho(X,Y)$

YOU CAN DO IT!

Fernando Trujano

# Math Review

## Derivatives · k constant

$\frac{d}{dx}(k \cdot f(x)) = k \cdot f'(x)$

$\frac{d}{dx}(f(x))^n = n f(x)^{n-1} f'(x)$

$\frac{d}{dx}(f(x) \pm g(x)) = f'(x) \pm g'(x)$

$\frac{d}{dx}(f(x) \cdot g(x)) = f(x) g'(x) + g(x) f'(x)$

$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x) f'(x) - f(x) g'(x)}{(g(x))^2}$ LODINI -HIDILO/LO²

$\frac{d}{dx}(\ln(f(x))) = \frac{1}{f(x)} \cdot f'(x)$

$\frac{d}{dx} e^{f(x)} = e^{f(x)} \cdot f'(x)$

$\frac{d}{dx} a^{f(x)} = a^{f(x)} \ln(a) f'(x)$

$\frac{d}{dx} \log(x) = 1/x$

## Integrals

$\int k f(x) dx = k \int f(x) dx$

$\int (f(x) + g(x)) dx = \int f(x) dx + \int g(x) dx$

$\int u^n du = \frac{u^{n+1}}{n+1} + C$

$\int \frac{1}{u} du = \ln|u| + C$

$\int e^u du = e^u + C$

$\int a^u du = \frac{a^u}{\ln a} + C$

$\int_a^b f(x) dx = F(b) - F(a)$

$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$

$\int e^{at} dt = \frac{1}{a} e^{at} dt$

Integration by parts

$\int u dv = uv - \int v du$

## Weak Law of Large Numbers

For $\epsilon > 0$ $P(|M_n - \mu| \geq \epsilon)$ ← sample mean

$= P\left(\left|\frac{X_1 + X_2 + \dots X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$ as $n \to \infty$

## Convergance in probability

Sequence $Y_n$ converges to a if for $\epsilon > 0$ $\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0$

↳ ie almost all of the PDF of $Y_n$ eventually gets concentrated close to a.

# Sample Problems

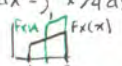
## Densities and Conditioning

Given: $f_X(x) = \begin{cases} x/4 & 1 < x \leq 3 \\ 0 & \text{else} \end{cases}$

$A = \{X \geq 2\}$

$E[X] = \int_1^3 x f_X(x) dx = \int_1^3 x^2/4 \, dx$

$= \frac{x^3}{12}\Big|_1^3 = 26/12$




$P(A) = P(X \geq 2)$

$= \int_2^3 f_X(x) = 5/8$

same as unconditional but scaled

$f_{X|A} = \begin{cases} f_X(x)/P(A) & 2 \leq x \leq 3 \\ 0 & \text{else} \end{cases}$ $\frac{2x}{5}$

$E[X|A] = \int_2^3 x f_{X|A}(x) dx$

$= 2x^3/15 \Big|_2^3 = \frac{2}{5}(27 - 8) = 38/15$

$Y = X^2$. $E[Y] = E[X^2] = \int_1^3 x^2 f_X(x) dx$

$= \int_1^3 x^3/4 \, dx = \frac{x^4}{16}\Big|_1^3 = \frac{1}{16}(81 - 1) = 5$

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

$E[Y^2] = E[X^4] = \int_1^3 x^4 f_X(x) = 91/3$

$\Rightarrow var(Y) = 91/3 - 5^2 = 16/3$

## Exponentials

Given: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$

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

$= \int_0^x \lambda e^{-\lambda t} dt = \lambda \frac{1}{-\lambda} e^{-\lambda t}\Big|_0^x$

$= -e^{-\lambda x} - (-e^{-\lambda 0}) = -e^{-\lambda x} + 1$

$F_X(x) = \begin{cases} -e^{-\lambda x} + 1 & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$

$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$

$= \int_0^{\infty} x \lambda e^{-\lambda x} dx$

Recall integration by parts

$\int u dv = uv - \int v du$

$u = x$
$dv = \lambda e^{-\lambda x} dx$ } choose

$v = \lambda \frac{1}{-\lambda} e^{-\lambda x} dx = -e^{-\lambda x}$

$du = 1 dx$ ← calculate

$\int u dv = \cancel{-x e^{-\lambda x}\Big|_0^{\infty}} - \int_0^{\infty} -e^{-\lambda x} dx$

$e^{-\infty} \to 0$

$= -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^{\infty} = 1/\lambda$

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

$E[X^2] = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx$

$u = x^2$ $du = 2x dx$
$dv = \lambda e^{-\lambda x} dx$ $v = -e^{-\lambda x}$

$\int u dv = \underbrace{-x^2 e^{-\lambda x}\Big|_0^{\infty}}_{0} - \int -2x e^{-\lambda x} dx$

$= -\int -2x e^{-\lambda x} dx = 2 \int x e^{-\lambda x} dx$

$= \frac{2}{\lambda} \underbrace{\int x \lambda e^{-\lambda x} dx}_{E[X]} = \left(\frac{2}{\lambda}\right)\left(\frac{1}{\lambda}\right) = \frac{2}{\lambda^2}$

$\Rightarrow var(X) = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2$

Given: $X_1, X_2, X_3 \sim exp(\lambda)$ and indep

$Z = max(X_1, X_2, X_3)$ $W = min(X_1, X_2)$

$F_Z(z)$: PDF $F_Z(z) = P(Z \leq z) = P(max(X_{..}) \leq z)$

$= P(X_1 \leq z, X_2 \leq z, X_3 \leq z)$ key insight

$= P(X_1 \leq z) P(X_2 \leq z) P(X_3 \leq z)$ by independence

$= F_X(z) F_X(z) F_X(z) = (1 - e^{-\lambda z})^3$

$\Rightarrow f_Z(z) = 0$ if $z < 0$ else $3(1 - e^{-\lambda z})^3$

$F_W(w)$: PDF $F_W(w) = P(W \leq w) = P(min(X_{..}) \leq w)$ ← derivative

$= 1 - P(min(X_1, X_2) > w)$

$= 1 - P(X_1 > w, X_2 > w)$

$= 1 - P(X_1 > w) P(X_2 > w) = 1 - e^{-2\lambda w}$

# Iterated Expectations

$E[E[X|Y]] = E[X]$ Total variance

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

## Sum of Random # of independent RVs

Ex: N # stores to visit — X's iid + independent of N

$X_i$: \$ spent on store i — $Y = X_1 + X_2 + \dots + X_N$ RV

Find $E[Y]$

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

$E[Y|N=n] = E[X_1 + X_2 + X_N | N=n]$

$= E[X_1 + X_2 + \dots + X_n] = nE[X]$

$E[Y|N] = NE[X]$ constant

$E[Y] = E[NE[X]] = E[N]E[X]$

Find $var(Y)$

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

$var(E[Y|N]) = var(NE[X])$

$= (E[X])^2 var(N)$

$E[var(Y|N)] =$

$var(Y|N=n) = var(X_1 + X_2 + \dots X_n) = n var(X)$

$\Rightarrow var(Y|N) = N var(X)$

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

$var(Y) = E[N] var(X) + (E[N])^2 var(N)$

# Central Limit Theorem (CLT)

$S_n = X_1 + X_2 + \dots + X_n$ iid — $S_n \sim N(n\mu, n\sigma^2)$, $Z \sim N(0,1)$

$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ : $\lim_{n\to\infty} P(Z_n \le z) = P(Z \le z) = \Phi(z)$

CDF converges to normal CDF

Problem types: $P(S_n \le a) = b$ — Given 2 (params $n, \mu, \sigma$) Find third

Given $\mu, n$ find $b$: Package weight $X_i \sim Exp(\lambda)$, $\lambda = 1/2$

Load container w/ $n = 100$ packages — max capacity — since

$P(S_n > 210)$? $Z_n = \frac{S_n - \mu n}{\sqrt{n}\sigma}$ $\mu = \sigma = 2$

$P\left(\frac{S_n - 200}{20} > \frac{210-200}{20}\right) = 1 - P(Z < .5) = 1 - \Phi(.5)$

$= P(Z_n \ge .5) \approx P(Z \ge .5)$

Given $b, n$ find $a$: Choose capacity s.t. $P(S_n \ge a) = .05$

$.05 = P\left(\frac{S_n - 200}{20} \ge \frac{a-200}{20}\right) = 1 - \Phi\left(\frac{a-200}{20}\right) \Rightarrow \Phi\left(\frac{a-200}{20}\right) = .95$

Given $a, b$ find $n$. How large can n be so that

$P(S_n \ge 210) = .05$ $\quad .05 = 1 - \Phi\left(\frac{210-2n}{2\sqrt{n}}\right)$ — look at chart or solve

$P\left(\frac{S_n - 2n}{2\sqrt{n}} \ge \frac{210-2n}{2\sqrt{n}}\right) \Rightarrow \Phi\left(\frac{210-2n}{2\sqrt{n}}\right) = .95$

# Bayesian Inference

- Unknown $\Theta$ — Prior $P_\Theta(\theta)$
- Observation $X$ — model $P_{X|\Theta}$

Bayes: $f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x|\theta)$

$$f_{\Theta|X}(\theta, x) = \frac{f_\Theta(\theta) f_{X|\Theta}(x|\theta)}{f_X(x)}$$

Posterior

MAP: $\hat\theta_{MAP} = \arg\max_\theta P_{\Theta|X}(\theta|x)$

LMS: $\hat\theta_{LMS} = E[\Theta | X = x] = \int \theta f_{\Theta|X}(\theta|x)$

LLMS: $\hat\theta_L = E[\Theta] + \frac{cov(\Theta, X)}{var(X)}(X - E[X])$

Conditional $P(error)$: $P(\hat\theta \ne \Theta | X = x)$

Overall $P(error) = P(\hat\theta \ne \Theta) =$

$\sum_x P(\hat\theta \ne \Theta | X = x) P_X(x) =$

$\sum_\theta P(\hat\theta \ne \Theta | \Theta = \theta) P_\Theta(\theta)$

MAP: maximizes posterior prob

If discrete minimizes $P(error|x)$ and overall $P(error)$ $\quad \hat\theta_{MAP} = \arg\max_\theta f_{\Theta|X}$

## Recognizing Normal PDFs

If RV X has PDF: $c \cdot e^{-(\alpha x^2 + \beta x + \gamma)}$

w/ $\alpha > 0 \Rightarrow X \sim N\left(-\frac{\beta}{2\alpha}, \frac{1}{2\alpha}\right)$ $\quad \frac{d}{dx} = 0$

If normal: Mean = peak of PDF

$\Rightarrow$ maximize PDF $\Rightarrow$ minimize exponent

Recall: $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \underbrace{\frac{1}{\sigma\sqrt{2\pi}}}_{c} e^{-(x-\mu)^2/2\sigma^2}$$

## Multiple Observations

$\Theta \sim N(x_0, \sigma_0^2)$ $\quad W_i \sim N(0, \sigma_i^2)$

$X_i = \theta + W_i \Rightarrow f_{\Theta|X}(\theta, x) \propto \prod_{i=0}^{n} e^{-\frac{1}{2\sigma_i^2}(x_i - \theta)^2}$

$$\hat\theta_{MAP} = \hat\theta_{LMS} = \sum_{i=0}^{n} \frac{x_i}{(\sigma_i)^2} \Big/ \sum_{i=0}^{n} \frac{1}{\sigma_i^2}$$

- Posterior is normal

LMS: Minimizes Mean Squared Error (MSE)

$MSE \triangleq E[(\Theta - \hat\theta)^2]$ $\quad \hat\theta = E[\Theta]$

note that w/ no observations: $MSE = var(\theta)$

Since $\hat\theta_{LMS} = E[\theta | X = x] =$ $\quad = var(\theta | X = x)$

$\Rightarrow MSE = E[(\Theta - E[\Theta | X = x])^2] = E\left[var_{\Theta|X}\right]$

LLMS Minimizes $E[(\theta - aX - b)^2]$

If $\hat\theta_{LMS}$ is linear $\Rightarrow \hat\theta_{LMS} = \hat\theta_{LLMS}$

$\hat\theta_{LLMS} = E[\Theta] + \rho\frac{\sigma_\theta}{\sigma_X}(X - E[X])$

MSE: $var(\theta | X = x)$ — coefficient next to $x^2$ term in $f_X(x)$

w/ same $\sigma \Rightarrow MSE = \sigma^2/(n+1)$

# Stochastic Processes

↳ Sequence of Random Variables

## Bernoulli Process

Independent and identically distributed — Param

↳ inf sequence of iid Bernoulli RVs w/ p

# successes: $S_n = X_1 + X_2 + \dots X_n$ $\quad S_n \sim Bin(n, p)$

Time till 1st success: $T_1 = \min\{i : X_i = 1\}$ $\quad T_1 \sim Geo(p)$

Fresh start property: causally determined RV ← look at past only

Fresh start after time N ← Independent of N

$\Rightarrow$ Process $X_{N+1}, X_{N+2} \dots$ is Bernoulli(p)

## Time till $k^{th}$ arrival:

0 0 0 1 0 1 0 0 1 1 0 0

$Y_1$ $Y_2$ $Y_3$

$T_1$ $T_2$ $T_3$

$Y_k$: time of $k^{th}$ arrival

$T_k$: $k^{th}$ interarrival time

$Y_k = T_1 + T_2 + \dots T_k$ $\quad Y_k \sim Pascal(k, p)$

$$P_{Y_k}(t) = \begin{cases} \binom{t-1}{k-1} p^k (1-p)^{t-k} & \text{if } t \in \{k, k+1 \dots\} \\ 0 & \text{else} \end{cases}$$

## Merging

$X_t$ · · x · x · · · x $\quad Ber(p)$

independent $Z_t$ · x x · x · x · x $\quad Ber(p+q-pq)$

$Y_t$ · x · · x · x · · $\quad Ber(q)$

## Splitting

$Ber(pq)$ $\quad$ · x · · · · · ·

$Ber(p)$ $\xrightarrow{q}$ · · x · x · · · x ·

$Ber(p(1-q))$ $\xrightarrow{1-q}$ · · · · x · · · x ·

# Poisson Process

continious version of Bernoulli process

$P(k, T)$. Probability of k arrivals in any interval of duration T. $\quad X \sim Poi(\lambda)$

$P(k, T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$ $\quad \Rightarrow P_X(i) = \frac{\lambda^i e^{-\lambda}}{i!}$

$E[N_T] = \lambda T$ $\quad var(N_T) = \lambda T$

$N_T$: # arrivals in $[0, T]$: $N_T \sim Poi(\lambda T)$ — arrival rate

* Time till 1st success: $Exp(\lambda)$ $\quad P(T_1 > t) = e^{-\lambda t}$

$\Rightarrow E[\text{interarrival}] = 1/\lambda$

## Memorylessness Property

causal RV

↳ Process starting at time $\tau$ also Poisson

## Time $Y_k$ of $k^{th}$ arrival

Sum of y iid exp RVs

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$$ Erlang distribution

## Sum of Poisson

$X \sim Poi(\lambda)$ $\quad Y \sim Poi(\mu)$ $\quad X, Y$ independent

$\Rightarrow X + Y \sim Poi(\lambda + \mu)$ arrivals independent

Merging: Merging X and $Y \sim Poi(\lambda + \mu)$

* $P\left(\text{arrival in } X \mid \text{arrival in merged}\right) = \lambda/(\lambda + \mu)$

$T_1 = \min\{X_1, Y_1\}$ independent

EX: 3 bulbs. lifetime $\sim Exp(\lambda)$

1st burnout? Merge all processes $\sim Poi(3\lambda)$

First arrival. $1/3\lambda$

All burnout? Merge + Find 1st burnout: $1/3\lambda$

Merge remaining + find 1st burnout: $1/2\lambda$

Arrival in remaining: $1/\lambda \Rightarrow T = \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda}$

## Splitting

$X \sim Poi(\lambda)$ split X into Y, Z with $p = q^{n^z}$

$\Rightarrow Z \sim Poi(\lambda q)$ $\quad Y \sim Poi(\lambda(q-1))$

Z and Y independent!

# Markov Chains

$1-p$ (A) $\xrightarrow{p}$ (B) $p$ 1

- RV's (states) with transition probabilities
- $P_{ij}$: Prob of going from i to j in 1 step

$P_{ij} = P(X_{n+1} = j | X_n = i)$. Transition prob. of node = 1

Markov Property: Given current state, Past does not matter

## n-step transition probability

(i) $\xrightarrow{n \text{ steps}}$ (j)

$r_{ij}(n)$: going from $i \to j$ in n steps

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1) P_{kj} = \sum_{k=1}^{m} P_{ik} r_{kj}(n-1)$$

(# states)

## Recurrent State:

state i is recurrent if starting from i, there is always a way to return to i.

Transient State: not recurrent

Recurrent Class: collection of recurrent states communicating only w/ each other

Periodic Class: states in recurrent class that can be grouped into $d > 1$ groups so that all transitions from one group lead to next.

self loop $\Rightarrow$ not periodic

## Steady state Probabilities

If: · only 1 recurrent class and recurrent class not periodic $\Rightarrow$ $\quad \sum_{k=1}^{m} \pi_k = 1$

$$r_{ij}(n) \xrightarrow[n\to\infty]{} \pi_j = \sum_{k=1}^{m} \pi_k P_{kj}$$

steady state probability of being in state j

## Birth death processes

(1) $\xrightarrow{p}$ (2) $\xrightarrow{p}$ (3) $\xrightarrow{p}$ (4) $\xrightarrow{p}$ (5) $\xrightarrow{1-p}$ (4) $\xrightarrow{1-p}$ (3) $\xrightarrow{1-p}$ (2) $\xrightarrow{1-p}$ (1)

Note that transitions through line must be equal so: $\pi_i P_i = \pi_{i+1}(1 - P_{i+1})$

$\pi_2 = \pi_1\left(\frac{p}{1-p}\right)$ $\quad \pi_3 = \pi_1\left(\frac{p}{1-p}\right)^2$ $\quad \pi_4 = \pi_1\left(\frac{p}{1-p}\right)^4$

$\pi_1 = \left(1 - \frac{p}{1-p}\right) \Big/ \left(1 - \left(\frac{p}{1-p}\right)^4\right)$

## Absorbing Probabilities

(1) $\xrightarrow{p}$ (k) $\xrightarrow{1}$ 1

Absorbing state: Recurrent state k w/ $P_{kk} = 1$

$a_i$: Probability of ending in a given you started in i. $\quad a_i = \sum_{j=1}^{m} P_{ij} a_j$ $\quad a_a = 1$

* Can think of a recurring class as one big absorbing state.

Expected time to absorption

$v_i$: Expected # of transitions given initial state i

(3) $\xrightarrow{1-p}$ (4) $\xrightarrow{1}$ 1

(2) $\xrightarrow{p}$ (4)

$v_2 = 1 + pv_4 + (1-p)v_3$

: Can solve system of equations

If more than one absorbing state:

- Combine all absorbing states + fix branches
- Solve as above

Mean First Passage: Expected # steps to get $i \to s$ for first time $= t_i$

- Remove outgoing arcs from $s \Rightarrow s$ is now absorbing

Mean Recurrence Time: If you leave s, how long until back in s. $= t^*_s$ $\quad t_s = 0$

(s) $\xrightarrow{p}$ (q) $\xrightarrow{(q^-)}$ (3) $\xrightarrow{1}$ (s)

$t_q^* = 1 + qt_3 + xt_5 + \cancel{p}\, q^0$

**Sample Proof:** If $X, Y$ independent then

$$\mathrm{var}(XY) = (E[X])^2\mathrm{var}(Y) + (E[Y])^2\mathrm{var}(X) + \mathrm{var}(X)\mathrm{var}(Y)$$

Total variance

$$\mathrm{var}(XY) = E[\mathrm{var}(XY|X)] + \mathrm{var}(E[XY|X])$$

$$= E[X^2\mathrm{var}(Y|X)] + \mathrm{var}(XE[Y|X])$$

independent

$$= E[X^2\mathrm{var}(Y)] + \mathrm{var}(XE[Y])$$

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

$$\mathrm{var}(X) = (E[X^2]) - (E[X])^2 \Rightarrow E[X^2] = \mathrm{var}(X) + (E[X])^2$$

$$\mathrm{var}(XY) = (E[X^2] + \mathrm{var}(X))\mathrm{var}(Y) + (E[Y])^2\mathrm{var}(X)$$

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

$\square$

**Sample: LMS and MAP:** $f_\Theta(\theta) = \begin{cases} c\theta^3 & \theta \in [0,1] \\ 0 & \text{else} \end{cases}$

$Y = \theta + 3W$ $\quad W \sim N(0,1)$

observation

$$\hat{\theta}_{MAP} = \mathrm{argmax}\, f_{\Theta|Y}(\theta|y)$$

$$f_{\Theta|Y}(\theta|y) = \frac{f_\Theta(\theta) f_{Y|\Theta}(y|\theta)}{f_Y(y)} \quad I \text{ need to maximize}$$

mean $w = 0$

Since $Y = \theta + 3W$ $\quad Y \sim N(\theta, 9)$ $\quad \mathrm{var}(3W) = 3^2\mathrm{var}(W) = 9$

$$\Rightarrow f_{Y|\Theta}(y|\theta) = \frac{1}{3\sqrt{2\pi}} e^{\frac{-(y-\theta)^2}{18}}$$

constant

$\Rightarrow$ will have to maximize $\log \theta^3 - \frac{(y-\theta)^2}{18}$

$$\hat{\theta}_{LMS} = E[\Theta|Y=y] = \int_0^1 \theta f_{\Theta|Y}(\theta|y)$$

$$f_{\Theta|Y}(\theta|y) = \frac{f_\Theta(\theta) f_{Y|\Theta}(y|\theta)}{\int_0^1 f_\Theta(\theta) f_{Y|\Theta}(y|\theta)\, d\theta}$$

**Estimating with Noise:**

$$X = \Theta + W \quad N(0,1)$$

var of $W$

$$= \theta + W \sim N(\theta, 1)$$

$$f_{\Theta|X}(\theta|x) = f_\Theta(\theta) f_{X|\Theta}(x|\theta) / f_X(x)$$

$$f_{\Theta|X}(\theta|x) = \frac{1}{f_X(x)} c e^{-1/2(\theta^2)} \cdot c e^{-1/2(x-\theta)^2}$$

$$= c(x) e^{-\text{quadratic}(\theta)}$$

Maximize Quadratic: $\frac{1}{2}\theta^2 + \frac{1}{2}(x-\theta)^2$

$\theta + \theta - x = 0 \Rightarrow \theta = x/2$

with multiple observations

$$f_{\Theta|X}(\theta|x) = \frac{1}{f_X(x)} c_0 e^{-(\theta - x_0)^2/2\sigma_0^2} \cdot \prod_{i=0}^{n} c_i e^{-(x_i-\theta)^2/2\sigma_i^2}$$

$$\Rightarrow c \cdot \exp\{\text{quad}(\theta)\}$$

$$\text{quad}(\theta) = \frac{(\theta - x_0)^2}{2\sigma_0^2} + \frac{(\theta - x_1)^2}{2\sigma_0^2} + \dots + \frac{(\theta - x_n)^2}{2\sigma_n^2}$$

**LLMS Coin Bias Example**

coin w/ unknown bias $\theta$ prior $f_\Theta(\theta) \sim \text{Unif}[0,1]$

flip $n$ times $X$ = # heads

$$\hat{\theta}_L = E[\theta] + \frac{\mathrm{cov}(\theta, X)}{\mathrm{var}(X)}(X - E[X])$$

$E[\theta] = \frac{1}{2}$ $\quad \mathrm{var}(\theta) = 1/12$ since uniform $\quad P_{X|\theta} \sim \text{Binomial}(n, \theta)$

$E[X|\theta] = n\theta$ $\quad \mathrm{var}(X|\theta) = n\theta(1-\theta)$ since

$\mathrm{var}(X) = E[X^2] - (E[X])^2$

$E[X^2] = E[E[X^2|\theta]]$

$E[X^2|\theta] = \mathrm{var}(X^2|\theta) + E[X]^2$

$= n\theta(1-\theta) + n^2\theta^2$

$E[X^2] = E[n\theta(1-\theta) + n^2\theta^2] = E[n\theta + (n^2-n)\theta^2]$

$E[\theta^2] = \mathrm{var}(\theta) + E[\theta]^2 = 1/3$

$E[X^2] = E[n\theta] + E[(n^2-n)\theta^2] = \frac{n}{2} + \frac{n^2-n}{3}$

$\mathrm{var}(X) = \frac{n}{6} + \frac{n^2}{3} - \frac{n^2}{4} =$

$\mathrm{cov}(X, \theta) = E[\theta X] - E[\theta]E[X]$

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

$= E[n\theta^2] = n/3$ $\quad n \quad x+1$

**Sample: Poisson Carwash:** Cars arrive as $\sim \text{Poi}(\lambda)$

wash takes .25 hrs $\rightarrow$ carwash occupied in mean time

a) PMF of # cars arriving in 3hrs

$P(k, 3) = (3\lambda)^k e^{-3\lambda}/k!$ (from cheat sheet)

b) $P(\text{car accepted} \mid t = 1/6) = P(\text{no cars arrived in } 1/6)$

$= P(0, 1/6) = e^{-\lambda/6}$

c) PDF of $T_2$: Time until $2^{nd}$ car leaves

$T_2 = X_1 + .25 + X_2 + .25$

First arrival $\sim \text{Exp}(\lambda)$ $\quad$ Service time

$T_2 - .5 = X_1 + X_2$ $\quad$ Time of $2^{nd}$ arrival $\sim \text{Erlang}(2)$ $\quad k=2$

shift erlang by $.5 \Rightarrow t = t - .5$

$$f_{T_2}(t) = \lambda^2(t-.5)e^{\lambda(t-.5)}$$

d) carwas closes at $T \sim \text{Exp}(\mu)$. Find PMF of profit $\quad$ Each arrival $\Rightarrow \$10$

$P_1$ = car arrivals $\sim \text{Poi}(\lambda)$ $\quad P_2$ = store close $\sim \text{Poi}(\mu)$

Merge $P_1$ and $P_2 \sim \text{Poi}(\lambda + \mu) = P_3$

$P(\text{arrival in } P_3 \text{ from } P_1) = \frac{\lambda}{\lambda+\mu}$ ...from $P_2 = \frac{\mu}{\lambda+\mu}$

Consider a Bernoulli process where each trial is an arrival in $P_3$ and a success occurs if that arrival is from $P_2$.

$\Rightarrow$ # cars from $0 \rightarrow T$ = # failures before $= N_T$ success

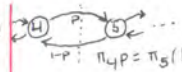
$P_{N_T}(n) = (1-p)^{n-1}p$ where $p = \frac{\mu}{\lambda+\mu}$

$M$ = profit $= 10N_T \Rightarrow N_T = \frac{M}{10}$

$P_M(m) = P_{N_T}(\frac{m}{10}) = (1-p)^{m/10}p$

**In General:** # Poison arrivals from $0 \rightarrow$ some event

- Make event a poisson process
- Merge processes
- Bernoulli process: arrival in merged = trial
  - $\hookrightarrow$ success = 1st arrival of some event
- Solve as above




$\pi_H p = \pi_S(1-p)$

Frequency balance