EEE 554: Probability and Stochastic Processes

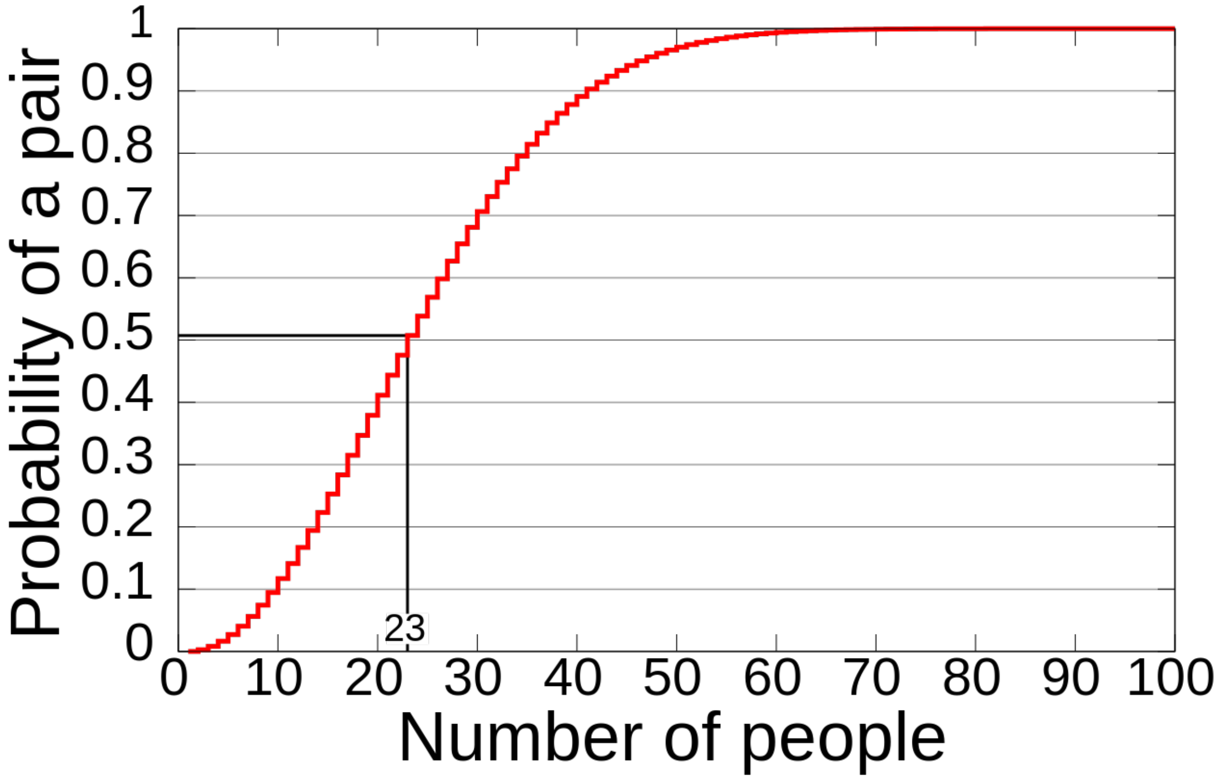
This file contains most of the slides we will be going through in Spring 2023. For all the details of what we went through, please check out the recorded lectures.

Example: Birthday Problem

- Birthday problem: Given *N* people in a room, what is the probability that 2 or more people will have the same birthday?
- 1 – "Probability that all *N* people have different bdays"

$$1 - \frac{365 \times 364 \times \cdots \times (365 - N + 1)}{365^N}$$

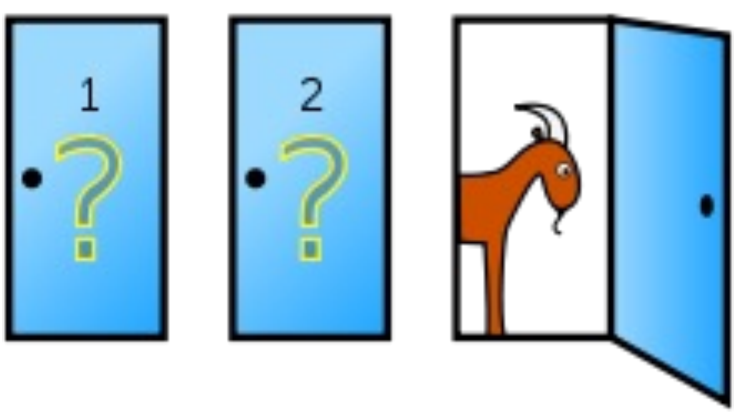
Example: Birthday Problem



http://en.wikipedia.org/wiki/Birthday_problem

Example: Monty Hall Problem

- Game show setting; 3 doors: 2 Goats, 1 Car
- You chose one door. Host opens one of the doors that you didn't pick ***and*** has a goat.
- You have the option to stick with your choice or switch
- What should you do??



http://en.wikipedia.org/wiki/Monty_Hall_problem

Example: Monty Hall Problem

- Assume you are committed to switching.
- If you had selected a C initially, then switching will lead to a G (why?)
- If you had selected a G initially, the switching will lead to a C.
- Since selecting a G initially is more likely (2/3 prob), then switching is the best policy
- Conclusion: Switching is twice more likely to win you a car than sticking with your original choice!
- The extra info that the host provides is useful!

Example: The Hat Problem

- *n* people throw their hats in a box and pick one at random
- How many people will get their own hats back?
- It is clearly a number in $\{1,2,\ldots,n\}$; Let's call this number X

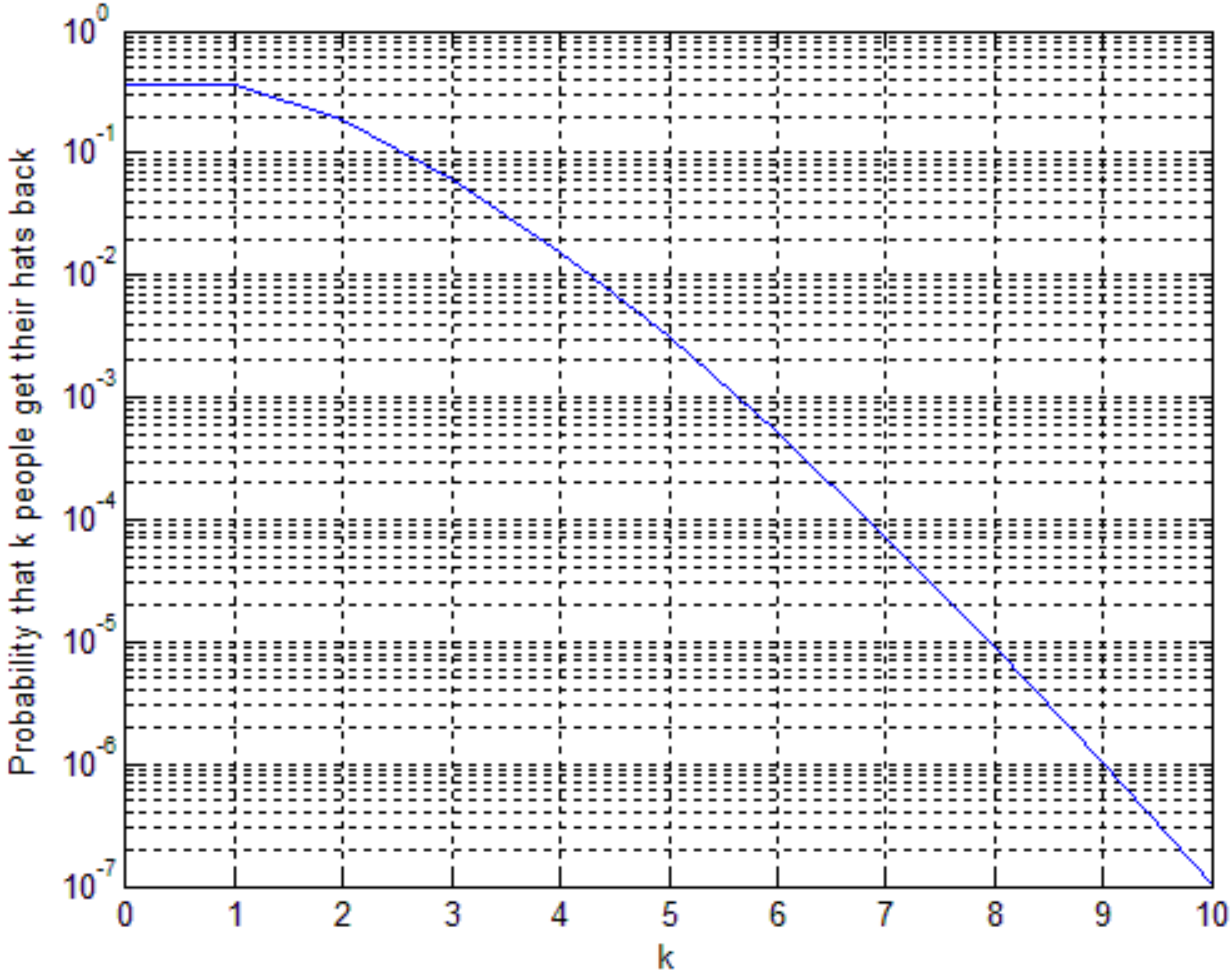
- Some interesting facts:

$$P[X = n] = 1/n!$$

$$P[X = 0] = \sum_{j=0}^{n} \frac{(-1)^j}{j!} \approx 1/e, \quad n \ \text{large}$$

$$E[X] = 1$$

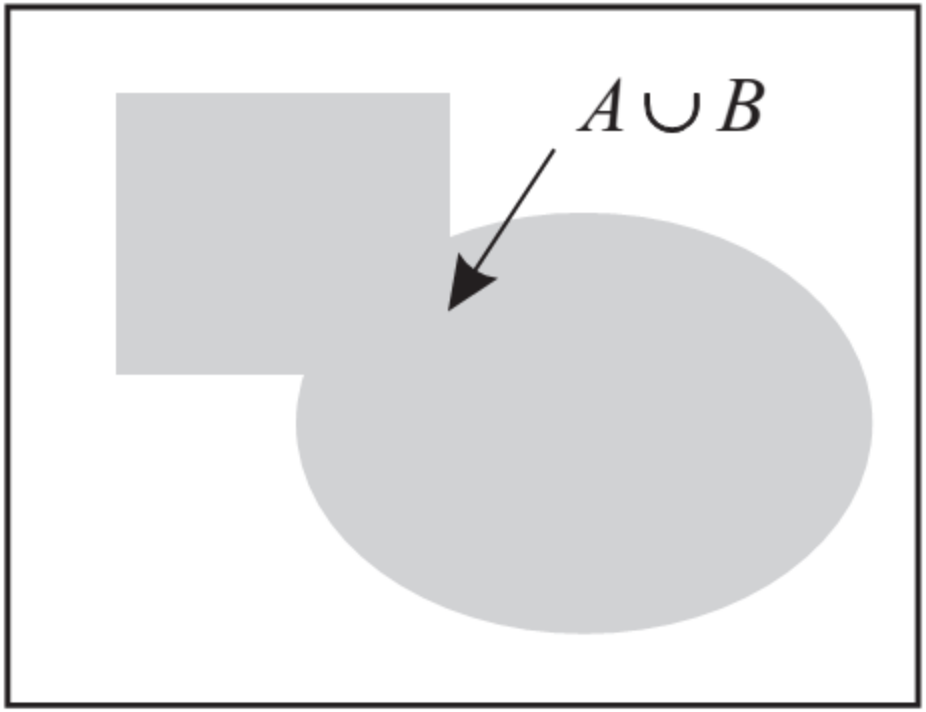
Example: The Hat Problem



Review of Set Theory

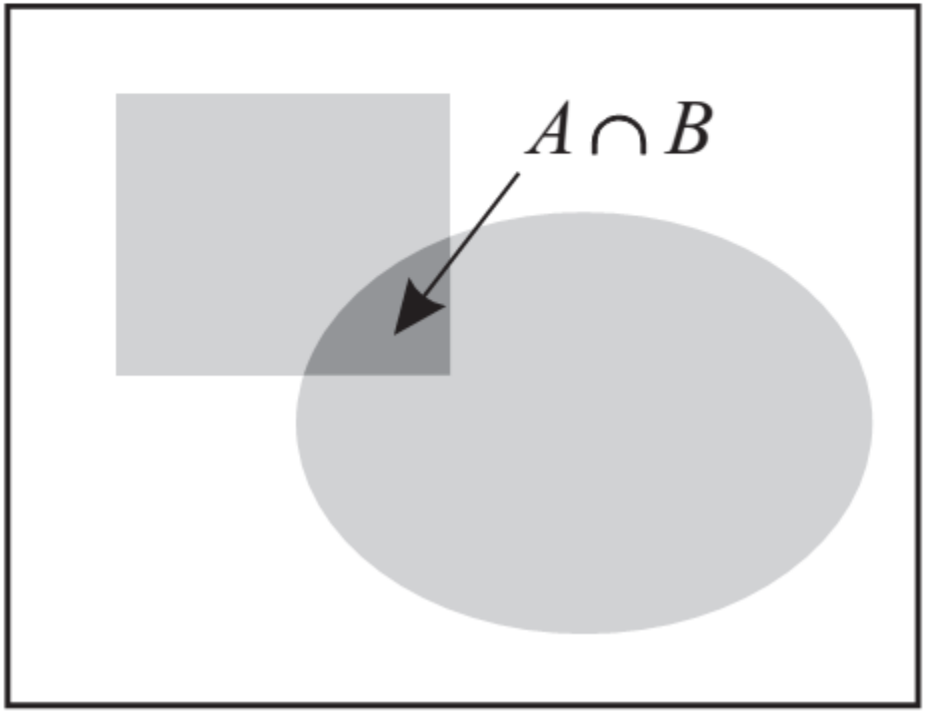
- Probability theory is grounded in set theory
- Union
- Intersection
- Compliment
- De Morgan's Law

Set Union



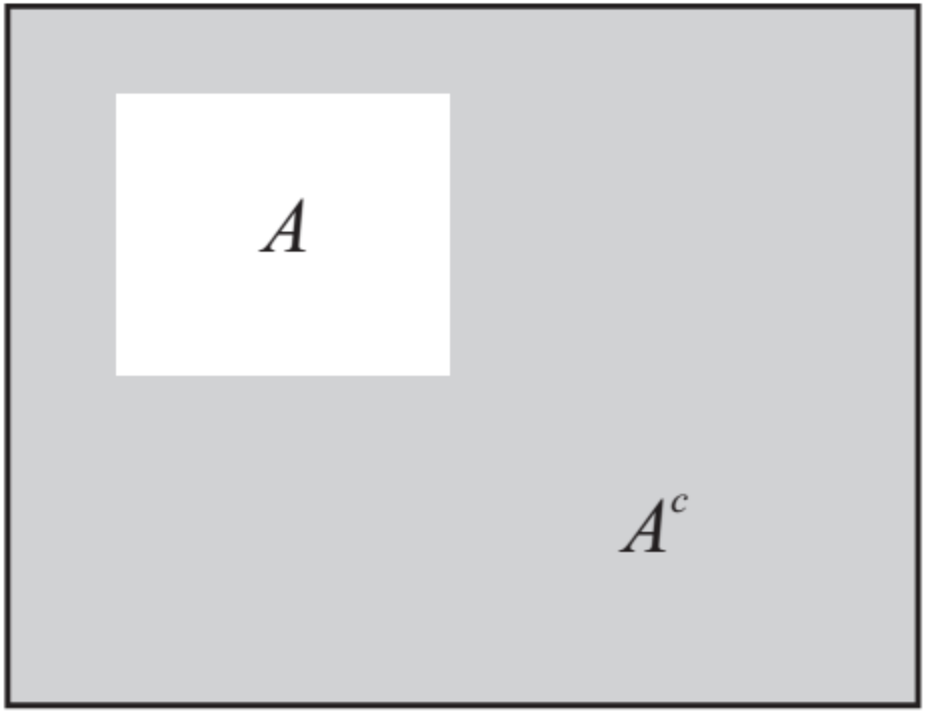
$x \in A \cup B$ if and only if $x \in A$ or $x \in B$

Set Intersection



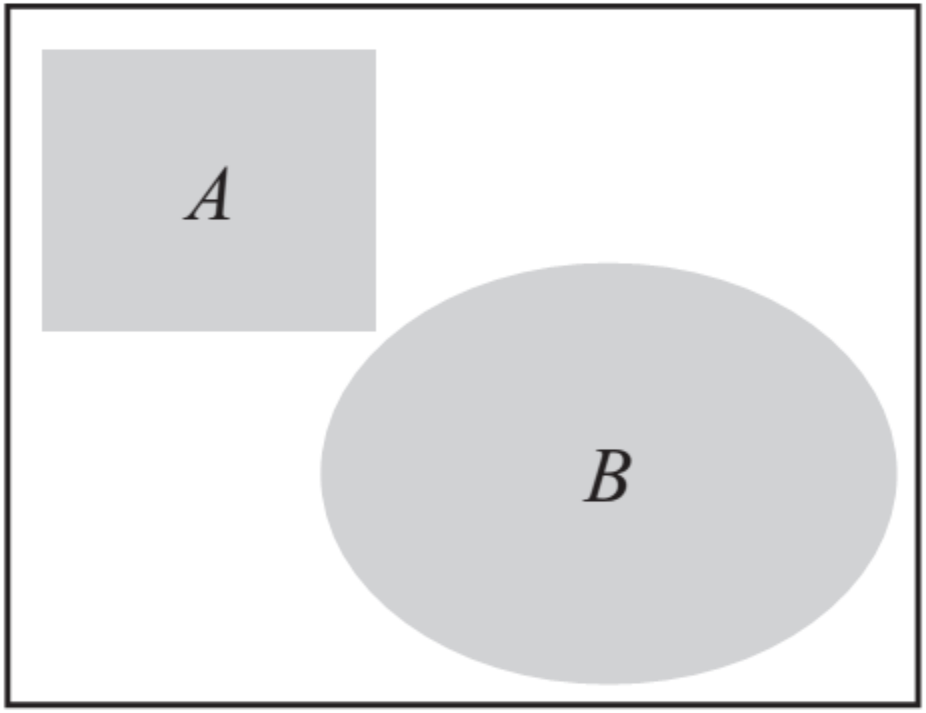
$x \in A \cap B$ if and only if $x \in A$ and $x \in B$

Set Compliment



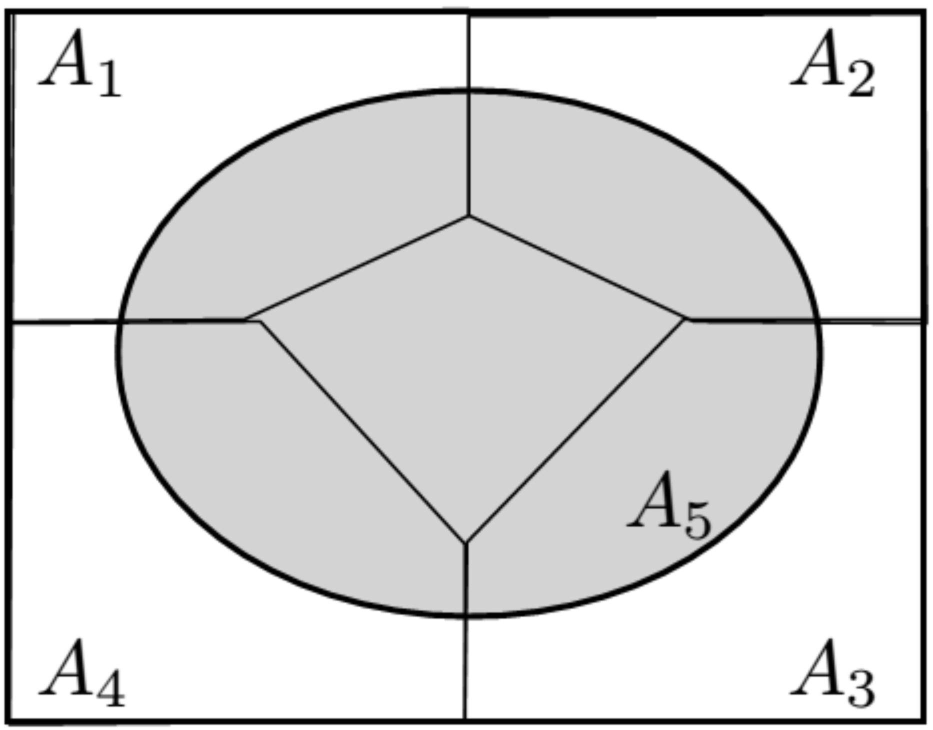
$x \in A^c$ if and only if $x \notin A$

Mutually Exclusive (Disjoint)



$$A \cap B = \varnothing$$

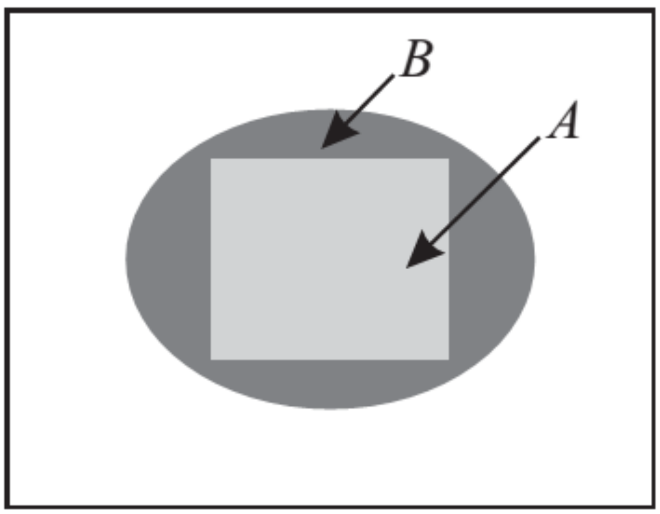
Collectively Exhaustive



$$A_1 \cup A_2 \cup \cdots \cup A_n = S$$

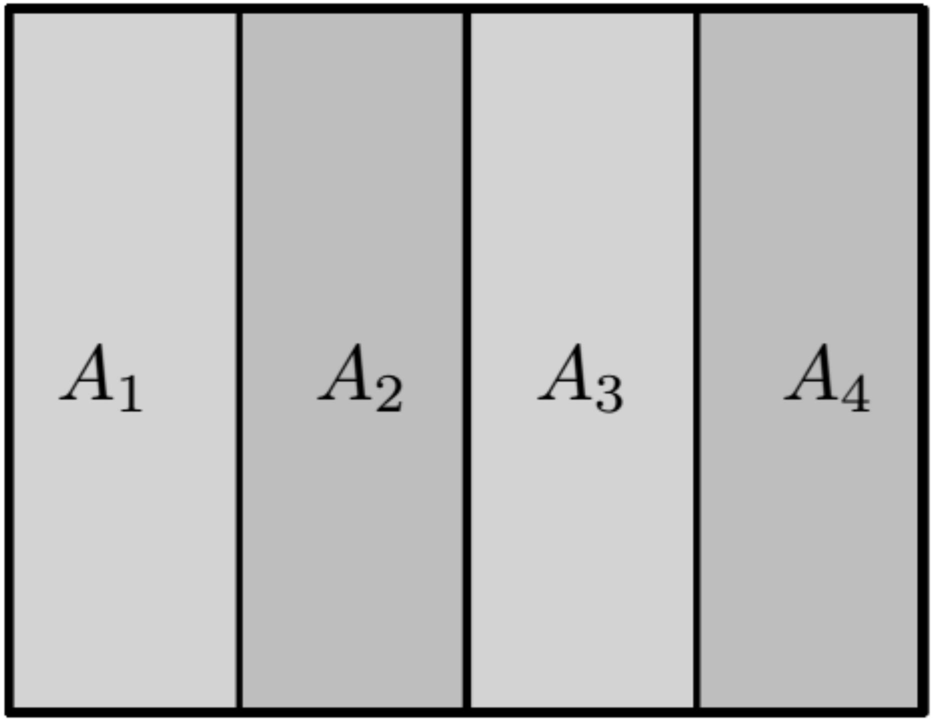
Other Basic Definitions

$A \subset B$ means if $x \in A$ then $x \in B$



$A = B$ if and only if $B \subset A$ and $A \subset B$

Partition



- Both mutually exclusive AND collectively exhaustive

De Morgan's Law

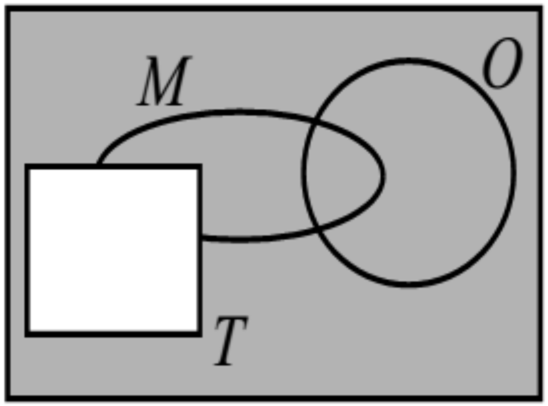
$$(A \cup B)^c = A^c \cap B^c$$

There are two parts to the proof:

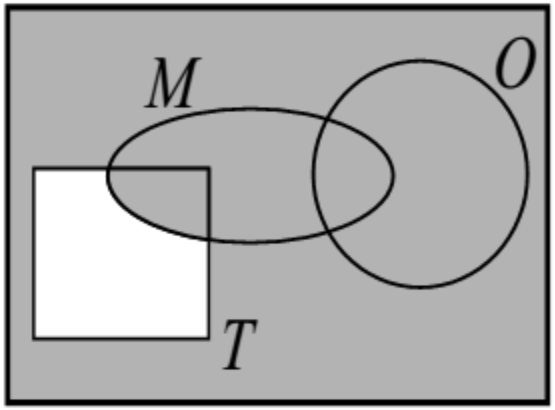
- To show $(A \cup B)^c \subset A^c \cap B^c$, suppose $x \in (A \cup B)^c$. That implies $x \notin A \cup B$. Hence, $x \notin A$ and $x \notin B$, which together imply $x \in A^c$ and $x \in B^c$. That is, $x \in A^c \cap B^c$.

- To show $A^c \cap B^c \subset (A \cup B)^c$, suppose $x \in A^c \cap B^c$. In this case, $x \in A^c$ and $x \in B^c$. Equivalently, $x \notin A$ and $x \notin B$ so that $x \notin A \cup B$. Hence, $x \in (A \cup B)^c$.

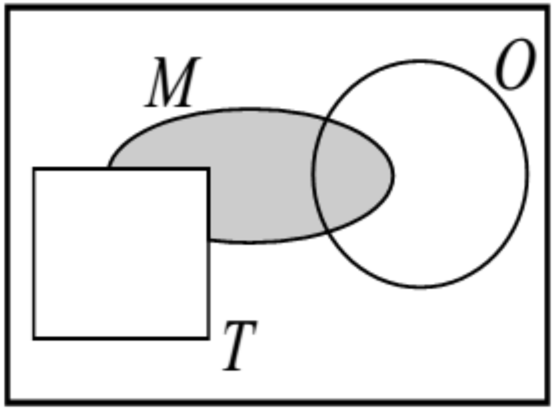
Example



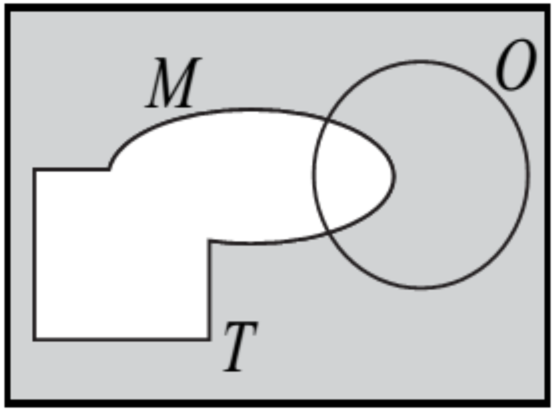
(a) $N = T^c$



(b) $N \cup M$



(c) $N \cap M$



(d) $T^c \cap M^c$

Experiment

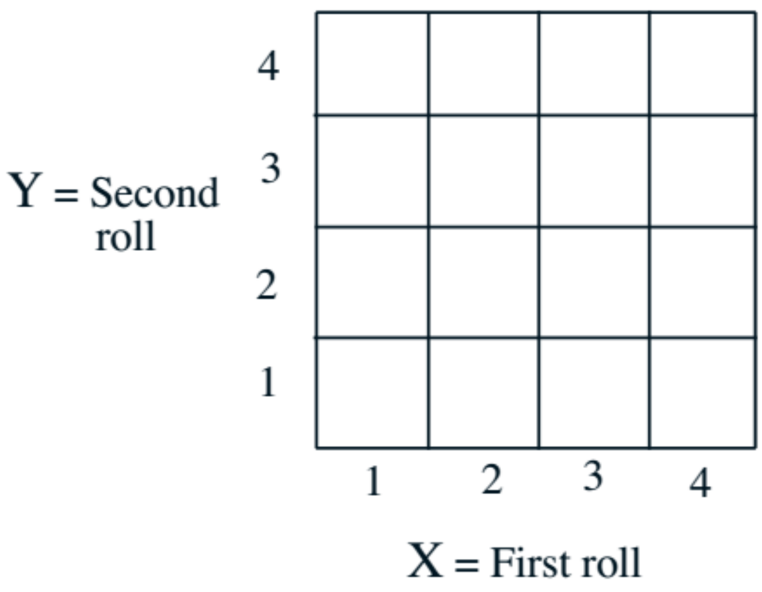
- A repeatable procedure that produces potentially different outcomes at every trial
- Set of ***all*** outcomes is the sample space
- *Events* are sets of outcomes
- Events are subsets of the sample space
- Probabilities will be assigned to events

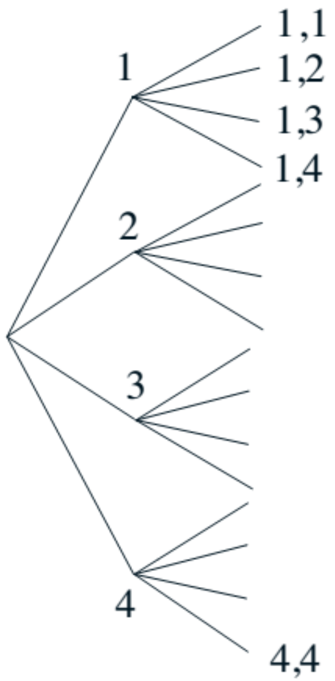
More on Sample Space

- “List” (set) of possible outcomes
- Denoted by S or Ω
- List must be:
  - Mutually exclusive
  - Collectively exhaustive
- Art: to be at the “right” granularity
- Do you just care about even/odd in dice?

Sample Space: Discrete Example

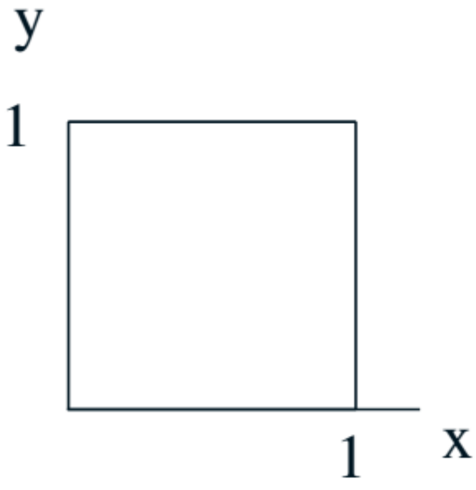
- Two rolls of a tetrahedral die
- Sample space vs. sequential description





Sample Space: Continuous Example

$$\Omega = \{(x, y) \mid \ 0 \leq x, y \leq 1\}$$



Probability Space

- Probability space consists of
    1. Sample space
    2. A collection of subsets of sample space called events
    3. Probability assignments to each event

- Events are subsets of sample space
- Collection of events satisfy certain requrements
- E.g. union of two events is an event; same with intersection
- There are other technical requirements
- How probabilities are assigned to events is subject to rules (axioms)

Axioms of Probability

- **Event:** a subset of the sample space
- Probability is assigned to events

Axioms:

1. **Nonnegativity:** $\mathbf{P}(A) \geq 0$
2. **Normalization:** $\mathbf{P}(\Omega) = 1$
3. **Additivity:** If $A \cap B = \emptyset$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$

- $$\begin{aligned} \mathbf{P}(\{s_1, s_2, \ldots, s_k\}) &= \mathbf{P}(\{s_1\}) + \cdots + \mathbf{P}(\{s_k\}) \\ &= \mathbf{P}(s_1) + \cdots + \mathbf{P}(s_k) \end{aligned}$$

- Axiom 3 needs strengthening
- Do weird sets have probabilities?

Example

- Experiment: single dice roll

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Outcomes are just 1,2,3,4,5,6.
- Events are the set of all subsets
- Closed under unions and intersections
- There are 64 events
- The probability of an event = (# of elements in the event) x 1/6
- E.g., $\mathbf{P}(\{1, 2, 4\}) = 1/2$

Alternate Example

- Experiment: Single dice roll (same as before)

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

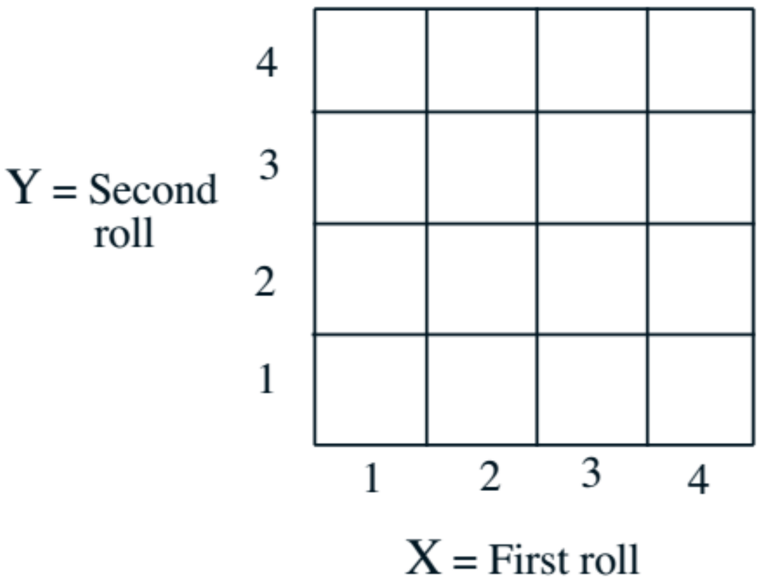
- There are only 4 events: (different than before)

$$\emptyset, \{2, 4, 6\}, \{1, 3, 5\}, \Omega$$

- Closed under unions and intersections
- Tells if outcome is even or odd
- The probability of an event = (# of elements in the event) x 1/6

Discrete Example Cont'

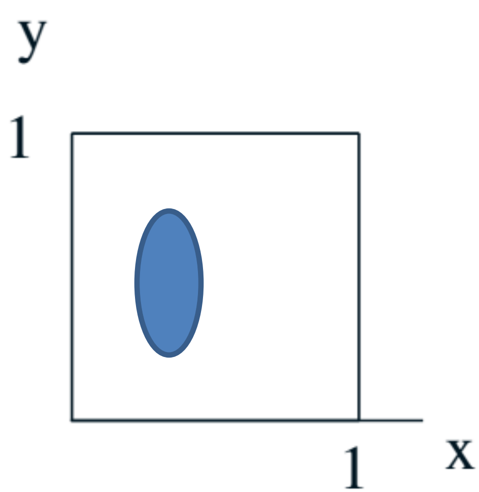
- Recall two rolls of a tetrahedral die



- Events are collections of boxes
- Probabilities of events = # of boxes / 16

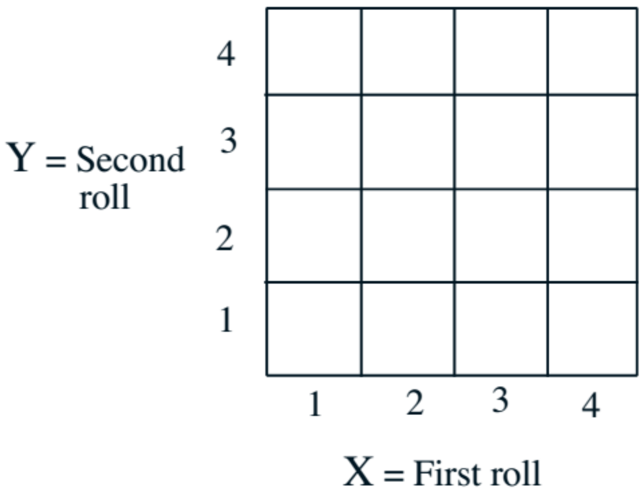
Continuous Example Cont'

$$\Omega = \{(x, y) \mid \ 0 \leq x, y \leq 1\}$$



- Events are regions in the box
- Probabilities of events = Area of region

Finite Sample Space Example



- Let every possible outcome have probability 1/16
  - $\mathrm{P}((X, Y)$ is $(1,1)$ or $(1,2)) =$
  - $\mathrm{P}(\{X = 1\}) =$
  - $\mathrm{P}(X + Y$ is odd$) =$
  - $\mathrm{P}(\min(X, Y) = 2) =$

Discrete Uniform Law

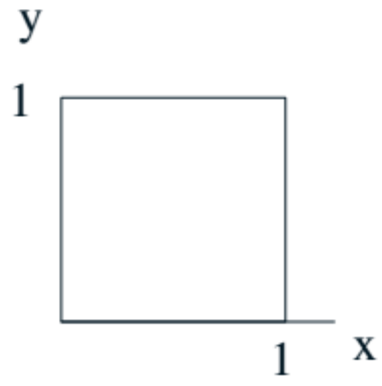
- Let all outcomes be equally likely
- Then,

$$\mathbf{P}(A) = \frac{\text{number of elements of } A}{\text{total number of sample points}}$$

- Computing probabilities $\equiv$ counting
- Defines fair coins, fair dice, well-shuffled decks

Continuous Uniform Law

- Two "random" numbers in $[0,1]$.



- **Uniform** law: Probability = Area
  - $\mathrm{P}(X+Y \le 1/2) = ?$
  - $\mathrm{P}(\,(X,Y) = (0.5, 0.3)\,)$

Geometric Sum Formula

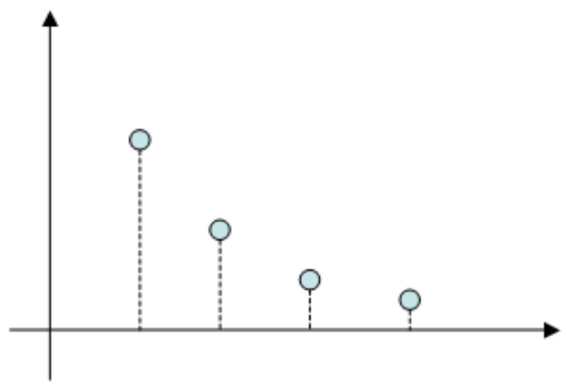
$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \quad , |r| < 1$$

- Will be useful throughout the class
- What happens when $|r| \geq 1$?
- What happens when we differentiate both sides wrt *r* ?
- Alternative way:

$$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r} \quad , |r| < 1$$

Countable Infinity

- Sample space: $\{1, 2, \ldots\}$
  - We are given $\mathbf{P}(n) = 2^{-n}$, $n = 1, 2, \ldots$
  - Find $\mathbf{P}(\text{outcome is even})$



$$\mathbf{P}(\{2,4,6,\ldots\}) = \mathbf{P}(2) + \mathbf{P}(4) + \cdots = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots = \frac{1}{3}$$

- Countable additivity axiom (needed for this calculation):
  If $A_1, A_2, \ldots$ are disjoint events, then:

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots$$

Countable Infinity Cont'

- Can verify $\sum_{n=0}^{\infty} \mathbf{P}(n) = 1$ using the geometric sum formula (*r*=1/2)

- Even outcome probability can be verified using same formula (*r*=1/4)

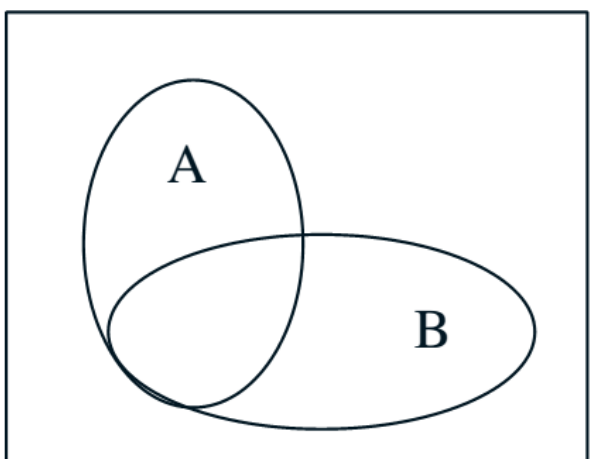
Conditional Probability

- $\mathbf{P}(A \mid B) =$ probability of A,
  given that B occurred
  - B is our new universe

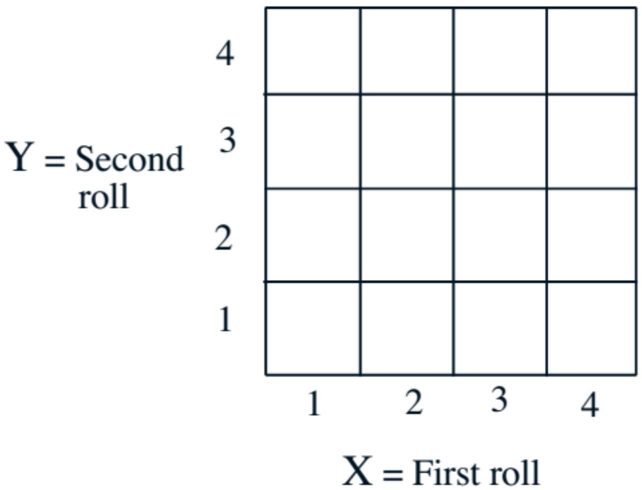
- **Definition:** Assuming $\mathbf{P}(B) \neq 0$,

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

$\mathbf{P}(A \mid B)$ undefined if $\mathbf{P}(B) = 0$



Die Roll Example

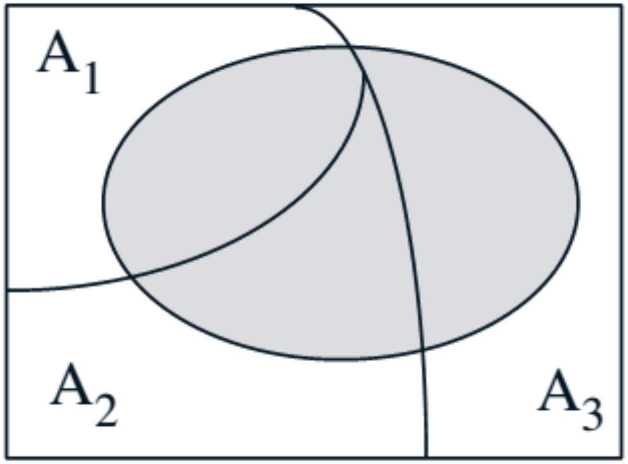


- Let B be the event: $\min(X, Y) = 2$
- Let $M = \max(X, Y)$
- $\mathbf{P}(M = 1 \mid B) =$
- $\mathbf{P}(M = 2 \mid B) =$

Total Probability Theorem

- Partition of sample space into A_1, A_2, A_3
- Have $\mathbf{P}(B \mid A_i)$, for every i

$$\begin{aligned}\mathbf{P}(B) = \quad & \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) \\ + \quad & \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) \\ + \quad & \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)\end{aligned}$$

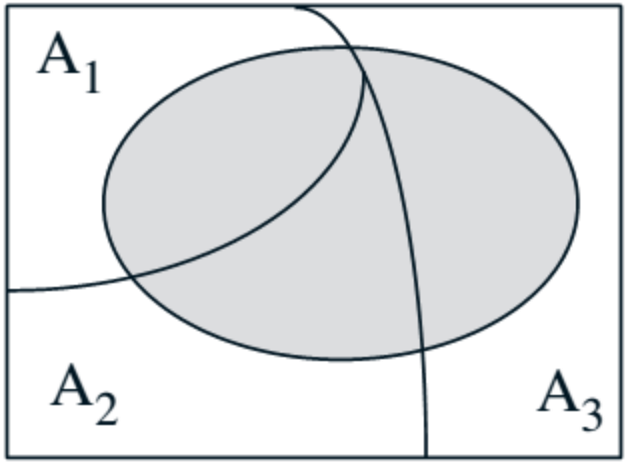


Bayes' Rule

- Systematic approach for incorporating new evidence
- Thomas Bayes (1701-1761)

- "Prior" probabilities $\mathbf{P}(A_i)$
  - initial "beliefs"
- We know $\mathbf{P}(B \mid A_i)$ for each i
- Wish to compute $\mathbf{P}(A_i \mid B)$
  - revise "beliefs", given that B occurred

Bayes' Rule Cont'



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

$$= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)}$$

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

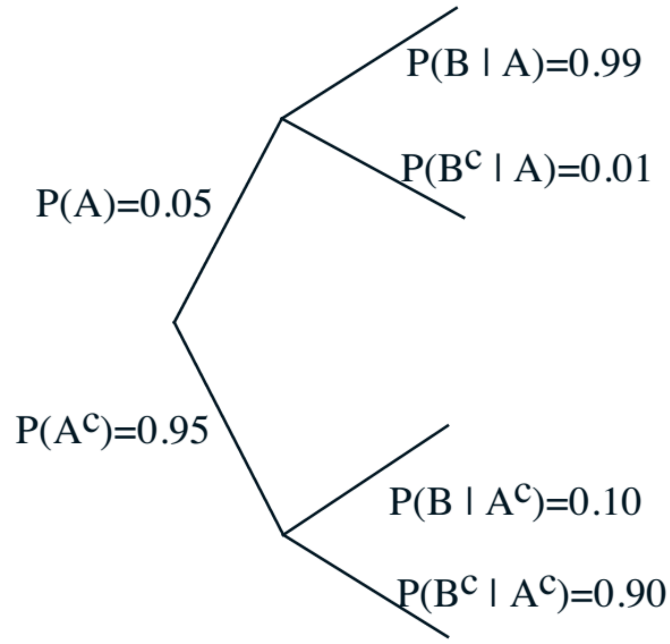
Models Based on Cond. Prob.

- Event A: Airplane is flying above
  Event B: Something registers on radar screen

$\mathbf{P}(A \cap B) =$

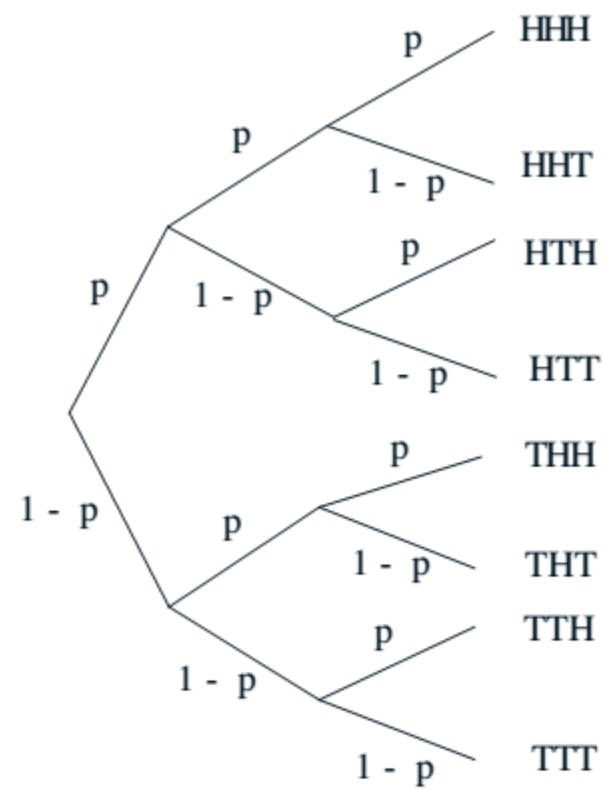
$\mathbf{P}(B) =$

$\mathbf{P}(A \mid B) =$



Sequence of Coin Tosses

- 3 tosses of a biased coin: $\mathbf{P}(H) = p$, $\mathbf{P}(T) = 1 - p$



$\mathbf{P}(THT) =$

$\mathbf{P}(\text{1 head}) =$

$\mathbf{P}(\text{first toss is H} \mid \text{1 head}) =$

Review

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}, \qquad \text{assuming } \mathbf{P}(B) > 0$$

- Multiplication rule:

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

- Total probability theorem:

$$\mathbf{P}(B) = \mathbf{P}(A)\mathbf{P}(B \mid A) + \mathbf{P}(A^c)\mathbf{P}(B \mid A^c)$$

- Bayes rule:

$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)}$$

Independence

- **"Defn:"** $\mathbf{P}(B \mid A) = \mathbf{P}(B)$
  - "occurrence of A
    provides no information
    about B's occurrence"

- Recall that $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B \mid A)$

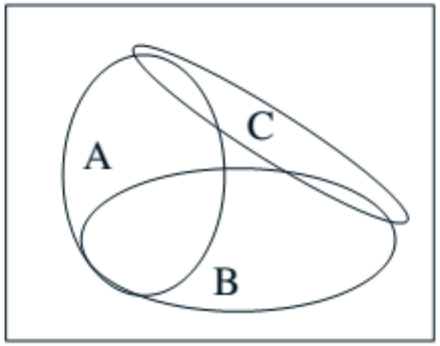
- **Defn:** $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$

- Symmetric with respect to A and B
  - applies even if $\mathbf{P}(A) = 0$
  - implies $\mathbf{P}(A \mid B) = \mathbf{P}(A)$

Conditioning May Ruin Independence

- Conditional independence, given C, is defined as independence under probability law $\mathbf{P}(\cdot \mid C)$

- Assume A and B are independent



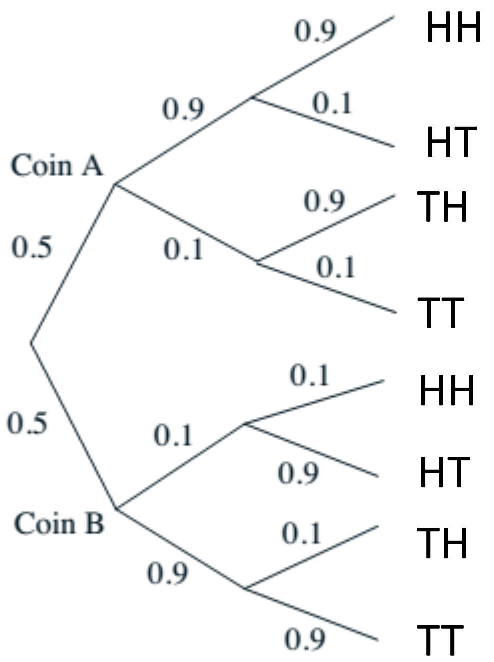
- If we are told that C occurred, are A and B independent?

Conditioning and Independence

- In contrast to prev. slide, removing conditioning may ruin independence

- Two unfair coins, A and B:
  $\mathbf{P}(H \mid \text{coin } A) = 0.9$, $\mathbf{P}(H \mid \text{coin } B) = 0.1$
  choose either coin with equal probability

- Given selected coin, flips assumed indep.



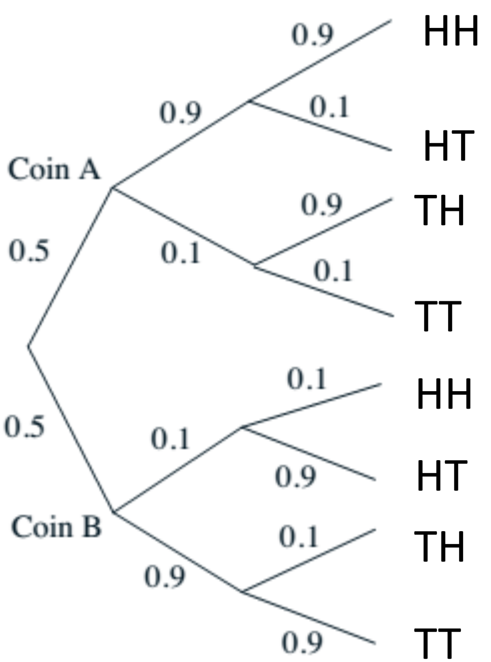
Conditioning and Independence Cont'

- Removing conditioning may ruin independence

- Recall given coin, flips are independent

- If we don't know which coin, are flips indep?

- No! Because, previous trials have info about which coin is selected.



Independence of a Collection of Events

- Intuitive definition:
  Information on some of the events tells us nothing about probabilities related to the remaining events
  - E.g.:

$$\mathbf{P}(A_1 \cap (A_2^c \cup A_3) \mid A_5 \cap A_6^c) = \mathbf{P}(A_1 \cap (A_2^c \cup A_3))$$

- Mathematical definition:
  Events $A_1, A_2, \ldots, A_n$
  are called **independent** if:

$$\mathbf{P}(A_i \cap A_j \cap \cdots \cap A_q) = \mathbf{P}(A_i)\mathbf{P}(A_j) \cdots \mathbf{P}(A_q)$$

  for any distinct indices $i, j, \ldots, q$,
  (chosen from $\{1, \ldots, n\}$)

Pairwise Independence

- Two independent fair coin tosses
  - A: First toss is H
  - B: Second toss is H
  - $\mathbf{P}(A) = \mathbf{P}(B) = 1/2$

HH	HT
TH	TT

  - C: First and second toss give same result
  - $\mathbf{P}(C) =$
  - $\mathbf{P}(C \cap A) =$
  - $\mathbf{P}(A \cap B \cap C) =$
  - $\mathbf{P}(C \mid A \cap B) =$
- Pairwise independence **does not** imply independence

Counting!

- Principles of counting
- Many examples
  - permutations
  - k-permutations
  - combinations
  - partitions
- Binomial probabilities

Discrete Uniform Law

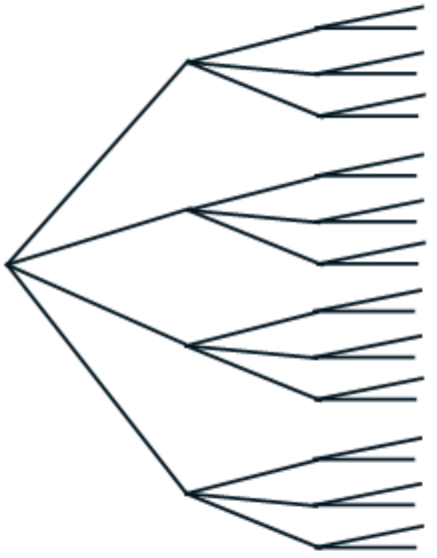
- Let all sample points be equally likely
- Then,

$$\mathbf{P}(A) = \frac{\text{number of elements of } A}{\text{total number of sample points}} = \frac{|A|}{|\Omega|}$$

- Just count...

Basics

- r stages
- n_i choices at stage i
- Number of choices is: $n_1 n_2 \cdots n_r$
- Number of license plates with 3 letters and 4 digits $=$
- ... if repetition is prohibited $=$
- **Permutations:** Number of ways of ordering n elements is:
- Number of subsets of $\{1, \ldots, n\}$ $=$



Example

- Probability that six rolls of a six-sided die all give different numbers?
  - Number of outcomes that make the event happen:
  - Number of elements in the sample space:
  - Answer: 0.0154
- Don’t bet your money on it!

Combinations

- $\binom{n}{k}$: number of k-element subsets of a given n-element set
- Two ways of constructing an ordered sequence of k **distinct** items:
  - Choose the k items one at a time: $n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!}$ choices
  - Choose k items, then order them ($k!$ possible orders)

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

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

Sum of Combinations

$$\sum_{k=0}^{n} \binom{n}{k} = \; 2^n$$

- Proof: Counting bit strings of length *n*
- *kth* term of the sum = # of length *n* bit strings with *k* ones
    - Both sides equal to total # of length *n* bit strings

Examples

- How many binary sequences of length 10 are there?

- How many 4-letter words can letters A-Z produce?

Splitting Deck into 4; one Ace Each

- 52-card deck, dealt to 4 players
- Find $\mathbf{P}$(each gets an ace)
- Outcome: a partition of the 52 cards
  - number of outcomes:

$$\frac{52!}{13!\,13!\,13!\,13!}$$

- Count number of ways of distributing the four aces: $4 \cdot 3 \cdot 2$
- Count number of ways of dealing the remaining 48 cards

$$\frac{48!}{12!\,12!\,12!\,12!}$$

Answer:

$$\frac{4 \cdot 3 \cdot 2 \dfrac{48!}{12!\,12!\,12!\,12!}}{\dfrac{52!}{13!\,13!\,13!\,13!}} = 10.5\%$$

Binomial Probabilities

- n independent coin tosses
  - $\mathbf{P}(H) = p$

- $\mathbf{P}(HTTHHH) =$

- $\mathbf{P}(\text{sequence}) = p^{\#\text{ heads}}(1-p)^{\#\text{ tails}}$

$$\mathbf{P}(k \text{ heads}) = \sum_{k\text{-head seq.}} \mathbf{P}(\text{seq.})$$

$$= (\#\text{ of } k\text{-head seqs.}) \cdot p^k(1-p)^{n-k}$$

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

Example

- event B: 3 out of 10 tosses were "heads" .
  - Given that B occurred,
    what is the (conditional) probability
    that the first 2 tosses were heads?

- All outcomes in set B are equally likely:
  probability $p^3(1-p)^7$
  - Conditional probability law is uniform

- Number of outcomes in B:

- Out of the outcomes in B,
  how many start with HH?

Example (Repetition Code)

- To communicate 1 we send 111
- To communicate 0 we send 000
- We receive 3 (possibly toggled) bits
- Bits toggle with probability q independently
- Receiver does majority combining
- What is the probability that receiver is fooled?

Random Variables

- An assignment of a value (number) to every possible outcome

- Mathematically: A function from the sample space Ω to the real numbers
  - discrete or continuous values

- Can have several random variables defined on the same sample space

- Notation:
  - random variable X
  - numerical value x

Probability Mass Function

$$\begin{aligned} p_X(x) &= \mathbf{P}(X = x) \\ &= \mathbf{P}(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\}) \end{aligned}$$

- $p_X(x) \geq 0 \qquad \sum_x p_X(x) = 1$

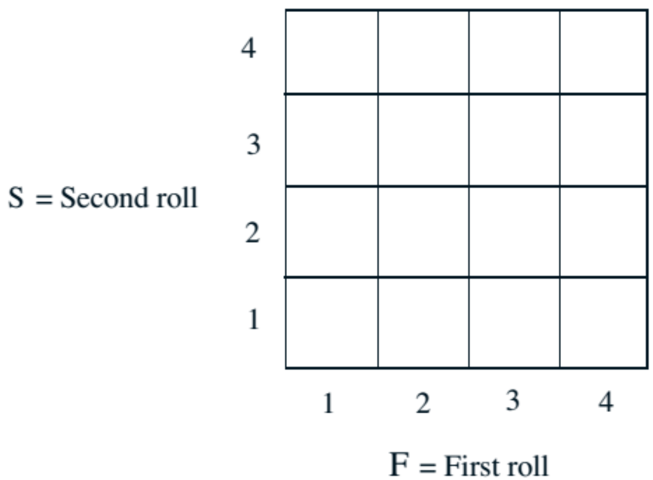
- **Example:** X=number of coin tosses until first head
  - assume independent tosses, $\mathbf{P}(H) = p > 0$

$$\begin{aligned} p_X(k) &= \mathbf{P}(X = k) \\ &= \mathbf{P}(TT \cdots TH) \\ &= (1-p)^{k-1}p, \qquad k = 1, 2, \ldots \end{aligned}$$

  - **geometric** PMF

Example

F: outcome of first throw
S: outcome of second throw
$X = \min(F, S)$



$$p_X(2) =$$

Bernoulli PMF

- Bernoulli(*p*) RV takes on values 0 or 1

$$p_X(x) = \begin{cases} 1-p & x=0, \\ p & x=1, \\ 0 & \textit{otherwise}, \end{cases} \qquad 0 < p < 1$$

- Same RV regardless of experiments producing
  - {0,1}
  - {H,T}
  - {Accept,Reject}

Binomial PMF

- *n* indep. coin flips, *p* is prob. of success; how many successes?
- Binomial (*n,p*) RV takes on values {0,1,…,*n*}

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad 0 < p < 1$$

- Can be viewed as a sum of *n* independent Bernoulli RVs

Binomial PMF Cont'

- X: number of heads in n independent coin tosses
- $\mathbf{P}(H) = p$
- Let $n = 4$

$$\begin{aligned} p_X(2) &= \mathbf{P}(HHTT) + \mathbf{P}(HTHT) + \mathbf{P}(HTTH) \\ &\quad + \mathbf{P}(THHT) + \mathbf{P}(THTH) + \mathbf{P}(TTHH) \\ &= 6p^2(1-p)^2 \\ &= \binom{4}{2} p^2 (1-p)^2 \end{aligned}$$

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$

Recall Geometric PMF

- X: number of independent coin tosses until first head

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

Pascal PMF

- Let *L* be the number of trials until *k* rejects
- "reject" same as "heads"
- *p* is probability of reject in each trial
- Generalization of geometric PMF

$$\mathsf{P}\left[L = l\right] = \mathsf{P}\left[\underbrace{k-1 \text{ rejects in } l-1 \text{ attempts}}_{A}, \underbrace{\text{reject on attempt } l}_{B}\right]$$

$$\mathsf{P}\left[A\right] = \binom{l-1}{k-1} p^{k-1} (1-p)^{l-1-(k-1)}.$$

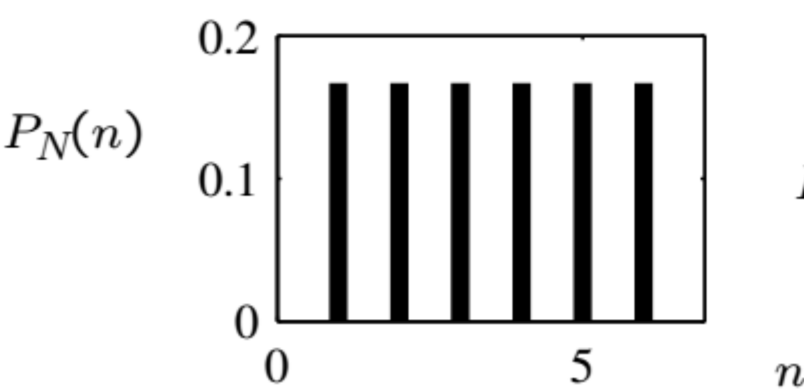
$$P_L(l) = \mathsf{P}\left[L = l\right] = \mathsf{P}\left[A\right]\mathsf{P}\left[B\right] = \binom{l-1}{k-1} p^{k} (1-p)^{l-k}$$

Discrete Uniform RV

- Equiprobable on consecutive integers *k, k*+1, …,*l*

$$P_X(x) = \begin{cases} 1/(l-k+1) & x = k, k+1, k+2, \ldots, l \\ 0 & \textit{otherwise} \end{cases}$$

- Roll a fair die and observe the number of points



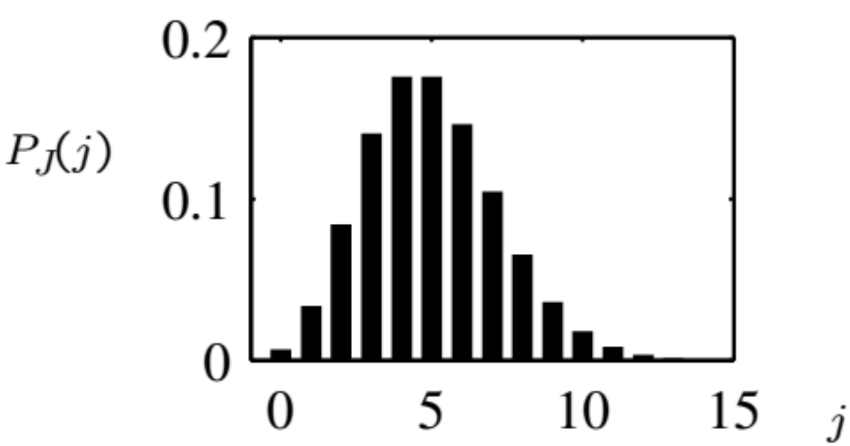
$$P_N(n) = \begin{cases} 1/6 & n = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise.} \end{cases}$$

Poisson PMF

- Used to model number of random arrivals

$$P_X(x) = \begin{cases} \alpha^x e^{-\alpha}/x! & x = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases}$$

- Example: Calls arrive randomly at 0.25 calls/sec
- The PMF of number of calls in a 20 sec. interval?



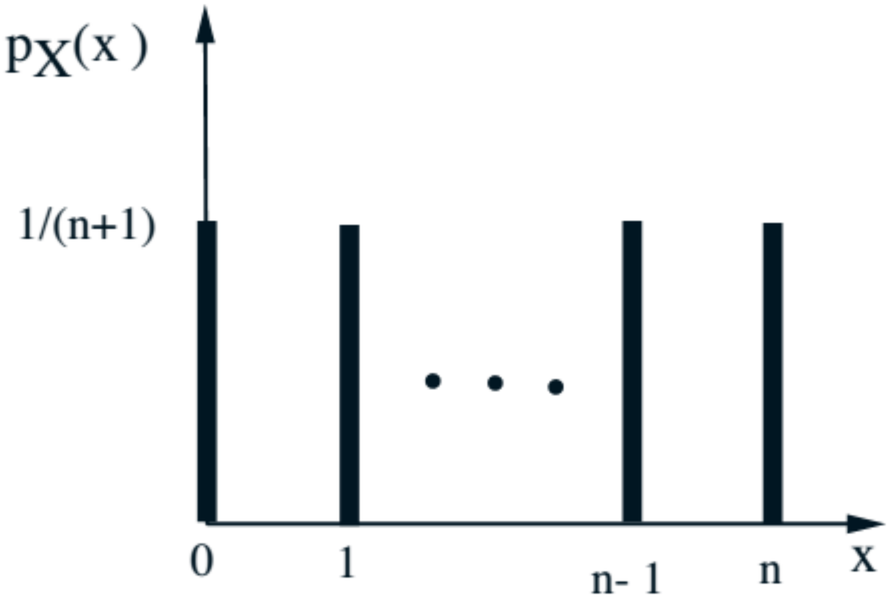
$$P_J(j) = \begin{cases} 5^j e^{-5}/j! & j = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Expectation

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

- AKA “mean” of the RV or the distribution
- Interpretations
  - Center of gravity of the PMF
  - Average in large number of repetitions of the experiment

Example: Uniform PMF



$$\mathbf{E}[X] = 0 \times \frac{1}{n+1} + 1 \times \frac{1}{n+1} + \cdots + n \times \frac{1}{n+1} =$$

Example: Geometric PMF

$$P_X(x) = \begin{cases} pq^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \qquad q = 1 - p$$

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

Pulling the *p* out of the sum and using math identity

$$\mathsf{E}[X] = p \sum_{x=1}^{\infty} xq^{x-1} = \frac{p}{q} \sum_{x=1}^{\infty} xq^{x} = \frac{p}{q} \frac{q}{1-q^2} = \frac{p}{p^2} = \frac{1}{p}$$

Math identity can be derived by differentiating the geometric sum formula

Example: Poisson PMF

$$P_X(x) = \begin{cases} \alpha^x e^{-\alpha}/x! & x = 0, 1, 2, \dots, \\ 0 & \textit{otherwise}, \end{cases}$$

$$\mathsf{E}[X] = \sum_{x=0}^{\infty} x P_X(x) = \sum_{x=0}^{\infty} x \frac{\alpha^x}{x!} e^{-\alpha}$$

$$= \alpha \sum_{x=1}^{\infty} \frac{\alpha^{x-1}}{(x-1)!} e^{-\alpha}$$

$$= \alpha \underbrace{\sum_{l=0}^{\infty} \frac{\alpha^l}{l!} e^{-\alpha}}_{1} = \alpha$$

- Last sum is 1 because $e^{\alpha} = \sum_{l=0}^{\infty} \alpha^l / l!$

Properties of Expectations

- Let X be a r.v. and let $Y = g(X)$
  - Hard: $\mathbf{E}[Y] = \sum_y y p_Y(y)$
  - Easy: $\mathbf{E}[Y] = \sum_x g(x) p_X(x)$
- Caution: In general, $\mathbf{E}[g(X)] \neq g(\mathbf{E}[X])$

Properties: If α, β are constants, then:

- $\mathbf{E}[\alpha] =$
- $\mathbf{E}[\alpha X] =$
- $\mathbf{E}[\alpha X + \beta] =$

Other Examples without Proofs

- Binomial *(n,p)*

$$\mathsf{E}[X] = np$$

- Pascal *(k,p)*

$$\mathsf{E}[X] = k/p$$

- Uniform *(k,l)*

$$\mathsf{E}[X] = (k + l)/2$$

Variance

- Measure of “spread” of the PMF

Recall: $\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$

- **Second moment:** $\mathbf{E}[X^2] = \sum_x x^2 p_X(x)$

- **Variance**

$$\begin{aligned}\text{var}(X) &= \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \\ &= \sum_x (x - \mathbf{E}[X])^2 p_X(x) \\ &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2\end{aligned}$$

Properties:

- $\text{var}(X) \geq 0$
- $\text{var}(\alpha X + \beta) = \alpha^2 \text{var}(X)$

Variance Examples

- Consider the effect of parameters on variance

Random Variable	Variance
Bernoulli	$p(1-p)$
Geometric	$(1-p)/p^2$
Binomial	$np(1-p)$
Pascal	$k(1-p)/p^2$
Poisson	α
Discrete Uniform	$(l-k)(l-k+2)/12$

- Proofs require summation tricks (see textbook)

Review of Mean and Variance

- Random variable X: function from sample space to the real numbers
- PMF (for discrete random variables): $p_X(x) = \mathbf{P}(X = x)$
- Expectation:

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$$

$$\mathbf{E}[\alpha X + \beta] = \alpha \mathbf{E}[X] + \beta$$

Review Cont'

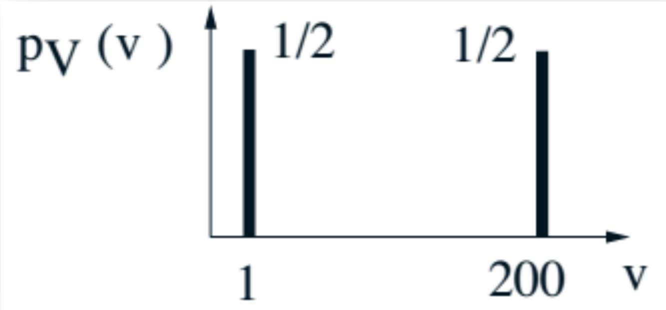
$$\mathbf{E}\Big[X - \mathbf{E}[X]\Big] =$$

$$\begin{aligned} \text{var}(X) &= \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \\ &= \sum_x (x - \mathbf{E}[X])^2 p_X(x) \\ &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \end{aligned}$$

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

Random Speed

- Traverse a 200 mile distance at constant but random speed V
- $d = 200$, $T = t(V) = 200/V$
- $\mathbf{E}[V] =$
- $\text{var}(V) =$
- $\sigma_V =$



Either riding a donkey (1 mile/hr)
OR a Ferrari (200 miles/hr)

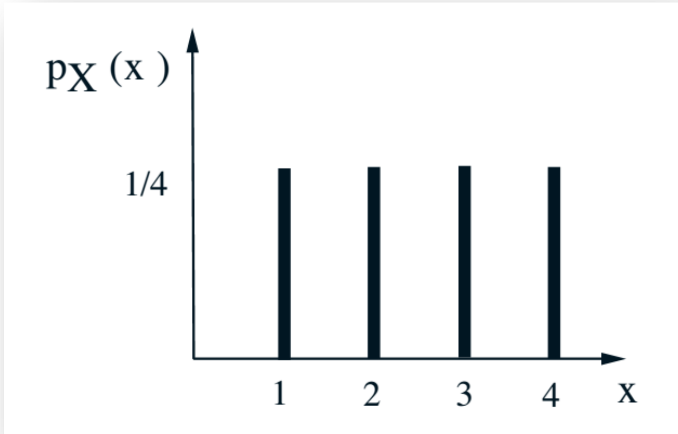
Average Speed vs Average Time

- time in hours $= T = t(V) =$
- $\mathbf{E}[T] = \mathbf{E}[t(V)] = \sum_v t(v) p_V(v) =$
- $\mathbf{E}[TV] = 200 \neq \mathbf{E}[T] \cdot \mathbf{E}[V]$
- $\mathbf{E}[200/V] = \mathbf{E}[T] \neq 200/\mathbf{E}[V]$.

Moral: Mean of the product not always the product of the means
Mean of the inverse not always the inverse of the means

Conditioning PMF on Events

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



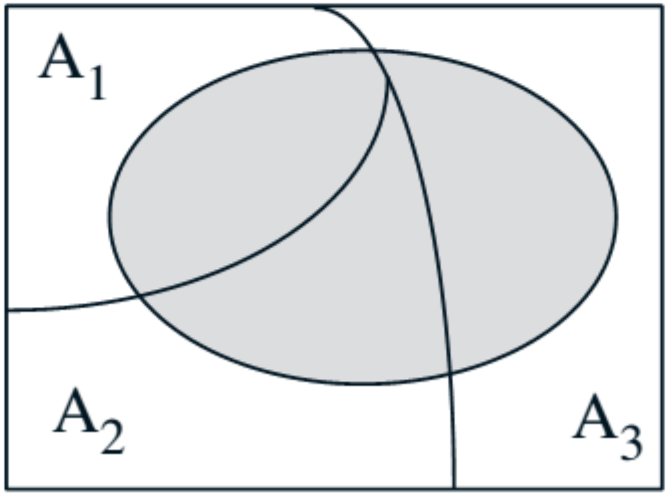
- Let $A = \{X \geq 2\}$

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

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

Total Expectation Theorem

- Partition of sample space
  into disjoint events $A_1, A_2, \ldots, A_n$



$$\mathbf{P}(B) = \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n)$$

$$p_X(x) = \mathbf{P}(A_1)p_{X|A_1}(x) + \cdots + \mathbf{P}(A_n)p_{X|A_n}(x)$$

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

Example

- X is geometrically distributed

$$A_1: \ \{X = 1\}, \quad A_2: \ \{X > 1\}$$

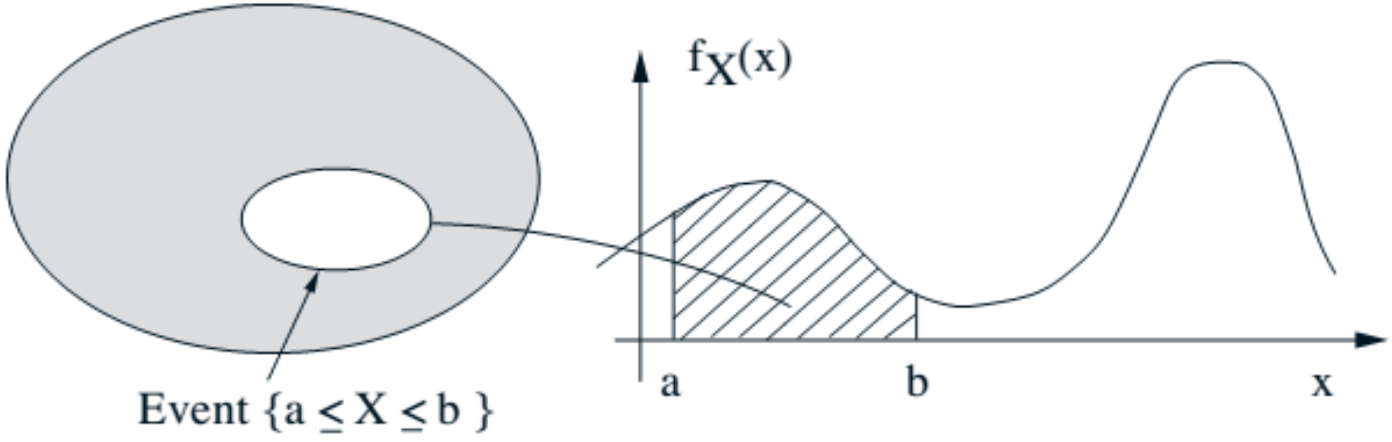
$$\begin{aligned} \mathbf{E}[X] \ = \ & \ \mathbf{P}(X = 1)\mathbf{E}[X \mid X = 1] \\ & + \mathbf{P}(X > 1)\mathbf{E}[X \mid X > 1] \end{aligned}$$

- Solve to get $\mathbf{E}[X] = 1/p$

- Compare with earlier derivation which required summation trick

Continuous RVs

- The RV can take on a continuum of values
- Its probability distribution is captured by a PDF



PDFs

- Probability density function
- Nonnegative and integrates to unity
- Integrate over a set to find probability of set
- If continuous, probability that the RV is any fixed value is zero

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

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

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

$$\mathbf{P}(X \in B) = \int_B f_X(x)\,dx, \quad \text{for ``nice'' sets } B$$

Means and Variances

- Replace sums with integrals:

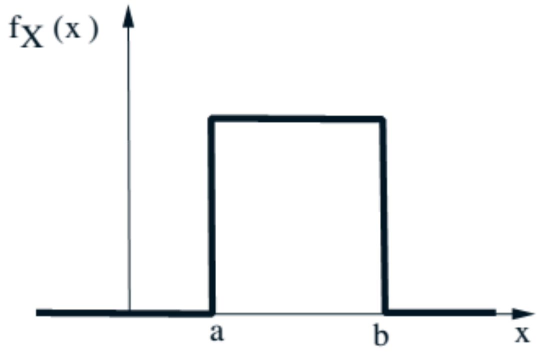
$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

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

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

Example

- **Continuous Uniform r.v.**



- $f_X(x) = \qquad\qquad a \leq x \leq b$

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

- $\sigma_X^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\,dx = \frac{(b-a)^2}{12}$

Cumulative Distribution Function

$$F_X(x) = \mathbf{P}(X \leq x)$$

- CDFs are defined for *any* RV (continuous or discrete)
- For continuous case

$$F_X(x) \; = \int_{-\infty}^{x} f_X(t)\, dt$$

- For discrete case

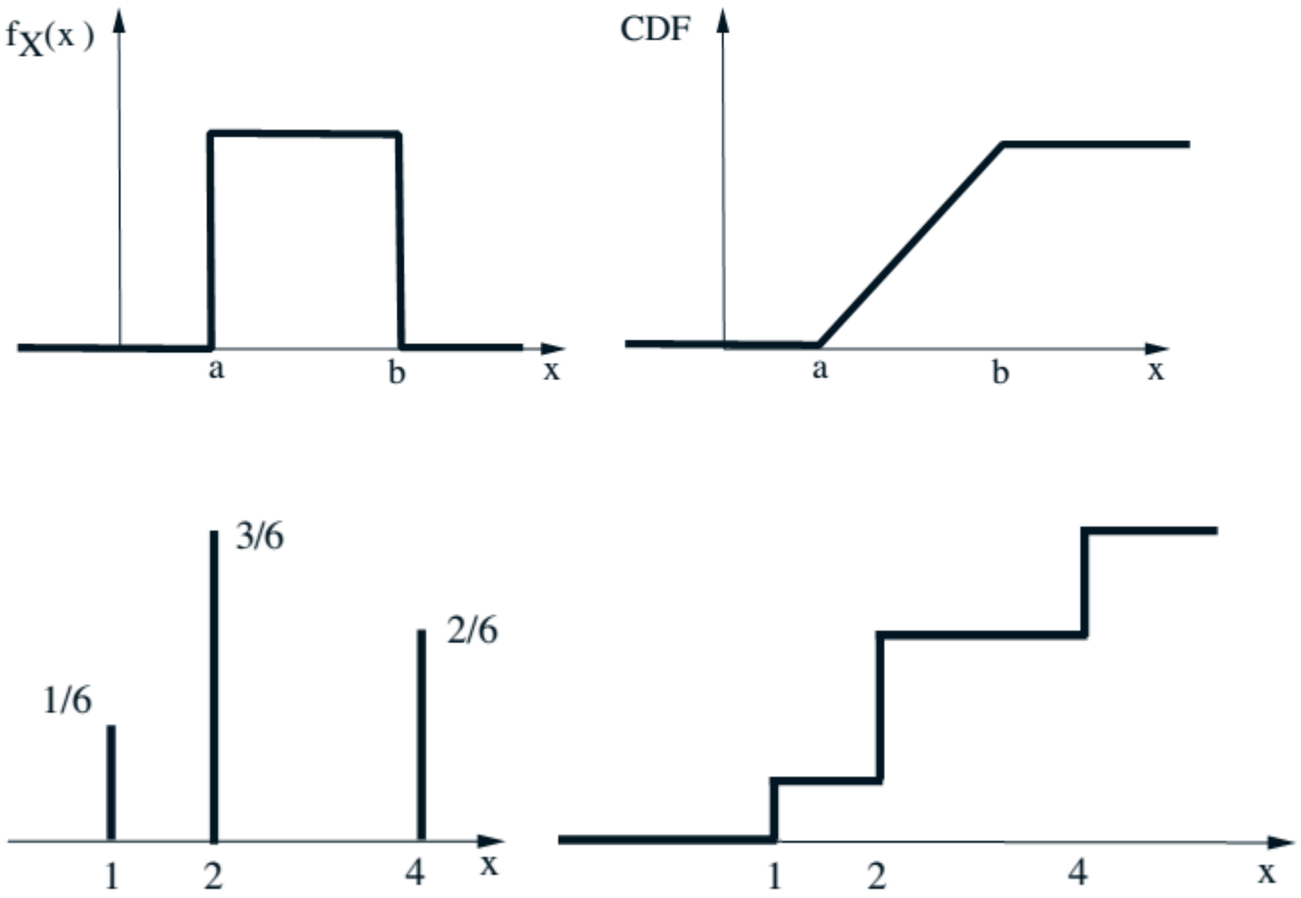
$$F_X(x) = \mathbf{P}(X \leq x) = \sum_{k \leq x} p_X(k)$$

CDF Properties

$$F_X(x) = \mathrm{P}(X \leq x)$$

- It is the probability of something, so between 0 and 1
- Monotone non-decreasing function
- $F_X(-\infty) = 0$; $F_X(\infty) = 1$
- Defined both for continuous and discrete RVs

Example



Example

- Consider a geometric RV *Y* with *p=1/4*
- CDF of *Y* ?

$$P_Y(y) = \begin{cases} (1/4)(3/4)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

$$F_Y(n) = \sum_{j=1}^{n} P_Y(j) = \sum_{j=1}^{n} \frac{1}{4} \left(\frac{3}{4}\right)^{j-1}.$$

- *Recall* $(1-x)\sum_{j=1}^{n} x^{j-1} = 1 - x^n$

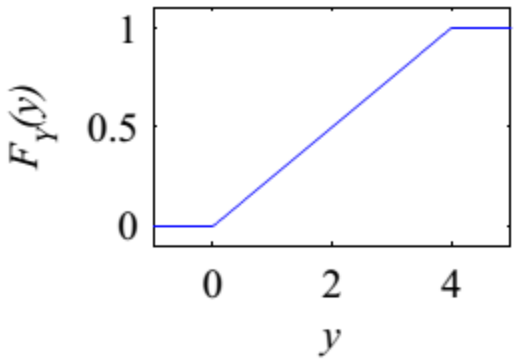
$$F_Y(n) = 1 - \left(\frac{3}{4}\right)^{n}.$$

- Here *n* is an integer but CDF is piecewise constant

Example

- Consider a RV with CDF

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ y/4 & 0 \le y \le 4 \\ 1 & y > 4. \end{cases}$$



(a) $\mathrm{P}[Y \le -1] = F_Y(-1) = 0$

(b) $\mathrm{P}[Y \le 1] = F_Y(1) = 1/4$

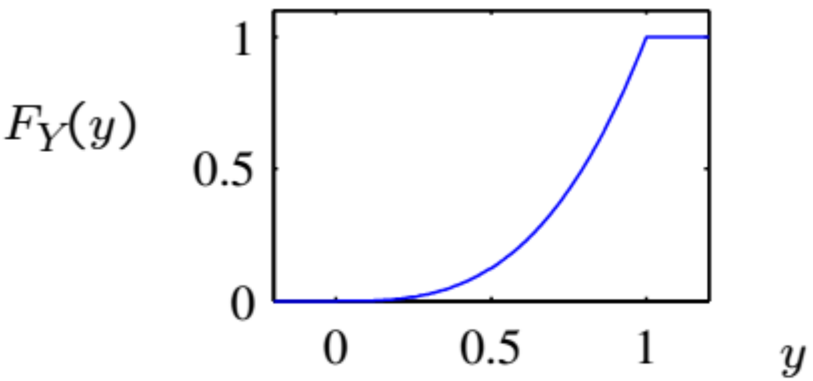
(c) $$\begin{aligned}\mathrm{P}[2 < Y \le 3] &= F_Y(3) - F_Y(2) \\ &= 3/4 - 2/4 = 1/4.\end{aligned}$$

(d) $$\begin{aligned}\mathrm{P}[Y > 1.5] &= 1 - \mathrm{P}[Y \le 1.5] \\ &= 1 - F_Y(1.5) \\ &= 1 - (1.5)/4 = 5/8.\end{aligned}$$

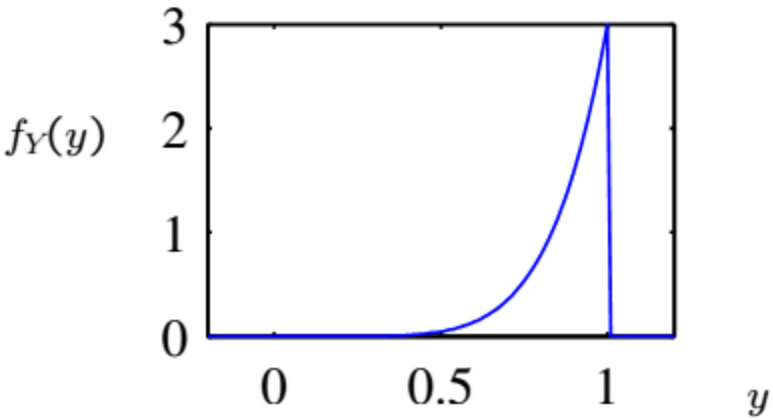
Example

- Consider a RV with CDF below; find PDF

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ y^3 & 0 \le y \le 1 \\ 1 & y > 1. \end{cases}$$



$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} 3y^2 & 0 < y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$



Example

- Consider a RV with PDF below

$$f_X(x) = \begin{cases} cxe^{-x/2} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

Sketch the PDF and find the following:

(a) the constant c

(b) the CDF $F_X(x)$

(c) $P[0 \leq X \leq 4]$

(d) $P[-2 \leq X \leq 2]$

Solution

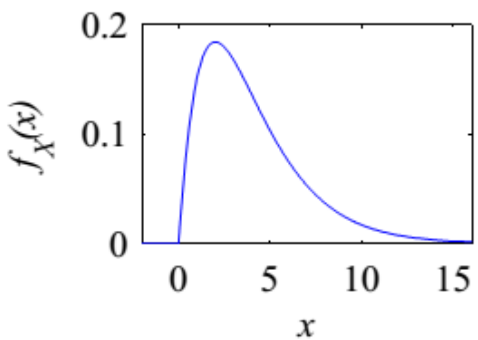
(a) First we will find the constant c and then we will sketch the PDF. To find c, we use the fact that

$$1 = \int_{-\infty}^{\infty} f_X(x)\, dx = \int_0^{\infty} cxe^{-x/2}\, dx. \tag{1}$$

We evaluate this integral using integration by parts:

$$\begin{aligned} 1 &= \underbrace{-2cxe^{-x/2}\Big|_0^{\infty}}_{=0} + \int_0^{\infty} 2ce^{-x/2}\, dx \\ &= -4ce^{-x/2}\Big|_0^{\infty} = 4c. \end{aligned} \tag{2}$$

Thus $c = 1/4$ and X has the Erlang $(n = 2, \lambda = 1/2)$ PDF



$$f_X(x) = \begin{cases} (x/4)e^{-x/2} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

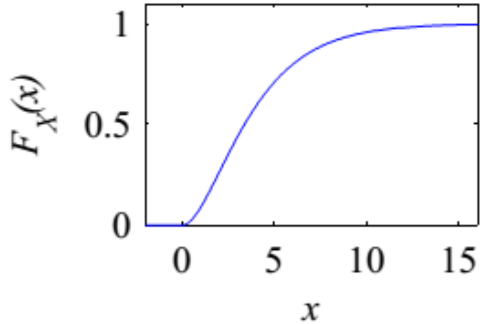
(b) To find the CDF $F_X(x)$, we first note X is a nonnegative random variable so that $F_X(x) = 0$ for all $x < 0$. For $x \geq 0$,

[Continued]

Solution Cont'

$$F_X(x) = \int_0^x f_X(y)\,dy = \int_0^x \frac{y}{4} e^{-y/2}\,dy$$
$$= -\frac{y}{2} e^{-y/2}\Big|_0^x + \int_0^x \frac{1}{2} e^{-y/2}\,dy$$
$$= 1 - \frac{x}{2} e^{-x/2} - e^{-x/2}.$$

The complete expression for the CDF is



$$F_X(x) = \begin{cases} 1 - \left(\frac{x}{2} + 1\right) e^{-x/2} & x \geq 0, \\ 0 & \text{ow.} \end{cases}$$

(c) From the CDF $F_X(x)$,

$$P[0 \leq X \leq 4] = F_X(4) - F_X(0)$$
$$= 1 - 3e^{-2}.$$

(d) Similarly,

$$P[-2 \leq X \leq 2] = F_X(2) - F_X(-2)$$
$$= 1 - 3e^{-1}.$$

Example

The probability density function of the random variable Y is

$$f_Y(y) = \begin{cases} 3y^2/2 & -1 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Sketch the PDF and find the following:

(a) the expected value $\mathrm{E}[Y]$

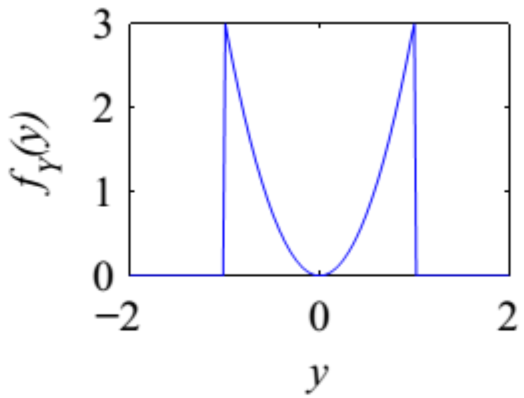
(b) the second moment $\mathrm{E}[Y^2]$

(c) the variance $\mathrm{Var}[Y]$

(d) the standard deviation σ_Y

Solution

The PDF of Y is



$$f_Y(y) = \begin{cases} 3y^2/2 & -1 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) The expected value of Y is

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_{-1}^{1} (3/2) y^3\, dy = (3/8) y^4 \big|_{-1}^{1} = 0.$$

(b) The second moment of Y is

$$E\left[Y^2\right] = \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy = \int_{-1}^{1} (3/2) y^4\, dy = (3/10) y^5 \big|_{-1}^{1} = 3/5.$$

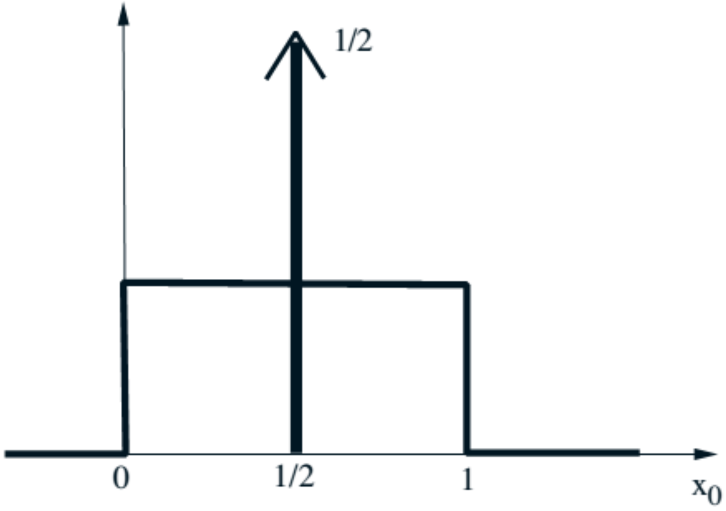
(c) The variance of Y is

$$\text{Var}[Y] = E\left[Y^2\right] - (E[Y])^2 = 3/5.$$

(d) The standard deviation of Y is $\sigma_Y = \sqrt{\text{Var}[Y]} = \sqrt{3/5}$.

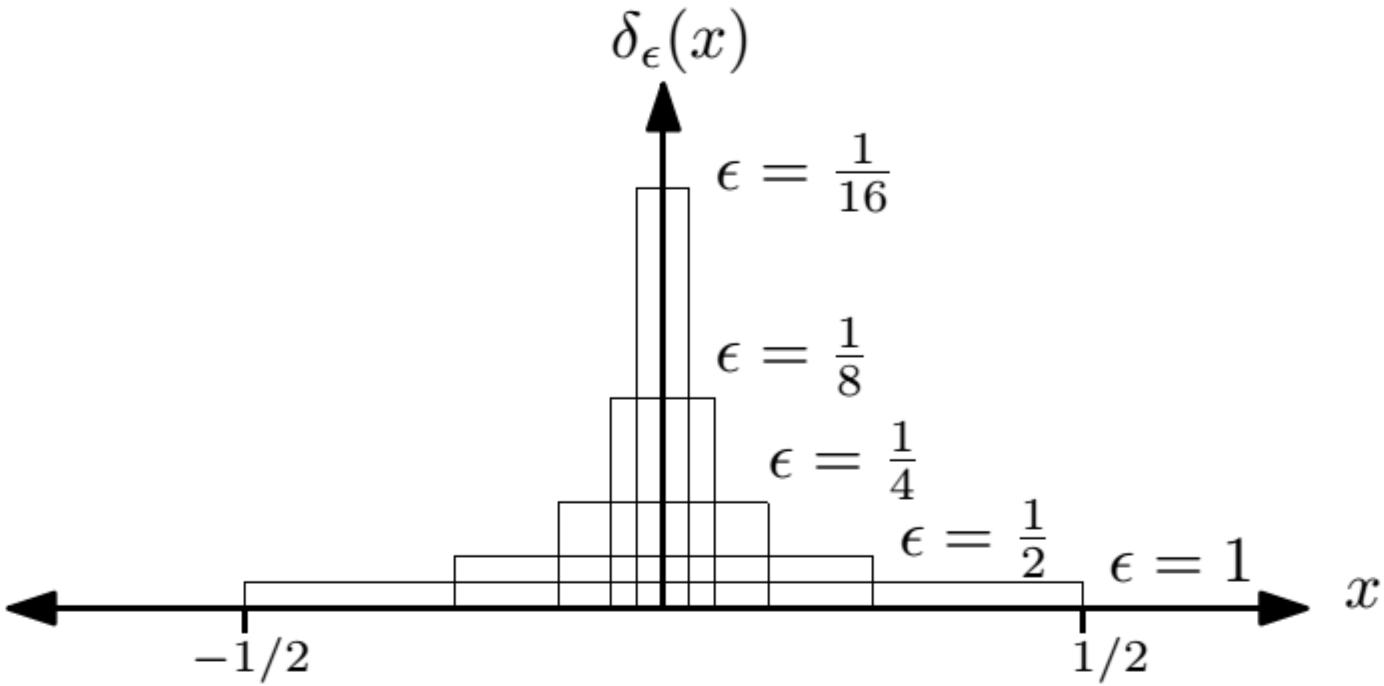
Mixed Distributions

- Schematic drawing of a combination of a PDF and a PMF



Dirac's delta function shows that $\mathbf{P}(X = 1/2) = 1/2$

Recalling Dirac's Delta

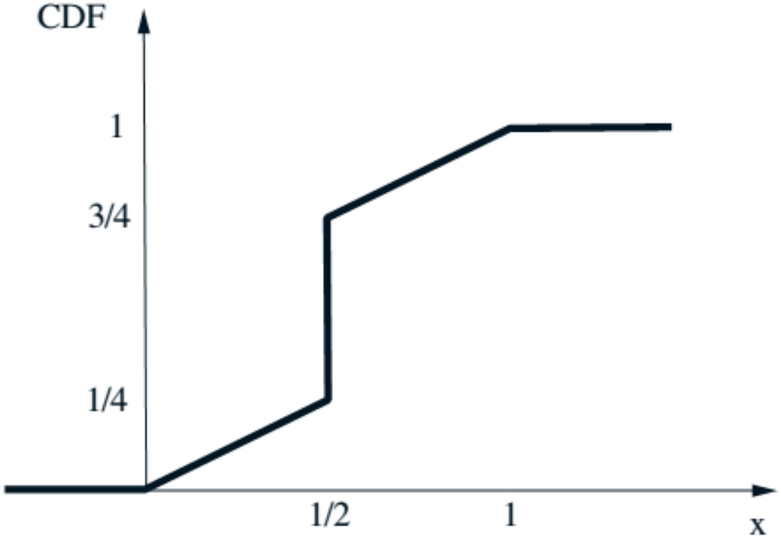


$$\int_{-\infty}^{\infty} g(x)\delta(x - x_0)\, dx = g(x_0)$$

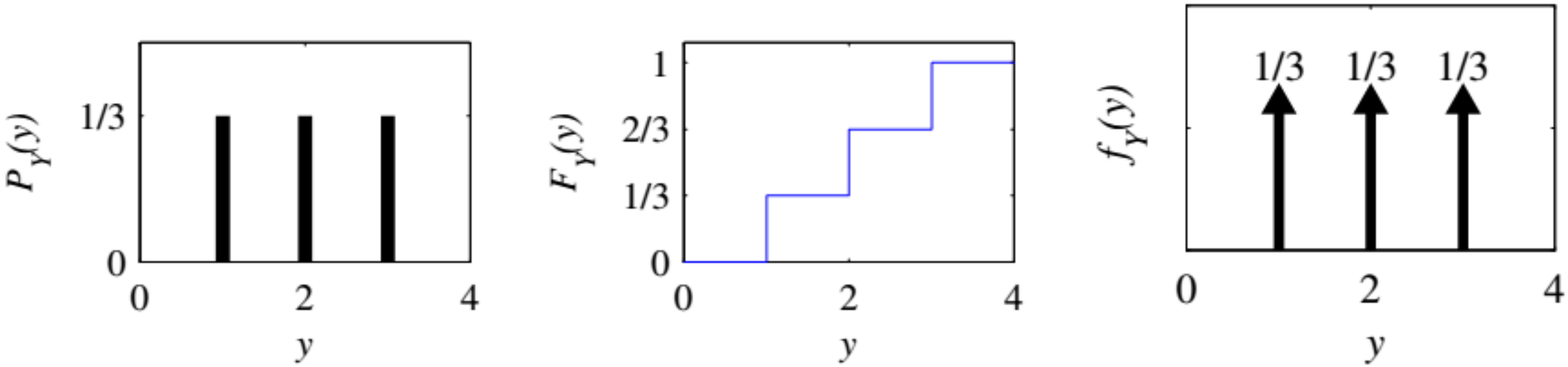
Mixed Distributions Cont'

- The corresponding CDF:

$$F_X(x) = \mathbf{P}(X \le x)$$



Example of Mixed Distributions



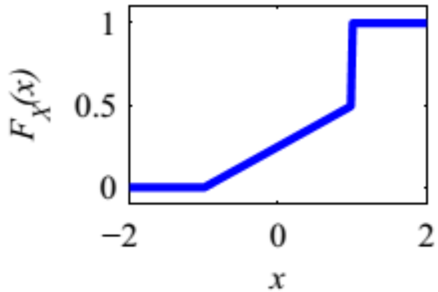
Worked Example

$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/4 & -1 \le x < 1 \\ 1 & x \ge 1. \end{cases}$$

Sketch the CDF and find the following:
(a) $\mathrm{P}[X \le 1]$
(b) $\mathrm{P}[X < 1]$
(c) $\mathrm{P}[X = 1]$
(d) the PDF $f_X(x)$

Solution

The CDF of X is



$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/4 & -1 \leq x < 1, \\ 1 & x \geq 1. \end{cases}$$

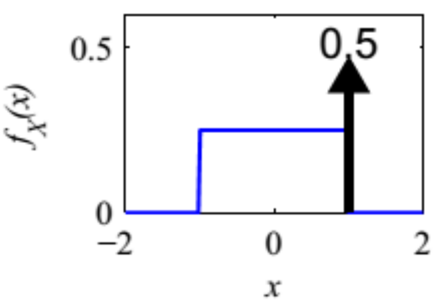
The following probabilities can be read directly from the CDF:

(a) $\mathrm{P}[X \leq 1] = F_X(1) = 1$.

(b) $\mathrm{P}[X < 1] = F_X(1^-) = 1/2$.

(c) $\mathrm{P}[X = 1] = F_X(1^+) - F_X(1^-) = 1/2$.

(d) We find the PDF $f_Y(y)$ by taking the derivative of $F_Y(y)$. The resulting PDF is



$$f_X(x) = \begin{cases} \frac{1}{4} & -1 \leq x < 1, \\ \frac{\delta(x-1)}{2} & x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Continuous vs Discrete vs Mixed

For any random variable X,

- X always has a CDF $F_X(x) = \mathsf{P}[X \leq x]$.

- If $F_X(x)$ is piecewise flat with discontinuous jumps, then X is discrete.

- If $F_X(x)$ is a continuous function, then X is continuous.

- If $F_X(x)$ is a piecewise continuous function with discontinuities, then X is mixed.

- When X is discrete or mixed, the PDF $f_X(x)$ contains one or more delta functions.

Uniform Distribution

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x < b, \\ 0 & \textit{otherwise}, \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x \le a, \\ (x-a)/(b-a) & a < x \le b \\ 1 & x > b. \end{cases}$$

$$\mathrm{E}[X] = (b+a)/2.$$
$$\mathrm{Var}[X] = (b-a)^2/12.$$

Exponential Distribution

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

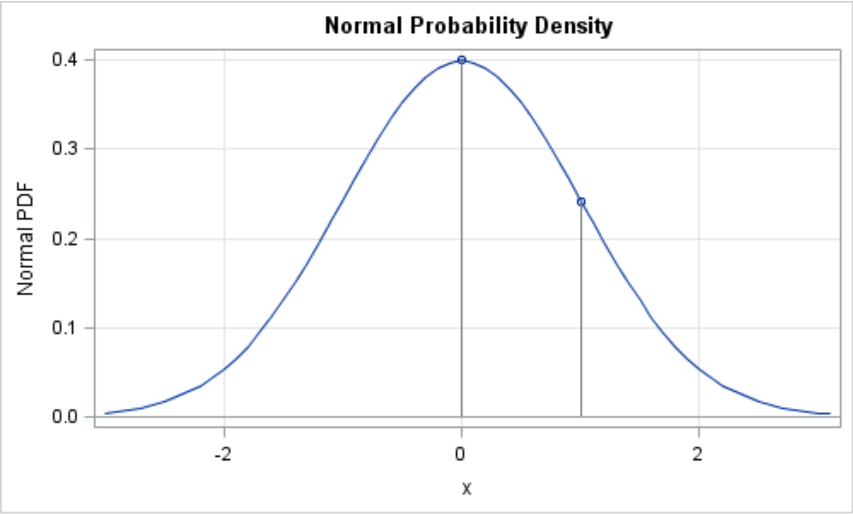
$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

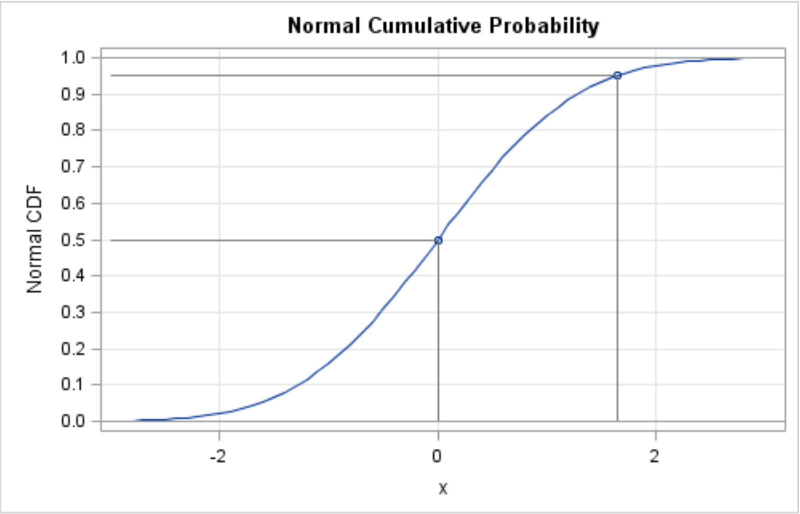
$$\begin{aligned} \mathsf{E}[X] &= 1/\lambda. \\ \mathsf{Var}[X] &= 1/\lambda^2. \end{aligned}$$

Gaussian (Normal) Distribution

- Standard normal $N(0,1)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$





Gaussian (Normal) Distribution

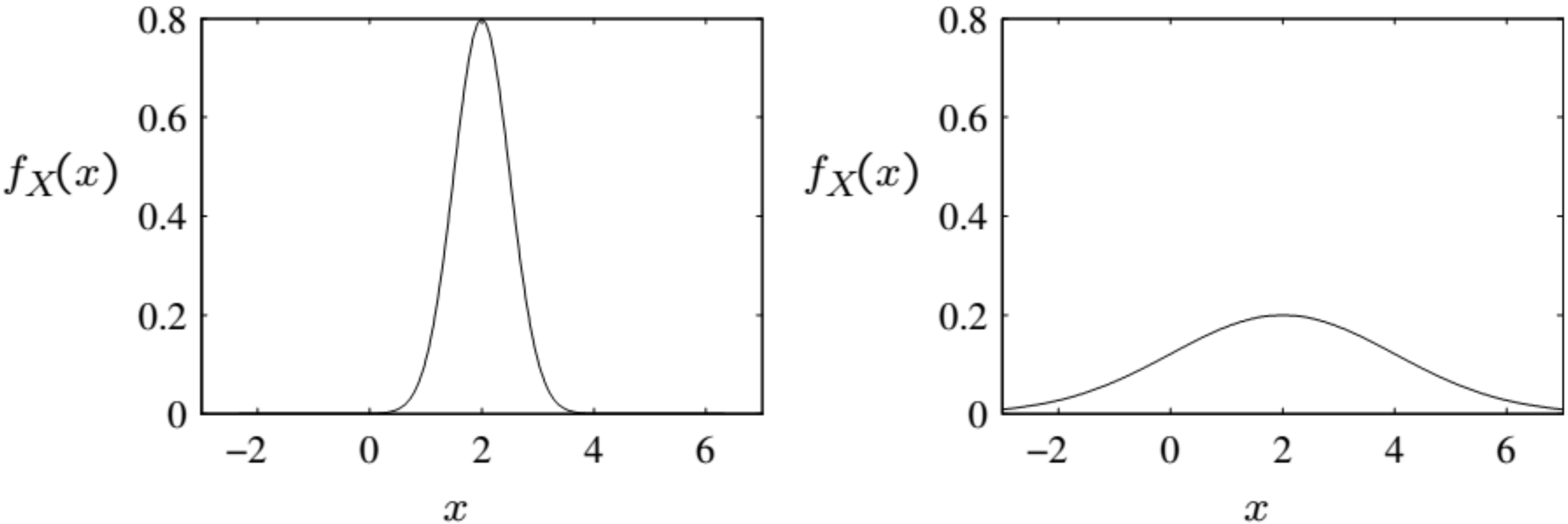
- General normal $N(\mu, \sigma^2)$:

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

- It turns out that:
  $\mathbf{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

- Let $Y = aX + b$
  - Then: $\mathbf{E}[Y] =$ $\mathrm{Var}(Y) =$
  - Fact: $Y \sim N(a\mu + b, a^2\sigma^2)$

Gaussian (Normal) Distribution



(a) $\mu = 2,\ \sigma = 1/2$ **(b)** $\mu = 2,\ \sigma = 2$

Calculating Normal Probabilities

- No closed form available for CDF
  - but there are tables (for standard normal)
- If $X \sim N(\mu, \sigma^2)$, then $\dfrac{X-\mu}{\sigma} \sim N(\qquad\qquad)$
- If $X \sim N(2, 16)$:

$$\mathbf{P}(X \leq 3) = \mathbf{P}\left(\frac{X-2}{4} \leq \frac{3-2}{4}\right) = \text{CDF}(0.25)$$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du$$

Calculating Normal Probabilities

- Tables exist
- Can be constructed using numerical methods
- No, you don't have to memorize them!

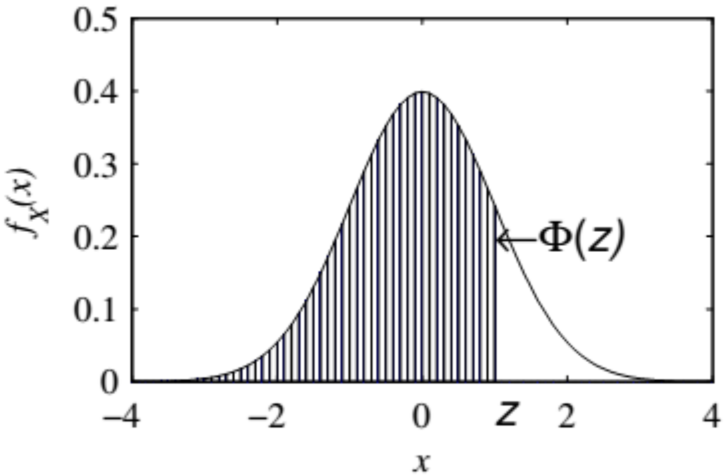
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.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

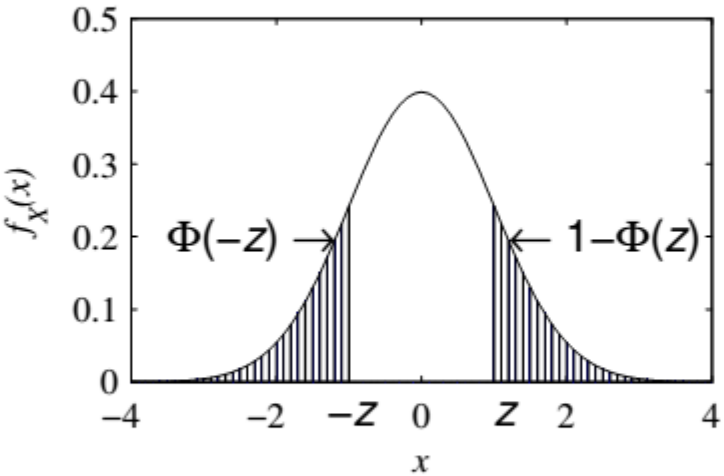
Calculating Normal Probabilities

Let X be Gaussian with mean 61 and std 10; What is $P[X \leq 46]$

$$\mathrm{P}\,[X \leq 46] = F_X(46) = \Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.933 = 0.067$$

Used $\Phi(-z) = 1 - \Phi(z)$ in 3rd inequality (see below)





Joint Distributions: CDF

$$F_{X,Y}(x, y) = \mathsf{P}[X \leq x, Y \leq y]$$

(a) $0 \leq F_{X,Y}(x, y) \leq 1$,
(b) $F_{X,Y}(\infty, \infty) = 1$,
(c) $F_X(x) = F_{X,Y}(x, \infty)$,
(d) $F_Y(y) = F_{X,Y}(\infty, y)$,
(e) $F_{X,Y}(x, -\infty) = 0$,
(f) $F_{X,Y}(-\infty, y) = 0$,
(g) If $x \leq x_1$ and $y \leq y_1$, then

$$F_{X,Y}(x, y) \leq F_{X,Y}(x_1, y_1)$$

- Applies to continuous or discrete

Example

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 5, \\ 0 & y < 6, \\ (x-5)(y-6) & 5 \le x < 6, 6 \le y < 7, \\ y-6 & x \ge 6, 6 \le y < 7, \\ x-5 & 5 \le x < 6, y \ge 7, \\ 1 & \text{otherwise.} \end{cases}$$

Find $F_X(x)$ and $F_Y(y)$.

Solution

$$F_X(x) = \begin{cases} 0 & x < 5, \\ x - 5 & 5 \le x < 6, \\ 1 & x \ge 6, \end{cases} \qquad F_Y(y) = \begin{cases} 0 & y < 6, \\ y - 6 & 6 \le y < 7, \\ 1 & y \ge 7. \end{cases}$$

- Both *X* and *Y* are uniformly distributed

Joint PMF

$$P_{X,Y}(x,y) = \mathsf{P}\left[X = x, Y = y\right]$$

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x,y)\,, \qquad P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x,y)$$

- Convenient for discrete random variables
- $P_X(x)$ and $P_Y(y)$ are called marginal PMFs

- A rectangular array of probabilities
- Row-wise sum? Column-wise sum?

Example

$P_{Q,G}(q,g)$	$g = 0$	$g = 1$	$g = 2$	$g = 3$
$q = 0$	0.06	0.18	0.24	0.12
$q = 1$	0.04	0.12	0.16	0.08

Calculate the following probabilities:

(a) $\text{P}[Q = 0]$

(b) $\text{P}[Q = G]$

(c) $\text{P}[G > 1]$

(d) $\text{P}[G > Q]$

Solution

(a) The probability that $Q = 0$ is

$$\begin{aligned} \mathrm{P}[Q=0] &= P_{Q,G}(0,0) + P_{Q,G}(0,1) + P_{Q,G}(0,2) + P_{Q,G}(0,3) \\ &= 0.06 + 0.18 + 0.24 + 0.12 = 0.6. \end{aligned}$$

(b) The probability that $Q = G$ is

$$\mathrm{P}[Q=G] = P_{Q,G}(0,0) + P_{Q,G}(1,1) = 0.18.$$

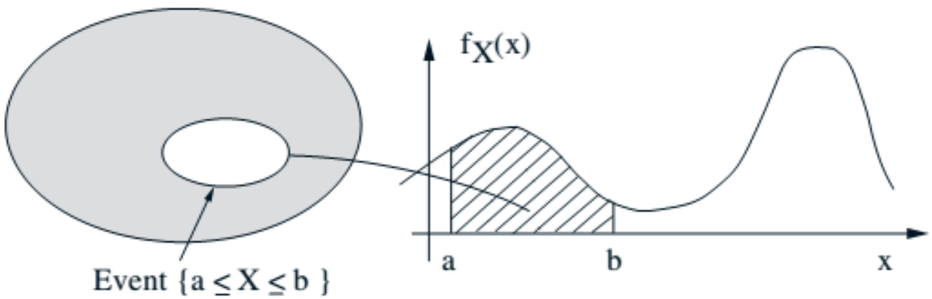
(c) The probability that $G > 1$ is

$$\begin{aligned} \mathrm{P}[G>1] &= \sum_{g=2}^{3}\sum_{q=0}^{1} P_{Q,G}(q,g) \\ &= 0.24 + 0.16 + 0.12 + 0.08 = 0.6. \end{aligned}$$

(d) The probability that $G > Q$ is

$$\begin{aligned} \mathrm{P}[G>Q] &= \sum_{q=0}^{1}\sum_{g=q+1}^{3} P_{Q,G}(q,g) \\ &= 0.18 + 0.24 + 0.12 + 0.16 + 0.08 = 0.78. \end{aligned}$$

Recall



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

- $\mathbf{P}(x \le X \le x + \delta) \approx f_X(x) \cdot \delta$
- $\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$

Joint PDF Definition

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\; dv\, du$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x\, \partial y}$$

$$f_{X,Y}(x,y) \geq 0$$

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

Joint PDF Interpretation

$$\mathbf{P}((X,Y) \in S) = \int\int_S f_{X,Y}(x,y)\,dx\,dy$$

- Interpretation:

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

- Expectations:

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

- From the joint to the marginal:

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

- X and Y are called independent if

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

Joint PDF Example

Random variables X and Y have joint PDF

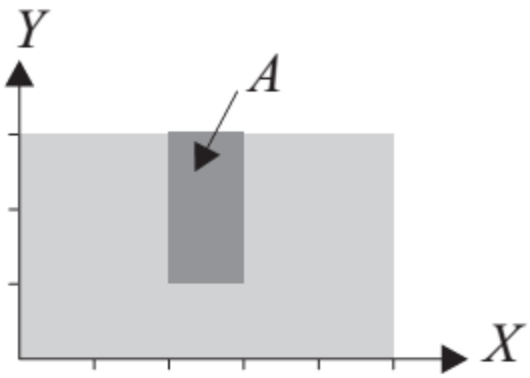
$$f_{X,Y}(x,y) = \begin{cases} c & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases}$$

Find the constant c and $\mathrm{P}[A] = \mathrm{P}[2 \le X < 3, 1 \le Y < 3]$.

Solution

- Finding constant:

$$1 = \int_0^5 \int_0^3 c \, dy \, dx = 15c \qquad c = 1/15$$



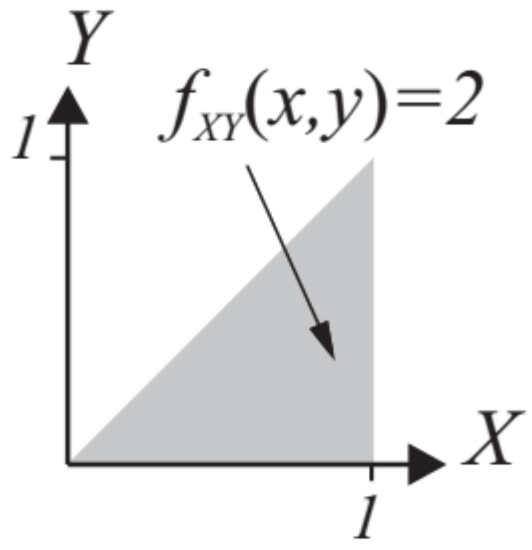
- Finding probability of event *A*

$$\mathsf{P}[A] = \int_2^3 \int_1^3 \frac{1}{15} \, dv \, du = 2/15$$

Example

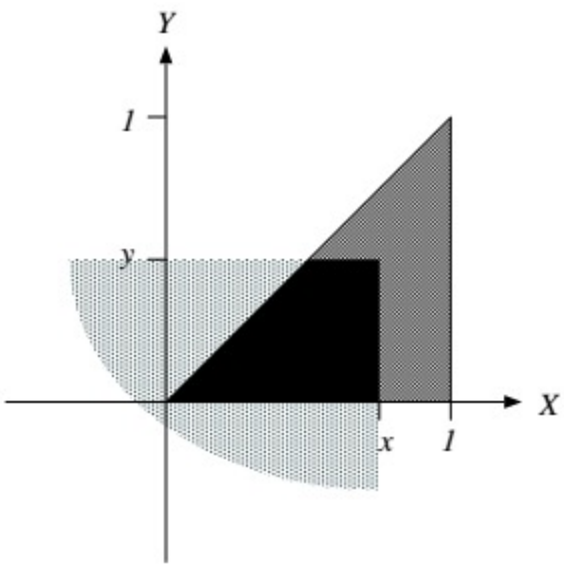
- Find joint CDF $F_{X,Y}(x, y)$

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$



Solution

- To perform integration to find CDF a diagram to find integral limits
- If either $x < 0$ or $y < 0$ $F_{X,Y}(x,y) = 0$egion to integrate)
- If both $x \geq 1$ $y \geq 1$hen $F_{X,Y}(x,y) = 1$
- Consider figures for all other cases

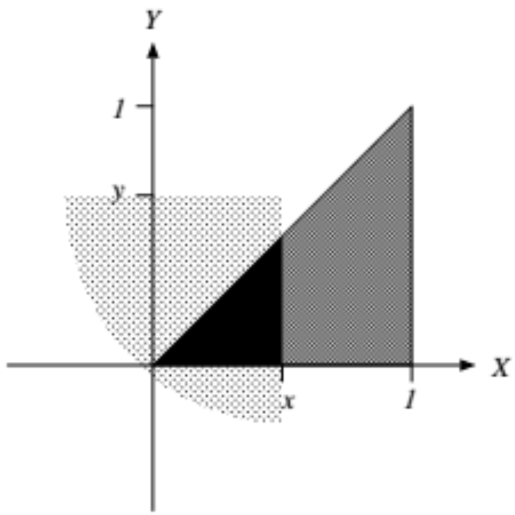


$$F_{X,Y}(x,y) = \int_0^y \int_v^x 2\, du\, dv = 2xy - y^2$$

$$0 \leq y \leq x \leq 1$$

Solution Cont'

- Consider figures for all other cases

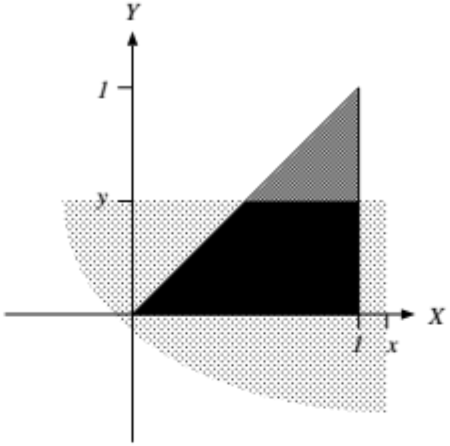


$$0 \le x < y$$
$$0 \le x \le 1$$

$$F_{X,Y}(x,y) = \int_0^x \int_v^x 2\, du\, dv = x^2$$

Solution Cont'

- Consider figures for all other cases



$$0 \leq y \leq 1$$
$$x > 1$$

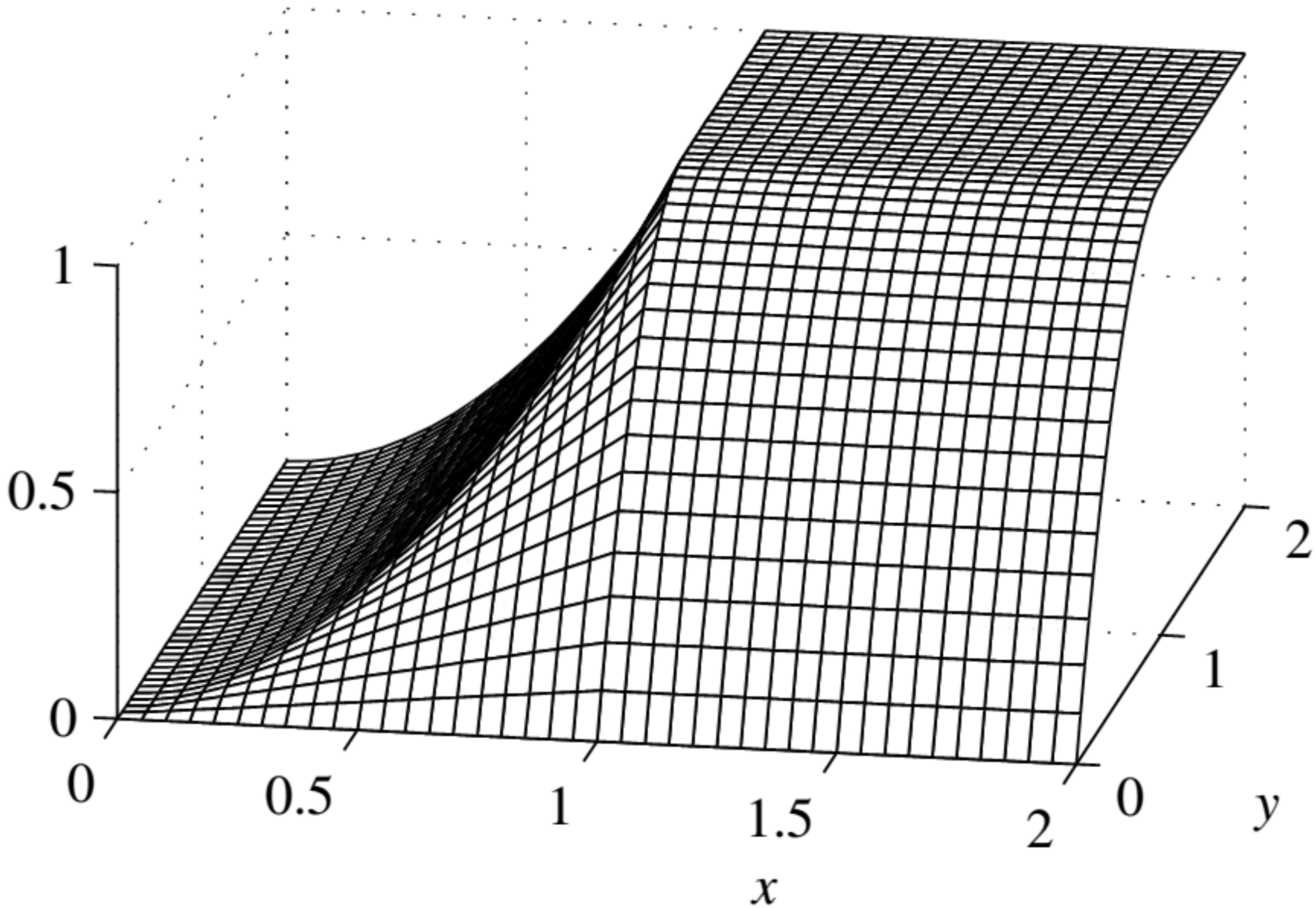
$$F_{X,Y}(x, y) = \int_0^y \int_v^1 2\, du\, dv = 2y - y^2$$

Solution Cont'

- Putting it all together:

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ 2xy - y^2 & 0 \leq y \leq x \leq 1 \\ x^2 & 0 \leq x < y, 0 \leq x \leq 1 \\ 2y - y^2 & 0 \leq y \leq 1, x > 1 \\ 1 & x > 1, y > 1 \end{cases}$$

Solution Cont'



Example

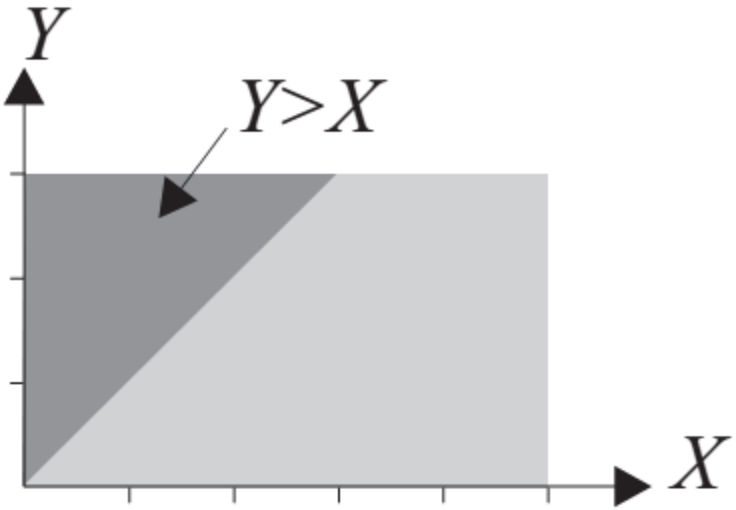
$$f_{X,Y}(x,y) = \begin{cases} 1/15 & 0 \leq x \leq 5, 0 \leq y \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathrm{P}[A] = \mathrm{P}[Y > X]?$$

Solution

$$\mathrm{P}[A] = \int_0^3 \left(\int_x^3 \frac{1}{15} \right) dy\, dx$$

$$= \int_0^3 \frac{3-x}{15}\, dx = -\frac{(3-x)^2}{30}\Big|_0^3 = \frac{3}{10}.$$



Marginal PDF

Fact:

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

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

The PDFs of *X* and *Y* are termed "marginal PDFs"

Example

Given:

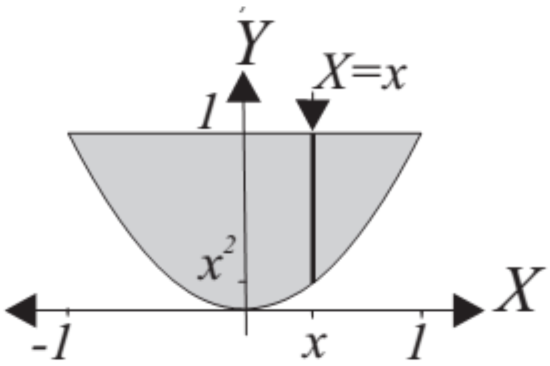
$$f_{X,Y}(x,y) = \begin{cases} 5y/4 & -1 \le x \le 1, x^2 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

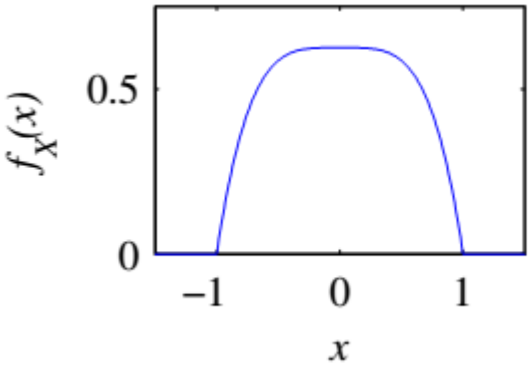
Find the marginal PDFs of *X* and *Y*

Solution

For $\quad -1 \le x \le 1,$

$$f_X(x) = \int_{x^2}^{1} \frac{5y}{4}\,dy = \frac{5(1-x^4)}{8}$$



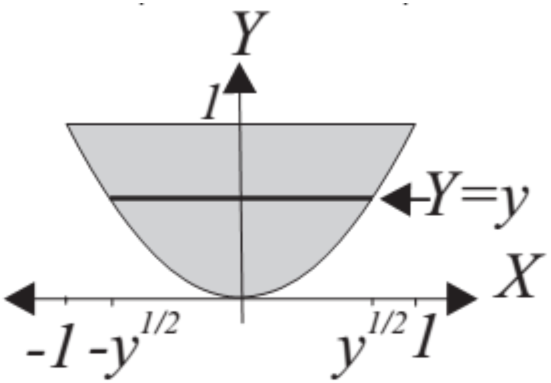


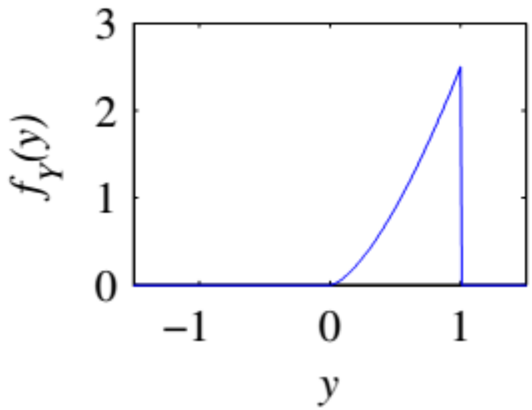
$$f_X(x) = \begin{cases} 5(1-x^4)/8 & -1 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution Continued

For $0 \le y \le 1,$

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{5y}{4}\,dx = \frac{5y}{4}x\Big|_{x=-\sqrt{y}}^{x=\sqrt{y}} = 5y^{3/2}/2.$$





$$f_Y(y) = \begin{cases} (5/2)y^{3/2} & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Expectations and Variance: 2 RVs

We already covered:

$$\mathsf{E}[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) P_{X,Y}(x, y)$$

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

Expectation is linear:

$$\mathsf{E}[X + Y] = \mathsf{E}[X] + \mathsf{E}[Y]$$

Variance: 2 RVs

- Can use definition to prove:

$$\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\,\mathsf{E}[(X-\mu_X)(Y-\mu_Y)]$$

- Last term is related to “covariance”:

$$\mathsf{Cov}[X,Y] = \mathsf{E}[(X-\mu_X)(Y-\mu_Y)]$$

- We will see more about (normalized) covariance later.

Variance Properties Cont'

- $\text{Var}(aX) = a^2\text{Var}(X)$
- $\text{Var}(X + a) = \text{Var}(X)$
- Let $Z = X + Y$.
  If X, Y are independent:

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

If $X = Y$, $\text{Var}(X + Y) =$

If $X = -Y$, $\text{Var}(X + Y) =$

If X, Y indep., and $Z = X - 3Y$,
$\text{Var}(Z) =$

Binomial Mean and Variance

- $X = \#$ of successes in n independent trials

$$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

- Not easy to calculate using definition of expectation

Binomial Mean and Variance

- $X_i = \begin{cases} 1, & \text{if success in trial } i, \\ 0, & \text{otherwise} \end{cases}$

- $\mathbf{E}[X_i] =$

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

- $\text{Var}(X_i) =$

- $\text{Var}(X) =$

The Hat Problem

- n people throw their hats in a box and then pick one at random.
  - X: number of people who get their own hat
  - Find $\mathbf{E}[X]$

The Hat Problem Cont'

$$X_i = \begin{cases} 1, & \text{if } i \text{ selects own hat} \\ 0, & \text{otherwise.} \end{cases}$$

- $X = X_1 + X_2 + \cdots + X_n$
- $\mathbf{P}(X_i = 1) =$
- $\mathbf{E}[X_i] =$
- Are the X_i independent?
- $\mathbf{E}[X] =$

Variance in the Hat Problem

- $\text{Var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \mathbf{E}[X^2] - 1$

$$X^2 = \sum_i X_i^2 + \sum_{i,j:i \neq j} X_i X_j$$

- $\mathbf{E}[X_i^2] =$

$$\mathbf{P}(X_1 X_2 = 1) = \mathbf{P}(X_1 = 1) \cdot \mathbf{P}(X_2 = 1 \mid X_1 = 1)$$

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

Conditioning RV on an Event

- One can define conditional probabilities (see earlier lectures)
- This involves conditioning on events $P[A|B]$
- Earlier we saw that PMFs can be conditioned on events
- One can also define conditional CDFs or PDFs.
- Example:

Given the event B *with* $\mathrm{P}[B] > 0$, *the* conditional cumulative distribution function *of* X *is*

$$F_{X|B}(x) = \mathrm{P}\left[X \leq x|B\right].$$

- As we saw earlier, discrete RVs one can define conditional PMFs:

$$P_{X|B}(x) = \mathrm{P}\left[X = x|B\right].$$

Conditioning RV on an Event (PMF Example)

- Before we do general RVs conditioned on an event, we revisit it for PMFs

A website distributes instructional videos on bicycle repair. The length of a video in minutes X has PMF

$$P_X(x) = \begin{cases} 0.15 & x = 1, 2, 3, 4, \\ 0.1 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Suppose the website has two servers, one for videos shorter than five minutes and the other for videos of five or more minutes. What is the PMF of video length in the second server?

Solution

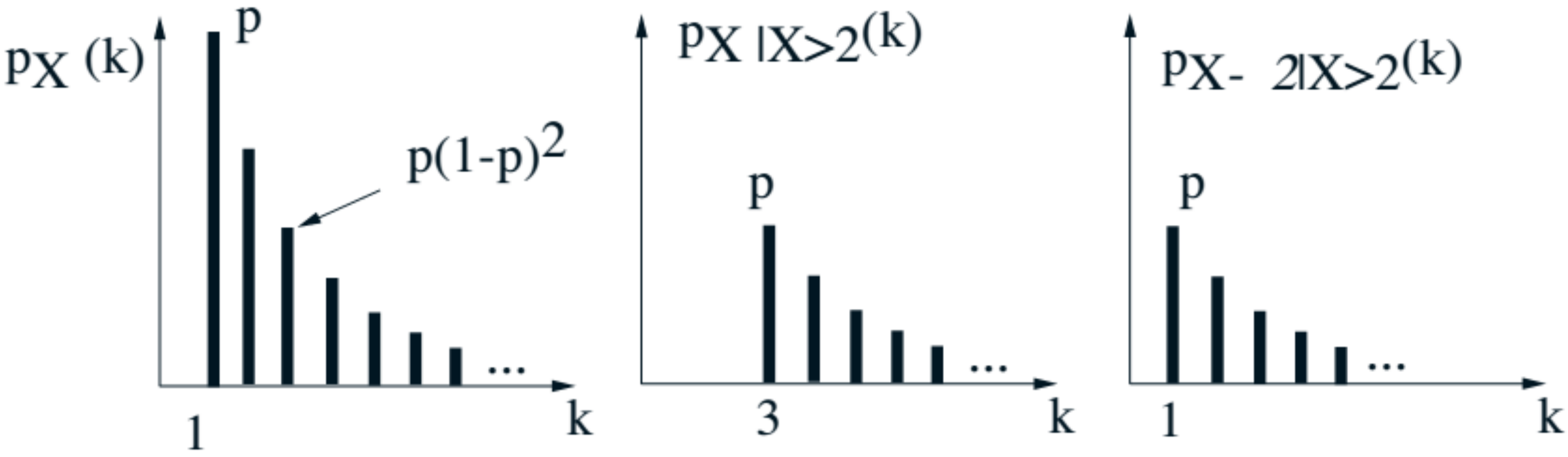
$$P_{X|L}(x) = \begin{cases} \dfrac{P_X(x)}{\mathrm{P}[L]} & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathrm{P}[L] = \sum_{x=5}^{8} P_X(x) = 0.4.$$

$$P_{X|L}(x) = \begin{cases} 0.1/0.4 = 0.25 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases}$$

Memoryless Geometric

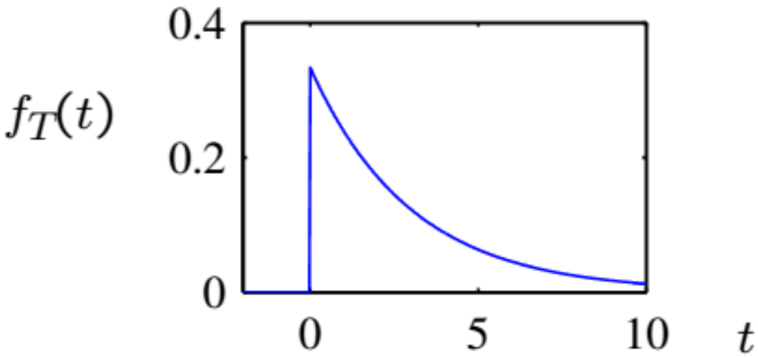
- X: number of independent coin tosses until first head
- Memoryless property: Given that $X > 2$, the r.v. $X - 2$ has same geometric PMF



Memoryless Exponential

- Exponential is the continuous analogue of the geometric

Suppose the duration T (in minutes) of a telephone call is an exponential (1/3) random variable:



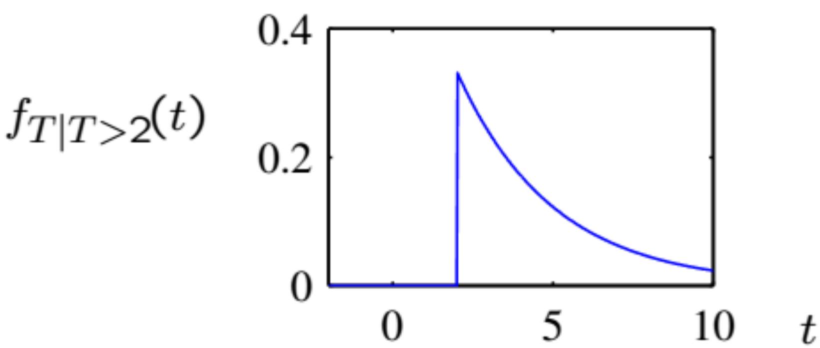
$$f_T(t) = \begin{cases} (1/3)e^{-t/3} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For calls that last at least 2 minutes, what is the conditional PDF of the call duration?

Memoryless Exponential

$$\mathsf{P}\left[T > 2\right] = \int_2^\infty f_T(t)\, dt = e^{-2/3}. \tag{1}$$

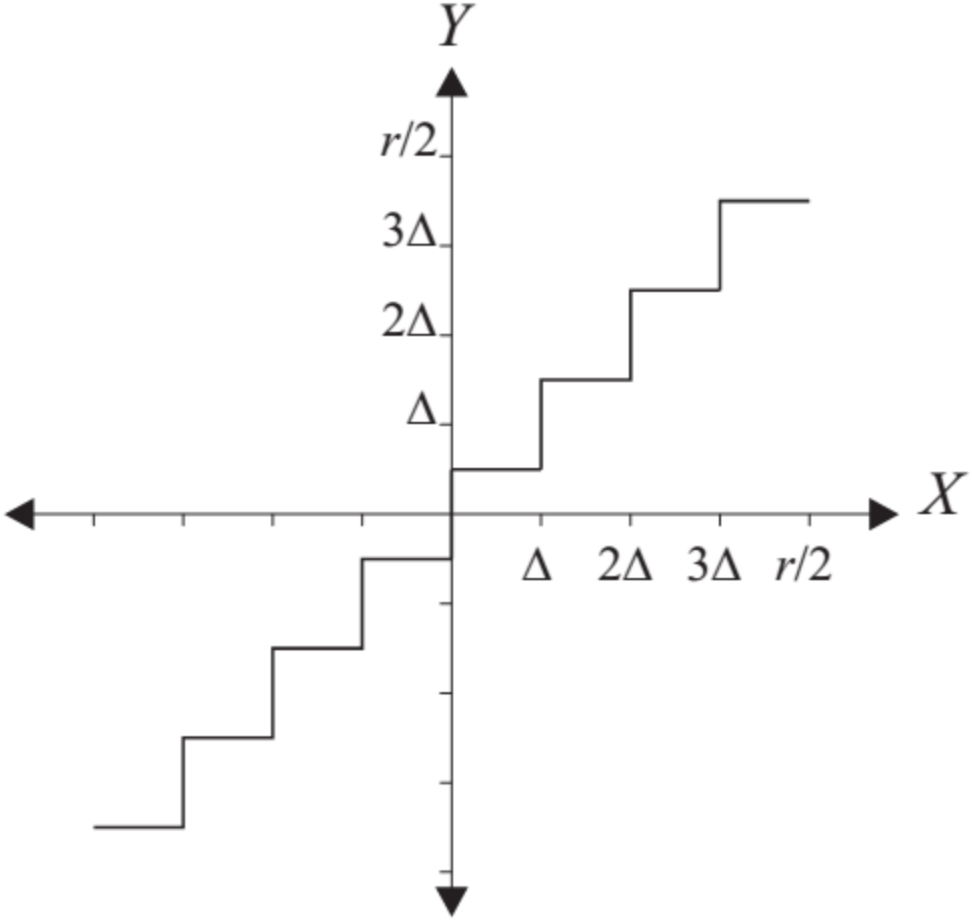
The conditional PDF of T given $T > 2$ is



$$f_{T|T>2}(t) = \begin{cases} \frac{f_T(t)}{\mathsf{P}[T>2]} & t > 2, \\ 0 & \text{otherwise,} \end{cases}$$
$$= \begin{cases} \frac{1}{3}e^{-(t-2)/3} & t > 2, \\ 0 & \text{otherwise.} \end{cases}$$

Quantizer Example

- X is continuous, Y is discrete to convert into bits
- A rounding operation that involves loss of info



Quantizer Example Continued

- If X is uniform over *(-r/2,r/2)* and we are given *Y*
- What is the conditional distribution of *X* given *Y*?
- Knowing the value of *Y* just means *X* is in a certain region:

$$B_i = \{i\Delta \le X < (i+1)\Delta\}$$

$$\mathsf{P}[B_i] = \int_{i\Delta}^{(i+1)\Delta} f_X(x)\, dx = \frac{\Delta}{r} = \frac{1}{n}.$$

$$f_{X|B_i}(x) = \begin{cases} \dfrac{f_X(x)}{\mathsf{P}[B_i]} & x \in B_i, \\ 0 & \text{otherwise}, \end{cases}$$

$$= \begin{cases} 1/\Delta & i\Delta \le x < (i+1)\Delta, \\ 0 & \text{otherwise}. \end{cases}$$

Total Probability Theorem for RVs

- Let $B_1, \ldots, B_m$ be a partition of the sample space
- Then we have:

Discrete: $$P_X(x) = \sum_{i=1}^{m} P_{X|B_i}(x)\,\mathsf{P}[B_i]$$

Continuous: $$f_X(x) = \sum_{i=1}^{m} f_{X|B_i}(x)\,\mathsf{P}[B_i]$$

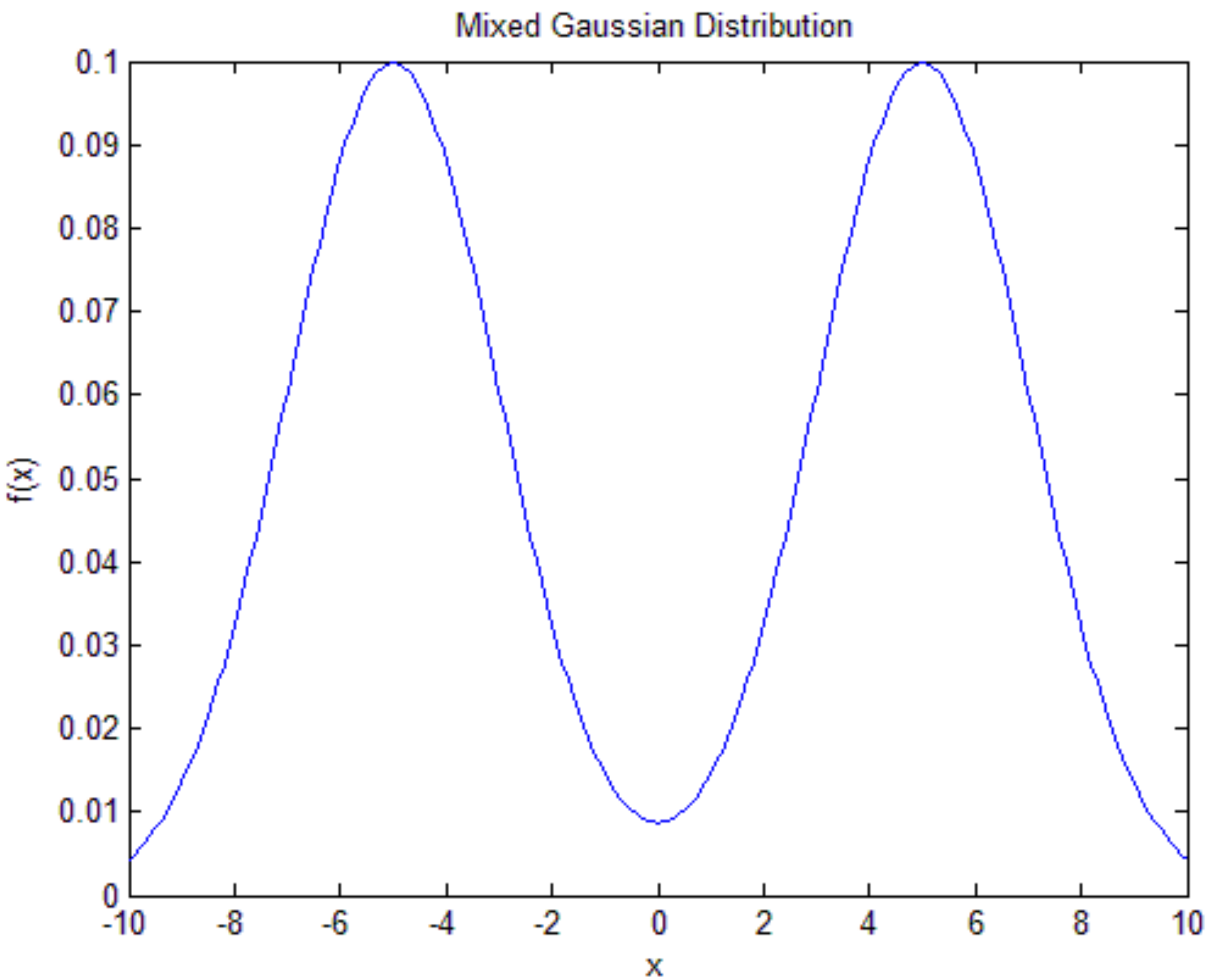
Example

Random variable X is a voltage at the receiver of a modem. When symbol "0" is transmitted (event B_0), X is the Gaussian $(-5, 2)$ random variable. When symbol "1" is transmitted (event B_1), X is the Gaussian $(5, 2)$ random variable. Given that symbols "0" and "1" are equally likely to be sent, what is the PDF of X?

$$\begin{aligned} f_X(x) &= f_{X|B_0}(x)\,\mathrm{P}[B_0] + f_{X|B_1}(x)\,\mathrm{P}[B_1] \\ &= \frac{1}{4\sqrt{2\pi}}\left(e^{-(x+5)^2/8} + e^{-(x-5)^2/8}\right). \end{aligned}$$

- Sum of two bell curves with different means and variances
- This is known as a "mixed Gaussian distribution"

Example Cont'



Conditional PMF of a RV Given Another

- Recall:
- $p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$
- $\sum_x \sum_y p_{X,Y}(x,y) =$
- $p_X(x) = \sum_y p_{X,Y}(x,y)$
- Conditional PMF of a RV given another: A version of Bayes' Rule:
- $p_{X|Y}(x \mid y) = \mathbf{P}(X = x \mid Y = y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$
- $\sum_x p_{X|Y}(x \mid y) =$

Conditional PDF of a RV Given Another

- Recall

$$\mathbf{P}(x \leq X \leq x + \delta) \approx f_X(x) \cdot \delta$$

- By analogy, would like:

$$\mathbf{P}(x \leq X \leq x + \delta \mid Y \approx y) \approx f_{X|Y}(x \mid y) \cdot \delta$$

- This leads us to the **definition**:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad \text{if } f_Y(y) > 0$$

- For given y, conditional PDF is a (normalized) "section" of the joint PDF

- If independent, $f_{X,Y} = f_X f_Y$, we obtain

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

Independent RVs

- Two RVs are independent if the joint dist. is the product of marginals
- Another definition: Conditioning makes no difference

$$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

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

$$P_{X|Y}(x|y) = P_X(x) \ , \forall y$$

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

Example

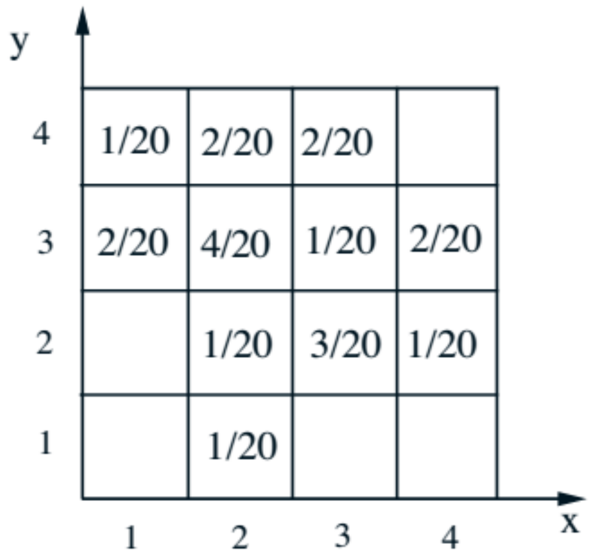
$$f_{X,Y}(x,y) = \begin{cases} 4xy & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 2y & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

YES!

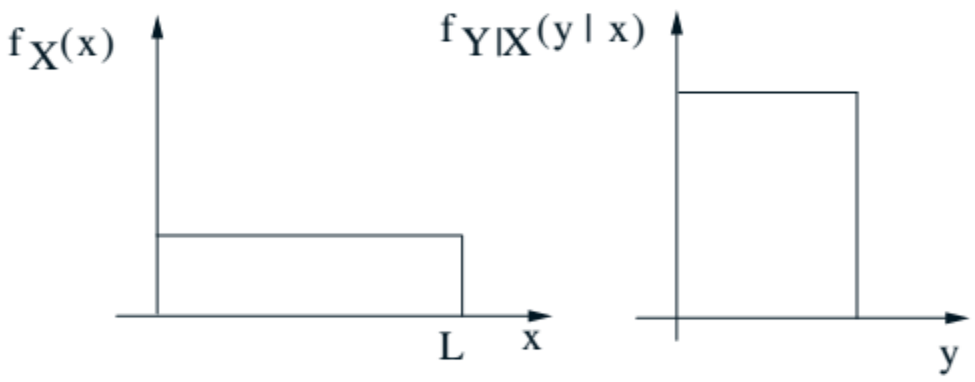
Example



- Independent?
- What if we condition on $X \leq 2$ and $Y \geq 3$?

Stick-Breaking

- Break a stick of length ℓ twice:
  break at X: uniform in $[0, 1]$;
  break again at Y, uniform in $[0, X]$



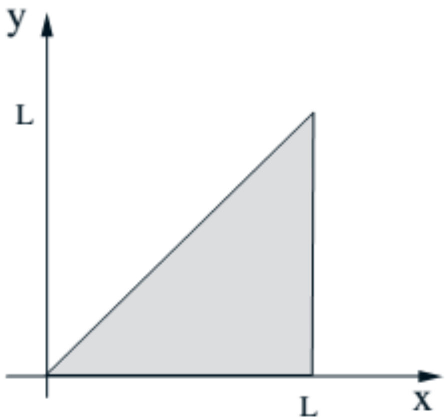
- What is the PDF of *Y* and its expectation?
- Example in conditional probability

Stick Breaking Cont'

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y \mid x) = \frac{1}{\ell x}, \qquad 0 \le y \le x \le \ell$$

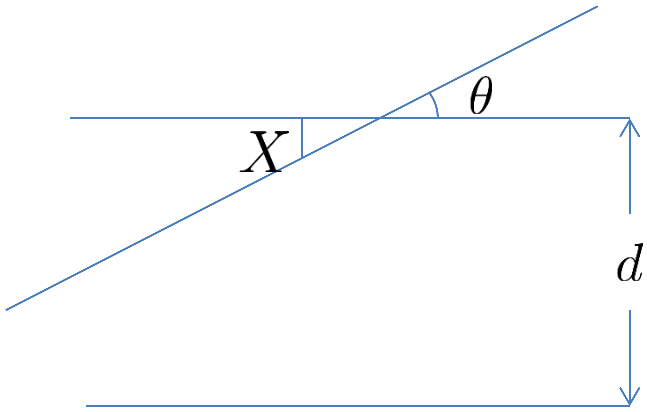
$$\begin{aligned} f_Y(y) &= \int f_{X,Y}(x,y)\,dx \\ &= \int_y^\ell \frac{1}{\ell x}\,dx \\ &= \frac{1}{\ell}\log\frac{\ell}{y}, \qquad 0 \le y \le \ell \end{aligned}$$

$$\mathbf{E}[Y] = \int_0^\ell y f_Y(y)\,dy = \int_0^\ell y \frac{1}{\ell}\log\frac{\ell}{y}\,dy = \frac{\ell}{4}$$



Buffon's Needle

- Parallel lines at distance d
  Needle of length ℓ (assume $\ell < d$)
- Find $\mathbf{P}$(needle intersects one of the lines)



- $X \in [0, d/2]$: distance of needle midpoint to nearest line

Buffon's Needle Cont'

- **Model:** X, Θ uniform, independent

  $f_{X,\Theta}(x,\theta) = \qquad 0 \le x \le d/2,\ 0 \le \theta \le \pi/2$

- Intersect if $X \le \frac{\ell}{2}\sin\Theta$

$$\begin{aligned}\mathbf{P}\left(X \le \frac{\ell}{2}\sin\Theta\right) &= \int\int_{x\le\frac{\ell}{2}\sin\theta} f_X(x) f_\Theta(\theta)\,dx\,d\theta \\ &= \frac{4}{\pi d}\int_0^{\pi/2}\int_0^{(\ell/2)\sin\theta} dx\,d\theta \\ &= \frac{4}{\pi d}\int_0^{\pi/2}\frac{\ell}{2}\sin\theta\,d\theta = \frac{2\ell}{\pi d}\end{aligned}$$

Recall Bayes Rule Discrete

- We have already seen

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

$$p_Y(y) = \sum_x p_X(x) p_{Y|X}(y \mid x)$$

Example:

- $X = 1, 0$: airplane present/not present
- $Y = 1, 0$: something did/did not register on radar

Bayes Rule Continuous

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

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

Example: X: some signal; "prior" $f_X(x)$

Y: noisy version of X

$f_{Y|X}(y \mid x)$: model of the noise

Discrete *X* Continuous *Y*

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

$$f_Y(y) = \sum_x p_X(x) f_{Y|X}(y \mid x)$$

Example:

- X: a discrete signal; "prior" $p_X(x)$
- Y: noisy version of X
- $f_{Y|X}(y \mid x)$: continuous noise model

Discrete *Y* Continuous *X*

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

$$p_Y(y) = \int_x f_X(x) p_{Y|X}(y \mid x)\, dx$$

Example:

- X: a continuous signal; "prior" $f_X(x)$ (e.g., intensity of light beam);
- Y: discrete r.v. affected by X (e.g., photon count)
- $p_{Y|X}(y \mid x)$: model of the discrete r.v.

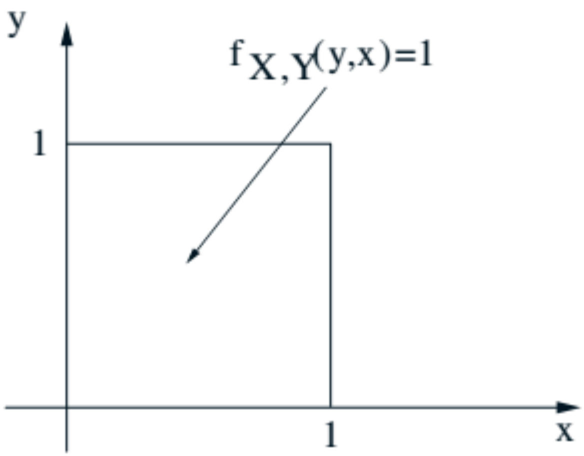
Derived RVs

- PMF or PDF of a function of one or more RVs
- Some texts categorize in terms of
  - How many RVs involved
  - Continuous vs discrete
- Example below is for two RVs, continuous

Obtaining the PDF for

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

involves deriving a distribution.
Note: $g(X,Y)$ is a random variable

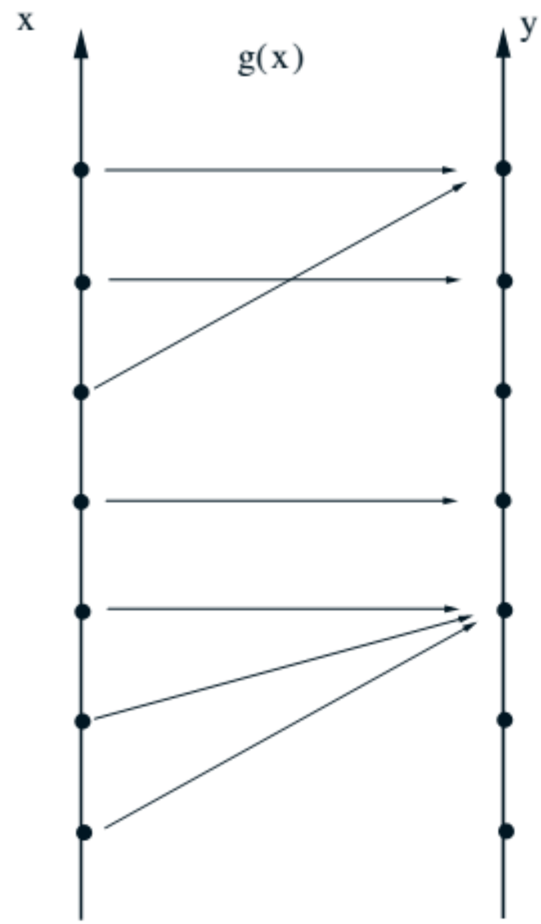


Derived RVs

- Do not need to find the distribution to compute
  - Expectations
  - Variances
  - Moments etc.

Discrete Case: One RV

- Obtain probability mass for each possible value of $Y = g(X)$



$$
\begin{aligned}
p_Y(y) &= \mathbf{P}(g(X) = y) \\
&= \sum_{x:\, g(x)=y} p_X(x)
\end{aligned}
$$

Discrete Case: Two RVs

- Similar to the single RV case:

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

$$P_W(w) = \sum_{(x,y):g(x,y)=w} P_{X,Y}(x, y)$$

- Multiple RV case can be similarly obtained

Continuous Case

- **Two-step procedure:**
  - Get CDF of Y: $F_Y(y) = \mathbf{P}(Y \leq y)$
  - Differentiate to get

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

Example

- X: uniform on [0,2]
- Find PDF of $Y = X^3$

- **Solution:**

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \le y) = \mathbf{P}(X^3 \le y) \\ &= \mathbf{P}(X \le y^{1/3}) = \frac{1}{2} y^{1/3} \end{aligned}$$

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{1}{6y^{2/3}}$$

Example: Squaring a Uniform RV

- *X* is a uniform RV over (-1,1), and $W = X^2$
- Note that *W* is always positive, even though *X* isn't

$$F_W(w) = \mathsf{P}\left[X^2 \le w\right] = \mathsf{P}\left[-\sqrt{w} \le X \le \sqrt{w}\right]$$

$$= \int_{-\sqrt{w}}^{\sqrt{w}} f_X(x)\ dx$$

Example: Square Root

- RV $Y = \sqrt{X}$; *X* is exponentially distributed $F_X(x) = 1 - e^{-\lambda x}.$

$$
\begin{aligned}
F_Y(y) &= \mathsf{P}\left[\sqrt{X} \le y\right] \\
&= \mathsf{P}\left[X \le y^2\right] = F_X\left(y^2\right).
\end{aligned}
$$

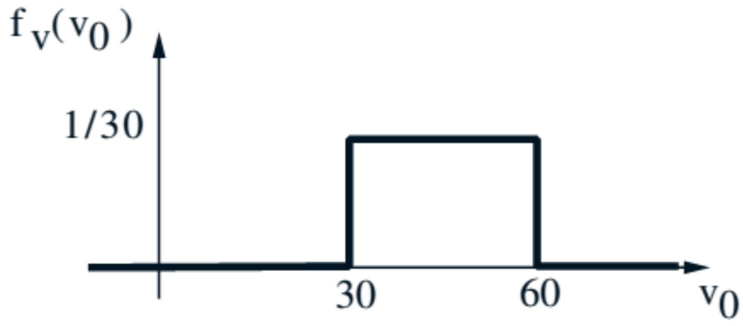
$$
F_Y(y) = \begin{cases} 1 - e^{-\lambda y^2} & y \ge 0 \\ 0 & \text{otherwise} \end{cases}
$$

$$
f_Y(y) = \begin{cases} 2\lambda y e^{-\lambda y^2} & y \ge 0 \\ 0 & \text{otherwise} \end{cases}
$$

- Application in wireless comms

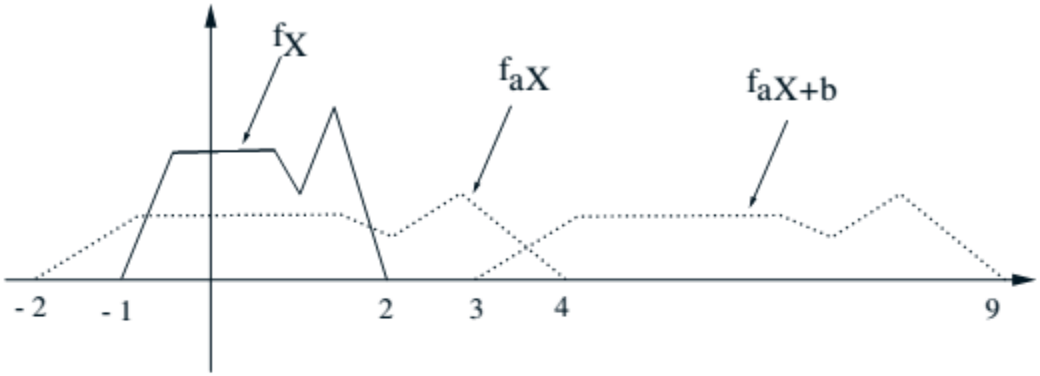
Example

- Joan is driving from Boston to New York. Her speed is uniformly distributed between 30 and 60 mph. What is the distribution of the duration of the trip?
- Let $T(V) = \dfrac{200}{V}$.
- Find $f_T(t)$



PDF of Linear and Affine Functions

$Y = 2X + 5$

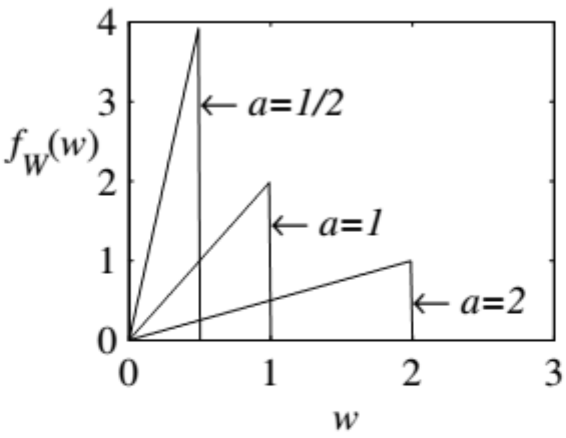


$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Example

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$W = aX$$



$$f_W(w) = \frac{1}{a} f_X(w/a)$$
$$= \begin{cases} 2w/a^2 & 0 \le w \le a, \\ 0 & \text{otherwise.} \end{cases}$$

As a increases, the PDF stretches horizontally.

What if X is Normal?

- $Y = aX + b$

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

- Y is also normal! What are the mean and variance of Y?

Stretching Other Distributions

$W = aX$, where $a > 0$.

(a) If X is uniform (b, c), then W is uniform (ab, ac).

(b) If X is exponential (λ), then W is exponential (λ/a).

(c) If X is Erlang (n, λ), then W is Erlang $(n, \lambda/a)$.

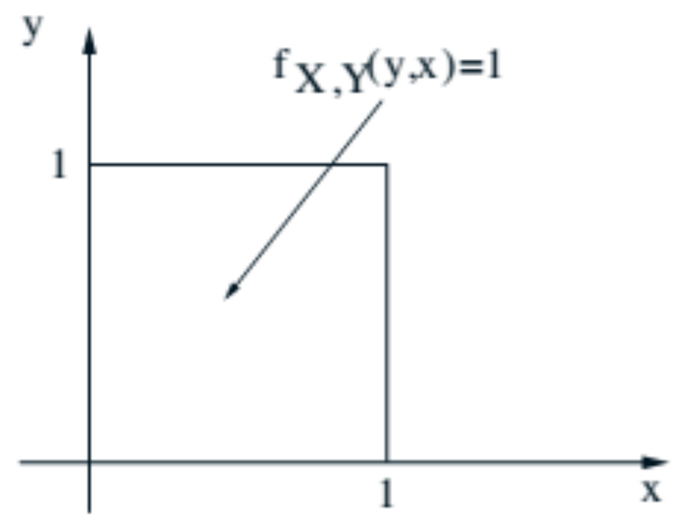
(d) If X is Gaussian (μ, σ), then W is Gaussian $(a\mu, a\sigma)$.

CDF of a Ratio of Two RVs

Find the PDF of $Z = g(X, Y) = Y/X$

$F_Z(z) =$ $\qquad z \leq 1$

$F_Z(z) =$ $\qquad z \geq 1$



A General Formula

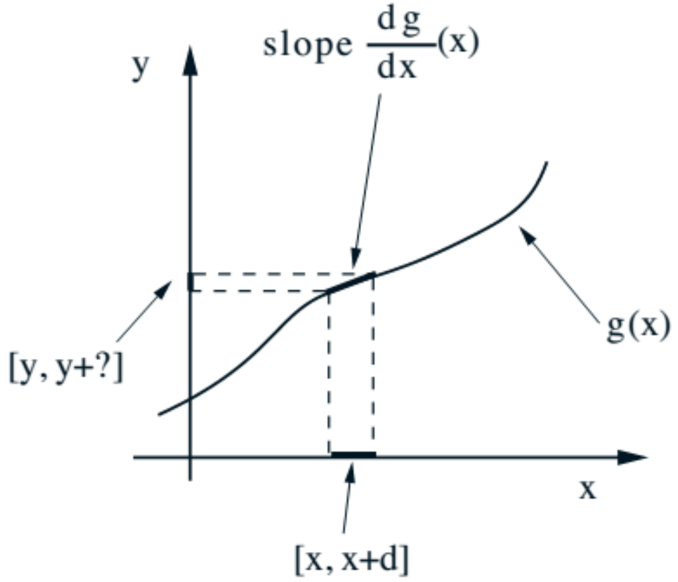
- Let $Y = g(X)$
  g strictly monotonic.

- Event $x \leq X \leq x + \delta$ is the same as
  $g(x) \leq Y \leq g(x + \delta)$
  or (approximately)
  $g(x) \leq Y \leq g(x) + \delta |(dg/dx)(x)|$

- Hence,

$$\delta f_X(x) = \delta f_Y(y) \left| \frac{dg}{dx}(x) \right|$$

where $y = g(x)$



Practical Example

- How to generate a RV with desired distribution
- *U* is uniform over [0,1], how do we turn it into an exponential RV?

$$X = g(U)$$

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-x} & x \geq 0. \end{cases}$$

$$g(U) = -\ln(1 - U) \quad = \quad F_X^{-1}(U)$$

- Can transform a uniform RV into ***any*** distribution using inverse CDF!

Example

- Ratio of two independent exponential RVs

$$f_{X,Y}(x,y) = \begin{cases} \lambda\mu e^{-(\lambda x+\mu y)} & x \geq 0, y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

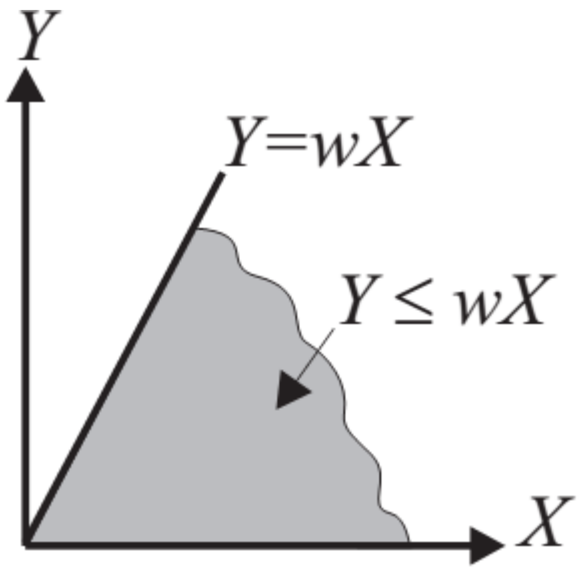
Find the PDF of $W = Y/X$

Solution

- Ratio of two independent exponential RVs

$$F_W(w) = \mathsf{P}[Y/X \le w] = \mathsf{P}[Y \le wX].$$

$$\begin{aligned}
\mathsf{P}[Y \le wX] &= \int_0^\infty \left(\int_0^{wx} f_{X,Y}(x,y)\, dy \right) dx \\
&= \int_0^\infty \lambda e^{-\lambda x} \left(\int_0^{wx} \mu e^{-\mu y}\, dy \right) dx \\
&= \int_0^\infty \lambda e^{-\lambda x} \left(1 - e^{-\mu w x}\right) dx \\
&= 1 - \frac{\lambda}{\lambda + \mu w}.
\end{aligned}$$



Solution Cont'

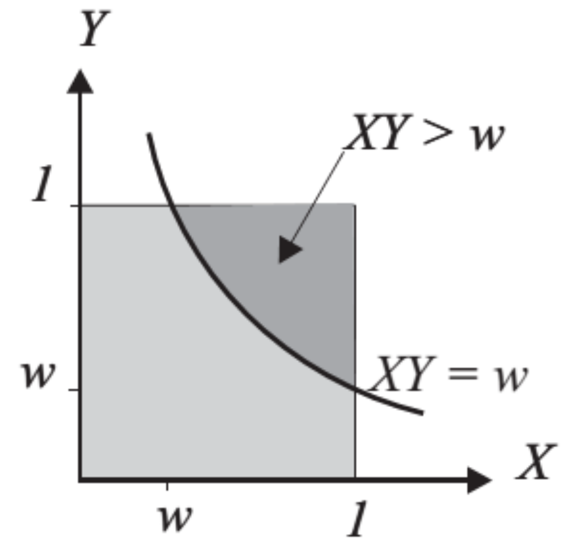
- Ratio of two independent exponential RVs

$$F_W(w) = \begin{cases} 0 & w < 0, \\ 1 - \dfrac{\lambda}{\lambda + \mu w} & \omega \geq 0. \end{cases}$$

$$f_W(w) = \begin{cases} \dfrac{\lambda\mu}{(\lambda + \mu w)^2} & w \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Example: Product of Two RVs

- Product of two RVs $W = XY$



$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

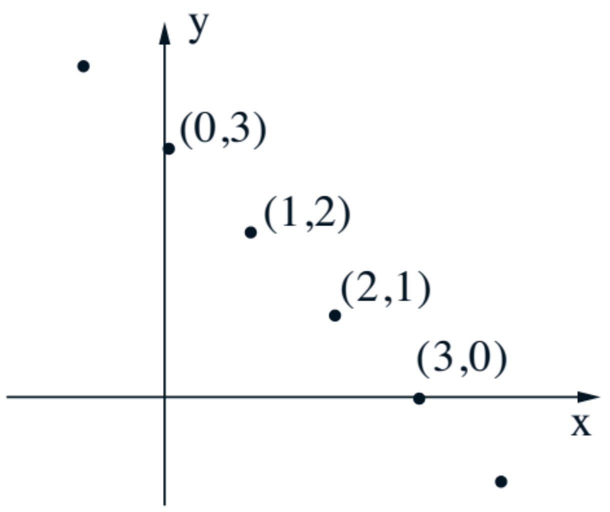
$$\begin{aligned} F_W(w) &= 1 - \mathrm{P}\left[XY > w\right] \\ &= 1 - \int_w^1 \int_{w/x}^1 dy\, dx \\ &= 1 - \int_w^1 (1 - w/x)\, dx \\ &= \left(x - w \ln x |_{x=w}^{x=1}\right) \\ &= w - w \ln w. \end{aligned}$$

$$f_W(w) = \frac{dF_W(w)}{dw} = -\ln w.$$

$$f_W(w) = \begin{cases} 0 & w < 0, \\ -\ln w & 0 \le w \le 1, \\ 0 & w > 1. \end{cases}$$

The Distribution of a Sum

- $W = X + Y$; X, Y independent



$$\begin{aligned} p_W(w) &= \mathbf{P}(X + Y = w) \\ &= \sum_x \mathbf{P}(X = x)\mathbf{P}(Y = w - x) \\ &= \sum_x p_X(x) p_Y(w - x) \end{aligned}$$

- This is convolution of the PMFs!

Recall Convolution

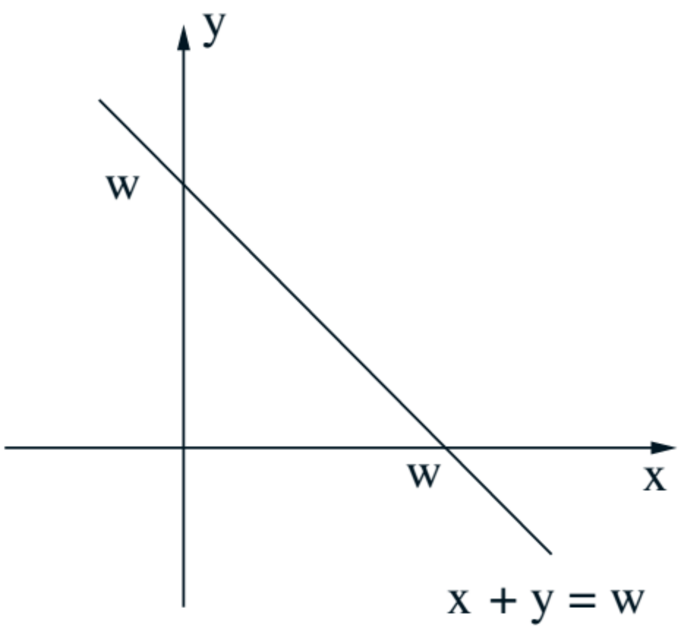
- Mechanics:
  - Put the pmf's on top of each other
  - Flip the pmf of Y
  - Shift the flipped pmf by w
    (to the right if $w > 0$)
  - Cross-multiply and add

Continuous Case

- $W = X + Y$; X, Y independent

- $f_{W|X}(w \mid x) = f_Y(w - x)$

- $f_{W,X}(w, x) = f_X(x) f_{W|X}(w \mid x)$
  $= f_X(x) f_Y(w - x)$



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

Example

$$f_X(x) = \begin{cases} 3e^{-3x} & x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 2e^{-2y} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Find the density of *X+Y* if *X* and *Y* are independent
- Convolution:

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} f_X(w-y)\, f_Y(y)\, dy \\ &= 6 \int_0^w e^{-3(w-y)} e^{-2y}\, dy = 6e^{-3w} \int_0^w e^y\, dy = 6e^{-3w}\left(e^w - 1\right) \end{aligned}$$

$$f_W(w) = \begin{cases} 6(e^{-2w} - e^{-3w}) & w \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

Two Independent Normals

- $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$, independent

$$
\begin{aligned}
f_{X,Y}(x, y) &= f_X(x) f_Y(y) \\
&= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{ -\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} \right\}
\end{aligned}
$$

- PDF is constant on the ellipse where

$$
\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2}
$$

is constant

- Ellipse is a circle when $\sigma_x = \sigma_y$

Their Sum

- $X \sim N(0, \sigma_x^2)$, $Y \sim N(0, \sigma_y^2)$, independent

- Let $W = X + Y$

$$
\begin{aligned}
f_W(w) &= \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\, dx \\
&= \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} e^{-x^2/2\sigma_x^2} e^{-(w-x)^2/2\sigma_y^2}\, dx \\
\text{(algebra)} &= c e^{-\gamma w^2}
\end{aligned}
$$

- Conclusion: W is normal
  - mean=0, variance=$\sigma_x^2 + \sigma_y^2$
  - same argument for nonzero mean case

Moment Generating Function

- The MGF of a RV X is defined as

$$\phi_X(s) = \mathsf{E}\left[e^{sX}\right]$$

- Continuous

$$\phi_X(s) = \int_{-\infty}^{\infty} f_X(x)\, e^{sx} dx$$

- Discrete

$$\phi_X(s) = \sum_{x=-\infty}^{\infty} p_X(x)\, e^{sx} dx$$

Generating Moments

- The MGF of a RV X is defined as

$$\phi_X(s) = \mathsf{E}\left[e^{sX}\right]$$

- Easy to see that

$$\mathsf{E}\left[X^n\right] = \left.\frac{d^n \phi_X(s)}{ds^n}\right|_{s=0}$$

Discrete MGFs

Bernoulli (p)	$P_X(x)$	$=$	$\begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{otherwise} \end{cases}$	$1-p+pe^s$
Binomial (n, p)	$P_X(x)$	$=$	$\binom{n}{x} p^x (1-p)^{n-x}$	$(1-p+pe^s)^n$
Geometric (p)	$P_X(x)$	$=$	$\begin{cases} p(1-p)^{x-1} & x=1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{pe^s}{1-(1-p)e^s}$
Pascal (k, p)	$P_X(x)$	$=$	$\binom{x-1}{k-1} p^k (1-p)^{x-k}$	$(\dfrac{pe^s}{1-(1-p)e^s})^k$
Poisson (α)	$P_X(x)$	$=$	$\begin{cases} \alpha^x e^{-\alpha}/x! & x=0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$	$e^{\alpha(e^s-1)}$
Disc. Uniform (k, l)	$P_X(x)$	$=$	$\begin{cases} \frac{1}{l-k+1} & x=k,k+1,\ldots,l \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{sk}-e^{s(l+1)}}{1-e^s}$

Continuous MGFs

Constant (a)	$f_X(x)$	$=$	$\delta(x-a)$	e^{sa}
Uniform (a, b)	$f_X(x)$	$=$	$\begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{bs} - e^{as}}{s(b-a)}$
Exponential (λ)	$f_X(x)$	$=$	$\begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{\lambda}{\lambda - s}$
Erlang (n, λ)	$f_X(x)$	$=$	$\begin{cases} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$(\dfrac{\lambda}{\lambda - s})^n$
Gaussian (μ, σ)	$f_X(x)$	$=$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$	$e^{s\mu + s^2\sigma^2/2}$

Example

- Find moments of exp. RV with parameter λ

$$\mathsf{E}[X] = \left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \left.\frac{\lambda}{(\lambda - s)^2}\right|_{s=0} = \frac{1}{\lambda}$$

$$\mathsf{E}\left[X^2\right] = \left.\frac{d^2\phi_X(s)}{ds^2}\right|_{s=0} = \left.\frac{2\lambda}{(\lambda - s)^3}\right|_{s=0} = \frac{2}{\lambda^2}$$

$$\mathsf{E}[X^n] = \left.\frac{d^n\phi_X(s)}{ds^n}\right|_{s=0} = \left.\frac{n!\lambda}{(\lambda - s)^{n+1}}\right|_{s=0} = \frac{n!}{\lambda^n}$$

MGF of a Sum of Independent RVs

- Let $W = X_1 + \cdots + X_n$ with X_i independent

$$\phi_W(s) = \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s)$$

- If also identically distributed

$$\phi_W(s) = [\phi_X(s)]^n$$

- Useful to prove asymptotic results about sums of RVs
- Useful to prove distributions of sums of RVs

Sum of Independent Poisson RVs

- Let $W = K_1 + \cdots + K_n$ with K_i independent, Poisson with $E[K_i] = \alpha_i$

$$
\begin{aligned}
\phi_{K_i}(s) &= e^{\alpha_i(e^s-1)} \\
\phi_W(s) &= e^{\alpha_1(e^s-1)} e^{\alpha_2(e^s-1)} \ldots e^{\alpha_n(e^s-1)} \\
&= e^{(\alpha_1+\cdots+\alpha_n)(e^s-1)} \\
&= e^{(\alpha_T)(e^s-1)}
\end{aligned}
$$

$$
\alpha_T = \alpha_1 + \cdots + \alpha_n
$$

- Sum of independent Poisson RVs is a Poisson

Sum of Independent Gaussian RVs

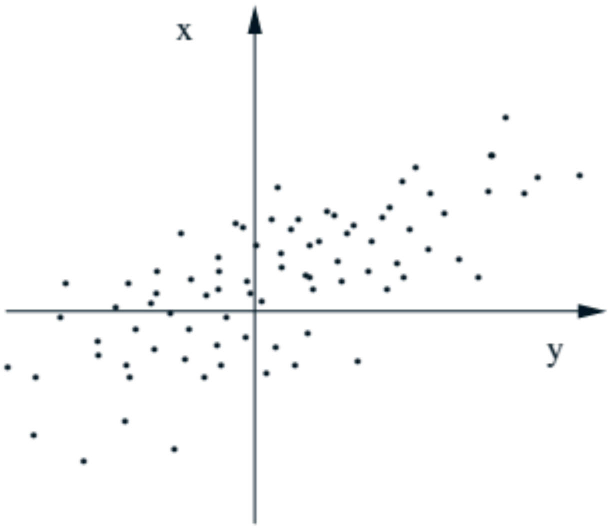
- Let $W = X_1 + \cdots + X_n$ with X_i independent, Gaussian with
- $E[X_i] = \mu_i \quad \sigma_i^2 = Var[X_i]$

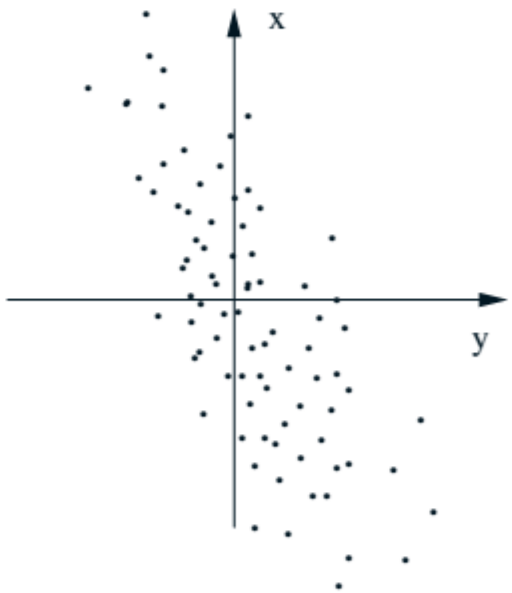
$$
\begin{aligned}
\phi_W(s) &= \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s) \\
&= e^{s\mu_1+\sigma_1^2 s^2/2} e^{s\mu_2+\sigma_2^2 s^2/2} \ldots e^{s\mu_n+\sigma_n^2 s^2/2} \\
&= e^{s(\mu_1+\cdots+\mu_n)+(\sigma_1^2+\cdots+\sigma_n^2)s^2/2}.
\end{aligned}
$$

- Sum of independent Gaussian RVs is a Gaussian

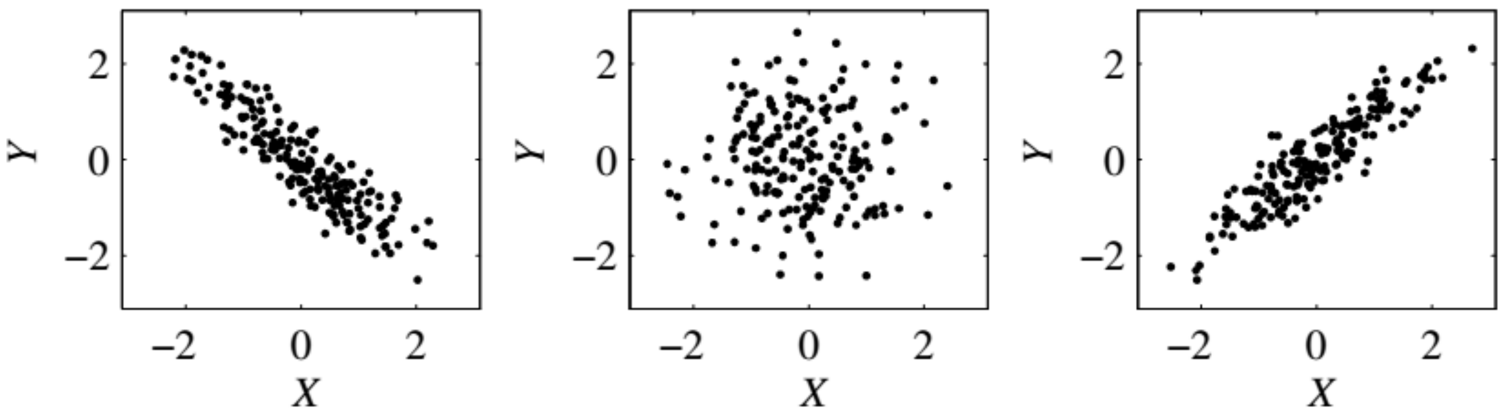
Covariance

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





Covariance



(a) $\rho_{X,Y} = -0.9$ **(b)** $\rho_{X,Y} = 0$ **(c)** $\rho_{X,Y} = 0.9$

Each graph has 200 samples, each marked by a dot, of the random variable pair (X, Y) such that $\mathsf{E}[X] = \mathsf{E}[Y] = 0$, $\mathsf{Var}[X] = \mathsf{Var}[Y] = 1$.

Covariance Cont'

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

- Zero-mean case: $\text{cov}(X,Y) = \mathbf{E}[XY]$
- $\text{cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$
- $\text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{var}(X_i) + 2 \sum_{(i,j):i \neq j} \text{cov}(X_i, X_j)$
- independent $\Rightarrow$ $\text{cov}(X,Y) = 0$
  (converse is not true)

Correlation Coefficient

- Dimensionless and normalized

$$\rho = \mathbf{E}\left[\frac{(X - \mathbf{E}[X])}{\sigma_X} \cdot \frac{(Y - \mathbf{E}[Y])}{\sigma_Y}\right] = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

- $-1 \leq \rho \leq 1$
- $|\rho| = 1 \Leftrightarrow (X - \mathbf{E}[X]) = c(Y - \mathbf{E}[Y])$
  (linearly related)
- Independent $\Rightarrow \rho = 0$
  (converse is not true)

Correlation Coefficient Practical Examples

- X is a student's height. Y is the same student's weight. $0 < \rho_{X,Y} < 1$.
- X is the distance of a cellular phone from the nearest base station. Y is the power of the received signal at the cellular phone. $-1 < \rho_{X,Y} < 0$.
- X is the temperature of a resistor measured in degrees Celsius. Y is the temperature of the same resistor measured in Kelvins. $\rho_{X,Y} = 1$.
- X is the gain of an electrical circuit measured in decibels. Y is the attenuation, measured in decibels, of the same circuit. $\rho_{X,Y} = -1$.
- X is the telephone number of a cellular phone. Y is the Social Security number of the phone's owner. $\rho_{X,Y} = 0$.

For Independent RVs

$$\mathsf{E}[g(X)h(Y)] = \mathsf{E}[g(X)]\,\mathsf{E}[h(Y)]$$

$$\mathsf{Cov}[X, Y] = \rho_{X,Y} = 0$$

$$\mathsf{Var}[X + Y] = \mathsf{Var}[X] + \mathsf{Var}[Y]$$

Covariance Between Gaussians

- Consider: signal X is Gaussian $(0, \sigma_X)$

  noise Z is Gaussian $(0, \sigma_Z)$

- *X* and *Z* are independent. Find $\rho_{X,Y}$ re $Y = X + Z$

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}$$

$$= \frac{\sigma_X^2}{\sqrt{\sigma_X^2(\sigma_X^2 + \sigma_Z^2)}} = \sqrt{\frac{\sigma_X^2/\sigma_Z^2}{1 + \sigma_X^2/\sigma_Z^2}}.$$

Bivariate Gaussians

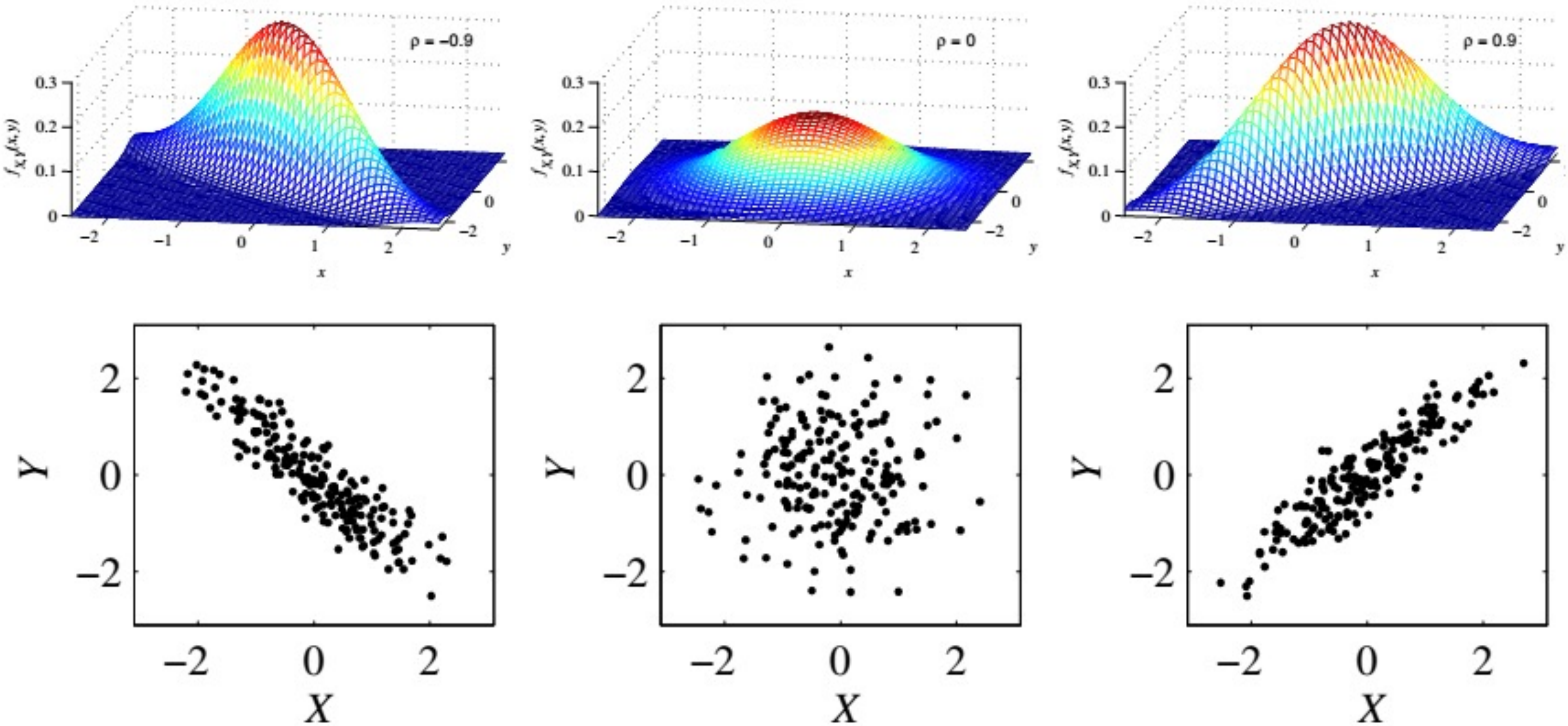
Random variables X *and* Y *have a* bivariate Gaussian PDF *with parameters* μ_X, μ_Y, $\sigma_X > 0$, $\sigma_Y > 0$, *and* $\rho_{X,Y}$ *satisfying* $-1 < \rho_{X,Y} < 1$ *if*

$$f_{X,Y}(x,y) = \frac{\exp\left[-\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho_{X,Y}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{2(1-\rho_{X,Y}^2)}\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}},$$

Fact:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}}e^{-(x-\mu_X)^2/2\sigma_X^2}, \qquad f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}}e^{-(y-\mu_Y)^2/2\sigma_Y^2}$$

Bivariate Gaussians



The Joint Gaussian PDF $f_{X,Y}(x,y)$ for $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and three values of $\rho_{X,Y} = \rho$. Next to each PDF, we plot 200 sample pairs (X, Y) generated with that PDF.

Multivariate Random Vectors

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \mathsf{P}\left[X_1 \leq x_1,\ldots,X_n \leq x_n\right]$$

Discrete:

$$P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \mathsf{P}\left[X_1 = x_1,\ldots,X_n = x_n\right]$$

Continuous:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{\partial x_1 \cdots \partial x_n}$$

I.I.D. Multivariate Random Vectors

$X_1, \ldots, X_n$ *are* independent and identically distributed (iid) *if*

Discrete: $P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P_X(x_1)P_X(x_2)\cdots P_X(x_n)$

Continuous: $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_X(x_1)f_X(x_2)\cdots f_X(x_n)$.

Example

The random variables $X_1, \ldots, X_n$ have the joint PDF

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \begin{cases} 1 & 0 \le x_i \le 1, i = 1,\ldots,n, \\ 0 & \text{otherwise.} \end{cases}$$

Let A denote the event that $\max_i X_i \le 1/2$. Find $\mathsf{P}[A]$.

$$\begin{aligned} \mathsf{P}[A] = \mathsf{P}\left[\max_i X_i \le 1/2\right] &= \mathsf{P}[X_1 \le 1/2, \ldots, X_n \le 1/2] \\ &= \int_0^{1/2} \cdots \int_0^{1/2} 1\, dx_1 \cdots dx_n = \frac{1}{2^n}. \end{aligned}$$

Conditional Expectation

- Given the value y of a r.v. Y:

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

  (integral in continuous case)

- Stick example: stick of length ℓ
  break at uniformly chosen point Y
  break again at uniformly chosen point X

- $\mathbf{E}[X \mid Y = y] = \dfrac{y}{2}$ (number)

$$\mathbf{E}[X \mid Y] = \frac{Y}{2} \quad \text{(r.v.)}$$

Gaussian Case

$$f_{X,Y}(x,y) = \frac{\exp\left[-\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho_{X,Y}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{2(1-\rho_{X,Y}^2)}\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}$$

$$f_{X|Y}(x|y) = \frac{1}{\tilde{\sigma}_X\sqrt{2\pi}} e^{-(x-\tilde{\mu}_X(y))^2/2\tilde{\sigma}_X^2}$$

$$\mathsf{E}[X|Y=y] = \tilde{\mu}_X(y) = \mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)$$

$$\mathsf{Var}[X|Y=y] = \tilde{\sigma}_X^2 = \sigma_X^2(1-\rho^2).$$

Gaussian Case Cont'

$$\mathrm{Var}\,[Y|X=x] = \sigma_Y^2(1-\rho_{X,Y}^2) \le \sigma_Y^2,$$
$$\mathrm{Var}\,[X|Y=y] = \sigma_X^2(1-\rho_{X,Y}^2) \le \sigma_X^2.$$

These formulas state that for $\rho_{X,Y} \neq 0$, learning the value of one of the random variables leads to a model of the other random variable with reduced variance.

Iterated Expectations

$$\mathbf{E}[\mathbf{E}[X \mid Y]] = \sum_y \mathbf{E}[X \mid Y = y] p_Y(y) = \mathbf{E}[X]$$

- In stick example:

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

Conditional Variance

- $\text{var}(X \mid Y = y) = \mathbf{E}\left[(X - \mathbf{E}[X \mid Y = y])^2 \mid Y = y\right]$
- $\text{var}(X \mid Y)$: a r.v.
  with value $\text{var}(X \mid Y = y)$ when $Y = y$

- **Law of total variance:**

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

Proof

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

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

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

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

Sum of right-hand sides of (c), (d):
$\mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \text{var}(X)$

Sum of a Random Number of RVs

- N: number of stores visited
  (N is a nonnegative integer r.v.)

- X_i: money spent in store i
  - X_i assumed i.i.d.
  - independent of N

- Let $Y = X_1 + \cdots + X_N$

Sum of a Random Number of RVs Cont'

$$\begin{aligned}
\mathbf{E}[Y \mid N = n] &= \mathbf{E}[X_1 + X_2 + \cdots + X_n \mid N = n] \\
&= \mathbf{E}[X_1 + X_2 + \cdots + X_n] \\
&= \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n] \\
&= n\,\mathbf{E}[X]
\end{aligned}$$

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

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

Variance of a Random Sum

- $\text{var}(Y) = \mathbf{E}[\text{var}(Y \mid N)] + \text{var}(\mathbf{E}[Y \mid N])$

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

- $\text{var}(Y \mid N = n) = n\,\text{var}(X)$
  $\text{var}(Y \mid N) = N\,\text{var}(X)$
  $\mathbf{E}[\text{var}(Y \mid N)] = \mathbf{E}[N]\,\text{var}(X)$

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

Law of Large Numbers and the Central Limit Theorem

- The LLN and CLT are about convergence of sums of a large # of RVs
- LLN is about the "average" (sum / # of RVs)
- CLT involves a different normalization
- Both are fundamental in probability theory and practice
- We first overview of the regular notion of "convergence of sequences"

Chebyshev's Inequality

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

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

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

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

Deterministic Limits

- Sequence a_n
  Number a

- a_n converges to a

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

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

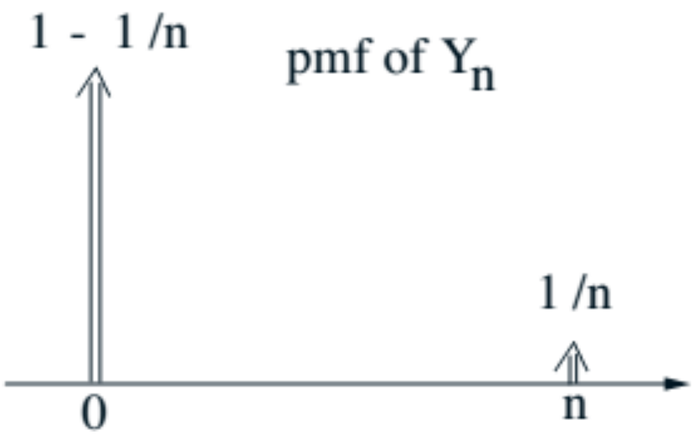
- For every $\epsilon > 0$,
  there exists n_0,
  such that for every $n \geq n_0$,
  we have $|a_n - a| \leq \epsilon$.

Convergence “in probability”

- Sequence of random variables Y_n
- converges in probability to a number a:
  “(almost all) of the PMF/PDF of Y_n ,
  eventually gets concentrated
  (arbitrarily) close to a”
- For every $\epsilon > 0$,

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

Example



- What is $E[Y_n]$?
- Does Y_n converge?

Weak Law of Large Numbers

- AKA convergence of the sample mean
- $X_1, X_2, \ldots$ are i.i.d. with finite mean μ and variance σ^2

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

- $\mathbf{E}[M_n] =$

- $\mathsf{Var}(M_n) =$

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

- M_n converges in probability to μ

Polling

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

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

- $M_n = (X_1 + \cdots + X_n)/n$
  fraction of "yes" in our sample
- Goal: 95% confidence of $\leq$1% error

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

Polling Cont'

- Use Chebyshev's inequality:

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

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

Different Scalings of the Sum

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

- Look at three variants of their sum:

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

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

- $\dfrac{S_n}{\sqrt{n}}$ constant variance σ^2
  - Asymptotic shape?

Central Limit Theorem

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

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

  - zero mean
  - unit variance

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

- **Theorem:** For every c:

$$\mathbf{P}(Z_n \leq c) \to \mathbf{P}(Z \leq c)$$

Profundity and Usefulness

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

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

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

Back to the Polls

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

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

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

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

Apply CLT

- Suppose we want:

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

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

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

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

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

Binomial is a Sum of RVs

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

Example

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

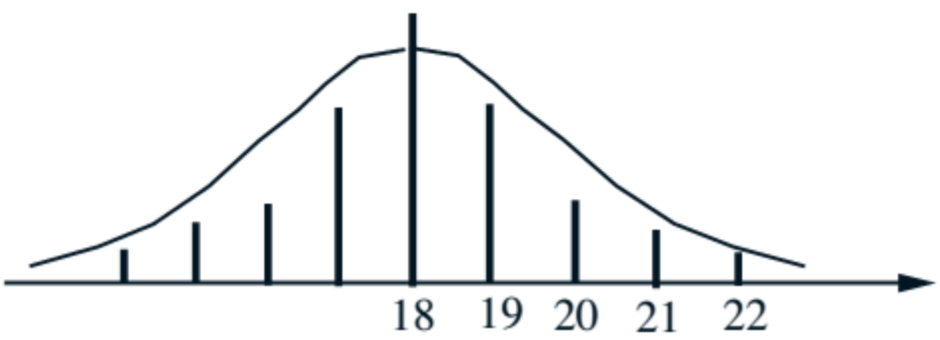
- Exact answer:

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

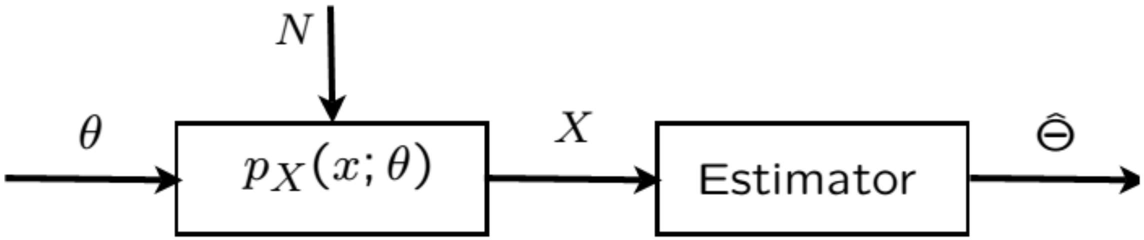
Correction for PMF Approximation

- $\mathbf{P}(S_n \leq 21) = \mathbf{P}(S_n < 22)$,
  because S_n is integer

- Compromise: consider $\mathbf{P}(S_n \leq 21.5)$



Introduction to Classical Statistics



- θ contains all unknown parameters
- X contains all *N* measurements
- "Estimator" processes the measurements to approximate θ
- If θ comes from a known discrete set then we have "detection"

Detection and Estimation

- **Problem types:**
  - Hypothesis testing:
    $H_0: \ \theta = 1/2$ versus $H_1: \ \theta = 3/4$
  - Composite hypotheses:
    $H_0: \ \theta = 1/2$ versus $H_1: \ \theta \neq 1/2$
  - Estimation: design an **estimator** $\hat{\Theta}$,
    to keep estimation **error** $\hat{\Theta} - \theta$ small

Maximum Likelihood Estimation

- Model, with unknown parameter(s):
  $X \sim p_X(x; \theta)$

- Pick θ that "makes data most likely"

$$\hat{\theta}_{\text{ML}} = \arg\max_{\theta} p_X(x; \theta)$$

- "argmax" means the value of θ that maximizes

Example

- $X_1, \ldots, X_n$: i.i.d., exponential(θ)

$$p_{X_i}(x_i) = \theta e^{-\theta x_i}$$

$$p_X(x) = \prod_{i=1}^{n} \theta e^{-\theta x_i}$$

$$\arg\max_\theta p_X(x) = \arg\max_\theta \prod_{i=1}^{n} \theta e^{-\theta x_i}$$

$$= \arg\max_\theta \left(n \log \theta - \theta \sum_{i=1}^{n} x_i \right)$$

$$\hat{\theta}_{\mathrm{ML}} = \frac{n}{x_1 + \cdots + x_n} \qquad \hat{\Theta}_n = \frac{n}{X_1 + \cdots + X_n}$$

Good Estimators

- **Unbiased:** $\mathbf{E}[\hat{\Theta}_n] = \theta$
  - exponential example, with $n = 1$:
    $\mathbf{E}[1/X_1] = \infty \neq \theta$
    (biased)
- **Consistent:** $\hat{\Theta}_n \to \theta$ (in probability)
  - exponential example:
    $(X_1 + \cdots + X_n)/n \to \mathbf{E}[X] = 1/\theta$
  - can use this to show that:
    $\hat{\Theta}_n = n/(X_1 + \cdots + X_n) \to 1/\mathbf{E}[X] = \theta$
- **"Small" mean squared error (MSE)**

$$\begin{aligned}\mathbf{E}[(\hat{\Theta} - \theta)^2] &= \text{var}(\hat{\Theta} - \theta) + (\mathbf{E}[\hat{\Theta} - \theta])^2 \\ &= \text{var}(\hat{\Theta}) + (\text{bias})^2\end{aligned}$$

Example: Estimating the Expectation

- $X_1, \ldots, X_n$: i.i.d., mean θ, variance σ^2

$X_i = \theta + W_i$

W_i: i.i.d., mean, 0, variance σ^2

$$\hat{\Theta}_n = \text{sample mean} = M_n = \frac{X_1 + \cdots + X_n}{n}$$

Properties of the "Sample Mean"

Properties:

- $\mathbf{E}[\widehat{\Theta}_n] = \theta$ (unbiased)
- WLLN: $\widehat{\Theta}_n \to \theta$ (consistency)
- MSE: σ^2/n

- Sample mean often turns out to also be the ML estimate.
  E.g., if $X_i \sim N(\theta, \sigma^2)$, i.i.d.

Confidence Intervals (CI)

- An estimate $\hat{\Theta}_n$ may not be informative enough

- An $1-\alpha$ **confidence interval** is a (random) interval $[\hat{\Theta}_n^-, \hat{\Theta}_n^+]$,

  s.t. $\quad \mathbf{P}(\hat{\Theta}_n^- \leq \theta \leq \hat{\Theta}_n^+) \geq 1-\alpha, \quad \forall\ \theta$

  – often $\alpha = 0.05$, or 0.25, or 0.01

Example

- CI in estimation of the mean

$$\hat{\Theta}_n = (X_1 + \cdots + X_n)/n$$

normal tables: $\Phi(1.96) = 1 - 0.05/2$

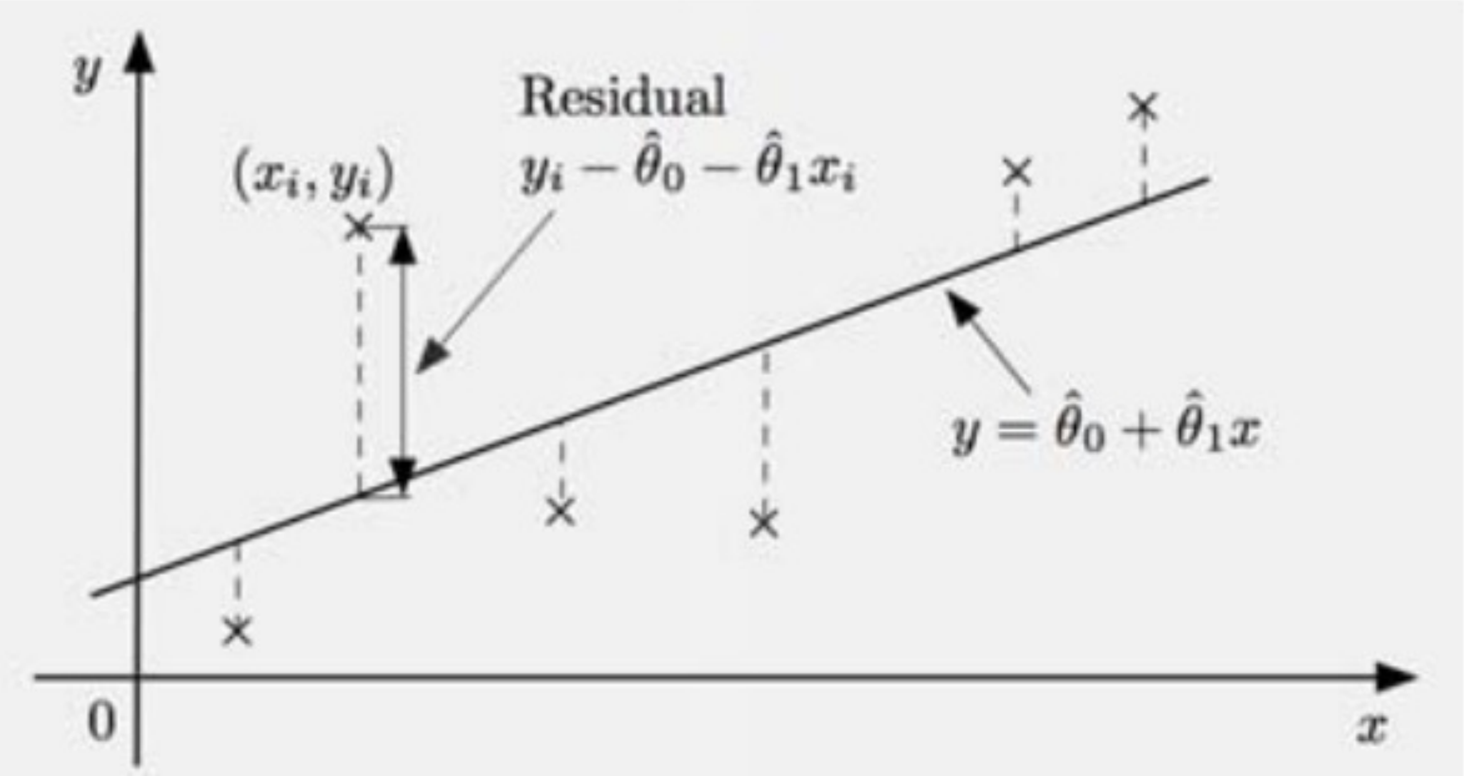
$$\mathbf{P}\left(\frac{|\hat{\Theta}_n - \theta|}{\sigma/\sqrt{n}} \leq 1.96\right) \approx 0.95 \quad \text{(CLT)}$$

$$\mathbf{P}\left(\hat{\Theta}_n - \frac{1.96\,\sigma}{\sqrt{n}} \leq \theta \leq \hat{\Theta}_n + \frac{1.96\,\sigma}{\sqrt{n}}\right) \approx 0.95$$

More generally: let z be s.t. $\Phi(z) = 1 - \alpha/2$

$$\mathbf{P}\left(\hat{\Theta}_n - \frac{z\sigma}{\sqrt{n}} \leq \theta \leq \hat{\Theta}_n + \frac{z\sigma}{\sqrt{n}}\right) \approx 1 - \alpha$$

Regression



y
Residual
(x_i, y_i)
$y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i$
$y = \hat{\theta}_0 + \hat{\theta}_1 x$
0
x

Regression Cont'

- Data: $(x_1, y_1),\ (x_2, y_2), \ldots, (x_n, y_n)$
- Model: $y \approx \theta_0 + \theta_1 x$

$$\min_{\theta_0, \theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \qquad (*)$$

- One interpretation:
  $Y_i = \theta_0 + \theta_1 x_i + W_i, \quad W_i \sim N(0, \sigma^2)$, i.i.d.
  - Likelihood function $f_{X,Y|\theta}(x, y; \theta)$ is:

$$c \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \right\}$$

  - Take logs, same as (*)
  - Least sq. $\leftrightarrow$ pretend W_i i.i.d. normal

Linear Regression

- **Model** $y \approx \theta_0 + \theta_1 x$

$$\min_{\theta_0, \theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$$

- **Solution** (set derivatives to zero):

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n}, \quad \overline{y} = \frac{y_1 + \cdots + y_n}{n}$$

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\hat{\theta}_0 = \overline{y} - \hat{\theta}_1 \overline{x}$$

Variations on Regression

- **Multiple linear regression:**
  - **data:** $(x_i, x_i', x_i'', y_i),\ i = 1, \ldots, n$
  - **model:** $y \approx \theta_0 + \theta x + \theta' x' + \theta'' x''$
  - **formulation:**

$$\min_{\theta,\theta',\theta''} \sum_{i=1}^{n} (y_i - \theta_0 - \theta x_i - \theta' x_i' - \theta'' x_i'')^2$$

Variations on Regression Cont’

- **Choosing the right variables**
  - model $y \approx \theta_0 + \theta_1 h(x)$
    e.g., $y \approx \theta_0 + \theta_1 x^2$
  - work with data points $(y_i, h(x))$
  - formulation:

$$\min_{\theta} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 h_1(x_i))^2$$

Binary Hypothesis Testing

- Binary θ; new terminology:
  - **null hypothesis** H_0:
    $$X \sim p_X(x; H_0) \qquad \text{[or } f_X(x; H_0)\text{]}$$
  - **alternative hypothesis** H_1:
    $$X \sim p_X(x; H_1) \qquad \text{[or } f_X(x; H_1)\text{]}$$

- Partition the space of possible data vectors
  Rejection region R:
  reject H_0 iff data $\in R$

Type I and Type II Errors

- Types of errors:
  - **Type I** (**false rejection**, false alarm):
    H_0 true, but rejected

$$\alpha(R) = \mathbf{P}(X \in R ; H_0)$$

  - **Type II** (**false acceptance**,
    missed detection):
    H_0 false, but accepted

$$\beta(R) = \mathbf{P}(X \notin R ; H_1)$$

Likelihood Ratio Test

- How to identify the rejection region?
- Accept H_1 if

$$\frac{f_X(x;H_1)}{f_X(x;H_0)} > \xi$$

- Threshold ξ rades off Type I and Type II errors

Example (Testing Mean)

- n data points, i.i.d.

  H_0: $X_i \sim N(0,1)$

  H_1: $X_i \sim N(1,1)$

- Likelihood ratio test; rejection region:

$$\frac{(1/\sqrt{2\pi})^n \exp\{-\sum_i (X_i - 1)^2/2\}}{(1/\sqrt{2\pi})^n \exp\{-\sum_i X_i^2/2\}} > \xi$$

$$\sum_i X_i > \xi'$$

How to find Threshold

- Find ξ' such that

$$\mathbf{P}\Big(\sum_{i=1}^{n} X_i > \xi'; H_0\Big) = \alpha$$

  - use normal tables

- Pick threshold to get a tolerable false alarm probability

Example (Testing Variance)

- n data points, i.i.d.
  H_0: $X_i \sim N(0,1)$
  H_1: $X_i \sim N(0,4)$

- Likelihood ratio test; rejection region:

$$\frac{(1/2\sqrt{2\pi})^n \exp\{-\sum_i X_i^2/(2\cdot 4)\}}{(1/\sqrt{2\pi})^n \exp\{-\sum_i X_i^2/2\}} > \xi$$

$$\sum_i X_i^2 > \xi'$$

How to Find Threshold

- Find ξ' such that

$$\mathbf{P}\Big(\sum_{i=1}^{n} X_i^2 > \xi'; H_0\Big) = \alpha$$

  - the distribution of $\sum_i X_i^2$ is known (derived distribution problem)
  - "chi-square" distribution; tables are available

Random Processes

- First view:
  sequence of random variables $X_1, X_2, \ldots$
- $\mathbf{E}[X_t] =$
- $\mathsf{Var}(X_t) =$
- Second view:
  what is the right sample space?

The Bernoulli Process

- A sequence of independent Bernoulli trials

- At each trial, i:
  - $\mathbf{P}(\text{success}) = \mathbf{P}(X_i = 1) = p$
  - $\mathbf{P}(\text{failure}) = \mathbf{P}(X_i = 0) = 1 - p$

- Examples:
  - Sequence of lottery wins/losses
  - Sequence of ups and downs of the Dow Jones
  - Arrivals (each second) to a bank
  - Arrivals (at each time slot) to server

Why is it Called Bernoulli Process?

- S is the number of successes in n time slots and is Bernoulli

- $\mathbf{P}(S = k) =$

- $\mathbf{E}[S] =$

- $\mathsf{Var}(S) =$

Interarrival Times

- T_1: number of trials until first success
  - $\mathbf{P}(T_1 = t) =$
  - Memoryless property
  - $\mathbf{E}[T_1] =$
  - $\text{Var}(T_1) =$

- If you buy a lottery ticket every day, what is the distribution of the length of the first string of losing days?

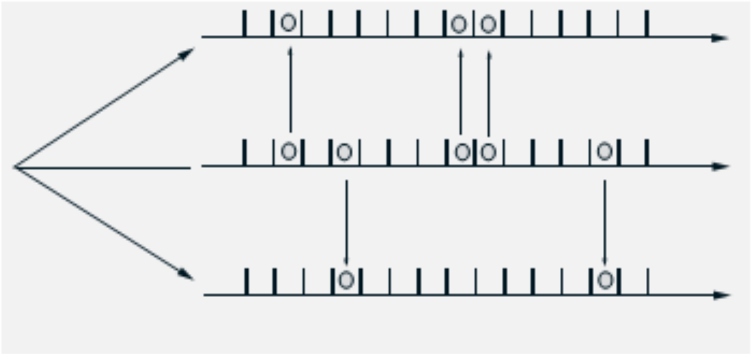
Time of the *k*th arrival

- Given that first arrival was at time t
  i.e., $T_1 = t$:
  additional time, T_2, until next arrival
  - has the same (geometric) distribution
  - independent of T_1

- Y_k: number of trials to kth success
  - $\mathbf{E}[Y_k] =$
  - $\mathrm{Var}(Y_k) =$
  - $\mathbf{P}(Y_k = t) =$

Splitting of a Bernoulli Process

- Using independent coin flips



- Still Bernoulli with a new success probability

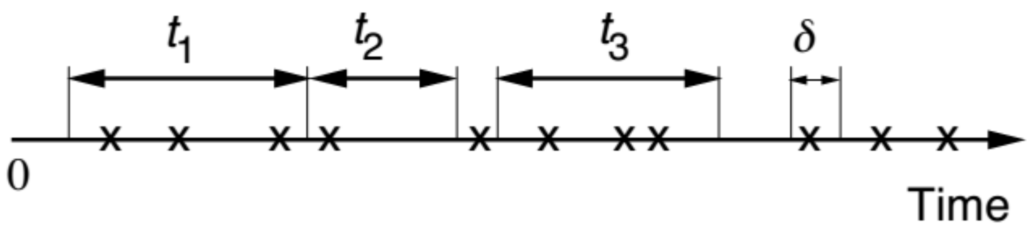
Merging of Independent Bernoulli Processes

- Collisions are counted as a single arrival

Bernoulli Process Review

- Discrete time; success probability p
- Number of arrivals in n time slots: binomial pmf
- Interarrival times: geometric pmf
- Time to k arrivals: Pascal pmf
- Memorylessness

Poisson Process



- **Time homogeneity:**
  $P(k,\tau)$ = Prob. of k arrivals in interval of duration τ

- Numbers of arrivals in disjoint time intervals are **independent**

- **Small interval probabilities:**
  For VERY small δ:

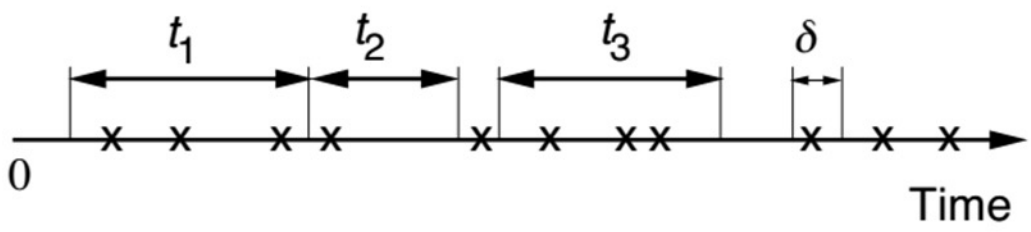
$$P(k,\delta) \approx \begin{cases} 1-\lambda\delta, & \text{if } k=0; \\ \lambda\delta, & \text{if } k=1; \\ 0, & \text{if } k>1. \end{cases}$$

  – λ: "arrival rate"

Binomial Poisson Relation

- Binomial *(n,p)* PMF converges to Poisson PMF
- Large *n*, small *p* ($n \to \infty$, $p \to 0$, $np \to \alpha$)
- AKA "Poisson Limit Theorem"
- Can be generalized to Binomial *process* and Poisson *process*

PMF of Number of Arrivals



- Finely discretize $[0, t]$: approximately Bernoulli

- N_t (of discrete approximation): binomial

- Taking $\delta \to 0$ (or $n \to \infty$) gives:

$$P(k, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, \qquad k = 0, 1, \ldots$$

- $\mathbf{E}[N_t] \;=\; \lambda t, \qquad\qquad \text{var}(N_t) \;=\; \lambda t$

Example

- You get email according to a Poisson process at a rate of $\lambda = 5$ messages per hour. You check your email every thirty minutes.

- Prob(no new messages) =

- Prob(one new message) =

Interarrival Times

- Y_k time of kth arrival
- **Erlang** distribution:

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}, \qquad y \geq 0$$

- Time of first arrival ($k = 1$):
  exponential: $f_{Y_1}(y) = \lambda e^{-\lambda y}, \quad y \geq 0$
  - Memoryless property: The time to the next arrival is independent of the past

Erlang PDF

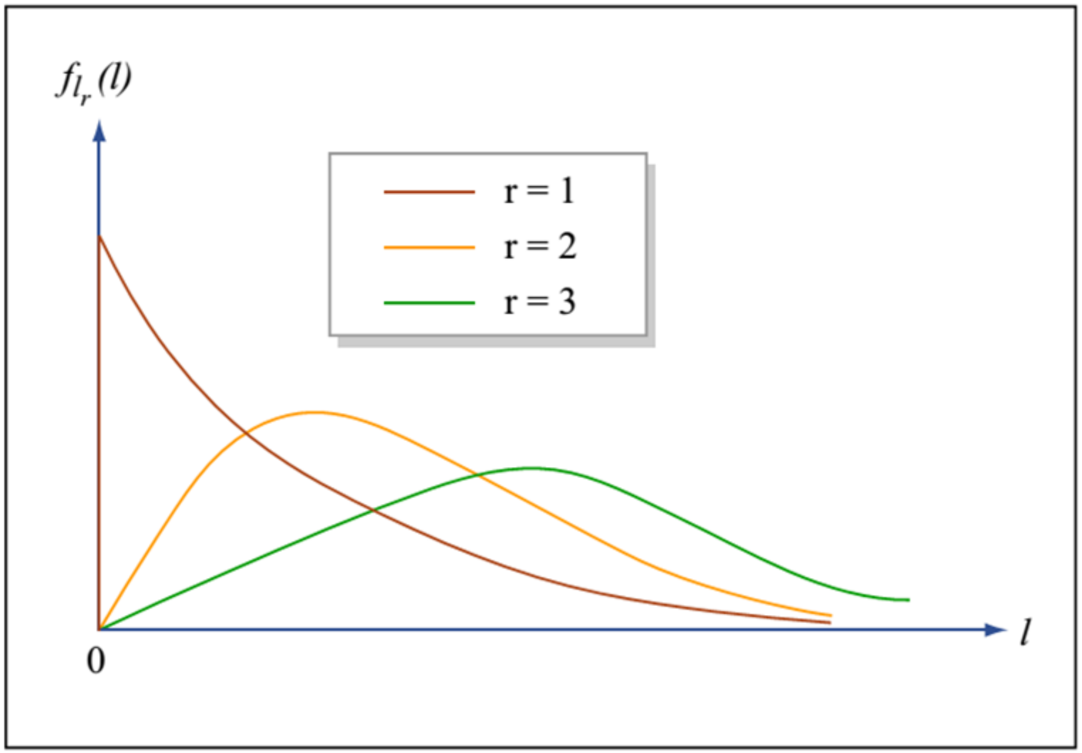
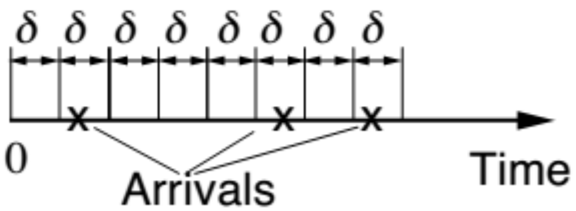


Image by MIT OpenCourseWare.

Bernoulli Poisson Relation



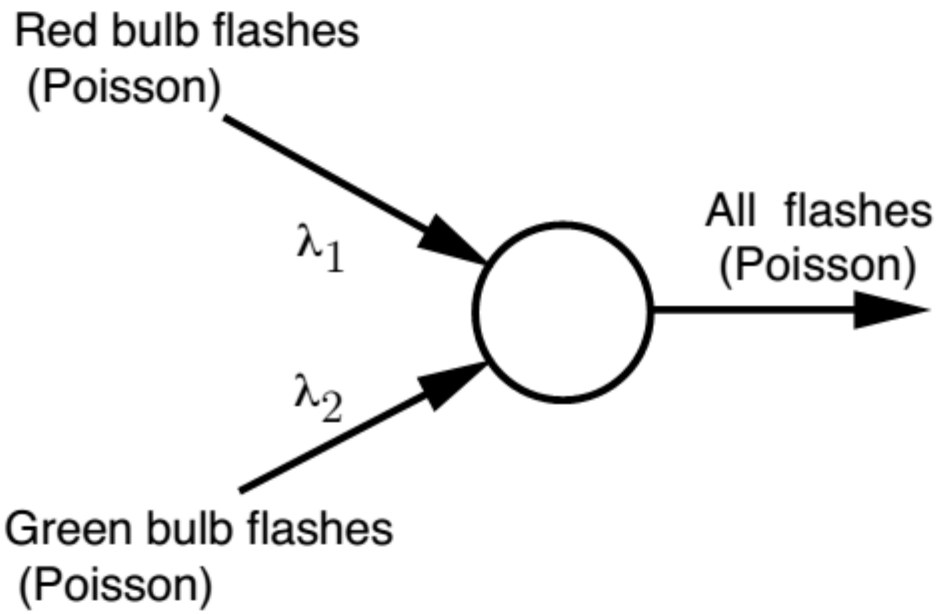
$n = t/\delta$

$p = \lambda\delta$

$np = \lambda t$

	POISSON	BERNOULLI
Times of Arrival	Continuous	Discrete
Arrival Rate	λ/unit time	p/per trial
PMF of # of Arrivals	Poisson	Binomial
Interarrival Time Distr.	Exponential	Geometric
Time to k-th arrival	Erlang	Pascal

Merging of Poisson Processes



Merging of Poisson Cont'

- Sum of independent Poisson **random variables** is Poisson

- Merging of independent Poisson **processes** is Poisson

What is the probability that the next arrival comes from the first process?

Review

- Defining characteristics
  - **Time homogeneity**: $P(k,\tau)$
  - **Independence**
  - **Small interval probabilities** (small δ):

$$P(k,\delta) \approx \begin{cases} 1-\lambda\delta, & \text{if } k=0, \\ \lambda\delta, & \text{if } k=1, \\ 0, & \text{if } k>1. \end{cases}$$

- N_τ is a Poisson r.v., with parameter $\lambda\tau$:

$$P(k,\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, \qquad k = 0, 1, \ldots$$

Review Cont'

$$\mathbf{E}[N_\tau] = \text{var}(N_\tau) = \lambda\tau$$

- Interarrival times ($k = 1$): exponential:

$$f_{T_1}(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \qquad \mathbf{E}[T_1] = 1/\lambda$$

- Time Y_k to kth arrival: Erlang(k):

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}, \qquad y \geq 0$$

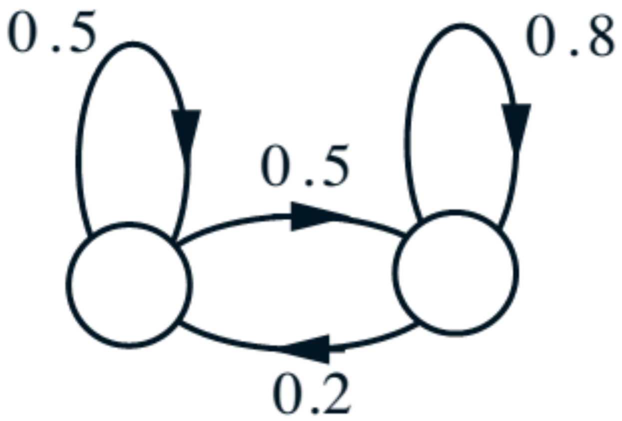
Markov Chains

- X_n: state after n transitions
  - belongs to a finite set, e.g., $\{1, \ldots, m\}$
  - X_0 is either given or random
- **Markov property/assumption:**
  (given current state, the past does not matter)

$$
\begin{aligned}
p_{ij} &= \mathbf{P}(X_{n+1} = j \mid X_n = i) \\
&= \mathbf{P}(X_{n+1} = j \mid X_n = i,\ X_{n-1}, \ldots, X_0)
\end{aligned}
$$

- Model specification:
  - identify the possible states
  - identify the possible transitions
  - identify the transition probabilities

Two State Example

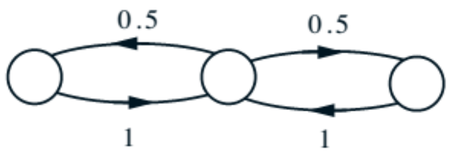


$$r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i)$$

	$n=0$	$n=1$	$n=2$	$n=100$	$n=101$
$r_{11}(n)$					
$r_{12}(n)$					
$r_{21}(n)$					
$r_{22}(n)$					

Convergence Issues

- Does $r_{ij}(n)$ converge to something?



n odd: $r_{22}(n)=$ n even: $r_{22}(n)=$

- Does the limit depend on initial state?

$r_{11}(n)=$

$r_{31}(n)=$

$r_{21}(n)=$

Steady-State Probabilities

- Do the $r_{ij}(n)$ converge to some π_j? (independent of the initial state i)

- Yes, if:
  - recurrent states are all in a single class, and
  - single recurrent class is not periodic

- Assuming "yes," start from key recursion

$$r_{ij}(n) = \sum_k r_{ik}(n-1)p_{kj}$$

  - take the limit as $n \to \infty$

$$\pi_j = \sum_k \pi_k p_{kj}, \qquad \text{for all } j$$

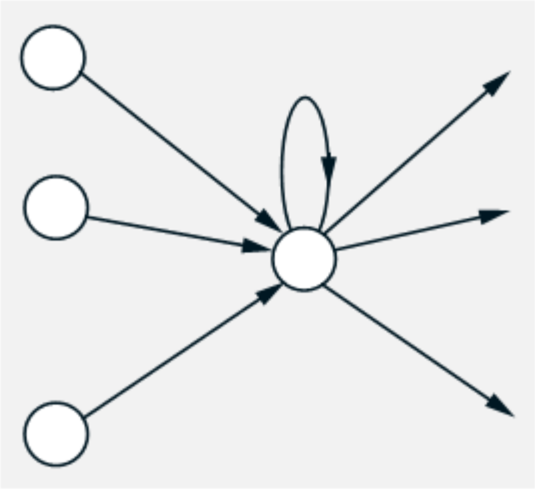
  - Additional equation:

$$\sum_j \pi_j = 1$$

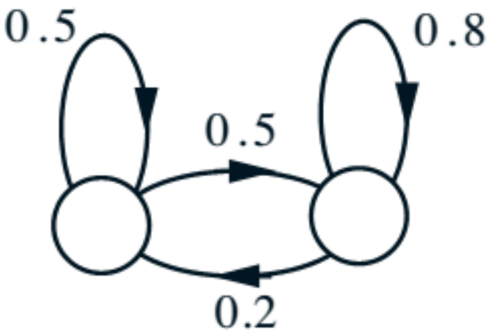
Frequency Interpretation

$$\pi_j = \sum_k \pi_k p_{kj}$$

- (Long run) frequency of being in j: π_j
- Frequency of transitions $k \to j$: $\pi_k p_{kj}$
- Frequency of transitions into j: $\sum_k \pi_k p_{kj}$



Back to Two-State Example



Review

Markov Chains

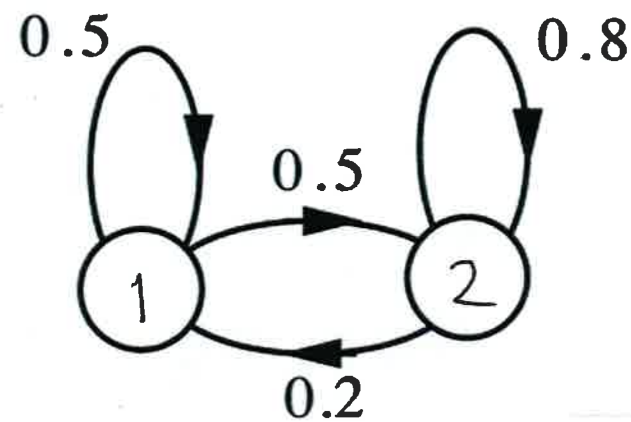
- Andrey (Andrei) Andreyevich Markov (1856-1922)
- Finite state machine randomly evolving in time
- Probability of next state only depends on current and not past
- Outcome is a sequence of states
- Simplest example of a random process that is not independent

- Examples
  - Board games played with dice
  - Random walk
  - Your phone guessing your next word in a text uses Markov models

Markov Chains

- X_n: state after n transitions
  - belongs to a finite set, e.g., $\{1, \ldots, m\}$
  - X_0 is either given or random
- **Markov property/assumption:**
  (given current state, the past does not matter)

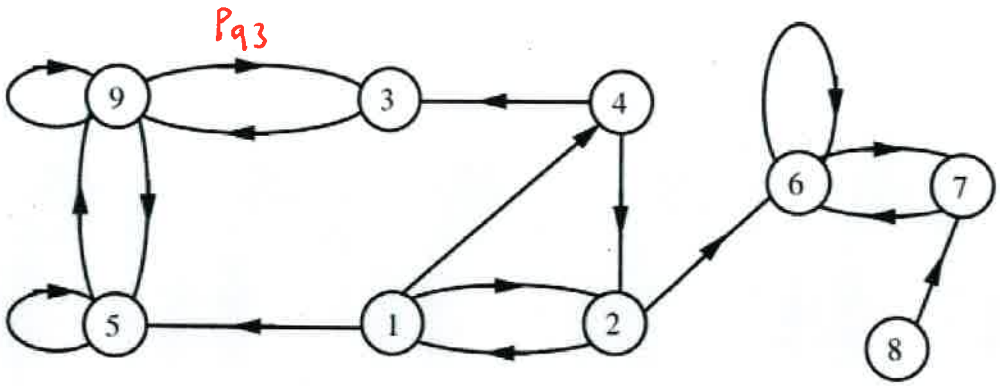
$$p_{ij} = \mathrm{P}(X_{n+1} = j \mid X_n = i) = \mathrm{P}(X_{n+1} = j \mid X_n = i,\ X_{n-1}, \ldots, X_0)$$



$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix}$$

- Model specification:
  - identify the possible states
  - identify the possible transitions
  - identify the transition probabilities
- Initial state pmf needs to be specified

Example



$\mathbf{P}(X_1 = 2, X_2 = 6, X_3 = 7 \mid X_0 = 1) = P(1 \to 2 \to 6 \to 7) = P_{12}\, P_{26}\, P_{67} > 0.$

$\Sigma P(2 \to ? \to ? \to ? \to 7)$

$\mathbf{P}(X_4 = 7 \mid X_0 = 2) = P_{21}\, P_{12}\, P_{26}\, P_{67}$

$+ \; P_{26}\, P_{67}\, P_{76}\, P_{67}$

$+ \; P_{26}\, P_{66}\, P_{66}\, P_{67}$

$+$ others)

Recurrent and Transient States

that a state j is **accessible** from a state i if for some n, the n-step transition probability $r_{ij}(n)$ is positive, i.e., if there is positive probability of reaching j, starting from i, after some number of time periods. An equivalent definition is that there is a possible state sequence $i, i_1, \ldots, i_{n-1}, j$, that starts at i and ends at j, in which the transitions $(i, i_1), (i_1, i_2), \ldots, (i_{n-2}, i_{n-1}), (i_{n-1}, j)$ all have positive probability. Let $A(i)$ be the set of states that are accessible from i. We say that i is **recurrent** if for every j that is accessible from i, i is also accessible from j; that is, for all j that belong to $A(i)$ we have that i belongs to $A(j)$.

- 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**

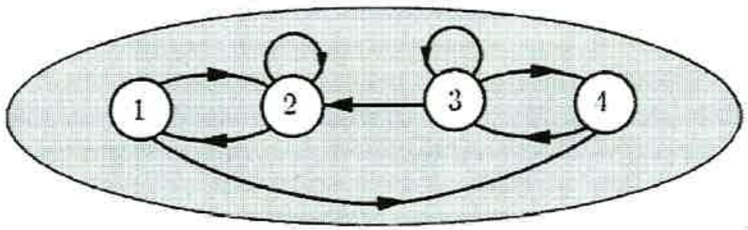
- **Recurrent class**:
  collection of recurrent states that
  "communicate" to each other
  and to no other state

Decomposition

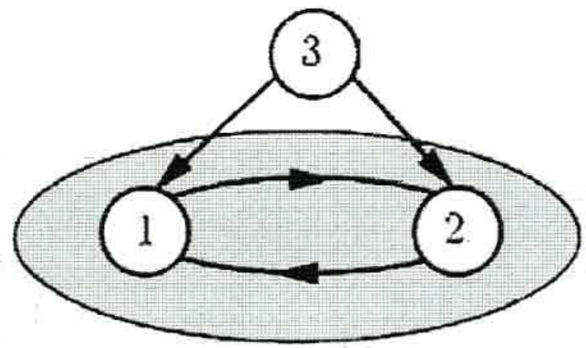
Markov Chain Decomposition

- A Markov chain can be decomposed into one or more recurrent classes, plus possibly some transient states.
- A recurrent state is accessible from all states in its class, but is not accessible from recurrent states in other classes.
- A transient state is not accessible from any recurrent state.
- At least one, possibly more, recurrent states are accessible from a given transient state.

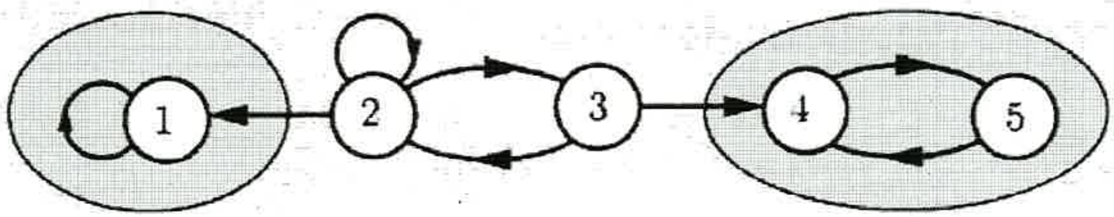
Decomposition Examples



Single class of recurrent states

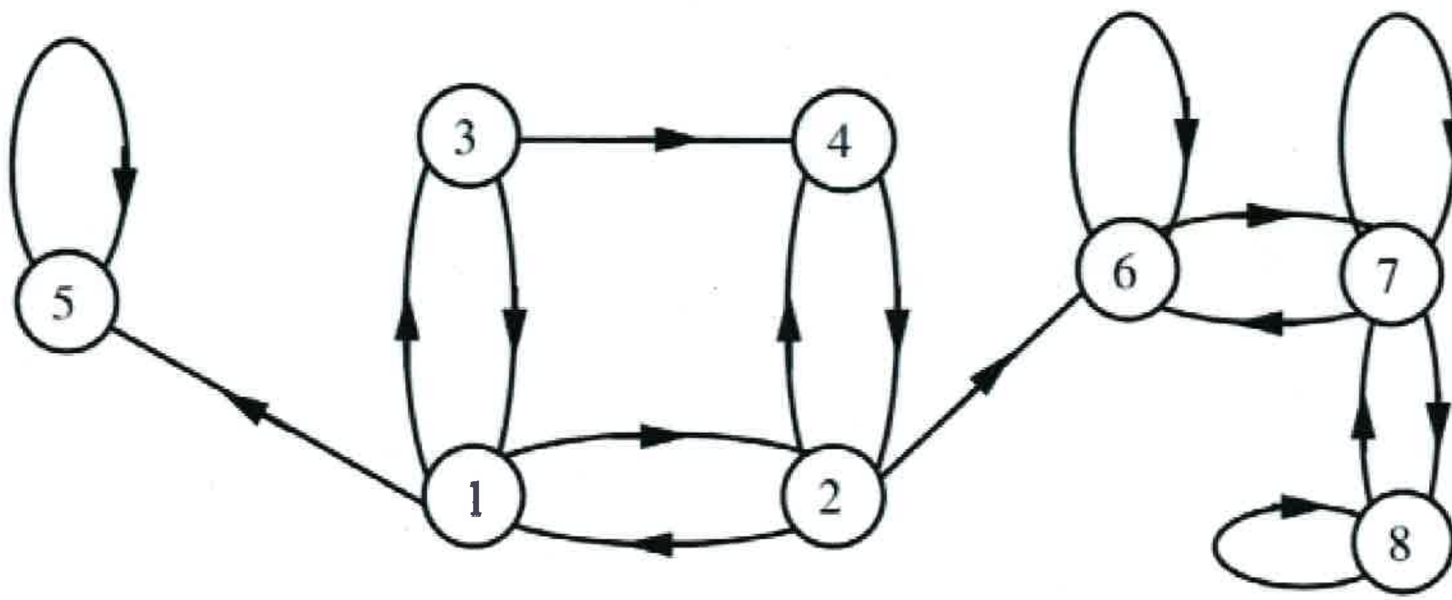


Single class of recurrent states (1 and 2)
and one transient state (3)



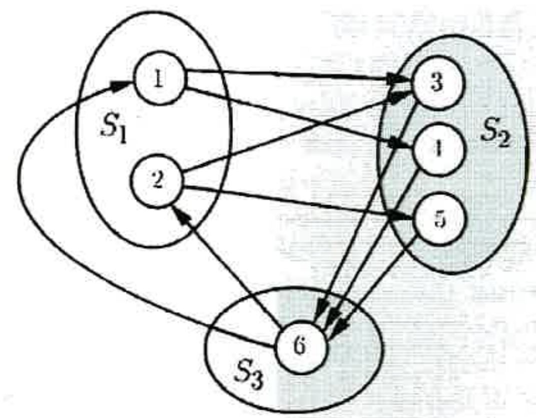
Two classes of recurrent states
(class of state 1 and class of states 4 and 5)
and two transient states (2 and 3)

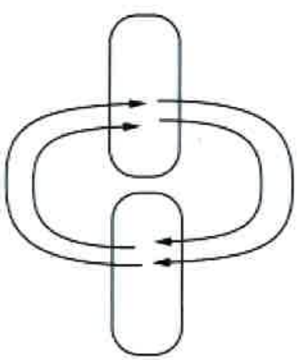
Recurrent and Transient Cont'

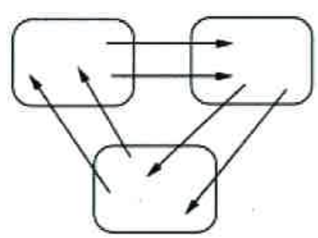


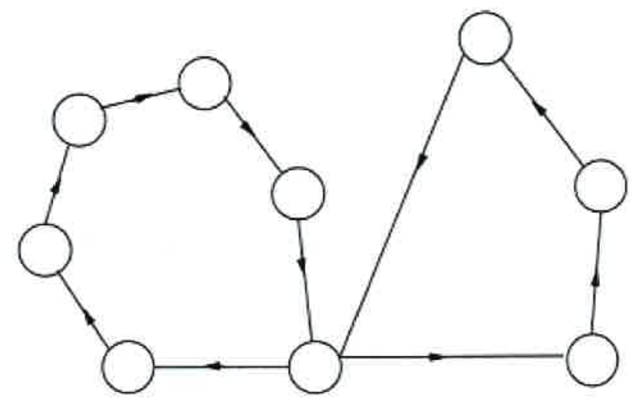
Periodic States

- 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

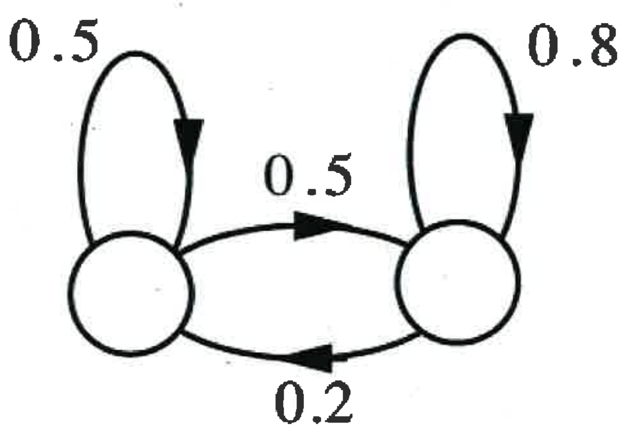








Two State Example



$$r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i)$$

Key recursion:

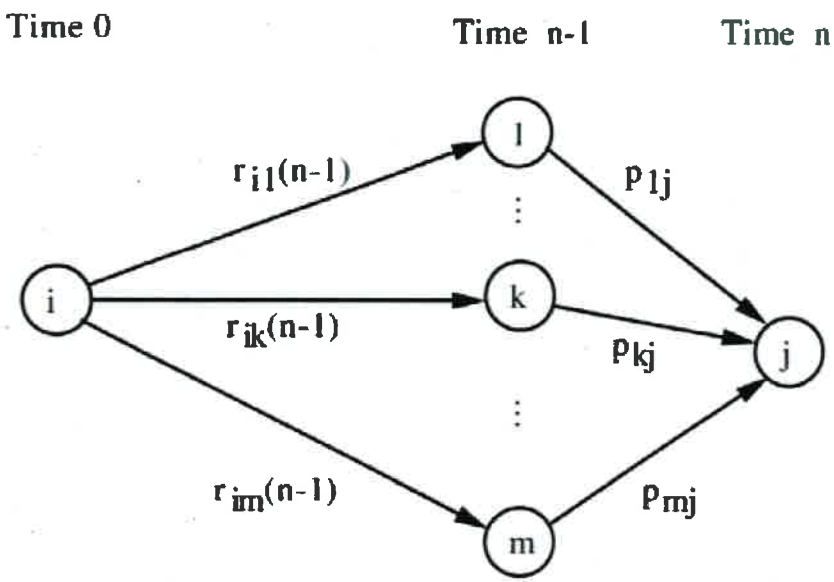
$$r_{ij}(n) = \sum_k r_{ik}(n-1)p_{kj}$$

$R(n) = R(n-1)\,P$ and $R(0) = I$

$\Rightarrow R(n) = P^n = \underbrace{P \cdot P \cdots P}_{n}$

	$n=0$	$n=1$	$n=2$	$n=100$	$n=101$
$r_{11}(n)$	1				
$r_{12}(n)$	0				
$r_{21}(n)$	0				
$r_{22}(n)$	1				

Key Recursion

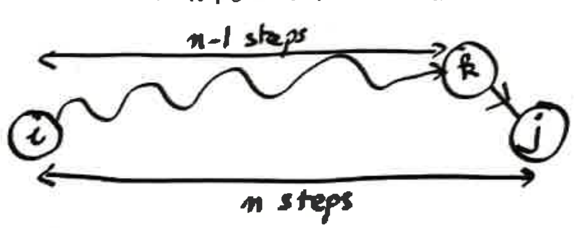


Key recursion:

$$r_{ij}(n) = \sum_k r_{ik}(n-1)p_{kj}$$

- state probabilities, given initial state i:

$$\begin{aligned} r_{ij}(n) &= \mathrm{P}(X_n = j \mid X_0 = i) \\ &= \mathrm{P}(X_{n+s} = j \mid X_s = i) \end{aligned}$$



Steady-State Probabilities

- Recall $r_{ij}(n)$ be the prob of going from state *i* to *j* in *n* steps

- If recurrent states are all in a single class AND the single class is not periodic $\lim_{n \to \infty} r_{ij}(n) = \pi_j$

- Taking the limit of $\boldsymbol{r_{ij}(n) = \sum_k r_{ik}(n-1) p_{kj}}$

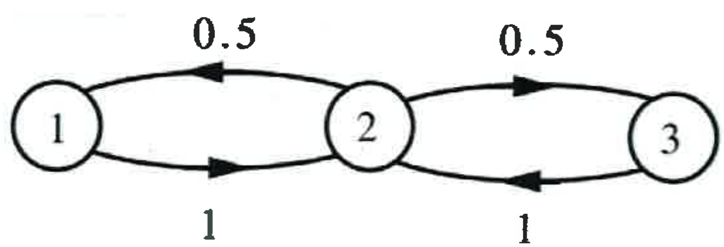
$$\pi_j = \sum_k \pi_k p_{kj} \quad \sum_j \pi_j = 1$$

- Independent of the initial state *i*

- Find the steady state probabilities by solving

Convergence Questions

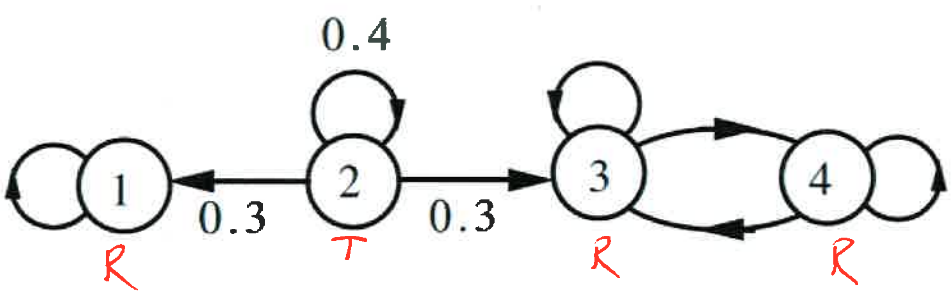
- does $r_{ij}(n)$ converge to something?



n odd: $r_{22}(n) =$ 0

n even: $r_{22}(n) =$ 1

- does the limit depend on initial state?



$r_{11}(n) =$ 1

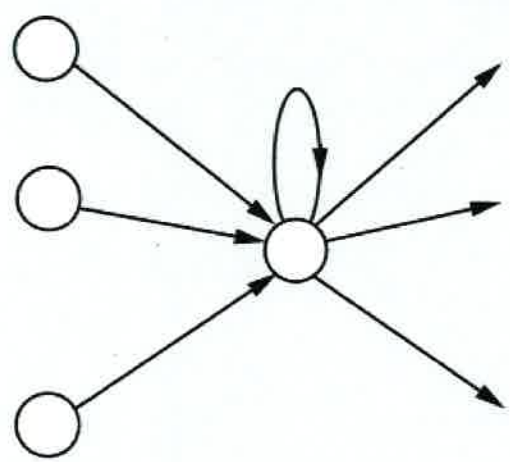
$r_{31}(n) =$ 0

$r_{21}(n) =$ $(0.4)^{n-1}\, 0.3$

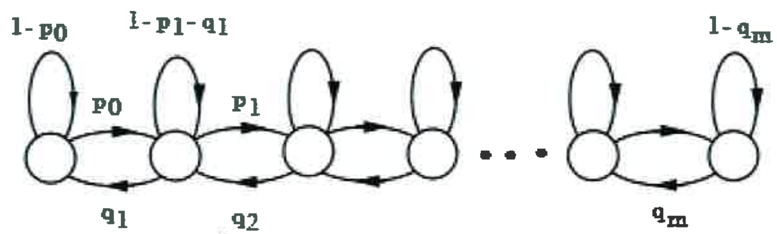
Frequency Interpretation

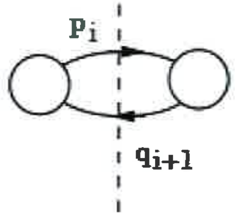
$$\pi_j = \sum_k \pi_k p_{kj}$$

- (Long run) frequency of being in j: π_j
- Frequency of transitions $k \to j$: $\pi_k p_{kj}$
- Frequency of transitions into j: $\sum_k \pi_k p_{kj}$



Birth-Death Processes





$$\pi_i p_i = \pi_{i+1} q_{i+1}$$

- Assume $p < q$ and $m \approx \infty$

$$\pi_0 = 1 - \rho$$

$$\mathbf{E}[X_n] = \frac{\rho}{1-\rho} \qquad \text{(in steady-state)}$$

- Special case: $p_i = p$ and $q_i = q$ for all i
  $\rho = p/q$ =load factor

$$\pi_{i+1} = \pi_i \frac{p}{q} = \pi_i \rho$$

$$\pi_i = \pi_0 \rho^i, \qquad i = 0, 1, \ldots, m$$

Acknowledgements and Citations

- Parts of these slides are adapted from the following resources:

- Power Point Lecture Slides for the Textbook http://bcs.wiley.com/hebcs/Books?action=index&bcsId=8676&itemId=1118324560
- John Tsitsiklis. *6.041 Probabilistic Systems Analysis and Applied Probability, Fall 2010*. (Massachusetts Institute of Technology: MIT OpenCourseWare),http://ocw.mit.edu (Accessed 13 Apr, 2015). License: Creative Commons BY-NC-SA

1.

You roll two fair six-sided dice; one die is red, the other is white. Let R_i be the event that the red die rolls i. Let W_j be the event that the white die rolls j.

(a) What is $\text{P}[R_3 W_2]$?

(b) What is the $\text{P}[S_5]$ that the sum of the two rolls is 5?

Solution:

The roll of the red and white dice can be assumed to be independent. For each die, all rolls in $\{1, 2, \ldots, 6\}$ have probability $1/6$.

(a) Thus

$$\text{P}[R_3 W_2] = \text{P}[R_3]\,\text{P}[W_2] = \frac{1}{36}.$$

(b) In fact, each pair of possible rolls $R_i W_j$ has probability $1/36$. To find $\text{P}[S_5]$, we add up the probabilities of all pairs that sum to 5:

$$\text{P}[S_5] = \text{P}[R_1 W_4] + \text{P}[R_2 W_3] + \text{P}[R_3 W_2] + \text{P}[R_4 W_1] = 4/36 = 1/9.$$

2.

You have a shuffled deck of three cards: 2, 3, and 4, and you deal out the three cards. Let E_i denote the event that ith card dealt is even numbered.

(a) What is $\text{P}[E_2|E_1]$, the probability the second card is even given that the first card is even?

(b) What is the conditional probability that the first two cards are even given that the third card is even?

(c) Let O_i represent the event that the ith card dealt is odd numbered. What is $\text{P}[E_2|O_1]$, the conditional probability that the second card is even given that the first card is odd?

(d) What is the conditional probability that the second card is odd given that the first card is odd?

Solution:

The sample outcomes can be written ijk where the first card drawn is i, the second is j and the third is k. The sample space is

$$S = \{234, 243, 324, 342, 423, 432\}. \tag{1}$$

and each of the six outcomes has probability 1/6. The events E_1, E_2, E_3, O_1, O_2, O_3 are

$$E_1 = \{234, 243, 423, 432\}, \qquad O_1 = \{324, 342\}, \tag{2}$$
$$E_2 = \{243, 324, 342, 423\}, \qquad O_2 = \{234, 432\}, \tag{3}$$
$$E_3 = \{234, 324, 342, 432\}, \qquad O_3 = \{243, 423\}. \tag{4}$$

(a) The conditional probability the second card is even given that the first card is even is

$$\mathrm{P}[E_2|E_1] = \frac{\mathrm{P}[E_2E_1]}{\mathrm{P}[E_1]} = \frac{\mathrm{P}[243, 423]}{\mathrm{P}[234, 243, 423, 432]} = \frac{2/6}{4/6} = 1/2. \tag{5}$$

(b) The conditional probability the first card is even given that the second card is even is

$$\mathrm{P}[E_1|E_2] = \frac{\mathrm{P}[E_1E_2]}{\mathrm{P}[E_2]} = \frac{\mathrm{P}[243, 423]}{\mathrm{P}[243, 324, 342, 423]} = \frac{2/6}{4/6} = 1/2. \tag{6}$$

(c) The probability the first two cards are even given the third card is even is

$$\mathrm{P}[E_1E_2|E_3] = \frac{\mathrm{P}[E_1E_2E_3]}{\mathrm{P}[E_3]} = 0. \tag{7}$$

(d) The conditional probabilities the second card is even given that the first card is odd is

$$\mathrm{P}[E_2|O_1] = \frac{\mathrm{P}[O_1E_2]}{\mathrm{P}[O_1]} = \frac{\mathrm{P}[O_1]}{\mathrm{P}[O_1]} = 1. \tag{8}$$

3.

In an experiment with equiprobable outcomes, the sample space is $S = \{1, 2, 3, 4\}$ and $\mathrm{P}[s] = 1/4$ for all $s \in S$. Find three events in S that are pairwise independent but are not independent.

Solution:

For a sample space $S = \{1, 2, 3, 4\}$ with equiprobable outcomes, consider the events

$$A_1 = \{1, 2\} \qquad A_2 = \{2, 3\} \qquad A_3 = \{3, 1\}. \tag{1}$$

Each event A_i has probability $1/2$. Moreover, each pair of events is independent since

$$\mathrm{P}[A_1 A_2] = \mathrm{P}[A_2 A_3] = \mathrm{P}[A_3 A_1] = 1/4. \tag{2}$$

However, the three events A_1, A_2, A_3 are not independent since

$$\mathrm{P}[A_1 A_2 A_3] = 0 \neq \mathrm{P}[A_1]\,\mathrm{P}[A_2]\,\mathrm{P}[A_3]. \tag{3}$$

4.

For independent events A and B, prove that

(a) A and B^c are independent.

(b) A^c and B are independent.

(c) A^c and B^c are independent.

Solution:

(a) For any events A and B, we can write the law of total probability in the form of

$$\mathrm{P}[A] = \mathrm{P}[AB] + \mathrm{P}[AB^c]. \tag{1}$$

Since A and B are independent, $\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B]$. This implies

$$\mathrm{P}[AB^c] = \mathrm{P}[A] - \mathrm{P}[A]\,\mathrm{P}[B] = \mathrm{P}[A](1 - \mathrm{P}[B]) = \mathrm{P}[A]\,\mathrm{P}[B^c]. \tag{2}$$

Thus A and B^c are independent.

(b) Proving that A^c and B are independent is not really necessary. Since A and B are arbitrary labels, it is really the same claim as in part (a). That is, simply reversing the labels of A and B proves the claim. Alternatively, one can construct exactly the same proof as in part (a) with the labels A and B reversed.

(c) To prove that A^c and B^c are independent, we apply the result of part (a) to the sets A and B^c. Since we know from part (a) that A and B^c are independent, part (b) says that A^c and B^c are independent.

5.

Prove that if X is a nonnegative integer-valued random variable, then

$$\mathrm{E}[X] = \sum_{k=0}^{\infty} \mathrm{P}[X > k].$$

Solution:

We write the sum as a double sum in the following way:

$$\sum_{i=0}^{\infty} \mathrm{P}[X > i] = \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} P_X(j). \tag{1}$$

At this point, the key step is to reverse the order of summation. You may need to make a sketch of the feasible values for i and j to see how this reversal occurs. In this case,

$$\sum_{i=0}^{\infty} \mathrm{P}[X > i] = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} P_X(j) = \sum_{j=1}^{\infty} j P_X(j) = \mathrm{E}[X]. \tag{2}$$

6.

X is the binomial $(5, 0.5)$ random variable.

(a) Find the standard deviation of X.

(b) Find $\mathrm{P}[\mu_X - \sigma_X \le X \le \mu_X + \sigma_X]$, the probability that X is within one standard deviation of the expected value.

Solution:

(a) The expected value of X is

$$\begin{aligned} \mathrm{E}[X] &= \sum_{x=0}^{5} x P_X(x) \\ &= 0\binom{5}{0}\frac{1}{2^5} + 1\binom{5}{1}\frac{1}{2^5} + 2\binom{5}{2}\frac{1}{2^5} \\ &\qquad + 3\binom{5}{3}\frac{1}{2^5} + 4\binom{5}{4}\frac{1}{2^5} + 5\binom{5}{5}\frac{1}{2^5} \\ &= [5 + 20 + 30 + 20 + 5]/2^5 = 5/2. \end{aligned} \tag{1}$$

The expected value of X^2 is

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= \sum_{x=0}^{5} x^2 P_X(x) \\
&= 0^2\binom{5}{0}\frac{1}{2^5} + 1^2\binom{5}{1}\frac{1}{2^5} + 2^2\binom{5}{2}\frac{1}{2^5} \\
&\qquad + 3^2\binom{5}{3}\frac{1}{2^5} + 4^2\binom{5}{4}\frac{1}{2^5} + 5^2\binom{5}{5}\frac{1}{2^5} \\
&= [5+40+90+80+25]/2^5 = 240/32 = 15/2. \qquad (2)
\end{aligned}$$

The variance of X is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}\left[X\right])^2 = 15/2 - 25/4 = 5/4. \qquad (3)$$

By taking the square root of the variance, the standard deviation of X is $\sigma_X = \sqrt{5/4} \approx 1.12$.

(b) The probability that X is within one standard deviation of its mean is

$$\begin{aligned}
\mathrm{P}\left[\mu_X - \sigma_X \le X \le \mu_X + \sigma_X\right] &= \mathrm{P}\left[2.5 - 1.12 \le X \le 2.5 + 1.12\right] \\
&= \mathrm{P}\left[1.38 \le X \le 3.62\right] \\
&= \mathrm{P}\left[2 \le X \le 3\right]. \qquad (4)
\end{aligned}$$

By summing the PMF over the desired range, we obtain

$$\begin{aligned}
\mathrm{P}\left[2 \le X \le 3\right] &= P_X(2) + P_X(3) \\
&= 10/32 + 10/32 = 5/8.
\end{aligned}$$

7.

Show that the variance of $Y = aX + b$ is $\mathrm{Var}[Y] = a^2 \mathrm{Var}[X]$.

Solution:

For $Y = aX + b$, we wish to show that $\mathrm{Var}[Y] = a^2 \mathrm{Var}[X]$. We begin by noting that Theorem 3.12 says that $\mathrm{E}[aX + b] = a\,\mathrm{E}[X] + b$. Hence, by the definition of variance.

$$\begin{aligned}
\mathrm{Var}[Y] &= \mathrm{E}\left[(aX + b - (a\,\mathrm{E}[X] + b))^2\right] \\
&= \mathrm{E}\left[a^2 (X - \mathrm{E}[X])^2\right] \\
&= a^2\,\mathrm{E}\left[(X - \mathrm{E}[X])^2\right]. \qquad (1)
\end{aligned}$$

Since $\mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{Var}[X]$, the assertion is proved.

8.

Given a random variable X with expected value μ_X and variance σ_X^2, find the expected value and variance of

$$Y = \frac{X - \mu_X}{\sigma_X}.$$

Given that

$$Y = \frac{1}{\sigma_x}(X - \mu_X), \qquad (1)$$

we can use the linearity property of the expectation operator to find the mean value

$$\mathrm{E}[Y] = \frac{\mathrm{E}[X - \mu_X]}{\sigma_X} = \frac{\mathrm{E}[X] - \mathrm{E}[X]}{\sigma_X} = 0. \qquad (2)$$

Using the fact that $\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X]$, the variance of Y is found to be

$$\mathrm{Var}[Y] = \frac{1}{\sigma_X^2} \mathrm{Var}[X] = 1. \qquad (3)$$

9.

Random variable K has a Poisson (α) distribution. Derive the properties $\mathrm{E}[K] = \mathrm{Var}[K] = \alpha$. Hint: $\mathrm{E}[K^2] = \mathrm{E}[K(K-1)] + \mathrm{E}[K]$.

Solution:

The PMF of K is the Poisson PMF

$$P_K(k) = \begin{cases} \lambda^k e^{-\lambda}/k! & k = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The mean of K is

$$\mathrm{E}[K] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda. \tag{2}$$

To find $\mathrm{E}[K^2]$, we use the hint and first find

$$\mathrm{E}[K(K-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!}. \tag{3}$$

By factoring out λ^2 and substituting $j = k - 2$, we obtain

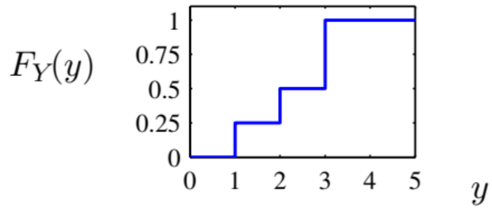
$$\mathrm{E}[K(K-1)] = \lambda^2 \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!}}_{1} = \lambda^2. \tag{4}$$

The above sum equals 1 because it is the sum of a Poisson PMF over all possible values. Since $\mathrm{E}[K] = \lambda$, the variance of K is

$$\begin{aligned} \mathrm{Var}[K] &= \mathrm{E}\left[K^2\right] - (\mathrm{E}[K])^2 \\ &= \mathrm{E}[K(K-1)] + \mathrm{E}[K] - (\mathrm{E}[K])^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned} \tag{5}$$

1.

Discrete random variable Y has the CDF $F_Y(y)$ as shown:



Use the CDF to find the following probabilities:

(a) $\mathrm{P}[Y < 1]$ and $\mathrm{P}[Y \le 1]$

(b) $\mathrm{P}[Y > 2]$ and $\mathrm{P}[Y \ge 2]$

(c) $\mathrm{P}[Y = 3]$ and $\mathrm{P}[Y > 3]$

(d) $P_Y(y)$

Solution:

Using the CDF given in the problem statement we find that

(a) $\mathrm{P}[Y < 1] = 0$ and $\mathrm{P}[Y \le 1] = 1/4$.

(b)

$$\mathrm{P}[Y > 2] = 1 - \mathrm{P}[Y \le 2] = 1 - 1/2 = 1/2. \tag{1}$$
$$\mathrm{P}[Y \ge 2] = 1 - \mathrm{P}[Y < 2] = 1 - 1/4 = 3/4. \tag{2}$$

(c)

$$P[Y = 3] = F_Y(3^+) - F_Y(3^-) = 1/2. \quad (3)$$

$$P[Y > 3] = 1 - F_Y(3) = 0. \quad (4)$$

(d) From the staircase CDF of Problem 3.4.1, we see that Y is a discrete random variable. The jumps in the CDF occur at at the values that Y can take on. The height of each jump equals the probability of that value. The PMF of Y is

$$P_Y(y) = \begin{cases} 1/4 & y = 1, \\ 1/4 & y = 2, \\ 1/2 & y = 3, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

2.

The cumulative distribution function of random variable X is

$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/2 & -1 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$

(a) What is $P[X > 1/2]$?

(b) What is $P[-1/2 < X \le 3/4]$?

(c) What is $P[|X| \le 1/2]$?

(d) What is the value of a such that $P[X \le a] = 0.8$?

Solution:

The CDF of X is

$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/2 & -1 \le x < 1, \\ 1 & x \ge 1. \end{cases} \quad (1)$$

Each question can be answered by expressing the requested probability in terms of $F_X(x)$.

(a)

$$\begin{aligned} \mathrm{P}[X > 1/2] &= 1 - \mathrm{P}[X \leq 1/2] \\ &= 1 - F_X(1/2) = 1 - 3/4 = 1/4. \end{aligned} \tag{2}$$

(b) This is a little trickier than it should be. Being careful, we can write

$$\begin{aligned} \mathrm{P}[-1/2 \leq X < 3/4] = \mathrm{P}[-1/2 < X \leq 3/4] \\ + \mathrm{P}[X = -1/2] - \mathrm{P}[X = 3/4]. \end{aligned} \tag{3}$$

Since the CDF of X is a continuous function, the probability that X takes on any specific value is zero. This implies $\mathrm{P}[X = 3/4] = 0$ and $\mathrm{P}[X = -1/2] = 0$. (If this is not clear at this point, it will become clear in Section 4.7.) Thus,

$$\begin{aligned} \mathrm{P}[-1/2 \leq X < 3/4] &= \mathrm{P}[-1/2 < X \leq 3/4] \\ &= F_X(3/4) - F_X(-1/2) = 5/8. \end{aligned} \tag{4}$$

(c)

$$\begin{aligned} \mathrm{P}[|X| \leq 1/2] &= \mathrm{P}[-1/2 \leq X \leq 1/2] \\ &= \mathrm{P}[X \leq 1/2] - \mathrm{P}[X < -1/2]. \end{aligned} \tag{5}$$

Note that $\mathrm{P}[X \leq 1/2] = F_X(1/2) = 3/4$. Since the probability that $\mathrm{P}[X = -1/2]$ 0, $\mathrm{P}[X < -1/2] = \mathrm{P}[X \leq 1/2]$. Hence $\mathrm{P}[X < -1/2] = F_X(-1/2) = 1/4$. This implies

$$\begin{aligned} \mathrm{P}[|X| \leq 1/2] &= \mathrm{P}[X \leq 1/2] - \mathrm{P}[X < -1/2] \\ &= 3/4 - 1/4 = 1/2. \end{aligned} \tag{6}$$

(d) Since $F_X(1) = 1$, we must have $a \leq 1$. For $a \leq 1$, we need to satisfy

$$\mathrm{P}[X \leq a] = F_X(a) = \frac{a+1}{2} = 0.8. \tag{7}$$

Thus $a = 0.6$.

3.

For constants a and b, random variable X has PDF

$$f_X(x) = \begin{cases} ax^2 + bx & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

What conditions on a and b are necessary and sufficient to guarantee that $f_X(x)$ is a valid PDF?

Solution:

$$f_X(x) = \begin{cases} ax^2 + bx & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

First, we note that a and b must be chosen such that the above PDF integrates to 1.

$$\int_0^1 (ax^2 + bx)\, dx = a/3 + b/2 = 1 \tag{2}$$

Hence, $b = 2 - 2a/3$ and our PDF becomes

$$f_X(x) = x(ax + 2 - 2a/3) \tag{3}$$

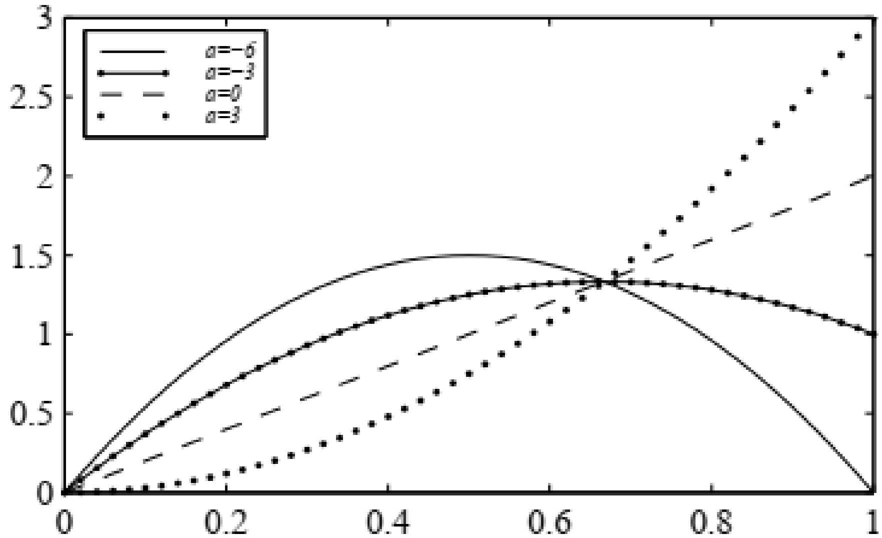
For the PDF to be non-negative for $x \in [0, 1]$, we must have $ax + 2 - 2a/3 \ge 0$ for all $x \in [0, 1]$. This requirement can be written as

$$a(2/3 - x) \le 2, \qquad 0 \le x \le 1. \tag{4}$$

For $x = 2/3$, the requirement holds for all a. However, the problem is tricky because we must consider the cases $0 \le x < 2/3$ and $2/3 < x \le 1$ separately because of the sign change of the inequality. When $0 \le x < 2/3$, we have $2/3 - x > 0$ and the requirement is most stringent at $x = 0$ where we require $2a/3 \le 2$ or $a \le 3$. When $2/3 < x \le 1$, we can write the constraint as $a(x - 2/3) \ge -2$. In this case, the constraint is most stringent at $x = 1$, where we must have $a/3 \ge -2$ or $a \ge -6$. Thus a complete expression for our requirements are

$$-6 \le a \le 3, \qquad b = 2 - 2a/3. \tag{5}$$

As we see in the following plot, the shape of the PDF $f_X(x)$ varies greatly with the value of a.



4.

Long-distance calling plan A offers flat-rate service at 10 cents per minute. Calling plan B charges 99 cents for every call under 20 minutes; for calls over 20 minutes, the charge is 99 cents for the first 20 minutes plus 10 cents for every additional minute. (Note that these plans measure your call duration exactly, without rounding to the next minute or even second.) If your long-distance calls have exponential distribution with expected value τ minutes, which plan offers a lower expected cost per call?

Solution:

Let X denote the holding time of a call. The PDF of X is

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

We will use $C_A(X)$ and $C_B(X)$ to denote the cost of a call under the two plans. From the problem statement, we note that $C_A(X) = 10X$ so that $\mathrm{E}[C_A(X)] = 10\,\mathrm{E}[X] = 10\tau$. On the other hand

$$C_B(X) = 99 + 10(X - 20)^+, \tag{2}$$

where $y^+ = y$ if $y \geq 0$; otherwise $y^+ = 0$ for $y < 0$. Thus,

$$\begin{aligned} \mathrm{E}\left[C_B(X)\right] &= \mathrm{E}\left[99 + 10(X-20)^+\right] \\ &= 99 + 10\,\mathrm{E}\left[(X-20)^+\right] \\ &= 99 + 10\,\mathrm{E}\left[(X-20)^+ | X \leq 20\right] \mathrm{P}\left[X \leq 20\right] \\ &\qquad + 10\,\mathrm{E}\left[(X-20)^+ | X > 20\right] \mathrm{P}\left[X > 20\right]. \end{aligned} \tag{3}$$

Given $X \leq 20$, $(X-20)^+ = 0$. Thus $\mathrm{E}[(X-20)^+ | X \leq 20] = 0$ and

$$\mathrm{E}\left[C_B(X)\right] = 99 + 10\,\mathrm{E}\left[(X-20) | X > 20\right] \mathrm{P}\left[X > 20\right]. \tag{4}$$

Finally, we observe that $\mathrm{P}[X > 20] = e^{-20/\tau}$ and that

$$\mathrm{E}\left[(X-20) | X > 20\right] = \tau \tag{5}$$

since given $X \geq 20$, $X - 20$ has a PDF identical to X by the memoryless property of the exponential random variable. Thus,

$$\mathrm{E}\left[C_B(X)\right] = 99 + 10\tau e^{-20/\tau} \tag{6}$$

Some numeric comparisons show that $\mathrm{E}[C_B(X)] \leq \mathrm{E}[C_A(X)]$ if $\tau > 12.34$ minutes. That is, the flat price for the first 20 minutes is a good deal only if your average phone call is sufficiently long.

5.

The voltage V at the output of a microphone is the continuous uniform $(-1, 1)$ random variable. The microphone voltage is processed by a clipping rectifier with output

$$L = \begin{cases} |V| & |V| \leq 0.5, \\ 0.5 & \text{otherwise}. \end{cases}$$

(a) What is $\mathrm{P}[L = 0.5]$?

(b) What is $F_L(l)$?

(c) What is $\mathrm{E}[L]$?

Solution:

Since the microphone voltage V is uniformly distributed between -1 and 1 volts, V has PDF and CDF

$$f_V(v) = \begin{cases} 1/2 & -1 \le v \le 1, \\ 0 & \text{otherwise}, \end{cases} \qquad F_V(v) = \begin{cases} 0 & v < -1, \\ (v+1)/2 & -1 \le v \le 1, \\ 1 & v > 1. \end{cases} \tag{1}$$

The voltage is processed by a limiter whose output magnitude is given by below

$$L = \begin{cases} |V| & |V| \le 0.5, \\ 0.5 & \text{otherwise}. \end{cases} \tag{2}$$

(a)

$$\begin{aligned} \mathrm{P}[L = 0.5] &= \mathrm{P}[|V| \ge 0.5] = \mathrm{P}[V \ge 0.5] + \mathrm{P}[V \le -0.5] \\ &= 1 - F_V(0.5) + F_V(-0.5) \\ &= 1 - 1.5/2 + 0.5/2 = 1/2. \end{aligned} \tag{3}$$

(b) For $0 \le l \le 0.5$,

$$\begin{aligned} F_L(l) = \mathrm{P}[|V| \le l] &= \mathrm{P}[-l \le v \le l] \\ &= F_V(l) - F_V(-l) \\ &= 1/2(l+1) - 1/2(-l+1) = l. \end{aligned} \tag{4}$$

So the CDF of L is

$$F_L(l) = \begin{cases} 0 & l < 0, \\ l & 0 \le l < 0.5, \\ 1 & l \ge 0.5. \end{cases} \tag{5}$$

(c) By taking the derivative of $F_L(l)$, the PDF of L is

$$f_L(l) = \begin{cases} 1 + (0.5)\delta(l - 0.5) & 0 \le l \le 0.5, \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

The expected value of L is

$$\begin{aligned}
\mathrm{E}[L] &= \int_{-\infty}^{\infty} l f_L(l)\, dl \\
&= \int_0^{0.5} l\, dl + 0.5 \int_0^{0.5} l(0.5)\delta(l - 0.5)\, dl = 0.375.
\end{aligned} \tag{7}$$

6.

X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & x \geq 0, y \geq 0, x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

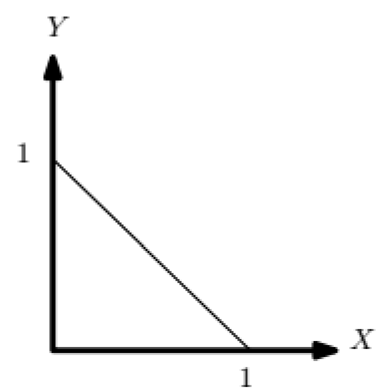
(a) Are X and Y independent?

(b) Let $U = \min(X,Y)$. Find the CDF and PDF of U.

(c) Let $V = \max(X,Y)$. Find the CDF and PDF of V.

Solution:

The key to the solution is to draw the triangular region where the PDF is nonzero:

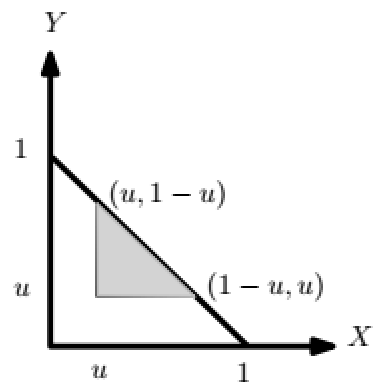


(a) X and Y are not independent. For example it is easy to see that $f_X(3/4) = f_Y(3/4) > 0$ and thus $f_X(3/4)f_Y(3/4) > 0$. However, $f_{X,Y}(3/4, 3/4) = 0$.

(b) First we find the CDF. Since $X \geq 0$ and $Y \geq 0$, we know that $F_U(u) = 0$ for $u < 0$. Next, for non-negative u, we see that

$$\begin{aligned} F_U(u) = \mathrm{P}\left[\min(X,Y) \leq u\right] &= 1 - \mathrm{P}\left[\min(X,Y) > u\right] \\ &= 1 - \mathrm{P}\left[X > u, Y > u\right]. \end{aligned} \tag{1}$$

At this point is is instructive to draw the region for small u:



We see that this area exists as long as $u \leq 1 - u$, or $u \leq 1/2$. This is because if both $X > 1/2$ and $Y > 1/2$ then $X + Y > 1$ which violates the constraint $X + Y \leq 1$. For $0 \leq u \leq 1/2$,

$$\begin{aligned} F_U(u) &= 1 - \int_u^{1-u} \int_u^{1-x} 2\, dy\, dx \\ &= 1 - 2\frac{1}{2}[(1-u) - u]^2 = 1 - [1 - 2u]^2. \end{aligned} \tag{2}$$

Note that we wrote the integral expression but we calculated the integral as c times the area of integration. Thus the CDF of U is

$$F_U(u) = \begin{cases} 0 & u < 0, \\ 1 - [1 - 2u]^2 & 0 \leq u \leq 1/2, \\ 1 & u > 1/2. \end{cases} \tag{3}$$

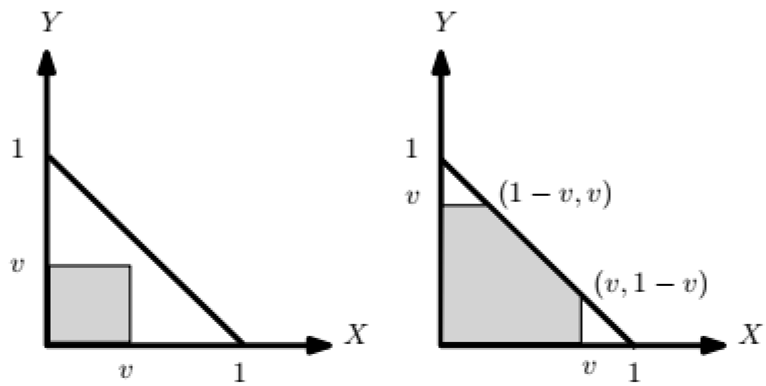
Taking the derivative, we find the PDF of U is

$$F_U(u) = \begin{cases} 4(1 - 2u) & 0 \leq u \leq 1/2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(c) For the CDF of V, we can write

$$\begin{aligned}
F_V(v) = \mathrm{P}[V \le v] &= \mathrm{P}[\max(X,Y) \le v] \\
&= \mathrm{P}[X \le v, Y \le v] \\
&= \int_0^v \int_0^v f_{X,Y}(x,y)\ dx, dy. \qquad (5)
\end{aligned}$$

This is tricky because there are two distinct cases:



$0 \le v \le 1/2$ $\qquad$ $1/2 \le v \le 1$

For $0 \le v \le 1/2$,

$$F_V(v) = \int_0^v \int_0^v 2\, dx\, dy = 2v^2. \qquad (6)$$

For $1/2 \le v \le 1$, you can write the integral as

$$\begin{aligned}
F_V(v) &= \int_0^{1-v} \int_0^v 2\, dy\, dx + \int_{1-v}^v \int_0^{1-x} 2\, dy\, dx \\
&= 2\left[v^2 - \frac{1}{2}[v-(1-v)]^2\right] \\
&= 2v^2 - (2v-1)^2 = 4v - 2v^2 - 1, \qquad (7)
\end{aligned}$$

where we skipped the steps of the integral by observing that the shaded area of integration is a square of area v^2 minus the cutoff triangle on the upper right corner. The full expression for the CDF of V is

$$F_V(v) = \begin{cases} 0 & v < 0, \\ 2v^2 & 0 \le v \le 1/2, \\ 4v - 2v^2 - 1 & 1/2 \le v \le 1, \\ 1 & v > 1. \end{cases} \tag{8}$$

Taking a derivative, the PDF of V is

$$f_V(v) = \begin{cases} 4v & 0 \le v \le 1/2, \\ 4(1-v) & 1/2 \le v \le 1. \end{cases} \tag{9}$$

7.

Random variables X and Y have the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} c & x \ge 0, y \ge 0, x + y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

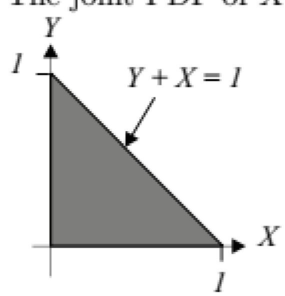
(a) What is the value of the constant c?

(b) What is $P[X \le Y]$?

(c) What is $P[X + Y \le 1/2]$?

Solution:

(a) The joint PDF of X and Y is



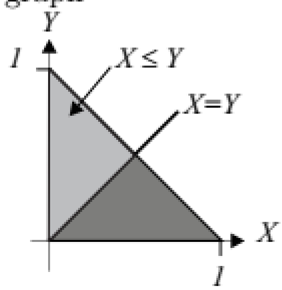
$$f_{X,Y}(x,y) = \begin{cases} c & x + y \le 1, x, y \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

To find the constant c we integrate over the region shown. This gives

$$\int_0^1 \int_0^{1-x} c\, dy\, dx = cx - \frac{cx}{2}\Big|_0^1 = \frac{c}{2} = 1. \tag{1}$$

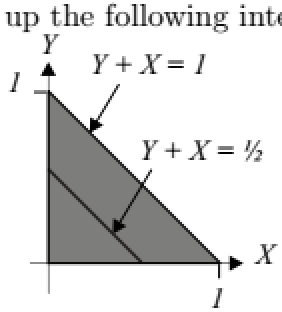
Therefore $c = 2$.

(b) To find the P$[X \leq Y]$ we look to integrate over the area indicated by the graph



$$\begin{aligned} \mathrm{P}[X \leq Y] &= \int_0^{1/2} \int_x^{1-x} dy\, dx \\ &= \int_0^{1/2} (2 - 4x)\, dx \\ &= 1/2. \end{aligned} \tag{2}$$

(c) The probability P$[X + Y \leq 1/2]$ can be seen in the figure. Here we can set up the following integrals



$$\begin{aligned} \mathrm{P}[X + Y \leq 1/2] &= \int_0^{1/2} \int_0^{1/2-x} 2\, dy\, dx \\ &= \int_0^{1/2} (1 - 2x)\, dx \\ &= 1/2 - 1/4 = 1/4. \end{aligned} \tag{3}$$

8.

Over the circle $X^2 + Y^2 \leq r^2$, random variables X and Y have the uniform PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/(\pi r^2) & x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the marginal PDF $f_X(x)$?

(b) What is the marginal PDF $f_Y(y)$?

Solution:

Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/(\pi r^2) & 0 \le x^2 + y^2 \le r^2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The marginal PDF of X is

$$f_X(x) = 2\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2}\,dy = \begin{cases} \frac{2\sqrt{r^2-x^2}}{\pi r^2} & -r \le x \le r, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) Similarly, for random variable Y,

$$f_Y(y) = 2\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{1}{\pi r^2}\,dx = \begin{cases} \frac{2\sqrt{r^2-y^2}}{\pi r^2} & -r \le y \le r, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

9.

Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = ce^{-(2x^2-4xy+4y^2)}.$$

(a) What are $\mathrm{E}[X]$ and $\mathrm{E}[Y]$?

(b) Find the correlation coefficient $\rho_{X,Y}$.

(c) What are $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$?

(d) What is the constant c?

(e) Are X and Y independent?

Solution:

For the joint PDF

$$f_{X,Y}(x,y) = ce^{-(2x^2-4xy+4y^2)}, \tag{1}$$

we proceed as in Problem 5.9.1 to find values for σ_Y, σ_X, $\mathrm{E}[X]$, $\mathrm{E}[Y]$ and ρ.

(a) First, we try to solve the following equations

$$\left(\frac{x - \mathrm{E}[X]}{\sigma_X}\right)^2 = 4(1-\rho^2)x^2, \tag{2}$$

$$\left(\frac{y - \mathrm{E}[Y]}{\sigma_Y}\right)^2 = 8(1-\rho^2)y^2, \tag{3}$$

$$\frac{2\rho}{\sigma_X\sigma_Y} = 8(1-\rho^2). \tag{4}$$

The first two equations yield $\mathrm{E}[X] = \mathrm{E}[Y] = 0$.

(b) To find the correlation coefficient ρ, we observe that

$$\sigma_X = 1/\sqrt{4(1-\rho^2)}, \qquad \sigma_Y = 1/\sqrt{8(1-\rho^2)}. \tag{5}$$

Using σ_X and σ_Y in the third equation yields $\rho = 1/\sqrt{2}$.

(c) Since $\rho = 1/\sqrt{2}$, now we can solve for σ_X and σ_Y.

$$\sigma_X = 1/\sqrt{2}, \qquad \sigma_Y = 1/2. \tag{6}$$

(d) From here we can solve for c.

$$c = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} = \frac{2}{\pi}. \tag{7}$$

(e) X and Y are dependent because $\rho \neq 0$.

10.

Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What are $\mathrm{E}[X]$ and $\mathrm{Var}[X]$?

(b) What are $\mathrm{E}[Y]$ and $\mathrm{Var}[Y]$?

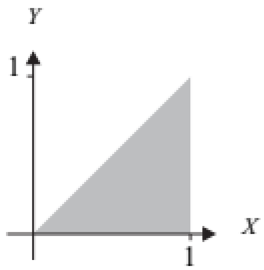
(c) What is $\mathrm{Cov}[X,Y]$?

(d) What is $\mathrm{E}[X+Y]$?

(e) What is $\mathrm{Var}[X+Y]$?

Solution:

Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Before finding moments, it is helpful to first find the marginal PDFs. For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_0^x 2\, dy = 2x. \tag{2}$$

Note that $f_X(x) = 0$ for $x < 0$ or $x > 1$. For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \int_y^1 2\, dx = 2(1-y). \tag{3}$$

Also, for $y < 0$ or $y > 1$, $f_Y(y) = 0$. Complete expressions for the marginal PDFs are

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} f_Y(y) \qquad = \begin{cases} 2(1-y) & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(a) The first two moments of X are

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^1 2x^2\, dx = 2/3, \tag{5}$$

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \int_0^1 2x^3\, dx = 1/2. \tag{6}$$

The variance of X is $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 1/2 - 4/9 = 1/18$.

(b) The expected value and second moment of Y are

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \int_0^1 2y(1-y)\,dy = y^2 - \left.\frac{2y^3}{3}\right|_0^1 = \frac{1}{3}, \tag{7}$$

$$\mathrm{E}\left[Y^2\right] = \int_{-\infty}^{\infty} y^2 f_Y(y)\ dy = \int_0^1 2y^2(1-y)\,dy = \left.\frac{2y^3}{3} - \frac{y^4}{2}\right|_0^1 = \frac{1}{6}. \tag{8}$$

The variance of Y is $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = 1/6 - 1/9 = 1/18$.

(c) Before finding the covariance, we find the correlation

$$\mathrm{E}[XY] = \int_0^1 \int_0^x 2xy\,dy\,dx = \int_0^1 x^3\,dx = 1/4 \tag{9}$$

The covariance is

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 1/36. \tag{10}$$

(d) $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] = 2/3 + 1/3 = 1$

(e) By Theorem 5.12,

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y] = 1/6. \tag{11}$$

1.

Let K be the number of heads in

$n = 100$ flips of a coin. Devise significance tests for the hypothesis H that the coin is fair such that

(a) The significance level $\alpha = 0.05$ and the rejection set R has the form $\{|K - \mathrm{E}[K]| > c\}$.

(b) The significance level $\alpha = 0.01$ and the rejection set R has the form $\{K > c'\}$.

Solution:

(a) We wish to develop a hypothesis test of the form

$$\mathrm{P}\left[\left|K - \mathrm{E}\left[K\right]\right| > c\right] = 0.05. \tag{1}$$

to determine if the coin we've been flipping is indeed a fair one. We would like to find the value of c, which will determine the upper and lower limits on how many heads we can get away from the expected number out of 100 flips and still accept our hypothesis. Under our fair coin hypothesis, the expected number of heads, and the standard deviation of the process are

$$\mathrm{E}\left[K\right] = 50, \qquad \sigma_K = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5. \tag{2}$$

Now in order to find c we make use of the central limit theorem and divide the above inequality through by σ_K to arrive at

$$\mathrm{P}\left[\frac{\left|K - \mathrm{E}\left[K\right]\right|}{\sigma_K} > \frac{c}{\sigma_K}\right] = 0.05. \tag{3}$$

Taking the complement, we get

$$\mathrm{P}\left[-\frac{c}{\sigma_K} \leq \frac{K - \mathrm{E}\left[K\right]}{\sigma_K} \leq \frac{c}{\sigma_K}\right] = 0.95. \tag{4}$$

Using the Central Limit Theorem we can write

$$\Phi\left(\frac{c}{\sigma_K}\right) - \Phi\left(\frac{-c}{\sigma_K}\right) = 2\Phi\left(\frac{c}{\sigma_K}\right) - 1 = 0.95. \tag{5}$$

This implies $\Phi(c/\sigma_K) = 0.975$ or $c/5 = 1.96$. That is, $c = 9.8$ flips. So we see that if we observe more then $50 + 10 = 60$ or less then $50 - 10 = 40$ heads, then with significance level $\alpha \approx 0.05$ we should reject the hypothesis that the coin is fair.

(b) Now we wish to develop a test of the form

$$\mathrm{P}[K > c] = 0.01. \tag{6}$$

Thus we need to find the value of c that makes the above probability true. This value will tell us that if we observe more than c heads, then with significance level $\alpha = 0.01$, we should reject the hypothesis that the coin is fair. To find this value of c we look to evaluate the CDF

$$F_K(k) = \sum_{i=0}^{k} \binom{100}{i} (1/2)^{100}. \tag{7}$$

Computation reveals that $c \approx 62$ flips. So if we observe 62 or greater heads, then with a significance level of 0.01 we should reject the fair coin hypothesis. Another way to obtain this result is to use a Central Limit Theorem approximation. First, we express our rejection region in terms of a zero mean, unit variance random variable.

$$\begin{aligned} \mathrm{P}[K > c] &= 1 - \mathrm{P}[K \leq c] \\ &= 1 - \mathrm{P}\left[\frac{K - \mathrm{E}[K]}{\sigma_K} \leq \frac{c - \mathrm{E}[K]}{\sigma_K}\right] = 0.01. \end{aligned} \tag{8}$$

Since $\mathrm{E}[K] = 50$ and $\sigma_K = 5$, the CLT approximation is

$$\mathrm{P}[K > c] \approx 1 - \Phi\left(\frac{c - 50}{5}\right) = 0.01. \tag{9}$$

From Table 4.2, we have $(c - 50)/5 = 2.35$ or $c = 61.75$. Once again, we see that we reject the hypothesis if we observe 62 or more heads.

2.

Random variables X and Y have

joint PMF given by the following table:

$P_{X,Y}(x,y)$	$y=-1$	$y=0$	$y=1$
$x=-1$	3/16	1/16	0
$x=0$	1/6	1/6	1/6
$x=1$	0	1/8	1/8

We estimate Y by $\hat{Y}_L(X) = aX + b$.

(a) Find a and b to minimize the mean square estimation error.

(b) What is the minimum mean square error e_L^*?

Hint: Use the following result:

Random variables X and Y have expected values μ_X and μ_Y, standard deviations σ_X and σ_Y, and correlation coefficient $\rho_{X,Y}$, The optimum linear mean square error (LMSE) estimator of X given Y is

$$\hat{X}_L(Y) = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y) + \mu_X.$$

This linear estimator has the following properties:

(a) The minimum mean square estimation error for a linear estimate is

$$e_L^* = \mathrm{E}\left[(X - \hat{X}_L(Y))^2\right] = \sigma_X^2(1 - \rho_{X,Y}^2).$$

(b) The estimation error $X - \hat{X}_L(Y)$ is uncorrelated with Y.

necessary moments of X and Y.

$$\begin{aligned}
\mathrm{E}[X] &= -1(1/4) + 0(1/2) + 1(1/4) = 0, && (1)\\
\mathrm{E}\left[X^2\right] &= (-1)^2(1/4) + 0^2(1/2) + 1^2(1/4) = 1/2, && (2)\\
\mathrm{E}[Y] &= -1(17/48) + 0(17/48) + 1(14/48) = -1/16, && (3)\\
\mathrm{E}\left[Y^2\right] &= (-1)^2(17/48) + 0^2(17/48) + 1^2(14/48) = 31/48, && (4)\\
\mathrm{E}[XY] &= 3/16 - 0 - 0 + 1/8 = 5/16. && (5)
\end{aligned}$$

The variances and covariance are

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 1/2, && (6)\\
\mathrm{Var}[Y] &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 493/768, && (7)\\
\mathrm{Cov}[X,Y] &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 5/16, && (8)\\
\rho_{X,Y} &= \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{5\sqrt{6}}{\sqrt{493}}. && (9)
\end{aligned}$$

By reversing the labels of X and Y in Theorem 12.3, we find that the optimal linear estimator of Y given X is

$$\hat{Y}_L(X) = \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}(X - \mathrm{E}[X]) + \mathrm{E}[Y] = \frac{5}{8}X - \frac{1}{16}. \qquad (10)$$

The mean square estimation error is

$$e_L^* = \mathrm{Var}[Y](1 - \rho_{X,Y}^2) = 343/768. \qquad (11)$$

3.

The random variables X and Y have the joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} 2(y+x) & 0 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is $\hat{X}_L(Y)$, the linear minimum mean square error estimate of X given Y?

Solution:

To solve this problem, we use Theorem 12.3. The only difficulty is in computing $\mathrm{E}[X]$, $\mathrm{E}[Y]$, $\mathrm{Var}[X]$, $\mathrm{Var}[Y]$, and $\rho_{X,Y}$. First we calculate the marginal PDFs

$$f_X(x) = \int_x^1 2(y+x)\,dy = y^2 + 2xy\Big|_{y=x}^{y=1} = 1 + 2x - 3x^2, \tag{1}$$
$$f_Y(y) = \int_0^y 2(y+x)\,dx = 2xy + x^2\Big|_{x=0}^{x=y} = 3y^2. \tag{2}$$

The first and second moments of X are

$$\mathrm{E}[X] = \int_0^1 (x + 2x^2 - 3x^3)\,dx = x^2/2 + 2x^3/3 - 3x^4/4\Big|_0^1 = 5/12, \tag{3}$$
$$\mathrm{E}\left[X^2\right] = \int_0^1 (x^2 + 2x^3 - 3x^4)\,dx = x^3/3 + x^4/2 - 3x^5/5\Big|_0^1 = 7/30. \tag{4}$$

The first and second moments of Y are

$$\mathrm{E}[Y] = \int_0^1 3y^3\,dy = 3/4, \qquad \mathrm{E}\left[Y^2\right] = \int_0^1 3y^4\,dy = 3/5. \tag{5}$$

Thus, X and Y each have variance

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{129}{2160}, \tag{6}$$
$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = \frac{3}{80}. \tag{7}$$

To calculate the correlation coefficient, we first must calculate the correlation

$$\begin{aligned}\mathrm{E}[XY] &= \int_0^1 \int_0^y 2xy(x+y)\,dx\,dy \\ &= \int_0^1 \left[2x^3y/3 + x^2y^2\right]\Big|_{x=0}^{x=y} dy = \int_0^1 \frac{5y^4}{3}\,dy = 1/3.\end{aligned} \tag{8}$$

Hence, the correlation coefficient is

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} \frac{\mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{5}{\sqrt{129}}. \tag{9}$$

Finally, we use Theorem 12.3 to combine these quantities in the optimal linear estimator.

$$\begin{aligned}\hat{X}_L(Y) &= \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y - \mathrm{E}[Y]) + \mathrm{E}[X] \\ &= \frac{5}{\sqrt{129}}\frac{\sqrt{129}}{9}\left(Y - \frac{3}{4}\right) + \frac{5}{12} = \frac{5}{9}Y.\end{aligned} \tag{10}$$

1.

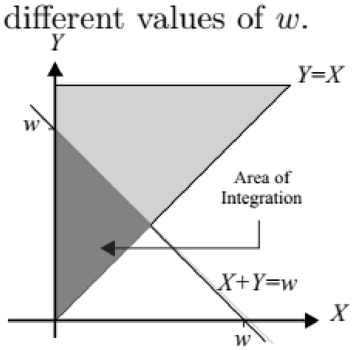
Find the PDF of $W = X + Y$ when X and Y have the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The joint PDF of X and Y is

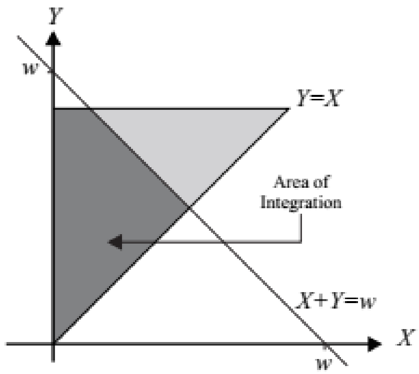
$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We wish to find the PDF of W where $W = X + Y$. First we find the CDF of W, $F_W(w)$, but we must realize that the CDF will require different integrations for different values of w.



For values of $0 \le w \le 1$ we look to integrate the shaded area in the figure to the right.

$$F_W(w) = \int_0^{\frac{w}{2}} \int_x^{w-x} 2\,dy\,dx = \frac{w^2}{2}. \tag{2}$$



For values of w in the region $1 \le w \le 2$ we look to integrate over the shaded region in the graph to the right. From the graph we see that we can integrate with respect to x first, ranging y from 0 to $w/2$, thereby covering the lower right triangle of the shaded region and leaving the upper trapezoid, which is accounted for in the second term of the following expression:

$$F_W(w) = \int_0^{\frac{w}{2}} \int_0^{y} 2\,dx\,dy + \int_{\frac{w}{2}}^{1} \int_0^{w-y} 2\,dx\,dy$$
$$= 2w - 1 - \frac{w^2}{2}. \quad (3)$$

Putting all the parts together gives the CDF

$$F_W(w) = \begin{cases} 0 & w < 0, \\ \frac{w^2}{2} & 0 \le w \le 1, \\ 2w - 1 - \frac{w^2}{2} & 1 \le w \le 2, \\ 1 & w > 2, \end{cases} \quad (4)$$

and (by taking the derivative) the PDF

$$f_W(w) = \begin{cases} w & 0 \le w \le 1, \\ 2 - w & 1 \le w \le 2, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

2.

Let X be a Gaussian $(0, \sigma)$ random variable. Use the moment generating function to show that

$$\mathrm{E}[X] = 0, \qquad \mathrm{E}[X^2] = \sigma^2,$$
$$\mathrm{E}[X^3] = 0, \qquad \mathrm{E}[X^4] = 3\sigma^4.$$

Let Y be a Gaussian (μ, σ) random variable. Use the moments of X to show that

$$\begin{aligned} \mathrm{E}\left[Y^2\right] &= \sigma^2 + \mu^2, \\ \mathrm{E}\left[Y^3\right] &= 3\mu\sigma^2 + \mu^3, \\ \mathrm{E}\left[Y^4\right] &= 3\sigma^4 + 6\mu\sigma^2 + \mu^4. \end{aligned}$$

Using the moment generating function of X, $\phi_X(s) = e^{\sigma^2 s^2/2}$. We can find the nth moment of X, $\mathrm{E}[X^n]$ by taking the nth derivative of $\phi_X(s)$ and setting $s = 0$.

$$\mathrm{E}[X] = \sigma^2 s e^{\sigma^2 s^2/2}\Big|_{s=0} = 0, \tag{1}$$

$$\mathrm{E}\left[X^2\right] = \sigma^2 e^{\sigma^2 s^2/2} + \sigma^4 s^2 e^{\sigma^2 s^2/2}\Big|_{s=0} = \sigma^2. \tag{2}$$

Continuing in this manner we find that

$$\mathrm{E}\left[X^3\right] = \left(3\sigma^4 s + \sigma^6 s^3\right) e^{\sigma^2 s^2/2}\Big|_{s=0} = 0, \tag{3}$$

$$\mathrm{E}\left[X^4\right] = \left(3\sigma^4 + 6\sigma^6 s^2 + \sigma^8 s^4\right) e^{\sigma^2 s^2/2}\Big|_{s=0} = 3\sigma^4. \tag{4}$$

To calculate the moments of Y, we define $Y = X + \mu$ so that Y is Gaussian (μ, σ). In this case the second moment of Y is

$$\mathrm{E}\left[Y^2\right] = \mathrm{E}\left[(X+\mu)^2\right] = \mathrm{E}\left[X^2 + 2\mu X + \mu^2\right] = \sigma^2 + \mu^2. \tag{5}$$

Similarly, the third moment of Y is

$$\begin{aligned} \mathrm{E}\left[Y^3\right] &= \mathrm{E}\left[(X+\mu)^3\right] \\ &= \mathrm{E}\left[X^3 + 3\mu X^2 + 3\mu^2 X + \mu^3\right] = 3\mu\sigma^2 + \mu^3. \end{aligned} \tag{6}$$

Finally, the fourth moment of Y is

$$\begin{aligned} \mathrm{E}\left[Y^4\right] &= \mathrm{E}\left[(X+\mu)^4\right] \\ &= \mathrm{E}\left[X^4 + 4\mu X^3 + 6\mu^2 X^2 + 4\mu^3 X + \mu^4\right] \\ &= 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4. \end{aligned} \tag{7}$$

3.

X and Y are independent random variables with PDFs

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1, \\ 0 & \text{otherwise}. \end{cases}$$

Let $A = \{X > Y\}$.

(a) What are $\mathrm{E}[X]$ and $\mathrm{E}[Y]$?

(b) What are $\mathrm{E}[X|A]$ and $\mathrm{E}[Y|A]$?

X and Y are independent random variables with PDFs

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise}, \end{cases} \qquad f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

For the event $A = \{X > Y\}$, this problem asks us to calculate the conditional expectations $\mathrm{E}[X|A]$ and $\mathrm{E}[Y|A]$. We will do this using the conditional joint PDF $f_{X,Y|A}(x, y)$. Since X and Y are independent, it is tempting to argue that the event $X > Y$ does not alter the probability model for X and Y. Unfortunately, this is not the case. When we learn that $X > Y$, it increases the probability that X is large and Y is small. We will see this when we compare the conditional expectations $\mathrm{E}[X|A]$ and $\mathrm{E}[Y|A]$ to $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

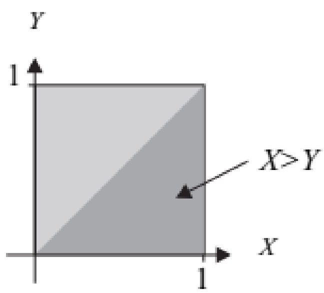
(a) We can calculate the unconditional expectations, $\mathrm{E}[X]$ and $\mathrm{E}[Y]$, using the marginal PDFs $f_X(x)$ and $f_Y(y)$.

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} f_X(x)\, dx = \int_0^1 2x^2\, dx = 2/3, \tag{2}$$

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} f_Y(y)\, dy = \int_0^1 3y^3\, dy = 3/4. \tag{3}$$

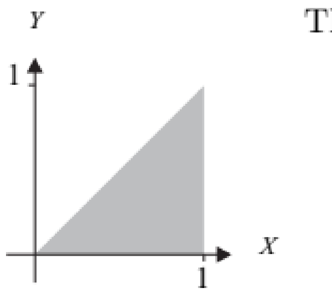
(b) First, we need to calculate the conditional joint PDF $ipdf X, Y|Ax, y$. The first step is to write down the joint PDF of X and Y:

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) = \begin{cases} 6xy^2 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$



The event A has probability

$$\begin{aligned} P[A] &= \iint_{x>y} f_{X,Y}(x,y)\, dy\, dx \\ &= \int_0^1 \int_0^x 6xy^2\, dy\, dx \\ &= \int_0^1 2x^4\, dx = 2/5. \end{aligned} \tag{5}$$



The conditional joint PDF of X and Y given A is

$$\begin{aligned} f_{X,Y|A}(x,y) &= \begin{cases} \frac{f_{X,Y}(x,y)}{P[A]} & (x,y) \in A, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 15xy^2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

The triangular region of nonzero probability is a signal that given A, X and Y are no longer independent. The conditional expected value of X given A is

$$\begin{aligned} E[X|A] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y|A}(x,y|a)\, x, y\, dy\, dx \\ &= 15 \int_0^1 x^2 \int_0^x y^2\, dy\, dx \\ &= 5 \int_0^1 x^5\, dx = 5/6. \end{aligned} \tag{7}$$

Instructor's note: The first equality in (7) is a few typos. It should be the double integral of x times

Double integral of f_X,Y|A(x,y) dy x dx

The conditional expected value of Y given A is

$$\begin{aligned} \mathrm{E}[Y|A] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y|A}(x,y)\, dy\, dx \\ &= 15\int_0^1 x \int_0^x y^3\, dy\, dx \\ &= \frac{15}{4}\int_0^1 x^5\, dx = 5/8. \end{aligned} \tag{8}$$

We see that $\mathrm{E}[X|A] > \mathrm{E}[X]$ while $\mathrm{E}[Y|A] < \mathrm{E}[Y]$. That is, learning $X > Y$ gives us a clue that X may be larger than usual while Y may be smaller than usual.

4.

This problem outlines the steps needed to show that the Gaussian PDF integrates to unity. For a Gaussian (μ, σ) random variable W, we will show that

$$I = \int_{-\infty}^{\infty} f_W(w)\, dw = 1.$$

(a) Use the substitution $x = (w - \mu)/\sigma$ to show that

$$I = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\, dx.$$

(b) Show that

$$I^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy.$$

(c) Change to polar coordinates to show that $I^2 = 1$.

First we note that since W has an $N[\mu, \sigma^2]$ distribution, the integral we wish to evaluate is

$$I = \int_{-\infty}^{\infty} f_W(w)\, dw = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(w-\mu)^2/2\sigma^2}\, dw. \tag{1}$$

(a) Using the substitution $x = (w - \mu)/\sigma$, we have $dx = dw/\sigma$ and

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx. \tag{2}$$

(b) When we write I^2 as the product of integrals, we use y to denote the other variable of integration so that

$$\begin{aligned} I^2 &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy. \end{aligned} \tag{3}$$

(c) By changing to polar coordinates, $x^2 + y^2 = r^2$ and $dx\, dy = r\, dr\, d\theta$ so that

$$\begin{aligned} I^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r\, dr\, d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} -e^{-r^2/2}\Big|_0^{\infty}\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1. \end{aligned} \tag{4}$$

5.

Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = ce^{-(x^2/8)-(y^2/18)}.$$

What is the constant c? Are X and Y independent?

X and Y have joint PDF

$$f_{X,Y}(x,y) = ce^{-(x^2/8)-(y^2/18)}. \tag{1}$$

The omission of any limits for the PDF indicates that it is defined over all x and y. We know that $f_{X,Y}(x,y)$ is in the form of the bivariate Gaussian distribution so we look to Definition 5.10 and attempt to find values for σ_Y, σ_X, $E[X]$, $E[Y]$ and ρ. First, we know that the constant is

$$c = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}. \tag{2}$$

Equating the exponent of (1) with the bivariate Gaussian, we get

rho = 0, sigma_X = 2, sigma_Y = 3, and c=1/(12 pi).

6.

Random variables X and Y have joint PDF

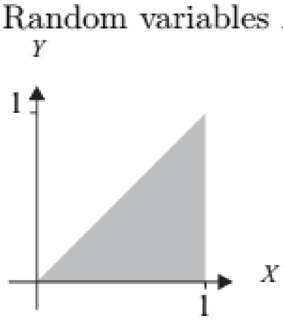
$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = Y/X$.

(a) What is S_W, the range of W?

(b) Find $F_W(w)$, $f_W(w)$, and $E[W]$.

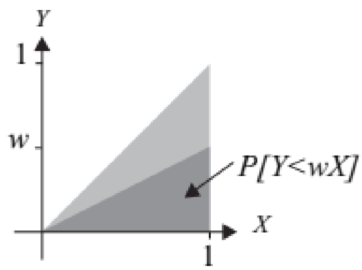
Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) Since X and Y are both nonnegative, $W = Y/X \geq 0$. Since $Y \leq X$, $W = Y/X \leq 1$. Note that $W = 0$ can occur if $Y = 0$. Thus the range of W is $S_W = \{w|0 \leq w \leq 1\}$.

(b) For $0 \leq w \leq 1$, the CDF of W is



$$F_W(w) = \mathrm{P}[Y/X \leq w] \\ = \mathrm{P}[Y \leq wX] = w. \tag{2}$$

The complete expression for the CDF is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ w & 0 \leq w < 1, \\ 1 & w \geq 1. \end{cases} \tag{3}$$

By taking the derivative of the CDF, we find that the PDF of W is

$$f_W(w) = \begin{cases} 1 & 0 \leq w < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We see that W has a uniform PDF over $[0,1]$. Thus $\mathrm{E}[W] = 1/2$.

7.

X and Y are independent identically distributed Gaussian $(0,1)$ random variables. Find the CDF of $W = X^2 + Y^2$.

Since X_1 and X_2 are iid Gaussian $(0,1)$, each has PDF

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{1}$$

For $w < 0$, $F_W(w) = 0$. For $w \geq 0$, we define the disc

$$\mathcal{R}(w) = \left\{(x_1, x_2) | x_1^2 + x_2^2 \leq w\right\}. \tag{2}$$

and we write

$$F_W(w) = \mathrm{P}\left[X_1^2 + X_2^2 \le w\right] = \iint_{\mathcal{R}(w)} f_{X_1}(x_1)\, f_{X_2}(x_2)\; dx_1\, dx_2$$
$$= \iint_{\mathcal{R}(w)} \frac{1}{2\pi} e^{-(x_1^2+x_2^2)/2}\, dx_1\, dx_2. \qquad (3)$$

Changing to polar coordinates with $r = \sqrt{x_1^2 + x_2^2}$ yields

$$F_W(w) = \int_0^{2\pi} \int_0^{\sqrt{w}} \frac{1}{2\pi} e^{-r^2/2} r\, dr\, d\theta$$
$$= \int_0^{\sqrt{w}} r e^{-r^2/2}\, dr = -e^{-r^2/2}\Big|_0^{\sqrt{w}} = 1 - e^{-w/2}. \qquad (4)$$

Taking the derivative of $F_W(w)$, the complete expression for the PDF of W is

$$f_W(w) = \begin{cases} 0 & w < 0, \\ \frac{1}{2} e^{-w/2} & w \ge 0. \end{cases} \qquad (5)$$

Thus W is an exponential $(\lambda = 1/2)$ random variable.

8.

For a constant $a > 0$, a Laplace random variable X has PDF

$$f_X(x) = \frac{a}{2} e^{-a|x|}, \quad -\infty < x < \infty.$$

Calculate the MGF $\phi_X(s)$.

For a constant $a > 0$, a zero mean Laplace random variable X has PDF

$$f_X(x) = \frac{a}{2} e^{-a|x|} \quad -\infty < x < \infty \qquad (1)$$

The moment generating function of X is

$$\begin{aligned}\phi_X(s) = \mathrm{E}\left[e^{sX}\right] &= \frac{a}{2}\int_{-\infty}^{0} e^{sx}e^{ax}\,dx + \frac{a}{2}\int_{0}^{\infty} e^{sx}e^{-ax}\,dx \\ &= \frac{a}{2}\frac{e^{(s+a)x}}{s+a}\bigg|_{-\infty}^{0} + \frac{a}{2}\frac{e^{(s-a)x}}{s-a}\bigg|_{0}^{\infty} \\ &= \frac{a}{2}\left(\frac{1}{s+a} - \frac{1}{s-a}\right) \\ &= \frac{a^2}{a^2 - s^2}. \qquad (2)\end{aligned}$$

9.

Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & x \geq 0, y \geq 0, x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is the variance of $W = X + Y$?

We can use the variance identity

$$\mathrm{Var}[W] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}\left[X,Y\right]. \qquad (1)$$

The first two moments of X are

$$\mathrm{E}\left[X\right] = \int_0^1 \int_0^{1-x} 2x\,dy\,dx = \int_0^1 2x(1-x)\,dx = 1/3, \qquad (2)$$

$$\mathrm{E}\left[X^2\right] = \int_0^1 \int_0^{1-x} 2x^2\,dy\,dx = \int_0^1 2x^2(1-x)\,dx = 1/6. \qquad (3)$$

Thus the variance of X is $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 1/18$. By symmetry, it should be apparent that $\mathrm{E}[Y] = \mathrm{E}[X] = 1/3$ and $\mathrm{Var}[Y] = \mathrm{Var}[X] = 1/18$. To find the covariance, we first find the correlation

$$\mathrm{E}\left[XY\right] = \int_0^1 \int_0^{1-x} 2xy\,dy\,dx = \int_0^1 x(1-x)^2\,dx = 1/12. \qquad (4)$$

The covariance is

$$\text{Cov}[X,Y] = \text{E}[XY] - \text{E}[X]\,\text{E}[Y] = 1/12 - (1/3)^2 = -1/36. \tag{5}$$

Finally, the variance of the sum $W = X + Y$ is

$$\begin{aligned}\text{Var}[W] &= \text{Var}[X] + \text{Var}[Y] - 2\,\text{Cov}[X,Y] \\ &= 2/18 - 2/36 = 1/18.\end{aligned} \tag{6}$$

For this specific problem, it's arguable whether it would easier to find $\text{Var}[W]$ by first deriving the CDF and PDF of W. In particular, for $0 \le w \le 1$,

$$\begin{aligned}F_W(w) &= \text{P}[X + Y \le w] \\ &= \int_0^w \int_0^{w-x} 2\,dy\,dx \\ &= \int_0^w 2(w-x)\,dx = w^2.\end{aligned} \tag{7}$$

Hence, by taking the derivative of the CDF, the PDF of W is

$$f_W(w) = \begin{cases} 2w & 0 \le w \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

From the PDF, the first and second moments of W are

$$\text{E}[W] = \int_0^1 2w^2\,dw = 2/3, \qquad \text{E}\left[W^2\right] = \int_0^1 2w^3\,dw = 1/2. \tag{9}$$

The variance of W is $\text{Var}[W] = \text{E}[W^2] - (\text{E}[W])^2 = 1/18$. Not surprisingly, we get the same answer both ways.

HW #4

1.

Suppose we flip a fair coin repeatedly. Let X_i equal 1 if flip i was heads (H) and 0 otherwise. Let N denote the number of flips needed until H has occurred 100 times. Is N independent of the random sequence $X_1, X_2, \ldots$? Define $Y = X_1 + \cdots + X_N$. Is Y an ordinary random sum of random variables? What is the PMF of Y?

Solution:

In this problem, $Y = X_1 + \cdots + X_N$ is not a straightforward random sum of random variables because N and the X_i's are dependent. In particular, given $N = n$, then we know that there were exactly 100 heads in N flips. Hence, given N, $X_1 + \cdots + X_N = 100$ *no matter what is the actual value of* N. Hence $Y = 100$ every time and the PMF of Y is

$$P_Y(y) = \begin{cases} 1 & y = 100, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

2.

Internet packets can be classified as video (V) or as generic data (D). Based on a lot of observations taken by the Internet service provider, we have the following probability model: $\mathrm{P}[V] = 3/4$, $\mathrm{P}[D] = 1/4$. Data packets and video packets occur independently of one another. The random variable K_n is the number of video packets in a collection of n packets.

(a) What is $\mathrm{E}[K_{100}]$, the expected number of video packets in a set of 100 packets?

(b) What is $\sigma_{K_{100}}$?

(c) Use the central limit theorem to estimate $\mathrm{P}[K_{100} \geq 18]$.

(d) Use the central limit theorem to estimate $\mathrm{P}[16 \leq K_{100} \leq 24]$.

Solution: There is an error in the solutions below since the author of the problem asks about video packets, but answers the question for data packets. The method is the same. I will ask the grader to give full marks for this question.

Knowing that the probability that voice call occurs is 0.8 and the probability that a data call occurs is 0.2 we can define the random variable D_i as the number of data calls in a single telephone call. It is obvious that for any i there are only two possible values for D_i, namely 0 and 1. Furthermore for all i the D_i's are independent and identically distributed withe the following PMF.

$$P_D(d) = \begin{cases} 0.8 & d = 0, \\ 0.2 & d = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

From the above we can determine that

$$\mathrm{E}[D] = 0.2, \qquad \mathrm{Var}[D] = 0.2 - 0.04 = 0.16. \tag{2}$$

With these facts, we can answer the questions posed by the problem.

(a) $\mathrm{E}[K_{100}] = 100\,\mathrm{E}[D] = 20$.

(b) $\mathrm{Var}[K_{100}] = \sqrt{100\,\mathrm{Var}[D]} = \sqrt{16} = 4$.

(c)

$$\begin{aligned} \mathrm{P}[K_{100} \geq 18] &= 1 - \Phi\left(\frac{18-20}{4}\right) \\ &= 1 - \Phi(-1/2) = \Phi(1/2) = 0.6915. \end{aligned} \tag{3}$$

(d)

$$\begin{aligned} \mathrm{P}[16 \leq K_{100} \leq 24] &= \Phi\left(\frac{24-20}{4}\right) - \Phi\left(\frac{16-20}{4}\right) \\ &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 = 0.6826. \end{aligned} \tag{4}$$

3.

Let $X_1, X_2, \ldots$ denote a sequence of independent samples of a random variable X with variance $\text{Var}[X]$. We define a new random sequence $Y_1, Y_2, \ldots$ as $Y_1 = X_1 - X_2$ and $Y_n = X_{2n-1} - X_{2n}$.

(a) Find $\text{E}[Y_n]$ and $\text{Var}[Y_n]$.

(b) Find the expected value and variance of $M_n(Y)$.

Solution:

(a) Since $Y_n = X_{2n-1} + (-X_{2n})$, Theorem 9.1 says that the expected value of the difference is

$$\text{E}[Y] = \text{E}[X_{2n-1}] + \text{E}[-X_{2n}] = \text{E}[X] - \text{E}[X] = 0. \tag{1}$$

By Theorem 9.2, the variance of the difference between X_{2n-1} and X_{2n} is

$$\text{Var}[Y_n] = \text{Var}[X_{2n-1}] + \text{Var}[-X_{2n}] = 2\,\text{Var}[X]. \tag{2}$$

(b) Each Y_n is the difference of two samples of X that are independent of the samples used by any other Y_m. Thus $Y_1, Y_2, \ldots$ is an iid random sequence. By Theorem 10.1, the mean and variance of $M_n(Y)$ are

$$\text{E}[M_n(Y)] = \text{E}[Y_n] = 0, \tag{3}$$

$$\text{Var}[M_n(Y)] = \frac{\text{Var}[Y_n]}{n} = \frac{2\,\text{Var}[X]}{n}. \tag{4}$$

4.

For an arbitrary random variable X, use the Chebyshev inequality to show that the probability that X is more than k standard deviations from its expected value $\text{E}[X]$ satisfies

$$\text{P}[|X - \text{E}[X]| \geq k\sigma] \leq \frac{1}{k^2}.$$

For a Gaussian random variable Y, use the $\Phi(\cdot)$ function to calculate the probability that Y is more than k standard deviations from its expected value $\text{E}[Y]$. Compare the result to the upper bound based on the Chebyshev inequality.

Solution:

We know from the Chebyshev inequality that

$$P[|X - E[X]| \geq c] \leq \frac{\sigma_X^2}{c^2}. \tag{1}$$

Choosing $c = k\sigma_X$, we obtain

$$P[|X - E[X]| \geq k\sigma] \leq \frac{1}{k^2}. \tag{2}$$

The actual probability the Gaussian random variable Y is more than k standard deviations from its expected value is

$$\begin{aligned} P[|Y - E[Y]| \geq k\sigma_Y] &= P[Y - E[Y] \leq -k\sigma_Y] + P[Y - E[Y] \geq k\sigma_Y] \\ &= 2\,P\left[\frac{Y - E[Y]}{\sigma_Y} \geq k\right] \\ &= 2Q(k). \end{aligned} \tag{3}$$

The following table compares the upper bound and the true probability:

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Chebyshev bound	1	0.250	0.111	0.0625	0.040
$2Q(k)$	0.317	0.046	0.0027	6.33×10^{-5}	5.73×10^{-7}

The Chebyshev bound gets increasingly weak as k goes up. As an example, for $k = 4$, the bound exceeds the true probability by a factor of 1,000 while for $k = 5$ the bound exceeds the actual probability by a factor of nearly 100,000.

5.

Let $X_1, \ldots, X_n$ be independent samples of a random variable X. Use the Chernoff bound to show that $M_n(X) = (X_1 + \cdots + X_n)/n$ satisfies

$$P[M_n(X) \geq c] \leq \left(\min_{s \geq 0} e^{-sc}\phi_X(s)\right)^n.$$

Solution:

Let $W_n = X_1 + \cdots + X_n$. Since $M_n(X) = W_n/n$, we can write

$$P\left[M_n(X) \geq c\right] = P\left[W_n \geq nc\right]. \tag{1}$$

Since $\phi_{W_n}(s) = (\phi_X(s))^n$, applying the Chernoff bound to W_n yields

$$P\left[W_n \geq nc\right] \leq \min_{s \geq 0} e^{-snc}\phi_{W_n}(s) = \min_{s \geq 0} \left(e^{-sc}\phi_X(s)\right)^n. \tag{2}$$

For $y \geq 0$, y^n is a nondecreasing function of y. This implies that the value of s that minimizes $e^{-sc}\phi_X(s)$ also minimizes $(e^{-sc}\phi_X(s))^n$. Hence

$$P\left[M_n(X) \geq c\right] = P\left[W_n \geq nc\right] \leq \left(\min_{s \geq 0} e^{-sc}\phi_X(s)\right)^n. \tag{3}$$

6.

Let X_A be the indicator random variable for event A with probability $P[A] = 0.8$. Let $\hat{P}_n(A)$ denote the relative frequency of event A in n independent trials.

(a) Find $E[X_A]$ and $\text{Var}[X_A]$.

(b) What is $\text{Var}[\hat{P}_n(A)]$?

(c) Use the Chebyshev inequality to find the confidence coefficient $1 - \alpha$ such that $\hat{P}_{100}(A)$ is within 0.1 of $P[A]$. In other words, find α such that

$$P\left[\left|\hat{P}_{100}(A) - P\left[A\right]\right| \leq 0.1\right] \geq 1 - \alpha.$$

(d) Use the Chebyshev inequality to find out how many samples n are necessary to have $\hat{P}_n(A)$ within 0.1 of $P[A]$ with confidence coefficient 0.95. In other words, find n such that

$$P\left[\left|\hat{P}_n(A) - P\left[A\right]\right| \leq 0.1\right] \geq 0.95.$$

Solution:

(a) Since X_A is a Bernoulli ($p = P[A]$) random variable,

$$E\left[X_A\right] = P\left[A\right] = 0.8, \qquad \text{Var}[X_A] = P\left[A\right]\left(1 - P\left[A\right]\right) = 0.16. \tag{1}$$

(b) Let $X_{A,i}$ to denote X_A on the ith trial. Since

$$\hat{P}_n(A) = M_n(X_A) = \frac{1}{n} \sum_{i=1}^{n} X_{A,i}, \tag{2}$$

is a sum of n independent random variables,

$$\text{Var}[\hat{P}_n(A)] = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}[X_{A,i}] = \frac{\text{P}[A](1 - \text{P}[A])}{n}. \tag{3}$$

(c) Since $\hat{P}_{100}(A) = M_{100}(X_A)$, we can use Theorem 10.5(b) to write

$$\begin{aligned} \text{P}\left[\left|\hat{P}_{100}(A) - \text{P}[A]\right| < c\right] &\geq 1 - \frac{\text{Var}[X_A]}{100c^2} \\ &= 1 - \frac{0.16}{100c^2} = 1 - \alpha. \end{aligned} \tag{4}$$

For $c = 0.1$, $\alpha = 0.16/[100(0.1)^2] = 0.16$. Thus, with 100 samples, our confidence coefficient is $1 - \alpha = 0.84$.

(d) In this case, the number of samples n is unknown. Once again, we use Theorem 10.5(b) to write

$$\begin{aligned} \text{P}\left[\left|\hat{P}_{n}(A) - \text{P}[A]\right| < c\right] &\geq 1 - \frac{\text{Var}[X_A]}{nc^2} \\ &= 1 - \frac{0.16}{nc^2} = 1 - \alpha. \end{aligned} \tag{5}$$

For $c = 0.1$, we have confidence coefficient $1-\alpha = 0.95$ if $\alpha = 0.16/[n(0.1)^2] = 0.05$, or $n = 320$.

1. Consider a Poisson point process with rate lambda=1 on the real line. Let N(t) be the random variable that the number of points over [0,t). Generate such a point process in Matlab and plot two realizations of N(t) versus t. Do the same for lambda =2.

Solution: Generate exponentially distributed random variables, and use those as interrarrival times to generate the Poisson points and N(t).

2.

Given a Poisson process $N(t)$, identify which of the following are Poisson processes..

(a) $N(2t)$, (b) $N(t/2)$,
(c) $2N(t)$, (d) $N(t)/2$,
(e) $N(t+2)$, (f) $N(t) - N(t-1)$.

Solution:

(a) and (b) are Poisson, the others are not. For (a) and (b), $N'(t) = N(\alpha t)$ is Poisson because for any interval (t_1, t_2),

$$N'(t_2) - N'(t_1) = N(\alpha t_2) - N(\alpha t_1), \tag{1}$$

which is Poisson with expected value $\alpha\lambda(t_2 - t_1)$. Also, non-overlapping intervals with respect to $N'(t)$ correspond to non-overlapping intervals for $N(t)$ and thus the number of arrivals in non-overlapping intervals are independent. For case (c), $2N(t)$ is not Poisson because $2N(t)$ is always even. For case (d), $N(t)/2$ takes on fractional sample values, which a Poisson process cannot. For case (e), $N(t+2)$ may be nonzero at time $t = 0$ but a Poisson process is always zero at $t = 0$. Finally, for (f), $N(t) - N(t-1)$ is not necessarily increasing.

3.

Customers arrive at the Veryfast Bank as a Poisson process of rate λ customers per minute. Each arriving customer is immediately served by a teller. After being served, each customer immediately leaves the bank. The time a customer spends with a teller is called the service time. If the service time of a customer is exactly

two minutes, what is the PMF of the number of customers $N(t)$ in service at the bank at time t?

Solution:

Note that it matters whether $t \geq 2$ minutes. If $t \leq 2$, then any customers that have arrived must still be in service. Since a Poisson number of arrivals occur during $(0, t]$,

$$P_{N(t)}(n) = \begin{cases} (\lambda t)^n e^{-\lambda t}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \qquad (0 \leq t \leq 2.) \tag{1}$$

For $t \geq 2$, the customers in service are precisely those customers that arrived in the interval $(t-2, t]$. The number of such customers has a Poisson PMF with mean $\lambda[t - (t-2)] = 2\lambda$. The resulting PMF of $N(t)$ is

$$P_{N(t)}(n) = \begin{cases} (2\lambda)^n e^{-2\lambda}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \qquad (t \geq 2.) \tag{2}$$

4.

Customers arrive at a casino as a Poisson process of rate 100 customers per hour. Upon arriving, each customer must flip a coin, and only those customers who flip heads actually enter the casino. Let $N(t)$ denote the process of customers entering the casino. Find the PMF of N, the number of customers who arrive between 5 PM and 7 PM.

Solution:

Customers entering (or not entering) the casino is a Bernoulli decomposition of the Poisson process of arrivals at the casino doors. By Theorem 13.6, customers entering the casino are a Poisson process of rate $100/2 = 50$ customers/hour. Thus in the two hours from 5 to 7 PM, the number, N, of customers entering the casino is a Poisson random variable with expected value $\alpha = 2 \cdot 50 = 100$. The PMF of N is

$$P_N(n) = \begin{cases} 100^n e^{-100}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

5.

Let Y_k denote the number of failures between successes $k-1$ and k of a Bernoulli (p) random process. Also, let Y_1 denote the number of failures before the first success. What is the PMF $P_{Y_k}(y)$? Is Y_k an iid random sequence?

Solution:

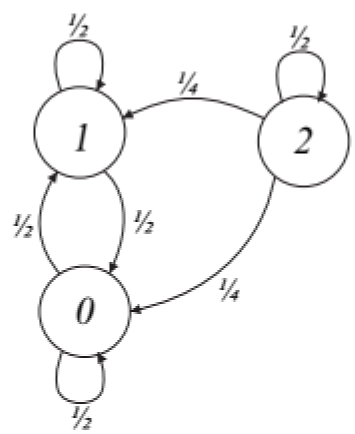
The number Y_k of failures between successes $k-1$ and k is exactly $y \geq 0$ iff after success $k-1$, there are y failures followed by a success. Since the Bernoulli trials are independent, the probability of this event is $(1-p)^y p$. The complete PMF of Y_k is

$$P_{Y_k}(y) = \begin{cases} (1-p)^y p & y = 0, 1, \dots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Since this argument is valid for all k including $k = 1$, we can conclude that $Y_1, Y_2, \dots$ are identically distributed. Moreover, since the trials are independent, the failures between successes $k-1$ and k and the number of failures between successes $k'-1$ and k' are independent. Hence, $Y_1, Y_2, \dots$ is an iid sequence.

6.

Find the state transition matrix $\mathbf{P}$ for the Markov chain:



From the given Markov chain, the state transition matrix is

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \tag{1}$$

7.

In a two-state discrete-time Markov chain, state changes can occur each second. Once the system is OFF, the system stays off for another second with probability 0.2. Once the system is ON, it stays on with probability 0.1. Sketch the Markov chain and find the state transition matrix $\mathbf{P}$.

Solution:

This problem is very straightforward if we keep in mind that P_{ij} is the probability that we transition from state i to state j. From Example 12.1, the state transition matrix is

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \tag{1}$$

8.

The state of a discrete-time Markov chain with transition matrix $\mathbf{P}$ can change once each second; X_n denotes the system state after n seconds. An observer examines the system state every m seconds, producing the observation sequence $\hat{X}_0, \hat{X}_1, \ldots$ where $\hat{X}_n = X_{mn}$. Is $\hat{X}_0, \hat{X}_1, \ldots$ a Markov chain? If so, find the state transition matrix $\hat{\mathbf{P}}$.

Solution:

To determine whether the sequence $\hat{X}_n$ forms a Markov chain, we write

$$\begin{aligned} P\left[\hat{X}_{n+1} = j | \hat{X}_n = i, \hat{X}_{n-1} = i_{n-1}, \ldots, \hat{X}_0 = i_0\right] & \\ = P\left[X_{m(n+1)} = j | X_{mn} = i, X_{m(n-1)} = i_{n-1}, \ldots, X_0 = i_0\right] & \qquad (1) \\ = P\left[X_{m(n+1)} = j | X_{mn} = i\right] & \qquad (2) \\ = P_{ij}(m) & \qquad (3) \end{aligned}$$

The key step is in observing that the Markov property of X_n implies that X_{mn} summarizes the past history of the X_n process. That is, given X_{mn}, $X_{m(n+1)}$ is independent of X_{mk} for all $k < n$.

Finally, this implies that the state $\hat{X}_n$ has one-step state transition probabilities equal to the m-step transition probabilities for the Markov chain X_n. That is, $\hat{\mathbf{P}} = \mathbf{P}^m$.

9.

A circular game board has K spaces numbered $0, 1, \ldots, K-1$. Starting at space 0 at time $n = 0$, a player rolls a fair six-sided die to move a token. Given the current token position X_n, the next token position is $X_{n+1} = (X_n + R_n) \mod K$ where R_n is the result of the player's nth roll. Find the stationary probability vector $\boldsymbol{\pi} = \begin{bmatrix}\pi_0 & \cdots & \pi_{K-1}\end{bmatrix}'$.

For this system, it's hard to draw the entire Markov chain since from each state n there are six branches, each with probability 1/6 to states $n+1, n+2, \ldots, n+6$. (Of course, if $n+k > K-1$, then the transition is to state $n+k \mod K$.) Nevertheless, finding the stationary probabilities is not very hard. In particular, the nth equation of $\boldsymbol{\pi}' = \boldsymbol{\pi}'\mathbf{P}$ yields

$$\pi_n = \frac{1}{6}\left(\pi_{n-6} + \pi_{n-5} + \pi_{n-4} + \pi_{n-3} + \pi_{n-2} + \pi_{n-1}\right). \tag{1}$$

Rather than try to solve these equations algebraically, it's easier to guess that the solution is

$$\boldsymbol{\pi} = \begin{bmatrix}1/K & 1/K & \cdots & 1/K\end{bmatrix}'. \tag{2}$$

It's easy to check that $1/K = (1/6) \cdot 6 \cdot (1/K)$